UNIVERSIDAD NACIONAL DE SANTIAGO DEL ESTERO

FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍAS

LICENCIATURA EN SISTEMAS DE INFORMACIÓN

TRABAJO FINAL DE GRADUACIÓN

APLICACIÓN DE TEXT MINING PARA LA DETECCIÓN DE HABILIDADES DE E-TUTORES DE APRENDIZAJE COLABORATIVO SOPORTADO POR COMPUTADORA

Autor:

Pablo Fernando Santana Mansilla

Profesor Guía:

Dra. Ing. Rosanna Costaguta

Profesor Asesor:

Lic. Daniela Missio

Marzo de 2013

UNIVERSIDAD NACIONAL DE SANTIAGO DEL ESTERO
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍAS

TRABAJO FINAL DE GRADUACIÓN DE LA LICENCIATURA EN SISTEMAS DE INFORMACIÓN

"APLICACIÓN DE TEXT MINING PARA LA DETECCIÓN DE HABILIDADES DE E-TUTORES DE APRENDIZAJE COLABORATIVO SOPORTADO POR COMPUTADORA"

Autor:

.....
Pablo Fernando Santana Mansilla

Profesor Guía:

.....
Dra. Ing. Rosanna Costaguta

Asesor:

.....
Lic. Daniela Missio

* ——————————————— * ——————————————— *

Aprobado el día ..... del mes de ..... del año 20.....

por el Tribunal integrado por

....

....

Santiago del Estero – Argentina

DEDICATORIAS

A MIS PADRES SUSANA Y LORENZO.

A MI HERMANA YESISA.

A MI ABUELA ARMINDA.

Pablo Fernando Santana Mansilla

AGRADECIMIENTOS

A la Dra. Ing. Rosanna Costaguta porque sus conocimientos, su apoyo, su confianza en mi trabajo, y su capacidad de guiar mis ideas han sido un aporte invaluable, no solamente en el desarrollo del presente trabajo sino también en mi formación como investigador.

A la Lic. Daniela Missio, por su disponibilidad, su paciencia, y las sugerencias brindadas para desarrollar el trabajo final.

A mis padres, por heberme apoyado a lo largo de mi carrera universitaria, por sus valores, por sus consejos, y por motivarme constantemente a tratar de ser una mejor persona día a día.

A la Ing. Diana Palliotto, la Lic. Susana Herrera y los estudiantes de Metodología de la Investigación II por haber participado en las experiencias de ACSC desde las que se obtuvo parte de los datos usados en la parte empírica de esta investigación.

A Lic. Elvio Suarez y los estudiantes de Estadistica por haber participado en las experiencias de ACSC desde las que se obtuvo parte de los datos usados en la parte empírica de esta investigación.

Al Lic. Manuel Juárez Pacheco, investigador del Centro Nacional de Investigación y Desarrollo Tecnológico de Mexico, por su colaboración en el suministro de parte de los datos necesarios para la realización de la parte empírica de esta investigación.

Pablo Fernando Santana Mansilla
Santiago del Estero, Argentina
Marzo 2013

INDICE DE CONTENIDOS

RESUMEN xxiv
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA 1
I.1. INTRODUCCIÓN 1
I.2. PLANTEAMIENTO DEL PROBLEMA 3
I.3. JUTIFICACIÓN DE LA RELEVANCIA DEL PROBLEMA 5
I.4. ANTECEDENTES 6
I.5. HIPÓTESIS Y OBJETIVOS 9
I.4.1. Objetivos Generales 9
I.4.2. Objetivos Específicos 9
I.6. SOLUCIÓN PROPUESTA 10
I.7. CONTRIBUCIONES 13
I.8. ESTRUCTURA DEL TRABAJO 13
CAPÍTULO II: NOCIONES PRELIMINARES 15
II.1. E-LEARNING 15
II.2 E-TUTOR 17
II.3 CONSTRUCTIVISMO 18
II.4 CONSTRUCTIVISMO SOCIAL 19
II.5 GRUPO DE APRENDIZAJE 19
II.6 APRENDIZAJE COLABORATIVO 20
II.7 APRENDIZAJE COOPERATIVO 22
II.8 APRENDIZAJE COLABORATIVO SOPORTADO POR COMPUTADORA 24
II.9 HABILIDADES Y COMPETENCIAS 25
II.9.1 Habilidad 25
II.9.2 Competencia 27
II.10 DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS 28
II.11 MINERÍA DE DATOS 30
II.12 APRENDIZAJE DE MÁQUINA 31
II.12.1 Aprendizaje Supervisado y no Supervisado 32
II.12.2 Conjuntos de Entrenamiento y de Prueba 32
II.13 PROBLEMAS LINEALMENTE SEPARABLES 34
II.14 REDES NEURONALES 35
II.14.1 Neuronas Biológicas 35
II.14.2 Neuronas Artificiales 36
II.14.3 Aprendizaje en Redes Neuronales Artificiales 37
II.14.4 El Perceptrón 39
II.14.4.1 Aprendizaje en el perceptrón 41
CAPÍTULO III: MINERÍA DE TEXTOS 43
III.1 MINERÍA DE TEXTOS 43

III.2 INFORMACIÓN ESTRUCTURADA Y DESESTRUCTURADA 45
III.3 OPERACIONES DE PRE PROCESAMIENTO 46
III.3.1Estandarización de los Documentos 47
III.3.2 Identificación de Token o Tokenización 47
III.3.3 Normalización 48
III.3.4 Remoción de Stop Words 49
III.3.5 Determinación del Límite de las Sentencias 49
III.3.6 Etiquetado de Parte del Discurso 50
III.3.7 Discriminación del Sentido de la Palabra 50
III.3.8 Parsing 51
III.4 REPRESENTACIÓN DE LOS DOCUMENTOS 51
III.5 FORMALIZACIÓN DE LA TAREA DE CLASIFICACIÓN 53
III.6 APROXIMACIONES A LA TAREA DE CLASIFICACIÓN 54
III.7 CLASIFICACIÓN HARD VERSUS SOFT 55
III.8 TECNICAS DE CONSTRUCCIÓN DE CLASIFICADORES 56
III.8.1 Clasificadores Probabilísticos 56
III.8.2 Clasificadores de Árbol de Decisión 57
III.8.2.1 Algoritmo C4.5 57
III.8.3 Clasificadores de Regla de Decisión 58
III.8.4 Clasificadores Basados en Ejemplo 59
III.8.5 Métodos Lineales de Puntuación 60
III.8.5.1 Máquinas de vector de soporte 62
III.8.5.2 Máquinas de vector de soporte con márgenes desiguales 63
III.8.6 Redes neuronales 65
III.8.6.1 Algoritmo PAUM 66
III.9 EVALUACION DE LA EFECTIVIDAD DE LOS CLASIFICADORES 68

III.9.1 Métodos de Estimación de la Efectividad 68
III.9.2 Medidas de Efectividad 69
CAPÍTULO IV: LAS HABILIDADES DE LOS DOCENTES EN GRUPOS DE APRENDIZAJE COLABORATIVO ONLINE 73
IV.1 IMPORTANCIA DEL E-TUTOR EN ENTORNOS DE ACSC 73
IV.2 INVESTIGACIONES SOBRE LA TAREA DE LOS E-TUTORES DE ACSC 75
IV.3 DIFERENCIAS ENTRE E-LEARNING Y CLASES PRESENCIALES 77
IV.4 CLASIFICACIÓN DE HABILIDADES PROPUESTA 78
IV.4.1 Habilidades Administrativas 80
IV.4.1.1 Planificación 80
IV.4.1.2 Administración del tiempo 81
IV.4.1.3 Coordinación 81
IV.4.1.4 Establecer y cumplir con las directrices para el proceso de aprendizaje 82

IV.4.1.5 Administrar el contenido de las conferencias 83
IV.4.1.6 Administrar programas de aprendizaje 83
IV.4.1.7 Gestionar información 83
IV.4.1.8 Proveer dirección y soporte administrativo 84
IV.4.1.9 Monitorear Actividades 84
IV.4.2 Habilidades Sociales 86
IV.4.2.1 Estimular la interacción entre estudiantes 86
IV.4.2.2 Estimular la participación 87
IV.4.2.3 Desarrollar, soportar y mantener una comunidad de aprendizaje 88
IV.4.2.4 Crear un ambiente de aprendizaje agradable 90
IV.4.2.5 Manejar comportamientos de discusión inapropiados 92
IV.4.2.5.1 Mirones 92
IV.4.2.5.2 No participantes 92
IV.4.2.5.3 Bloqueadores 92
IV.4.2.5.4 Dominadores 93
IV.4.2.5.5 Buscadores de reconocimiento 93
IV.4.2.6 Manejar contribuciones inapropiadas 93
IV.4.2.7 Alabar y modelar el comportamiento de discusión bueno 94
IV.4.2.8 Manejar conflictos 95
IV.4.2.9 Trabajar con sentimientos y sensaciones 96
IV.4.2.10 Motivar 96
IV.4.3 Habilidades Pedagógicas 98
IV.4.3.1 Diseño instruccional 98
IV.4.3.2 Creación de actividades de aprendizaje 98
IV.4.3.3 Creación de recursos de aprendizaje 99
IV.4.3.4 Facilitar la comprensión del material de aprendizaje 100
IV.4.3.5 Flexibilidad 101
IV.4.3.6 Promover el debate/ la discusión 102
IV.4.3.7 Gestionar el debate 102
IV.4.3.8 Vincular 103
IV.4.3.9 Resumir 104
IV.4.3.10 Asistir a los estudiantes en la gestión de su propio aprendizaje 104
IV.4.3.11 Gestionar el acceso de los estudiantes a los recursos de aprendizaje 105
IV.4.3.12 Desarrollar las habilidades de aprendizaje de los estudiantes 105
IV.4.3.13 Monitorear el proceso de aprendizaje 106
IV.4.3.14 Resolver problemas de aprendizaje 107
IV.4.3.15 Familiarizar a los estudiantes con el ambiente online 107
IV.4.3.16 Marcar el ritmo del proceso de aprendizaje 107
IV.4.4 Habilidades Técnicas 111
IV.4.4.1 Seleccionar la tecnología 111

IV.4.4.2 Usar la tecnología 111
IV.4.4.3 Proveer soporte técnico 112
IV.4.4.4 Desarrollar habilidades técnicas de los estudiantes 112
IV.4.5 Habilidades de comunicación 113
IV.4.5.1 Preguntar 114
IV.4.5.1.1 Preguntas relevantes para el ámbito social 114
IV.4.5.1.2 Preguntas relevantes para el ámbito de evaluación 115
IV.4.5.1.3 Preguntas relevantes para el ámbito pedagógico 116
IV.4.5.2 Escuchar 116
IV.4.5.3 Responder 116
IV.4.5.4 Dialogar con los estudiantes 118
IV.4.5.5 Crear el ambiente de comunicación 119
IV.4.5.6 Gestionar problemas de comunicación 119
IV.4.6 Habilidades relativas al rol de evaluación 120
IV.4.6.1 Evaluar a los estudiantes 121
IV.4.6.2 Evaluar el proceso de enseñanza 121
IV.4.6.3 Reportar los resultados 122

CAPÍTULO V: ETAPAS DE UN ESTUDIO DE ANÁLISIS DE CONTENIDO 125

V.1 DEFINICIÓN DE ANÁLISIS DE CONTENIDO 125
V.2 MARCO DE TRABAJO PARA ANÁLISIS DE CONTENIDO 127
V.2.1 Cuerpo de Textos 127
V.2.2 Preguntas de Investigación 128
V.2.3 Contexto 129
V.2.4 Constructo Analítico 130
V.2.5 Inferencias 130
V.2.6 Evidencia de validación 132
V.3 ELEMENTOS DEL ANÁLISIS DE CONTENIDO 132
V.3.1 Separación en Unidades 134
V.3.1.1 Unidades de muestreo 134
V.3.1.2 Unidades de grabación/codificación 135
V.3.1.3 Unidades de contexto 135
V.3.2 Muestreo 135
V.3.3 Grabación 138
V.3.4 Reducción 139
V.3.5 Inferencia 140
V.3.6 Narración 140
V.4 FIABILIDAD 140
V.4.1 Tipos de Fiabilidad 141
V.4.2 Cálculo de la Fiabilidad 142

CAPÍTULO VI: ETAPAS DE UN PROYECTO DE MINERÍA DE TEXTOS.....146

VI.1 METODOLOGÍA CRISP-DM 146
VI.1.1 Fase1: Comprensión del Negocio 147
VI.1.1.1 Determinar los objetivos del negocio 147
VI.1.1.2 Evaluar la situación 148
VI.1.1.3 Determinar los objetivos de la minería de datos 149
VI.1.1.4 Producir el plan del proyecto 149
VI.1.2 Fase 2: Comprensión de los Datos 150
VI.1.2.1 Recolectar datos iniciales 150
VI.1.2.2 Describir los Datos 150
VI.1.2.3 Explorar los datos 150
VI.1.2.4 Verificar la calidad de los datos 151
VI.1.3 FASE 3: Preparación de los Datos 151
VI.1.3.1 Seleccionar los datos 151
VI.1.3.2 Limpiar los datos 152
VI.1.3.3 Construir los datos 152
VI.1.3.4 Integrar los datos 152
VI.1.3.5 Formatear los datos 153
VI.1.4 FASE 4: Modelado 153
VI.1.4.1 Seleccionar la técnica de modelado 153
VI.1.4.2 Generar el diseño de la prueba 154
VI.1.4.3 Construir el modelo 154
VI.1.4.4 Evaluar el modelo 155
VI.1.5 Fase 5 : Evaluación 155
VI.1.5.1 Evaluar resultados 155
VI.1.5.2 Revisar el proceso 156
VI.1.5.3 Determinar próximos pasos 156
VI.1.6 FASE 6: Despliegue 157
VI.1.6.1 Plan de despliegue 157
VI.1.6.2 Plan de monitoreo y mantenimiento 157
VI.1.6.3 Producir el reporte final 157
VI.1.6.4 Revisar el Proyecto 158
VI.2 APLICACIÓN DE CRISP-DM A PROBLEMAS DE MINERÍA DE TEXTOS 158
VI.2.1 Enfoques para Descubrimiento de Conocimiento en Textos 158
VI.2.1.1 Enfoque de Tan 159
VI.2.1.2 Enfoque de Palazzo 160
VI.2.1.3 Enfoque de Halliman 161
VI.2.2 Adaptación de CRISP-DM a Problemas de Minería de Textos 163

CAPÍTULO VII: DOCUMENTACIÓN DEL ESTUDIO DE ANÁLISIS DE CONTENIDO 165

VII.1 CUERPOS DE TEXTOS 165
VII.2 PREGUNTAS DE INVESTIGACIÓN 166
VII.3 CONTEXTO 166
VII.4 CONSTRUCTO ANALÍTICO 167
VII.4.1 Estimular la Interacción entre Estudiantes 168
VII.4.2 Estimular la Participación 168
VII.4.3 Desarrollar, Soportar y Mantener una Comunidad de Aprendizaje 169
VII.4.3.1 Desarrollar la cohesión grupal 169
VII.4.3.2 Promover la colaboración 169
VII.4.3.3 Promover el respeto 169
VII.4.4 Crear un Ambiente de Aprendizaje Agradable 170
VII.4.5 Manejar Comportamientos de Discusión Inapropiados 171
VII.4.5.1 Mirones 171
VII.4.5.2 No participantes 171
VII.4.5.3 Bloqueadores 171
VII.4.5.4 Dominadores 172
VII.4.5.5 Buscadores de reconocimiento 173
VII.4.6 Manejar Contribuciones Inapropiadas 173
VII.4.6.1 Actuar contra las críticas 173
VII.4.6.2 Actuar contra la intimidación 173
VII.4.6.3 Actuar contra la discriminación 174
VII.4.7 Alabar y Modelar el Comportamiento de Discusión Bueno 174
VII.4.8 Manejar Conflictos 175
VII.4.9 Trabajar con Sentimientos y Sensaciones 175
VII.4.10 Motivar 175
VII.5 ELEMENTOS DEL ANÁLISIS DE CONTENIDO 176
VII.5.1 Separación en Unidades 176
VII.5.1.1 Unidad de muestreo 176
VII.5.1.2 Unidad de grabación 176
VII.5.1.3 Unidad de contexto 176
VII.5.2 Muestreo 177
VII.5.2.1 Plan de muestreo 177
VII.5.2.2 Muestreo resultante 178
VII.5.3 Grabación/codificación 178
VII.5.3.1 Capacidades de los codificadores 178
VII.5.3.1.1 Habilidades cognitivas 178
VII.5.3.1.2 Formación/ antecedentes 178
VII.5.3.2 Entrenamiento de los codificadores 179
VII.5.3.3 Semántica del lenguaje de datos 179
VII.5.3.4 Sintaxis del lenguaje de datos 180

VII.5.3.4.1 Variables 180
VII.5.3.4.2 Valores 180
VII.5.3.4.3 Constantes 182
VII.5.3.4.4 Gramática 182
VII.5.3.4.5 Lógica 183
VII.5.3.5 Naturaleza y administración de registros 183
VII.5.3.5.1 Información administrativa 183
VII.5.3.5.2 Información de organización 185
VII.5.3.5.3 Información sobre el fenómeno 185
VII.5.4 Reducción del Texto a Representaciones Manejables 186
VII.5.4.1 Primer proceso de codificación 186
VII.5.4.2 Segundo proceso de codificación 189
VII.5.5 Inferencias y Narración 191
VII.6 FIABILIDAD 197

CAPÍTULO VIII: ADAPTACIONES REALIZADAS A LOS RECUROS DE GATE 199

VIII.1 NECESIDAD DE ADAPTAR GATE 199
VIII.2 EXPRESIONES REGULARES 199
VIII.3 ANOTACIONES 200
VIII.4 GRAMÁTICA JAPE 201
VIII.5 ADAPTACIONES EFECTUADAS A ANNIE SENTENCE SPLITTER 202
VIII.5.1 Fase Prepare 204
VIII.5.2 Fase preProcesarTokens 204
VIII.5.2.1 Regla ignorarSignosDentroPreg 205
VIII.5.2.2 Regla ignorarSignosDentroExcl 205
VIII.5.3 Fase modificarTokens 205
VIII.5.3.1 Regla modifTokensInterrogacion 205
VIII.5.3.2 Regla modifTokensExclamacion 205
VIII.5.4 Fase find1 205
VIII.5.4.1 Regla DotNetStop 206
VIII.5.4.2 Regla DotFileName 206
VIII.5.4.3 Regla pregExclamPtoSuspensivoNoCierraEnunciado 206
5.4.4 Regla pregExclamPtoSuspensivoCierraEnunciado 206
VIII.5.4.5 Regla ignorarPuntoSuspensivos 207
VIII.5.4.6 Regla Number 207
VIII.5.4.7 Regla knownAbbrev 207
VIII.5.4.8 Regla upperCaseAbbrev 207
VIII.5.4.9 Regla abreviaturaMinuscula 207
VIII.5.4.10 Regla abreviaturaMinMay 207
5.4.11 Regla abreviaturaPlural 208
VIII.5.4.12 Regla abreviaturaAcentuada 208

VIII.5.4.13 Regla abreviaturaCaracterEspec 208
VIII.5.4.14 Regla abreviaturaEtcetera 208
VIII.5.4.15 Regla direccionEmail 209
VIII.5.4.16 Regla direccionURI 209
VIII.5.4.17 Regla identificadorJava 209
VIII.5.4.18 Regla Ldots 209
VIII.5.4.19 Regla Lpunct 209
VIII.5.4.20 Regla Split1 209
VIII.5.4.21 Regla CR 210
VIII.5.4.22 Regla puntosEnEnumeracion 210
VIII.5.4.23 Regla preguntasexclamacionesConsec 210
VIII.5.4.24 Regla preguntaexclamacionConCont 210
VIII.5.4.25 Regla dosPuntosAntesListado 211
VIII.5.4.26 Regla parentesisAclaratorio 211
VIII.5.4.27 Regla corcheteAclaratorio 211
VIII.5.4.28 Regla signoLlaves 211
VIII.5.4.29 Regla signoComillas 211
VIII.5.5 Fase find2 212
VIII.5.5.1 Regla puntosSuspensivosSucedenaEtc 212
VIII.5.5.2 Regla reconocerDivisorconetc 213
VIII.5.5.3 Regla preguntaExclamacionSucedenaEtc 214
VIII.5.5.4 Regla etcSeguidodedosPuntos 215
VIII.5.5.5 Regla ignorarPtosSuspensivosInterrogacionExclamacion 215
VIII.5.5.6 Regla abreviaturaMinusculaMayusculaSeguidaPto 216
VIII.5.5.7 Regla ptoDeEnumeracionNoDivisor 217
VIII.5.5.8 Regla parentesisEnEnumeracion 218
VIII.5.5.9 Regla preguntaexclamacionNoSegundaParte 220
VIII.5.5.10 Regla eliminaranotacionExclIntePtosSuspensivos 220
VIII.5.6 Fase find3 221
VIII.5.7 Fase find4 223
VIII.5.7.1 Regla eliminarAnotSplitdePtoEnumeracion 223
VIII.5.7.2 Regla omitirAnotSplitdePtoEnumeracion 223
VIII.5.8 Fase Split 224
VIII.5.8.1 Regla internalSplits 225
VIII.5.8.2 Regla externalSplits 225
VIII.5.9 Fase reestablecerTokens 225
VIII.5.10 Fase postProcesarTokens 225
VIII.5.11 Fase cleanup 225
VIII.6 ELIMINACIÓN DE *STOP WORDS* 226
VIII.6.1 Definición de *stop words* con *ANNIE Gazetteer* 226

VIII.6.2 Eliminación de stop words con *JAPE Transducer* 226
VIII.6.2.1Regla stopWordsSinAcento 226
VIII.6.2.2Regla stopWordsAcentuadas 226
VIII.7 PRUEBAS APLICADAS A LAS ADAPTACIONES DE GATE 227

CAPÍTULO IX: DOCUMENTACIÓN DE LA APLICACIÓN DE CRISP-DM 228

IX.1 FASE 1: COMPRENSIÓN DEL NEGOCIO 228
IX.1.1 Determinar los Objetivos del Negocio 228
IX.1.1.1 Antecedentes 228
IX.1.1.2 Objetivos de negocio 228
IX.1.1.3 Criterio de éxito de negocio 229
IX.1.2 Evaluar la Situación 229
IX.1.2.1 Inventario de recursos 229
IX.1.2.1.1 Personal 229
IX.1.2.1.2 Datos 229
IX.1.2.1.3 Hardware 230
IX.1.2.1.4 Software 230
IX.1.2.2 Requerimientos, supuestos, y restricciones 231
IX.1.2.2.1 Requerimientos 231
IX.1.2.2.2 Supuestos 231
IX.1.2.2.3 Restricciones 231
IX.1.2.3 Riesgos y contingencias 232
IX.1.2.4 Terminología 233
IX.1.2.5 Costos y beneficios 236
IX.1.3 Determinar los Objetivos de la Minería de Datos 236
IX.1.3.1 Objetivos minería de datos 236
IX.1.3.2 Criterios de éxito de la minería 237
IX.1.4 Producir el Plan del Proyecto 237
IX.1.4.1 Plan del proyecto 237
IX.1.4.2 Evaluación inicial de técnicas y herramientas 238
IX.1.4.2.1 Evaluación de técnicas 238
IX.1.4.2.2 Evaluación de herramientas de software disponibles 241
IX.2 FASE 2: COMPRENSIÓN DE LOS DATOS 244
IX.2.1 Recolectar Datos Iniciales 244
IX.2.2 Describir los Datos 246
IX.2.3 Explorar los Datos 248
IX.2.4 Verificar la Calidad de los Datos 252
IX.3 FASE 3:PREPARACIÓN DE LOS DATOS 254
IX.3.1 Seleccionar los Datos 254
IX.3.2 Limpiar los Datos 255
IX.3.2.1 Modificación de archivos 255

IX.3.2.2 Exclusión de archivos 255
IX.3.3 Construir los Datos 256
IX.3.3.1 Identificación de *token* 256
IX.3.3.2 Identificación de sentencias 257
IX.3.3.3 Normalización 257
IX.3.3.4 Etiquetado de parte del discurso 258
IX.3.3.5 Supresión de *stop words* 258
IX.3.3.6 Combinación de operaciones de construcción de datos 259
IX.3.4 Integrar los Datos 261
IX.3.5 Formatear los Datos 262
IX.4 FASE 4: MODELADO 264
IX.4.1 Seleccionar la Técnica de Modelado 264
IX.4.1.1 K-esimo vecino más cercano (KNN) 264
IX.4.1.2 Máquina de vector de soporte (SVM) 265
IX.4.1.3 Perceptrón con márgenes desiguales (PAUM) 265
IX.4.1.4 Naïve bayes 265
IX.4.1.5 Algoritmo C4.5 266
IX.4.2 Generar el Diseño de la Prueba 266
IX.4.2.1 Método hold-out 266
IX.4.2.2 Método de validación cruzada n-fold 267
IX.4.3 Construir el Modelo 268
IX.4.3.1 Construcción del modelo con KNN 268
IX.4.3.2 Construcción del modelo con SVM 275
IX.4.3.3 Construcción del modelo con PAUM 279
IX.4.3.4 Construcción del Modelo con Naïve Bayes 283
IX.4.3.5 Construcción del modelo con C4.5 284
IX.4.4 Evaluar el Modelo 284
IX.4.4.1 Resultados de la evaluación con hold-out 287
IX.4.4.2 Resultados de la evaluación con validación cruzada 295
IX.5 FASE 5: EVALUACIÓN 300
IX.5.1 Evaluar Resultados 300
IX.5.2 Revisar el Proceso 302
IX.5.3 Determinar Próximos Pasos 305
IX.6 FASE 6: DESPLIEGUE 306
IX.6.1 Revisar el Proyecto 306
CAPÍTULO X: CONCLUSIONES 309
REFERENCIAS BIBLIOGRÁFICAS 316
ANEXO A: ANÁLISIS DE CONTENIDO 324
ANEXO B: MINERÍA DE TEXTOS 389

INDICE DE TABLAS

Tabla 2.1. Comparación entre aprendizaje cooperativo y aprendizaje colaborativo……....23
Tabla 2.2. Tabla de verdad paras las operaciones OR y OR exclusivo……34
Tabla 2.3. Analogía entre red neuronal biológica y red neuronal artificial……..37
Tabla 3.1. Tabla de Contingencia para la categoría c_i……..……..……..70
Tabla 3.2. Tabla de contingencia global……..……..……..71
Tabla 4.1. Diferencias entre clases presenciales y clases *online*……..……77
Tabla 4.2. Sub habilidades y atributos de habilidad administrativos de e-tutores de ACSC ……..……..……..……..……..……..……..84
Tabla 4.3. Sub habilidades y atributos de habilidad sociales de e-tutores de ACSC……..97
Tabla 4.4. Habilidades y sub habilidades pedagógicas de e-tutores de ACSC……..108
Tabla 4.5. Habilidades y sub habilidades técnicas de e-tutores de ACSC……113
Tabla 4.6. Habilidades y sub habilidades de comunicación de e-tutores de ACSC……...120
Tabla 4.7. Habilidades y sub habilidades de evaluación de e-tutores de ACSC……122
Tabla 5.1.Tipos de fiabilidad…….141
Tabla 5.2. Ejemplo de matriz de datos de fiabilidad…….143
Tabla 5.3. Estructura de la matriz de coincidencias observadas (Krippendorff, 2004)…144
Tabla 5.4. Ejemplo de matrices de coincidencias observadas…….145
Tabla 7.1. Atributos de habilidad social y sus abreviaturas…….180
Tabla 7.2. Ejemplos de expresiones bien y mal formadas……..182
Tabla 7.3. Cantidad de veces que los codificadores identificaron el mismo atributo de habilidad social ……..187
Tabla 7.4. Pares de atributos de habilidad social en conflicto y cantidad de veces que se repite la situación……..188
Tabla 7.5. Cantidad de veces que se manifiestan los atributos de habilidad social……...190
Tabla 9.1. Planificación del proyecto de minería de textos…….238
Tabla 9.2. Comparación entre herramientas de minería de textos *open sourc*……..243
Tabla 9.3. Archivos en su formato original junto con los archivos de texto plano derivados…….244
Tabla 9.4. Ejemplo de múltiples habilidades identificadas en un mensaje……..247
Tabla 9.5. Resumen de archivos de texto plano……..248
Tabla 9.6. Cantidad total y cantidad media de palabras por atributo de habilidad……249

Tabla 9.7. Palabras frecuentes y atributos de habilidad asociados…….250
Tabla 9.8. Resumen de la supresión de archivos en cada experiencia de ACSC……256
Tabla 9.9. Comparación de resultados de lematización y *stemming to a root...* ……257
Tabla 9.10. Parámetros de tiempo de ejecución de *Batch Learning PR*...…….270
Tabla 9.11. Tiempos de construcción y de clasificación para el algoritmo KNN……275
Tabla 9.12: Tiempos de construcción y de clasificación para el algoritmo SVM……278
Tabla 9.13. Resultados del ajuste de los parámetros del algoritmo PAUM……280
Tabla 9.14. Tiempos de construcción y de clasificación para el algoritmo PAUM……..283
Tabla 9.15. Medidas de efectividad para los algoritmo Naïve Bayes y C4.5……..284
Tabla 9.16. Medidas de efectividad para el algoritmo KNN……..287
Tabla 9.17. Medidas de medidas de efectividad para el algoritmo SVM……...290
Tabla 9.18. Medidas de medidas efectividad para el algoritmo PAUM……293
Tabla 9.19. Resultados de la evaluación mediante 10-fold……296

INDICE DE FIGURAS

Figura 2.1. Elementos del *e-learning* (Foix & Zavando, 2002)……...16
Figura 2.2. Relación entre habilidades y actividades realizadas (Barker, 2010)……..26
Figura 2.3. Relación entre competencias y habilidades (PICF, 2009) ……..28
Figura 2.4. Resumen de los pasos de un proceso de KDD (Fayyad *et al*., 1996b)……30
Figura 2.5. Representación bi dimensional de las operaciones OR y OR exclusivo……...35
Figura 2.6. Estructura de una red neuronal Biológica (Negnevitsky, 2005)……...36
Figura 2.7. Estructura de un red neuronal artificial típica (Negnevitsky, 2005)……..37
Figura 2.8. Diagrama de una neurona (Negnevitsky, 2005)……....38
Figura 2.9. Funciones de activación de una neurona (Negnevitsky, 2005)……..39
Figura 2.10. Perceptrón de capa simple con dos entradas (Negnevitsky, 2005)……..40
Figura 2.11. Separabilidad lineal en (a) un perceptrón de dos entradas y (b) un perceptrón de tres entradas (Negnevitsky, 2005)……...41
Figura 2.12. Algoritmo de entrenamiento del perceptrón (Negnevitsky, 2005)……..42
Figura 13.1 Etapas de un proceso de minería de textos (Aranha & Passos, 2008)……..45
Figura 3.2 Árbol de análisis sintáctico simple (Weiss *et al.*, 2005)……51
Figura 3.3. Modelo de "bolsa de palabras" con pesos binarios……52
Figura 3.4. Cálculo de la puntuación ponderada para un documento……..61
Figura 3.5. Representación grafica de clasificador SVM bi dimensional (Sebastiani2002)…….....63
Figura 3.6 Problema de clasificación bi dimensional con sus clasificadores SVM (Li & Shawe-taylor, 2003)……64
Figura 3.7. Algoritmo PAUM (Li *et al.*, 2002)……67
Figura 4.1. Niveles de la clasificación de habilidades……79
Figura 4.2. Relación entre las categorías de la clasificación de habilidades……79
Figura 5.1. *framework* para análisis de contenido (Krippendorff, 2004)……127
Figura 5.2. Diagrama de Toulmin (Krippendorff, 2004)……...131
Figura5.3. Diagrama de Toulmin aplicado al análisis de contenido (Krippendorff, 2004)……...131
Figura 5.4 Componentes del análisis de contenido (Krippendorff, 2004)……133
Figura 6.1 Fases del modelo de referencia CRISP-DM (Chapman *et al.*, 2001) ……..147
Figura 6.2 Marco de trabajo para minería de textos (Tan, 1999)…….……..160

Figura 14 Marco de trabajo de Palazzo para minería de textos (Magalhães, 2002)…….161
Figura 6.4. Marco de trabajo de Halliman para minería de textos (Magalhães, 2002)…..162
Figura 7.1. Modelo de hoja de trabajo para codificadores……186
Figura 7.2. Porcentaje de ocurrencia de las sub habilidades sociales……...193
Figura 7.3. Distribución de porcentaje entre los atributos de las sub habilidades sociales……...194
Figura 8.1. Modelo de grafo de anotación (Cunningham *et al.*, 2011)…..……..…201
Figura 8.2. Estructura del archivo main.jape de *ANNIE Sentence Splitter*……204
Figura 8.3. Ejemplo de aplicación de regla puntosSuspensivosSucedenaEtc……212
Figura 8.4. Ejemplo de aplicación de regla reconocerDivisorconetc……213
Figura 8.5. Ejemplo de aplicación de regla preguntaExclamacionSucedenaEtc……..214
Figura 8.6. Ejemplo de aplicación de regla etcSeguidodedosPuntos……215
Figura 8.7. Ejemplo de aplicación de regla ignorarPtosSuspensivosInterrogacionExclamacion…..…….216
Figura 8.8. Ejemplo de aplicación de regla abreviaturaMinusculaMayusculaSeguidaPto……217
Figura 8.9. Ejemplo de aplicación de regla ptoDeEnumeracionNoDivisor……..218
Figura 8.10. Ejemplo de aplicación de regla parentesisEnEnumeracion……..219
Figura 8.11. Ejemplo de aplicación de regla preguntaexclamacionNoSegundaParte……220
Figura 8.12. Ejemplo de aplicación de regla eliminaranotacionExclIntePtosSuspensivos……..221
Figura 8.13. Ejemplo de aplicación de regla ptosSuspensivosdosPuntosAntesListado….222
Figura 8.14. Ejemplo de aplicación de regla eliminarAnotSplitdePtoEnumeracion…….223
Figura 8.15. Ejemplo de aplicación de regla omitirAnotSplitdePtoEnumeracion……224
Figura 9.1. Ejemplo de mensaje de e-tutor……247
Figura 9.2. Diagrama de barra con la frecuencia de los atributos de habilidad..……...247
Figura 9.3. Incidencia de las palabras frecuentes en atributos de habilidad……..251
Figura 9.4. Primera combinación de operaciones de pre procesamiento……...259
Figura 9.5. Segunda combinación de operaciones de pre procesamiento……..259
Figura 9.6. Tercera combinación de operaciones de pre procesamiento……..……..259
Figura 9.7. Cuarta combinación de operaciones de pre procesamiento……..……...260
Figura 9.8. Quinta combinación de operaciones de pre procesamiento……..……...260
Figura 9.9. Sexta combinación de operaciones de pre procesamiento……..……260
Figura 9.10. Séptima combinación de operaciones de pre procesamiento….……261

Figura 9.11. Octava combinación de operaciones de pre procesamiento……..261
Figura 9.12. Corpus de mensajes, base de datos, y mensajes del Corpus……..262
Figura 9.13. Ejemplo de anotación “Sentence” con la característica “atributo_hab”……264
Figura 9.14. Archivo de configuración xml para el algoritmo KNN……268
Figura 9.15. Definición del elemento NGRAM cuando se usa lematización en la tercera combinación de operaciones de pre procesamiento……..……..……..…….271
Figura 9.16. Definición del elemento NGRAM cuando se usa *stemming to a root* en la tercera combinación de operaciones de pre procesamiento……..……..……...272
Figura 9.17. Definición del elemento NGRAM para la cuarta combinación de operaciones de pre procesamiento……..……..……..……..……..……..272
Figura 9.18. Definición del elemento NGRAM cuando se usa lematización en la séptima combinación de operaciones de pre procesamiento……..……..……..…….273
Figura 9.19. Definición del elemento NGRAM cuando se usa *stemming to a root* en la séptima combinación de operaciones de pre procesamiento……..……...273
Figura 9.20. Archivo de configuración xml para el algoritmo SVM……..……..276
Figura 9.21. Archivo de configuración xml para el algoritmo PAUM……..……...282
Figura 9.22. Conjuntos de anotación usados en *Corpus Quality Assurance*……..……286
Figura 9.23. Variación de la efectividad de los clasificadores construidos con K=5……288
Figura 9.24. Variación de la efectividad de los clasificadores construidos con K=10…..288
Figura 9.25. Variación de la efectividad de los clasificadores construidos con K=15…..289
Figura 9.26. Variación de la precisión de los clasificadores construidos con SVM……..291
Figura 9.27. Variación del recall de los clasificadores construidos con SVM……...291
Figura 9.28. Variación de la precisión de los clasificadores construidos con PAUM…….……..……..……..……..……..……..……..……..…...294
Figura 9.29. Variación del recall de los clasificadores construidos con PAUM……295
Figura 9.30. Cambio la precisión y recall de acuerdo al método de evaluación para los modelos construidos con KNN.…..……..……..……..……..…..…….297
Figura 9.31. Cambio la precisión y recall de acuerdo al método de evaluación para los modelos construidos con SVM.……..……..……..……..……297
Figura 9.32. Cambio la precisión y recall de acuerdo al método de evaluación para los modelos construidos con PAUM.……..……..……..……..……...298
Figura 9.33. Síntesis de la metodología usada en el proyecto……..……..…...302
Figura 9.34. Interfaz de usuario con el resultado del reconocimiento automático de habilidades…..……..……..……..……..……..……..…306

RESUMEN

Los sistemas de Aprendizaje Colaborativo Soportado por Computadora (ACSC) facilitan alcanzar el aprendizaje grupal con independencia del tiempo y espacio donde estén localizados los estudiantes. Sin embargo, la tecnología por sí misma no garantiza la creación de un ambiente donde la colaboración se produzca. Para colaborar efectivamente los estudiantes necesitan (entre otras cosas) de un e-tutor (docente) que coordine las discusiones, promueva la participación y gestione adecuadamente los conflictos que surjan durante la dinámica grupal. Teniendo en cuenta que la figura del e-tutor es clave para el éxito de las experiencias de ACSC, y que se sabe poco sobre la manera en que conviene que éste intervenga en las actividades de aprendizaje colaborativo de los estudiantes, se planteó como primer objetivo para este trabajo, el crear un esquema de clasificación de las habilidades que se considera deberían poseer los e-tutores de ACSC para desempeñarse de manera efectiva. Considerando que el análisis manual las interacciones registradas en entornos de ACSC requiere mucho tiempo y esfuerzo, el segundo objetivo consistió en desarrollar métodos automáticos de análisis de interacciones y reconocimiento de habilidades.

La taxonomía de habilidades fue creada a partir de una integración entre consideraciones vinculadas con las habilidades de e-tutores en e-learning no colaborativo, y algunas recomendaciones para docentes existentes en bibliografía sobre ACSC. Se espera que la clasificación de habilidades planteada no solo ayude a mejorar la comprensión de la tarea de los docentes de ACSC, sino que también sirva de guía tanto para diseñar actividades de capacitación que permitan a los e-tutores desarrollar sus habilidades, como para seleccionar o diseñar herramientas de software que brinden un soporte adecuado a la tarea de los mismos.

Una vez planteada la taxonomía de habilidades se siguieron los lineamientos de la metodología CRISP-DM para construir, mediante técnicas de minería de textos o *text mining*, clasificadores que identifican automáticamente las habilidades manifestadas por los e-tutores al interactuar con estudiantes, en entornos de ACSC que utilizan interfaces de texto libre en español. Para construir los clasificadores se realizó una serie de experiencias de ACSC con docentes y estudiantes de la FCEyT. Las contribuciones docentes obtenidas en estas experiencias fueron objeto de un doble análisis, mediante la técnica de análisis de contenido y también mediante los clasificadores logrados por minería de textos. La

contrastación de los resultados alcanzados por ambos análisis permitió determinar que el *text mining* es una técnica efectiva para reconocer automáticamente las habilidades manifestadas por los e-tutores.

Palabras Clave: *Aprendizaje Colaborativo Soportado por Computadora, e-tutor, Clasificación de habilidades de e-tutor, Text mining, Reconocimiento automático de habilidades.*

CAPÍTULO I
PLANTEAMIENTO DEL PROBLEMA

I.1. INTRODUCCIÓN

El término Aprendizaje Colaborativo Soportado por Computadoras (ACSC) se usa para designar situaciones en las cuales el aprendizaje se organiza mediante actividades colaborativas, mediadas por computadora, que involucran a grupos de estudiantes. Los sistemas de ACSC no sólo permiten a las personas aprender en grupo, con independencia del tiempo y espacio donde estén localizadas, sino que brindan un entorno apropiado para el desarrollo de habilidades de solución de problemas (Suh & Lee, 2006) y de liderazgo (Day, 2009). Pero, generalmente en los ambientes de ACSC las interacciones que son ideales para el aprendizaje no ocurren espontáneamente (Tchounikine *et al.*, 2010; Schwarz & Asterhan, 2011; Onrubia & Engel, 2012), y el uso de computadoras incluso puede llegar a inhibirlas (Orvis & Lassiter, 2006), por lo que resulta imprescindible la oportuna intervención del e-tutor.

Si se tiene en cuenta que los e-tutores de ACSC usualmente no saben cuál es la manera más efectiva de asistir a los estudiantes (Kukulska-Hulme, 2004) y que existen pocas chances de éxito sin docentes con las habilidades necesarias para sacarle provecho a las herramientas tecnológicas disponibles (Hughes *et al.*, 2002; Kukulska-Hulme, 2004; Orvis & Lassiter, 2006), es evidente la necesidad de plantear mecanismos que le permitan a los e-tutores adquirir las habilidades necesarias para desempeñarse adecuadamente en entornos de ACSC. De acuerdo con Barker (2002) una alternativa consiste en desarrollar sistemas de software que entrenen a los e-tutores. Sin embargo, tanto la falta de conocimiento sobre las habilidades requeridas en un e-tutor colaborativo (Kukulska-Hulme, 2004; Orvis & Lassiter, 2006; Zumbach *et al.*, 2004) como los efectos negativos sobre el proceso de comunicación que tienen las técnicas de análisis de interacciones basadas en interfaces estructuradas o semi estructuras que se han utilizado comúnmente en las investigaciones sobre ACSC (Chen, 2006; Constantino-González & Suthers, 2001; Soller *et al.*, 2005), se presentan como principales obstáculos a vencer si se pretende desarrollar aplicaciones que permitan a los e-tutores adquirir las habilidades que manifiestan con deficiencia.

Con el propósito de contribuir a la superación de los dos obstáculos antes mencionados, en el presente trabajo de investigación se definió un esquema de clasificación de las habilidades que debería poseer un e-tutor de ACSC. De forma complementaria a la taxonomía de habilidades, se aplicaron técnicas de clasificación de minería de textos para construir clasificadores con la capacidad de identificar automáticamente las habilidades manifestadas por los e-tutores al interactuar con sus alumnos. Se considera que de esta manera se contribuirá a la viabilidad futura del desarrollo de sistemas informatizados que permitan entrenar a los e-tutores de ACSC.

La creación del esquema de clasificación de habilidades se hizo en base a una recopilación bibliográfica sobre competencias y habilidades para e-tutores de entornos de e-learning no colaborativo, es decir entornos donde los estudiantes no trabajan en grupo e interactúan solo con el docente. En el proceso de selección de las habilidades que formarían parte de la taxonomía presentada en este trabajo se tuvieron en cuenta lineamientos y recomendaciones para docentes existentes en la literatura propia del ACSC. Cabe aclarar que la clasificación creada es necesaria para aplicar las técnicas de minería de textos en el análisis de las interacciones, pero de ninguna manera puede considerarse de carácter definitivo. Por el contrario, por ser este un trabajo interdisciplinario con la pedagogía, se piensa que la misma podrá ser tomada como punto de partida por los profesionales del área que se interesen en realizar futuras investigaciones que permitan su validación científica.

La construcción de los clasificadores automáticos mediante minería de textos sigue la aproximación de aprendizaje de máquina, es decir, los clasificadores o modelos creados pueden considerarse como una función que analizando las características de contribuciones docentes clasificadas a mano por personas expertas determina de manera inductiva las condiciones que deben cumplir los mensajes para ser asignados a un determinado tipo de habilidad (Sebastiani, 2002). Para obtener el conjunto de interacciones clasificadas por expertos se realizó una serie de experiencias de ACSC donde docentes de la FCEyT guiaron a grupos de estudiantes durante el desarrollo de sus tareas. Sobre los mensajes publicados por los e-tutores se aplicó la técnica de análisis de contenido de modo tal que, un psicopedagogo y un docente con experiencia en ACSC (expertos) hagan corresponder a cada oración de los mensajes de los e-tutores un atributo de habilidad del esquema de clasificación propuesto. En tanto que, en la construcción de los clasificadores o modelos se siguieron los lineamientos de la metodología CRISP-DM (Chapman *et al*., 2001). Para evaluar los clasificadores fue preciso que los mensajes docentes se dividieran en dos

conjuntos, uno de entrenamiento y el otro de prueba. Los algoritmos de aprendizaje de máquina construyen los clasificadores en base a las características de los documentos del conjunto de entrenamiento, en tanto que la comparación entre las habilidades automáticamente asignadas al conjunto de prueba (mediante los clasificadores) con las decisiones de clasificación efectuadas por los expertos durante el análisis de contenido, permitió determinar el grado de efectividad de la minería de textos en la detección de habilidades.

I.2. PLANTEAMIENTO DEL PROBLEMA

El uso de medios computacionales en el dominio del Aprendizaje Colaborativo (AC) originó los sistemas de ACSC, estos sistemas permiten alcanzar el aprendizaje grupal con independencia del tiempo y espacio donde estén localizados los estudiantes. Sin embargo, disponer de herramientas de software que soporten el aprendizaje en grupo no garantiza que los estudiantes colaboren, porque no son las características de la tecnología utilizada sino la forma en que esta se utiliza lo que soporta la colaboración (Constantino-González & Suthers, 2001; Hughes *et al*., 2002; Olivares, 2007; Orvis & Lassiter, 2006). Así, el éxito de la colaboración *online* depende (entre otras cosas) de las habilidades con las que cuentan los e-tutores, puesto que son los encargados de coordinar y mediar las discusiones grupales (Borges & Baranauskas, 2003; Kukulska-Hulme, 2004; McPherson & Baptista Nunes, 2004).

La exploración bibliográfica realizada desafortunadamente reveló que no existe una base teórica sólida sobre las habilidades de los e-tutores de ACSC (Kukulska-Hulme, 2004; Orvis & Lassiter, 2006; Zumbach *et al.*, 2004; Greiffenhagen, 2012) y que los pocos estudios donde se evaluó la efectividad de diferentes estrategias de enseñanza han conducido a resultados inconsistentes (Olivares, 2007). A esto habría que sumarle la falsa creencia de que los docentes con experiencia en clases presenciales solo necesitan aprender a usar la tecnología (Awouters & Jans, 2009; Gerrard, 2002; Olivares, 2007; Orvis & Lassiter, 2006) porque en entornos de ACSC pueden seguir utilizando técnicas propias del aula tradicional con las cuales están familiarizados (Barker, 2002; Orvis & Lassiter, 2006; Dewar & Whittington, 2000; Williams & Roberts, 2002). Si se tiene en cuenta además que los e-tutores usualmente no tienen experiencia con las técnicas usadas en ACSC (Williams & Roberts, 2002; Kukulska-Hulme, 2004), y que las iniciativas de este tipo tienen pocas chances de ser exitosas sin docentes que posean habilidades necesarias para sacarle provecho a las herramientas disponibles (Hughes *et al*., 2002; Kukulska-Hulme, 2004;

McPherson & Baptista Nunes, 2004; Orvis & Lassiter, 2006), es evidente que los docentes necesitan capacitación y formación para participar en procesos de ACSC. Surge así el primer interrogante, ¿Qué habilidades debería poseer un e-tutor de ACSC para cumplir su rol de manera efectiva?

Una vez especificadas las habilidades lo siguiente sería brindar a los e-tutores mecanismos que les permitan adquirir aquellas que manifiestan con deficiencia. Una de las alternativas es el entrenamiento mediante herramientas de software (Barker, 2002; do Prado & Ferneda, 2008). En el contexto del ACSC lo anterior, sin embargo, no es una empresa sencilla tanto por la falta de estudios en relación a las habilidades como por la dificultad de analizar las contribuciones de los docentes. En referencia al primer punto, conviene tener en cuenta que las circunstancias influyen en la forma en que tiene lugar la intervención del docente (De Smet *et al*., 2008; Kukulska-Hulme, 2004), por lo cual si al analizar los mensajes no se identifican determinadas habilidades esto no necesariamente significa que el e-tutor no cuente con tales habilidades sino que quizás no eran necesarias en la situación bajo análisis. Lo antedicho implica que otro de los problemas existentes (si bien no fue abordado en este trabajo) consiste en identificar las circunstancias bajo las cuales se debería realizar el entrenamiento.

Con respecto al análisis de las contribuciones docentes en un ambiente ACSC, la identificación de habilidades puede realizarse de forma manual o automática, pero en ambos casos no es una tarea sencilla. Por un lado, cuando se tiene una cantidad considerable de interacciones las aproximaciones manuales son prácticamente inviables debido al tiempo, esfuerzo y dinero que demandan (Chen & Wasson, 2002; Rosé *et al.*, 2008). Por otro lado, para el análisis automático de las interacciones *online* se ha recurrido tanto a interfaces estructuradas y semi estructuradas como a texto libre. Las interfaces estructuradas y semi estructuradas (oraciones de apertura, contribuciones dirigidas por menú, contribuciones basadas en diagramas, y grafos de conversación) (Barros & Verdejo, 2001; Chen, 2006; Soller *et al.*, 2005) permiten simplificar el proceso de análisis al no tener que usar técnicas de procesamiento de lenguaje natural (Constantino-González & Suthers, 2001; Soller *et al*., 2005; Tchounikine *et al.*, 2010). Sin embargo, existen investigaciones que demuestran que este tipo de interfaces socavan el proceso de colaboración porque restringen las posibilidades de interacción, provocan que la comunicación sea lenta, y crean stress relacional (Chen, 2006; Constantino-González & Suthers, 2001; Olivares, 2007; Soller *et al.*, 2005). En relación al texto libre, al permitir que la comunicación se realice sin restricciones ya no es posible usar la interfaz para hacer

inferencias sobre el proceso de interacción, y por consiguiente, es necesario recurrir a técnicas de procesamiento de lenguaje natural. Hasta el momento, este tipo de tecnología se usó de manera limitada en tareas como clasificación de los tópicos de una conversación, caracterización de patrones de discusión, o evaluación de la precisión de los estudiantes en el uso de sentencias de apertura (Tchounikine *et al.*, 2010).

Dadas las deficiencias asociadas con el análisis manual y automático de interacciones se necesita una técnica capaz de detectar las habilidades que poseen los e-tutores sin afectar negativamente la dinámica de trabajo de los estudiantes o docentes, ni representar un alto costo o sobrecarga. En este contexto la minería de textos o *text mining* (en particular sus técnicas de clasificación) se muestra a priori como una alternativa viable para cumplir con las condiciones mencionadas, dada su capacidad para manejar la incertidumbre, vaguedad, borrosidad, diversidad de estructuras, y gran cantidad de palabras que caracterizan al lenguaje natural (Hotho *et al.*, 2005). Así nace la segunda pregunta de investigación ¿Cómo se puede utilizar *text mining* en la identificación de las habilidades que posee un e-tutor de ACSC?

I.3. JUTIFICACIÓN DE LA RELEVANCIA DEL PROBLEMA

Con el rápido desarrollo de la sociedad basada en conocimiento ha crecido la importancia que se le da a la creación colaborativa de conocimientos. En este contexto, el ACSC se ha vuelto una nueva forma de educación donde los estudiantes crean conocimientos por medio de la interacción con personas localizadas en cualquier punto del planeta (Suh & Lee, 2006). Por otro parte, las investigaciones realizadas demuestran que los sistemas de ACSC brindan un entorno apropiado para el desarrollo de habilidades tales como solución de problemas (Suh & Lee, 2006), pensamiento crítico, mayor retención, establecimiento de metas, interpretación, y análisis, todas las cuales son cualidades de un líder exitoso (Day, 2009). Además los entornos de ACSC pueden mejorar el aprendizaje, reducir el sentimiento de aislamiento, e incrementar la motivación (Hughes *et al*., 2002). Sin embargo, la tecnología por sí misma no garantiza la creación de un ambiente donde la colaboración pueda ocurrir (Hughes *et al*., 2002; Olivares, 2007; Orvis & Lassiter, 2006; Onrubia & Engel, 2012; Kreijns *et al.*, 2003) porque para colaborar efectivamente los estudiantes necesitan (entre otras cosas) de un e-tutor que coordine las discusiones, promueva la participación, y ayude a enfrentar los problemas que surjan durante el trabajo grupal (Awouters & Jans, 2009; Borges & Baranauskas, 2003; Guzdial *et al*., 2002; Hughes *et al*., 2002; Kreijns *et al.*, 2003; Orvis & Lassiter, 2006; Soller, 2001; Schwarz &

Asterhan, 2011). Según Borges y Baranauskas (2003) y Kukulska-Hulme (2004) tanto la calidad como la profundidad del aprendizaje dependen de la forma en que la colaboración sea coordinada pero, la mayoría de los e-tutores no están familiarizados con las técnicas que se requieren en ACSC, ni se conocen con precisión cuáles habilidades deberían poseer para desempeñarse adecuadamente.

La taxonomía de habilidades creada durante el desarrollo del presente trabajo contribuirá a tener una mejor comprensión de la tarea de los e-tutores de ACSC por lo cual su utilidad es tanto pedagógica como tecnológica. En el plano pedagógico, la clasificación podría ser considerada al diseñar actividades de capacitación que permitan a los e-tutores desarrollar las habilidades necesarias para desempeñarse efectivamente. En el plano tecnológico, la taxonomía de habilidades tiene un gran valor ya que antes de seleccionar o diseñar herramientas de software que brinden soporte a la tarea de los e-tutores, es necesario comprender los comportamientos de los docentes que facilitan el aprendizaje grupal (Olivares, 2007; Greiffenhagen, 2012)

Para analizar las interacciones registradas en entornos de ACSC una persona podría utilizar el esquema de clasificación de habilidades creado pero, esto es un proceso tedioso que demanda mucho tiempo, esfuerzo y dinero, tanto considerando el entrenamiento en el uso del esquema como su aplicación. Por lo tanto, la taxonomía de habilidades contribuye parcialmente a solucionar el problema de determinar las habilidades que posee un e-tutor. El uso de minería de textos en la identificación automática de habilidades no solo hará menos trabajoso el análisis de interacciones mediante interfaces basadas en texto libre, sino que además permitirá que el proceso de análisis se lleve a cabo sin tener que restringir las posibilidades de comunicación entre estudiantes y docentes.

I.4. ANTECEDENTES

En la literatura sobre ACSC se reconoce la importancia de la figura del e-tutor en el éxito del aprendizaje, sin embargo, poco se sabe sobre la manera en la que los docentes intervienen en las actividades de aprendizaje colaborativo de los estudiantes (Greiffenhagen, 2012). A pesar de que las habilidades propias de un e-tutor de ACSC no han sido descriptas detalladamente, entre los investigadores que han estudiado esta temática se pueden identificar dos enfoques. Por un lado, se tienen trabajos como los de (Fåhræus, 2000; Juárez *et al.*, 2008; Lim & Cheah, 2003; Olivares, 2007; Orvis & Lassiter, 2006; Palloff & Pratt, 2001) donde se plantean roles o se proporcionan guías y recomendaciones para e-tutores de ACSC pero sin brindar mayores precisiones. Por el

otro, los trabajos de (Bonk *et al.*, 2004; Lund, 2004; Palloff & Pratt, 2007; Teles *et al*., 2001) analizan el rol de los e-tutores de ACSC tomando como base el esquema de organización por roles, heredado de la literatura sobre habilidades y competencias para docentes de e-learning que no recurren al aprendizaje colaborativo como estrategia pedagógica sino que se basan en el modelo de interacción docente-estudiante.

Teniendo en cuenta que los investigadores agrupados bajo el segundo enfoque solo consideraron a Berge (1995) y Mason (1991), se asume que para ampliar la perspectiva sobre la tarea de los e-tutores podría ser de utilidad indagar fuentes como (Barker, 2002; Blackburn, 2004; Chang, 2004; Cornelius & Higgison, 2001; Denis *et al.*, 2004; Dewar & Whittington, 2000; Gerrard, 2002; González & Salmon, 2004; Goodyear *et al.*, 2001; Higgison, 2001; IITT, 2007; McPherson & Baptista Nunes, 2004 ; Reid, 2002; Reid, 2003; Smith, 2005; Sulčič & Sulčič, 2007) donde también se plantean esquemas de clasificación de roles, competencias, y habilidades relativos a e-tutores de ambientes donde los estudiantes aprenden de manera individual. No obstante, no se debería perder de vista que en la dinámica grupal surgen problemas (personas que nunca participan, personas que dominan la discusión, discusiones sin fin, etc.) que no están presentes cuando los alumnos estudian solos.

En lo que respecta a la minería de textos, aunque no se localizaron antecedentes sobre su uso en el proceso de identificación de habilidades, puede ser de ayuda el considerar trabajos de investigación donde el análisis de las interacciones en sistemas de ACSC, que cuentan con interfaces de texto libre, se realizó mediante técnicas de procesamiento de lenguaje natural, aprendizaje de máquina, minería de datos y minería de textos:

- En el sistema ARGUNAUT las discusiones entre los estudiantes ocurren en un espacio de trabajo gráfico compartido donde para hacer una contribución los estudiantes seleccionan formas geométricas pre definidas, completan con texto las formas seleccionadas para expresar sus aportes, y vinculan las formas geométricas a las contribuciones de otros estudiantes (Tchounikine *et al.*, 2010). Teniendo en cuenta que la carga cognitiva de moderar simultáneamente múltiples discusiones puede ser demasiado alta para los docentes, ARGUNAUT intenta resumir las discusiones de los estudiantes y alertar a los e-tutores sobre aspectos y eventos críticos que requieran su intervención (Witten & Frank, 2005). El análisis de las contribuciones de los estudiantes y las discusiones se hace por medio de técnicas de

aprendizaje de máquina y procesamiento de texto (Witten & Frank; Tchounikine *et al*., 2010).

- El trabajo de Kumar *et al.* (2007) en torno al sistema CycleTalk tiene por finalidad brindar asistencia en tiempo real a estudiantes colaborativos en tareas de resolución de problemas. Mediante clasificadores de aprendizaje de máquina estos investigadores identifican los tópicos de las discusiones que mantienen los estudiantes por medio de una herramienta de chat. Cada vez que se detecta uno de los tópicos pre establecidos un agente de software interactúa con los estudiantes para estimular la reflexión sobre conceptos vinculados al tópico. Además, en CycleTalk se trata de lograr que los alumnos mantengan su compromiso y participen de manera equitativa en las conversaciones.
- La aproximación de Padilha *et al.* (2004) consiste en combinar técnicas de minería de textos con técnicas de minería de datos. La minería de textos permite identificar y clasificar las contribuciones de los estudiantes en un chat o foro. A su vez, la tipificación de los diálogos sirve de base para, mediante técnicas de minería de datos, producir una serie de informes (a nivel grupal o de cada estudiante) que permiten comparar el estado actual de la interacción con un estado ideal.
- Partiendo de la idea de que para brindar soporte a la escritura colaborativo se tienen que integrar los productos de software que proveen realimentación sobre los documentos producidos con las herramientas que proveen información sobre el proceso de redacción, Reimann y sus colegas desarrollaron CWE (Reimann *et al*., 2010). La arquitectura de CWE tiene tres componentes principales: Google Docs, Glosser y WriteProc. Google Docs provee la funcionalidad para las actividades de escritura además de encargarse de mantener un registro de todas las versiones de los documentos y de la interacción de los estudiantes con los documentos. WriteProc aplica una combinación de técnicas estadísticas y técnicas de minería de procesos (sobre los datos de Google Docs) para extraer conocimientos sobre la cooperación e interacción grupal. En tanto que, Glosser recurre a técnicas de minería de textos para comprender cómo se desarrollaron los conceptos e ideas durante el proceso de escritura.

Finalmente, conviene tener en cuenta que si bien hasta el momento no existe una metodología estándar para conducir estudios de minería de textos, en Magalhaes (2002) se demostró que la metodología de minería de datos CRISP-DM puede aplicarse

exitosamente si se hacen algunos ajustes. Esto coincide con los planteos de investigadores tales como (Feldman & Sanger, 2007; García Adeva & Calvo, 2006; Weiss et *al*., 2005) que consideran a la minería de textos como una especialización de la minería de datos.

I.5. HIPÓTESIS Y OBJETIVOS

Teniendo en cuenta lo tedioso del análisis manual de las interacciones registradas en entornos de ACSC, y los efectos negativos de las aproximaciones automatizadas basadas en interfaces estructuradas o semi estructuradas, se plantea la siguiente hipótesis que actuará de guía durante el desarrollo del presente trabajo:

Las técnicas de clasificación de la minería de textos permiten identificar de manera efectiva las habilidades manifestadas por los e-tutores en los mensajes publicados en entornos de ACSC.

I.4.1. Objetivos Generales

- Contribuir a la comprensión de la tarea de los e-tutores de ACSC.
- Promover la mejora de la capacitación de los e-tutores de ACSC.
- Auxiliar en el proceso de selección y/o desarrollo de herramientas de software que provean soporte adecuado a la tarea de los e-tutores.
- Favorecer la ampliación del área de aplicación de la minería de textos.
- Contribuir a la reducción del tiempo, costo, y esfuerzo requeridos para identificar las habilidades manifestadas por e-tutores de ACSC.

I.4.2. Objetivos Específicos

- Elaborar un esquema de clasificación de las habilidades que requieren los e-tutores para desempeñarse adecuadamente en ambientes de ACSC.
- Identificar automáticamente mediante clasificadores (modelos) las habilidades manifestadas por los e-tutores de ACSC.
- Identificar las operaciones de pre procesamiento que permiten lograr el mayor grado de efectividad en la identificación automática de las habilidades de e-tutores de ACSC.

- Determinar el grado de efectividad de diversos algoritmos de clasificación de minería de textos en la identificación automática de las habilidades manifestadas por los e-tutores de ACSC.

I.6. SOLUCIÓN PROPUESTA

La primera parte de la solución consistió en la creación de una clasificación de habilidades de e-tutores que se organiza en tres niveles (categorías de habilidad, sub habilidades, y atributos de habilidad), donde cada nivel proporciona más detalles sobre su predecesor.

Para obtener el conjunto de datos (entrenamiento y prueba) que se usó para aplicar minería de textos se realizaron varias experiencias de ACSC en el entorno de aprendizaje Moodle. Pero las interacciones por sí mismas no son suficientes para construir los clasificadores porque se necesita conocer los atributos de habilidad manifestados por los e-tutores. Mediante la técnica de análisis de contenido, expertos humanos hicieron corresponder uno o más atributos de habilidad a los mensajes de los docentes. De acuerdo con Krippendorff (2004) los elementos que forman parte de un estudio de análisis de contenido son:

- Separación en unidades: consiste en distinguir sistemáticamente los segmentos de texto que son de interés para llevar a cabo el análisis. Es necesario que la información necesaria para el análisis esté representada en la colección de unidades y no en las relaciones entre unidades, porque las relaciones no son tenidas en cuenta.
- Muestreo: permite economizar el esfuerzo de investigación limitando las observaciones a un subconjunto manejable de unidades que es estadísticamente o conceptualmente representativo del conjunto de todas las unidades posibles, la población o universo de interés. De forma ideal, el análisis de la población completa y el análisis de una muestra representativa de la población deberían conducir a los mismos resultados.
- Grabación: los analistas que conducen el estudio confeccionan instrucciones de grabación que son utilizadas por terceros (codificadores) para leer e interpretan los textos. Las instrucciones de grabación permiten a los codificadores expresar sus experiencias con los textos en un lenguaje de datos que pueda ser analizado con las técnicas apropiadas.

- Reducción: cobra relevancia cuando se trabaja con volúmenes grandes de datos ya que consiste en aplicar técnicas estadísticas u otros métodos para resumir o simplificar los mismos de modo tal de lograr una representación eficiente. Una lista de los tipos de token y su frecuencia de ocurrencia es una representación más eficiente que la tabulación de todas las ocurrencias. La lista reemplaza duplicaciones por una frecuencia y dado que una representación puede ser creada desde otra nada se pierde.
- Inferencia: a partir de los textos analizados se generan inferencias que buscan responder a las preguntas de investigación que motivaron el estudio de análisis de contenido.
- Narración: se relatan los resultados de la investigación de una manera que sean comprensibles para otras personas. Esto puede significar explicar el significado práctico de los descubrimientos, discutir lo apropiado del uso del análisis de contenido en lugar de técnicas observacionales directas, o bien hacer recomendaciones para acciones legales, prácticas, o investigación posterior.

Una vez que los expertos humanos identificaron los atributos de habilidad en el conjunto de interacciones se procedió a construir los clasificadores siguiendo la metodología CRISP-DM (Chapman *et al*., 2001) cuyas fases se describen brevemente a continuación:

- Comprensión del negocio: se busca entender los objetivos y requerimientos del proyecto desde una perspectiva de negocios, para luego convertir este conocimiento en la definición de un problema de minería de datos y un plan preliminar para lograr los objetivos.
- Comprensión de los datos: luego de la recolección de datos se procede con una serie de actividades que permiten familiarizarse con los datos, identificar problemas de calidad de los datos, y detectar sub conjuntos interesantes para elaborar hipótesis sobre información oculta.
- Preparación de los datos: cubre todas las actividades para construir el conjunto de datos final (sobre el que se aplica la herramienta de modelado) a partir del conjunto inicial de datos no procesados. Las actividades de preparación de datos incluyen la selección de atributos así como la transformación y limpieza de los datos para las herramientas de modelado. Estas tareas usualmente se realizan varias veces y sin respetar un orden prescripto.

- Modelado: se seleccionan y aplican varias técnicas de modelado para luego calibrar sus parámetros a los valores óptimos. Algunas técnicas tienen requerimientos específicos sobre los datos por lo que usualmente se necesita volver a la fase de preparación de datos.
- Evaluación: antes de proceder a la fase de despliegue del modelo (o los modelos) construido es fundamental evaluarlo exhaustivamente y revisar los pasos ejecutados para construir el modelo para estar seguros de que permite cumplir con los objetivos del negocio. Al finalizar esta fase se debe decidir si se usará los resultados de la minería de datos.
- Despliegue: la creación del modelo generalmente no es el fin del proyecto. Incluso cuando la finalidad del modelo es incrementar el conocimiento sobre los datos, el conocimiento ganada necesitar ser organizado y presentado en una manera que el cliente lo pueda usar. La fase de despliegue puede ser tan simple como generar un reporte o compleja como implementar un proceso de minería de datos repetible en toda organización.

Durante la aplicación de la metodología CRISP-DM se recurrió a la herramienta de minería de textos GATE (General Architecture for Text Engineering) para automatizar actividades de preparación de datos (identificación de oraciones, identificación de *Tokens*, supresión de *stop words*, etc.), construir los modelos con diversos algoritmos de clasificación (KNN, SVM, PAUM, C4.5, y Näive Bayes), y calcular métricas típicamente utilizadas para evaluar la efectividad de clasificadores automáticos. Cabe resaltar que se realizaron numerosas modificaciones sobre los recursos programados en GATE a fin de adecuar la herramienta al uso específico que se le dio en este trabajo, y al hecho de aplicarla sobre textos escritos en español (GATE fue programada para reconocer textos en lengua inglesa, la cual difiere mucho de nuestra ortografía y gramática).

Teniendo en cuenta que el presente trabajo es de carácter académico y que la construcción de los clasificadores mediante minería de textos es un proceso complejo que demanda un tiempo y esfuerzo considerables, se decidió trabajar solo con los atributos de la habilidad social, de la taxonomía propuesta, a fin de demostrar la viabilidad del uso de la minería de textos en el reconocimiento automático de habilidades. El interés en las habilidades sociales, vinculadas con la gestión del proceso de interacción entre estudiantes, se debe a que la adquisición de conocimiento en un ambiente de ACSC está influenciada

por las interacciones grupales, relaciones interpersonales y comunicación con otros (Dillenbourg, 1999; Kukulska-Hulme, 2004; Olivares, 2007; Orvis & Lassiter, 2006).

I.7. CONTRIBUCIONES

Al área del ACSC:

- Creación de un esquema de clasificación de las habilidades deseables en los e-tutores de ACSC.

Al área de la minería de textos:

- Definición de reglas que permiten que el recurso de procesamiento *ANNIE Sentence Splitter* de GATE pueda reconocer oraciones al procesar textos en español.
- Definición de reglas que permiten que en GATE pueda aplicarse la eliminación de *stop words* (preposiciones, artículos, conjunciones, etc.) al procesar textos en español.
- Creación de modelos que pueden usarse para reconocer automáticamente las habilidades sociales manifestadas por los e-tutores en entornos de ACSC que cuentan con herramientas de interacción basadas en texto libre.

I.8. ESTRUCTURA DEL TRABAJO

Este resto de los capítulos del presente trabajo se organizan en cuatro partes: marco teórico, marco metodológico, marco empírico, y conclusiones. En el marco teórico (capítulos II, III, IV) se describen los conceptos, las teorías, y la clasificación de habilidades de e-tutores que sustentan esta investigación. El marco metodológico se refiere a la técnica de analisis de contenido, la metodologia CRISP-DM, y las adaptaciones que se deberian hacer a CRISP-DM para utilizarla en proyectos de mineria de textos. En el marco empírico (capítulos VII, VIII, y IX) se documentan las adaptaciones realizadas a los recursos de procesamiento de GATE, los resultados del análisis de contenido, y la construccion de los clasificadores automáticos. Finalmente, en el capítulo X se presentan las conclusiones del trabajo realizado.

MARCO TEÓRICO

Para definir los conceptos de relevancia en el marco del presente trabajo de investigación, se decidió descomponer el marco teórico en los tres capítulos cuyos contenidos se describen brevemente.

Capítulo II: Nociones Preliminares

En este capítulo se definen términos de uso frecuente a lo largo del trabajo de modo tal de no solo facilitar la comprensión del lector sino también de evitar discrepancias que suelen presentarse en la literatura. En primer lugar, especifica el significado de conceptos tales como aprendizaje cooperativo, aprendizaje colaborativo, *e-learnig*, e-tutor, y grupo de aprendizaje, ya que son necesarios para comprender el campo del ACSC. Luego se establecen las diferencias entre lo que son las habilidades y las competencias dado que comúnmente existe confusión en cuanto a sus significados. Teniendo en cuenta que la minería de textos usa técnicas de la minería de datos, se comenta en que consisten el descubrimiento de conocimiento desde bases de datos y la minería de datos en particular. También se aborda el área del aprendizaje de máquina y las redes neuronales.

Capítulo III: Minería de Textos

En el segundo capítulo se enuncia una definición de míneria de textos y se plantean las etapas constitutivas de un proyecto de minería de textos. En particular, se pone enfasis en la fase de pre procesamiento dado que permite convertir los documentos de texto a un formato intermedio donde se pueden aplicar algoritmos de aprendizaje. Por último, se centra la atención en la tarea de clasificación y los algoritmos de aprendizaje que pueden utilizase para construir clasificadores automáticos.

Capítulo IV: Las Habilidades de los Docentes en Grupos de Aprendizaje Colaborativo *Online*

Luego de analizar la importancia del rol del e-tutor en entornos de ACSC y la falta de conocimiento con respecto a la tarea de los docentes en este tipo de ambientes, se plantea un esquema de clasificación de la habilidades que se considera deberian poseer los e-tutores de ACSC para desempeñarse de manera efectiva como tales.

CAPÍTULO II
NOCIONES PRELIMINARES

II.1. E-LEARNING

En términos generales el *e-learning* ha sido descrito como el uso de tecnología electrónica para entregar, soportar, y mejorar la enseñanza y el aprendizaje (David *et. al*, 2004). Pero, en el marco de este trabajo de investigación se tendrá en cuenta a (Rosenberg, 2001) que restringe el *e-learning* a los procesos de enseñanza y de aprendizaje que hacen uso de la tecnología de Internet puesto que considera que:

a) El *e-learning* se produce en red. Esto permite actualización en forma instantánea, almacenamiento/recuperación, distribución y compartición de la instrucción o información. Este requerimiento provoca que sistemas de aprendizaje basados en tecnologías tales como los CD o DVD no sean clasificados como *e-learning*, ya que carecen de la capacidad de red que permite que la información y la instrucción sean distribuidas y actualizadas en forma instantánea.
b) En *e-learning* se usa tecnología de Internet estándar (como el protocolo TCP/IP y los navegadores web) para crear la plataforma de entrega del material de aprendizaje.
c) El *e-learning* se enfoca en una visión amplia de soluciones de aprendizaje que van más allá de los paradigmas tradicionales de entrenamiento. Esto significa que el *e-learning* no incluye solo el entrenamiento sino también la entrega de información y el uso de herramientas para mejorar la performance.

El *e-learning* no consiste solamente en subir a la web los materiales de un curso, más bien se trata de combinar recursos, interactividad, soporte y actividades de aprendizaje estructuradas. Por lo cual, es acertado definir al *e-learning* como una actividad que utiliza de manera integrada y pertinente computadores y redes de comunicación, en la formación de un ambiente propicio para la construcción de la experiencia de aprendizaje (Foix & Zavando, 2002). Siguiendo los lineamientos de estos autores los elementos del *e-learning* serían:

- **Sistema de administración de aprendizaje (LMS)**: es el núcleo alrededor del cual giran todos los demás elementos. Consiste de software para servidores de Internet/Intranet que se ocupa de:
  - Gestionar los usuarios: inscripción, control de sus aprendizajes e historial, generación de informes, etc.
  - Gestionar y lanzar los cursos: manteniendo un registro de la actividad de los usuarios como ser los resultados de los tests y evaluaciones, así como también los tiempos y accesos al material formativo.
  - Gestionar los servicios de comunicación: programar y ofrecer recursos para la comunicación (foros de discusión, chat, videoconferencia, etc.) conforme sean necesarios.
- ***Courseware* o Contenidos**: los contenidos pueden ser entregados en formatos diversos, y dependiendo de si son adecuados o no para un curso en particular. Los formatos más habituales son video conferencia, presentaciones *Power Point*¸ elementos multimedia e interactivos, pizarra virtual, documentos disponibles para descarga, y foros de discusión.
- **Sistemas de comunicación sincrónica y asincrónica:** los sistemas de comunicación sincrónica (como chat de voz y videoconferencias) permiten la comunicación estudiante-estudiante o estudiante-docente en tiempo real. Por su parte, los sistemas asincrónicos (por ejemplo, e-mail, wiki, y foros) no brindan comunicación en tiempo real pero los aportes quedan registrados pudiendo ser consultados cuando sea necesario.

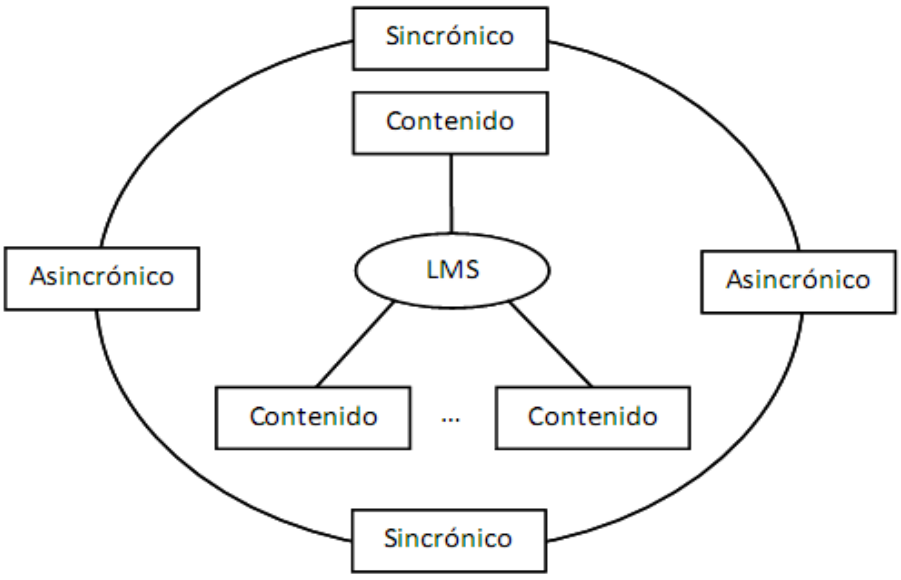


Figura 2.15. Elementos del *e-learning* (Foix & Zavando, 2002)

II.2 E-TUTOR

La función de tutoría comprende aquellos aspectos del trabajo de un profesor que involucran administrar y "animar" interacciones con y entre estudiantes, especialmente con respecto a su participación en actividades de aprendizaje en red (CSALT, 2001). Así entonces, se puede decir que un e-tutor o tutor *online* es alguien que apoya los procesos de aprendizaje por medio de la interacción directa con estudiantes que están frecuentemente separados unos de otros y del tutor en tiempo y espacio (Sulčič & Sulčič, 2007). Al soportar las interacciones un e–tutor busca que los estudiantes puedan aprender de manera efectiva en ambientes *online* (Cornelius & Higgison, 2001).

En la bibliografía sobre *e-learning* a los e-tutores se les asignan múltiples roles que van desde aquellos vinculados con dar consejos a los estudiantes, a los relacionados con proveer asistencia ante problemas con la tecnología. Aunque no todos los autores coinciden, básicamente se pueden identificar cuatro roles de los e-tutores:

- **Rol Social**: involucra la creación de un ambiente de aprendizaje cómodo, agradable y amigable en el cual los estudiantes sientan que el aprendizaje es posible (Mason, 1991; Llorente Cejudo, 2006; Bláquez Entonado & Alonso Díaz, 2006; Benito Osorio, 2009). Para crear este ambiente propicio para el aprendizaje el e-tutor debe promover relaciones humanas, desarrollar la cohesión grupal, ayudar a los estudiantes a trabajar juntos, animar y motivar a los estudiantes, etc. (Berge, 1995; Bláquez Entonado & Alonso Díaz, 2006).
- **Rol Pedagógico/Intelectual**: para cumplir con este rol el e-tutor actúa como un facilitador del proceso de aprendizaje que clarifica y explica los contenidos; focaliza la discusión en conceptos críticos, principios y habilidades; plantea preguntas e indaga en las respuestas de los estudiantes; interviene para promover conversaciones productivas e interesantes; da coherencia a la discusión; y sintetiza los debates (CSALT ,2001; Benito Osorio, 2009; Llorente Cejudo, 2006; Berge, 1996 ; Blackburn, 2004; Mason, 1991).
- **Rol Administrativo/Organizativo:** gran parte de la tarea organizativa de un e-tutor pasa por establecer la agenda de trabajo (objetivos, horarios, reglas de procedimiento, y normas de toma de decisión) de modo tal de mantener a los estudiantes trabajando hacia sus objetivos (Mason, 1991; Bláquez Entonado & Alonso Díaz, 2006; Benito Osorio, 2009; Llorente Cejudo, 2006). Pero, también es necesario que maneje las discusiones con dirección y liderazgo; se encargue de

recordarle a los estudiantes la planificación, reglas, procedimientos, etc.; organice y coordine el trabajo grupal; y maneje la sobrecarga de información (Mason, 1991; Berge, 1995; Cornelius & Higgison, 2001; Llorente Cejudo, 2006; CSALT ,2001).

- **Rol Técnico:** en este aspecto la tarea del e-tutor consiste en lograr que los alumnos se sientan confortables con el sistema de *software* de modo que se concentren en sus tareas (Berge, 1995; Cornelius & Higgison, 2001; CSALT ,2001; Benito Osorio, 2009; Llorente Cejudo, 2006). Para cumplir con su propósito el e-tutor debe ser competente con el sistema de *software* que forma parte del ambiente de *e-learning* y proveer soporte ante problemas técnicos (Llorente Cejudo, 2006).

Más allá de cuál sea el rol concreto que un e-tutor desempeñe, es importante tener en cuenta que los ambientes de ACSC requieren que el e-tutor actúe como un facilitador, encargándose de coordinar y mediar las discusiones grupales hacia el logro de las metas y los objetivos de aprendizaje de las actividades realizadas (Borges & Baranauskas, 2003a). El facilitador tiene la responsabilidad de crear una atmósfera de aprendizaje colaborativa promoviendo la participación y colaboración entre los estudiantes (Borges & Baranauskas, 2003a; Borges & Baranauskas, 2003b).

II.3 CONSTRUCTIVISMO

En el marco de esta teoría de aprendizaje se considera que el estudiante juega un rol activo y no pasivo, ya que el conocimiento no es recibido desde afuera o desde alguien sino que es la propia interpretación y el procesamiento de lo que se recibe a través de los sentidos lo que crea el conocimiento (Ally, 2004; Borges & Baranauskas, 2003a). Esto transforma al profesor en un facilitador o guía y convierte al estudiante en centro del aprendizaje, ya que debe construir su propio conocimiento en lugar de aceptar lo dado por el docente (Hiltz & Benbunan-Fich, 1997; McInnerney & Roberts, 2004; de Laat, 2006). Así, el aprendizaje es visto como un proceso activo en el cual los estudiantes se comprometen con y construyen nuevas ideas o conceptos en base a su conocimiento actual y/o pasado (experiencias pasadas, enseñanza formal, lectura, compartir con pares, y sus esfuerzos creativos). Son los estudiantes quienes seleccionan, transforman, integran, y hacen una elección informada (en base a sus modelos mentales únicos) en el desarrollo de su entendimiento (Jaques & Salmon, 2007).

II.4 CONSTRUCTIVISMO SOCIAL

Teoría propuesta por Vygotsky, y extensión de la teoría de Piaget sobre el desarrollo cognitivo individual, que pone énfasis en el contexto social-cultural donde tiene lugar el proceso de aprendizaje y en su impacto sobre lo que se aprende (Huang, 2002; Kumar, 2000). La aproximación socio-constructivista se centra en el desarrollo individual con respecto a la interacción social tanto entre estudiantes como entre el docente y los estudiantes. Esto significa que el e-tutor tiene que intervenir para ayudarle a los estudiantes en el manejo de los recursos de aprendizaje así como también en el manejo de las interacciones con sus pares y con el e-tutor mismo (Denis *et al.*, 2004).

II.5 GRUPO DE APRENDIZAJE

Antes de hacer referencia a los grupos de aprendizaje conviene dejar en claro que se entiende por grupo. De acuerdo a Souto (2009) un grupo se puede definir como aquello que no es la suma de individuos sino un plus que surge en la dinámica de la interacción. Así entonces, si bien los individuos son la materia prima de los grupos, para que una colección de individuos sea considerada un grupo debe poseer todas las cualidades siguientes en mayor o menor medida (Jaques & Salmon, 2007):

- **Percepción colectiva**: los miembros son colectivamente conscientes de su existencia como grupo.
- **Necesidades**: las personas se unen a un grupo porque creen que este va a satisfacer algunas de sus necesidades o les dará algunas recompensas como reconocimiento o comprensión de sí mismos a través de la realimentación.
- **Objetivos compartidos**: los miembros mantienen o desarrollan rápidamente objetivos o ideales compartidos que los unen. El logro de los objetivos será una de las recompensas.
- **Interdependencia**: los miembros son interdependientes en la medida en que son afectados por y responden a cualquier evento que afecta a cualquier miembro del grupo.
- **Organización social**: esto comprende una unidad social con normas, reglas, estatus, poder y relaciones emocionales.
- **Interacción**: los miembros influencian y responden a cada uno en el proceso de comunicación. El sentido de grupo existe incluso cuando los miembros no están reunidos en el mismo lugar, tal como cuando son parte de un grupo *online*.

- **Cohesión:** los miembros quieren permanecer en el grupo para contribuir a su bienestar y objetivos, y para participar en sus actividades.
- **Membrecía:** dos o más personas que interactúan explícita o implícitamente por más de un par de minutos constituyen un grupo si hay reconocimiento de vínculos mutuos.

Una definición más orientada al contexto de la educación es la dada por Jaques y Salmon (2007), quienes consideran que en un grupo las personas se reúnen para compartir conocimiento, para el desarrollo personal o para aprender a través de la discusión. Un tipo particular de grupo lo constituyen los grupos de aprendizaje, y de acuerdo con Souto (2009), se lo puede entender como una estructura formada por personas que interactúan en un espacio y tiempo común para lograr ciertos y determinados aprendizajes en los alumnos a través de su participación.

En el marco de este trabajo de investigación se considerará que un grupo de aprendizaje es un tipo particular de grupo cuya característica distintiva es el buscar el aprendizaje de los estudiantes por medio de la interacción mutua, pero sin imponer la restricción de que los estudiantes deben compartir un mismo espacio y tiempo puesto que, una de las características de los ambientes de ACSC es el posibilitar el aprendizaje grupal independientemente de la ubicación espacio-temporal de sus miembros.

II.6 APRENDIZAJE COLABORATIVO

El aprendizaje colaborativo es un modelo instruccional basado en el constructivismo social que puede ser definido como un proceso grupal desestructurado, orientado hacia el exterior donde el objetivo primario del grupo es generar a través de la interacción creativa la mejor solución o soluciones (Olivares, 2007). En el aprendizaje colaborativo se cultiva la independencia de los estudiantes, una atmósfera de disenso, la carencia de estructura, el libre intercambio de ideas (Olivares, 2007), y el respeto por las contribuciones de los pares (McInnerney & Roberts, 2004). Una de las características propias del aprendizaje colaborativo es lo que Dillenbourg (1999) denomina división horizontal del trabajo. Esto es, los miembros de un grupo no dividen el trabajo en tareas independientes sino más bien en capas de razonamiento, donde las capas están entretejidas (un individuo monitorea al otro) y los roles cambian a cada instante (con el regulador volviéndose regulado).

Según Tinzmann *et al*. (1990) las clases colaborativas se diferencian de las clases tradicionales por las cuatro características indicadas a continuación:

- **Conocimiento compartido entre profesores y estudiantes**: los docentes colaborativos valoran y aprovechan el conocimiento, la experiencia personal, estrategias, y la cultura que los estudiantes poseen. De esta manera, cuando los estudiantes ven que sus experiencias y conocimientos son valorados, se sienten motivados y es más probable que establezcan conexiones entre su propio aprendizaje y el aprendizaje escolar.
- **Autoridad compartida entre profesores y estudiantes**: a diferencia de las clases tradicionales los docentes colaborativos invitan a los estudiantes a establecer objetivos, sugerir actividades, y proveer opciones de evaluación de modo tal de capturar sus intereses y objetivos. Como consecuencia los estudiantes tienen oportunidades para preguntar e investigar cuestiones de interés personal, y tienen la voz en el proceso de toma de decisión. Esto es sumamente importante para el aprendizaje auto-regulado y la motivación.
- **Profesores como mediadores**: como consecuencia del conocimiento y la autoridad compartida los docentes pasan a ser mediadores del aprendizaje.
- **Grupos heterogéneos de estudiantes**: la perspectiva, la experiencia, y el historial de todos los estudiantes es importante para el aprendizaje enriquecido. Es por esto que, en clases colaborativas ningún estudiante es privado de la oportunidad de contribuir y realizar apreciaciones de las contribuciones de sus compañeros.

Los cuatro puntos antes mencionados traen aparejados cambios, tanto en los roles de los estudiantes y de los docentes, como en la manera en que estos interactúan. Para los docentes su nuevo rol se define en términos de la mediación o moderación del aprendizaje a través del diálogo y la colaboración (Tinzmann *et al*., 1990). Los moderadores ya no ponen énfasis en proveer la respuesta correcta o en decir cuales miembros del grupo están en lo cierto, sino en realizar una mínima intervención pedagógica para re-direccionar el trabajo grupal en una dirección productiva o monitorear cuales miembros quedan fuera de la interacción (Dillenbourg, 1999). En este sentido para Dillenburg (1999) el aprendizaje colaborativo alude a una situación en la cual se espera que ocurran formas particulares de interacción entre las personas, lo cual daría lugar a mecanismos de aprendizaje, pero sin que existan garantías de que tales interacciones esperadas realmente ocurran. Por

consiguiente, parte de la tarea del docente consiste en buscar formas de promover, y en lo posible garantizar, que las interacciones se dirijan a la construcción colaborativa de conocimiento entre los estudiantes (Orvis & Lassiter ,2006; McInnerney & Roberts, 2004).

II.7 APRENDIZAJE COOPERATIVO

Se entenderá que el aprendizaje cooperativo es el uso educativo de pequeños grupos que permitan a los estudiantes trabajar juntos para maximizar su propio aprendizaje y el de los demás (Jhonson *et al*., 1999). Esta definición quedaría incompleta si no se tiene en cuenta que el aprendizaje cooperativo es un proceso social motivado por el trabajo de grupo estructurado, y se ocupa de promover salidas tanto académicas como sociales; esto es, los estudiantes aprenden nuevas habilidades sociales y como trabajar juntos para lograr objetivos académicos. A lo largo de este proceso, el profesor actúa como un *manager* que impone la estructura y usa estrategias instruccionales para engendrar habilidades sociales, interdependencia positiva, cooperación y responsabilidad (Olivares, 2007). Según Johnson y Johnson (2008) para entender la dinámica de la cooperación y la tarea de los docentes es preciso considerar los siguientes cinco elementos:

- **Interdependencia positiva**: existe cuando los individuos perciben que pueden alcanzar sus objetivos si y solo si los individuos con los que están cooperativamente vinculados logran sus propios objetivos.
- **Responsabilidad individual**: la interdependencia positiva busca crear "fuerzas de responsabilidad" que incrementan el sentimiento de responsabilidad los miembros del grupo para: completar su parte del trabajo, y facilitar el trabajo de los otros miembros del grupo. En la medida que el desempeño de una persona afecta los resultados de los colaboradores, esta persona se siente responsable por el bienestar de sus pares y el propio. La responsabilidad individual existe cuando el desempeño de cada miembro del grupo se evalúa, los resultados se entregan a cada integrante y al grupo de modo tal de comparar contra un estándar de desempeño, y cada miembro es responsable ante sus compañeros de contribuir con su parte del trabajo al éxito del grupo.
- **Promoción de la interacción**: ocurre cuando los miembros del grupo estimulan y facilitan el esfuerzo de cada uno de manera de lograr los objetivos grupales.
- **Uso apropiado de habilidades sociales**: los miembros de un grupo deben tener o se les debe enseñar las habilidades interpersonales necesarias para la cooperación

de alta calidad y deben ser motivados para usarlas. Para coordinar sus esfuerzos en el logro de objetivos mutuos los integrantes de un grupo deben: conocerse y confiar en el otro, comunicarse con exactitud y sin ambigüedad, aceptarse y apoyarse entre sí, y resolver los conflictos de manera constructiva.

- **Procesamiento de Grupo**: puede ser definido como una reflexión en una sesión grupal para: describir las acciones de los miembros que son útiles y las acciones que son inútiles, y decidir qué acciones cambiar o continuar. El propósito del procesamiento de grupo es clarificar y mejorar la efectividad de los miembros en su contribución a los esfuerzos conjuntos para lograr los objetivos grupales.

Como se puede apreciar en la Tabla 2.1 el aprendizaje cooperativo se diferencia del colaborativo en cuanto al grado de control del docente, la estructura impuesta a la tarea, la forma de dividir el trabajo, los objetivos, y el rol de los estudiantes. No obstante, el hecho de que ambos sean procesos grupales orientados al aprendizaje y que el aprendizaje cooperativo haya contribuido al desarrollo del ACSC (Borges & Baranauskas, 2003a; Olivares, 2007; Lehtinen *et al.*, 1999) justifica el realizar un examen a la literatura de aprendizaje cooperativo vinculada con las habilidades de e-tutores.

Características	**Aprendizaje Cooperativo**	**Aprendizaje Colaborativo**
Orientación Epistemológica	Instrucción estructurada	Construcción social
Estructura Grupal	Alta Interdependencia positiva	Baja Individualista
Rol del Docente	Administrador	Moderador/ Facilitador/Guía
Rol de los estudiantes	Cooperativo/ Agradable	Disidente/Independiente
Objetivos	Desarrollo de habilidades sociales y aprendizaje para todos los miembros	Construcción de conocimiento mediante la conversación. Preocupación por resolución de problemas
División de las Tareas	Vertical/ Independientes	Horizontal/ Vinculadas

Tabla 2.2. Comparación entre aprendizaje cooperativo y aprendizaje colaborativo

La división vertical del trabajo en sub tareas independientes implica que los resultados parciales de las sub tareas deben ensamblarse para conducir a una salida o solución final (Dillenbourg, 1999).

II.8 APRENDIZAJE COLABORATIVO SOPORTADO POR COMPUTADORA

El ACSC surge en la década de 1990 como reacción al software que forzaba a los alumnos a aprender como individuos aislados (Olivares, 2007; Lehtinen *et al.*, 1999). Dos corrientes de investigación que hicieron grandes contribuciones para el desarrollo de las ideas de ACSC fueron el Trabajo Cooperativo/Colaborativo Soportado por Computadora (TCSC) y el aprendizaje cooperativo (Borges & Baranauskas, 2003a; Olivares, 2007; Lehtinen *et al.*, 1999). En lo que respecta al aprendizaje cooperativo, las investigaciones estuvieron encaminadas a explicar los beneficios del aprendizaje cooperativo en relación al aprendizaje tradicional centrado en el profesor. Por su parte, las investigaciones en TCSC han revelado muchas cuestiones sobre la naturaleza cooperativa del trabajo en un contexto computarizado, y de hecho muchas de las ideas y herramientas usadas en ACSC fueron creadas en el ámbito del TCSC (Lehtinen *et al.*, 1999).

El campo del ACSC es una especie de sombrilla bajo la que se reúne a personas de diferentes disciplinas y antecedentes (Borges & Baranauskas, 2003b), por lo tanto para evitar confusiones conviene mencionar algunas definiciones de ACSC:

- Campo de estudio que centralmente tiene que ver con el significado y las prácticas de construcción de significado en el contexto de la actividad conjunta, y las formas en las cuales estas prácticas son mediadas a través de artefactos diseñados (Olivares, 2007).
- Se refiere al estudio y propuestas sobre las formas de diseñar tecnología para soportar procesos de aprendizaje promovidos a través de los esfuerzos colaborativos de estudiantes trabajando en una situación dada (Borges & Baranauskas, 2003a).
- Uso educacional de la tecnología de computadoras para facilitar el aprendizaje grupal. El ACSC es la unión de la teoría e investigación en aprendizaje colaborativo basado en clase y comunicación mediada por computadora (CMC) en la búsqueda de entender como los proyectos de grupo basados en CMC pueden mejorar el aprendizaje (McWhaw *et al.*, 2003).

En las definiciones precedentes si bien se menciona la tecnología, el hincapié esta puesto en el uso de la misma como soporte para la colaboración en grupos de estudiantes, y en consecuencia el campo del ACSC no puede ser restringido al simple desarrollo de productos de software. En su lugar, el énfasis debe ser puesto en el estudio de problemas y cuestiones sobre el aprendizaje en configuraciones colaborativas en las cuales, como afirman Borges & Baranauskas (2003b) la computadora tiene un rol esencial.

II.9 HABILIDADES Y COMPETENCIAS

Conviene establecer diferencias entre los términos habilidad y competencia porque en la literatura sobre *e-learning* y ACSC han sido considerados como vocablos intercambiables pero también como antónimos.

II.9.1 Habilidad

Una habilidad puede ser entendida como "una capacidad física o mental (o destreza) que tiene una persona y que permite que alguna tarea u objetivo puede ser alcanzado. Más precisamente, una habilidad (o conjunto de habilidades) es esencialmente una colección de uno o más procedimientos ejecutables que un trabajador usa para completar exitosamente una tarea o trabajo en una manera efectiva y/o eficiente, de modo tal de producir alguna salida deseada o artefacto" (Barker, 2010, p. 14).

En la Figura 2.2 se representa la relación que existe entre los objetivos perseguidos, las actividades que llevan a la consecución de los objetivos y las habilidades requeridas por las actividades. El procedimiento usual para convertir un propósito en una salida deseada consiste en: 1) identificar la actividad que necesaria para lograr el objetivo deseado; 2) identificar el conjunto de habilidades necesario para ejecutar las tareas involucradas; 3) adquirir las habilidades necesarias; 4) practicar las habilidades para lograr el nivel requerido de desempeño; 5) usar las habilidades en la actividad; y 6) crear la salida, resultado o producto.



Figura 16.2. Relación entre habilidades y actividades realizadas (Barker, 2010)

En (Barker, 2010) se clasifican a las habilidades en dos tipos básicos, las innatas y las adquiridas.

- **Habilidades innatas**: son las habilidades que posee una persona como consecuencia natural de su existencia. Ejemplos de habilidades innatas incluyen: la capacidad de observar el ambiente usando técnicas visuales, la percepción de estímulos acústicos, el sentido del tacto, y la generación de expresiones sonoras de varias clases.
- **Habilidades adquiridas**: son las habilidades desarrolladas mediante procesos de aprendizaje y entrenamiento, por lo tanto, usualmente pueden ser ajustadas y mejoradas como resultado de la práctica. Por ejemplo, una persona puede aprender a tocar la guitarra y luego practicar para lograr el nivel requerido de destreza.

Teniendo en cuenta la categorización anterior, se puede decir que las habilidades de un e-tutor de ACSC pertenecen a la categoría de habilidades adquiridas, por cuanto no son habilidades que posean todas las personas por el simple hecho de existir. De lo contrario investigadores tales como (Kukulska-Hulme2004; Williams2002; Greiffenhagen, 2012; Kukulska-Hulme, 2004; Orvis & Lassiter, 2004; Zumbach *et al.*, 2004) no hablarían de la carencia de habilidades en los docentes de ACSC, ni de la falta de conocimiento sobre la manera en la que los docentes intervienen en las actividades de aprendizaje colaborativo de los estudiantes.

II.9.2 Competencia

En general se asocia el término competencia con el desempeño que se espera tengan las personas en su ámbito laboral, como ejemplo se pueden citar las siguientes definiciones:

- Conjunto de conocimientos, habilidades y destrezas, tanto especificas como transversales, que debe reunir un titulado para satisfacer plenamente las exigencias sociales (RIACES, 2004).
- Un conjunto identificable y evaluable de conocimientos, actitudes, valores y habilidades relacionados entre sí que permiten desempeños satisfactorios en situaciones reales de trabajo, según estándares utilizados en el área ocupacional (Consejo Federal de Cultura y Educación, 1996, citado en Barchni *et al.*, 2008).
- Conjunto de conocimientos, habilidades, actitudes, comprensiones y disposiciones cognitivas, metacognitivas, socioafectivas y psicomotoras apropiadamente relacionadas entre sí para facilitar el desempeño flexible, eficaz y con sentido de una actividad o de cierto tipo de tareas en contextos nuevos y retadores (Vasco,2003, citado en Barchni *et al.*, 2008).
- Una competencia involucra un conjunto relacionado de conocimientos, habilidades y actitudes que permiten a una persona realizar efectivamente las actividades de una ocupación o función, de tal manera que satisface o excede los estándares esperados en una profesión particular o entorno de trabajo (Spector & De la teja, 2001).

En base a lo planteado en las definiciones precedentes, es evidente que los vocablos competencia y habilidad no son sinónimos, sino que por el contrario el primero abarca al segundo. Además, para que una persona sea considerada competente no basta con poseer las habilidades, conocimientos, valores y actitudes a las que hace referencia la competencia en cuestión, sino que debe ponerlos en práctica en una tarea o grupo de tareas especificas.

La Figura 2.3 ilustra la relación existente entre los términos mencionados en las definiciones de competencia antes citadas.



Figura 2.17. Relación entre competencias y habilidades (PICF, 2009)

II.10 DESCUBRIMIENTO DE CONOCIMIENTO EN BASES DE DATOS

El proceso de descubrimiento de conocimiento en bases de datos (KDD) abarca todos los pasos necesarios para obtener conocimiento útil a partir de datos y se lo puede definir como el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y últimamente entendibles en los datos (Fayyad *et al.*, 1996a). De acuerdo a los autores de esta definición, el término "patrón" debe ser entendido como una expresión en algún lenguaje que describe un subconjunto de datos o un modelo aplicable al subconjunto. Así entonces, la extracción del patrón hace referencia a ajustar un modelo a los datos, encontrar una estructura a partir de los datos, o en términos generales elaborar una descripción de alto nivel de un conjunto de datos. Por su parte, con "no trivial" se hace referencia a que está involucrado algún proceso de búsqueda o inferencia, y no se trata del simple cálculo de una cantidad predefinida como el promedio de un conjunto de números.

Tal como se menciona en (Marbán *et al.*, 2009) se han planteado varios modelos de proceso de KDD, como ser: 5 A's, Human-Centerek, SEMMA, Two Crows, KDD Roadmap, DMIE, RAMSYS, y CRISP-DM. Su importancia radica en que especifican las entradas y salidas de cada una de las tareas requeridas a lo largo de un proyecto de KDD (desde la determinación de objetivos hasta la entrega y mantenimiento del conocimiento descubierto) y permiten que este proceso de lleve a cabo de una manera repetible, medible,

y controlable. Los modelos de proceso existentes son bastante similares aunque existen ciertas diferencias en cuanto a la terminología, orientación hacia un contexto industrial o académico, y énfasis en los diferentes pasos. No obstante, es posible plantear un modelo genérico que consiste de las 8 etapas mostradas en la Figura 2.4, las cuales se detallan a continuación (Fayyad *et al*., 1996a; Fayyad *et al*., 1996b; Raynor, 1999).

- **Aprender el dominio de aplicación:** consiste en desarrollar un entendimiento del dominio de aplicación y el conocimiento previo relevante, e identificar los objetivos desde el punto de vista del cliente.
- **Creación del conjunto de datos objetivo**: consiste en seleccionar un conjunto de datos, o centrase en un subconjunto de variables o una muestra de datos, en donde se realizará el descubrimiento.
- **Limpieza de datos y pre-procesamiento**: operaciones básicas incluyen la remoción del ruido, recolectar la información necesaria para modelar o representar el ruido, decidir las estrategias para manejar campos de datos faltantes, y contabilidad de información de secuencias temporales y cambios conocidos.
- **Reducción y proyección de datos**: consiste en encontrar características útiles para representar los datos dependiendo del objetivo de la tarea. Con métodos de transformación o reducción de la dimensión el número de variables bajo consideración puede reducirse.
- **Seleccionar la función de minería de datos**: se busca hacer coincidir los objetivos del proceso de KDD con un método de minería de datos particular (clasificación, *clustering*, regresión, etc.).
- **Seleccionar el algoritmo o algoritmos de minería de datos**: se selecciona el o los algoritmos de minería de datos y el o los métodos a ser usados para buscar patrones en los datos. Este proceso incluye decidir que modelos y parámetros parecen ser apropiados y ajustar un método de minería de datos particular con el criterio general del proceso de KDD (por ejemplo el usuario puede estar más interesado en la comprensión del modelo que en sus capacidades predictivas).
- **Minería de datos**: consiste en buscar patrones de interés en una forma representacional particular o conjunto de representaciones como ser reglas y árboles de clasificación, regresión y *clustering.*
- **Interpretación**: abarca la interpretación de los patrones minados, y posiblemente regresar a cualquiera de los pasos anteriores para una iteración posterior, así como

también visualización de los patrones extraídos, remoción de los patrones redundantes o irrelevantes, y traducir los patrones útiles en términos entendibles por el usuario.

- **Uso del conocimiento descubierto**: se usa el conocimiento descubierto incorporándolo en un sistema para una acción posterior, o simplemente documentándolo y reportándolo a las partes interesadas.

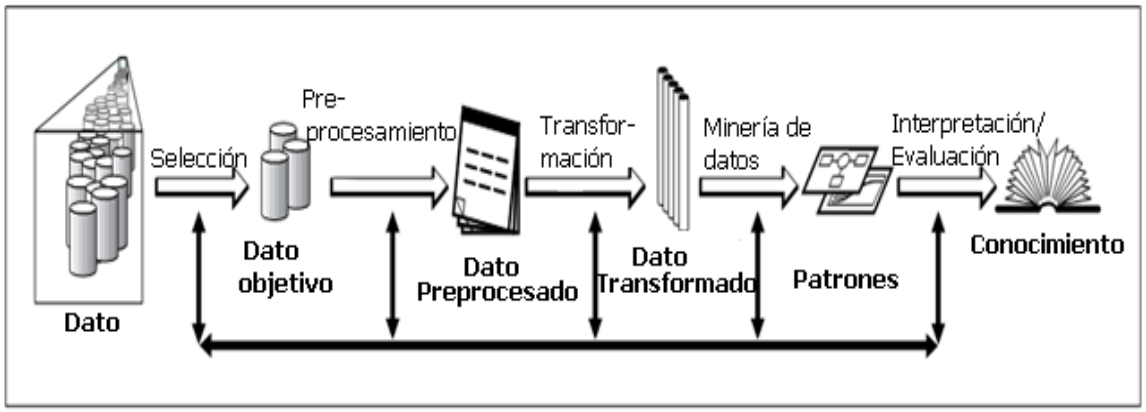


Figura 2.18. Resumen de los pasos de un proceso de KDD (Fayyad *et al*., 1996b)

Si se observa la Figura 2.4 es evidente que el KDD es un proceso iterativo y que la minería de datos es solo una de las múltiples etapas involucradas.

II.11 MINERÍA DE DATOS

La minería de datos tiene como objetivo resolver el problema de tratar con el flujo creciente de datos que es recolectado continuamente en bases de datos electrónicas puesto que, los seres humanos tenemos capacidades analíticas limitadas para procesar conjuntos de datos grandes y complejos. La expresión minería de datos o *Data Mining* se usa en los campos de estadística y descubrimiento de conocimiento para describir la aplicación de procedimientos automáticos o semiautomáticos sobre grandes cantidades de datos con el propósito de descubrir patrones o reglas significativas y previamente no conocidas (Raynor, 1999). Algunas definiciones, entre las múltiples existentes, que resultan más significativas son:

- Minería de datos es el análisis de (frecuentemente grandes) conjuntos de datos observacionales para encontrar relaciones no sospechadas y resumir los datos en formas novedosas que son tanto comprensibles y útiles para el propietario de los datos (Hand *et al*., 2001, citado en Athaudal *et al.*, 2009).

- La minería de datos es un campo interdisciplinario que reúne técnicas desde aprendizaje de máquina, reconocimiento de patrones, estadística, base de datos y visualización para tratar la extracción de información desde grandes bases de datos (Cabena *et al*., 1998, citado en Athaudal *et al.*, 2009).
- La minería de datos consiste en la extracción de patrones interesantes (no triviales, previamente desconocidos, y potencialmente útiles) o conocimiento desde una gran cantidad de datos (Han & Kamber, 2001, citado en Athaudal *et al.*, 2009).

En una economía orientada a los servicios, centrada en los consumidores, y altamente competitiva la minería de datos se convierte en una herramienta clave para el crecimiento de las empresas ya que permite lograr ventajas competitivas a partir del aprovechamiento de los datos sobre los consumidores, competencia, entidades de regulación, el gobierno, etc. (Witten *et al.*, 2005).

II.12 APRENDIZAJE DE MÁQUINA

Antes de dar una definición de aprendizaje de máquina se debería tener en claro qué se entiende por aprendizaje en el dominio de la Inteligencia Artificial. Para Mitchell (1997) el término aprendizaje se puede definir de la siguiente manera: se dice que un programa de computadora aprende desde la experiencia *E* con respecto a alguna clase de tareas *T* y medida de performance *P*, si su performance en las tareas *T*, medida en base a *P*, mejora con la experiencia *E*.

Para ilustrar la definición se puede considerar un programa que aprende a jugar a las damas que podría mejorar su desempeño, medido por su habilidad de ganar en la clase de tareas que involucran jugar a las damas, a través de la experiencia obtenida al jugar contra sí mismo. Considerando los elementos que aparecen en la definición de aprendizaje, para tener bien especificado un problema de aprendizaje se debe identificar tres características: la clase de las tareas, la medida de desempeño a mejorar, y la fuente de experiencia.

Partiendo de la definición de Mitchell, se puede decir que el aprendizaje de máquina comprende mecanismos adaptables que permiten a las computadoras aprender desde la experiencia, aprender en base a ejemplos y aprender por analogía (Negnevitsky, 2005). Puesto que la capacidad de aprendizaje puede mejorar el desempeño de un sistema

a lo largo del tiempo, los mecanismos de aprendizaje de máquina son la base para los sistemas adaptables. En la definición de Negnevitsky el término "experiencia" debe ser entendido como refiriéndose a información sobre el pasado disponible para el algoritmo de aprendizaje, la cual típicamente toma la forma de datos electrónicos recolectados y disponibles para su análisis (Mohri *et al*., 2012). El principal objetivo práctico del aprendizaje de máquina consiste en generar predicciones precisas para datos no procesados y diseñar algoritmos tanto eficientes como robustos para producir esas predicciones (Mohri *et al*., 2012).

II.12.1 Aprendizaje Supervisado y no Supervisado

Teniendo en cuenta el tipo de instancias o ítems de datos (ejemplos) usados para el aprendizaje se puede distinguir entre aprendizaje supervisado y no supervisado. Bajo el aprendizaje supervisado el algoritmo aprende a partir de ejemplos pre clasificados o etiquetados a mano y hace predicciones para datos nunca antes procesados (Coppin, 2004). Este tipo de aprendizaje se relaciona con los problemas de clasificación, regresión, y ranking (Mohri *et al*., 2012). Bajo aprendizaje no supervisado el algoritmo aprende sin necesidad de intervención humana ya que a partir de ejemplos no clasificados o sin etiquetas puede hacer predicciones para datos nunca antes procesados (Mohri *et al*., 2012; Coppin, 2004). Un buen ejemplo de aprendizaje no supervisado son los mapas de Kohonen. Un mapa de Kohonen es una red neuronal con la capacidad de clasificar un conjunto de datos de entrada sin que se le indique las categorías de los datos y sin un conjunto de datos para el entrenamiento (Coppin, 2004). El *clustering* es también un ejemplo de aprendizaje no supervisado (Mohri *et al*., 2012).

II.12.2 Conjuntos de Entrenamiento y de Prueba

Tal como se comentó anteriormente los algoritmos de aprendizaje de máquina aprenden a realizar una tarea o a resolver un problema a partir de datos sobre experiencias pasadas. Comúnmente se denomina ejemplos a los ítems o instancias de datos usados para el aprendizaje y la evaluación de la efectividad del aprendizaje. Cada uno de los ejemplos se representa mediante un conjunto de características o atributos. A su vez, a cada atributo se le asocia una o más etiquetas, es decir valores o categorías. Para ilustrar los conceptos anteriores se puede considerar el problema de clasificar automáticamente e-mails como spam y no spam:

- Ejemplos: los mensajes de e-mail que se usarán para el aprendizaje y la evaluación de la efectividad del aprendizaje.
- Características: pueden ser de relevancia la longitud del mensaje, el nombre del emisor, la presencia de determinadas palabras clave, rasgos del encabezado, etc.
- Etiquetas: las categorías spam y no spam asignadas a cada uno de los mensajes de e-mail.

Una vez finalizada la etapa de entrenamiento es deseable poder evaluar su efectividad. En este caso, antes de aplicar los algoritmos de aprendizaje el conjunto de datos tiene que dividirse en dos conjuntos o muestras, no necesariamente del mismo tamaño:

- **Conjunto de entrenamiento**: son los ejemplos usados para entrenar el algoritmo (Mohri *et al.,* 2012). Es decir, el algoritmo aprende a realizar la tarea o resolver un problema teniendo en cuenta las características de las instancias de datos pertenecientes a este conjunto. Para el problema de clasificación de e-mail este conjunto estaría constituido por mensajes de e-mail con su respectiva etiqueta.
- **Conjunto de prueba**: son los ejemplos usados para evaluar el desempeño del algoritmo de aprendizaje (Mohri *et al*., 2012). En el problema de detección de spam el conjunto de prueba consiste de una colección de ejemplos para los cuales el algoritmo debe predecir las etiquetas en base al valor de sus características. La comparación entre las predicciones y las etiquetas manualmente asignadas al conjunto de prueba permite medir el desempeño del algoritmo.

Las instancias de datos pertenecientes al conjunto de prueba de ninguna manera pueden participar en la fase de entrenamiento puesto que, si esta condición no se cumple los resultados experimentales obtenidos es probable que sean irrealmente buenos, y la evaluación no tendría carácter científico (Sebastiani, 2002).

Los algoritmos de aprendizaje comúnmente tienen uno o más parámetros que deben ser optimizados en la etapa de entrenamiento y, los ejemplos usados para seleccionar el valor apropiado de tales parámetros forman parte del conjunto de validación (Mohri *et al.*, 2012). Así como no se prueba un clasificador con los documentos usados para su entrenamiento, tampoco es correcto probarlo con los documentos usados en la optimización: el conjunto de validación y el conjunto de prueba deben mantenerse

separados (Sebastiani, 2002). De esta manera el conjunto de datos inicial tendría que dividirse en tres sub conjuntos: entrenamiento, validación, y prueba.

II.13 PROBLEMAS LINEALMENTE SEPARABLES

Algunos algoritmos de aprendizaje de máquina requieren que el conjunto de datos sea linealmente separable y como consecuencia conviene dejar en claro que significa esta expresión.

En geometría un hiperplano de un espacio de *n* dimensiones es un sub conjunto plano de *n*-1 dimensiones que separa el espacio en dos mitades. Se dice que dos puntos son linealmente separables en un espacio de *n* dimensiones si pueden ser separados por un hiperplano. En términos matemáticos lo anterior se puede expresar como:

Sean X_0 y X_1 dos conjuntos de puntos en un espacio de *n* dimensiones. Se dice que X_0 y X_1 son linealmente separables si existen *n+1* números reales $w_1, w_2, \ldots, w_n, K$ tales que para cada punto $x \in X_0$ se cumple que $\sum_{i=1}^{n} x_i w_i \geq k$ y para cada punto $x \in X_1$ se cumple que $\sum_{i=1}^{n} x_i w_i < k$ donde *xi* es el i-esimo componente de *x* (Gruzling, 2006).

En la definición anterior al producto $x_i w_i$ se lo conoce como combinación lineal. En el caso de un espacio bi-dimensional se dice que dos puntos son linealmente separables si pueden ser completamente separados mediante un línea recta.

Para visualizar las diferencias entre un problema o función linealmente separable y uno linealmente no separable se puede considerar el planteo de Coppin (2004) sobre las operaciones lógicas disyunción (OR) y disyunción exclusiva (OR exclusivo). En la Tabla 2.2 se muestra la combinación de valores de verdad para las dos variables, x1 y x2, y los resultados de las operaciones mencionadas.

Variables de Entrada		**OR**	**OR Exclusivo**
x1	x2	**x1 ∪ x2**	**x1 ⊕ x2**
0	0	0	0
0	1	1	1
1	0	1	1
1	1	1	0

Tabla 2.2. Tabla de verdad paras las operaciones OR y OR exclusivo

En la Figura 2.5 se representan las operaciones OR y OR exclusivo en un espacio bi dimensional en base a los valores de sus entradas. Los puntos en el espacio de entrada donde el resultado de aplicar la operación es 1 se indican mediante puntos negros, y los

puntos donde el resultado de aplicar la operación es 0 se indican mediante puntos blancos. En la Figura 2.5 (a) se puede trazar una recta de modo que todos los puntos negros queden a un lado y los puntos blancos queden al otro lado. En cambio, con la operación OR exclusivo no se puede trazar la recta puesto que es linealmente no separable.

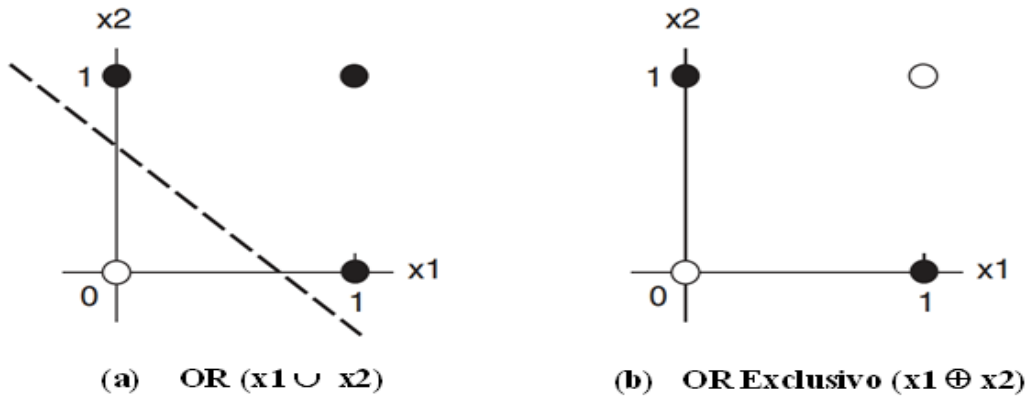


Figura 2.19. Representación bi dimensional de las operaciones OR y OR exclusivo

II.14 REDES NEURONALES

II.14.1 Neuronas Biológicas

El cerebro contiene una red densamente interconectada de células nerviosas, o unidades básicas de procesamiento de información, llamadas neuronas. Aunque una neurona es una estructura muy simple, usando simultáneamente múltiples neuronas el cerebro puede procesar información más rápido que las computadoras más veloces existentes en la actualidad (Negnevitsky, 2005).

Una neurona consiste de un cuerpo celular (soma), un número de fibras llamadas dendritas, y una fibra larga llamada axón (Negnevitsky, 2005). Una neurona recibe entradas desde otras neuronas por medio de sus dendritas, y cuando la señal de entrada supera un determinado umbral la neurona se "dispara". Más precisamente, ocurre una reacción química que causa un pulso eléctrico (conocido como acción potencial) que es enviado por medio de su axón (la salida de la neurona) hacia las sinapsis que conectan la neurona a las dendritas de otras neuronas (Coppin, 2004). La Figura 2.6 muestra gráficamente la estructura de una red de neuronas biológicas.

El cerebro humano tiene una propiedad conocida como plasticidad, lo cual significa que las neuronas pueden cambiar la naturaleza y el número de sus conexiones con otras neuronas en respuesta a los eventos que ocurren (Coppin, 2004). Debido a esta plasticidad, las conexiones entre las neuronas que conducen a la respuesta correcta son fortalecidas

mientras que las conexiones que conduce a la respuesta errónea son debilitadas. Como consecuencia las redes neuronales tienen la habilidad de aprender a partir de la experiencia. La facilidad y naturalidad con la que pueden aprender las redes neuronales biológicas es el motivo por el cual se intenta emular su comportamiento en las computadoras mediante las redes neuronales artificiales (Negnevitsky, 2005).

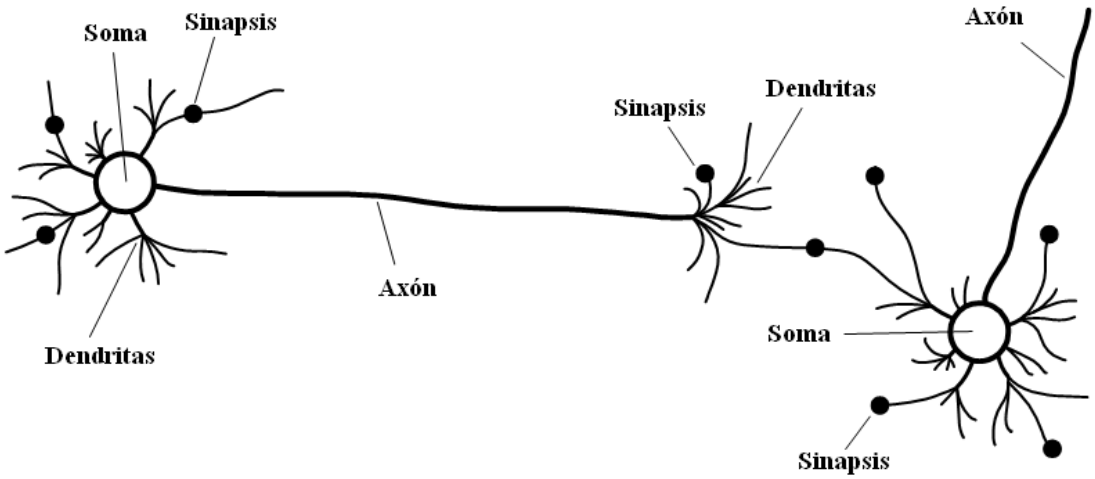


Figura 2.20. Estructura de una red neuronal Biológica (Negnevitsky, 2005)

II.14.2 Neuronas Artificiales

Las redes neuronales artificiales se inspiran en el cerebro humano y consisten de procesadores simples altamente interconectados llamados neuronas artificiales. Las neuronas en una red neuronal artificial (RNA) tienen menor cantidad de conexiones que las neuronas biológicas, y las redes neuronales son significativamente más pequeñas en comparación al número de neuronas en el cerebro humano (Coppin, 2004).

En una RNA *"las neuronas se conectan mediante vínculos con peso que llevan las señalas desde una neurona a otra. Cada neurona recibe varias señales de entrada a través de sus conexiones, sin embargo, nunca produce más de una señal de salida simple. La señal de salida se transmite por medio de la conexión de salida de la neurona (el equivalente al axón en la neurona biológica). La conexión de salida a su vez se divide en una serie de brazos que transmiten la misma señal (la señal no se divide entre los brazos). Los brazos de salida de cada neurona terminan en las conexiones de entrada de las otras neuronas en la red"* (Negnevitsky, 2005, p. 167). La Figura 2.7 representa las conexiones en una RNA, en tanto que en la Tabla 2.3 se muestra la analogía entre las RNAs y las biológicas.

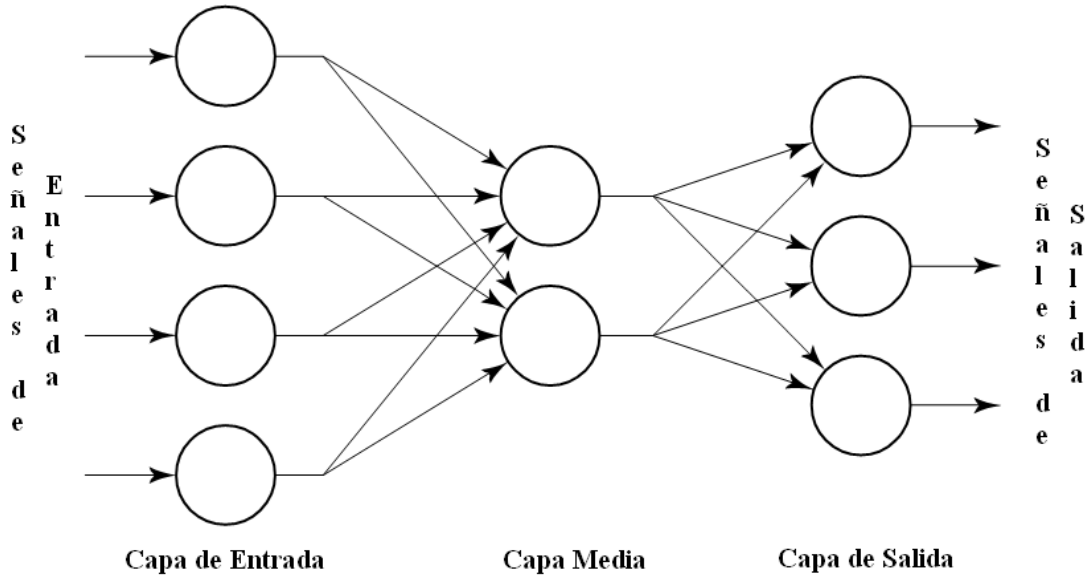


Figura 2.21. Estructura de un red neuronal artificial típica (Negnevitsky, 2005)

Red Neuronal Biológica	Red Neuronal Artificial
Soma	Neurona
Dendrita	Entrada
Axon	Salida
Sinapis	Peso

Tabla 2.3. Analogía entre red neuronal biológica y red neuronal artificial

II.14.3 Aprendizaje en Redes Neuronales Artificiales

Como se puede ver en la Figura 2.7 una red neuronal artificial típica se compone de una jerarquía de capas, y las neuronas están organizadas a lo largo de estas capas. Las neuronas se conectan a su ambiente externo mediante las capas de entrada y de salida.

Las conexiones entre las neuronas se materializan mediante enlaces y cada enlace tiene asociado un peso numérico. Estos pesos indican la fortaleza o importancia de cada neurona y son la base de la memoria a largo plazo de las RNAs (Negnevitsky, 2005). Una red aprende ajustando repetidamente el valor de los pesos. Este proceso de ajuste tiene por objetivo que el comportamiento de entrada/salida de la red este en línea con el ambiente (Negnevitsky, 2005).

Cada neurona funciona como una unidad de procesamiento básica que recibe varias señales desde sus conexiones de entrada, computa un nuevo nivel de activación y lo envía como una señal de salida mediante sus conexiones de salida (Negnevitsky, 2005). La señal de entrada a una neurona puede estar compuesta de datos no procesados o ser las salidas de

otras neuronas. A su vez, la señal de salida puede ser la solución para el problema o la entrada para otras neuronas de la red.

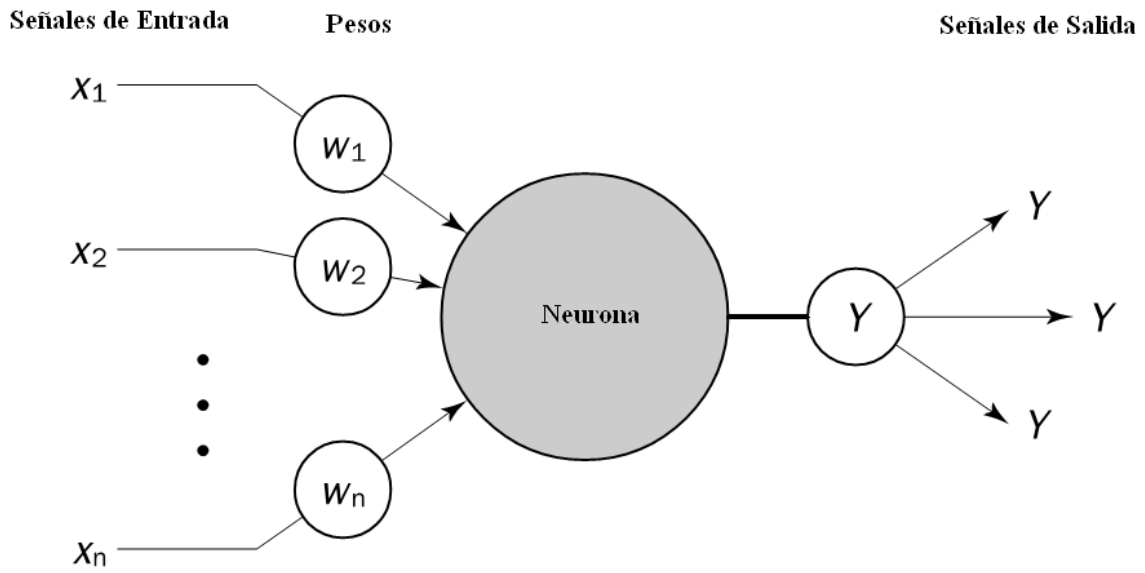


Figura 2.22. Diagrama de una neurona (Negnevitsky, 2005)

De acuerdo con Negnevitsky (2005) para determinar el valor de la señal de salida de cada neurona se procede de la siguiente manera: se calcula la suma ponderada de la señales de entrada y se la compara con un valor umbral θ. Si la entrada de la red es menor que el umbral, la salida de la neurona es -1. Pero si la entrada en mayor o igual al umbral, la neurona se activa y la salida es igual a +1. Es decir que las neuronas usan función de activación o transferencia indicada a continuación:

$$X = \sum_{i=1}^{n} x_i w_i$$

$$Y = \begin{cases} +1 & si\ X \geq \theta \\ -1 & si\ X < \theta \end{cases}$$

En la expresión anterior X es la entrada ponderada de la neurona, x_i representa el valor de la entrada *i*, w_i es el peso para la entrada *i*, *n* es el número de neuronas de entrada y, la Y corresponde a la salida de la neurona. Este tipo de función de activación se conoce comúnmente como función *sign*. En términos de la función *sign* la salida de la neurona puede expresarse como sigue:

$$Y = sign\left[\sum_{i=1}^{n} x_i w_i - \theta\right]$$

Además de la función de activación *sign* las funciones *step*, lineal, y sigmoide sea han utilizado comúnmente dado que tienen numerosas aplicaciones prácticas. La representación grafica de las cuatro funciones de activación mencionadas se muestra en la Figura 2.9. En esta figura el eje *x* de cada gráfico representa el valor de entrada de la neurona, y el eje *y* representa la salida o nivel de activación.

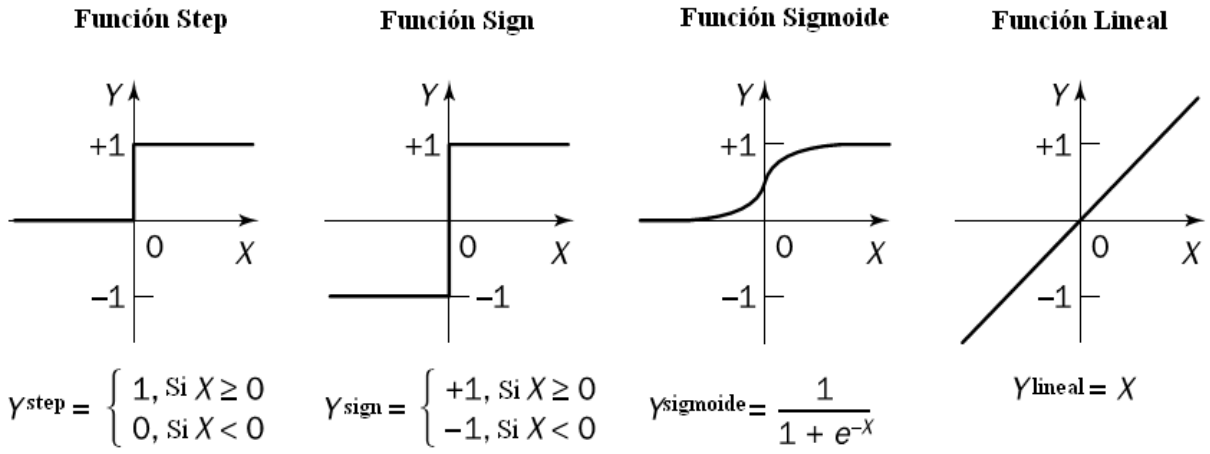


Figura 2.23. Funciones de activación de una neurona (Negnevitsky, 2005)

Las funciones de activación *sign* y *step* se usan por lo general en tareas de clasificación y reconocimiento de patrones. Por su parte, la función lineal provee una salida igual al peso de la entrada de la neurona y se usa para aproximación lineal. Por último, la función sigmoide tiene la capacidad de transformar valores de entrada (comprendidos entre menos y más infinito) en un valor en el rango 0 y 1.

II.14.4 El Perceptrón

El perceptrón es la forma más simple de red neuronal. Consiste de una neurona con pesos sinápticos ajustables. Un perceptrón de una capa simple con dos entradas se muestra en la Figura 2.10.

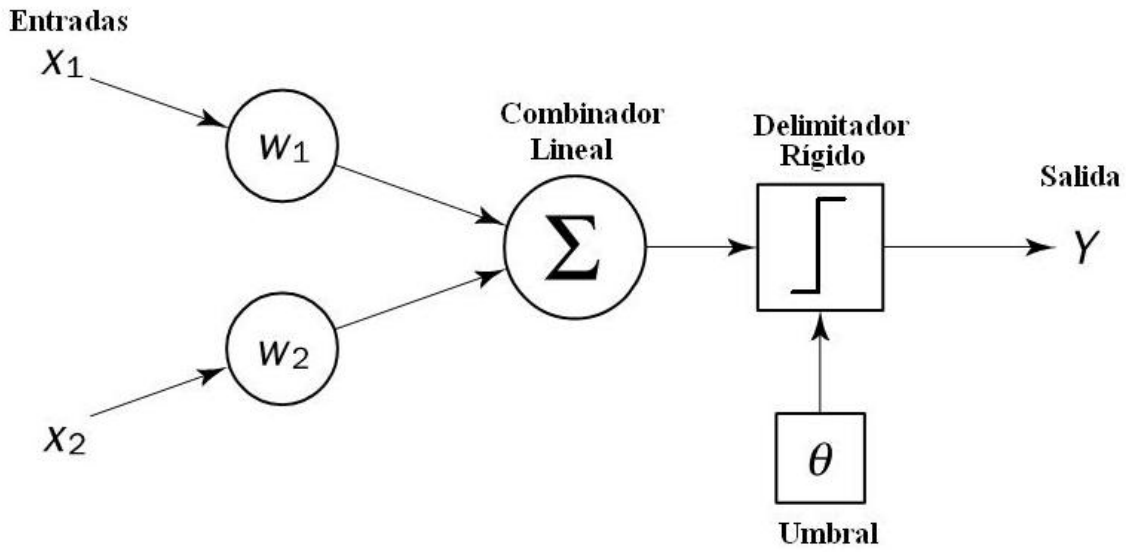


Figura 2.24. Perceptrón de capa simple con dos entradas (Negnevitsky, 2005)

En un perceptrón simple, como el de la Figura 2.10 "la suma ponderada de las entradas se aplica a un delimitador rígido, el cual produce una salida igual a +1 si la entrada es positiva y -1 si la entrada es negativa. El propósito del perceptrón es clasificar entradas, o en otras palabras estímulos $x_1, x_2, \ldots, x_n$ aplicados externamente en una de dos clases, A1 y A2." (Negnevitsky, 2005, p. 170). Así entonces, para el perceptrón elemental el espacio n-dimensional de entradas es dividido por un hiperplano en dos regiones de decisión.

Para ilustrar las ideas anteriores se puede considerar el funcionamiento del algoritmo perceptrón cuando se tienen dos y tres entradas. Para el caso de dos entradas, x_1 y x_2, el límite de decisión toma la forma de una línea recta que se muestra en negrita en la Figura 2.11 (a). El punto 1 que cae arriba de la línea de límite pertenece a la clase A1 en tanto, que el punto 2 pertenece a la clase A2 puesto que cae debajo de la línea. El umbral θ se usa para cambiar el límite de decisión.

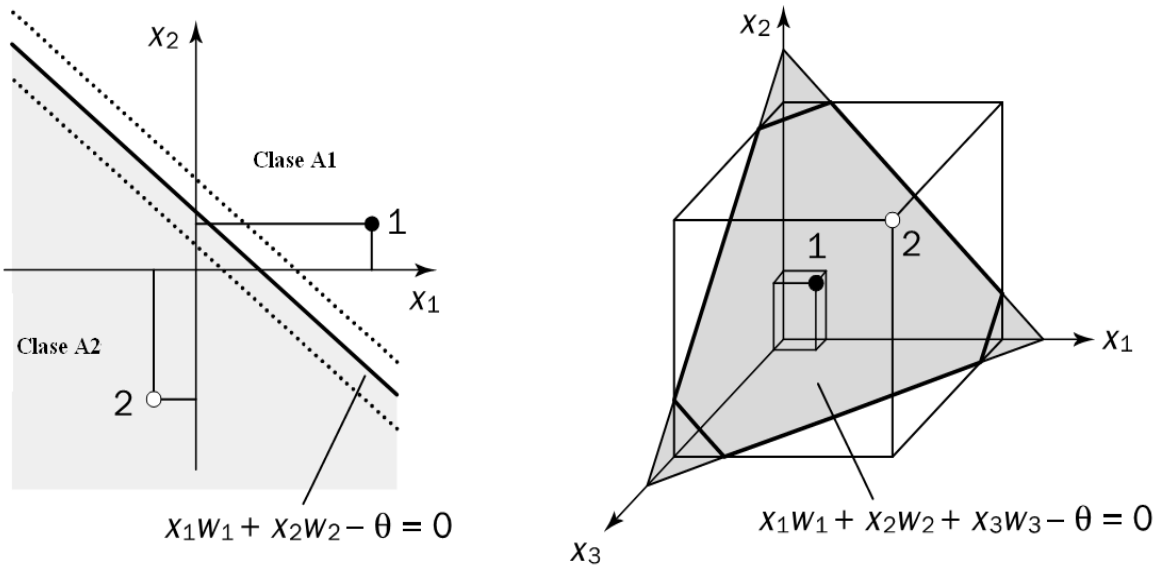


Figura 2.25. Separabilidad lineal en (a) un perceptrón de dos entradas y (b) un perceptrón de tres entradas (Negnevitsky, 2005)

Para el caso de tres variables de entrada el hiperplano aun se puede visualizar. La Figura 2.11 (b) muestra las tres dimensiones para un perceptrón de tres entradas. En este caso el plano de separación se define mediante la ecuación: $x_1w_1 + x_2w_2 + x_3w_3 - \theta = 0$.

II.14.4.1 Aprendizaje en el perceptrón

En el algoritmo perceptrón el aprendizaje de la tarea de clasificación se realiza mediante pequeños ajustes a los pesos de manera de reducir la diferencia entre la salida deseada y la obtenida con el perceptrón. Luego de asignar aleatoriamente a los pesos valores del rango [-0.5, 0.5], se aplica una serie de actualizaciones hasta que se obtiene una salida consistente con los ejemplos de entrenamiento. El proceso de ajuste de los pesos es bastante simple. Para la iteración p sea $Y_{(p)}$ la salida la salida real y sea $Y_{d(p)}$ la salida obtenida, luego el error se obtiene como:

$$e(p)=Y_d(p)-Y(p) \quad \text{donde } p=1,2,3\ldots$$

La iteración p debe ser entendida como el p-esimo ejemplo de entrenamiento presentado al algoritmo perceptrón. Si el error $e(p)$ es positivo lo que se necesita es incrementar la salida $Y_{(p)}$ del perceptrón pero , si es negativo lo que corresponde es decrementar $Y_{(p)}$. Teniendo en cuenta que cada perceptrón contribuye $x_i(p)$ x $w_i(p)$ a la entrada total X(p), si el valor de $x_i(p)$ es positivo, un incremento en su peso $w_i(p)$ tiende a incrementar la salida Y(p) del perceptrón. En cambio, si el peso $x_i(p)$ es negativo, un

incremento en $w_i(p)$ tiende a decrementar el valor Y(p). En base al planteamiento anterior, la regla de aprendizaje del perceptrón se define de la siguiente manera:

$$wi(p+1)=wi(p)+ \alpha \text{ x } xi(\text{p}) \text{ x } e(p)$$

En la regla de aprendizaje el signo α es una constante positiva menor a uno conocida como ratio de aprendizaje. Partiendo de esta regla, se deriva un algoritmo de entrenamiento del perceptrón para tareas de clasificación que se detalla en la Figura 2.12.

Paso 1: Inicialización

A lo pesos w_1, w_2, …, w_n y el umbral θ se les asigna inicialmente números aleatorios del rango [0.5,0.5]

Paso 2: Activación

Se activa el perceptrón aplicando las entradas $x_1(p)$, $x_2(p)$,…, $x_n(p)$ y la salida deseada $Y_d(p)$. Se calcula la salida para la iteración p=1

$$Y(p) = step\left[\sum_{i=1}^{n} xi(p)wi(p) - \theta\right]$$

Donde n es el número de entradas del perceptrón y *step* la función de activación aplicada.

Paso 3: Entrenamiento de los pesos

Se actualiza los pesos del perceptrón de la siguiente manera:

$$w_i(p+1)=w_i(p)+ \Delta w_i(p)$$

En la expresión anterior Δw_i(p) es la corrección del peso en la iteración p. La corrección se calcula mediante la regla delta que se define como:

$$\Delta w_i(p)= \alpha \text{ x } x_i(p) \text{ x } e(p)$$

Paso 4: Iteración

Se incrementa la iteración p en uno, se vuelve al paso 2 y se repite el proceso hasta su convergencia.

Figura 2.26. Algoritmo de entrenamiento del perceptrón (Negnevitsky, 2005)

En el algoritmo de entrenamiento del perceptrón la convergencia se da cuando todos los pesos son correctos y los errores son iguales a cero. Cada iteración del algoritmo se conoce comúnmente con el nombre de época. Una de las principales limitaciones del algoritmo perceptrón es que puede modelar solamente funciones linealmente separables.

CAPÍTULO III
MINERÍA DE TEXTOS

III.1 MINERÍA DE TEXTOS

La era de la información facilitó el almacenamiento de grandes cantidades de datos. La proliferación de documentos disponibles en la web, en *intranets* corporativas, en cables de noticias, y en cualquier lugar es abrumadora. Pero, aunque la cantidad de datos disponible crece constantemente, nuestra habilidad para procesarlos se mantiene constante. La minería de textos o *text mining* es un área de investigación que trata de resolver el problema de sobre carga de información usando técnicas tales como minería de datos, aprendizaje de máquina, procesamiento de lenguaje natural, recuperación de la información, y gestión del conocimiento (Feldman & Sanger, 2007).

De manera análoga a la minería de datos el *text mining* busca extraer información útil desde de datos mediante la identificación y exploración de patrones interesantes. Sin embargo, en la minería de textos las fuentes de datos son colecciones de documentos y los patrones no se buscan en registros de datos formalizados como sucede con la minera de datos sino en datos textuales no estructurados (Feldman & Sanger, 2007). Así entonces el *text mining*, también conocido como minería de datos de texto o descubrimiento de conocimiento desde base de datos textuales (KDT), se refiere al proceso de extracción de patrones interesantes y no triviales o conocimiento desde documentos de texto (Tan, 1999; Hotho *et al.*, 2005; César, 2007; García & Calvo, 2006). Es decir, su objetivo es descubrir cosas tales como tendencias, desviaciones y asociaciones entre "gran" cantidad de información textual (César, 2007). Las técnicas de la minería de textos intentan revelar la información oculta por medio de métodos que son capaces de hacer frente a la gran cantidad de palabras y estructuras, vaguedad, incerteza, y la borrosidad que caracterizan al lenguaje natural (Hotho *et al*., 2005). Según do Prado y Ferneda (2008) las técnicas de minería de textos se pueden organizar en cuatro categorías:

- Técnicas de clasificación: consisten en asignar objetos a clases o categorías pre definidas.
- Técnicas de Análisis de asociación: pueden ser aplicadas para ayudar a identificar palabras o conceptos que ocurren juntos y entender el contenido de un documento o colección de documentos.

- Extracción de información: son capaces de encontrar datos o expresiones relevantes dentro de documentos.
- *Clustering*: se aplica para descubrir estructuras subyacentes en un conjunto de documentos.

Teniendo en cuenta lo planteado por Aranha y Passos (2008) un proceso de minería de textos contiene 4 macro pasos: reunión, pre procesamiento, indexado, y minería.

- Reunión: el objetivo es determinar y seleccionar el dominio en el cual la técnica de minería de textos se aplicará, de modo tal de poder recolectar los datos que van a componer la base de datos textual.
- Pre procesamiento: es responsable de obtener una representación estructurada o semi estructurada del documento, generalmente en la forma de una tabla atributo-valor, que pueda ser enviada a los algoritmos de descubrimiento de conocimiento. Aunque puede ser bastante costoso, el pre procesamiento es absolutamente necesario para el éxito del proceso de minería de textos.
- Indexado: se caracteriza por el uso de métodos de recuperación de información, tales como indexado, que permiten incrementar la performance del proceso.
- Minería: tiene por finalidad encontrar patrones o información útil y previamente desconocida presente en los documentos. Para la extracción de patrones se usa técnicas similares a las del proceso de minería de datos como clasificación, *clustering*, y aprendizaje de máquina.

Un paso extra que se puede visualizar en la Figura 3.1 es el análisis realizado por un humano. Esto puede consistir en validar la información minada para construir un reporte o para navegación. En el proceso de navegación, un analista humano dirige las herramientas de minería para lograr la información deseada.

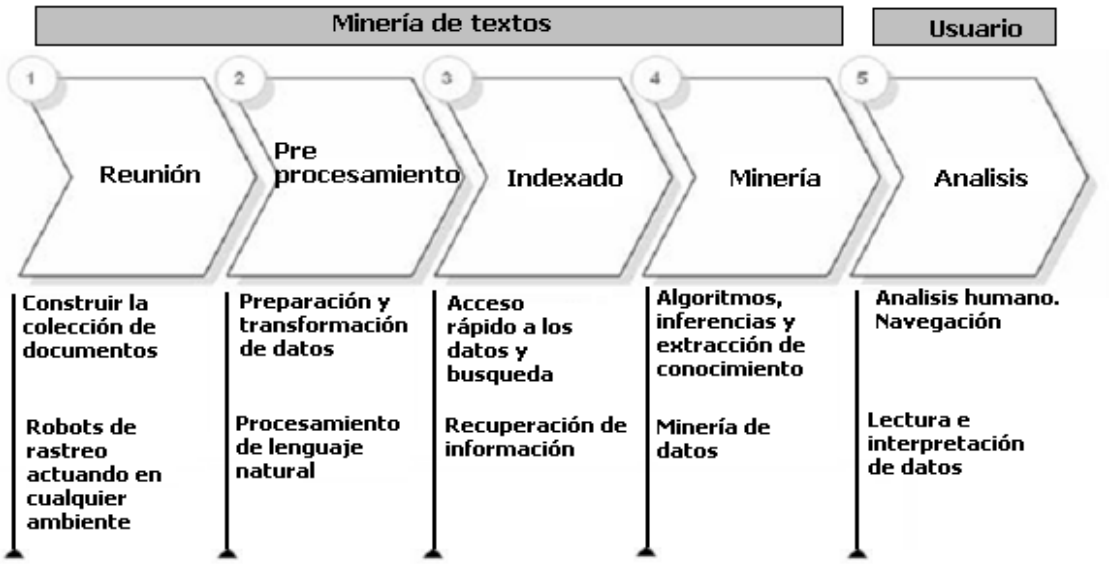


Figura 27.1 Etapas de un proceso de minería de textos (Aranha & Passos, 2008)

III.2 INFORMACIÓN ESTRUCTURADA Y DESESTRUCTURADA

Las personas tratamos con dos tipos de información: estructurada y desestructurada. Por un lado, la información desestructurada es usualmente generada por humanos para humanos, aunque las computadoras requieren algún nivel de interpretación para que la información no estructurada tenga significado (García Adeva & Calvo, 2006). Así por ejemplo, si una empresa tuviera una serie de memorándums y quisiera identificar aquellos donde se menciona una ganancia trimestral igual o superior al 25%, una consulta a un motor de búsqueda para que analice los textos desestructurados no será suficiente. Por otro lado, se tiene la información estructurada, cuyo significado está inequívocamente implicado en su formato (García Adeva & Calvo, 2006).

Las aplicaciones típicas de la minería de datos asimilan solo información estructurada que es cuidadosamente preparada antes de aplicar cualquier método de aprendizaje (Weiss *et al.*, 2005). Los ítems de datos (atributos) se describen claramente y se almacenan de manera uniforme para cada ejemplo (registro) que se considera en el proceso de aprendizaje. Generalmente se esperan dos tipos de información: numérica ordenada y categórica. Los atributos numéricos ordenados tienen valores donde las comparaciones mayor que o menor que tienen sentido. Por su parte, los atributos de categorías son códigos numéricos no ordenados cuyo significado se define en un libro de códigos no para ser usados por el programa de aprendizaje sino, para ser interpretados por las personas que analizan los resultados del aprendizaje (Weiss *et al.*, 2005).

Mientras existe una gran cantidad de herramientas de software que capturan y analizan información estructurada, en ambientes organizacionales una gran cantidad de información desestructurada permanece sin ser explorada (García Adeva & Calvo, 2006). La principal razón de esta situación es el gran desafío que supone el análisis y procesamiento de este tipo de información. Si se tiene en cuenta que la información desestructurada representa alrededor del 80% de la información disponible en una organización, es evidente la importancia que tiene el tratamiento de la información desestructurada (García Adeva & Calvo, 2006). En este escenario, conforme crece la importancia de la información no estructurada, crece la relevancia de la minería de textos debido a que cuenta con métodos capaces de revelar información oculta en documentos textuales.

III.3 OPERACIONES DE PRE PROCESAMIENTO

Los documentos de texto son por naturaleza desestructurados pero dado que los sistemas de minería de textos no aplican sus algoritmos de descubrimiento de conocimiento a colecciones de documentos no estructurados, en la minería de textos se pone considerable énfasis a lo que se conoce como operaciones de pre procesamiento (Weiss *et al*., 2005; Feldman & Sanger, 2007). Las operaciones de pre procesamiento son responsables de transformar datos no estructurados almacenados en colecciones de documentos en un formato intermedio estructurado más explícitamente (Feldman & Sanger, 2007). Estas operaciones se centran en la identificación, extracción, refinamiento y adición de características representativas de los documentos en lenguaje natural de modo tal que, las características más representativas sean usadas para el *text mining* mientras que las restantes descartadas (Feldman & Sanger, 2007). En cambio, en la minería de datos se asume que los datos han sido almacenados en un formato estructurado, por lo cual gran parte del pre procesamiento consiste en: la depuración y normalización de los datos, y la creación de tablas a partir de la unión de dos o más tablas (Feldman & Sanger, 2007).

El pre procesamiento si bien puede ser costoso es absolutamente necesario para el éxito del proceso de minería de textos (Aranha & Passos, 2008). Las operaciones de pre procesamiento incluyen una variedad de técnicas seleccionadas y adaptadas de campos como la recuperación de información, extracción de información, y lingüística computacional (Feldman & Sanger, 2007). Si bien las técnicas de pre procesamiento se pueden usar de manera aislada, los común es que se las use de manera combinada para crear las representaciones estructuradas o semi estructuradas de los documentos de textos

no procesados. A continuación se incluye una serie de operaciones de preparación de datos que han sido usadas con frecuencia en la fase de pre procesamiento de proyectos de minería de textos.

III.3.1Estandarización de los Documentos

Cuando se recolecta los documentos sobre los que se aplicará minería de textos es bastante común que estén almacenados en una variedad de formatos, dependiendo de la manera en que hayan sido generados (Weiss *et al*., 2005). Así por ejemplo, algunos documentos pueden haber sido generados por un procesador de textos que trabaja con un formato de archivo propietario, otros pueden haber sido generados usando un editor de texto simple y guardados como texto codificado en ASCII, e incluso puede haber documentos escaneados y almacenados como imágenes. Si se va usar todos los documentos, puede ser de ayuda convertirlos a un formato estándar.

La principal ventaja de la estandarización es que las herramientas de minería pueden aplicarse sin tener que considerar el tipo de cada documento (Weiss *et al*., 2005). Por lo tanto, para extraer información de un documento es irrelevante el editor que fue usado para crearlo o cual fue su formato original.

III.3.2 Identificación de Token o Tokenización

El primer paso en el pre procesamiento de textos suele ser dividir la cadena de caracteres en palabras o más exactamente en *tokens* (Weiss *et al*., 2005; Feldman & Sanger, 2007; Aranha & Passos, 2008). Como resultado de esta operación el texto se representa por una secuencia de *tokens* delimitados por caracteres primitivos tales como espacio en blanco, coma, punto, etc. (Aranha & Passos, 2008). La identificación de *token* es fundamental puesto que es difícil extraer información de alto nivel de un documento sin antes conocer los *tokens* que lo componen (Weiss *et al*., 2005). Hablando con propiedad, uno debería referirse a tipos de *token* ya que cada *token* en un documento es una instancia de un tipo de *token*. Así por ejemplo en la oración "el número de *tokens* es mayor al número de tipos de token" se tiene dos instancias del tipo de *token* "de", aunque comúnmente se acepta que se diga que hay dos instancias del *token* "de".

Es bastante usual que las herramientas de software que identifican *token* (conocidos como *tokenizer*) extraigan características superficiales de la secuencia de caracteres que componen un *token* (Feldman & Sanger, 2007). Entre estas propiedades se tiene: uso de mayúsculas, inclusión de dígitos, signos de puntuación, caracteres especiales, etc.

La identificación de un *token* es un proceso dependiente del lenguaje en el que están escritos los documentos por lo cual, puede ser necesario personalizar el *tokenizer* para obtener las mejores características posibles (Weiss *et al*., 2005).

III.3.3 Normalización

Una vez que la cadena de caracteres se dividió en *tokens* el siguiente paso puede ser convertir cada *token* a una forma estándar, un proceso usualmente denominado normalización (Weiss *et al*., 2005; Aranha & Passos, 2008). La idea central de esta operación de pre procesamiento es evitar múltiples formas de una palabra asociadas con el mismo concepto (Aranha & Passos, 2008). Así por ejemplo, "el concepto de «libro» es «una colección de objetos que consisten de números de páginas unidas entre sí » y puede ser representado por las palaras «libro» y «libros »" (Aranha & Passos, 2008, p.61). Puesto que la normalización propone que estas dos formas deben ser agrupadas en una que tiene el mismo significado, podría usarse la palabra "libro" para representar los términos "libro" y "libros". Por consiguiente, el propósito de la normalización es reducir el número de tipos de *tokens* distintos e incrementar la frecuencia de ocurrencia de algunos tipos de *token* individuales (Weiss *et al*., 2005).

Para tareas de clasificación (en particular con algoritmos que consideran la frecuencia) la normalización provee pequeños beneficios (Weiss *et al*., 2005) y en algunos casos puede llegar a dañar la efectividad (Sebastiani, 2002). Por lo tanto, la decisión de si conviene o no aplicar normalización depende del tipo de tarea o problema. Teniendo en cuenta la estrategia seguida para reducir la variación de los tipos de *token* se puede hablar de *inflectional stemming*, de s*temming to a root*, y de lematización (Aranha & Passos, 2008).

a) ***Inflectional stemming*:** Tiene en cuenta los tiempos verbales porque busca regularizar las variaciones gramaticales del tipo singular/plural y pasado/presente (Weiss *et al*., 2005). Este proceso trunca los *tokens* por lo que pueden llegar a ser difíciles de entender (Aranha & Passos, 2008). Así por ejemplo los *tokens* "regla", "reglas", y "reglamentación" serían reemplazadas por "regl".
b) ***Stemming to a root*:** Es una estrategia más agresiva que el *inflectional stemming* ya que consiste en obtener la forma raíz de cada *token* sin prefijos ni sufijos (Weiss *et al*., 2005). Con este enfoque el *token* "desmoralización" se reduciría a "moral". El *stemnnig to a root* reduce drásticamente el número de tipos de *token* por lo cual la distribución estadística se vuelve un poco más confiable (Weiss *et al*., 2005).

c) **Lematización:** Las múltiples formas de representar un *token* son reducidas a su forma primitiva (Aranha & Passos, 2008). Esto significa que los *tokens* "analizar", "analizador" y "analizado" tendrían que ser sustituidos por el lexema "análisis". Esta estrategia tiene la ventaja de capturar las intenciones de búsqueda de los usuarios puesto que luego de hacer los reemplazos el texto sigue siendo fácil de entender (Aranha & Passos, 2008).

III.3.4 Remoción de *Stop Words*

Las *stop words* (o palabras vacías) son palabras de uso frecuente (artículos, preposiciones, conjunciones, pronombres, etc.) en un lenguaje que usualmente no contribuyen a la semántica de los documentos y que no tienen ningún valor añadido (Feldman & Sanger, 2007; Weiss *et al*., 2005). Las *stop words* pueden ser descartadas dadas que ocurren con tanta frecuencia que tienen muy poco contenido de información para distinguir entre categorías y una capacidad predictiva escasa (Hotho *et al*., 2005; Weiss *et al*., 2005)

La supresión de *stop words* puede aplicarse en cualquier momento de la etapa de pre procesamiento pero suele ser más efectivo generar las características aplicando las otras operaciones de preparación de datos y luego rechazar aquellas que se correspondan con *stop words* (Weiss *et al*., 2005).

III.3.5 Determinación del Límite de las Sentencias

La determinación de límite de las sentencias es básicamente un problema de decidir cuales instancias de un signo de puntuación (punto, puntos suspensivos,!,?, etc.) marcan el fin de una oración y cuáles no (Weiss *et al*., 2005). Si bien intuitivamente se podría pensar que es suficiente con considerar como oración a todas las cadenas de caracteres encerradas entre signos de puntuación (punto, puntos suspensivos, !,?, etc.), esta aproximación no es muy efectiva ya que a lo sume puede lograrse un porcentaje de efectividad del 85% (Aranha & Passos, 2008). Por lo tanto, para mejorar la efectividad en el reconocimiento de sentencias se tiene que determinar cuáles signos de puntuación son delimitadores y cuáles no. Sin embargo, esta no es una tarea sencilla ya que signos de puntuación como el signo de punto no constituyen delimitadores cuando forman parte, por ejemplo, de direcciones URL, números decimales, direcciones IP, etc. (Aranha & Passos, 2008).

Además del problema de ambigüedad de los delimitadores, otro gran desafío es el de la ocurrencia de comillas, paréntesis, corchetes, y llaves. En el caso de las comillas,

cuando una oración contiene frases dentro de sí, los signos de puntuación dejan de funcionar como delimitadores de sentencias. Por ejemplo, los signos de punto entre comillas deben ser ignorados en la oración: "No iré contigo. Cuando llegues llámame por teléfono así me quedo tranquilo." dijo Ricardo.

III.3.6 Etiquetado de Parte del Discurso

Una vez que se divide el texto en *tokens* y sentencias el próximo paso depende de lo que se hará con el texto. Si no es necesario continuar con el análisis lingüístico, las características pueden generarse directamente desde los *tokens*. Sin embargo, si se tiene un objetivo más específico como ser el reconocer nombres de personas, lugares, y organizaciones, es conveniente realizar un análisis lingüístico adicional del texto y extraer características más sofisticadas (Weiss *et al*., 2005). Teniendo en cuenta que en todos los lenguajes las palabras se organizan en clases gramaticales o partes del discurso, el siguiente paso lógico sería determinar la parte del discurso (POS) de cada *token* (Weiss *et al*., 2005).

La operación de etiquetado de la parte del discurso o *Part-of-Speech Tagging* (POST) consiste en clasificar los palabras de acuerdo al rol que desempeñan en la sentencia en la que aparecen (Feldman & Sanger, 2007). Las etiquetas POS proveen información sobre el contenido semántico de una palabra. El conjunto de etiquetas comúnmente usado incluye artículo, sustantivo, verbo, adjetivo, preposición, número y nombre propio (Feldman & Sanger, 2007), aunque las categorías en un lenguaje no es algo intrínseco sino que depende de la manera en que un lingüista determinado analiza el lenguaje (Weiss *et al*., 2005).

Usualmente los etiquetadores POS en alguna etapa de su procesamiento realizan análisis morfológico de las palabras, por lo que una salida adicional suele ser la secuencia de *stems* o lexemas que forman parte del texto de entrada (Feldman & Sanger, 2007).

III.3.7 Discriminación del Sentido de la Palabra

La discriminación o desambiguación del sentido de una palabra consiste en determinar el significado de una palabra en un contexto determinado. Este es un problema que existe debido al fenómeno de la polisemia que caracteriza al lenguaje natural, es decir la misma secuencia de caracteres puede tener diferentes significados (Aranha & Passos, 2008). La discriminación del sentido de una palabra es el problema más sofisticado en el procesamiento de lenguaje natural y hasta el momento no existe un algoritmo que pueda desambiguar completamente un texto a pesar de la gran cantidad de esfuerzo que se ha

dedicado a esta tarea (Aranha & Passos, 2008; Weiss *et al*., 2005). Por consiguiente, a menos que un proyecto de minería de textos requiera discriminación del sentido, es mejor proceder sin esta operación (Weiss *et al*., 2005).

III.3.8 Parsing

Un *parser* o analizador sintáctico determina las relaciones entre las la palabras que forman parte de una oración y la función que desempeñan en la sentencia (sujeto, objeto, etc.) (Weiss *et al.*, 2005). Existe un gran diversidad de analizadores sintácticos (cada uno asociado con una teoría lingüística determinada) pero comúnmente vinculan cada sentencia a una estructura simple, usualmente un árbol o grafo acíclico dirigido (Weiss *et al.*, 2005). A modo de ejemplo se puede considerar lo que comúnmente se denomina *parser* de contexto libre. En este tipo de análisis sintáctico los nodos hoja del árbol son las palabras de la oración, las frases en las que se agrupan las palabras son los nodos internos, y el nodo raíz suele estar etiquetado con la letra S (Weiss *et al.*, 2005). En la Figura 3.2 se muestra el resultado del análisis sintáctico para la oración "Johnson fue reemplazado en XYZ Corp. por Smith."

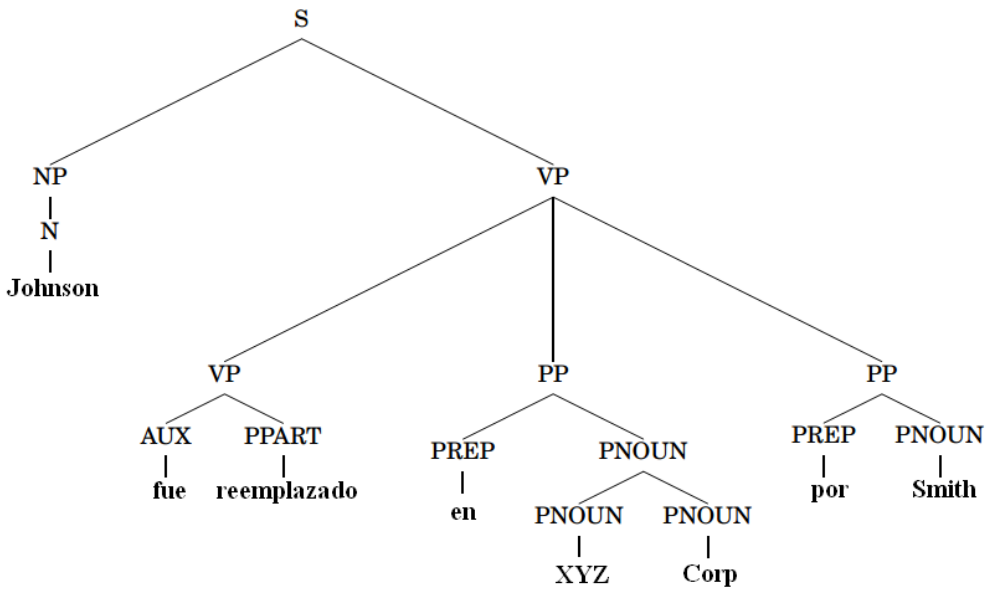


Figura 3.2 Árbol de análisis sintáctico simple (Weiss *et al.*, 2005)

III.4 REPRESENTACIÓN DE LOS DOCUMENTOS

Una vez aplicadas las operaciones de pre procesamiento para generar las características (*token*, lemma, *stem*, etc.) de relevancia para el problema entre manos se tiene que decidir como representar los documentos. Si se considera que cada característica es una dimensión en un espacio de características se pueden representar a los documentos

por medio de un vector en este espacio, esto es una secuencia de características con sus pesos correspondientes (Feldman & Sanger, 2007).

El modelo de representación más común se conoce como "bolsa de palabras" y consiste en usar todas las palabras que contiene un documento como características (Feldman & Sanger, 2007). De esta manera la dimensión del espacio de características es igual al número de palabras diferentes presentes en todos los documentos. La Figura 3.3 muestra un ejemplo del modelo de "bolsa de palabras" donde las filas representan los vectores de características de cada documento y las columnas las palabras presentes en la colección de documentos.

Documento 1

A Mario le gusta ver películas. A María le gusta también.

Documento 2

A Mario también le gusta ver partidos de fútbol.

	A	Mario	le	gusta	ver	películas	María	también	partidos	de	fútbol
Documento 1	1	1	1	1	1	1	1	1	0	0	0
Documento2	1	1	1	1	1	0	0	1	1	1	1

Figura 3.3. Modelo de "bolsa de palabras" con pesos binarios

En el modelo "bolsa de palabras" puede recurrirse a diversos métodos para asignar peso a las características. El método más simple es el binario que asigna uno a la característica si la palabra correspondiente está presente y cero de lo contrario. Existen métodos más complejos que tienen en cuenta las frecuencias de las palabras en el documento, en la categoría, y en la colección completa. Entre ellos un esquema que se usa con frecuencia es el tf-idf. Dada una palabra *w* en un documento *d*, el peso que le asigna tf-idf es igual a:

$$tf - idf(w) = tf(w) * idf(w)$$

$$idf(w) = log\left(\frac{N}{df(w)}\right)$$

De acuerdo a las ecuaciones anteriores el peso tf-idf de la palabra *w* es igual al número de veces que aparece una palabra o frecuencia del término (tf) multiplicado por un factor de escala de la importancia de la palabra (Weiss *et al.*, 2005). Este factor de escale se denomina inversa de la frecuencia del documento (idf). Para calcular idf se cuenta la cantidad de documentos que contienen la palabra *w* (df) y se divide esta cantidad por el

número total de documentos. Usando esta métrica las palabras que ocurren con frecuencia en un documento pero son relativamente infrecuentes en la colección de documentos reciben un valor de peso alto para ese documento, indicando que las palabras son muy informativas sobre los rasgos que son especiales para ese documento (Kumar *et al.*, 2007). Pero cuando una palabra aparece en muchos documentos se considera como de poca importancia y el factor de escala asume valores próximos a cero. Cuanto más grande es el valor de tf-idf mayor es la importancia de la palabra para el método de aprendizaje (Weiss *et al.*, 2005).

La metrica tf-idf puede calcularse no solo para palabras sino también para características tales como lexemes, *stem,* palabras que siempre aparecen juntas en los documentos, y los N-gramas. Un N-grama es una secuencia de términos (por lo general palabras o caracteres) de longitud N (Náther, 2005). Así por ejemplo la palabra "Texto" puede ser descompuesta en los siguientes N-gramas de caracteres:

bi-gramas:_T, Te, ex, xt, to, o_
tri-gramas:_Te, Tex, ext, xto, to_, o__
cuatri-gramas:_Tex,Text, exto, xto_,to__,o____

En los ejemplos anteriores el carácter de guión bajo representa un espacio en blanco. El agregar los espacios en blanco al comienzo y al final de las palabras ayuda con las situaciones de coincidencias al principio y al final de las palabras (Cavnar & Trenkle, 1994).

III.5 FORMALIZACIÓN DE LA TAREA DE CLASIFICACIÓN

La tarea de clasificación puede definirse formalmente como la tarea de aproximar una función de asignación de categoría desconocida F: DxC -> {0,1}, donde D es el conjunto de documentos y C es el conjunto de categorías pre definidas (Feldman & Sanger, 2007). El valor de F(*d*,*c*) es 1 si el documento *d* pertenece a la categoría *c* de otra manera el valor es 0. La función de aproximación M: DxC -> {0,1} se conoce como clasificador, modelo, o hipótesis, y el objetivo es construir un clasificador que produzca resultados tan cercanos como sea posible a la función de asignación de categoría F (Sebastiani, 2002; Feldman & Sanger, 2007).

Dependiendo de las características de F, se puede distinguir entre clasificación de etiqueta simple y clasificación de etiqueta múltiple (Feldman & Sanger, 2007). En la

clasificación de etiqueta simple cada documento pertenece exactamente a una categoría. La clasificación binaria es un caso especial de categorización de etiqueta simple donde el número de categorías es dos. El caso de etiquetas múltiples puede resolverse por medio de |C| clasificadores binarios, uno para cada una de las C categorías, siempre y cuando las decisiones de asignar los documentos a las diversas categorías sean independientes entre sí.

III.6 APROXIMACIONES A LA TAREA DE CLASIFICACIÓN

Como con cualquier otra tarea de Inteligencia Artificial hay dos aproximaciones que pueden usarse para la clasificación de textos: ingeniería de conocimiento y aprendizaje de máquina (Feldman & Sanger, 2007).

En la aproximación de ingeniera de conocimiento con la asistencia de un experto del domino (un experto en la membrecía de los documentos al conjunto de categorías seleccionadas) se definen las condiciones suficientes para que un documento sea clasificado bajo una u otra categoría (Feldman & Sanger, 2007; Sebastiani, 2002). Luego este conocimiento se codifica en un sistema ya sea manera declarativa o como reglas de un procedimiento de clasificación (Feldman & Sanger, 2007). La principal desventaja de este enfoque es lo que se conoce como cuello de botella de la adquisición de conocimiento, esto es la gran cantidad de mano de obra altamente calificada y conocimiento experto requerido para crear y mantener las reglas de clasificación (Feldman & Sanger, 2007). Esto significa que si las categorías se actualizan tanto el experto en el dominio como el ingeniero de conocimiento deben intervenir nuevamente, y si el clasificador se migra a un dominio diferente (es decir nuevas categorías) un nuevo experto en el dominio tiene que intervenir y el trabajo tiene que repetirse desde cero (Sebastiani, 2002).

A partir de la década de 1990 la aproximación de ingeniería de conocimiento ha ido perdiendo popularidad en comparación al paradigma de aprendizaje de máquina. Con esta segunda aproximación, un proceso inductivo general construye automáticamente un clasificador para la categoría *c* observando las características de un conjunto de documentos clasificados como *c* o $\bar{c}$ por un experto de dominio, y a partir de estas características, el proceso inductivo infiere las condiciones que documentos previamente no examinados deberían cumplir para ser clasificados bajo la categoría *c* (Sebastiani, 2002). En terminología de aprendizaje de máquina, el problema de clasificación es una instancia de aprendizaje supervisado puesto que el proceso es guiado por la aplicación de la función de asignación de categoría conocida en el conjunto de entrenamiento (Feldman & Sanger, 2007).

Las ventajas de usar aprendizaje de máquina son que puede lograrse una precisión comparable a la de expertos humanos, y una reducción considerable en el esfuerzo de trabajo de los expertos, puesto que no se necesita intervención de los ingenieros de conocimiento ni de los expertos de dominio para construir el clasificador o para migrarlo a un nuevo conjunto de categorías (Sebastiani, 2002). Cuando se recurre a la aproximación de aprendizaje de máquina el esfuerzo no se centra en construir el clasificador sino en obtener un constructor automático de clasificadores. Esto significa que teniendo el constructor, todo lo que se necesita es el proceso inductivo, es decir la construcción de un clasificador a partir de un conjunto de documentos clasificados a mano. Es más conveniente clasificar documentos de manera manual que construir y ajustar un conjunto de reglas, puesto que es más fácil caracterizar un concepto extensionalmente (es decir seleccionar instancias de él) que intencionalmente (es decir describir el concepto en palabras o describir el procedimiento para reconocer las instancias) (Sebastiani, 2002).

Para (Feldman & Sanger, 2007) existen cuatro cuestiones fundamentales que deben abordarse cuando se usan técnicas de aprendizaje de máquina para desarrollar un clasificador automático. Primero, se tienen que decidir las categorías usadas para catalogar las instancias. Segundo, es preciso contar con un conjunto de entrenamiento para cada categoría. Como una regla práctica, se necesitan alrededor de 30 ejemplos por cada categoría. Tercero, se deben decidir las características (palabras, *token*, lexema, etc.) que representan a las instancias. Por último, se debe seleccionar el algoritmo a usar para la clasificación.

III.7 CLASIFICACIÓN HARD VERSUS SOFT

Cuando un sistema de clasificación automatizado toma una decisión binaria para cada par documento-categoría se dice que tal sistema efectúa una clasificación *hard* (Feldman & Sanger, 2007). En cambio en una aproximación semi automatizada al problema de clasificación un experto humano es el encargado de asignar el documento a la categoría que le corresponde. La decisiones del experto humano se basan en una lista de categorías organizada por el sistema de clasificación en base a lo apropiada que se estima las categorías son para el documento (Feldman & Sanger, 2007). En este tipo de situaciones se dice que el sistema realiza una clasificación *soft* o de *ranking*. Cuando la salida de un clasificador para cada par documento-categoría no es un valor binario sino un número real comprendido en el rango [0,1] a este valor se lo denomina valor del estado de la clasificación (CSV). Este tipo de clasificadores conducen naturalmente a una

clasificación de ranking pero si se necesita una decisión binaria puede ser producida chequeando el CSV contra un umbral específico.

III.8 TECNICAS DE CONSTRUCCIÓN DE CLASIFICADORES

En esta sección se comentan algunos de los métodos que pueden utilizarse para la construcción inductiva de clasificadores.

III.8.1 Clasificadores Probabilísticos

Los clasificadores probabilísticos tienen en cuenta la probabilidad de ocurrencia de las palabras en un documento y computan el valor del estado de la clasificación (CSV) como P(*c*|*d*), esto es la probabilidad de que un documento representado por el vector $d=(w_1, w_2,\ldots,w_n)$ de términos (binarios o con peso) pertenezca a la categoría *c* (Feldman & Sanger, 2007; Sebastiani, 2002). Considerando que los documentos usualmente tienen una gran cantidad de términos, el cálculo de la probabilidad anterior es imposible (Weiss *et al.*, 2005). Así por ejemplo con 100 términos se tendrían 2^{100} combinaciones posibles. Por lo tanto, lo que se hace es estimar P(*c* | *d*) mediante el teorema de Bayes dado por:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)}$$

P(*d*) es la probabilidad de que un documento aleatoriamente seleccionado este representado por el vector *d* en tanto que, P(*c*) es la probabilidad de que un documento aleatoriamente seleccionado pertenezca a la clase *c*. P(*c*) se calcula como la cantidad de documentos pertenecientes a la categoría *c* divida entre la cantidad total de documentos. La probabilidad marginal P(*d*) no se necesita ser computada porque es constante para todas las categorías. Sin embargo, para calcular P(*d*|*c*) se supone que las coordenadas del vector *d* son estadísticas independientes una de otras (Feldman & Sanger, 2007; Sebastiani, 2002). Así entonces, P(*d*|*c*) se calcula como:

$$P(d|c) = \prod_i P(w_i|c)$$

Los clasificadores probabilísticos que se basan en este supuesto se conocen como clasificadores Naïve Bayes. El carácter "Naïve" se debe a que el supuesto de independencia no se verifica en práctica y frecuentemente es falso (Sebastiani, 2002).

El desempeño de un clasificador construido con Naïve Bayes es por lo general inferior a los que se pueden obtener con los otros métodos descriptos más adelante. Sin embargo tiene partidarios puesto que consume muy poca memoria y requiere escaso poder de cómputo. El mejor rendimiento se consigue cuando se tiene un pequeño diccionario con las palabras clave necesarias para tomar las decisiones de clasificación (Weiss *et al.*, 2005).

III.8.2 Clasificadores de Árbol de Decisión

Un clasificador de árbol de decisión es un árbol donde los nodos internos están etiquetados con características, los arcos con pruebas que se aplican al peso de las características en los documentos, y los nodos hoja con categorías (Feldman & Sanger, 2007). Para asignar la categoría a un documento se comienza por el nodo raíz y se continúa de manera descendente por los arcos cuyas condiciones son satisfechas por el documento hasta que se alcanza un nodo hoja. Al documento se le asigna la categoría que corresponde al nodo hoja.

La mayoría de los sistemas basados en árbol de decisión usan un procedimiento general para inducir el árbol tal como ID3, C4.5, y CART (Feldman & Sanger, 2007). Por lo común el árbol se construye recursivamente seleccionando una característica *f* en cada paso y dividiendo la colección de documentos de entrenamiento en dos sub colecciones (una conteniendo *f* y la otra sin *f*) hasta tener solo documentos de una categoría (Feldman & Sanger, 2007). En este punto se genera un nodo hoja y se lo etiqueta con la categoría. En cada paso la selección de la característica *f* se basa en una medida teórica de información tal como entropía o ganancia de información (Feldman & Sanger, 2007). Para evitar que el árbol generado sufra del problema de sobreajuste a la colección de documentos de entrenamiento se aplica lo que se conoce como poda, es decir se remueve las ramas del árbol que son demasiado especificas.

Los clasificadores construidos con el método de árbol de decisión por lo general tienen un desempeño inferior a los obtenidos mediante otras técnicas pero la ventaja es que pueden ser fácilmente entendidos por las personas (Sebastiani, 2002).

III.8.2.1 Algoritmo C4.5

El algoritmo C4.5 recursivamente visita cada nodo de decisión, seleccionando la división óptima, hasta que no se requieren posteriores divisiones. Para seleccionar la división óptima este algoritmo se basa en el concepto de ganancia de la información o entropía (Larose, 2005) que se explicará brevemente.

Para un evento con probabilidad *p*, la cantidad promedio de información en bits requerida para trasmitir el resultado es $-\log_2(p)$. Como ser, el resultado de arrojar una moneda con probabilidad 0.5 puede ser trasmitido usando $-\log_2(0.5)=1$ bit, esto es cero o uno dependiendo del resultado. Para una variable *x* cuyos *k* valores posibles tienen las probabilidades $p_1,p_2,\ldots, p_k$ simplemente se usa la suma ponderada de los $\log_2(p_k)$, y la formula se reescribe como:

$$H(x) = -\sum_j p_j log_2(p_j)$$

De acuerdo con Larose (2005) en C4.5 se usa el concepto de entropía de la siguiente manera: supóngase que se tiene un candidato *S* para la división, el cual divide el conjunto *T* de datos de entrenamiento en los subconjuntos $T_1,T_2,\ldots,T_k$. El requerimiento de información medio puede calcularse como la suma ponderada de las entropías de las sub conjuntos individuales, de la siguiente manera:

$$H_s(T) = \sum_{i=1}^{k} P_i H_s(T_i)$$

Donde P_i representa la proporción de registros en el subconjunto *i*. La ganancia de información se define como ganancia(S)=H(T) - H_s(T), esto es, el incremento en la información producida al dividir el conjunto *T* de datos de entrenamiento de acuerdo al candidato de división *S*. En cada nodo de decisión, C4.5 selecciona como división optima a la división que tiene la ganancia de información más grande.

III.8.3 Clasificadores de Regla de Decisión

Los clasificadores de regla de decisión son simbólicos al igual que los árboles de decisión, por lo cual son fáciles de comprender para las personas. Un clasificador de regla de decisión consiste de reglas en Forma Normal Disyuntiva (FND) que se construyen a partir de un conjunto de documentos de entrenamiento mediante un método de aprendizaje inductivo de regla (Sebastiani, 2002; Feldman & Sanger, 2007). Una formula en FND es una disyunción de clausulas conjuntivas por lo cual un documento se clasifica bajo una determinada categoría si y solo si satisface la formula, es decir si y solo si satisface al menos una de las clausulas (Sebastiani, 2002). Las reglas en FND son similares a los arboles de decisión porque codifican cualquier decisión Booleana. Sin embargo, la ventaja de las reglas en FND es que los métodos de construcción buscan generar clasificadores más compactos que los de árbol de decisión.

Usualmente los métodos de aprendizaje intentan seleccionar la mejor regla desde el conjunto de todas las reglas de cobertura posibles (es decir reglas que clasifican correctamente todos los ejemplos de entrenamiento) teniendo en cuenta algún criterio de optimización (Feldman & Sanger, 2007). Mientras en los árboles de decisión se usa una estrategia de divide y vencerás, por lo común las reglas en FND se construyen de manera ascendente. El clasificador inicial se obtiene desde el conjunto de entrenamiento viendo a cada documento *d* como una clausula del tipo:

$$d_1 \wedge d_2 \wedge \ldots \wedge d_n \rightarrow c$$

En la clausula los d_i son las características del documento en tanto que *c* su categoría. El algoritmo de aprendizaje aplica una serie de generalizaciones (por ejemplo eliminando términos desde las claúsulas y uniendo reglas) de manera de lograr las reglas más compactas posibles mientras se mantiene la propiedad de cobertura (Feldman & Sanger, 2007). Al final del proceso un paso de poda similar a los árboles de decisión se aplica de modo de sacrificar cobertura en beneficio de generalidad.

III.8.4 Clasificadores Basados en Ejemplo

Los clasificadores basados en ejemplo no construyen una representación declarativa explicita de las categorías sino que computan directamente la similitud entre el documento a ser clasificado y los documentos de entrenamiento (Feldman & Sanger, 2007). Estos métodos han sido denominados "estudiantes perezosos" porque difieren la decisión de cómo generalizar más allá de los datos de entrenamiento hasta que se requiere una nueva clasificación. El entrenamiento para estos clasificadores consiste simplemente en almacenar representaciones de los documentos de entrenamiento junto con sus categorías (Feldman & Sanger, 2007).

El ejemplo más prominente de los clasificadores basados en ejemplo es el algoritmo KNN (K-ésimo vecino más cercano). Para decidir si el documento *d* pertenece a la categoría *c* KNN chequea si los K documentos de entrenamiento más similares a *d* pertenecen a la categoría *c* (Feldman & Sanger, 2007). Si la respuesta es positiva para una proporción suficientemente grande se toma una decisión positiva, caso contrario una decisión negativa.

Una cuestión central para el algoritmo KNN es determinar cuándo los documentos son similares. Una estrategia posible consiste comparar dos documentos midiendo la distancia con una ecuación de la forma:

$$distancia(x,y) = (x_1 - y_1)^2 + (x_2 - y_2)^2 \ldots + (x_n - y_n)^2$$

Esta medida de distancia simplemente considera la diferencia entre cada atributo al cuadrada. De este modo cuanto mayor sea la distancia, menor será la conexión entre el par de documentos. Sin embargo, una medida más natural de similitud entre documentos consiste en contar la cantidad de palabras que tienen en común (Weiss *et al.*, 2005). En base a este conteo se pueden ordenar los documentos por similitud de modo que los documentos más similares tengan la mayor cantidad de palabras en común. Esta métrica difiere de las métricas tradicionales de cálculo de distancia puesto que ignora las palabras que ocurren en un documento pero no en el otro (Weiss *et al.*, 2005). En la práctica una aproximación que ha conducido a mejores resultados predictivos es la similitud del coseno, donde para calcular la distancia se tiene en cuenta tf-idf.

La construcción de un clasificador con KNN requiere que se especifique la cantidad K de documentos del conjunto de entrenamiento a tener en cuenta al momento de clasificar un nuevo documento. El determinar un valor de K para cada categoría es un proceso tedioso por lo cual se suele usar un único valor de K para casi todas las categorías, con un valor inferior a K para las categorías donde la cantidad de ejemplos sea menor a K (Weiss *et al.*, 2005). El valor de K se puede optimizar en un conjunto de validación pero algunos investigadores consideran que con K=20 o bien con valores 30<=K<=45 se obtienen los mejores resultados (Sebastiani, 2002; Feldman & Sanger, 2007). Además varios experimentos han mostrado que incrementar el valor de K no degrada el desempeño de manera significativa (Feldman & Sanger, 2007).

Los clasificadores de KNN suelen tener muy buen desempeño para tareas de clasificación de textos y son bastante robustos ya que no requieren que el conjunto de clases sea separable linealmente (Feldman & Sanger, 2007; Sebastiani, 2002). El único inconveniente de KNN es su ineficiencia en cuanto al tiempo de clasificación: mientras que por ejemplo un clasificador lineal solo tiene que calcular el producto interior para clasificar un nuevo documento, KNN requiere que el conjunto de entrenamiento completo sea jerarquizado en función a la similitud con el documento a clasificar, lo cual es mucho más costoso computacionalmente (Sebastiani, 2002).

III.8.5 Métodos Lineales de Puntuación

Para lograr una buena performance en la clasificación de documentos puede ser necesario trabajar con vectores de características de grandes dimensiones. Si bien es

probable que muchas de las características no sean relevantes, para una persona puede ser un tanto complicado decidir cuáles son de utilidad y cuáles no. Por consiguiente, sería conveniente que un algoritmo de clasificación tenga la capacidad tomar un conjunto grande de características y de seleccionar solo las que sean de utilidad. Una manera de lograr esto es usando puntuación lineal.

Para mostrar el funcionamiento de este método se puede considerar el problema de distinguir entre dos clases planteado en (Weiss, 2005). El método de puntuación asigna una puntuación positiva para predecir la clase positiva y puntuación negativa para predecir la clase negativa. En la Figura 3.4 se muestra la manera de usar los pesos para establecer la puntuación de un documento. Una vez que se conoce el peso de cada palabra en el documento se los suma para determinar la puntuación del documento.

Modelo Lineal

Palabra	**Peso**
Dividendo	0.8741
Ganancias	0.4988
Ocho	-0.0866
Extraordinario	-0.0267
Meses	-0.1801
Pago	0.6141
Creció	-0.0253
Separación	0.9050
York	-0.1908
….	

Nuevo Documento

Palabras	**Puntuación**
dividendo, pago, creció	1.4629

Figura 3.4. Cálculo de la puntuación ponderada para un documento

Matemáticamente este método es una función de puntuación lineal que se puede definir mediante la siguiente ecuación:

$$Puntuación(D) = \sum_j w_j x_j + b = w \cdot x + b$$

En la ecuación anterior *D* representa el documento, w_j es el peso de la j-ésima palabra de un diccionario de las palabras de la colección de documentos, *b* es una constante denominada sesgo, x_j es uno o cero dependiendo de la ausencia o presencia de la j-ésima palabra en el documento.

En términos geométricos los métodos de puntuación lineal producen una línea o hiperplano en el espacio de características. Aunque una línea no puede ajustarse a superficies complejas, y una forma curva puede ser necesaria, frecuentemente se puede crear características no lineales apropiadas de manera que la curva en el espacio original se situé en un hiperplano en el espacio ampliado con las características no lineales adicionales (Weiss, 2005). De esta manera la no linealidad puede capturarse explícitamente construyendo características no lineales. Con la aproximación de puntuación lineal la tarea de modelado es bastante simple porque uno debe concentrarse en crear características útiles y dejar que el algoritmo determine la manera de asignar peso a cada característica creada. Por lo tanto, la cuestión clave en los métodos de puntuación ponderada es el aprender los pesos de las características a partir de la colección de documentos ya que una vez establecidos los pesos las categorías se asignan aplicando la ecuación definida para *Puntuacion(D)* (Weiss, 2005).

Una de las maneras de determinar los pesos del vector *w* es mediante el algoritmo Máquina de Vector de Soporte (SVM) que se describe a continuación en su versión estándar y con márgenes desiguales.

III.8.5.1 Máquinas de vector de soporte

En términos geométricos un clasificador SVM binario (cuando se tienen solo dos categorías o clases) puede ser visto como un hiperplano en el espacio de características que separa los puntos que representan instancias positivas de la categoría de los puntos que representan instancias negativas (Feldman & Sanger, 2007). El hiperplano de clasificación se selecciona durante el entrenamiento como el único hiperplano que separa las instancias positivas conocidas de las instancias negativas conocidas con el máximo margen. El margen es la distancia desde el hiperplano al puto más cercano de los conjuntos positivo y negativo. Los hiperplanos de SVM son completamente determinados por un conjunto relativamente pequeño de instancias de entrenamiento conocidas como vectores de soporte.

Para definir formalmente el método SVM, siguiendo los lineamientos de Li y Shawe-taylor (2003), supóngase que se tiene un conjunto de entrenamiento Z=((x_1,y_1), (x_2, y_2), …, (x_m,y_m)), donde los x_i es un vector de características de *n* dimensiones y y_i=(+1 o bien -1) es su etiqueta. El objetivo es aprender los parámetros $w \in R^n$ y $b \in R$ de un clasificador lineal en el espacio de características, esto es:

$$h_{w,b}(x) = sign(\langle w, x \rangle + b)$$

Donde $\langle , \rangle$ representa en producto interior en R^n, es decir el producto interior en el espacio vectorial real de *n* dimensiones.

La Figura 3.5 ilustra que en el caso bi-dimensional (se consideran dos características) se pueden seleccionar varias rectas como hiperplanos de clasificación. Del conjunto de rectas paralelas el método SVM selecciona la recta resaltada en negrita puesto que separa las instancias positivas de las negativas con el máximo margen. El la Figura 5 las cruces representan las instancias positivas, los círculos corresponden a las instancias negativas, y las rectas son los hiperplanos de clasificación. En tanto que, los cuadrados señalan los vectores de soporte.

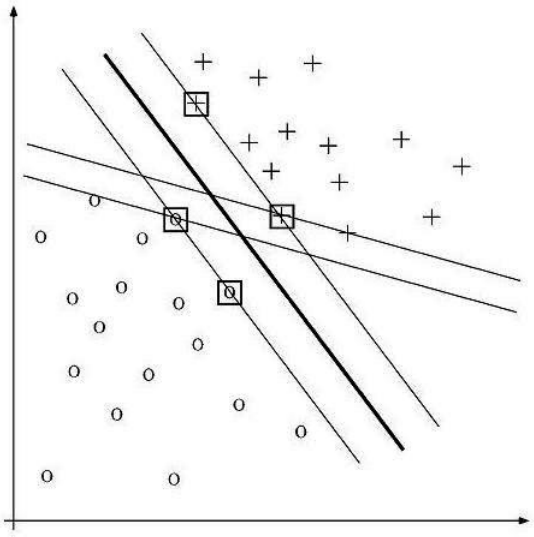


Figura 3.5. Representación grafica de clasificador SVM bi dimensional (Sebastiani2002)

El algoritmo SVM puede aplicarse con espacios de características de grandes dimensiones puesto que es bastante robusto ante el problema de sobre ajuste a los datos del conjunto de entrenamiento (Feldman & Sanger, 2007). Tampoco necesita selección de características ya que el algoritmo tiene la capacidad de seleccionar las características de utilidad (Sebastiani, 2002). Otra de las ventajas de SVM es que se puede ampliar el espacio de características original mediante características no lineales de modo tal de poder modelar las situaciones donde las instancias positivas y negativas no sean linealmente separables (Sebastiani, 2002).

III.8.5.2 Máquinas de vector de soporte con márgenes desiguales

El algoritmo SVM con márgenes desiguales fue propuesto por Li y Shawe-Taylor (2003) debido que el desempeño del algoritmo SVM para categorías pequeñas (esto es

categorías con número pequeño de documentos relevantes en la colección) es bastante pobre.

En el algoritmo SVM el margen puede ser considerado como una medida de la habilidad del clasificador para tolerar errores puesto que, un clasificar probablemente clasificará correctamente una instancia de prueba si tiene el margen más grande. Si el conjunto de entrenamiento es representativo del conjunto de datos completo, un clasificador con el margen más grande con respecto al conjunto de entrenamiento tendrá el mejor desempeño en cuanto a generalización. Sin embargo, si el conjunto de entrenamiento no es representativo el máximo margen aprendido puede llegar a tener pobre performance de generalización tal como se ilustra en la Figura 3.6.

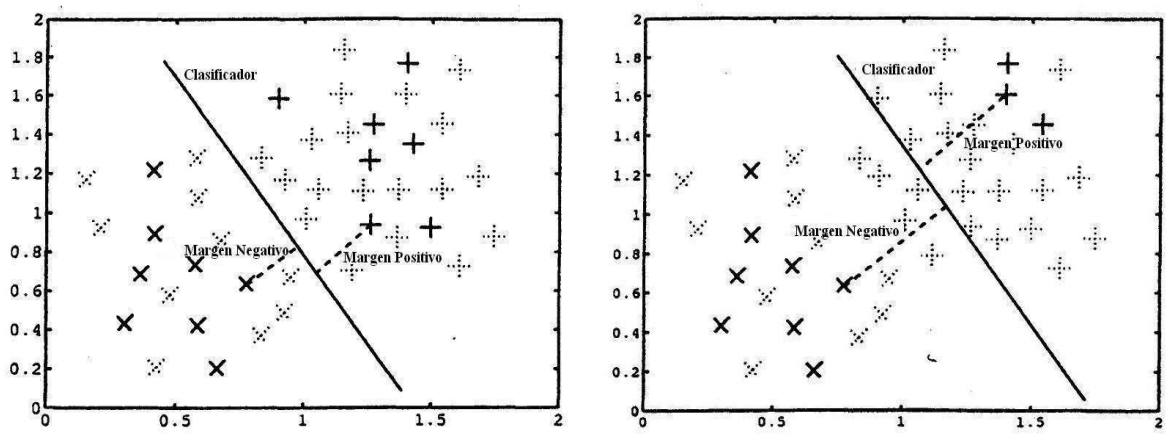


Figura 3.6 Problema de clasificación bi dimensional con sus clasificadores SVM (Li & Shawe-taylor, 2003)

En la Figura 3.6 se muestra un problema de clasificación binario con dos dimensiones junto con dos tipos de conjuntos de entrenamiento y sus respectivos clasificadores SVM. En los dos gráficos cada "+" representa una instancia positiva y cada "x" representa una instancia negativo, en tanto que los símbolos "+" y "x" con línea solida son los que forman parte del conjunto de entrenamiento.

Los ejemplos de entrenamiento del gráfico de la izquierda en la Figura 3.6 son representativos del conjunto de datos por lo cual el máximo margen aprendido tiene buena capacidad de generalización. En cambio, en el gráfico de la derecha el número de ejemplos de entrenamiento positivos es muy pequeño y los ejemplos de entrenamiento se acumulan en una pequeña región del espacio de características. En este caso el clasificador SVM con máximo margen no es una buena predicción puesto que clasifica erróneamente algunos ejemplos positivos. No obstante, si el límite de clasificación fuera alejado apropiadamente

de los ejemplos de entrenamiento positivos, el clasificador tendría una mejor performance de generalización.

Teniendo en cuenta el planteo anterior, si un clasificador basado en margen va a ser aprendido desde un conjunto de entrenamiento desbalanceado que tiene poco ejemplos positivos, convendría que el algoritmo de aprendizaje establezca un margen con respecto a los ejemplos positivos (el margen positivo) que sea más grande que el margen con respecto a los ejemplos negativos (el margen negativo). Para distinguir entre el margen positivo y el margen negativo Li y Shawe-taylor proponen introducir el parámetro τ en el algoritmo SVM estándar. Con el parámetro τ se controla el ratio del margen positivo sobre el margen negativo. De este modo el clasificador SVM puede ser encontrado resolviendo el siguiente problema de optimización:

minimizar $_{w,b,\xi}\ \langle w,w\rangle + C_\tau \sum_{i=1}^{n} {}_i$

sujeto a $\langle w,x_i\rangle + {}_i + b \geq 1 \quad si\ y_i = +1$

$\langle w,x_i\rangle - {}_i + b \leq -\tau$ si y_i=-1

${}_i \geq 0$ para i=1,…,m

La solución al problema de optimización expresado en la ecuación anterior corresponde al clasificador SVM con márgenes desiguales. No se entrará en los detalles matemáticos de la ecuación de definición del clasificador pero se debe aclarar que ξ_i corresponde a los multiplicadores de Lagrange que se utilizan para resolver el problema de optimización mediante técnicas de programación cuadrática. En tanto que factor de costo C_τ mide el costo de los ejemplos clasificados incorrectamente en el conjunto de entrenamiento.

Los experimentos realizados por Li y Shawe-taylor mostraron que efectivamente existe una relación entre el parámetro de margen τ y el número de ejemplos de entrenamiento positivos, y que cuando se tiene una cantidad pequeña de ejemplos positivos corresponde utilizar un valor pequeño del parámetro τ.

III.8.6 Redes neuronales

Un clasificador de red neuronal es una red de neuronas o unidades de procesamiento donde las unidades de entrada representan características, la o las unidades de salida representan la categoría o categorías de interés, y los pesos en los arcos que conectan las unidades representan relaciones de dependencia (Sebastiani, 2002). Para

clasificar un documento los pesos de sus características son cargados en las unidades de entrada, la activación se propaga hacia delante a lo largo de la red, y los valores finales de los unidades de salida determinan las decisiones de categorización (Feldman & Sanger, 2007).

Usualmente las redes neuronales se entrenan con propagación hacia atrás. Con esta estrategia los documentos se cargan en las unidades de entrada, si una clasificación errónea ocurre el error se propaga hacia atrás de las red modificando los pesos de los arcos para minimizar el error (Sebastiani, 2002; Feldman & Sanger, 2007).

III.8.6.1 Algoritmo PAUM

El Algoritmo Perceptrón con Márgenes Desiguales (PAUM) es una extensión del algoritmo Peceptrón, descripto en la sección II.14.4, que fue propuesto en (Li *et al.*, 2002) para ser aplicado en problemas de clasificación donde los ejemplos positivos son muy raros en relación a los ejemplos negativos.

Para comprender en funcionamiento de PAUM supóngase que se tiene una muestra de entrenamiento $z=((x_1,y_1), (x_2, y_2), \ldots, (x_m,y_m)) \in (X \text{ x } \{-1,+1\})^m$ de tamaño *m* junto con un mapeo de característica $\emptyset: X \rightarrow K \subseteq R^n$ en un espacio vectorial *K* de *n* dimensiones. El objetivo es aprender los parámetros $\mathbf{w} \in K \; y \; b \in R$ de un clasificador lineal:

$$h_{\mathbf{w},b}(x) = sign(f_{\mathbf{w},b}(x)), \qquad f_{\mathbf{w},b}(x) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

En las ecuaciones precedentes $\mathbf{x}=\emptyset(x)$ en tanto $\langle , \rangle$ representa en producto interior en *K*.

Con el algoritmo Perceptrón en cada paso se chequea si el ejemplo de entrenamiento actual $(x_i, y_i) \in z$ es correctamente clasificado por el clasificador $(y_i(\langle \mathbf{w}_t, \mathbf{x}_i \rangle + b_t) > 0)$ y se actualiza el peso del vector $\mathbf{w}_t$ de otro modo. En el paso de actualización los pesos del vector $\mathbf{w}_t$ y el termino de sesgo b_t son cambiados por $\mathbf{w}_{t+1}=\mathbf{w}_t+\eta y_i \mathbf{x}_i$ y $b_t+1=b_t+\eta$. Este algoritmo converge siempre y cuando los datos de entrenamiento sean linealmente separables en el espacio de características. La velocidad de convergencia, es decir la cantidad de actualizaciones hasta la convergencia, la controla el margen máximo de los datos de entrenamiento. El margen máximo $\gamma(\mathbf{w},b,z)$ de un clasificador $f_{\mathbf{w},b}$ es el valor real mínimo de la muestra de entrenamiento, y se lo define como:

$$\gamma(\mathbf{w}, b, z) = min_{(x_i, yi) \in z} \frac{y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b)}{\|\mathbf{w}\|}$$

Cuanto mayor sea el margen máximo $\gamma(z) = max_{\mathbf{w},b}\gamma(\mathbf{w},b,z)$para una muestra de entrenamiento particular *z* menor es la cantidad de actualizaciones que realiza Perceptrón hasta la convergencia.

Considerando que en los problemas de clasificación es mucho más importante clasificar correctamente los ejemplos positivos que clasificar correctamente los ejemplos negativos, en el algoritmo PAUM se trata a los ejemplos positivos de manera diferente a los ejemplos negativos. Este tratamiento diferencial se logra incorporando un margen positivo γ_{+1} y un margen negativo γ_{-1}. Estos márgenes se definen de la siguiente manera:

$$\gamma_{+1}(\mathbf{w},b,z) = min_{(x_i,+1)\in z}\frac{+(\langle \mathbf{w},\mathbf{x}_i\rangle + b)}{\|\mathbf{w}\|}$$

$$\gamma_{-1}(\mathbf{w},b,z) = min_{(x_i,-1)\in z}\frac{-(\langle \mathbf{w},\mathbf{x}_i\rangle + b)}{\|\mathbf{w}\|}$$

El algoritmo que se obtiene usando las formulas anteriores para el cálculo de los márgenes se denomina algoritmo PAUM y se lo detalla en la Figura 3.7.

Los resultados experimentales de Li *et al.* (2002) han demostrada que el algoritmo PAUM tiene una performance igual o mejor que el SVM estándar y que permite reducir el tiempo de entrenamiento.

```
Requiere: una muestra de entrenamiento linealmente separable
z=(x,y) ∈(X x {-1, +1})^m
Requiere: un rato de aprendizaje η ∈ R+
Requiere: dos parámetros de margen τ-1,τ+1 ∈ R+

W0=0; b0=0; t=0; R=max_{xi∈x}||xi||
repeat
for i=1 to m do
     if yi(⟨wt,xi⟩ + bt) ≤ τyi  then
          wt+1=wt+ηyixi
          bt+1=bt+ηyiR^2
          t← t+1
     end if
end for
until no se hagan más actualizaciones en el ciclo for
return (wt,bt)
```

Figura 3.7. Algoritmo PAUM (Li *et al.*, 2002)

III.9 EVALUACION DE LA EFECTIVIDAD DE LOS CLASIFICADORES

III.9.1 Métodos de Estimación de la Efectividad

Cuando se dispone de una gran cantidad de datos la estimación de la efectividad de los clasificadores en una nueva colección de datos no es tarea complicada: el clasificador se construye utilizando una muestra o conjunto de entrenamiento grande, y se lo evalúa en un conjunto de prueba grande (Witten & Frank, 2005). En general, cuanto más grande sean los conjuntos de prueba y entrenamiento, mejores clasificadores y estimaciones de su efectividad se obtienen. Ahora bien, cuando la cantidad de datos disponibles para el entrenamiento y la evaluación es limitada, se pueden emplear dos métodos para estimar la efectividad de los clasificadores: hold-out y n-fold.

Con el método hold-out se reserva una cierta cantidad de datos para la prueba y el resto se usa para el entrenamiento. Comúnmente se suele destinar un tercio de los datos para la prueba y usar los dos tercios restantes para el entrenamiento (Witten & Frank, 2005).

Podría suceder que el conjunto de entrenamiento (o de prueba) no sea representativo. Si bien, en general, no se puede decir si una muestra es representativa o no, un simple chequeo que vale la pena consiste en asegurarse que cada clase del conjunto de datos este representada en aproximadamente la misma proporción en los conjuntos de entrenamiento y de prueba (Witten & Frank, 2005). Si, por mala suerte, en el conjunto de entrenamiento no se tuvieran ejemplos de una clase C_1, el clasificador no tendría forma de aprender a partir de estos datos como clasificar las instancias que pertenecen a la clase C_1. Pero además, la clase C_1 estaría sobre representada en el conjunto de prueba puesto que no existen instancias de la misma en el conjunto de entrenamiento. Para evitar este tipo de situaciones se tiene que asegurar que el muestreo aleatorio se hago de manera que cada clase este representada correctamente tanto en el conjunto de entrenamiento como en el conjunto de prueba. Este procedimiento se denomina estratificación, y cuando se lo aplica junto con el método hold-out se habla de hold-out estratificado (Witten & Frank, 2005).

Si bien la estratificación es una estrategia útil, provee solo un reguardo primitivo contra la representación desigual en los conjuntos de entrenamiento y de prueba. Una manera de mitigar el sesgo causado por una selección de muestra particular para el método hold-out consiste en repetir el proceso completo, entrenamiento y prueba, varias veces con diferentes muestras aleatorias (Witten & Frank, 2005). En cada iteración una cierta

proporción (como ser dos tercios) de los datos se selecciona de manera aleatorio para el entrenamiento (posiblemente con estratificación) y el resto se usa para las pruebas. Las métricas de efectividad de las diversas iteraciones se promedian para obtener una estimación de la efectividad total.

Con el método hod-out también podría considerarse el intercambiar el rol de los datos de entrenamiento y los datos de prueba (es decir, entrenar el clasificador en los datos de prueba y probar el clasificador en los datos de entrenamiento) y promediar los dos resultados, de manera de reducir el efecto de la representación desigual de los conjuntos de entrenamiento y de prueba (Witten & Frank, 2005). Desafortunadamente, esto es solo posible con una separación 50:50 entre datos de entrenamiento y de prueba, lo cual generalmente no es lo ideal porque es mejor usar más de la mitad de los datos para el entrenamiento incluso a expensas de los datos de prueba.

Una aproximación alternativa al método hold-out pero basada en la idea de intercambiar el rol de los datos de entrenamiento y de prueba la proporciona el método de validación cruzada *n*-fold. Con la validación cruzada "la colección de datos se divide en *n folds* o subconjuntos mutuamente exclusivos (D_1, D_2,…, D_n) y aproximadamente del mismo tamaño. El entrenamiento y la prueba se realizan *n* veces. En la iteración *i*, la porción D_i se reserva como el conjunto de prueba, y las porciones restantes se usan para entrenar el clasificador. Así en la primera iteración, los subconjuntos D_2, …,D_n sirven como conjunto de entrenamiento para obtener el primer clasificador, el cual se prueba en el conjunto D_1; en la segunda iteración el entrenamiento se hace en los subconjuntos D_1,D_3,…, D_n y la prueba en D_2, y así sucesivamente." (Han & Kamber, 2006). A diferencia del método hold-out, con la validación cruzada cada muestra se usa el mismo número de veces para entrenamiento y una sola vez para prueba. Los valores de las métricas de efectividad se obtienen promediando los resultados de las *n* iteraciones. En validación cruzada lo común es que se usen 10 *folds* o muestras puesto que pruebas extensivas sobre numerosas bases de datos, con diferentes técnicas de aprendizaje, han mostrado que con 10 subconjuntos se obtiene la mejor estimación de métricas de efectividad (Witten & Frank, 2005).

III.9.2 Medidas de Efectividad

Los clasificadores de documentos típicamente son evaluados experimentalmente en vez de analíticamente. El motivo es que para evaluar un sistema analíticamente (por ejemplo, probar que el sistema es completo y correcto), se necesita una especificación

formal del problema que el sistema está tratando de resolver (por ejemplo, con respecto a que la correctitud y la completitud son definidas), y la noción central de la clasificación de textos (es decir, la pertenencia de un documento a una categoría) es, debido a su carácter subjetivo, inherentemente no formalizable (Sebastiani, 2002). La evaluación experimental de un clasificador usualmente mide su habilidad para tomar decisiones de clasificación correctas, esto es mide su efectividad en lugar de su eficiencia.

La efectividad de los clasificadores usualmente se mide en términos de dos nociones clasificas para la recuperación de información, la precisión (π) y el recall (ρ). Para definir formalmente estos dos términos supóngase que F es la función de asignación de categorías verdadera y que F* es el modelo o clasificador que trata de aproximar la función F. La precisión $c_i(\pi_i)$ se define como la probabilidad condicional P(F*(d_x, c_i)= T | F(d_x, c_i)=T) , esto es, la probabilidad de que si un documento aleatorio d_x se clasifica bajo la categoría c_i, la decisión es correcta (Sebastiani, 2002). De manera análoga el recall $c_i(\rho_i)$ se define como P(F(d_x, c_i)= T | F*(d_x, c_i)=T), esto es, la probabilidad de que si un documento d_x debería ser clasificado como c_i, la decisión se toma (Sebastiani, 2002).

La precision y el recall pueden ser vistas como probabilidades subjetivas, es decir como una medida de la expectativa de los usuarios de que el sistema se comportara correctamente cuando clasifique bajo c_i documentos previamente no procesados (Sebastiani, 2002). Siguiendo esta línea de pensamiento, estas probabilidades pueden estimarse de la tabla de contingencia para la categoría c_i (ver Tabla 3.1). En Tabla 3.1 se definen FP_i (falsos positivos o errores por comisión) es el número de documentos incorrectamente clasificados bajo c_i, TN_i (verdaderos negativos), TP_i (verdaderos positivos), y FN_i (falsos negativos o errores por omisión).

Categoría *Ci*		**Decisión del experto**	
		SI	**NO**
Decisión del Clasificador	**SI**	TP_i	FP_i
	NO	FN_i	TN_i

Tabla 3.1. Tabla de Contingencia para la categoría c_i

Teniendo en cuenta la tabla de contingencia la estimación de la precisión y el recall para la categoría c_i se calcula de la siguiente manera:

$$\pi_i = \frac{TP_i}{TP_i + FP_i} \qquad \rho_i = \frac{TP_i}{TP_i + FN_i}$$

Para estimar la precisión y el recall de todas las categorías se puede recurrir a dos métodos diferentes: micro promedio y macro promedio. Cuando se aplica el método de micro promedio (μ) la precisión y el recall se obtienen sumando todas las decisiones individuales:

$$\pi_\mu = \frac{TP}{TP + FP} = \frac{\sum_{i=1}^{|C|} TP_i}{\sum_{i=1}^{|C|} (TP_i + FP_i)}$$

$$\rho_\mu = \frac{TP}{TP + FN} = \frac{\sum_{i=1}^{|C|} TP_i}{\sum_{i=1}^{|C|} (TP_i + FN_i)}$$

Con el método de macro promedio (*M*) la precisión y el recall se evalúan primero de manera local, y luego globalmente por un promedio de los resultados de las diversas categorías.

$$\mu_M = \frac{\sum_{i=1}^{|C|} \pi_i}{|C|} \qquad \rho_M = \frac{\sum_{i=1}^{|C|} \rho_i}{|C|}$$

La tabla de contingencia global (ver Tabla 3.2) que se usa en el calculo del micro y macro promedio se obtiene sumando las tablas de contingencia individuales.

Conjunto de Categorías $C=\{c_1,\ldots,c_{\|C\|}\}$		**Decisiones del experto**	
		SI	**NO**
Decisiones del Clasificador	**SI**	$TP = \sum_{i=1}^{\|C\|} TP_i$	$FP = \sum_{i=1}^{\|C\|} FP_i$
	NO	$FN = \sum_{i=1}^{\|C\|} FN_i$	$TN = \sum_{i=1}^{\|C\|} TN_i$

Tabla 3.2. Tabla de contingencia global

Al evaluar clasificadores automáticos se tiene que buscar una solución intermedia entre precisión y recall puesto que, un clasificador podría fácilmente lograr el 100 % precisión sin tomar decisiones de clasificación (y así no cometer errores en lo que identifica), o podría lograr 100% de recall clasificando todo (y así no perder nada de lo que debería ser clasificado) **(**Cunningham *et al.*, 2011). La medida F se suele utilizar de manera conjunto con la precisión y el recall, y representa un promedio ponderado de los dos:

$$medida\ F = \frac{(\beta^2 + 1)\pi * \rho}{(\beta^2 \pi) + \rho}$$

El parámetro β en la definición de la medida F refleja la ponderación de precisión versus recall **(**Cunningham *et al.*, 2011). Si a β se le asigna el valor 1, la precisión y el recall pesan lo mismo. Con β fijado en 0.5, la precisión pesa dos veces más que el recall. En cambio con β igual a 2, es el recall el que pesa dos veces más que la precisión.

CAPÍTULO IV
LAS HABILIDADES DE LOS DOCENTES EN GRUPOS DE APRENDIZAJE COLABORATIVO ONLINE

IV.1 IMPORTANCIA DEL E-TUTOR EN ENTORNOS DE ACSC

La incorporación de medios computacionales al dominio del Aprendizaje Colaborativo (AC) ha permitido definir nuevos escenarios de enseñanza y de aprendizaje, originando los sistemas de Aprendizaje Colaborativo Soportado por Computadoras (ACSC). Estos sistemas ofrecen versiones electrónicas de muchas actividades y recursos presentes en las aulas de enseñanza tradicional (presencial). Así, es posible disponer de espacios para trabajo compartido, lecturas y presentaciones *online*, resultados de evaluaciones, listados de bibliografía, repositorios de ejercicios y materiales, etc. Contando también con herramientas de comunicación síncrona y/o asíncrona (chat, foro, e-mail, etc.) que posibilitan tanto la comunicación como la colaboración entre los estudiantes.

El énfasis que se pone actualmente en el aprendizaje en grupo está motivado tanto por la importancia que se le confiere en el ámbito empresarial al trabajo en equipo, como por la comprensión de que en una sociedad de la información tan cambiante las habilidades de comunicación son cada vez más importantes (Mcwhaw *et al*., 2003). Los sistemas de ACSC no solo brindan el soporte tecnológico necesario para un aprendizaje grupal con estudiantes localizados en distintos puntos en el tiempo y espacio sino que, son el entorno apropiado para el desarrollo de habilidades de solución de problemas (Suh & Lee, 2006), pensamiento crítico, mayor retención, establecimiento de metas, interpretación, y análisis, todas las cuales son cualidades de un líder exitoso (Day, 2009). Además, este tipo de sistemas permiten resolver problemas como el sentimiento de aislamiento, insatisfacción, y desmotivación, que suelen llevar a los estudiantes a abandonar sus cursos *online* (Chang, 2004; Hughes *et al*., 2002). Sin embargo, soportar el aprendizaje grupal mediante sistemas de software no garantiza la obtención de estos beneficios potenciales, ni que los estudiantes colaboren puesto que no son las características de la tecnología utilizada sino su forma de implementación lo que soporta la colaboración (Hughes2002; Bonk *et al*., 2004; Orvis & Lassiter, 2006; Olivares2007). Para colaborar efectivamente los estudiantes necesitan (entre otras cosas) de un docente o e-tutor que, actuando como un facilitador, coordine las

discusiones, promueva la participación, estimule la interacción, y ayude a enfrentar cualquier problema o dificultad que surja en el grupo (Kukulska-Hulme, 2004; Orvis & Lassiter, 2006; Awouters & Jans, 2009; Borges & Baranauskas, 2003; Guzdial *et al*., 2002; Hughes *et al.* , 2002; Kreijns *et al*., 2003; Soller, 2001: Schwarz & Asterhan , 2011).

La necesidad de los estudiantes de ACSC de recibir guía y asistencia por parte de un e-tutor se hace evidente cuando se considera que aun cuando los medios interactivos le dan una gran libertad de control a los estudiantes, pueden originarse dificultades cuando los mismos tengan que organizar la información que reciben, estructurar las discusiones, y desarrollar perspectivas personales, dando como resultado bajos niveles de construcción de conocimiento (De Smet *et al*., 2008). Además, el uso de computadoras puede inhibir o causar problemas en las interacciones entre estudiantes (Orvis & Lassiter, 2006; Olivares, 2007) por lo que la intervención del docente para hacer crecer los niveles de interacción y evitar que haya estudiantes que no participen es primordial (De Smet *et al*., 2008). Por último, si bien la discusión entre estudiantes es un método parcial excelente de aprendizaje, los estudiantes necesitan consultar al e-tutor (Kukulska-Hulme, 2004).

En base a lo expuesto en los párrafos anteriores, se puede afirmar que la colaboración *online* tiene pocas posibilidades de ser exitosa sin docentes con las habilidades necesarias para sacarle provecho a las herramientas disponibles (Hughes *et al.*, 2002; Kukulska-Hulme, 2004; McPherson & Baptista, 2004; Orvis & Lassiter, 2006) ya que tanto la calidad como la profundidad del aprendizaje dependen (en gran parte) de la forma en que la colaboración sea coordinada por los e-tutores (Borges & Baranauskas, 2003; Kukulska-Hulme, 2004).

La exploración bibliográfica realizada desafortunadamente reveló que no existe una base teórica sólida sobre las habilidades de los e-tutores de ACSC (Kukulska-Hulme, 2004; Orvis & Lassiter, 2006; Zumbach *et al*., 2004; Greiffenhagen, 2012) y que los pocos estudios donde se evaluó la efectividad de diferentes estrategias de enseñanza han conducido a resultados inconsistentes (Olivares, 2007). La falta de claridad en lo que respecta a las habilidades de un e-tutor incrementa el riesgo de invertir tiempo, esfuerzo, y dinero en iniciativas de ACSC que fracasan debido a que el cuerpo docente seleccionado no posee las habilidades requeridas (Gerrard, 2002). Un elemento que complica aún más el panorama es que la mayoría de los e-tutores no tienen experiencia con las técnicas usadas en ACSC (Williams & Roberts, 2002; Kukulska-Hulme2004), siendo práctica común que a docentes acostumbrados a dar clases presenciales se les asigne la tarea de dictar clases *online*. Así, se asume erróneamente que los profesores sólo necesitan aprender a usar la

tecnología (Awouters & Jans, 2009; Gerrard, 2002; Orvis & Lassiter, 2006; Olivares, 2007) ya que se considera que las técnicas utilizadas en el aula tradicional son igualmente aplicables en la enseñanza mediada por computadora (Barker, 2002; Williams & Roberts, 2002; Orvis & Lassiter, 2006). En este sentido, se debería tener en cuenta que:

- Las funciones de un docente en ambientes virtuales no son una extensión y/o transferencia de las funciones requeridas en clases presenciales (Bonk *et al*., 2004; De Smet *et al*. 2008; Guasch *et al*., 2009).
- La tecnología debería ser vista tan sólo como una herramienta al alcance del e-tutor, siendo en primer lugar necesario comprender los comportamientos de los docentes que facilitan el aprendizaje grupal, para luego buscar aquellas tecnologías que sirvan de soporte a su labor (Olivares, 2007; Greiffenhagen, 2012).

Para tratar de contribuir a rellenar el vacío teórico existente en torno a las habilidades de los docentes de ACSC se decidió plantear un esquema de clasificación de las habilidades que se considera debería poseer un e-tutor de ACSC para coordinar el aprendizaje grupal. Se piensa que la taxonomía de habilidades no solo contribuirá a tener una mejor comprensión de la tarea de los e-tutores de ACSC sino que podría ser considerada al diseñar actividades de capacitación que permitan a los e-tutores desarrollar las habilidades necesarias para desempeñarse efectivamente. Asimismo, el comprender los comportamientos de los docentes que facilitan el aprendizaje grupal es un paso previo necesario para seleccionar o diseñar herramientas de software que brinden soporte a la tarea de los e-tutores (Olivares2007, Greiffenhagen2012).

IV.2 INVESTIGACIONES SOBRE LA TAREA DE LOS E-TUTORES DE ACSC

A pesar de que en la literatura de ACSC se reconoce la importancia del e-tutor para el éxito del aprendizaje colaborativo, hay una ausencia de estudios detallados sobre la manera en la cual los e-tutores intervienen en las actividades de aprendizaje colaborativo, que de cierto modo refleja el estado ambivalente de los docentes en un campo que trata de transferir la autoridad desde los docentes a los estudiantes (Greiffenhagen, 2012). Por una parte mucha de la innovación tecnológica ha sido dirigida con el propósito de transferir la enseñanza y el aprendizaje desde clases dirigidas por los docentes a prácticas centradas en los estudiantes. Desde esta perspectiva los estudiantes no son vistos como recipientes a ser

llenados por la información impartida por los profesores, sino más bien como constructores activos de conocimiento. Sin embargo por otro lado, se reconoce que los docentes aún juegan un rol crucial, aunque nuevo, en las actividades de aprendizaje colaborativo.

Para (Greiffenhagen, 2012) la tensión existente puede apreciarse en la literatura pues los docentes han sido conceptualizados tanto en términos positivos como negativos: como figuras de autoridad que amenazan las oportunidades de aprendizaje auto dirigido de los estudiantes y, como facilitadores que guían y estimulan las actividades de aprendizaje de sus alumnos.

Teniendo en cuenta la postura de (Greiffenhagen, 2012) no es de extrañar que las habilidades propias de un e-tutor de ACSC no hayan sido descriptas detalladamente. Sin embargo, entre los investigadores que han estudiado esta temática se pueden identificar dos enfoques. Por un lado, se tiene trabajos como los de (Fåhræus, 2000; Lim & Cheah, 2003; Orvis & Lassiter, 2006; Olivares, 2007; Juárez *et al*., 2008; Palloff & Pratt, 2001) donde se plantea la forma en la que se considera deberían agruparse las actividades, tareas, o cuestiones de incumbencia para un e-tutor. Por el otro, (Teles *et al*., 2001; Bonk *et al*., 2004; Lund, 2004; Palloff & Pratt, 2007) toman como base el esquema de organización por roles heredado de la literatura sobre habilidades y competencias para docentes de *e-learning* que no recurren al aprendizaje colaborativo como estrategia pedagógica. Si se opta por seguir este segundo enfoque, uno se encuentra con que no existe consenso respecto a la manera de denominar la labor de los docentes. Algunos de los términos utilizados son: e-tutor, e-teacher, cyber moderador, e-moderador, tutor virtual, etc. Todos ellos se usan como sinónimos pero también con significados divergentes.

En el campo del *e-learning*, donde no se utiliza el aprendizaje colaborativo, también existe una dualidad a la hora de referirse a las actividades de los e-tutores. Desde la década de 1990 se han sugerido pautas generales o bien descripto los roles potenciales de los e-tutores con el fin de orientar a los docentes nuevos en la enseñanza *online* (Bonk *et al*., 2004). La descripción de los roles usualmente viene acompañada con sus competencias y/o habilidades correspondientes. Estos dos términos se usan de manera intercambiable por muchos investigadores sin embargo, tal como se comento en el capítulo II los términos habilidad y competencia no son equivalentes sino que las habilidades están incluidas dentro de las competencias. Otra cuestión a considerar es que, las habilidades que son de relevancia para e-tutores de entornos donde los estudiantes aprenden de manera individual no pueden transferirse directamente a los docentes de ACSC puesto que en la dinámica grupal surgen problemas (personas que nunca participan, personas que tienden a

dominar la discusión, discusiones sin fin, etc.) que no están presentes cuando los estudiantes aprenden de manera individual.

IV.3 DIFERENCIAS ENTRE E-LEARNING Y CLASES PRESENCIALES

Aun cuando exista cierta similitud entre la enseñanza en ambientes de *e-learning* y las clases presenciales, los considerados buenos docentes no necesariamente serán buenos *e*-tutores ya que los ambientes *online* requieren por parte de los docentes un nuevo conjunto de habilidades (Higgison, 2001; Duggleby *et al.*, 2002; Dewar & Whittington, 2000). La enseñanza en *e-learning* no es algo tan simple como digitalizar documentos escritos o usar el correo electrónico para enviar las actividades o calificaciones a los estudiantes (Llorente Cejudo, 2006). Las diferencias existentes entre las clases presenciales y las clases en ambientes *online* pueden servir de ayuda para comprender las causas por las cuales las habilidades de un e-tutor no son una simple transferencia y/o extensión de las requeridas en los docentes tradicionales (Guasch *et al.*, 2009). Tomando como base a (Cornelius & Higgison, 2001; CSALT ,2001; McPherson & Baptista Nunes, 2004; Duggleby *et al.*, 2002) las principales diferencias se muestran en la siguiente Tabla 4.1.

Característica	**Clases presenciales**	**Clases *online***
Tiempo	Se requiere asistencia regular a clases.	Los estudiantes manejan su propio tiempo.
Distancia	El docente obtiene feedback verbal y no verbal de forma inmediata.	Este tipo de feedback no está disponible.
Tecnología	La tecnología no es la característica principal.	El e-tutor debe ser capaz de tratar con problemas técnicos con la interfaz o su computadora para que el aprendizaje tenga lugar.
Sentido de control del docente	Alto de grado de liderazgo del docente que no es posible ignorar.	Existe menor grado de control del docente y los alumnos lo pueden ignorar fácilmente.
Condiciones de la reunión	Frecuentemente se debe esperar a que los participantes lleguen.	No se debe esperar a que los participantes lleguen.
Dinámica grupal	Escuchar sin participación puede ser mal visto.	Puede haber escucha activa sin participación.
	Discusiones/ interacciones inmediatas	Discusiones/interacciones con delay.

Tabla 4.1. Diferencias entre clases presenciales y clases *online*

Característica	Clases presenciales	Clases *online*
Marcadores culturales (apariencia física, voz, raza, etc.)	Presencia de marcadores culturales	Ausencia de marcadores culturales
Secuencia de la discusión	Sigue una secuencia lineal	No sigue una secuencia lineal y en su lugar se promueven múltiples discusiones.

Tabla 4.1. (Cont.)

IV.4 CLASIFICACIÓN DE HABILIDADES PROPUESTA

Como parte de la creación del esquema de clasificación de habilidades de docentes de ACSC se consideró conveniente integrar las dos líneas de investigación existentes en la literatura de ACSC que se describieron en la sección IV.2, considerando además los planteos pertenecientes al área del *e-learning* no colaborativo. Por otra parte, también fue necesario llevar a cabo una revisión de bibliografía sobre aprendizaje cooperativo y aprendizaje colaborativo. La indagación en el terreno del aprendizaje cooperativo se justifica por ser una de las corrientes de investigación que contribuyó al desarrollo de las ideas de ACSC (Olivares, 2007).

El análisis de todo este material dejó en evidencia en primer lugar, que existen categorías y habilidades equivalentes, y en segundo, que en algunos trabajos se plantean categorías que abarcan a una o más categorías de otras propuestas. Luego de examinar cada una de las habilidades identificadas en la literatura revisada, se decidió agrupar a aquellas que parecían estar estrechamente vinculadas, y se rotuló a cada grupo con un término que reflejara el sentido de las habilidades que contiene. Conjuntamente con este proceso de agrupamiento de habilidades relacionadas, se optó por tomar como principal criterio de organización los roles (Administrativo, Pedagógico, Social, Técnico, Comunicación y Evaluación) que desempeñan los e-tutores al coordinar el aprendizaje.

Con el propósito de facilitar la comprensión de la labor de los docentes en entornos de ACSC se decidió organizar la taxonomía en los tres niveles que se pueden ver en la Figura 4.1. El primero respondiendo a las habilidades vinculadas a los roles arriba mencionados, el segundo considerando las sub habilidades relacionados con cada habilidad, y finalmente, el tercer nivel para especificar los atributos de cada sub habilidad.

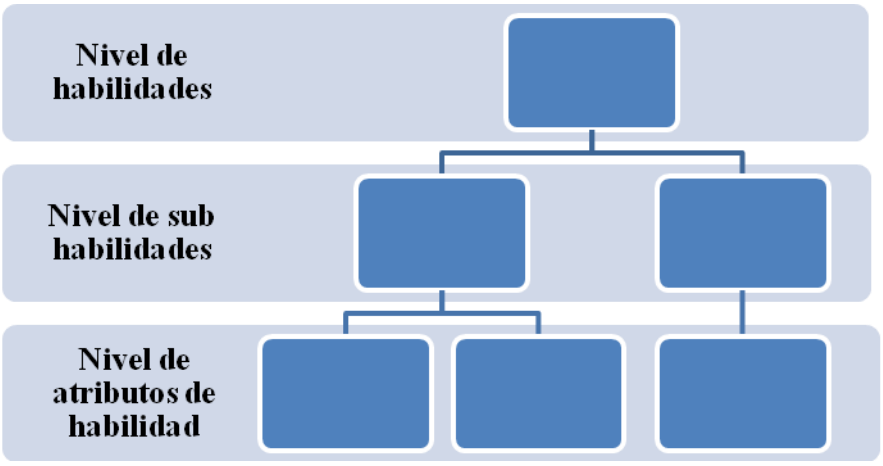


Figura 4.1. Niveles de la clasificación de habilidades

A pesar de que las habilidades se hayan separado en categorías no se las debe considerar de manera aislada ya que como se podrá apreciar en los párrafos venideros existen habilidades que se vinculan o tienen influencia sobre una o más habilidades. Así por ejemplo la sub habilidad de preguntar tiene incidencia en las categorías sociales, pedagógicas y de evaluación. En la Figura 4.2 las flechas entre los cuadrados muestran la idea de que las categorías de habilidad se vinculan entre sí.

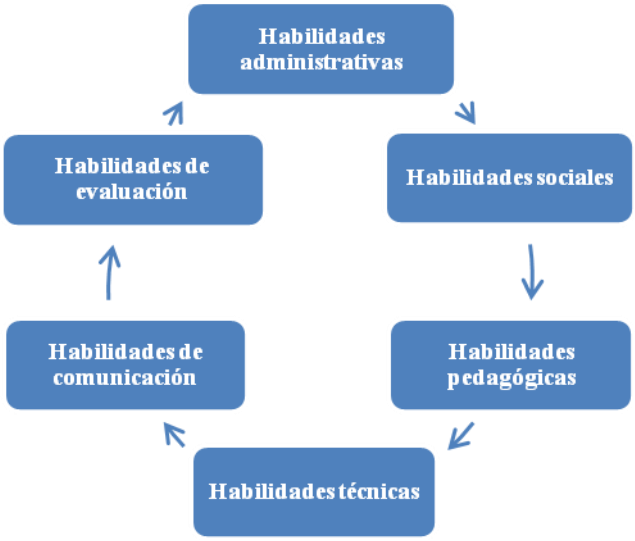


Figura 4.2. Relación entre las categorías de la clasificación de habilidades

A continuación se describe cada una de las categorías de habilidad de la Figura 4.2 en términos de sus sub habilidades y atributos de habilidad.

IV.4.1 Habilidades Administrativas

En el plano administrativo los e-tutores se encargan de organizar su propio trabajo y las actividades de los estudiantes, así como también de mantener la estructura del curso (Barker, 2002; Bonk *et al*., 2004; Llorente Cejudo, 2006; Lund, 2004). La importancia de las habilidades administrativas es tal que si un e-tutor no establece una agenda y mantiene al grupo trabajando en pos de sus objetivos la experiencia de aprendizaje puede fracasar (Mason, 1991). Además, si un e-tutor muestra deficiencia de este tipo de habilidades (por medio de alteraciones constantes en los horarios, cambios continuos en las fechas de presentación de las tareas, pobre administración de tiempo, etc.), puede provocar sentimientos de incertidumbre, confusión y frustración en la comunidad de estudiantes (Packham *et al*., 2004). Las sub habilidades que se considera forman parte de categoría de habilidades administrativas se detalla a continuación.

IV.4.1.1 Planificación

En términos generales la planificación consiste en establecer el camino más apropiado para dirigir las necesidades de los estudiantes (TCOL, 2003). La planificación es un aspecto de la labor de un e-tutor que está presente a lo largo de un curso, dado que frecuentemente se necesita hacer ajustes o recurrir a planes de contingencia cuando falla la tecnología o se presentan otro tipo de problemas que provocan desfasajes temporales en la ejecución de las actividades previstas (Varvel, 2007). La sub habilidad de planificación se puede describir en términos de los siguientes atributos de habilidad:

- Especificar los objetivos de la experiencia de aprendizaje (Mason, 1991; Bláquez Entonado & Alonso Díaz, 2006; Benito Osorio, 2009; Llorente Cejudo, 2006; Goodyear *et al*., 2001; Adell & Sales, 1999; Lim & Cheah, 2003; Palloff & Pratt, 2007; Bonk *et al*., 2004).
- Especificar el calendario de actividades y los horarios teniendo en cuenta que se debe dar a los estudiantes el tiempo suficiente para gestionar sus discusiones, leer, escribir y reflexionar (Mason, 1991; CSALT, 2001; Bláquez Entonado & Alonso Díaz, 2006; Benito Osorio, 2009; Ulster, 2002; Llorente Cejudo, 2006; Adell & Sales, 1999; Fåhræus, 2000; Lim & Cheah, 2003; Palloff & Pratt, 2007; Bonk *et al*., 2004; Juárez *et al*., 2008).
- Identificar tareas que pueden ser agrupadas o reasignadas (Kemshal Bell, 2001).
- Establecer las prioridades para la enseñanza (Kemshal Bell, 2001).

IV.4.1.2 Administración del tiempo

Si se tiene en cuenta que una regla practica comúnmente usada es que un día clase en un aula se traduce en una semana *online* (Cornelius & Higgison, 2001), es evidente que en *e-learning* los procesos de enseñanza y de aprendizaje consumen un tiempo considerable. Además frecuentemente los estudiantes tienen que cumplir plazos (Duggleby *et al*., 2002; Varvel, 2007). Por lo anterior se considera que un e-tutor requiere de los atributos de habilidad de administración de tiempo indicados a continuación:

- Ayudar a los estudiantes en la administración de su propio tiempo para evitar que tengan dificultadas para organizar sus estudios efectivamente (Hootstein, 2002; Jones *et al*., 2004).
- Mantener a los estudiantes al tanto de la planificación y tiempo disponible (CSALT, 2001; Kemshal Bell, 2001; Duggleby *et al*., 2002; Smith, 2005; Juárez *et al*., 2008).
- Renegociar fechas de vencimiento si emergen conflictos (Palloff & Pratt, 2007; Bonk *et al.*, 2004).
- Lograr que se cumpla el calendario establecido (Juárez *et al*., 2008).

IV.4.1.3 Coordinación

La sub habilidad de coordinación permite dar y mantener en el curso una estructura acorde con los entornos de ACSC (Lund, 2004). Esta sub habilidad se relaciona con los atributos de habilidad de:

- Organizar a los estudiantes para actividades grupales especificando el tamaño de los grupos, los roles de los participantes, el tiempo de duración de las discusiones, el medio de comunicación que se usará, los tópicos a tratar, etc. (IITT, 2007; Barker, 2002; Llorente Cejudo, 2006; Lim & Cheah, 2003; Bonk *et al.*, 2004; Lund, 2004; Teles *et al*., 2001).
- Coordinar las actividades de modo tal de ayudar a los estudiantes a conocer qué necesita ser hecho y cuando (CSALT, 2001; Teles *et al*., 2001; Lund, 2004; Juárez *et al*., 2008; Bonk *et al.*, 2004).
- Sincronizar y re sincronizar: asegurar que todos los estudiantes comienzan al mismo tiempo y que sus actividades estén sincronizadas (Berge, 1995; Llorente Cejudo, 2006).

IV.4.1.4 Establecer y cumplir con las directrices para el proceso de aprendizaje

Es necesario establecer y mantener directrices que sirvan de guía para los estudiantes durante el proceso de aprendizaje (ION, 2005), ya que caso contrario los mismos estarán confundidos y desorganizados, y los procesos de enseñanza y de aprendizaje podrían verse obstaculizados (Palloff & Pratt, 2007). Los atributos de habilidad que contribuyen a fijar y mantener las normas que rigen el proceso de aprendizaje son:

- Definir guías de comportamiento para los estudiantes: sobre todo cuando interactúan estudiantes de diferentes culturas se debe proveer una guía de lo que es apropiado y lo que es inapropiado puesto que, pueden existir discrepancias sobre lo que constituye un comportamiento aceptable (Kemshal Bell, 2001; Hootstein, 2002; Duggleby *et al*., 2002; Smith, 2005; Lim & Cheah, 2003; Palloff & Pratt, 2007).
- Especificar procedimientos de toma de decisión (Mason, 1991; Benito Osorio, 2009; Llorente Cejudo, 2006; Adell & Sales, 1999; Palloff & Pratt, 2007).
- Establecer convenciones para la comunicación que enfaticen la interacción y minimicen el conflicto afectivo (Kemshal Bell, 2001; eTTNet TWG 2, 2005; Orvis & Lassiter, 2006; Bonk *et al*., 2004).
- Especificar reglas de procedimiento para hacer contribuciones (Bonk *et al*., 2004, Lim & Cheah, 2003).
- Indicar el tipo de lenguaje que es aceptable (Olivares, 2007).
- Conducir a los estudiantes en base a las restricciones éticas, institucionales y legales que tienen influencia sobre el curso: es preciso que se tenga en cuenta reglas de reporte académico y privacidad de los estudiantes, regulaciones respecto a los estudiantes con discapacidades, y procedimientos de honestidad académica (Varvel, 2007; eTTNet TWG 2, 2005). La habilidad del e-tutor para mantener a un curso funcionando dentro de la configuración institucional, legal, y de la ética general tiene tanta importancia como su grado de capacitación ya que si un docente deja de lado estas cuestiones puede perder la oportunidad de demostrar cuan capacitado esta (Varvel, 2007).
- Recordarle a los estudiantes las guías y procedimientos establecidos sobre todo cuando no cumplen con alguno de estos (CSALT, 2001; Lim & Cheah, 2003).
- Modificar directrices en caso de ser necesario (Palloff & Pratt, 2001).

IV.4.1.5 Administrar el contenido de las conferencias

El administrar el contenido de las conferencias comprende los atributos de habilidad de:

- Cambiar encabezados de tema incorrectos: en el caso de que una contribución de los estudiantes se haya hecho bajo un encabezado erróneo se lo debe cambiar de lugar (Berge, 1995; CSALT ,2001; Llorente Cejudo, 2006).
- Mantener la estructura de la conferencia en una forma lógica (Barker, 2002).
- Mantener el espacio *online* libre de basura tal como post duplicados, pornografía, publicidad, o cualquier tipo de contenido que las normas del grupo no habiliten (Full Circ, 2001).

IV.4.1.6 Administrar programas de aprendizaje

Es necesario que un e-tutor de un inicio adecuado a los programas o eventos de aprendizaje. Esto implica invitar a los estudiantes a participar, publicar un mensaje de bienvenida indicando que el curso está abierto, o tomar cualquier acción necesaria para que los estudiantes inicien sus actividades (CSALT ,2001; eTTNet TWG 2, 2005). Por otro parte, también es conveniente contar con una actividad que, una vez alcanzados los objetivos deseados, proporcione un cierre apropiado para el curso o actividad de aprendizaje (eTTNet TWG 2, 2005; Berge, 1995; Lim & Cheah, 2003; McPherson & Baptista Nunes, 2004). La necesidad de finalizar las actividades puede parecer trivial pero es importante porque el entusiasmo de algunos estudiantes puede conducirlos a exigirle a los docentes cuestiones que están fuera del ámbito del curso (Cornelius & Higgison, 2001).

IV.4.1.7 Gestionar información

Para llevar a cabo la tarea de gestionar la información un e-tutor debería contar con atributos de habilidad que le permitan:

- Recolectar y archivar trabajos de los estudiantes para evaluación (Barker, 2002; Bonk *et al.*, 2004).
- Archivar y mantener recursos de enseñanza (Barker, 2002).
- Mantener proformas administrativas y plantillas de documentos que contienen, por ejemplo datos demográficos de los estudiantes (nombre, antecedentes educacionales, necesidades educativas especiales, etc.), asistencia de los alumnos, notas, etc. (IITT, 2007; Barker, 2002; TCOL, 2003; Chang, 2004).

- Mantener registros de los logros de los estudiantes y resultados de la aplicación de lo aprendido (eTTNet TWG 2, 2005).
- Comunicar a los estudiantes información sobre plazos, procedimientos y el programa del curso (TCOL, 2003; Palloff & Pratt, 2007).
- Ayudar a los estudiantes a manejar la sobrecarga de información ya que muchos pueden no estar acostumbrados a la gran cantidad de información disponible en ambientes de *e-learning* (Hootstein, 2002; McPherson & Baptista Nunes, 2004).
- Asegurar que la información acerca de los estudiantes se transmite a la persona correcta en el momento correcto (TCOL, 2003).

IV.4.1.8 Proveer dirección y soporte administrativo

Las cuestiones administrativas (revisión y cambio de materias, acceso a las planificaciones, enrolamiento, etc.) no deberían ser una pesadilla para los estudiantes (IITT, 2007; Lefoe *et al*., 2002; de Laat, 2006) si un e-tutor cuenta con los atributos de habilidad de: establecer mecanismos apropiados para proveer dirección y soporte (Kemshal Bell, 2001; IITT, 2007), identificar cuándo la dirección y el soporte son requeridos (Kemshal Bell, 2001), y proveer dirección y soporte de acuerdo con las necesidades de los estudiantes (ION, 2005; Kemshal Bell, 2001).

IV.4.1.9 Monitorear Actividades

El monitorear las actividades de los estudiantes para conocer que están haciendo en cada momento y asegurar la autenticidad de su trabajo (Full Circ, 2001; Cornelius & Higgison, 2001; Guasch *et al.*, 2009; de Laat, 2006) implica que los e-tutores posean los atributos de habilidad de:

- Registrar cuándo el estudiante se comunico (Duggleby *et al*., 2002).
- Registrar cuándo se dio *feedback* (Duggleby *et al*., 2002).
- Identificar si un estudiante se está quedando retrasado (Duggleby *et al*., 2002).
- Identificar si un estudiante está inactivo (Duggleby *et al*., 2002).

Sub habilidades	Atributos de habilidad
Planificación	• Especificar objetivos de la experiencia de aprendizaje. • Especificar el calendario de actividades y los horarios. • Identificar tareas que pueden ser agrupadas o reasignadas. • Establecer prioridades para la enseñanza.

Tabla 4.2. Sub habilidades y atributos de habilidad administrativos de e-tutores de ACSC

Sub habilidades	Atributos de habilidad
Administración del tiempo	• Ayudar a los estudiantes a administrar su tiempo. • Mantener a los estudiantes al tanto de la planificación y el tiempo disponible. • Renegociar fechas de vencimiento. • Lograr que se cumpla el calendario establecido.
Coordinación	• Coordinar el trabajo grupal. • Coordinar las actividades. • Sincronizar y re sincronizar.
Establecer y cumplir con las directrices para el proceso de aprendizaje	• Definir guías de comportamiento. • Especificar procedimientos de toma de decisión. • Establecer convenciones para la comunicación. • Especificar reglas de procedimiento para hacer contribuciones. • Indicar el tipo de lenguaje aceptable. • Conducir al curso en base a las restricciones éticas, institucionales y legales. • Recordarle a los estudiantes las reglas y procedimientos establecidos. • Modificar directrices.
Administrar el contenido de las conferencias	• Cambiar encabezados de tema incorrectos. • Mantener la estructura de la conferencia. • Mantener el espacio *online* libre de basura.
Administrar programas de aprendizaje	• Abrir programas o eventos de aprendizaje. • Finalizar programas o eventos de aprendizaje.
Gestionar información	• Recolectar y archivar trabajo de los estudiantes. • Archivar y mantener recursos de enseñanza. • Mantener proformas administrativas y plantillas de documentos. • Mantener registros de los logros de los estudiantes. • Comunicar a los estudiantes información sobre plazos, procedimientos, y el programa del curso. • Ayudar a manejar la sobrecarga de información. • Asegurar que la información de los estudiantes se transmita a la persona correcta.
Proveer dirección y soporte administrativo	• Determinar los medios apropiados para proveer dirección y soporte. • Identificar cuando se requiere dirección y soporte. • Proveer dirección y soporte de acuerdo a las necesidades de los estudiantes.
Monitorear Actividades	• Registrar cuando los estudiantes se comunican. • Registrar cuando se dio *feedback* . • Identificar estudiantes retrasados. • Identificar estudiantes inactivos.

Tabla 4.2. (Cont.)

IV.4.2 Habilidades Sociales

La faceta social de la labor de un e-tutor se vincula con incorporación de los aspectos sociales y personales de modo tal de desarrollar un ambiente confortable y amigable donde los estudiantes sientan que están interactuando con otras personas y no con una máquina (Lim & Cheah, 2003; Blackburn, 2004; Lund, 2004; Llorente Cejudo, 2006; Palloff & Pratt, 2007). Las sub habilidades sociales impactan directamente en la efectividad de la colaboración *online* porque:

- La carencia de señales no verbales y el contexto social significan que la construcción de una atmosfera adecuada para la educación depende de las herramientas virtuales e interacciones (Bonk *et al*., 2004; Teles *et al*., 2001).
- Permiten minimizar situaciones como el aislamiento y falta de motivación que suelen producirse cuando un alumno trabaja solo en este tipo de entornos (Llorente Cejudo, 2006).

Los atributos de habilidad que se considera forman parte de la faceta social de la tarea de los e-tutores son los especificados en los próximos párrafos.

IV.4.2.1 Estimular la interacción entre estudiantes

Si los integrantes de un grupo hablan uno después del otro, sin tener en cuenta lo que plantean sus pares, ni mucho menos hacer comentarios al respecto, tal diálogo no puede considerarse colaborativo (Ingram & Hathorn, 2004). Es preciso no confundir interacción con participación porque la interacción requiere que los miembros del grupo se respondan activamente entre sí, explicitando ideas y generando devoluciones (Cenich & Santos, 2006). Uno de los grandes problemas de los entornos de *e-learning*, en los que los estudiantes apenas pueden interactuar con sus pares o el *e-tutor*, es el alto índice de abandono debido a la sensación de soledad (Goméz Gutiérrez, 2008; Sulčič & Sulčič ,2007; Chang, 2004). El lograr un adecuado nivel de interacción entre los estudiantes contribuye a disminuir el índice de abandono de los mismos e influye de manera positiva sobre su motivación, satisfacción y aprendizaje (Goméz Gutiérrez, 2008). Dado que la interacción entre estudiantes es un elemento clave en el aprendizaje grupal y que la tecnología frecuentemente causa problemas o inhibe la interacción estudiante-estudiante, le corresponde a los e-tutores la tarea de moderar el trabajo grupal posibilitando las interacciones (Kreijns *et al*., 2003; Kukulska-Hulme, 2004; Orvis & Lassiter, 2006; Olivares, 2007). Los atributos de habilidad vinculados con incitar la interacción entre los estudiantes son:

- Usar una diversidad de técnicas (trabajo entre pares, discusiones informales, debates, etc.) para estimular a los estudiantes a ponerse en contacto e interactuar (Berge, 1995; Blackburn, 2004; Llorente Cejudo, 2006; Goodyear *et al*., 2001; ION, 2005; Hootstein, 2002; Cornelius & Higgison, 2001; Salmon, 2003).
- Mejorar las relaciones entre los estudiantes (ION, 2005; Hootstein, 2002; Guasch *et al*., 2009).
- Moderar el trabajo grupal de modo de permitirle a los estudiantes interactuar pero teniendo en cuenta que se debe minimizar las intervenciones para que los estudiantes se puedan conocer y puedan gestionar sus discusiones en forma independiente (Oren *et al.*, 2002). Demasiada participación del *e-tutor* puede tener el efecto de reducir la cantidad de interacción entre los estudiantes y crear dependencia hacia él (Oren *et al.*, 2002; Cornelius & Higgison, 2001). La participación excesiva del *e-tutor* es evidente cuando los alumnos esperan que el docente realice una contribución antes de continuar y cuando los estudiantes dirigen sus preguntas o comentarios a él en lugar de a otros miembros del grupo (Palloff & Pratt, 2001).
- Discutir con los estudiantes las formas de facilitar la creación de interacciones sociales (Oren *et al.*, 2002).
- Estimular a los estudiantes a responderse mutuamente recurriendo a estrategias tales como redireccionar preguntas, es decir que cuando un estudiante haga un pregunta en lugar de responderla se le puede pedir a otro miembro del grupo que lo haga (Jaques & Salmon, 2007).

La interacción social y construcción de relaciones interpersonales es de gran importancia en los ambientes de ACSC por lo cual se le debería dar a los estudiantes el tiempo necesario para que interactúen (Orvis & Lassiter, 2006).

IV.4.2.2 Estimular la participación

En las clases *online* es probable que un número limitado de alumnos esten activos dado que muchas veces se sienten inhibidos y se limitan a hacer contribuciones formales (Cornelius & Higgison, 2001; Jaques & Salmon, 2007). De hecho, comúnmente se espera que alrededor de un tercio de los estudiantes estén activos, un tercio simplemente observe y el tercio restante no tome parte de la actividad *online* (Cornelius & Higgison, 2001). De aquí que sea necesario promover la participación de los estudiantes cuando exista carencia

de contribuciones (Lim & Cheah, 2003; Olivares, 2007; Juárez *et al*., 2008; Bonk *et al*., 2004). Los atributos de habilidad que se requieren en este sentido son:

- Solicitar a los estudiantes que realicen comentarios o expresen su opinión sobre en un tópico o pregunta (Berge, 1995; CSALT ,2001; Llorente Cejudo, 2006; Jaques & Salmon, 2007).
- Impulsar a los estudiantes a explicar o dar más detalles sobre sus ideas o las ideas presentadas por sus compañeros (Mason, 1991; Teles *et al.*, 2001; Llorente Cejudo, 2006) mediante recursos tales como preguntas de elaboración del tipo "¿Puedes decirme más?", "¿Puedes explicar eso?", y "¿Qué más?" (Jaques & Salmon, 2007).
- Asegurar que todos los miembros del grupo contribuyan de manera igualitaria porque la colaboración no ocurre dentro de un grupo a menos que haya una participación aproximadamente igual entre los integrantes (McWhaw *et al*., 2003; Ingram & Hathorn, 2004). El grupo no está realmente colaborando si algunos miembros hacen la mayor parte del trabajo mientras otros apenas contribuyen (Ingram & Hathorn, 2004).
- Cambiar el tema: cuando hay carencia de contribuciones de los estudiantes el *e-tutor* necesita ser lo suficientemente flexible para modificar el tema o tópico de discusión (Lim & Cheah, 2003).
- Plantear tópicos o tareas significativas que lleven a los estudiantes a involucrarse activamente en las discusiones (Olivares, 2007).
- Regresar preguntas: la estrategia de realizar a los alumnos las preguntas que ellos mismos plantean puede dar lugar a mayor participación de los estudiantes (Jaques & Salmon, 2007).
- Indagar sobre las causas de la reducción en el nivel de participación: una disminución en el nivel de participación usualmente indica que los estudiantes están atravesando por dificultades y con simplemente preguntarles si existe algún problema se puede originar respuestas (Jaques & Salmon, 2007; Full Circ, 2001).

IV.4.2.3 Desarrollar, soportar y mantener una comunidad de aprendizaje

Una comunidad de aprendizaje que apoye y anime a los estudiantes es un prerrequisito para el éxito del aprendizaje online porque permite reducir los sentimientos de aislamiento y de alienación que son típicos en entornos de *e-learning* (Varvel, 2007; Cornelius & Higgison, 2001; Hootstein, 2002; ION, 2005). Para contribuir a la creación de una comunidad de aprendizaje se precisan de los atributos de habilidad de:

- Desarrollar la cohesión grupal: es tarea del e-tutor tratar de proveer a los estudiantes oportunidades para desarrollar la cohesión grupal y mantenerse unidos, ya que la distancia en tiempo y espacio propia de los ambientes de *e-learning* puede impedir el desarrollo de la cohesión grupal (Berge, 1995; Collins & Berge, 1996; Palloff & Pratt, 2007; Lim & Cheah, 2003; Orvis & Lassiter, 2006; Fåhræus, 2000). La cohesión contribuye a que los estudiantes estén motivados a destinar más tiempo en realizar contribuciones reflexivas y tomar parte en un diálogo significativo, en vez de cumplir sin entusiasmo alguno con su parte de la tarea (Orvis & Lassiter, 2006). Comenzar un curso *online* con una discusión de los objetivos de aprendizaje y trabajar en pos de lograr un objetivo común crea los cimientos para una comunidad de aprendizaje puesto que, la cohesión grupal es positivamente reforzada si los objetivos del grupo coinciden con los objetivos de sus miembros (Palloff & Pratt, 2007; Fåhræus, 2000). También, compartir detalles de la vida personal (viajes, observaciones, emociones, quien se es como persona, etc.) promueve la cohesión grupal (Palloff & Pratt, 2007).
- Promover la colaboración: un e-tutor debería poder crear oportunidades para que los estudiantes colaboren de modo tal que compartan experiencias y conocimientos (Denis *et al*., 2004; Kemshal Bell, 2001; Cornelius & Higgison, 2001; Blackburn, 2004; McPherson & Baptista Nunes, 2004; Varvel, 2007; eTTNet TWG 2, 2005; Smith, 2005). El atributo de habilidad de promover la colaboración es un elemento crítico en la formación de una comunidad de aprendizaje *online*, a tal punto que, la carencia de este atributo puede dar como resultado bajos niveles de participación y una interacción bidireccional entre el docente y los estudiantes (Palloff & Pratt, 2007). La colaboración se ve fuertemente afectada cuando los estudiantes comienzan a competir entre sí, por lo cual si en los grupos se observa la emergencia de la competencia es necesario buscar la forma de mantener un balance entre competencia y colaboración (Kemshal Bell, 2001).
- Promover el respeto: para funcionar como una comunidad de aprendizaje los estudiantes necesitan sentir que son respetados como personas y que son participantes iguales en el proceso de aprendizaje (Palloff & Pratt, 2007; Varvel, 2007). Los e-tutores deberían prestarle atención a los insultos y las burlas porque constituyen dos formas de falta de respeto bastante comunes. Cuando los alumnos están frustrados, confundidos o no concuerdan con el punto de vista de sus pares pueden expresar estas emociones inapropiadamente arremetiendo contra el curso,

insultando, usando un tono irrespetuoso o incluso mediante comentarios derogatorios (Berge, 1995; Cornelius & Higgison, 2001; CSALT ,2001; Blackburn, 2004; Duggleby *et al*., 2002; Palloff & Pratt, 2001; Lim & Cheah, 2003; Olivares, 2007). Asimismo, no hay que perder de vista que los estudiantes pueden incluir bromas ofensivas que atacan a otros individuos o al grupo (Duggleby *et al*., 2002).

Cuando un estudiante insulta o se burla de alguno de sus compañeros el e-tutor dispone de varias alternativas para hacer frente a esta situación como ser: remover al estudiante del curso (CSALT ,2001), si el mensaje es extremo eliminarlo e informar a los alumnos las razones de tal determinación (Duggleby *et al*., 2002), editar el mensaje si no es extremo e informar las razones de tal determinación (Duggleby *et al*., 2002), y recordarle al estudiante que la contribución esta fuera de las reglas de comportamiento aceptadas (TCOL, 2003; McPherson & Baptista Nunes, 2004; Bailey & Luetkehans, 1998).

Una comunidad de aprendizaje no puede existir al menos que los miembros del grupo se respondan entre sí (Palloff & Pratt, 2007), por lo que el atributo de habilidad de impulsar a los estudiantes a responderse mutuamente (mencionado en la sub habilidad de estimular la interacción) también tiene influencia en la existencia de la comunidad de aprendizaje. Por último, la formación de una comunidad de aprendizaje necesita que los grupos interactúen armoniosamente (Fåhræus, 2000) pero esto no es una tarea sencilla sino que se necesita toda una gama de atributos de habilidad como ser: tratar con comportamientos de discusión inapropiados, manejar contribuciones inapropiadas, manejar conflictos, y motivar a los estudiantes.

IV.4.2.4 Crear un ambiente de aprendizaje agradable

La creación de un ambiente calmo, relajado, y positivo favorece a que los estudiantes se sientan incluidos y que perciban que están participando en una red con otras personas en vez de interactuando con una máquina (Lim & Cheah, 2003; CSALT, 2001; Duggleby *et al*., 2002; Palloff & Pratt, 2007; Goodyear *et al*., 2001; Kemshal Bell, 2001; Ulster, 2002; Oren *et al.*, 2002). Un ambiente con las características mencionadas debería impulsar más participación, menos conflicto, y una fuerte cohesión social en el grupo (Orvis & Lassiter, 2006). Los atributos de habilidad de un *e-tutor* que contribuyen a la creación de un ambiente con las características antes mencionadas son:

- Dar la bienvenida a los estudiantes que participan en el curso (Cornelius & Higgison, 2001; Llorente Cejudo, 2006; Palloff & Pratt, 2001; Bonk *et al*., 2004; Lund, 2004).
- Usar rompe hielos: siempre es bueno comenzar con los llamados rompe hielos que pueden ser tan simple como pedirles a los estudiantes que se presenten a sí mismos o que planteen sus expectativas para el curso (Duggleby *et al*., 2002; Goodyear *et al*., 2001; Palloff & Pratt, 2001; Palloff & Pratt, 2007). El impulsar a los participantes a introducirse o presentarse por sí mismos suele ser una forma conveniente de ayudarles a familiarizarse (Berge, 1995; Cornelius & Higgison, 2001; CSALT ,2001; Blackburn, 2004; Llorente Cejudo, 2006). Por su parte, cuando los estudiantes comenten sus expectativas para el curso se puede chequear que todos estén comenzando desde el mismo punto y, cuanto mayor sea la coincidencia en expectativas mayor es la probabilidad de que el proceso de aprendizaje sea colaborativo (Palloff & Pratt, 2007).
- Fomentar comentarios sobre las presentaciones: para que los estudiantes se sientan bienvenidos conviene que se los estimule a responder a la presentación o introducción publicada por sus pares ya que caso contrarío puede darse que los alumnos publiquen sus presentaciones pero nadie las tenga en cuenta (Palloff & Pratt, 2007).
- Proteger contra el miedo al ridículo: la creación de una atmósfera donde los estudiantes se sientan confiados, confortables, y libres de compartir sus pensamientos y sentimientos ayuda a que no tengan miedo al ridículo ni a la retribución de su pares (Duggleby *et al*., 2002; Palloff & Pratt, 2007). El miedo al ridículo público frecuentemente sofoca la participación de los estudiantes (Berge, 1995; Llorente Cejudo, 2006).
- Adoptar un estilo amigable e informal en las contribuciones como por ejemplo deseando felicidades a los estudiantes en fechas festivas (De Smet *et al.*, 2008; Lim & Cheah, 2003).
- Impulsar a los estudiantes a actuar de forma amigable entre sí (Palloff & Pratt, 2007; Oren *et al.*, 2002).

IV.4.2.5 Manejar comportamientos de discusión inapropiados

Los comportamientos de discusión de los estudiantes que pueden impedir su progreso en el logro de sus objetivos (Jaques & Salmon, 2007) a los que un e-tutor debe estar atento son: mirones, no participantes, bloqueadores, dominadores, y buscadores de reconocimiento.

IV.4.2.5.1 Mirones

Los mirones son aquellos estudiantes que se loguean en el curso pero se limitan a observar la actividad de sus pares sin tomar parte activa. Un e-tutor debe aceptar mirones, monitorear su participación, y proveer soporte e impulso cuando sea apropiado ya que algunas personas aprenden observando a otros (Kemshal Bell, 2001). Esto implica que no se debe asumir que el aprendizaje no está teniendo lugar si no participan (Berge, 1995; CSALT ,2001; Blackburn, 2004; Llorente Cejudo, 2006). Además, los estudiantes podrían tener buenas razones para no verse involucrados en discusiones y esto debe ser respetado (Cornelius & Higgison, 2001). El decidir la manera de intervenir ante un alumno mirón es un poco más complicado en entornos de *e-learning* que en clases presenciales. Si un estudiante esta frecuentemente en línea pero en silencio un e-tutor puede asumir que eligió estar en silencio, de modo que cualquier intervención por parte del *e-tutor* para llamarle la atención puede ser vista por el mismo como dirigiéndose al estudiante en una manera poco útil. Sin embargo, puede que el alumno no haya precisamente elegido no participar sino que existen múltiples motivos que los llevaron a este comportamiento y cualquier oportunidad que se le brinde para participar será bienvenida por su parte (CSALT, 2001).

IV.4.2.5.2 No participantes

Cuando se detecta que un estudiante no estuvo en contacto o no se logueo en el sitio del curso por un largo periodo de tiempo, es conveniente contactarlo para determinar las causas de su ausencia e invitarlo a participar más activamente (Duggleby *et al*., 2002; Jaques & Salmon, 2007; Palloff & Pratt, 2001). En las clases *online* la ausencia de un estudiante es bastante perceptible y tiene un efecto perjudicial en los otros miembros del grupo (Palloff & Pratt, 2001).

IV.4.2.5.3 Bloqueadores

Los bloqueadores son aquellos estudiantes que tienden a interferir con el progreso de sus pares y del grupo citando experiencias personales no relacionadas al problema del grupo; hablando de su esposa, el hombre común, de su mascota, etc.; insistiendo en la discusión de temas ya resueltos por el resto del grupo; rechazando ideas sin consideración; o evitando la votación de algún tema a resolver (Jaques & Salmon, 2007).

IV.4.2.5.4 Dominadores

Es preciso que un e-tutor intervenga cuando un estudiante intente dominar la discusión en forma consciente o inconsciente ya que las tareas de aprendizaje son más gratificantes si cada uno tiene la oportunidad de contribuir y nadie domina la discusión (TCOL, 2003; IITT, 2007; Berge, 1995; CSALT ,2001; Salmon, 2007; Llorente Cejudo, 2006; Duggleby *et al*., 2002; Lim & Cheah, 2003; Palloff & Pratt, 2001; Kukulska-Hulme, 2004). Además, el tratar con dominadores contribuye a que los estudiantes menos participativos no se saturen al tener que escucharlos en todo momento (Jaques & Salmon, 2007). Un estudiante domina la discusión inconscientemente cuando suele ser el primero en contribuir y lo hace de modo tal que deja poco para decir a sus pares (Duggleby *et al*., 2002). En este caso podría ser conveniente agradecerle al estudiante por su contribución y decirle que sería bueno escuchar a otros (Duggleby *et al*., 2002; Jaques & Salmon, 2007), o alternativamente se podría solicitarle al estudiante que espere las respuestas de sus pares antes de contribuir (Berge, 1995; CSALT ,2001; Llorente Cejudo, 2006). Finalmente, un alumno intenta dominar la discusión en forma consciente cuando da direcciones autoritariamente, rebaja el rango/status de sus pares, interrumpe contribuciones de otros, etc., (Jaques & Salmon, 2007).

IV.4.2.5.5 Buscadores de reconocimiento

También es necesario tratar con aquellos alumnos que intentan llamar la atención ya sea hablando demasiado, con ideas extremas, haciendo alarde o jactándose (Jaques & Salmon, 2007).

Cuando un e-tutor se enfrenta a un estudiante con alguno de los comportamientos mencionados debe tratar de solicitarle un cambio en su forma de actuar (Lim & Cheah, 2003), pero en caso de no obtener respuesta puede ser conveniente suspender su acceso al curso (Palloff & Pratt, 2001). Con respecto a los estudiantes mirones y no participantes, suele ser también de ayuda el tratar de hacerlos sentir como parte del grupo (Orvis & Lassiter, 2006).

IV.4.2.6 Manejar contribuciones inapropiadas

Para lograr que el tono y contenido de las comunicaciones sean apropiados de modo tal que los estudiantes no se sientan incómodos (Bailey & Luetkehans, 1998; Duggleby *et al*., 2002) se requiere que un e-tutor posea atributos de habilidad vinculados con:

- Actuar contra las críticas: el e-tutor debe prestar atención y desalentar a los estudiantes que tienden a criticar o culpar a otros ya que este tipo de comentarios no es constructivo (Cornelius & Higgison, 2001; Bailey & Luetkehans, 1998; Jaques & Salmon, 2007; Kukulska-Hulme, 2004).
- Actuar contra la intimidación: es necesario mantener el ambiente libre de intimidación porque esto puede llevar a la interrupción del curso (Cornelius & Higgison, 2001; Varvel, 2007).
- Actuar contra la discriminación: es preciso que se excluya cualquier tipo de comportamiento o lenguaje discriminatorio (eTTNet TWG 2, 2005).

Ante cualquiera de las contribuciones inapropiadas antes mencionadas un e-tutor dispone de varias alternativas de acción, entre ellas: indicarle al estudiante que la contribución esta fuera de las reglas de comportamiento aceptadas y recordarle cual es la forma de comportamiento apropiada (TCOL, 2003; McPherson & Baptista Nunes, 2004; Bailey & Luetkehans, 1998), si el mensaje es extremo eliminarlo e informar a los estudiantes las razones de tal determinación (Duggleby *et al*., 2002), editar el mensaje (en el caso de que no sea extremo) e informar las razones de tal determinación (Duggleby *et al*., 2002), y remover al estudiante del curso (CSALT ,2001)

IV.4.2.7 Alabar y modelar el comportamiento de discusión bueno

Es conveniente que un e-tutor elogie y demuestre por medio de ejemplos (modele) los comportamientos de discusión que se espera muestren los estudiantes cuando interactúan *online* (Berge, 1995; Kemshal Bell, 2001; Llorente Cejudo, 2006; Full Circ, 2001; Smith, 2005; Lim & Cheah, 2003; Bonk *et al*., 2004; CSALT ,2001). Así por ejemplo, el dar las gracias a los estudiantes que responden puede ser útil para impulsar cortesía e interacción (Berge, 1995; Llorente Cejudo, 2006).

Si bien en las fuentes consultadas no se deja en claro que se entiende por comportamiento de discusión bueno, en base a lo tratado en apartados anteriores se puede inferir que el comportamiento de los estudiantes se puede calificar como bueno si: cumple con las reglas de comportamiento establecidas para el curso, no se incurre en alguna forma de comportamiento de discusión inapropiado, no se realizan contribuciones inapropiadas, y en general se actúa de forma cortés.

IV.4.2.8 Manejar conflictos

Conviene permitir cierto grado de desacuerdos, controversias, o conflictos dentro de los grupos porque su presencia indica que la dinámica del grupo se está desarrollando satisfactoriamente (Palloff & Pratt, 2001; McWhaw *et al*., 2003; Bailey & Luetkehans, 1998; Jaques & Salmon, 2007). De hecho, un conflicto bien manejado puede volverse una oportunidad para unir al grupo mucho más rápido y dar vía libre en el grupo a más intimidad, expresión emocional, y habilidad de resolver conflictos futuros (Lim & Cheah, 2003). Sin embargo, si el conflicto no se trata de forma adecuada puede destruir la comunidad de aprendizaje, reducir la satisfacción de los estudiantes, reducir el aprendizaje, conducir a una baja retención (ION, 2005; Blackburn, 2004; Varvel, 2007), e incluso impedir que las tareas sean completadas (Palloff & Pratt, 2001). Nada sofoca la participación en un curso *online* como el conflicto no resuelto porque el miedo a ser atacado reduce la participación y el interés en la clase (Palloff & Pratt, 2001). Para hacer frente a los conflictos que surgen entre los estudiantes un e-tutor necesita recurrir a los atributos de habilidad de:

- Facilitar el movimiento de los estudiantes a lo largo del conflicto de modo que puedan desarrollar normas para hacer frente a conflictos sucesivos (Palloff & Pratt, 2001).
- Mediar diferencias entre estudiantes en conflicto por ejemplo señalando temas de interés común o áreas de acuerdo (Kemshal Bell, 2001; Bailey & Luetkehans, 1998; Kukulska-Hulme, 2004).
- Asistir a los miembros del grupo cuando están luchando para lograr consenso (Bailey & Luetkehans, 1998).
- Excluir a personas conflictivas: se puede excluir a uno o más de los estudiantes involucrados en el conflicto de modo que el proceso de aprendizaje continúe para los otros miembros del grupo. En estos casos es conveniente tener una conversación con el resto de los estudiantes para explicar porque se tomó tal decisión (Palloff & Pratt, 2001).

No se debe perder de vista que cuando emerge el conflicto puede ser conveniente observar sin involucrarse ya que los estudiantes generalmente son capaces de tratar con el conflicto y lograr una resolución exitosa. Una intervención temprana del e-tutor con el propósito de sofocar la controversia y las disputas puede desaprovechar la posibilidad de

usar el conflicto para promover el aprendizaje (McWhaw *et al*., 2003). Sin embargo, se debe intervenir si el conflicto crece, interfiere con el progreso de la clase, o se torna en un ataque personal (Palloff & Pratt, 2001; Berge, 1995; Kemshal Bell, 2001; CSALT, 2001; Blackburn, 2004; Bailey & Luetkehans, 1998; Llorente Cejudo, 2006).

IV.4.2.9 Trabajar con sentimientos y sensaciones

Ante la falta de contacto real para humanizar el entorno de aprendizaje un e-tutor tiene que ser capaz tanto de expresar sus sentimientos y sus sensaciones, como de lograr que los estudiantes también expresen sus sentimientos y sus sensaciones (Benito Osorio, 2009; Bláquez Entonado & Alonso Díaz, 2005; Adell & Sales, 1999; Salmon, 2003). Además, debe ser capaz de confrontar conflictos afectivos o emocionales que pudieran surgir entre los estudiantes, ya que el clima afectivo o emocional del grupo puede influir en la efectividad, el procesamiento de información, y la creatividad (Orvis & Lassiter, 2006).

IV.4.2.10 Motivar

Es fundamental que un e-tutor pueda motivar a los estudiantes porque no basta con que los alumnos conozcan como interactuar y dispongan de los recursos tecnológicos, también se necesita que tengan el deseo y predisposición para coordinar sus esfuerzos (Orvis & Lassiter, 2006; Fåhræus, 2000; Goodyear *et al*., 2001; Kemshal Bell, 2001; ION, 2005; Llorente Cejudo, 2006; Bláquez Entonado & Alonso Díaz, 2005; Hathaway *et al.*, 2007; Cornelius & Higgison, 2001; Palloff & Pratt, 2001; Goméz Gutiérrez, 2008). Por otro lado, los estudiantes necesitan que los e-tutores los motiven cuando tienen que hacer un esfuerzo para superar las sensaciones de pérdida y aislamiento que se producen principalmente cuando enfrentan problemas que no dominan y que no pueden solucionar (TCOL, 2003; Llorente Cejudo, 2006). Por último, la motivación influye en el aprendizaje de los estudiantes sobre todo cuando tienen que esforzarse para adquirir habilidades y conocimientos complejos (Kukulska-Hulme, 2004). Los atributos de habilidad que entran en juego a la hora de motivar a los estudiantes son:

- Proveer palabras de aliento y apoyo (Llorente Cejudo, 2006; Packham *et al*., 2004).
- Gratificar a los estudiantes por sus contribuciones ya sea reconociendo o elogiando aportes en forma pública y privada, asignando parte de la nota en función a las contribuciones, etc. (Kemshal Bell, 2001; Varvel, 2007; Cornelius & Higgison, 2001; CSALT, 2001; Palloff & Pratt, 2007; Bonk *et al*., 2004; Teles *et al.*, 2001; Lund, 2004).

- Mostrar confianza en que los estudiantes pueden cumplir con las tareas (Kemshal Bell, 2001; Orvis & Lassiter, 2006).
- Establecer objetivos desafiantes pero que los estudiantes puedan lograr (Shepherd, 1999).

Sub habilidades	Atributos de habilidad
Estimular la interacción entre estudiantes	• Usar técnicas para estimular la interacción. • Mejorar las relaciones. • Promover la interacción sin dominar la discusión. • Discutir formas de facilitar la interacción. • Estimular respuestas mutuas.
Estimular la participación	• Solicitar comentarios sobre tópicos tratados. • Fomentar explicaciones. • Asegurar contribuciones igualitarias. • Cambiar el tema de discusión. • Plantear tópicos o tareas que mejoren el involucramiento. • Regresar preguntas. • Indagar causas de la participación reducida.
Desarrollar, soportar y mantener una comunidad de aprendizaje	• Desarrollar la cohesión grupal. • Promover la colaboración. • Promover el respeto.
Crear un ambiente de aprendizaje agradable	• Dar la bienvenida. • Usar rompe hielos. • Fomentar comentarios sobre presentaciones. • Proteger contra el miedo al ridículo. • Usar un estilo amigable e informal. • Promover la amistad.
Manejar comportamientos de discusión inapropiados	• Tratar con mirones. • Tratar con no participantes. • Tratar con bloqueadores. • Tratar con dominadores. • Tratar con buscadores de reconocimiento.
Manejar contribuciones inapropiadas	• Actuar contra las críticas. • Actuar contra la intimidación. • Actuar contra la discriminación.
Alabar y modelar el comportamiento de discusión bueno	• Alabar comportamientos de discusión bueno. • Modelar comportamientos de discusión buenos.
Manejar conflictos	• Fomentar el desarrollo de normas para manejar conflictos. • Mediar diferencias. • Lograr consenso. • Excluir a personas conflictivas.

Tabla 4.3. Sub habilidades y atributos de habilidad sociales de e-tutores de ACSC

Sub habilidades	Atributos de habilidad
Trabajar con sentimientos y sensaciones	• Expresar sentimientos y sensaciones. • Estimular la expresión de sentimientos y sensaciones. • Manejar conflicto afectivo.
Motivar	• Proveer palabras de aliento y apoyo. • Valorar el esfuerzo. • Mostrar confianza. • Establecer objetivos desafiantes.

Tabla 4.3. (Cont.)

IV.4.3 Habilidades Pedagógicas

Esta categoría de habilidades supone una participación activa del e-tutor en las actividades de un curso, desde el diseño hasta el dictado de clases, a fin de lograr una experiencia educacional efectiva y de alta calidad donde los estudiantes logren los objetivos de aprendizaje del curso (Varvel, 2007; Cornelius & Higgison, 2001; Teles *et al*., 2001; Lund, 2004; Bonk *et al*., 2004). Las sub habilidades que se consideran de carácter pedagógico se comentan a continuación.

IV.4.3.1 Diseño instruccional

La sub habilidad de diseño instruccional puede describirse en términos de los atributos de habilidad de:

- Diseñar el programa de aprendizaje (eTTNet TWG 2, 2005; Goméz Gutiérrez, 2008; Palloff & Pratt, 2007).
- Diseñar estrategias metodológicas que se usarán para lograr los objetivos planteados (Goméz Gutiérrez, 2008).
- Diseñar *feedback*: especificar los tipos de *feedback* que se proveerán, como se suministrarán y cuando (eTTNet TWG 2, 2005).
- Diseñar salidas de aprendizaje apropiadas (Ulster, 2002).
- Diseñar criterios, métodos, estrategias y recursos para evaluar a los estudiantes y al programa de aprendizaje (Goodyear *et al*., 2001; Ulster, 2002; Barker, 2002; Smith, 2005; eTTNet TWG 2, 2005).

IV.4.3.2 Creación de actividades de aprendizaje

Las tareas o actividades demasiado obvias y que carecen de ambigüedad atentan contra la colaboración al no dejar espacio para el cuestionamiento, la negociación, la explicación, y la argumentación (Lim & Cheah, 2003). En consecuencia, para que las

tareas sirvan de soporte a la colaboración es necesario que un e-tutor posea los atributos de habilidad de:

- Crear tareas que estimulen a los estudiantes a comprometerse activamente en los diálogos (Lim & Cheah, 2003; Lund, 2004).
- Crear tareas de aprendizaje relevantes al tópico y a las salidas de aprendizaje deseadas (Goodyear *et al*., 2001; Kemshal Bell, 2001; Ulster, 2002; CSALT ,2001).
- Crear tareas novedosas y desafiantes pero acordes con las capacidades de los estudiantes (Ulster, 2002; Kukulska-Hulme, 2004).
- Crear tareas relevantes a los intereses personales y a las experiencias de los estudiantes de modo tal de motivarlos (Berge, 1995; Llorente Cejudo, 2006; CSALT ,2001; eTTNet TWG 2, 2005; Kukulska-Hulme, 2004; Ulster, 2002).
- Crear tareas consistentes con las restricciones y capacidades tecnológicas (Goodyear *et al*., 2001).
- Crear tareas con escala de tiempo apropiada (Goodyear *et al*., 2001; Ulster, 2002).

IV.4.3.3 Creación de recursos de aprendizaje

Los atributos de habilidad de creación de recursos de aprendizaje (glosarios, ejemplos, enlaces, lecturas, resúmenes, crucigramas, cuestionarios, videos, etc.) son indispensables, ya que las características propias de la enseñanza *online* impiden que se utilicen los recursos usados en clases presenciales (Goméz Gutiérrez, 2008). La creación de los recursos de aprendizaje se vincula con los atributos de habilidad especificados a continuación:

- Seleccionar contenidos (Adell & Sales, 1999).
- Organizar y secuenciar los contenidos (señalizar, vincular, y separar en unidades o módulos) de modo tal de que los estudiantes puedan concentrarse en los aspectos que están dentro de su capacidad (Goodyear *et al.*, 2001; Barker, 2002; Goméz, 2008).
- Diseñar recursos especificando formato, contenido, y cualquier otro detalle necesario (Adell & Sales, 1999; eTTNet TWG 2, 2005; Ulster, 2002).
- Desarrollar los recursos de aprendizaje en base a problemas y escenarios del mundo real para que estos resulten significativos para los estudiantes (Berge, 1995; Duggleby *et al*., 2002; Hootstein, 2002; eTTNet TWG 2, 2005; Palloff & Pratt, 2007; Llorente Cejudo, 2006). Cuanto mayor sea la relación que puedan establecer

los alumnos entre sus experiencias de vida y conocimientos con el contexto de la clase *online*, mayor será su compromiso en el proceso de aprendizaje y su entendimiento de lo que aprenden (Palloff & Pratt, 2007).

- Seleccionar recursos apropiados a cada curso de modo que conduzcan a salidas de aprendizaje efectivas (Adell & Sales, 1999; eTTNet TWG 2, 2005; Varvel, 2007).

IV.4.3.4 Facilitar la comprensión del material de aprendizaje

Para facilitarle a los estudiantes el entendimiento del contenido del curso un e-tutor precisa de los atributos de habilidad de:

- Ayudar a los estudiantes a identificar puntos y conceptos clave centrando la discusión en ellos (Berge, 1995; Mason, 1991; Llorente Cejudo, 2006; CSALT ,2001; Benito Osorio, 2009; Blackburn, 2004; Adell & Sales, 1999) mediante estrategias tales como las preguntas de final abierto (que conducen a respuestas diferentes a un sí o no) (Kemshal Bell, 2001).
- Focalizar la discusión en principios y habilidades (Berge, 1995; Llorente Cejudo, 2006; CSALT ,2001; Benito Osorio, 2009; Blackburn, 2004). De acuerdo a (Kemshal Bell, 2001) el uso de preguntas de final abierto puede ser útil para enfocar las respuestas de los estudiantes en principios y habilidades.
- Presentar la información en una manera que ayude a los estudiantes en su interpretación, entendimiento, e internalización (Guasch *et al*., 2009; Shepherd, 1999; Varvel, 2007).
- Guiar a los estudiantes en la exploración del material del curso, y del material relacionado (links, listas de lectura, puntos de comienzo para investigación, etc.), para ayudarlos a organizar sus ideas (TCOL, 2003; Ulster, 2002; Duggleby *et al*., 2002; Goodyear *et al*., 2001; Salmon, 2003; Palloff & Pratt, 2007).
- Explicar y clarificar los contenidos cuando haya malos entendidos (Olivares, 2007; IITT, 2007; Varvel, 2007; Llorente Cejudo, 2006; Benito Osorio, 2009; ION, 2005).
- Ayudar a los estudiantes a establecer relaciones entre los contenidos nuevos y sus conocimientos previos (Hootstein, 2002).
- Ayudar a los estudiantes a analizar el contenido del curso, a hacer conexiones entre el contenido y sus objetivos de aprendizaje, y a entender las aplicaciones potenciales del contenido (TCOL, 2003; Lim & Cheah, 2003).

- Responder por medio de analogías, ejemplos, y explicaciones a las preguntas de los estudiantes concernientes a los contenidos del curso (IITT, 2007; Chang, 2004; Lim & Cheah, 2003). No obstante, un e-tutor no debería proveer todas las respuestas ya que los estudiantes se benefician cuando responden las preguntas por su propia cuenta y comparten su conocimiento (TCOL, 2003).
- Reforzar las contribuciones de los estudiantes proporcionando conocimientos avanzados y diferentes perspectivas (de Laat, 2006; Goodyear *et al*., 2001; Lim & Cheah, 2003).
- Realizar preguntas de final abierto para estimular el pensamiento crítico y la reflexión sobre los tópicos en discusión (Teles *et al*., 2001; Lund, 2004; Palloff & Pratt, 2007; Lim & Cheah, 2003).

IV.4.3.5 Flexibilidad

Un e-tutor flexible puede responder ante diversas necesidades, modalidades y estilos de aprendizaje de los estudiantes (Berge, 1995; Varvel, 2007; Llorente Cejudo, 2006; CSALT ,2001; Kukulska-Hulme, 2004). Para que un docente sea considerado flexible tendría que contar con los atributos de habilidad siguientes:

- Construir el curso teniendo en cuenta diferentes estilos de aprendizaje (Palloff & Pratt, 2001).
- Utilizar diversas estrategias como ser debate, investigación, juego de roles y simulaciones (Kemshal Bell, 2001).
- Adaptar el contenido y las actividades a las necesidades, los estilos y las modalidades de aprendizaje de los estudiantes (Duggleby *et al*., 2002; ION, 2005; Varvel, 2007; Gunn, 2001; Blackburn, 2004).
- Modificar las actividades cuando no contribuyan a una salida de aprendizaje positiva, o cuando no resulten apropiadas para la colaboración (Kukulska-Hulme, 2004; Palloff & Pratt, 2007).
- Adaptar recursos de aprendizaje (Blackburn, 2004; eTTNet TWG 2, 2005).
- Utilizar medios de comunicación (e-mail, foros, chat, etc.) acordes con las necesidades del curso y de los estudiantes (Varvel, 2007; Palloff & Pratt, 2001)
- Usar estrategias y enfoques de enseñanza para incluir a todos los estudiantes (Jaques & Salmon, 2007; Gunn, 2001).
- Ajustar las estrategias y los enfoques de enseñanza en función al contenido, las necesidades educacionales, y las inquietudes de los estudiantes (ION, 2005;

Salmon, 2003; Gunn, 2001; Palloff & Pratt, 2001; Fåhræus, 2000; Ulster, 2002; Varvel, 2007).

- Adaptarse ante cambios en los contextos de enseñanza, métodos, y audiencias (Salmon, 2003).

IV.4.3.6 Promover el debate/ la discusión

En los entornos de *e-learning* el debate o discusión juega un papel fundamental ya que para los alumnos los apuntes o resúmenes de las clases presenciales subidas al sitio del curso no son más que un articulo adicional para leer (Palloff & Pratt, 2007). Además el debate contribuye a incrementar la unión de los estudiantes y reforzar el progreso individual (Llorente Cejudo, 2006). Los atributos de habilidad vinculados con promover el debate son:

- Impulsar a los estudiantes a reflexionar sobre su postura presentando perspectivas conflictivas, opuestas, y alternativas que pueden conducir a debates (Berge, 1995; TCOL, 2003; Varvel, 2007; Llorente Cejudo, 2006; Lim & Cheah, 2003).
- Refinar las preguntas realizadas, o plantear otras, para estimular el debate cuando se detecta que el nivel de discusión se viene reduciendo progresivamente (Palloff & Pratt, 2001).
- Usar mensajes que impulsen la discusión en dos o más temas a la vez (Cornelius & Higgison, 2001).
- Apoyar y valorar: un ingrediente clave para la discusión efectiva es la creación de una atmosfera de apertura y confianza donde cada persona sea valorado por lo que es y no tema hacer el ridículo (Jaques & Salmon, 2007).

IV.4.3.7 Gestionar el debate

Una vez generado el debate es necesario que el e-tutor lo gestione para luchar contra la tendencia de los estudiantes y del grupo a desviar la discusión desde su tópico original (Lim & Cheah, 2003). A la hora de gestionar las discusiones, para sacar provecho de las mismas y que no sean un desperdicio de tiempo, un e-tutor precisa de atributos de habilidad para:

- Guiar la discusión manteniéndola con los objetivos y metas de la clase (Mason, 1991; Goodyear *et al*., 2001; de Laat, 2006; Palloff & Pratt, 2007).
- Prevenir que las discusiones se salgan de su curso, ya sea reestructurando actividades, dando información adicional, formulando preguntas clarificadoras y

expansivas, reformulando preguntas realizadas, o realizando resúmenes (Palloff & Pratt, 2001; IITT, 2007; Mason, 1991; Hathaway *et al.*, 2007; Smith, 2005; Duggleby *et al*., 2002; Lim & Cheah, 2003; Palloff & Pratt, 2007). La necesidad de dar información adicional surge debido a que muchas veces los estudiantes no reciben toda la información o comienzan en la dirección errónea y en consecuencia es necesario aclarar que se espera de ellos (Duggleby *et al*., 2002).

- Reencaminar discusiones cuando avancen en la dirección errónea (Palloff & Pratt, 2001; IITT, 2007).
- Dar ritmo y coherencia a la discusión (Benito Osorio, 2009; Salmon, 2003).
- Cerrar discusiones no productivas o que ya no despiertan interés (Salmon, 2003; Shepherd, 1999; Full Circ, 2001).
- Cerrar la discusión cuando se llegue a una conclusión y no se requieren contribuciones posteriores (Duggleby *et al*., 2002; Lim & Cheah, 2003).
- Guiar a los estudiantes en los detalles técnicos de la discusión indicando por ejemplo la longitud adecuada de cada mensaje, cómo justificar las afirmaciones, y cómo estructurar cada mensaje (Lim & Cheah, 2003; Bonk *et al*., 2004; Teles *et al*., 2001; Lund, 2004).
- Dirigir a los estudiantes a diferentes tópicos y carpetas para contribución (Bonk *et al*., 2004; Teles *et al.*, 2001; Lund, 2004).

IV.4.3.8 Vincular

Cuando los estudiantes realizan un número considerable de contribuciones, las sub habilidades de resumir y de vincular son fundamentales (Hathaway *et al.*, 2007). El vincular o entrelazar consiste en reunir varias contribuciones de los estudiantes para proporcionar una perspectiva unificada, y una síntesis momentánea, que sirva como punto de comienzo de una nueva ronda de debate (Cornelius & Higgison, 2001; Mason, 1991).

Al momento de vincular contribuciones es conveniente que un e-tutor cuente con los atributos de habilidad de:

- Reunir fragmentos de diferentes mensajes y presentarlos con una nueva perspectiva (Salmon, 2007; Jaques & Salmon, 2007).
- Señalar un asunto o tópico de una discusión que se vincula con otros (Jaques & Salmon, 2007);
- Entrelazar comentarios y conceptos dispares para permitirle a los estudiantes lograr un sentido de cumplimiento y dirección (Mason, 1991; Bonk *et al*., 2004).

- Vincular varias ramas de conversación a fin de impulsar a los estudiantes a reflexionar sobre la discusión y a continuar tratando el tópico más a fondo (Berge, 1995; Llorente Cejudo, 2006; CSALT ,2001).
- Enfocar a los estudiantes en las cuestiones clave de la discusión para impulsar contribuciones finales antes de resumir (Jaques & Salmon, 2007; Lim & Cheah, 2003; Teles *et al*., 2001; Bonk *et al*., 2004).

IV.4.3.9 Resumir

Si bien las sub habilidades de resumir y de vincular se relacionan estrechamente, la diferencia principal reside en que los resúmenes suelen plantearse al finalizar las discusiones o clases, mientras que se pueden realizar vinculaciones varias veces y en diferentes momentos durante una discusión. Los atributos de habilidad relacionados con hacer resúmenes son:

- Resumir los puntos clave de una discusión identificando temas unificadores y señalando desacuerdos (IITT, 2007; Mason, 1991; Goodyear *et al*., 2001; Salmon, 2003; Benito Osorio, 2009; Blackburn, 2004; Lim & Cheah, 2003; de Laat, 2006; Lund, 2004; Bonk *et al*., 2004).
- Subrayar contribuciones de los estudiantes que añaden ideas frescas o miran al tópico de una manera interesante (Salmon, 2007; Jaques & Salmon, 2007).
- Añadir comentarios o críticas y señalar omisiones, otras perspectivas o aplicaciones (Salmon, 2007; Jaques & Salmon, 2007).
- Resumir el material de aprendizaje (Berge, 1995; Llorente Cejudo, 2006).
- Ayudar a los estudiantes a sintetizar el contenido (Lim & Cheah, 2003).
- Ayudar a los estudiantes a sintetizar sus ideas relacionando diferentes contribuciones (CSALT ,2001).

IV.4.3.10 Asistir a los estudiantes en la gestión de su propio aprendizaje

Los atributos de habilidad que contribuyen a que los alumnos asuman responsabilidad por su propio aprendizaje y se organicen siguiendo sus propias reglas, son:

- Proporcionar tareas que permitan explorar, investigar, y trabajar colaborativamente (Palloff & Pratt, 2001).
- Auxiliar a los estudiantes en la identificación de sus necesidades y estilos de aprendizaje (eTTNet TWG 2, 2005).

- Asistir a los alumnos en la articulación de sus necesidades y preocupaciones de aprendizaje (Goodyear *et al*., 2001; Jaques & Salmon, 2007).
- Ayudar a los alumnos en la articulación de sus expectativas (Goodyear *et al*., 2001).
- Brindar asistencia a los estudiantes en la selección de oportunidades de aprendizaje (eTTNet TWG 2, 2005).
- Ayudar a los estudiantes a desarrollar y revisar sus planes o trayectorias futuras de aprendizaje (eTTNet TWG 2, 2005).

IV.4.3.11 Gestionar el acceso de los estudiantes a los recursos de aprendizaje

Para gestionar el acceso de los estudiantes a los recursos de aprendizaje es conveniente que un e-tutor posea los atributos de habilidad de:

- Referir a los estudiantes a recursos disponibles: si el e-tutor responde detalladamente a cada pregunta de los estudiantes existe el peligro de que se vuelvan dependientes (Goodyear *et al*., 2001; Salmon, 2003; Shepherd, 1999). En este sentido, una base de datos de las comunicaciones con los estudiantes y las respuestas suministradas puede servir para reducir parte de la carga de trabajo porque se podría direccionar a los estudiantes a los comentarios realizados cuando se vuelva a tratar un mismo tema (Lund, 2004).
- Ayudar a los estudiantes a localizar recursos específicos en el ambiente de *e-learning* (Chang, 2004).
- Impulsar a los estudiantes a consultar recursos y fuentes de información adicionales (librerías universitarias, *e-journals*¸ sitios web sobre los contenidos estudiados, etc.) o expertos en el campo de manera tal que tengan más y mejor conocimiento del material y los asuntos discutidos (Chang, 2004; Bonk *et al*., 2004; Teles *et al*., 2001; Lund, 2004; Palloff & Pratt, 2007; Lim & Cheah, 2003; TCOL, 2003).

IV.4.3.12 Desarrollar las habilidades de aprendizaje de los estudiantes

Una de las tareas de un e-tutor es que los estudiantes desarrollen y apliquen habilidades tanto específicas de cada materia como genéricas (TCOL, 2003; Ulster, 2002). Los atributos de habilidad que se precisan para esto son:

- Ayudar a los estudiantes a desarrollar las habilidades de organización de conceptos y de desarrollo de mapas mentales de modo que puedan estructurar su aprendizaje en una manera que tenga sentido para ellos (TCOL, 2003).

- Asistir a los estudiantes en el desarrollo de las habilidades para articular ideas en forma escrita o verbal, y debatirlas productivamente (TCOL, 2003).
- Ayudar a los estudiantes a desarrollar la habilidad de resumir (Hathaway *et al.*, 2007).
- Fomentar el desarrollo de las habilidades de colaboración de manera que los estudiantes puedan lograr objetivos de aprendizaje a través de interacciones con sus pares (TCOL, 2003). Las habilidades de colaboración son imprescindibles en los entornos de ACSC dado que sin estas habilidades o con ellas desarrolladas a niveles inadecuados, los estudiantes son incapaces de: compartir y discutir sus ideas y opiniones; argumentar, justificar y negociar sus puntos de vista; construir conocimiento a través de la discusión, la reflexión y la toma de decisiones; etc. (Costaguta, 2008).
- Ayudar a los estudiantes a desarrollar la habilidad de administración de discusiones (TCOL, 2003).
- Demostrar por medio de ejemplos las habilidades de aprendizaje que deberían poseer los estudiantes (Goodyear *et al*., 2001; Ulster, 2002; Shepherd, 1999).
- Descomponer habilidades complejas en unidades pequeñas de modo tal que los estudiantes puedan desarrollar progresivamente aquellas habilidades que están dentro de sus capacidades (Ulster, 2002).

IV.4.3.13 Monitorear el proceso de aprendizaje

Para ayudar a los estudiantes a lograr un aprendizaje exitoso un e-tutor tendría que ser capaz de monitorear la comprensión así como los malentendidos. Para ello se requiere que el e-tutor pueda: identificar cuáles son las áreas de contenido que los estudiantes han entendido bien y cuáles han quedado menos claras (TCOL, 2003; Salmon, 2003; Olivares, 2007), determinar las razones por las que se produjeron interpretaciones erróneas (Olivares, 2007), y aconsejar a los estudiantes sobre la trayectoria que están siguiendo (Goodyear *et al*., 2001; Duggleby *et al*., 2002;Bonk *et al.*, 2004; Teles *et al*., 2001; Lund, 2004). El monitoreo de la comprensión puede hacerse mediante preguntas de clarificación, como ser, "¿Puedes explicar eso de otra manera?", "¿Qué quieres decir con…?", y "¿Puedes darme un ejemplo?" (Jaques & Salmon, 2007).

IV.4.3.14 Resolver problemas de aprendizaje

Los e-tutores tienen que identificar problemas que obstaculizan el proceso de aprendizaje de los estudiantes para brindar asistencia en el momento adecuado y con los métodos apropiados (Varvel, 2007). Entre las dificultades de aprendizaje que son de incumbencia para los docentes se tiene: conocimiento insuficiente en el área de contenido, falta de acceso a recursos apropiados, carencia de conocimiento en cómo usar recursos, déficit de habilidades, y problemas con los materiales del curso (TCOL, 2003).

IV.4.3.15 Familiarizar a los estudiantes con el ambiente *online*

Uno de los aspectos de la labor pedagógica de un e-tutor comprende familiarizar a los estudiantes con el ambiente de *e-learning* y las prácticas de trabajo esperadas (Goodyear *et al.*, 2001; ION, 2005; Blackburn, 2004). Pero también se requiere que familiarice a los estudiantes con las convenciones de la disciplina, esto es con la forma en que las ideas son desarrolladas y presentadas, estrategias aceptables e inaceptables para la investigación, y requerimientos de escritura y reporte (TCOL, 2003).

IV.4.3.16 Marcar el ritmo del proceso de aprendizaje

El e-tutor debe marcar el ritmo del proceso de aprendizaje de modo que no queden estudiantes retrasados (Salmon, 2007; de Laat, 2006). Además, muchas veces se necesita que los alumnos incorporen ciertos conocimientos antes de que se les presenten nuevos temas o contenidos (Varvel, 2007). Para mantener a los estudiantes avanzando a un ritmo confortable un e-tutor debería contar con los siguientes atributos de habilidad:

- Hacer reflexionar a los estudiantes que aprenden rápidamente y aquellos a los que les cuesta aprender sobre las consecuencias que esto tiene (Salmon, 2007).
- Impulsar a los estudiantes avanzados a recordar los temas o discusiones que han completado y proveer respuestas o sugerencias a los estudiantes que quedaron retrasados (Salmon, 2007).
- Resumir con frecuencia de modo de permitirle a los estudiantes ponerse al día rápidamente cuando ingresan al sistema (Salmon, 2007).

Habilidades	Sub habilidades
Diseño instruccional	• Diseñar el programa de aprendizaje. • Diseñar estrategias metodológicas. • Diseñar *feedback*. • Diseñar salidas de aprendizaje apropiadas. • Diseñar criterios, métodos, estrategias, y recursos de evaluación.
Creación de actividades de aprendizaje	• Crear actividades que estimulen a los estudiantes a comprometerse activamente en discusiones. • Crear actividades de aprendizaje relevantes al tópico y salida de aprendizaje deseada. • Crear actividades de aprendizaje que sean desafiantes. • Crear actividades relevantes a los intereses personales y experiencias de los estudiantes. • Crear actividades consistentes con las restricciones y capacidades tecnológicas. • Crear actividades con escala de tiempo apropiada.
Creación de recursos de aprendizaje	• Selección de contenidos. • Estructurar el contenido disponible para los estudiantes. • Diseñar recursos de aprendizaje. • Desarrollar recursos de aprendizaje basado en problemas y escenarios del mundo real. • Seleccionar recursos de aprendizaje.
Facilitar la comprensión del material de aprendizaje	• Centrar la discusión en puntos y conceptos claves. • Focalizar la discusión en principios y habilidades. • Presentar la información en una manera que ayude a su interpretación. • Guiar en la exploración del material del curso. • Explicar y clarificar los contenidos. • Ayudar a los estudiantes a conectar el contenido con el conocimiento previo. • Ayudar a los estudiantes a hacer conexiones entre el contenido del curso y sus objetivos de aprendizaje. • Responder las preguntas de los estudiantes concernientes a los contenidos del curso. • Reforzar contribuciones de los estudiantes. • Estimular el pensamiento crítico y reflexión.
Promover el debate/discusión	• Impulsar a los estudiantes a reflexionar sobre su postura. • Refinar preguntas para estimular el debate. • Usar mensajes que impulsen la discusión varios temas. • Apoyar y valorar.

Tabla 4.4. Habilidades y sub habilidades pedagógicas de e-tutores de ACSC

Habilidades	Sub habilidades
Gestionar el debate	• Guiar la discusión manteniéndola con los objetivos y metas de la clase. • Prevenir que las discusiones se salgan de su curso. • Reencaminar discusiones fuera de curso. • Dar ritmo y coherencia a la discusión. • Cerrar hilos de discusión no-productivos. • Cerrar la discusión cuando se llegue a una conclusión. • Guiar a los estudiantes en los detalles técnicos de la discusión. • Dirigir a los estudiantes a diferentes tópicos y carpetas para contribución.
Flexibilidad	• Construir el curso considerando diferentes estilos de aprendizaje. • Usar actividades diversas. • Adaptar el contenido y las actividades a las necesidades y los estilos de aprendizaje. • Modificar actividades no apropiadas para la colaboración. • Adaptar recursos de aprendizaje. • Utilizar medios de comunicación acordes con las necesidades del curso y los estudiantes. • Usar estrategias de enseñanza para incluir a todos los estudiantes. • Adaptar las estrategias de enseñanza en función al contenido y las necesidades educacionales. • Adaptarse ante cambios en los contextos de enseñanza, métodos, y audiencias.
Vincular	• Reunir fragmentos de diferentes mensajes y presentarlos con una nueva perspectiva. • Señalar asuntos o tópicos vinculados. • Entrelazar comentarios y conceptos dispares. • Vincular ramas de conversación en una. • Subrayar cuestiones clave de las discusiones.
Resumir	• Resumir los puntos clave de una discusión. • Subrayar contribuciones de los estudiantes que añaden ideas frescas. • Añadir comentarios o críticas. • Resumir el material de aprendizaje. • Ayudar a los estudiantes a sintetizar el contenido. • Ayudar a los estudiantes a sintetizar sus ideas.
Asistir a los estudiantes en la gestión de su propio aprendizaje	• Proporcionar tareas que permitan trabajar colaborativamente. • Auxiliar a los estudiantes a identificar sus necesidades y estilos de aprendizaje.

Tabla 4.4. (Cont.)

Habilidades	Sub habilidades
Asistir a los estudiantes en la gestión de su propio aprendizaje	• Asistir a los estudiantes en la articulación de sus necesidades y preocupaciones de aprendizaje. • Ayudar a los estudiantes a articular sus expectativas. • Asistir a los estudiantes en la selección de oportunidades de aprendizaje. • Ayudar a los estudiantes a desarrollar sus planes de aprendizaje.
Gestionar el acceso de los estudiantes a los recursos de aprendizaje	• Referir a los estudiantes a los recursos disponibles. • Ayudar a los estudiantes a localizar recursos. • Impulsar a los estudiantes a consultar recursos y fuentes de información adicionales.
Desarrollar las habilidades de aprendizaje de los estudiantes	• Ayudar a los estudiantes a desarrollar la habilidad de organización de conceptos. • Ayudar a los estudiantes a desarrollar las habilidades para articular ideas y debatirlas productivamente. • Ayudar a los estudiantes a desarrollar la habilidad de resumir. • Fomentar el desarrollo de las habilidades de colaboración de los estudiantes. • Ayudar a los estudiantes a desarrollar sus habilidades de administración de discusiones. • Demostrar las habilidades mediante ejemplos. • Descomponer habilidades complejas en unidades pequeñas.
Monitorear el proceso de aprendizaje	• Monitorear la comprensión así como malentendidos. • Determinar razones de la interpretación errónea. • Aconsejar a los estudiantes sobre la trayectoria que están siguiendo.
Resolver problemas de aprendizaje	• Identificar problemas que bloquean el aprendizaje de los estudiantes. • Brindar asistencia a los estudiantes que enfrenten problemas de aprendizaje.
Familiarizar a los estudiantes con el ambiente *online*	• Familiarizar a los estudiantes con el ambiente y las prácticas de trabajo esperadas. • Familiarizar a los estudiantes con las convenciones de la disciplina.
Marcar el ritmo del proceso de aprendizaje	• Hacer reflexionar a los estudiantes sobre las consecuencias de su facilidad o dificultad para aprender. • Impulsar a los estudiantes avanzados a recordar los temas completados y proveer respuestas a los estudiantes retrasados. • Resumir con frecuencia.

Tabla 4.4. (Cont.)

IV.4.4 Habilidades Técnicas

Las habilidades técnicas son algo propio de los ambientes de *e-learning* por lo cual no se observan en aulas presenciales (chang, 2004). Con estas habilidades se pretende lograr que la tecnología sea transparente para los alumnos de modo que ellos se concentren sólo en sus tareas (Berge, 1995; Hootstein, 2002; Benito Osorio, 2009; Cornelius & Higgison, 2001). Las cuestiones técnicas merecen atención porque si los estudiantes tienen problemas para comunicarse, o no saben cómo usar la tecnología, las actividades grupales se verán perjudicadas (Bonk *et al*., 2004; Olivares, 2007). El quehacer técnico de un e-tutor comprende cuatro aspectos: seleccionar tecnología, usar herramientas de software, proveer soporte a los alumnos ante los problemas técnicos, y desarrollar las habilidades de los estudiantes para usar la tecnología.

IV.4.4.1 Seleccionar la tecnología

En conjunción con expertos técnicos, con las políticas y con los procedimientos de la organización, el e-tutor toma o ayuda a tomar decisiones tecnológicas que mejoren el ambiente disponible para los estudiantes (IITT, 2007; Reid, 2003; Goodyear *et al*., 2001). Al momento de decidir sobre la tecnología, un e-tutor requiere de los atributos de habilidad de: entender las capacidades y limitaciones de la tecnología disponible (Goodyear *et al*., 2001; Salmon, 2003), seleccionar medios tecnológicos concordantes con las salidas y los objetivos de aprendizaje perseguidos (Goodyear *et al*., 2001; Bonk *et al*., 2004), y seleccionar herramientas para desarrollar recursos de aprendizaje acordes con sus habilidades y con la complejidad de los recursos (eTTNet TWG 2, 2005).

IV.4.4.2 Usar la tecnología

Es preciso que un e-tutor pueda usar de manera efectiva y eficiente la tecnología de computadoras y los programas de software que típicamente son requeridos en *e-learning* para mejorar la enseñanza y el aprendizaje, la productividad personal, y la administración de información (Goodyear *et al*., 2001; Guasch *et al*. ,2009; Varvel, 2007; Reid, 2003; Barker, 2002; Llorente Cejudo, 2006; Smith, 2005). Un e-tutor que se siente confortable con la tecnología usada puede transferir ese mismo nivel de confort a los estudiantes y hacer de la tecnología algo transparente para ellos (Palloff & Pratt, 2007). En relación al uso de la tecnología se pueden distinguir los siguientes atributos de habilidad:

- Usar herramientas de comunicación sincrónica (chat, conferencias, y video conferencias) y asincrónica (e-mail, pizarrones de anuncios, y foros) (ION, 2005;

Kemshal Bell, 2001; Salmon, 2003; Barker, 2002; Chang, 2004; Llorente Cejudo, 2006).

- Desarrollar y editar páginas web (ION, 2005; Kemshal Bell, 2001; Barker, 2002).
- Utilizar procesadores de texto, hojas de cálculo y bases de datos (Barker, 2002).
- Editar y actualizar el contenido en sistemas de administración de curso (Goodyear *et al*., 2001; Chang, 2004; ION, 2005; Llorente Cejudo, 2006).
- Utilizar diversos recursos (calendario, grupos, wiki, journal, etc.) de los sistemas de administración de curso (ION, 2005).

IV.4.4.3 Proveer soporte técnico

Cuando el soporte técnico no se proporciona como un servicio independiente es el e-tutor quien ayuda a los estudiantes a enfrentar los problemas técnicos porque los alumnos suelen tener dificultades y sentirse perdidos en las primeras semanas de clase hasta que se acostumbran a usar la tecnología (Duggleby *et al*., 2002; Llorente Cejudo, 2006; Palloff & Pratt, 2001). El proveer soporte técnico a los estudiantes implica los atributos de habilidad de: resolver preguntas técnicas rápida y eficientemente (Berge, 1995; IITT, 2007); guiar a los estudiantes que experimentan problemas técnicos (Varvel, 2007; Lund, 2004; Palloff & Pratt, 2007; Ulster, 2002; Chang, 2004; Goodyear *et al*., 2001), y contactar a los estudiantes con el personal de soporte técnico cuando no se pueden resolver sus problemas (Palloff & Pratt, 2001; Lund, 2004; Palloff & Pratt, 2007; Chang, 2004).

IV.4.4.4 Desarrollar habilidades técnicas de los estudiantes

El e-tutor debería asistir a los alumnos en volverse usuarios competentes y confortables de las herramientas de software disponibles en el entorno de *e-learning* (Berge, 1995; Llorente Cejudo, 2006; Teles *et al*., 2001; Bonk *et al*., 2004; Lund, 2004; Palloff & Pratt, 2007). Para lograr este cometido son fundamentales los atributos de habilidad que permitan:

- Determinar cuan familiarizados y confortables están los estudiantes con la tecnología ya que muchos estudiantes pueden no saber utilizar o tener dificultades con los paquetes de software empleados para interactuar y comunicarse en entornos de ACSC (Olivares, 2007; Williams & Roberts, 2002).
- Ayudar a los estudiantes a desarrollar sus habilidades técnicas fundamentales (Chang, 2004; Lefoe *et al*., 2002; Gunn, 2001; eTTNet TWG 2, 2005).

- Mostrar mediante ejemplos (modelar) prácticas éticas y legales relacionadas al uso de tecnología (Varvel, 2007).

Habilidades	Sub habilidades
Seleccionar la tecnología	• Entender las capacidades y limitaciones de la tecnología disponible. • Seleccionar medios tecnológicos concordantes con las salidas y objetivos de aprendizaje. • Seleccionar herramientas para desarrollar recursos de aprendizaje.
Usar la tecnología	• Usar herramientas de comunicación sincrónica y asincrónica. • Desarrollar páginas web. • Utilizar procesadores de texto, hojas de cálculo y bases de datos. • Editar el contenido en sistemas de administración de curso. • Utilizar recursos de los sistemas de administración de curso.
Proveer soporte técnico	• Resolver preguntas técnicas. • Guiar a los estudiantes con problemas técnicos. • Contactar a los estudiantes con el personal de soporte técnico.
Desarrollar habilidades técnicas de los estudiantes	• Determinar cuan familiarizados están los estudiantes con la tecnología. • Ayudar a los estudiantes a desarrollar sus habilidades técnicas. • Modelar prácticas éticas y legales sobre el uso de tecnología.

Tabla 4.5. Habilidades y sub habilidades técnicas de e-tutores de ACSC

IV.4.5 Habilidades de comunicación

La comunicación *online* es claramente diferente de la comunicación cara a cara ya que el lenguaje corporal, movimiento de las manos, señales visuales y tono de voz están ausentes en los mensajes *online* (Dewar & Whittington, 2000). Las habilidades de comunicación son esenciales para lograr una comunicación efectiva con y entre estudiantes porque en los ambientes *online* la palabra escrita puede ser malinterpretada (Packham *et al*., 2004; Hathaway *et al.*, 2007; Fåhræus, 2000). Un e-tutor es usualmente catalogado como infectivo cuando incurre en prácticas de comunicación inefectivas y limitadas como ser: respuesta demorada o falta de respuesta a las comunicaciones por parte del alumno; respuesta breve, que no provee beneficios a los estudiantes; respuesta tardía, después que

los estudiantes se han desplazado a otras actividades; y acceso irregular al entorno de *e-learning* para comunicarse con los estudiantes (Packham *et al*., 2004).

IV.4.5.1 Preguntar

Las preguntas le pueden servir a un e-tutor al menos para cuatro propósitos en una discusión: testear el conocimiento de los estudiantes, clarificar información, estimular a los estudiantes a expresar sus ideas y construir argumentos, y estimular a los estudiantes a relacionar sus pensamientos con los de sus compañeros. Frecuentemente una misma pregunta puede cumplir con todos los propósitos anteriores, todo depende del clima y el mensaje subyacente en la pregunta (Jaques & Salmon, 2007).

Los cuatro propósitos anteriores revelan la naturaleza multifacética de la habilidad de preguntar. El testeo del conocimiento puede ser visto como vinculado a las habilidades de evaluación, la clarificación de la información es una cuestión pedagógica que favorece la comprensión del contenido, el estimular a los estudiantes a expresar sus ideas tiene componentes sociales en la medida que contribuye a la participación, y finalmente el fomentar la compartición de pensamientos es una cuestión social puesto que ayuda a la interacción entre estudiantes.

En base a lo planteado en los dos párrafos precedentes queda en evidencia que la habilidad de preguntar no puede ser catalogada como algo exclusivo del rol de comunicación de un e-tutor, sino que esta habilidad tiene incidencia en otros roles. Es por ello que, a continuación se decidió dedicar un apartado a cada rol donde la habilidad de realizar preguntas interviene.

IV.4.5.1.1 Preguntas relevantes para el ámbito social

El uso de preguntas de final abierto promueve la participación dado que pueden incitar a respuestas variadas (Lim & Cheah, 2003; Jaques & Salmon, 2007; Full Circ, 2001; Cornelius & Higgison, 2001). Algunas situaciones en las que el uso de preguntas se encuentra dentro de la esfera de las habilidades sociales son:

- Mirar alrededor del grupo: en las clases presenciales el tutor no solo debe prestar atención al estudiante que habla, sino que tiene que hacer un paneo a toda la clase porque esto contribuye a que los estudiantes presten atención a sus pares y no se crea un patrón de comunicación uno a uno. El habito de hacer un paneo de los estudiantes puede trasladarse a clases *online* mediante preguntas del tipo “gracias

por tan valiosa contribución, xxx. ¿Pensamientos/respuestas/reacciones, gente?" (Jaques & Salmon, 2007).

- Regresar las preguntas: un e-tutor que responde a todas las preguntas de sus estudiantes se coloca a sí mismo en una posición de superioridad, para evitar esto puede ser mejor hacerle a los estudiantes las mismas preguntas que ellos plantean. Es usualmente mejor dejar que los alumnos formulen sus propias ideas, pero cuando el e-tutor es la única persona que conoce la respuesta o los estudiantes proceden lentamente puede que convenga dejar de lado esta estrategia (Jaques & Salmon, 2007).
- Redireccionar: en lugar de regresar la pregunta al estudiante que la planteó puede convenir redireccionarla a otros estudiantes. Por ejemplo, si un estudiante dice "No entiendo lo que el autor está tratando de decir ¿Qué significa todo esto?", puede redireccionarse por medio de "¿Qué piensan ustedes?", "¿Todos los demás tienen el mismo problema?" o bien "¿Tienen alguna idea de lo que significa esto?" (Jaques & Salmon, 2007).

Las preguntas de elaboración tratadas en (Jaques & Salmon, 2007) también son de utilidad ya que ayudan a los estudiantes a expresar sus sentimientos y pensamientos. Ejemplos de este tipo de preguntas son: "¿Puedes decirme más?", "¿Puedes explicar eso?", "¿Qué más?", y "¿Cómo te hace sentir eso?".

IV.4.5.1.2 Preguntas relevantes para el ámbito de evaluación

Durante el proceso de evaluación de un estudiante un e-tutor puede recurrir a las preguntas de testeo dado que su propósito es obtener información específica. Estas preguntas que usualmente comienzan con palabras como "qué", 'donde, 'cuando', y 'cual', pueden dividirse en las siguientes sub-categorías (Jaques & Salmon, 2007):

- Chequear conocimiento: "¿Cuál es el mejor catalizador para……?" "¿Qué críticos han descrito a…… cómo……?".
- Comprensión: "Cómo justificas……?".
- Aplicación: "Cómo predices que eso trabajará en……" "¿Qué relevancia tiene eso en……?".
- Análisis: "¿Qué cualidades tienen en común……?""¿Qué hubiera pasado si……?".
- Síntesis: "¿Cómo se conecta eso con……?""¿Podrías resumir lo que hemos discutido hasta aquí?".

- Evaluación: "¿Cuál piensas es el mejor……?" "¿Cómo te sientes acerca de eso?".

IV.4.5.1.3 Preguntas relevantes para el ámbito pedagógico

Tal como se mencionó al tratar las habilidades pedagógicas, las preguntas le sirven a los e-tutores para múltiples propósitos, como ser: el enfocar la discusión en conceptos críticos, principios y habilidades; prevenir que la discusión se salga de su curso; y chequear la comprensión de los estudiantes.

Independientemente del tipo de pregunta que se utilice se debe tratar que la misma sea concisa y que los estudiantes puedan responderla (Kemshal Bell, 2001). También hay que tener cuidado de no ser intrusivo o dominante realizando una cantidad excesiva de preguntas porque esto puede contribuir a establecer un patrón de comunicación focal o lineal (Jaques & Salmon, 2007; Palloff & Pratt, 2001). Cuanto más se apoderen los estudiantes de la tarea realizándose preguntas entre ellos mayor será la colaboración y la responsabilidad que tomarán por su propio aprendizaje (Jaques & Salmon, 2007).

Finalmente, si bien el uso de preguntas es una estrategia que funciona en clases presenciales donde el docente puede guiar la discusión siguiendo un orden de preguntas, en los entornos de *e-learning* podría dar origen a una compleja mezcla de mensajes ya que los estudiantes podrían elegir responder a todas las preguntas asincrónicamente y en cualquier orden (Salmon, 2003).

IV.4.5.2 Escuchar

Si un e-tutor no es capaz de escuchar nunca entenderá la posición y las necesidades de los estudiantes (ION, 2005; Kemshal Bell, 2001; Jaques & Salmon, 2007; Blackburn, 2004). Para escuchar apropiadamente conviene que un e-tutor cuente con los atributos de habilidad de: atender a todos los estudiantes dándoles a conocer que están siendo escuchados (Kemshal Bell, 2001), interpretar los mensajes explícitos e implícitos (TCOL, 2003), y agradecer a los estudiantes que dan realimentación por su preocupación y ayuda (Jaques & Salmon, 2007).

IV.4.5.3 Responder

Un e-tutor debe identificar y satisfacer las necesidades de los estudiantes relativas al feedback y clarificación puesto que si no reciben respuestas los alumnos tendrán dificultades en desarrollar un sentido de participación en el curso (Fåhræus, 2000; Cornelius & Higgison, 2001; Kemshal Bell, 2001; ION, 2005). El proveer feedback sobre

las contribuciones de los estudiantes, el contenido del curso, las interacciones, o el proceso de resolución de problemas tiene un impacto positivo en la motivación, la interacción, las habilidades de resolución de problemas, y el aprendizaje grupal (Kukulska-Hulme, 2004; Cornelius & Higgison, 2001). Además, los estudiantes usualmente encuentran a la comunicación unidireccional y sin respuesta, como algo no satisfactorio (Fåhræus, 2000). Sin embargo, el feedback excesivo o muy frecuente puede tener efectos alienantes (ION, 2005; Kemshal Bell, 2001; Bláquez Entonado & Alonso Díaz, 2006). Para que un e-tutor proporcione respuestas efectivas se requieren los atributos de habilidad de:

- Proveer *feedback* oportuno y frecuente: generalmente es mejor dar *feedback* tan pronto como sea posible porque la mayoría de los alumnos no toleran las demoras (Kemshal Bell, 2001; Duggleby *et al*., 2002; Packham *et al*., 2004; Hathaway *et al.*, 2007; Cornelius & Higgison, 2001; TCOL, 2003; Fåhræus, 2000; Lim & Cheah, 2003). En lo posible se debería contestar a las preguntas formuladas por los alumnos en un plazo máximo de entre 36 a 48 horas, ya que en un ambiente de comunicación mediado por computadora el *feedback* atrasado puede desalentar la actividad y producir sentimientos de insatisfacción, inquietud, frustración, y soledad (Cornelius & Higgison, 2001; Llorente Cejudo, 2006; Packham *et al*., 2004; Fåhræus, 2000; de Laat, 2006).
- Proveer *feedback* claro y constructivo: no basta con decir que algo está incompleto sino que el e-tutor debe tratar de explicar de forma no ambigua los pasos que debe seguir el estudiante para asegurar que la actividad cumpla con el estándar requerido y pueda mejorar su rendimiento (Kemshal Bell, 2001; Duggleby *et al*., 2002; Hootstein, 2002; Dewar & Whittington, 2000; Varvel, 2007; TCOL, 2003; Shepherd, 1999; Fåhræus, 2000).
- Proveer *feedback* detallado: es necesario enfocar el *feedback* en cuestiones específicas dado que si es general no le servirá al estudiante para mejorar (Kemshal Bell, 2001; Varvel, 2007; Packham *et al*., 2004; Hathaway *et al.*, 2007).
- Proveer *feedback* positivo: aun cuando la actividad recibida no sea lo esperado conviene comenzar diciendo algo positivo y estimulante o alabando algún aspecto de los logros de los estudiantes. De esta manera, el estudiante nunca percibirá el *feedback* como algo injusto o duro (Duggleby *et al*., 2002; Packham *et al*., 2004; Hathaway *et al.*, 2007; Goodyear *et al.*, 2001; Shepherd, 1999; Fåhræus, 2000; Bonk *et al.*, 2004).

- Proveer *feedback* descriptivo en vez de evaluativo: el e-tutor debería describir lo que realmente ve o escucha para reducir la reacción negativa por parte del receptor (Jaques & Salmon, 2007). Antes de referirse a una contribución de los estudiantes es conveniente que un e-tutor considere el tono y contenido, cuestiones relativas al autor (habilidades, conocimiento y actitudes) que pueda conocer a partir de contribuciones previas, y el tiempo de la contribución en relación al desarrollo de la discusión (Berge, 1995; Llorente Cejudo, 2006; CSALT, 2001).
- Dirigir el *feedback* hacia el comportamiento que los alumnos pueden cambiar o controlar (Jaques & Salmon, 2007).
- Chequear que el receptor ha entendido: en la medida de lo posible intentar que el receptor comente con sus propias palabras el *feedback* recibido para chequear si coincide con lo que se tenía en mente (Jaques & Salmon, 2007).
- Usar una variedad de métodos de *feedback*: incluyendo texto, gráficos, y audio (Kemshal Bell, 2001).

IV.4.5.4 Dialogar con los estudiantes

Los atributos de habilidad relacionadas con mantener un diálogo efectivo con los estudiantes son:

- Usar técnicas de dialogo envolventes, interactivas, personales, y cálidas (Kemshal Bell, 2001).
- Comunicarse con los estudiantes en forma clara, concisa, y evitando el uso excesivo del lenguaje propio de la disciplina o profesión (IITT, 2007; Salmon, 2003).
- No dar conferencias: evitar secuencias de comentarios largas y elaboradas ya que son difíciles de leer en la pantalla, resultan tediosas, impiden la discusión y pueden conducir al silencio (Berge, 1995; Llorente Cejudo, 2006; CSALT, 2001; Palloff & Pratt, 2001; Bonk *et al.*, 2004; Lefoe *et al*., 2002). Por ello, es recomendable usar mensajes múltiples y pequeños, sobre todo cuando se trata más de un tema a la vez (Full Circ, 2001).
- Usar un estilo de escritura que compense la carencia de contacto visual y lenguaje corporal (Salmon, 2003; Kemshal Bell, 2000). En este contexto los emoticones son de gran utilidad para, por ejemplo, trasmitir emociones (Kemshal Bell, 2000; Garrison *et al.*, 1999).

IV.4.5.5 Crear el ambiente de comunicación

Es primordial crear un ambiente donde los estudiantes se sientan confortables expresándose ellos mismos ya que si perciben el proceso de interacción como una amenaza esto puede tener un efecto distorsivo en la comunicación (Palloff & Pratt, 2001; Jaques & Salmon, 2007). Para asegurar que el diálogo se mantenga constructivo es conveniente que el e-tutor posea atributos de habilidad que le permitan:

- Crear un ambiente *online* de soporte donde los estudiantes se sientan confortables preguntando y respondiendo preguntas (Kemshal Bell, 2001).
- Proporcionar un estilo cortés y respetuoso en la comunicación (Salmon, 2003).
- Impulsar libertad de expresión y de pensamiento (Kemshal Bell, 2001; Dewar & Whittington, 2000).
- Usar el humor como una forma de ayudar a los estudiantes a sentirse bienvenidos y seguros (Palloff & Pratt, 2001). No obstante, la ironía y el sarcasmo deben utilizarse cuidadosamente dado que las diferencias culturales o étnicas pueden llevar a la confusión o interpretación errónea (Berge, 1995; Kemshal Bell, 2001; Salmon, 2003; Blackburn, 2004; Fåhræus, 2000; Bonk *et al.*, 2004; Lund, 2004; Lim & Cheah, 2003).
- Evitar interpretaciones erróneas del silencio y falta de respuesta (Orvis & Lassiter, 2006).

IV.4.5.6 Gestionar problemas de comunicación

Entre los problemas de comunicación a los cuales debe hacer frente un e-tutor se encuentran los conflictos, los hilos paralelos, y las estrategias defensivas de los estudiantes (ION, 2005; Jaques & Salmon, 2007; Fåhræus, 2001; Fåhræus, 2000). Los conflictos ya fueron tratados en la sección dedicada a las habilidades sociales. Con respecto a los hilos paralelos, la demora de tiempo propia de los ambientes de *e-learning* asincrónicos ocasiona que distintos tipos de contribuciones sean mezclados. Para tratar el problema de los hilos paralelos y eliminar la sobrecarga de información que estos generan se puede colocar la discusión de cada tópico en diferentes foros y fomentar el hábito de escribir contribuciones cortas y etiquetas (Fåhræus, 2000; Fåhræus, 2001).

Según (Jaques & Salmon, 2007) las estrategias defensivas de los estudiantes que pueden actuar como barreras para la comunicación son: evitar que otros expresen sus propias ideas, tratar constantemente de cambiar las opiniones de sus pares, implicar superioridad, y evitar la expresión de sentimientos.

Habilidades	Sub habilidades
Preguntar	• Realizar preguntas concisas. • Realizar preguntas que los estudiantes puedan responder. • Realizar preguntas de manera no intrusiva.
Escuchar	• Indicarle a los estudiantes que están siendo escuchados. • Interpretar los mensajes explícitos e implícitos. • Agradecer al estudiante que da el *feedback*.
Responder	• Proveer *feedback* oportuno y frecuente. • Proveer *feedback* claro y constructivo. • Proveer *feedback* detallado. • Proveer *feedback* positivo. • Proveer *feedback* descriptivo. • Dirigir el *feedback* al comportamiento que se puede cambiar. • Chequear que el receptor ha entendido. • Usar una variedad de métodos de *feedback*.
Dialogar con los estudiantes	• Usar técnicas de dialogo envolventes, interactivas, personales, y cálidas. • Comunicarse con los estudiantes en forma clara y concisa. • No dar conferencias. • Compensar la carencia de lenguaje corporal.
Crear el ambiente de comunicación	• Crear un ambiente *online* de soporte. • Proporcionar un estilo cortés y respetuoso en la comunicación. • Impulsar libertad de expresión y de pensamiento. • Usar el humor para que los estudiantes se sientan bienvenidos. • Evitar interpretaciones erróneas del silencio y falta de respuesta.
Gestionar problemas de comunicación	• Manejar hilos paralelos. • Tratar con estudiantes que buscan que otros no expresen sus ideas. • Tratar con estudiantes que buscan cambiar las opiniones de sus pares. • Tratar con estudiantes que tienen aires de superioridad. • Tratar con estudiantes que buscan evitar la expresión de sentimientos.

Tabla 4.6. Habilidades y sub habilidades de comunicación de e-tutores de ACSC

IV.4.6 Habilidades relativas al rol de evaluación

Las tres sub habilidades que se pueden distinguir en esta faceta de la tarea de un e-tutor, se vinculan con la evaluación de los estudiantes y del proceso de enseñanza, de manera tal de poder decidir sobre el éxito de la experiencia de aprendizaje (Reid, 2003; eTTNet TWG 2, 2005).

IV.4.6.1 Evaluar a los estudiantes

El e-tutor debe poder evaluar a los estudiantes usando técnicas de evaluación (pruebas, cuestionarios, simulaciones, etc.) que permitan no sólo chequear que los estudiantes hayan logrado sus objetivos de aprendizaje, sino también, facilitar su progreso a lo largo del curso (Cornelius & Higgison, 2001; Varvel, 2007; Shepherd, 1999; IITT, 2007). En discusiones *online* existe un fuerte vínculo existente entre evaluación, motivación y participación. Esto es, "la evaluación puede diseñarse para proveer la fuerza motivadora para participar *online* en una manera productiva, al mantener el ímpetu, influir en la calidad de las contribuciones y prescribir la dirección y ritmo de la participación" (Cornelius & Higgison, 2001). Además, la evaluación puede ser usada para reconocer el valor de la experiencia de los estudiantes (Cornelius & Higgison, 2001). Para evaluar a los alumnos, los e-tutores precisan los atributos de habilidad vinculados con:

- Aplicar técnicas de evaluación Inicial, Formativa o de Proceso, y Sumativa de manera de determinar lo que los estudiantes han aprendido y logrado (Fåhræus, 2001; Palloff & Pratt, 2007; ION, 2005; Goodyear *et al*., 2001; IITT, 2007; Llorente Cejudo, 2006; eTTNet TWG 2, 2005; Adell & Sales, 1999).
- Interrogar a los estudiantes recurriendo por ejemplo a preguntas de testeo del tipo "Cómo justificas…?", "¿Qué cualidades tienen en común…?", y "¿Qué hubiera pasado si…?" (Shepherd, 1999; Jaques & Salmon, 2007).
- Otorgar calificaciones justas a los estudiantes por su participación y contribuciones (Salmon, 2003; Shepherd, 1999; Bonk *et al.*, 2004).
- Asegurar la autenticidad del trabajo de los estudiantes (Goodyear *et al*., 2001).
- Identificar fortalezas y debilidades de los estudiantes (eTTNet TWG 2, 2005).

IV.4.6.2 Evaluar el proceso de enseñanza

En caso de ser necesario un e-tutor debería modificar el proceso instruccional en base a los datos que dispone sobre el éxito de la enseñanza *online*, y lo que aprende sobre las necesidades de los estudiantes (Goodyear *et al*., 2001; Duggleby *et al*., 2002). En este sentido los atributos de habilidad que conviene que un e-tutor posea, son:

- Determinar donde la tutoría *online* fue exitosa y donde una aproximación diferente hubiera sido más exitosa (Goodyear *et al*., 2001; Salmon, 2003; Reid, 2002; IITT, 2007; Palloff & Pratt, 2007). Una forma de hacer esto es obteniendo *feedback* y sugerencias por parte de los estudiantes con respecto a la tutoría que han recibido (IITT, 2007).

- Monitorear y revisar en forma continua el curso (eTTNet TWG 2, 2005).
- Realizar cambios y mejoras al curso (Duggleby *et al*., 2002; eTTNet TWG 2, 2005).
- Evaluar los recursos usados en el curso (Goodyear *et al*., 2001; Varvel, 2007).
- Chequear que las adaptaciones del curso o tutoría realmente satisfagan las necesidades de los estudiantes (Kemshal Bell, 2001).

IV.4.6.3 Reportar los resultados

El brindarles a los estudiantes información sobre su rendimiento permite darles a conocer donde están parados, guiarlos en lo que necesitan hacer para mejorar su posición en la clase, y motivarlos a participar (Olivares, 2007). Los atributos de habilidad que se corresponden con informar a los estudiantes sobre su desempeño son:

- Criticar amablemente el trabajo de los estudiantes (TCOL, 2003; Mason, 1991; Duggleby *et al*., 2002).
- Proveer críticas constructivas con sugerencias para mejora (Varvel, 2007).
- Distribuir calificaciones conservando estatutos legales (Goodyear *et al*., 2001).
- Proveer un tiempo de respuesta corto al realizar evaluaciones (Varvel, 2007; IITT, 2007).

Al proporcionarles a los estudiantes información sobre su desempeño no se deben perder de vista los principios sobre el *feedback* que se trataron en el apartado de habilidades de comunicación.

Habilidades	Sub habilidades
Evaluar a los estudiantes	• Aplicar técnicas de evaluación inicial, formativa y sumativa. • Interrogar a los estudiantes. • Otorgar calificaciones justas. • Asegurar la autenticidad del trabajo de los estudiantes. • Identificar fortalezas y debilidades de los estudiantes.
Evaluar el proceso de enseñanza	• Determinar si la tutoría fue exitosa. • Monitorear y revisar de forma continua el curso. • Mejorar el curso. • Evaluar los recursos usados en el curso. • Chequear que las adaptaciones del curso satisfagan a los estudiantes.

Tabla 4.7. Habilidades y sub habilidades de evaluación de e-tutores de ACSC

Habilidades	Sub habilidades
Reportar los resultados	• Criticar amablemente el trabajo de los estudiantes. • Proveer críticas constructivas. • Distribuir calificaciones conservando estatutos legales. • Proveer un tiempo de respuesta corto.

Tabla 4.7. (Cont.)

MARCO METODOLÓGICO

La descripción de las metodologías, métodos, técnicas y herramientas usadas durante el desarrollo del presente trabajo de investigación se estructura en dos capítulos.

Capítulo V: Etapas de un Estudio de Análisis de Contenido

Se parte de la definición de análisis de contenido para luego abordar las etapas y elementos que forman parte de un estudio de análisis de contenido. También, se muestra mediante un ejemplo el proceso para calcular un coeficiente de acuerdo entre los observadores o analistas involucrados en un proyecto de análisis de contenido.

Capítulo VI: Etapas de un Proyecto de Minería de Textos

Se describen las actividades y resultados esperados de cada una de las etapas de la metodología de minería de datos CRISP-DM. Partiendo de la visión de que la minería de textos es una especialización de la minería de datos, se abordan las adaptaciones que pueden realizarse a la metodología CRISP-DM de modo tal de que CRISP-DM pueda ser utilizada en proyectos de minería de textos.

CAPÍTULO V
ETAPAS DE UN ESTUDIO DE ANÁLISIS DE CONTENIDO

V.1 DEFINICIÓN DE ANÁLISIS DE CONTENIDO

Al análisis de contenido se lo puede definir como una técnica de investigación para hacer inferencias replicables y válidas desde textos (u otra materia significativa) a los contextos de su uso (Krippendorff, 2004). En la definición anterior la expresión "otra materia significativa" indica que el análisis de contenido no debe restringirse al material escrito. Trabajos de arte, imágenes, mapas, sonidos, signos, símbolos, e incluso registros numéricos pueden ser incluidos como datos siempre y cuando los mismos hablen a alguien sobre un fenómeno exterior a lo que puede ser censado u observado.

Tal como lo plantea Krippendorff los resultados a los que conduce el análisis de contenido deberían ser tanto replicables como válidos. La replicabilidad hace referencia a que investigadores trabajando en diferentes puntos del tiempo, y quizás bajo circunstancias diferentes, deberían obtener los mismos resultados cuando aplican la misma técnica a los mismos datos. Por su parte, la validez de los resultados indica que el proceso de investigación puede examinarse cuidadosamente y que las afirmaciones resultantes pueden ser sostenidas frente a evidencia disponible independientemente.

De acuerdo con Krippendorff (2004) las siguientes 6 características de los textos son relevantes a la definición de análisis de contenido dada:

- Los textos no tienen objetivo, es decir no hay cualidades independientes del lector. No existe nada inherente en un texto y los significados de un texto son siempre asignados por alguien.
- Los textos no tienen significados simples que puedan ser encontrados, identificados, y descriptos por lo que son. Así es posible ofrecer interpretaciones siquiátricas, sociológicas, políticas o poéticas de los textos. A pesar de que estas versiones son diferentes pueden ser igualmente válidas.
- El significado invocado por los textos no necesita ser compartido. Aunque el acuerdo intersubjetivo sobre lo que un autor quiere decir o lo que un texto significa simplificaría el análisis de contenido, tal consenso rara vez existe. Demandar que los analistas encuentren una "base común" restringiría el dominio empírico del

análisis de contenido al aspecto manifiesto de la comunicación, o podría restringir el uso de análisis de contenido a una pequeña comunidad de productores, receptores, y analistas quienes miran el mundo desde la misma perspectiva.

- Los significados (contenidos) hablan de algo más que los textos dados, incluso donde la convención sugiere que los mensajes los contienen o que los textos los tienen. Los textos pueden proveer información sobre eventos en localizaciones distantes, sobre objetos que ya no existen, sobre ideas en la mente de las personas, y acciones disponibles. Los textos también pueden conducir a respuestas de diversos tipos. Todos estos fenómenos vinculan la lectura de los textos a algo más.
- Los textos tienen significados relativos a contextos, discursos, o propósitos particulares. Profesionales de distintos campos tienen sus propias razones basadas en la disciplina para interpretar de modo diferente las afirmaciones dadas. Sin embargo, las diferencias en interpretaciones no excluyen la posibilidad de acuerdo dentro de contextos particulares. De hecho una vez que los analistas han elegido el contexto dentro del cual intentan asignar significados a un texto dado, la diversidad de interpretaciones puede ser reducida a un número manejable, algunas veces a una. Un contexto transforma los datos perceptuales en textos legibles y sirve como una justificación conceptual para interpretaciones razonables e incluso para los resultados del análisis de contenido. Una vez que el analista ha seleccionado un contexto para un cuerpo particular de textos y claramente entiende el contexto, ciertos tipos de preguntas admiten una respuesta y otras no tienen sentido. Para que el análisis de contenido sea replicable los analistas deben explicar el contexto que guía sus inferencias.
- La naturaleza del texto demanda que los analistas de contenido extraigan inferencias desde el cuerpo de textos a su contexto seleccionado. Para los analistas de contenido la lectura sistemática de textos reduce el rango de posibles inferencias concernientes a hechos no observados, intenciones, estados mentales, efectos, prejuicios, acciones planeadas, y condiciones antecedentes o consecuentes. Los analistas de contenido infieren respuestas a preguntas de investigación particulares desde sus textos. Las inferencias de los analistas son más sistemáticas, explícitamente informadas, e idealmente verificables que lo que lectores ordinarios hacen con textos.

V.2 MARCO DE TRABAJO PARA ANÁLISIS DE CONTENIDO

Como toda técnica el análisis de contenido involucra una serie de procedimientos que pueden ser aprendidos y que son independientes del investigador. Con el propósito de proveer una guía para el diseño, desarrollo y revisión crítica de estudios de análisis de contenido Krippendorff (2004) propone un *framework* conceptual cuyos componentes son: un cuerpo de textos, preguntas de investigación, un contexto, constructo analítico, inferencias y evidencia de validación. Estos componentes se explican en las secciones incluidas a continuación.

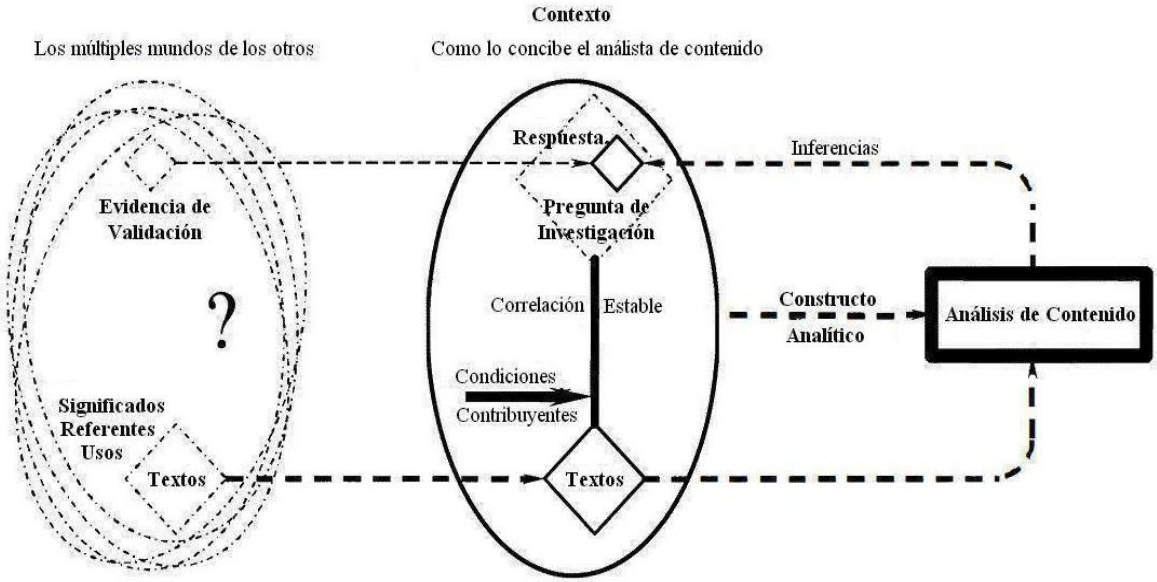


Figura 5.1. *framework* para análisis de contenido (Krippendorff, 2004)

V.2.1 Cuerpo de Textos

Los textos son el punto de comienzo del proceso de análisis de contenido pero se debe tener en cuenta que los textos usualmente no han sido previstos para responder a preguntas de investigación específicas. Los textos son textos en la medida que son escritos para ser leídos, interpretados, y entendidos por otras personas además de los analistas. La lectura de un analista (las unidades, sintaxis, y estructuras narrativas que constituyen las unidades para el analista) difieren de las lecturas que inician las interpretaciones de lectores ordinarios, incluyendo los autores de los textos. Sin embargo, la lectura del analista nunca debería ser tomada como la única legitima ni deberían los analistas de contenido asumir el poder de determinar la forma de los textos que analizan. Porque en este caso se estarían examinando ellos mismos.

V.2.2 Preguntas de Investigación

La formulación de preguntas de investigación es fundamental para crear un diseño de investigación exitoso. Las preguntas de investigación son el objetivo de las inferencias del analista desde los textos disponibles y generalmente delinean varias respuestas posibles e inicialmente inciertas.

En cierta medida las preguntas de investigación son análogas a las hipótesis. Sin embargo, mientras las hipótesis se confrontan con la evidencia observacional directa, las preguntas de investigación del análisis de contenido deben ser respondidas mediante inferencias extraídas desde textos. Las hipótesis científicas son aceptadas a causa de la preponderancia de evidencia en favor de una a expensas de otras hipótesis, en tanto que las inferencias desde los textos conciernen a un fenómeno que no se observa durante un análisis de contenido, fenómeno que esta fuera de los textos y en consecuencia mantiene su carácter hipotético hasta que se confirma por las incidencias de validación. Existen dos razones para comenzar un análisis de contenido con preguntas de investigación:

- Eficiencia: cuando la investigación está motivada por preguntas específicas los analistas avanzan más rápidamente desde el muestreo de los textos disponibles porque leen los textos con un propósito. En cambio sin preguntas de investigación un analista puede dejarse llevar por los significados que vienen a su mente mientras lee un texto y seguir hilos de inferencias a donde quiera que los mismos puedan conducir.
- Fundamento empírico: la búsqueda de respuestas a las preguntas de investigación también sustenta el estudio empíricamente. Las respuestas a las preguntas planteadas implican afirmaciones verdaderas que pueden ser soportadas mediante observación directa o al menos por argumentación plausible desde observaciones relacionadas.

No todas las preguntas que se pueden formular califican como preguntas de investigación para un estudio de análisis de contenido, para ello es necesario que las preguntas reúnan las siguientes características:

- Se deben poder responder (inferir de manera abductiva) examinando el cuerpo de textos.
- Delinean un conjunto de posibles (hipotéticas) respuestas entre las cuales el analista selecciona.

- Conciernen a un fenómeno actualmente inaccesible o no observable en el contexto problematizado de los textos disponibles.
- Permiten validación: no basta con solo responder una pregunta de investigación sino que se debe buscar como mínimo otra manera de confirmar, al menos en principio, la ocurrencia del fenómeno inferido: observaciones adicionales, correlaciones con medidas que se conoce son válidas, e incluso observaciones del éxito logrado cuando se actúa en base a la información que el análisis de contenido provee. El formular preguntas de modo que las respuestas puedan ser validadas protege al analista de contenido de perderse en meras abstracciones o categorizaciones auto convenientes.

V.2.3 Contexto

Los textos adquieren significancia (significado, contenidos, cualidades simbólicas, e interpretaciones) en los contextos de su uso. El contexto especifica el mundo en el cual los textos pueden relacionarse a las preguntas de investigación del analista. Un contexto es siempre la construcción de alguien, el ambiente conceptual de un texto, la situación en la que juega un rol.

En el curso de un análisis de contenido el contexto abarca todo el conocimiento que el analista aplica a los textos, ya sea en la forma de teorías científicas, proposiciones sostenidas plausivamente, evidencia empírica, intuiciones sustentadas o conocimientos de hábitos de lectura. El conocimiento del contexto puede separarse en dos tipos:

- Red de correlaciones estables: las cuales se cree conectan los textos disponibles a las respuestas posibles a las preguntas de investigación dadas, ya sea que las correlaciones sean establecidas empíricamente, derivadas desde teoría aplicable o meramente asumidas para el propósito de un análisis.
- Condiciones contribuyentes: consisten de todos aquellos factores que se sabe afectan a la red de correlaciones estables en maneras previsibles.

El contexto es siempre uno de muchos, porque personas de diferentes disciplinas tienden a colocar a los mismos textos en diferentes contextos como consecuencia de que siguen sus propias agendas de investigación, y de que se aproximan a los textos con sus propias preguntas, conceptos, modelos y herramientas analíticas. En la Figura 5.1 la idea

de múltiples contextos se muestra por medio de óvalos que rodean a los textos y sus múltiples significados.

En la medida que el contexto reemplaza a lo que es momentáneamente inaccesible a la observación directa, no hay un límite al número de contextos que pueden aplicarse en un análisis dado. Por consiguiente, los analistas de contenido tienen que hacer explícitos los contextos seleccionados de manera que los resultados del análisis sean claros para sus pares científicos y los beneficiarios, y no surjan interpretaciones erróneas.

V.2.4 Constructo Analítico

Un constructo analítico permite operacionalizar lo que el analista conoce, sospecha o asume sobre el contexto de los textos, esto es la red de correlaciones estables y las condiciones bajo las cuales estas correlaciones pueden cambiar. Los constructos analíticos representan esta red en forma computable y aseguran que el análisis no proceda en violación de lo que se conoce de las condiciones que rodean a los textos.

Los constructos analíticos equivalen a reglas de inferencia que guían al analista en una serie de pasos desde los textos disponibles a las respuestas buscadas. Pero además, transforman el conocimiento del contexto en portable para llevar a cabo el análisis de contenido en contextos similares y permiten que otras personas puedan examinar los procedimientos que un analista ha utilizado. En su forma más simple los constructos analíticos se representan por medio de una función, una colección de afirmaciones "if-then" o un programa de computadora.

V.2.5 Inferencias

Las inferencias que se obtienen en un estudio de análisis de contenido se pretende que respondan a las preguntas de investigación. El análisis de contenido no se centra en inferencias inductivas (desde particularidades a generalidades) o deductivas (desde generalidades a particularidades) sino en un tipo particular denominado inferencias abductivas. Este tipo de inferencias se puede considerar como una forma de razonamiento que se mueve desde textos particulares, a través de explicaciones sensibles al contexto de esos textos, a respuestas particulares a las preguntas de investigación, es decir desde particularidades a particularidades. A modo de ejemplo se puede considerar la edad y competencia lingüística. Si bien ninguna implica a la otra, teniendo experiencia con adquisición del lenguaje en niños se podría inferir la edad un niño en base a los sonidos que hace o el vocabulario que usa. Las inferencias se hacen con cierta probabilidad que

puede ser fortalecida si se toman en cuenta otras variables (las condiciones contribuyentes).

Para Krippendorff (2004) el movimiento desde los textos a las respuestas de las preguntas de investigación puede ilustrarse por medio del modelo de la teoría de la argumentación de Toulmin. Según este modelo el movimiento desde datos (D) a conclusiones o afirmaciones (C) debe ser justificado por una garantía apropiada (G). Así por ejemplo, sí se sabe una persona es Sueca la inferencia de que esta persona es probablemente protestante está garantizada por el conocimiento que la mayoría de los Suecos son protestantes. Pero dado que puede haber excepciones en relación a la inferencia, se incluye a calificación (Q) de la conclusión. Además se tiene una base (B) sobre la cual la garantía puede ser justificada o sustentada.

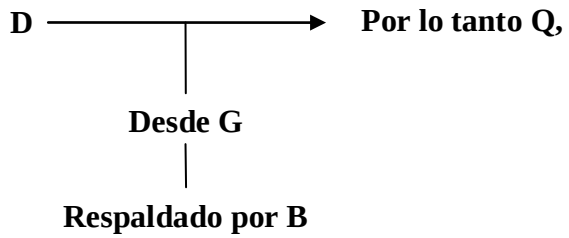


Figura 5.2. Diagrama de Toulmin (Krippendorff, 2004)

En un estudio de análisis de contenido en el camino desde los textos a la respuesta a la pregunta de investigación es el constructo analítico más la seguridad que el análisis se realizo confiablemente lo que garantiza la inferencia. A su vez la inferencia es sustentada por el conocimiento que tiene el analista del contexto, lo cual asegura que el camino seguido conduce a conclusiones validas.

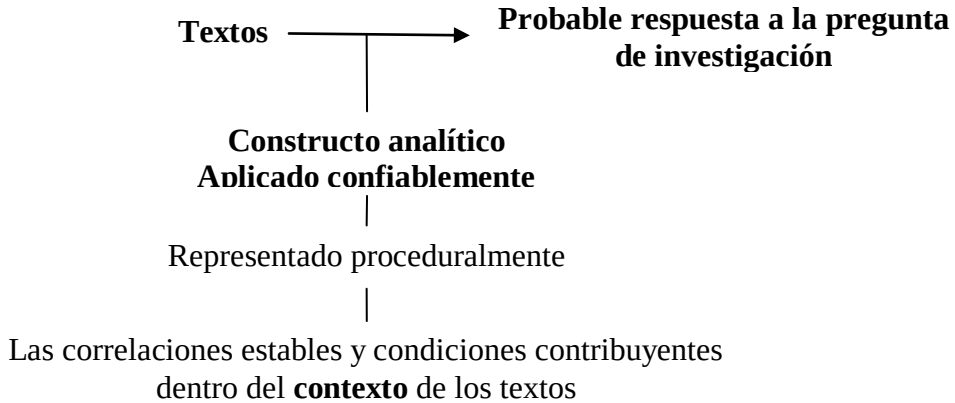


Figura5.3. Diagrama de Toulmin aplicado al análisis de contenido (Krippendorff, 2004)

V.2.6 Evidencia de validación

Cualquier análisis de contenido debería ser validable en principio. Dado que la razón de ser del análisis de contenido es la ausencia de evidencia observacional directa, la validación puede ser difícil e incluso imposible en la práctica. Es difícil cuando el análisis de contenido se lleva a cabo en la ausencia de evidencia observacional directa como ser el análisis en tiempo de guerra de actividades militares planeadas a partir de propaganda doméstica (uso de los medios comunicación para influenciar una comunidad). Es imposible cuando las preguntas de investigación conciernen a sucesos pasados o futuros.

El propósito de requerir que el análisis de contenido sea validable en principio es prevenir al analista de perseguir preguntas de investigación que no permiten validación empírica o que conducen a resultados cuyo único soporte es la autoridad del investigador. A modo de ejemplo, si un analista de contenido infiere lo que un grupo de votantes particular aprendió sobre un candidato desde avisos de campaña en televisión, mediante una encuesta a las personas expuestas a los avisos se podría validar o invalidar la inferencia. No obstante, si el analista insiste en que los avisos tienen cierto contenido, no hay manera de validar este descubrimiento.

Es primordial que los analistas de contenido sean específicos sobre lo que puede validar o invalidar sus resultados, incluso si después no se lleva a cabo el esfuerzo de validación, porque de esta manera vinculan el análisis de contenido a una realidad que se puede confirmar de múltiples formas.

V.3 ELEMENTOS DEL ANÁLISIS DE CONTENIDO

Si se abre la caja correspondiente a "Análisis de contenido" de la Figura 5.1 se pueden identificar 6 componentes que necesitan los analistas para proceder desde los textos a los resultados: separación en unidades, muestreo, grabación, reducción, inferencia y narración. Los primeros 4 componentes constituyen lo que puede ser llamado creación de los datos, es decir la construcción de datos computables desde textos no editados o sin tratar.

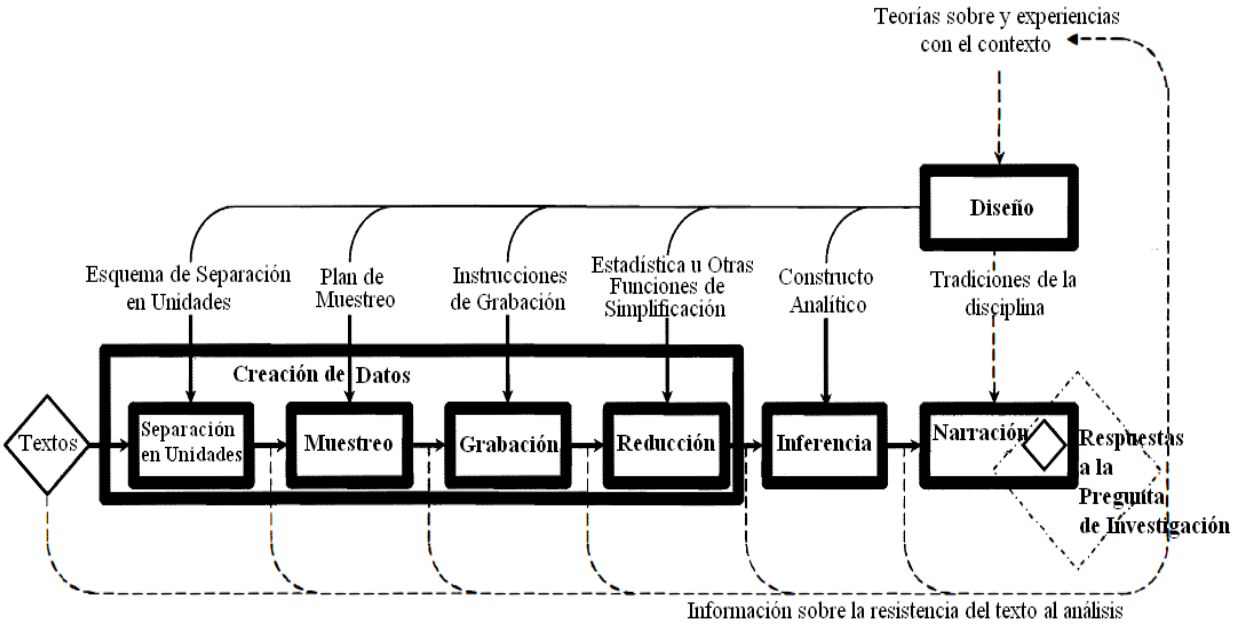


Figura 5.4 Componentes del análisis de contenido (Krippendorff, 2004)

Los componentes de análisis de contenido no necesitan estar organizados tan linealmente como se muestra en la Figura 5.4. Un diseño de análisis de contenido (la secuencia de pasos que sigue un investigador en un estudio concreto) puede incluir ciclos iterativos, la repetición de un proceso particular hasta que cierta calidad se logre. Los componentes pueden también repetirse de varias formas. Por ejemplo, la separación en unidades puede preceder al muestreo de los documentos completos, pero puede ser necesaria para describir los detalles de su contenido. Así las instrucciones de grabación pueden incluir esquemas de separación en unidades. Por otro parte, un análisis de contenido puede usar componentes que no están representados en la Figura 5.4.

Las instrucciones escritas del analista de contenido (representadas con los cuadros en negrita en la Figura 5.4) tienen que incluir toda la información que el analista pueda comunicar a otros de manera que puedan replicar el estudio o evaluarlo críticamente.

Cualquier conjunto de instrucciones impone una estructura a los textos disponibles. Idealmente esta estructura debería sentirse como natural pero puede suceder que se sienta inapropiado o forzada. Durante el desarrollo de alguna fase del análisis de contenido, un analista sensible podría "resistir a la violencia" que instrucciones pobres pueden infligir en los textos e intentar reformular las instrucciones de manera que sean apropiadas para los textos en mano. Esta aproximación sensible se ilustra en la Figura 5.4 mediante líneas discontinuas, que muestran otro flujo de información que es motivado por la resistencia del analista a pasos analíticos inapropiados.

Finalmente, a pesar de que el contexto dirija el análisis de un texto, los resultados del análisis contribuyen a la conceptualización o re conceptualización del contexto, redirigen el análisis y así sucesivamente. Esto revela la naturaleza esencialmente recursiva del proceso de análisis de contenido.

V.3.1 Separación en Unidades

Consiste en identificar dentro de un medio (un continuo inicialmente no diferenciado) secciones continuas que contienen información relevante a la pregunta de investigación. Estas secciones se convierten en unidades de análisis o grabación, y las secciones entre las unidades identificadas se descartan. Además, es necesario que la información necesaria para el análisis esté representada en la colección de unidades, no en las relaciones entre unidades, porque la separación en unidades desecha las mismas. Ejemplos de separación en unidades incluyen el recortar artículos relevantes en un diario, señalar las secciones pertinentes en una cinta de video para su posterior análisis, aislar movimientos de conversación dentro de una conversación, e identificar eventos históricamente significativos en el tiempo.

En análisis de contenido se puede distinguir entre unidades de muestreo, unidades de grabación/codificación y unidades de contexto. En investigación de encuestas, donde las unidades tienden a ser los individuos que responden las preguntas, el muestreo desde una población es sencillo porque la diferencia entre unidades de muestreo y unidades de grabación desaparece, y las unidades de contexto son irrelevantes. Pero en análisis de contenido los tres tipos de unidades antes mencionados tienen funciones diferentes.

V.3.1.1 Unidades de muestreo

Son unidades que se distinguen para su inclusión selectiva en un análisis. Así por ejemplo alguien que sostiene haber analizado un cierto número de temas de un diario, está hablando de unidades de muestreo. Al momento de definir unidades de muestreo se debe tener en cuenta que:

- Todas las conexiones a través de las unidades de muestreo, si existen, no tengan influencia en el análisis.
- Toda la información relevante está contenida en las unidades de muestreo individuales, o en caso contrario las omisiones no empobrecen el análisis.

V.3.1.2 Unidades de grabación/codificación

Son unidades que se distinguen para la descripción, transcripción, grabación o codificación separada. Una vez que un analista describe las unidades de grabación, las descripciones y categorías asignadas son comparadas, analizadas, resumidas y usadas como la base para hacer inferencias. Las unidades de grabación están típicamente contenidas en las unidades de muestreo, a lo sumo coinciden con ellas, pero nunca las exceden. Una buena razón para seleccionar unidades de grabación que sean más pequeñas que las unidades de muestreo es que las unidades de muestreo son frecuentemente demasiado ricas o complejas para ser descriptas confiablemente.

V.3.1.3 Unidades de contexto

Son unidades de materia textual que establecen límites en la información a ser considerada en la descripción de unidades de grabación. Si se estuviera interesado en conocer si la palabra "ir" se refiere a un juego, a una acción, o a una orden, lo natural sería considerar la sentencia como unidad de contexto y la palabra como unidad de grabación. Las sentencias son la mínima unidad de contexto para las palabras individuales pero pueden no ser suficientes ya que, por ejemplo, para identificar el referente de un pronombre personal puede ser necesario examinar un par de sentencias que preceden al pronombre.

A diferencia de las unidades de muestreo y grabación, las unidades de contexto no se cuentan, no necesitan ser independientes entre sí, pueden solaparse, y pueden consultarse en la descripción de varias unidades de grabación. Aunque las unidades de contexto generalmente rodean a las unidades de grabación que ayudan a identificar, pueden preceder a la ocurrencia de una unidad de grabación o estar localizadas en cualquier lugar, como ser pies de nota, índices, glosarios o introducciones. Si bien no hay un límite en relación al tamaño de una unidad de contexto, las unidades de contexto grandes suelen conducir a explicaciones más específicas y semánticamente más adecuadas de las unidades de grabación, pero requieren más esfuerzo del analista.

V.3.2 Muestreo

El muestreo permite al analista economizar el esfuerzo de investigación limitando las observaciones a un subconjunto manejable de unidades que es estadísticamente o conceptualmente representativo del conjunto de todas las unidades posibles, la población o

universo de interés. De forma ideal, el análisis de la población completa y el análisis de una muestra representativa de la población deberían conducir a los mismos resultados.

La teoría de muestreo se refiere a la capacidad de generalizar las propiedades encontradas en una muestra a la población desde la cual la muestra se obtiene. Se basa en la ley de los grandes números para estimar el sesgo introducido por generalizaciones desde tamaños de muestra inadecuados, y provee justificaciones para varias técnicas estadísticas intentando minimizar tal sesgo. Crear muestras representativas para análisis de contenido es más complejo que crear muestras para por ejemplo experimentos psicológicos o investigación de consumidor, ya que la teoría de muestreo no es directamente aplicable en los problemas de muestreo que un analista de contenido enfrenta. Son 4 los supuestos que impiden la aplicación completa de la teoría de muestreo al muestreo de textos:

- En la teoría de muestreo las unidades de muestreo son individuos (reales o metafóricos) lo cual implica que son unidades indivisibles, independientes entre sí, y en consecuencia estadísticamente contables por sus propiedades, opiniones, o comportamientos. En cambio los textos pueden ser conceptualizados y separados en unidades de diversas maneras.
- En la teoría de muestreo las unidades muestreadas son las unidades contadas, en tanto que en análisis de contenido este es raramente el caso. Los analistas podrían muestrear cartas, temas de los diarios, o períodos de tiempo de producción de una película, pero encuentran respuestas a sus preguntas de investigación enumerando sentencias, categorizando referencias, o interpretando imágenes visuales.
- Los investigadores de encuesta controlan las preguntas realizadas a los entrevistados y deciden sobre la legitimidad de las respuestas de los entrevistados. En virtud de su membrecía a una población de interés, todos los individuos muestreados son considerados igualmente informativos sobre las preguntas de los investigadores de encuestas. En contraste, es raro que las diferentes unidades textuales tengan igual relevancia para una pregunta de investigación.
- La teoría de muestreo tradicional es una teoría de representación, en el sentido de que una muestra obtenida desde una población tiene la misma distribución de propiedades que la población. Esto sugiere que todos los miembros de la población tienen igual oportunidad de ser incluidos en la muestra. En cambio los analistas de contenido tienen que considerar al menos dos poblaciones a la vez: la población de respuestas a una pregunta de investigación y la población de textos que contienen o

conducen a las respuestas a esa pregunta. Por lo tanto los analistas de contenido están rara vez interesados en representaciones precisas del universo textual, en su lugar su preocupación es que los textos de interés sean relevantes a la pregunta de investigación y ayuden a responderla con imparcialidad.

Solo cuando todas las unidades muestreadas son igualmente informativas con respecto a una pregunta de investigación el muestreo en análisis de contenido es el mismo que el muestreo en investigación de encuestas. Para estas situaciones, pueden aplicarse técnicas de muestreo probabilísticas (muestreo aleatorio, sistemático y estratificado) que aseguran que todas las unidades tengan la misma oportunidad de ser incluidas en la muestra. Sin embargo, cuando las unidades de muestreo son desigualmente informativas, lo cual es bastante típico en análisis de contenido, el muestreo de textos se vuelve una función de lo que se conoce sobre la distribución de información dentro del universo textual. Las técnicas de muestreo que responden a esta situación son las indicadas a continuación:

- **Muestreo de probabilidad variable**: esta técnica reconoce que todas las unidades textuales no son igualmente informativas en relación a las respuestas a las preguntas de investigación y asigna a cada unidad de muestreo una probabilidad individual de contribuir a una respuesta. Si por ejemplo se buscara responder preguntas sobre opinión pública, un analista podría muestrear diarios de acuerdo a sus cifras de circulación. En tal muestra los diarios de mayor circulación (que afectan más gente) estarán sobre representados en relación a diarios de baja circulación.
- **Muestreo de *cluster***: es una técnica aplicable cuando un analista no puede enumerar todas las unidades de análisis pero encuentra grupos grades de tales unidades o *clusters*. En este caso se comienza listando los *clusters* disponibles, luego se selecciona entre ellos aplicando muestreo aleatorio, sistemático, o estratificado y todas las unidades contenidas en los *clusters* seleccionados son analizadas.
- **Muestreo de bola de nieve**: es una técnica multi etapa cuya aplicación comienza con una muestra inicial de unidades a la cual repetidamente se le aplica un conjunto de criterios de muestreo. Esta recursión produce una secuencia de adiciones a las unidades de muestreo que causan que la muestra crezca en tamaño hasta que se cumple el criterio de terminación.

- **Muestreo de relevancia**: el uso de muestras aleatorias implica reconocer que no se tienen pistas respecto a la población de interés o donde encontrar la información necesaria. Cuando se usa el muestreo de relevancia, los analistas proceden examinando (al menos de manera superficial) los textos de modo tal de seleccionar unidades textuales que contribuyan a responder las preguntas de investigación. Las unidades de texto resultantes no tienen la intención de ser representativas de la población de textos sino que son la población de textos relevantes de donde se excluyen las unidades textuales que no poseen información de relevancia.

V.3.3 Grabación

La grabación tiene lugar cuando los observadores, lectores, o analistas interpretan lo que ven, leen, o encuentran y luego expresan sus experiencias en términos de un lenguaje de datos que pueda ser analizado por medio del uso de técnicas apropiadas. Las razones de la necesidad de este componente analítico son:

- Crear registros durables de un fenómeno de otra manera efímero, tal como las palabras habladas o imágenes que pasan. Una vez que los fenómenos están grabados, los analistas pueden comparar los mismos, aplicarle diferentes métodos, y replicar el análisis de otros investigadores.
- El transformar texto no editado, imágenes originales y/o sonidos no estructurados en representaciones analizables.

En la práctica la grabación es la parte del análisis de contenido que representa el mayor problema para los analistas, quienes deben formular instrucciones de grabación que ellos y otros investigadores puedan seguir confiablemente. Las instrucciones de grabación no solo especifican el proceso de grabación sino que también permiten asegurarse que el significado de los registros resultantes esté disponible para otros. Al especificar las instrucciones de grabación los analistas buscan por un lado que las mismas contengan reglas explicitas y detalladas que los codificadores pueden aplicar confiablemente, y por el otro el minimizar el uso de juicios subjetivos.

Para poder replicar el análisis de contenido se recomienda que las instrucciones de grabación incluyan información específica en cuatro áreas principales:

- Las capacidades (habilidades cognitivas y formación) que los codificadores necesitan tener.

- El entrenamiento que los codificadores deben recibir en preparación para la tarea de grabación.
- La sintaxis y semántica del lenguaje de datos, preferentemente incluyendo los procedimientos cognitivos que los codificadores deben aplicar para grabar los textos e imágenes en forma eficiente y confiable.
- La naturaleza y administración de los registros a ser producidos. Los registros son la representación más básica y explicita del fenómeno analizado.

En el análisis de contenido un lenguaje de datos sirve mediador entre las experiencias de leer un texto, interpretar imágenes, y observar un fenómeno social efímero de interés por un lado y las demandas formales hechas por los procedimientos formales y analíticos por el otro lado. El lenguaje de datos permite representar la salida del proceso de grabación y se lo define en base a su semántica y sintaxis. La semántica del lenguaje de datos delinea el significado de los términos usados en el proceso codificación o grabación por lo cual cuando la semántica está mal definida no se puede interpretar las marcas o números que los codificadores dejan. Por su porte, la sintaxis satisface las operaciones formales requeridas en el manejo científico de datos. Cuando la sintaxis es incompatible con las demandas hechas por la técnica analítica empleada, los resultados analíticos son ininteligibles. La sintaxis consiste de: variables, valores, constantes, una gramática y una lógica. Las constantes especifican de que manera los valores de las variables se relacionan entre sí, las reglas de la gramática gobiernan la construcción de expresiones bien formadas en tanto que, la lógica define las relaciones (igualdad, vinculación, u ordenamiento) entre las combinaciones de valores de las diferentes expresiones del lenguaje de datos.

V.3.4 Reducción

Este componente cobra relevancia cuando se trabaja con volúmenes grandes de datos, ya que consiste en aplicar técnicas estadísticas u otros métodos para resumir o simplificar los datos de manera de lograr una representación eficiente. Una lista de los tipos de token y su frecuencia es una representación más eficiente que la tabulación de todas las ocurrencias. La lista reemplaza duplicaciones por una frecuencia y dado que una representación puede ser creada desde otra nada se pierde. No obstante, en muchas técnicas estadísticas para agregación de unidades de análisis (coeficientes de correlación, parámetros de distribuciones, e índices) la información se pierde. En búsquedas

cualitativas los resúmenes y rearticulaciones tienen un efecto similar puesto que reducen la diversidad del texto a lo que realmente importa.

V.3.5 Inferencia

Las inferencias abductivas distinguen el análisis de contenido desde otras formas de indagación. El inferir abductivamente un fenómeno contextual desde los textos mueve el análisis fuera de los datos. Esto cierra la brecha entre las explicaciones descriptivas de los textos y lo que estos significan, refieren, implican, provocan o causan. A diferencia de las inferencias inductivas y deductivas, las inferencias abductivas requieren garantías que a su vez pueden ser sustentadas por evidencia. En análisis de contenido, las garantías las proveen los constructos analíticos que están sustentados por todo lo que se conoce sobre el contexto.

V.3.6 Narración

Para que los resultados de un estudio de análisis de contenido sean comprensibles por terceros es necesario que el analista haga una narración de las respuestas a sus preguntas de investigación. Esta narración puede consistir en: explicar el significado práctico de los descubrimientos, explicar las contribuciones que hacen los descubrimientos a la literatura disponible, discutir lo apropiado del uso del análisis de contenido en lugar de técnicas observacionales directas, o bien hacer recomendaciones para acciones legales, prácticas, o investigación posterior.

V.4 FIABILIDAD

Los datos son el elemento base en el que se confía para el razonamiento, discusión y cálculo. Para sustentarse sobra una base indiscutible es necesario estar seguros de que los datos se han generado con todas las precauciones concebibles contra las distorsiones, contaminaciones o sesgos intencional o accidental, y que significan lo mismo para cualquiera que los use. La fiabilidad provee la garantía que los resultados de una investigación pueden ser duplicados, esto es que ningún (o solo una cantidad insignificante) ruido extraño ha tenido influencia en el proceso y contaminado los datos o perturbado los resultados de la investigación (Krippendorff, 2004).

V.4.1 Tipos de Fiabilidad

La Tabla 5.1 muestra los 3 tipos de fiabilidad (estabilidad, reproductibilidad, y exactitud) que se pueden distinguir si se considera la manera de generar los datos de fiabilidad. Los datos de fiabilidad son datos adicionales a los datos cuya fiabilidad está en duda. Los datos de fiabilidad se obtienen duplicando el esfuerzo de investigación bajo diversas condiciones (por ejemplo, usando investigadores con personalidades diferentes).

Fiabilidad	Diseño	Causas de Desacuerdos	Solidez
Estabilidad	test-retest	Inconsistencias intra observador	Debil
Reproductibilidad	test-test	Inconsistencias intra observador + desacuerdos inter observadores	Media
Precisión	test-standard	Inconsistencias intra observador + desacuerdos inter observadores + deviaciones desde el estandár	Fuerte

Tabla 5.1.Tipos de fiabilidad

La estabilidad es el grado en el cual un proceso se mantiene sin cambios a lo largo del tiempo. Se la mide como la extensión hasta la cual un proceso de codificación o medición conduce a los mismos resultados en pruebas repetidas. Los datos para evaluar la estabilidad se crean bajo las condiciones test-retest, es decir que el mismo observador relee, re categoriza o re analiza el mismo texto (usualmente luego de un tiempo) o el mismo dispositivo de medición se aplica repetidamente sobre un conjunto de objetos. Bajo las condiciones test-retest la falta de fiabilidad se manifiesta por las variaciones en el desempeño del observador o el instrumento de medición. En el caso de seres humanos las variaciones de desempeño (también conocidas como desacuerdo intra observador o inconsistencia interna) se deben a la falta de seguridad, descuidos, distracciones, dificultades para comprender instrucciones escritos, o la tendencia a relajar los estándares de desempeño ante el cansancio. La estabilidad es insuficiente como criterio para aceptar los datos como fiables.

La reproductibilidad es el grado en el cual un proceso puede ser replicado por analistas diferentes trabajando en condiciones diversas, en diferentes localizaciones, o usando instrumentos de medición distintos pero funcionalmente equivalentes. La reproductibilidad requiere que los datos de fiabilidad se obtengan bajo condiciones test-test, como por ejemplo dos o más individuos trabajando de manera independiente aplican el mismo conjunto de instrucciones a las mismas unidades de análisis. El desacuerdo entre los observadores se debe tanto a las inconsistencias intra observador como a las diferencias

inter observadores en la interpretación y aplicación de las instrucciones de grabación. La reproductibilidad es un indicador más fuerte de fiabilidad que la estabilidad.

La precisión es el grado con el cual un proceso cumple sus especificaciones y conduce a los resultados para los que fue diseñado a conducir. Para establecer la precisión los datos tienen que obtenerse bajo condiciones test-estandar, es decir se debe comparar el desempeño de uno o más procesos de producción de datos con el desempeño de un proceso tomado como correcto. El desacuerdo observado entre los dos tipos de desempeño se debe a las inconsistencias intra observador, diferencias inter observadores, y desviaciones desde el estándar. La precisión es la prueba de fiabilidad más fuerte puesto que responde a las tres fuentes de variación.

V.4.2 Cálculo de la Fiabilidad

Para medir el desacuerdo entre los codificadores puede usarse el coeficiente α propuesto por Krippendorff (2004) que de manera general se define como:

$$\alpha = \frac{Do}{De}$$

Donde D_o es una medida del desacuerdo observado y D_e es una medida del desacuerdo que puede esperarse cuando el azar prevalece. Cuando el acuerdo es perfecto $D_o=0$ y α asume el valor 1 indicando fiabilidad perfecta. Cuando tanto el acuerdo como el desacuerdo son una cuestión de azar y los desacuerdos observados y esperados son iguales, los valores $D_o=D_e$ y $\alpha=0$ indican ausencia de fiabilidad. El coeficiente α puede asumir valores negativos sin embargo, dado que el objetivo es la máxima fiabilidad posible los valores negativos están demasiado alejados de donde las diferencias en fiabilidad importan. Los valores negativos son consecuencia de dos tipos de errores:

- Errores de muestreo: suceden por una variedad de razones pero para el coeficiente α se consideran solo los errores causados por muestras de tamaño pequeño.
- Desacuerdo sistemático: ocurre cuando los observadores acuerdan desacordar o seguir interpretaciones opuestas de las instrucciones dadas a ellos. El desacuerdo observado desvía de la fiabilidad perfecta pero el desacuerdo sistemático puede causar que los valores sean inferiores a los que se esperaría si el azar prevalece.

Los pasos para el cálculo del coeficiente α dependen del tipo de dato y la cantidad de observadores. A continuación se ilustra por medio de un ejemplo tomado de

Krippendorff (2004) la manera de proceder cuando se tienen múltiples observadores, varias categorías nominales y valores faltantes. El interés en este caso particular del cálculo del coeficiente α se debe a que en el estudio de análisis de contenido documentado en el capítulo VII se trabajó con dos codificadores, una variable nominal que puede asumir 43 valores diferentes, y donde además para algunas unidades de grabación, los codificadores inicialmente no especificaban el valor de la variable nominal.

Primer paso

Se construye la matriz de datos de fiabilidad que contiene los valores asignados por los observadores a las unidades de análisis. Con *m* codificadores y *r* unidades la matriz de fiabilidad sería de *m*x*r* elementos.

Unidades de grabación *u*:	1	2	3	4	5	6	7	8	9	10	11	12
Observador A:	☺	☼	♠	♠	☼	☺	■	☺	☼			
Observador B:	☺	☼	♠	♠	☼	☼	■	☺	☼	♫		
Observador C:		♠	♠	♠	☼	♠	■	☼	☼	♫	☺	♠
Observador D:	☺	☼	♠	♠	☼	■	■	☺	☼	♫	☺	
Número de valores *mu*:	3	4	4	4	4	4	4	4	4	3	2	1

Tabla 5.2. Ejemplo de matriz de datos de fiabilidad

En la matriz de ejemplo de la Tabla 5.2 se tiene 4 observadores y 12 unidades de análisis por lo cual la dimensión de la matriz es 4x12=48. Para considerar los valores faltantes en la matriz de fiabilidad (7 de un total de 48 celdas) se agrega una fila que lista el número m_u de valores asignados a la unidad *u*.

Segundo paso

Consiste en construir la matriz de coincidencias observadas que cuenta el número de valores emparejables dentro de las unidades de análisis. Considerando que a partir de m_u valores se pueden formar $m_u(m_u\text{-}1)$ pares de valores, para que los valores de una unidad sean emparejables el valor de *mu* debe ser mayor a 1. En la unidad 1 existen m_1 $(m_1\text{-}1)=$ 3(3-1)=6 pares del tipo ☺-☺. La unidad 2 contiene 3 ☼ y 1 ♠ por lo que se puede formar un total de $m_2(m_2\text{-}1)=4(4\text{-}1)=12$ pares, de los cuales 3(3-1)=6 son del tipo ☼-☼, 3*1=3 del tipo ☼-♠ y finalmente 1*3=3 del tipo ♠-☼. Un caso extremo se puede

ver en la unidad 6 donde se tiene $m_6(m_6\text{-}1)=4(4\text{-}1)=12$ pares y todos los valores son diferentes. El otro extremo lo constituye la unidad 12 ya que ♠ no participa en ningún par, tal como lo indica $m_{12}(m_{12}\text{-}1)=1(1\text{-}1)=0$.

Para que cada valor contribuya exactamente una entrada en la matriz de coincidencia observada cada uno de los *mu*(*mu*-1) pares de valores posibles contenidos en la unidad *u* deben contribuir 1/(*mu*-1) a la matriz de coincidencia. La estructura genérica de una matriz de coincidencias observadas se muestra en la Tabla 5.3.

Valores:	1	.	k	.	.	v	
1	O_{11}	.	O_{1k}	.	.	O_{1v}	n_1
.	.	.					.
.	.	.					.
c	O_{c1}	.	O_{ck}	.	.	O_{cv}	$n_c=\sum_k o_{ck}$
.	.	.					
v	O_{v1}	.	O_{vk}	.	.	Ovv	n_v
	n_1	.	n_k	.	.	n_v	$n=\sum_c \sum_k n_{ck}$

Donde $o_{ck} = \sum_u \frac{Número\ de\ pares\ c-k\ en\ u}{m_u - 1}$

Tabla 5.3. Estructura de la matriz de coincidencias observadas (Krippendorff, 2004)

Las 5 coincidencias ☺-☺ de la unidad 1 contribuyen 6/(3-1)= 3 a la celda ☺-☺ de la matriz. Por su parte, los tres tipos de pares en la unidad 2 contribuyen 6/(4-1)=2 a la celda ☼-☼, 3/(4-1)=1 a la celda ☼-♠, y 1 a la celda ♠-☼ de la matriz. Los 12 pares de valores de la unidad 6 añaden 1/(4-1)=1/3 a cada una de las 12 celdas de la matriz. Las contribuciones de las unidades 1, 2 y 6 se muestran en las primeras tres matrices debajo. La cuarta corresponde a la suma de las contribuciones de todas las unidades, aquí se debe notar que el valor ♠ de la unidad 12 no contribuye a la matriz porque no hay nada contra que compararlo.

	Unidad 1					Unidad 2					Unidad 6					Suma de las 12					
	☺	☼	♠	■	♫	☺	☼	♠	■	♫	☺	☼	♠	■	♫	☺	☼	♠	■	♫	
☺	3											1/3	1/3	1/3		7	4/3	1/3	1/3		9=n☺
☼							2	1			1/3		1/3	1/3		4/3	10	4/3	1/3		13=n☼
♠							1				1/3	1/3		1/3		1/3	4/3	8	1/3		10=n♠
■											1/3	1/3	1/3			1/3	1/3	1/3	4		5=n■
♫																				3	3=n♫
																9	13	10	5	3	40=n

Tabla 5.4. Ejemplo de matrices de coincidencias observadas

Tercer paso

Se calcula coeficiente α de la siguiente manera:

$$\alpha_{nominal}=1-\frac{Do}{De}=1-(n-1)\frac{n-\sum_c Occ}{n^2-\sum_c n_c^2}=1-(40-1)\frac{40-(7+10+8+4+3)}{40^2-(9^2+13^2+10^2+5^2+3^2)}=0.734$$

Lo primero a notar en la ecuación anterior es que la definición genérica de α puede reformularse de manera tal que no sea necesario construir la matriz de coincidencia esperada.

Antes de poder establecer un valor de juicio con respecto al resultado alcanzado para el coeficiente α, en la situación ejemplo desarrollada, se debe tener en cuenta que Krippendorff (2004) recomienda:

- Confiar en variables con valores de fiabilidad superiores a α=0.800.
- Las variables con valores de fiabilidad entre α=0.667 y α=0.800 tendrían que ser consideradas solo para elaborar conclusiones de carácter tentativo.

Así, en base a las dos recomendaciones de Krippendorff, para el ejemplo desarrollado en los párrafos previos, se puede decir que el nivel de acuerdo entre los cuatro observadores no es suficiente para aceptar los datos como fiables. Es decir, es poco probable que los valores de la matriz de datos de fiabilidad obtenidos con los observadores A, B, C y D se puedan repetir si el estudio de análisis de contenido se lleva a cabo nuevamente.

CAPÍTULO VI
ETAPAS DE UN PROYECTO DE MINERÍA DE TEXTOS

VI.1 METODOLOGÍA CRISP-DM

De acuerdo con Athaudal *et al.* (2009) la metodología CRISP-DM (Proceso Estándar Trans-Industrial para Minería de Datos) se ha convertido en un estándar para el desarrollo de proyectos de descubrimiento de conocimiento en bases de datos (KDD) y una de sus principales ventajas es que puede usarse con cualquier herramienta de software y para resolver cualquier problema de minería de datos.

CRISP-DM fue concebida en 1996, a partir de la iniciativa de tres grandes empresas vinculadas a la minería de datos (Daimler-Benz, ISL, y NCR), con la intención de proveer un modelo de proceso estándar, no propietario y disponible libremente para todos los practicantes de minería de datos (Chapman *et al*., 2000). Si bien en 2006 se anunció que se trabajaría en pos de desarrollar la versión 2.0 de CRISP-DM, hasta el momento no existen novedades al respecto por lo cual para su descripción se tendrá en cuenta (Chapman *et al*., 2000), que es el documento oficial que describe CRISP-DM 1.0.

El ciclo de vida de un proyecto de minería de datos consiste de 6 fases tal como se muestra en la Figura 6.1. La secuencia entre las fases no es rígida porque usualmente se requieren movimientos hacia adelante y hacia atrás de cada fase. Esto depende de la salida de cada fase, y de la fase o tarea de la fase que tiene que realizarse a continuación.

El círculo externo de la Figura 6.1 simboliza la naturaleza cíclica de la minería de datos. La minería de datos no finaliza con el despliegue o aplicación de la solución porque las lecciones aprendidas a lo largo del proceso pueden dar lugar a nuevas preguntas de negocio o a un refinamiento de las ya existentes. De esta manera los procesos de minería de datos subsecuentes se benefician de las experiencias previas.

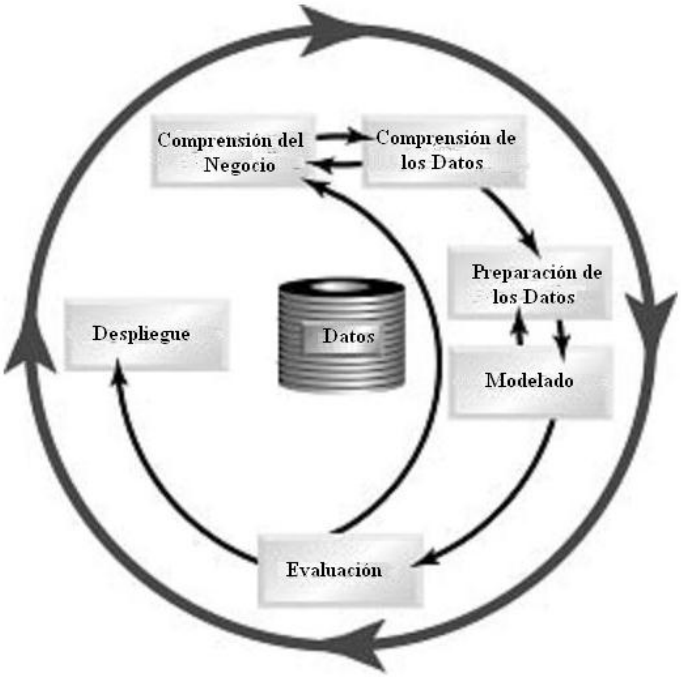


Figura 6.1 Fases del modelo de referencia CRISP-DM (Chapman *et al.*, 2001)

A continuación se describen de manera breve las actividades involucradas y las salidas esperadas en cada una de las fases de CRISP-DM.

VI.1.1 Fase1: Comprensión del Negocio

VI.1.1.1 Determinar los objetivos del negocio

Descripción de la tarea

El primer objetivo del analista de datos es comprender a fondo, desde una perspectiva de negocio, lo que el cliente quiere lograr. Frecuentemente el cliente tiene muchos objetivos y restricciones compitiendo entre sí que deben ser balanceadas. Una posible consecuencia de descuidar esta etapa es el destinar mucho esfuerzo a la producción de respuestas correctas para preguntas equivocadas.

Salidas

Antecedente: se registra la información conocida sobre la situación de negocio de la organización al comienzo del proyecto.

Objetivos de negocio: se describe el objetivo primario del cliente, desde una perspectiva de negocio. Además del objetivo de negocio primario típicamente hay preguntas de negocio que al cliente le gustaría tratar.

Criterio de éxito de negocio: consiste en describir (desde una perspectiva de negocio) el criterio para decidir si un resultado es exitoso o de utilidad para el proyecto. El criterio de éxito puede ser específico y capaz de ser medido objetivamente o puede ser general y subjetivo. En el último caso se debe indicar quien es la persona encargada de decidir si se cumplió con el criterio de éxito.

VI.1.1.2 Evaluar la situación

Descripción de la tarea

Esta tarea comprende una especificación más detallada sobre todos los recursos, restricciones, y otros factores que deberían ser considerados al determinar el objetivo del análisis de datos y el plan del proyecto.

Salidas

Inventario de recursos: se listan todos los recursos del proyecto, incluyendo: personal (expertos en negocios, expertos en datos, soporte técnico, y personal de minería de datos), datos (datos extraídos o acceso a datos operacionales), recursos de computación (plataformas de hardware) y software (herramientas de minería de datos y cualquier otro software relevante).

Requerimientos, supuestos y restricciones: elaboración de una lista de todos los requerimientos para el proyecto incluyendo fecha de finalización, comprensibilidad y calidad de los recursos, seguridad y asuntos legales. La lista de supuestos puede incluir cuestiones sobre los datos que pueden chequearse durante la minería de datos pero, también puede comprender supuestos no comprobables sobre el área de negocio en la que se desarrolla el proyecto. Por su parte, las restricciones pueden referirse a la disponibilidad de recursos pero también a restricciones tecnológicas tales como el tamaño de los datos que es práctico para modelado.

Riesgos y contingencias: comprende la enumeración de todos los riesgos o eventos que podrían causar que el proyecto se retrase o fracase. Es necesario que también se indiquen los planes de contingencia correspondientes, esto es las acciones a tomarse si los riesgos se convierten en una realidad.

Terminología: es un glosario de la terminología relevante al proyecto que comprende tanto términos necesarios para comprender el área de negocio en la cual se desarrolla el proyecto, como la terminología de minería de datos.

Costos y beneficios: se construye un análisis de costo-beneficio donde se comparan los costos del proyecto con los potenciales beneficios para el negocio si el proyecto es exitoso.

VI.1.1.3 Determinar los objetivos de la minería de datos

Descripción de la tarea

Un objetivo de negocio define los objetivos en terminología de negocios. Un objetivo de minería define los objetivos del proyecto en términos técnicos. Así por ejemplo un objetivo de negocio podría ser "Incrementar las ventas por catálogos a los clientes existentes". El objetivo de minería de datos asociado podría ser "Predecir cuantos aparatos pequeños un cliente comprará, teniendo en cuenta sus compras en los últimos tres años, información demográfica (edad, salario, ciudad, etc.) y el precio de cada ítem".

Salidas

Objetivos de minería de datos: descripción de las salidas del proyecto que permiten lograr los objetivos de negocio.

Criterio de éxito de la minería de datos: descripción en términos técnicos del criterio para decidir si un resultado del proyecto es exitoso, por ejemplo un cierto nivel de exactitud predictiva.

VI.1.1.4 Producir el plan del proyecto

Descripción de la tarea

Se debe describir el plan para lograr los objetivos de minería de datos y en consecuencia alcanzar los objetivos de negocio. El plan debería especificar los pasos a seguir durante el resto del proyecto incluyendo una selección inicial de técnicas y herramientas.

Salidas

Plan del proyecto: especifica todas las etapas a ser ejecutadas junto con la duración, recursos necesarios, entradas, salidas y dependencias. Cuando es posible se deben hacer explicitas las iteraciones en el proceso de minería de datos, como ser repeticiones de las fases de modelado y evaluación. Como parte del plan del proyecto es importante hacer un análisis de las dependencias entre el cronograma de actividades y los riesgos. Una cuestión importante a tener en cuenta es que el plan del proyecto es un documento dinámico, es decir que al finalizar cada fase se debe hacer una revisión del progreso y los logros de manera tal de actualizar el plan del proyecto en función de esta evaluación.

Evaluación inicial de técnicas y herramientas: la evaluación de técnicas y herramientas se realiza al principio del proyecto puesto que las herramientas y técnicas seleccionadas influyen en el resto del proyecto.

VI.1.2 Fase 2: Comprensión de los Datos

VI.1.2.1 Recolectar datos iniciales

Descripción de la tarea

Se adquiere o accede a los datos listados en los recursos del proyecto. Cuando se adquieren datos desde múltiples fuentes la integración se convierte en una cuestión adicional a ser considerada, aquí o posteriormente en la fase de preparación de datos.

Salidas

Reporte de la recolección de datos inicial: se lista la base de datos (o bases de datos) adquirida junto con su localización dentro del proyecto, los métodos usados en la adquisición y cualquier problema encontrado. El registrar los problemas y sus soluciones ayuda con futuras replicaciones del proyecto o bien con la ejecución de proyectos similares en el futuro.

VI.1.2.2 Describir los Datos

Descripción de la tarea

Esta tarea consiste en examinar las propiedades superficiales de los datos adquiridos y presentar los resultados.

Salida

Reporte de descripción de los datos: de los datos adquiridos se describen la cantidad (como ser el numero de registros y campos en cada tabla), las identidades de los campos y otras características superficiales que han sido descubiertas.

VI.1.2.3 Explorar los datos

Descripción de la tarea

Esta tarea tiene en cuenta las preguntas de minería de datos que pueden ser abordadas mediante herramientas de consulta, visualización y reporte. Esto incluye: distribución de atributos clave, relación entre pares o un pequeño número de atributos, resultados de agregaciones simples, propiedades de sub poblaciones significantes, y análisis estadístico simple. Estos tipos de análisis pueden abordar directamente los

objetivos de minería de datos, pueden contribuir a refinar los reportes de calidad o descripción de los datos, o pueden orientar la transformación u otro tipo de preparación de datos necesaria para proseguir con el análisis.

Salida

Reporte de exploración de los datos: describe los resultados de la exploración incluyendo primeros descubrimientos o hipótesis iniciales y su impacto en el resto del proyecto. Puede incluir gráficos que indican características de los datos o llevan a sub conjuntos de datos interesantes para ser examinados posteriormente.

VI.1.2.4 Verificar la calidad de los datos

Descripción de la tarea

Se hace un análisis de la calidad de los datos abordando preguntas tales como: ¿Los datos son completos (cubren todos los casos requeridos)? ¿Los datos son correctos o contienen errores? y si hay errores ¿qué tan comunes son? ¿Hay valores faltantes en los datos? Si hay valores faltantes ¿cómo se los representa, dónde ocurren y cuán comunes son?

Salidas

Reporte de calidad de los datos: si se identifican problemas de calidad tiene que indicarse las posibles soluciones. Las soluciones generalmente dependen del conocimiento de los datos y del negocio.

VI.1.3 FASE 3: Preparación de los Datos

VI.1.3.1 Seleccionar los datos

Descripción de la tarea

Se tiene que decidir los datos a usar en el análisis. Los criterios a tener en cuenta incluyen relevancia para los objetivos de minería de datos, calidad y restricciones técnicas tales como límites en el volumen o tipo de datos.

Salidas

Razones para la inclusión/ exclusión: hay que enumerar los datos incluidos/excluidos junto con los motivos para tales decisiones.

VI.1.3.2 Limpiar los datos

Descripción de la tarea

La limpieza permite incrementar la calidad de los datos al nivel requerido por las técnicas de análisis. Esto puede involucrar la selección de subconjuntos de datos limpios, la inserción de valores por defecto o técnicas más ambiciosas como la estimación de datos faltantes mediante modelado.

Salidas

Reporte de limpieza de datos: en este reporte de describen las decisiones y acciones tomadas para tratar los problemas de calidad reportados durante la tarea de verificación de la calidad de los datos de la fase de comprensión de los datos. Las transformaciones de los datos para propósitos de limpieza y el posible impacto en los resultados del análisis deberían ser considerados.

VI.1.3.3 Construir los datos

Descripción de la tarea

La tarea incluye operaciones de preparación de datos constructivas tales como la producción de atributos derivados, registros nuevos o valores transformados para atributos existentes.

Salidas

Atributos derivados: son atributos construidos a partir de uno o más atributos existentes en el mismo registro. Como por ejemplo: área= largo*ancho.

Registros generados: se refiere a la creación de registros completamente nuevos. Por ejemplo se pueden crear registros para clientes que no hayan realizado compras durante un determinado año o período. Si bien estos registros no forman parte del conjunto de datos inicial para propósitos de modelado puede ser necesario representar el hecho de que ciertos clientes no realizaron compras.

VI.1.3.4 Integrar los datos

Descripción de la tarea

Comprende los métodos para combinar información desde múltiples tablas o registros para crear nuevos registros o valores.

Salidas

Datos Combinados: la combinación de las tablas se refiere a unir dos o más tablas que tienen información sobre los mismos objetos. Por ejemplo: una cadena de ventas por menor tiene una tabla con información sobre las características generales de cada tienda (espacio en el suelo, tipo de centro comercial, etc.), otra tabla con datos resumidos de las ventas (ganancia, porcentaje de cambios en las ventas en relación a años anteriores, etc.) y una tercera tabla con información sobre la demografía del área circundante. Cada una de estas tablas contiene un registro para cada tienda. Las tablas pueden combinarse en una nueva tabla con un registro por tienda que contiene los campos combinados de las tablas fuentes.

Agregaciones: incluye las operaciones donde nuevos valores se computan resumiendo información desde tablas y/o registros. Un ejemplo puede ser el convertir una tabla de compras de consumidores donde se tiene un registro por compra en una nueva tabla que contiene un registro por consumidor con campos tales como número de compras, porcentaje promedio de compras, porcentaje de ítems en promoción, etc.

VI.1.3.5 Formatear los datos

Descripción de la tarea

Las transformaciones de formateo se refieren principalmente a modificaciones sintácticas de los datos que no cambian su significado pero, que podrían ser requeridas por la herramienta de modelado.

Salidas

Datos reformateados: algunas herramientas tienen requerimientos en el orden de los atributos, como ser el primer campo tiene que ser un identificador único para cada registro o el último campo debe ser la salida que el modelo va predecir. También podría ser necesario modificar el orden de los registros en el conjunto de datos ya sea porque la herramienta necesita registros ordenados de acuerdo a un determinado atributo o bien un ordenamiento totalmente aleatorio.

VI.1.4 FASE 4: Modelado

VI.1.4.1 Seleccionar la técnica de modelado

Descripción de la tarea

Como primer paso de la fase de modelado se tiene que decidir la técnica de modelado que se usará. Mientras la selección de la herramienta de minería sucede en la

fase de comprensión del negocio, esta tarea se refiere a seleccionar la técnica de modelado específica, por ejemplo construcción de un árbol de decisión con C4.5 o generación de una red neuronal con propagación hacia atrás. Si múltiples técnicas se aplican al problema, la tarea debe realizarse de manera separada para cada técnica.

Salidas

Técnica de modelado: se documenta la técnica de modelado que se usará.

Supuestos de modelado: se deben registrar los supuestos específicos de cada técnica de modelado, como ser: los atributos deben tener distribución uniforme, no se permiten valores faltantes, el atributo de clase debe ser simbólico, etc.

VI.1.4.2 Generar el diseño de la prueba

Descripción de la tarea

Antes de construir el modelo se necesita generar un procedimiento o mecanismo para probar la calidad y validez del modelo. Por ejemplo en tareas de minería de datos supervisadas tales como la clasificación, es común usar el ratio de error como una medida de la calidad de los modelos. Por consiguiente, se suele separar el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba, el modelo se construye con el conjunto de entrenamiento y la estimación de la calidad se hace con el conjunto de prueba.

Salidas

Diseño de prueba: describe el plan previsto para el entrenamiento, la prueba y validación de los modelos. Un componente primario del plan es decidir cómo dividir el conjunto de datos en los conjuntos de entrenamiento, prueba y validación.

VI.1.4.3 Construir el modelo

Descripción de la tarea

Consiste en ejecutar la herramienta de modelado en un conjunto de datos preparado para crear uno o más modelos.

Salidas

Ajustes de los parámetros: con cualquier herramienta de modelado frecuentemente existe un gran número de parámetros que pueden ser ajustados. Se tienen que listar los parámetros, los valores seleccionados, y los motivos en los que se basó la selección.

Modelos: son los modelos producidos por la herramienta de modelado.

Descripción del modelo: consiste en describir el o los modelos obtenidos, la interpretación que se hace de ellos y cualquier dificultad encontrada con respecto al significado.

VI.1.4.4 Evaluar el modelo

Descripción de la tarea

El ingeniero de minería de datos interpreta los modelos de acuerdo a su conocimiento del dominio, el criterio de éxito de la minería de datos y el diseño de prueba. Esta tarea se vincula con la fase de evaluación posterior. Mientras el ingeniero juzga el éxito de la aplicación de las técnicas de modelado y descubrimiento más técnicamente, en la fase de evaluación contacta analistas de negocio para discutir los resultados de la minería de datos en el contexto de negocio. Además, esta tarea solo considera los modelos mientras que la fase de evaluación también tiene en cuenta otros resultados producidos a lo largo del proyecto.

El ingeniero de minería de datos trata de crear un ranking de los modelos de acuerdo al criterio de evaluación establecido. En la medida de lo posible el ingeniero de minería de datos tiene en cuenta los objetivos de negocio y el criterio de éxito de negocio.

Salidas

Evaluación del modelo: se resumen los resultados de la tarea de modelado haciendo una lista de las cualidades de los modelos generados y ordenando su calidad el uno en relación al otro.

Ajustes de los parámetros revisados: se revisan los ajustes realizados a los parámetros y se los refina para la próxima ejecución de la tarea de Construir el Modelo. Es preciso que se documenten todas las iteraciones de la construcción y evaluación del modelo hasta que encontrar el o los mejores modelos.

VI.1.5 Fase 5 : Evaluación

VI.1.5.1 Evaluar resultados

Descripción de la tarea

Los pasos de evaluación anteriores tratan con factores tales como la exactitud y generalidad del modelo. Este paso evalúa en qué grado el modelo satisface los objetivos de negocio y busca determinar si hay algunas razones de negocio por las cuales el modelo es deficiente. Otra opción de evaluación es probar el o los modelos en aplicaciones de prueba en el mundo real si las restricciones de tiempo y presupuesto lo permiten. Además del modelo en la evaluación se valoran otros resultados de minería de datos que no

necesariamente se relacionan a los objetivos de negocio originales y descubrimientos que pueden revelar desafíos adicionales, información o pistas para trabajos futuros.

Salidas

Evaluación de los resultados de minería de datos con respecto a los criterios de éxito de negocio: se tienen que resumir los resultados de la evaluación en términos de criterios de éxito de negocio incluyendo una declaración final de si el proyecto cumplió con los objetivos de éxito iniciales.

Modelos aprobados: los modelos generados que satisfacen los criterios de éxito de negocio se convierten en modelos aprobados.

VI.1.5.2 Revisar el proceso

Descripción de la tarea

En este punto con suerte el modelo es satisfactorio y satisface las necesidades de negocio. Lo que se necesita ahora es una revisión minuciosa del proceso de minería para determinar si hay algún factor importante o tarea que de alguna manera ha sido pasada por alto. Esta revisión cubre también cuestiones de aseguramiento de la calidad, por ejemplo: ¿El modelo se construyó correctamente? ¿Se consideraron únicamente atributos cuyo uso estaba permitido y que están disponibles para análisis futuros?

Salidas

Revisión del proceso: es un resumen del proceso de revisión donde se señalan las actividades que se han omitido y/o deberían ser repetidas.

VI.1.5.3 Determinar próximos pasos

Descripción de la tarea

Teniendo en cuenta los resultados de la evaluación y la revisión del proyecto se tiene que decidir si finalizar el proyecto y continuar con la fase de despliegue, si iniciar nuevas iteraciones, o si establecer nuevos proyectos de minería de datos. El análisis de los recursos y el presupuesto restante influye en las decisiones tomadas.

Salidas

Lista de posibles acciones: enumera todas las posibles acciones futuras junto con los argumentos a favor y en contra de cada opción.

Decisión: describe la decisión de cómo proceder junto con el motivo que llevo a la decisión.

VI.1.6 FASE 6: Despliegue

VI.1.6.1 Plan de despliegue

Descripción de la tarea

Esta tarea toma los resultados de la evaluación y establece una estrategia para utilizar o desplegar en la empresa los resultados de la minería de datos. Si se identificó un procedimiento para crear modelos relevantes, este procedimiento se documento para utilización posterior.

Salida

Plan de despliegue: es un resumen de la estrategia de despliegue que incluye los pasos necesarios y como realizarlos.

VI.1.6.2 Plan de monitoreo y mantenimiento

Descripción de la tarea

El monitoreo y mantenimiento son dos cuestiones importantes si el resultado de la minería de datos se vuelve parte del día a día del negocio y su ambiente. Una preparación cuidadosa de una estrategia de mantenimiento ayuda a evitar largos periodos innecesarios de uso incorrecto de los resultados de la minería de datos. Para monitorear la utilización de los resultados de la minería de datos se necesita un plan detallado del proceso de monitoreo.

Salidas

Plan de monitoreo y mantenimiento: resume la estrategia de monitoreo y mantenimiento incluyendo los pasos necesarios y como realizarlos.

VI.1.6.3 Producir el reporte final

Descripción de la tarea

Al final del proyecto, el líder del proyecto y su equipo escriben un reporte final. Dependiendo del plan de despliegue este reporte puede ser solo un resumen del proyecto y sus experiencias (si no ha sido documentado durante su desarrollo) o puede ser presentación final y exhaustiva de los resultados de la minería de datos.

Salidas

Reporte final: es el reporte escrito final del proceso de minería de datos. Este incluye todos los productos previos y resúmeme los resultados para presentarlos de manera organizada.

Presentación final: al finalizar el proyecto se realiza una reunión donde los resultados se presentan de manera verbal al cliente.

VI.1.6.4 Revisar el Proyecto

Descripción de la tarea

Evaluar lo que salió bien y lo que salió mal, lo que se hizo bien y lo que necesita ser mejorado.

Salidas

Documentación de la experiencia: se tienen que resumir las experiencias más importantes del proyecto. Por ejemplo: dificultades, aproximaciones erróneas o sugerencias para seleccionar técnicas de minería de datos en situaciones similares pueden ser parte de esta documentación.

VI.2 APLICACIÓN DE CRISP-DM A PROBLEMAS DE MINERÍA DE TEXTOS

En la exploración bibliográfica realizada en el desarrollo del presente trabajo de investigación no se localizaron antecedentes sobre una metodología para conducir estudios de minería de textos. Si bien la metodología CRISP-DM fue concebida para servir de guía durante el desarrollo de proyectos de minería de datos, en su tesis de maestría Magalhães (2002) demostró que la metodología CRISP-DM puede utilizarse para abordar problemas de minería de textos si se la complementa con las recomendaciones realizadas por Palazzo (Loh, 2002), Tan (Tan, 1999) y Halliman (Magalhães, 2002). Esto coincide con los planteos de autores tales como (Feldman & Sanger, 2007; García & Calvo, 2006; Weiss *et al.*, 2005) que consideran a la minería de textos como una especialización de la minería de datos que puede usar los mismos métodos que la minería de datos en la medida que los datos textuales se transformen en un formato estructurado o semi estructurado.

VI.2.1 Enfoques para Descubrimiento de Conocimiento en Textos

La tareas planteadas por Palazzo, Tan y Halliman permiten aplicar métodos de minería de datos a características (como ser valores de atributos o palabras clave)

extraídas desde documentos de texto. Sin embargo, estas estrategias (surgidas de la experiencia de sus autores) no pueden considerarse un metodología sino más bien un marco de trabajo o esbozo que guía a los analistas a lo largo del proceso de minería de textos indicando los principales pasos (Magalhães, 2002).

VI.2.1.1 Enfoque de Tan

Según Tan (1999) un proceso de minería de textos consiste de dos fases: refinamiento de textos y destilación de conocimiento.

El refinamiento de textos transforma documentos de texto de formato libre en un formato intermedio que puede ser semi estructurado (Gráfico conceptual) o estructurado (Base de datos). Se dice que el formato intermedio se basa en documento cuando cada entidad representa un documento, y que se basa en concepto cuando cada entidad representa un objeto o concepto de interés en un dominio especifico.

La fase de destilación de conocimiento deduce patrones o conocimiento desde el formato intermedio. Cuando se aplica minería a un formato intermedio basado en documento se deducen patrones y relaciones a través de los documentos. La clasificación, *clustering* y visualización de documentos son ejemplos de minería a partir de un formato intermedio basado en documento. Cuando se mina un formato intermedio basado en concepto se deduce patrones y relaciones entre los objetos o conceptos. Operaciones de minería tales como modelado predictivo y descubrimiento de asociación pertenecen a esta categoría.

En la Figura 6.2 la flecha desde el cuadro "Forma intermedia basada en documento" a "Forma intermedia basada en concepto" señala que una forma intermedia basada en documento se puede transformar a una basada en concepto si se extrae información relevante de acuerdo a los objetos de interés en un dominio específico.

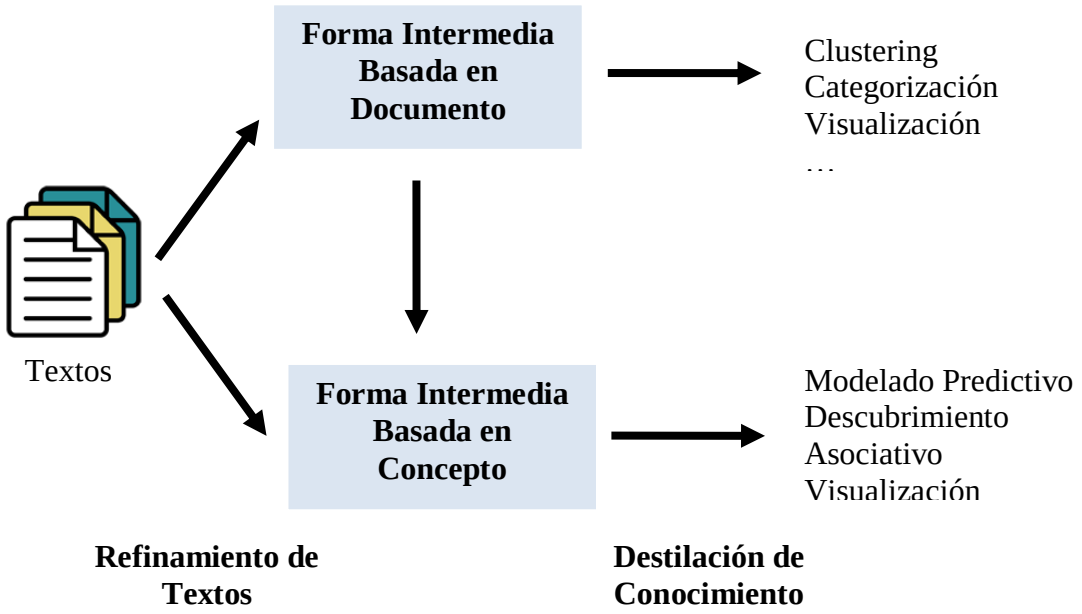


Figura 6.2 Marco de trabajo para minería de textos (Tan, 1999)

VI.2.1.2 Enfoque de Palazzo

Durante el proceso de descubrimiento de conocimiento en textos puede seguirse un enfoque reactivo o un enfoque proactivo (Loh *et al.*, 2000). En el primer caso el conocimiento se adquiere para resolver un problema específico del usuario, una necesidad resultante de un estado anómalo de conocimiento. En estos casos el usuario sabe lo que quiere y puede identificar la solución a su problema cuando la encuentra. En cambio, con el modo pro activo el propósito de la adquisición de conocimiento es exploratoria ya que se busca detectar problemas potenciales o bien oportunidades. El usuario no tiene un problema o necesidad definida, por lo cual a lo largo del proceso se debe descubrir, explorar, estudiar y probar hipótesis de una solución al problema.

De acuerdo con Loh *et al.* (2000) el enfoque de Palazzo para descubrimiento de conocimiento en textos es del tipo proactivo y consta de los siguientes pasos:

- **Selección de textos**: en primer lugar se tiene que seleccionar una colección de textos ya sea de manera manual o mediante técnicas automáticas tales como recuperación de información (localiza textos en base a palabras clave o términos) y clasificación (separa textos por asunto).
- **Análisis de la colección (todo o parte)**: se debe decidir si aplicar las técnicas de descubrimiento a toda o parte de la colección de textos. La separación en partes puede hacerse de manera automática o según un criterio definido por el usuario.

- **Análisis de grupos de textos (todo o parte)**: una buena manera de realizar esta tarea es analizando y extrayendo una lista de términos comunes en los textos (técnica de listado de conceptos clave).
- **Comparación de sub colecciones entre o en relación a la colección completa**: los resultados de aplicar las técnicas de listado de conceptos clave, diferencia y asociación a cada sub colección pueden compararse con los resultados de otras sub colecciones o de la colección como un todo.
- **Validación de hipótesis**: se aplican técnicas de resúmenes para leer e interpretar los resultados.
- **Realimentación**: se realimenta y refuerza el proceso completo hasta lograr el objetivo esperado.

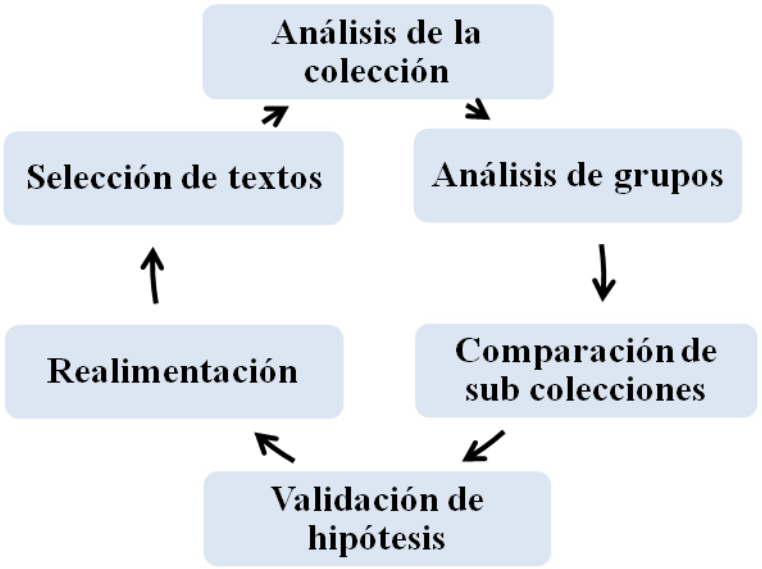


Figura 28 Marco de trabajo de Palazzo para minería de textos (Magalhães, 2002)

VI.2.1.3 Enfoque de Halliman

Para analizar informaciones textuales externas a las empresas Halliman divide el ambiente en partes denominadas fuerzas de ambiente. El uso de la minería de textos para estudiar las tendencias y la distribución de las fuerzas permite detectar amenazas y fortalecer oportunidades para las empresas (Magalhães, 2002). El marco de trabajo de Halliman se compone de las etapas mostradas en la Figura 6.4.

El proceso comienza con una comprensión del dominio de la empresa por medio del análisis de los competidores y las fuerzas que estos ejercen sobre sus actividades que pueden mejorar las tácticas y estrategias desarrolladas. Para cada fuerza ambiental se

asocian palabras clave. Por ejemplo a la fuerza gubernamental le correspondería “federal”, “senado”, y “congreso”; a la fuerza tecnológica puede corresponderle “tecnología” y “patente”; en tanto que a la fuerza de marketing se le puede asociar las palabras “mercado” y “marketing”.

Teniendo el dominio y las palabras clave el siguiente paso es recuperar los textos para verificar y analizar su contenido. Se realiza una selección mediante una búsqueda con palabras clave.

Se identifican y excluyen los textos no pertenecientes al dominio. Los archivos restantes se analizan, tomando las palabras clave extraídas para componer una planilla de cálculo. El resultado es clasificado de acuerdo a la cantidad de palabras clave encontradas en cada texto. Con los resultados de las planillas de cálculo se elaboran gráficos (distribución, palabras más usadas, tendencias, etc.) que facilitan el proceso de análisis de las informaciones obtenidas.

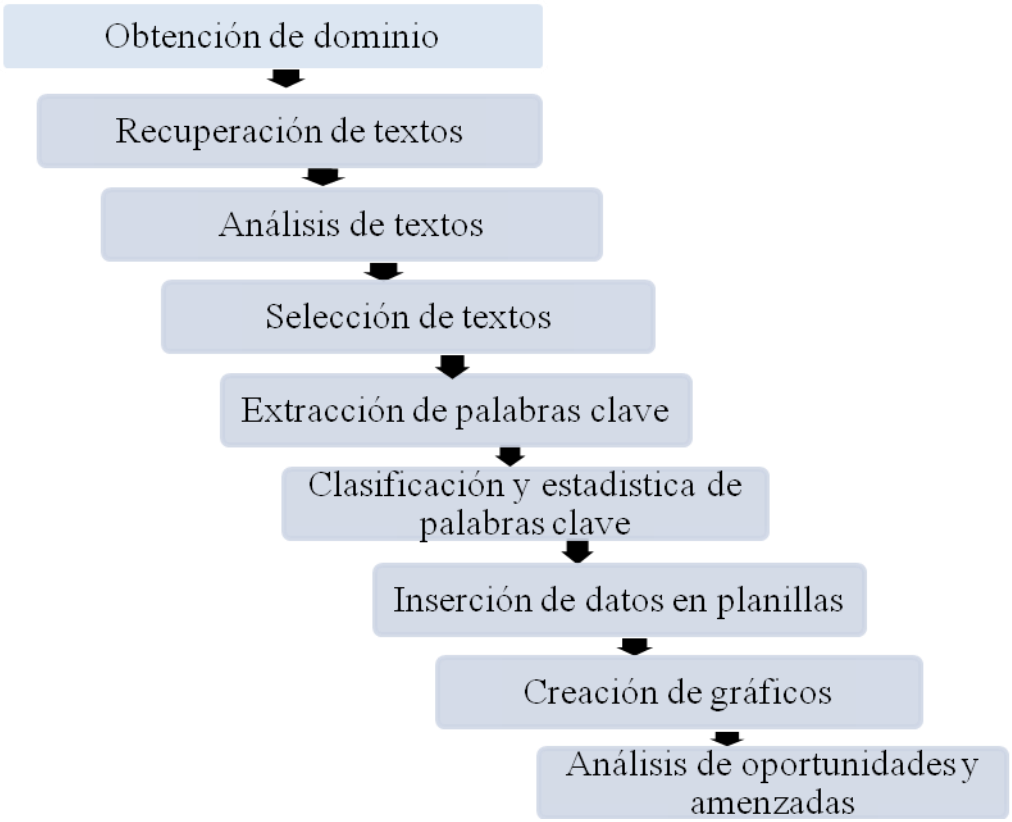


Figura 6.4. Marco de trabajo de Halliman para minería de textos (Magalhães, 2002)

Los pasos mostrados en la Figura 6.4 deberían conducir a una disminución del tiempo necesario para adquirir la información relevante, aunque esto depende de las habilidades y del conocimiento del analista.

VI.2.2 Adaptación de CRISP-DM a Problemas de Minería de Textos

En su trabajo de tesis Magalhães (2002) comprobó que la metodología CRISP-DM puede usarse para conducir proyectos de minería de textos en la medida que se la complemente con las estrategias de descubrimiento de conocimiento en textos propuestas por Tan, Halliman, y Palazzo. Los aportes que pueden hacer las estrategias tratadas en la sección 6.2.1 a la metodología CRISP-DM son los siguientes:

- **Fase Comprensión de datos**: el abordaje de Tan es de utilidad para importar los datos textuales a una base de datos estructurada sobre todo cuando la información esta almacenada en múltiples formatos de archivos. Este paso facilita la obtención de los primeros gráficos y descubrimientos.
- **Fase de Modelado**: siguiendo a Palazzo se pueden identificar grupos en los datos en base a la presencia de palabras clave para luego realizar un análisis minucioso de cada grupo.
- **Fases evaluación y Despliegue**: el marco de trabajo de Halliman puede usarse para clasificar los datos y para obtener diversos gráficos que les permiten a los administradores de la empresa analizar los resultados del proyecto de minería de una manera más ágil.

MARCO EMPÍRICO

El trabajo empírico desarrollado se describe a lo largo de tres capítulos cuyo contenido se indica de manera breve a continuación.

Capítulo VII: Documentación del Estudio de Análisis de Contenido

Se describe la forma en que se aplicó la técnica de análisis de contenido de modo que expertos humanos identifiquen las habilidades manifestadas por e-tutores al comunicarse con sus estudiantes en entornos de ACSC .

Capítulo VIII: Adaptaciones realizadas a los Recursos de GATE

Se describe las adaptaciones realizadas al recurso *ANNIE Sentence Splitter* de modo tal de poder reconocer oraciones en textos en español. También se indica como puede implementarse la eliminación de *stop words* ya que GATE no cuenta con un recurso dedicado exclusivamente a esa operación de pre procesamiento.

Capítulo IX: Documentación de la Aplicación de CRISP-DM

En este capítulo se describen las actividades de la metodología CRISP-DM que se realizaron para construir clasificadores que detecten automáticamente las habilidades manifestadas por docentes de ACSC al interactuar con sus estudiantes mediante herramientas de comunicación basadas en texto libre.

CAPÍTULO VII
DOCUMENTACIÓN DEL ESTUDIO DE ANÁLISIS DE CONTENIDO

VII.1 CUERPOS DE TEXTOS

El cuerpo de textos utilizado en el presente estudio se compone de la siguiente manera:

- Los mensajes que mediante la herramienta de foro de la plataforma Moodle intercambiaron los estudiantes y docentes de la asignatura Metodología de la investigación II, de la carrera Licenciatura en Sistemas de Información de la UNSE, durante una experiencia de ACSC realizada en el año 2010.
- Los diálogos que a través de la herramienta de foro de la plataforma Moodle, mantuvieron los estudiantes y docentes de la asignatura Estadística de la carrera Licenciatura en Sistemas de Información de la UNSE, durante una experiencia de ACSC realizada en el año 2010.
- Las comunicaciones vía chat entre docentes mexicanos, quienes participaron de un seminario de formación desarrollado en el marco del proyecto "Trabajo y aprendizaje colaborativo con tecnologías de información y comunicación en ciencias" (Tactics) en el periodo 2004-2005.
- Los mensajes intercambiados vía chat y foro entre estudiantes mexicanos que participaron del "Curso Preliminar de Fundamentos en Ciencias de la Computación" y el equipo docente a cargo del mismo. El curso se dictó en el año 2007 en el marco del proyecto Tactics.
- Las comunicaciones vía chat entre los alumnos de la Maestría en Ciencias computacionales y los docentes a cargo del dictado del curso "Propedéutico de Matemáticas", desarrollado en el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) de México en el año 2008.

De las experiencias de ACSC desarrolladas con docentes y estudiantes de la FCEyT se obtuvo tan solo el 18% de las unidades de análisis empleadas en el presente estudio de análisis de contenido. En particular, con la asignatura estadística la participación tanto de

los estudiantes como de los docentes fue prácticamente nula. Al ser consultados sobre esta situación, los miembros del equipo catedrá manifestaron que si bien saben como trabajar con grupos de estudiantes en las aulas, no les resulta fácil hacerlo en un entorno de ACSC como Moodle. Esta falta de experiencia en la coordinación del aprendizaje grupal mediante herramientas de *e-learning* era lo común entre los docentes de otros asignaturas.

Teniendo en cuenta la falta de experiencia de los docentes de la FCEyT con el ACSC y que la cantidad datos que se tenia era prácticamente insignificante para poder aplicar métodos de aprendizaje de máquina, era evidente la necesidad de adquirir mayor cantidad de datos pero sin realizar nuevas experiencias de ACSC. Ante esta situación se decidió buscar registros de interacciones de experiencias de ACSC desarrolladas por otros investigadores. De este modo se logro contactar a investigadores que participaron del proyecto Tactics y a investigadores del CENIDET, quienes brindaron acceso a las sesiones de chat y comunicaciones vía foro de las experiencias de ACSC en las que estuvieron involucrados.

VII.2 PREGUNTAS DE INVESTIGACIÓN

- ¿Cuáles son las habilidades sociales que manifestaron las personas que se desempeñaron como e-tutores en las experiencias de ACSC consideradas?

VII.3 CONTEXTO

El contexto está formado por la teoría existente sobre las habilidades de los e-tutores tanto en el campo del *e-learning* general como en el campo del ACSC, ya que en la misma se establecen las bases para comprender cuál es la tarea de un e-tutor de ACSC y se indican cuáles son las acciones o comportamientos típicos asociados a cada una de las habilidades de un e-tutor.

En cuanto a la separación entre correlaciones estables y condiciones contribuyentes, se considera que las primeras están establecidas principalmente en la literatura de *e-learning* no colaborativo sobre habilidades de e-tutores y en menor grado en la literatura sobre ACSC. En la bibliografía de *e-learning* no colaborativo las habilidades planteadas se han derivado de la experiencia (Salmon, 2003; Varvel, 2007), se han contrastado con la realidad mostrando como se cumplen en la práctica (Mason, 1991), o su planteamiento se realizó o fue impulsado por organizaciones reconocidas en el área de la enseñanza y capacitación mediante herramientas de *e-learning* (Goodyear *et al*., 2001; Kemshal Bell, 2001; TCOL, 2002; Ulster, 2002; eTTNet TWG 2, 2005; Duggleby *et al*., 2002; IITT,

2007; ION, 2005). Si bien la coincidencia no es total, en general existe consenso sobre cuál es la tarea de un e-tutor de entornos donde estudiantes aprenden de manera individual, es decir que no se recurre al aprendizaje colaborativo como estrategia pedagógica. En tanto que, en el área del ACSC a pesar de que el panorama no es del todo claro se pueden identificar dos enfoques. Por un lado, investigadores tales como (Bonk *et al*., 2004; Lund, 2004; Palloff y Pratt, 2007; Teles *et al*., 2001) toman como base el esquema de organización por roles heredado de la literatura sobre habilidades y competencias para docentes de *e-learning* no colaborativo. Por el otro, (Fåhræus, 2000; Palloff & Pratt, 2001; Lim y Cheah, 2003; Orvis y Lassiter, 2006; Olivares, 2007; Juárez *et al.*, 2008) realizan sus propios planteamientos sobre la manera en que consideran se deberían agrupar las actividades, tareas, o cuestiones de incumbencia para un e-tutor de ACSC. A pesar de la existencia de estos intentos de sistematizar las habilidades o tareas propias de un e-tutor de ACSC, dado que los mismos se formularon siguiendo enfoques completamente diferentes es difícil encontrar puntos en común.

Tomando como base el último punto, en la taxonomía de habilidades que se presentó en el capítulo IV se buscaron conciliar las diferencias entre los dos enfoques existentes en el campo del ACSC, como también encontrar los puntos en común con las clasificaciones del e-learning no colaborativo, de manera tal de crear un marco unificado de las habilidades de un e-tutor de ACSC. De esta manera, gran parte de las correlaciones estables estarían plasmadas en el esquema de clasificación obtenido durante esta investigación.

Finalmente, siguiendo a (Denis *et al.*, 2004; Cornelius & Higgison, 2001; Kukulska-Hulme, 2004) se puede afirmar que los rasgos propios de cada materia, las características de los docentes y alumnos, las salidas de aprendizaje buscadas, la naturaleza de la tarea, y la tecnología usada, conforman las condiciones contribuyentes que influyen tanto en las habilidades manifestadas por los e-tutores como en las relaciones que se puedan trazar entre éstas y las acciones de los mismos.

VII.4 CONSTRUCTO ANALÍTICO

En el presente estudio se decidió usar constructos analíticos conocidos como aplicación de estándar (Krippendorff, 2004) porque comparando los mensajes o contribuciones de los e-tutores contra el esquema de clasificación de habilidades presentado en el capítulo IV se determinan las habilidades sociales que se manifiestan. Los constructos se especificaron mediante sentencias del tipo “si (condición) entonces (atributo

de habilidad)", donde la "(condición)" indica la o las acciones de los e-tutores vinculadas a cada atributo de habilidad especificado en "(atributo de habilidad)".

Para facilitar la comprensión de los constructos analíticos se decidió agruparlos de acuerdo con las sub habilidades en las que se dividieron las habilidades sociales en la clasificación planteada en el capítulo IV. De este modo, fue posible establecer los constructos que se enuncian en las próximas sub secciones.

VII.4.1 Estimular la Interacción entre Estudiantes

- Si (se usa técnicas como trabajo en pares, discusiones informales, debates, etc., de modo que los estudiantes se pongan en contacto) entonces (se manifiesta la el atributo de habilidad de usar técnicas para estimular la interacción).
- Si (se busca mejorar las relaciones entre los estudiantes) entonces (se manifiesta el atributo de habilidad de mejorar las relaciones).
- Si (se modera el trabajo grupal pero minimizando las intervenciones de modo que los estudiantes gestionen sus propias discusiones) entonces (se manifiesta el atributo de habilidad de promover la interacción sin dominar la discusión).
- Si (se discute con los estudiantes las formas de facilitar la creación de interacciones) entonces (se manifiesta el atributo de habilidad de discutir formas de interacción).
- Si (se impulsa a los estudiantes a que se respondan entre sí por ejemplo realizando a un estudiante una pregunta planteada por su compañero) entonces (se manifiesta el atributo de habilidad de estimular respuestas mutuas).

VII.4.2 Estimular la Participación

- Si (se solicita a los estudiantes un comentario u opinión sobre un tópico o pregunta) entonces (se manifiesta el atributo de habilidad de solicitar comentarios sobre tópicos tratados).
- Si (se impulsa a los estudiantes a explicar o dar más detalles de sus ideas o las ideas presentadas por sus compañeros como ser mediante preguntas del tipo ¿Puedes decirme más?, ¿Puedes explicar eso?, y ¿Qué más?) entonces (se manifiesta el atributo de habilidad de fomentar explicaciones).
- Si (se busca asegurar que todos los miembros del grupo contribuyan de manera equilibrada) entonces (se manifiesta el atributo de habilidad de asegurar contribuciones igualitarias).

- Si (se modifica el tema o tópico de discusión cuando hay una carencia de contribuciones de los participantes) entonces (se manifiesta el atributo de habilidad de cambiar el tema de discusión).
- Si (se plantean tópicos o tareas que lleven a los estudiantes a que se involucren activamente) entonces (se manifiesta el atributo de habilidad de plantear tópicos o tareas que mejoren el involucramiento).
- Si (se realiza a los estudiantes las mismas preguntas que ellos plantean) entonces (se manifiesta el atributo de habilidad de regresar preguntas).
- Si (ante una reducción en el nivel de participación grupal se interroga a los estudiantes tratando de determinar las causas) entonces (se manifiesta el atributo de habilidad de indagar causas de la participación reducida).

VII.4.3 Desarrollar, Soportar y Mantener una Comunidad de Aprendizaje

VII.4.3.1 Desarrollar la cohesión grupal

- Si (se fomenta la discusión sobre los objetivos de aprendizaje en pos de lograr el consenso sobre los mismos) entonces (se manifiesta el atributo de habilidad de desarrollar la cohesión grupal).
- Si (se impulsa a los estudiantes a compartir detalles de la vida personal como ser viajes, observaciones, emociones, quien se es como persona, etc.) entonces (se manifiesta el atributo de habilidad de desarrollar la cohesión grupal).

VII.4.3.2 Promover la colaboración

- Si (se fomenta la colaboración entre los estudiantes) entonces (se manifiesta el atributo de habilidad de promover la colaboración).
- Si (se toman medidas de modo tal de que los estudiantes no compitan entre sí) entonces (se manifiesta el atributo de habilidad de promover la colaboración).

VII.4.3.3 Promover el respeto

- Si (se remueve del curso a los estudiantes que arremeten contra el curso y sus pares por medio de insultos) entonces (se manifiesta el atributo de habilidad de promover el respeto).
- Si (se elimina los mensajes que contienen insultos y se informa a los estudiantes la razón de tal determinación) entonces (se manifiesta el atributo de habilidad de promover el respeto).
- Si (se edita los mensajes que contienen insultos y se informa a los estudiantes la razón de tal determinación) entonces (se manifiesta el atributo de habilidad de promover el respeto).

- Si (a los estudiantes que recurren a insultos se les indica que sus contribuciones están fuera de las reglas de comportamiento aceptadas) entonces (se manifiesta el atributo de habilidad de promover el respeto).
- Si (se remueve del curso a los estudiantes que hacen bromas ofensivas que atacan a sus pares o el grupo) entonces (se manifiesta el atributo de habilidad de promover el respeto).
- Si (se elimina los mensajes donde los estudiantes hacen bromas ofensivas que atacan a sus pares o el grupo y se informa las razones de tal determinación) entonces (se manifiesta el atributo de habilidad de promover el respeto).
- Si (se edita los mensajes donde los estudiantes hacen bromas ofensivas que atacan a sus pares o el grupo y se informa las razones de tal determinación) entonces (se manifiesta el atributo de habilidad de promover el respeto).
- Si (a los estudiantes que hacen bromas ofensivas que atacan a sus pares o el grupo se les indica que sus contribuciones están fuera de las reglas de comportamiento aceptadas) entonces (se manifiesta el atributo de habilidad de promover el respeto).

VII.4.4 Crear un Ambiente de Aprendizaje Agradable

- Si (se da la bienvenida a los estudiantes que participan en el curso) entonces (se manifiesta el atributo de habilidad de dar la bienvenida).
- Si (se pide a los estudiantes que se presenten a sí mismos, que planteen sus expectativas para el curso, o se recurre a otras estrategias para romper con el silencio inicial) entonces (se manifiesta el atributo de habilidad de usar rompe hielos).
- Si (se estimula a los estudiantes a comentar las presentaciones o introducciones de su pares) entonces (se manifiesta el atributo de habilidad de fomentar comentarios sobre presentaciones).
- Si (se trata de lograr que los estudiantes se sientan confiados, confortables, y libres de compartir sus pensamientos y sentimientos sin miedo al ridículo) entonces (se manifiesta el atributo de habilidad de protección contra el miedo al ridículo).
- Si (se uso un estilo amigable e informal en las contribuciones) entonces (se manifiesta el atributo de habilidad de usar un estilo amigable).
- Si (se impulsa a los estudiantes a actuar de forma amigable entre sí) entonces (se manifiesta el atributo de habilidad de promover amistad).

VII.4.5 Manejar Comportamientos de Discusión Inapropiados

VII.4.5.1 Mirones

- Si (a los estudiantes que se limitan a observar las actividades de sus pares sin realizar contribuciones se les pide que cambien su comportamiento) entonces (se manifiesta el atributo de habilidad de tratar con mirones).
- Si (se trata de lograr que los estudiantes que se limitan a observar las actividades de sus pares sin realizar contribuciones se sientan parte del grupo) entonces (se manifiesta el atributo de habilidad de tratar con mirones).
- Si (se suspende el acceso al curso a los estudiantes que se limitan a observar las actividades de sus pares sin realizar contribuciones) entonces (se manifiesta el atributo de habilidad de tratar con mirones).

VII.4.5.2 No participantes

- Si (se solicita una participación más activa a los estudiantes que no estuvieron en contacto con sus pares durante largo tiempo) entonces (se manifiesta el atributo de habilidad de tratar con no participantes)
- Si (se trata de lograr que los estudiantes que no estuvieron en contacto con sus pares durante largo tiempo se sientan parte del grupo) entonces (se manifiesta el atributo de habilidad de tratar con no participantes)
- Si (se busca determinar las causas por las que algún estudiante no estuvo en contacto con sus pares durante largo tiempo) entonces (se manifiesta el atributo de habilidad de tratar con no participantes)

VII.4.5.3 Bloqueadores

- Si (se pide un cambio de comportamiento a los estudiantes que citan experiencias personales no relacionadas al problema del grupo) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores).
- Si (se pide un cambio de comportamiento a los estudiantes que hablan de su esposa, el hombre común, de su mascota, y en general de cuestiones no relacionadas al tema que el grupo está tratando) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores).
- Si (se pide un cambio de comportamiento a los estudiantes que discuten demasiado en un punto que el resto del grupo ha resuelto) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores).
- Si (se pide un cambio de comportamiento a los estudiantes que rechazan ideas sin consideración) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores)

- Si (se pide un cambio de comportamiento a los estudiantes que evitan que sus pares lleguen a una votación) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores)
- Si (se suspende el acceso al curso a los estudiantes que citan experiencias personales no relacionadas al problema del grupo) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores)
- Si (se suspende el acceso al curso a los estudiantes que hablan de su esposa, el hombre común, de su mascota, y en general de cuestiones no relacionadas al tema que el grupo está tratando) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores)
- Si (se suspende el acceso al curso a los estudiantes que discuten demasiado en un punto que el resto del grupo ha resuelto) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores)
- Si (se suspende el acceso al curso a los estudiantes que rechazan ideas sin consideración) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores)
- Si (se suspende el acceso al curso a los estudiantes que evitan que sus pares lleguen a una votación) entonces (se manifiesta el atributo de habilidad de tratar con bloqueadores)

VII.4.5.4 Dominadores

- Si (a los estudiantes que siempre son los primeros en contribuir dejando poco para decir a sus pares se les indica que sería bueno escuchar a otros) entonces (se manifiesta el atributo de habilidad de tratar con dominadores).
- Si (a los estudiantes que siempre son los primeros en contribuir dejando poco para decir a sus pares se les pide que esperen respuestas de sus pares antes de contribuir) entonces (se manifiesta el atributo de habilidad de tratar con dominadores).
- Si (se suspende el acceso al curso a los estudiantes que siempre son los primeros en contribuir dejando poco para decir a sus pares) entonces (se manifiesta el atributo de habilidad de tratar con dominadores).
- Si (a los estudiantes que buscan dominar la discusión dando direcciones autoritariamente, rebajando rango/status, interrumpiendo contribuciones de otros, etc., se les pide que cambien su comportamiento) entonces (se manifiesta la sub habilidad de tratar con dominadores).

- Si (se suspende el acceso al curso a los estudiantes que buscan dominar la discusión dando direcciones autoritariamente, rebajando rango/status, interrumpiendo contribuciones de otros, etc.) entonces (se manifiesta el atributo de habilidad de tratar con dominadores).

VII.4.5.5 Buscadores de reconocimiento

- Si (se pide que cambien su comportamiento a los alumnos que intentan llamar la atención ya sea hablando demasiado, con ideas extremas, haciendo alarde o jactándose, o bien siendo bulliciosos) entonces (se manifiesta el atributo de habilidad de tratar con buscadores de reconocimiento).
- Si (se suspende el acceso al curso a los alumnos que intentan llamar la atención ya sea hablando demasiado, con ideas extremas, haciendo alarde o jactándose, o bien siendo bulliciosos) entonces (se manifiesta el atributo de habilidad de tratar con buscadores de reconocimiento).

VII.4.6 Manejar Contribuciones Inapropiadas

VII.4.6.1 Actuar contra las críticas

- Si (se remueve del curso a los estudiantes que tienden a criticar o culpar a sus pares) entonces (se manifiesta el atributo de habilidad de actuar contra las críticas).
- Si (se elimina los mensajes donde los estudiantes critican o culpan a sus pares y se informa las razones de tal determinación) entonces (se manifiesta el atributo de habilidad de actuar contra las críticas).
- Si (se edita los mensajes donde los estudiantes critican o culpan a sus pares y se informa las razones de tal determinación) entonces (se manifiesta el atributo de habilidad de actuar contra las críticas).
- Si (a los estudiantes que critican o culpan a sus pares se les indica que sus contribuciones están fuera de las reglas de comportamiento aceptadas) entonces (se manifiesta el atributo de habilidad de actuar contra las críticas).

VII.4.6.2 Actuar contra la intimidación

- Si (se remueve del curso a los estudiantes que intimidan a sus pares) entonces (se manifiesta el atributo de habilidad de actuar contra la intimidación).
- Si (se elimina los mensajes donde los estudiantes intimidan a sus pares y se informa las razones de tal determinación) entonces (se manifiesta el atributo de habilidad de actuar contra la intimidación).

- Si (se edita los mensajes donde los estudiantes intimidan a sus pares y se informa las razones de tal determinación) entonces (se manifiesta el atributo de habilidad de actuar contra la intimidación).
- Si (a los estudiantes que intimidan a sus pares se les indica que sus contribuciones están fuera de las reglas de comportamiento aceptadas) entonces (se manifiesta el atributo de habilidad de actuar contra la intimidación).

VII.4.6.3 Actuar contra la discriminación

- Si (se remueve del curso a los estudiantes que discriminan a sus pares) entonces (se manifiesta el atributo de habilidad de actuar contra la discriminación).
- Si (se elimina los mensajes donde los estudiantes discriminan a sus pares y se informa las razones de tal determinación) entonces (se manifiesta el atributo de habilidad de actuar contra la discriminación).
- Si (se edita los mensajes donde los estudiantes discriminan a sus pares y se informa las razones de tal determinación) entonces (se manifiesta el atributo de habilidad de actuar contra la discriminación).
- Si (a los estudiantes que discriminan a sus pares se les indica que sus contribuciones están fuera de las reglas de comportamiento aceptadas) entonces (se manifiesta el atributo de habilidad de de actuar contra la discriminación).

VII.4.7 Alabar y Modelar el Comportamiento de Discusión Bueno

- Si (se elogia los comportamientos de discusión buenos de los estudiantes) entonces (se manifiesta el atributo de habilidad de alabar comportamientos de discusión bueno).
- Si (se no se incurre en alguna forma de comportamiento inapropiado) entonces (se manifiesta el atributo de habilidad de modelar comportamientos de discusión bueno).
- Si (no se realizan contribuciones que pueden calificarse como inapropiadas) entonces (se manifiesta el atributo de habilidad de modelar comportamientos de discusión bueno).
- Si (en la comunicación con los estudiantes se usa un estilo cortés) entonces (se manifiesta el atributo de habilidad de modelar comportamientos de discusión buenos).

VII.4.8 Manejar Conflictos

- Si (se facilita el movimiento de los estudiantes a lo largo del conflicto de modo que puedan desarrollar normas para hacer frente a conflictos sucesivos) entonces (se manifiesta el atributo de habilidad de fomentar el desarrollo de normas para manejar conflictos).
- Si (se trata de mediar las diferencias entre los estudiantes en conflicto) entonces (se manifiesta el atributo de habilidad de mediar diferencias).
- Si (se asiste a los miembros del grupo cuando están luchando para lograr consenso) entonces (se manifiesta el atributo de habilidad de lograr consenso).
- Si (se excluye del curso al o los estudiantes conflictos) entonces (se manifiesta el atributo de habilidad de excluir a personas conflictivas).

VII.4.9 Trabajar con Sentimientos y Sensaciones

- Si (se expresa libremente emociones y sensaciones) entonces (se manifiesta el atributo de habilidad de expresar sentimientos y sensaciones).
- Si (se impulsa a los estudiantes a que expresen sus sentimientos y sensaciones) entonces (se manifiesta el atributo de habilidad de estimular la expresión de sentimientos y sensaciones).
- Si (se hace frente a las emociones negativos o conflicto afectivo entre los estudiantes) entonces (se manifiesta el atributo de habilidad de manejar conflicto afectivo).

VII.4.10 Motivar

- Si (se provee palabras de alieno y apoyo) entonces (se manifiesta el atributo de habilidad de proveer palabras de aliento y apoyo).
- Si (se gratifica las contribuciones de los estudiantes reconociéndolas en forma pública o privada, asignando parte de la nota en función de las mismas, etc.) entonces (se manifiesta el atributo de habilidad de valorar el esfuerzo).
- Si (se muestra confianza en que los estudiantes pueden cumplir con las tareas) entonces (se manifiesta el atributo de habilidad de mostrar confianza).
- Si (se establece objetivos desafiantes pero que los estudiantes puedan lograr) entonces (se manifiesta el atributo de habilidad de establecer objetivos desafiantes).

VII.5 ELEMENTOS DEL ANÁLISIS DE CONTENIDO

VII.5.1 Separación en Unidades

VII.5.1.1 Unidad de muestreo

Tanto para el chat como para el foro se considera como unidad de muestreo a las contribuciones o mensajes publicados por los e-tutores.

VII.5.1.2 Unidad de grabación

Las oraciones que forman parte de los mensajes publicados por los e-tutores en el foro y chat son las unidades de grabación. Para identificar las oraciones en las contribuciones de los docentes se tuvo en cuenta las reglas sobre el uso de signos de puntuación especificadas por la Real Academia Española (RAE) en el capítulo 5 del libro *Ortografía de la lengua española* (RAE, 1999). En el Anexo A se pueden consultar las reglas sobre signos de puntuación de la RAE.

La reglamentación de la RAE sobre los signos de puntuación fue complementada con las siguientes indicaciones sobre el uso del signo de interrogación y del signo de exclamación extraídas de (Maqueo, 2004):

Signo de Interrogación

- Si lo que va entre signos de interrogación es una oración completa, tanto la interrogativa como la oración que le sigue se escriben con mayúscula. Por ejemplo: "¿No lo conoces? Yo estaba seguro de que sí".
- Si la interrogativa es la segunda parte de la oración, empieza con minúscula y, a veces, va precedida de una coma. Por ejemplo: "Pero, ¿no lo conocías? ", "Pero ¿no lo conocías?", y "Ya te lo habían preguntado, ¿verdad?".
- Si la oración interrogativa tiene una continuación, ésta se escribe con mayúscula y la oración que le sigue empieza con minúscula. Por ejemplo: "¿No lo conocías?- le preguntaron todos".

Signo de Exclamación

Se usa de igual forma que el signo de interrogación, salvo que con oraciones exclamativas.

VII.5.1.3 Unidad de contexto

- Foro: los mensajes publicados tanto por los e-tutores como por los estudiantes, en relación a un mismo asunto o tema, que preceden al mensaje del cual forma parte la oración bajo análisis se consideran como la unidad del contexto de la misma.

- Chat: se considera que la unidad de contexto está constituida por todos los mensajes que preceden a la oración bajo consideración y que se generaron durante cada una de las sesiones de chat o lapso de tiempo de la sesión de chat dedicada a cada grupo de estudiantes, ya sea que los mensajes fueron producidos por los e-tutores y los alumnos o automáticamente por herramienta de chat. Los mensajes de la herramienta de chat usualmente son notificaciones de que una persona ingreso al chat o lo abandono.

VII.5.2 Muestreo

VII.5.2.1 Plan de muestreo

Se decidió aplicar un plan de muestreo de relevancia dado que no todas las unidades de muestreo, en este caso los mensajes de los e-tutores, son igualmente relevantes a la hora de tratar de responder a la pregunta de investigación del presente estudio. Lo anterior significa que se incluyeron solamente los mensajes que los e-tutores enviaron a un miembro del grupo, al grupo completo o a la clase, y que el analista (responsable del trabajo final) considera que se vinculan con la manifestación de alguno de los atributos de habilidad social. Es decir que, se tuvieron en cuenta las unidades de muestreo relacionados a cualquiera de las siguientes cuestiones:

- Estimulación de la interacción entre estudiantes.
- Estimulación de la participación.
- Desarrollo, soporte y mantenimiento de la comunidad de aprendizaje.
- Creación de un ambiente de aprendizaje agradable.
- Manejo de comportamientos de discusión inapropiados.
- Manejo de contribuciones inapropiadas.
- Alabar y modelar comportamientos de discusión buenos.
- Manejo de conflictos.
- Expresión de sentimientos y sensaciones.
- Motivación.

Asimismo se dejaron de lado los mensajes publicados por las personas que se desempeñaron en carácter de soporte técnico, observador o evaluador del curso ya que el centro de interés lo constituyen los docentes encargados de facilitar el proceso de aprendizaje grupal.

VII.5.2.2 Muestreo resultante

En la Tabla A.1 del Anexo A se pueden encontrar los mensajes de los e-tutores que se seleccionaron al aplicar el plan de muestreo de relevancia.

VII.5.3 Grabación/codificación

Se llevó a cabo un proceso de codificación donde la tarea de las personas encargadas de analizar los mensajes de los e-tutores (un psicopedagogo y un docente con experiencia en ACSC) consistió en catalogarlos siguiendo reglas previamente establecidas y en cuya formulación no tuvieron participación alguna. A continuación se incluye información sobre las capacidades de los codificadores, el entrenamiento que recibieron los codificadores, la sintaxis del lenguaje de datos, la semántica del lenguaje de datos y el manejo de los registros de manera que la investigación de análisis de contenido realizada puede ser replicada.

VII.5.3.1 Capacidades de los codificadores

VII.5.3.1.1 Habilidades cognitivas

- Habilidad de leer.
- Habilidad de seguir instrucciones.
- Habilidad de discernir entre varias categorías.
- Habilidad de tomar decisiones aun en situaciones donde se pueda presentar cierta ambigüedad.
- Habilidad de seguir una secu
- encia de mensajes en una conversación.
- Habilidad de interpretar el contenido latente y no solo manifiesto.
- Habilidad de escribir.

VII.5.3.1.2 Formación/ antecedentes

- Familiarizado con la enseñanza en entornos el *e-learning* y preferentemente de ACSC.
- Familiarizado con la terminología propia de los medios de comunicación sincrónicos y asincrónicos que ofrecen las computadoras.
- Haber alcanzado como mínimo un nivel de educación secundaria.
- Hablante nativo del idioma español.

VII.5.3.2 Entrenamiento de los codificadores

Para asegurarse que los codificadores entendían y podían aplicar correctamente las instrucciones de grabación se realizaron las siguientes actividades:

- Explicar a los codificadores en qué consiste la tarea de codificación y porqué es necesario llevarla a cabo.
- Se le suministró a los codificadores una descripción verbal de las habilidades sociales consideradas en este estudio de modo tal que los mismos puedan plantear las dudas y dificultades que pudieran tener.
- Se proporcionó a los codificadores un glosario con los términos que no son de uso generalizado y que se incluyeron en la descripción verbal de las habilidades sociales.
- Se suministró a los codificadores las instrucciones de codificación de modo que se familiaricen con las mismas.
- Se seleccionó de forma aleatoria un hilo de discusión en un foro y dos sesiones de chat para que los codificadores apliquen las instrucciones de grabación. Este paso permitió identificar dificultades que tenían los codificadores sobre el uso de las instrucciones de codificación al igual que algunas dudas sobre las habilidades sociales.

VII.5.3.3 Semántica del lenguaje de datos

Se optó por especificar la semántica del lenguaje de datos mediante una combinación de designaciones verbales y esquema de decisión.

Las designaciones verbales no son otra cosa más que el esquema de clasificación de habilidades de e-tutor de ACSC que se propone en el marco de este trabajo de investigación. La idea es que el esquema sirva como una fuente de consulta en los casos en que las decisiones a las que conduce el esquema de decisión o los pasos que se siguieron para llegar a las mismas puedan causar cierta duda en los codificadores.

Por su parte, se uso un esquema de decisión puesto que el número de habilidades y sub habilidades en que descompuso la categoría de habilidades sociales de un e-tutor es bastante grande como para que un codificador pueda tenerlas a todas en mente a la hora de asignar significado a cada una de las unidades de grabación bajo análisis. Además, se considera que el uso del esquema de decisión agiliza el proceso de codificación porque:

- En cada paso del esquema de decisión se trató de descartar la mayor cantidad de alternativas no aplicables a la unidad de grabación bajo análisis.

- El esquema se confeccionó con la intención de minimizar o en el mejor de los casos eliminar la necesidad de que el codificador deba consultar el esquema de habilidades para asignar significado a la unidad analizada.
- Reduce la confusión que se puede generar al tener que considerar múltiples alternativas simultáneamente.

En el Anexo A se incluye el esquema de decisión que siguieron los codificadores al igual que una explicación sobre cómo usarlo.

VII.5.3.4 Sintaxis del lenguaje de datos

VII.5.3.4.1 Variables

Se trabajo con dos variables nominales de valor simple denominadas *varUnidadGrab* y *varHabilidad*. La primera corresponde a las oraciones que forman parte de los mensajes de los e-tutores que se obtuvieron al realizar el muestreo en tanto que, *varHabilidad* indica el atributo de habilidad social presente en cada una de las oraciones muestreadas.

VII.5.3.4.2 Valores

Los valores de la variable *varHabilidad* corresponden a la abreviatura de los atributos de habilidad social que forman parte de la clasificación de habilidades un e-tutor de ACSC considerada en este estudio. Se decidió usar la abreviatura de modo que los codificadores no se vean obligados a tener que escribir el nombre completo del atributo de habilidad al clasificar las unidades de grabación. En la Tabla 7.1 se muestra para cada atributo de habilidad social la abreviatura que le corresponde.

Atributo de Habilidad	Abreviatura
Atributo de habilidad de usar técnicas para estimular la interacción.	AHUTEI
Atributo de habilidad de mejorar relaciones.	AHMR
Atributo de habilidad de promover la interacción sin dominar la discusión.	AHPISDD
Atributo de habilidad de discutir formas de interacción.	AHDFI
Atributo de habilidad de estimular respuestas mutuas.	AHERM
Atributo de habilidad de solicitar comentarios sobre tópicos tratados.	AHSCTT
Atributo de habilidad de fomentar explicaciones.	AHFE
Atributo de habilidad de asegurar contribuciones igualitarias.	AHACI
Atributo de habilidad de cambiar el tema de discusión.	AHCTD

Tabla 7.1. Atributos de habilidad social y sus abreviaturas

Atributo de Habilidad	Abreviatura
Atributo de habilidad de plantear tópicos o tareas que mejoren el involucramiento.	AHPTMI
Atributo de habilidad de regresar preguntas.	AHRP
Atributo de habilidad de indagar sobre las causas de la participación reducida.	AHICPR
Atributo de habilidad de desarrollar la cohesión grupal.	AHDCG
Atributo de habilidad de promover la colaboración.	AHPC
Atributo de habilidad de promover el respeto.	AHPR
Atributo de habilidad de dar la bienvenida.	AHDB
Atributo de habilidad de usar rompe hielos.	AHURH
Atributo de habilidad de fomentar comentarios sobre presentaciones.	AHFCP
Atributo de habilidad de protección contra el miedo al ridículo.	AHPCMR
Atributo de habilidad de usar un estilo amigable.	AHUEA
Atributo de habilidad de promover la amistad.	AHPA
Atributo de habilidad de tratar con mirones.	AHTM
Atributo de habilidad de tratar con no participantes.	AHTNP
Atributo de habilidad de tratar con bloqueadores.	AHTB
Atributo de habilidad de tratar con dominadores.	AHTD
Atributo de habilidad de tratar con buscadores de reconocimiento.	AHTBR
Atributo de habilidad de actuar contra las críticas.	AHACC
Atributo de habilidad de actuar contra la intimidación.	AHACIN
Atributo de habilidad de actuar contra la discriminación.	AHACD
Atributo de habilidad de alabar comportamientos de discusión buenos.	AHACDB
Atributo de habilidad de modelar comportamientos de discusión buenos.	AHMCDB
Atributo de habilidad de fomentar el desarrollo de normas para manejar conflictos.	AHFDNMC
Atributo de habilidad de mediar diferencias.	AHMD
Atributo de habilidad de lograr consenso.	AHLC
Atributo de habilidad de excluir a personas conflictivas.	AHEPC
Atributo de habilidad de expresar sentimientos y sensaciones.	AHESS
Atributo de habilidad de estimular la expresión de sentimientos y sensaciones.	AHEESS
Atributo de habilidad de manejar conflicto afectivo.	AHMCA
Atributo de habilidad de proveer palabras de aliento y apoyo.	AHPPAA
Atributo de habilidad de valorar el esfuerzo.	AHVE
Atributo de habilidad de mostrar confianza.	AHMC
Atributo de habilidad de establecer objetivos desafiantes.	AHEOD
Otro Atributo de habilidad.	OAH

Tabla 7.1. (cont)

El atributo OAH se asigna a las unidades de grabación cuando los codificadores no identifican alguno de los atributos de habilidad social que forman parte de la clasificación de habilidades propuesta en el capítulo IV.

Los valores de la variable *varUnidadGrab* corresponden al contenido textual de las oraciones que forman parte de la muestra.

VII.5.3.4.3 Constantes

La relación existente entre las dos variables con las que se trabaja es de asociación, en el sentido de que uno o más valores de la variable *varUnidadGrab* se vinculan con un determinado valor de la variable *varHabilidad*. Esta relación podría representarse por medio de una flecha de la siguiente manera: X ->Y, donde X representa a las oraciones en tanto que Y corresponde a los atributos de habilidad social.

VII.5.3.4.4 Gramática

Una gramática indica que combinaciones de valores de las variables se aceptan como expresiones bien formadas o legitimas. Para expresar la relación de asociación entre las dos variables mediante el operador "->" se debería considerar que:

- La relación siempre va desde la variable *varUnidadGrab* a la variable *varhabilidad* pero no en sentido inverso.
- La variable *varhabilidad* es de valor simple por lo que no puede asumir más de un valor a la vez.
- La variable *varUnidadGrab* es de valor simple por lo que no puede asumir más de un valor a la vez.
- Los valores de la variable *varUnidadGrab* tienen que estar asociados al menos a un valor de la variable *varhabilidad* porque caso contrario la expresión carece de sentido.
- Un valor de la variable *varhabilidad* puede estar asociado con cero o más valores de *varUnidadGrab.*

En la Tabla 7.2 se muestran ejemplos de expresiones bien formadas y expresiones que no cumplen con las reglas gramaticales. La letra X corresponde a la variable *varUnidadGrab* mientras que la letra Y a la variable *varhabilidad*. Los sub índices se usan para hacer referencia a valores concretos de las dos variables.

Expresiones bien formadas	Expresiones mal formadas
$->y_1$ $x_1 -> y_1$ $x_1,x_2,\ldots,x_n -> y_1$ $x_1 -> y_1, y_2,\ldots y_n$	x1 <- y1 x1-> $x1 -> z_1$

Tabla 7.2. Ejemplos de expresiones bien y mal formadas

Teniendo en cuenta las reglas anteriores, para que los registros obtenidos mediante el proceso de codificación se consideren correctos es preciso que se cumplan las siguientes pautas:

- A cada unidad de grabación (oración analizada) debe corresponderle solo una abreviatura de habilidad, por lo que no se aceptan aquellas a las que se le asocie más de una abreviatura de habilidad.
- No se aceptan unidades de grabación en las que no se especifique la abreviatura de habilidad asociada.
- No se aceptan unidades de grabación a las que se les asigna una abreviatura de habilidad que no pertenezca al conjunto de abreviaturas de habilidad permitidas.

VII.5.3.4.5 Lógica

Al tener un solo tipo de expresión bien formado no se necesitan especificar reglas que indiquen de qué manera se relacionan las combinaciones de valores de diferentes expresiones.

VII.5.3.5 Naturaleza y administración de registros

Los registros creados como parte del proceso de grabación se representaron mediante un arreglo de registros donde cada fila del arreglo o tabla corresponde a una unidad de grabación u oración que forma parte de los mensajes muestreados.

VII.5.3.5.1 Información administrativa

El arreglo de registros contiene la información administrativa indicada a continuación, que se repite en cada hoja que se necesitó para representar al conjunto de registros generados:

a) Nombre del proyecto

Especifica el nombre del proyecto o estudio de análisis de contenido.

b) Versión de instrucciones de grabación

Indica la versión de las instrucciones de grabación que utilizada por los codificadores para crear los registros.

c) Tipo de texto

Indica el propósito para el cual se utilizan las unidades de grabación y puede asumir los siguientes valores:

- Muestra de prueba: cuando las unidades de grabación no se utilizan para responder a las preguntas de investigación del estudio sino como base para entrenar a los codificadores en el uso de las instrucciones de grabación, chequear que las instrucciones de grabación sean fáciles de comprender, etc.
- Muestra: cuando las unidades de grabación constituyen una muestra del cuerpo total de textos analizados y se usan para responder a las preguntas de investigación del estudio.
- Cuerpo principal de textos: cuando las unidades de grabación se usan para responder a las preguntas de investigación pero no se aplicó ninguna técnica de muestreo sino que en su conjunto las unidades de grabación representan el cuerpo de textos completo.

d) Nombre del Codificador

Nombre de la persona que crea los registros a partir de la aplicación de las instrucciones de grabación.

e) Nombre del verificador

Nombre de la persona que se encarga de chequear que los registros creados por los codificadores cumplan con las reglas gramaticales especificadas con anterioridad.

f) Estado

Indica de qué manera se proceso las unidades de grabación que forman parte del arreglo de registros en cuestión. Los estados posibles son los siguientes:

- Completado: cuando las unidades de grabación han sido generadas por el codificador.
- Bajo modificación: las unidades de grabación se consideran bajo modificación cuando la persona encargada de la verificación detecta que no se cumple alguna de las reglas gramaticales mencionadas previamente y las entrega al codificador para que realice todas las modificaciones necesarias.
- Verificado: las unidades de grabación se consideran verificadas una vez que la persona encargada de la verificación comprueba que se cumplieron todas las reglas gramaticales.

g) Código de unidad de grabación

Código alfanumérico que permite identificar y localizar a las unidades de grabación. El código se forma con la letra "U" seguida de tantos dígitos como sean necesarios para identificar a todas las unidades de grabación que resultaron del proceso de separación en unidades. El Anexo A incluye la Tabla A.1 donde para cada unidad de grabación se indica el código de unidad de grabación y la experiencia de ACSC a la que pertenece, esto último por si fuera necesario consultar el contexto en el que aparece cada unidad de grabación.

VII.5.3.5.2 Información de organización

Dado que no se trabaja con unidades de grabación anidadas ni vinculadas circularmente no es necesario que se especifique información sobre la manera en la que las categorías aplicadas a un nivel se relacionan con las que se aplican a los niveles vinculados.

VII.5.3.5.3 Información sobre el fenómeno

Cada registro que forma parte del arreglo de registros contiene información sobre la unidad de grabación analizada y el atributo de habilidad que se considera vinculado a la unidad en cuestión. En la Figura 7.1 se muestra un modelo de la hoja entregada a los codificadores para que registren tanto información administrativa como información sobre el fenómeno de interés. Además en la hoja se indica a los codificadores la manera de proceder para asignar el atributo de habilidad a cada unidad de grabación.

Nombre del proyecto

Versión de instrucciones de grabación:……
Tipo de texto: muestra de prueba/ muestra/ cuerpo principal de textos.
Nombre del Codificador:……
Nombre del verificador:……
Estado: completado/ bajo modificación/verificado

Instrucciones
Para cada una de las filas de la tabla siguiente, que representan oraciones de mensajes que los e-tutores envían a sus alumnos, se pide completar la columna *Atributo de habilidad identificado*. Para determinar el valor de la columna deberá seguir las instrucciones de grabación que le fueron suministradas previamente.

Código de unidad de grabación	Unidad de grabación	Atributo de habilidad identificado

Figura 7.1. Modelo de hoja de trabajo para codificadores

VII.5.4 Reducción del Texto a Representaciones Manejables

En el presente estudio de análisis de contenido se llevó a cabo un doble proceso de codificación. En el primero, los codificadores (psicopedagogo y experto en ACSC) trabajaron de manera separada y sin consultarse, para asignar un atributo de habilidad a cada una de las unidades de grabación. Puesto que no hubo un acuerdo del 100 % en los atributos de habilidad identificados por los dos codificadores, fue necesaria una segunda reunión para conciliar las diferencias. En este segundo proceso de codificación también participó el analista de contenido (responsable del trabajo final).

VII.5.4.1 Primer proceso de codificación

El análisis de las hojas de trabajo completadas por los codificadores dejó en evidencia que puede presentarse diversas situaciones en relación a las unidades de grabación, a saber:

- Unidades de grabación a la que los codificadores le asignaron los mismos atributos de habilidad.
- Unidades de grabación que difieren en los atributos de habilidad asignados por los codificadores.
- Unidades de grabación a las que los codificadores no le asignaron un atributo de habilidad.
- Unidades de grabación donde solo uno de los analista le asignó el atributo de habilidad.
- Unidades de grabación descartadas (tachadas en la hoja de trabajo) por los codificadores por considerarlas que no deberían formar parte del análisis.

Sobre un total de 891 unidades de grabación los codificadores coincidieron en el atributo de habilidad asignado en 816 oportunidades. La cantidad de veces que se identificó cada atributo de habilidad (considerando solamente las coincidencias entre los codificadores) se muestra en la Tabla 7.3. Cabe aclarar que no figuran los atributos de habilidad AHCTD, AHRP, AHPR, AHPCMR, AHPA, AHTBR, AHACC, AHACIN, AHACD, AHEPC, y AHEESS porque no fueron identificados simultáneamente por los dos codificadores.

Atributo de habilidad	Cantidad de Veces Identificado
AHFE	114
AHMCDB	108
OAH	90
AHDB	82
AHPPAA	79
AHPC	71
AHUTEI	37
AHDFI	36
AHVE	23
AHUEA	21
AHSCTT	21
AHURH	17
AHPTMI	16
AHDCG	15
AHACI	13
AHPISDD	11
AHTM	10

Tabla 7.3. Cantidad de veces que los codificadores identificaron el mismo atributo de habilidad social

Atributo de habilidad	Cantidad de Veces Identificado
AHESS	8
AHTNP	6
AHTD	5
AHERM	5
AHFDNMC	4
AHACDB	3
AHEOD	3
AHFCP	3
AHMC	3
AHMCA	3
AHICPR	3
AHLC	2
AHTB	2
AHMD	1
AHMR	1

Tabla 7.3. (Cont.)

Las discrepancias en los atributos de habilidad asignados por los codificadores se dieron en 49 unidades de grabación tal como se detalla en Tabla 7.4. La primera columna de cada fila corresponde al par de atributos de habilidad que se asignó a la misma unidad de grabación, mientras que la segunda columna indica la cantidad de veces que se repitió la situación con el mismo par de atributos de habilidad.

Atributo de habilidad	Cantidad de Discrepancias
Experto en ACSC:AHFE Psicopedagogo: AHPC	1
Experto en ACSC:AHPC Psicopedagogo: AHUTEI	34
Experto en ACSC: AHUTEI Psicopedagogo: AHPC	2
Experto en ACSC: AHMCDB Psicopedagogo: AHUEA	2
Experto en ACSC: AHPC Psicopedagogo: AHSCTT	2
Experto en ACSC: AHLC Psicopedagogo: AHPPAA	1
Experto en ACSC: AHPISDD Psicopedagogo: AHACI	1
Experto en ACSC: AHPC Psicopedagogo: AHURH	1

Tabla 7.4. Pares de atributos de habilidad social en conflicto y cantidad de veces que se repite la situación.

Atributo de habilidad	Cantidad de Discrepancias
Experto en ACSC: AHESS Psicopedagogo: AHPPAA	1
Experto en ACSC: AHESS Psicopedagogo: OAH	1
Experto en ACSC: AHUEA Psicopedagogo: AHDB	1
Experto en ACSC: AHMCDB Psicopedagogo: AHURH	1
Experto en ACSC: AHESS Psicopedagogo: AHEESS	1

Tabla 7.4. (Cont.)

Dejando de lado las 865 unidades de grabación (sobre el total de 891) a las que los codificadores le asignaron el mismo o diferente atributo de habilidad, en las 26 unidades de grabación restantes se tienen: 2 sin atributo de habilidad, 3 donde solo uno de los codificadores identificó el atributo de habilidad, y 21 que fueron tachadas por los analistas en las hojas de trabajo. El contenido de las unidades que los codificadores descartaron por considerar que no tendrían que ser analizadas es el nombre del e-tutor que publicó el mensaje. Esta situación se dio cuando la comunicación se realizó mediante la herramienta de foro, ya que los docentes tendieron en ese caso a colocar su nombre al final del mensaje enviado a los estudiantes, a pesar de que el software de ACSC colocaba de manera automática el nombre del autor como encabezado de cada mensaje.

Los atributos de habilidad asignados por los dos codificadores a cada una de las 891 unidades de grabación durante el primer proceso de codificación se documentan en Tabla A.1 del Anexo A.

VII.5.4.2 Segundo proceso de codificación

Para construir los clasificadores automáticos mediante minería de textos es necesario conocer el atributo de habilidad asociado con cada oración que forma parte de las contribuciones docentes que se usarán en el proceso de aprendizaje. Si se trabajase con el conjunto de habilidades que identificó cada uno de los codificadores por separado se estarían construyendo dos modelos cuyas decisiones de clasificación reflejarían las discrepancias existentes entre los codificadores. Por consiguiente, es necesario lograr consenso entre los codificadores con respecto al atributo de habilidad que le corresponde a las unidades de grabación en las que no hubo coincidencia en el código de habilidad asignado en el primer proceso de codificación.

Por otro parte, para incrementar la cantidad de datos con la que se trabaja, y por ende mejorar la capacidad de generalización de los clasificadores automáticos, los codificadores tienen que volver a analizar y hacer corresponder un atributo de habilidad en los casos en que las unidades de grabación habían sido descartadas, no se les asigno atributo de habilidad, o que solo fueron consideradas por uno de los codificadores.

La Tabla 7.5 se generó luego de haber resuelto las diferencias existentes entre los codificadores y haber incluido en el análisis las unidades de grabación descartadas.

Atributo de Habilidad	Cantidad de Veces Identificado
OAH	114
AHFE	114
AHMCDB	110
AHDB	83
AHPC	82
AHPPAA	79
AHUTEI	65
AHDFI	36
AHSCTT	23
AHVE	23
AHUEA	22
AHURH	18
AHPTMI	16
AHDCG	15
AHACI	13
AHPISDD	12
AHESS	10
AHTM	10
AHTNP	6
AHERM	5
AHTD	5
AHFDNMC	4
AHACDB	3
AHEOD	3
AHFCP	3
AHICPR	3
AHLC	3
AHMCA	3
AHMC	3
AHTB	2
AHEESS	1
AHMD	1
AHMR	1

Tabla 7.5. Cantidad de veces que se manifiestan los atributos de habilidad social

En la tabla precedente no se incluyen los atributos de habilidad AHCTD, AHRP, AHPR, AHPCMR, AHPA, AHTBR, AHACC, AHACIN, AHACD, y AHEPC porque no se reconocieron sentencias donde se manifiesten.

VII.5.5 Inferencias y Narración

En base a los resultados del segundo proceso de codificación, se puede afirmar que los atributos de habilidad social manifestados por los e-tutores que participaron de las experiencias de ACSC analizadas, son:

- Atributo de habilidad de usar técnicas para estimular la interacción (AHUTEI).
- Atributo de habilidad de mejorar relaciones (AHMR).
- Atributo de habilidad de promover la interacción sin dominar la discusión (AHPISDD).
- Atributo de habilidad de discutir formas de interacción (AHDFI).
- Atributo de habilidad de estimular respuestas mutuas (AHERM).
- Atributo de habilidad de solicitar comentarios sobre tópicos tratados (AHSCTT).
- Atributo de habilidad de fomentar explicaciones (AHFE).
- Atributo de habilidad de asegurar contribuciones igualitarias (AHACI).
- Atributo de habilidad de plantear tópicos o tareas que mejoren el involucramiento (AHPTMI).
- Atributo de habilidad de indagar sobre las causas de la participación reducida (AHICPR).
- Atributo de habilidad de desarrollar la cohesión grupal (AHDCG).
- Atributo de habilidad de promover la colaboración (AHPC).
- Atributo de habilidad de dar la bienvenida (AHDB).
- Atributo de habilidad de usar rompe hielos (AHURH).
- Atributo de habilidad de fomentar comentarios sobre presentaciones (AHFCP).
- Atributo de habilidad de usar un estilo amigable (AHUEA).
- Atributo de habilidad de tratar con mirones (AHTM).
- Atributo de habilidad de tratar con no participantes (AHTNP).
- Atributo de habilidad de tratar con bloqueadores (AHTB).
- Atributo de habilidad de tratar con dominadores (AHTD).
- Atributo de habilidad de alabar comportamientos de discusión buenos (AHACDB).

- Atributo de habilidad de modelar comportamientos de discusión buenos (AHMCDB).
- Atributo de habilidad de fomentar el desarrollo de normas para manejar conflictos (AHFDNMC).
- Atributo de habilidad de mediar diferencias (AHMD).
- Atributo de habilidad de lograr consenso (AHLC).
- Atributo de habilidad de expresar sentimientos y sensaciones (AHESS).
- Atributo de habilidad de estimular la expresión de sentimientos y sensaciones (AHEESS).
- Atributo de habilidad de manejar conflicto afectivo (AHMCA).
- Atributo de habilidad de proveer palabras de aliento y apoyo (AHPPAA).
- Atributo de habilidad de valorar el esfuerzo (AHVE).
- Atributo de habilidad de mostrar confianza (AHMC).
- Atributo de habilidad de establecer objetivos desafiantes (AHEOD).

La Figura 7.2 muestra el porcentaje de ocurrencia de los atributos de habilidad agrupados de acuerdo a la sub habilidad social a la que pertenecen. La distribución de las unidades de grabación es bastante equilibrada ya que en 7 de las 11 sub habilidades el porcentaje es superior al 10 %. Al parecer tanto el manejar conflictos como el trabajar con sentimientos y sensaciones no han sido de relevancia en los cursos, a los cuales pertenecen las interacciones analizadas, ya que en forma conjunta apenas reúnen el 2% de las 891 unidades de grabación. La sub habilidad de "Manejar contribuciones inapropiadas" no aparece en la Figura 7.2 ya que los docentes no tuvieron que actuar ante la presencia de críticas, intimidación o discriminación entre los miembros de los grupos de estudiantes. Tal como se mencionó en el punto VII.3, el ambiente (los estudiantes, la tarea, la materia, y la tecnología) tienen influencia en las habilidades manifestadas por los e-tutores y por lo tanto en una experiencia de ACSC no es necesario que se manifiesten todos los atributos de habilidad social que conforman la taxonomía planteada en el capítulo IV.

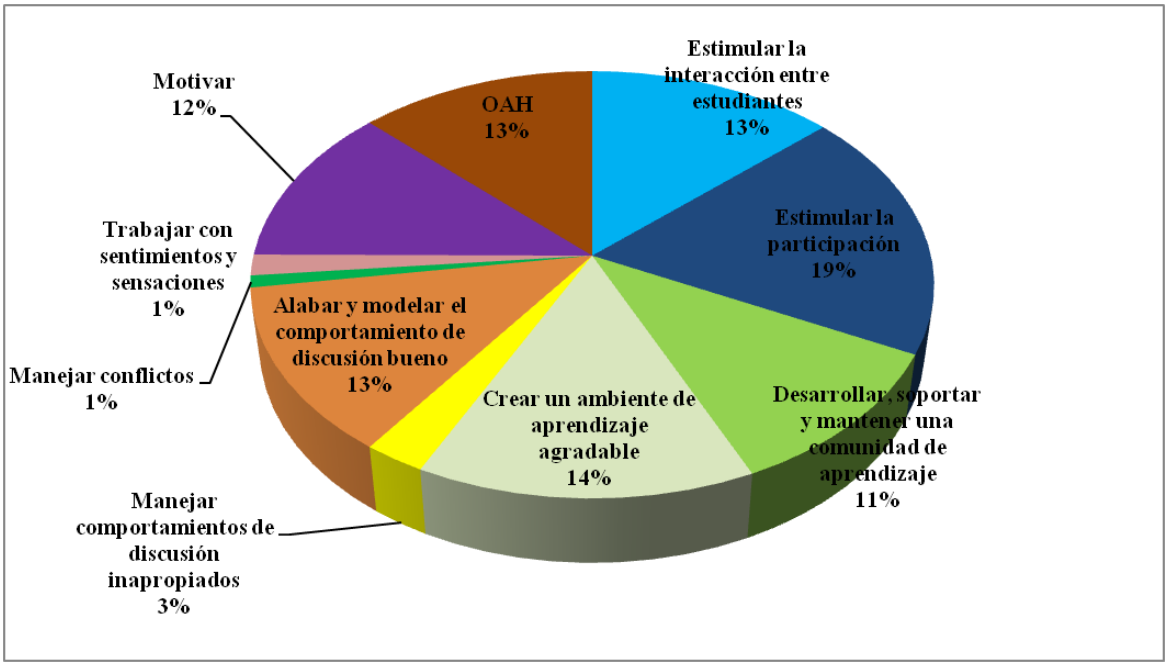


Figura 7.2. Porcentaje de ocurrencia de las sub habilidades sociales

De cierto modo es lógico que los atributos de habilidad, de la sub habilidad "Estimular la Participación", se presenten con mayor frecuencia porque si los estudiantes no participan no interactúan y si no interactúan es imposible que colaboren.

La abreviatura OAH fue introducida en el esquema de decisión de modo que los codificadores la asignen a las unidades de grabación en las que no pueden identificar correspondencia con ninguno de los atributos de habilidad social. El atributo de habilidad OAH fue asignado al 13% de las unidades de grabación, pero esto no significa que el proceso de muestreo selectivo se haya realizado de manera incorrecta, puesto que para incluir un mensaje de los e-tutores en la muestra, bastaba con que al menos una de las oraciones que lo componen se corresponda con un atributo de habilidad social. Es decir que al incluir en la muestra una contribución de los docentes no se puede garantizar que en todas las oraciones constituyentes se vaya a identificar un atributo de habilidad social.

En la Figura 7.3 se puede visualizar la distribución de porcentajes para los atributos de habilidad que componen cada una de las sub habilidades incluidas en la Figura 7.2.

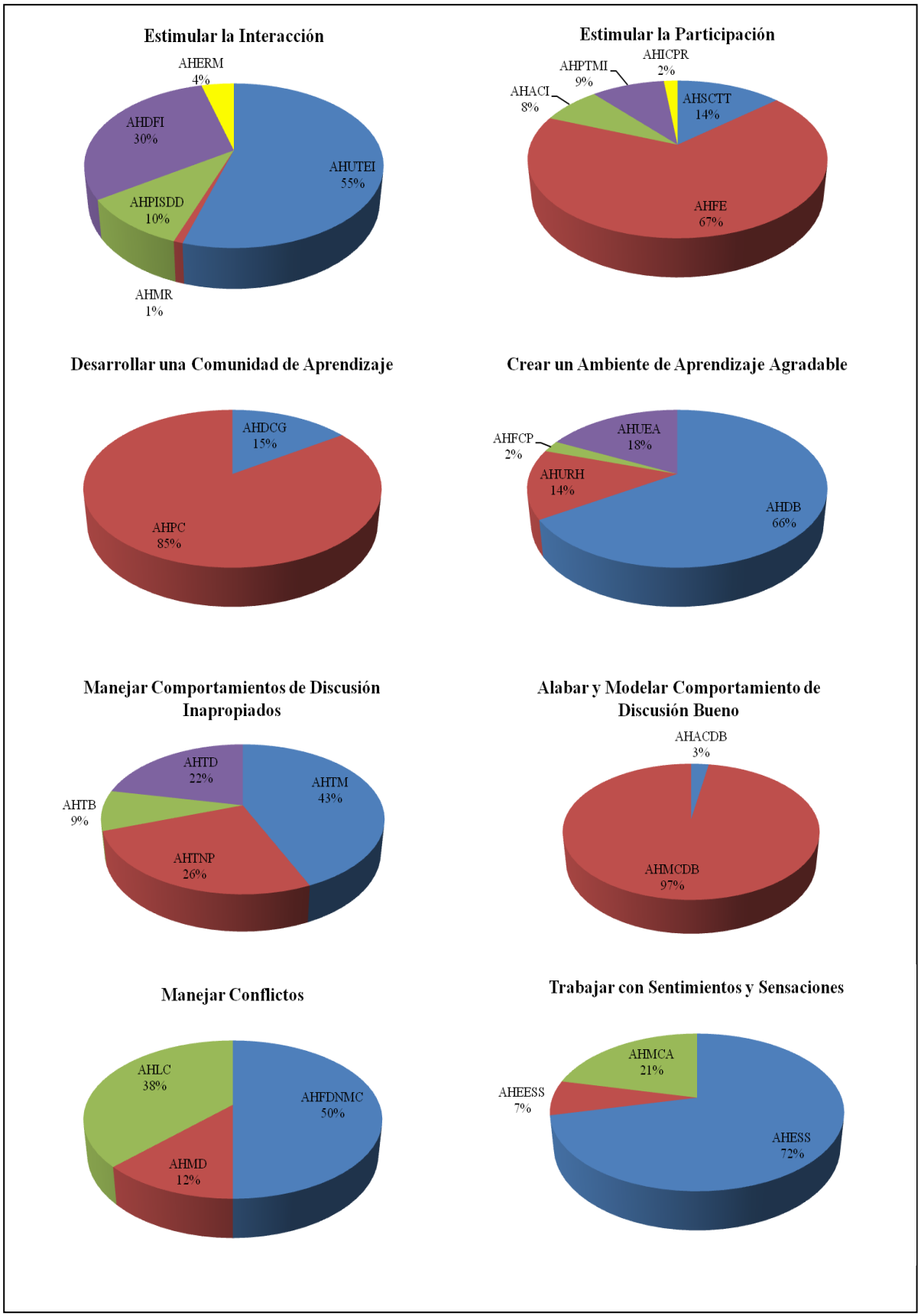


Figura 7.3. Distribución de porcentaje entre los atributos de las sub habilidades sociales

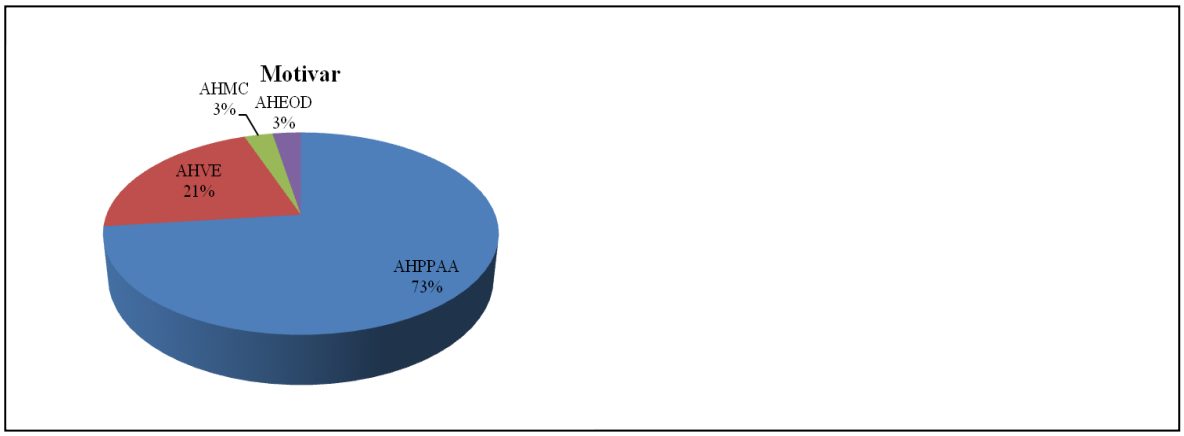


Figura 7.3. (Cont.)

Como se demuestra en la Figura 7.3, para la sub habilidad estimular la interacción, los docentes usaron técnicas para estimular la interacción y definieron normas que facilitan el contacto entre los alumnos pero, sin tratar de jugar un papel dominante en el proceso de interacción. Esto se ve reflejado en el 59% de las contribuciones catalogadas como AHERM y AHUTEI, el 30% asignadas a AHDFI, y el 10 % obtenido para AHPISDD. El escaso 1% del atributo de habilidad AHMR da a entender que la relación entre los miembros de los diversos grupos en general fue buena.

Teniendo en cuenta que los atributos de habilidad de AHFE, AHPTMI, y AHSCTT reúnen el 90% de las 169 unidades de grabación correspondientes a la sub habilidad estimular la participación, es evidente que la estrategia más usada por los e-tutores para lograr que los estudiantes participen consiste en enfocar los aportes de los estudiantes en torno a los temas o tópicos del curso.

Para fomentar el desarrollo de la comunidad de aprendizaje es lógico que el atributo de habilidad más utilizado haya sido AHPC puesto que en los entornos de ACSC se espera que los estudiantes compartan experiencias y conocimientos. El éxito de las intervenciones buscando promover la colaboración lo confirma la ausencia de competencia dentro de los grupos de estudiantes en las cinco experiencias de ACSC consideradas. Esta última afirmación se puede corroborar examinando el registro de interacción de las experiencias de ACSC que se encuentra en la subcarpeta *Experiencias_de_ ACSC* de la carpeta *Analisis_ de_ contenido* del CD que acompaña al presente documento.

Tal como lo muestra el gráfico de torta de la Figura 7.3, el atributo de habilidad de dar bienvenida es fundamental para creación de un ambiente de aprendizaje agradable ya que reúne el 66% de las unidades de grabación que se consideraron vinculadas con la sub habilidad de crear el ambiente de aprendizaje. No obstante, este atributo de habilidad no se

observó en las interacciones de Metodología de la investigación y de Estadística, probablemente esto se deba a que los docentes combinaron clases presenciales con clases *online*. Tanto en Estadística como en Metodología, las actividades grupales en Moodle se desarrollaron a mitad de cuatrimestre y lo más seguro es que la bienvenida por parte de los docentes haya tenido lugar en la primera clase presencial. Lo sucedido con el atributo de habilidad de AHDB permite confirmar lo que se planteó sobre las condiciones contribuyentes, esto es, el contexto en el que se desarrollan las actividades de los docentes tiene influencia en las habilidades manifestadas.

Los atributos de habilidad relacionados con tratar con comportamientos de discusión inapropiados se observaron apenas en el 2.6 % de las unidades de grabación. De los cinco atributos de habilidad pertenecientes a esta sub habilidad el único que no se manifestó fue el de tratar con buscadores de reconocimiento (AHTBR).

En la sub habilidad vinculada con los comportamientos de discusión buenos el 97% de oraciones se corresponden con el atributo de habilidad AHMCDB y el 3% restante con el atributo de habilidad AHACDB. Estos valores dan a entender que los docentes no elogiarion los comportamientos de discusión buenos de los estudiantes pero si trataron de mostrar con su conducta cual es la forma correcta de comportarse en el entorno de ACSC.

De la Figura 7.3 se desprende que a la hora de tratar con los conflictos los e-tutores no excluyeron a los estudiantes conflictivos sino que buscaron mediar las conflictos (12 % de las unidades de grabación con AHMD) de modo de lograr consenso (38 % de las unidades de grabación con AHLC). También fomentaron el desarrollo de normas para que los grupos puedan tratar por su cuenta con situaciones conflictivas que se presenten en el futuro, como lo demuestra 50 % las sentencias asignadas al atributo AHFDNMC.

Al analizar la Figura 7.2 se mencionó que la sub habilidad de trabajar con sentimientos y sensaciones no había sido de relevancia. De acuerdo a las porcentajes consigandos en la Figura 7.3, la mayoría de las intervenciones de los docentes en esta categoría (más exactamente el 72 %) estuvieron encaminadas a expresar sentimientos y sensaciones (AHESS). Si bien se trató con el conflicto afectivo y se estimuló a los estudiantes a expresar sus sentimientos y experiencias, las unidades de grabación catalogadas con estos dos atributos de habilidad representan menos del 30% del total de oraciones vinculadas con tratar con sentimientos y sensaciones.

El hecho de que el 12% de las oraciones que se pueden identificar en las contribuciones docentes se vinculen con el atributo de habilidad de motivar (Figura 7.2), de cierto modo confirma que para lograr la participación de los estudiantes no basta con

que sepan usar los recursos tecnológicos sino que deben estar predispuestos a interactuar con sus pares y con el docente. En base a los porcentajes mostrados en la Figura 7.3 la estrategia más usada para motivar a los alumnos consistió en proveer palabras de aliento ya que el 73% de las 108 unidades de grabación con vinculadas con motivar a los estudiantes se les asignó el atributo AHPPAA. En segundo lugar con un 21% de las manifestaciones se tiene al atributo AHVE vinculado con hacer que los estudiantes sientan que sus contribuciones son valoradas.

VII.6 FIABILIDAD

De acuerdo con Krippendorff (2004) los resultados del análisis de contenido pueden replicarse cuando investigadores diferentes aplican las mismas técnicas sobre los mismos datos y en circunstancias cambiantes. Para chequear esta posibilidad los datos de fiabilidad tienen que generarse como mínimo bajo condiciones test-retest.

En el presente estudio se calculó la reproductibilidad del proceso de codificación usando el procedimiento explicado en la sección V.4.2 porque intervinieron dos codificadores (múltiples observadores), se tenían varias categorías nominales (43 atributos de habilidad) y había unidades de grabación a las que solo un codificar o ambos no le asignaron el atributo de habilidad que le correspondía. Esto significa que los cálculos de reproductibilidad se hicieron con los resultados del primero proceso de codificación. Si bien es lógico pensar que con el segundo proceso de codificación la negociación entre los codificadores puede incrementar la fiabilidad más allá de los datos de fiabilidad generados individualmente por los codificadores, este es un supuesto que no tiene una evidencia mensurable (Krippendorff, 2004). Por consiguiente, los datos de fiabilidad deben calcularse antes del proceso de reconciliación de diferencias.

A continuación se incluye el último paso del cálculo de la reproductibilidad pero no se debería perder de vista que a lo largo del proceso fue necesario construir 891 matrices de coincidencias observadas, una por cada unidad de grabación. Las matrices de coincidencia observada se incluyen en la subcarpeta *Calculo_de_Fiabilidad* de la carpeta *Analisis_ de_ contenido* del CD que acompaña al presente trabajo.

$$\alpha_{nominal}=1-\frac{Do}{De}=1-(n-1)\frac{n-\sum_c Occ}{n^2+\sum_c n_c^2}=1-(1730-1)\frac{1730-1632}{1730^2-248386}=0.938261565$$

El valor 0.938 indica que el proceso de codificación condujo a resultados fiables puesto que según los lineamientos de Krippendorff (2004) se deben aceptar los datos cuyo nivel de fiabilidad sea como mínimo de 0.80.

A partir del valor del parámetro α también se puede inferir que las instrucciones de grabación permiten reducir el nivel de discrepancia entre los codificadores y en consecuencia incrementar el grado de acuerdo. Si bien en un principio se pensó en la posibilidad de que el responsable del trabajo final participe en el proceso de codificación esta opción tuvo que ser descartada ya que también sería el encargado de confeccionar las instrucciones de grabación. Cuando la persona que crea las instrucciones de grabación se encarga de su aplicación lo más probable es que los resultados obtenidos no se deban a la calidad de las instrucciones sino más bien, a la experiencia y conocimiento que tiene esa persona sobre el problema que se pretende resolver con análisis de contenido.

CAPÍTULO VIII
ADAPTACIONES REALIZADAS A LOS RECUROS DE GATE

VIII.1 NECESIDAD DE ADAPTAR GATE

GATE es una herramienta de software que proporciona recursos para realizar múltiples operaciones de pre procesamiento en un proyecto de minería de textos. Si se tiene en cuenta que muchos de estos recursos de procesamiento fueron creados teniendo en mente textos escritos en idioma inglés, es evidente que se necesita adaptarlos para que trabajen correctamente al procesar textos en español. En particular, en este trabajo fue necesario adaptar el recurso de procesamiento *ANNIE Sentence Splitter* para que pueda reconocer oraciones en textos escritos en español. También se tuvo que recurrir al recurso de procesamiento *JAPE Transducer* para implementar la operación de eliminación de *stop words*, puesto que GATE no cuenta con un recurso específico desarrollado para realizar esta tarea.

Antes de entrar de lleno en las adaptaciones realizadas a *ANNIE Sentence Splitter* y en la descripción de la implementación del nuevo recurso que permita la supresión de *stop words*, se presentan algunas nociones preliminares (expresiones regulares, anotaciones, y gramática JAPE) necesarias para poder comprender la descripción de las reglas empleadas tanto en el caso de *ANNIE Sentence Splitter* como en el de *JAPE Transducer*.

VIII.2 EXPRESIONES REGULARES

Las expresiones regulares permiten describir un conjunto de cadenas de caracteres en base a las características en común compartidas por las cadenas del conjunto (Oracle, 2013). En su concepción inicial las expresiones regulares fueron pensadas como un recurso para buscar texto y datos, pero en los lenguajes de programación también se las usa para la editar y manipular tanto texto como datos. La forma más simple de expresión regular es un carácter literal como por ejemplo “a”. Si el texto analizado fuera “María es una buena persona” la expresión regular “a” identificaría cada una de las cinco letras “a”. Cuando se trabaja con expresiones regulares existen operadores para describir clases de caracteres, rangos, repeticiones, etc. Así por ejemplo la expresión regular “[A-Za-z]” sirve para

reconocer cualquier letra en mayúscula o minúscula comprendida entre la letra "A" y la letra "Z". En GATE la definición de expresiones regulares debe hacerse conforme a la sintaxis del *API regex* de la plataforma JAVA.

VIII.3 ANOTACIONES

Una anotación es una forma de meta asociada con una sección particular del contenido de un documento (Cunningham *et al.*, 2011). La conexión entre la anotación y el contenido al que se refiere se logra a través de dos punteros que representan las localizaciones de principio y fin del contenido cubierto. Una anotación se define mediante:

- Un nodo de comienzo: una localización en el contenido del documento definida por un offset, es decir, por un desplazamiento en relación a la primera anotación del documento.
- Un nodo de fin: una localización en el contenido del documento definida por un offset.
- Tipo: es una cadena de caracteres.
- Características: se describen mediante pares <nombre-atributo, valor-atributo>. El nombre del atributo es una cadena de caracteres mientras que los valores pueden ser cualquier objeto Java.
- ID: un valor entero único para cada anotación de un mismo conjunto de anotación.

Las anotaciones se agrupan en conjuntos de anotación. Un conjunto de anotación puede ser visto como un grafo acíclico dirigido, en donde los nodos representan localizaciones particulares en el contenido del documento, y los arcos están hechos de anotaciones que van desde la posición indicada por el nodo de inicio hasta la posición apuntada por el nodo final (Cunningham *et al.*, 2011). La Figura 8.1 ilustra esta metáfora de conjuntos de anotación como grafos.

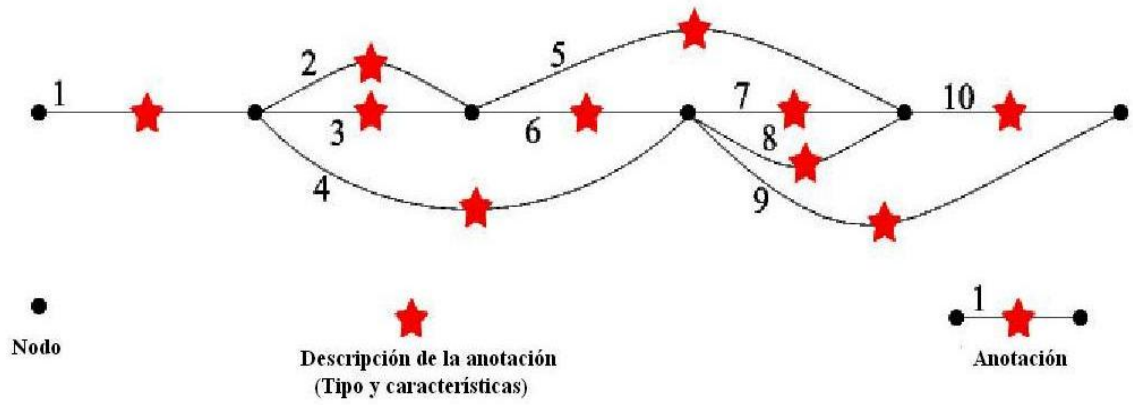


Figura 8.1. Modelo de grafo de anotación (Cunningham *et al.*, 2011)

VIII.4 GRAMÁTICA JAPE

El Motor de Patrones de Anotación en Java (JAPE) permite reconocer expresiones regulares en las anotaciones de los documentos (Cunningham *et al.*, 2011). Una gramática de JAPE consiste de un conjunto de fases compuestas de reglas patrón-acción. Cada una de las reglas tiene una parte izquierda (LHS) y una parte derecha (RHS) separadas mediante "-->". En LHS se describe el patrón a buscar en las anotaciones del documento, en tanto que RHS indica las acciones a realizar si el documento contiene el patrón. En LHS el patrón buscado se encierra entre paréntesis en tanto que cada anotación que forma parte del patrón se encierra entre llaves. El patrón descripto en LHS puede consistir de una anotación, una anotación con un valor de atributo especifico, o una expresión regular que debe satisfacer una anotación. Para clarificar los conceptos mencionados considérese la regla desarrollada para reconocer en textos español abreviaturas formadas por letras minúsculas:

```
Phase: find1
Input: Token SpaceToken
Options: control = appelt

Rule: abreviaturaMinuscula
({Token.orth=="lowercase",Token.length=="1"}{Token.string=="."})+
:abreviaturaMin
--> {}
```

Lo primero a especificar en una gramática es su nombre, en el ejemplo "Phase: find1". Se usa la palabra "Phase" porque las gramáticas de JAPE pueden combinarse formando una secuencia donde cada gramática constituye una fase. También es necesario indicar los tipos de anotaciones usados por la gramática, en este caso anotaciones "Token"

y “SpaceToken” según se indica en “Input: Token SpaceToken”. Las anotaciones de entrada utilizadas por una gramática deben generarse por un recurso de procesamiento ejecutado con anterioridad. En el caso de “find1” las anotaciones “Token” y “SpaceToken” se obtienen utilizando *GATE Unicode Tokeniser*. En cuanto a la opción “Control”, la misma precisa la manera de proceder cuando más de una regla dentro de una fase coincide con la misma región de texto.

La gramática “find1” tiene como única regla a “abreviaturaMinuscula”. Esta regla permite identificar abreviaturas compuestas por una secuencia de una o más letras en minúscula separadas mediante el signo de punto. El proceso de reconocimiento de las abreviaturas tiene éxito cada vez que en el texto analizado se detecta una anotación del tipo *“*Token*”* escrita en minúscula (Token.orth=="lowercase"), de longitud igual a uno (Token.length=="1"), y que tiene a continuación una anotación del tipo *“*Token*”* igual al signo de punto (Token.string=="."). Cuando una secuencia de anotaciones del texto analizado coincide con la regla, a la secuencia completa se la etiqueta con “abreviaturaMin”. Luego en RHS se usa la etiqueta pero al fragmento de texto cubierto por “abreviaturaMin” no se le aplica ningún tipo de operación tal como lo señalan las llaves vacías luego de “--->”.

En el Anexo B se proporcionan más detalles sobre JAPE, el uso de gramáticas y los operadores que pueden intervenir en la construcción de reglas.

VIII.5 ADAPTACIONES EFECTUADAS A ANNIE SENTENCE SPLITTER

Al momento de trabajar con análisis de contenido se descubrió que en cada una de las contribuciones de los docentes se podía identificar más de una habilidad y que por consiguiente convenía tomar a las oraciones como unidad de grabación. Por este motivo, como paso previo al uso de los algoritmos de aprendizaje de la minería de textos se necesita dividir en oraciones a los mensajes de los e-tutores. La descomposición de los mensajes de los e-tutores en oraciones tiene influencia directa en los resultados que se obtienen con los clasificadores de habilidades ya que una identificación errónea de sentencias llevaría a que las habilidades no se reconozcan apropiadamente. Teniendo en cuenta que uno de los propósitos del presente trabajo es simplificar el proceso de reconocimiento de las habilidades que manifiestan los docentes de ACSC, reduciendo el esfuerzo de las personas interesadas en realizar ese reconocimiento, se optó por trabajar en la segmentación automática de los mensajes utilizando GATE.

GATE cuenta con dos recursos de procesamiento para identificar oraciones en un documento (*RegEx Sentence Splitter* y *ANNIE Sentence Splitter*) pero los mismos fueron creados para ser aplicados a textos en inglés. El proceso de adaptación de los dos recursos al español se hizo conforme a las reglas de puntuación especificadas por la Real Academia Española (RAE) que se incluyen en el Anexo A, considerando además las reglas adicionales sobre los signos de exclamación e interrogación mencionadas en el apartado VII.5.1.2.

Ya sea que se utilice *RegEx Sentence Splitter* o bien *ANNIE Sentence Splitter*, a cada oración identificada se le asigna una anotación del tipo "Sentence", en tanto que a los caracteres que permiten distinguir oraciones le corresponde una anotación "Split". Las anotaciones "Split" tienen una característica denominada "kind" con dos valores posibles: "internal" para todos los signos de puntuación que señalan el fin de una oración, y "external" para los caracteres que señalan el fin de una oración pero que no forman parte de la misma.

RegEx Sentence Splitter es un divisor de sentencias basado en expresiones regulares que deben ser especificadas siguiendo la sintaxis del API regex de Java. Si bien se utilizó este recurso en primera instancia, tuvo que ser descartado dado que genera errores en la identificación de oraciones cuando una abreviatura tiene a continuación un signo de puntuación del cual no está separada por caracteres de espacio.

Habiendo descartado *RegEx Sentence Splitter* el trabajo de adaptación se concentró en torno a *ANNIE Sentence Splitter*. Este recurso de procesamiento requiere especificar dos parámetros, a saber:

- gazetterListsURL: la dirección de localización de una lista de las abreviaturas usadas comúnmente que permite distinguir puntos que señalan el fin de oraciones de aquellos que forman parte de las abreviaturas. Las abreviaturas que componen esta lista se incluyen en el Anexo B.
- transducerURL: la dirección de la localización de una gramática JAPE que indica cuando un signo de puntuación marca el fin de una oración.

Para identificar sentencias en textos en español se necesitó de una gramática multifase compuesta de once fases. Cuando se trabaja con JAPE la lista de fases y su orden de ejecución se especifican en un archivo índice que por convención se denomina "main.jape". La estructura del archivo índice utilizado en el presente trabajo de investigación se muestra en la Figura 8.2.

```
MultiPhase:     main
Phases:

prepare
preProcesarTokens
modificarTokens
find1
find2
find3
find4
split
reestablecerTokens
postProcesarTokens
cleanup
```

Figura 8.2. Estructura del archivo main.jape de *ANNIE Sentence Splitter*

En los apartados venideros se describirá cada una de las fases del archivo "main.jape" junto con las reglas que las componen pero sin entrar en detalles sobre el código fuente ya que el mismo puede consultarse en el Anexo B. Cabe aclarar que fue necesario crear 41 reglas para identificar oraciones en textos en español, y se pudieron reutilizar 13 de las reglas incluidas en GATE para el reconocimiento de sentencias en inglés. Las reglas de la gramática incluidas en GATE son usadas por defecto cada vez que se crea una instancia de *ANNIE Sentence Splitter*. A continuación se describe la funcionalidad de todas la reglas utilizadas en este trabajo.

VIII.5.1 Fase Prepare

Consiste de la regla *cleanUp* que se encarga de borrar los datos temporarios creados por ejecuciones previas de *ANNIE Sentence Splitter*. Esta regla forma parte de la gramática que por defecto usa *ANNIE Sentence Splitter* y no fue necesario modificación alguna.

VIII.5.2 Fase preProcesarTokens

En esta fase se identifican oraciones interrogativas y exclamativas que tienen en su interior al menos un signo de punto, de cierre de interrogación, o de cierre de exclamación. Se compone de dos reglas que crean anotaciones temporales de tipo "preguntaModif" y del tipo "exclamacionModif" que son utilizadas por la fase modificarTokens.

VIII.5.2.1 Regla ignorarSignosDentroPreg (Regla nueva)

Cuando entre el signo de apertura y de cierre de interrogación se identifica al menos un signo de punto o un signo de cierre de exclamación esta regla asigna a una anotación del tipo "preguntaModif" a la oración interrogativa.

VIII.5.2.2 Regla ignorarSignosDentroExcl (Regla nueva)

Cuando entre el signo de apertura y de cierre de exclamación se identifica al menos un signo de punto o un signo de cierre de interrogación esta regla asigna a una anotación del tipo "exclamacionModif" a la oración exclamativa.

VIII.5.3 Fase modificarTokens

Se modifican las anotaciones del tipo "Token" incluidas en una oración interrogativa o exclamativa y cuya característica "string" sea igual al signo de cierre de interrogación, signo de cierre de exclamación o signo de punto. Los cambios se efectúan por medio de las dos reglas indicadas a continuación.

VIII.5.3.1 Regla modifTokensInterrogacion (Regla nueva)

En las anotaciones del tipo "preguntaModif" creadas en la fase preProcesarTokens se modifican las características "string" y "kind" de las anotaciones del tipo "Token" cuya característica "string" sea igual al signo de punto o al signo de cierre de exclamación. De este modo se evita que los signos de punto y de cierre de exclamación incluidos en una oración interrogativa sean considerados como divisores de sentencia.

VIII.5.3.2 Regla modifTokensExclamacion (Regla nueva)

En las anotaciones del tipo "exclamacionModif" creadas en la fase preProcesarTokens se modifican las características "string" y "kind" de las anotaciones del tipo "Token" cuya característica "string" sea igual al signo de punto o al signo de cierre de interrogación. De este modo se evita que los signos de punto y de cierre de interrogación incluidos en una oración exclamativa sean considerados como divisores de sentencia.

VIII.5.4 Fase find1

Es la fase más compleja porque está compuesta de 29 reglas que permiten asignar una anotación del tipo "Split" a los signos de puntuación que señalan el fin de una oración. Las reglas que forman parte de esta fase son las indicadas en los próximos puntos.

VIII.5.4.1 Regla DotNetStop (Regla reusada)

Detecta una secuencia de caracteres que se corresponde con ".net", ".Net", o ".NET" pero no lleva a cabo ninguna acción cuando esto sucede. Esto significa que la tarea de la regla es evitar que el signo de punto se considere como divisor de sentencias cuando aparece en alguna de las tres formas de escribir el nombre de la plataforma de desarrollo de Microsoft.

VIII.5.4.2 Regla DotFileName (Regla reusada)

Identifica un signo de punto que forma parte de un nombre de extensión de archivo tal como ".EXE" o ".exe". No se le asigna una anotación del tipo "Split" al signo de punto puesto que en este tipo de situaciones no se señala el fin de una oración.

VIII.5.4.3 Regla pregExclamPtoSuspensivoNoCierraEnunciado (Regla nueva)

Permite ignorar como divisores de sentencia a los signos de cierre de interrogación, cierre de exclamación, y puntos suspensivos, cuando en el texto analizado se identifica una secuencia de caracteres compuesta por el signo de cierre de interrogación o el signo de cierre de exclamación y puntos suspensivos que no cierran un enunciado. Los puntos suspensivos no cierran un enunciado cuando tienen a continuación una palabra en minúscula, una coma, un punto y coma, o el signo de dos puntos. Para clarificar un poco las ideas anteriores supóngase que se tiene la oración "Uno de los aspectos más críticos de la misión fue el aterrizaje a tal punto que fue necesario ¡¡construir el paracaídas más grande de todos los tiempos!! ... de manera de asegurar que no se dañarán piezas de este laboratorio de 900 kilogramos". En este caso los dos signos de cierre de exclamación no marcan el fin de la oración porque tienen a continuación los puntos suspensivos, y esos puntos suspensivos tampoco señalan el fin de la oración porque están sucedidos de una palabra en minúscula.

5.4.4 Regla pregExclamPtoSuspensivoCierraEnunciado (Regla nueva)

Detecta una secuencia de caracteres formada por el singo de cierre de interrogación o cierre de exclamación, puntos suspensivos y una palabra que comienza con mayúscula. Se le asigna una anotación del tipo "Split" a los puntos suspensivos porque en este contexto son los puntos suspensivos y no el signo de cierre de exclamación o de cierre de interrogación lo que señala el fin de una oración.

VIII.5.4.5 Regla ignorarPuntoSuspensivos (Regla nueva)

Permite ignorar como divisores de sentencia a los puntos suspensivos cuando tienen a continuación una palabra en minúscula, una coma, un punto y coma, o el signo de dos puntos.

VIII.5.4.6 Regla Number (Regla reusada)

Reconoce direcciones IP, números donde la parte entera y la parte decimal se separan con punto, y números donde las unidades de mil y de millón se separan mediante el signo de punto. En este contexto el signo de punto no marca el fin de una oración y por lo tanto no se le asigna una anotación del tipo "Split".

VIII.5.4.7 Regla knownAbbrev (Regla reusada)

Mediante esta regla se ignora como delimitador de sentencia al signo de punto cuando forma parte de alguna de las abreviaturas de uso convencional incluidas en la lista señalada por el parámetro gazetterListsURL de *ANNIE Sentence Splitter*. No fue necesario modificar esta regla en la gramática por defecto que usa *ANNIE Sentence Splitter* pero si se tuvo que crear la lista de abreviaturas en español. Esta lista se encuentra en el Anexo B e incluye todas las abreviaturas que para ser reconocidas deben definirse explícitamente ya que no pueden ser descriptas por medio de un patrón.

VIII.5.4.8 Regla upperCaseAbbrev (Regla reusada)

Reconoce abreviaturas constituidas por una secuencia de letras en mayúscula separadas mediante el signo de punto, como ser "C.C." por "Casilla de Correo" o "D.F." por "Distrito Federal". Al signo de punto no se le hace corresponder una anotación del tipo "Split" puesto que funciona como delimitador de oraciones.

VIII.5.4.9 Regla abreviaturaMinuscula (Regla nueva)

Permite identificar abreviaturas constituidas por una secuencia de letras en minúscula separadas mediante el signo de punto como ser "a.m." por "ante meridiem". En este caso el signo de punto tiene que ser ignorado como delimitador de oraciones.

VIII.5.4.10 Regla abreviaturaMinMay (Regla nueva)

Permite reconocer abreviaturas formadas por una letra en minúscula seguida de punto y una secuencia de una o más letras en mayúscula separadas por medio del punto, tal es el caso de "a.J.C." por "antes de Jesucristo". Ante esta situación lo que corresponde es no considerar al signo de punto como señalador del fin de una oración y es por ello que no se le asigna una anotación del tipo "Split".

5.4.11 Regla abreviaturaPlural (Regla nueva)

Mediante esta regla se identifica el plural de las abreviaturas constituidas por una sola letra, tal es el caso de "pp." por "páginas". El plural de una abreviatura formada por una sola letra se obtiene duplicando la letra. El signo de punto del plural de una abreviatura no marca el fin de una oración.

VIII.5.4.12 Regla abreviaturaAcentuada (Regla nueva)

Mediante esta regla se logran reconocer las abreviaturas que tienen tilde o acento en alguna de las letras que la componen, como ser "admón." por "administración" y "cód." por "código". Las abreviaturas acentuadas no pueden ser incluidas en el archivo .lst señalado por el parámetro *gazetterListsURL* de *ANNIE Sentence Splitter* porque este tipo de archivos no admite caracteres acentuados. Además para reconocer un carácter acentuado en una gramática JAPE no se lo debe escribir literalmente sino mediante su código Unicode. Teniendo en cuenta que el código Unicode para "ó" es "\u00F3", a la abreviatura "cód." se la debe escribir como "c\u00F3d.". Como es de esperarse la regla abreviaturaAcentuada no asigna una anotación del tipo "Split" al signo de punto que forma parte de la abreviatura con acento.

VIII.5.4.13 Regla abreviaturaCaracterEspec (Regla nueva)

La regla fue definida para tratar con las abreviaturas que si bien no contienen caracteres acentuados están formadas con un carácter que para ser reconocido debe ser reemplazado por su código Unicode. En esta versión de la gramática JAPE de *ANNIE Sentence Splitter* se considera solamente la letra "ñ" por lo cual mediante abreviaturaCaracterEspec se puede identificar las cadenas de caracteres "Dña.", "Dñas.", "dña." y "dñas." que son abreviaturas de las palabras "Doña", "Doñas", "doña" y "doñas". Cuando esta regla entra en acción no le asigna una anotación del tipo "Split" al signo de punto.

VIII.5.4.14 Regla abreviaturaEtcetera (Regla nueva)

Permite identificar las situaciones donde el signo de punto de la abreviatura "etc." no funciona como separador de oraciones. Por consiguiente, con esta regla al signo de punto de la abreviatura "etc." no se le asigna una anotación del tipo "Split".

VIII.5.4.15 Regla direccionEmail (Regla nueva)

Detecta las direcciones de correo electrónico que forman parte del texto analizado y de esta manera evita que los signos de punto de los correos electrónicos identificados se consideren como divisores de sentencia.

VIII.5.4.16 Regla direccionURI (Regla nueva)

Las direcciones URI son una cadena de caracteres que identifica inequívocamente un recurso (página web, documento, servicio, enciclopedia, etc.) por lo tanto el o los signos de punto que forman parte de una dirección URI no deben ser considerados como delimitadores del fin de una oración. Aunque se acostumbra llamar URLs a todas las direcciones Web, URI es un identificador más completo que se diferencia de un URL en que permite incluir en la dirección una subdirección. Mediante la regla direccionURI se logran reconocer las direcciones URI en el texto analizado y se evita que al o a los signos de punto que la componen se les asigne una anotación del tipo "Split".

VIII.5.4.17 Regla identificadorJava (Regla nueva)

Mediante esta regla se detecta el o los signos de punto que forman parte de un nombre de identificador válido en Java. Cuando en el texto de entrada hay una cadena de caracteres que coincide con el patrón de esta regla no se lleva a cabo ningún tipo de acción y se evita que el signo de punto del identificador se tome como delimitador de sentencia.

VIII.5.4.18 Regla Ldots (Regla reusada)

Con esta regla se impide que cuatro o más signos de punto consecutivos se tomen como separados de oraciones, porque cuando en el texto analizado se identifica una cadena de caracteres que coincide con el patrón especificado por esta regla no se ejecuta ninguna acción.

VIII.5.4.19 Regla Lpunct (Regla reusada)

Mediante esta regla una secuencia de seis o más signos de cierre de interrogación o de cierre de exclamación (posiblemente alternados) es ignorada como señalador del fin de una oración.

VIII.5.4.20 Regla Split1 (Regla reusada)

Permite reconocer una secuencia compuesta por uno a cinco signos de cierre de interrogación, uno a cinco signos de cierre de exclamación, uno a cinco signos de cierre de interrogación y signos de cierre de exclamación alternados, y uno a tres signos de punto. En este tipo de situaciones el o los signos de cierre de interrogación, signos de cierre de

exclamación y punto, se desempeñan como delimitadores del fin de oraciones por lo cual se les asigna una anotación del tipo "Split". A pesar de que la RAE establece que los puntos suspensivos son tres, en entornos virtuales es bastante común que se escriban dos en lugar de tres. Por este motivo se decidió tratar como suspensivos a la cadena de caracteres "…" y a la cadena de caracteres "..".

VIII.5.4.21 Regla CR (Regla reusada)

Regla incluida en la gramática que por defecto usa *ANNIE Sentence Splitter* para identificar dos caracteres de nueva línea consecutivos o dos caracteres de nueva línea separados por caracteres de espacio. A cualquiera de estas dos cadenas de caracteres se les asigna una anotación del tipo "Split" con el valor de la característica "kind" igual a "external".

VIII.5.4.22 Regla puntosEnEnumeracion (Regla nueva)

Mediante esta regla se evita que los signos de punto en los ítems de una lista numerada se tomen como delimitadores de sentencia. Es necesaria una regla separada para tratar las listas numeradas porque pueden existir ítems del tipo "1.1.", "1.a.", o "a.1." que no podrían ser reconocidos por la regla Number.

VIII.5.4.23 Regla preguntasexclamacionesConsec (Regla nueva)

De acuerdo con las reglas de la RAE una secuencia de oraciones interrogativas o exclamativas separadas mediante el signo de coma, o punto y coma, tiene que considerarse como una sola oración y solamente el signo de cierre de la última interrogación o exclamación funciona como delimitador de oración. Con la regla preguntasexclamacionesConsec se evita que se asigne una anotación del tipo "Split" al signo de cierre de interrogación o al signo de cierre de exclamación cuando tiene a continuación una coma, o un punto y coma.

VIII.5.4.24 Regla preguntaexclamacionConCont (Regla nueva)

Esta regla identifica cuando en el texto analizado se tiene el signo de cierre de interrogación o el signo de cierre de exclamación sucedido por una letra o palabra en minúscula. En este tipo de casos se dice que la oración interrogativa o la oración exclamativa tiene una continuación y por lo tanto al signo de cierre no se lo toma como delimitador de fin de oración.

VIII.5.4.25 Regla dosPuntosAntesListado (Regla nueva)

Cuando el signo de dos puntos precede a los ítems de un listado se lo toma como señal del fin de una oración a pesar de no existir una regla de la RAE al respecto. Se decidió considerar al signo de dos puntos como delimitador de oraciones puesto que cada uno de los ítems del listado puede ser bastante extenso e incluir, por ejemplo, explicaciones o preguntas que estimulen la participación o reflexión. Si cada ítem del listado no se tratase como un párrafo separado sería imposible reconocer las situaciones donde se manifiesten múltiples intenciones en cada ítem. Cuando la regla dosPuntosAntesListado reconoce que el signo de dos puntos precede a los ítems de un listado le asigna una anotación del tipo "Split".

VIII.5.4.26 Regla parentesisAclaratorio (Regla nueva)

Mediante esta regla se impide que los signos de puntuación incluidos entre paréntesis se lleguen a tomar como delimitadores de oraciones, puesto que el fragmento de texto entre paréntesis tiene una puntuación independiente del texto donde está incluido. Con esta regla se puede reconocer hasta dos niveles de anidamiento de paréntesis, esto es paréntesis que en su interior contienen un fragmento entre paréntesis.

VIII.5.4.27 Regla corcheteAclaratorio (Regla nueva)

Mediante esta regla se impide que los signos de puntuación incluidos entre corchetes se lleguen a tomar como delimitadores de oraciones, puesto que el fragmento de texto recogido entre corchetes tiene una puntuación independiente del texto donde está incluido. Con esta regla se pueden reconocer corchetes que en su interior contienen un fragmento entre corchetes.

VIII.5.4.28 Regla signoLlaves (Regla nueva)

Esta regla evita que los signos de puntuación incluidos entre llaves se consideren como separadores de oraciones puesto que el fragmento de texto recogido entre llaves tiene una puntuación independiente del texto donde está incluido. Con esta regla se pueden reconocer llaves que en su interior contienen un fragmento entre llaves.

VIII.5.4.29 Regla signoComillas (Regla nueva)

La regla permite ignorar a los signos de puntuación incluidos entre comillas ya que el texto entre comillas tiene puntuación independiente y lleva sus propios signos ortográficos. Usualmente se distinguen tres tipos de comillas: la comillas angulares, latinas o españolas (« »); las comillas inglesas (“ ”); y las comillas simples (‘ ’). En GATE se

debe contemplar un cuarto tipo de comillas (" ") muy similar a las comillas inglesas pero que tienen un código Unicode diferente.

VIII.5.5 Fase find2

Las reglas de esta fase actúan de manera complementaría a las reglas de la fase find1 para resolver las situaciones donde se asignó erróneamente una anotación del tipo "Split", y situaciones donde correspondía asignar una anotación del tipo "Split" pero no se hizo. Las falencias en la asignación de las anotaciones "Split" se deben a que algunas de las reglas de la fase find1 tienen en cuenta el contexto donde ocurren las anotaciones y por lo tanto causan que ese contexto no pueda ser utilizado por las otras reglas de la fase. La fase find2 está constituida por las reglas detalladas a continuación.

VIII.5.5.1 Regla puntosSuspensivosSucedenaEtc (Regla Nueva)

Cuando la regla abreviaturaEtcetera de la fase find1 identifica la cadena de caracteres "etc.", examina y como consecuencia consume la anotación del tipo "Token" que tiene inmediatamente a su derecha. De esta manera la anotación "Token" que sucede a "etc." se vuelve no disponible para el resto de las reglas de la fase find1. Si la cadena de caracteres "etc." tuviera a su derecha cuatro signos de punto consecutivos y una palabra en mayúscula, a los últimos tres puntos se les asignaría erróneamente una anotación del tipo "Split" porque de acuerdo a la regla Ldots una secuencia de 4 o más signos de punto no señalan el fin de una oración. Para ilustrar la situación antes mencionada considérese el ejemplo de la Figura 8.3.

Texto analizado
"Las palabras como oquedad, orfandad, orfanato, osamenta, osario, óseo, oval, óvalo, ovarios, oscense, oler, etc….. Se escriben sin h porque no la tienen en su origen."

Oraciones identificadas en find1
"Las palabras como oquedad, orfandad, orfanato, osamenta, osario, óseo, oval, óvalo, ovarios, oscense, oler, etc….."

"Se escriben sin h porque no la tienen en su origen."

Oraciones identificadas con la incorporación de puntosSuspensivosSucedenaEtc
"Las palabras como oquedad, orfandad, orfanato, osamenta, osario, óseo, oval, óvalo, ovarios, oscense, oler, etc….. Se escriben sin h porque no la tienen en su origen."

Figura 8.3. Ejemplo de aplicación de regla puntosSuspensivosSucedenaEtc

Con la regla puntosSuspensivosSucedenaEtc se busca una cadena de caracteres igual a la abreviatura "etc.", cuatro signos de puntos consecutivos, y una palabra en mayúscula. Cuando el texto analizado contiene la cadena de caracteres mencionada, la regla elimina la anotación del tipo "Split" que, en la fase find1, se le asigna a los últimos tres de los cuatro puntos consecutivos.

VIII.5.5.2 Regla reconocerDivisorconetc (Regla Nueva)

Cuando la regla abreviaturaEtcetera de la fase find1 identifica la cadena de caracteres "etc.", examina y como consecuencia consume la anotación del tipo "Token" que tiene inmediatamente a su derecha. De esta manera la anotación "Token" que sucede a "etc." se vuelve no disponible para el resto de las reglas de la fase find1. Si la cadena de caracteres "etc." tuviera a su derecha el signo de cierre de interrogación o el signo de cierre de exclamación y una palabra que comienza con mayúscula, lo que correspondería es asignar una anotación del tipo "Split" al signo de cierre de exclamación o signo de cierre de interrogación. Sin embargo, con las reglas de find1 al signo de cierre de interrogación o al signo de cierre de exclamación se lo omitiría como señalador del fin de una oración puesto que al ser consumido por la regla abreviaturaEtcetera no estaría disponible para el resto de las reglas de la fase. Con la regla reconocerDivisorconetc se reconoce la situación antes comentada y al signo de cierre de interrogación o el signo de cierre de exclamación se le asigna una anotación del tipo "Split". Un ejemplo de los resultados con y sin la regla reconocerDivisorconetc se muestra en la Figura 8.4.

Texto analizado
"¿Doña Cecilia tenía autorización para restaurar la pintura, alguien sabia de su labor, fue un atentado a la obra, etc.? Son interrogantes que surgen ante esta inusual situación."

Oraciones identificadas en find1
"¿Doña Cecilia tenía autorización para restaurar la pintura, alguien sabia de su labor, fue un atentado a la obra, etc.? Son interrogantes que surgen ante esta inusual situación."

Oraciones identificadas con la incorporación de reconocerDivisorconetc
"¿Doña Cecilia tenía autorización para restaurar la pintura, alguien sabia de su labor, fue un atentado a la obra, etc.?"

"Son interrogantes que surgen ante esta inusual situación."

Figura 8.4. Ejemplo de aplicación de regla reconocerDivisorconetc

VIII.5.5.3 Regla preguntaExclamacionSucedenaEtc (Regla Nueva)

Si la cadena de caracteres "etc." tuviera a su derecha una secuencia de seis signos de cierre de exclamación o seis signos de cierre de interrogación, a los últimos cinco signos de la secuencia se les asignaría erróneamente una anotación del tipo "Split" porque de acuerdo a la regla Lpunct una secuencia de más de cinco signos de cierre de interrogación o de cierre de exclamación debe ser ignorada como divisor de sentencias. Esta situación se da porque cuando la regla abreviaturaEtcetera de la fase find1 identifica la cadena de caracteres "etc." examina y como consecuencia consume el primer signo de cierre de exclamación o de cierre de interrogación que tiene a su derecha. Cuando la regla preguntaExclamacionSucedenaEtc encuentra en el texto analizado la cadena de caracteres "etc." seguida de seis signos de cierre de interrogación o de cierre de exclamación lo que hace es suprimir la anotación del tipo "Split" que se asigna en la fase find1 a los últimos cinco signos de cierre de exclamación o de cierre de interrogación. Un ejemplo de la aplicación de la regla preguntaExclamacionSucedenaEtc se incluye en la Figura 8.5.

Texto analizado
"El Ayuntamiento de Borja decidió contratar los servicios una empresa de una seguridad privada y de este modo garantizar la protección del patrimonio cultural. ¿¿¿¿¿¿Cuál será la empresa de seguridad que se contratará, cuántos miembros de la empresa participarán, cuanto saldrá el operativo, etc.?????? Son detalles que hasta el momento nadie ha querido precisar."

Oraciones identificadas en find1
"El Ayuntamiento de Borja decidió contratar los servicios una empresa de una seguridad privada y de este modo garantizar la protección del patrimonio cultural."

"¿¿¿¿¿¿Cuál será la empresa de seguridad que se contratará, cuántos miembros de la empresa participarán, cuanto saldrá el operativo, etc.??????"

"Son detalles que hasta el momento nadie ha querido precisar."

Oraciones identificadas con la incorporación de preguntaExclamacionSucedenaEtc
"El Ayuntamiento de Borja decidió contratar los servicios una empresa de una seguridad privada y de este modo garantizar la protección del patrimonio cultural."

"¿¿¿¿¿¿Cuál será la empresa de seguridad que se contratará, cuántos miembros de la empresa participarán, cuanto saldrá el operativo, etc.?????? Son detalles que hasta el momento nadie ha querido precisar."

Figura 8.5. Ejemplo de aplicación de regla preguntaExclamacionSucedenaEtc

VIII.5.5.4 Regla etcSeguidodedosPuntos (Regla Nueva)

Cuando las reglas de la fase find1 procesan la abreviatura "etc." seguida de dos signos de punto consecutivos que tienen a continuación una palabra que comienza con minúscula, una coma, un punto y coma, o el signo de dos puntos, lo que hacen es asignar una anotación del tipo "Split" al último de los dos puntos. Sin embargo, en esta situación los dos signos de punto deben ser tratados como puntos suspensivos que no cierran un enunciado y por lo tanto no corresponde asignar la anotación del tipo "Split" al segundo signo de punto. Por consiguiente la intervención de la regla etcSeguidodedosPuntos consiste en eliminar la anotación del tipo "Split" que las reglas de la fase find1 asignan al segundo punto de la secuencia. Los resultados que se obtienen con y sin la intervención de la regla etcSeguidodedosPuntos se muestran en la Figura 8.6.

Texto analizado
"Por ello es bastante común verlas viajar miles de km para dar conferencias sin importar el tamaño o prestigio de la institución que las invite. El año pasado recorrieron Alemania, Austria, Bélgica, Italia, Suiza, etc... para dar a conocer su labor y conseguir patrocinadores."

Oraciones identificadas en find1
"Por ello es bastante común verlas viajar miles de km para dar conferencias sin importar el tamaño o prestigio de la institución que las invite. El año pasado recorrieron Alemania, Austria, Bélgica, Italia, Suiza, etc..."

"para dar a conocer su labor y conseguir patrocinadores."

Oraciones identificadas con la incorporación de etcSeguidodedosPuntos
"Por ello es bastante común verlas viajar miles de km para dar conferencias sin importar el tamaño o prestigio de la institución que las invite. El año pasado recorrieron Alemania, Austria, Bélgica, Italia, Suiza, etc... para dar a conocer su labor y conseguir patrocinadores."

Figura 8.6. Ejemplo de aplicación de regla etcSeguidodedosPuntos

VIII.5.5.5 Regla ignorarPtosSuspensivosInterrogacionExclamacion (Regla Nueva)

Mediante esta regla se elimina la anotación del tipo "Split" que las reglas de la fase find1 asignan a los puntos suspensivos cuando preceden al signo de cierre de interrogación o el signo de cierre de exclamación. En estos casos no son los puntos suspensivos los que actúan como divisores de sentencias sino los signos de cierre de interrogación o los signos de cierre de exclamación. Un ejemplo de los resultados que se consiguen con y sin la intervención de la regla ignorarPtosSuspensivosInterrogacionExclamacion se muestra en la Figura 8.7.

Texto analizado
"Las prácticamente infinitas las opciones al trabajar con arreglos…!! El límite lo establece el ingenio, la creatividad, y la imaginación."

Oraciones identificadas en find1
"Las prácticamente infinitas las opciones al trabajar con arreglos…"

!!

"El límite lo establece el ingenio, la creatividad, y la imaginación."

Oraciones identificadas con la incorporación de ignorarPtosSuspensivosInterrogacionExclamacion
"Las prácticamente infinitas las opciones al trabajar con arreglos…!!

"El límite lo establece el ingenio, la creatividad, y la imaginación."

Figura 8.7. Ejemplo de aplicación de regla ignorarPtosSuspensivosInterrogacionExclamacion

Cabe aclarar que la regla ignorarPtosSuspensivosInterrogacionExclamacion intervine cuando una oración interrogativa o exclamativa no tiene el signo de apertura. Las reglas de las fases preProcesarTokens y modificarTokens son las que actúan cuando una oración interrogativa con signo de apertura y de cierre tiene en su interior uno o más signos de punto, o cuando una oración exclamativa con signo de apertura y de cierre tiene en su interior uno o más signos de punto.

VIII.5.5.6 Regla abreviaturaMinusculaMayusculaSeguidaPto (Regla Nueva)

Cuando el texto analizado contiene una letra en mayúscula o minúscula seguida de dos signos de punto con las reglas de la fase find1 se considera que la letra junto con el primer punto es una abreviatura y que el segundo signo de punto marca el fin de la oración. Esto no es correcto porque no debe escribirse un punto luego de las abreviaturas cuando las abreviaturas coinciden con el fin de una oración. Es el punto de la abreviatura el que opera como delimitador de la sentencia. Por lo tanto, con la regla abreviaturaMinusculaMayusculaSeguidaPto se elimina la anotación del tipo "Split" que se le asigna en la fase find1 al segundo de los dos signos punto que suceden a la letra minúscula o mayúscula. La Figura 8.8 describe las situaciones comentadas y el resultado que se obtiene cuando interviene abreviaturaMinusculaMayusculaSeguidaPto.

Texto analizado
"Para definir una variable del tipo arreglo se debe darle un nombre teniendo en cuenta las reglas para identificadores validos de PHP y.. luego asignarle elementos mediante la función array()."

Oraciones identificadas en find1
"Para definir una variable del tipo arreglo se debe darle un nombre teniendo en cuenta las reglas para identificadores validos de PHP y..

"luego asignarle elementos mediante la función array()."

Oraciones identificadas con la incorporación de abreviaturaMinusculaMayusculaSeguidaPto
"Para definir una variable del tipo arreglo se debe darle un nombre teniendo en cuenta las reglas para identificadores validos de PHP y.. luego asignarle elementos mediante la función array()."

Figura 8.8. Ejemplo de aplicación de regla abreviaturaMinusculaMayusculaSeguidaPto

VIII.5.5.7 Regla ptoDeEnumeracionNoDivisor (Regla Nueva)

Esta regla permite detectar una cadena de caracteres con las siguientes características: un carácter de nueva línea, ítems de un listado separados mediante un signo de punto y que termina en punto, y una anotación del tipo "Split" asignada al signo de punto del final de la secuencia. Una vez identificada la cadena de caracteres antes descripta, la regla ptoDeEnumeracionNoDivisor elimina la anotación "Split" asignada en la fase1 al signo de punto. De acuerdo con la regla puntosEnEnumeracion de la fase find1, cuando los ítems de un listado están separados mediante el signo de punto, al último punto no se le debe asignar una anotación del tipo "Split", pero esto no sucede con el primer ítem de una enumeración porque cuando la regla dosPuntosAntesListado de la fase find1 identifica un signo de dos puntos que tiene a continuación el carácter de nueva línea, lo que hace es consumir el carácter de nueva línea y volverlo no disponible para el resto de las reglas de la fase find1. La Figura 8.9 contiene un ejemplo del funcionamiento de *ANNIE Sentence Splitter* con y sin la regla ptoDeEnumeracionNoDivisor.

Texto analizado
"También pueden comunicarse directamente con los miembros del Instituto abajo listados:

1. Lic. Rodrigo Díaz, diaz.rod@daus.usal.es.
2. Dra. Magda Delva, magda-del@daus.usal.es.
3. Ing. Stepanova, stapaneva_marina@daus.usal.es.
4. Dr. Joaquín V. González, joacaogon@daus.usal.es."

Oraciones identificadas en find1
"También pueden comunicarse directamente con los miembros del Instituto abajo listados:"

"1. "

"Lic. Rodrigo Díaz, diaz.rod@daus.usal.es."

"2. Dra. Magda Delva, magda-del@daus.usal.es."

"3. Ing. Stepanova, stapaneva_marina@daus.usal.es."

"4. Dr. Joaquín V. González, joacaogon@daus.usal.es."

Oraciones identificadas con la incorporación de ptoDeEnumeracionNoDivisor
"También pueden comunicarse directamente con los miembros del Instituto abajo listados:"

"1. Lic. Rodrigo Díaz, diaz.rod@daus.usal.es."

"2. Dra. Magda Delva, magda-del@daus.usal.es."

"3. Ing. Stepanova, stapaneva_marina@daus.usal.es."

"4. Dr. Joaquín V. González, joacaogon@daus.usal.es."

Figura 8.9. Ejemplo de aplicación de regla ptoDeEnumeracionNoDivisor

VIII.5.5.8 Regla parentesisEnEnumeracion (Regla Nueva)

La regla da la posibilidad de identificar una sucesión de uno a cinco signos de apertura de interrogación o apertura de exclamación que tiene a su izquierda un paréntesis de cierre, ítems de un listado separados mediante el signo de punto y que puede o no terminar en punto, y el carácter de nueva línea. En tanto que, a su derecha la sucesión de signos de apertura de interrogación y apertura de exclamación tiene una palabra que comienza con mayúscula. La regla no asigna anotaciones pero permite evitar que a los signos de apertura de interrogación o de exclamación se los considere como marcadores

del comienzo de una nueva oración, y que por lo tanto, se le asigne una anotación del tipo "Split" al paréntesis. La Figura 8.10 ejemplifica el resultado obtenido con y sin la regla parentesisEnEnumeracion.

Texto analizado
"El viaje de Curiosity es un proyecto requirió una inversión de casi 2.500 millones de dólares pero que permitirá responder a múltiples interrogantes sobre el planeta rojo:
1) ¿Existe o existió vida en Marte?
2) Si hubo vida pero se extinguió ¿qué condiciones llevaron provocaron la extinción?
3) ¿Existen o existieron fuentes de agua en Marte?
4) ¿Están dadas las condiciones para enviar una misión con astronautas?"

Oraciones identificadas en find1
"El viaje de Curiosity es un proyecto requirió una inversión de casi 2.500 millones de dólares pero que permitirá responder a múltiples interrogantes sobre el planeta rojo:"

"1)"

"¿Existe o existió vida en Marte?"

"2) Si hubo vida pero se extinguió ¿qué condiciones llevaron provocaron la extinción?"

"3)"

" ¿Existen o existieron fuentes de agua en Marte?"

"4)"

"¿Están dadas las condiciones para enviar una misión con astronautas?"

Oraciones identificadas con la incorporación de parentesisEnEnumeracion
"El viaje de Curiosity es un proyecto requirió una inversión de casi 2.500 millones de dólares pero que permitirá responder a múltiples interrogantes sobre el planeta rojo:"

"1) ¿Existe o existió vida en Marte?"

"2) Si hubo vida pero se extinguió ¿qué condiciones llevaron provocaron la extinción?
"

"3) ¿Existen o existieron fuentes de agua en Marte?"

"4) ¿Están dadas las condiciones para enviar una misión con astronautas?"

Figura 8.10. Ejemplo de aplicación de regla parentesisEnEnumeracion.

VIII.5.5.9 Regla preguntaexclamacionNoSegundaParte (Regla Nueva)

Cuando una oración interrogativa o exclamativa comienza con mayúscula se dice que constituye una oración independiente o equivalentemente que no es una continuación de la oración que tiene a su izquierda. Esto significa que los signos de apertura de interrogación o de exclamación estarían demarcando el fin de la oración de la izquierda. Pero dado que con las reglas de la fase find1 no se podría reconocer la situación antes descripta, se tuvo que crear la regla preguntaexclamacionNoSegundaParte. Mediante esta regla lo que se hace es buscar en el texto una secuencia de uno a cinco signos de apertura de interrogación o apertura de exclamación que tenga a su derecha una palabra en mayúscula y a la izquierda una oración sin un signo de puntuación que marque su fin. Cuando el texto analizado contiene este patrón se procede a asignar una anotación del tipo "Split" a la palabra o carácter de la izquierda más cercano a los signos de apertura de la interrogación o la exclamación. El efecto de la incorporación de la regla preguntaexclamacionNoSegundaParte se puede visualizar en la Figura 8.11.

Texto analizado
"También pueden descargar el documento ca4f6_manualphp.pdf que es un mini tutorial sobre PHP (creado por Alexander Garzon) ¡¡Pueden presentar los ejercicios en ASP!! aquellos fanáticos de este lenguaje pero se prefiere PHP en lugar de las opciones para desarrollo web ofrecidas en .NET."

Oraciones identificadas en find1
"También pueden descargar el documento ca4f6_manualphp.pdf que es un mini tutorial sobre PHP (creado por Alexander Garzon) ¡¡Pueden presentar los ejercicios en ASP!! aquellos fanáticos de este lenguaje pero se prefiere PHP en lugar de las opciones para desarrollo web ofrecidas en .NET."

Oraciones identificadas con la incorporación de preguntaexclamacionNoSegundaParte
"También pueden descargar el documento ca4f6_manualphp.pdf que es un mini tutorial sobre PHP (creado por Alexander Garzon)"

"¡¡Pueden presentar los ejercicios en ASP!! aquellos fanáticos de este lenguaje pero se prefiere PHP en lugar de las opciones para desarrollo web ofrecidas en .NET."

Figura 8.11. Ejemplo de aplicación de regla preguntaexclamacionNoSegundaParte

VIII.5.5.10 Regla eliminaranotacionExclIntePtosSuspensivos (Regla Nueva)

La regla fue incorporada para tratar con situaciones donde una secuencia de uno a cinco signos de cierre de interrogación o de cierre de exclamación tenga a su derecha únicamente puntos suspensivos. Con las reglas de la fase find1 se asignaría de manera

equivocada una anotación del tipo "Split" tanto a los signos de cierre de exclamación o de cierre de interrogación como a los puntos suspensivos. Si bien en la fase1 mediante las reglas pregExclamPtoSuspensivoNoCierraEnunciado y pregExclamPtoSuspensivoCierra Enunciado se considera el caso en que una oración interrogativa o exclamativa puede estar sucedida por puntos suspensivos, las dos reglas examinan la anotación del tipo "Token" a la derecha de los puntos suspensivos por lo que fallan cuando a la derecha no se tiene anotaciones del tipo "Token". Para facilitar la comprensión de las ideas expuestas se incluyó un caso de ejemplo en la Figura 8.12.

Texto analizado
"Cualquier duda, sugerencia, o comentario lo pueden plantear enviando un e-mail a admin.sistemas@etronfarmacia.com, en http://www.etronfarmacia.com/contact.htm, por Teléf. al 011- 51281111 de 9 A.M. a 6 P.M. de Lunes a Viernes. En la próxima semana habilitaremos un buzón de sugerencias para que el proceso sea ¡totalmente anónimo! ..."

Oraciones identificadas en find1
"Cualquier duda, sugerencia, o comentario lo pueden plantear enviando un e-mail a admin.sistemas@etronfarmacia.com, en http://www.etronfarmacia.com/contact.htm, por Teléf. al 011- 51281111 de 9 A.M. a 6 P.M. de Lunes a Viernes."

"En la próxima semana habilitaremos un buzón de sugerencias para que el proceso sea ¡totalmente anónimo!"

"..."

Oraciones identificadas con la incorporación de eliminaranotacionExclIntePtosSuspensivos.
"Cualquier duda, sugerencia, o comentario lo pueden plantear enviando un e-mail a admin.sistemas@etronfarmacia.com, en http://www.etronfarmacia.com/contact.htm, por Teléf. al 011- 51281111 de 9 A.M. a 6 P.M. de Lunes a Viernes."

"En la próxima semana habilitaremos un buzón de sugerencias para que el proceso sea ¡totalmente anónimo! ..."

Figura 8.12. Ejemplo de aplicación de regla eliminaranotacionExclIntePtosSuspensivos.

VIII.5.6 Fase find3

La fase se compone solo de la regla ptosSuspensivosdosPuntosAntesListado que permite identificar situaciones donde los puntos suspensivos tienen a continuación el signo de dos puntos y el carácter de nueva línea, para luego asignar al signo de dos puntos una anotación del tipo "Split". Si bien la regla dosPuntosAntesListado de la fase find1 establece que cuando el signo de dos puntos precede al carácter de nueva línea lo que

corresponde es considerar al signo de dos puntos como divisor de oraciones, ello no sucede, porque cuando la regla ignorarPuntoSuspensivos identifica puntos suspensivos seguidos del signo de dos puntos consume el signo de dos puntos y lo convierte en no disponible para el resto de las reglas de la fase find1. No se puede incluir la regla ptosSuspensivosdosPuntosAntesListado en la fase find2 porque entraría en conflicto con las reglas ptoDeEnumeracionNoDivisor y parentesisEnEnumeracion. En la Figura 8.13 se muestran los resultados a los que se arriba cuando interviene ptosSuspensivosdosPuntosAntesListado y cuando no interviene.

Texto analizado

"Desde la comisión organizadora se ha puntualizado que…:

a) Se podrán hacer fotos a la pintura, pero sin traspasar la zona acotada que estará a un metro de la pared.
b) Se colocarán carteles que recuerden al público que se encuentran en un espacio religioso.
c) ¡Queda totalmente prohibido! el expendio de bebidas en un área de 500 metros a la redonda del santuario."

Oraciones identificadas en find1 y find2

"Desde la comisión organizadora se ha puntualizado que…:

a) Se podrán hacer fotos a la pintura, pero sin traspasar la zona acotada que estará a un metro de la pared."

"b) Se colocarán carteles que recuerden al público que se encuentran en un espacio religioso."

"c) ¡Queda totalmente prohibido! el expendio de bebidas en un área de 500 metros a la redonda del santuario."

Oraciones identificadas con la incorporación de ptosSuspensivosdosPuntosAntesListado

"Desde la comisión organizadora se ha puntualizado que…:"

"a) Se podrán hacer fotos a la pintura, pero sin traspasar la zona acotada que estará a un metro de la pared."

"b) Se colocarán carteles que recuerden al público que se encuentran en un espacio religioso.

"c) ¡Queda totalmente prohibido! el expendio de bebidas en un área de 500 metros a la redonda del santuario."

Figura 8.13. Ejemplo de aplicación de regla ptosSuspensivosdosPuntosAntesListado

VIII.5.7 Fase find4

Esta fase contiene dos reglas que tratan con ítems de una enumeración o listado que se encuentran al principio del texto analizado.

VIII.5.7.1 Regla eliminarAnotSplitdePtoEnumeracion (Regla Nueva)

Si al principio del texto analizado se tuviera un ítem de un listado o enumeración las reglas de las fases find1, find2, y find3 considerarían que el número del ítem junto con el punto constituyen una oración independiente de lo que se tiene a la derecha. Sin embargo, lo correcto es tratar al número de ítem junto al signo de punto como parte de la oración que tiene a su derecha. Con la regla eliminarAnotSplitdePtoEnumeracion se suprime la anotación del tipo "Split" que en las fases precedentes se le asigna al punto del ítem del listado. En la Figura 8.14 se muestra por medio de un ejemplo los resultados que arroja *ANNIE Sentence Splitter* cuando interviene eliminarAnotSplitdePtoEnumeracion y cuando no interviene.

Texto analizado
"3. Resistencia a antibióticos
Aumento de la resistencia a los antibióticos y antivirales que puede provocar que en el futuro no serán eficaces para hacer frente a otro tipo de enfermedades (neumonía, gripe, VIH, infecciones estafilocócicas de la sangre, etc.)."

Oraciones identificadas en find1, fin2, y find3
"3."

"Resistencia a antibióticos
Aumento de la resistencia a los antibióticos y antivirales que puede provocar que en el futuro no serán eficaces para hacer frente a otro tipo de enfermedades (neumonía, gripe, VIH, infecciones estafilocócicas de la sangre, etc.)."

Oraciones identificadas con la incorporación de eliminarAnotSplitdePtoEnumeracion.
"3. Resistencia a antibióticos
Aumento de la resistencia a los antibióticos y antivirales que puede provocar que en el futuro no serán eficaces para hacer frente a otro tipo de enfermedades (neumonía, gripe, VIH, infecciones estafilocócicas de la sangre, etc.)."

Figura 8.14. Ejemplo de aplicación de regla eliminarAnotSplitdePtoEnumeracion

VIII.5.7.2 Regla omitirAnotSplitdePtoEnumeracion (Regla Nueva)

Para tratar con los ítems de un listado o enumeración localizados al comienzo de un texto la regla eliminarAnotSplitdePtoEnumeracion busca un número seguido de un signo de punto con una anotación del tipo "Split" y luego elimina la anotación del tipo "Split"

del signo de punto. Sin embargo, si el número seguido de punto estuviera localizado en el medio o al final del texto analizado, el signo de punto si marcaría el fin de la oración. La regla omitirAnotSplitdePtoEnumeracion impide que se elimine la anotación del tipo "Split" del signo de punto cuando la secuencia número más signo de punto con anotación "Split" se ubica en el medio o al final del texto. Para clarificar las nociones precedentes se muestra un caso de ejemplo en la Figura 8.15.

Texto analizado
"4. De acuerdo con Aracely Argueta, jefa de residentes del Hospital Nuestra Señora de Fátima de Cojutepeque, la cantidad de personas atendidas en el centro de salud fue 57. De las cuales cinco fueron diagnosticadas de mayor gravedad.
5. Según manifestó la vocera del centro asistencial, dentro de los 57 lesionados no figuraba el conductor del autobús."

Oraciones identificadas en find1, fin2, y find3
"4. De acuerdo con Aracely Argueta, jefa de residentes del Hospital Nuestra Señora de Fátima de Cojutepeque, la cantidad de personas atendidas en el centro de salud fue 57. De las cuales cinco fueron diagnosticadas de mayor gravedad."

"5. Según manifestó la vocera del centro asistencial, dentro de los 57 lesionados no figuraba el conductor del autobús."

Oraciones identificadas con la incorporación de omitirAnotSplitdePtoEnumeracion
"4. De acuerdo con Aracely Argueta, jefa de residentes del Hospital Nuestra Señora de Fátima de Cojutepeque, la cantidad de personas atendidas en el centro de salud fue 57."

"De las cuales cinco fueron diagnosticadas de mayor gravedad."

"5. Según manifestó la vocera del centro asistencial, dentro de los 57 lesionados no figuraba el conductor del autobús."

Figura 8.15. Ejemplo de aplicación de regla omitirAnotSplitdePtoEnumeracion

VIII.5.8 Fase Split

La fase Split se encarga de crear las anotaciones del tipo "Sentence" en base a las anotaciones generadas en las fases find1, find2, find3, y fin4. No es necesario modificar las reglas de la fase Split para que trabajen con textos en español puesto que su funcionamiento depende exclusivamente de las anotaciones del tipo "Split" generadas con anterioridad. La fase Split se compone de dos reglas, internalSplits y externalSplits, que se describen a continuación.

VIII.5.8.1 Regla internalSplits (Regla Nueva)

Cada vez que en el texto analizado identifica una anotación del tipo "Split" con la característica "kind" igual a "internal" asigna una anotación del tipo "Sentence" al fragmento de texto comprendido entre la anotación del tipo "Split" actual y la anotación del tipo "Split" anterior. La anotación del tipo "Split" con la característica "kind" igual a "internal" forma parte del fragmento de texto cubierto por la anotación del tipo "Sentence" creada.

VIII.5.8.2 Regla externalSplits (Regla Nueva)

Cada vez que en el texto analizado identifica una anotación del tipo "Split" con la característica "kind" igual a "external" asigna una anotación del tipo "Sentence" al fragmento de texto comprendido entre la anotación del tipo "Split" actual y la anotación del tipo "Split" anterior. La anotación del tipo "Split" con la característica "kind" igual a "external" no forma parte del fragmento de texto cubierto por la anotación del tipo "Sentence" creada.

VIII.5.9 Fase reestablecerTokens

La fase contiene solo la regla reestablecerString que fue creada para restablecer el valor original de las características "string" y "kind" de las anotaciones del tipo "Token" editadas en la fase modificarTokens.

VIII.5.10 Fase postProcesarTokens

Esta fase por medio de la regla removerAnotTemporal se encarga de eliminar las anotaciones del tipo "preguntaModif" y "exclamacionModif" que son creadas por las reglas de la fase preProcesarTokens.

VIII.5.11 Fase cleanup

Está formada por la regla cleanUp que elimina las anotaciones temporales creadas por *ANNIE Sentence Splitter* para poder reconocer las sentencias y las abreviaturas definidas explícitamente en el archivo señalado por el parámetro transducerURL. La regla cleanUp forma parte de la gramática que por defecto usa *ANNIE Sentence Splitter* y no fue necesario modificarla.

VIII.6 ELIMINACIÓN DE *STOP WORDS*

GATE no cuenta con un recurso dedicado exclusivamente a la supresión de *stop words* pero esta operación de pre procesamiento puede implementarse recurriendo a *JAPE Transducer* y *ANNIE Gazetteer* como se muestra a continuación.

VIII.6.1 Definición de *stop words* con *ANNIE Gazetteer*

El rol de un *gazetteer* es identificar en el texto nombres de entidades definidos en un archivo de texto plano denominado lista. Cada lista representa un conjunto de nombres de entidades con un nombre de entidad por línea (por ejemplo: nombre de ciudades, organizaciones, días de la semana, etc.). En el presente trabajo de investigación se creó una lista *gazetteer* de nombre "stopwords.lst" compuesta por *stop words* de uso común en el idioma español. Las *stop words* utilizadas en este trabajo se pueden consultar en el Anexo B y fueron extraídas del paquete tm (ver punto IX.1.4.2.2).

VIII.6.2 Eliminación de *stop words* con *JAPE Transducer*

Con *ANNIE Gazetteer* se reconoce las *stop words* pero para eliminarlas se tuvo que recurrir a *JAPE Transducer*. Este recurso de procesamiento permite ejecutar una gramática JAPE sobre el texto analizado. La gramática definida para suprimir las *stop words* se compone de la fase de nombre suprimirStopWords. Las reglas que forman parte de la fase suprimirStopWords se describen en los próximos dos apartados.

VIII.6.2.1Regla stopWordsSinAcento (Regla nueva)

Para identificar *stop words* en el texto analizado utiliza la lista de *stop words* definidas explícitamente para *ANNIE Gazetteer*. Cuando se reconoce una *stop word* se elimina la anotación del tipo "Token" asociada con el fragmento de texto que corresponde a la *stop word* y no el fragmento de texto. Debido a que en los archivos .lst no pueden incluirse caracteres acentuados, con la regla stopWordsSinAcento solo se puede identificar y eliminar las *stop words* sin acento.

VIII.6.2.2Regla stopWordsAcentuadas (Regla nueva)

Mediante esta regla se comprueba si la característica "string" de las anotaciones del tipo "Token" del texto analizado es igual a alguna de las *stop words* acentuadas incluidas

en el Anexo B. Cuando se reconoce una *stop word* acentuada se procede a eliminar la anotación del tipo "Token" y por consiguiente no se altera el texto analizado.

VIII.7 PRUEBAS APLICADAS A LAS ADAPTACIONES DE GATE

Para comprobar el funcionamiento correcto de las gramáticas JAPE creados tanto para adaptar *ANNIE Sentece Splitter* como para implementar la operación de supresión de *stop words* se aplicaron técnicas de pruebas de caja negra (conjetura de errores, particiones de equivalencia y análisis de valores límite) y de caja blanca (cobertura de sentencia, cobertura de decisión y cobertura de condición múltiple). Al momento de chequear que no existieran conflictos entre las reglas de una misma fase y entre las reglas de las diferentes fases se siguió una estrategia de integración ascendente.

La documentación de las pruebas aplicadas se incluyó en la carpeta *Adaptaciones_de_GATE* del CD que acompaña al presente documento.

CAPÍTULO IX

DOCUMENTACIÓN DE LA APLICACIÓN DE CRISP-DM

IX.1 FASE 1: COMPRENSIÓN DEL NEGOCIO

IX.1.1 Determinar los Objetivos del Negocio

IX.1.1.1 Antecedentes

Los entornos de *e-learning* son una fuente rica en información para estudiar los procesos de enseñanza y aprendizaje que tienen lugar en cursos virtuales, dado que permiten contar con un registro completo de las interacciones estudiante-estudiante como también estudiante-docente. Sin embargo, cuando se dispone de una gran cantidad de interacciones las aproximaciones de análisis manual son prácticamente inviables debido al tiempo, dinero, y esfuerzo que requieren (Chen & Wasson, 2002; Rosé *et al.*, 2008). De allí que sea necesario el uso de aproximaciones automatizadas que permitan liberar a los investigadores de parte de las tareas.

Para modelar las interacciones en la práctica se han utilizado tanto interfaces estructuradas como semi-estructuradas (oraciones de apertura, contribuciones dirigidas por menú, contribuciones basadas en diagramas) y texto libre. Si bien las interfaces estructuradas y semi-estructuradas simplifican el proceso de análisis de interacciones ya que no es preciso recurrir a técnicas de procesamiento de lenguaje natural, también pueden tener efectos negativos en el proceso de interacción puesto que restringen el tipo de actos comunicativos y hacen que la comunicación sea más lenta (Chen, 2006; Constantino-González & Suthers, 2001; Olivares, 2007; Soller *et al.*, 2005). En lo que respecta al texto libre, la minería de textos se muestra a priori como una herramienta adecuada para analizar este tipo de interacciones dada su capacidad para hacer frente a la vaguedad, incerteza, y gran cantidad de estructuras que caracterizan al lenguaje natural (Hotho *et al.*, 2005).

Teniendo en cuenta lo planteado en los párrafos anteriores, en este proyecto se decidió utilizar la técnica de minería de textos para identificar las habilidades sociales manifestadas por e-tutores de ACSC a partir del análisis de los mensajes publicados por ellos en foros y chat.

IX.1.1.2 Objetivos de negocio

- Identificar automáticamente las habilidades sociales manifestadas por los e-tutores de ACSC.

Partiendo del objetivo anterior se pueden derivar las dos preguntas de negocio indicadas a continuación:

- ¿Cuáles son las características lingüísticas de los mensajes que permiten lograr mayores grados de efectividad en la identificación automática de las habilidades sociales de e-tutores?
- ¿Cuál es el grado de efectividad de los algoritmos de minería de textos en la identificación automática de las habilidades sociales manifestadas por los e-tutores de ACSC?

IX.1.1.3 Criterio de éxito de negocio

- Identificar como mínimo el 60 % de las habilidades sociales manifestadas por los e-tutores en los mensajes publicados en entornos de ACSC.
- Identificar habilidades sociales manifestadas por los e-tutores en los mensajes publicados en entornos de ACSC con un margen de error menor o igual al 30 %.

IX.1.2 Evaluar la Situación

IX.1.2.1 Inventario de recursos

IX.1.2.1.1 Personal

- Un analista encargado de aplicar la metodología CRISP-DM y de interactuar con el software de minería de textos (Analista).
- Un experto en minería de datos y en ACSC para brindar asistencia en el desarrollo de las actividades de CRISP-DM como también para evaluar los resultados obtenidos (Experto).
- Un psicopedagogo para evaluar los resultados obtenidos (Psicopedagogo).

IX.1.2.1.2 Datos

- Cinco archivos .doc que registran lo sucedido en cinco sesiones de chat de un seminario de formación para los docentes mexicanos que participaron del proyecto "Trabajo y aprendizaje colaborativo con tecnologías de información y comunicación en ciencias" (Tactics).
- Seis archivos .doc que registran lo acaecido en seis asesorías vía chat desarrolladas en el marco del "Curso Preliminar de Fundamentos en Ciencias de la Computación" destinado a docentes mexicanos que participaron del proyecto "Trabajo y aprendizaje colaborativo con tecnologías de información y

comunicación en ciencias" (Tactics). Para este curso, se dispone además de treinta y un archivos .pdf de las comunicaciones estudiante-estudiante y estudiante-docente mediante la herramienta de foro.

- Once archivos .pdf que registran lo sucedido en diez sesiones de chat de un curso denominado "Propedéutico de Matemáticas" destinado a aspirantes de una maestría en ciencias de la computación o mecánica. El curso se dictó en el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) de México.
- Treinta archivos .htm que registran las comunicaciones estudiante-estudiante y estudiante-docente de la cátedra Metodología de la Investigación II de la carrera Licenciatura en Sistemas de Información de la UNSE. La plataforma de ACSC usada fue Moodle y la comunicación se hizo mediante la herramienta de foro.
- Veinte tres archivos .htm que registran las comunicaciones estudiante-estudiante y estudiante-docente de la cátedra Estadística correspondiente a la carrera Licenciatura en Sistemas de Información de la UNSE. La plataforma de ACSC usada fue Moodle y la comunicación se hizo mediante la herramienta de foro.

IX.1.2.1.3 Hardware

- Notebook Lenovo B570: Procesador Pentium B950 de 2.1 Ghz, 4 GB RAM, 500 GB disco.
- PC: Procesador AMD Athlon X2 7750 Black Edition de 2.7 Ghz, 2 GB RAM, 500 GB de disco.

IX.1.2.1.4 Software

- Windows xp server pack 2 de 32 bit.
- Windows 7 de 32 bit.
- GATE versión 6.0.
- Tree Tagger.
- Cygwin version 1.5.
- Weka version 3.6.4.
- ActivePerl version 5.12.3.1204.

IX.1.2.2 Requerimientos, supuestos, y restricciones

IX.1.2.2.1 Requerimientos

- El proyecto debe finalizar antes del 15 de Febrero de 2013.
- Los resultados arrojados por los clasificadores automáticos (habilidades identificadas) deben ser comprensibles por personas no familiarizadas con las habilidades de los e-tutores de ACSC.
- Los resultados arrojados por los clasificadores automáticos (habilidades identificadas) deben ser comprensibles por personas no familiarizadas con la minería de textos.
- Como parte del proceso de clasificación los mensajes de los e-tutores deben ser descompuestos oraciones sin necesidad de intervención del usuario.

IX.1.2.2.2 Supuestos

- En el presente proyecto no se tendrán en cuenta datos tales como fecha, hora, identidad del emisor, tópico bajo el cual se hace una contribución, etc., que suelen ser almacenados por entornos de ACSC sino que solo se analizará el contenido textual de los mensajes publicados por los e-tutores. Por lo cual, se asume que para identificar las habilidades manifestadas por los e-tutores basta con analizar los mensajes publicados.
- En lo relación al proceso de enseñanza y de aprendizaje se asume que se desarrolló completamente en un entorno de ACSC. Es decir, se considera que no hubo contacto cara a cara para resolver la actividad o tarea a la que pertenecen el conjunto de mensajes docentes usados en este proyecto.

IX.1.2.2.3 Restricciones

Una cuestión que tendrá gran impacto en el tiempo necesario para completar este proyecto es que se dispone de una sola persona para llevar a cabo la mayoría de las actividades de minería de textos. Durante este proceso, se priorizarán las acciones que no involucren costos porque al ser un trabajo de carácter académico no se cuenta con recursos financieros.

En lo que respecta a los datos disponibles, al analizar los mensajes publicados por los e-tutores (mediante la técnica de análisis de contenido) se identificaron 891 oraciones en las que potencialmente se manifiestan habilidades sociales de e-tutor. Esta cantidad de datos podría llegar a ser insuficiente para construir un modelo que reconozca

automáticamente al menos un 60 % de las habilidades sociales que fueron identificadas por expertos humanos y que, simultáneamente no asigne erróneamente atributos de habilidad a más del 30 % de las oraciones clasificadas.

IX.1.2.3 Riesgos y contingencias

Riesgo 1: La descomposición automática en oraciones de textos en español requiere que se tengan en cuenta múltiples reglas que entran en conflicto.

Plan de contingencia

- Crear casos de prueba para chequear cada regla por separado.
- Crear casos de prueba para chequear la interacción entre las reglas y asegurar que los conflictos se hayan resuelto.

Riesgo 2: La descomposición automática en oraciones de textos en español no puede realizarse correctamente con los recursos brindados por la herramienta de minería de textos seleccionada.

Plan de contingencia

- Buscar componentes de software de terceros que descompongan automática en oraciones de textos en español y que puedan ser integrados en la herramienta de minería de textos seleccionada.

Riesgo 3: La descomposición automática en oraciones de textos en español no puede realizarse correctamente con los recursos brindados por la herramienta de minería de textos seleccionada, ni tampoco existe un componente de software de terceros que realice la descomposición y sea compatible con la herramienta de minería de textos.

Plan de contingencia

- Buscar una herramienta de minería de textos alternativa.
- Solicitar a los usuarios que ingresen oración por oración el texto a ser analizado.

Riesgo 4: Los datos recolectados no son suficientes para construir un modelo que permita identificar automáticamente las habilidades sociales de los e-tutores de ACSC.

Plan de contingencia

- Solicitar la colaboración de investigadores que tengan a su disposición un registro de las interacciones estudiante-docente en entornos de ACSC de manera de extraer (mediante análisis de contenido) aquellas interacciones donde se manifiesten habilidades sociales.
- Llevar a cabo nuevas experiencias de ACSC con docentes y estudiantes de la FCEyT de la UNSE.

Riesgo 5: Los resultados presentados por los clasificadores automáticos (habilidades identificadas) son incomprensibles para los usuarios.

Plan de contingencia

- Realizar cambios en la interfaz de la herramienta de minería de textos de modo que los resultados se muestren de una manera más amigable para el usuario final.
- En el caso de que no sean posibles adaptaciones a la herramienta de minería de textos, confeccionar un manual de ayuda donde se explique cómo interpretar los resultados mostrados arrojados por los clasificadores automáticos.

IX.1.2.4 Terminología

Anotación: es un tipo de meta dato asociado a una sección particular del contenido de un documento (Cunningham *et al*., 2011). Las anotaciones existen siempre como miembros de los conjuntos por lo que no pueden ser creadas de forma independiente al conjunto al cual pertenecen.

Aprendizaje Colaborativo Soportado por Computadora (ACSC): uso educacional de la tecnología de computadoras para facilitar el aprendizaje grupal. El ACSC es la unión de la teoría e investigación en aprendizaje colaborativo basado en clase y comunicación mediada por computadora (CMC) en la búsqueda de entender como los proyectos de grupo basados en CMC pueden mejorar el aprendizaje (McWhaw *et al.*, 2003).

Atributo de habilidad: para identificar las habilidades sociales manifestadas por los e-tutores de ACSC se tomará como base una clasificación de habilidades descripta en el capítulo IV. La mencionada taxonomía de habilidades se organiza en tres niveles

(habilidad, sub habilidad, y atributo de habilidad) donde cada nivel proporciona mayores detalles del nivel que le precede.

Clasificación: la tarea de clasificación puede definirse formalmente como la tarea de aproximar una función de asignación de categoría desconocida F: DxC -> {0,1}, donde D es el conjunto de documentos y C es el conjunto de categorías pre definidas (Feldman & Sanger, 2007). El valor de F(d,c) es 1 si el documento d pertenece a la categoría c de otra manera el valor es 0. La función de aproximación M: DxC -> {0,1} se conoce como clasificador, modelo, o hipótesis, y el objetivo es construir un clasificador que produzca resultados tan cercanos como sea posible a la función de asignación de categoría F (Sebastiani, 2002; Feldman & Sanger, 2007).

Característica de anotación: las anotaciones tienen asociados pares del tipo <nombre-atributo, valor-atributo>. El nombre del atributo o la característica es del tipo *String* mientras los valores pueden ser cualquier tipo de objeto Java (Cunningham *et al*., 2011). Así por ejemplo si se aplica el recurso de procesamiento *GATE Unicode Tokeniser* cada *token* tendrá las características o atributos “kind”, “length”, “orth” y “string”.

Conjunto de anotación: contiene anotaciones y mantiene una serie de índices para proveer rápido acceso al contenido de las anotaciones (Cunningham *et al*., 2011).

Corpus: es una entidad conceptual similar a una Base de Datos que permite almacenar y manipular una colección de uno o más documentos de texto.

e-learning: se lo puedo definir como aquella actividad que utiliza de manera integrada y pertinente computadores y redes de comunicación, en la formación de un ambiente propicio para la construcción de la experiencia de aprendizaje (Foix & Zavando, 2002).

e-tutor: es la persona que apoya los procesos de aprendizaje por medio de la interacción directa con los estudiantes que frecuentemente están separados unos de otros y del e-tutor en tiempo y espacio (Sulčič & Sulčič, 2007). Al soportar las interacciones un e–tutor busca que los estudiantes puedan aprender de manera efectiva en ambientes online.

Habilidad: capacidad física o mental que una persona tiene y que le permite completar una tarea o lograr un objetivo. De manera más precisa, una habilidad (o conjunto de habilidades) es esencialmente una colección de uno o más procedimientos ejecutables que un operario usa para completar exitosamente una tarea o trabajo de un modo efectivo y/o eficiente, de manera de producir alguna salida o artefacto deseado (Barker, 2010).

Normalización: consiste en convertir los tokens a un formato estándar de modo tal de reducir el número de tipos de token e incrementar la frecuencia de ocurrencia de algunos tipos individuales (Weiss *et al.*, 2005). Cuando el proceso de normalización se limita a regularizar variaciones gramaticales tales como singular/plural y pasado/presente, el proceso se conoce como *inflectional stemming* (Weiss *et al.*, 2005). Una forma de normalización más agresiva es el *stemming to a root* que busca lograr la forma raíz de los tokens sin prefijos y sufijos inflexionales o derivacionales (Weiss *et al.*, 2005). Una aproximación de normalización diferente al *stemming* es la lematización ya que busca que las múltiples formas de representar un *token* sean reducidas a su forma primitiva (Aranha & Passos, 2008). Al aplicar lematización los *tokens* "analizar", "analizador" y "analizado" tendrían que ser sustituidos por el lexema "análisis".

Medida F: una manera de combinar la precisión y el recall es mediante la medida F que corresponde a un promedio ponderado de los dos valores. Cuando se le da igual importancia a la precisión y el recall la medida se denomina F_1 y se define como: 2 /(1/precisión + 1/recall) (Feldman & Sanger 2007).

Precisión: mide el número de documentos clasificados correctamente como un porcentaje del número de documentos clasificados. Es decir, mide cuantos de los documentos que el clasificador identificó fueron correctos independientemente de si también fallo en recuperar documentos correctos (Sebastiani, 2002; Cunningham *et al.*, 2011). La precisión se puede definir mediante la siguiente ecuación:

$$\frac{\text{Número de predicciones positivas correctas}}{\text{Número de predicciones positivas}}$$

Recall: mide el número de documentos clasificados correctamente como un porcentaje del número total de documentos correctos. En otras palabras, mide cuantos de los documentos

que deberían haber sido identificados fueron identificados, independientemente de cuantas clasificaciones falsas se hayan hecho (Sebastiani, 2002; Cunningham *et al*., 2011).

$$\frac{\text{Número de predicciones positivas correctas}}{\text{Número de documentos de la clase}}$$

Stop words: son palabras (artículos, preposiciones, conjunciones, etc.) que no contribuyen al contenido semántico de los documentos y que al no tener un valor añadido real pueden ser eliminadas (Feldman & Sanger, 2007).

Tokenización: la tokenizacion o identificación de Token consiste en descomponer un texto en tipos de Token tales como palabras, números, signos de puntuación, etc. Un Token debe ser entendido como la unidad lexica básica de un lenguaje.

Para hacer referirse a los atributos de habilidad social de la taxonomía presentada en el capítulo IV se utilizaran las abreviaturas definidas en Tabla 7.1.

IX.1.2.5 Costos y beneficios

Dado que el presente proyecto forma parte del desarrollo de un trabajo final de graduación, no son relevantes los costos en términos financieros y solo se requiere invertir tiempo. En relación a los beneficios, se considera que de tener éxito en la identificación automática de las habilidades sociales de los e-tutores de ACSC se estaría contribuyendo a la reducción del costo, tiempo, y esfuerzo que le puede suponer a una persona el analizar las interacciones registradas en entornos de interfaces no estructuradas (texto libre) con el fin de conocer las habilidades que poseen los docentes que tienen a su cargo el dictado de clases.

IX.1.3 Determinar los Objetivos de la Minería de Datos

IX.1.3.1 Objetivos minería de datos

Construir un modelo que permita identificar automáticamente las habilidades sociales manifestadas por los e-tutores de ACSC en los mensajes que intercambian con estudiantes o grupos de estudiantes mediante una herramienta de chat o foro.

IX.1.3.2 Criterios de éxito de la minería

- El modelo debe permitir identificar las habilidades sociales manifestadas por los e-tutores de ACSC con una precisión de 0.70.
- El modelo debe permitir identificar las habilidades sociales manifestadas por los e-tutores de ACSC con un recall de 0.60.
- Determinar la o las operaciones de pre procesamiento que permiten lograr características lingüísticas que conducen a los más altos grados de efectividad en la identificación automática de las habilidades sociales de e-tutores de ACSC.
- Establecer el grado de efectividad de diversos algoritmos de clasificación en la identificación automática de las habilidades sociales manifestadas por los e-tutores de ACSC.

IX.1.4 Producir el Plan del Proyecto

IX.1.4.1 Plan del proyecto

En la Tabla 9.1 se muestran las principales actividades del proyecto, el tiempo estimado, y los recursos humanos que se necesitan. El orden temporal muestra la dependencia entre las actividades y el paralelismo refleja que dos o más actividades se pueden desarrollar al mismo tiempo. Las entradas y salidas de cada actividad, así como aquellas actividades que es necesario repetir serán especificadas conforme avance el proyecto

Actividades	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Recursos
I-Entendimiento del Negocio						
1. Determinar objetivos de negocio						Analista
2. Evaluar la situación						
3. Determinar objetivos de la minería de datos						
4. Producir plan del proyecto						
5. Evaluar técnicas y herramientas						
II-Entendimiento de datos						Analista
III- Preparación de datos						
1. Limpiar los datos						Analista
2. Construir los datos						
3. Formatear los datos						
IV- Modelado						
1. Seleccionar técnica de modelado						Analista
2. Generar diseño de prueba						
3. Construir el modelo						
4. Evaluar el modelo						Analista
V-Evaluación						Experto-Psicopedagogo-Analista
VI-Despliegue						Analista

Tabla 9.1. Planificación del proyecto de minería de textos

IX.1.4.2 Evaluación inicial de técnicas y herramientas

IX.1.4.2.1 Evaluación de técnicas

Tal como se comento anteriormente en el presente proyecto se busca construir un modelo que detecte automáticamente las habilidades manifestadas por e-tutores al analizar los mensajes que intercambian con sus estudiantes en un entorno de ACSC. El problema de

clasificación de textos puede resolverse mediante un proceso inductivo que automáticamente construye un clasificador observando las características de un conjunto de documentos manualmente clasificados por expertos de dominio. En base a estas características el proceso inductivo infiere las condiciones que documentos previamente no examinados deberían cumplir para ser clasificados bajo una u otra categoría. La construcción inductiva de clasificadores puede ser abordada con una diversidad de métodos como los que se comentan a continuación.

Clasificadores probabilísticos

Tienen en cuenta la probabilidad de ocurrencia de las palabras en un documento. Si consideramos que C es la etiqueta de clase en la que estamos interesados y que x es un vector de características que denota la presencia o ausencia de palabras en los documentos, el objetivo es estimar Pr(C|x). Es decir, que se busca estimar la probabilidad de una clase dada la presencia o ausencia de determinadas palabras (Weiss *et al.*, 2005).

Clasificadores de árbol de decisión

Se trabaja con árboles donde los nodos internos están etiquetados con características, los arcos con pruebas que se aplican al peso de las características en los documentos, y los nodos hoja con categorías (Sebastiani, 2002; Feldman & Sanger, 2007). Para asignar la categoría a un documento se comienza por el nodo raíz y se continúa de manera descendente por los arcos cuyas condiciones son satisfechas por los documentos hasta que se alcanza un nodo hoja. Al documento se le asigna la categoría que corresponde al nodo hoja (Feldman & Sanger, 2007).

Clasificadores de regla de decisión

Un clasificador construido mediante el método de aprendizaje de regla inductiva consiste de una regla condicional con la premisa en forma normal disyuntiva (Sebastiani, 2002; Feldman & Sanger, 2007). En la premisa se representa la presencia o ausencia de palabras clave mientras que la clausula denota la decisión de clasificación (Sebastiani, 2002).

Redes neuronales

Un clasificador de red neuronal es una red de unidades donde las unidades de entrada representan características, la o las unidades de salida representan la categoría o categorías de interés, y los pesos en los arcos que conectan las unidades representan relaciones de

dependencia (Sebastiani, 2002). Para clasificar un documento los pesos de sus características son cargados en los nodos de entrada, la activación se propaga hacia delante a lo largo de la red, y los valores finales de los nodos de salida determinan las decisiones de categorización (Feldman & Sanger, 2007).

Clasificadores basados en ejemplo

No se construye una representación declarativa explicita de las categorías sino que se computa directamente la similitud entre el documento a ser clasificado y los documentos de entrenamiento. La fase de entrenamiento para estos clasificadores consiste simplemente en almacenar representaciones de los documentos de entrenamiento junto con sus categorías. El ejemplo más prominente de los clasificadores basados en ejemplo es el algoritmo KNN donde para decidir si el documento *d* pertenece a la categoría *c* se chequea si los K documentos similares a *d* pertenecen también a la categoría *c*. Si la respuesta es positiva para una proporción suficientemente se toma una decisión positiva, caso contrario una decisión negativa. (Feldman & Sanger, 2007)

Máquinas de Vector de Soporte

En términos geométricos un clasificador de Maquinas de Vector de Soporte (SVM) binario (cuando se tienen solo dos categorías o clases) puede ser visto como un híper plano en el espacio de características que separa los puntos que representan instancias positivas de la categoría de los puntos que representan instancias negativas. El híper plano de clasificación se selecciona durante el entrenamiento como el único híper plano que separa las instancias positivas conocidas de las instancias negativas conocidas con el máximo margen. El margen es la distancia desde el híper plano al puto más cercano de los conjuntos positivo y negativo. Los híper planos SVM son completamente determinados por un conjunto relativamente pequeño de instancias de entrenamiento conocidas como vectores de soporte. (Feldman & Sanger, 2007)

Cualquiera de los métodos anteriores podría utilizarse para construir un clasificador, pero en el marco de este proyecto los criterios que se tendrán en cuenta para decidir cual usar son:

- La efectividad de la clasificación (precisión y recall) que puede lograrse con el modelo obtenido.
- La facilidad con la que se puede comprender el modelo construido.

- El soporte que brinda la herramienta de minería de textos seleccionada al método de construcción del modelo.

IX.1.4.2.2 Evaluación de herramientas de software disponibles

En una búsqueda realizada en Internet se descubrió que varias herramientas de software pueden ser utilizadas para aplicar minería de textos pero, teniendo en cuenta que el proyecto forma parte de un trabajo final de graduación se decidió analizar solamente las herramientas *open source* porque se las puede usar de manera gratuita. En la Tabla 9.2 se comparan las principales características de las siguientes herramientas *open source* de minería de textos:

- GATE: es una herramienta de minería de textos que incluye recursos que permiten desarrollar y utilizar componentes de software que procesan lenguaje natural (Cunningham *et al.*, 2011). En el sitio http://gate.ac.uk se puede descargar la última versión de GATE así como también acceder a novedades, tutoriales, ejemplos de código fuente, etc. sobre esta herramienta.
- Pimiento: es una librería que soporta el desarrollo de aplicaciones de minería de textos, no una aplicación por sí misma (García Adeva & Calvo, 2006). Si bien Pimiento es *open source* no se la puede descargar directamente de la web y para obtenerla uno debe comunicarse con sus creadores. El subdominio http://erabaki.ehu.es/jjga/pimiento perteneciente a la Universidad del País Vasco contiene información sobre el procedimiento para acceder a Pimiento.
- R/tm: R es un lenguaje y un entorno de programación para análisis estadístico y grafico. Sin embargo, los métodos de R están limitados a datos de entrada estructurados. El paquete tm proporciona la infraestructura básica necesaria para organizar, transformar, y analizar datos textuales (Feinerer, 2008). El trabajo de Feinerer (2008) explica cómo utilizar el paquete tm en el entorno R, muestra ejemplos de aplicaciones prácticas, e incluye el manual de usuario del paquete tm. Desde el sitio http://tm.r-forge.r-project.org se puede descargar el paquete tm y acceder a material relacionado con el mismo.
- Weka/Kea: Weka es una herramienta de minería de datos ampliamente utilizada que ofrece funcionalidades de minería de textos por medio de la librería Kea (Witten *et al.*, 2005). Weka se puede descargar desde http://www.cs.waikato.ac.nz/ml/weka/index.html en tanto que la librería Kea desde http://www.nzdl.org/Kea/download.html.

Analizadas las herramientas antes citadas, fue posible establecer tanto las similitudes como las diferencias que enuncian a continuación y se sintetizan en Tabla 9.2:

- Todas las herramientas de minería de textos brindan soporte a las operaciones de pre-procesamiento (remoción de stop words, POST, lematización, etc.) que permiten convertir el texto en un formato semi estructurado o estructurado.
- Todas las herramientas de minería de textos cuentan con un API que le permiten interactuar con otras aplicaciones de software.
- Las herramientas de minería de textos utilizan básicamente los mismos algoritmos (SVM, KNN, y Naïve Bayes) para la construcción de clasificadores.
- Todas las herramientas de minería de textos ofrecen soporte a una variedad de formatos de archivos pero, GATE se destaca porque puede trabajar con nueve formatos de archivo distintos.
- Si bien es cierto que mediante la API se podría incorporar funciones adicionales a todas las herramientas de minería de textos, GATE cuenta con facilidades que permiten adaptar sus recursos de procesamiento a las necesidades de los usuarios. Así por ejemplo, se pueden adaptar los recursos de procesamiento a un idioma o dominio de aplicación mediante el uso de expresiones regulares o del lenguaje JAPE.
- De las cuatro herramientas de minería de textos consideradas solo GATE y R/tm cuentan con manual de usuario. GATE se destaca por sobre R/tm no solo porque su guía de usuario es bastante completa y detallada, sino además, porque en la página web oficial se puede encontrar una wiki, videos tutoriales, consejos, ejemplos de código fuente, y una recopilación de preguntas realizadas frecuentemente. Todas estas cuestiones sobre la documentación son fundamentales cuando se tiene que aprender a usar un producto de software.

Producto	**Características**					
	Pre procesamiento	API	Algoritmos de clasificación	Formatos de archivo	Personalizar	Documentación
GATE	•	•	KNN, SVM, Naïve Bayes, PAUM, C4.5	texto plano , rtf, pdf, Microsoft Word, OpenOffice , e-mail , html, xml, sgml	•	•
Pimiento	•	•	KNN, SVM, Naïve Bayes, Rocchio	rtf, pdf, Microsoft Word, OpenOffice , html,		
R/tm	•	•	KNN, SVM, Red neuronal , Árbol de decisión, Naïve Bayes	texto plano, Microsoft Word , pdf, e-mail, csv texto, html, gmane rss		•
Weka/Kea	•	•	KNN, Naïve Bayes, C4.5			

Tabla 9.2. Comparación entre herramientas de minería de textos *open source*

En base al análisis realizado sobre las herramientas disponibles, se decidió utilizar GATE para la construcción de los clasificadores puesto que cuenta con facilidades para adaptar sus recursos de procesamiento a las necesidades de cada proyecto, dispone de documentación más completa y detallada sobre cómo usarlo, y permite trabajar con mayor cantidad de formatos de archivo. Un aspecto que no se puede pasar por alto al considerar a GATE es que existe una comunidad de desarrollados que constantemente incorpora nuevas funcionalidades y características. Así, por ejemplo, se pueden destacar las adaptaciones realizadas a la infraestructura de GATE para poder utilizar las capacidades de almacenamiento y cómputo de la *Cloud Computing* de modo tal que, se puedan llevar a cabo proyectos de minería de textos en bases de datos de gran tamaño que sería imposible procesar en un tiempo razonable en computadoras de escritorio.

IX.2 FASE 2: COMPRENSIÓN DE LOS DATOS

IX.2.1 Recolectar Datos Iniciales

Los datos sobre los que se aplicará técnicas de clasificación de minería de textos provienen de sesiones de aprendizaje colaborativo donde se utilizó chat o foro como herramienta de comunicación. Las sesiones contienen los mensajes publicados tanto por e-tutores como por estudiantes pero, teniendo en cuenta que el objetivo de negocio consiste en identificar automáticamente las habilidades sociales manifestadas por los e-tutores de ACSC, es evidente que no todos los mensajes publicados son igualmente relevantes. Tal como se detalla en el capítulo VII esto llevó a que se aplicará la técnica de muestreo de relevancia de manera tal de seleccionar, en los 106 archivos originales, los mensajes de los docentes en los que se considera se manifiestan habilidades sociales.

Los archivos de las sesiones de ACSC están almacenados en cuatro formatos diferentes: pdf, htm, rtf, y doc. Para evitar problemas de compatibilidad en la representación de caracteres se decidió reemplazar los formatos anteriores por archivos de texto plano y que los caracteres sean almacenados siguiendo el sistema de codificación Unicode. A cada mensaje de los docentes le corresponde un archivo txt separado. En la Tabla 9.3 se muestra la cantidad de archivos en texto plano que se derivan para cada uno de los archivos en formato pdf, htm, rtf, y doc.

Archivo Fuente	Formato Original	Total Archivos Texto Plano Derivados
Seminario de formación para los docentes que participaron del proyecto Tactics		
Sesión chat 1	doc	40
Sesión chat 2	rtf	34
Sesión chat 3	rtf	39
Sesión chat 4a	rtf	15
Sesión chat 4b	rtf	10
Curso Preliminar de Fundamentos en Ciencias de la Computación		
Como queda la negacion del complemento de un conjunto	pdf	1
Donde estan los ejercicios de la primera parte	pdf	1
Lógica Difusa	pdf	1
Asesora01-mate-chat-5-07-07	doc	25
asesoria02-mate-chat-10-07-07	doc	13
asesoria03-mate-chat-12-07-07	doc	35
asesoria04-mate-chat-17-07-07	doc	37
asesoria05-mate-chat-19-07-07	doc	32
asesoria07-mate-chat-31-07-07	doc	22

Tabla 9.3. Archivos en su formato original junto con los archivos de texto plano derivados

Archivo Fuente	Formato Original	Total Archivos Texto Plano Derivados
Curso Propedéutico de Matemáticas		
sesión1_chat1	pdf	37
sesión2_chat2	pdf	75
sesión3_chat3	pdf	25
sesión_chat4	pdf	6
sesion3_sabado_5jul08	pdf	30
sesión_chat_5	pdf	29
sesión_chat_6	pdf	24
sesión_chat_7	pdf	31
sesión_chat_09	pdf	21
sesión_chat_10	pdf	31
sesión_chat_11	pdf	20
Metodología de la Investigación		
Esquemas2	htm	1
Planteo del problema	htm	1
Planteo del problema 2	htm	1
Matriz de Consistencia	htm	1
esperamos correcciones y comentarios no duden en acerlas	htm	1
marcos metodologico y empirico	htm	1
Matrices de consistencia	htm	2
Cómo van con la Matriz de Consistencia	htm	1
el nombre de la aplicación y la hipótesis	htm	1
Matriz Consistencia de "Acceso indebido a los datos mediante WebApp"	htm	1
Borrador de Matriz de Consistencia	htm	1
Cómo van con la Matriz de Consistencia G2	htm	1
Formato de la Matriz	htm	1
Interacción en el foro	htm	1
Resultado de Evaluaciones y Cambio Fecha Semin 2	htm	1
Estadística	htm	1
Fecha limite para presentación del trabajo	htm	1
Foro para todo el grupo	htm	1
Libro	htm	1
Definición del protocolo	htm	1

Tabla 9.3 (Cont.)

Los archivos de las sesiones de ACSC en su formato original se pueden consultar en la subcarpeta *Experiencias_de_ ACSC* de la carpeta *Analisis_ de_ contenido* del CD incluido en este trabajo. Por su parte, los archivos de texto plano derivados se encuentran en la subcarpeta *Archivos_de_Texto_Plano* de la carpeta *Mineria_de_Textos*.

De la observación de la Tabla 9.3 se desprende que para los archivos htm se derivaron muy pocos archivos de texto plano en comparación con lo que sucede con los archivos doc, rtf, y pdf. La razón es que los archivos htm corresponden a interacciones estudiante-estudiante y estudiante-docente mediante la herramienta de foro. La participación de los docentes en cada hilo de discusión en un foro fue inferior a la cantidad de participaciones en cada sesión de chat. Téngase en cuenta que en cada hilo de discusión de un foro se trata por lo general un tema y las contribuciones suelen ser pocas en cantidad pero extensas en contenido, en cambio en una sesión de chat se discuten múltiples tópicos y los mensajes suelen ser numerosos en cantidad pero cortos en contenido.

En los archivos .txt los mensajes de los e-tutores no se almacenaron con toda la información asociada que existe en los archivos originales (pdf, doc, htm, rtf) porque se asume que para identificar las habilidades es suficiente con considerar el mensaje en sí mismo. Esto significa que de los mensajes publicados mediante la herramienta de foro se eliminaron atributos como la fecha, la hora, el autor, y el hilo de discusión al cual pertenece el mensaje; en tanto que, para los mensajes publicados mediante la herramienta de chat se optó por descartar el autor y la hora.

Para construir un clasificador no basta con conocer el contenido de los mensajes publicados por los e-tutores sino que también se necesita conocer cuál atributo de habilidad social se manifiesta en las oraciones que componen cada mensaje. A los atributos de habilidad que los expertos en pedagogía identificaron en los mensajes mediante la técnica de análisis de contenido, se los puede representar con un campo rotulado “Habilidad manifestada”. En GATE este campo será reproducido mediante la característica de anotación “atributo_hab” que tendrá cada anotación del tipo “Sentence”.

IX.2.2 Describir los Datos

Los datos del proyecto consisten de los mensajes publicados por los e-tutores y los atributos de habilidad social vinculados a cada una de las oraciones que forman parte de los mensajes. El atributo de habilidad social se representa mediante el campo “Habilidad manifestada” cuyos valores se corresponden con las abreviaturas de los atributos de habilidad social del apartado IX.1.2.4. Téngase en cuenta que a la abreviatura OAH se la usa cuando en la contribución del docente no se identifica la manifestación de alguno de los atributos de habilidad social.

En la Figura 9.1 y la Tabla 9.4 se muestra como en un mensaje publicado por un e-tutor se puede reconocer más de un atributo de habilidad social.

Mauricio...
Estaría bueno que digas qué es ESFE y las características... Fijate... qué variables informáticas aparecen en tu hipótesis??? Buscá y no vas a encontrar ninguna... Cuáles son las variables? correctitud?... Piensen un poco más... QUÉ OPINAN EL RESTO DE INTEGRANTES DEL FORO?

Figura 9.1. Ejemplo de mensaje de e-tutor

Oraciones Identificadas	Habilidad Manifestada
Mauricio...	OAH
Estaría bueno que digas qué es ESFE y las características...	AHFE
Fijate... qué variables informáticas aparecen en tu hipótesis???	AHPTMI
Buscá y no vas a encontrar ninguna...	OAH
Cuáles son las variables? correctitud?...	AHSCTT
Piensen un poco más...	AHUTEI
QUÉ OPINAN EL RESTO DE INTEGRANTES DEL FORO?	AHPC

Tabla 9.4. Ejemplo de múltiples habilidades identificadas en un mensaje

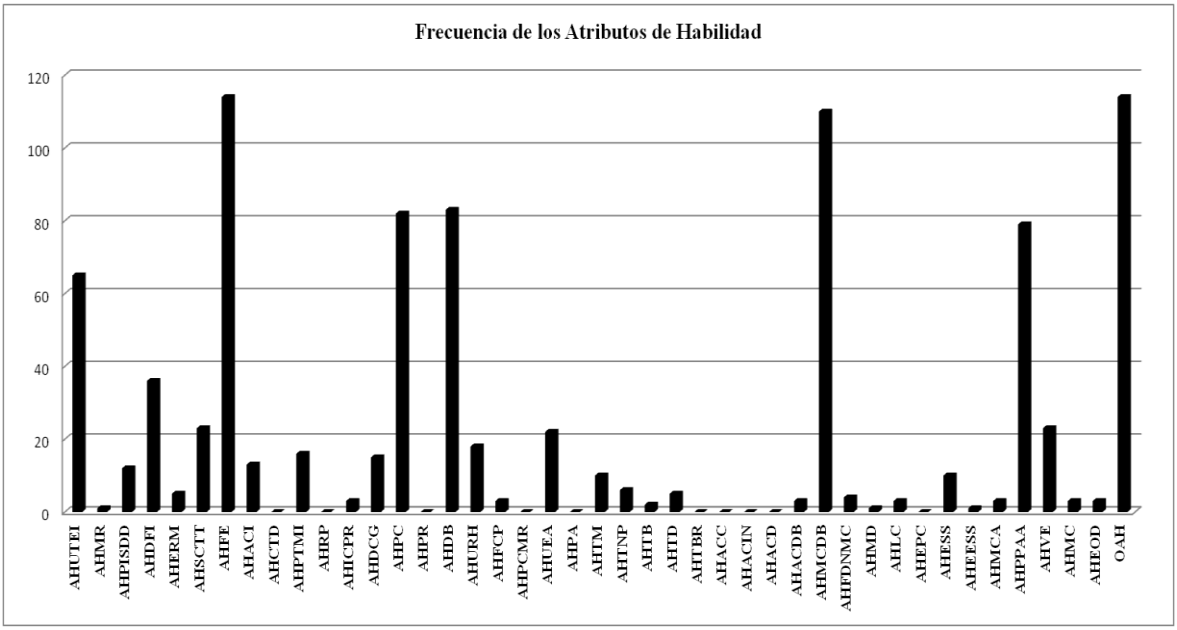


Figura 9.2. Diagrama de barra con la frecuencia de los atributos de habilidad

La Figura 9.2 indica la frecuencia con la que se identificó cada atributo de habilidad social en el conjunto de 981 oraciones analizadas. A simple vista puede comprobarse que una pequeña porción de los atributos de habilidad reúne la mayor cantidad de oraciones.

Por su parte, en la Tabla 9.5 para cada experiencia de aprendizaje colaborativo analizada se detalla: cantidad de archivos de texto plano, el tamaño total de los archivos, tamaño medio de los archivos, y cantidad total de oraciones identificadas.

Experiencia de ACSC	Cantidad de Archivos	Tamaño Total de Archivos (Bytes)	Tamaño Medio de Archivos (Bytes)	Total de Oraciones
Seminario de formación para los docentes	138	16942	122,768	161
Curso Preliminar de Fundamentos en Ciencias de la Computación	167	13060	78,203	171
Curso Propedéutico de Matemáticas	329	33560	102,006	398
Metodología de la investigación II	16	19510	1219,375	140
Estadística	5	2032	406,5	21
Totales	655	85104	129,929	891

Tabla 9.5. Resumen de archivos de texto plano

IX.2.3 Explorar los Datos

La Tabla 9.6 muestra el número de oraciones en las que se identificó cada atributo de habilidad, la cantidad de palabras vinculadas a cada atributo de habilidad, y la cantidad media de palabras en las oraciones donde se identificó cada atributo de habilidad. Teniendo en cuenta este último dato se puede ver que la situación es bastante dispar para los distintos atributos de habilidad ya que existen atributos de habilidad para los cuales las oraciones asociadas tienen una longitud media de 16 palabras mientras que para otros atributos de habilidad la cantidad media de palabras es apenas de 3 o 4.

Atributo de Habilidad	Cantidad de Oraciones	Cantidad Total de Palabras	Cantidad Media de Palabras por Oración
AHFDNMC	4	65	16,25
AHMCA	3	48	16
AHACDB	3	46	15,3333
AHDCG	15	201	13,4
AHDFI	36	480	13,33
AHVE	23	300	13,0434
AHPTMI	16	206	12,875
AHTD	5	64	12,8
AHICPR	3	38	12,66
AHPISDD	12	135	11,25
AHESS	10	106	10,6
AHTNP	6	63	10,5
AHFCP	3	31	10,3333
OAH	114	1173	10,2894
AHERM	5	51	10,2
AHURH	18	183	10,1666
AHUTEI	65	617	9,4923
AHPC	82	693	8,4512
AHSCTT	23	194	8,4347
AHACI	13	105	8,0769
AHFE	114	848	7,4385
AHMR	1	7	7
AHEOD	3	19	6,3333
AHLC	3	18	6
AHUEA	22	126	5,7272
AHMC	3	15	5
AHEESS	1	5	5
AHPPAA	79	394	4,9873
AHTM	10	42	4,2
AHDB	83	335	4,0361
AHMCDB	110	435	3,9545
AHTB	2	6	3
AHMD	1	3	3
AHCTD	0	0	0
AHRP	0	0	0
AHPR	0	0	0
AHPCMR	0	0	0
AHPA	0	0	0
AHTBR	0	0	0
AHACC	0	0	0
AHACIN	0	0	0
AHACD	0	0	0
AHEPC	0	0	0
Totales	891	7052	7,9147

Tabla 9.6. Cantidad total y cantidad media de palabras por atributo de habilidad

En la tabla anterior la columna *Cantidad de Oraciones* es igual a cero para los atributos de habilidad de: AHCTD, AHRP, AHPR, AHPCMR, AHPA, AHTBR, AHACC, AHACIN, AHACD, y AHEPC, por cuanto como ya se dijo en el Capítulo VII, no fue posible reconocer instancias de las mismos en los diálogos analizados. Como consecuencia, con el conjunto de mensajes de los e-tutores con los que se está trabajando, no se podrán construir clasificadores automáticos para estos diez atributos de habilidad. Esto es así porque no se dispone de ejemplos que puedan ser utilizados por los algoritmos de minería de textos para aprender las características que deben reunir las contribuciones de los e-tutores para ser clasificados como manifestando alguno de los diez atributos mencionados.

De los valores mostrados en Tabla 9.6 también queda en evidencia que algunos atributos de habilidad (AHFE, AHMCDB y OAH) fueron identificados en más de 100 oraciones, mientras que para otros (AHMR, AHERM, AHICPR, AHFCP, AHTNP, AHTB, AHTD, AHACDB, AHFDNMC, AHMD, AHLC, AHEESS, AHMCA, AHMC, AHEOD) la cantidad de oraciones no llega a 10. Esto significa que los algoritmos de construcción de clasificadores difícilmente puedan inferir las características que las contribuciones docentes deben reunir para ser catalogados bajo alguna de las categorías para las que se dispone de menos de 10 contribuciones.

Considerando las palabras que aparecen al menos 5 veces en el conjunto de datos analizado (excepto las *stop words*) y el atributo de habilidad al que se vinculan estas palabras, se obtuvieron los valores de frecuencia que se muestran en la Tabla 9.7 y en la Figura 9.3.

Palabra	Frecuencia	Atributo
Hola	75	AHDB
Saludos	19	AHMCDB
Ok	18	AHPPAA
Gracias	15	AHMCDB
Así	13	AHPPAA
Luego	13	AHMCDB
Foro	10	AHUTEI
Susana	10	OAH
Cuál	9	AHFE
Entiendo	9	AHFE
Foro	9	AHPC
Ok	9	AHMCDB
Rosamari	9	AHDB
Herrera	8	OAH
Correcto	6	AHPPAA

Tabla 9.7. Palabras frecuentes y atributos de habilidad asociados

Palabra	Frecuencia	Atributo
Hugo	6	OAH
Investigación	6	OAH
Metodología	6	OAH
Pregunta	6	AHUTEI
Vemos	6	AHUEA
Decir	5	AHFE
Equipo	5	AHDFI
Función	5	AHFE
Grupo	5	AHDFI
Grupo	5	AHPC
Guadalupe	5	AHPPAA
José	5	OAH
Metodología	5	AHUTEI
Ok	5	AHPC
Preguntas	5	AHURH
Quieres	5	AHFE
Seguimos	5	AHMCDB
Sergio	5	AHPPAA
X	5	AHFE
Z	5	AHFE

Tabla 9.7 (Cont.)

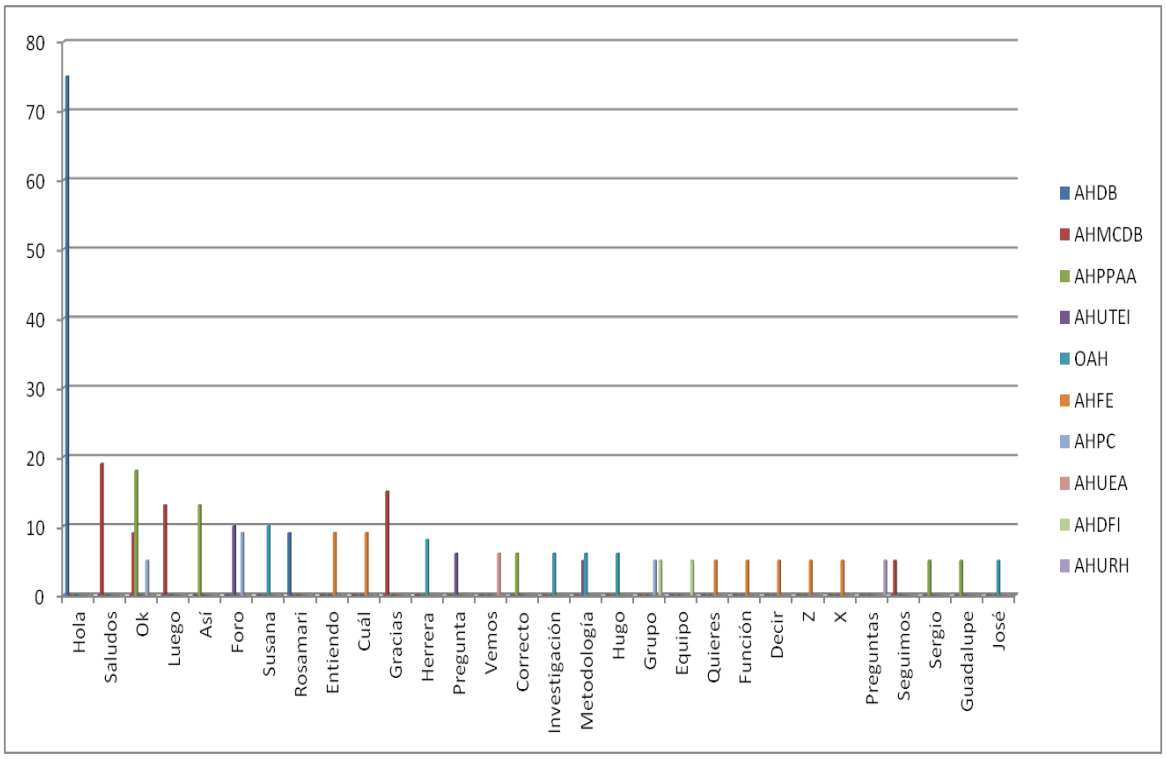


Figura 9.3. Incidencia de las palabras frecuentes en atributos de habilidad

La Figura 9.3 revela que casi todas las palabras de uso frecuente están asociadas a un solo atributo de habilidad ya que tan solo a cuatro (“Ok,” Metodología”, “Foro”, y “Grupo”) les corresponde más de una columna en los segmentos del eje horizontal

destinados a las palabras. Así por ejemplo, como se ve en la Tabla 9.7, a la palabra "Foro" se la puede encontrar en contribuciones docentes catalogadas como AHUTEI pero también en contribuciones catalogadas como AHPC. Esto podría indicar que para construir los clasificadores no son necesarias operaciones de pre procesamiento complicadas (tales como parsing o POST) ya que las palabras dan buenos indicios sobre la categoría de habilidad que le corresponde a cada contribución de los docentes.

Entre las palabras que ocurren con frecuencia en el conjunto de datos analizados se tiene "Susana" y "Herrera". Las mismas corresponden al nombre y apellido de uno de los e-tutores que participó en las experiencias de ACSC analizadas. El mismo recurrentemente escribía su nombre y/o apellido al final de sus contribuciones en el foro a pesar de que la herramienta de software usada (Moodle) colocaba automáticamente el nombre del autor como encabezado de cada mensaje. Se considera apropiado eliminar de las contribuciones docentes las oraciones formadas solo por el nombre y/o apellido del docente que lo público porque el nombre y apellido seguramente van a cambiar de un curso a otro. Por lo tanto, al procesar nuevas interacciones de ACSC el nombre y apellido analizado van a ser totalmente distintos a los que se tuvo en cuenta en la fase de construcción de los clasificadores. Esto significa que a menos que el nombre y/o apellido estén acompañados de otras palabras, no serán características de gran ayuda para construir clasificadores que automáticamente asignen atributos de habilidad a contribuciones docentes provenientes de nuevas experiencias de ACSC.

Una situación análoga al del nombre de los docentes se da con el nombre de los estudiantes y de lugares. Los nombres no son características lingüísticas de relevancia para construir los modelos porque difícilmente se mantengan al analizar un nuevo conjunto de interacciones de ACSC. Por consiguiente, se tienen que eliminar los mensajes formados solo por nombres de estudiantes y de lugares.

Por otro lado, al examinar los archivos en texto plano se pudo descubrir que si un archivo origen (pdf, htm, rtf, y doc) contenía un emoticon representado mediante una imagen, el emoticon desaparecía en el archivo de texto plano.

IX.2.4 Verificar la Calidad de los Datos

Como ya se estableció en la sección anterior, en el conjunto de contribuciones docentes con las que se cuenta para este proyecto no existen mensajes de los e-tutores a los cuales se les haya hecho corresponder los atributos de habilidad de: AHCTD, AHRP, AHPR, AHPCMR, AHPA, AHTBR, AHACC, AHACIN, AHACD, y AHEPC. Por

consiguiente, el modelo que se obtenga al finalizar el proyecto no permitirá identificar automáticamente aquellos atributos de habilidad de los que no se dispone de ejemplos que pueden usarse para la fase de entrenamiento. Esto significa que si se aplican los clasificadores a un nuevo conjunto de contribuciones docentes no se podrían identificar los diez atributos de habilidad antes listados aun cuando estuvieran presentes. Una cuestión que no se debería perder de vista es que los atributos de habilidad a los que recurren los docentes de ACSC dependen de la situación que les toca enfrentar, por lo que en una clase o curso no es obligación que manifiesten todos los atributos de habilidad de la clasificación de habilidades descripta en el Capítulo IV. Por consiguiente, la ausencia de algunos atributos de habilidad en el conjunto de datos usado en este proyecto no debe considerarse como un problema de discrepancia entre lo que sucede en la realidad y la información recabada sobre esta realidad.

Los atributos de habilidad que los expertos identificaron en el análisis de contenido no están representados de manera equilibrada en el conjunto de datos porque por un lado los atributos de habilidad OAH, AHMCDB, y AHFE aparecen más de 100 veces, pero por el otro existen 15 atributos de habilidad (AHMR, AHERM, AHICPR, AHFCP, AHTNP, AHTB, AHTD, AHACDB, AHFDNMC, AHMD, AHLC, AHEESS, AHMCA, AHMC, AHEOD) cuya cantidad de apariciones no llega a 10.

Una posible solución a las dos situaciones mencionadas consistiría en obtener un conjunto de mensajes de e-tutores donde se manifiesten todos los atributos de habilidad de la clasificación de habilidades utilizada y además los atributos estén representados en una cantidad más o menos equilibrada. Para el caso de los atributos con escasa cantidad de oraciones se podría también optar por descártalos de modo tal que la distribución desequilibrada de atributos de habilidad no afecte la efectividad del reconocimiento automático.

En los textos analizados las palabras aparecen con la letra inicial en mayúscula, con todas sus letras en minúscula, o con todas sus letras en mayúscula. Sin embargo, no debe tomarse a esto como un error porque al comienzo de una oración la letra inicial suele escribirse en mayúscula y en los entornos online el texto escrito con todas las letras en mayúscula corresponde a un grito. Por consiguiente no es necesario convertir todas las palabras a minúsculas o mayúsculas.

En cuanto a la ortografía, se tienen que corregir los errores de acentuación, de omisión de letras, de inserción de letras, y de unión de palabras. Las palabras con errores ortográficos corresponden apenas al 2,9% de las palabras que forman el conjunto de

mensajes de los e-tutores, pero es preciso que se hagan las correcciones necesarias. De esta manera se evitaría que, por ejemplo, "resumir" y "resiumir" se tomen como dos palabras diferentes debido a que en la segunda se agregó una letra "i" accidentalmente.

Tal como se mencionó al finalizar el punto IX.2.3, los archivos en texto plano no contienen los emoticones representados por imágenes en los archivos a partir de los cuales fueron obtenidos. Como consecuencia, lo que corresponde es insertar en los archivos de texto plano el o los caracteres Unicode equivalentes al emoticon representado mediante una imagen. Así entonces, se tendría que reemplazar  por ":)", por ":(", por ";(", etc.

Al final de algunos de los mensajes de los e-tutores pertenecientes al Curso Propedéutico de Matemáticas aparecían cadenas de caracteres como "[SIO 1]", "[SIAI 1]", "[SIAI 3]", etc. Dado que estas cadenas de caracteres no forman parte del mensaje del e-tutor, ni tienen un significado definido con anterioridad, se procederá a suprimirlas de los archivos en texto plano.

Una última cuestión a considerar es la existencia de archivos duplicados, esto es archivos que coinciden tanto en su contenido textual como en los atributos de habilidad asignados a cada oración que los componen. Esta situación surge porque en una misma sesión de chat los e-tutores publican más de una vez el mismo mensaje o bien cuando publican el mismo mensaje en el foro de cada uno de los grupos de estudiantes.

IX.3 FASE 3:PREPARACIÓN DE LOS DATOS

IX.3.1 Seleccionar los Datos

Los archivos obtenidos en las cinco experiencias de aprendizaje colaborativo contienen una gran cantidad de contribuciones o mensajes publicados por los e-tutores mediante la herramienta de foro o chat, pero para este proyecto son de relevancia solamente los mensajes donde expertos (mediante la técnica de análisis de contenido) identificaron atributos de habilidad social porque el objetivo es reconocer automáticamente dichas habilidades. Teniendo en cuenta que los clasificadores se construyen mediante un proceso inductivo que observando instancias de cada categoría determina las características que las sentencias deberían reunir para ser asignados a tal o cual categoría, es poco probable que con 1 o 2 ejemplos se pueda inferir los rasgos que distinguen a cada uno de los atributos de habilidad. Por lo expuesto, se decidió excluir aquellas oraciones cuyos atributos de habilidad asociados tengan una frecuencia de ocurrencia inferior a 5. En este sentido, se podría haber sido más restrictivo porque hay que recordar que en el apartado

III.6 se mencionó que para problemas de clasificación lo recomendable es contar con al menos 30 instancias de cada categoría con la que se trabaja.

No se debe perder de vista que con este criterio de exclusión adaptado se reduce a 21 la cantidad de atributos de habilidad usados en el presente proyecto. Es decir, que solo se podrán construir clasificadores automáticos para 21 de los 42 atributos de habilidad social que forman parte de la taxonomía adoptada.

IX.3.2 Limpiar los Datos

Como parte de la limpieza de los datos se procedió tanto a modificar los archivos de texto plano para solucionar los problemas de calidad de datos mencionados en la sección IX.2.4, así como también a excluir los archivos que no son de utilidad para la construcción de los clasificadores.

IX.3.2.1 Modificación de archivos

Como primera medida se modificaron las palabras que presentaban errores de acentuación, de omisión de letras y de inserción de letras. En segunda instancia, se separaron aquellas palabras que se habían escrito como si fueran una sola.

Para resolver el problema de los emoticones lo que se hizo fue insertar en los archivos de texto plano los caracteres Unicode equivalentes al emoticon representado mediante una imagen.

En los archivos correspondientes al Curso Propedéutico de Matemáticas se eliminaron los caracteres incluidos al final de algunos mensajes ya que al no formar parte de los mismos no hacían otra cosa más que contaminar el conjunto de datos. También se suprimieron las oraciones formadas solo por el nombre y/o apellido del e-tutor que público el mensaje.

La última tarea de edición del contenido de los archivos consistió en eliminar las oraciones a las que se les había asociado alguno de los atributos de habilidad que tenían menos de 5 instancias en el conjunto de datos.

IX.3.2.2 Exclusión de archivos

Para evitar la existencia de archivos duplicados se eliminaron los archivos que contenían exactamente el mismo texto y cuyas oraciones tenían asignados los mismos atributos de habilidad. También fueron eliminados los archivos que contenían solo nombres de lugares o de personas, y los archivos constituidos únicamente por sentencias

asociadas con atributos de habilidad cuya frecuencia era inferior a 5. Los archivos de texto plano descartados se colocaron en el directorio *Mineria_de_Textos\Archivos_de_Texto_Plano\Archivos_Descartados* del CD que acompaña a este documento de modo tal de que, si fuera necesario, puedan ser utilizados en estudios futuros. En la Tabla 9.8 se muestra la cantidad final de archivos para cada experiencia de ACSC luego de efectuada la supresión descripta en este apartado.

Experiencia	**Cantidad original de Archivos**	**Cantidad de Archivos descartados**	**Cantidad resultante de Archivos**	**Porcentaje de Archivos descartados**
Curso Propedéutico de Matemáticas	329	40	289	12,15
Curso Preliminar de Fundamentos en Ciencias de la Computación	167	16	151	9,58
Metodología de la investigación II	16	1	15	6,25
Seminario de formación para los docentes	138	5	133	3,62
Estadística	5	0	5	0
Totales	655	62	593	9,46

Tabla 9.8. Resumen de la supresión de archivos en cada experiencia de ACSC

IX.3.3 Construir los Datos

Las operaciones de construcción de datos permiten obtener los datos sobre los que se aplicará los algoritmos de clasificación pero en este proyecto no generaron nuevos atributos ni registros como sucede con la minería de datos. En su lugar, las operaciones de construcción seleccionadas permiten generar anotaciones que contienen información previamente no conocida sobre los mensajes de los e-tutores. Concretamente las operaciones de construcción de datos usadas son: identificación de token, identificación de sentencias, normalización, etiquetado de parte del discurso (POST), y supresión de *stop words*.

IX.3.3.1 Identificación de *token*

El conocer los diferentes tipos de *token* contenidos en los mensajes de los docentes es un paso previo necesario para identificar las oraciones que forman parte de los mismos.

Para dividir los mensajes de los e-tutores en tipos de *token* se uso el recurso de procesamiento *GATE Unicode Tokeniser*.

IX.3.3.2 Identificación de sentencias

Teniendo en cuenta que en cada oración de los mensajes de los e-tutores se manifiesta un atributo de habilidad, como paso previo a la construcción del modelo, es necesario identificar las oraciones que forman parte de cada mensaje publicado por los e-tutores. La descomposición en sentencias de las contribuciones docentes se puede realizar de manera automática por medio de las adaptaciones efectuadas al recurso de procesamiento *ANNIE Sentence Splitter* que se comentaron en el capítulo VIII.

IX.3.3.3 Normalización

GATE cuenta con los recursos de procesamiento *Tree Tagger* y *Snowball Based Stemmers* para reducir el número de los distintos tipos de *token* y simultáneamente incrementar la frecuencia de ocurrencia de algunos tipos de *token* individuales. Con *Tree Tagger* no solo se obtienen las etiquetas POST correspondiente a cada *token* sino que también se realiza lematización. Por su parte, *Snowball Based Stemmers* implementa la operación de *stemming to a root* mediante la supresión de los sufijos de los *tokens*. La Tabla 9.9 contiene una muestra de las características de anotación "lemma" y "stem" generadas por *Tree Tagger* y *Snowball Based Stemmers* respectivamente.

Token Original	**Resultado de *Tree Tagger* (*lemma*)**	**Resultado de *Snowball Based Stemmers* (*stem*)**
chica	chico	chic
chico	chico	chic
chicas	chico	chic
chicago	chicago	chicag
chicles	chicle	chicl
cheque	cheque	chequ
chequeo	chequear	cheque
tormenta	tormenta	torment
tornado	tornar	torn
tornar	tornar	torn
toronto	toronto	toront
torpeza	torpeza	torpez

Tabla 9.9. Comparación de resultados de lematización y *stemming to a root.*

IX.3.3.4 Etiquetado de parte del discurso

Se optó por aplicar esta operación de modo tal de determinar si la inclusión de información sobre el contenido semántico de los *tokens* permite mejorar la efectividad de los clasificadores en la identificación automática de habilidades sociales. Para identificar la etiqueta POST correspondiente a cada *token* se recurrió a:

- *Generic Tagger*: la gramática JAPE que debe aplicarse a cada texto antes de ejecutar *Tree Tagger* se incluye en el Anexo B. La gramática crea la característica de anotación "string" en cada anotación del tipo "Sentence" creada por el recurso de procesamiento *ANNIE Sentence Splitter*. El valor de "string" es el contenido textual de cada oración identificada por *ANNIE Sentence Splitter*. Las reglas para esta gramática (denominada *Sentence_String*) no fueron creadas durante el desarrollo del trabajo final ya que se las puede descargar desde la página oficial de GATE.
- *Tree Tagger*: este es el recurso usado por *Generic Tagger* para poder asignar las etiquetas POST porque permite procesar textos escritos en español. Puesto que se decidió que *Tree Tagger* reconozca los *tokens* por su cuenta, las anotaciones del tipo "Token" generadas por *GATE Unicode Tokeniser* son omitidas.

IX.3.3.5 Supresión de *stop words*

Es una operación que no genera nuevos campos, registros o anotaciones sino que por el contrario suprime de un texto aquellas palabras que coincidan con alguna de las *stop words* conocidas para el lenguaje español. En este proyecto se trabaja con la lista de *stop words* incluidas en el paquete tm (ver punto IX.1.4.2.2). La eliminación de *stop words* no tiene que hacerse manualmente sino que puede automatizarse con los recursos de procesamiento *Jape Transducer* y *ANNIE Gazetteer* disponibles en GATE. En el Anexo B se incluyen la lista de *stop words* así como la gramática usada por *Jape Transducer*. La gramática de *Jape Transducer* consiste de las siguientes 2 reglas:

- stopWordsSinAcento: suprime todas las anotaciones del tipo token cuya característica "string" coincide con una *stop word* incluida en la lista de *ANNIE Gazetteer*.
- stopWordsAcentuadas: suprime todas las anotaciones del tipo token cuya característica "string" coincide con una de las *stop words* acentuadas incluidas en la regla.

Es preciso distinguir entre *stop words* acentuadas y no acentuadas puesto que en la lista *gazetteer* usada por *ANNIE Gazetteer* no pueden incluirse caracteres acentuados.

IX.3.3.6 Combinación de operaciones de construcción de datos

Las operaciones de construcción de datos o pre procesamiento (como se las conoce en el campo del *text mining*) no se pueden aplicar sobre los archivos de texto plano porque los mismos no pueden almacenar las anotaciones que se van a generar, y por ello es preciso crear un corpus en GATE como se explicará en la sección 3.4. De las cinco operaciones mencionadas en los apartados previos solo las dos primeras son requeridas antes de construir los clasificadores automáticos. Las últimas tres (normalización, etiquetado de parte del discurso, y supresión de *stop words*) se combinan con la identificación de *token*, el reconocimiento de oraciones, y entre ellas, durante la etapa de modelado de manera tal de comprobar si se obtienen diferencias en la performance de los clasificadores construidos. Lo anterior da lugar a las combinaciones de operaciones de pre procesamiento mostradas en las Figuras 9.4 a 9.11.

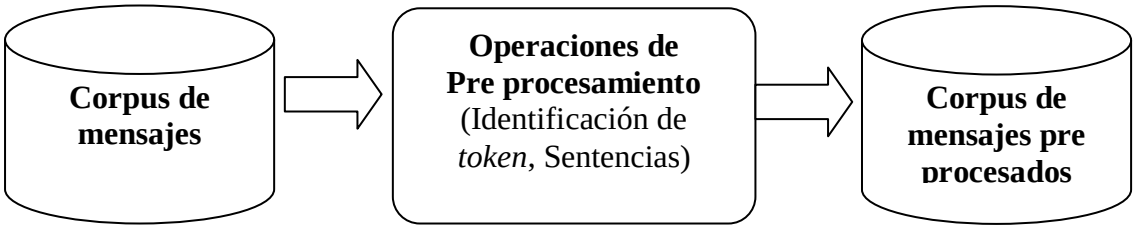


Figura 9.4. Primera combinación de operaciones de pre procesamiento

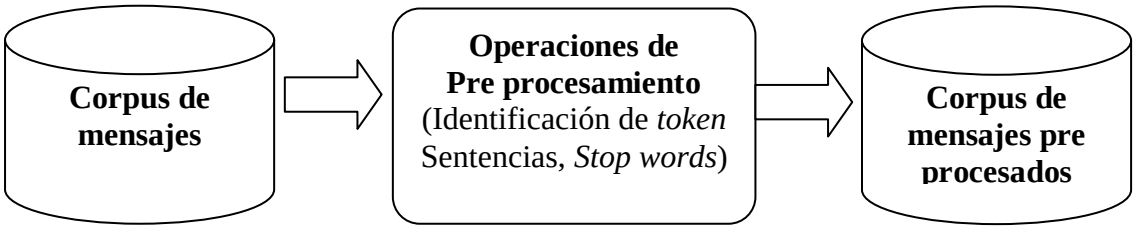


Figura 9.5. Segunda combinación de operaciones de pre procesamiento

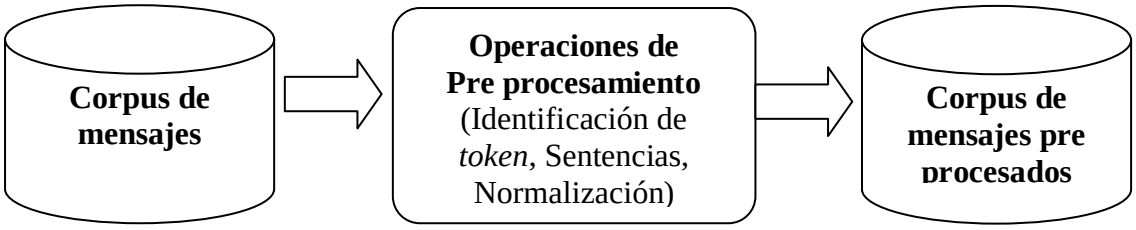


Figura 9.6. Tercera combinación de operaciones de pre procesamiento

Para llevar a cabo la normalización se cuenta con los recursos de procesamiento *Tree Tagger* y *Snowball Based Stemmers* por lo cual en este paso se tendría dos sub combinaciones. Por un lado, identificación de *token*, reconocimiento de sentencias, y lematización; por el otro, identificación de *token*, reconocimiento de sentencias, y *stemming to a root*.



Figura 9.7. Cuarta combinación de operaciones de pre procesamiento

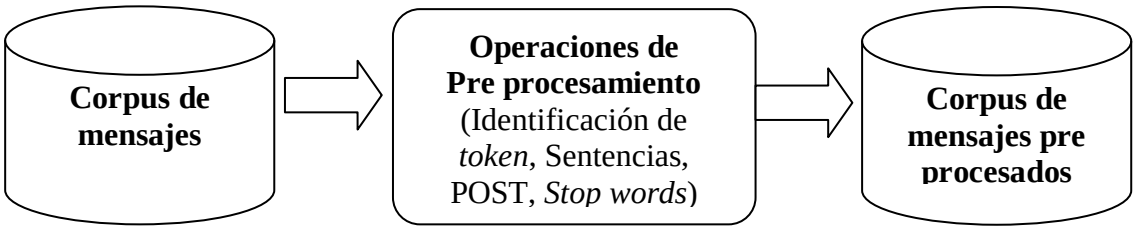


Figura 9.8. Quinta combinación de operaciones de pre procesamiento

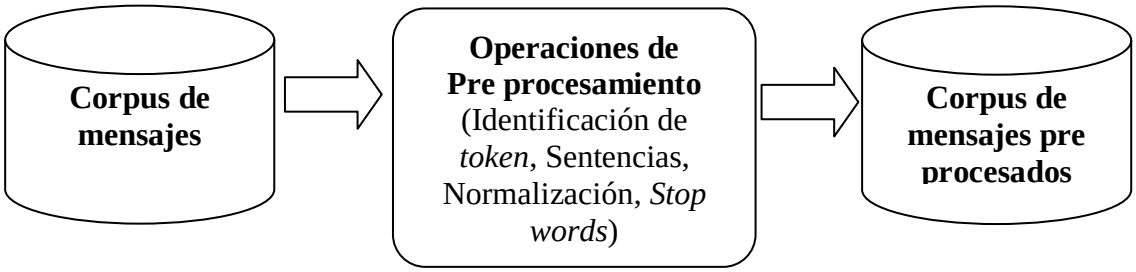


Figura 9.9. Sexta combinación de operaciones de pre procesamiento

De manera análoga a la tercera combinación de operaciones de pre procesamiento, si se utiliza *Tree Tagger* se tendrá la sub combinación identificación de *token*, identificación de sentencias, lematización, y *supresión de stop words*; en cambio si se recurre a *Snowball Based Stemmers* se tendrá la secuencia identificación de *token*, reconocimiento de sentencias, *stemming to a root*, y eliminación de *stop words*.

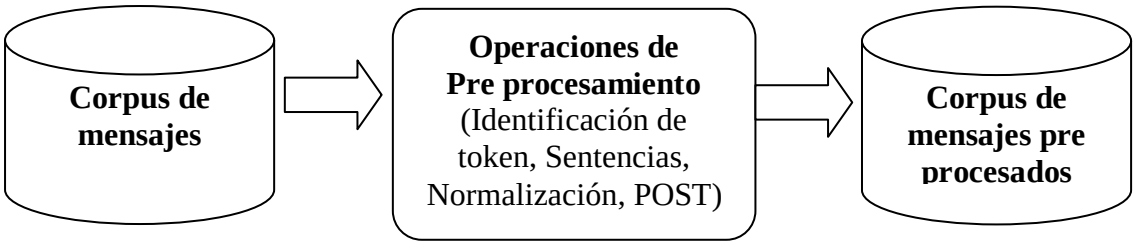


Figura 9.10. Séptima combinación de operaciones de pre procesamiento

Debido a que *Generic Tagger* y *Tree Tagger* difieren en cuanto a sus estrategias de normalización, para la séptima combinación de operaciones de pre procesamiento se tienen dos opciones: en primer lugar identificación de *token*, reconocimiento de sentencias, lematización, y POST; y en segundo lugar identificación de *token*, reconocimiento de sentencias, *stemming to a root*, y POST.

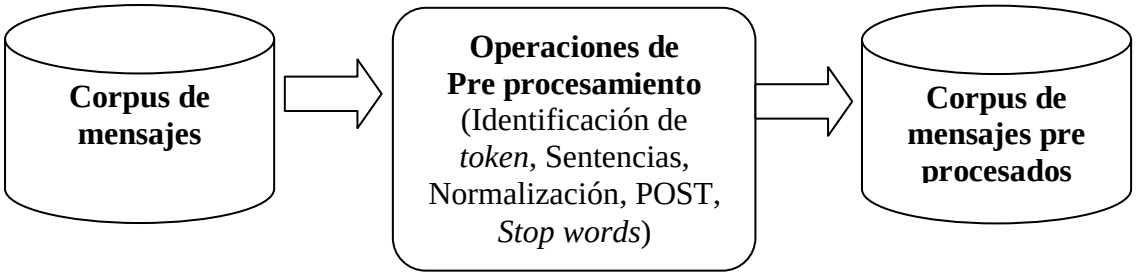


Figura 9.11. Octava combinación de operaciones de pre procesamiento

Tal como sucede con las combinaciones 3,6 y 7 de operaciones de preparación de datos, en la combinación 8 se tiene dos sub combinaciones:

- Identificación de *token*, identificación de sentencias, lematización, POST, y supresión de *stop words*.
- Identificación de *token*, identificación de sentencias, *stemming to a root*, POST, y supresión de *stop words*.

IX.3.4 Integrar los Datos

En el contexto de este trabajo de investigación la integración de datos no se aplica a las tablas de una base de datos sino que consiste en integrar los archivos de texto plano que contienen los mensajes publicados por los e-tutores. Para ello se creó un corpus denominado *corpusMensajes* que contiene todos los archivos de texto plano relevantes para la construcción de los clasificadores automáticos. Para que el corpus se almacene de manera persistente se creó la base *BDmensajes* que es del tipo *SerialDataStore*.

En la parte izquierda de la Figura 9.12 aparecen el corpus *corpusMensajes* y la base de datos *BDmensajes* donde se almacena el corpus. En tanto que en la parte derecha debajo de *GATE Document* se listan los archivos que componen *corpusMensajes*. Este corpus contiene los 593 archivos que quedaron luego de aplicar la operación de exclusión de archivos descripta en el punto IX.3.2.2.

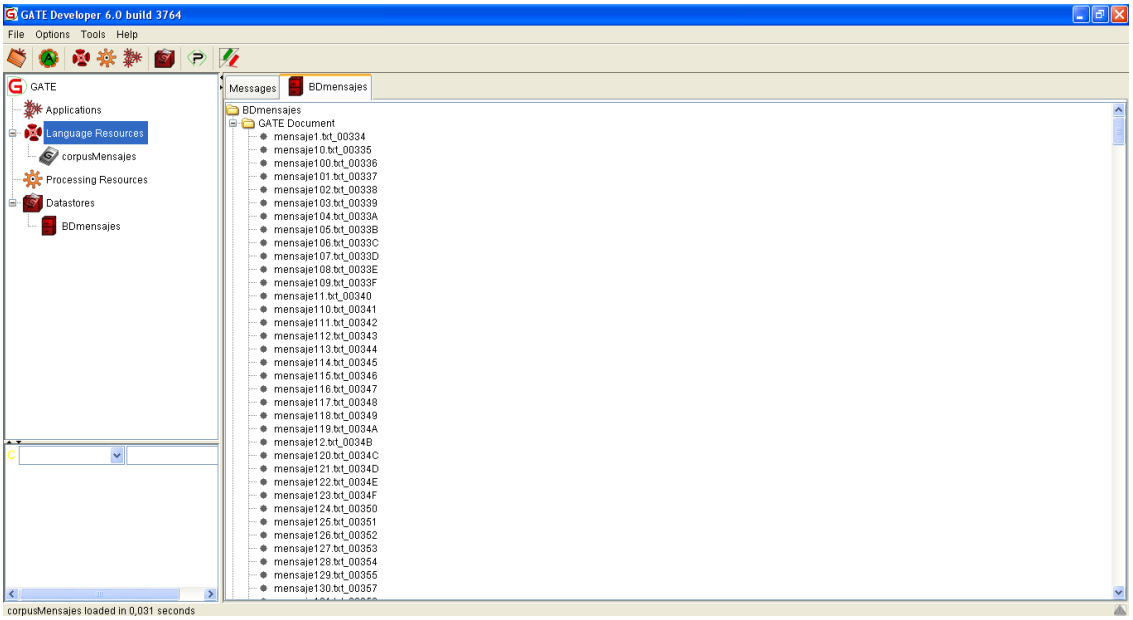


Figura 9.12. Corpus de mensajes, base de datos, y mensajes del Corpus

IX.3.5 Formatear los Datos

No es necesario hacer modificaciones sintácticas al contenido textual de los mensajes de los e-tutores pero para GATE las oraciones estarán representadas mediante uni-gramas dado que el objetivo es que las oraciones de las contribuciones docentes se clasifiquen en función a las características lingüísticas de las anotaciones del tipo "Token" que las componen. Los n-gramas son un recurso de GATE que permite caracterizar a una anotación (en este caso del tipo "Sentence") en términos de las características lingüísticas de las anotaciones (en este caso del tipo "Token") que la componen. Las características lingüísticas de las anotaciones del tipo "Token" usadas para la construcción y evaluación de los clasificadores automáticos cambian de acuerdo a la combinación de operaciones de pre procesamiento aplicadas por lo que se las detallan a continuación:

- Primera combinación: de cada anotación del tipo "Token" se consideran las características "string" y "kind".

- Segunda combinación: de cada anotación del tipo "Token*"* se consideran las características "string" y "kind".
- Tercera combinación: de cada anotación del tipo "Token" se considera la característica "lemma" para el caso de la lematización. En tanto que para el *stemming to a root* se tienen en cuenta las características "stem" y "kind".
- Cuarta combinación: de cada anotación del tipo "Token" se tienen en cuenta las características "string" y "category".
- Quinta combinación: de cada anotación del tipo "Token" se tienen en cuenta las características "string" y "category".
- Sexta combinación: de cada anotación del tipo "Token" se considera la característica "lemma" para el caso de la lematización. En tanto que para el *stemming to a root* se usan las características "stem" y "kind".
- Séptima combinación: de cada anotación del tipo "Token" se utilizan las características "lemma" y "category" para el caso de la lematización. En tanto que con el *stemming to a root* se usan las características "stem" y "category".
- Octava combinación: de cada anotación del tipo "Token" se consideran las características "lemma" y "category" cuando se recurre a la lematización. En tanto que para el *stemming to a root* se tienen en cuenta las características "stem" y "category".

Para indicar la importancia o relevancia de cada anotación del tipo "Token" que forma parte de un uni-grama se optó por usar tf-idf, es decir la inversa de la frecuencia de ocurrencia de la anotación "Token" en el documento de texto.

Para tareas de clasificación es necesario conocer la clase o categoría de cada instancia o ejemplo usado para la construcción y evaluación de los clasificadores automáticos. Teniendo en cuenta que se busca clasificar son las oraciones que forman parte de los mensajes de los e-tutores, es necesario que a cada anotación del tipo "Sentence" (creada mediante *ANNIE Sentence Splitter*) se le asocie el atributo de habilidad que le corresponde. Por ello, para cada anotación del tipo "Sentence" se debe crear y completar de manera manual la característica de anotación "atributo_hab". Esta característica de anotación se corresponde con la segunda columna de la Tabla 9.4 y sus valores son las abreviaturas de los atributos de habilidad.

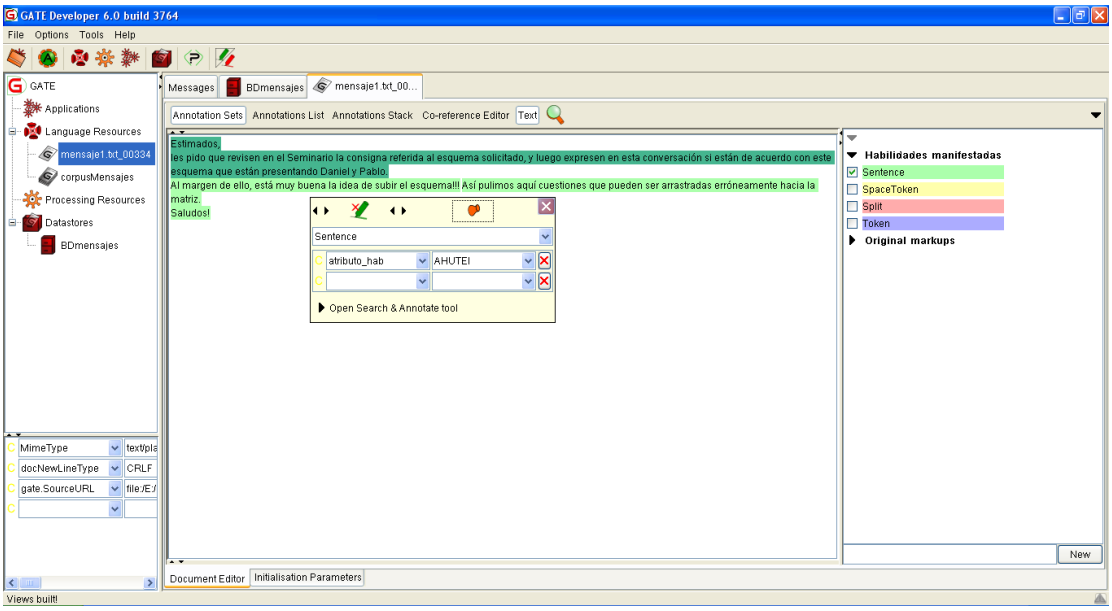


Figura 9.13. Ejemplo de anotación “Sentence” con la característica “atributo_hab”

IX.4 FASE 4: MODELADO

IX.4.1 Seleccionar la Técnica de Modelado

Considerando que el objetivo es identificar de manera automática las habilidades sociales manifestadas por los e-tutores en los mensajes que publican mediante herramientas de ACSC, es evidente que se tiene entre manos un problema de clasificación. Cuando se recurre a la aproximación de aprendizaje de máquina para resolver un problema de clasificación se asume que existe un conjunto de objetos pertenecientes a diferentes clases, que los objetos se pueden caracterizar con diversos atributos, y que se conoce de antemano la categoría de cada uno de los objetos. En GATE se pueden utilizar los algoritmos KNN, SVM, PAUM, Näive Bayes y C4.5 para construir un clasificador que permita determinar la categoría correspondiente a objetos desconocidos.

IX.4.1.1 K-esimo vecino más cercano (KNN)

Este método no construye representaciones declarativas explicitas de las diversas clases o categorías sino que se basa en el cálculo de la similitud entre el texto a ser clasificado y los textos del conjunto de entrenamiento. En el presente proyecto la tarea consiste en clasificar automáticamente cada anotación del tipo “Sentence” contenida en una contribución de un e-tutor, por lo tanto no se debe perder de vista que en GATE dos anotaciones del tipo “Sentence” se consideran iguales si cubren el mismo fragmento de texto y el valor de cada una de sus características coinciden.

IX.4.1.2 Máquina de vector de soporte (SVM)

Para construir un clasificador mediante el algoritmo SVM es preciso que el problema sea linealmente separable, es decir que pueda encontrarse un hiperplano que separe las instancias negativas de las positivas. Para el caso de los problemas binarios (cuando se tienen dos clases o categorías) lo anterior significa que existe una recta que separa los puntos que representan instancias positivas de la clase de los puntos que representan instancias negativas. Puesto que para los 21 atributos de habilidad usados en este estudio la cantidad de ejemplos positivos es bastante pequeña en relación a la cantidad de ejemplos negativos, se decidió recurrir a la versión de SVM con márgenes desiguales. Esta variante del algoritmo SVM fue introducida por Li y Shawe-taylor (2003) dado que el algoritmo SVM estándar no tiene buen desempeño en conjuntos de datos de entrenamiento desbalanceadas con solo una pequeña fracción de ejemplos positivos. En estos casos suele ser beneficioso que el algoritmo de aprendizaje establezca el margen con respecto a los ejemplos positivos un poco más grande que el margen con respecto a los ejemplos negativos. La relación entre el margen y el numero de ejemplos de entrenamiento positivos se la expresa mediante el parámetro τ. A un número pequeño de ejemplos positivos le corresponde un valor pequeño del parámetro τ.

IX.4.1.3 Perceptrón con márgenes desiguales (PAUM)

El perceptrón es una red neuronal simple que clasifica sus entradas en una de dos categorías posibles. Por lo tanto, los clasificadores construidos con el algoritmo perceptrón son aplicables solamente a problemas linealmente separables. El algoritmo PAUM es una extensión del perceptrón especialmente diseñado para tratar con problemas de dos clases donde los ejemplos positivos son muy raros en relación a los ejemplos negativos (Li *et al.*, 2002). Basándose en la idea de que en problemas de clasificación es más importante clasificar correctamente ejemplos positivos que clasificar correctamente los ejemplos negativos, en PAUM los ejemplos positivos y negativos son tratados de manera diferente. Este trato diferencial se logra mediante la incorporación de un margen positivo (τ_{+1}) y un margen negativo (τ_{-1}).

IX.4.1.4 Naïve bayes

En un clasificador construido mediante Naïve Bayes, para determinar el atributo de habilidad que le corresponde a una anotación del tipo “Sentence”, se tiene en cuenta la probabilidad de ocurrencia de cada atributo de habilidad dada la presencia o ausencia de

determinadas características. El carácter "Naïve" de los clasificadores se debe a que se asume que las características que representan a las oraciones son independientes entre sí, aunque en la práctica no se verifica este supuesto y frecuentemente es falso (Sebastiani, 2002).

IX.4.1.5 Algoritmo C4.5

C4.5 es una implementación de los clasificadores basados en árbol de decisión. Es decir que para asignar un atributo de habilidad a una anotación del tipo "Sentence" se comienza por un nodo raíz y se continúa de manera descendente por los arcos cuyas condiciones son satisfechas por la anotación "Sentence" hasta que se alcanza un nodo hoja. A la anotación del tipo "Sentence" se le asigna el atributo de habilidad que corresponde al nodo hoja. No existen supuestos sobre la calidad, formato o distribución de los datos que deban cumplirse para construir un clasificador con C4.5

IX.4.2 Generar el Diseño de la Prueba

Para evaluar la efectividad en la identificación automática de habilidades se calcularon las métricas de precisión, recall, y F_1 en primera instancia, con el método hold-out, y luego con el método de validación cruzada n-fold.

IX.4.2.1 Método hold-out

Con el método hold-out los documentos se dividen aleatoriamente en dos conjuntos independientes, uno de entrenamiento y uno de prueba. Al conjunto de entrenamiento se suele destinar 2/3 de los documentos en tanto que para el conjunto de prueba se destina el tercio restante. En ocasiones suelen ejecutarse varias iteraciones del método hold-out para luego promediar las métricas de efectividad de todas las corridas pero, en este proyecto se decidió usar una sola iteración porque de acuerdo con (Sebastiani, 2002) para comparar los resultados obtenidos con los diversos algoritmos es preciso que se cumplan las siguientes condiciones:

- Los experimentos deben realizarse en la misma colección de documentos, es decir con los mismos documentos y las mismas categorías.
- Los experimentos tienen que realizarse con la misma división entre conjuntos de entrenamiento y prueba.
- En los experimentos debe usarcé la misma métrica de evaluación, y si la métrica depende de algunos parámetros, con los mismos valores.

Con varias iteraciones del método hold-out los documentos de los conjuntos de prueba y de entrenamiento serían asignados de manera aleatoria por GATE en cada iteración por lo cual no se puede garantizar que cuando se aplique KNN, PAUM, SVM, Naïve Bayes, y C4.5 la composición de los conjuntos de prueba y entrenamiento sea idéntica.

Para establecer cómo dividir *corpusMensajes* en los dos subconjuntos se procedió a ejecutar el recurso de procesamiento *Batch Learning PR*, con el algoritmo KNN, especificando que el 66% de los documentos debían destinarse al conjunto de entrenamiento. De los 593 documentos pertenecientes a *corpusMensajes*, 391 se destinaron para el entrenamiento y 202 para la prueba. Paso seguido *corpusMensajes* se separó en *corpusEntrenamiento* y *corpusPrueba*. Para determinar los documentos que pertenecían a *corpusEntrenamiento* y *corpusPrueba* se tuvo que examinar la lista de documentos destinados al entrenamiento y la lista de documentos destinados a la prueba en el archivo *logFileForNLPLearning.save* generado al ejecutar *Batch Learning PR*.

IX.4.2.2 Método de validación cruzada n-fold

Si bien el método hold-out permite evaluar la efectividad de un clasificador automático, el mismo no garantiza que los atributos de habilidad estén representados en la misma proporción en los conjuntos de entrenamiento y de prueba. Así por ejemplo, si todas las oraciones de *corpusMensajes* que tuvieran asociadas el atributo de habilidad AHUTEI fueran a parar a *corpusPrueba*, no puede esperarse que un modelo aprenda a catalogar correctamente oraciones de este tipo. Una posible solución a este problema es el utilizar el método conocido como validación cruzada n-fold. Lo más común es usar una versión de n-fold conocida como 10-fold que consiste en repetir el procedimiento de evaluación 10 veces. En cada iteración 9/10 de los documentos se usan para entrenamiento y el restante 1/10 para prueba. Los valores de las métricas precisión, recall, y F_1 se obtienen promediando los resultados de las 10 repeticiones.

A diferencia de lo que sucede con el método hold-out, cuando en GATE se usa el método de validación cruzada no se necesita hacer la separación entre *corpusEntrenamiento* y *corpusPrueba* porque en cada iteración GATE determina automáticamente que documentos forman el conjunto prueba y cuales el conjunto de entrenamiento.

IX.4.3 Construir el Modelo

Se uso el recurso de procesamiento *Batch Learning PR* para construir los modelos con los algoritmos mencionados en el apartado de selección de la técnica de modelado. A continuación se describe de manera separada el proceso de construcción de los modelos para cada uno de los algoritmos y los rasgos de los modelos obtenidos.

IX.4.3.1 Construcción del modelo con KNN

Para poder construir un modelo mediante el recurso de procesamiento *Batch Learning PR* se debe especificar una serie de parámetros en su archivo de configuración y en tiempo de ejecución. Los elementos definidos en el archivo de configuración xml para el algoritmo KNN se incluyen en la Figura 9.14 en tanto que, el valor de los parámetros a especificar en tiempo de ejecución se muestra en la Tabla 9.10.

```
<?xml version="1.0"?>
<ML-CONFIG>
  <VERBOSITY level="1"/>
  <IS-LABEL-UPDATABLE value="true"/>
  <IS-NLPFEATURELIST-UPDATABLE value="true"/>

  <EVALUATION method="holdout" runs="1" ratio="0.66"/>
  <multiClassification2Binary method="one-vs-another"/>
  <PARAMETER name="thresholdProbabilityClassification" value=
  "0.5"/>

  <ENGINE nickname="KNN" implementationName="KNNWeka" options=
  "-k 5"/>

  <DATASET>
     <INSTANCE-TYPE>Sentence</INSTANCE-TYPE>
     <NGRAM>
        <NAME>Sentencegram</NAME>
        <NUMBER>1</NUMBER>
        <CONSNUM>2</CONSNUM>
        <CONS-1>
              <TYPE>Token</TYPE>
             <FEATURE>string</FEATURE>
         </CONS-1>
        <CONS-2>
              <TYPE>Token</TYPE>
             <FEATURE>kind</FEATURE>
          </CONS-2>
         <WEIGHT>1.0</WEIGHT>
         <ValueTypeNgram>3</ValueTypeNgram>
     </NGRAM>
```

Figura 9.14. Archivo de configuración xml para el algoritmo KNN

```
      <ATTRIBUTE>
            <NAME>habilidad_etutor</NAME>
            <SEMTYPE>NOMINAL</SEMTYPE>
            <TYPE>Sentence</TYPE>
            <FEATURE>atributo_hab</FEATURE>
            <POSITION>0</POSITION>
            <CLASS/>
       </ATTRIBUTE>
    </DATASET>
</ML-CONFIG>
```

Figura 9.14. (Cont.)

Esta configuración es válida para la primera y segunda combinación de operaciones de preparación de datos mencionadas en IX.3.3.6. A continuación se explica brevemente el significado de los diversos ítems.

- **VERBOSITY**: se escogió el nivel 1 por lo cual en la ventana de mensajes de GATE Developer se muestra la información de configuración de mayor importancia.
- El valor "true" de **IS-LABEL-UPDATABLE** y **IS-NLPFEATURELIST-UPDATABLE** indica que la lista de abreviaturas de atributos de habilidad y de características lingüísticas se actualiza en función a los valores del conjunto de entrenamiento.
- **EVALUATION:** indica que el 66 % de los documentos forman parte del conjunto de entrenamiento y que el 34% se asignó al conjunto de prueba. Se podría haber dejado en blanco este elemento porque el método hold-out no se aplicó con el recurso *Batch Learning PR* sino mediante *Corpus Quality Assurance* tal como se explica en la sección IX.4.4.
- **multiClassification2Binary**: dada la existencia de 21 atributos de habilidad se tiene un problema donde a las oraciones que forman los mensajes de los e-tutores se les debe asignar una entre múltiples clases, por consiguiente se optó por convertirlo en múltiples problemas binarios con la estrategia "one-vs-another". Esto implica que cada uno de los modelos construidos contendrá 21 clasificadores, uno para cada atributo de habilidad.
- **thresholdProbabilityClassification:** de acuerdo al valor dado a este parámetro, para asignar un atributo de habilidad a una oración es necesario que probabilidad de que la oración se corresponda con ese atributo de habilidad sea como mínimo del 50%.

- **ENGINE:** indica que se usa la implementación de Weka del algoritmo KNN y que para clasificar una sentencia se tiene en cuenta el atributo de habilidad de las 5 sentencias más similares.
- **DATASET:** señala que las anotaciones del tipo "Sentence" se usan como instancias de entrenamiento para la construcción de los clasificadores. Para cada anotación del tipo "Sentence" se considera dos características: un uni-grama de nombre "Sentencegram" y un atributo de nombre "habilidad_etutor".
    a) Sentencegram: las anotaciones del tipo "Sentence" se representan mediante un uni-grama compuesto de las anotaciones del tipo "Token" abarcadas por el tipo de anotación "Sentence". A su vez, para cada anotación del tipo "Token" se considera su "string*"* y "kind". A **WEIGHT** se le asignó el valor "1.0" porque al momento de catalogar las anotaciones del tipo "Sentence" la única característica que se tiene en cuenta es el uni-grama de anotaciones del tipo "Token". Finalmente el valor de **ValueTypeNgram** especifica que a cada anotación del tipo "Token" se le asocia la inversa de su frecuencia en el documento.
    b) habilidad_etutor: mediante <POSITION>0</POSITION> se indica que el valor de la característica "atributo_hab" de una anotación del tipo "Sentence" depende solamente de la anotación "Sentence" y que no se debe tener en cuenta las anotaciones que le preceden o suceden. Además con el elemento vacio <CLASS/> se marca a la característica "atributo_hab" como un atributo de clase. Es decir que el valor de "atributo_hab" debe estar presente en cada oración del conjunto de entrenamiento y ausente en las oraciones del conjunto de prueba.

Parámetro	Valor
URL	E:\Proyecto_Identificacion_Habilidades\KNN\KNN-5\pre Procesamiento1\knnweka.xml
Corpus	corpusEntrenamiento
inputASName	Habilidades manifestadas
outputASName	
learningMode	TRAINING

Tabla 9.10. Parámetros de tiempo de ejecución de *Batch Learning PR*

Los valores de los parámetros de tiempo de ejecución que figuran en Tabla 9.10 son válidos para todas las combinaciones de operaciones de pre procesamiento y solo debe

modificarse la URL en función de donde se guarde el archivo xml de configuración. Dado que se busca construir el modelo el valor de *learningMode* es igual a "TRAINING", las operaciones de pre procesamiento deben aplicarse sobre el *Corpus* denominado *corpusEntrenamiento*, y las anotaciones que se consideran en la construcción de los clasificadores pertenecen al conjunto de anotación "Habilidades manifestadas" tal como lo señala el parámetro *inputASName*. El parámetro *outputASName* señala el conjunto de anotación donde se colocarán las anotaciones creadas al aplicar los clasificadores y solo es necesario cuando *learningMode* es igual a "EVALUATION" o "APPLICATION".

Si el modelo se construye con la tercera combinación de operaciones de pre procesamiento, y recurriendo a la lematización, en la Figura 9.14 se tendría que modificar la definición del elemento <NGRAM> tal como se muestra en la Figura 9.15.

```
<NGRAM>
   <NAME>Sentencegram</NAME>
   <NUMBER>1</NUMBER>
   <CONSNUM>1</CONSNUM>
   <CONS-1>
         <TYPE>Token</TYPE>
        <FEATURE>lemma</FEATURE>
    </CONS-1>
    <WEIGHT>1.0</WEIGHT>
    <ValueTypeNgram>3</ValueTypeNgram>
</NGRAM>
```

Figura 9.15. Definición del elemento NGRAM cuando se usa lematización en la tercera combinación de operaciones de pre procesamiento

Para el caso del *stemming to a root*, de cada anotación del tipo *"Token"* se tienen en cuenta las características "stem" y "kind" por lo cual al elemento <NGRAM> habría que reemplazarlo por el de la Figura 9.16.

```
<NGRAM>
    <NAME>Sentencegram</NAME>
    <NUMBER>1</NUMBER>
    <CONSNUM>2</CONSNUM>
    <CONS-1>
          <TYPE>Token</TYPE>
           <FEATURE>stem</FEATURE>
     </CONS-1>
     <CONS-2>
          <TYPE>Token</TYPE>
           <FEATURE>kind</FEATURE>
     </CONS-2>
     <WEIGHT>1.0</WEIGHT>
     <ValueTypeNgram>3</ValueTypeNgram>
 </NGRAM>
```

Figura 9.16. Definición del elemento NGRAM cuando se usa *stemming to a root* en la tercera combinación de operaciones de pre procesamiento.

Para ajustar el archivo de configuración xml a la cuarta combinación de operaciones de preparación de datos basta con reemplazar la definición del elemento <NGRAM> de la Figura 9.14 por la definición de la Figura 9.17.

```
<NGRAM>
   <NAME>Sentencegram</NAME>
   <NUMBER>1</NUMBER>
   <CONSNUM>2</CONSNUM>
   <CONS-1>
         <TYPE>Token</TYPE>
        <FEATURE>string</FEATURE>
   </CONS-1>
    <CONS-2>
         <TYPE>Token</TYPE>
        <FEATURE>category</FEATURE>
    </CONS-2>
    <WEIGHT>1.0</WEIGHT>
    <ValueTypeNgram>3</ValueTypeNgram>
</NGRAM>
```

Figura 9.17. Definición del elemento NGRAM para la cuarta combinación de operaciones de pre procesamiento.

La quinta combinación de operaciones de pre procesamiento se diferencia de la cuarta solamente por la inclusión de la remoción de *stop words* por lo cual la definición para el <NGRAM> es la misma. De manera similar la sexta combinación de operaciones de pre procesamiento se diferencia de la tercera por el uso de la supresión de *stop words* y en consecuencia las definiciones del <NGRAM> de las Figuras 9.15 y 9.16 se aplican en la sexta combinación.

Para la séptima combinación de operaciones de pre procesamiento corresponde utilizar las definiciones de <NGRAM> que muestran las Figuras 9.18 y 9.19 dependiendo de si se recurre a la lematización o bien al *stemming to a root*.

```
<NGRAM>
   <NAME>Sentencegram</NAME>
   <NUMBER>1</NUMBER>
   <CONSNUM>2</CONSNUM>
   <CONS-1>
         <TYPE>Token</TYPE>
        <FEATURE>lemma</FEATURE>
    </CONS-1>
    <CONS-2>
         <TYPE>Token</TYPE>
        <FEATURE>category</FEATURE>
     </CONS-2>
    <WEIGHT>1.0</WEIGHT>
    <ValueTypeNgram>3</ValueTypeNgram>
</NGRAM>
```

Figura 9.18. Definición del elemento NGRAM cuando se usa lematización en la séptima combinación de operaciones de pre procesamiento

```
<NGRAM>
   <NAME>Sentencegram</NAME>
   <NUMBER>1</NUMBER>
   <CONSNUM>2</CONSNUM>
   <CONS-1>
         <TYPE>Token</TYPE>
        <FEATURE>stem</FEATURE>
    </CONS-1>
    <CONS-2>
         <TYPE>Token</TYPE>
        <FEATURE>category</FEATURE>
     </CONS-2>
    <WEIGHT>1.0</WEIGHT>
    <ValueTypeNgram>3</ValueTypeNgram>
</NGRAM>
```

Figura 9.19. Definición del elemento NGRAM cuando se usa *stemming to a root* en la séptima combinación de operaciones de pre procesamiento

En la octava combinación de operaciones de preparación de datos se usan las definiciones de <NGRAM> de las dos figuras anteriores porque se distingue de la séptima combinación únicamente por la inclusión de la eliminación de *stop words*.

De los resultados arrojados por el método hold-out para los modelos obtenidos con el algoritmo KNN para k=5 y las diversas combinaciones de operaciones de pre procesamiento (mostrados en las Tablas 9.11 y 9.16), se pueden mencionar los siguientes aspectos:

- A medida que se aplica mayor cantidad de operaciones de preparación de datos, es decir se reduce la variabilidad en el espacio de características lingüísticas a considerar de los documentos, se mejora la efectividad en la identificación de habilidades.
- Cuando se utiliza el recurso de procesamiento *Tree Tagger* para conocer la categoría POST o el *lemma* de cada anotación del tipo "Token" se requiere un poco más de dos minutos para la construcción del modelo. En cambio, con los otros recursos el tiempo no supera los 21 segundos.
- En los modelos obtenidos recurriendo al recurso de procesamiento *Tree Tagger* se necesitó un poco más de 73 segundos para clasificar los 202 documentos del conjunto de prueba, en cambio los modelos donde no intervino *Tree Tagger* requirió como máximo 25 segundos.

Finalizada la etapa de experimentación inicial con el algoritmo KNN para K=5, se procedió a combinar los valores de K=10 y K=15 con los 8 tipos de operaciones pre procesamiento del apartado IX.3.3.6 para comprobar si el efecto de las operaciones de preparación de datos se mantenía ante variaciones del parámetro K. Teniendo en cuenta que en la práctica para valores de K mayores a 20 no se han producido mejoras significativas en la performance de los clasificadores (Sebastiani, 2002) inicialmente se fijo a 20 como un valor límite para K. Pero debido a que, como se puede comprobar en la Tabla 9.16, con K=10 y K=15 se reduce la efectividad de los clasificadores en el reconocimiento de habilidades, no se construyeron modelos con K=20. Es probable que la disminución de la efectividad por el aumento de los valores de K se deba a las características del conjunto de documentos usados en el proyecto, pues como ya se mencionó en el apartado IX.2.3 existen atributos de habilidad que no superan las 10 instancias mientras otros llegan a las 100 instancias. En este contexto, lo natural sería que los documentos de mayor grado de similitud al documento a clasificar pierdan peso en relación a los otros documentos que quizás no sean tan similares pero son mayoría.

La Tabla 9.16 muestra que los incrementos de precisión y recall que se logran mediante las operaciones de pre procesamiento se siguen manteniendo ante variaciones del parámetros K, aunque su impacto tiende a reducirse si se lo compara con las mejoras de la efectividad logradas para K=5. Por otro lado, en los experimentos realizados con K=10 y K=15, se observó que los tiempos para la construcción de los modelos (incluyendo las operaciones de pre procesamiento) y los tiempos de clasificación no varían

significativamente en relación a los tiempos requeridos con K=5. En la Tabla 9.11 aparecen expresados en segundos los tiempos de construcción de los modelos (T1) y los tiempos de clasificación (T2).

Parámetros / Procesamiento	K=5		K=10		K=15	
	T_1	T_2	T_1	T_2	T_1	T_2
Pre procesamiento 1	14,687	24,125	18,532	19,781	11,36	19,11
Pre procesamiento 2	13,593	20,016	14,937	20,781	14,968	19,391
Pre procesamiento 3 A	140,094	76,672	140,204	74,969	140,501	76,453
Pre procesamiento 3 B	15,579	14,797	14,657	17,547	14,72	15,812
Pre procesamiento 4	137,454	79,234	140,766	82,297	137,781	79,656
Pre procesamiento 5	137,374	79,359	136,796	80,875	138,468	80,406
Pre procesamiento 6 A	135,39	76,219	136,219	76,36	135,89	77,609
Pre procesamiento 6 B	20,937	19,547	20,64	10,16	20,327	20,735
Pre procesamiento 7 A	145,032	78,531	190,798	80,61	186,704	73,813
Pre procesamiento 7 B	134,609	80,453	135,141	78,125	133,125	85,015
Pre procesamiento 8 A	141,97	76,593	143,251	78,469	142,032	81,188
Pre procesamiento 8 B	133,265	81,922	130,063	79,89	130,484	93,578

Tabla 9.11. Tiempos de construcción y de clasificación para el algoritmo KNN

Para construir los modelos mediante el algoritmo KNN para K=10 y K=15 se tiene que modificar el valor de *options* en la definición del elemento <ENGINE/> que aparece en la Figura 9.14. Además de acuerdo a la combinación de operaciones de pre procesamiento utilizada se debe ajustar el archivo de configuración xml tal como se ha comentado.

En el directorio *Mineria_de_Textos\Modelos\KNN* del CD que acompaña a este documento se incluyen los archivos de configuración xml junto con los modelos que se pueden obtener considerando las diversas operaciones de procesamiento y los parámetros K=5, K=10 y K=15.

IX.4.3.2 Construcción del modelo con SVM

En la construcción de los modelos mediante el algoritmo SVM se usó la implementación en Java del paquete LibSVM con la que cuenta GATE. A continuación, en la Figura 9.20 se muestra el archivo de configuración que debe ser utilizado con la primera y segunda combinación de operaciones de pre procesamiento.

```
<?xml version="1.0"?>
<ML-CONFIG>
  <VERBOSITY level="0"/>
  <IS-LABEL-UPDATABLE value="true"/>
  <IS-NLPFEATURELIST-UPDATABLE value="true"/>

   <EVALUATION method="holdout" runs="1" ratio="0.66"/>
   <multiClassification2Binary method="one-vs-others"/>
   <PARAMETER name="thresholdProbabilityClassification" value=
   "0.5"/>

   <ENGINE nickname="SVM" implementationName="SVMLibSvmJava"
   options="-s 0 -c 0.8 -t 0 -m 100 -tau 0.5 "/>

   <DATASET>
     <INSTANCE-TYPE>Sentence</INSTANCE-TYPE>
     <NGRAM>
        <NAME>Sentencegram</NAME>
        <NUMBER>1</NUMBER>
        <CONSNUM>2</CONSNUM>
        <CONS-1>
           <TYPE>Token</TYPE>
            <FEATURE>string</FEATURE>
         </CONS-1>
         <CONS-2>
           <TYPE>Token</TYPE>
            <FEATURE>kind</FEATURE>
         </CONS-2>
        <WEIGHT>1.0</WEIGHT>
     </NGRAM>
     <ATTRIBUTE>
           <NAME>habilidad_etutor</NAME>
           <SEMTYPE>NOMINAL</SEMTYPE>
           <TYPE>Sentence</TYPE>
           <FEATURE>atributo_hab</FEATURE>
           <POSITION>0</POSITION>
           <CLASS/>
    </ATTRIBUTE>
   </DATASET>
</ML-CONFIG>
```

Figura 9.20. Archivo de configuración xml para el algoritmo SVM

Este archivo xml difiere del usado con KNN solamente en la especificación del elemento ENGINE y por ello las explicaciones se centrarán en este elemento. Mediante *implementationName* y *Nickname* se establece que se usa la implementación "SVMLibSvmJava" del algoritmo SVM y que a este algoritmo se le asignó el sobrenombre “SVM”. Para construir el modelo es necesario también especificar el valor para los siguientes parámetros:

- -s: el valor “0” indica que se aplica SVM a problemas de clasificación binarios.

- -c: por defecto el costo de los errores por clasificación es de 1 pero con el fin de permitir un pequeño margen de error y obtener resultados más generalizables se decidió usar "0.8".
- -t: puesto que se asume que los ejemplos del conjunto de entrenamiento se pueden separar en dos grupos al momento de construir el clasificador para cada atributo de habilidad, el valor "0" en t indica que se usa un kernel del tipo lineal.
- -m: el espacio de memoria cache usado para construir los clasificadores es de 100 MB.
- tau: teniendo en cuenta que para todos los atributos de habilidad la cantidad de ejemplos de entrenamiento positivos es menor a la cantidad de ejemplos negativos se prefirió por usar la versión de SVM con márgenes desiguales. Esto significa que los valores de tau usados van a diferir de "1" que es valor que corresponde a la versión estándar de SVM.

Para construir los modelos con SVM y las otras 6 combinaciones de operaciones de pre procesamiento se tiene que modificar el elemento NGRAM en base a las observaciones realizadas al explicar la construcción mediante el algoritmo KNN. En (Li & Shawe-taylor, 2003) se sugiere que los valores óptimos para τ usualmente se encuentran dentro de las siguientes 15 alternativas: {-0.1, 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 2}. En este proyecto no se construyeron modelos con los 15 valores listados ya que al combinarlos con las 8 alternativas de preparación de datos se tendría un total de 120 modelos posibles. Lo que se hizo fue comenzar con τ=0.5 para luego incrementar o reducir el valor en función al efecto que la variación tenia sobre le efectividad en el reconocimiento de habilidades. Siguiendo esta estrategia solo se probaron cuatro valores de τ (0.3, 0.4, 0.5 y 0.6) puesto que fuera de los mismos la precisión y el recall asumían valores que no permitirían cumplir con los criterios de éxito definidos para este proyecto. Los resultados de las pruebas realizadas con los diversos valores de τ se incluyen en la Tabla 9.17.

Para cada uno de los cuatro valores de τ se aplicaron las diversas operaciones de pre procesamiento con el fin de establecer si tenían influencia en el reconocimiento de las habilidades. Los resultados de los experimentos (resumidos en Tabla 9.17) muestran que si bien la eliminación *stop words* tiene un efecto casi imperceptible sobre la precisión de los clasificadores, el asignar POST reduce el valor de la precisión hasta en un décimo. Por su

parte, la normalización puede ocasionar una disminución de la precisión en un décimo cuando se opta por el *stemming to a root*. Lo anterior sin embargo no significa que la combinación de la normalización con el POST causen una reducción de la precisión de dos décimos porque incluso aplicando las dos operaciones la reducción se mantiene dentro de un décimo. La disminución del valor de precisión conforme se reduce el espacio de características lingüísticas a ser tenidas en cuenta muestra que SVM se desempeña mejor cuando el espacio de características es amplio.

Con relación al valor de τ a medida que se acerca a "1" se produce un incremento del valor de precisión pero una reducción en el recall, en tanto que para valores próximos a cero es el valor de recall el que crece y el de precisión el que disminuye.

Parámetros / Procesamiento	τ=0,3		τ=0,4		τ=0,5		τ=0,6	
	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
Pre procesamiento 1	10,078	6,829	10,735	6,672	17,953	12,531	14,469	8,90
Pre procesamiento 2	21,75	10,625	21,531	10,578	16,454	7,203	18,641	10,40
Pre procesamiento 3 A	141,173	70,11	141,891	66,672	134,375	69,297	139,845	65,5
Pre procesamiento 3 B	14,859	6,969	13,016	6,875	17,516	9,5	12,626	7,43
Pre procesamiento 4	131,469	69,922	127,234	69,45	133,875	69,25	126,422	66,47
Pre procesamiento 5	129,934	68,985	129,122	67,656	129,17	66,64	128,137	67,01
Pre procesamiento 6 A	128,125	67,453	128,078	66,938	127,5	64,844	127,906	65,81
Pre procesamiento 6 B	17,594	8,828	18,563	7,187	18,735	7,562	18,641	6,84
Pre procesamiento 7 A	131,938	70,5	129,937	82,4	131,75	67,156	130,093	66,95
Pre procesamiento 7 B	133,407	73,266	132,407	68	133,407	67,375	132,047	66,29
Pre procesamiento 8 A	130,828	70,5	129,844	65,031	132,797	72,593	129,594	66,36
Pre procesamiento 8 B	132	66,812	132,391	66,281	135,36	68,672	132,157	66,47

Tabla 9.12: Tiempos de construcción y de clasificación para el algoritmo SVM

La Tabla 9.12 muestra los segundos que se necesitan para construir los modelos (T1) y para clasificar nuevas instancias (T2) para las diversas combinaciones de valores del parámetro τ y las operaciones de pre procesamiento. En general se observa que a medida que crece el parámetro τ decrece el tiempo de clasificación. Por ejemplo, para la combinación 7B de operaciones de preparación de datos se produce una reducción de casi 7 segundos en el tiempo de clasificación cuando el parámetro τ cambia desde 0.3 a 0.6.

Los archivos de configuración xml así como los modelos construidos al combinar las diversas operaciones de procesamiento con los cuatro valores del parámetro τ mostrados en la Tabla 9.12 se incluyen en el directorio *Mineria_de_Textos\Modelos\SVM* del CD que complementa a este documento.

IX.4.3.3 Construcción del modelo con PAUM

Con el algoritmo PAUM se tiene que indicar un valor para el margen positivo (τ_{+1}) y un valor para el margen negativo (τ_{-1}). De acuerdo con Li *et al.* (2002) los valores óptimos de τ_{-1} suelen pertenecer al conjunto {-1.5,-1,-0.5, 0,0.1,0.5, 1.0} en tanto que los valores óptimos de τ_{+1} pertenecen a {-1,-0.5,0,0.1,0.5,1,2,5,10,50}. Teniendo en cuenta que existen 70 combinaciones de los valores de los márgenes y que se tienen 8 operaciones de pre procesamiento, el número total de modelos posibles es de 560. No se construyeron todos los modelos posibles para establecer el valor apropiado de los parámetros sino que, mediante la validación cruzada 10-fold se combino los valores de los dos márgenes con la primera operación de preparación de datos. De todas combinaciones posibles entre los valores de los dos márgenes y la primera operación de pre procesamiento (expuestas en Tabla 9.13) se seleccionaron aquellas que permitían obtener el valor más alto de precisión, llegar al valor más alto de recall, tener la menor diferencia entre los valores de recall y precisión, y cumplir con al menos uno de los criterios de éxito del proyecto pero sin alejarse demasiado del valor establecido por el otro criterio de éxito. Los valores que permitieron cumplir la primera condición son τ_{+1}=1 y τ_{-1}=-1.5, en tanto que la segunda y tercera condición la cumplen los parámetros τ_{+1}= 10 y τ_{-1}=0, y para la cuarta condición se tiene que usar τ_{+1}=2 y τ_{-1}=-1.5. Del proceso de ajuste de los parámetros realizado con la validación cruzada se pueden hacer las siguientes observaciones:

- Para valores positivos de τ_{+1} un crecimiento del valor de τ_{-1} en el sentido positivo significa un aumento en el valor de la precisión y una disminución en el valor del recall.
- Para valores positivos de τ_{+1} un decrecimiento del valor de τ_{-1} en el sentido negativo significa un aumento en el valor de la precisión y una disminución en el valor del recall.
- Para $\tau+_1$=0 un crecimiento del valor de τ_{-1} en el sentido positivo significa un aumento en el valor de la precisión y una disminución en el valor del recall.
- Para $\tau+_1$= 0 un decrecimiento del valor de τ_{-1} en el sentido negativo significa un aumento en el valor de la precisión y una disminución en el valor del recall.
- Para τ_{-1}= 0 un aumento en el valor de τ_{+1} en el sentido positivo significa una disminución de precisión y un pequeño aumento del recall.
- Para τ_{-1}=0 a medida que el valor τ_{+1} se alejo del cero en sentido negativo los valores del recall y precisión disminuyen.

- Para valores negativos de τ_{-1} a medida que el valor de τ_{+1} crece en el sentido positivo se produce una disminución de precisión y un aumento del recall.
- Para valores negativos de τ_{-1} a medida que τ_{+1} se alejo del cero en sentido negativo se produce una disminución tanto de precisión como de recall.

τ_{-1} \ τ_{+1}	**-1,5**		**-1**		**-0.5**		**0**		**0.1**		**0.5**		**1**	
	π	ρ	π	ρ	π	ρ	π	ρ	π	ρ	π	ρ	π	ρ
-1	0.773	0.138	0.716	0.213	0.217	0.206	0.296	0.303	0.293	0.302	0.217	0.198	0.674	0.214
-0,5	0.779	0.175	0.607	0.253	0.714	0.165	0.539	0.276	0.342	0.355	0.324	0.258	0.568	0.246
0	0.764	0.179	0.577	0.284	0.364	0.324	0.375	0.388	0.373	0.386	0.364	0.324	0.577	0.284
0,1	0.708	0.192	0.726	0.195	0.686	0.209	0.462	0.340	0.448	0.291	0.703	0.213	0.723	0.194
0,5	0.758	0.141	0.706	0.187	0.663	0.222	0.425	0.322	0.462	0.314	0.657	0.227	0.692	0.226
1	0.770	0.193	0.635	0.237	0.625	0.245	0.433	0.371	0.411	0.333	0.649	0.244	0.667	0.237
2	0.569	0.300	0.485	0.350	0.599	0.304	0.405	0.380	0.373	0.384	0.439	0.376	0.497	0.354
5	0.555	0.319	0.429	0.397	0.422	0.356	0.389	0.373	0.389	0.402	0.389	0.402	0.436	0.391
10	0.408	0.387	0.397	0.398	0.400	0.411	0.390	0.403	0.393	0.405	0.393	0.404	0.405	0.405
50	0.360	0.375	0.358	0.372	0.346	0.360	0.347	0.362	0.347	0.361	0.345	0.360	0.354	0.369

Tabla 9.13. Resultados del ajuste de los parámetros del algoritmo PAUM

Una vez definidos los valores de τ_{+1} y τ_{-1} que cumplen las cuatro condiciones mencionadas se procedió a construir modelos con los valores establecidos y las 8 combinaciones de operaciones de pre procesamiento a fin de determinar el efecto que la modificación del espacio de características tenía sobre la efectividad de los modelos. En este paso si se tuvo en cuenta la separación entre *corpusEntrenamiento* y *corpusPrueba*, pero cuando se seleccionó los valores de los parámetros con la validación cruzada de 10-fold la distinción entre los conjuntos de entrenamiento y de prueba la realizaba el software automáticamente. En los modelos obtenidos al combinar los valores de τ_{-1} y τ_{+1} con las operaciones de pre procesamiento se pudo comprobar que la reducción del espacio de características no incrementa la precisión sino que la disminuye (ver Tabla 9.18). Esto indica que cuanto más grande sea el espacio de características menor será la cantidad de errores que cometerán los clasificadores obtenidos mediante PAUM.

La Figura 9.21 corresponde al archivo de configuración para el algoritmo PAUM con τ_{+1}=1 y τ_{-1} =-1.5. El valor "0.3" de –optB indica que el termino de sesgo b tiene un valor igual a "b+0.3". Para construir los modelos con τ_{+1}= 10 y τ_{-1}=0 como también con τ_{+1}=2 y τ_{-1}=-1.5 se deben modificar -p y –n en la definición del elemento ENGINE, en tanto que para aplicar las diversas operaciones de preparación de datos es preciso editar la definición del elemento NGRAM siguiendo los lineamientos de la sección IX.4.3.1.

```
<?xml version="1.0"?>
<ML-CONFIG>
  <VERBOSITY level="0"/>
  <IS-LABEL-UPDATABLE value="true"/>
  <IS-NLPFEATURELIST-UPDATABLE value="true"/>
  <EVALUATION method="holdout" runs="1" ratio="0.66"/>
  <multiClassification2Binary method="one-vs-others"/>
  <PARAMETER name="thresholdProbabilityClassification" value
  ="0.5"/>
  <ENGINE nickname="PAUM" implementationName="PAUM" options="-p
  1-n -1.5 -optB 0.3 "/>
   <DATASET>
     <INSTANCE-TYPE>Sentence</INSTANCE-TYPE>
     <NGRAM>
        <NAME>Sentencegram</NAME>
        <NUMBER>1</NUMBER>
        <CONSNUM>2</CONSNUM>
        <CONS-1>
            <TYPE>Token</TYPE>
            <FEATURE>string</FEATURE>
         </CONS-1>
         <CONS-2>
            <TYPE>Token</TYPE>
            <FEATURE>kind</FEATURE>
         </CONS-2>
        <WEIGHT>1.0</WEIGHT>
     </NGRAM>
     <ATTRIBUTE>
           <NAME>habilidad_etutor</NAME>
           <SEMTYPE>NOMINAL</SEMTYPE>
           <TYPE>Sentence</TYPE>
           <FEATURE>atributo_hab</FEATURE>
           <POSITION>0</POSITION>
           <CLASS/>
      </ATTRIBUTE>
   </DATASET>
</ML-CONFIG>
```

Figura 9.21. Archivo de configuración xml para el algoritmo PAUM

En el directorio *Mineria_de_Textos\Modelos\PAUM* del CD que sirve de complemento al presente trabajo de investigación, se incluyen los archivos de configuración y los clasificadores automáticos construidos al considerar las ocho combinaciones de operaciones de pre procesamiento y, las tres combinaciones de valores de los parámetros τ_{+1} y τ_{-1} de la Tabla 9.14.

Parámetros / Procesamiento	τ_{+1}=1 τ_{-1}=-1,5		τ_{+1}=2 τ_{-1}= -1,5		τ_{+1}=10 τ_{-1}= 0	
	T_1	T_2	T_1	T_2	T_1	T_2
Pre procesamiento 1	14,687	9,484	14,438	10,687	14,141	10,063
Pre procesamiento 2	15,797	6,813	15,703	6,922	14,297	7,891
Pre procesamiento 3 A	139,126	68,171	140,173	69,031	144,173	69,406
Pre procesamiento 3 B	16,854	8,062	14,454	6,078	14,751	6,844
Pre procesamiento 4	138,219	65,813	140,626	68,812	137,953	66,141
Pre procesamiento 5	137,547	68,109	137,796	66,172	137,937	66,656
Pre procesamiento 6 A	135,266	66,843	135,156	66,75	135,077	66,265
Pre procesamiento 6 B	20,89	6,813	21,187	6,062	20,093	6,609
Pre procesamiento 7 A	139,579	65,64	139,797	64,203	139,641	66,281
Pre procesamiento 7 B	133,5	74,204	136,235	66,485	132,266	64,75
Pre procesamiento 8 A	144,5	70,375	142,328	66,312	141,235	67,765
Pre procesamiento 8 B	130,75	68,593	130,203	65,906	130,062	67,407

Tabla 9.14. Tiempos de construcción y de clasificación para el algoritmo PAUM

De la comparación de la tabla anterior con los valores de tiempo incluidos en las Tablas 9.11 y 9.12 se desprende que el tiempo promedio necesario para construir un modelo con PAUM es apenas inferior a lo que requiere el algoritmo KNN y superior en 4 segundos si se lo compara con el algoritmo SVM. Por otro lado, para clasificar nuevas instancias con un modelo obtenido mediante PAUM o SVM se necesitan en promedio 10 segundos menos que en los clasificadores del tipo KNN.

IX.4.3.4 Construcción del Modelo con Naïve Bayes

La Tabla 9.15 contiene las medidas de efectividad (precisión, recall, y F_1) para los modelos construidos con Naïve Bayes. Un análisis superficial podría llevar a concluir que el algoritmo Naïve Bayes es la mejor aproximación al problema de identificación automática de habilidades de e-tutor puesto que permite lograr una precisión de 1.00 (la máxima posible) con todas las combinaciones de operaciones de pre procesamiento. Sin embargo, esto no es correcto porque el valor de recall igual a cero indica que en realidad no se pudo asignar el atributo de habilidad a ninguna de las contribuciones de los docentes. Una revisión a la ecuación del cálculo de precisión presentada en la sección IX.1.2.4 permitirá comprender mejor esta situación. De acuerdo con esta ecuación la precisión es el cociente entre el número de predicciones positivas correctas y el número de predicciones positivas. Para Naïve Bayes el numerador de la ecuación es cero puesto que no se identificó correctamente el atributo de habilidad para ninguna de las sentencias del conjunto de prueba, y el denominador también es cero porque no se realizaron decisiones de clasificación. El numerador y el denominador pueden ser cancelados por ser iguales y en consecuencia el resultado del cociente es 1.

$$\frac{Número\ de\ predicciones\ positivas\ correctas}{Número\ de\ predicciones\ positivas} = \frac{0}{0} = 1$$

De lo expuesto en el párrafo anterior se concluye que el algoritmo Naïve Bayes no permite construir clasificadores para la detección automática de las habilidades de los docentes de entornos de ACSC.

Para mayores detalles de las medidas de efectividad de los modelos construidos con Naïve Bayes se puede revisar el directorio *Mineria_de_Textos\Modelos\ Naive_Bayes* del CD entregado junto con el presente documento.

IX.4.3.5 Construcción del modelo con C4.5

En GATE se puede utilizar la implementación de Weka del algoritmo C4.5, pero la aproximación de árboles de decisión no es de utilidad para la identificación automática de habilidades de un e-tutor, ya que como se muestra en Tabla 9.15, el valor de recall es igual a cero independientemente de las operaciones de preparación aplicadas al conjunto de datos.

Algoritmo / Procesamiento	Naïve Bayes			C4.5		
	π	ρ	F_1	π	ρ	F_1
Pre procesamiento 1	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 2	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 3 A	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 3 B	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 4	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 5	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 6 A	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 6 B	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 7 A	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 7 B	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 8 A	1.00	0.00	0.00	1.00	0.00	0.00
Pre procesamiento 8 B	1.00	0.00	0.00	1.00	0.00	0.00

Tabla 9.15. Medidas de efectividad para los algoritmo Naïve Bayes y C4.5

Los modelos obtenidos con el algoritmo C4.5 y las medidas de efectividad se pueden consultar en el directorio *Mineria_de_Textos\Modelos\C4.5* del CD que acompaña al presente documento.

IX.4.4 Evaluar el Modelo

Para evaluar los modelos construidos inicialmente se usó la técnica hold-out destinando el 66 % de los documentos para entrenamiento y el 44 % para pruebas. La

comparación entre los atributos de habilidad asignados por expertos y los atributos de habilidad reconocidos automáticamente por los clasificadores se hizo por medio de un recurso de GATE denominado *Corpus Quality Assurance*. Este recurso permite calcular precisión (π), recall (ρ), y F_1 entre dos conjuntos de anotación de un mismo corpus.

Teniendo en cuenta que *Corpus Quality Assurance* necesita dos conjuntos de anotación y que "atributo_hab" debe estar presente en cada anotación del tipo "Sentence" del conjunto de entrenamiento pero ausente en el conjunto de prueba, se decidió trabajar con tres conjuntos de anotación sobre *corpusPrueba:* "Habilidades manifestadas", "preProcesamiento", y "Habilidades reconocidas". "Habilidades manifestadas" contiene el atributo de habilidad que los expertos asignaron a cada oración del *corpusPrueba*. Por su parte, "preProcesamiento" es un conjunto intermedio que contiene las anotaciones creadas por las operaciones de preparación de datos que se tiene que aplicar a un documento antes de poder clasificar automáticamente las sentencias que lo componen. En tanto que, "Habilidades reconocidas" tiene solamente anotaciones del tipo "Sentence" donde el valor de la característica de anotación "atributo_hab" es igual al atributo de habilidad que los clasificadores asignan a las sentencias de *corpusPrueba*. De esta manera para calcular precisión, recall, y F_1 las comparaciones se hacen entre las características de anotación "atributo_hab" de los conjuntos "Habilidades manifestadas" y "Habilidades reconocidas" de cada documento de *corpusPrueba*.

El proceso anterior se puede completar sin necesidad de recurrir al conjunto de anotación "preProcesamiento" porque el conjunto "Habilidades reconocidas" puede contener las anotaciones de las operaciones de pre procesamiento pero, esto implicaría la creación de dos anotaciones solapadas del tipo "Sentence" en "Habilidades reconocidas" para las oraciones de *corpusPrueba*.

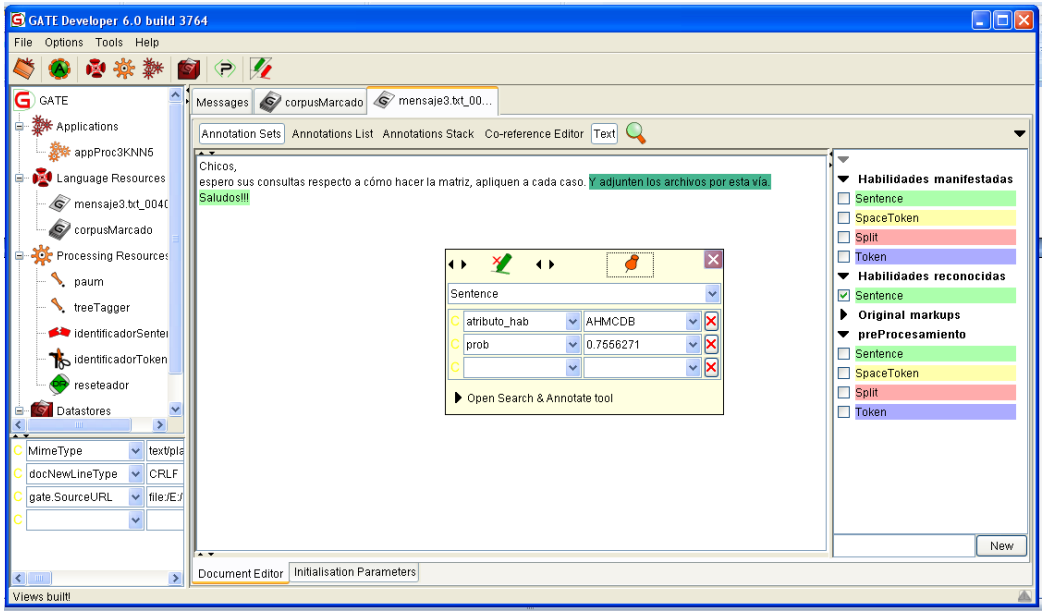


Figura 9.22. Conjuntos de anotación usados en *Corpus Quality Assurance*

Al calcular precisión, recall (ρ), y F_1 mediante *Corpus Quality Assurance* se generan las siguientes salidas:

- **Tabla de estadísticas del corpus**: muestra los valores de efectividad de la clasificación calculados sobre el conjunto de prueba completo.
- **Tabla de estadísticas de los documentos**: contiene una fila para cada documento en el corpus. Las columnas de la tabla señalan los valores precisión, recall y F_1 obtenidos para cada documento.
- **Tablas de contingencia**: se tiene una tabla de contingencia para la colección de documentos completa y tantas tablas de contingencia individuales como documentos existan en la colección.

Las salidas generadas mediante *Corpus Quality Assurance* se almacenaron en un archivo de nombre "resultados_CorpusQA.html". En la subcarpeta *Modelos* de la carpeta *Mineria_de_Textos* del CD que se adjunta a este trabajo se tiene un archivo "resultados_CorpusQA.html" para cada uno de los modelos construidos con los cinco algoritmos de aprendizaje (KNN, PAUM, Naïve Bayes, C4.5, y SVM) considerados en este trabajo de investigación.

IX.4.4.1 Resultados de la evaluación con hold-out

La Tabla 9.16 contiene los valores de precisión, recall y F_1 para el algoritmo KNN cuando se considera 5,10 o 15 oraciones del conjunto de entrenamiento para clasificar una oración del conjunto de pruebas.

Parámetros / Procesamiento	KNN con K=5			KNN con K=10			KNN con K=15		
	π	ρ	F_1	π	ρ	F_1	π	ρ	F_1
Pre procesamiento 1	0,36	0,27	0,31	0,29	0,25	0,27	0,22	0,22	0,22
Pre procesamiento 2	0,40	0,30	0,34	0,32	0,29	0,30	0,24	0,24	0,24
Pre procesamiento 3 A	0,42	0,31	0,36	0,34	0,31	0,32	0,27	0,27	0,27
Pre procesamiento 3 B	0,45	0,35	0,39	0,33	0,31	0,32	0,25	0,25	0,25
Pre procesamiento 4	0,36	0,29	0,32	0,27	0,25	0,26	0,21	0,21	0,21
Pre procesamiento 5	0,39	0,31	0,34	0,33	0,31	0,32	0,23	0,23	0,23
Pre procesamiento 6 A	0,42	0,32	0,36	0,38	0,33	0,35	0,29	0,28	0,29
Pre procesamiento 6 B	0,46	0,36	0,40	0,35	0,33	0,34	0,28	0,27	0,28
Pre procesamiento 7 A	0,42	0,32	0,37	0,31	0,29	0,30	0,24	0,24	0,24
Pre procesamiento 7 B	0,43	0,33	0,37	0,31	0,30	0,31	0,24	0,24	0,24
Pre procesamiento 8 A	0,42	0,32	0,36	0,34	0,31	0,33	0,27	0,27	0,27
Pre procesamiento 8 B	0,46	0,36	0,41	0,37	0,34	0,35	0,24	0,24	0,24

Tabla 9.16. Medidas de efectividad para el algoritmo KNN

El aumento de los valores de las tres medidas de efectividad de los clasificadores conforme se aplica mayor cantidad de operaciones de pre procesamiento muestra que para el algoritmo KNN, cuando se reduce el espacio de características lingüísticas, se pueden identificar las oraciones que tienen mayor grado de similitud con la oración a ser clasificada. El efecto de la preparación de datos se mantiene a pesar de las variaciones en el parámetro K, aunque su impacto en la precisión y el recall tiende a disminuir cuando K aumenta. Esto considerando que con K=5 se pueden lograr incrementos de precisión y recall de hasta un 0.10 en tanto que para K=15 el crecimiento es a lo sumo de 0.07. En las Figuras 9.23, 9.24, y 9.25 se pueden apreciar las variaciones producidas en precisión, recall, y F_1 en función a las operaciones de pre procesamiento aplicadas para los modelos construidos con K=5, k=10 y K=15.

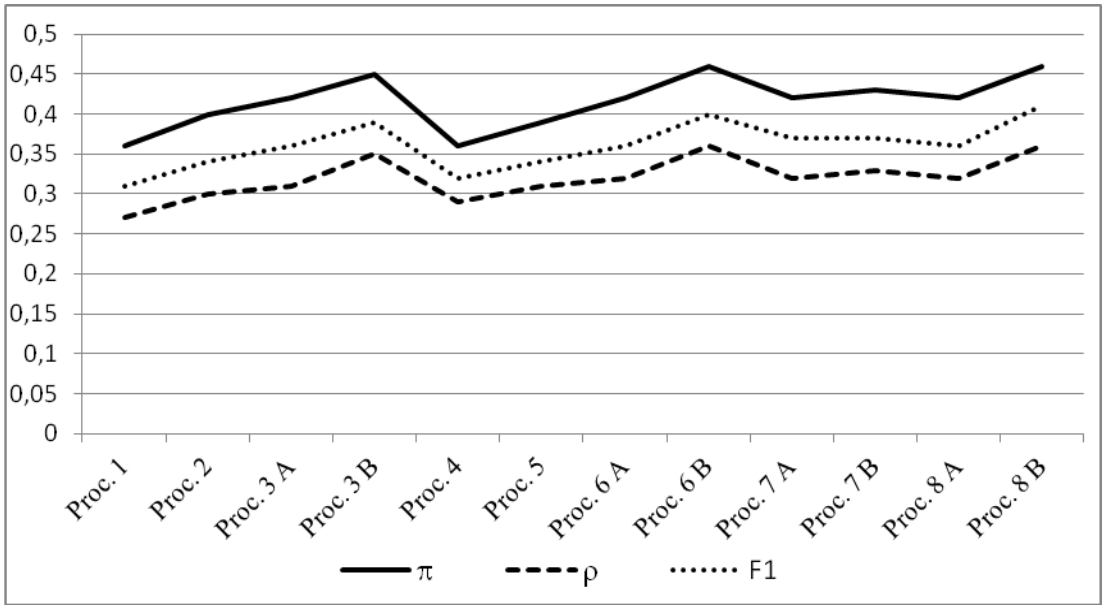


Figura 9.23. Variación de la efectividad de los clasificadores construidos con K=5

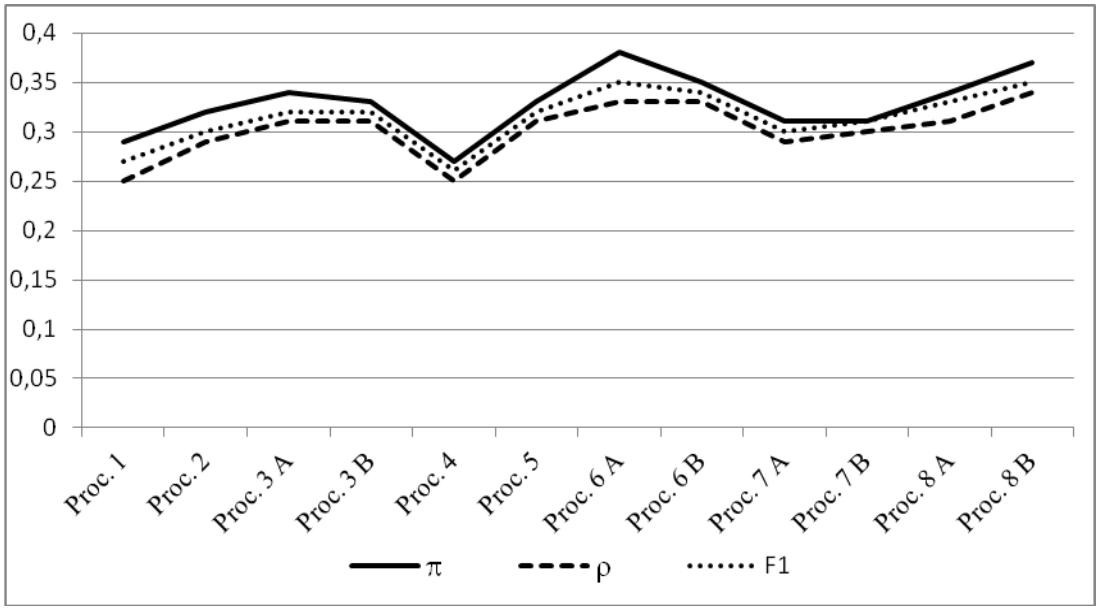


Figura 9.24. Variación de la efectividad de los clasificadores construidos con K=10

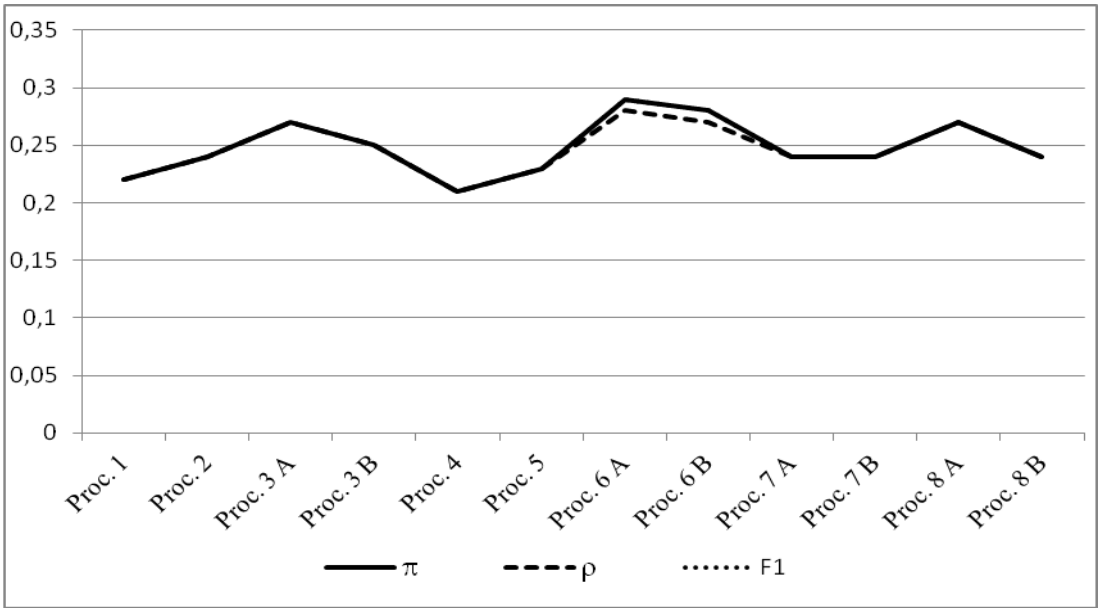


Figura 9.25. Variación de la efectividad de los clasificadores construidos con K=15

De la Tabla 9.16 también se desprende que conforme crecen los valores de K se incrementa la cantidad de errores cometidos por los clasificadores, es decir disminuye su precisión. En particular, para el atributo AHFE se pudo comprobar, en las tablas de contingencia producidas por *Corpus Quality Assurance*, que con K=5 se cometen 15 errores de clasificación en tanto que para K=15 la cantidad de errores casi se duplica. Tal como se había señalado en IX.4.3.1 un incremento del valor de K implica una pérdida de peso de las sentencias que guardan mayor grado de similitud con la sentencia que se busca clasificar.

Las celdas resaltadas con amarrillo en la Tabla 9.16 muestran que los mejores valores para las tres medidas de efectividad se consiguen con K=5 y la segunda variante de la octava combinación de operaciones de pre procesamiento (identificación de Token, identificación de sentencias, POST, *stemming to a root*, y eliminación de *stop words*). Sin embargo, los valores no permiten cumplir con ninguno de los dos criterios de éxito establecidos para el proyecto. Por un lado, se pretendía reconocer al menos el 60% de las habilidades sociales manifestadas por los docentes pero el valor de recall igual a 0.36 señala que solo se pudo reconocer el 36 % de las habilidades presentes. Por otro lado, π =0,46 indica que se clasificó correctamente el 46 % de los atributos de habilidad identificados por lo cual el margen de error es bastante superior al 30% que se había establecido como límite.

Un factor que influye en la efectividad de los clasificadores automáticos basados en el algoritmo KNN es la similitud entre las sentencias catalogadas bajo un mismo atributo

de habilidad. Si se examinan las tablas de contingencia de los archivos "resultados_CorpusQA.html" del directorio *Mineria_de_Textos\Modelos\KNN* en el CD que se adjunta, se puede comprobar que para el atributo de habilidad OAH a pesar de existir 77 ejemplos de entrenamiento, los clasificadores construidos pudieron identificar correctamente a lo sumo 2 de las 14 instancias del conjunto de prueba. Recuérdese que este atributo de habilidad fue asignado por los expertos a todas aquellas oraciones en las que no se podía identificar un atributo de habilidad social, por lo cual es de esperar que difícilmente se puedan encontrar puntos en común en la diversidad de oraciones que reúne. Téngase en cuenta también que las sentencias etiquetadas con OAH pueden pertenecer a cualquiera de las otras 5 categorías de habilidad del esquema de clasificación construido para este proyecto. La situación anterior se hace más evidente cuando se considera un atributo de habilidad como AHUEA, ya que con solo 13 ejemplos de entrenamiento se pudo construir un clasificador mucho más preciso que el clasificador correspondiente a OAH.

Parámetros / Procesamiento	**τ=0,3**			**τ=0,4**			**τ=0,5**			**τ=0,6**		
	π	ρ	F1	π	ρ	Φ_1	π	ρ	F_1	π	ρ	F_1
Pre procesamiento 1	0,67	0,31	0,42	0,80	0,25	0,38	0,84	0,22	0,35	0,84	0,20	0,32
Pre procesamiento 2	0,65	0,29	0,40	0,78	0,24	0,37	0,84	0,21	0,34	0,83	0,20	0,32
Pre procesamiento 3 A	0,61	0,27	0,38	0,70	0,25	0,37	0,78	0,22	0,34	0,83	0,21	0,34
Pre procesamiento 3 B	0,62	0,28	0,39	0,72	0,24	0,36	0,81	0,22	0,34	0,84	0,19	0,31
Pre procesamiento 4	0,61	0,27	0,38	0,70	0,24	0,35	0,77	0,21	0,33	0,75	0,19	0,30
Pre procesamiento 5	0,62	0,29	0,39	0,70	0,22	0,34	0,76	0,22	0,34	0,77	0,19	0,31
Pre procesamiento 6 A	0,64	0,27	0,38	0,71	0,24	0,36	0,77	0,22	0,34	0,84	0,20	0,33
Pre procesamiento 6 B	0,64	0,28	0,39	0,72	0,24	0,36	0,82	0,22	0,35	0,86	0,20	0,32
Pre procesamiento 7 A	0,63	0,29	0,40	0,72	0,25	0,37	0,76	0,22	0,35	0,80	0,21	0,33
Pre procesamiento 7 B	0,65	0,28	0,39	0,71	0,25	0,38	0,76	0,22	0,34	0,80	0,20	0,32
Pre procesamiento 8 A	0,62	0,28	0,39	0,74	0,25	0,37	0,75	0,22	0,34	0,80	0,21	0,33
Pre procesamiento 8 B	0,66	0,29	0,40	0,72	0,26	0,39	0,75	0,22	0,34	0,80	0,20	0,32

Tabla 9.17. Medidas de medidas de efectividad para el algoritmo SVM

La Tabla 9.17 muestra los valores de precisión, recall, y F_1 para diferentes valores de parámetro τ del algoritmo SVM con márgenes desiguales introducido por Li y Shawe-taylor (2003). Como puede observarse las operaciones de pre procesamiento no incrementan la efectividad de los clasificadores automáticos tal como sucede con el algoritmo KNN sino todo lo contrario. Es posible que esto se deba a que es más fácil determinar los valores de los vectores W que permiten distinguir entre atributos de habilidad cuando se tiene un espacio grande de características (ver sección III.8.5). La reducción del espacio de características (en particular con operaciones como la

normalización) causa un aumento en la similitud entre las sentencias y es de esperarse que sea más complicado determinar el peso de los elementos del vector W de cada atributo de habilidad de modo que se lo pueda distinguir fácilmente de otros atributos de habilidad. En la Figura 9.26 se puede ver el decrecimiento de la precisión en función a las operaciones de pre procesamiento aplicadas para construir los modelos con τ =0.3, 0.4, 0.5 y 0.6. En tanto que, la Figura muestra 9.27 como la preparación de datos afecta al recall.

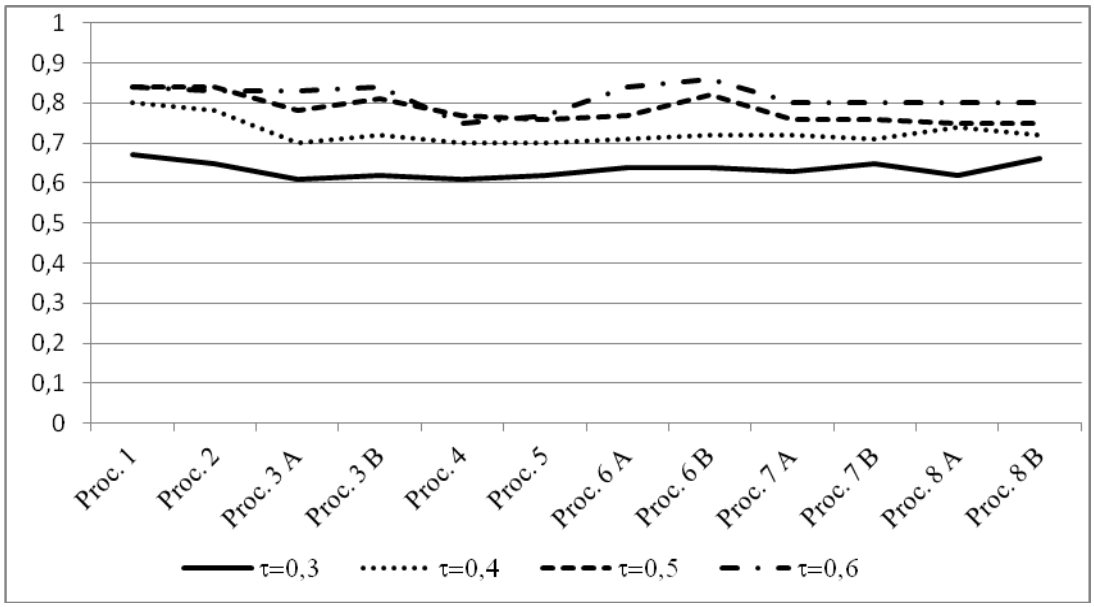


Figura 9.26. Variación de la precisión de los clasificadores construidos con SVM

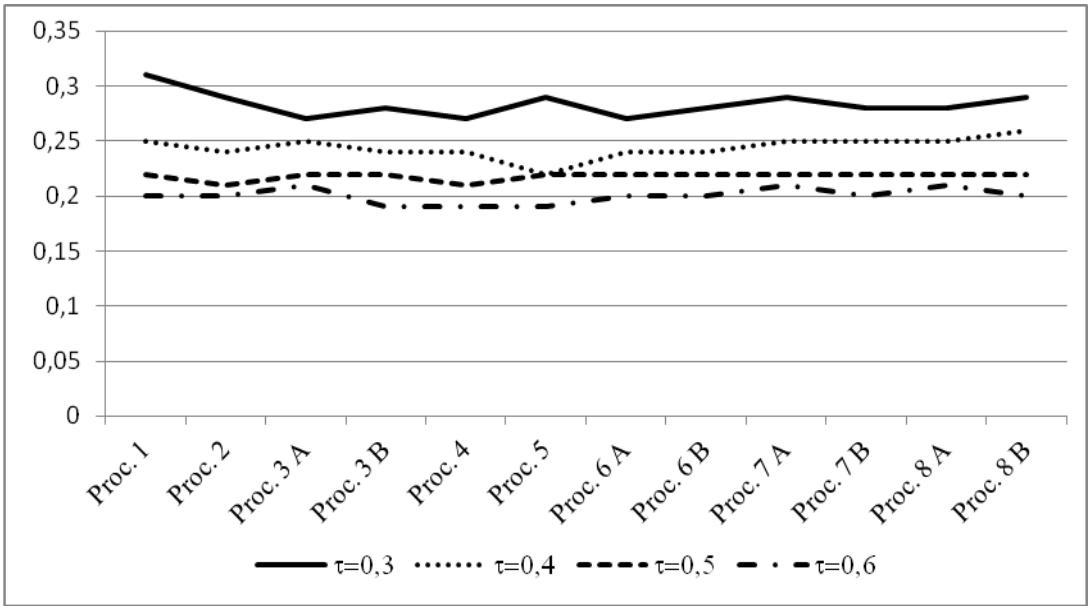


Figura 9.27. Variación del recall de los clasificadores construidos con SVM

Las ganancias en precisión conforme crece el valor del parámetro τ se pueden asociar a dos aspectos. En primera instancia, con el algoritmo SVM los incrementos en el valor de τ llevan a que progresivamente se vayan descartado los clasificadores de los atributos de habilidad con pocos ejemplos en el conjunto de entrenamiento y en consecuencia, se eliminan los errores que se pueden llegar a cometer debido a la incapacidad de generalización en base a una cantidad reducida de ejemplos. En segunda instancia, los clasificadores construidos para los atributos con mayor cantidad de instancias de entrenamiento se vuelven más rígidos en cuanto a las características que deben satisfacer las oraciones para ser agrupadas bajo uno u otro atributo de habilidad. Estos incrementos en precisión que se pueden lograr a medida que crece τ tienen como efecto colateral la reducción de la cantidad de habilidades manifestadas que se puede llegar a reconocer, es decir un recall bajo. Por consiguiente, cuando se dispone de pocos ejemplos de entrenamiento conviene utilizar valores de τ próximos al cero tal como se manifiesta en (Li & Shawe-taylor, 2003).

En cuanto a los criterios de éxito establecidos para el proyecto, los valores de la Tabla 9.17 muestran que el segundo de ellos se satisface ya sea con τ=0.4, τ=0.5 o τ=0.6 porque en todos los casos la precisión es mayor o igual al 0.70. Pero, desafortunadamente ninguna de estas soluciones cumple con el primer criterio de éxito puesto que el porcentaje máximo de habilidades manifestadas que se puede llegar a reconocer es del 26%, un 10% menos de lo logrado con el algoritmo KNN para K=5. Una manera de identificar la mayor cantidad de habilidades manifestadas por los e-tutores pero sin que esto implique tener un margen de error demasiado alto es construyendo los clasificadores con τ=0.3 y la primera combinación de operaciones de pre procesamiento (identificación de token y reconocimiento de sentencias). Con un modelo obtenido de esta manera los máximos valores de π y ρ a los que se puede arribar son 0.67 y 0.31 respectivamente (ver celdas resaltadas en Tabla 9.17). Esto significa que se podría reconocer el 31% de las habilidades manifestadas y que el margen de error del reconocimiento sería de apenas del 33%.

Parámetros / Procesamiento	τ_{+1}=1		τ_{-1}=-1.5	τ_{+1}=10		τ_{-1}=0	τ_{+1}=2		τ_{-1}=-1,5
	π	ρ	F1	π	ρ	F1	π	ρ	F1
Pre procesamiento 1	0,84	0,16	0,27	0,44	0,43	0,43	0,86	0,22	0,34
Pre procesamiento 2	0,84	0,19	0,31	0,42	0,42	0,42	0,82	0,20	0,32
Pre procesamiento 3 A	0,82	0,18	0,30	0,43	0,42	0,43	0,67	0,24	0,36
Pre procesamiento 3 B	0,80	0,15	0,26	0,40	0,40	0,40	0,76	0,18	0,29
Pre procesamiento 4	0,76	0,15	0,25	0,42	0,41	0,42	0,70	0,20	0,31
Pre procesamiento 5	0,81	0,17	0,28	0,43	0,42	0,42	0,67	0,20	0,30
Pre procesamiento 6 A	0,77	0,19	0,31	0,43	0,42	0,43	0,70	0,22	0,33
Pre procesamiento 6 B	0,80	0,15	0,26	0,42	0,41	0,41	0,70	0,21	0,33
Pre procesamiento 7 A	0,79	0,16	0,27	0,42	0,41	0,42	0,74	0,21	0,32
Pre procesamiento 7 B	0,79	0,17	0,28	0,43	0,42	0,43	0,76	0,20	0,32
Pre procesamiento 8 A	0,77	0,16	0,27	0,44	0,43	0,43	0,75	0,21	0,33
Pre procesamiento 8 B	0,80	0,18	0,29	0,42	0,42	0,42	0,76	0,22	0,34

Tabla 9.18. Medidas de medidas efectividad para el algoritmo PAUM

El algoritmo PAUM ofrece un enfoque apropiado para tratar con el problema de reconocimiento automático de habilidades de e-tutores porque utilizando un margen positivo y un margen negativo puede hacer frente a problemas de clasificación donde la cantidad de ejemplos positivos son muy raros o escasos en relación a la cantidad de ejemplos negativos. En el *corpusEntrenamiento* la disparidad entre ejemplos positivos y negativos es crítica porque para algunos atributos de habilidad la cantidad de ejemplos negativos es hasta 50 veces mayor a la cantidad de positivos.

Los niveles de precisión de la Tabla 9.18 son comparables con el estudio sobre clasificación de documentos reportado en (Li *et al.*, 2002). De hecho la precisión para (τ_{+1}=1, τ_{-1}=-1.5) y (τ_{+1}=2 , τ_{-1}=-1,5) es superior a los valores π= 0,751 y π= 0.636 obtenidos por los clasificadores construidos en (Li *et al.*, 2002) para las 90 categorías que forman parte la de la colección Mod-Apte y para las 30 categorías con menor cantidad de ejemplos en el conjunto de entrenamiento. Si se considera que en Mod-Apte (una muestra de la colección Reuters-21578) se destinan 9603 documentos a la fase de entrenamiento y que en *corpusEntrenamiento* se tiene apenas 542 oraciones, es destacable que se hayan logrado valores altos de precisión a partir de un conjunto de entrenamiento mucho más pequeño. Conviene mencionar dos aspectos sobre el estudio de Li y sus colegas, en primer lugar no se reportan los valores de recall, y segundo lugar, los valores de precisión fueron obtenidos mediante validación cruzada 10-fold. Al no contar con información sobre el recall en Mod-Apte no se puede determinar si las cifras de recall de Tabla 9.18 son menores, iguales o mayores a lo que se podría esperar usualmente para los diversos niveles de precisión. Pero lo más importante, no se puede establecer una estrategia a seguir para incrementar los valores de recall sin que ello implique sacrificar la precisión.

Para cumplir con el segundo criterio de éxito definido para el presente estudio podrían construirse modelos con (τ_{+1}=1, τ_{-1}=-1.5) o bien (τ_{+1}=2, τ_{-1}=-1,5), pero en ninguno de estos casos se podría llegar a reconocer el 60% de las habilidades manifestadas por los e-tutores porque el máximo valor de recall alcanzado es 0.24. Utilizar los parámetros τ_{+1}=10 y τ_{-1}=0 permite identificar un poco más del 40% de las habilidades a las que recurren los e-tutores pero el margen de error se mantendría superior al 55 % ya que los valores de precisión oscilan alrededor de 0.40.

Una manera de lograr que el margen de error se mantenga cercano al 30 % y que el recall sea lo más cercano posible al 0.6 es utilizando los parámetros τ_{+1}=2 y τ_{-1}=-1,5 en forma conjunta con la primera combinación de operaciones de pre procesamiento o la primera alternativa de la tercera combinación (ver celdas resaltadas en Tabla 9.18). Con la tercera combinación de operaciones de preparación de datos se pueden tener clasificadores con un margen de error inferior en apenas un 3% al máximo fijado pero el porcentaje de habilidades identificadas sería del 24 %. Esta cifra si bien es bastante inferior al 0.60 que se estableció como mínimo para recall, es el máximo valor que pudo lograrse con las tres combinaciones de τ_{+1} y τ_{-1} seleccionadas para algoritmo PAUM en la sección IX.4.3.3. Por su parte, con la primera combinación de operaciones de pre procesamiento si bien el porcentaje de habilidades reconocidas cae un 2%, se logra un incremento de la precisión superior a un decimo.

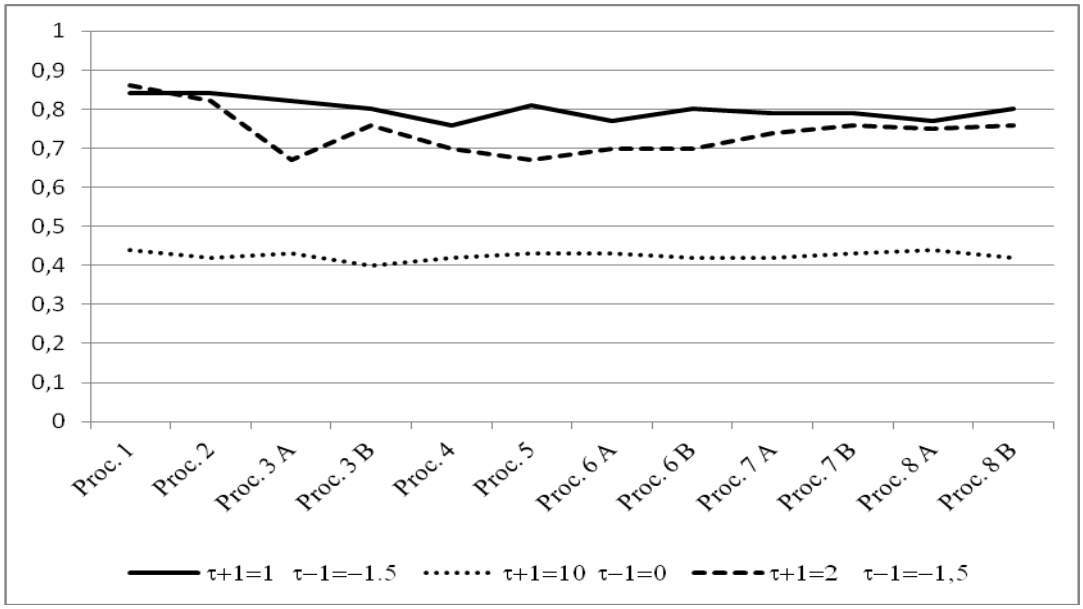


Figura 9.28. Variación de la precisión de los clasificadores construidos con PAUM

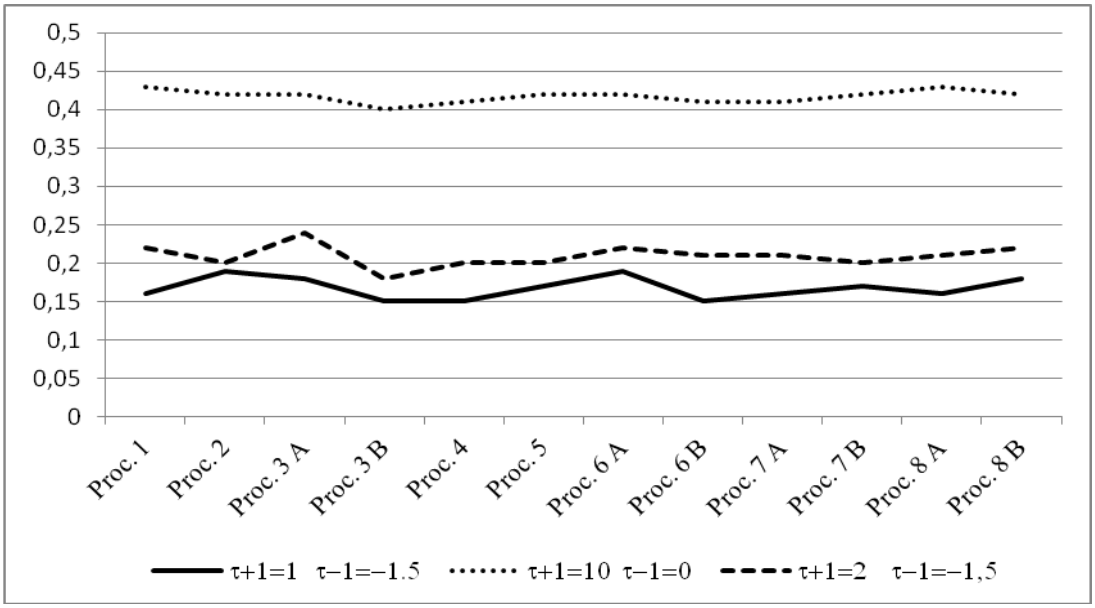


Figura 9.29. Variación del recall de los clasificadores construidos con PAUM

Las Figuras 9.28 y 9.29 reflejan los cambios en la precisión y el recall en función a la preparación de datos para los modelos construidos con PAUM. Todas las operaciones de pre procesamiento ocasionan una reducción de la precisión aunque el recall tiende a mantenerse estable.

IX.4.4.2 Resultados de la evaluación con validación cruzada

Los resultados de la evaluación con hold-out pueden estar influenciados por los elementos que fueron asignados al conjunto de entrenamiento y al conjunto de prueba al momento de descomponer el conjunto de datos. Para comprobar si esto había sucedido en este proyecto se aplico la técnica de validación cruzada de 10-fold para los modelos obtenidos mediante:

- El algoritmo KNN con K=5.
- El algoritmo SVM con τ=0,3.
- El algoritmo PAUM con τ_{+1}=2 y τ_{-1}=-1,5.

Algoritmo / Procesamiento	KNN con K=5			SVM con τ=0,3			PAUM τ_{+1}=2 τ_{-1}=−1,5		
	π	ρ	F1	π	ρ	F1	π	ρ	F1
Pre procesamiento 1	0.297	0.279	0.285	0.579	0.317	0.408	0.720	0.251	0.368
Pre procesamiento 2	0.298	0.287	0.292	0.583	0.318	0.410	0.706	0.252	0.368
Pre procesamiento 3 A	0.339	0.317	0.327	0.602	0.322	0.416	0.700	0.258	0.374
Pre procesamiento 3 B	0.350	0.335	0.342	0.574	0.305	0.396	0.667	0.269	0.380
Pre procesamiento 4	0.300	0.275	0.286	0.559	0.307	0.394	0.668	0.238	0.349
Pre procesamiento 5	0.313	0.282	0.296	0.561	0.304	0.392	0.692	0.234	0.345
Pre procesamiento 6 A	0.363	0.330	0.345	0.592	0.308	0.403	0.698	0.239	0.353
Pre procesamiento 6 B	0.340	0.329	0.334	0.587	0.300	0.394	0.687	0.249	0.361
Pre procesamiento 7 A	0.343	0.323	0.332	0.596	0.318	0.412	0.701	0.242	0.357
Pre procesamiento 7 B	0.338	0.315	0.326	0.594	0.307	0.404	0.692	0.240	0.353
Pre procesamiento 8 A	0.368	0.336	0.350	0.596	0.318	0.412	0.702	0.248	0.363
Pre procesamiento 8 B	0.373	0.340	0.355	0.594	0.307	0.404	0.674	0.249	0.361

Tabla 9.19. Resultados de la evaluación mediante 10-fold

En base a la tabla anterior se puede afirmar que los resultados del método hold-out si están influenciados por la composición de los conjunto de prueba y de entrenamiento. Los valores de precisión arrojados por la validación cruzada de 10-fold son inferiores a los del método hold-out y en algunos casos las diferencias llegan hasta un decimo, es decir que con hold-out los resultados de la métrica precisión son bastante optimistas. En cuanto al recall, para el algoritmo KNN se observa un decrecimiento en sintonía con la precisión, lo cual indica que disminuye el porcentaje de atributos de habilidad identificados correctamente al igual que el porcentaje de habilidades manifestadas que se logra reconocer. Por su parte, para los algoritmos SVM y PAUM la variación del recall es en sentido contrario al de la precisión. Esto significa que crece el porcentaje de habilidades manifestadas que se logra identificar pero que también crece la cantidad de habilidades identificadas erróneamente. En las Figuras 9.30, 9.31 y 9.32 se puede apreciar claramente las diferencias en los valores de precisión y recall que obtuvieron con los métodos hold-out y 10-fold para los modelos construidos con KNN, SVM y PAUM.

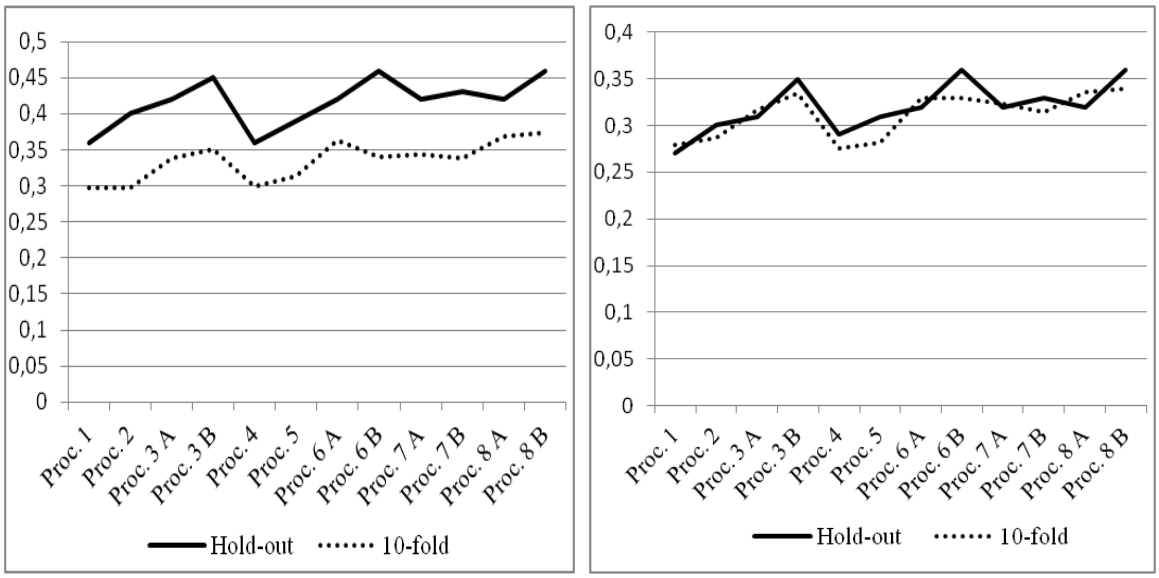


Figura 9.30. El gráfico de la izquierda muestra el cambio la precisión de acuerdo al método de evaluación para los modelos construidos con KNN para K=5. El gráfico de la derecha muestra cómo cambia el recall.

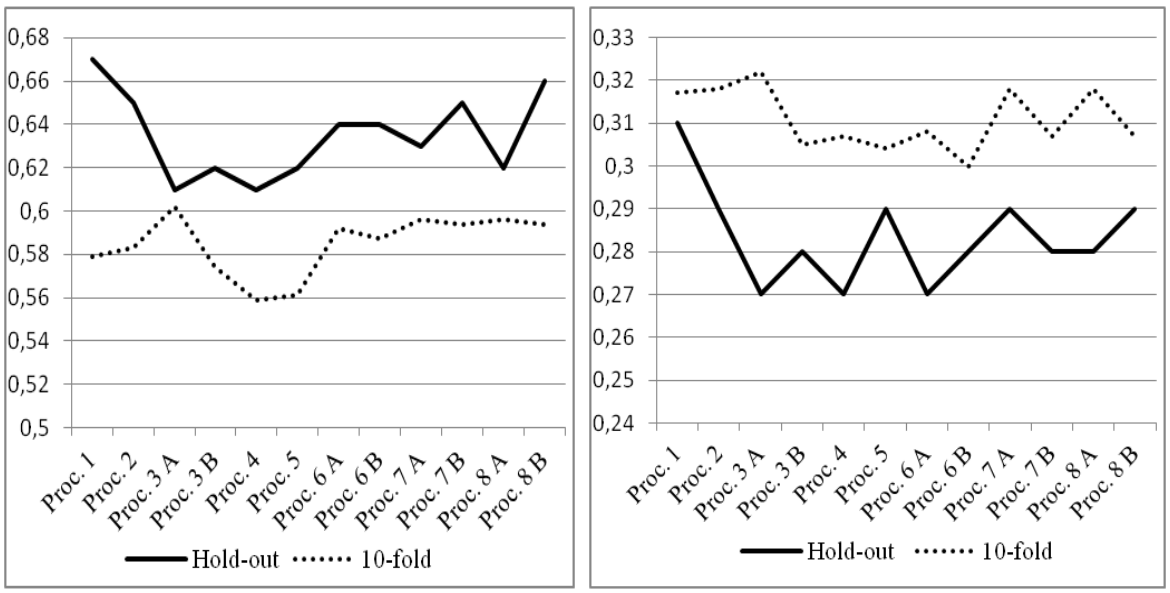


Figura 9.31. El gráfico de la izquierda muestra el cambio la precisión de acuerdo al método de evaluación para los modelos construidos con SVM para τ=0,3. El gráfico de la derecha muestra cómo cambia el recall.

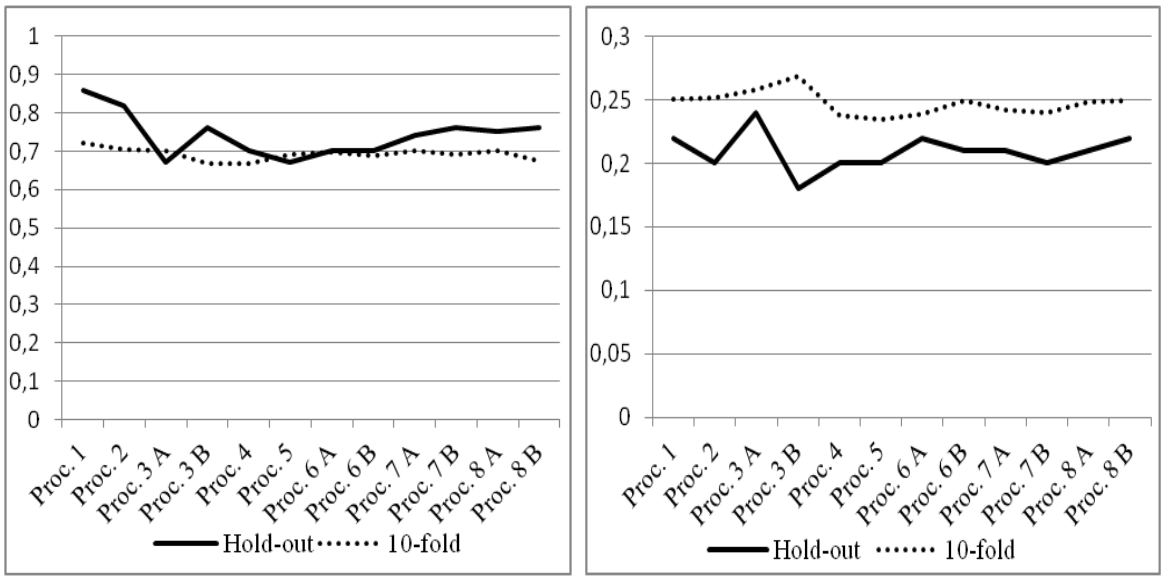


Figura 9.32. El gráfico de la izquierda muestra el cambio de la precisión de acuerdo al método de evaluación para los modelos construidos con PAUM para $\tau_{+1}=2$ $\tau_{-1}=-1{,}5$. El gráfico de la derecha muestra cómo cambia el recall.

Considerando que la validación cruzada protege contra resultados que pueden estar influenciados por una distribución desbalanceada de las categorías en los conjuntos de entrenamiento y prueba, conviene tomar la decisión de cual algoritmo utilizar en base a los valores de precisión y recall mostrados en Tabla 9.19. Recordando que en el apartado 1.1.3 se establecieron los dos siguientes criterios de éxito para el proyecto: "Identificar como mínimo el 60 % de las habilidades sociales manifestadas por los e-tutores en los mensajes publicados en entornos de ACSC" e "Identificar habilidades sociales manifestadas por los e-tutores en los mensajes publicados en entornos de ACSC con un margen de error menor o igual al 30 %", se puede afirmar que estos objetivos compiten entre sí porque mientras uno apunta a lograr valores de recall superiores al 0.60, el otro busca que la precisión sea como mínimo del 0.70. Pero, se sabe que en un clasificador a medida que crece el porcentaje de decisiones de clasificación correctas para una determinada categoría, en relación a la cantidad total de elementos a los que se les asignó la misma categoría (precisión), también decrece el porcentaje de elementos pertenecientes a esa categoría clasificados correctamente, en relación a la cantidad total de elementos que pertenecen la categoría (recall). Así, teniendo en cuenta la relación inversa entre precisión y recall, se debe decidir si se privilegia la precisión sobre el recall o a la inversa.

Supóngase que se dispone de un registro de las interacciones que un e-tutor ha mantenido con sus estudiantes en experiencias de ACSC pasadas y que se debe decidir si se contrata a ese docente para hacerse cargo de un nuevo curso. Cuanto mayor sea la precisión de un modelo mayor será la garantía de que cuando se identifique una habilidad la decisión se toma correctamente, por su parte un valor alto de recall indicaría que se puede reconocer gran parte de las habilidades a las que recurrió el e-tutor. Teniendo en cuenta que las habilidades sociales se vinculan con la coordinación del proceso de interacción entre los estudiantes y que las habilidades que posee un docente son un factor que tiene gran influencia en el éxito o fracaso de las experiencias de ACSC, se considera que debe privilegiarse el cumplimiento del segundo criterio de éxito sobre la satisfacción del primero. El costo que puede pagarse por recurrir a clasificadores que reconozcan muchas habilidades pero sin ofrecer garantías de que el reconocimiento se hizo correctamente es el fracaso de la experiencia de ACSC debido a que se contrato a un e-tutor no capacitado.

De acuerdo a los resultados de la validación cruzada 10-fold, mostrados en Tabla 9.19, el único algoritmo que permite lograr un margen de error menor o igual al 30 % o equivalentemente valores de precisión iguales o superiores a 0.70 es PAUM con los parámetros τ_{+1}=2 τ_{-1}=-1,5. Para la preparación de los datos podría recurrirse a pre procesamiento 1, pre procesamiento 2, pre procesamiento 3 A, pre procesamiento 7 A, o pre procesamiento 8 A. Con la primera combinación de operaciones de pre procesamiento se logra el máximo valor de precisión (π=0,72) y un recall igual a 0,25 en tanto que, con pre procesamiento 3 A se logra el máximo nivel de recall (ρ=0,26) pero la precisión es de 0,70. Puesto que con pre procesamiento 3 A se satisface el segundo criterio de éxito fijado para el proyecto y el primer objetivo queda un poco menos alejado, se prefiere la tercera combinación de operaciones de pre procesamiento sobre la primera. Por consiguiente para el problema de reconocimiento automático de habilidades de e-tutores de ACSC se estima apropiado construir modelos que cumplan las siguientes condiciones:

- Algoritmo: PAUM.
- Parámetros: τ_{+1}=2 τ_{-1}=-1,5
- Pre procesamiento: identificación de *token*, reconocimiento de sentencias y lematización.

Retomando lo expuesto en el apartado IX.4.4.1 sobre los experimentos de Li *et al.* (2002) con la colección de documentos Mod-Apte, se puede afirmar que un modelo construido de acuerdo a los lineamientos anteriores permitiría obtener una precisión apenas inferior al 0,751 de las 90 categorías de la colección Mod-Apte, pero que se encuentra por arriba del 0.636 logrado para las 30 categorías con menor cantidad de ejemplos en el conjunto de entrenamiento.

El tiempo necesario para clasificar nuevas instancias no ha sido considerado como un factor crítico en el presente estudio pero si hubiera sido así, la selección de las operaciones de pre procesamiento hubiera cambiado. Las cifras de la Tabla 9.14 indican que con la primera o la segunda combinación de operaciones de preparación de datos se necesito menos de 11 segundos para clasificar las sentencias de los 202 documentos del conjunto de prueba pero con la primera opción de la tercera combinación se necesita 1 minuto. Si bien no es un tiempo demasiado alto se debe tener en cuenta que podría incrementarse en el caso de que se tenga que clasificar un conjunto mucho grande de documentos, donde cada documento tenga una cantidad de sentencias superior a las 2,68 que en promedio forman parte de los documentos de *corpusPrueba*.

IX.5 FASE 5: EVALUACIÓN

IX.5.1 Evaluar Resultados

Durante el desarrollo del presente estudio se pudo constatar que si se utilizan técnicas de clasificación de minería de textos en el análisis de los mensajes publicados por e-tutores en entornos de ACSC, basados en interfaces no estructuras, se pueden reconocer automáticamente las habilidades sociales manifestadas por los docentes al interactuar con sus estudiantes. De esta manera la minería de textos se convierte en una alternativa viable para conocer el desempeño de los e-tutores sin demandar el tiempo y el esfuerzo requeridos por el análisis manual de las interacciones. Además, con la minería de textos desaparecerían los efectos colaterales sobre la comunicación que tienen las interfaces estructuradas o semi-estructuradas porque los usuarios de herramientas de ACSC podrían comunicarse sin tener restringidas sus alternativas de interacción ni seguir modelos de dialogo pre establecidos.

La construcción de los modelos se realizó con cinco algoritmos pero solo PAUM, SVM, y KNN ofrecen una aproximación viable para el problema de reconocimiento de habilidades de e-tutores puesto que con Naïve Bayes y C4.5 los niveles de recall son iguales a cero con independencia de las operaciones de pre procesamiento aplicadas. Las 8

combinaciones de operaciones de pre procesamiento utilizadas para obtener las características lingüísticas tuvieron un efecto variable en los clasificadores obtenidos. Para KNN a medida que se aplica mayor cantidad de operaciones de pre procesamiento mejores cifras de precisión y recall se consiguen mientras que, para SVM y PAUM las operaciones de pre procesamiento tienen como efecto una reducción de la precisión y el recall.

Además de lo comentado en el párrafo anterior, la construcción de los modelos permitió responder a las dos preguntas derivadas del objetivo de negocio. En lo que respecta a la primera pregunta, los valores en las Tablas 9.16, 9.17, 9.18 y 9.19 indican que para los algoritmos KNN y SVM las características "lemma", "lemma" y "category", o bien "stem" y "category" permiten lograr mejor efectividad en la identificación de las habilidades. En tanto que para el caso del algoritmo PAUM las características "string", "lemma", o "lemma" y "category" son las de mayor utilidad para distinguir entre atributos de habilidad. La respuesta a la segunda pregunta de investigación también la brindan las Tablas 9.16 a 9.19 que contienen los valores de precisión, recall y F_1 arrojados por los métodos hold-out y 10-fold para los clasificadores obtenidos al combinar los algoritmos KNN, SVM, y PAUM con las operaciones de preparación de datos.

A partir del análisis realizado en el apartado IX.4.4 sobre el cumplimiento de los criterios de éxito por parte de los 80 modelos construidos en el transcurso del proyecto (24 con KNN, 32 con SVM, y 24 con PAUM) se llego a la conclusión de que la mejor opción sería construir un modelo con el algoritmo PAUM. Los clasificadores automáticos (uno por cada atributo de habilidad) que se obtienen con PAUM para τ_{+1}=2, τ_{-1}=-1.5, y la tercera combinación de operaciones de preparación de datos permiten lograr un valor de precisión de 0.70 y de recall igual a 0.26. Esto significa que se cumple en un 100% el segundo criterio de éxito definido para el proyecto y que el primer criterio de éxito queda cubierto en un 43%.

En cuanto a los requerimientos del proyecto definidos en IX.1.2.2.1, se puede afirmar que se ha cumplido con dos de los cuatro porque el proyecto finalizo dos semanas antes del tiempo límite fijado y mediante el recurso de procesamiento *ANNIE Sentence Splitter* las oraciones se pueden identificar sin necesidad de que intervenga una persona. Cabe aclarar que la definición de las reglas que permiten adaptar *ANNIE Sentence Splitter* para procesar textos en español y la comprobación del funcionamiento correcta de las reglas no se considero al momento de elaborar el plan del proyecto. La necesidad de que la descomposición en sentencias se realice de modo automático había sido identificada antes de aplicar la metodología CRISP-DM por lo que, al momento de comenzar el proyecto de

minería ya se habían definido y probado las reglas usadas por *ANNIE Sentence Splitter*. De haberse considerado las adaptaciones realizadas a *ANNIE Sentence Splitter* como parte de la aplicación de CRISP-DM, el tiempo necesario para completar todas las actividades se hubiera incrementado en un poco más de 3 meses. Los pasos necesarios para satisfacer los requerimientos referidos a la facilidad de comprensión de los resultados de la clasificación se abordarán en la sección en IX.5.3.

Como evaluación final se puede decir que los resultados conseguidos en el transcurso del proyecto son muy positivos puesto que:

- Se cumplió con todos los requerimientos establecidos.
- Se respondió a todas las preguntas de investigación.
- Se pudo cumplir uno de los criterios de éxito y se logro una cobertura del 43% para el otro.
- Se logró comprender de que manera las operaciones de pre procesamiento afectan a los algoritmos de aprendizaje y como la incidencia cambia de acuerdo al valor de los parámetros propios de cada algoritmo.

IX.5.2 Revisar el Proceso

Una síntesis de todos los pasos considerados y aplicados a lo largo del presente proyecto se muestra en el Figura 9.33.

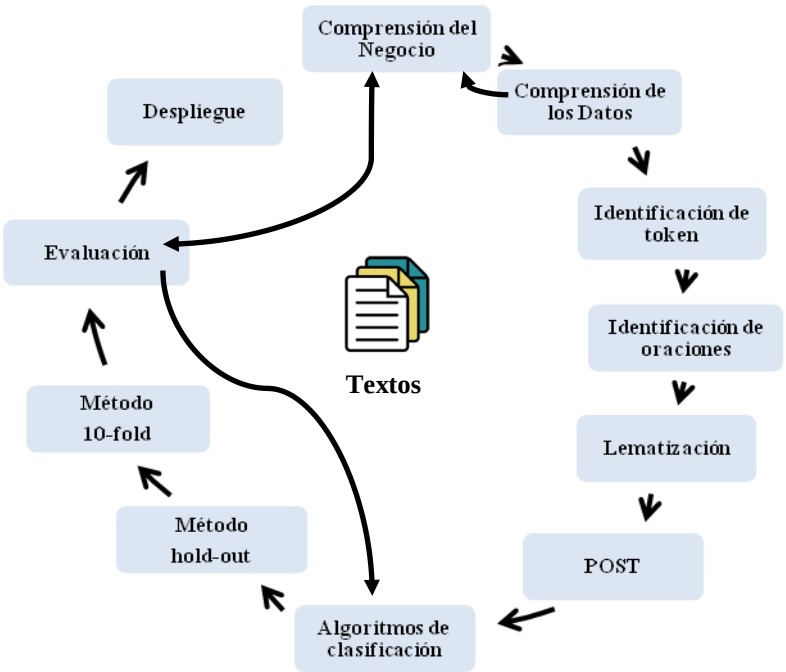


Figura 9.33. Síntesis de la metodología usada en el proyecto

De la fase de entendimiento del negocio lo más destacable es la comparación realizada entre herramientas de minería de textos, que llevo a seleccionar a GATE ya que cuenta con documentación más completa y una diversidad de recursos que permiten adaptar la herramienta a las necesidades de cada proyecto.

En el marco de la fase de entendimiento de datos, luego de haber descartado los archivos donde no se habían identificado habilidades sociales, se aplicó una conversión de formato. Los datos usados en el proyecto estaban almacenados en cuatro formatos de archivo diferentes (pdf, htm, rtf, y doc) y para evitar problemas de compatibilidad se optó por convertirlos en archivos texto plano donde los caracteres se almacenen en base al sistema de codificación Unicode. Tratando de mejorar la comprensión del conjunto de datos se analizó la frecuencia de ocurrencia de las palabras y la cantidad de oraciones en la que manifestaba cada atributo de habilidad. El examen de la frecuencia de las palabras permitió notar que los nombres (ya sea de personas o lugares) no eran características lingüísticas de relevancia para construir los modelos. En tanto que, la frecuencia de ocurrencia de los atributos de habilidad reveló que había 10 con cero ocurrencias y otros 15 con 10 o menos ocurrencias. Para los atributos con cero ocurrencias quedó claro que no se podrían construir clasificadores porque no se disponía de ejemplos a partir de los cuales inferir las características que las sentencias debían reunir para ser catalogadas bajo alguno de los 10 atributos.

Al inicio de la etapa de preparación de datos se estableció que para ser incluidos en el estudio los atributos de habilidad tendrían que haberse manifestado como minino 5 veces en el conjunto de datos. Esto llevó a reducir a 21 los atributos de habilidad con los que se hizo el estudio. Si bien se denominaron operaciones de pre procesamiento a la identificación de token, el reconocimiento de oraciones, normalización, POST y supresión de stop words, las actividades de preparación de datos necesarias para construir los modelos también incluyeron:

- Modificación de archivos de texto plano para corregir los problemas de calidad de los datos como consecuencia de la presencia de errores de ortografía, emoticones, y oraciones formadas solo por el nombre y/o apellido del e-tutor.
- Supresión o eliminación de archivos duplicados, archivos formados solo por nombres de personas o lugares, y archivos cuyos atributos de habilidad tenían una frecuencia inferior a 5.
- Creación del *corpusMensajes* para integrar los archivos de texto plano.

- Definición del uni-grama de modo tal que cada sentencia se analice en términos de las anotaciones del tipo "Token" que la componen.

La eliminación de *stop words* se hizo teniendo en cuenta la característica "string*"* de las anotaciones del tipo *Token* pero podría haberse basado en la característica "lemma*"*. Si la supresión de *stop words* se hiciera considerando el "lemma" se eliminaría una anotación del tipo "Token" con la característica "string" igual a "conseguimos" puesto que el "lemma" de "conseguimos" es "conseguir" y "conseguir" forma parte de la lista de abreviaturas de R/tm.

En la fase de modelado se destinaron 593 oraciones al conjunto de entrenamiento y 202 al conjunto de prueba para evaluar los modelos con el método hold-out. De manera complementaria a hold-out se aplicó el método 10-fold porque permite asegurar que los resultados de la evaluación no estén influenciados por la composición de los conjuntos de entrenamiento y de prueba. Como parte del modelado también se definieron los archivos de configuración necesarios para construir los clasificadores con KNN, SVM, PAUM, C4.5 y Naïve Bayes. Para cada una de las operaciones de pre procesamiento se mostraron las adaptaciones que tienen que realizarse a los archivos de configuración. Seguidamente, se realizó un examen del efecto de la preparación de datos sobre las medidas de efectividad, el tiempo de construcción de los modelos y el tiempo de clasificación. Finalmente, la consideración de las métricas de efectividad, arrojadas tanto por el método hold-out como 10-fold, permitió determinar que con τ_{+1}=2, τ_{-1}=-1.5, y la tercera combinación de operaciones de pre procesamiento se logran los clasificadores que mejor cumplen con los objetivos de negocio.

La etapa de evaluación trasciende los aspectos técnicos de la efectividad de los clasificadores (precisión, recall, F1) considerados en la fase de modelado, para transformarse en una evaluación más global donde se consideran los requerimientos, preguntas, y objetivos de negocio definidos para el proyecto. En tanto que, el apartado destinado a la etapa de despliegue se refiere a las dificultadas que tuvieron que resolverse y las lecciones aprendidas.

A pesar de que los resultados a los que se arribó en el proyecto son alentadores en lo que se refiere al uso de la minería de textos para identificar las habilidades de los e-tutores, queda claro que se debería seguir trabajando para encontrar una manera de mejorar los niveles de recall de los clasificadores pero sin que esto implique sacrificar precisión. En este sentido, un posible curso de acción sugerido en (Li *et al.*, 2002) consiste

en ajustar los parámetros de los algoritmos KNN, SVM, y PAUM a la cantidad de ejemplos que se disponga de cada atributo de habilidad.

IX.5.3 Determinar Próximos Pasos

En vistas de que se pudo cumplir tanto con los criterios de éxito como con los requerimientos, y de que se obtuvieron respuestas para las preguntas de investigación, no se considera necesario realizar nuevas iteraciones de la metodología CRISP-DM. Antes de proceder con la fase de despliegue se tendrían que completar dos acciones:

- Crear una aplicación que utilice el modelo construido con PAUM para clasificar contribuciones de los e-tutores nunca antes procesadas. Esta aplicación trabajará con dos conjuntos de anotación denominados "preProcesamiento" y "Habilidades e-tutor". En "preProcesamiento" se crearán todas las anotaciones necesarias para poder clasificar una oración en tanto que "Habilidades e-tutor" contendrá una anotación del tipo "Sentence" por cada atributo de habilidad identificado. Cada anotación del tipo "Sentence" tendrá a su vez una característica denominada "atributo_hab" cuyos valores son las abreviaturas del nombre de los atributos de habilidad. De esta manera una persona que use la aplicación no necesita estar familiarizado con la minería de textos ni tampoco tener un conocimiento a fondo sobre las habilidades de los e-tutores, lo único que necesita es saber el significado de cada abreviatura o tener a mano una lista con el significado.
- Utilizar el recurso "GATECloud.net" de manera que la aplicación que usa el modelo creado con PAUM pueda ejecutarse en una computadora distinta a la usada durante la construcción de los clasificadores. Para probar que la aplicación permite reconocer atributos de habilidad social en una computadora donde no se tiene instalado GATE se usará la notebook mencionada en el inventario de recursos. La necesidad de que los clasificadores automáticos puedan ejecutarse en una computadora sin GATE no está documentada pero dado que GATE cuenta con la facilidad "GATECloud.net" se considera pertinente darle al usuario esta posibilidad.

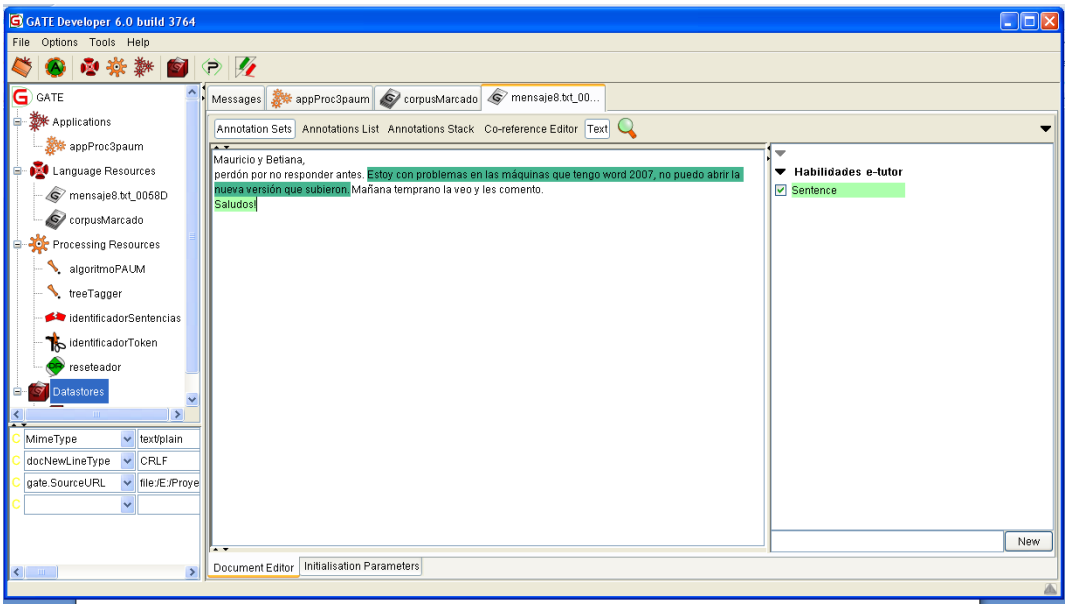


Figura 9.34. Interfaz de usuario con el resultado del reconocimiento automático de habilidades

La Figura 9.34 muestra el resultado de ejecutar la aplicación que usa los clasificadores construidos con PAUM tal como lo vería el usuario. Los recursos de procesamiento usados para crear la aplicación son *GATE Unicode Tokeniser*, *ANNIE Sentence Splitter*, *Generic Tagger*, *Tree Tagger*, *Batch Learning PR*, y *Document Reset PR*. Para el usuario será visible únicamente el conjunto de anotación "Habilidades e-tutor" porque una vez clasificadas todas las oraciones se suprime "preProcesamiento" mediante *Document Reset PR*. La aplicación fue incluida en el directorio *Mineria_de_Textos\Aplicación* del CD que se adjunta.

IX.6 FASE 6: DESPLIEGUE

IX.6.1 Revisar el Proyecto

De acuerdo a la metodología CRISP-DM la revisión del proyecto consiste en hacer un resumen de las experiencias vividas, lecciones aprendidas, errores y aciertos cometidos durante el desarrollo del proyecto. Entre las cuestiones que tuvieron que ser resueltas y que provocaron el retraso en las actividades planificadas se tiene:

- La creación de la gramática JAPE a ser utilizada por *Jape Transducer* para eliminar las *stop words.*
- El calcula de precisión, recall, y F_1 para los modelos construidos. Inicialmente se tenía pensando usar el recurso *Corpus Bechmark* que compara las anotaciones entre dos corpus. Sin embargo, por causas que no se pudieron establecer, en todo momento este recurso arrojó como resultado el valor "1.0" para las tres medidas de efectividad, indicando además que no había coincidencias entre el corpus que

contenía las habilidades identificadas por expertos humanos y el corpus que con las habilidades reconocidas por los clasificadores automáticos. Por este motivo finalmente se utilizó *Corpus Quality Assurance.*

- *Tree Tagger* necesita de librerías propias de Linux que no están disponibles en ninguna de las versiones del sistema operativo Windows. Para poder usar *Tree Tagger* en Windows se tuvo que instalar Cygwin que actúa como una capa API Linux proveyendo gran parte de la funcionalidad de Linux. Pero además fue necesario crear en el directorio de instalación de GATE un archivo de configuración de nombre "build.properties.xml" de manera que GATE pueda usar las librerías de Cygwin.
- En las primeras pruebas realizadas con *Tree Tagger* no se identificaba el "lemma" de las anotaciones del tipo "Token" y en la documentación de GATE no había comentarios al respecto. Luego de varios intentos se descubrió que la causa del problema era el valor del parámetro "encoding" del recurso *GenericTagger*. Si al parámetro "encoding" se le asigna un valor distinto a "ISO-8859-1" *Tree Tagger* no funciona correctamente. El valor de "encoding" se refiere al sistema de codificación usado en las anotaciones temporales creadas por *Tree Tagger* por lo que no tiene necesariamente que coincidir con el sistema de codificación de los documentos analizados, que para el presente estudio es Unicode.

La inclusión del atributo de habilidad OAH puede haber sido una decisión equivocada porque reúne oraciones muy diferentes entre sí. Este atributo se uso durante el estudio de análisis de contenido descripto en el capítulo VII de modo de agrupar todas las sentencias en las que los expertos no identificaban alguno de los 42 atributos de habilidad social. De allí que es difícil que en la fase de aprendizaje los clasificadores puedan determinar el patrón o características en común de todas las oraciones catalogadas como OAH. De hecho con los 80 modelos construidos se pudo llegar a reconocer correctamente tan solo 2 de los 14 ejemplos etiquetados con OAH en el conjunto de entrenamiento. Por consiguiente las cifras de recall para este atributo de habilidad se mantuvieron siempre debajo del 0.15 y esto a su vez influyó en el valor del recall calculado para la totalidad de los atributos de habilidad.

Por último, en los experimentos realizados se comprobó que cuando en el archivo de configuración de los algoritmos de aprendizaje el parámetro **thresholdProbabilityClassification** toma valores próximos a "1.0" se produce un

aumento de la precisión y una disminución del recall. En tanto, que si el valor del parámetro se aproxima a cero, se produce un aumento del recall pero disminuye la precisión. Esto significa que el valor de **thresholdProbabilityClassification** puede ser ajustado para incrementar la precisión o el recall en función de cuál de estos dos aspectos del desempeño de los clasificadores sea de mayor relevancia para el problema entre manos.

CAPÍTULO X
CONCLUSIONES

Para facilitar la comprensión de las conclusiones a las que se arribó producto del desarrollo de la diversidad de tareas necesarias para finalizar este trabajo de investigación, se decidió formularlas considerando tres áreas troncales: las habilidades de los e-tutores, el análisis de contenido en las interacciones, y la minería de textos. Finalizando la sección se enuncian algunas líneas futuras de investigación.

a) <u>Habilidades de los e-tutores</u>

La clasificación de habilidades planteada en este trabajo de investigación se creó con la intención de mejorar la comprensión de la tarea de los e-tutores en entornos de ACSC. Esta es una cuestión primordial para el éxito de las experiencias de ACSC porque no basta con que los docentes tengan experiencia en los contenidos académicos del curso y en cómo usar la tecnología, además es necesario que comprendan cuáles son sus roles y las habilidades necesarias para desempeñarlos de manera efectiva. Si bien existen otros factores que influyen en el éxito de una experiencia de ACSC, tales como las contribuciones de los miembros del grupo y las habilidades de colaboración manifestadas por los alumnos, el rol de los docentes es fundamental. La taxonomía de habilidades desarrollada durante la realización de este trabajo (descripta en el Capítulo IV) originó dos artículos que fueron aceptados en dos eventos internacionales. Primero, en el *Segundo Congreso Internacional de Educación en Ciencia y Tecnología* (Santana *et al*., 2011) y luego, en el *Cuarto Simposio Internacional de investigación "La investigación en la universidad: experiencias innovadoras en investigación aplicada"* (Santana *et al*., 2012). Es de destacar el interés demostrado por los comités académicos evaluadores y por las audiencias de los eventos citados, resaltándose la valoración que de esta taxonomía realizó la comunidad científica especializada por cuanto sostienen que este trabajo de investigación aporta conocimientos hasta el momento inexistentes o poco formalizados en el tema específico de las habilidades de los e-tutores de ACSC. No obstante, de ninguna manera puede considerarse que la clasificación de habilidades propuesta es de carácter definitivo sino que, por ser el fruto de un trabajo interdisciplinario con la pedagogía, se

piensa que la misma podrá ser tomada como punto de partida por los profesionales del área interesados en realizar futuras investigaciones que permitan su validación científica.

A lo largo del proceso de elaboración del esquema de habilidades presentado quedó en evidencia que la tarea de los e-tutores de ACSC es multifacética ya que requiere que se tengan en consideración factores sociales, administrativos, pedagógicos, técnicos, comunicacionales, y de evaluación. Sin embargo, las diferentes categorías de habilidad no deberían ser consideradas en forma separada puesto que como se pudo observar existen sub habilidades que están presentes en varias categorías de habilidad, como ocurre, por ejemplo, con la sub habilidad de preguntar. Otro claro ejemplo de las vinculaciones existentes entre diferentes sub habilidades y atributos de habilidad, es la sub habilidad de desarrollar, soportar y mantener una comunidad de aprendizaje, ya que en una comunidad de aprendizaje para lograr una interacción armoniosa entre los estudiantes se necesitan de atributos de habilidad vinculados con: tratar con comportamientos de discusión inapropiados, manejar contribuciones inapropiadas, manejar conflictos, y motivar a los estudiantes.

Retomando el tema de la validación de la clasificación de habilidades planteada, los resultados del análisis de contenido son auspiciosos porque en las experiencias de ACSC analizadas los expertos humanos reconocieron el 75% de los atributos de habilidad pertenecientes a la categoría social. Esto significa que en la práctica efectivamente se manifiestan los comportamientos y las conductas contempladas en la taxonomía de habilidades elaborada. Si bien en los registros de interacciones analizadas no se identificó el 100% de los atributos de la habilidad social, se debe tener en cuenta que existen aspectos tales como características de la tarea, acciones y necesidades de los estudiantes, la tecnología utilizada, y rasgos propios del e-tutor que influyen en la forma en que tiene lugar la intervención del docente. Por lo tanto, como se afirmó en otras secciones de este trabajo, no es obligatorio que se manifiesten todas las habilidades ya que cada e-tutor recurre a las habilidades que demandan las circunstancias, y puede incluso que alguna de las categorías, sub habilidades, o atributos de habilidad que forman parte de la clasificación planteada no sean requeridos.

En el planteamiento del problema se mencionó que disponer de herramientas de software que soporten el aprendizaje en grupo no garantiza que los estudiantes colaboren, porque no son las características de la tecnología utilizada sino la forma en que ésta se utiliza lo que soporta la colaboración. En la taxonomía de habilidades descripta en el Capítulo IV, las habilidades tecnológicas son escasas en relación a las categorías social,

pedagógica, y administrativa, pero esto de ninguna manera implica que el aspecto técnico de la labor de un e-tutor deba ser menospreciado, puesto que si los estudiantes tienen problemas para comunicarse o no saben cómo usar la tecnología las actividades grupales se verán perjudicadas.

b) Análisis de contenido de las interacciones

Con la aplicación de la técnica de análisis de contenido, se pudo comprobar que es correcto considerar como unidad de grabación a las oraciones o sentencias que forman parte de los mensajes de los e-tutores, porque en cada contribución se pueden manifestar múltiples habilidades. Asimismo, es conveniente usar instrucciones de codificación en la forma de esquema de decisión para hacer corresponder atributos de habilidad a las unidades de grabación ya que los codificadores pudieron seguir el esquema de decisión incluido en el Anexo A sin grandes dificultades.

El valor de 94 % calculado para el porcentaje de acuerdo entre los dos codificadores que participaron del estudio de análisis de contenido, indica que las unidades de grabación que forman parte del Anexo A pueden servir de base para formar una futura colección de unidades de grabación mucho más amplia. Esta colección podría incluir solamente atributos de la habilidad social, o bien atributos de habilidad de todas las categorías del esquema de clasificación planteado en este trabajo de investigación, a fin de ser utilizada por algoritmos de aprendizaje de máquina en la construcción de clasificadores automáticos. Teniendo en cuenta que los datos de fiabilidad se generaron bajo la condición test-test, es de esperar que los atributos de habilidad social asignados a las unidades de grabación (incluidas en el Anexo A) sean identificados por otros analistas interesados en las habilidades sociales de e-tutores de ACSC. De esta manera, las unidades de grabación usadas en el análisis de contenido representan correctamente la función de asignación de categorías (es decir, representan conocimiento experto sobre como asignar las categorías) que los algoritmos de aprendizaje de máquina buscan aproximar.

c) Minería de textos

Durante el desarrollo del presente trabajo de investigación se pudo demostrar que las técnicas de clasificación de minería de textos permiten reconocer las habilidades sociales manifestadas por los e-tutores cuando se aplican a los mensajes publicados por los mismos en entornos de ACSC basados en interfaces no estructuras.

La evaluación de la efectividad de los clasificadores construidos con los cinco algoritmos de aprendizaje se hizo en base a los métodos hold-out y 10-fold de modo de asegurar que los resultados no estuvieran influenciados por la composición de los conjuntos de entrenamiento y de prueba. Los resultados de la evaluación mostraron que los algoritmos PAUM y SVM con márgenes desiguales permiten construir los clasificadores automáticos de mejor desempeño (considerando las métricas precisión y recall) en la identificación de atributos de habilidad. Si se tiene en cuenta que en la colección de datos usada la cantidad de instancias positivas de cada atributo de habilidad es menor a la cantidad de instancias negativas, es lógico que PAUM y SVM con márgenes desiguales tengan mejor desempeño puesto que son algoritmos especialmente diseñados para tratar con colecciones de documentos donde existe un desbalance entre las instancias negativas y las instancias positivas. Las diferencias más notables se dan entre el algoritmo KNN y PAUM, puesto que tanto con el método 10-fold como con hold-out, los valores de precisión de los modelos obtenidos con PAUM duplican a la máxima precisión que puede lograrse con KNN. En cuanto a los algoritmos Naïve Bayes y C4.5, se pudo comprobar que no proporcionan un enfoque apropiado para tratar con el problema de reconocimiento automático de habilidades de e-tutores de ACSC porque los valores de recall iguales a cero señalan que los clasificadores construidos no tienen la capacidad de asignar atributos de habilidad a los mensajes publicados por los e-tutores.

Si se considera el método de evaluación hold-out los niveles de precisión que se pueden lograr tanto con SVM con márgenes desiguales, como con PAUM, serían superiores al 0.80, en cambio para el método 10-fold los valores de precisión son superiores al 0.7. A pesar de que estas cifras indican que, ya sea con SVM con márgenes desiguales o PAUM, se pueden lograr buenos niveles precisión, aún queda trabajo por hacer puesto que los valores de recall con SVM con márgenes desiguales rondan el 0.30 en tanto que con PAUM se aproximan al 0.25. En trabajos de investigación futuros, con el propósito de incrementar el valor calculado para recall pero sin sacrificar la precisión, se podría buscar una colección de documentos que contenga mayor cantidad de ejemplos para cada atributo de habilidad, o en su defecto, ajustar los parámetros de los algoritmos en base a la cantidad de instancias de cada atributo de habilidad. La primera estrategia se deriva del hecho de que en la colección de mensajes de e-tutores empleada existen atributos de habilidad con apenas cinco instancias, pero, Feldman y Sanger (2007) sugieren que para lograr un buen desempeño los algoritmos de clasificación deberían trabajar con al menos 30 ejemplos de cada categoría. En particular, el algoritmo KNN mostró mayor

sensibilidad ante la cantidad de ejemplos de cada atributo de habilidad puesto que, cuando el parámetro K tomaba valores superiores a 5, tanto la precisión como el recall de los clasificadores disminuyó. Los algoritmos SVM con márgenes desiguales y PAUM no mostraron el mismo comportamiento ante la cantidad de instancias de los atributos de habilidad, puesto que como ya se mencionó los mismos se crearon teniendo en cuenta la existencia de pocos ejemplos positivos de entrenamiento. La segunda estrategia para mejorar el desempeño de los algoritmos se deriva del trabajo de Li *et al.* (2002) y consiste en ajustar los parámetros de los algoritmos KNN, SVM, y PAUM a la cantidad de ejemplos que se disponga de cada atributo de habilidad.

A pesar de que se necesita seguir trabajando para mejorar el valor calculado para recall, los valores de precisión conseguidos con SVM y PAUM son más que destacables, considerando que la cantidad de ejemplos utilizada para el entrenamiento de los algoritmos es muy inferior a la cantidad empleada en otros trabajos de investigación donde se recurre a la minería de textos para tareas de clasificación. Así por ejemplo, en el trabajo de Li *et al.* (2002), con la colección Mod-Apte, se reportaron valores de precisión de 0.751 en los clasificadores construidos para las 90 categorías que forman parte la de la colección. Cabe aclarar que en la colección Mod-Apte se destinan 9603 documentos para el entrenamiento, mientras que en el presente trabajo de investigación los valores de precisión del 0.7 del algoritmo PAUM se lograron con apenas 542 oraciones en el conjunto de entrenamiento.

Al comenzar el trabajo de investigación se creía que las operaciones de pre procesamiento permitían mejorar la efectividad de los clasificadores automáticos pero, como se pudo verificar al aplicar la metodología CRISP-DM, el efecto que tienen las operaciones de pre procesamiento depende del algoritmo de aprendizaje. Con el algoritmo KNN, a medida que se aplica una mayor cantidad de operaciones de pre procesamiento se obtienen mejores valores de efectividad (precisión y recall) puesto que, las operaciones de pre procesamiento reducen la diferencias entre los documentos, y el algoritmo KNN se basa en el cálculo de la similitud entre los documentos a ser clasificados y los documentos del conjunto de entrenamiento. En cambio, para los algoritmos SVM y PAUM las operaciones de preparación de datos causan la reducción tanto de la precisión como del recall, esto se debe a que estos algoritmos no calculan la similitud entre las instancias a clasificar y las del conjunto de entrenamiento. En los modelos construidos con SVM, al igual que con PAUM, se les asigna un peso a las características de todas las sentencias del conjunto de entrenamiento y al momento de clasificar una nueva sentencia se tiene en cuenta el peso de las características que la representan. Esto implica que cuanto mayores

diferencias existan entre las oraciones pertenecientes a cada atributo de habilidad, más fácilmente se podrá establecer el atributo de habilidad que le corresponde a una oración a ser clasificada a partir del cálculo del peso de sus características.

Para poder aplicar minería de textos la mayor carga de trabajo estuvo representada por las adaptaciones realizadas a GATE, de modo tal de poder identificar oraciones en textos en español e implementar la operación de eliminación de *stop words*. En particular, las adaptaciones realizadas a *ANNIE Sentence Splitter* requirieron mucho tiempo y esfuerzo debido a que las reglas de puntuación aplicables a textos en español son completamente diferentes a las reglas válidas para textos en ingles. Sin embargo, dado que las reglas seguidas para descomponer los textos en oraciones son las especificadas por la Real Academia Española, la gramática creada para *ANNIE Sentence Splitter* puede aplicarse sin inconvenientes a cualquier texto en español independientemente de su país de procedencia. De la misma manera, la gramática JAPE para la implementación de la eliminación de *stop words* podría ser reusada en proyectos de minería de textos donde se trabaje con textos en español. Si se considera que JAPE se basa en expresiones regulares es lógico suponer que las reglas de creadas para *ANNIE Sentence Splitter* podrían ser usadas incluso en proyectos donde no se utilice a GATE como herramienta de minería de textos, siempre y cuando se adapten las expresiones regulares de *ANNIE Sentence Splitter* al software de minería de textos empleado. Otro aspecto a destacar de JAPE es que al basarse en expresiones regulares simplifica considerablemente la lógica de decisión necesaria para determinar cuando un signo de puntuación marca el fin de una oración, esto se debe a que evita combinaciones complejas de instrucciones condicionales y cíclicas.

La metodología CRISP-DM proporciona un marco de trabajo apropiado para desarrollar proyectos de minería de textos, aun cuando existan diferencias en cuanto a la terminología y las tareas propias de la minería de datos y la minería de textos. Las mayores diferencias surgen en la fase de preparación de los datos. En la minería de datos se asume que los datos han sido almacenados en un formato estructurado, por lo cual el pre procesamiento consiste casi exclusivamente en la depuración de los datos y la creación de tablas a partir de la unión de otras dos o más tablas. En cambio, en minería de textos el foco de las operaciones de pre procesamiento consiste en crear representaciones estructuradas o semi estructuradas de una colección de documentos de texto no estructurado. Como consecuencia de lo anterior, es lógico que cuestiones tales como la eliminación de *stop words*, lematización, POST, etc., no sean consideradas explícitamente por la metodología CRISP-DM. En relación a las adaptaciones para usar CRISP-DM en

proyectos de minería de textos, propuestas por Magalhães (2002) en su tesis de maestría, se verificó que son de utilidad ya que por ejemplo, el análisis de la frecuencia de ocurrencia de los atributos de habilidad permitió reconocer en una etapa temprana del proyecto que no se podrían construir clasificadores para los atributos con cero instancias.

Considerando que los clasificadores automáticos pueden identificar habilidades en los mensajes de los e-tutores en cuestión de segundos o minutos, dependiendo de la cantidad de mensajes procesados y de las operaciones de pre procesamiento aplicadas, y que con el algoritmo PAUM puede lograrse una precisión del orden del 0.70, se puede afirmar que la minería de textos se convierte en una alternativa viable para conocer el desempeño de los e-tutores de ACSC sin demandar el tiempo y el esfuerzo requeridos por el análisis manual de las interacciones. Además, con la minería de textos desaparecerían los efectos colaterales sobre la comunicación que tienen las interfaces estructuradas o semi-estructuradas porque, los usuarios de herramientas de ACSC podrían comunicarse con interfaces basadas en texto libre sin tener restringidas sus alternativas de interacción, ni seguir modelos de diálogo pre establecidos. En el futuro, para construir clasificadores automáticos para todas las categorías de habilidad de la clasificación creada en el desarrollo del presente trabajo, lo único que se necesitaría sería una colección de mensajes o contribuciones docentes clasificadas a mano por expertos humanos.

Líneas Futuras de Investigación

En base a los resultados obtenidos en esta investigación se piensa que la minería de textos podría ser utilizada como parte de una herramienta de software de entrenamiento de habilidades para e-tutores de ACSC. Una aplicación de este tipo tendría que analizar las interacciones registradas en grupos de aprendizaje colaborativo de manera de identificar conflictos que requieran la intervención de los e-tutores para resolverse. En estos casos, el sistema se encargaría de sugerir a los e-tutores las acciones adecuadas a llevar a cabo para mejorar el aprendizaje grupal y simultáneamente entrenar de manera personalizada las habilidades que no hayan manifestado adecuadamente.

REFERENCIAS BIBLIOGRÁFICAS

Adell J. y Sales A. (1999) *El profesor online: elementos para la definición de un nuevo rol docente*. Procedente de EDUTEC'99. Universidad de Sevilla, 1999.

Aranha C. y Passos E. (2008) Automatic NLP for Competitive Intelligence. En: do Prado H. A. y Ferneda E. (eds) *Emerging Technologies of Text Mining: Techniques and Applications*. Information Science Reference, pp. 54-76.

Athaudal R., Tissera M., y Fernando C. (2009) Data Mining Applications: Promise and Challenges. En: Ponce J. y Karahoca A. (eds) *Data Mining and Knowledge Discovery in Real Life Applications*. Croacia, In-Tech, pp. 201-214.

Awouters V. y Jans S. (2009) E-learning competencies for teachers in secondary and higher education. *International Journal of Emerging Technologies in learning* (iJET), 4 (2), 58-60.

Barros B. y Verdejo M. F. (2001) Entornos para la realización de actividades de aprendizaje colaborativo a distancia. *Inteligencia Artificial: Revista Iberoamérica de Inteligencia Artificial*, 5 (12), 39-49.

Barker P. (2002) *Skill set for online teaching*. Procedente de Word Conference on educational, Multimedia, Hipermedia, & Telecomunicatios. Denver, USA.

Barker P. (2010) Introduction. En: Barker Philip y Van Schaik P. (eds) *Electronic Performance Support. Using Digital Technology to Enhance Human Ability*. Ashgate Gower, pp. 3-29.

Benito Osorio D. (2009) Aprendizaje en el entorno del e-learning: estrategias y figura del e-moderador. *Revista de Universidad y Sociedad del Conocimiento*, 6 (2).

Berge Z. L. (1995) The Role of the Online Instructor/ Facilitator. *Educational Technology*, 35 (1), 22-30.

Blackburn K. (2004) *Is online teaching for you?* Accedido Marzo 2013, Disponible en http://knowledgetree.flexiblelearning.net.au/edition06/download/ Blackburn.pdf

Bláquez Entonado F. y Alonso Díaz L. (2005) Apuntes para la formación del docente de e-learning. *Enseñanza & Teaching: Revista interuniversitaria de didáctica* , 23, pp. 65-86.

Bláquez Entonado F. y Alonso Díaz L. (2006) *A Training Proposal for e-Learning Teachers*. European Journal of Open, Distance and E-learning. Accedido Marzo 2013, Disponible en http://www.eurodl.org/?article=242

Bonk C. J., Wisher R. A., y Lee J. (2004) Moderating Learner-Centered E-Learning: Problems and Solutions, Benefits and Implications. En: Roberts T. S. (ed) *Online Collaborative Learning: Theory and Practice*.USA, Information Science Publishing, pp. 54-85

Borges M. A. F. y Baranauskas M. C. C. (2003) CollabSS: a Tool to Help the Facilitator in Promoting Collaboration among Learners. *Educational Technology & Society*, 6 (1), 64-69.

Cavnar W. B. y Trenkle J. M. (1994) *N-Gram-Based Text Categorization*. Procedente de SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval.

Chapman P., Clinton J., Kerber R., Khabaza T., Reinartz T., Shearer C., y Wirth R. (2001) *CRISP-DM 1.0 Step-by-step data mining guide*. USA, SPSS Inc.

Chang S. (2004) The roles of mentors in electronic learning environments. *AACE Journal*, 12(3), 331-342.

Chen W. y Wasson B. (2002) An Instructional Assistant Agent for Distributed Collaborative Learning. En: Cerri S.A., Gouardères G., y Paraguaçu F. (eds) *Intelligent Tutoring Systems, Lecture Notes in Computer Science*: *Proceeding of ITS 2002, LNCS 2363*. Alemania, Springer-Verlag, pp. 609–618.

Chen W. (2006) Supporting teachers' intervention in collaborative Knowledge building. *Journal of Network and Computer Applications*, 29, 200-215.

Constantino-González M. y Suthers D. (2001) Coaching Web-based Collaborative Learning based on Problem Solution Differences and Participation. En: Moore J. D., Redfield C. L., y Johnson Lewis (eds) *Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future: Proceeding of AI&ED 2001*. IOS Press, pp. 176-187.

Coppin B. (2004) Introduction to Machine Learning. En: Coppin B. *Artificial Intelligence Illuminated*. Jones and Bartlett Publishers, pp. 267-289

Cornelius S. y Higgison C. (2001) The tutor's role and Effective Strategies for Online Tutoring. En: Higgison Carol (ed) *Online Tutoring e–Book: Proceeding of OTiS International e-Workshop: Developing Online Tutors*, Cap. 2.

CSALT (2001) *Effective Networked Learning in Higher Education: notes and guideline. Networked Learning in Higher Education Project (JCALT)*. Publicado por Centre for Studies in Advanced Learning Technologies, University of Lancaster. Accedido Marzo 2013, Disponible en http://csalt.lancs.ac.uk/jisc/guidelines_final.doc

Cunningham H., Maynard D., Bontcheva K., Tablan V., Aswani N., Roberts I., y Gorrell G. (2011) *Developing Language Processing Components with GATE Version 6 (a User Guide)*. Reino Unido, Departamento de Ciencia de la Computación, Universidad de Sheffield.

Day T.W. (2009) Online Collaborative Learning and Leadership Development. En: *Encyclopedia of Distance Learning, 2nd Ed.* Information Science Reference, pp. 1488 - 1492.

Denis B., Watland P., Sébastien P., y Verday N. (2004) *Roles and Competencies of the e-Tutor*. Procedente de Networked Learning Conference. Universidad de Lancaster, Inglaterra.

De Smet M., Van K. H., y Valcke M. (2008) Blending asynchronous discussion groups and peer tutoring in higher education: An exploratory study of online peer tutoring behavior. *Computers & Education*, 50, 207–223.

Dewar T. y Whittington D. (2000) *Teaching Online: A New Skill Set*. Procedente de Knowledge - Productive Learning at Work Conference. Sydney, Australia.

Dillenbourg P. (1999) What do you mean by collaborative learning? En: P. Dillenbourg (ed) *Collaborative-learning: Cognitive and Computational Approaches*. Oxford, Elsevier, pp.1-19.

do Prado H. A. y Ferneda E. (2008) *Emerging Technologies of Text Mining: Techniques and Applications*. Information Science Reference.

Fåhræus Eva R. (2000) *Growing Knowledge How to Support Collaborative Learning e-Discussions in Forum Systems*. No publicado, tesis de licenciatura Departamento de Ciencias de Sistemas y Computadoras, Universidad de Estocolmo.

Fayyad U., Piatetsky-Shapiro G., y Smyth P. (1996a) *From Data Mining to Knowledge Discovery in Databases*. AI Magazine.

Fayyad U., Piatetsky-Shapiro G., y Smyth P. (1996b) The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communiactions of the ACM*, 39(11), 27-34

Feinerer I. (2008) *A Text Mining Framework in R and Its Applications*. No publicado, Tesis Doctoral, WU Vienna University of Economics and Business.

Feldman R. y Sanger J. (2007) *The text mining handbook. Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

García Adeva J. J. y Calvo R. A. (2006) Mining Text with Pimiento. *IEEE Internet Computing*, 10 (4), 27-35.

Foix C. y Zavando S. (2002) *Estándares e-learning.Estado del Arte*. Centro de investigación y tecnología de Chile (INTEC). Accedido Marzo 2013, Disponible en http://empresas.sence.cl/documentos/elearning/INTEC%20-%20Estandares%20e-learning.pdf

Garrison D. R., Anderson T., y Archer W. (1999) Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, 2, 87-105.

Gerrard C. (2002) *Promoting Best Practice for E-tutoring through Staff Development*. Procedente de 9th Networked Learning International Conference. Sheffield, University of Sheffield.

González F. y Salmon G. (2004) *La función y la formación del e-moderator: clave del éxito en los nuevos entornos de aprendizaje*. Procedente de Online Educa Barecelona. España.

Goodyear P., Salmon G., Spector M. J., Steeples C., y Tickner S. (2001) Competences for online teaching: a especial report. *Educational Technology, Research and Development*, 49 (1), 65-72.

Greiffenhagen C. (2012) Making rounds: The routine work of the teacher during collaborative learning with computers. *International Journal of Computer-Supported Collaborative Learning*, 7(1), 11-42.

Guasch T., Alvarez I., y Espasa A. (2009) University teacher competencies in a virtual teaching/ learning environment: Analysis of a teacher training experience. *Teaching and Teacher Education*, 26(2), 199-206.

Gruzling N. (2006) *Linear separability of the vertices of an n-dimensional hypercube*. No publicado, M.Sc Thesis, University of Northern British Columbia.

Guzdial M., Ludovice P., Realff M., Morley T., y Carroll K. (2002) *When collaboration doesn't work*. Procedente de International Conference of the Learning Sciences. Seattle, USA.

Han J. y Kamber M. (2006) Classification and Prediction. En: *Data Mining: Concepts and Techniques*. San Franciso, Elsevier, pp. 285-382

Higgison C. (2001) Staff Development. En: Higgison Carol (ed) *Online Tutoring e–Book: Proceeding of OTiS International e-Workshop: Developing Online Tutors*. Cap. 8.

Hotho A., Nürnberger A., y Paaß G. (2005) Brief survey of text mining. *LDV Forum GLDV Journal for Computational Linguistics and Language Technology*, 20 (1), 19-62.

Hughes S.C., Wickersham L., Ryan-Jones D.L., y Smith S.A. (2002) Overcoming Social and Psychological Barriers to Effective On-line Collaboration. *Educational Technology & Society*, 5 (1), pp. 86-92.

Institute of IT Training (2007) Competency framework for e-tutores. A guide document on competency framework for e-tutors

Jaques D. y Salmon G. (2007) Learning in groups. A handbook for face-to-face and online environments. Londres, Routledge.

Johnson D. W. y Johnson R. T. (2008) Social Interdependence Theory and Cooperative Learning: The Teacher's Role. En: Gillies R. M., Ashman A. F., y Terwel J. (eds) *The Teacher's Role in Implementing Cooperative Learning in the Classroom*. Springer, pp. 9-37.

Juárez M., Buenfil R. N., y Trigueros M. (2008) De las prácticas convencionales a los ambientes de aprendizaje colaborativo a distancia. Un estudio con profesores de ciencias de bachillerato desde la Teoría de la actividad. *Revista Mexicana de Investigación Educativa*, 13 (39), 1055-1083.

Kemshal B. G. (2001) *The online teacher*. Reporte final de Steering Committee of the VET Teachers and Online Learning Project, TAFE NSW. Australia, Departamento de Educación y Entrenamiento.

Kreijns K., Kirschner P. A., y Jochems W. (2003) Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior*, 19, 335–353.

Krippendorff Klaus (2004) *Content analysis: an introduction to its methodology , 2*da edición. USA, SAGE Publications.

Kukulska-Hulme A. (2004): Do Online Collaborative Groups Need Leaders? En: *Online Collaborative Learning: Theory and Practice*. Information Science Publishing, pp. 262-280.

Kumar R., Rosé C.P., Wang Y.C., Joshi M., y Robinson A. (2007) Tutorial dialogue as adaptive collaborative learning support. En: Luckin R., Koedinger K.R., Greer J. (eds) *Proceedings of C*onference on Artificial Intelligence in Education: *Building Technology Rich Learning Contexts That Work.* Amsterdam, IOS Press, pp. 383–390

Larose D. T. (2005) Decision trees. En: Discovering knowledge in data. An introduction to Data Mining. John Wiley & Sons , pp. 107-127

Li Y. y Shawe-Taylor J. (2003) The SVM with uneven margins and Chinese document categorization. Procedente de 17th Pacific Asia Conference on Language Information and Computation (PACLIC17). Singapur.

Li Y., Zaragoza H., Herbrich R., Shawe-Taylor J., y Kandola J. (2002) The Perceptron Algorithm with Uneven Margins. Procedente de Nineteenth International Conference on Machine Learning.

Lim C. P. y Cheah P. T. (2003) The Role of the Tutor in Asynchronous Discussion Boards: A Case Study of a Pre-Service Teacher Course. *Education Media International*, 40 (1-2), 33-47.

Loh S., Krug Wives L., y Palazzo J. (2000) Descoberta Proativa de Conhecimento em Coleções Textuais: Iniciando sem Hipóteses. En: IV Oficina de Inteligência Artificial, 2000, Pelotas, RS. IV Oficina de Inteligência Artificial, pp. 143-154.

Llorente Cejudo M. del C. (2006). *El tutor en E-learning: aspectos a tener en cuenta*. Edutec. *Revista Electrónica de Tecnología Educativa, nro. 20*.

Lund K. (2004) Human support in CSCL. What, for Whom, and by Whom? En: Strijbos J., Kirschner P. A., y Martens R. L. (eds) *What We Know About CSCL And Implementing It In Higher Education*. Kliuwer Academic Publishers, pp.167-198.

Mason R. (1991) Moderating educational computer conferencing. DEOSNEWS, 1(19).

Mcwhaw K., Schnackenberg H., Sclater J., y Abrami P. C. (2003). From co-operation to collaboration helping students Become collaborative learners. En: Ashman A. y Gillies R.

(eds) *Co-operative Learning. The social and intellectual outcomes of learning in groups*. Routledge Falmer, pp.69-86.

McPherson M. y Baptista N. M. (2004) *The role of tutors as an integral part of online learning support*. European Journal of Open, Distance and E-learning. Accedido en Marzo de 2013, Disponible en http://www.eurodl.org/materials/contrib/2004/ Maggie_MsP.html

Mitchell T. M. (1997) Machine Learning. McGraw-Hill Science, pp. 1-19

Mohri M., Rostamizadeh A., y Talwalkar A. (2012): Foundations of Machine Learning. *MIT Press*, pp. 1-8

Náther P. (2005) N-gram based Text Categorization. En: N-gram based Text Categorization. No publicado, Diploma thesis, Faculty of Mathematics, Physics and informatics, Comenius University, pp. 10-13

Negnevitsky Michael (2005) Artificial Intelligence A Guide to Intelligent Systems. Addison-Wesley, pp. 165-217

Olivares O. J. (2007) Collaborative vs. Cooperative Learning: The Instructor's Role in Computer Supported Collaborative Learning. En: Orvis K. L. y Lassiter A. L.R (eds) *Computer-Supported Collaborative Learning: Best Practices and Principles for Instructors*. USA, Information Science Publishing, pp. 20-39.

Onrubia J. y Engel A. (2012) The role of teacher assistance on the effects of a macro-script in collaborative writing tasks. *International Journal of Computer-Supported Collaborative Learning*, 7(1), pp. 161-186.

Oracle (2013) Regular Expressions. Accedido en Marzo de 2013, Disponible en http://docs.oracle.com/javase/tutorial/essential/regex/intro.html

Orvis K.L. y Lassiter A.L.R. (2006): Computer-Supported Collaborative Learning: The Role of the Instructor. En: *Teaching and learning with virtual teams*. Information Science Publishing, pp. 158-179.

Padilha T.P. P., Almeida L. M., y Alves J. B. M. (2004) Mining Techniques for Models of Collaborative Learning. Procedente de 2nd International Workshop on Designing Computational Models of Collaborative Learning Interaction. Brasil.

Palloff R.M. y Pratt K. (2001) *Lessons from the Cyberspace Classroom.The Realities of Online Teaching*. USA, Jossey-Bass.

Palloff R. M., y Keith P. (2007) Moving Teaching and Learning Online. En: *Building Online Learning Communities-Effective strategies for the virtual classroom*. USA, Jossey-Bass, pp. 105-126.

Marbán Oscar, Mariscal Gonzalo y Segovia Javier (2009) *A Data Mining & Knowledge Discovery Process Model*. En : Ponce Julio y Karahoca Adem (eds.) *Data mining from data to knowledge in real applications*. In-tech. Croacia, pp.1-16

RAE (1999) Ortografía de la Lengua Española. España, Real Academia Española.

Reid D. (2002) A Classification Schema of Online Tutor Competencies. Procedente de International Conference on Computers in Education (ICCE).

Reid D. (2003) Was she smiling as she typed that: an exploratory study into online tutor competencies and the factors that affect those competencies. En: Crisp G., Thiele D., Scholten I., Barker S., y Baron J. (eds) *Interact, Integrate, Impact: Proceeding of 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*. Adelaide, Australia.

Reimann P., Calvo R.A., Yacef K., y Southavilay V. (2010) *Comprehensive Computational Support for Collaborative Learning from Writing*. Procedente de International Conference on Computers in Education (ICCE). Putrajaya, Malaysia.

Raynor W (1999) *The International Dictionary of Artificial Intelligence*. Glenlake Publishing Company Ltd. USA.

Rosé C., Wang Y., Cui Y., Arguello J., Stegmann K., Weinberger A., y Fischer F. (2008) Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3 (3), 237–271.

Rosenberg M. J. (2001) *E-learning: Strategies for Delivering Knowledge in the Digital Age*. USA, McGraw-Hill Professional.

Santana Mansilla P., Costaguta R., y Missio D. (2011) *Clasificación de habilidades de e-tutores en aprendizaje colaborativo soportado por computadora*. Procedente de Segundo Congreso Internacional de Eduación en Ciencia y Tecnología. Universidad Nacional de Catamarca, Catamarca.

Santana Mansilla P., Costaguta R., y Missio D. (2012) Habilidades de E-tutores en Grupos Colaborativos. En: Peñaranda, N., Zazarini, S., y Bejarano, I. F. (eds) *Experiencias Innovadoras en Investigación Aplicada*. Jujuy, Ediciones DASS-UCSE, pp. 687-704

Schwarz B.B. y Asterhan C.S. (2011) E-Moderation of Synchronous Discussions in Educational Settings: A Nascent Practice. *Journal of the Learning Sciences*, 20 (3), pp. 395-442.

Sebastiani F. (2002) Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1–47.

Magalhaes S. E. (2002) *Descoberta de Conhecimento com o uso de text mining: Cruzando o Abismo de Moore*. No publicado, Tesis de Master, Universidad Católica de Brasil.

Smith T.C. (2005) Fifty-One Competencies for Online Instruction. *The Journal of Educators Online* (JEO), 2(2).

Soller A. L. (2001) Supporting Social Interaction in an Intelligent Collaborative Learning System. *International Journal of Artificial Intelligence in Education*, 12, 40-62.

Soller A., Martínez M. A., Jermann P., y Muehlenbrock M. (2005) From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning. *International Journal of Artificial Intelligence in Education*, 15 (4), 261-290.

Suh H. y Lee S. (2006) Collaborative Learning Agent for Promoting Group Interaction. *ETRI Journal*, 28(4), 461-474.

Sulčič V. y Sulčič A. (2007) Can Online Tutors Improve the Quality of E-Learning? *Issues in Informing Science and Information Technology*, 4, 210-210.

Tchounikine P., Rummel N., McLaren B.M. (2010) Computer Supported Collaborative Learning and Intelligent Tutoring Systems. En: *Advances in Intelligent Tutoring Systems.* Springer, pp. 447-463.

Teles L., Ashton S., Roberts T. , y Tzoneva I. (2001) The role of the instructor in e-learning collaborative environments. *TechKnowLogia*, May/Jun.

Universidad de Ulster. *Skills and competences for staff.* En: *Skills for e-learning*. Accedido Marzo 2013, Disponible en http://www2.plymouth.ac.uk/ed/ELT%20documents/developers/Skills-competences.pdf

Weiss S. M., Indurkhya N., Zhang T., y Damerau F. J. (2005) *Text Mining Predictive Methods for Analyzing Unstructured Information*. USA, Springer.

Williams S. y Roberts T. (2002) Computer Supported Collaborative Learning: Strengths and Weaknesses. Procedente de International Conference on Computers in Education (ICCE'02).

Witten I.H. y Frank E. (2005) Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed. San Francisco, Morgan Kaufmann.

Witten I. H., Paynter G. W., Frank E., Gutwin C., y Nevill-Manning C. G. (2005) KEA: Practical automatic keyphrase extraction. En: Theng Y. y Foo S. (eds) *Design and Usability of Digital Libraries: Case Studies in the Asia Pacific*. London, Information Science Publishing, pp. 129-152.

Zumbach J., Hillers A., y Reimann P. (2004) Supporting Distributed Problem-Based Learning: The Use of Feedback Mechanisms in Online Learning. En: Roberts T.S. (ed) *Online Collaborative Learning: Theory and Practice*. USA, Information Science Publishing, pp. 86-102.

ANEXO A
ANÁLISIS DE CONTENIDO

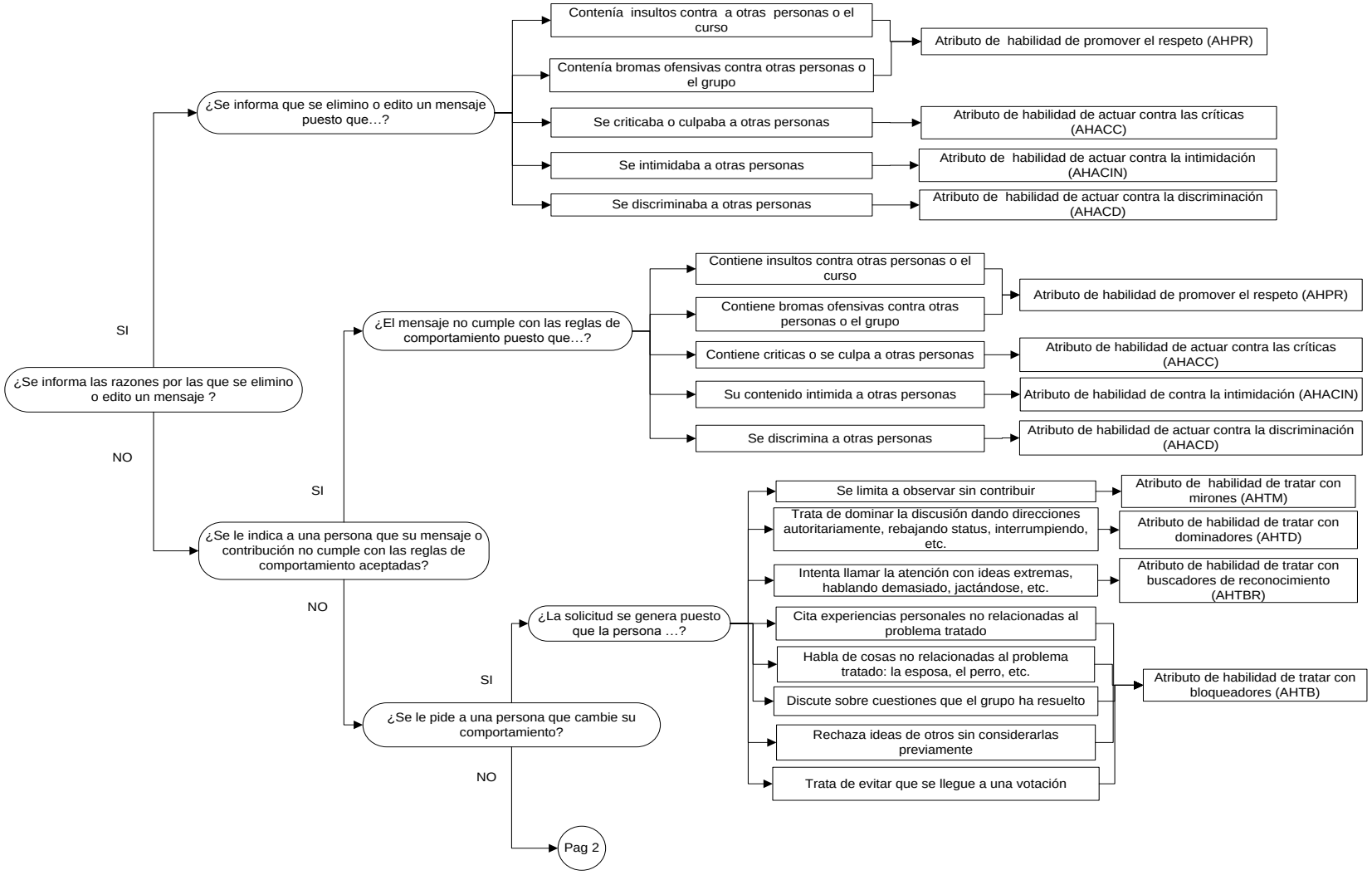
A.1 Forma de usar el esquema de decisión

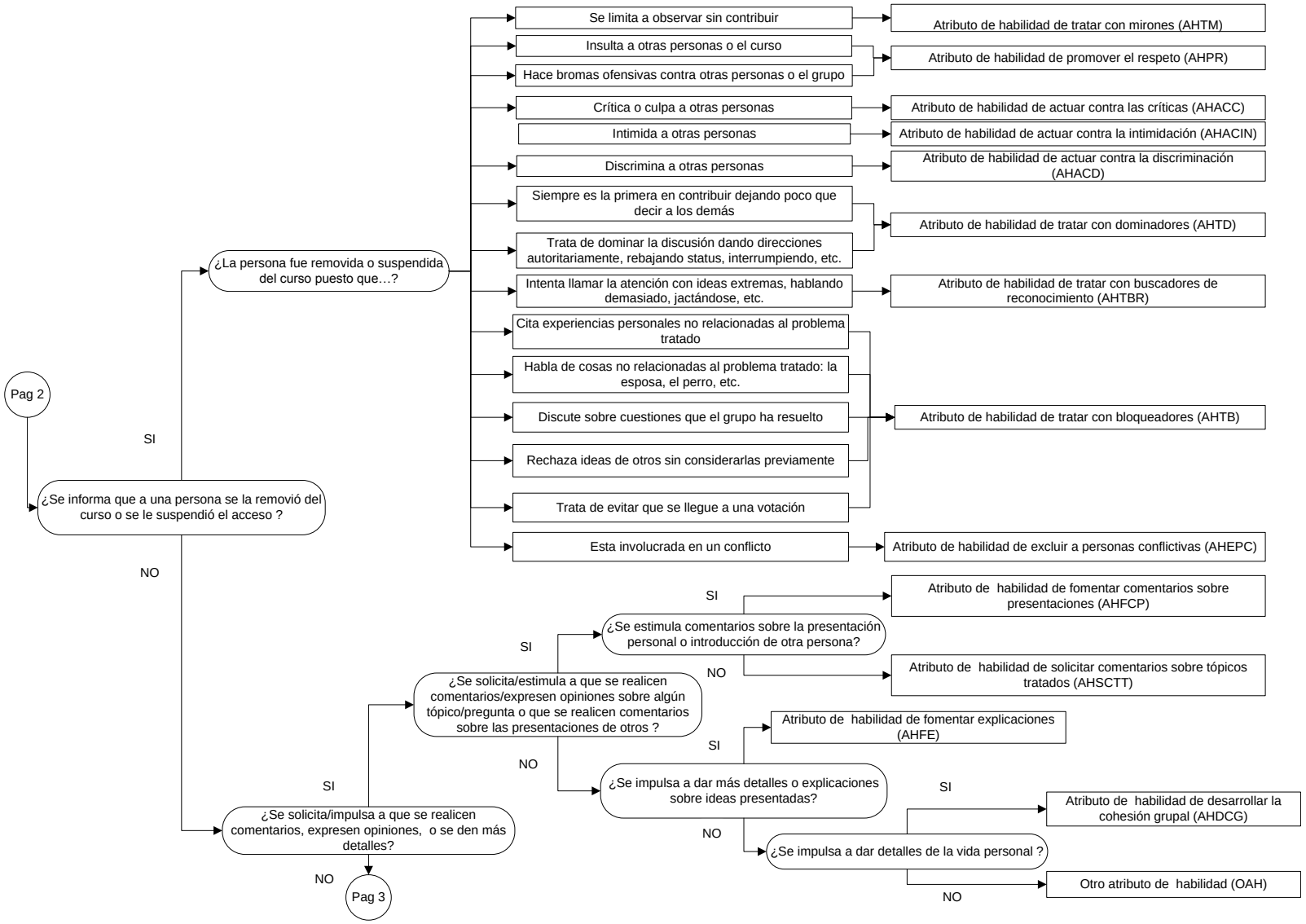
El esquema de decisión permite identificar si cuando un e-*tutor* se comunica con un estudiante o grupo de estudiantes se manifiesta alguna de las habilidades sociales y en caso de ser así cual es el atributo de habilidad que corresponde a la habilidad manifestada. Al aplicar el esquema de decisión es preciso tener en cuenta que:

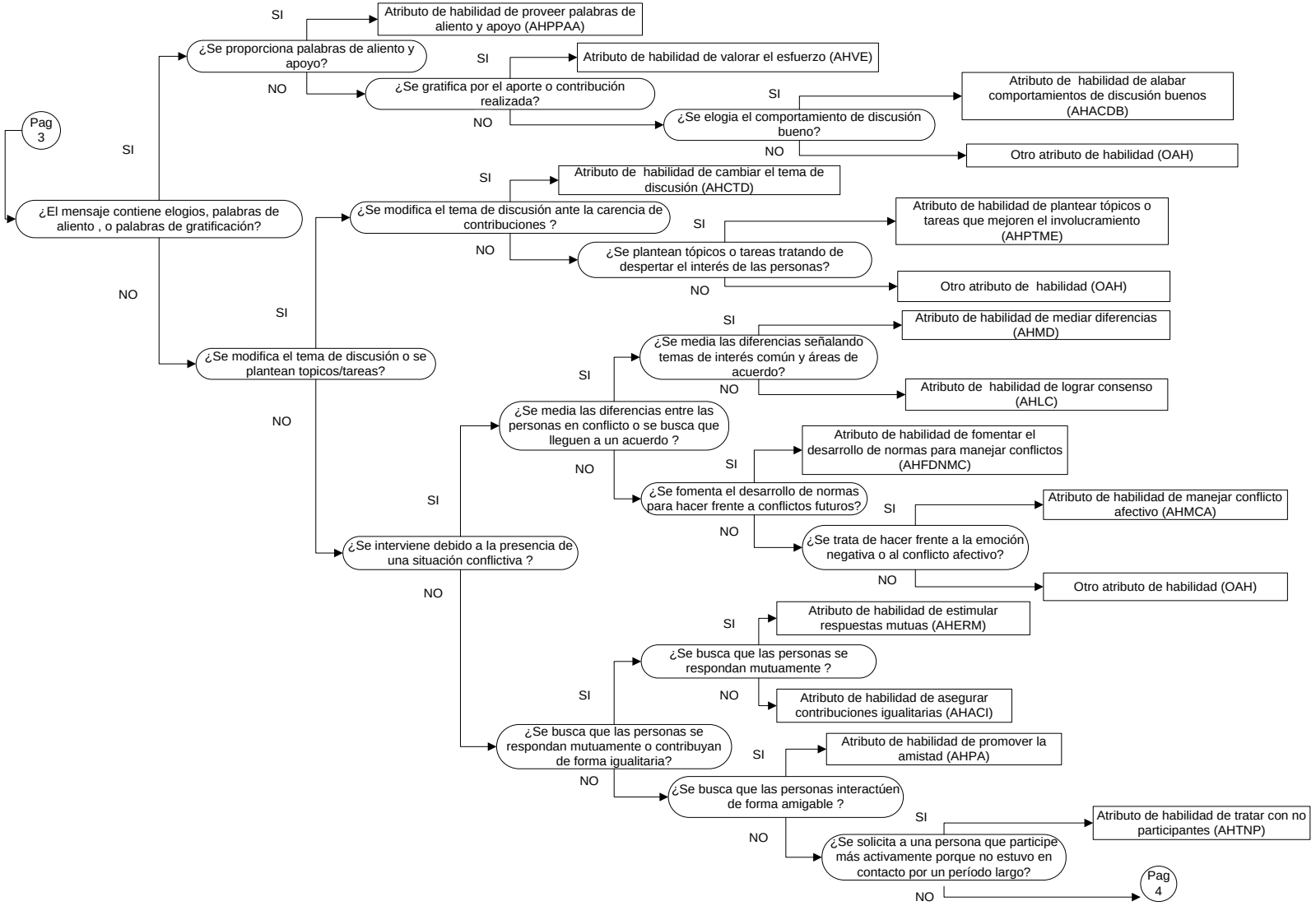
- Los óvalos representan puntos de decisión o preguntas que guían en la identificación del atributo de habilidad manifestado.
- Las fechas que emergen de la parte superior de los óvalos indican el camino a seguir en el caso de que la respuesta a la pregunta del óvalo sea afirmativa.
- Las fechas que emergen de la parte inferior de los óvalos indican el camino a seguir en el caso de que la respuesta a la pregunta del óvalo sea negativa.
- Los cuadrados indican la el atributo de habilidad que se manifiesta en el mensaje analizado.
- Los círculos con un número en su interior y que tienen una flecha de salida indican que para continuar analizando el mensaje es preciso dirigirse a una página que contiene un círculo con el mismo número pero con la flecha de entrada.
- Los círculos con un número en su interior y que tienen una flecha de llegada o entrada indican que ese es el punto donde continuar analizando el mensaje bajo consideración.

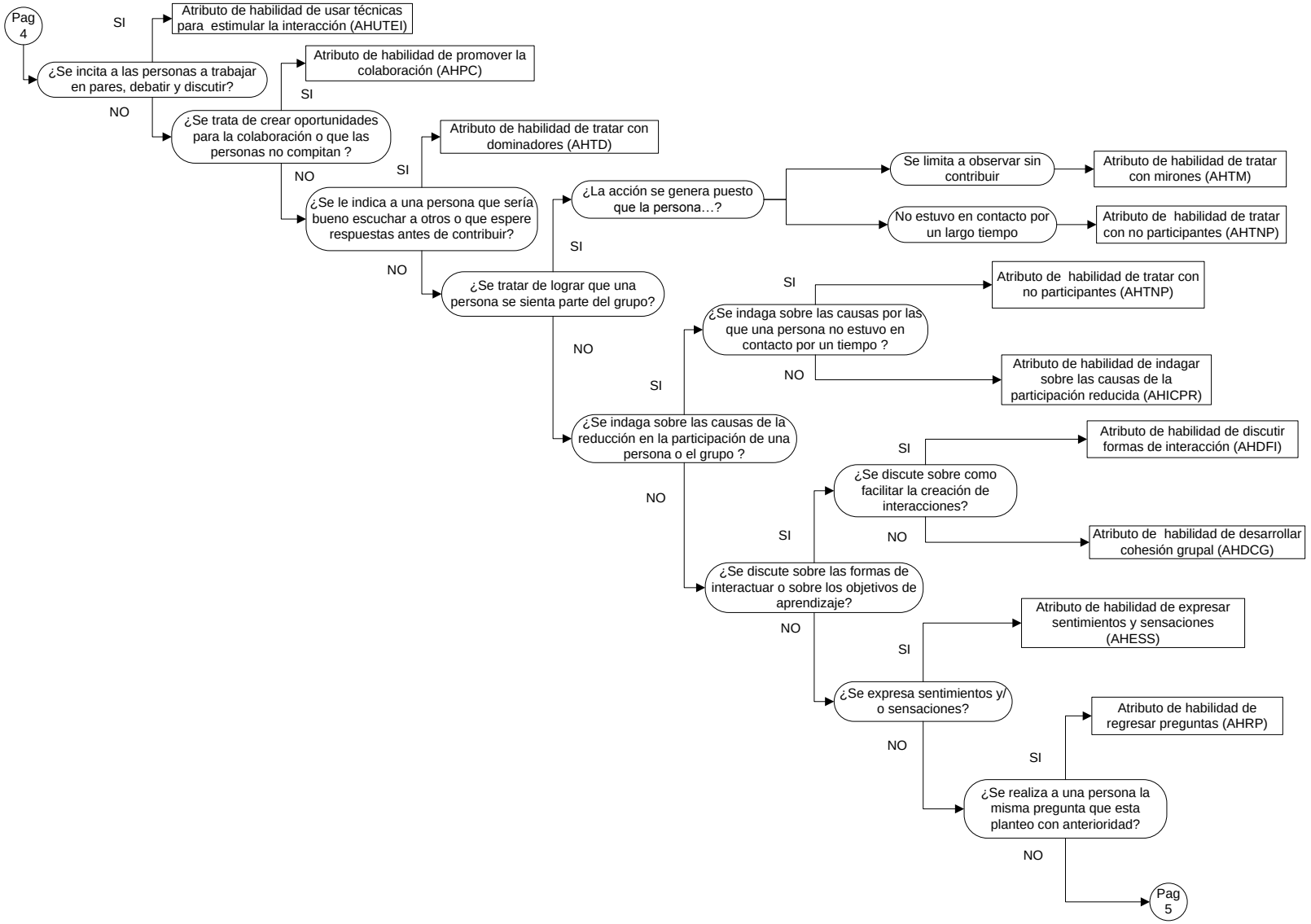
Para asignar un atributo de habilidad al mensaje del e-tutor que se analiza es necesario comenzar por la pregunta "¿Se informa las razones por las que se elimino o edito un mensaje?" del esquema de decisión para luego seguir el camino señalado por las flechas en función de las respuestas a los puntos de decisión. Este proceso finaliza cuando se alcanza uno de los cuadros de atributo de habilidad ya que este indica el atributo de habilidad que corresponde asignar al mensaje considerado. Entre paréntesis se incluyo el código alfabético o abreviatura de cada atributo de habilidad que tiene que usarse al catalogar un mensaje en lugar del nombre completo del atributo de habilidad.

A.2 Esquema de decisión




Pag 2
¿Se informa que a una persona se la removió del curso o se le suspendió el acceso ?
SI
NO
¿La persona fue removida o suspendida del curso puesto que…?
Se limita a observar sin contribuir
Insulta a otras personas o el curso
Hace bromas ofensivas contra otras personas o el grupo
Crítica o culpa a otras personas
Intimida a otras personas
Discrimina a otras personas
Siempre es la primera en contribuir dejando poco que decir a los demás
Trata de dominar la discusión dando direcciones autoritariamente, rebajando status, interrumpiendo, etc.
Intenta llamar la atención con ideas extremas, hablando demasiado, jactándose, etc.
Cita experiencias personales no relacionadas al problema tratado
Habla de cosas no relacionadas al problema tratado: la esposa, el perro, etc.
Discute sobre cuestiones que el grupo ha resuelto
Rechaza ideas de otros sin considerarlas previamente
Trata de evitar que se llegue a una votación
Esta involucrada en un conflicto
Atributo de habilidad de tratar con mirones (AHTM)
Atributo de habilidad de promover el respeto (AHPR)
Atributo de habilidad de actuar contra las críticas (AHACC)
Atributo de habilidad de actuar contra la intimidación (AHACIN)
Atributo de habilidad de actuar contra la discriminación (AHACD)
Atributo de habilidad de tratar con dominadores (AHTD)
Atributo de habilidad de tratar con buscadores de reconocimiento (AHTBR)
Atributo de habilidad de tratar con bloqueadores (AHTB)
Atributo de habilidad de excluir a personas conflictivas (AHEPC)
¿Se solicita/impulsa a que se realicen comentarios, expresen opiniones, o se den más detalles?
SI
NO
Pag 3
¿Se solicita/estimula a que se realicen comentarios/expresen opiniones sobre algún tópico/pregunta o que se realicen comentarios sobre las presentaciones de otros ?
SI
NO
¿Se estimula comentarios sobre la presentación personal o introducción de otra persona?
SI
NO
Atributo de habilidad de fomentar comentarios sobre presentaciones (AHFCP)
Atributo de habilidad de solicitar comentarios sobre tópicos tratados (AHSCTT)
¿Se impulsa a dar más detalles o explicaciones sobre ideas presentadas?
SI
NO
Atributo de habilidad de fomentar explicaciones (AHFE)
¿Se impulsa a dar detalles de la vida personal ?
SI
NO
Atributo de habilidad de desarrollar la cohesión grupal (AHDCG)
Otro atributo de habilidad (OAH)


Pag 3
¿El mensaje contiene elogios, palabras de aliento , o palabras de gratificación?
SI
NO
¿Se proporciona palabras de aliento y apoyo?
SI
Atributo de habilidad de proveer palabras de aliento y apoyo (AHPPAA)
NO
¿Se gratifica por el aporte o contribución realizada?
SI
Atributo de habilidad de valorar el esfuerzo (AHVE)
NO
¿Se elogia el comportamiento de discusión bueno?
SI
Atributo de habilidad de alabar comportamientos de discusión buenos (AHACDB)
NO
Otro atributo de habilidad (OAH)
¿Se modifica el tema de discusión o se plantean topicos/tareas?
SI
¿Se modifica el tema de discusión ante la carencia de contribuciones ?
SI
Atributo de habilidad de cambiar el tema de discusión (AHCTD)
NO
¿Se plantean tópicos o tareas tratando de despertar el interés de las personas?
SI
Atributo de habilidad de plantear tópicos o tareas que mejoren el involucramiento (AHPTME)
NO
Otro atributo de habilidad (OAH)
NO
¿Se interviene debido a la presencia de una situación conflictiva ?
SI
¿Se media las diferencias entre las personas en conflicto o se busca que lleguen a un acuerdo ?
SI
¿Se media las diferencias señalando temas de interés común y áreas de acuerdo?
SI
Atributo de habilidad de mediar diferencias (AHMD)
NO
Atributo de habilidad de lograr consenso (AHLC)
NO
¿Se fomenta el desarrollo de normas para hacer frente a conflictos futuros?
SI
Atributo de habilidad de fomentar el desarrollo de normas para manejar conflictos (AHFDNMC)
NO
¿Se trata de hacer frente a la emoción negativa o al conflicto afectivo?
SI
Atributo de habilidad de manejar conflicto afectivo (AHMCA)
NO
Otro atributo de habilidad (OAH)
NO
¿Se busca que las personas se respondan mutuamente o contribuyan de forma igualitaria?
SI
¿Se busca que las personas se respondan mutuamente ?
SI
Atributo de habilidad de estimular respuestas mutuas (AHERM)
NO
Atributo de habilidad de asegurar contribuciones igualitarias (AHACI)
NO
¿Se busca que las personas interactúen de forma amigable ?
SI
Atributo de habilidad de promover la amistad (AHPA)
NO
¿Se solicita a una persona que participe más activamente porque no estuvo en contacto por un período largo?
SI
Atributo de habilidad de tratar con no participantes (AHTNP)
NO
Pag 4


Pag 4
SI
Atributo de habilidad de usar técnicas para estimular la interacción (AHUTEI)
¿Se incita a las personas a trabajar en pares, debatir y discutir?
SI
Atributo de habilidad de promover la colaboración (AHPC)
NO
¿Se trata de crear oportunidades para la colaboración o que las personas no compitan ?
SI
Atributo de habilidad de tratar con dominadores (AHTD)
NO
¿Se le indica a una persona que sería bueno escuchar a otros o que espere respuestas antes de contribuir?
SI
¿La acción se genera puesto que la persona…?
Se limita a observar sin contribuir
Atributo de habilidad de tratar con mirones (AHTM)
No estuvo en contacto por un largo tiempo
Atributo de habilidad de tratar con no participantes (AHTNP)
NO
¿Se tratar de lograr que una persona se sienta parte del grupo?
SI
Atributo de habilidad de tratar con no participantes (AHTNP)
¿Se indaga sobre las causas por las que una persona no estuvo en contacto por un tiempo ?
NO
Atributo de habilidad de indagar sobre las causas de la participación reducida (AHICPR)
NO
SI
¿Se indaga sobre las causas de la reducción en la participación de una persona o el grupo ?
SI
Atributo de habilidad de discutir formas de interacción (AHDFI)
¿Se discute sobre como facilitar la creación de interacciones?
NO
SI
Atributo de habilidad de desarrollar cohesión grupal (AHDCG)
¿Se discute sobre las formas de interactuar o sobre los objetivos de aprendizaje?
NO
SI
Atributo de habilidad de expresar sentimientos y sensaciones (AHESS)
¿Se expresa sentimientos y/ o sensaciones?
NO
SI
Atributo de habilidad de regresar preguntas (AHRP)
¿Se realiza a una persona la misma pregunta que esta planteo con anterioridad?
NO
Pag 5

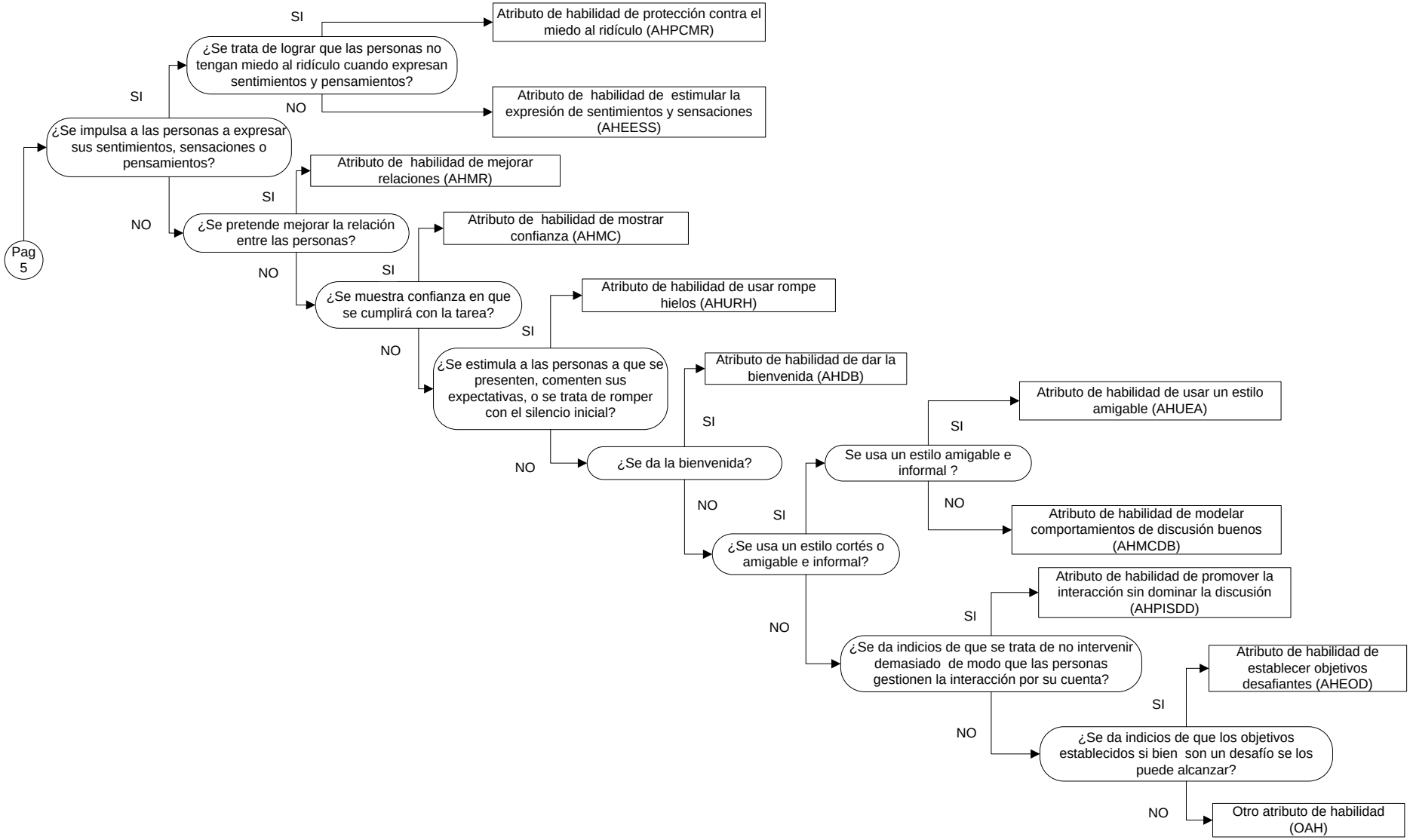

Pag 5
¿Se impulsa a las personas a expresar sus sentimientos, sensaciones o pensamientos?
SI
NO
¿Se trata de lograr que las personas no tengan miedo al ridículo cuando expresan sentimientos y pensamientos?
SI
NO
Atributo de habilidad de protección contra el miedo al ridículo (AHPCMR)
Atributo de habilidad de estimular la expresión de sentimientos y sensaciones (AHEESS)
¿Se pretende mejorar la relación entre las personas?
SI
NO
Atributo de habilidad de mejorar relaciones (AHMR)
¿Se muestra confianza en que se cumplirá con la tarea?
SI
NO
Atributo de habilidad de mostrar confianza (AHMC)
¿Se estimula a las personas a que se presenten, comenten sus expectativas, o se trata de romper con el silencio inicial?
SI
NO
Atributo de habilidad de usar rompe hielos (AHURH)
¿Se da la bienvenida?
SI
NO
Atributo de habilidad de dar la bienvenida (AHDB)
¿Se usa un estilo cortés o amigable e informal?
SI
NO
Se usa un estilo amigable e informal ?
SI
NO
Atributo de habilidad de usar un estilo amigable (AHUEA)
Atributo de habilidad de modelar comportamientos de discusión buenos (AHMCDB)
¿Se da indicios de que se trata de no intervenir demasiado de modo que las personas gestionen la interacción por su cuenta?
SI
NO
Atributo de habilidad de promover la interacción sin dominar la discusión (AHPISDD)
¿Se da indicios de que los objetivos establecidos si bien son un desafío se los puede alcanzar?
SI
NO
Atributo de habilidad de establecer objetivos desafiantes (AHEOD)
Otro atributo de habilidad (OAH)

Tabla A.1 Unidades de Grabación y atributos de habilidad social identificados por los codificadores
Tabla A.1

Código de unidad de grabación/ codificación	Unidad de grabación/codificación	Experiencia de ACSC	Atributo de habilidad identificado Analista 1	Atributo de habilidad identificado Analista 2
U1	Chicos, espero sus consultas respecto a cómo hacer la matriz, apliquen a cada caso.	Metodología II	AHPC	AHPC
U2	Y adjunten los archivos por esta vía.	Metodología II	OAH	OAH
U3	Saludos!!!	Metodología II	AHMCDB	AHMCDB
U4	Mauricio y resto del grupo del foro, recuerden que cuando investigamos estamos involucrando ciertos aspectos de la realidad que son las VARIABLES (informáticas) que se investigan y que deben aparecer en el planteo del problema, en los objetivos, en la hipótesis.	Metodología II	AHUTEI	AHUTEI
U5	Por lo tanto no es cómo queda mejor sino tienen que plantear de manera que los que leamos (el resto de los integrantes del foro) tenemos que identificar rápidamente esos aspectos o variables.	Metodología II	AHUTEI	AHUTEI
U6	De paso ya les dije que en la matriz, una de las cuestiones que tienen que tener en cuenta para verificar la "consistencia" son las variables.	Metodología II	OAH	OAH
U7	Ah!	Metodología II	OAH	OAH
U8	Daniel y resto del grupo, cuando opinemos tenemos que decir por qué pensamos de esa manera!!! no basta con "me gusta" o "no me gusta".	Metodología II	AHDFI	AHDFI
U9	Saludos a todos y me alegro mucho los que intervinieron!!!	Metodología II	AHVE	AHVE
U10	Susana H.	Metodología II	xxx	xxx
U11	Daniel, Está bien tienes que poner algo relacionado con tu problema...	Metodología II	AHFE	AHFE
U12	Pero tené en cuenta por favor el comentario que hice en la conversación de Diaz Navarro.	Metodología II	AHERM	AHERM
U13	Intentá escribir algo más preciso, subilo a este foro, y te contestamos entre todos.	Metodología II	AHFE	AHPC

U14	Ok?	Metodología II	AHFE	AHFE
U15	Saludos!!!	Metodología II	AHMCDB	AHMCDB
U16	Susana	Metodología II	xxx	xxx
U17	Mauricio...	Metodología II	xxx	xxx
U18	Estaría bueno que digas qué es ESFE y las características...	Metodología II	AHFE	AHFE
U19	Fijate... qué variables informáticas aparecen en tu hipótesis???	Metodología II	AHPTME	AHPTME
U20	Buscá y no vas a encontrar ninguna...	Metodología II	OAH	OAH
U21	Cuáles son las variables? correctitud?...	Metodología II	AHSCTT	AHSCTT
U22	Piensen un poco más...	Metodología II	AHUTEI	AHUTEI
U23	QUÉ OPINAN EL RESTO DE INTEGRANTES DEL FORO?	Metodología II	AHPC	AHPC
U24	Susana Herrera	Metodología II	xxx	xxx
U25	Mauricio, antes que nada me pareció muy buena tu idea de poner en esta conversación los ensayos de matrices que vayan haciendo!!!	Metodología II	AHVE	AHVE
U26	Ojalá el otro grupo la suba también.	Metodología II	AHESS	AHESS
U27	Esto permite ir revisando y corrigiendo...	Metodología II	AHVE	AHVE
U28	Te comento...	Metodología II	OAH	OAH
U29	A esta altura no se exige tanto, después deberán pulir más detalles...	Metodología II	OAH	OAH
U30	Pero en general...	Metodología II	xxx	xxx
U31	El problema está bien.	Metodología II	AHVE	AHVE
U32	Les dije varias veces que "Desarrollar..." es un objetivo de un proyecto de desarrollo o de ingeniería.	Metodología II	OAH	OAH
U33	Que un proyecto de investigación se tiene que investigar para obtener un conocimiento inédito o nuevo en la disciplina.	Metodología II	OAH	OAH
U34	Entonces revisen el objetivo general.	Metodología II	AHSCTT	AHSCTT
U35	Los verbos usados en los objetivos específicos están bien.	Metodología II	OAH	OAH
U36	Podrían también revisar el contenido o sentido de lo que pusieron, se puede mejorar.	Metodología II	AHSCTT	AHSCTT
U37	Respecto a la hipótesis...	Metodología II	AHSCTT	AHSCTT
U38	En qué asignatura manejaron "correctitud" como variable?	Metodología II	AHUTEI	AHUTEI
U39	Talvez precisión, etc...	Metodología II	xxx	xxx

U40	Tienen que revisar ello.	Metodología II	AHSCTT	AHSCTT
U41	Saludos!!!	Metodología II	AHMCDB	AHMCDB
U42	Mauricio y Betiana, perdón por no responder antes.	Metodología II	AHMCDB	AHMCDB
U43	Estoy con problemas en las máquinas que tengo word 2007, no puedo abrir la nueva versión que subieron.	Metodología II	AHMCDB	AHMCDB
U44	Mañana temprano la veo y les comento.	Metodología II	AHMCDB	AHMCDB
U45	Saludos!	Metodología II	AHMCDB	AHMCDB
U46	Susana Herrera	Metodología II	xxx	xxx
U47	Estimada Betiana, En el marco metodológico tienes que referirte a cuáles son las metodologías de investigación y de desarrollo que te guiarán en la definición de la metodología que ustedes proponen para su investigación.	Metodología II	AHUTEI	AHUTEI
U48	En concreto, en el marco metodológico van metodologías ya existentes mientras que la metodología que ustedes proponen va en el apartado referido a "metodología o planificación".	Metodología II	OAH	OAH
U49	Ustedes mezclaron y en este punto ponen algo muy general sobre metodología de la investigación y sobre los pasos que ustedes seguirán para su investigación.	Metodología II	OAH	OAH
U50	Respecto a metodologías de investigación, está bien que hagan referencia a un autor determinado, después tendrán que referenciar en su metodología y justificar qué toman de ese autor, o qué etapas.	Metodología II	AHVE	AHVE
U51	Luego, para desarrollar un simulador, ustedes deberán seguir alguna metodología de una agil o rápida o la que deseen.	Metodología II	AHUTEI	AHUTEI
U52	Eso ya lo habíamos conversado en la presentación oral del seminario 1.	Metodología II	OAH	OAH
U53	Recuerdan?	Metodología II	AHPC	AHPC
U54	Esas metodologías deberían estar incluidas en el marco metodológico.	Metodología II	OAH	OAH
U55	Si aún no definieron la metodología (que no sería muy atinado que esté ocurriendo eso) deberían describir las metodologías que serán estudiadas.	Metodología II	OAH	OAH
U56	Luego en la etapa de metodología o planificación, ustedes determinan cada etapa que tendrá su investigación, integrando o mencionando lo que pusieron en el marco metodológico.	Metodología II	OAH	OAH
U57	RESPECTO AL MARCO EMPÍRICO	Metodología II	OAH	OAH

	Comentan lo de la asignatura MI 2, que no es necesario.			
U58	Y luego, "como por arte de magia" comienzan a hablar de la UNSE.	Metodología II	OAH	OAH
U59	Es necesario una breve introducción que justifique qué es lo que se considera en el marco empírico y por qué.	Metodología II	OAH	OAH
U60	Además, debería describirse sobre todo la asignatura Simulación en la cual se trabajará.	Metodología II	OAH	OAH
U61	Saludos y esperemos a ver que justifiquen los otros integrantes del grupo que opinaron que les parece correcto.	Metodología II	AHPC	AHPC
U62	Será que están trabajando bien en este item de la nota???	Metodología II	AHPISDD	AHPISDD
U63	Tu consulta fue muy oportuna Betiana para que revisen tus compañeros este tema!	Metodología II	AHVE	AHVE
U64	Susana Herrera	Metodología II	xxx	xxx
U65	Estimados miembros de este grupo 1, antes que nada los felicito por la interacción y participación que tienen.	Metodología II	AHACDB	AHACDB
U66	Muy interesantes los intercambios de opiniones.	Metodología II	AHACDB	AHACDB
U67	Me gustaría que Pablo participe un poco más!	Metodología II	AHACI	AHACI
U68	Y Betiana también!	Metodología II	AHACI	AHACI
U69	Ahora paso a comentar la matriz de Daniel-Pablo:	Metodología II	AHURH	AHURH
U70	- Primero que todo, es una matriz muy bien lograda.	Metodología II	AHPPAA	AHPPAA
U71	Comprendieron lo que se pedía en el ejercicio.	Metodología II	AHPPAA	AHPPAA
U72	- Respecto a los objetivos generales, considero que están bien!	Metodología II	AHPPAA	AHPPAA
U73	No hagas caso a la opinión de Mauricio esta vez, jajajaja...	Metodología II	AHLC	AHLC
U74	Recuerden que los objetivos generales son los aspectos que "a la larga, en el futuro" contribuirá la investigación y en este caso, está bien lo que pusieron.	Metodología II	AHPPAA	AHPPAA
U75	Además, Mauricio reitero que los objetivos "desarrollar..." corresponden a un trabajo de aplicación y no de investigación.	Metodología II	AHTD	AHTD
U76	- Respecto al tercer objetivo específico, al contrario, considero muy acertada la opinión de Mauricio.	Metodología II	AHPPAA	AHPPAA
U77	Llama la atención que allí se hable de confiabilidad...	Metodología II	AHUTEI	AHUTEI
U78	Por qué no seguridad?	Metodología II	AHUTEI	AHUTEI

U79	Además respecto a ese objetivo, no se debería poner allí algo tan acertado como un número.	Metodología II	OAH	OAH
U80	Es muy arriesgado!	Metodología II	OAH	OAH
U81	Y con la investigación se intenta pero nunca se está seguro, sino no haría falta corroborar la hipótesis.	Metodología II	OAH	OAH
U82	Conviene poner "optimizar, aumentar,..."	Metodología II	OAH	OAH
U83	- Respecto a los objetivos, en general, usaron los verbos indicados en la asignatura.	Metodología II	AHVE	AHVE
U84	MUY BIEN!!!	Metodología II	AHPPAA	AHPPAA
U85	- Y, por último, respecto a los indicadores, considero muy atinada la opinión de Betiana.	Metodología II	AHVE	AHVE
U86	Creo que falta pensar un poquito más cómo se medirían esos indicadores...	Metodología II	AHUTEI	AHUTEI
U87	Por ejemplo, no está claro el tema de poner como indicador "lenguajes de programación"... y los que siguen posteriormente.	Metodología II	OAH	OAH
U88	Ojo! puede ser que en la nota completa, al describir el proceso, se entienda todo!	Metodología II	OAH	OAH
U89	Pero por las dudas... verifiquen esa parte.	Metodología II	AHUTEI	AHUTEI
U90	Intenté entrar en el sitio indicado por Daniel pero me fue imposible.	Metodología II	OAH	OAH
U91	No sé si es mi conexión o pasa algo en el sitio.	Metodología II	OAH	OAH
U92	Saludos a todos!!!	Metodología II	AHMCDB	AHMCDB
U93	Susana Herrera	Metodología II	xxx	xxx
U94	Estimados, les pido que revisen en el Seminario la consigna referida al esquema solicitado, y luego expresen en esta conversación si están de acuerdo con este esquema que están presentando Daniel y Pablo.	Metodología II	AHUTEI	AHUTEI
U95	Al margen de ello, está muy buena la idea de subir el esquema!!!	Metodología II	AHVE	AHVE
U96	Así pulimos aquí cuestiones que pueden ser arrastradas erróneamente hacia la matriz.	Metodología II	AHPC	AHPC
U97	Saludos!	Metodología II	AHMCDB	AHMCDB
U98	Susana Herrera	Metodología II	xxx	xxx
U99	Pablo y compañero, creo que el problema está coherente con el planteo global de la matriz y con lo que vienen presentando en las clases anteriores.	Metodología II	AHVE	AHVE

U100	Saludos, Susana Herrera	Metodología II	AHMCDB	AHMCDB
U101	Pablo, ya opiné en otra conversación posterior.	Metodología II	AHMCDB	AHMCDB
U102	Saludos!susana herrera	Metodología II	AHMCDB	AHMCDB
U103	Chicos, espero sus consultas respecto a cómo hacer la matriz, apliquen a cada caso	Metodología II	AHSCTT	AHSCTT
U104	Y adjunten los archivos por esta vía.	Metodología II	AHUTEI	AHUTEI
U105	Saludos!!!	Metodología II	AHMCDB	AHMCDB
U106	Hola Cecilia y Gabriela!	Metodología II	AHDB	AHDB
U107	Perdón por no responder antes, estuve ocupada con otras actividades.	Metodología II	AHMCDB	AHMCDB
U108	Respecto a la matriz de consistencia.	Metodología II	AHVE	AHVE
U109	En general está bien para el poco tiempo que tuviero para trabajar PERO voy a referirme a las falencias más grave que estoy detectando.	Metodología II	AHVE	AHVE
U110	Ya les comenté anteriormente por correo (y veo que no corrigieron) que respecto al problema, si la solución es una metodología ágil, tienen que hacer referencia a cuáles problemas existen con respecto a las metodologías convencionales.	Metodología II	OAH	OAH
U111	Por favor verifiquen el objetivo general.	Metodología II	AHUTEI	AHUTEI
U112	Debe ser algo con lo que la investigación desea colaborar, mejorar, en el futuro.	Metodología II	OAH	OAH
U113	Respecto a los objetivos específicos, no están utilizando los verbos sugeridos en las asignaturas MI 1 y Mi 2 (lo cual lleva a que no están planteando correctamento los obj específicos=).	Metodología II	OAH	OAH
U114	Cuando hablan de CALIDAD es necesario que justifiquen con qué ESTANDARES DE CALIDAD se están manejando, tienen que ser estándares actuales!	Metodología II	OAH	OAH
U115	Además hay que ver con mucho cuidado la relación entre calidad y usabilidad.	Metodología II	OAH	OAH
U116	Medir la usabilidad ya es un gran desafío!!!	Metodología II	OAH	OAH
U117	Qué pasó con la eficiencia?	Metodología II	AHUTEI	AHUTEI
U118	Está en la hipótesis y no está en la operacionalización.	Metodología II	OAH	OAH
U119	De qué manera relaciono en la medición la metodología propuesta con la	Metodología II	AHUTEI	AHUTEI

	calidad y eficiencia de las aplicaciones?			
U120	Quieren medir la eficiencia de la metodología o la eficiencia de una aplicación desarrollada con la metodología que se propondría?	Metodología II	AHUTEI	AHUTEI
U121	Espero que estas observaciones las orienten y las guíen.	Metodología II	AHMCDB	AHMCDB
U122	Esperemos a ver qué opina el resto de integrantes de este grupo 2!!!	Metodología II	AHPC	AHPC
U123	Saludos y seguimos en contacto!	Metodología II	AHMCDB	AHMCDB
U124	Susana Herrera	Metodología II	xxx	xxx
U125	Chicos, No olviden que Pablo ya subió la plantilla de la Matriz de Consistencia.	Metodología II	AHPC	AHPC
U126	Saludos!	Metodología II	AHMCDB	AHMCDB
U127	Estimados alumnos de MI2, Recuerden que prorrogamos la fecha de presentación del Seminario 2 para el dia 30 de Noviembre.	Metodología II	AHURH	AHURH
U128	Sin embargo, a pesar de ello, no interactúan en el foro, no hay consultas... qué pasa?	Metodología II	AHICPR	AHICPR
U129	Por un lado la falta de consultas indicaría que está todo bien...	Metodología II	OAH	OAH
U130	Por otro lado piden prórroga...	Metodología II	OAH	OAH
U131	Es necesario que todos interactúen en el foro, todos los integrantes de cada grupo.	Metodología II	AHUTEI	AHUTEI
U132	Tenemos esta posibilidad y debemos aprovecharla.	Metodología II	AHEOD	AHEOD
U133	Pedí que por este medio me envíen sus avances pero no recibí aún nada.	Metodología II	AHICPR	AHICPR
U134	Sigo esperando.	Metodología II	AHPISDD	AHPISDD
U135	Saludos!	Metodología II	AHMCDB	AHMCDB
U136	Susana Herrera	Metodología II	xxx	xxx
U137	Están disponibles en el sitio de la asignatura los resultados del Seminario 1.	Metodología II	AHURH	AHURH
U138	Además la fecha de presentación del Seminario 2 es el 24/11/10.	Metodología II	OAH	OAH
U139	Saludos!	Metodología II	AHMCDB	AHMCDB
U140	Susana Herrera	Metodología II	xxx	xxx
U141	Gracias señorita Noelia.	Estadística	AHMCDB	AHMCDB
U142	Lo vamos a ver.	Estadística	AHPC	AHPC

U143	Elvio	Estadística	xxx	xxx
U144	A pedido de un grupo de alumnos se ha creado un foro general, donde pueden participar todos los alumnos.	Estadística	AHPC	AHPC
U145	Aprovechen este foro, para informar acerca de las dudas que vayan surgiendo, para que podamos superarlas entre todos.	Estadística	AHPC	AHPC
U146	Estimados.	Estadística	AHMCDB	AHMCDB
U147	Se encuentra disponible en este sitio, la versión digital del libro Los secretos de los éxitos de la naturaleza, que ha inspirado el trabajo de Taller.	Estadística	AHURH	AHURH
U148	Es interesante que lo lean.	Estadística	AHUTEI	AHUTEI
U149	Tengo entendido que solo leyeron un capitulo.	Estadística	OAH	OAH
U150	Saludos.	Estadística	AHMCDB	AHMCDB
U151	Elvio	Estadística	xxx	xxx
U152	La fecha limite para la presentacion del trabajo es el 19 de noviembre.	Estadística	AHPTME	AHPTME
U153	No habría mucho tiempo para modificaciones razón por la cual deberan tratar de que el Informe sea ya el definitivo.	Estadística	AHPTME	AHPTME
U154	Saludos.	Estadística	AHMCDB	AHMCDB
U155	Elvio	Estadística	xxx	xxx
U156	Estimados alumnos.	Estadística	AHMCDB	AHMCDB
U157	En base a lo que conversamos propongan o definan las características (protocolo) de la interacción entre computadoras.	Estadística	AHPC	AHPC
U158	Que opinan de la propuesta del amigo Pablo Santana?.	Estadística	AHPC	AHPC
U159	Recuerden la importancia del debate "hasta las últimas consecuencias".	Estadística	AHEOD	AHEOD
U160	Saludo afectuoso.	Estadística	AHMCDB	AHMCDB
U161	Elvio	Estadística	xxx	xxx
U162	hola Janet, Alicia, Malu y una persona nueva tambien	Tactics 2004	AHDB	AHDB
U163	hola Alicia, janet ya esta aquí	Tactics 2004	AHDB	AHDB
U164	si y me dice hidalia que no tiene los contactos, asi que comienza a invitar tu Janet y alicia y yo	Tactics 2004	AHDCG	AHDCG
U165	malu dice queno sirve su msn y lo estan componiendo	Tactics 2004	AHPC	AHPC
U166	Mayra acaba de entrar	Tactics 2004	AHPC	AHPC

U167	Hola a todos!!!	Tactics 2004	AHDB	AHDB
U168	Hola a todos	Tactics 2004	AHDB	AHDB
U169	empezamos por pachuca, les parece?	Tactics 2004	AHURH	AHURH
U170	tu	Tactics 2004	AHPC	AHPC
U171	Pachuca, Cuernavaca y México	Tactics 2004	OAH	OAH
U172	Pachuca, Morelos y DF.	Tactics 2004	OAH	OAH
U173	quien más de pachuca?	Tactics 2004	AHPC	AHPC
U174	Bienvenidos Carmen y Ricardo	Tactics 2004	AHDB	AHDB
U175	Quiénes están de Morelos?	Tactics 2004	AHPC	AHPC
U176	hola soy Manuel Juárez, vivo en Cuernavaca, Morelos y he sido acompañante del proyecto tactics por tres años, estoy muy entusiasmado con la participación de nuevos elementos y espero que trabajemos colaborativamente, de a de veras!!	Tactics 2004	OAH	OAH
U177	Hola, vivo en el D.F. y fuí designada coordinadora del serminario y en esta ocasión como moderadora de la sesión.	Tactics 2004	AHDCG	AHDCG
U178	Me estoy integrando al trabajo colaborativo.	Tactics 2004	AHPC	AHPC
U179	Bien, somos Chely y Ricardo de Pachuca, Mayra, Manuel y Janet de Morelos, Alicia y yo del D. F.	Tactics 2004	AHDCG	AHDCG
U180	hola Tony!!	Tactics 2004	AHDB	AHDB
U181	Hola Tony, nos estamos presentando	Tactics 2004	AHDB	AHDB
U182	te puedes presentar	Tactics 2004	AHDCG	AHDCG
U183	Efectivamente, es muy caótico, pero debemos terminar la presentación de todos.	Tactics 2004	AHDCG	AHDCG
U184	Mayra esta jugando	Tactics 2004	AHTB	AHTB
U185	podemos pasar a comentar las experiencias brevemente, porque debemos nombrar al moderador de la siguiente semana	Tactics 2004	AHDFI	AHDFI
U186	Vamos a las experiencias de quienes han trabajado en tactics	Tactics 2004	AHUTEI	AHUTEI
U187	Les parece que mayra inicie?	Tactics 2004	AHDFI	AHDFI
U188	Rosamaría y Ricardo son los únicos con experiencia en el proyecto?	Tactics 2004	AHPC	AHPC
U189	Manuel y Malú, nos pueden comentar su experiencia?	Tactics 2004	AHPC	AHPC

U190	cuántos están de acuerdo con que sea Manuel?	Tactics 2004	AHDFI	AHDFI
U191	votemos ahora	Tactics 2004	AHDFI	AHDFI
U192	estamos votando por el moderador, dspués la fecha	Tactics 2004	AHDCG	AHDCG
U193	Tony y yo ya votamos por manuel	Tactics 2004	OAH	OAH
U194	sugiero que la fecha del chat, la acordemos por correo	Tactics 2004	AHDFI	AHDFI
U195	me uno a la propuesta, haremos un poll en tactics-mexico	Tactics 2004	AHPC	AHPC
U196	lo discutimos junto con los otros de Johnson y Johnson	Tactics 2004	AHDFI	AHDFI
U197	y la discusión comienza en el foro	Tactics 2004	AHDFI	AHDFI
U198	Se enviarían comentarios a través del grupo a partir de mañana, les parece?	Tactics 2004	AHDFI	AHDFI
U199	a partir de que pongamos todos nuestros comentarios	Tactics 2004	AHDFI	AHDFI
U200	Entonces, se sugiere revisar todos los documentos y mandar comentarios	Tactics 2004	AHDFI	AHDFI
U201	Janet, podrías grabar esta conversación y mandarmela mañana por correo electrónico, estoy en un cibercafe	Tactics 2004	AHPC	AHPC
U202	que en los primeros 5 minutos hagamos comentarios sobre el documento de presentación y el resto concluyamos el foro sobre las bases teoricas del aprendizaje colaborativo	Tactics 2004	AHDFI	AHDFI
U203	Hola buenas noches	Tactics 2004	AHDB	AHDB
U204	Déjame ver	Tactics 2004	OAH	OAH
U205	No tengo a Ricardo pero no a Hugo	Tactics 2004	OAH	OAH
U206	Hola Rosamaría	Tactics 2004	AHDB	AHDB
U207	Hola a todos	Tactics 2004	AHDB	AHDB
U208	Hola Rosamari	Tactics 2004	AHDB	AHDB
U209	Manuel: me resulta difícil leer con ese color.	Tactics 2004	AHUEA	AHUEA
U210	Puedes cambiarlo	Tactics 2004	AHUEA	AHUEA
U211	Mientran se nos unen los demás, se reciben propuestas para la orden del chat	Tactics 2004	AHDFI	AHDFI
U212	Gracias.	Tactics 2004	AHMCDB	AHMCDB
U213	Moción de orden por qué no empezamos a proponer la secuencia del chat mientras solucionamos la incorporación de los miembros que faltan del grupo.	Tactics 2004	AHDFI	AHDFI

U214	Constantino todavía la letra está muy pequeña .	Tactics 2004	OAH	OAH
U215	Perdón.....	Tactics 2004	AHMCDB	AHMCDB
U216	Ahora si que chino	Tactics 2004	OAH	OAH
U217	Por favor propongan el primer tema a discutir porque sino se nos va el tiempo como la vez pasada	Tactics 2004	AHPISDD	AHPISDD
U218	Se reciben propuestas	Tactics 2004	AHPC	AHPC
U219	Me parece bien lo que propone Rosamari que se ajusta a la lectura del segundo capítulo de Jhonson y nos serviría para poder proponer el tipo de equipo para el nuevo el ejercicio de Tactics	Tactics 2004	AHVE	AHVE
U220	Opinen por favor sobre la propuesta	Tactics 2004	AHPC	AHPC
U221	Bueno entonces tenemos dos aspectos el tamaño del equipo y los temas, ¿qué deciden?	Tactics 2004	AHUTEI	AHUTEI
U222	OK, ya hay tres propuestas , los demás podrían apoyar a alguna o proponer otras	Tactics 2004	AHDFI	AHDFI
U223	Entonces debemos proponer	Tactics 2004	AHPC	AHPC
U224	tenemos que escoger uno para desglosarlo en las partes que vamos a discutir	Tactics 2004	AHDFI	AHDFI
U225	Por supuesto que tenemos que tomar en cuenta a todosç	Tactics 2004	AHDCG	AHDCG
U226	Si le estoy diciendo a Hugo que con respecto al chat no está muy atrasado aunque haya entrado hoy por primera vez , si ha leído las lecturas	Tactics 2004	AHTNP	AHTNP
U227	Vuelvo a proponer que comencemos a discutir , alguno de los temas ya casi llevamos 20 min.y no nos hemos puesto de acuerdo	Tactics 2004	AHDFI	AHDFI
U228	Me parece bien, podemos por turno emitir nuestra opinión sobre cada una de las lecturas	Tactics 2004	AHDFI	AHDFI
U229	Sería bueneo entre todos resiumir ya que hemos leído lo que han escrito los demás lo que cada lectura nos aporta para planear esta nueva etapa del proyecto	Tactics 2004	AHUTEI	AHUTEI
U230	¡Por qué no comienzas Manuel y vamos abundando sobre lo que resumas?	Tactics 2004	AHPISDD	AHPISDD
U231	Bueno comenzaríamos sobre la primera , se reciben comentarios.	Tactics 2004	AHUTEI	AHUTEI
U232	Estoy de acuerdo	Tactics 2004	AHDCG	AHDCG
U233	Por el momento no puedo permanecer en el foro pero, como secretario, me	Tactics 2004	OAH	OAH

	comprometo a subir el resumen el dia de mañana.			
U234	Adios	Tactics 2004	AHMCDB	AHMCDB
U235	podías ser más explícita	Tactics 2004	AHFE	AHFE
U236	Hola Rosamari , estamos hablando de cas características del profesor como facilitadopr	Tactics 2004	AHDB	AHDB
U237	Hola Mayra	Tactics 2004	AHDB	AHDB
U238	Si como dice Carmen Aracely hay que por lo menos sacar conclusiones de lo que hemos discutido sobre el rol del maestro	Tactics 2004	AHDFI	AHDFI
U239	O.K entonces mandamos los comentarios al for	Tactics 2004	AHPC	AHPC
U240	Muy bien byeç	Tactics 2004	AHUEA	AHUEA
U241	hola Hidalia, si pero ya vino la secre de este asunto	Tactics 2004	AHDB	AHDB
U242	Hola a todos, Buenas Noches!!!	Tactics 2004	AHDB	AHDB
U243	Hola, ya entre a esta otra ventana (creo)	Tactics 2004	AHDB	AHDB
U244	Hola Rosa María, si ya estas con nosotros	Tactics 2004	AHDB	AHDB
U245	NO importa Ricardo, de la Prepa 1 no hay nadie	Tactics 2004	OAH	OAH
U246	o quieren ya comenzar con la primera pregunta?	Tactics 2004	AHDFI	AHDFI
U247	Vale, comienza Mayra o Janet	Tactics 2004	AHPC	AHPC
U248	Vale mientras se encuentra podría comenzar alguien del Madrid, del Die o Ricardo	Tactics 2004	AHPC	AHPC
U249	Hidalia por que identificas trabajo colaborativo con resolución de problemas?	Tactics 2004	AHFE	AHFE
U250	La experiencia resultó muy rica, pero efectivamente faltó tiempo para una discusión y un resumen final, que espero hacer la siguiente clase.	Tactics 2004	AHPPAA	AHPPAA
U251	Lo cierto es que les gustó tener que ver con lo que hacían todos los demás	Tactics 2004	AHVE	AHVE
U252	quieren hacer una última aportación?	Tactics 2004	AHPC	AHUTEI
U253	Hola Monstruo de Gila	Tactics 2004	AHDB	AHDB
U254	¿Quien eres?	Tactics 2004	AHURH	AHURH
U255	vamos preparando respuestas mientras Rosa Maria termina de escribir el tomo II	Tactics 2004	AHUTEI	AHPC
U256	Fin del tomo II (Hacer resumenes de tanta gente no es fácil)	Tactics 2004	OAH	OAH

U257	hola!	Tactics 2004	AHDB	AHDB
U258	¿no les parece?	Tactics 2004	AHPC	AHUTEI
U259	Hidalia, tu como ves?	Tactics 2004	AHFE	AHFE
U260	No me lo van a creer pero estamos SOBRE EL TIEMPO NUEVAMENTE, una última ronda y seguimos con la pregunta siguiente, le podríamos ayudar a Rosa María.	Tactics 2004	AHPISDD	AHACI
U261	Justamente a la distancia es todavía mucho más difícil.	Tactics 2004	AHESS	AHESS
U262	Con todo y que conozco a varios esto cuesta mucho de procesar, al menos para nosotros que no somos de la generación cibernética.	Tactics 2004	AHESS	AHESS
U263	A lo mejor para los chavos es más fácil, pero mi experiencia con los grupos no lo demuestra así.	Tactics 2004	AHESS	AHESS
U264	La proxima semana nos toca ver a tactics, creo que esta pregunta se podría ampliar para la semana siguiente	Tactics 2004	AHPTME	AHPTME
U265	Estamos NUEVAMENTE SOBRE EL TIEMPO, alguien quiere aportar algo más	Tactics 2004	AHUTEI	AHUTEI
U266	o seguimos con la última pregunta	Tactics 2004	AHDFI	AHDFI
U267	¿Qué fue lo que más me impactó de este texto?	Tactics 2004	AHFE	AHFE
U268	¿por qué	Tactics 2004	AHFE	AHFE
U269	Lo largo (que sangrona).	Tactics 2004	AHUEA	AHUEA
U270	El trabajo sistemático del autor y los múltiples indicadores que tiene para fundamentar cada una de sus aseveraciones.	Tactics 2004	OAH	OAH
U271	De verdad creen que son tan diferentes?	Tactics 2004	AHFE	AHFE
U272	Es bueno considerar todo eso, porque Tactics - Mexico se supone que trabaja con la diversidad de escuelas, estilos, planes, culturas y estudiantes!!	Tactics 2004	OAH	OAH
U273	VAle pues si no tienen otra cosa que agregar	Tactics 2004	AHUEA	AHUEA
U274	OK Bye bye.	Tactics 2004	AHUEA	AHUEA
U275	Saludos a todos	Tactics 2004	AHMCDB	AHMCDB
U276	Gracias Rosa María!!	Tactics 2004	AHMCDB	AHMCDB
U277	De nada	Tactics 2004	AHMCDB	AHMCDB
U278	Gracias Manuel y hasta la proxima.	Tactics 2004	AHMCDB	AHMCDB

U279	Ora si ya me tengo que ir	Tactics 2004	AHMCDB	AHMCDB
U280	Sales, lo mando en cuanto pueda.	Tactics 2004	OAH	OAH
U281	Chau	Tactics 2004	AHUEA	AHUEA
U282	Vale, pues creo que la semblanza la podriamos hacer entre varios, recuerden que yo no estuve desde el inicio	Tactics 2004	AHPISDD	AHPISDD
U283	Asi que si Hugo, Mayra y Janet me ayudan o alguien más, pues creo que la podríamos hacer	Tactics 2004	AHPC	AHPC
U284	Aun esta la propuesta de trabajar en subequipos	Tactics 2004	AHDFI	AHDFI
U285	y ya con base en su retroalimentación podríamos hacer la semblanza	Tactics 2004	AHDFI	AHDFI
U286	Ricardo, tu les puedes pasar el chat a tus compañeros?	Tactics 2004	AHDCG	AHDCG
U287	Vale, entonces por favor diles que consulten sus correos electrónicos para incluir sus experiencias en del proyecto en la semblanza, bueno en realidad solo la tuya y la de Hugo	Tactics 2004	AHDCG	AHDCG
U288	Vale, adios.	Tactics 2004	AHMCDB	AHMCDB
U289	hola	Tactics 2004	AHDB	AHDB
U290	Hola Hugo.	Tactics 2004	AHDB	AHDB
U291	Nada.	Tactics 2004	OAH	OAH
U292	Yo la pongo, gracias.	Tactics 2004	AHMCDB	AHMCDB
U293	Manuel, tu estas invitando a ALicia y Pati?	Tactics 2004	AHACI	AHACI
U294	Si, invita a todos como van llegando.	Tactics 2004	AHACI	AHACI
U295	Si no llega Tere puedes ser secre hasta que llegue ella?	Tactics 2004	AHDFI	AHDFI
U296	hola Alicia y Pati.	Tactics 2004	AHDB	AHDB
U297	Pero puedes dejar tus aportaciones con Alicia!	Tactics 2004	AHACI	AHACI
U298	La primera pregunta podemos empezar con Hugo y Pati:	Tactics 2004	AHUTEI	AHUTEI
U299	¿Cuál fue su motivación para entrar al proyecto?	Tactics 2004	AHPC	AHPC
U300	¿Cumplió con sus expectativas?	Tactics 2004	AHPC	AHPC
U301	Hugo, quieres empezar?	Tactics 2004	AHPC	AHPC
U302	Manuel, no puedes llenar el vacio?	Tactics 2004	AHPC	AHPC
U303	Cuales son tus expectativas, Hugo?	Tactics 2004	AHPC	AHPC
U304	Crees que ha salido bien?	Tactics 2004	AHESS	AHEESS

U305	Rosa, tu puedes contestar despues como Pati ya salió.	Tactics 2004	AHPC	AHPC
U306	Quieres especificar mas?	Tactics 2004	AHFE	AHFE
U307	Hugo, tú piensas lo mismo que Rosa Mari sobre el trabajo de los alumnos	Tactics 2004	OAH	OAH
U308	muy bien.	Tactics 2004	AHPPAA	AHPPAA
U309	me sentí muy solita!	Tactics 2004	AHESS	AHESS
U310	Ya estas, Hugo.	Tactics 2004	OAH	OAH
U311	Ahora que hacemos?	Tactics 2004	AHPC	AHPC
U312	Quieren esperar un rato para reiniciar las preguntas?	Tactics 2004	AHDFI	AHDFI
U313	Hugo, tu eres el unico maestro!	Tactics 2004	OAH	OAH
U314	Pero donde ven las fallas principales?		AHFE	AHFE
U315	Perdon la interrupcion, pero que quieren hacer como somos tan pocos ahora?	Tactics 2004	AHDFI	AHDFI
U316	Seguir o pedimos que cada quien escriba sus respuestas y subirlas al grupo?	Tactics 2004	AHDFI	AHDFI
U317	Hugo, mejor hacemos una respuesta a cada pregunta que les mande y cada quien y subirla al chat.	Tactics 2004	AHDCG	AHDCG
U318	Gracias Alicia pero yo puedo mandar un mensaje porque suspendimos la sesion y que hacer.	Tactics 2004	OAH	OAH
U319	Manuel, tu tienes el chat completo para grabarlo?	Tactics 2004	OAH	OAH
U320	Gracias.	Tactics 2004	AHMCDB	AHMCDB
U321	y estamos en contacto.	Tactics 2004	AHMCDB	AHMCDB
U322	igualmente.	Tactics 2004	AHMCDB	AHMCDB
U323	Hola Luisa, los ejercicios y retos están en el contenido de cada uno de los documentos.	Tactics 2007	AHDB	AHDB
U324	José Luis	Tactics 2007	xxx	xxx
U325	Hola Luisa, la negación es correcta, aunque se vea un poco rara, pero con el diagrama de Veen correspondiente se puede apreciar mejor que ¬(A') = A.	Tactics 2007	OAH	OAH
U326	José Luis	Tactics 2007	xxx	xxx
U327	Hola Christian, te agradezco la información que nos has proporcionado, es interesante y que bueno que tenemos un experto en conjuntos difusos.	Tactics 2007	AHPPAA	AHPPAA
U328	Ya aprovecharemso tu experiencia cunado estes por aquí.	Tactics 2007	AHMCDB	AHMCDB

U329	José Luis.	Tactics 2007	xxx	xxx
U330	Hola a todos	Tactics 2007	AHDB	AHDB
U331	Hola Cinthya, Maribel, Lazaro,Omar, Yanet.	Tactics 2007	AHDB	AHDB
U332	Hola Cinthya	Tactics 2007	AHDB	AHDB
U333	Hola Ana, te aclaro que NO SOY la Maestra NDREA, sino el profesor JOSE LUIS.	Tactics 2007	AHDB	AHDB
U334	Hasrta luego aLázaro	Tactics 2007	AHMCDB	AHMCDB
U335	Si Cinthya, deben poner las respuestas en el folder EJERCICOS Y RETOS de su grupo	Tactics 2007	AHFDNMC	AHFDNMC
U336	Hola Maribel y Ana	Tactics 2007	AHDB	AHDB
U337	Si Ana, pero puedes preguntar las dudas que tengas desde hoy.	Tactics 2007	AHUTEI	AHUTEI
U338	Hola It	Tactics 2007	AHDB	AHDB
U339	Acaración para It que acaba de entrar al chat	Tactics 2007	AHTNP	AHTNP
U340	eso espero Omar	Tactics 2007	AHMC	AHMC
U341	Si Maribel	Tactics 2007	OAH	OAH
U342	Maribel: puedes subir tus dudas y preguntas al foro general, no al de tu equipo.	Tactics 2007	AHDFI	AHDFI
U343	Pero primero deben tratar de resolver las duda entre ustedes	Tactics 2007	AHDFI	AHDFI
U344	El procedimeinto es el siguiente: leer y tratar de resolver los ejercicos y retos, si hay dudas, consultar con mi compañero de equipo, si no se resuelven, hacer la pregunat en el foro, si aún hay dudas, les daré pistas para resolverlas	Tactics 2007	AHPTME	AHPTME
U345	Y DE SOBRE LAS PREGUNTAS QUE HAYAN PLATEADO LOS EQUIPOS	Tactics 2007	AHPTME	AHPTME
U346	Hola Rubi, si tienes compañero es: Francisco Leon en el equipo GMMA	Tactics 2007	AHPC	AHPC
U347	NO se les olvide poner su presentación personal en su FOLDER, así nos conoceremos mejor	Tactics 2007	AHFCP	AHFCP
U348	Si, pero pondremos una carpeta de presentaciones con todos los participantes.	Tactics 2007	AHFCP	AHFCP
U349	ESTOY ESPERANDO SUS PREGUNTAS	Tactics 2007	AHUTEI	AHUTEI
U350	Hasta luego Cinthya	Tactics 2007	AHMCDB	AHMCDB

U351	Hasta Luego Cristian	Tactics 2007	AHMCDB	AHMCDB
U352	Hola de nuevo	Tactics 2007	AHDB	AHDB
U353	LES INFORMO que a partir de este moemnto dejo el chat.	Tactics 2007	AHMCDB	AHMCDB
U354	RECIBAN UN CORDIAL SALUDO.	Tactics 2007	AHMCDB	AHMCDB
U355	Hola a todos	Tactics 2007	AHDB	AHDB
U356	Espero que tengan preguntas concretras sobre los ejercicios y retos	Tactics 2007	AHPC	AHURH
U357	O sobre la forma de trabajo	Tactics 2007	AHPC	AHSCTT
U358	O quizá tengan comentarios sobre los materiales???	Tactics 2007	AHPC	AHSCTT
U359	¿Nadie se atreve a preguntar?	Tactics 2007	AHUTEI	AHUTEI
U360	Hola Itzel	Tactics 2007	AHDB	AHDB
U361	comentame tus dudas	Tactics 2007	AHFE	AHFE
U362	¿por qué te confundes?	Tactics 2007	AHFE	AHFE
U363	¿qué opinas ahora??	Tactics 2007	AHSCTT	AHSCTT
U364	¿qué te pareció el material sobre predicados?	Tactics 2007	AHSCTT	AHSCTT
U365	hasta luego	Tactics 2007	AHMCDB	AHMCDB
U366	Ok, espero que en la próxima asesoria haya más preguntas.	Tactics 2007	AHPPAA	AHPPAA
U367	hasta la próxima.	Tactics 2007	AHMCDB	AHMCDB
U368	Hola a todos	Tactics 2007	AHDB	AHDB
U369	Buenso días	Tactics 2007	AHDB	AHDB
U370	Hola Cinthya	Tactics 2007	AHDB	AHDB
U371	Hola Ana	Tactics 2007	AHDB	AHDB
U372	Hola Andrea	Tactics 2007	AHDB	AHDB
U373	Hola Omar	Tactics 2007	AHDB	AHDB
U374	Hola Oscar	Tactics 2007	AHDB	AHDB
U375	A ver Ana, comienza	Tactics 2007	AHURH	AHURH
U376	Nos puedes dar algun ejemplo?	Tactics 2007	AHFE	AHFE
U377	Hoal Yanet	Tactics 2007	AHDB	AHDB
U378	si, vas bien	Tactics 2007	AHPPAA	AHPPAA
U379	pero necsitamos aclarar bien cual es el enunciaod que quieres formalizar	Tactics 2007	AHFE	AHFE
U380	No entiendo lo de textual ,cinthya	Tactics 2007	AHFE	AHFE

U381	Las dos son corrctas Cinthya	Tactics 2007	AHLC	AHPPAA
U382	Hola Luisa	Tactics 2007	AHDB	AHDB
U383	para mejorar la comunicación les prpongo lo siguiente: Que cada uno por uno platee una duda y entre todos la aclaremos	Tactics 2007	AHDFI	AHDFI
U384	si Luisa	Tactics 2007	AHPPAA	AHPPAA
U385	¿Quién le responde a Rubí?	Tactics 2007	AHPISDD	AHPISDD
U386	Es correcto Christian	Tactics 2007	AHPPAA	AHPPAA
U387	y despues vemos el de Rubí.	Tactics 2007	OAH	OAH
U388	ahora ¿qué podemos hacer?	Tactics 2007	AHPC	AHPC
U389	si, y qué más?	Tactics 2007	AHFE	AHFE
U390	deja revisar tus ejercicos israel, espera un minuto	Tactics 2007	AHTD	AHTD
U391	Israel, No puedo revisar tus ejercicos si no los has subido, he visto los de Maribel.	Tactics 2007	OAH	OAH
U392	LES RECURSO QUE ES MUY IMPORTANTE QUE COMENTE LAS SOLUCIONES ENTRE USTEDES	Tactics 2007	AHPC	AHPC
U393	oK, LUISA, CUÁL ES TU DUDA	Tactics 2007	AHFE	AHFE
U394	claro Rubí,	Tactics 2007	AHPPAA	AHPPAA
U395	Ok israel, te toca, gracias por tu paciencia	Tactics 2007	AHMCDB	AHMCDB
U396	Ok israel, pero no los podre revisar en este momento, porque tendría que salir del chat y despues volver.	Tactics 2007	AHTD	AHTD
U397	Ok Cinthya, mañana en la mañana tendrán algunas observaciones, aún no he terminado de revisar.	Tactics 2007	AHTD	AHTD
U398	RECUERDE QUE EN ESTE CURSO NOS RETROALIMENTAMOS DE LAS DUDAS Y PREGUNTAS DE TODOS	Tactics 2007	AHPC	AHPC
U399	oK, ME TENGO QUE retirar, seguimos en comunicación	Tactics 2007	AHMCDB	AHMCDB
U400	Así es Luisa.	Tactics 2007	AHPPAA	AHPPAA
U401	Saludos a todos.	Tactics 2007	AHMCDB	AHMCDB
U402	hasta el próximo martes.	Tactics 2007	AHMCDB	AHMCDB
U403	Hola a todos	Tactics 2007	AHDB	AHDB
U404	Hola Cinthya	Tactics 2007	AHDB	AHDB

U405	o preguntas o comentarios	Tactics 2007	AHURH	AHURH
U406	claro, Cinthya es tu turno	Tactics 2007	AHACI	AHACI
U407	Hola israel, este es el turno de los grupos Beta y Delta	Tactics 2007	AHMCDB	AHMCDB
U408	si, Cinthya	Tactics 2007	AHPPAA	AHPPAA
U409	aclara eso que dices	Tactics 2007	AHFE	AHFE
U410	hola Omar, este es el turno de los equipos Beta y Delta	Tactics 2007	AHMCDB	AHMCDB
U411	¿de qué forma has dibujado los círculos?	Tactics 2007	AHFE	AHFE
U412	Hola Yanet	Tactics 2007	AHDB	AHDB
U413	Yanet, te has podido comunicar con Ana?	Tactics 2007	AHMR	AHMR
U414	Te sugiero que comentes con ella el análisis de las definiciones	Tactics 2007	AHPC	AHPC
U415	a lo mejor hasta logran una presentación dinámica de una de las igualdades en le Power Point o en Flash	Tactics 2007	AHPC	AHPC
U416	Estoy seguro de que pueden dejar fluir su imaginación	Tactics 2007	AHESS	AHESS
U417	qué otra duda o pregunta o comentario tienes?	Tactics 2007	AHSCTT	AHSCTT
U418	¿y qué te pareció lo que has leído?	Tactics 2007	AHFE	AHFE
U419	Hola Marible, estmos en el turo de los grupos Dseta y Eta	Tactics 2007	AHMCDB	AHMCDB
U420	Maribel, no s e si leíste el coreo, pero te hemos reasignado de equipo,	Tactics 2007	AHPC	AHPC
U421	ok Maribel	Tactics 2007	AHPPAA	AHPPAA
U422	¿qué opinas Yanet?	Tactics 2007	AHTM	AHTM
U423	Ok Itzel, y cúales son tus preguntas, dudas o comentarios?	Tactics 2007	AHFE	AHFE
U424	Ok Omar	Tactics 2007	AHPPAA	AHPPAA
U425	si, Itzel continua	Tactics 2007	AHPC	AHPC
U426	Itzel, puedes decirnos con más precisión qué dice el texto al que haces referencia?	Tactics 2007	AHFE	AHFE
U427	puedes citar el párrafo completo?	Tactics 2007	AHFE	AHFE
U428	que consta de:???	Tactics 2007	AHFE	AHFE
U429	Omar, cuál es tu pregunta	Tactics 2007	AHFE	AHFE
U430	Maribell, hoy pondré en tu nuevo folder tu autoevaluación.	Tactics 2007	OAH	OAH
U431	¿qué opinan?	Tactics 2007	AHUTEI	AHUTEI
U432	a otro tipo de representaciones ¿cómo cuáles?	Tactics 2007	AHFE	AHFE

U433	No capto lo de las 256 posibles combinaciopnes	Tactics 2007	AHFE	AHFE
U434	Ok, gracias Sergio	Tactics 2007	AHPPAA	AHPPAA
U435	Ok Omar,	Tactics 2007	AHPPAA	AHPPAA
U436	Seguiremos en cominicación Sergio	Tactics 2007	AHMCDB	AHMCDB
U437	Hola Maribell	Tactics 2007	AHDB	AHDB
U438	Hasta luego, maribell	Tactics 2007	AHMCDB	AHMCDB
U439	Nos vemos el próximo jueves.	Tactics 2007	AHMCDB	AHMCDB
U440	Hola a Todos	Tactics 2007	AHDB	AHDB
U441	Hola Israel,	Tactics 2007	AHDB	AHDB
U442	espero que hayn podido comentar algo entre los integrantes del equipo	Tactics 2007	AHURH	AHURH
U443	y tengan preguntas grupales	Tactics 2007	AHURH	AHURH
U444	describela en lenguaje natural	Tactics 2007	AHFE	AHFE
U445	aprovechando: ¿qué te pareció el material sobre análisis de las definciones?	Tactics 2007	AHSCTT	AHSCTT
U446	me gustaría saber tu opinión	Tactics 2007	AHFE	AHFE
U447	en qué sentido?	Tactics 2007	AHFE	AHFE
U448	por qué crees que se complicaría su aplicación?	Tactics 2007	AHFE	AHFE
U449	si puedes hacer el análisis de la definicón estaría muy bien	Tactics 2007	AHPPAA	AHPPAA
U450	aprovecho para comentarte que revice lo que hiciste con el análisis de la pre y post condición d elos algoritmos	Tactics 2007	OAH	OAH
U451	y me pareció muy interesante.	Tactics 2007	AHESS	AHESS
U452	Ok israel, seguimos en comunicación	Tactics 2007	AHMCDB	AHMCDB
U453	Hasta luego	Tactics 2007	AHMCDB	AHMCDB
U454	Hola Rubi	Tactics 2007	AHDB	AHDB
U455	Ok, dime	Tactics 2007	AHPC	AHPC
U456	qué números?	Tactics 2007	AHFE	AHFE
U457	Hola Cristian, estamos comentando sobre los conjuntos borrosos	Tactics 2007	AHDB	AHDB
U458	Si, comentas cuál es?	Tactics 2007	AHFE	AHFE
U459	Ok Rubí	Tactics 2007	AHPPAA	AHPPAA
U460	Cuéntanos Christian	Tactics 2007	AHPC	AHPC
U461	¿qué información queires compartir?, Christian	Tactics 2007	AHFE	AHFE

U462	sobre qué fue?	Tactics 2007	AHFE	AHFE
U463	y comentanos lo que has pensado al respecto	Tactics 2007	AHFE	AHFE
U464	¿por qué necesitas convertirlo a nítido?	Tactics 2007	AHFE	AHFE
U465	ok, ha sido interesante tu paticiapación, Christian,	Tactics 2007	AHVE	AHVE
U466	gracias tambien a Rubi	Tactics 2007	AHPPAA	AHPPAA
U467	segundo no entiendo lo de las matrices, pues aclarar la pregunta?	Tactics 2007	AHFE	AHFE
U468	Ok, veamos Cinthya	Tactics 2007	AHERM	AHERM
U469	Algún comentario sobre las lecturas?	Tactics 2007	AHFE	AHFE
U470	¿qué les pareció el material sobre "análisis de las definiciones?	Tactics 2007	AHFE	AHFE
U471	Hasta luego.	Tactics 2007	AHMCDB	AHMCDB
U472	Hola a todos	Tactics 2007	AHDB	AHDB
U473	Hola Cinthya	Tactics 2007	AHDB	AHDB
U474	¿Qué comentarios, dudas o preguntas tienes?	Tactics 2007	AHFE	AHFE
U475	Ok, a ver Maribel, ,	Tactics 2007	AHUTEI	AHUTEI
U476	por ejemplo...	Tactics 2007	AHFE	AHFE
U477	Cintjya revisa tu coreo ya te envié el documento	Tactics 2007	AHPPAA	AHPPAA
U478	especifica Maribell	Tactics 2007	AHFE	AHFE
U479	Ok Maribel, espero que se vayna resolviendo tus dudas	Tactics 2007	AHMCDB	AHMCDB
U480	Hola Sergio	Tactics 2007	AHDB	AHDB
U481	qué comentarios, dudas o preguntas tienes?	Tactics 2007	AHSCTT	AHSCTT
U482	a ver, cuál?	Tactics 2007	AHFE	AHFE
U483	¿cuál sería la función?	Tactics 2007	AHFE	AHFE
U484	¿qué dominio estás considerando?	Tactics 2007	AHFE	AHFE
U485	que consta de ??	Tactics 2007	AHFE	AHFE
U486	es decir, sus componentes son:	Tactics 2007	AHFE	AHFE
U487	un elemento de la función tiene la siguiente forma: (nombre completo, ...) ¿cuales son los otros componentes	Tactics 2007	AHFE	AHFE
U488	Hola Omar	Tactics 2007	AHDB	AHDB
U489	no entiendo lo que me quieres decir	Tactics 2007	AHFE	AHFE
U490	Ok, nos vemos el jueves,	Tactics 2007	AHUEA	AHUEA

U491	Alguien más desea preguntar o comentar algo?	Tactics 2007	AHSCTT	AHSCTT
U492	Ok, nos "vemos" el proximo jueves 2 de agosto	Tactics 2007	AHUEA	AHUEA
U493	reciban un cordial saludo.	Tactics 2007	AHMCDB	AHMCDB
U494	Buenos dias a todos	Cenidet 2008	AHDB	AHDB
U495	¡¡BIENVENIDOS!!	Cenidet 2008	AHDB	AHDB
U496	Les invitamos a participar activamente en el curso, eso les permitirá prepararse mejor para el curso de Matemáticas Discretas.	Cenidet 2008	AHURH	AHURH
U497	Hola a todos, BIENVENIDOS a esta nueva forma de aprender!	Cenidet 2008	AHDB	AHDB
U498	Vaya, que gusto ver que hay mucha audiencia	Cenidet 2008	AHPPAA	AHPPAA
U499	buenas tardes Miguel Angel	Cenidet 2008	AHDB	AHDB
U500	¿Qué opinan de esa presentación?	Cenidet 2008	AHFCP	AHFCP
U501	¿qué significa para ustedes comprender un concepto matemático o computacional?	Cenidet 2008	AHFE	AHFE
U502	Algo más básico Guadalupe	Cenidet 2008	OAH	OAH
U503	Que deben tratar de responder primero de forma individual y si tienen dudas consultar con su compañero de equipo	Cenidet 2008	AHDFI	AHDFI
U504	si aún tienen dudas o no llegan a acuerdos, deben plantear sus preguntas o dudas al grupo en general	Cenidet 2008	AHDFI	AMDFI
U505	Todos cuentan con las "Herramientas de colaboración grupal"	Cenidet 2008	AHACI	AHACI
U506	Espero que entre hoy y mañana lo hagan, de lo contrario, modificaremos los equipos para puedan trabajar en pareja.	Cenidet 2008	AHACI	AHACI
U507	Fernanda: que desea saber?	Cenidet 2008	AHFE	AHFE
U508	Hola Fernando, no te preocupes, nosotros somos los que te indicaremos a que grupo perteneces	Cenidet 2008	AHDCG	AHDCG
U509	RECUERDEN: primero debe haber un trabajo individual, intentanto reslover los ejercicios y retos, despues deben trabajar en equipo.	Cenidet 2008	AHDCG	AHDCG
U510	No se preocupen, como fue muy rápido el proceso, algunos alumnos entran después, los esperaremos (no en el chat) pero si a que se pongan en contacto con nosotros	Cenidet 2008	AHMCA	AHMCA
U511	en caso de no ser así, se volveran a distribuir los alumnos que si estan activos en los grupos, para que realmente puedan trabajar en equipo	Cenidet 2008	AHFDNMC	AHFDNMC

U512	Debe QUEDAR EVIDENCIA DE SU TRABAJO EN EQUIPO.	Cenidet 2008	AHACI	AHACI
U513	Recuerden que el ÚLTIMO RECURSO para resolver dudas son los profesores.	Cenidet 2008	AHPISDD	AHPISDD
U514	Bien, me gusta ese ánimo	Cenidet 2008	AHESS	AHPPAA
U515	Hola Ana	Cenidet 2008	AHDB	AHDB
U516	Cómo esta?	Cenidet 2008	AHMCDB	AHMCDB
U517	Quisieras presentarte con nuestros alumnos?	Cenidet 2008	AHDCG	AHDCG
U518	Ok, Jóvenes, espramos su participación activa.	Cenidet 2008	AHMC	AHMC
U519	OK, Jóvenes, esperamos su participación activa.	Cenidet 2008	AHMC	AHMC
U520	Ok, los dejo con Ana.	Cenidet 2008	AHMCDB	AHMCDB
U521	Nos vemos el próximo jueves.	Cenidet 2008	AHUEA	AHUEA
U522	Hasta luego a todos.	Cenidet 2008	AHMCDB	AHMCDB
U523	Gracias Ana, te mando un gran abrazo desde cuernavaca	Cenidet 2008	AHUEA	AHUEA
U524	Un abrazo y buenas noches barcelona!!	Cenidet 2008	AHUEA	AHUEA
U525	Hasta luego Ana.	Cenidet 2008	AHMCDB	AHMCDB
U526	Ahora si, jóvenes, nos vemos el Jueves.	Cenidet 2008	AHMCDB	AHMCDB
U527	Antes de que empiecen a despedirse los felicito a todos y BIENVENIDOS!!	Cenidet 2008	AHVE	AHVE
U528	No te preocupes antonio, esperemos uqe se conecte en la próxima sesión	Cenidet 2008	AHMCA	AHMCA
U529	Bien, cuidense, trabajen y nos vemos el jueves	Cenidet 2008	AHUEA	AHUEA
U530	perdón Aquino	Cenidet 2008	OAH	OAH
U531	de nada.	Cenidet 2008	AHMCDB	AHMCDB
U532	Me despido, hasta luego	Cenidet 2008	AHMCDB	AHMCDB
U533	hola a todos	Cenidet 2008	AHDB	AHDB
U534	Hola a todos	Cenidet 2008	AHDB	AHDB
U535	Ok Lizeth, es tu turno de preguntar	Cenidet 2008	AHPC	AHPC
U536	¿Esa sería una oración equivalente a la primera?	Cenidet 2008	AHFE	AHFE
U537	Hola MArino.	Cenidet 2008	AHDB	AHDB
U538	Gracias por tu respuesta.	Cenidet 2008	AHVE	AHVE
U539	Eres del grupo Alfa?	Cenidet 2008	AHPC	AHPC

U540	permitenos un momento MArino	Cenidet 2008	AHMCDB	AHMCDB
U541	Qué opinas Lizeth??	Cenidet 2008	AHTNP	AHTNP
U542	Observa que se trata de un or exclusivo .	Cenidet 2008	OAH	OAH
U543	Se nos termino el tiempo.	Cenidet 2008	OAH	OAH
U544	Te propongo que nos envies tus dudas al foro general.	Cenidet 2008	AHPC	AHPC
U545	Marino, estuviste en la presentaciòn el dia martes?	Cenidet 2008	AHICPR	AHICPR
U546	Son Fernando y Maribel.	Cenidet 2008	AHPC	AHPC
U547	Perdón, me perdí.	Cenidet 2008	AHMCDB	AHMCDB
U548	Quien es el equipo beta?	Cenidet 2008	AHPC	AHPC
U549	Son Fernando y Maribel	Cenidet 2008	AHPC	AHPC
U550	Marino, ya no supe si viste mi mensaje.	Cenidet 2008	AHTNP	AHTNP
U551	Todos ustedes tienen un compañero para trabajar en equipo	Cenidet 2008	AHDCG	AHDCG
U552	Les recuerdo que deben realizar un trabajo individual y trabajo en equipo.	Cenidet 2008	AHDFI	AHDFI
U553	Y debe quedar constancia de ese trabajo en los foros de trabajo grupal.	Cenidet 2008	AHDFI	AHDFI
U554	No se si hantenido problemas para comunicarse, pero no se ven sus discusiones en el for respectivo	Cenidet 2008	AHUTEI	AHPC
U555	Beta no tienen más preguntas?	Cenidet 2008	AHPC	AHUTEI
U556	Ok.	Cenidet 2008	AHPPAA	AHPPAA
U557	Gracias por su participación.	Cenidet 2008	AHMCDB	AHMCDB
U558	Guadalupe, por qué pusieron esa oración?	Cenidet 2008	AHFE	AHFE
U559	En otras palabras ¿como supieron que son equivalentes?	Cenidet 2008	AHFE	AHFE
U560	Ok Carlos, gracias por aclarar	Cenidet 2008	AHVE	AHVE
U561	Estan de acuerdo carlos y Guadalupe en que participe Norma, aunque no sea su turno?	Cenidet 2008	AHERM	AHERM
U562	Ok Norma, Cuál es tu aportación??	Cenidet 2008	AHPC	AHPC
U563	hola de nuevo	Cenidet 2008	AHDB	AHDB
U564	exacto	Cenidet 2008	AHMCDB	AHMCDB
U565	OK.	Cenidet 2008	AHPPAA	AHPPAA
U566	Carlos, perdon por no contestar antes.	Cenidet 2008	AHESS	OAH

U567	eso es lo que se tiene que hacer.	Cenidet 2008	AHPTME	AHPTME
U568	Tu, que opinas Jenifer?	Cenidet 2008	AHTNP	AHTNP
U569	Suena un poco raro, no te parece?	Cenidet 2008	AHFE	AHFE
U570	¿Qué opinas Jenifer?	Cenidet 2008	AHPC	AHPC
U571	dime	Cenidet 2008	AHUTEI	AHUTEI
U572	Deben entregar un solo trabajo, pero debe quedar evidencia de que estuvieron discutiendo las respuestas	Cenidet 2008	AHDFI	AHDFI
U573	Los invitamos a que discutan en los foros	Cenidet 2008	AHUTEI	AHUTEI
U574	exacto	Cenidet 2008	AHMCDB	AHMCDB
U575	estan de acuerdo en...	Cenidet 2008	AHPISDD	AHPISDD
U576	ok.	Cenidet 2008	AHPPAA	AHPPAA
U577	retomando la sugerencia de Jose Luis,les recomiendo que discutan la doble implicación en los foros, de acuerdo?	Cenidet 2008	AHFDNMC	AHFDNMC
U578	ok Toño te enviamos un correo electrónico, sale?	Cenidet 2008	OAH	OAH
U579	qué propone?	Cenidet 2008	AHLC	AHLC
U580	¿cómo?	Cenidet 2008	AHFE	AHFE
U581	Así es Iván.	Cenidet 2008	AHPPAA	AHPPAA
U582	¿estas de acuerdo Luisa?	Cenidet 2008	AHTM	AHTM
U583	Muy rápido y la discutimos en el foro.	Cenidet 2008	AHPISDD	AHPISDD
U584	Bien planteada.	Cenidet 2008	AHPPAA	AHPPAA
U585	Le seguimos en el foro.	Cenidet 2008	AHMCDB	AHMCDB
U586	Hola	Cenidet 2008	AHDB	AHDB
U587	por qué piensas eso?	Cenidet 2008	AHFE	AHFE
U588	si Segio a si es	Cenidet 2008	AHPPAA	AHPPAA
U589	Sergio: por que piensas que es un o y no un y	Cenidet 2008	AHFE	AHFE
U590	era dirigida a Sergio Norma.	Cenidet 2008	AHTD	AHTD
U591	Tu estas de acuerdo con Sergio?	Cenidet 2008	AHERM	AHERM
U592	Sergio: Te convencimos?	Cenidet 2008	AHMD	AHMD
U593	El preceso es : trabajo individual y luego en equipo, entregan el resultado del equipo.	Cenidet 2008	AHDFI	AHDFI

U594	Les recomiendo que como equipo traigan sus preguntas ya preparadas,ok?	Cenidet 2008	AHERM	AHERM
U595	por fin Henry	Cenidet 2008	AHUTEI	AHUTEI
U596	Por qué es falsa?	Cenidet 2008	AHFE	AHFE
U597	por qué?	Cenidet 2008	AHFE	AHFE
U598	no entiendo Adair	Cenidet 2008	AHFE	AHFE
U599	por qué?	Cenidet 2008	AHFE	AHFE
U600	No entiendo	Cenidet 2008	AHFE	AHFE
U601	exacto	Cenidet 2008	AHMCDB	AHMCDB
U602	por fin omar	Cenidet 2008	AHUTEI	AHUTEI
U603	Hola adan	Cenidet 2008	AHDB	AHDB
U604	ciao henry	Cenidet 2008	AHUEA	AHUEA
U605		Cenidet 2008	AHUEA	AHUEA
U606	Adan: obtines eso por que aplicas doble negación ciersto?	Cenidet 2008	AHSCTT	AHSCTT
U607	adan	Cenidet 2008	xxx	xxx
U608	No entiendo como aplica el condicional.	Cenidet 2008	AHFE	AHFE
U609	Ok, seguimos discutiendo en el foro general.	Cenidet 2008	AHMCDB	AHMCDB
U610	Quén falta??	Cenidet 2008	AHPC	AHPC
U611	OK, adelante	Cenidet 2008	AHPPAA	AHPPAA
U612	Hoy se arregla.	Cenidet 2008	OAH	OAH
U613	gracias por la observación.	Cenidet 2008	AHMCDB	AHMCDB
U614	Los que faltan de Preguntar??	Cenidet 2008	AHPC	AHPC
U615	Adelanta Adair	Cenidet 2008	AHUTEI	AHUTEI
U616	despues consulta y discute con tu compañero.	Cenidet 2008	AHUTEI	AHUTEI
U617	Si, el trabajo final echo en equipo.	Cenidet 2008	AHMCDB	AHMCDB
U618	Ok.	Cenidet 2008	AHPPAA	AHPPAA
U619	Jóves, ha sido un placer el conversar con Ustedes, Nos comunicamnos el próximo sábado a las 10:00 hrs.	Cenidet 2008	AHESS	AHESS
U620	Hasta luego.	Cenidet 2008	AHMCDB	AHMCDB
U621	Hola a todos	Cenidet 2008	AHDB	AHDB
U622	Tienen la palabra Lizeth, Carlos morales, Fernando y Luisa	Cenidet 2008	AHPC	AHPC

U623	Qué opinan los miembros de los dos equipos?	Cenidet 2008	AHPC	AHUTEI
U624	Buenos dias a todos	Cenidet 2008	AHDB	AHDB
U625	Recuerde el foro como medio para comentar con otros equipos sus dudas.	Cenidet 2008	AHPC	AHPC
U626	Buena pregunta	Cenidet 2008	AHPPAA	AHPPAA
U627	Así es Jenifer.	Cenidet 2008	AHPPAA	AHPPAA
U628	Les propongo que planteen en el foro cómo queda este ejercicio.	Cenidet 2008	AHSCTT	AHSCTT
U629	Perdón, ya vi cual es.	Cenidet 2008	OAH	OAH
U630	No se si Guadalupe tenga alguna pregunta	Cenidet 2008	AHTM	AHTM
U631	a ti, como a que te suena?	Cenidet 2008	AHFE	AHFE
U632	y a ti Guadalupe?	Cenidet 2008	AHTM	AHTM
U633	Ok Guadalupe, gracias	Cenidet 2008	AHPPAA	AHPPAA
U634	Ok, Guadalupe.	Cenidet 2008	AHPPAA	AHPPAA
U635	quién la hace?	Cenidet 2008	AHUTEI	AHUTEI
U636	Ivan?	Cenidet 2008	AHPC	AHPC
U637	Hola Norma.	Cenidet 2008	AHDB	AHDB
U638	Así es, es correcto.	Cenidet 2008	AHPPAA	AHPPAA
U639	Norma, esta bien la segunda respuesta, ya que usa los dos cuantificadores, de acuerdo al enunciado.	Cenidet 2008	OAH	OAH
U640	Ok.	Cenidet 2008	AHPPAA	AHPPAA
U641	Sergio, eso es correcto.	Cenidet 2008	AHPPAA	AHPPAA
U642	No entiendo esta parte: "z es menor que 0, y a mayor que 0,a=z"?	Cenidet 2008	AHSCTT	AHSCTT
U643	Por qué 1/ z???	Cenidet 2008	AHFE	AHFE
U644	Esto está bien.	Cenidet 2008	AHPPAA	AHPPAA
U645	Así es henry.	Cenidet 2008	AHPPAA	AHPPAA
U646	Hasta luego.	Cenidet 2008	AHMCDB	AHMCDB
U647	Hola a todos	Cenidet 2008	AHDB	AHDB
U648	Saludos para todos	Cenidet 2008	AHDB	AHDB
U649	Así, es Adair César.	Cenidet 2008	AHPPAA	AHPPAA
U650	Ahora cuál es el orden de los cuantificadores?	Cenidet 2008	AHFE	AHFE
U651	Qué opinas Carlos?	Cenidet 2008	AHTM	AHTM

U652	Y tú César?	Cenidet 2008	AHTM	AHTM
U653	Por problemas técnicos no puedo Utilizar este editor, podemos pasar a la dirección: http://132.254.39.242/moodle/ Los que deseen pueden seguir comentando sus dudas y propuestas de formalización en este chat.	Cenidet 2008	OAH	OAH
U654	Hola a todos	Cenidet 2008	AHDB	AHDB
U655	Hola chicos bienos dias	Cenidet 2008	AHDB	AHDB
U656	Elías, Adair, Cesar, Henry?????	Cenidet 2008	AHUTEI	AHUTEI
U657	"En este foro Hay alguien que nos ve a todos" vrs. "Todos ven a alguin en este foro" qué opinan??	Cenidet 2008	AHPC	AHPC
U658	Ok Henry, y ese enunciado sería verdadero o falso??	Cenidet 2008	AHFE	AHFE
U659	Elías, tú que deseas preguntar?	Cenidet 2008	AHPC	AHPC
U660	muy buena pregunta	Cenidet 2008	AHPPAA	AHPPAA
U661	a que te refieres?	Cenidet 2008	AHFE	AHFE
U662	Recuerden que tiene el for para seguir con las discusiones o plantear otras dudas.	Cenidet 2008	AHPC	AHUTEI
U663	lo que importa es que tu tengas claro qué interpretación estas usando.	Cenidet 2008	AHPTME	AHPTME
U664	perdon, no	Cenidet 2008	AHPPAA	AHPPAA
U665	Jenifer: entonces no contestaste todos igual	Cenidet 2008	AHPTME	AHPTME
U666	ok jeni	Cenidet 2008	AHPPAA	AHPPAA
U667	Alguna pregunta Carlos ??	Cenidet 2008	AHPC	AHUTEI
U668	Qué propones Fernando?	Cenidet 2008	AHPC	AHUTEI
U669	quien es Q y quie es T?	Cenidet 2008	AHFE	AHFE
U670	por que scogiste doble implicacion y no un aimplicacion?	Cenidet 2008	AHFE	AHFE
U671	lo comentamos en el foro Fer	Cenidet 2008	AHPC	AHUTEI
U672	hola delta beta y alfa	Cenidet 2008	AHDB	AHDB
U673	ok fer sigamos entonces.... que te hace pesar que un cuantif universal te lleva a una doble implicacion?	Cenidet 2008	AHFE	AHFE
U674	fer: te cambio el ejemplo	Cenidet 2008	OAH	OAH
U675	Pero no respondiste la pregunta de Azucena.	Cenidet 2008	AHPC	AHUTEI

U676	que propusiste fernado	Cenidet 2008	AHFE	AHFE
U677	si, me parece correctta	Cenidet 2008	AHPPAA	AHPPAA
U678	Ok Jenifer, así es.	Cenidet 2008	AHPPAA	AHPPAA
U679	gracias a ustedes por preguntar	Cenidet 2008	AHMCDB	AHMCDB
U680	No te preocupes Toño, te encontraremos compañero	Cenidet 2008	AHMCA	AHMCA
U681	si no hay mas dudas les deseo un buen fin de semana a todos haciendo ejercicios, claro!!	Cenidet 2008	AHUEA	AHUEA
U682	Nos vemos el próximo Martes.	Cenidet 2008	AHMCDB	AHUEA
U683	Que la pasen bien.	Cenidet 2008	AHUEA	AHUEA
U684	Hola a todos, buen día.	Cenidet 2008	AHDB	AHDB
U685	HOla Carlos: Dejame revisar tu grupo y base de datos.	Cenidet 2008	AHUEA	AHDB
U686	Hola carlos, no se que estará pasando, espero que hoy se resuelva.	Cenidet 2008	AHUEA	AHUEA
U687	Espero las preguntas de los equipos Alfa y Beta.	Cenidet 2008	AHURH	AHURH
U688	Lizeth o Luisa , tienen alguna pregunta?	Cenidet 2008	AHPTME	AHPTME
U689	En la pregunta en la respuesta que pusimos del 5 c) ¿les parece correcto?, espero su comentario en el foro.	Cenidet 2008	AHFE	AHFE
U690	y puede participar Lizeth ya que no pudo particpar en su momento.	Cenidet 2008	AHTNP	AHTNP
U691	Lo que dices es correcto Lizeth.	Cenidet 2008	AHPPAA	AHPPAA
U692	Pero de forma más general se interpreta com multiplo de 2.	Cenidet 2008	OAH	OAH
U693	es decir, cómo decimos que dos conjuntos se interctamn en LN	Cenidet 2008	AHPTME	AHPTME
U694	¿Pueden indicarme algunas espresiones?	Cenidet 2008	AHFE	AHFE
U695	Alguna sugerencia?	Cenidet 2008	AHPC	AHUTEI
U696	En lenguaje más coloquial, por favor...	Cenidet 2008	AHFE	AHFE
U697	Ok, me parece bien Lizeth.	Cenidet 2008	AHPPAA	AHPPAA
U698	Ok Gracias, espero ver muchas de estas expresiones en su trabajo.	Cenidet 2008	AHMCDB	AHMCDB
U699	Si es correcto Carlos.	Cenidet 2008	AHPPAA	AHPPAA
U700	No se a qué te refieras Carlos.	Cenidet 2008	AHFE	AHFE
U701	Ok, eso tiene más sentido.	Cenidet 2008	AHPPAA	AHPPAA
U702	Ok Ahora y le pregunto a Juan Antonio:	Cenidet 2008	AHTM	AHTM
U703	Hagan el dibujo y vean.	Cenidet 2008	AHPTME	AHPTME

U704	Después nos comentan.	Cenidet 2008	AHPC	AHPC
U705	Asi es Sergio.	Cenidet 2008	AHPPAA	AHPPAA
U706	Ok.	Cenidet 2008	AHPPAA	AHPPAA
U707	Adair, Henry, César y Elías es su turno.	Cenidet 2008	AHPC	AHUTEI
U708	Y cuales son los predicados??	Cenidet 2008	AHFE	AHFE
U709	Y cuál es el universo?	Cenidet 2008	AHFE	AHFE
U710	Si, eso está bien.	Cenidet 2008	AHPPAA	AHPPAA
U711	En la expresión Vx,y ¬P(x) Vy RC(y)| ¬T(x,y) la y está cuantificada universalmente dos veces.	Cenidet 2008	OAH	OAH
U712	Por otro lado supongo que RC significa "y" es raiz caudrada y si es así, el predicado T(x,y) qué significa?	Cenidet 2008	AHFE	AHFE
U713	En la expresión Vx,y ¬P(x) Vy RC(y)| ¬T(x,y) la y está cuantificada universalmente dos veces.	Cenidet 2008	OAH	OAH
U714	Por otro lado supongo que RC significa "y" es raiz caudrada y si es así, el predicado T(x,y) qué significa?	Cenidet 2008	AHFE	AHFE
U715	Sus respuestas son correctas, ustedes consideraron los elemntos de A de forma directa.	Cenidet 2008	AHPPAA	AHPPAA
U716	Esa respuesta es correcta.	Cenidet 2008	AHPPAA	AHPPAA
U717	La primera puede considerarse como correcta, si se aclara el uso de la "r".	Cenidet 2008	OAH	OAH
U718	la segunda no.	Cenidet 2008	OAH	OAH
U719	Ok.	Cenidet 2008	AHPPAA	AHPPAA
U720	Si tienen más preguntas, por favor, envielas al foro.	Cenidet 2008	AHPC	AHUTEI
U721	Yo revisaré en la tarde y reponderé.	Cenidet 2008	AHMCDB	AHMCDB
U722	HASTA LUEGO.	Cenidet 2008	AHMCDB	AHMCDB
U723	Hola a todos, les recuerdo que el orden de asesoría será el siguiente: Primero equipos Tau y Sigmma; después equipos Psi y Lambda etc..	Cenidet 2008	AHPC	AHPC
U724	¿Qué quieres decir con " más entendible"?	Cenidet 2008	AHFE	AHFE
U725	Esa es una buena observación, Elías.	Cenidet 2008	AHPPAA	AHPPAA
U726	OK se los dejo para discutir en el foro.	Cenidet 2008	AHPC	AHUTEI
U727	Dejame ver Norma..	Cenidet 2008	OAH	OAH

U728	Eso es correcto, Norma.	Cenidet 2008	AHPPAA	AHPPAA
U729	Pero observa que la pregunta de Elías es muy sutil y merece una bunea reflexión para responderla.	Cenidet 2008	AHSCTT	AHSCTT
U730	Sergio, no se que estas hablando.	Cenidet 2008	AHFE	AHFE
U731	Puedes aclararlo, por favor?	Cenidet 2008	AHFE	AHFE
U732	Gracias Norma, ya entendí.	Cenidet 2008	AHPPAA	AHPPAA
U733	Los ejercicos del "link" no se debían hacer, pero que bueno que los intentaron resolver.	Cenidet 2008	AHVE	AHVE
U734	Es cierto Sergio.	Cenidet 2008	AHPPAA	AHPPAA
U735	hace afalta una definición de I(O).	Cenidet 2008	OAH	OAH
U736	Pero la interpretación que propones es adecuada.	Cenidet 2008	AHPPAA	AHPPAA
U737	Sergio, propón el ejemplo y proponlo en el foro.	Cenidet 2008	AHPISDD	AHPISDD
U738	Ivan, No entiendo.	Cenidet 2008	AHFE	AHFE
U739	Dejame ver Jenifer.	Cenidet 2008	AHMCDB	AHMCDB
U740	Marino, Alguna otra pregunta	Cenidet 2008	AHACI	AHACI
U741	Manuel, Qué haces ?	Cenidet 2008	AHTB	AHTB
U742	No si tengan alguna otra pregunta.	Cenidet 2008	AHPC	AHUTEI
U743	¡¡¡??	Cenidet 2008	OAH	OAH
U744	Qué opinas María José?	Cenidet 2008	AHPC	AHPC
U745	Norma nos ha proporsionado la siguiente liga: http://www.ia.urjc.es/grupo/docencia/automatas_itis/hojas/hoja2.pdf	Cenidet 2008	AHPC	AHPC
U746	Lizeth, tienes alguna pregunta??	Cenidet 2008	AHPC	AHUTEI
U747	POR OTRO lado, recuerden la dinámica de participación, debe quedar constancia de las discusiones, dudas, inquietudes de cada equipo.	Cenidet 2008	AHPC	AHPC
U748	LOS FELICITO por su participación activa.	Cenidet 2008	AHVE	AHVE
U749	En general van bien, lo que deben preguntarse Ustedes es que tanto están mejorando en la comprensión del lenguaje matemático.	Cenidet 2008	AHPPAA	AHPPAA
U750	Los invito a leer nuevamente los propósitos del curso.	Cenidet 2008	AHSCTT	AHSCTT
U751	Nos vemos el sábado.	Cenidet 2008	AHUEA	AHUEA
U752	Hasta luego.	Cenidet 2008	AHMCDB	AHMCDB

U753	Hola a todos, buenos días.	Cenidet 2008	AHDB	AHDB
U754	Hola Buenos días	Cenidet 2008	AHDB	AHDB
U755	Algún comentario???	Cenidet 2008	AHURH	AHURH
U756	Qué dudas o comentarios tienen?	Cenidet 2008	AHFE	AHFE
U757	Cuál es tu confisuón??	Cenidet 2008	AHFE	AHFE
U758	Lizeth, Cuál la confusión?	Cenidet 2008	AHFE	AHFE
U759	Así es Carlos.	Cenidet 2008	AHPPAA	AHPPAA
U760	Ok Marino, ¿cual es tu respuesta?	Cenidet 2008	AHSCTT	AHSCTT
U761	La pregunta también es para los equipos Delta y Epsilon: la relación {(1,1), (2,2), (3,3)} será transitiva?	Cenidet 2008	AHFE	AHFE
U762	No se entiende esta parte: $\forall$ a, b, c $\in$ C;(aRb) $\wedge$ (bRc) $\Rightarrow$ (aRc)	Cenidet 2008	AHFE	AHFE
U763	Por uqe dices que se convierte en reflexiva?	Cenidet 2008	AHFE	AHFE
U764	¿cóm aplicaste al defición de transitividad?	Cenidet 2008	AHFE	AHFE
U765	Bien Jeny!!	Cenidet 2008	AHVE	AHVE
U766	Bien Miguel!	Cenidet 2008	AHVE	AHVE
U767	Ok ahora digan si esta relación es transitiva: {(1,2), (3,4), (5,6)} ???	Cenidet 2008	AHFE	AHFE
U768	Ok Jenifer, eso es correcto, y Cuál es el valor de verdad de la implicación?	Cenidet 2008	AHSCTT	AHSCTT
U769	seguimnos con tu duda Martín	Cenidet 2008	OAH	OAH
U770	Tienen razón, Guadalupe y Juan Antonio.	Cenidet 2008	AHPPAA	AHPPAA
U771	Así es Guadalupe.	Cenidet 2008	AHPPAA	AHPPAA
U772	Carlos, observa tu implicación	Cenidet 2008	OAH	OAH
U773	Se los dejo para discutirlo en el foro.	Cenidet 2008	AHPISDD	AHPISDD
U774	Ok, comenten	Cenidet 2008	AHPC	AHUTEI
U775	Si, por favor.	Cenidet 2008	AHMCDB	AHMCDB
U776	Así es Sergio.	Cenidet 2008	AHPPAA	AHPPAA
U777	Cuales son tus definiciones.	Cenidet 2008	AHFE	AHFE
U778	qué dice el enunciado?	Cenidet 2008	AHSCTT	AHSCTT
U779	El equipo Tau puede participar.	Cenidet 2008	AHPC	AHUTEI
U780	Muy bien, se nos acabo el tiempo.	Cenidet 2008	AHMCDB	AHMCDB
U781	Pueden plantear sus preguntas en el foro.	Cenidet 2008	AHPC	AHUTEI

U782	Haremos una sesión especial con la herramienta de "CONSTRUCCIÓN CONJUNTA" EL DÍA DE HOY POR LA TARDE de 19:00 a 21:00, quienes deseen participar están invitados.	Cenidet 2008	AHMCDB	AHMCDB
U783	Les envío la invitación en unos minutos.	Cenidet 2008	AHMCDB	AHMCDB
U784	Saludos para todos.	Cenidet 2008	AHMCDB	AHMCDB
U785	Azucena y José Luis.	Cenidet 2008	xxx	xxx
U786	Hola a todos, iniciamos esta asesoría sobre la cuarta unidad del curso: FUNCIONES.	Cenidet 2008	AHDB	AHDB
U787	Asi que tienen la palabra Lizeth, Carlos Morales, Luisa y Fernando	Cenidet 2008	AHPC	AHUTEI
U788	Carlos Morales, es tu turno de preguntar.	Cenidet 2008	AHACI	AHACI
U789	Así queJenifer, Miguel Angel, Marino y Alejandro pueden plantear sus pregunats.	Cenidet 2008	AHACI	AHACI
U790	Así es Jenifer.	Cenidet 2008	AHPPAA	AHPPAA
U791	Marino, Alguna pregunta?	Cenidet 2008	AHTM	AHTM
U792	Ok Marino, Miguel Angel y Jenifer?	Cenidet 2008	AHPC	AHPC
U793	Les recuerdo que pueden plantear sus dudas en el FORO.	Cenidet 2008	AHPC	AHUTEI
U794	Así es Juan Antonio.	Cenidet 2008	AHPPAA	AHPPAA
U795	eso quedaría mejor.	Cenidet 2008	AHPPAA	AHPPAA
U796	Carlos Martín, no entiendo que quieres decir con "necesariamente el conjunto de B tiene que tener a la par uno de A".	Cenidet 2008	AHFE	AHFE
U797	Te propongo que lo comentemos en el foro.	Cenidet 2008	AHPC	AHUTEI
U798	Perdón, Norma no me quedo claro el ejemplo.	Cenidet 2008	AHFE	AHFE
U799	Cuál es el domino y cuáles la imgén de la función que propones?,	Cenidet 2008	AHFE	AHFE
U800	Y cuál es el dominio y el conjunto imgén de tu función?	Cenidet 2008	AHFE	AHFE
U801	Norma.	Cenidet 2008	xxx	xxx
U802	No me queda clara la función.	Cenidet 2008	AHFE	AHFE
U803	¿Cu´l esel dominio?	Cenidet 2008	AHFE	AHFE
U804	¿Cuál es el conjunto imgen?	Cenidet 2008	AHFE	AHFE
U805	Tienes razón Hugo, eso daría la idea de una composición de funciones.	Cenidet 2008	OAH	OAH
U806	Buena pregunta Hugo, te propongo que la plantees en el foro, ahi lo	Cenidet 2008	AHPC	AHPC

	comentamos.			
U807	Algo más que quieran comentar??	Cenidet 2008	AHPC	AHPC
U808	Así es Henry.	Cenidet 2008	AHPPAA	AHPPAA
U809	Ok Sergio, Esa sería una interpretación interesantecuando no se puede llevara a cabo la composición de funciones.	Cenidet 2008	OAH	OAH
U810	Saludos para todos.	Cenidet 2008	AHMCDB	AHMCDB
U811	Hasta Luego.	Cenidet 2008	AHMCDB	AHMCDB
U812	Hola a todos	Cenidet 2008	AHDB	AHDB
U813	Hola Mari José, ponme en contexto: cual es el ejercicio?	Cenidet 2008	AHURH	AHURH
814	(me disculpo por qué no puedo acceder a los documentos)	Cenidet 2008	AHMCDB	AHMCDB
U815	Hola Jose Luis	Cenidet 2008	AHDB	AHDB
U816	Hola A todos, Les ofrezco una disculpa por el retraso-	Cenidet 2008	AHDB	AHDB
U817	Hola Azucena, ¿cuáles han sido las preguntas?	Cenidet 2008	AHMCDB	AHURH
U818	OK.	Cenidet 2008	AHPPAA	AHPPAA
U819	Gracias, Azucena.	Cenidet 2008	AHPPAA	AHPPAA
U820	🙄	Cenidet 2008	AHUEA	AHUEA
U821	Así es María José, en realidad son dos teoremas de la forma P --> Q	Cenidet 2008	OAH	OAH
U822	que opinan los demás?	Cenidet 2008	AHPC	AHUTEI
U823	Para Norma: La idea está bien expresada, pero observa que mezclaste mucho el LN y el Lenguaje matemático.	Cenidet 2008	OAH	OAH
U824	Adair, ¿cómo justificarías tu afirmación?	Cenidet 2008	AHSCTT	AHSCTT
U825	cuál ley?	Cenidet 2008	AHFE	AHFE
U826	María José, por favor escribe tu pregunta en el foro y la responderemso hoy.	Cenidet 2008	AHPC	AHUTEI
U827	Disculpa el retraso en responderte.	Cenidet 2008	AHMCDB	AHMCDB
U828	Ante la duda, propongo que lo discutan en el foro!	Cenidet 2008	AHPC	AHPC
U829	Norma, Puedes mejorarun poco.	Cenidet 2008	AHEOD	AHEOD
U830	Te propongo que lo coemntemso en el foro.	Cenidet 2008	AHPC	AHUTEI
U831	o sea que estaban de colados?!!	Cenidet 2008	AHFDNMC	AHFDNMC

U832	Ok una conclusion: Adair y María José Tienen razón.	Cenidet 2008	AHPPAA	AHPPAA
U833	Entre todos han hecho El analisis de la estructura lógica del teorema.	Cenidet 2008	AHVE	AHVE
U834	Ok, no se que pregunta tengan Hugo, Adan o Henry	Cenidet 2008	AHACI	AHACI
U835	perdon, entendi mal, adelante!	Cenidet 2008	AHMCDB	AHMCDB
U836	3. - Si son funciones totales, no puede resultar que la composición sea una función parcial, cómo lo justificarias?	Cenidet 2008	OAH	OAH
U837	Ivan, por favor, plantea tu pregunta.	Cenidet 2008	AHTM	AHTM
U838	Norma, esta me paerece más adecuada V x,y,z | ((x,y) ∈ f ^ (y,z) ∈ g) ⇒ (x,z) ∈ (g °f) pero falta aclarar de donde son x, y y z.	Cenidet 2008	AHFE	AHFE
U839	Ok, Norma eso me aclara más la propuesta y me parece adecuada	Cenidet 2008	AHPPAA	AHPPAA
U840	Hola, creo que es tiempo de pasar a los siguientes	Cenidet 2008	AHPC	AHUTEI
U841	gracias.	Cenidet 2008	AHMCDB	AHMCDB
U842	Equipos siguinetes	Cenidet 2008	AHPC	AHUTEI
U843	Gracias	Cenidet 2008	AHMCDB	AHMCDB
U844	No se qué opine Guadalupe.	Cenidet 2008	OAH	OAH
U845	Así es Guadalupe.	Cenidet 2008	AHPPAA	AHPPAA
U846	No entiendo que quieres decir con: "esta relacionada con ella misma"	Cenidet 2008	AHFE	AHFE
U847	Ok.	Cenidet 2008	AHPPAA	AHPPAA
U848	Lizeth, espero que te haya sido de utilidad esta sesión.	Cenidet 2008	AHMCDB	AHMCDB
U849	Hasta luego.	Cenidet 2008	AHMCDB	AHMCDB
U850	Hola a todos, buenos días.	Cenidet 2008	AHDB	AHDB
U851	Hola, buenos dias a todos	Cenidet 2008	AHDB	AHDB
U852	¿Qué haremos el día de hoy?	Cenidet 2008	AHPC	AHUTEI
U853	Esperamos que nos comenten su autoevalaución del curso.	Cenidet 2008	AHURH	AHURH
U854	Considerando los aspectros que les señalamos el día de ayer y algunas preguntas específicas que deseamos plantearles.	Cenidet 2008	AHURH	AHURH
U855	Además daremos las indicaciones para el inico del curso presencial.	Cenidet 2008	AHURH	AHURH
U856	Iniciaremos con la pregunta:	Cenidet 2008	AHPC	AHUTEI
U857	¿Consideran que se cumplieron los propósitos del curos?	Cenidet 2008	AHPC	AHUTEI
U858	¿en qué medida?	Cenidet 2008	AHPC	AHUTEI

U859	Iniciaremos con la pregunta:	Cenidet 2008	AHPC	AHUTEI
U860	¿Consideran que se cumplieron los propósitos del curso?	Cenidet 2008	AHPC	AHUTEI
U861	¿en qué medida?	Cenidet 2008	AHPC	AHUTEI
U862	Ok, gracias Miguel Angel.	Cenidet 2008	AHMCDB	AHMCDB
U863	Bien Miguel	Cenidet 2008	AHPPAA	AHPPAA
U864	Les recordamos que todos pueden participar, se vale hacer los comentarios que consideren pertinentes.	Cenidet 2008	AHPC	AHPC
U865	Sergio?	Cenidet 2008	AHTM	AHTM
U866	Ok, alguien más?	Cenidet 2008	AHPC	AHUTEI
U867	Me gustaría saber su opinión sobre los materiales:	Cenidet 2008	AHPTME	AHPTME
U868	a. “traducción del LN al lenguaje de la Lógica proposicional.	Cenidet 2008	AHPTME	AHPTME
U869	b. “traducción del LN al lenguaje de la Lógica de predicados.	Cenidet 2008	AHPTME	AHPTME
U870	¿qué les parerecieron?, ¿creen que les fueron de utilidad?	Cenidet 2008	AHFE	AHFE
U871	Ok, alguien más?	Cenidet 2008	AHPC	AHPC
U872	¿Qué les pareció el tiempo para leer los materiales?	Cenidet 2008	AHFE	AHFE
U873	¿y el tiempo en la asesoria general?	Cenidet 2008	AHFE	AHFE
U874	Algún otro comentario??	Cenidet 2008	AHPC	AHPC
U875	Ok, gracias por sus observaciones.	Cenidet 2008	AHMCDB	AHMCDB
U876	Trabajaremos para mejorar esos aspectos.	Cenidet 2008	AHMCDB	AHMCDB
U877	los invito a reflexionar un poco y que me comenten: ¿cómo consideran que se podrían mejorar los materiales:	Cenidet 2008	AHFE	AHFE
U878	a. Traducción del LN al lenguaje matemático y viceversa.	Cenidet 2008	AHFE	AHFE
U879	b. Análisis de las definiciones.	Cenidet 2008	AHFE	AHFE
U880	Ok, tomaremos en cuenta sus observaciones para la nueva edición y algunos aspectos del curso presencial.	Cenidet 2008	AHMCDB	AHMCDB
U881	Como han visto, no tenemos un material específico para el análisis de las definiciones formalizadas de Ciencias de la computación.	Cenidet 2008	AHFE	AHFE
U882	¿qué debería contener ese material?, ¿cuál debería ser su estructura?	Cenidet 2008	AHFE	AHFE
U883	Dedicaremos los próximos 30 minutos a otros comentarios y preguntas	Cenidet 2008	AHPTME	AHPTME

	sobre el curso que desen plantear.			
U884	Tienes razón Omar: Lunes 22 8:00 Bien, por mi parte me despido no sin antes desearles un buen inicio en el CENIDET.	Cenidet 2008	AHMCDB	AHMCDB
U885	Hasta pronto!!!	Cenidet 2008	AHMCDB	AHUEA
U886	Saludos a todos y hasta pronto	Cenidet 2008	AHUEA	AHUEA
U887	Ok, Hugo, comentaré con la parte técnica y te avisamos mañana.	Cenidet 2008	OAH	OAH
U888	Ok, aclarada la duda, no me queda más que FELICITARLOS por su participación activa durante el curso y sus los logros individuales.	Cenidet 2008	AHACDB	AHACDB
U889	Nos vemos el día 25 de agosto.	Cenidet 2008	AHMCDB	AHMCDB
U890	RECIBAN UN CORDIAL SALUDO.	Cenidet 2008	AHMCDB	AHMCDB
U891	Reciban un cordial saludo.	Cenidet 2008	AHMCDB	AHMCDB

En la tabla anterior los caracteres “xxx” en la columna “Atributo de habilidad identificado Analista 1” y en la columna “Atributo de habilidad identificado Analista 2” señalan las unidades de grabación en las que los codificadores no asignaron un atributo de habilidad social.

CAPÍTULO V

PUNTUACIÓN

La puntuación de los textos escritos, con la que se pretende reproducir la entonación de la lengua oral, constituye un capítulo importante dentro de la ortografía de cualquier idioma. De ella depende en gran parte la correcta expresión y comprensión de los mensajes escritos. La puntuación organiza el discurso y sus diferentes elementos y permite evitar la ambigüedad en textos que, sin su empleo, podrían tener interpretaciones diferentes.

El español cuenta con los siguientes signos de puntuación:

punto .	signos de exclamación ¡!
coma ,	paréntesis ( )
punto y coma ;	corchetes [ ]
dos puntos :	raya —
puntos suspensivos ...	comillas «»; " "; ' '[32]
signos de interrogación ¿?	

El punto, la coma, el punto y coma, los dos puntos y los puntos suspensivos se escriben siempre sin dejar un espacio de separación con respecto a la palabra o el signo que precede, y separados por un espacio de la palabra o el signo que sigue, a no ser que este sea de cierre.

Los signos dobles, como los de interrogación y exclamación, los paréntesis, los corchetes, las comillas y las rayas que encierran aclaraciones e incisos, todos ellos compuestos por un signo de apertura y uno de cierre, se escriben de la manera siguiente:

a) Los de apertura se separan por medio de un espacio de la palabra o signo al que siguen, y se escriben sin espacio de separación con respecto a la palabra a la que anteceden[33].

b) Por el contrario, los signos de cierre se escriben sin espacio de separación con respecto a la palabra o signo al que siguen, separados por un espacio de la palabra a la que preceden y sin este espacio si lo que sigue es un signo de puntuación.

Ejemplos:

¿Se trató el asunto en la reunión?; ¿concretaron algo?

Ya conoces su refrán preferido: «Agua que no has de beber...».

La boda se celebrará en la parroquia de la Asunción (plaza de la Prosperidad, n.º 3), a la una de la tarde.

Cuando decidimos trasladarnos a Buenos Aires —una decisión muy meditada—, el negocio iba viento en popa.

---

[32] Existen otros signos ortográficos, que se verán en el apartado 5.11.

[33] Véase, sin embargo, *5.7.3* y *5.7.4*. La raya que introduce la intervención de un personaje en un diálogo se escribe también sin espacio de separación de la palabra que sigue.

5.1. Uso del punto

El punto (.) señala la pausa que se da al final de un enunciado. Después de punto —salvo en el caso del utilizado en las abreviaturas— siempre se escribe mayúscula[34].

Hay tres clases de punto: el punto y seguido, el punto y aparte y el punto final.

a) El punto y seguido separa enunciados que integran un párrafo. Después de un punto y seguido se continúa escribiendo en la misma línea. Si el punto está al fin de renglón, se empieza en el siguiente sin dejar margen. Por ejemplo:

Salieron a dar un breve paseo. La mañana era espléndida.

b) El punto y aparte separa dos párrafos distintos, que suelen desarrollar, dentro de la unidad del texto, contenidos diferentes. Después de punto y aparte se escribe en una línea distinta. La primera línea del nuevo párrafo debe tener un margen mayor que el resto de las líneas que lo componen, es decir, ha de quedar sangrada. Por ejemplo:

El mar estaba embravecido aquel día. Los barcos bailaban sobre el agua sorteando las olas con dificultad.

Miguel, sentado en el muelle, esperaba el regreso de su padre. Atisbaba el horizonte buscando ansioso su barco con la mirada.

c) El punto final es el que cierra un texto.

5.1.1. Otro uso del punto

El punto se utiliza también después de las abreviaturas. Ejemplos: *Sra., Excmo., cf.*[35].

5.1.2. Combinación del punto con otros signos

A menudo es necesario combinar el punto con otros signos que también cierran períodos, como son los paréntesis o las comillas. En estos casos, se coloca el punto siempre detrás de las comillas, corchetes o paréntesis de cierre. Por ejemplo:

Sus palabras fueron estas: «No quiero volver a verte». Después cerró de golpe la puerta de su casa. (Creo que estaba muy enojada).

Sobre la colocación del punto en comentarios y aclaraciones señaladas con raya, véase *5.9.3.*

5.1.3. Uso incorrecto del punto

Los títulos y los subtítulos de libros, artículos, capítulos, obras de arte, etc., cuando aparecen aislados, no llevan punto final. Ejemplos:

El llano en llamas
La Venus del espejo

5.2. Uso de la coma

La coma (,) indica una pausa breve que se produce dentro del enunciado.

---

[34] Véase *3.3.1*a.
[35] Véase el apartado *6.1.2* para los casos de abreviaturas y símbolos que no llevan punto.

5.2.1. Se emplea para separar los miembros de una enumeración, salvo los que vengan precedidos por alguna de las conjunciones *y, e, o, u*. Ejemplos:

Es un chico muy reservado, estudioso y de buena familia.

Acudió toda la familia: abuelos, padres, hijos, cuñados, etc.

¿Quieres café, té o un refresco?

Cuando los elementos de la enumeración constituyen el sujeto de la oración o un complemento verbal y van antepuestos al verbo, no se pone coma detrás del último. Ejemplos:

El perro, el gato y el ratón son animales mamíferos.

De gatos, de ratones y de perros no quiere ni oír hablar.

5.2.2. Se usa coma para separar miembros gramaticalmente equivalentes dentro de un mismo enunciado, a excepción de los casos en los que medie alguna de las conjunciones *y, e, ni, o, u*. Ejemplos:

Estaba preocupado por su familia, por su trabajo, por su salud.

Antes de irte, corre las cortinas, cierra las ventanas, apaga las luces y echa la llave.

Sin embargo, se coloca una coma delante de la conjunción cuando la secuencia que encabeza expresa un contenido (consecutivo, de tiempo, etc.) distinto al elemento o elementos anteriores. Por ejemplo:

Pintaron las paredes de la habitación, cambiaron la disposición de los muebles, y quedaron encantados.

También cuando esa conjunción está destinada a enlazar con toda la proposición anterior, y no con el último de sus miembros. Por ejemplo:

Pagó el traje, el bolso y los zapatos, y salió de la tienda.

Siempre será recomendable su empleo, por último, cuando el periodo sea especialmente largo. Por ejemplo:

Los instrumentos de precisión comenzaron a perder su exactitud a causa de la tormenta, y resultaron inútiles al poco tiempo.

5.2.3. En una relación cuyos elementos están separados por punto y coma, el último elemento, ante el que aparece la conjunción copulativa, va precedido de coma o punto y coma. Ejemplos:

En el armario colocó la vajilla; en el cajón, los cubiertos; en los estantes, los vasos, y los alimentos, en la despensa.

Con gran aplomo, le dijo a su familia que llegaría a las tres; a sus amigos, que lo esperasen a las cinco; y consiguió ser puntual en los dos casos.

5.2.4. Se escribe una coma para aislar el vocativo del resto de la oración. Ejemplos:

Julio, ven acá.

He dicho que me escuchéis, muchachos.

Cuando el vocativo va en medio del enunciado, se escribe entre dos comas. Por ejemplo:

Estoy alegre, Isabel, por el regalo.

5.2.5. Los incisos que interrumpen una oración, ya sea para aclarar o ampliar lo dicho, ya sea para mencionar al autor u obra citados, se escriben entre comas[16]. Son incisos casos como los siguientes:

a) Aposiciones explicativas. Por ejemplo:

En ese momento Adrián, el marido de mi hermana, dijo que nos ayudaría.

b) Las proposiciones adjetivas explicativas. Por ejemplo:

Los vientos del Sur, que en aquellas abrasadas regiones son muy frecuentes, incomodan a los viajeros.

c) Cualquier comentario, explicación o precisión a algo dicho. Ejemplos:

Toda mi familia, incluido mi hermano, estaba de acuerdo.

Ella es, entre mis amigas, la más querida.

Nos proporcionó, después de tantos disgustos, una gran alegría.

d) La mención de un autor u obra citados. Por ejemplo:

La verdad, escribe un político, se ha de sustentar con razones y autoridades.

5.2.6. Cuando se invierte el orden regular de las partes de un enunciado, anteponiendo elementos que suelen ir pospuestos, se tiende a colocar una coma después del bloque anticipado. No es fácil establecer con exactitud los casos en que esta anteposición exige el uso de la coma. Pero frecuentemente puede aplicarse esta norma práctica:

a) Si el elemento antepuesto admite una paráfrasis con «en cuanto a», es preferible usar coma. Por ejemplo:

Dinero, ya no le queda. (Es posible decir *En cuanto al dinero, ya no le queda*).

b) Si, por el contrario, admite una paráfrasis con «es lo que» o «es el que», no se empleará coma. Por ejemplo:

Vergüenza debería darte. (Equivalente a *Vergüenza es lo que debería darte*).

5.2.7. También suele anteponerse una coma a una conjunción o locución conjuntiva que une las proposiciones de una oración compuesta, en los casos siguientes:

a) En las proposiciones coordinadas adversativas introducidas por conjunciones como *pero, mas, aunque, sino*. Ejemplos:

Puedes llevarte mi cámara de fotos, pero ten mucho cuidado.

Cogieron muchas cerezas, aunque todas picadas por los pájaros.

b) Delante de las proposiciones consecutivas introducidas por *conque, así que, de manera que...* Ejemplos:

Prometiste acompañarle, conque ya puedes ir poniéndote el abrigo.

[16] Véanse también 5.7.1, 5.7.2, 5.8.1 y 5.9.1.

El sol me está dando en la cara, así que tendré que cambiarme de asiento.

c) Delante de proposiciones causales lógicas y explicativas. Ejemplos:

Es noble, porque tiene un palacio.

Están en casa, pues tienen la luz encendida.

5.2.8. Los enlaces como *esto es, es decir, o sea, en fin, por último, por consiguiente, sin embargo, no obstante, además, en tal caso, por lo tanto, en cambio, en primer lugar*, y también, a veces, determinados adverbios o locuciones que desempeñan la función de modificadores oracionales, como *generalmente, posiblemente, efectivamente, finalmente, en definitiva, por regla general, quizá*, colocados al principio de una oración, se separan del resto mediante una coma. Ejemplos:

Por consiguiente, no vamos a tomar ninguna resolución precipitada.

No obstante, es necesario reformar el estatuto.

Efectivamente, tienes razón.

Cuando estas expresiones van en medio de la oración, se escriben entre comas. Ejemplos:

Estas dos palabras son sinónimas, es decir, significan lo mismo.

Tales incidentes, sin embargo, no se repitieron.

Este tipo de accidentes están causados, generalmente, por errores humanos.

Si los bloques relacionados mediante estos enlaces forman parte de la misma oración compuesta escrita entre puntos, se suelen separar con punto y coma colocado delante del enlace, al que seguirá una coma. Véase el apartado *5.4.3.*

5.2.9. En los casos en que se omite un verbo, porque ha sido anteriormente mencionado o porque se sobrentiende, se escribe en su lugar una coma. Ejemplos:

El árbol perdió sus hojas; el viejo, su sonrisa.

Los niños, por aquella puerta.

En matemáticas, un genio; para la música, bastante mediocre.

5.2.10. En las cabeceras de las cartas, se escribe coma entre el lugar y la fecha. Por ejemplo:

Santiago, 8 de enero de 1999

5.2.11. Se escribe coma para separar los términos invertidos del nombre completo de una persona o los de un sintagma que integran una lista (bibliografía, índice...). Ejemplos:

BELLO, Andrés: *Gramática de la lengua castellana destinada al uso de los americanos.*

CUERVO, Rufino José: *Diccionario de construcción y régimen de la lengua castellana.*

— *construcción, materiales de*
— *papelería, artículos de*

5.2.12. Uso incorrecto de la coma

Debe evitarse separar el sujeto y el predicado mediante coma. Ejemplos de incorrección:

**Las estanterías del rincón, estaban perfectamente organizadas.*

**Un desgraciado incidente, ocasionó la dimisión de la junta directiva.*

Se exceptúan, como ya hemos visto (*5.2.5*), los casos en que media un inciso entre sujeto y predicado. Ejemplos:

La medicina preventiva, como ya ha quedado apuntado anteriormente, permitirá evitar la enfermedad en breve plazo.

La presencia de la protagonista de la película, que vestía un espectacular traje de noche, produjo muy diferentes comentarios.

5.3. Uso de los dos puntos

Los dos puntos (:) detienen el discurso para llamar la atención sobre lo que sigue.

Se usan los dos puntos en los casos siguientes:

5.3.1. Después de anunciar una enumeración. Ejemplos:

Van a subastar tres manuscritos: uno de Borges, otro de Alfonso Reyes y un tercero de Antonio Machado.

Tres son las provincias aragonesas: Huesca, Zaragoza y Teruel.

También para cerrar una enumeración, antes del anafórico que los sustituye, se utilizan los dos puntos. Ejemplos:

Natural, sana y equilibrada: así debe ser una buena alimentación.

Terremotos, inundaciones y erupciones volcánicas: esas son las principales catástrofes naturales.

5.3.2. Los dos puntos preceden a las citas textuales. En este caso, después de los dos puntos se suele escribir la primera palabra con inicial mayúscula. Ejemplos:

Las palabras del médico fueron: «Reposo y una alimentación equilibrada».

Ya lo dijo Descartes: «Pienso, luego existo».

5.3.3. Se emplea este signo de puntuación tras las fórmulas de saludo en las cartas y documentos. También en este caso la palabra que sigue a los dos puntos se escribe con mayúscula y, generalmente, en un renglón aparte. Ejemplos:

Querido amigo:
Te escribo esta carta para comunicarte...

Muy señor mío:
Le agradeceré se sirva tomar a su cargo...

5.3.4. Se emplean los dos puntos para conectar oraciones o proposiciones relacionadas entre sí sin necesidad de utilizar otro nexo. Son varias las relaciones que se pueden expresar:

a) Relación causa-efecto. Ejemplos:

Se ha quedado sin trabajo: no podrá ir de vacaciones este verano.

No necesitaba correr: aún era pronto.

b) Conclusión o resumen de la proposición anterior. Por ejemplo:

Varios vecinos monopolizaron la reunión con problemas particulares: no llegaron a ponerse de acuerdo.

c) Verificación o explicación de la proposición anterior, que suele tener un sentido más general. Por ejemplo:

La paella es un plato de la cocina española muy completo desde el punto de vista nutritivo: cuenta con la fécula del arroz, las proteínas de sus carnes y pescados, y la fibra de sus verduras.

5.3.5. Se utilizan los dos puntos para separar la ejemplificación del resto de la oración. Ejemplos:

De vez en cuando tiene algunos comportamientos inexplicables: hoy ha venido a la oficina con las zapatillas de andar por casa.

Puedes escribir un texto sobre algún animal curioso: el ornitorrinco, por ejemplo.

5.3.6. En textos jurídicos y administrativos —decretos, sentencias, bandos, edictos, certificados o instancias—, se colocan dos puntos después del verbo, escrito con todas sus letras mayúsculas (véase *3.2.5*), que presenta el objetivo fundamental del documento. La primera palabra del texto que sigue a este verbo se escribe siempre con inicial mayúscula y el texto forma un párrafo diferente. Por ejemplo:

CERTIFICA:

Que D. José Álvarez García ha seguido el Curso de Técnicas Audiovisuales celebrado en la Escuela de Cine durante los meses de abril y mayo del presente año.

5.4. Uso del punto y coma

El punto y coma (;) indica una pausa superior a la marcada por la coma e inferior a la señalada por el punto.

Se utiliza en estos casos:

5.4.1. Para separar los elementos de una enumeración cuando se trata de expresiones complejas que incluyen comas. Ejemplos:

La chaqueta es azul; los pantalones, grises; la camisa, blanca; y el abrigo, negro.

Cada uno de los grupos de trabajo preparará un taller: el primer grupo, el taller de cerámica; el segundo, el taller de telares; el tercero, el taller de cestería.

5.4.2. Para separar proposiciones yuxtapuestas, especialmente cuando en estas se ha empleado la coma. Ejemplos:

Era necesario que el hospital permaneciese abierto toda la noche; hubo que establecer turnos.

La muchacha, gozosa, corría hacia su casa; sus padres acababan de llegar.

La situación económica de la empresa, agravada en los últimos tiempos, era preocupante; se imponía una acción rápida y contundente si se deseaba salvar los puestos de trabajo.

A las cinco de la madrugada aún había luz en su habitación; seguramente, se había quedado dormido leyendo.

En muchos de estos casos, se podría optar por separar los períodos con punto y seguido. La elección del punto y seguido o del punto y coma depende de la vinculación semántica que exista entre las oraciones o proposiciones. Si el vínculo es débil, se prefiere usar un punto y seguido; mientras que, si es más sólido, es conveniente optar por el punto y coma.

También sería posible separar los mencionados períodos con dos puntos, puesto que casi siempre subyacen las mismas relaciones expresadas en el apartado *5.3.4*.

5.4.3. Se suele colocar punto y coma, en vez de coma, delante de conjunciones o locuciones conjuntivas como *pero, mas* y *aunque*, así como *sin embargo, por tanto, por consiguiente, en fin*, etc., cuando los períodos tienen cierta longitud y encabezan la proposición a la que afectan. Ejemplos:

Su discurso estuvo muy bien construido y fundamentado sobre sólidos principios; pero no consiguió convencer a muchos de los participantes en el congreso.

Los jugadores se entrenaron intensamente durante todo el mes; sin embargo, los resultados no fueron los que el entrenador esperaba.

Las últimas pruebas que han llegado de la imprenta parecen muy limpias; por tanto, creo que no tardaré mucho en revisarlas y devolverlas[37].

Si los bloques no son muy largos, se prefiere la coma (véase el apartado *5.2.7*). Ejemplos:

Vendrá, pero tarde.

Lo hizo, aunque de mala gana.

Si los períodos tienen una longitud considerable, es mejor separarlos con punto y seguido. Por ejemplo:

Este verano, varios vecinos del inmueble tienen previsto poner en venta sus respectivas viviendas. Por consiguiente, son previsibles numerosas visitas de posibles compradores.

5.5. Uso de los puntos suspensivos

Los puntos suspensivos (...) suponen una interrupción de la oración o un final impreciso.

Después de los puntos suspensivos, cuando cierran un enunciado, se escribe mayúscula. Por ejemplo:

El caso es que si lloviese... Mejor no pensar cosa tan improbable.

Estamos ante un bosque mediterráneo de encinas, alcornoques, pinos... Bajo estos árboles es fácil encontrar níscalos en otoños lluviosos.

Cuando los puntos suspensivos no cierran un enunciado y este continúa tras ellos, se escribe minúscula. Por ejemplo:

Estoy pensando que... aceptaré; en esta ocasión debo arriesgarme.

Se usan los puntos suspensivos en los siguientes casos:

5.5.1. Al final de enumeraciones abiertas o incompletas, con el mismo valor que la palabra *etcétera*.

[37] Sobre el uso de la coma después de algunas de estas locuciones conjuntivas, véase el apartado *5.2.8*.

Ejemplos:

Su tienda es como las de los pueblos, donde se vende de todo: comestibles, cacharros, ropas, juguetes...

Puedes hacer lo que te apetezca más: leer, ver la televisión, escuchar música...

5.5.2. Cuando se quiere expresar que antes de lo que va a seguir ha habido un momento de duda, temor o vacilación. Ejemplos:

Iré; no iré... Debo decidirme pronto.

Espero una llamada del hospital... Seguro que son buenas noticias.

No sé... Creo que... bueno, sí, me parece que voy a ir.

En ocasiones, la interrupción del enunciado sirve para sorprender al lector con lo inesperado de la salida. Por ejemplo:

Se convocó a una junta, se distribuyeron centenares de papeles anunciándola y, al final, nos reunimos... cuatro personas.

5.5.3. Para dejar un enunciado incompleto y en suspenso se utilizan los puntos suspensivos. Por ejemplo:

Fue todo muy violento, estuvo muy desagradable... No quiero seguir hablando de ello.

5.5.4. También se emplea este signo de puntuación cuando se reproduce una cita textual, sentencia o refrán, omitiendo una parte. Ejemplos:

En ese momento de indecisión, pensé: «Más vale pájaro en mano...» y acepté el dinero.

El escolar recitaba muy solemne: «Con diez cañones por banda...».

5.5.5. Se escriben tres puntos dentro de paréntesis (...) o corchetes [...] cuando al transcribir literalmente un texto se omite una parte de él. Por ejemplo:

Yo fui loco y ya soy cuerdo; fui don Quijote de la Mancha y soy agora [...] Alonso Quijano el Bueno.

(Cervantes: *Quijote*, II, LXXIV)

5.5.6. Combinación de los puntos suspensivos con otros signos

Tras los puntos suspensivos no se escribe nunca punto. Sin embargo, sí pueden colocarse otros signos de puntuación, como la coma, el punto y coma y los dos puntos. Ejemplos:

Cuando decidas los colores, las telas, el tipo de mobiliario..., ven a verme y haremos un presupuesto.

Pensándolo bien...: mejor que no se presente.

Ya habían llegado los libros, los ordenadores, el papel...; al día siguiente empezaría a trabajar.

Los signos de interrogación o exclamación se escriben delante o detrás de los puntos suspensivos, dependiendo de que el enunciado que encierran esté completo o incompleto. Ejemplos:

¿Me habrá traído los libros?... Seguro que sí.

¡Si te dije que...! Es inútil, nunca haces caso a nadie.

Tanto la coma, el punto y coma y los dos puntos como los signos de interrogación y exclamación se escribirán inmediatamente, sin un espacio que los separe de los puntos suspensivos, tal y como muestran los ejemplos anteriores[38].

5.6. Uso de los signos de interrogación y de exclamación

Los signos de interrogación (¿?) y exclamación (¡!) encierran enunciados que, respectivamente, interrogan o exclaman. Los primeros se utilizan para delimitar enunciados interrogativos directos; los segundos demarcan enunciados exclamativos, también en estilo directo, e interjecciones. Ejemplos:

¿Comisteis ayer en casa?

¿Dónde has comprado ese traje?

¡Eso es una injusticia!

¡Qué magnífica pintura!

¡Ay! ¡Eh! ¡Oh!

En la utilización de tales signos es preciso tener en cuenta estas consideraciones generales:

5.6.1. Los signos de interrogación y de exclamación son dos en cada caso: los signos que indican apertura (¿ ¡) y los signos que indican cierre (? !); se colocan al principio y al final del enunciado interrogativo y exclamativo respectivamente.

En nuestra lengua es obligatorio poner siempre el signo de apertura, que no deberá suprimirse a imitación de lo que ocurre en la ortografía de otras lenguas, en las que solo se usa el signo final porque tienen otras marcas gramaticales que suplen el primero. No obstante, véase *5.6.5.*

5.6.2. Después de los signos que indican cierre de interrogación o exclamación (? !) no se escribe nunca punto.

5.6.3. El signo de principio de interrogación (¿) o de exclamación (¡) se ha de colocar donde empieza la pregunta o la exclamación, aunque no comience con él el enunciado. Ejemplos:

Con respecto al impacto ambiental, ¿se ha previsto algún tipo de medidas para que su efecto sea el menor posible?

Si consigues la plaza, ¡qué alegría se va a llevar tu padre!

Obsérvese como los vocativos y las proposiciones subordinadas, cuando ocupan el primer lugar en el enunciado, se escriben fuera de la pregunta o de la exclamación. Sin embargo, si están colocados al final, se consideran dentro de ellas. Ejemplos:

Susana, ¿has decidido qué vas a hacer? / ¿Has decidido qué vas a hacer, Sonia?

Pepe, ¡cuánto me alegro de que hayas venido! / ¡Cuánto me alegro de que hayas venido, Pepe!

Si no responde, ¿qué le vamos a decir? / ¿Qué le vamos a decir si no responde?

[38] Tales signos son los que imponen en estos casos el uso de la mayúscula o la minúscula en la palabra siguiente.

5.6.4. Cuando se escriben varias preguntas o exclamaciones seguidas y estas son breves, se puede optar por considerarlas oraciones independientes, con sus correspondientes signos de apertura y cierre, y con mayúscula al comienzo de cada una de ellas. Ejemplos:

¿Dónde estás? ¿A qué hora piensas volver?

¡Quedan cinco minutos! ¡Llegamos tarde! ¡Date prisa!

Pero también es posible considerar el conjunto de las preguntas o exclamaciones como un único enunciado. En este caso hay que separarlas por comas o por puntos y comas, y solo en la primera se escribirá la palabra inicial con mayúscula. Ejemplos:

¿Cómo te llamas?, ¿en qué trabajas?, ¿cuándo naciste?, ¿dónde?

¡Cómo ha nevado esta noche!; ¡qué blanco está todo!; ¡qué frío vamos a pasar hoy!

5.6.5. En ocasiones, se utilizan los signos de final de interrogación (?) o de exclamación (!) entre paréntesis.

a) El signo de final de interrogación entre paréntesis expresa duda o ironía. Ejemplos:

Andrés Sánchez López es el presidente (?) de la asociación.

Tendría mucha gracia (?) que llegara a la cita con un día de retraso.

b) El signo de final de exclamación entre paréntesis expresa sorpresa o ironía. Ejemplos:

Un joven de treinta y seis años (!) fue el ganador del concurso de composición.

Está más gordo que nunca, pero dice que solo pesa ochenta kilos (!) en la báscula de su casa.

5.7. Uso de los paréntesis

Los paréntesis **( )** son signos que encierran elementos incidentales o aclaratorios intercalados en un enunciado.

Los paréntesis se usan en los siguientes casos:

5.7.1. Cuando se interrumpe el sentido del discurso con un inciso aclaratorio o incidental, sobre todo si este es largo o de escasa relación con lo anterior o posterior. Ejemplos:

El abuelo de Alberto (en su juventud fue un brillante cirujano) parecía una estatua sentado en aquel sillón.

Las asambleas (la última duró cuatro horas sin ningún descanso) se celebran en el salón de actos [39].

5.7.2. Para intercalar algún dato o precisión: fechas, lugares, significado de siglas, el autor u obra citados... Ejemplos:

El año de su nacimiento (1616) es el mismo en que murió Cervantes.

Toda su familia nació en Córdoba (Argentina).

Una ONG (organización no gubernamental) ha de ser, por principio, una asociación sin ánimo de lucro.

---

[39] Para estos incisos también se pueden utilizar rayas (véase 5.9.1).

«Más obran quintaesencias que fárragos» (Gracián)[40].

5.7.3. En ocasiones se utilizan los paréntesis para evitar introducir una opción en el texto. En estos casos se puede encerrar dentro del paréntesis una palabra completa o solo uno de sus segmentos[41]. Ejemplos:

En el documento se indicarán el (los) día(s) en que haya tenido lugar la baja.

Se necesita chico(a) para repartir pedidos[42].

5.7.4. Cuando se reproducen o transcriben textos, códices o inscripciones con abreviaturas, se pueden utilizar los paréntesis para reconstruir las palabras completas o los elementos que faltan en el original y se suplen[43]. Por ejemplo:

Imp(eratori) Caes(ari).

5.7.5. En la transcripción de textos se utilizan tres puntos entre paréntesis para dejar constancia de que se omite en la cita un fragmento del texto[44]. Por ejemplo:

Hasta aquí (...) la obra visible *de Menard, en su orden cronológico. Paso ahora a la otra: la subterránea, la interminablemente heroica, la impar. También, ¡ay de las posibilidades del hombre!, la inconclusa. Esa obra (...) consta de los capítulos noveno y trigésimo octavo de la primera parte del don Quijote y de un fragmento del capítulo veintidós.*

(Jorge Luis Borges: *Ficciones*)

5.7.6. Las letras o números que encabezan clasificaciones, enumeraciones, etc. pueden situarse entre paréntesis o seguidas del paréntesis de cierre. Ejemplos:

Estos libros podrán encontrarse en los lugares siguientes:

***(a)** En los estantes superiores de la sala de juntas.*
***(b)** En los armarios de la biblioteca principal.*

O bien:
Estos libros podrán encontrarse en los lugares siguientes:

***a)** En los estantes superiores de la sala de juntas.*
***b)** En los armarios de la biblioteca principal.*

5.7.7. Combinación de los paréntesis con otros signos

Los signos de puntuación correspondientes al período en el que va inserto el texto entre paréntesis se colocan siempre después del de cierre. Ejemplos:

Tenía varios hermanos (era el primogénito de una numerosa familia), pero no los veía desde hacía años.

---

[40] En algunos de estos casos también se pueden utilizar comas o rayas. Véanse *5.2.5* y *5.9.1*.
[41] En este uso puede alternar con la barra. Véase *5.11.3d*.
[42] Fuera de anuncios, circulares o algunos textos de tipo técnico, se recomienda evitar este uso, especialmente en documentos personalizados.
[43] También es posible utilizar en este caso los corchetes. Véase *5.8.3*.
[44] En este uso suelen alternar con los corchetes. Véase *5.5.5*.

¿En qué año se creó la ONU (Organización de las Naciones Unidas)?

El texto recogido dentro de los paréntesis tiene una puntuación independiente. Por eso, si el enunciado entre paréntesis es interrogativo o exclamativo, los signos de interrogación y exclamación se colocan dentro de los paréntesis. Ejemplos:

Ángel Peláez (si supieses lo que opina sobre la ecología y los ecologistas, no te caería tan bien) es el nuevo responsable del proyecto.

La manía de Ernesto por el coleccionismo (lo colecciona todo: sellos, monedas, relojes, plumas, llaveros...) ha convertido su casa en un almacén.

Su facilidad para los idiomas (¡habla con fluidez siete lenguas!) le ha abierto muchas puertas.

5.8. Uso de los corchetes

Los corchetes [ ] se utilizan por regla general de forma parecida a los paréntesis que incorporan información complementaria o aclaratoria.

La combinación de los corchetes con otros signos ortográficos es idéntica a la de los paréntesis (véase *5.7.7*).

Los corchetes se utilizan en las ocasiones siguientes:

5.8.1. Cuando dentro de un enunciado o texto que va entre paréntesis es preciso introducir alguna nota aclaratoria o precisión. Por ejemplo:

Una de las últimas novelas que publicó Benito Pérez Galdós (algunos estudiosos consideran su obra Fortunata y Jacinta *[1886-87] la mejor novela española del siglo XIX) fue* El caballero encantado *(1909).*

5.8.2. En poesía se coloca un solo corchete de apertura delante de las últimas palabras de un verso para indicar que no caben en la línea anterior. Ejemplo:

Soñaba en ese entonces en forjar un poema,
de arte nervioso y nueva obra audaz y suprema,

escogí entre un asunto grotesco y otro trágico,
llamé a todos los ritmos con un conjuro
[mágico
y los ritmos indóciles vinieron acercándose,
juntándose en las sombras, huyéndose y
[buscándose.

(José Asunción Silva: *El libro de versos*)

5.8.3. Cuando, en un texto transcrito, el copista o editor quiere incorporar alguna parte que falta, aclaración, nota, desarrollo de una abreviatura o cualquier interpolación ajena al texto original[43], se usan los corchetes. Por ejemplo:

[43] También es posible utilizar los paréntesis con esta función. Véase *5.7.4.*

La nieve hermoseaba [texto tachado: los parques y edificios de] la ciudad aquella fría mañana de diciembre.

5.8.4. También se utilizan los corchetes que encierran tres puntos suspensivos [...] cuando en un texto transcrito se omite una parte de él, ya sea una sola palabra o un fragmento[40]. Por ejemplo:

Pensé en lo que usted me había enseñado: que nunca hay que odiar a nadie. Le sonreí para decírselo; pero después pensé que él no pudo ver mi sonrisa [...] por lo negra que estaba la noche.

(Juan Rulfo: *Pedro Páramo*)

5.9. Uso de la raya

La raya o guion largo (—) se puede usar aisladamente, o bien, como en el caso de otros signos de puntuación, para servir de signo de apertura y cierre que aísle un elemento o enunciado.

Este signo se utiliza con los fines siguientes:

5.9.1. Para encerrar aclaraciones o incisos que interrumpen el discurso. En este caso se coloca siempre una raya de apertura antes de la aclaración y otra de cierre al final. Por ejemplo:

Llevaba la fidelidad a su maestro —un buen profesor— hasta extremos insospechados.

Esperaba a Emilio —un gran amigo—. Lamentablemente, no vino.

En este uso, las rayas pueden ser sustituidas por los paréntesis e incluso por comas (véanse *5.7.1, 5.7.2* y *5.2.5*). La diferencia entre una u otra opción depende de cómo perciba quien escribe el grado de conexión que el inciso mantiene con el resto del enunciado.

5.9.2. Para señalar cada una de las intervenciones de un diálogo sin mencionar el nombre de la persona o personaje al que corresponde. En este caso se escribe una raya delante de las palabras que constituyen la intervención. Por ejemplo:

—¿Qué has hecho esta tarde?
—Nada en especial. He estado viendo la televisión un rato.

5.9.3. Para introducir o encerrar los comentarios o precisiones del narrador a las intervenciones de los personajes. Se coloca una sola raya delante del comentario del narrador, sin necesidad de cerrarlo con otra, cuando las palabras del personaje no continúan inmediatamente después del comentario. Por ejemplo:

—Espero que todo salga bien —dijo Azucena con gesto ilusionado.

Se escriben dos rayas, una de apertura y otra de cierre, cuando las palabras del narrador interrumpen la intervención del personaje y esta continúa inmediatamente después. Por ejemplo:

—Lo principal es sentirse viva —añadió Pilar—. Afortunada o desafortunada, pero viva.

Tanto en un caso como en el otro, si fuese necesario poner detrás de la intervención del narrador un signo de puntuación, una coma o un punto, por ejemplo, se colocará después de sus palabras y tras la raya de cierre (si la hubiese). Por ejemplo:

[40] En este uso suelen alternar con los paréntesis. Véanse *5.5.5* y *5.7.5*.

—¿Deberíamos hablar con él? —preguntó Juan—. Es el único que no lo sabe.
—Sí —respondió la secretaria—, pero no podemos decirle toda la verdad.

5.9.4. En algunas listas, como índices alfabéticos de libros o bibliografías, la raya sirve para indicar que en ese renglón se omite una palabra, ya sea un concepto antes citado o el nombre de un autor que se repite. Ejemplos:

Verbos intransitivos
— transitivos
— irregulares
— regulares

Ortega y Gasset, J.: *España invertebrada* (1920-22).
— La rebelión de las masas (1930).
— Idea del teatro (1946).

5.10. Uso de las comillas

Hay diferentes tipos de comillas: las comillas angulares, también llamadas latinas o españolas (« »), las inglesas (“ ”) y las simples (‘ ’). Por lo general, es indistinto el uso de uno u otro tipo de comillas dobles; pero suelen alternarse cuando hay que utilizar comillas dentro de un texto ya entrecomillado. Por ejemplo:

Al llegar el coche deportivo, Lola susurró: «Vaya “cacharro” que se ha comprado Tomás».

Se utilizan comillas en los casos siguientes:

5.10.1. Para reproducir citas textuales de cualquier extensión[47]. Ejemplos:

Fue entonces cuando la novia dijo: «Sí».

Sus palabras fueron: «Por favor, el pasaporte».

Dice Miguel de Unamuno en La novela de don Sandalio:
«He querido sacudirme del atractivo del Casino, pero es imposible; la imagen de Don Sandalio me seguía a todas partes. Ese hombre me atrae como el que más de los árboles del bosque; es otro árbol más, un árbol humano, silencioso, vegetativo. Porque juega al ajedrez como los árboles dan hoja».

Cuando se ha de intercalar un comentario o intervención del narrador o transcriptor de la cita, no es imprescindible cerrar las comillas para volver a abrirlas después del comentario, pero puede hacerse. Para intercalar tales intervenciones, es preferible encerrarlas entre rayas. Por ejemplo:

«Los días soleados como este —comentó Silvia— me encantan».

5.10.2. En textos narrativos, a veces se utilizan las comillas para reproducir los pensamientos de los personajes, en contraste con el uso de la raya, que transcribe sus intervenciones propiamente dichas. Por

[47] En textos largos fue práctica, ahora inhabitual, colocar comillas de cierre (») al principio de cada línea para recordar que continuaba la cita. Por ejemplo:

En opinión de un escritor célebre: «El hombre tiene aptitud, por su naturaleza, para
»habitar en todos los países del mundo: en los arenales del desierto, en los
»montes más encumbrados, en los climas polares puede vivir y propagarse. No
»así los animales, que, sujetos a más estrechos límites, perecen fuera de ellos o
»arrastran vida penosa».

ejemplo:

—Es una mujer hermosa, hermosísima; si ustedes quieren, de talento, digna de otro teatro, de volar más alto...; si ustedes me apuran, diré que es una mujer superior —si hay mujeres así— pero al fin es mujer, et nihil humani...

No sabía lo que significaba ese latín, ni adónde iba a parar, ni de quién era, pero lo usaba siempre que se trataba de debilidades posibles.

Los socios rieron a carcajadas.

«¡Hasta en latín sabía maldecir el pillastre!», pensó el padre, más satisfecho cada vez de los sacrificios que le costaba aquel enemigo.

(Clarín: *La Regenta*, cap. IV)

5.10.3. Para indicar que una palabra o expresión es impropia, vulgar o de otra lengua, o que se utiliza irónicamente o con un sentido especial. Ejemplos:

Dijo, cargado de razón, que el asunto tenía algunas «prorrogativas».

En el salón han puesto una «boiserie» que les ha costado un dineral[48].

Últimamente está muy ocupado con sus «negocios».

5.10.4. Para citar títulos de artículos, poemas, cuadros...[49]. Ejemplos:

El artículo de Amado Alonso titulado «Noción, emoción, acción y fantasía en los diminutivos» está recogido junto con otros en un volumen antológico: Estudios lingüísticos. Temas españoles.

Nos leyó en voz alta el «Romance sonámbulo» del Romancero gitano.

En esta sala se puede ver el «Esopo» de Velázquez.

5.10.5. Cuando en un texto se comenta o se trata una palabra en particular, esta se aísla escribiéndola entre comillas[50]. Por ejemplo:

Como modelo de la primera conjugación, se utiliza usualmente el verbo «amar».

5.10.6. Cuando se aclara el significado de una palabra, este se encierra entre comillas. En tal caso se prefiere utilizar comillas simples. Por ejemplo:

«Espiar» ('acechar') no significa lo mismo que «expiar» las faltas.

5.10.7. Combinación de las comillas con otros signos

Los signos de puntuación correspondientes al período en el que va inserto el texto entre comillas se colocan siempre después de las comillas de cierre. Ejemplos:

Sus palabras fueron: «No lo haré»; pero al final nos ayudó.

¿De verdad ha dicho: «Hasta luego»?

---

[48] En textos impresos en letra redonda es más frecuente reproducir los extranjerismos con cursiva. Véase 2.12.

[49] En los textos impresos, los títulos de los libros, sin embargo, suelen escribirse con letra cursiva; en los textos manuscritos o mecanografiados es frecuente subrayarlos.

[50] También se puede distinguir la palabra escribiéndola subrayada o con un tipo de letra diferente, si esto es posible. Por ejemplo: *La palabra* entre, *incluida tradicionalmente en la lista de preposiciones, no funciona a veces como tal.*

¡Menudo «collage»!

El texto recogido dentro de las comillas tiene una puntuación independiente y lleva sus propios signos ortográficos. Por eso, si el enunciado entre comillas es interrogativo o exclamativo, los signos de interrogación y exclamación se colocan dentro de estas. Ejemplos:

Se dirigió al dependiente: «Por favor, ¿dónde puedo encontrar cañas de pescar?».

«¡Qué ganas tengo de que lleguen las vacaciones!», exclamó.

5.11. Otros signos ortográficos

La escritura del español también cuenta con otros signos auxiliares, que se exponen a continuación.

5.11.1. Diéresis o crema

La diéresis o crema (¨) es un signo que se coloca encima de las vocales en las siguientes ocasiones[51]:

a) Para indicar que ha de pronunciarse la vocal *u* en las combinaciones *gue* y *gui*. En este caso, el uso de la diéresis es preceptivo. Ejemplos: *vergüenza, pingüino, argüir.*

b) En textos poéticos, la diéresis puede usarse colocada sobre la primera vocal de un posible diptongo, para indicar que no existe. De esa forma la palabra a la que afecta y el verso en que se incluye cuentan con una sílaba más. Ejemplos:

El dulce murmurar deste rüido,
el mover de los árboles al viento,
el suave olor del prado florecido.

(Garcilaso de la Vega: *Égloga II*)

5.11.2. Guion

El guion (-) es un trazo horizontal de menor longitud que el signo llamado raya. Se utiliza básicamente cuando es necesario hacer divisiones dentro de una palabra y no se escribe entre espacios en blanco[52].

Tiene diferentes usos.

a) Se utiliza para separar, en determinados casos, los dos elementos que integran una palabra compuesta. Hay dos situaciones dignas de mención:

1.º Los compuestos de nueva creación formados por dos adjetivos, el primero de los cuales conserva invariable la terminación masculina singular, mientras el segundo concuerda en género y número con el nombre al que se refiere. Ejemplos: *tratado teórico-práctico, lección teórico-práctica, cuerpos técnico-administrativos.*

2.º Cuando dos gentilicios forman una palabra compuesta, esta se puede escribir separando o no ambos elementos con un guion. Si el compuesto resultante se siente como consolidado, lo escribiremos sin guion. Ejemplos: *hispanoárabe, francocanadiense.* Si el compuesto no es sentido como unidad, puede escribirse con guion. Ejemplos: *luso-japonés, hispano-ruso.*

---

[51] Para su empleo en textos antiguos, véase *2.2.3.*

[52] Sin embargo, cuando se emplea metalingüísticamente para separar las sílabas de una palabra, se suele escribir entre espacios, como aparece en los ejemplos de separación de sílabas registrados en este mismo apartado.

b) El guion sirve para dividir una palabra al final de renglón cuando no cabe en él completa. Para realizar esta división hay que tener en cuenta las consideraciones que se detallan más adelante (véase 5.12).

1.° Cuando la palabra contenga una *h* intercalada precedida de consonante, el guion se colocará siempre delante de la *h*, tratándola como principio de sílaba. Ejemplos: *des - hidratar, in - humano.*

2.° Los dígrafos *ll, rr* y *ch* no se pueden dividir con guion, por representar cada uno de ellos un único fonema. Ejemplos: *ca - llar, ca - rro, le - chuga.*

3.° Cuando en una palabra aparecen dos consonantes seguidas, generalmente la primera pertenece a la sílaba anterior y la segunda a la sílaba siguiente[33]. Ejemplos: *in - ten - tar, es - pal - da, es - to, suc - ción.*

Los grupos consonánticos formados por una consonante seguida de *l* o *r*, como *bl, cl, fl, gl, kl, pl, br, cr, dr, fr, gr, kr, pr, tr,* no pueden separarse y siempre inician sílaba. Ejemplos: *ha - blar, su - primir, de - trás, re - clamar, in - flamar*[34].

4.° Cuando tres consonantes van seguidas en una palabra, se reparten entre las dos sílabas respetando la inseparabilidad de los siguientes grupos consonánticos: los constituidos por una consonante más *l* o *r* vistos en el párrafo anterior, que siempre encabezan sílaba, y los grupos *st, ls, ns, rs, ds,* que siempre cierran la sílaba. Ejemplos: *ham - bre, am - plio, in - flar, en - trar, des - gracia, ist - mo, sols - ticio, cons - tante, abs - tenerse, supers - tición.*

5.° Cuando son cuatro las consonantes consecutivas en una palabra, las dos primeras forman parte de la primera sílaba y las dos restantes de la segunda. Ejemplos: *cons - treñir, abs - tracto, ads - cripción.*

6.° Es preferible no segmentar las palabras de otras lenguas al final de renglón, a no ser que se conozcan las reglas vigentes en los idiomas respectivos.

7.° Las siglas y acrónimos, así como las abreviaturas, no pueden dividirse al final de renglón. Así sucede, por ejemplo, con *UNESCO*. Se admite la división en los acrónimos que han pasado a incorporarse al léxico general, escritos, en consecuencia, con minúscula. Ejemplos: *lá - ser, ra - dar.*

c) Cuando se antepone el guion a una parte de una palabra (sílaba, morfema, etc.), indica que esta va en posición final. Ejemplos: *-illo, -idad, -ar*[35]. Cuando se pospone el guion a esa misma parte, indica que esta va en posición inicial. Ejemplos: *post-, re-, cant-.* Si el elemento en cuestión se coloca entre guiones, se entiende que está en interior de palabra. Ejemplos: *-ec-, -in-, -bl-.*

d) El guion también se emplea para unir palabras con un valor de enlace similar al de una preposición o una conjunción. Ejemplos: *la línea de metro Ventas-Cuatro Caminos, el partido Peñarol-Nacional, una relación amor-odio.*

---

[33] En América, Canarias y algunas áreas peninsulares, la secuencia *tl* forma grupo inseparable (se pronuncia, por ejemplo, *a-tlas*). En otras zonas de España tiende a producirse corte silábico entre las dos consonantes (se pronuncia *at-las*).

[34] Recuérdese, sin embargo, lo dicho en *2.9.1*, nota 20.

[35] Si a la sílaba anterior a un sufijo o elemento compositivo le corresponde llevar tilde, esta aparecerá sobre el guion. Ejemplos: *-ico, ca (alcohólico), -il (portátil).*

5.11.3. Barra

La barra (*/*) tiene los siguientes usos[56]:

a) Sirve para señalar el límite de los versos en los textos poéticos reproducidos en línea seguida. En este caso se escribe entre espacios. Por ejemplo:

¡Y si después de tantas palabras, / no sobrevive la palabra! /¡Si después de las alas de los pájaros, / no sobrevive el pájaro parado! / ¡Más valdría, en verdad, / que se lo coman todo y acabemos!

(César Vallejo: *Poemas póstumos*)

b) En algunas transcripciones de textos, se utiliza para señalar el cambio de línea en el original y también se escribe entre espacios. Por ejemplo:

QVINTA / PARTE DE FLOR / DE ROMANCES NVE / uos, nu[n]ca hasta agora impressos: / Llamado Ramillete de Flores: / De muchos, graues, y diuer / sos Autores. Recopi / lados no co[n] po / co traba / jo

c) Tiene valor preposicional en expresiones como *120 km/h, Real Decreto Legislativo 1/1995 de 24 de marzo, salario bruto 220 000 pts./mes*. En este uso se escribe sin separación alguna de los signos gráficos que une.

d) Colocada entre dos palabras o entre una palabra y un morfema, puede indicar también la existencia de dos o más opciones posibles[57]. En este caso no se escribe entre espacios. Ejemplos:

El/los día/s detallado/s.

Es el tipo de bromas y/o mentiras piadosas que Inés no soportaba.

(Alfredo Bryce Echenique: *La vida exagerada de Martín Romaña*)

e) Forma parte de abreviaturas como *c/* (por *calle*), *c/c* (por *cuenta corriente*), etc.

5.11.4. Apóstrofo

El apóstrofo (') no tiene prácticamente vigencia en el español actual.

Este signo solía emplearse antiguamente, sobre todo en poesía, para indicar la omisión o elisión de una vocal. Ejemplos: *d'aquel*, por *de aquel; l'aspereza*, por *la aspereza; qu'es*, por *que es*. Este uso se conserva en algunas ediciones actuales de obras antiguas.

Los nombres propios de otras lenguas que incluyen apóstrofo, lo conservan. Ejemplos: *O'Donnell, D'Ors*, etc.

5.11.5. Signo de párrafo

El signo de párrafo (§) se emplea, seguido de un número, para indicar divisiones internas dentro de los capítulos. Ejemplos: *§ 12, § 23*.

También se utiliza en las remisiones y citas de estas mismas divisiones. Por ejemplo: *Véase § 16*.

---

[56] En informática se utiliza también una barra inclinada hacia la izquierda (\).

[57] En este uso puede alternar con los paréntesis. Véase *5.7.3*. Recuérdese lo allí expuesto acerca de la oportunidad de su utilización.

5.11.6. Asterisco

El asterisco (*) es un signo en forma de estrella que se utiliza con estas funciones:

a) Como signo de llamada de nota al margen o a pie de página dentro de un texto. En este caso se pueden situar uno, dos, tres y hasta cuatro asteriscos en llamadas sucesivas dentro de una misma página. En ocasiones, estos asteriscos se encierran entre paréntesis: (*).

b) Se antepone a una palabra o expresión para señalar su incorrección, bien en su uso o bien en su ortografía. Ejemplos: *pienso *de que volverá tarde* (forma correcta: *pienso que volverá tarde*); **cocreta* (forma correcta: *croqueta*); **a grosso modo* (forma correcta: *grosso modo*).

c) En Lingüística histórica, antepuesto a una palabra, indica que se trata de un vocablo hipotético, resultado de una reconstrucción. Ejemplos: **bava*, **appariculare*.

d) El asterisco puede tener también otros valores circunstanciales especificados en un escrito o publicación concretos.

5.11.7. Llaves

Las llaves **{ }** constituyen un signo doble que encierra texto, aunque también se pueden utilizar de forma aislada.

En ambos casos, estos signos se utilizan en cuadros sinópticos y esquemas para agrupar opciones diferentes, establecer clasificaciones, desarrollar lo expresado inmediatamente antes o después, etc.

5.12. División de palabras al final de renglón

Para dividir palabras al final de una línea se utilizará el guion (véase *5.11.2*b) de acuerdo con estas normas:

a) El guion no debe separar letras de una misma sílaba. Por ejemplo, la palabra *teléfono* podría dividirse de las siguientes maneras:

te- *léfono*	*telé-* *fono*	*teléfo-* *no*

Sin embargo, cuando una palabra está integrada por otras dos que funcionan independientemente en la lengua, o por una de estas palabras y un prefijo, será potestativo dividir la voz resultante separando sus componentes, aunque la división no coincida con el silabeo de la palabra. Ejemplos:

no- *sotros*	*nos-* *otros*	*de-* *samparo*	*des-* *amparo*

b) Dos o más vocales seguidas no pueden separarse, tanto si constituyen un diptongo o un triptongo, por ejemplo en *can - ción, tiem - po, santi - güéis*, como si no lo constituyen y están en hiato, por ejemplo en *tea - tro, poé - tico*. Excepto si forman parte de dos segmentos distintos de una palabra compuesta, como se explicó en el párrafo anterior. Por ejemplo: *contra - espionaje*.

c) Cuando la primera sílaba de una palabra es una vocal, se evitará dejar esta letra sola al final de la

línea. Si la vocal está precedida de una *h*, se invalida la norma. Ejemplos: *amis - ta - des, he - re - de - ros.*

5.13. Usos no lingüísticos de algunos signos de puntuación

Aunque no constituyen materia estrictamente ortográfica, existen ciertos usos no lingüísticos de los signos de puntuación, generalmente referidos a notaciones o expresiones científicas y técnicas[58].

5.13.1. Usos no lingüísticos del punto

a) Aunque todavía es práctica común separar los millares, millones, etc., mediante un punto (o una coma en algunos lugares de América), la norma internacional establece que se prescinda de él. Para facilitar la lectura de estas expresiones, especialmente cuando constan de muchas cifras, se recomienda separarlas mediante espacios por grupos de tres. Por ejemplo: *4 829 430*. Sin embargo, no se utiliza nunca esta separación en la expresión de los años, en la numeración de páginas ni en los números de artículos, decretos o leyes. Ejemplos: *año 1942, página 1162, Real Decreto 1099/1986.*

b) Es aceptable, de acuerdo con la normativa internacional, el uso del punto para separar la parte entera de la parte decimal en las expresiones numéricas escritas con cifras. Por ejemplo: *3.1416*. Pero en este caso es preferible el uso de la coma (véase *5.13.2*).

c) En la expresión numérica del tiempo, el punto separa las horas de los minutos. Ejemplos: *15.30 h, 12.00 h*. En este uso pueden emplearse también los dos puntos (véase *5.13.3*a).

d) En Matemáticas, el punto indica la multiplicación de dos cantidades o expresiones, y se coloca siempre a media altura[59]. Ejemplo: $5 \cdot 4 = 20$; $2 \cdot (x + y) = 30$.

5.13.2. Usos no lingüísticos de la coma

Se utiliza la coma para separar la parte entera de la parte decimal en las expresiones numéricas escritas con cifras. Por ejemplo: *3,1416*. No obstante, la normativa internacional acepta también el uso del punto en este caso (véase *5.13.1*b).

5.13.3. Usos no lingüísticos de los dos puntos

a) Se utilizan dos puntos para separar las horas de los minutos en la expresión del tiempo[60], alternando con el uso del punto (véase *5.13.1*c). Ejemplos: *15:30 h, 12:00 h*.

b) También indican una división en expresiones matemáticas. Ejemplo: $8 : 2 = 4$. En este uso alternan con la barra (véase *5.13.5)* y con el símbolo ÷.

5.13.4. Usos no lingüísticos del guion

Se utiliza el guion con valor de enlace para unir dos números, sean consecutivos o no. De esta forma se hace referencia al intervalo que existe entre uno y otro. Este uso es válido tanto en el caso de números

---

[58] Este apartado sigue los usos del SI (Sistema Internacional de Unidades), reconocido oficialmente por la mayor parte de los países.

[59] Este uso alterna con el símbolo tradicional en forma de aspa. En la notación matemática se prescinde muchas veces de signos para indicar el producto de dos expresiones. Por ejemplo: $E = mc^2$ equivale a $E = m \cdot c^2$ (y también a $E = m \times c^2$).

[60] Tal empleo se ha extendido por su generalización en los relojes digitales.

arábigos como en el de números romanos. Ejemplos: *Alfonso Reyes (1889-1959); las páginas 23-45; los capítulos X-XIII; durante los siglos X-XI.*

Cuando, en la expresión de un período de tiempo, los dos primeros dígitos del año final del intervalo coinciden con los dos primeros dígitos del año inicial, aquellos pueden suprimirse. Por ejemplo: *la guerra civil española (1936-39).*

Pueden suprimirse también los dos primeros dígitos en la expresión de fechas (por ejemplo: *12-5-99*), otro de los usos del guion.

También se emplea el guion para unir una letra con el valor de abreviatura a una cifra o número romano y formar así expresiones que hacen referencia a una carretera. Ejemplos: *N-260, N-IV.*

5.13.5. Usos no lingüísticos de la barra

Los cocientes de magnitudes y unidades de medida se expresan mediante la barra. Ejemplos: *120 km/h, 10 m^3/s* (véase *5.11.3*c). En los cocientes de otras expresiones matemáticas, se podrán utilizar también los dos puntos (véase antes *5.13.3*b) y el símbolo ÷. Ejemplos: *x/y, 8/2.*

ANEXO B

MINERÍA DE TEXTOS

B.1 Mini Tutorial de JAPE

B.1.1 Patrones de anotación con JAPE

Si bien las expresiones regulares típicamente se aplican a cadenas de caracteres, el Motor de Patrones de Anotación en Java o Java Annotation Patterns Engine (JAPE) permite usar expresiones regulares sobre una estructura de datos más compleja como lo son las anotaciones.

Una gramática JAPE consiste de un conjunto de fases, cada una de las cuales está constituida por un conjunto de reglas patrón-acción. Las fases se ejecutan secuencialmente constituyendo una especie de cascada. El lado izquierdo de la regla (LHS) es una descripción de un patrón de anotación que se busca en los documentos analizados. En tanto que el lado derecha de la regla (RHS) especifica las acciones a realizar cuando el documento contiene el patrón descripto en LHS. La parte que precede a "-->" es LHS y la parte que le sigue es RHS. Las anotaciones que coinciden con el patrón de LHS pueden referenciarse en RHS mediante las etiquetas ligadas a los elementos del patrón.

Como puede verse a continuación, en el ejemplo "Phase: find1", para definir una gramática se comienza asignándole un nombre de fase. Se usa la palabra clave "Phase" porque como ya se mencionó una gramática se compone de fases que se ejecutan secuencialmente.

```
Phase:      find1
Input: Token SpaceToken Lookup DEFAULT_TOKEN
Options: control = appelt

Macro: FULLSTOP
({Token.string=="."})


Macro: PUNCT
({Token.string == "!"} | {Token.string == "?"})


Rule: Ldots
(FULLSTOP FULLSTOP FULLSTOP(FULLSTOP)+)
:variosPtos
--> {}

Rule: Split1
((PUNCT)[1,5]|(FULLSTOP) [1,3])
:split
-->
```

```
:split.Split = {kind = "internal"}
```

El nombre de una fase solo puede contener caracteres alfanuméricos y guión bajo, y no puede comenzar con número. Luego de darle un nombre a la gramática se recurre a la palabra reservada "Input" para proveer una lista de todos los tipos de anotación que usa la gramática. En la gramática find1 (del ejemplo enunciado antes) las anotaciones usadas son "Token SpaceToken Lookup DEFAULT_TOKEN". Si no se definen las anotaciones usadas, se considerarán todas las anotaciones. Luego del nombre y de los tipos de anotación se pueden especificar varias "Options":

- Control: define el método usada para decidir que regla aplicar cuando más de una regla coincide con un fragmento de texto (ver punto B.1.4 de este anexo).
- Debug: cuando se le asigna del valor "true", si la gramática se ejecuta con el modo "Appelt" y más de una regla coincide con la misma región de texto, los conflictos se muestran por pantalla.

El operador "." se usa tanto en LHS como en RHS para acceder a los atributos o propiedades de una anotación.

B.1.2 Lado Izquierdo (LHS)

El lado izquierdo de una gramática JAPE busca emparejamientos entre el patrón descripto y los fragmentos de texto, a la vez que trata de evitar emparejamientos no deseados. El patrón descripto en LHS puede consistir de una anotación, una anotación con un valor de atributo especifico, o una expresión regular. Los diversos recursos y operadores que pueden usarse en LHS para definir patrones se comentan a continuación.

B.1.2.1 Meta propiedades

Además de poder referenciar las características o atributos de una anotación, JAPE permite acceder a las meta propiedades de una anotación. El acceso a las meta propiedades se realiza mediante el símbolo "@" en lugar del símbolo "." después del nombre del tipo de anotación. En la actualidad se soportan tres meta propiedades:

- length: retorna la longitud del fragmento de texto cubierto por una anotación.
- string: retorna la cadena de caracteres cubierta por la anotación en el documento.
- cleanString: retorna la cadena de caracteres cubierta por la anotación en el documento pero descartando los espacios en blanco.

B.1.2.2 Macros

Permiten crear una definición que puede usarse múltiples veces en una gramática JAPE. En la definición de la gramática "find1" incluida en el apartado 1 de este anexo, se especificaron los macros "PUNCT" y "FULLSTOP" para ser usados posteriormente por las reglas Ldots y Split1.

B.1.2.3 Uso del contexto

El patrón a ser anotado por una regla se encierra entre paréntesis. Si el contexto precedente va ser incluido, se lo coloca antes del paréntesis. Si se necesita considerar el contexto siguiente, se lo coloca después de la etiqueta dada a la anotación. El contexto se usa cuando un patrón necesita ser reconocido solo si ocurre en determinadas situaciones pero, el contexto no forma parte del patrón a ser anotado por la regla. El contexto se describe exactamente de la misma forma que el patrón cuyas coincidencias se buscan en el documento. En la regla parentesisAclaratorio que se incluye a continuación, el contexto precedente está subrayado y el contexto siguiente se indica mediante letra cursiva.

```
Rule:parentesisAclaratorio
({Token.string =="("})
({Token.string != ")" }|{SpaceToken}|entreParentesis)+
:signoParent
({Token.string==")"})
-->{}
```

Se debe tener en cuenta que cuando en un documento se encuentra alguna coincidencia de una regla que usa contexto la regla consume el contexto y por lo tanto no puede ser reusado por otras reglas dentro de la misma fase.

B.1.2.4 Negación

En JAPE el caracter "!" se usa para señalar restricciones negativas. Una restricción negativa especifica la ausencia de anotaciones.

B.1.2.5 Operadores

Los operadores de composición y de coincidencia permiten combinar anotaciones y características de anotación en la definición de las reglas de una gramática JAPE. En la Tabla 1 se resume los operadores disponibles en JAPE.

Operador	Significado
!	Negación
|	Disyunción lógica
==	Igualdad
!=	Desigualdad
>	Mayor
<	Menor
>=	Mayor o igual
<=	Menor o igual
=~	Coincidencia parcial para expresiones regulares
!~	Desigualdad parcial para expresiones regulares
==~	Coincidencia para expresiones regulares
!=~	Desigualdad para expresiones regulares
{X contains Y}	Chequea si la anotación X contiene a la anotación Y
{X within Y}	Chequea si la anotación X está contenida dentro de la anotación Y

Tabla B.1. Operadores disponibles en JAPE

B.1.2.5.1 Operadores de composición

B.1.2.5.1.1 Operadores de unión y Kleene

Están disponibles los siguientes operadores de unión y kleene:

- "|": disyunción lógica.
- "*": cero o más ocurrencias.
- "?": cero o una ocurrencia.
- "+" una o más ocurrencias.

B.1.2.5.1.2 Notación de rango

Para especificar un rango se usa la notación $[x_1,x_2]$ donde x_1 es el límite inferior del rango y x_2 es el límite superior el rango. A modo de ejemplo se puede considerar los siguientes casos:

- ({Token}) [1,3] coincide con una secuencia de una a tres anotaciones del tipo Token en fila.
- ({Token.kind == number})[3] coincide con exactamente tres anotaciones Token del tipo number consecutivos.

B.1.2.5.2 Operadores de coincidencia

Los operadores de coincidencia se usan para especificar de qué manera debe tener lugar la coincidencia entre una especificación y una anotación en el documento.

B.1.2.5.2.1 Operadores de igualdad

Los operadores de igualdad en JAPE son "==" y "!=". El operador "!=" coincide en los casos en que el operador "=" no lo hace.

B.1.2.5.2.2 Operadores de comparación

Los operadores de comparación son "<", "<=", "=>" y ">". El significado de los operadores de comparación es bastante intuitivo, como ser {Token.length > 3} coincide con una anotación del tipo Token cuyo atributo length es mayor a 3.

B.1.2.5.2.3 Operadores de expresiones regulares

Los operadores de expresiones regulares son "=~", "==~", "!~" y "!=~". Estos operadores se usan para describir un patrón mediante expresiones regulares. Los operadores "=~" y "!~" admiten una coincidencia parcial de la anotación con la expresión regular, en cambio los operadores "==~" y "!=~" requieren que la expresión regular completa coincida o no coincida. Las expresiones regulares se deben especificar conforme a la sintaxis del API regex de la plataforma Java. En el patrón {Token.string==~"[a-z](\\S)*"} la expresión regular "[a-z](\\S)*" permite restringir la búsqueda a anotaciones del tipo Token cuya característica "string" sea igual a una letra o palabra en minúscula.

B.1.2.5.2.4 Operadores contextuales

Los operadores contextuales "contains" y "within" se usan para buscar anotaciones en el contexto de otras anotaciones.

- contains: se escribe de la forma {X contains Y}, devuelve "true" si una anotación del tipo X contiene completamente a una anotación del tipo Y.
- within: se escribe de la forma {X within Y}, devuelve "true" si una anotación del tipo X es cubierta completamente por una anotación del tipo Y.

Para cualquiera de los dos operadores el valor de Y puede ser una restricción completa.

B.1.3 Lado Derecho (RHS)

El RHS de una regla contiene información sobre la anotación a ser creada o manipulada. Al momento de crear una anotación la información sobre el fragmento de texto a ser anotado se transfiere desde el LHS de la regla por medio de las etiquetas. Luego los atributos y sus correspondientes valores son añadidos a la anotación. En la regla Split1

del apartado 1 al fragmento de texto referenciado por la etiqueta “split” se le asigna una anotación del tipo “Split” con la característica “kind” igual a “internal”. El RHS de una regla puede contener código Java para crear o manipular las anotaciones.

B.1.4 Prioridad

Todas las gramáticas tienen 5 posibles estilos de control: “brill”, “all”, “first”, “once” y “appelt”. Esta información se especifica al comienzo de la gramática.

- brill: significa que todas las reglas se activan cuando más de una regla coincide con la misma región del documento. Como resultado un segmento de texto puede tener asignado más de un tipo de entidad y no se necesita ordenamiento de prioridades. Con brill se ejecutan todas las reglas desde de la posición actual y se avanza buscando coincidencias desde la posición en el documento donde la coincidencia más larga finaliza.
- all: de manera similar al estilo brill ejecuta todas las reglas que coinciden en la misma región de texto pero, la búsqueda de coincidencias continúa desde el offset más próximo al actual.
- first: se activa la regla para la cual se encuentra la primera coincidencia en el documento y no se intenta obtener la coincidencia más larga.
- once: una vez que una regla se activa la fase completa finaliza su ejecución sin comprobar si el documento coincide con alguna de las otras reglas.
- appelt: solo una regla puede ser activada en la misma región de texto. Para decidir cual regla activar se tiene en cuenta las siguientes reglas de prioridad:
    - Para todas las reglas que coinciden con una región de texto comenzando desde un punto X, la regla que coincide con la región más larga se activa.
    - Si más de un regla coincide con la misma región de texto, se activa la que tiene la prioridad más alta.
    - Si se tiene más de una regla con la misma prioridad, se activa la definida primero en la gramática.

Para cada regla se puede definir explícitamente su prioridad mediante un número entero, y cuanto más grande es el número mayor es la prioridad de la regla. Si no existe una definición explícita de prioridad se considera que todas las reglas tienen la prioridad -1. Una cuestión importante a no perder de vista es que la priorización opera solo dentro de los límites de una fase de la gramática.

B.2 Lista de *stop words* en español

al
algo
algún
alguna
algunas
alguno
algunos
ambos
ampleamos
ante
antes
aquel
aquellas
aquellos
aqui
arriba
atras
bajo
bastante
bien
cada
cierta
ciertas
cierto
ciertos
como
con
conseguimos
conseguir
consigo
consigue
consiguen
consigues
contra
cual
cuando
de
del
dentro
desde
donde
dos
durante
e
el
él
ella
ellas
ellos
empleais
emplean
emplear
empleas
empleo
en
encima
entonces
entre
era
erais
eramos
éramos
eran
eras
eres
es
esa
esas
ese
eso
esos
esta
está
estaba
estabais
estábamos
estaban
estabas
estad
estada
estadas
estado
estados
estais
estáis
estamos
estan
están
estando
estar
estará
estarán
estarás
estaré
estaréis
estaremos
estaría
estaríais
estaríamos
estarían
estarías
estas
estás
este
esté
estéis
estemos
estén
estés
esto
estos
estoy
estuve
estuviera
estuvierais
estuviéramos
estuvieran
estuvieras
estuvieron
estuviese
estuvieseis
estuviésemos
estuviesen
estuvieses
estuvimos
estuviste
estuvisteis
estuvo
fin
fue
fuera
fuerais
fuéramos
fueran
fueras
fueron
fuese
fueseis
fuésemos
fuesen
fueses

fui
fuimos
fuiste
fuisteis
gueno
ha
habéis
había
habíais
habíamos
habían
habías
habida
habidas
habido
habidos
habiendo
habrá
habrán
habrás
habré
habréis
habremos
habría
habríais
habríamos
habrían
habrías
hace
haceis
hacemos
hacen
hacer
haces
hago
han
has
hasta
hay
haya
hayáis
hayamos
hayan
hayas
he
hemos
hube
hubiera
hubierais
hubiéramos
hubieran
hubieras
hubieron
hubiese
hubieseis
hubiésemos
hubiesen
hubieses
hubimos
hubiste
hubisteis
hubo
incluso
intenta
intentais
intentamos
intentan
intentar
intentas
intento
ir
la
largo
las
le
les
lo
los
más
me
mi
mí
mía
mías
mientras
mio
mío
míos
mis
modo
mucho
muchos
muy
nada
ni
no
nos
nosotras
nosotros
nuestra
nuestras
nuestro
nuestros
o
os
otra
otras
otro
otros
para
pero
poco
podeis
podemos
poder
podria
podriais
podriamos
podrian
podrias
por
porque
por qué
primero
puede
pueden
puedo
que
qué
quien
quienes
sabe
sabeis
sabemos
saben
saber
sabes
se
sea
seáis
seamos
sean
seas
sentid
sentida
sentidas
sentido
sentidos
ser
será

serán
serás
seré
seréis
seremos
sería
seríais
seríamos
serían
serías
si
sí
siendo
siente
sin
sintiendo
sobre
sois
solamente
solo
somos
son
soy
su
sus
suya
suyas
suyo
suyos
también
tanto
te
tendrá
tendrán
tendrás
tendré
tendréis
tendremos
tendría
tendríais
tendríamos
tendrían
tendrías
tened
teneis
tenéis
tenemos
tener
tenga
tengáis
tengamos
tengan
tengas
tengo
tenía
teníais
teníamos
tenían
tenías
tenida
tenidas
tenido
tenidos
teniendo
ti
tiempo
tiene
tienen
tienes
todo
todos
trabaja
trabajais
trabajamos
trabajan
trabajar
trabajas
trabajo
tras
tu
tú
tus
tuve
tuviera
tuvierais
tuviéramos
tuvieran
tuvieras
tuvieron
tuviese
tuvieseis
tuviésemos
tuviesen
tuvieses
tuvimos
tuviste
tuvisteis
tuvo
tuya
tuyas
tuyo
tuyos
ultimo
un
una
unas
uno
unos
usa
usais
usamos
usan
usar
usas
uso
va
vais
valor
vamos
van
vaya
verdad
verdadera
verdadero
vosostras
vosostros
vosotras
vosotros
voy
vuestra
vuestras
vuestro
vuestros
y
ya
yo

B.3 Gramática para Adaptar ANNIE Sentence Splitter

```
MultiPhase: main
Phases:

prepare
preProcesarTokens
modificarTokens
find1
find2
find3
find4
split
reestablecerTokens
postProcesarTokens
cleanup


                        ----------------------------------
Phase:      prepare
Input: Token
Options: control = once

//Makes sure there is no temporary data created by a previous run //of
the sentence splitter
Rule: cleanUp
{Token}
-->
{
  //if there were any sentences created, then we need to remove
  //the document  feature -> useful for future runs
  doc.getFeatures().remove("temp-last-sentence-end");
}


                        ----------------------------------
Phase: preProcesarTokens
Input: Token SpaceToken
Options: control = appelt debug = true

Rule:ignorarSignosDentroPreg
(({Token.string ==~"\\u00BF"}
({Token.string==~"[\\S&&[^?]](\\S)*"}|{SpaceToken})*)
({Token.string =="!"}|{Token.string ==".")})
(({Token.string==~"[\\S&&[^?]](\\S)*"}|{SpaceToken})*
{Token.string == "?"}))
:signoOmitir
-->:signoOmitir.preguntaModif={}

Rule:ignorarSignosDentroExcl
(({Token.string ==~"\\u00A1"}
({Token.string==~"[\\S&&[^!]](\\S)*"}|{SpaceToken})*)
({Token.string =="?"}|{Token.string ==".")})
(({Token.string==~"[\\S&&[^!]](\\S)*"}|{SpaceToken})*
{Token.string == "!"}))
:signoOmitir
-->:signoOmitir.exclamacionModif={}


                        ----------------------------------
```

```
Phase: modificarTokens
Input: Token preguntaModif exclamacionModif
Options: control = appelt debug = true

Rule: modifTokensInterrogacion
({preguntaModif})
:sentenceModificar
-->
{
      AnnotationSet tempSentenceAs =    (gate.AnnotationSet)bindings.
      get("sentenceModificar");
      AnnotationSet tempSentenceTokensAs = inputAS.get("Token").
      getContained (tempSentenceAs.firstNode().getOffset()
      ,tempSentenceAs.lastNode().getOffset());
      for(Annotation tempSentenceTokenAnn : tempSentenceTokensAs)
      {
            String tokenString=(String)tempSentenceTokenAnn.
            getFeatures().get("string");
            if("!".equals(tokenString)||".".equals(tokenString))
           {
                  FeatureMap nuevasCaracteristicas = Factory.
                  newFeatureMap();
                  nuevasCaracteristicas= tempSentenceTokenAnn.
                  getFeatures();
                  nuevasCaracteristicas.put("auxString",
                  nuevasCaracteristicas.get("string"));
                  nuevasCaracteristicas.put("auxKind",
                  nuevasCaracteristicas.get("kind"));
                  nuevasCaracteristicas.put("kind", "123");
                  nuevasCaracteristicas.put("string", "xyz");
            }
      }
}


Rule: modifTokensExclamacion
({exclamacionModif})
:sentenceModificar
-->
{AnnotationSet tempSentenceAs = (gate.AnnotationSet) bindings.get
("sentenceModificar");
AnnotationSet tempSentenceTokensAs = inputAS.get("Token").
getContained(tempSentenceAs.firstNode().getOffset(),tempSentenceAs.lastNo
de().getOffset());
      for(Annotation tempSentenceTokenAnn : tempSentenceTokensAs)
      {
            String tokenString=(String) tempSentenceTokenAnn.
           getFeatures().get("string");
            if(".".equals(tokenString)||"?".equals(tokenString)){
               FeatureMap nuevasCaracteristicas =Factory.newFeatureMap();
                  nuevasCaracteristicas= tempSentenceTokenAnn.
                  getFeatures();
                  nuevasCaracteristicas.put("auxString",
                  nuevasCaracteristicas.get("string"));
                  nuevasCaracteristicas.put("auxKind",
                  nuevasCaracteristicas.get("kind"));
                  nuevasCaracteristicas.put("kind", "123");
                  nuevasCaracteristicas.put("string", "xyz");
            }
      }
}
```

```
                                   ----------------------------------
Phase:      find1
Input: Token SpaceToken Lookup DEFAULT_TOKEN
Options: control = appelt

Macro: FULLSTOP
({Token.string=="."})

Macro: PUNCT
( {Token.string == "!"} |  {Token.string == "?"})


Macro: NEWLINE
(
  {SpaceToken.string == "\n"} |
  {SpaceToken.string=="\n\r"} |
  ({SpaceToken.string=="\n"}{SpaceToken.string=="\r"}) |
  {SpaceToken.string=="\r\n"} |
  ({SpaceToken.string=="\r"}{SpaceToken.string=="\n"})
)

Macro: entreParentesis
(
  {Token.string =="("}
  ({Token.string != ")"}|{SpaceToken})+
  {Token.string==")"}
)

Macro: entreCorchetes
(
  {Token.string =="["}
  ({Token.string != "]"}|{SpaceToken})+
  {Token.string=="]"}
)

Macro: entreLlaves
(
{Token.string =="{"}({Token.string != "}"} |{SpaceToken})+
{Token.string=="}"}
)

Macro: PROTOCOLO
(
{Token.string=="fax"}|{Token.string=="file"}|{Token.string=="ftp"}|{Token
.string=="gopher"}|{Token.string=="http"}|{Token.string=="https"}|{Token.
string=="ldap"}| {Token.string=="mailto"}
|{Token.string=="news"}|{Token.string=="nfs"}|
{Token.string=="nntp"}|{Token.string=="sip"}|
{Token.string=="sips"}|{Token.string=="tel"}|{Token.string=="telnet"}|{To
ken.string=="urn"}|{Token.string=="wais"}|
{Token.string=="xmpp"}
)

Macro:caracteresdeURI
(
{Token.kind==word}|{Token.kind==number}|{Token.string=="."}|
{Token.string=="/"}|{Token.string=="_"}| {Token.string=="-"}|
{Token.string=="%"}|{Token.string=="="}|{Token.string=="&"}|
{Token.string=="@"}
)
```

```
Rule:DotNetStop
 FULLSTOP
 (
   {Token.string == "NET"} |
   {Token.string == "net"} |
   {Token.string == "Net"}
 )
:puntoNet
--> {}

Rule:DotFileName
( FULLSTOP ({Token.orth == "lowercase"} |
{Token.orth == "allCaps"} | {Token.orth == "mixedCaps"}))
:extensionArchivo
--> {}

Rule:PregExclamPtoSuspensivoNoCierraEnunciado
(((PUNCT)[1,5]({SpaceToken})*)(FULLSTOP)[2,3])
:signosIgnorar
(({SpaceToken})*({Token.string==~"[a-z](\\S)*"}|
{Token.string==","}|{Token.string==";"}| {Token.string==":"}))
--> {}

Rule:PregExclamPtoSuspensivoCierraEnunciado
((PUNCT)[1,5]({SpaceToken})*)
((FULLSTOP)[2,3])
:coincidencia
(({SpaceToken})* {Token.string==~"[A-Z](\\S)*"})
-->:coincidencia.Split={kind="internal"}
Rule:ignorarPuntoSuspensivos
(FULLSTOP)[2,3]
:ignorarPuntos
(
 ({SpaceToken.string !="\n"})*
 ({Token.string==~"[a-z](\\S)*"}| {Token.string==","}|
{Token.string==";"}| {Token.string==":"})
)
-->{}

Rule:Number
({Token.kind == number}(FULLSTOP {Token.kind == number})+)
:numeroDecimalIP
--> {}

Rule: knownAbbrev
({SpaceToken})[0,1]
(
  {Lookup.majorType == "splitter_abbreviation"}
  {Token.string == "."}
)
:abreviatura
--> {}

Rule: upperCaseAbbrev
({Token.orth=="upperInitial", Token.length=="1"} FULLSTOP)+
:abreviatura
--> {}
```

```
Rule: AbreviaturaMinuscula
({Token.orth=="lowercase", Token.length=="1"} FULLSTOP)+
:abreviatura
--> {}

Rule: abreviaturaMinMay
(({Token.orth=="lowercase", Token.length=="1"} FULLSTOP)
({Token.orth=="upperInitial", Token.length=="1"} FULLSTOP)+)
:abreviatura
--> {}

Rule: abreviaturaPlural
({Token.string==~"([a-zA-Z])\\1"} FULLSTOP)
:abreviatura
-->{}

Rule:AbreviaturaAcentuada
(({Token.string==~"Acr\\u00F3n"}|
{Token.string==~"acr\\u00F3n"}|
{Token.string==~"Ac\\u00FAs"}|
{Token.string==~"ac\\u00FAs"}|
{Token.string==~"Adm\\u00F3n"}|
{Token.string==~"adm\\u00F3n"}|
{Token.string==~"\u00E1l"}|
{Token.string==~"C\\u00E9nt"}|
{Token.string==~"c\\u00E8nt"}|
{Token.string==~"C\\u00EDa"}|
{Token.string==~"c\\u00EDa"}|
{Token.string==~"C\\u00EDas"}|
{Token.string==~"c\\u00EDas"}|
{Token.string==~"Cl\\u00E1s"}|
{Token.string==~"cl\\u00E1s"}|
{Token.string==~"C\\u00F3d"}|
{Token.string==~"c\\u00F3d"}|
{Token.string==~"C\\u00F3ds"}|
{Token.string==~"c\\u00F3ds"}|
{Token.string==~"F\\u00EDs"}|
{Token.string==~"f\\u00EDs"}|
{Token.string==~"F\\u00F3rm"}|
{Token.string==~"f\\u00F3rm"}|
{Token.string==~"F\\u00F3rms"}|
{Token.string==~"f\\u00F3rms"}|
{Token.string==~"M\\u00E1x"}|
{Token.string==~"m\\u00E1x"}|
{Token.string==~"M\\u00E1xs"}|
{Token.string==~"m\\u00E1xs"}|
{Token.string==~"Mi\\u00E9r"}|
{Token.string==~"mi\\u00E9r"}|
{Token.string==~"M\\u00EDn"}|
{Token.string==~"m\\u00EDn"}|
{Token.string==~"M\\u00EDns"}|
{Token.string==~"m\\u00EDns"}|
{Token.string==~"N\\u00E1ut"}|
{Token.string==~"n\\u00E1ut"}|
{Token.string==~"N\\u00FAm"}|
{Token.string==~"n\\u00FAm"}|
{Token.string==~"N\\u00FAms"}|
{Token.string==~"n\\u00FAms"}|
{Token.string==~"\\u00F3p.Cit"}|
{Token.string==~"\\u00F3p.cit"}|
{Token.string==~"P\\u00E1g"}|
```

```
{Token.string==~"p\\u00E1g"}|
{Token.string==~"P\\u00E1gs"}|
{Token.string==~"p\\u00E1gs"}|
{Token.string==~"P\\u00E1rr"}|
{Token.string==~"p\\u00E1rr"}|
{Token.string==~"P\\u00E1rrs"}|
{Token.string==~"p\\u00E1rrs"}|
{Token.string==~"Pr\\u00F3l"}|
{Token.string==~"pr\\u00F3l"}|
{Token.string==~"Pr\\u00F3ls"}|
{Token.string==~"pr\\u00F3ls"}|
{Token.string==~"Qu\\u00EDm"}|
{Token.string==~"qu\\u00EDm"}|
{Token.string==~"Qu\\u00EDms"}|
{Token.string==~"qu\\u00EDms"}|
{Token.string==~"s\\u00E1b"}|
{Token.string==~"S\\u00E1"}|
{Token.string==~"S\\u00c1"}|
{Token.string==~"s\\u00E1b"}|
{Token.string==~"S\\u00E1b"}|
{Token.string==~"S\\u00c1B"}|
{Token.string==~"s\\u00E1bs"}|
{Token.string==~"S\\u00E1bs"}|
{Token.string==~"S\\u00c1BS"}|
{Token.string==~"S\\u00EDmb"}|
{Token.string==~"s\\u00EDmb"}|
{Token.string==~"S\\u00EDmbs"}|
{Token.string==~"s\\u00EDmbs"}|
{Token.string==~"Tel\\u00E9f"}|
{Token.string==~"tel\\u00E9f"}|
{Token.string==~"Tel\\u00E9fs"}|
{Token.string==~"tel\\u00E9fs"}|
{Token.string==~"T\\u00EDt"}|
{Token.string==~"t\\u00EDt"}|
{Token.string==~"T\\u00EDts"}|
{Token.string==~"t\\u00EDts"})
FULLSTOP)
:abreviatura
--> {}

 (({Token.string==~"D\\u00f1a"}|
{Token.string==~"d\\u00f1a"}|
{Token.string==~"D\\u00f1as"}|
{Token.string==~"d\\u00f1as"}) FULLSTOP)
:abreviatura
--> {}

Rule:abreviaturaEtcetera
({Token.string==~"etc"} FULLSTOP)
:abreviatura
(({SpaceToken})* {Token.string!=~"[A-Z0-9](\\S)*"})
--> {}

Rule:direccionEmail
(({Token.kind==word}|{Token.kind==number}| {Token.string=="."}|
{Token.string=="_"}| {Token.string=="%"}| {Token.string=="+"}
|{Token.string=="-"})+ {Token.string=="@"}
(({Token.kind==word}|{Token.kind==number}|{Token.string=="-"})+
({Token.string=="."}))+
({Token.kind==word}))
:validedEmail--> {}
```

```
Rule:direccionURI
(((((PROTOCOLO{Token.string==":"}{Token.string=="/"}
{Token.string=="/"})|(PROTOCOLO{Token.string==":"})|( {Token.string=="/"}
{Token.string=="/"})) |({Token.string=="/"}
))(((caracteresdeURI)+{Token.string=="?"}(caracteresdeURI|
{Token.string=="+"})+{Token.string=="#"}((caracteresdeURI|
{Token.string=="+"})*({Token.kind==word}|{Token.kind==number}|
{Token.string=="/"}|{Token.string=="_"}| {Token.string=="-"}|
{Token.string=="%"}|{Token.string=="="}|{Token.string=="&"}
|{Token.string=="@"}))))|((caracteresdeURI)+{Token.string=="?"}
(caracteresdeURI|{Token.string=="+"})*({Token.kind==word}|
{Token.kind==number}|{Token.string=="/"}|{Token.string=="_"}|
{Token.string=="-"}| {Token.string=="%"}|{Token.string=="="}
|{Token.string=="&"}|{Token.string=="@"}))|((caracteresdeURI)+
{Token.string=="#"}(caracteresdeURI|{Token.string=="+"})*
({Token.kind==word}|{Token.kind==number}|{Token.string=="/"}|
{Token.string=="_"}| {Token.string=="-"}| {Token.string=="%"}|
{Token.string=="="}|{Token.string=="&"}|{Token.string=="@"})
)|((caracteresdeURI)*({Token.kind==word}|{Token.kind==number}|
{Token.string=="/"}|{Token.string=="_"}| {Token.string=="-"}|
{Token.string=="%"}|{Token.string=="="}|{Token.string=="&"}|
{Token.string=="@"})) |({Token.string=="?"} (caracteresdeURI|
{Token.string=="+"})+{Token.string=="#"} (caracteresdeURI|
{Token.string=="+"})* ({Token.kind==word}|{Token.kind==number}|
{Token.string=="/"}|{Token.string=="_"}| {Token.string=="-"}|
{Token.string=="%"}|{Token.string=="="}|{Token.string=="&"}|
{Token.string=="@"}))|({Token.string=="?"}(caracteresdeURI|
{Token.string=="+"})*({Token.kind==word}|{Token.kind==number}|
{Token.string=="/"}|{Token.string=="_"}| {Token.string=="-"}|
{Token.string=="%"}|{Token.string=="="}|{Token.string=="&"}|
{Token.string=="@"})) | ({Token.string=="#"}(caracteresdeURI|
{Token.string=="+"})*({Token.kind==word}|{Token.kind==number}|
{Token.string=="/"}|{Token.string=="_"}| {Token.string=="-"}|
{Token.string=="%"}| {Token.string=="="}|{Token.string=="&"}|
{Token.string=="@"})))))|(({Token.string==~"(www|WWW)"}
{Token.string=="."} (caracteresdeURI)+ {Token.string=="?"}
(caracteresdeURI|{Token.string=="+"})+{Token.string=="#"}
(caracteresdeURI|{Token.string=="+"})*({Token.kind==word}|
{Token.kind==number}|{Token.string=="/"}| {Token.string=="_"}|
{Token.string=="-"}| {Token.string=="%"}|{Token.string=="="}|
{Token.string=="&"}|{Token.string=="@"})) |
({Token.string==~"(www|WWW)"}{Token.string=="."}
(caracteresdeURI)+{Token.string=="?"}(caracteresdeURI|
{Token.string=="+"})*({Token.kind==word}|{Token.kind==number}|{Token.stri
ng=="/"}|{Token.string=="_"}| {Token.string=="-"}|
{Token.string=="%"}|{Token.string=="="}|{Token.string=="&"}|
{Token.string=="@"}))|( {Token.string==~"(www|WWW)"} {Token.string=="."}
(caracteresdeURI)+ {Token.string=="#"}
(caracteresdeURI|{Token.string=="+"})*({Token.kind==word}|
{Token.kind==number}|{Token.string=="/"}
|{Token.string=="_"}| {Token.string=="-"}|{Token.string=="%"}
|{Token.string=="="}|{Token.string=="&"}|{Token.string=="@"})) |(
{Token.string==~"(www|WWW)"}{Token.string=="."}
(caracteresdeURI)*({Token.kind==word}|{Token.kind==number}|
{Token.string=="/"}|{Token.string=="_"}| {Token.string=="-
"}|{Token.string=="%"}|{Token.string=="="}|{Token.string=="&"}|
{Token.string=="@"})))))
:uriValida
--> {}
```

```
Rule:identificadorJava
((({Token.string==~"([a-zA-z])+"}|{Token.string=="$"}|
{Token.string=="_"}) ({Token.kind==number})*)+(FULLSTOP
(({Token.string==~"([a-zA-z])+"}|{Token.string=="$"}|
{Token.string=="_"}) ({Token.kind==number})*)+)+)
:campoPOO
--> {}



Rule: Ldots
(
  FULLSTOP
  FULLSTOP
  FULLSTOP
  (FULLSTOP)+
)
:variosPtos
-->{}

Rule:Lpunct
(PUNCT
PUNCT
PUNCT
PUNCT
PUNCT
(PUNCT)+)
:masDeSeisSignos
-->{}

Rule: Split1
( (PUNCT)[1,5]|(FULLSTOP) [1,3])
:split
-->
:split.Split = {kind = "internal"}

Rule: CR
(
  NEWLINE
  ({SpaceToken.kind == space})*
  NEWLINE
  ({SpaceToken.kind == space})*
):cr
-->
:cr.Split = {kind = "external"}

Rule:puntosEnEnumeracion
({SpaceToken.string=="\n"})
(({Token.string==~"(\\w){1,3}"} FULLSTOP)+)
:ptoEnum
-->{}

Rule:preguntasexclamacionesConsec
((PUNCT)[1,5] ({SpaceToken})* ({Token.string=="."}|{Token.string==";"}))
:variasPregExcl
-->{}
```

```
Rule:preguntaexclamacionConCont
(PUNCT)[1,5]
:continuacionPregExcl
(({SpaceToken.string!="\n"})* ({Token.string==~"[a-z]
(\\S)*"}|{Token.string=="-"} ({SpaceToken.string!="\n"})*
{Token.string==~"[a-z](\\S)*"}))
-->{}

Rule:dosPuntosAntesListado
({Token.string==":"})
:dosPuntos
( ({SpaceToken.string!="\n"})*
(({SpaceToken.string=="\n"})|({SpaceToken.string=="\r"})|({SpaceToken.str
ing=="\n"} {SpaceToken.string=="\r"})|
({SpaceToken.string=="r"} {SpaceToken.string=="\n"})))
-->
:dosPuntos.Split={kind="internal"}

Rule:parentesisAclaratorio
({Token.string =="("})
({Token.string != ")" }|{SpaceToken}|entreParentesis)+
:signoParent
({Token.string==")"})
-->{}

Rule:corcheteAclaratorio
({Token.string =="["})
({Token.string != "]" }|{SpaceToken}|entreCorchetes)+
:signocorch
({Token.string=="]"})
-->{}

Rule:signoLlaves
({Token.string =="{"})
({Token.string != "}" }|{SpaceToken}|entreLlaves)+
:sigLlave
({Token.string=="}"})
-->{}


Rule:signoComillas
(
({Token.string==~"\\u201C"}
 ({Token.string!=~"\\u201D"}|{SpaceToken})+
 {Token.string==~"\\u201D"})|
({Token.string==~"\\u0022"}
 ({Token.string!=~"\\u0022"}|{SpaceToken})+
 {Token.string==~"\\u0022"})|
({Token.string==~"\\u2018"}
 ({Token.string!=~"\\u2019"}|{SpaceToken})+
 {Token.string==~"\\u2019"})|
({Token.string==~"\\u00AB"}
 ({Token.string!=~"\\u00BB"}|{SpaceToken})+
 {Token.string==~"\\u00BB"})
):textoComillas
-->{}
```

```
                                    ----------------------------------
Phase: find2
Input: Token SpaceToken Split abrevMinuscula abrevMayuscula
Options: control = appelt debug = true

Macro: PUNCT
( {Token.string == "!"} | {Token.string == "?"})

Rule:puntosSuspensivosSucedenaEtc
({Token.string=="etc"} {Token.string=="."}({SpaceToken})*
{Token.string=="."})
({Split@string=="..."})
:contenido
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("contenido");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}

Rule:reconocerDivisorconetc
({Token.string=="etc"} {Token.string=="."}({SpaceToken})*)
(PUNCT)
:reconocerDivisorconetc
(({SpaceToken})* {Token.string==~"[A-Z](\\S)*"})
-->
:reconocerDivisorconetc.Split={kind="internal"}


Rule:preguntaExclamacionSucedenaEtc
({Token.string=="etc"} {Token.string=="."} ({SpaceToken})* PUNCT)
({Split@string==~"(\\?|\\!){5}"})
:contenido
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("contenido");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}

Rule:etcSeguidodedosPuntos
({Token.string=="etc"} {Token.string=="."} ({SpaceToken})*
{Token.string=="."})
({Split@string=="."})
:anotEliminar
( ({SpaceToken.string !="\n"})* ({Token.string==~"[a-z](\\S)*"}|
{Token.string==","}|{Token.string==";"}|   {Token.string==":"})
)
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("anotEliminar");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}
```

```
Rule:ignorarPtosSuspensivosInterrogacionExclamacion
({Split@string==~"\\.{2,3}"})
:contenido
(({SpaceToken})* (PUNCT))
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("contenido");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}


Rule:abreviaturaMinusculaMayusculaSeguidaPto
({abrevMinuscula@length==2}|{abrevMayuscula@length==2})
({Split@string=="."})
:contenido
(({SpaceToken})*{Token.string==~"[a-z](\\S)*"})
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("contenido");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}

Rule:ptoDeEnumeracionNoDivisor
(({SpaceToken.string=="\n"})({Token.string==~"(\\w){1,3}"}
({Token.string=="."} {Token.string==~"(\\w){1,3}"})*))
({Split@string=="."})
:identificado
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("identificado");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}

Rule:parentesisEnEnumeracion
(({SpaceToken.string=="\n"}|{SpaceToken.string=="\\A"})
(({Token.string==~"(\\w){1,3}"}
FULLSTOP)+|{Token.string==~"(\\w){1,3}"}(FULLSTOP
{Token.string==~"(\\w){1,3}"})*))
({Token.string==")"})
:parentEnumeracion
(({SpaceToken})*
({Token.string==~"\\u00a1"}|{Token.string==~"\\u00BF"})[1,5]
({SpaceToken})* {Token.string==~"[A-Z](\\S)*"})
-->{}

Rule:preguntaexclamacionNoSegundaParte
({Token.string
=~"[\\S&&[^.?!\\u00a1\\u00BF\\u2022\\u002A\\u2013\\uF0D8\\uF0A7\\uF02A\\u
F0FC\\u006F\\uF076\\uF02D\\u25A1\\u25AB\\u002D\\u25AA]]"})
:anotModificar
(({SpaceToken})* ({Token.string==~"\\u00BF"}|

{Token.string==~"\\u00a1"})[1,5] ({SpaceToken})* {Token.string==~"[A-
Z](\\S)*"})
-->:anotModificar.Split={kind="internal"}
```

```
Rule:eliminaranotacionExclIntePtosSuspensivos
({Split@string==~"(\\?|\\!){1,5}"})
:anotacionEliminar
(({SpaceToken})*({Split@string==".."}|{Split@string=="..."}))
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("anotacionEliminar");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}
```

```
Phase:      find3
Input: Token SpaceToken
Options: control = appelt

Rule:ptosSuspensivosdosPuntosAntesListado

({Token.string!="."}({Token.string=="."})[2,3] ({SpaceToken})*)
({Token.string==":"})
:dosPuntos
(({SpaceToken.string!="\n"})*(({SpaceToken.string=="\n"})|
({SpaceToken.string=="\r"})|({SpaceToken.string=="\n"}
{SpaceToken.string=="\r"})|
({SpaceToken.string=="r"} {SpaceToken.string=="\n"})))
-->
:dosPuntos.Split={kind=internal}
```

```
Phase: find4
Input: Token SpaceToken Split
Options: control = appelt debug = true


Rule: omitirAnotSplitdePtoEnumeracion
({Token}({SpaceToken})* {Token.kind=="number"})
({Token.string=="."})
:anoteEliminar
-->
{}

Rule: eliminarAnotSplitdePtoEnumeracion
{Token.kind=="number"}
({Split@string=="."})
:anoteEliminar
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("anoteEliminar");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}
```

```
Phase:split
Input: Split TempNoSplitText
Options: control = first

//sentence that consumes a split
Rule: internalSplits
({Split.kind == "internal"}):isplit
-->
{
  Long endOffset = ((AnnotationSet)bindings.get("isplit")).
      lastNode().getOffset();
  //find the end offset of previous sentences
  Long lastOffset = (Long)doc.getFeatures().get("temp-last-sentence-
end");
  if(lastOffset == null) lastOffset = new Long(0);
//
//  AnnotationSet sentences = outputAS.get("Sentence");
//  Long lastOffset = sentences == null || sentences.isEmpty() ?
//          new Long(0) :
//          sentences.lastNode().getOffset();
  //get the start offset of the first token.kind==word
  AnnotationSet tokens = inputAS.getContained(lastOffset, endOffset);
  if(tokens != null) tokens = tokens.get("Token");
  if(tokens != null && tokens.size() > 0){
    List<Annotation> tokList = new ArrayList<Annotation>(tokens);
    Collections.sort(tokList, new OffsetComparator());
    for(Annotation token : tokList){
      //String tokenKind = (String)token.getFeatures().get("kind");
      //if("word".equals(tokenKind)){
        Long startOffset = token.getStartNode().getOffset();
        if(startOffset.compareTo(endOffset) < 0){
          //create the new sentence
          try{
            outputAS.add(startOffset, endOffset, "Sentence",
                    Factory.newFeatureMap());
            //save the new end offset
            doc.getFeatures().put("temp-last-sentence-end", endOffset);
          }catch( InvalidOffsetException ioe){
            throw new GateRuntimeException(ioe);
          }
        }
        return;
     // }
    }
  }
}
```

```
//sentence that doesn't consume a split
Rule: externalSplits
({Split.kind == "external"}):esplit
-->
{
  Long endOffset = ((AnnotationSet)bindings.get("esplit")).
      firstNode().getOffset();
//  //get the end offset of the previous sentence
//  AnnotationSet sentences = outputAS.get("Sentence");
//  Long lastOffset = sentences == null || sentences.isEmpty() ?
//          new Long(0) :
//          sentences.lastNode().getOffset();
  //find the end offset of previous sentences
  Long lastOffset = (Long)doc.getFeatures().get("temp-last-sentence-
end");
  if(lastOffset == null) lastOffset = new Long(0);

  //get the start offset of the first token.kind==word
  AnnotationSet tokens = inputAS.getContained(lastOffset, endOffset);
  if(tokens != null) tokens = tokens.get("Token");
  if(tokens != null && tokens.size() > 0){
    //we have a more precise end offset
    endOffset = tokens.lastNode().getOffset();
    List<Annotation> tokList = new ArrayList<Annotation>(tokens);
    Collections.sort(tokList, new OffsetComparator());
    for(Annotation token : tokList){
      String tokenKind = (String)token.getFeatures().get("kind");
      if("word".equals(tokenKind)){
        Long startOffset = token.getStartNode().getOffset();
        if(startOffset.compareTo(endOffset) < 0){
          //create the new sentence
          try{
            outputAS.add(startOffset, endOffset, "Sentence",
                    Factory.newFeatureMap());
            doc.getFeatures().put("temp-last-sentence-end", endOffset);
          }catch( InvalidOffsetException ioe){
            throw new GateRuntimeException(ioe);
          }
        }
        return;
      }
    }
  }
}
```

```
Phase: reestablecerTokens
Input: Token
Options: control = appelt debug = true

Rule: reestblecerString
(
      {Token.kind=="123"}
)
:anotModificar
-->
{
      AnnotationSet tokenModificar=bindings.get("anotModificar");
      Annotation aux=tokenModificar.iterator().next();
      FeatureMap nuevasCaracteristicas = Factory.newFeatureMap();
      nuevasCaracteristicas=aux.getFeatures();
      nuevasCaracteristicas.put("string", nuevasCaracteristicas.get
      ("auxString"));
      nuevasCaracteristicas.put("kind", nuevasCaracteristicas.get
      ("auxKind"));
      nuevasCaracteristicas.put("tuneado", "YES");
      nuevasCaracteristicas.remove("auxString");
      nuevasCaracteristicas.remove("auxKind");
}

                        ---------------------------------

Phase: postProcesarTokens
Input: preguntaModif exclamacionModif
Options: control = appelt debug = true

Rule:removerAnotTemporal
({preguntaModif}|{exclamacionModif})
:deletetempAnn
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("deletetempAnn");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}

                        ---------------------------------
Phase:      cleanup
Input: Token
Options: control = once

//Removes temporary data created by the sentence splitter
Rule: cleanUp
{Token}
-->
{
  //if there were any sentences created, then we need to remove the
//documentfeature -> useful for future runs
 // doc.getFeatures().remove("temp-last-sentence-end");
  //remove all lookups used for abbreviations
  FeatureMap constraints = Factory.newFeatureMap();
  constraints.put("majorType", "splitter_abbreviation");
  AnnotationSet toRemove = inputAS.get("Lookup", constraints);
  if(toRemove != null) inputAS.removeAll(toRemove);
}
```

B.4 Gramática para Implementar la Eliminación de *Stop Words*

```
MultiPhase: main
Phases:

suprimirStopWords


Phase:      suprimirStopWords
Input: Token SpaceToken Lookup
Options: control = appelt

                        ---------------------------------

Rule: stopWordsSinAcento
( {Token, Lookup.majorType =="suprimir_stopw"})
:stopWord
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("stopWord");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}


Rule: stopWordsAcentuadas
(
      {Token.string==~"alg\\u00FAn"}|
      {Token.string==~"Alg\\u00FAn"}|
      {Token.string==~"ALG\\u00DAN"}|
      {Token.string==~"\\u00E9l"}|
      {Token.string==~"\\u00E9l"}|
      {Token.string==~"\\u00C9L"}|
      {Token.string==~"\\u00E9ramos"}|
      {Token.string==~"\\u00C9ramos"}|
      {Token.string==~"\\u00C9RAMOS"}|
      {Token.string==~"est\\u00E1"}|
      {Token.string==~"Est\\u00E1"}|
      {Token.string==~"EST\\u00C1"}|
      {Token.string==~"est\\u00E1bamos"}|
      {Token.string==~"Est\\u00E1bamos"}|
      {Token.string==~"EST\\u00C1BAMOS"}|
      {Token.string==~"est\\u00E1is"}|
      {Token.string==~"Est\\u00E1is"}|
      {Token.string==~"EST\\u00C1IS"}|
      {Token.string==~"est\\u00E1n"}|
      {Token.string==~"Est\\u00E1n"}|
      {Token.string==~"EST\\u00C1N"}|
      {Token.string==~"estar\\u00E1"}|
      {Token.string==~"Estar\\u00E1"}|
      {Token.string==~"ESTAR\\u00C1"}|
      {Token.string==~"estar\\u00E1n"}|
      {Token.string==~"Estar\\u00E1n"}|
      {Token.string==~"ESTAR\\u00C1N"}|
      {Token.string==~"estar\\u00E1s"}|
      {Token.string==~"Estar\\u00E1s"}|
```

```
{Token.string==~"ESTAR\\u00C1S"}|
{Token.string==~"estar\\u00E9"}|
{Token.string==~"Estar\\u00E9"}|
{Token.string==~"estar\\u00C9"}|
{Token.string==~"estar\\u00E9is"}|
{Token.string==~"Estar\\u00E9is"}|
{Token.string==~"ESTAR\\u00C9IS"}|
{Token.string==~"estar\\u00EDa"}|
{Token.string==~"Estar\\u00EDa"}|
{Token.string==~"ESTAR\\u00CDA"}|
{Token.string==~"estar\\u00EDais"}|
{Token.string==~"Estar\\u00EDais"}|
{Token.string==~"ESTAR\\u00CDAIS"}|
{Token.string==~"estar\\u00EDamos"}|
{Token.string==~"Estar\\u00EDamos"}|
{Token.string==~"ESTAR\\u00CDAMOS"}|
{Token.string==~"estar\\u00EDan"}|
{Token.string==~"estar\\u00EDan"}|
{Token.string==~"ESTAR\\u00CDAN"}|
{Token.string==~"estar\\u00EDas"}|
{Token.string==~"Estar\\u00EDas"}|
{Token.string==~"ESTAR\\u00CDAS"}|
{Token.string==~"est\\u00E1s"}|
{Token.string==~"Est\\u00E1s"}|
{Token.string==~"EST\\u00C1S"}|
{Token.string==~"est\\u00E9"}|
{Token.string==~"Est\\u00E9"}|
{Token.string==~"EST\\u00C9"}|
{Token.string==~"est\\u00E9is"}|
{Token.string==~"Est\\u00E9is"}|
{Token.string==~"EST\\u00C9IS"}|
{Token.string==~"est\\u00E9n"}|
{Token.string==~"Est\\u00E9n"}|
{Token.string==~"EST\\u00C9N"}|
{Token.string==~"est\\u00E9s"}|
{Token.string==~"Est\\u00E9s"}|
{Token.string==~"EST\\u00C9S"}|
{Token.string==~"estuvi\\u00E9ramos"}|
{Token.string==~"Estuvi\\u00E9ramos"}|
{Token.string==~"ESTUVI\\u00C9RAMOS"}|
{Token.string==~"estuvi\\u00E9semos"}|
{Token.string==~"Estuvi\\u00E9semos"}|
{Token.string==~"ESTUVI\\u00C9SEMOS"}|
{Token.string==~"fu\\u00E9ramos"}|
{Token.string==~"Fu\\u00E9ramos"}|
{Token.string==~"FU\\u00C9RAMOS"}|
{Token.string==~"fu\\u00E9semos"}|
{Token.string==~"Fu\\u00E9semos"}|
{Token.string==~"FU\\u00C9SEMOS"}|
{Token.string==~"hab\\u00E9is"}|
{Token.string==~"Hab\\u00E9is"}|
{Token.string==~"HAB\\u00C9IS"}|
{Token.string==~"hab\\u00EDa"}|
{Token.string==~"Hab\\u00EDa"}|
{Token.string==~"HAB\\u00CDA"}|
{Token.string==~"hab\\u00EDais"}|
{Token.string==~"Hab\\u00EDais"}|
```

```
{Token.string==~"HAB\\u00CDAIS"}|
{Token.string==~"hab\\u00EDamos"}|
{Token.string==~"Hab\\u00EDamos"}|
{Token.string==~"HAB\\u00CDAMOS"}|
{Token.string==~"hab\\u00EDan"}|
{Token.string==~"Hab\\u00EDan"}|
{Token.string==~"HAB\\u00CDAN"}|
{Token.string==~"hab\\u00EDas"}|
{Token.string==~"Hab\\u00EDas"}|
{Token.string==~"HAB\\u00CDAS"}|
{Token.string==~"habr\\u00E1"}|
{Token.string==~"Habr\\u00E1"}|
{Token.string==~"HABR\\u00C1"}|
{Token.string==~"habr\\u00E1n"}|
{Token.string==~"Habr\\u00E1n"}|
{Token.string==~"HABR\\u00C1N"}|
{Token.string==~"habr\\u00E1s"}|
{Token.string==~"Habr\\u00E1s"}|
{Token.string==~"HABR\\u00C1S"}|
{Token.string==~"habr\\u00E9"}|
{Token.string==~"Habr\\u00E9"}|
{Token.string==~"HABR\\u00C9"}|
{Token.string==~"habr\\u00E9is"}|
{Token.string==~"Habr\\u00E9is"}|
{Token.string==~"HABR\\u00C9IS"}|
{Token.string==~"habr\\u00EDa"}|
{Token.string==~"Habr\\u00EDa"}|
{Token.string==~"HABR\\u00CDA"}|
{Token.string==~"habr\\u00EDais"}|
{Token.string==~"Habr\\u00EDais"}|
{Token.string==~"HABR\\u00CDAIS"}|
{Token.string==~"habr\\u00EDamos"}|
{Token.string==~"Habr\\u00EDamos"}|
{Token.string==~"HABR\\u00CDAMOS"}|
{Token.string==~"habr\\u00EDan"}|
{Token.string==~"Habr\\u00EDan"}|
{Token.string==~"HABR\\u00CDAN"}|
{Token.string==~"habr\\u00EDas"}|
{Token.string==~"Habr\\u00EDas"}|
{Token.string==~"HABR\\u00CDAS"}|
{Token.string==~"hay\\u00E1is"}|
{Token.string==~"Hay\\u00E1is"}|
{Token.string==~"HAY\\u00C1IS"}|
{Token.string==~"hubi\\u00E9ramos"}|
{Token.string==~"Habi\\u00E9ramos"}|
{Token.string==~"HABI\\u00C9RAMOS"}|
{Token.string==~"hubi\\u00E9semos"}|
{Token.string==~"Habi\\u00E9semos"}|
{Token.string==~"HABI\\u00C9SEMOS"}|
{Token.string==~"m\\u00E1s"}|
{Token.string==~"M\\u00E1s"}|
{Token.string==~"M\\u00C1S"}|
{Token.string==~"m\\u00ED"}|
{Token.string==~"M\\u00ED"}|
{Token.string==~"M\\u00CD"}|
{Token.string==~"m\\u00EDa"}|
```

```
{Token.string==~"M\\u00EDa"}|
{Token.string==~"M\\u00CDA"}|
{Token.string==~"m\\u00EDas"}|
{Token.string==~"M\\u00EDas"}|
{Token.string==~"M\\u00CDAS"}|
{Token.string==~"m\\u00EDo"}|
{Token.string==~"M\\u00EDo"}|
{Token.string==~"M\\u00CDO"}|
{Token.string==~"m\\u00EDos"}|
{Token.string==~"M\\u00EDos"}|
{Token.string==~"M\\u00CDOS"}|
({Token.string=="por"} {SpaceToken} {Token.string==~"qu\\u00E9"})|
({Token.string=="Por"} {SpaceToken} {Token.string==~"qu\\u00E9"})|
({Token.string=="POR"} {SpaceToken} {Token.string==~"QU\\u00C9"})|
{Token.string==~"qu\\u00E9"}|
{Token.string==~"Qu\\u00E9"}|
{Token.string==~"QU\\u00C9"}|
{Token.string==~"se\\u00E1is"}|
{Token.string==~"Se\\u00E1is"}|
{Token.string==~"SE\\u00C1IS"}|
{Token.string==~"ser\\u00E1"}|
{Token.string==~"Ser\\u00E1"}|
{Token.string==~"SER\\u00C1"}|
{Token.string==~"ser\\u00E1n"}|
{Token.string==~"Ser\\u00E1n"}|
{Token.string==~"SER\\u00C1N"}|
{Token.string==~"ser\\u00E1s"}|
{Token.string==~"Ser\\u00E1s"}|
{Token.string==~"SER\\u00C1S"}|
{Token.string==~"ser\\u00E9"}|
{Token.string==~"Ser\\u00E9"}|
{Token.string==~"SER\\u00C9"}|
{Token.string==~"ser\\u00E9is"}|
{Token.string==~"Ser\\u00E9is"}|
{Token.string==~"SER\\u00C9IS"}|
{Token.string==~"ser\\u00EDa"}|
{Token.string==~"Ser\\u00EDa"}|
{Token.string==~"SER\\u00CDA"}|
{Token.string==~"ser\\u00EDais"}|
{Token.string==~"Ser\\u00EDais"}|
{Token.string==~"SER\\u00CDAIS"}|
{Token.string==~"ser\\u00EDamos"}|
{Token.string==~"Ser\\u00EDamos"}|
{Token.string==~"SER\\u00CDAMOS"}|
{Token.string==~"ser\\u00EDan"}|
{Token.string==~"Ser\\u00EDan"}|
{Token.string==~"SER\\u00CDAN"}|
{Token.string==~"ser\\u00EDas"}|
{Token.string==~"Ser\\u00EDas"}|
{Token.string==~"SER\\u00CDAS"}|
{Token.string==~"s\\u00ED"}|
{Token.string==~"S\\u00ED"}|
{Token.string==~"S\\u00CD"}|
{Token.string==~"tambi\\u00E9n"}|
{Token.string==~"Tambi\\u00E9n"}|
{Token.string==~"TAMBI\\u00C9N"}|
{Token.string==~"tendr\\u00E1"}|
```

```
{Token.string==~"Tendr\\u00E1"}|
{Token.string==~"TENDR\\u00C1"}|
{Token.string==~"tendr\\u00E1n"}|
{Token.string==~"Tendr\\u00E1n"}|
{Token.string==~"TENDR\\u00C1N"}|
{Token.string==~"tendr\\u00E1s"}|
{Token.string==~"Tendr\\u00E1s"}|
{Token.string==~"TENDR\\u00C1S"}|
{Token.string==~"tendr\\u00E9"}|
{Token.string==~"Tendr\\u00E9"}|
{Token.string==~"TENDR\\u00C9"}|
{Token.string==~"tendr\\u00E9is"}|
{Token.string==~"Tendr\\u00E9is"}|
{Token.string==~"TENDR\\u00C9IS"}|
{Token.string==~"tendr\\u00EDa"}|
{Token.string==~"Tendr\\u00EDa"}|
{Token.string==~"TENDR\\u00CDA"}|
{Token.string==~"tendr\\u00EDais"}|
{Token.string==~"Tendr\\u00EDais"}|
{Token.string==~"TENDR\\u00CDAIS"}|
{Token.string==~"tendr\\u00EDamos"}|
{Token.string==~"Tendr\\u00EDamos"}|
{Token.string==~"TENDR\\u00CDAMOS"}|
{Token.string==~"tendr\\u00EDan"}|
{Token.string==~"Tendr\\u00EDan"}|
{Token.string==~"TENDR\\u00CDAN"}|
{Token.string==~"tendr\\u00EDas"}|
{Token.string==~"Tendr\\u00EDas"}|
{Token.string==~"TENDR\\u00CDAS"}|
{Token.string==~"ten\\u00E9is"}|
{Token.string==~"Ten\\u00E9is"}|
{Token.string==~"TEN\\u00C9IS"}|
{Token.string==~"teng\\u00E1is"}|
{Token.string==~"Teng\\u00E1is"}|
{Token.string==~"TENG\\u00C1IS"}|
{Token.string==~"ten\\u00EDa"}|
{Token.string==~"Ten\\u00EDa"}|
{Token.string==~"TEN\\u00CDA"}|
{Token.string==~"ten\\u00EDais"}|
{Token.string==~"Ten\\u00EDais"}|
{Token.string==~"TEN\\u00CDAIS"}|
{Token.string==~"ten\\u00EDamos"}|
{Token.string==~"Ten\\u00EDamos"}|
{Token.string==~"TEN\\u00CDAMOS"}|
{Token.string==~"ten\\u00EDan"}|
{Token.string==~"Ten\\u00EDan"}|
{Token.string==~"TEN\\u00CDAN"}|
{Token.string==~"ten\\u00EDas"}|
{Token.string==~"Ten\\u00EDas"}|
{Token.string==~"TEN\\u00CDAS"}|
{Token.string==~"t\\u00FA"}|
{Token.string==~"T\\u00FA"}|
{Token.string==~"T\\u00DA"}|
{Token.string==~"tuvi\\u00E9ramos"}|
{Token.string==~"Tuvi\\u00E9ramos"}|
{Token.string==~"TUVI\\u00C9RAMOS"}|
```

```
      {Token.string==~"tuvi\\u00E9semos"}|
      {Token.string==~"Tuvi\\u00E9semos"}|
      {Token.string==~"TUVI\\u00C9SEMOS"}

)
:stopWord2
-->
{
      AnnotationSet deletetemporalAnn=bindings.get("stopWord2");
      Annotation auxAnn=deletetemporalAnn.iterator().next();
      outputAS.remove(auxAnn);
}
```

B.5 Gramática para Crear la Caracteristica *String* en Anotaciones *Sentence*

La regla es utilizada por *Generic Tagger* para crear una característica de anotación "string" en cada anotación del tipo "Sentence" identificada en un documento. El valor de la característica "string" es igual al contenido textual del fragmento de texto cubierto por la anotación del tipo "Sentece".

```
Phase: Sentence_String
Input: Sentence
Options: control=brill

Rule: sent
(
      {Sentence}
):s
-->
{
gate.AnnotationSet sAS = (gate.AnnotationSet)bindings.get("s");
gate.Annotation sentence = (gate.Annotation)sAS.iterator().next();
gate.FeatureMap sMap = sentence.getFeatures();
try {
    sMap.put( gate.taggerframework.GenericTagger.STRING_FEATURE_NAME,
doc.getContent().getContent(sentence.getStartNode().getOffset(),
sentence.getEndNode().getOffset()).toString());
    } catch (Exception ex) {
      ex.printStackTrace();
    }
}
```