

TESIS DOCTORAL

Recuperación Inteligente de
Contenidos Digitales Educativos



VNiVERSIDAD
D SALAMANCA

Departamento de Informática
y Automática

Facultad de Ciencias

AUTORA

Ana Belén Gil González

DIRECTORES

Dr. D. Juan Manuel Corchado Rodríguez

Dra. Dña. Sara Rodríguez González

DEPARTAMENTO DE INFORMÁTICA Y AUTOMÁTICA
FACULTAD DE CIENCIAS



**VNiVERSiDAD
D SALAMANCA**

TESIS DOCTORAL

RECUPERACIÓN INTELIGENTE DE CONTENIDOS DIGITALES
EDUCATIVOS

AUTORA

D^a. Ana Belén Gil González

DIRECTORES

Dr. D. Juan Manuel Corchado Rodríguez

Dra. Dña. Sara Rodríguez González

Noviembre de 2011

La memoria titulada “RECUPERACIÓN INTELIGENTE DE CONTENIDOS DIGITALES EDUCATIVOS” que presenta D^a. Ana Belén Gil González para optar al Grado de Doctor por la Universidad de Salamanca ha sido realizada bajo la dirección del profesor Dr. D. Juan Manuel Corchado Rodríguez, Profesor Catedrático del Departamento de Informática y Automática de la Universidad de Salamanca, y por la profesora Dra. Dña. Sara Rodríguez González, Profesora Ayudante Doctor del Departamento de Informática y Automática de la Universidad de Salamanca.

Salamanca, Noviembre de 2011

Los Directores

El Doctorando

Fdo.: Dr. D. Juan M. Corchado Rodríguez
Profesor Catedrático de Universidad
Informática y Automática
Universidad de Salamanca

Fdo. D^a. Ana Belén Gil González

Fdo.: Dra. Dña. Sara Rodríguez González
Profesora Ayudante Doctor
Informática y Automática
Universidad de Salamanca



Ana Belén Gil González: *Recuperación Inteligente de Contenidos Digitales Educativos*, PhD on Computers and Automation, © November 2011

Supervisors:

Dr. D. Juan Manuel Corchado Rodríguez
Dra. Dña. Sara Rodríguez González

Location: Salamanca

Si añades un poco a lo poco y lo haces así con frecuencia, pronto llegará a ser mucho.

Hesíodo (S. VIII AC-?) Poeta griego.

RESUMEN

A lo largo de la última década, los métodos para la gestión y organización de grandes volúmenes de contenidos digitales han sufrido una importante evolución. Internet como vehículo y contenedor de un gran número de dichos almacenes de datos contribuye al fenómeno de la dispersión de contenidos. La información, caracterizada de forma diversa está contenida en repositorios de datos especializados a los que es preciso acceder mediante métodos propios en cada uno, haciendo uso de inagotables mecanismos y protocolos de comunicación entre las partes implicadas en el proceso. Estas razones motivan la necesidad de investigar en técnicas, herramientas y metodologías que den una solución tecnológica hacia mecanismos eficaces de adaptación en la búsqueda y recuperación de contenidos digitales en un entorno caracterizado por (i) ser suma de contextos abiertos (ii) heterogéneos y (iii) altamente dinámicos.

El sector educativo es un significativo generador, depositario y consumidor de contenidos digitales educativos. Se trata por tanto de un sector relevante en el desarrollo e integración de soluciones emergentes para los procesos de gestión, localización y distribución de sus contenidos digitales. Los usuarios tienen acceso a tecnologías en continua evolución que les permiten obtener dichos contenidos educativos de una manera ubicua y en tiempo real. Esta apertura en la comunicación y la urgente necesidad de alcanzar una verdadera interoperabilidad entre los diferentes entornos de aplicaciones educativas hacen de este sector el escenario perfecto para investigar los procesos de búsqueda, recuperación e integración de contenidos digitales en repositorios digitales en entornos heterogéneos. Los problemas emergentes de la integración de contenidos digitales educativos son causados generalmente por las múltiples caracterizaciones de los contenidos, la gran cantidad de contenidos distribuidos y el acceso a ellos que requieren los distintos usuarios. Las diferentes aplicaciones de la Web manejan información distribuida haciendo que la integración de los datos y la seguridad de las aplicaciones sean aún más complicadas y, en muchas ocasiones, imposible de garantizar plenamente.

Finalmente, hay que añadir los problemas relativos a los canales de comunicación por donde viaja la información, como es el caso de Internet. En el campo de los sistemas multi-agente es conocida la capacidad de estos de

adaptarse de forma rápida y eficaz a los cambios que se producen en su entorno. Este trabajo propone un modelo para el desarrollo de sistemas de recuperación de contenidos digitales basado en el paradigma de las organizaciones virtuales de agentes. El modelo permite el desarrollo de una arquitectura abierta y flexible que soporta los servicios necesarios para llevar a cabo una la búsqueda de contenidos digitales distribuidos de manera dinámica. Basándose en el modelo propuesto, se desarrolla la arquitectura AIREH (*Architecture for Intelligent Recovery of Educational content in Heterogeneous environments*); una arquitectura multi-agente para la búsqueda e integración de contenidos educativos heterogéneos a través de un modelo de recuperación mediante búsquedas federadas. Este trabajo prueba la posibilidad de modelar un sistema eficiente para la gestión de sistemas abiertos a partir de un modelo de organizaciones adaptativo que facilita una planificación distribuida, flexible y dinámica. Tanto el modelo como las tecnologías presentadas en esta investigación, son un ejemplo del potencial para desarrollar sistemas de recuperación de contenidos digitales basados en el paradigma de las organizaciones virtuales de agentes. Las ventajas de la arquitectura propuesta son principalmente su flexibilidad, su personalización, la capacidad integradora de la solución y la eficiencia.

ABSTRACT

The last decade has seen a significant evolution in the methods for managing and organizing large volumes of digital content. The use of the Internet as a vehicle and container for such a large volume of data warehouses has led to a phenomenon of globalization and decentralization. The information is characterized in various ways, and contained in specialized data repositories each of which must be accessed by their own methods, resulting in dozens of communication protocols. Therefore, there is a need for research on the techniques, tools and methodologies that provide a technological solution that enables the adaptation of mechanisms in an environment that is (i) the sum of open contexts (ii) highly dynamic, and (iii) heterogeneous.

The education sector is a significant generator, consumer and depository for educational content. These aspects position the current educational environment in a relevant sector in the development and integration of emerging solutions to management processes for the location and distribution of digital content. Users have access to technologies that allow them to obtain information ubiquitously and in real-time. This openness in communication and the urgent need to achieve true interoperability between educational application environments have created the need for research in the search, retrieval and integration of heterogeneous educational content. The problems arising from the integration of educational content are usually caused by the multiple characterizations of the content, the vast amount of educational content distributed, and the access to them required by different users. In addition, new Web applications handle information distributed in different geographical areas making the integration of data and application security even more complicated and in many cases, impossible to fully guarantee. Finally, there is the additional problem of the communication channels through which information travels, such as Internet.

Multi-agent systems are known for their ability to adapt quickly and effectively to changes in their environment. This work proposes a model for the development of digital content retrieval based on the paradigm of virtual organizations of agents. The model allows the development of an open and flexible architecture that supports the services necessary to conduct a search for distributed digital content dynamically. The AIREH (Architecture for Intelligent Recovery of Educational content in Heterogeneous Environments) is based on the proposed model; it is a multi-agent architecture that can search and integrate heterogeneous educational content through a recovery model

that uses a federated search. This study demonstrate the possibility of modeling an efficient system for managing open systems from a model of an adaptive organization that provides distributed planning, and is flexible and dynamic . The model and the technologies presented in this research are an example of the potential for developing recovery systems for digital content based on the paradigm of virtual organizations of agents. The advantages of the proposed architecture are its flexibility, customization, integrative solution and efficiency.

Agradecimientos

Durante la realización de esta tesis he recibido ayuda y apoyo de muchas personas a las que quisiera expresar mi más sincero agradecimiento.

A mi familia, por su comprensión y su apoyo incondicional. Ellos han sido mi oasis y la motivación más importante para persistir y mejorar, caminando siempre hacia adelante.

Al grupo BISITE, por brindarme un punto de apoyo definitivo para seguir trabajando con ilusión, rigor, dedicación y con mucha calidad humana. Muy en especial a mis directores, Juan Manuel por su convicción y generosidad y Sara por guiarme sabiamente desde su mar de trabajo sin perder jamás ninguno la sonrisa. A Fernando, Beatriz y Deana por su inestimable ayuda y su mirada crítica. Gracias por vuestro tiempo y los muchos consejos en el desarrollo de este trabajo, sin vosotros no sería posible. A todos por hacerme partícipe de tantos momentos de duro trabajo y de tan buenos ratos compartidos, así es un placer trabajar.

A Zahia, Lilia y Angélica que me brindaron la oportunidad de trabajar igualmente en una gran familia en LIP6, estando fuera de casa.

A mis compañeros del Departamento de Informática y Automática, porque el día a día en un departamento en constante movimiento propicia el trabajo permanente que lleva al estar hoy aquí. Al club del café, por esos pequeños momentos diarios. Y muy especialmente a Ana, porque resulta estupendo contar siempre y entre tanto vaivén con su mirada serena, sincera y cuerda de cualquier asunto.

Gracias a todos.

ÍNDICE DE CONTENIDO

Resumen.....	v
Abstract.....	vii
Índice de Contenido.....	1
Índice de Figuras.....	7
Índice de Tablas.....	12
Lista de Acrónimos.....	14
Parte I. Introducción.....	17
1 Introducción.....	19
1.1 Descripción del problema y motivación.....	20
1.2 Hipótesis y Objetivos.....	23
1.3 Metodología.....	24
1.4 Estructura de la memoria.....	25
Parte II. Estado del Arte y Enfoques.....	29
2 Tecnologías y Estándares para Contenidos Digitales Educativos.....	31
2.1 Objetos de Aprendizaje.....	32
2.1.1 Concepto y Cometido.....	32
2.2 Estándares para el Mercado de los Objetos de Aprendizaje.....	34
2.3 Repositorios de Objetos de Aprendizaje.....	37
2.3.1 Funcionalidades de un ROA.....	38
2.3.2 Soluciones Tecnológicas de los ROA.....	39
2.3.3 Estándares para la Interoperabilidad de ROAs.....	40
2.4 Protocolos de Comunicación en la Interoperabilidad de Contenidos.....	43
2.5 Conclusiones del Capítulo.....	46

3	Recuperación Eficiente de Contenidos Educativos	49
3.1	Recuperación de Contenidos en Web.....	50
3.2	Medida de la efectividad de la recuperación.....	52
3.2.1	Medida del Rendimiento en la Recuperación de Contenidos Educativos	53
3.2.2	Evaluación del Rendimiento en Recuperación de Objetos de Aprendizaje.....	56
3.3	Sistemas de recomendación	60
3.3.1	Recomendación Basada en Contenido.....	62
3.3.2	Recomendación Basada en Filtrado colaborativo	63
3.4	Tecnologías de Soporte a los Sistemas de Recomendación.....	70
3.4.1	CBR aplicados a la Recomendación.....	70
3.4.2	Minería de Datos y Sistemas de Recomendación.....	75
3.5	Trabajos relacionados	78
3.5.1	Aproximaciones a la Recuperación Automática de Contenidos Educativos	78
3.5.2	Aproximaciones a la Recomendación de Objetos de Aprendizaje.....	80
3.6	Conclusiones Del Capítulo	84
4	Sistemas Multi-agente Basados en Organizaciones	85
4.1	El paradigma de Agente Inteligente	86
4.1.1	Clasificación de agentes	87
4.1.2	Arquitecturas	91
4.2	Agentes Deliberativos. El modelo BDI.....	93
4.3	Modelos de razonamiento en el agente	96
4.3.1	Agentes CBR- BDI, CBP-BDI.....	97
4.4	Sistemas Multi-agente	101

4.4.1	Interacción entre agentes en un SMA.....	102
4.5	Organizaciones de Agentes	103
4.5.1	Concepto de Organización Virtual.....	106
4.5.2	Propiedades de una organización virtual.....	107
4.5.3	Entornos de desarrollo de Organizaciones Virtuales de Agentes.....	113
4.5.4	Metodologías de desarrollo de Organizaciones Virtuales.....	114
4.6	Conclusiones del capítulo	116
Parte III. Propuesta		119
5	Modelo de Arquitectura Basada en Organizaciones Virtuales.....	121
5.1	Motivación.....	122
5.1.1	Contexto actual.....	122
5.1.2	Funcionalidad de los ROA para Búsqueda Federada	127
5.1.3	Búsqueda de la solución.....	131
5.2	Formalización de la Arquitectura Propuesta	133
5.2.1	Diseño del Modelo de Organización Virtual.....	135
5.2.2	Secuencia Funcional.....	156
5.2.3	Descripción del agente CBR-BDI para la recomendación	163
5.3	Conclusiones del Capítulo.....	171
Parte IV. Evaluación y Research overview.....		173
6	Resultados Experimentales y Evaluación.....	175
6.1	Resultados	176
6.1.1	Análisis del Caso de Estudio.....	178
6.1.2	Rendimiento para la Arquitectura de Recuperación.....	179

Índice

6.1.3	Técnicas de Clasificación Multietiquetado de Objetos de Aprendizaje.....	189
6.2	Generalización del Modelo Propuesto	193
6.3	Conclusiones.....	195
6.3.1	Contribuciones a la investigación.....	198
6.3.2	Trabajo Futuro	199
7	Research Overview.....	201
7.1	Introduction.....	201
7.1.1	Analysis of the problem context.....	203
7.1.2	Over the Solution.....	206
7.2	Previous Work.....	208
7.2.1	Approaches to Automatic Recovery Educational Content.....	208
7.2.2	Approaches to Learning Object Recommendation.....	210
7.3	Model Overview	213
7.3.1	Design for the Virtual Organization Model.....	215
7.4	Evaluation for the proposal	226
7.5	Experimental Results.....	227
7.5.1	Performance Evaluation for Architecture Recovery	228
7.5.2	Multilabel Classification Techniques for Learning Object	237
7.6	Proposed Model Generalization.....	239
7.7	Conclusions	240
7.8	Contributions	241
7.9	Future Work	243
	Bibliografía.....	245
	Referencias bibliográficas	247

Anexo A	Estándares para la Interoperabilidad de Contenidos Digitales Educativos	269
A.1	Estándares para el Mercado de Objetos de Aprendizaje	269
A.1.1	Dublin Core	269
A.1.2	IEEE LOM	272
A.1.3	Perfiles de Aplicación	279
A.1.4	Otros	281
A.2	Modelos de Agregación de Contenido	282
A.2.1	AICC (Aviation Industry CBT Committee).....	282
A.2.2	SCORM.....	283
A.2.3	Otros	286
A.3	Arquitecturas para la Interoperabilidad de ROAs	287
A.3.1	CORDRA.....	287
A.3.2	IMS Digital Repositories Specification - IMS DRI (Digital Repository Interoperability).....	290
A.3.3	IMS Learning Object Discovery & Exchange (LODE).....	296
A.4	Protocolos de Comunicación Cliente- ROA	299
A.4.1	Protocolo Z39.50	299
A.4.2	Search/Retrieve via URL (SRU)	301
A.4.3	The Open Archives Initiative Protocol for Metadata Harvesting (OAI-MPH).....	301
A.4.4	CEN SQI (Simple Query Interface).....	304
Anexo B	Estudio preliminar de condiciones	317
B.1	Funcionamiento de los ROAs Reales.....	319
B.2	Rendimiento de los ROA.....	321

Índice

B.2.1	Análisis de Errores	322
B.2.2	Tiempos vs. Número de resultados obtenidos.....	326
B.3	Estudio Sobre el Etiquetado de los OA en IEEE LOM	332
Anexo C	HERRAMIENTAS UTILIZADAS	347
C.1	EMFGormas.....	347
C.2	NetBeans	348
C.3	THOMAS	351
C.3.1	Arquitectura	351
C.3.2	Componentes Técnicos.....	356
C.4	MULAN	360
Anexo D	Proyectos, Publicaciones y Trabajos Relacionados	363

ÍNDICE DE FIGURAS

Figura 2-1. Taxonomía de los ROA en función de topología relacional	39
Figura 2-2. Capas definidas por SQL. Adaptada de [CWA 15454, 2005]	45
Figura 3-1. Conjuntos de estimación en la recuperación de documentos	52
Figura 3-2. Valores de Precisión de los Repositorios Analizados.....	58
Figura 3-3. Valores de la Exhaustividad Relativa de los Repositorios Analizados	60
Figura 3-4. Toma de datos del usuario para recomendación en un sitio web	64
Figura 3-5. Matriz de valoraciones.....	65
Figura 4-1. Clasificación de agentes según característica primaria [Nwana, 1995]	89
Figura 4-2. Clasificación de agentes software según Nwana [Nwana, 1995].....	90
Figura 4-3. Arquitectura básica de un agente BDI.....	94
Figura 4-4. Topología de organización. De izquierda-derecha y de arriba-abajo (1) Jerarquía, (2) Holarquía, (3) Coaliciones, (4) Grupos, (5) Congregaciones. (6) Federaciones y (7) Organizaciones Matriciales	108
Figura 4-5. Aprendizaje por Facilitación social (1) y por Imitación (2)	111
Figura 5-1. Resultado de búsqueda en MERLOT a través de su interfaz Web. 123	
Figura 5-2. Ejemplo de la descripción en LOM para un OA	125
Figura 5-3. Promedio de resultados obtenidos para consultas en dos momentos de tiempo.....	128
Figura 5-4. Tiempos de Acceso para evaluación de consultas en ROAs	129
Figura 5-5. Esquema general de distribución de tareas en la arquitectura	132
Figura 5-6. Fases de la metodología	136
Figura 5-7. Diagrama del modelo de organización (vista funcional). Misión de la arquitectura.....	142

Figura 5-8. Diagrama del modelo de organización de la arquitectura (Vista estructural).....	145
Figura 5-9. Diagrama del modelo de organización de la arquitectura (Vista funcional).....	145
Figura 5-10. Diagrama del modelo de organización (Vista funcional, Funcionalidad externa)	146
Figura 5-11. Diagrama del modelo de actividad del servicio Buscar OA (Perfil del servicio)	147
Figura 5-12. Diagrama del modelo de actividad del servicio BuscarOA (Relaciones entre A-Tareas)	148
Figura 5-13. Diagrama del modelo de actividad del servicio RevisarResultados (Perfil del servicio).....	149
Figura 5-14. Diagrama del modelo de actividad del servicio RevisarResultados (Relaciones entre A-Tareas que lo componen)	150
Figura 5-15. Diagrama del modelo de entorno de la arquitectura.....	150
Figura 5-16. Diagrama actualizado del modelo de organización, en el que se muestran los agentes internos, externos y los roles que juegan	151
Figura 5-17. Ejemplo de adquisición de rol por parte de un agente externo...	152
Figura 5-18. Diagrama actualizado del modelo de organización al adoptar una estructura de tipo congregación (Vista funcional).....	153
Figura 5-19. Diagrama actualizado del modelo de organización al adoptar una estructura de tipo congregación (Funcionalidad externa).....	154
Figura 5-20. Diagrama del modelo de interacción para el servicio "Buscar OA"	154
Figura 5-21. Diagrama de colaboración de la interacción BuscarOA	155
Figura 5-22. Parámetros generales configurables para la búsqueda	157
Figura 5-23. Diagrama de secuencia: Consulta.....	158
Figura 5-24. Interfaz de Usuario de Búsqueda de Contenidos en los Repositorios	159
Figura 5-25. Esquema funcional Búsqueda de OA	160

Figura 5-26. Diagrama funcional: catalogación de los OA recuperados.....	162
Figura 5-27. Ejemplo de Representaciones de casos en el dominio de los OA	164
Figura 5-29. Representación del espacio \mathcal{R}^3 de experiencias de usuario.....	167
Figura 5-28. Matriz de valoración de OA por los votos de usuarios.....	166
Figura 5-30. Presentación de resultados.....	170
Figura 5-31. OA en la opción “Ver más” del usuario.....	171
Figura 6-1. Documentos de Metadatos recuperados por el sistema en la fase de pruebas.....	180
Figura 6-2. Valores de la Ganancia Temporal.....	180
Figura 6-3. Comparativa de Documentos de Metadatos relevantes.....	181
Figura 6-4. Comparativas de precisión a bajo nivel para el modelo AIREH	182
Figura 6-5. Comparativas de la Exhaustividad	182
Figura 6-6. Comparativa de tiempos de consulta en ms. Entre las distintas fuentes.....	183
Figura 6-7. Presentación de resultados de una búsqueda al usuario	184
Figura 6-8. Evaluación de las recomendaciones del CBR.....	185
Figura 6-9. Evolución del CBR.....	186
Figura 6-10 Buscador de OA en GLOBE	187
Figura 6-11. Buscador de OA PALOMA	188
Figura 6-12. Número de ejemplos anotados con cada etiqueta	192
Figura 6-13. Ejemplo de la clasificación de Gramática LO.....	192
Figure 7-1. General scheme of distribution of tasks in architecture.....	207
Figure 7-2. The organizational model Diagram (functional view). Mission architecture	220
Figure 7-3. Organizational architecture model diagram (structural view)	221

Figure 7-4. Organizational architecture model diagram (functional view).....	221
Figure 7-5. Organizational architecture model diagram (functional view, external functionality).....	222
Figure 7-6. Model diagram of the architecture environment.....	222
Figure 7-7. Updated diagram of the organizational model that shows both internal and external agents, and the roles they play.....	223
Figure 7-8. Example acquisition role by an external agent.....	224
Figure 7-9. Adapted organizational model diagram according to a congregational structure (functional View).....	225
Figure 7-10. Adapted organizational model diagram according to a congregational structure (external functionality)	226
Figure 7-11. Metadata documents retrieved by the system in the testing phase	230
Figure 7-12. Temporary Gain Values	230
Figure 7-14. Comparative Precision Results for the Architecture Proposed ...	231
Figure 7-15. Relative Recall Comparative.....	232
Figure 7-16. Average consulting time.....	232
Figure 7-17. Evaluation of the recommendations of the CBR.....	234
Figure 7-18. Evolution of the CBR	235
Figura A-1. El modelo de datos de IEEE LOM.....	274
Figura A-2. Modelo CORDRA.....	289
Figura A-3. Arquitectura funcional de la especificación DRI [DRI, 2003].....	291
Figura A-4. Diagrama del Modelo General de Referencia diagrama con los distintos actores [DRI, 2003].....	293
Figura A-5. Función de núcleo: Search/Expose [DRI, 2003].....	294
Figura A-6. Función de núcleo Gather/Expose [DRI, 2003].....	294
Figura A-7. Funcionamiento de IMS LODE [LODE, 2010]	297
Figura A-8. Servicios de núcleo y de aplicación [CWA 15454, 2005]	304

Figura A-9. Intercambio de información entre dos repositorios a través de SQI [CWA 15454, 2005].....	306
Figura A-10. Modo de consulta síncrono en la especificación SQI.....	309
Figura A-11. SQI: Modo de consulta asíncrona [CWA 15454, 2005].....	311
Figura B-1. Crecimiento de los repositorios según OpenDOAR.....	318
Figura B-2. Errores en los repositorios Junio de 2010.....	325
Figura B-3. Comparativa de recuperación de OA sin errores en las dos tomas.....	325
Figura B-4. Tiempos (ms.) de consulta para el repositorio Agrega.....	327
Figura B-5. Tiempos (ms.) de consulta par el repositorio LORNET.....	327
Figura B-6. Tiempos de consulta (ms.) de los repositorios Acknowledge y Merlot.....	328
Figura B-7. Análisis de rendimiento ROA, Comparación de tiempos stateless y statefull.....	329
Figura B-8. Número medio de resultados.....	330
Figura B-9. Resultados recuperados para una misma consulta en dos tomas.....	330
Figura B-10. Resultados recuperados para las 60 consultas en Lornet en las dos tomas.....	331
Figura B-11. Análisis de los metadatos en IEEE-LOM de los OA por Etiquetas.....	333
Figura B-12. Análisis campos OA, Categoría 1. General.....	334
Figura B-13. Análisis campos OA, Campo 1.6 Keywords.....	335
Figura B-14. Análisis campos OA, Categorías 2. Life Cycle y 3. Meta-Metadata.....	336
Figura B-15. Análisis campos OA, Campo 4.3 Location.....	337
Figura B-16. Análisis campos OA, Categoría 5. Educational.....	338
Figura B-17. Análisis campos OA, Categoría 6. Rights.....	338
Figura B-18. Análisis campos OA, Categoría 7. <i>Relation</i>	339

Índice

Figura B-19. Análisis campos OA, Categoría 8. <i>Annotation</i>	340
Figura B-20. Análisis campos OA, Categoría 9. <i>Classification</i>	340
Figura B-21. Correspondencia IEEE LOM con Dublin Core	341
Figura B-22. Análisis campos OA, Porcentaje de uso de elementos	344
Figura B-23. Elementos de IEEE LOM más utilizados por los autores	345
Figura C-1. Interfaz EMFGormas e Eclipse.....	347
Figura C-2. Estructura de plugins de Eclipse	348
Figura C-3. Esquema de comunicación con servicios web	349
Figura C-4. Arquitectura THOMAS [GTIIA, 2009].....	352
Figura C-5. Aplicación THOMAS	357
Figura C-6. Interfaz THOMAS.....	358
Figura C-7. Interfaz manejo de SF en THOMAS.....	359
Figura C-8. Interfaz manejo de OMS en THOMAS	359
Figura C-9. Interfaz manejo de agentes e interacciones en THOMAS.....	360
Figura C-10. Interfaz manejo de servicios en THOMAS	360

ÍNDICE DE TABLAS

Tabla 3-1. Tópicos de Consultas realizadas.....	56
Tabla 3-2. Resultados de Ganancias Temporales en las dos muestras	57
Tabla 5-1. Conjunto de ROA Públicos con soporte a SQI	127
Tabla 5-2. Misión Organizativa de la arquitectura	142
Tabla 5-3. Grupos de interés de la arquitectura.....	143
Tabla 5-4. Condiciones de Entorno de la arquitectura.....	144

Tabla 5-5. Tecnología de Unidad de Trabajo. Descripción de las características del servicio Buscar OA.....	146
Tabla 5-6. Tecnología de Unidad de Trabajo. Descripción de las características del servicio RevisarResultados.....	148
Tabla 5-7. Transformación de la consulta lenguaje natural a PLQL ó vsql	161
Tabla 5-8. Atributos de los Casos.....	165
Tabla 6-1. Términos Seleccionados para la Batería de Consultas.....	178
Tabla 6-2. Características Comparativas de buscadores Federados de OA	189
Tabla 6-3. Resultados de Rendimiento.....	191
Table 7-1. Comparative characteristics of LO federated search engine	236
Tabla A-1. Elementos de la especificación Dublin Core Simple.....	270
Tabla A-2. Descripción del modelo de datos de IEEE LOM.....	275
Tabla A-3. Correspondencia IEEE LOM y Dublin Core	278
Tabla A-4. Método setQueryLanguage - Especificación SQI.....	307
Tabla A-5. Método setResultsFormat - Especificación SQI.....	308
Tabla A-6. Método setMaxQueryResults - Especificación SQI.....	308
Tabla A-7. Método setMaxDuration - Especificación SQI	309
Tabla A-8. Método setResultsSetSize - Especificación SQI	309
Tabla A-9. Método synchronuousQuery - Especificación SQI.....	310
Tabla A-10. Método getTotalResultsCount - Especificación SQI.....	310
Tabla A-11. Método asynchronousQuery - Especificación SQI.....	312
Tabla A-12. Método setSourceLocation - Especificación SQI	312
Tabla A-13. Método queryResultsListener - Especificación SQI	312
Tabla A-14. Método createSession - Especificación SQI	313
Tabla A-15. Método createAnonymousSession - Especificación SQI	314

Tabla A-16. Método destroySession - Especificación SQI	314
Tabla A-17. Errores en la especificación SQI	314
Tabla B-18. Características básicas de enlace del ROA	319
Tabla B-2. Análisis de ROA - Implementación de métodos	320
Tabla B-3. OA recuperados y tiempos medios	322
Tabla B-4. Códigos de error SQI.....	323

LISTA DE ACRÓNIMOS

ACM Association for Computing Machinery

ADL Advanced Distributed Learning

AICC Aviation Industry CBT Committee

AIREH Architecture for Intelligent Recovery of Educational content in Heterogeneous environments

AITP Association for Information Technology Professionals

ANSI American National Standards Institute

CAM Content Aggregation Model

CBR Case Based Reasoning

CBT Computer-Based Training

CCSDS Consultative Committee for Space Data Systems

CNRI Corporation for National Research Initiatives

CORDRA Content Object Repository Discovery and Registration/Resolution Architecture

CQL Contextual Query Language

DoD Departamento de Defensa (de los EEUU)

DSS Decision Support System

FIPA Foundation for Intelligent Physical Agents
IEEE Institute for Electrical and Electronic Engineers
IFIP International Federation on Information Processing
ILOX Information for Learning Object eXchange
IMS Information Management System
ISO International Standard Organization
KKD Knowledge Discovery from Databases
LCMS Learning Content Management System
LMS Learning Management System
LO Learning Object
LODE Learning Object Discovery & Exchange
LQS *Lexicon Query Service*
LTSC Learning Technology Standards Committee
OA Objeto de Aprendizaje
OAI-PMH Open Archives Initiative Protocol for Metadata Harvesting
OAIS Open Archival Information Systems
OMS Organization Management Systems
PLQL ProLearn Query Language
RIO Reusable Information Objects
ROA Repositorio de objeto de aprendizaje
RTE Run-Time Environment
SCORM Sharable Content Object Reference Model
SMA Sistema Multi-Agente
SN Sequencing and Navigation

SQI Simple Query Interface

SRU Search/Retrieve with URL

VSQI Very Simple Query Language.

PARTE I. INTRODUCCIÓN

1 INTRODUCCIÓN

El desarrollo de Internet ha causado una sociedad global de la información con un número creciente de contenidos y usuarios en el mundo entero. El sector educativo, poseedor y generador de grandes volúmenes de información, se está posicionando como uno de los más activos en esta revolución provocada por Internet y la aplicación de las nuevas tecnologías a los diversos contenidos digitales. Por un lado las diversas instituciones educativas, culturales y científicas del sector se encuentran en la ingente tarea de digitalizar y gestionar de forma automatizada los objetos de información y hacerlos accesibles a un usuario más o menos especializado. Por otro lado, dentro del sector educativo, el mayor esfuerzo que están llevando a cabo las instituciones se centra en la mejora de los procesos de creación y gestión eficaz de sus recursos didácticos. En el pasado, los trabajos de creación, adaptación o incluso de simple actualización de estos recursos derivaban en procesos largos y complejos y a menudo podían extenderse durante semanas o meses. Sin embargo en la actualidad, este proceso editorial de creación tradicional está siendo objeto de importantes cambios basándose en la máxima de la *reutilización* de contenidos soportados por un gran despliegue de tecnologías y estándares que configuran el motor de evolución actual del llamado e-learning.

Uno de los enfoques más ampliamente aceptados en el contexto de la educación a distancia se basa en fragmentar los contenidos en unidades modulares y autocontenidas de forma que puedan ser reutilizadas en distintos contextos educacionales así como en diferentes plataformas. Este planteamiento constituye el paradigma educacional basado en los denominados objetos de aprendizaje (OA) [Lujara *et al.*, 2007] y que será el contexto en el que se evaluará el desarrollo de este trabajo de investigación. La principal ventaja que se deriva del uso de estos elementos es la reutilización de los recursos educativos, ya que dichos objetos de aprendizaje están formados por dos elementos conectados. Por una parte el recurso en sí, destinado a la educación y por otro el etiquetado del mismo, que a través de una serie de metadatos los describe completamente. La utilización sistemática de dicho etiquetado permite automatizar procesos de búsqueda, catalogación y recuperación de dichos recursos.

Las tareas de uso, recuperación y catalogación de recursos educativos digitales se desarrollan en un entorno que implica el acceso a un número elevado de sistemas de información heterogéneos y dispersos. Los accesos a dichos sistemas de información se realizan a través de diferentes protocolos, dependientes de cada proveedor de contenidos a los que se accede de manera aislada. Estas dificultades convierten la tarea de recolección y clasificación de

dichos contenidos en una labor muy compleja. El paradigma de objeto de aprendizaje abre nuevas posibilidades en la gestión de los materiales digitales para los sistemas educativos a través de diferentes tecnologías y protocolos de comunicación. El uso de dicho paradigma establece un contexto semántico, determinado por los estándares empleados en su etiquetado, que posibilita el desarrollo de nuevos servicios automáticos relacionados con su uso. En este contexto es necesario el desarrollo de herramientas que gestionen de una manera centralizada el proceso de búsqueda y selección de los contenidos educativos almacenados en las colecciones dispersas por internet. En esta memoria se presenta una arquitectura que soporta la reutilización de dichos recursos educativos facilitando los procesos de búsqueda y filtrado personalizado de contenidos educativos digitales a través de búsquedas federadas en repositorios de objetos de aprendizaje distribuidos. El modelo propuesto está sustentado en las capacidades auto-adaptativas en entornos dinámicos y heterogéneos que proporcionan los sistemas multi-agente con capacidades organizacionales.

1.1 DESCRIPCIÓN DEL PROBLEMA Y MOTIVACIÓN

En aras de la reutilización, la aplicación de las tecnologías de la información en el sector educativo basa su esfuerzo de desarrollo actual en la fragmentación de los contenidos en unidades modulares e independientes que puedan ser utilizadas en distintos entornos y por diferentes aplicaciones en diferentes experiencias formativas. La reutilización permite reducir notablemente el tiempo de creación, ya no sólo de los recursos educativos, sino también de su agregación conformando cursos completos a demanda. El cambio en el proceso de creación se produce al basarse éste en la búsqueda y agregación de estas pequeñas piezas modulares, como simples recursos didácticos específicamente diseñados para ser utilizados en diferentes contextos. Con el propósito de facilitar dicha reutilización de recursos educativos, se ha formalizado el concepto de *objeto de aprendizaje* (OA) definido como “entidad digital o no digital que puede ser utilizada, reutilizada o referenciada mientras el aprendizaje sea soportado por tecnologías” [IEEE 1484.12.1-2002]. No puede ser considerado en sí mismo como una auténtica tecnología, pero es el germen de una metodología de trabajo y a su alrededor surgen significativos avances tecnológicos. El empleo de objetos de aprendizaje trae consigo toda una serie de nuevos conceptos, marcos de trabajo, problemas y cuestiones abiertas a la investigación, así como un conjunto de técnicas y tecnologías asociadas.

Habitualmente, los objetos de aprendizaje se almacenan en repositorios digitales, los *repositorios de objetos de aprendizaje (ROA)*; éstos suelen ser altamente heterogéneos, teniendo diferentes sistemas de almacenamiento, acceso a los objetos y métodos de consulta propios que exigen en muchas ocasiones la búsqueda manual de recursos, etc. El problema de la

heterogeneidad en los sistemas de bases de datos es un tema ampliamente estudiado desde hace tiempo [Abello *et al*, 2000], pero sigue siendo crucial para la recuperación de datos referentes a los objetos de aprendizaje y la homogenización y tratamiento de los resultados posteriormente.

El concepto de federación existe en diversos contextos; sin embargo, en el entorno de los sistemas de información consiste en la integración de sistemas de cómputo heterogéneos y distribuidos, donde la característica de la federación es la cooperación entre los sistemas independientes permitiendo una integración controlada [Sheth *et al*, 1990]. En cada uno de estos sistemas, se encuentra almacenada la información que no puede ser indexada por los buscadores web tradicionales, debido a la incompatibilidad de los formatos de dicha información con los procesos de indexación del buscador, así como de la disponibilidad en línea de la misma. La búsqueda federada es el mecanismo encargado de proveer así un acceso unificado a dichos sistemas de información, permitiendo el acceso a información que se encuentra en un formato diferente o incompatible a otros medios de búsqueda. Dicha búsqueda permite que los diversos sistemas de información sean capaces de interoperar entre ellos y de esta manera acceder a información residente en los mismos que de cualquier otra manera no podría ser consultada.

El hecho de que los repositorios de objetos de aprendizaje no posean ningún sistema en un nivel superior que permita la abstracción entre su funcionamiento interno y los usuarios finales de los datos almacenados resulta un problema del que adolecen la mayoría de repositorios en la actualidad. Con el fin de evitar el problema señalado algunos repositorios poseen una capa de abstracción que conecta las características internas de los mismos con el exterior permitiendo una mayor automatización y tratamiento informático de los OA que contienen. La existencia de dicha capa de abstracción da paso a la generación de un nuevo tipo de aplicaciones, que permiten la búsqueda de información en entornos distribuidos a través de diferentes formatos, servidores y redes, a través de sistemas de búsqueda federada de objetos de aprendizaje.

Para dar soporte al proceso de reutilización (Rapidez, Flexibilidad y Uniformidad) el objeto de aprendizaje cuenta con unos descriptores externos, a modo de etiqueta, que permitirán su identificación, organización, búsqueda y recuperación. Dicha descripción se realiza a través de metadatos basados en estándares a fin de asegurar la interoperabilidad entre sistemas y conformando el conjunto de atributos necesarios para describir adecuadamente el recurso. A través ellos se tiene un primer acercamiento al objeto y permiten conocer rápidamente sus características, siendo especialmente útiles en recursos no textuales, en los que su contenido no puede ser indexado por sistemas automáticos. Son varios los estándares que conviven en la actualidad para el etiquetado de objetos de aprendizaje. Los más remarcables son Dublin Core

[DCMI, 2010], MPEG-7[MPEG-7, 2010] y el más importante el Learning Object Metadata, LOM [IEEE 1484.12.1-2002]. En el etiquetado de estos OA, conviven distintos estándares con el uso de vocabularios, idiomas diferentes, etc. Por otro lado en ningún caso existen criterios de calidad que verifiquen que los descriptores que los autores vinculan al OA en cuestión sean ni válidos ni ricos en la información que aportan sobre él.

En este contexto ha surgido la aparición de un nuevo tipo de aplicaciones dedicadas a la búsqueda de contenidos en repositorios de objetos de aprendizaje [Totschnig *et al.*, 2010], [Duval *et al.*, 2011]. La realización de consultas simultáneas en diferentes repositorios, permite una mayor interoperabilidad y por lo tanto una mejor reusabilidad de los recursos que en ellos se almacenan. Gracias a las posibilidades que ofrecen este tipo de aplicaciones de búsqueda, las dinámicas de gestión para el almacenamiento, la recuperación y el uso de los OA en el contexto de la educación con soporte de materiales digitales on-line se encuentra en un importante proceso de evolución.

Con esta creciente avalancha de información al alcance aparece la dificultad de encontrar lo que se quiere, cuando se necesita y en la manera que mejor se adapte a nuestras necesidades y/o exigencias. El usuario se enfrenta con situaciones en las cuales tiene demasiadas opciones de elección y necesitará una ayuda para explorar, eliminar información no válida o seleccionarla filtrando la miríada de opciones con sus preferencias. Los motores de Búsqueda de Internet están diseñados para ser provechosos sobre información visible directamente en la Web, perdiendo su utilidad para búsquedas en repositorios de sectores específicos que albergan bases de datos especializadas, como las de contenidos educativos. Este hecho origina la necesidad de utilizar otro tipo de aplicaciones de búsqueda y gestión de la información (objetos de aprendizaje) que sean capaces de acceder y elaborar convenientemente todos los datos para dichos contenidos educativos generando como resultado una información útil para el usuario que le facilite el encontrar aquellos que son relevantes de acuerdo a las singularidades y los criterios fijados por el interesado.

Como se sucede de esta descripción del contexto educativo actual, la recuperación de objetos de aprendizaje válidos para un contexto educativo o un usuario final resulta una tarea muy compleja y en evolución, lo que le confiere una dinámica abierta y cambiante. Por ello, es necesario investigar en técnicas, herramientas y metodologías que permitan adaptarse de forma rápida y eficaz a los cambios que se produzcan, como por ejemplo nuevos estándares, diferentes lenguajes de consulta, sistemas de almacenamiento, etc. La adaptación al entorno en contextos abiertos y heterogéneos es un reto conocido en el campo de los sistemas multi-agente [Zambonelli *et al.*, 2004] [Reitbauer *et al.*, 2004] [Weyns *et al.*, 2004] [Capera *et al.*, 2003] [Razavi *et al.*, 2005] con el desarrollo de sistemas software capaces de responder y actuar por sí mismos ante cambios en su entorno. Estos requisitos definen un conjunto de

características que se abordan en este trabajo haciendo uso del paradigma de los sistemas abiertos y de las organizaciones virtuales. Dichos sistemas multi-agente basados en organizaciones [Rodríguez, 2009] disponen de teorías, modelos, mecanismos, métodos y herramientas que permiten desarrollar sistemas con capacidad de reorganización y que pueden adaptarse de esta forma a futuros cambios en su entorno. El apartado siguiente detallará la hipótesis de partida así como los objetivos marcados para la obtención de resultados.

1.2 HIPÓTESIS Y OBJETIVOS

Consideradas en el apartado anterior las carencias existentes actualmente para hacer efectivos los procesos de recuperación de contenidos en el contexto educativo, el presente trabajo de investigación aborda el estudio y desarrollo de una solución partiendo de la siguiente hipótesis:

La hipótesis de partida del presente trabajo de investigación se formaliza en que es posible modelar un sistema eficiente para la búsqueda, recuperación y gestión de contenido digital educativo en sistemas heterogéneos a partir de un modelo de organizaciones adaptativo que facilite una planificación distribuida y dinámica.

Se plantea la reutilización de contenidos educativos desde la perspectiva de su búsqueda y recuperación eficaz. Se establece la necesidad de incorporar criterios de calidad de los contenidos recuperados, así como una gestión eficaz de la información en el contexto. Las particularidades del desarrollo de SMA desde el punto de vista organizacional posibilitarán la formalización de un modelo para la búsqueda y extracción óptima de recursos digitales en entornos heterogéneos.

El objetivo principal que se pretende conseguir es modelar una arquitectura multi-agente basada en organizaciones como mecanismo eficaz para la búsqueda, recuperación y catalogación de recursos digitales en entornos heterogéneos.

En este trabajo de investigación confluyen aspectos de diferentes áreas de investigación tales como la recuperación de información y la teoría y desarrollo de agentes. Para alcanzar dicho objetivo, se estudian los conceptos y formalismos utilizados tanto por agentes deliberativos BDI como en las estructuras organizacionales de las sociedades de agentes. El desarrollo del modelo en un entorno educativo exigirá el estudio de estándares y protocolos de comunicación de repositorios en dicho entorno. Abordaremos el estudio de los repositorios de contenidos educativos orientando la solución del modelo

específicamente a solucionar los problemas de búsqueda y recuperación eficaz de los metadatos de objetos de aprendizaje (OA).

Para la realización de este trabajo es necesario llevar a cabo otros objetivos más específicos que permitan la consecución del objetivo final. Entre ellos:

- ❖ Realizar un análisis de los problemas concretos de la recuperación de contenidos educativos de calidad en entornos heterogéneos para detectar las carencias existentes.
- ❖ Realizar un estudio de las tecnologías de almacenamiento y recuperación en entornos heterogéneos.
- ❖ Buscar mecanismos formales útiles para la resolución de la búsqueda automática de información en entornos heterogéneos y dinámicos en tiempo de ejecución. En este sentido, en entornos abiertos, cambiantes y heterogéneos, los SMA dotados de capacidades organizacionales posibilitan mecanismos de generación de soluciones de forma autónoma en los que se profundizará.
- ❖ Proponer un modelo de arquitectura multi-agente para la recuperación y gestión de contenidos digitales en bases de datos heterogéneas sobre la base de organizaciones virtuales. Dicha arquitectura debe de ser flexible y adaptable a las necesidades del entorno de aplicación así como a los usuarios.
- ❖ Implementar la teoría de agentes como base de la arquitectura multi-agente en la búsqueda, filtrado y catalogación de objetos de aprendizaje.
- ❖ Desarrollar el prototipo de un sistema para un escenario de recuperación de contenido educativo específico mediante una organización de agentes adaptada al problema definido.
- ❖ Evaluar empíricamente los resultados obtenidos en entornos reales de aplicación.

1.3 METODOLOGÍA

El proceso de investigación que regirá este trabajo se basa en la metodología investigación-acción (*Accion-Research*). Esta metodología permite la práctica de una investigación empírica. Se parte de la identificación de un problema real enumerando un conjunto de posibles hipótesis que permitan en su estudio seleccionar una de ellas que permita el desarrollo de una propuesta enfocada a la obtención de una solución. Se ejecuta la acción que permita comprobar la hipótesis seleccionada. Finalmente, se formularán las conclusiones respectivas en la evaluación de los resultados del desarrollo e investigación sobre dicha propuesta. Para formalizar dicho modelo de investigación se han definido seis actividades descritas brevemente a continuación y que han sido completadas

de acuerdo a los objetivos planteados en el transcurso del presente trabajo de investigación:

1. Identificación y descripción de las características del problema de la recuperación de objetos de aprendizaje en entornos heterogéneos. A lo largo de esta actividad se presenta la problemática, se definen las características del entorno de trabajo y se proponen diferentes hipótesis para la solución total o parcial del problema. Este punto permite plantear los objetivos necesarios para la consecución de la solución.
2. Estudio y revisión constante e incremental del estado del arte. A lo largo de esta fase se ha analizado el estado del arte de las áreas, tecnologías y desarrollos relacionados con la presente investigación que ha permitido la obtención de un marco teórico y sus posibles desarrollos, esta fase ha posibilitado el enriquecer el conocimiento en la materia y mejorar el desempeño del trabajo de investigación presentado.
3. Diseño gradual e iterativo del modelo de acción propuesto. Partiendo de la información obtenida en las actividades anteriores, se ha diseñado y desarrollado un modelo que integra los componentes necesarios para proponer el desarrollo de acción que permite una solución innovadora y original a la problemática definida.
4. Desarrollo e implementación del modelo a través del desarrollo de un sistema prototipo. En esta fase se han formalizado las funcionalidades, componentes, interacciones, etc. del prototipo y su implementación en escenarios de aplicación real en el contexto de la recuperación de contenidos educativos en entornos heterogéneos. La implementación del prototipo ha permitido la generación de pruebas y obtención de resultados que han servido para la evaluación de la solución del modelo propuesto.
5. Análisis de resultados y formulación de conclusiones.
6. Diseminación constante del conocimiento, resultados y experiencias a la comunidad científica. Esta actividad continua a lo largo del todo el proceso de investigación planteado ha consistido en la elaboración y presentación de diversas publicaciones en revistas, congresos, talleres, generación de informes técnicos, etc. que han permitido validar y dar a conocer los avances y resultados parciales de los distintos hitos de la investigación en ámbitos de expertos con la consiguiente realimentación al trabajo.

1.4 ESTRUCTURA DE LA MEMORIA

Este documento se ha estructurado en siete capítulos y diversos apéndices distribuidos en cuatro bloques que conforman la presente memoria. Dicha estructura atiende al criterio de abordar por contenidos temáticos sucesivos el trabajo realizado. Los distintos bloques se detallan a continuación.

La **primera parte** de esta memoria corresponde al *capítulo 1*, en el cual se hace una introducción a este trabajo de investigación. Se describe la problemática en torno a la búsqueda y recuperación de contenidos digitales en el contexto educativo actual. Se señala la necesidad de obtener una mayor flexibilidad para los modelos actuales que permitan una solución en los modelos de recuperación de contenidos y que faciliten la reutilización de los recursos y posibiliten la adaptación de estos sistemas frente a cambios en el entorno. Se presentan los objetivos, la hipótesis y la motivación que han llevado al desarrollo del modelo. Finalmente, se detalla la metodología de investigación aplicada y se realiza una breve descripción de la estructura de la memoria.

La **segunda parte**, formada por los capítulos 2, 3 y 4, establece el marco en el que se desarrolla el trabajo. Presenta el estado del arte y diferentes enfoques existentes al problema planteado.

En el *capítulo 2*, se analiza el estado del arte del contexto actual del e-learning. Partiendo del paradigma de objeto de aprendizaje se expone su aplicación práctica en el contexto de la generación y soporte a los contenidos digitales. El capítulo analiza las especificaciones y estándares de metadatos educativos así como el papel de los Repositorios de Objetos de Aprendizaje como soporte a la interoperabilidad de los recursos almacenados.

El *capítulo 3* analiza las particularidades de los sistemas de recuperación eficiente de contenidos en grandes volúmenes de información. Se analizarán mecanismos y tecnologías complementarias para la evaluación y filtrado de los contenidos recuperados con el objetivo de adecuar los resultados obtenidos a las preferencias de los usuarios. Este capítulo propone además medidas de evaluación de recuperación eficiente de objetos de aprendizaje que serán utilizadas en el desarrollo posterior.

El *capítulo 4* analiza los sistemas multi-agente, haciendo hincapié en agentes que utilizan el modelo deliberativo BDI y que son capaces de cooperar con otros agentes dentro de una "sociedad". Especial importancia tendrán conceptos como adaptación, coordinación y planificación de agentes que sostienen las capacidades organizativas de los mismos y configuran el estudio de los modelos sociales de agentes. En este capítulo se estudiará la integración en los agentes de tecnologías complementarias, como son los sistemas de razonamiento basados en casos, para centrarse en aspectos importantes para el desarrollo posterior de la propuesta.

En la **tercera parte** de la memoria que corresponde al *capítulo 5* presenta en detalle el modelo de arquitectura propuesta, denominada AIREH (*Architecture for Intelligent Recovery of Educational content in Heterogeneous environments*). Dicha arquitectura está basada en organizaciones virtuales para la búsqueda y recuperación de contenidos digitales educativos. Se definen las principales funcionalidades que desempeña y los componentes que la integran tales como

los tipos de agentes y servicios utilizados, así como las tecnologías base para explotar su potencial.

La **cuarta parte** de la memoria presenta la evaluación del modelo mediante un caso de estudio a través de la funcionalidad de la arquitectura propuesta AIREH. Este apartado agrupa los resultados experimentales y las conclusiones obtenidas. Incorpora además un resumen del trabajo en inglés a modo de capítulo adicional. Esta última parte está formada por los capítulos 6 y 7. En el *capítulo 6* se plantea un problema real enmarcado en un entorno dinámico, que sirve como banco de pruebas para la aplicación del modelo. El problema trata de la búsqueda y recuperación de objetos de aprendizaje residentes en repositorios junto con aspectos de la gestión eficaz de los mismos. Entre estos aspectos destaca que el *sistema proporciona un sistema de búsqueda federada, lo que permite realizar* búsquedas en varios repositorios dispersos de forma simultánea a la vez que una gestión integral de todo el proceso para satisfacer las necesidades y requisitos de los usuarios. La implementación de la propuesta estará formada por una organización virtual en la que se pondrá a prueba del modelo de organización planteado en el capítulo anterior en un entorno real. El *capítulo 6* recoge las pruebas realizadas sobre la implementación particular de AIREH junto con los resultados obtenidos. Además se presentan las conclusiones obtenidas, se resumen las contribuciones de este trabajo al campo de la recuperación inteligente en entornos heterogéneos y se indican algunas de las posibles líneas de investigación futura. El *capítulo 7* del documento se corresponde con un resumen en inglés del trabajo realizado, tanto estudios iniciales, como propuesta, resultados y conclusiones obtenidas.

Para finalizar se proporciona una lista de las **referencias bibliográficas** utilizadas en este trabajo.

Los cuatro **anexos** incluidos permiten completar y detallar diferentes aspectos de la memoria no recoge y que han sido estudiados a lo largo del trabajo y conforman un recopilatorio sobre el entorno en el que se desarrolla el mismo. El *anexo A* recoge un estudio pormenorizado de los estándares que sustentan la interoperabilidad de contenidos educativos. El *anexo B* detalla un estudio preliminar realizado sobre los repositorios de objetos de aprendizaje y su funcionamiento real en el contexto educativo. El *anexo C* describe herramientas relevantes utilizadas en el desarrollo de este trabajo de investigación. Finalmente, el *anexo D* enumera proyectos, publicaciones y trabajos en relación con el desarrollo en este trabajo de investigación.

PARTE II. ESTADO DEL ARTE Y ENFOQUES

2 TECNOLOGÍAS Y ESTÁNDARES PARA CONTENIDOS DIGITALES EDUCATIVOS

En la última década, está cobrando una especial relevancia la necesidad de mejora de los procesos de acceso a contenidos digitales, donde la experiencia de usuario se contempla como una medida de calidad en el acceso y uso de dicha información. La correcta gestión del conocimiento implica una mejora en los sistemas de recuperación y tratamiento de la información, permitiendo una mayor adaptabilidad a las necesidades reales y a los modelos de búsqueda que utilizan los usuarios.

Existen gran variedad de grupos encargados del desarrollo de especificaciones y estándares orientados a identificar, definir y comunicar los recursos involucrados en el entorno de aprendizaje electrónico. Por un lado, los mecanismos de etiquetado mediante metadatos educativos describen de manera homogénea contenidos, métodos y elementos relacionados con la formación. El objetivo del uso de dichos metadatos es identificar dichos componentes y facilitar los procesos de comunicación, acceso, intercambio, distribución y reutilización en diferentes plataformas, niveles educativos y materias. Mediante diferentes especificaciones de metadatos se pueden identificar todos los elementos existentes en el entorno educativo tales como los recursos de aprendizaje, ámbitos educativos, perfiles de alumnos, tipos de pruebas y evaluaciones. Por otro lado, los procesos de comunicación, intercambio y reutilización de los contenidos educativos, van asociados al desarrollo de estándares sobre las arquitecturas y tecnologías que permiten el modelo de despliegue de los repositorios digitales interoperables que los almacenan, gestionan y permiten su uso a través de múltiples funcionalidades.

Las interoperabilidad, en el contexto de este trabajo, será el resultado de agregar de manera adecuada los distintos elementos implicados. Es necesaria la existencia de interoperabilidad entre los propios objetos de aprendizaje, los repositorios que los albergan y de sus metadatos, por lo que a continuación se describen en detalle los estándares que proporcionan las herramientas necesarias para cumplir con la integración de todos estos componentes definiendo el complejo contexto actual de los contenidos educativos en el que se plantea este trabajo.

Los metadatos educativos surgen como resultado de diferentes iniciativas desde distintos ámbitos a la propuesta de soluciones bajo la forma de

estándares, especificaciones o perfiles de aplicación, según sea el caso. Los metadatos educativos establecen así cómo etiquetar adecuadamente elementos educativos por medio de la definición de sus atributos o características.

2.1 OBJETOS DE APRENDIZAJE

Uno de los enfoques más ampliamente aceptados en el contexto de la tecnología educativa relacionada con el desarrollo de contenidos se basa en fragmentar estos en unidades modulares independientes de manera que puedan ser reutilizadas en distintos entornos y por diferentes aplicaciones. Las unidades así concebidas se denominan en inglés *learning objects* (LO). Dicho término puede traducirse por objetos de aprendizaje (OA). Un objeto de aprendizaje puede decirse que es un recurso digital especialmente preparado para formar parte de cursos u otras experiencias formativas. No obstante, parece útil explicar con mayor detalle cuáles son los conceptos y las definiciones que manejaremos así como sus límites y expectativas. A lo largo de estos apartados se intentará clarificar qué características hacen que un contenido educativo pueda ser denominado objeto de aprendizaje así como las implicaciones conceptuales y tecnológicas que de ello se derivan.

2.1.1 Concepto y Cometido

El *Learning Technology Standards Committee* (LTSC) del *Institute of Electrical and Electronics Engineers* (IEEE) define un objeto de aprendizaje como “entidad digital o no digital que puede ser utilizada, reutilizada o referenciada mientras el aprendizaje sea soportado por tecnologías”. El LTSC proporciona ejemplos de dichos objetos, incluyendo diferentes tipos de contenidos y medias tales como “multimedia content, instructional content, learning objectives, instructional software and software tools, and persons, organizations, or events referenced during technology supported learning” [IEEE 1484.12.1-2002]. Debido a esta definición tan general, casi cualquier cosa en el contexto educativo podría ser considerada un objeto de aprendizaje desde un libro de texto tradicional, una página web, un fragmento de contenido multimedia, un instrumento de software y aún una persona, un acontecimiento, o incluso un lugar. Debido a la vaguedad de la definición de IEEE, ésta ha sido sumamente criticada, al no permitir discernir e identificar claramente qué es un objeto de aprendizaje lo que la invalida como definición.

No consolidándose como la definición de objeto de aprendizaje mundialmente aceptada aparecen múltiples discusiones en el intento de definir y acotar el alcance de dicho concepto ([Friesen, 2004], [Sosteric *et al.*, 2002], [Wiley, 2001], [Polsani, 2003], [Wiley, 2000], [Moreno & Bailly-Baillièrè, 2002]). Parece entonces más práctico centrar el concepto de OA entorno a las características que lo hacen efectivo para el cometido de contenido educativo reutilizable.

Las definiciones ([Moreno & Bailly-Baillièrre, 2002] [Polsani, 2003]; [Wiley, 2000]) que apuntan a que la principal característica de un OA es que tenga un *nivel de granularidad adecuado* para el intercambio de contenidos. En esta misma línea [Kottler *et al.*, 2000] incorpora que pueden ser “reutilizados” con *independencia del medio* (Internet, Intranets, CD-ROM, clases presenciales, etc.) y personalizados según las necesidades instructivas”. Polsani (2003) incorpora a esta línea connotaciones pedagógicas y define un OA como una unidad de aprendizaje *independiente y autocontenida* que está predispuesta a ser reutilizada en *múltiples contextos instruccionales*.

Wiley (2001) explicita la necesidad de que los OA sean recursos digitales con *finés educativos* lo que completaría la definición del IEEE. Sin embargo no incluye restricciones a otra serie de aspectos y elementos que derivarían en una definición más clara del término como el de la granularidad.

L’Allier elabora una estructura para el OA en NETg [L’Allier, 1997], donde lo define con un modelo estructural que *contiene tres elementos*: un objetivo, una actividad de aprendizaje y una valoración que permite determinar si se ha alcanzado el objetivo de aprendizaje propuesto. Esta definición es una de las que más se acerca a un contexto pedagógico, pero no se pronuncia acerca de la interactividad con el OA y ha sido cuestionada porque según autores como Polsani “cualquier definición que estipula el uso deseado, método y mecanismo de medida de un OA previamente, restringe su usabilidad porque la metodología intención y valoración son determinadas por la situación instruccional y no por el OA en sí mismo”.

Siguen apareciendo distintas definiciones del concepto, que no aportando información relevante al respecto y reiteran tantas otras características deseables de los OA. En este sentido se toma la definición de [Chiappe *et al.*, 2007], donde describe el objeto de aprendizaje como una “entidad digital, autocontenible y reutilizable, con un claro propósito educativo, constituido por al menos tres componentes internos editables: contenidos, actividades de aprendizaje y elementos de contextualización. A manera de complemento, los objetos de aprendizaje han de tener una estructura (externa) de información que facilite su identificación, almacenamiento y recuperación: los metadatos”. En el contexto de este trabajo de investigación, la principal característica de los objetos de aprendizaje junto con todas las demás características descritas ha de ser el contener información clara para que puedan ser recuperados y tratados adecuadamente de forma automática de modo que la reutilización sea un proceso eficiente.

La idea es que un OA sea una unidad mínima de contenido con la intención de enseñar algo y que pueda ser reutilizado para otras situaciones educativas sin problemas de compatibilidad entre distintas plataformas. Esto significa que dichos materiales educativos se deben desarrollar, organizar y se distribuir de

una manera uniforme [Hamel&Ryan-Jones, 2002]. Es así como se podría reutilizar de manera razonable un mismo objeto en distintos niveles y disciplinas permitiendo la automatización de todos los procesos implicados.

2.2 ESTÁNDARES PARA EL MERCADO DE LOS OBJETOS DE APRENDIZAJE

El uso de los estándares es indispensable ante la masificación de cualquier tecnología: permiten su crecimiento, expansión y generalización. En los últimos años el uso de los *metadatos* (del griego *μετα*, *meta*, «después de» y del latín *datum*, «lo que se da», «dato» es decir, datos que describen otros datos) se ha acrecentado debido a la necesidad de identificar, organizar y almacenar las grandes cantidades de información en formato digital generadas. A principio de la década de los noventa, el uso creciente de metadatos crea la necesidad de estandarizar su uso de modo que se pudiesen identificar mediante descriptores consensuados archivos digitales de cualquier ámbito, facilitando todas las labores de clasificación y recuperación relacionadas. Un registro de metadatos, consiste en una serie de atributos o elementos necesarios para describir un recurso en cuestión.

En el ámbito educativo, la gran generación de contenidos digitales, conlleva un uso especializado de estándares para el etiquetado de dichos contenidos que incorpore todas las tareas relacionadas con dicho sector.

Existen una serie de elementos reglados, los cuales formalizan el funcionamiento de los metadatos educativos. Dichos elementos sobre los que se fundamenta la interoperabilidad de los contenidos educativos son los estándares, especificaciones, perfiles de aplicación, elementos de contenido, vocabularios controlados, etc. Un estándar para la definición de metadatos educativos resulta de un acuerdo en las características que un elemento de aprendizaje, método o técnica, independientemente del sistema informático y el hardware que se utilice. El estándar consiste finalmente en una tecnología, formato o método que ha sido ratificado por un organismo oficial de estandarización. Si éste no ha sido aprobado por ningún organismo oficial de estandarización se denomina especificación. Por otro lado aparecen también los llamados perfiles de aplicación, cuando la implementación de estándares o especificaciones en contextos educativos concretos exige considerar diferentes particularidades. En estos perfiles de aplicación se combinan y utilizan uno o más estándares o especificaciones para crear nuevos esquemas dirigidos a situaciones específicas, donde no se agregan elementos nuevos para asegurar la interoperabilidad con los estándares o especificaciones originales.

El desarrollo de estándares en e-learning está todavía en avance. En la medida que se vaya generalizando y afianzando, su uso permitirá una mayor

interoperabilidad, reusabilidad, manejabilidad, accesibilidad, durabilidad, escalabilidad y confiabilidad.

En el caso de los OA, los metadatos describen mediante etiquetas los atributos de dichos elementos instructivos con el fin de garantizar la interoperabilidad técnica. Potencian la utilidad de los OA, haciendo más fácil su almacenamiento, recuperación, localización e intercambio y por tanto su reutilización. Su importancia radica en el hecho de que a través de estos metadatos se puede hacer el primer acercamiento con el propósito de conocer rápidamente sus principales características. Para que los OA satisfagan tan amplia expectativa, apoyados en los metadatos, es indispensable que éstos hagan uso de los estándares que establecen normas tipificadas a dichos descriptores soportando la interoperabilidad de dichos recursos educativos.

Los metadatos de los OA describen dichos contenidos educativos proporcionando información dispar con distintos propósitos. Dicha información debe de atender a fines educativos, administrativos y estructurales, así como sobre el contexto, contenido y control [Caplan, 2005]. En este sentido, [Caplan, 2003] define tres grandes grupos de metadatos:

- **Metadatos descriptivos.** Su propósito es identificar de forma unívoca a un recurso y gracias a ello permite distinguir un recurso de otro. Este tipo de metadatos sirve también para formar colecciones de recursos similares, para la realización de funciones de evaluación, para la relación con otros recursos y para fomentar la usabilidad.
- **Metadatos administrativos.** Su fin es anotar información que facilite la gestión de los recursos, lo que incluye información como cuándo y dónde fue creado, quién es su responsable, así como información técnica, etc.
- **Metadatos estructurales.** Su objetivo es identificar cada una de las partes que componen el recurso, definiendo la estructura que le da forma.

Al utilizar un estándar de metadatos es posible organizarlos, localizarlos, recuperarlos, procesarlos, evaluarlos, intercambiarlos y reutilizarlos; es decir, garantizar la interoperabilidad con otros sistemas que manejen metadatos compatibles. Pero es importante destacar, que más allá del esquema de metadatos están los datos que se asignan a cada descriptor. La forma en que éstos se introduzcan facilitará o dificultará la identificación y recuperación de los recursos [Sicilia *et al.*, 2005].

Las especificaciones más utilizadas para la descripción de recursos digitales son el estándar Dublin Core [DCMI, 2010] y el IEEE *Learning Object Metadata* (IEEE LOM) [IEEE, 2002] que hasta la fecha es la única especificación

formalmente aprobada y que por lo tanto constituye el único estándar existente.

El origen de la especificación Dublin Core tiene su origen en la 2ª Conferencia Internacional de la World Wide Web Conference, en Chicago en octubre de 1994 y el taller conjunto organizado para discutir la semántica de metadatos en Dublin, Ohio, en marzo de 1995 (de ahí el nombre del estándar). Se desarrolló fundamentalmente en el ámbito bibliotecario, por lo que no está específicamente pensando para ser utilizado en un contexto educativo. Fue elaborado por la DMCI (*Dublin Core Metadata Initiative*), una organización dedicada a fomentar la adaptación de estándares con el objetivo de lograr interoperabilidad entre sistemas y a desarrollar vocabularios especializados para la descripción de recursos; todo ello para poder crear sistemas de búsqueda de información más inteligentes. Al tratarse de un sistema de marcado muy genérico y que está muy extendido, es habitual utilizarlo también en cualquier otro contexto, incluido el educativo. Así aunque este estándar no surge específicamente para la descripción de los OA, puede ser aplicado a la mayoría de los recursos, llegando a estar integrado en un número importante de especificaciones de OA como las definidas en IEEE LOM, EdNa, GEM, que describiremos en los siguientes apartados de esta memoria.

El estándar IEEE LOM (*IEEE Learning Object Metadata*), o simplemente LOM, es un modelo de datos utilizado para describir los OA. Resulta ser la única especificación reconocida internacionalmente desde el año 2002 a través del estándar de IEEE 1484.12.1:2002 [IEEE 1484.12.1-2002]. El propósito de este esquema es facilitar la búsqueda, evaluación, adquisición y uso de los OA, facilitando su intercambio y uso compartido, así como el desarrollo de catálogos e inventarios. Dicho estándar tiene en consideración la diversidad cultural y los diferentes contextos educativos en los que los objetos educativos y sus metadatos serán reutilizados generando perfiles de aplicación y vocabularios controlados relacionados con los distintos contextos incluyendo la internacionalización del mismo.

Los OAs son las unidades mínimas de contenido reutilizable, que en la definición de una unidad de aprendizaje pueden agruparse o secuenciarse en el modo que se estime oportuno para conseguir el propósito educativo para el que se destine. En definitiva, los OAs son recursos digitales de cualquier tipo (páginas web, vídeos, animaciones, etc.). Sin embargo, la construcción de OAs hace referencia al uso de estándares para dotar a esos recursos de propiedades que permitan su acceso, agregación e intercambio sin problemas de interoperabilidad. La agregación de los OA para configurar cursos o unidades más complejas se realizará mediante una estructura que permita alcanzar las especificaciones de las diferentes situaciones de enseñanza. Con los estándares actuales existen diferentes modelos que permiten dicha agregación de contenidos a partir de los OAs. Los dos modelos principales son el desarrollado

por AICC (*Aviation Industry CBT Committee*) y el que forma parte de SCORM (*Sharable Content Object Reference Model*).

El Anexo A, hace una revisión detallada de los estándares aquí enumerados para el etiquetado de los contenidos educativos que describen así como analiza sus relaciones.

2.3 REPOSITORIOS DE OBJETOS DE APRENDIZAJE

El término "Repositorio Digital" se utiliza para describir una amplia gama de sistemas que proporcionan la infraestructura para el almacenamiento, conservación, gestión, desarrollo y oferta de todo tipo de contenido electrónico.

Con el objetivo de facilitar la reutilización en el sector educativo están los denominados Repositorios de Objetos de Aprendizaje (ROA) que almacenan recursos educativos y sus metadatos (o sólo estos últimos). Las definiciones existentes de lo que es un ROA [DRI, 2003] [Stockley y Laverty, 2003] [RLG, 2002] [CANARIE, 2001], son muy similares. En general, concretan el ROA especificando de forma clara el modo en el que los repositorios actúan como almacenes de datos o catálogos de recursos. Dichos repositorios están creados para dar soporte a un proceso educativo basado en la reutilización y reorientación de los recursos de aprendizaje con el objetivo de cubrir las necesidades del usuario final [Porter *et al.*, 2004]. Dichos repositorios proporcionan acceso on-line a colecciones de recursos educativos generalmente en formato electrónico. Aunque los repositorios proporcionan un vínculo entre los contenidos y sus metadatos, no necesariamente tienen que ser almacenados ambos en el mismo lugar. Es importante resaltar entonces que un ROA puede indicar las ubicaciones de los objetos almacenados en cualquier otro lugar, tanto en línea como en ubicaciones locales [JORUM+ project, 2004].

Este tipo de sistemas de almacenamiento son por tanto la infraestructura clave para almacenar y administrar dichos recursos educativos, pero fundamentalmente en el objetivo de este trabajo de investigación son las fuentes clave para localizar y recuperar toda información relativa a dichos recursos digitales. Para permitir dichas funcionalidades de búsqueda, localización y recuperación de contenidos un ROA proporcionará algún tipo de interfaz de búsqueda bien con funcionalidades para la interacción directa con humanos a través de interfaces de búsqueda interactiva o bien a través de interfaces que permitan la búsqueda a otros sistemas software, por ejemplo a través de interfaces de consulta mediante servicios web.

2.3.1 Funcionalidades de un ROA

Las funcionalidades básicas proporcionadas por un repositorio se describen detalladamente en el modelo de referencia OAIS (*Open Archival Information Systems*). Este modelo de referencia se define por la recomendación CCSDS 650.0-B-1[OAIS, 2002] del Comité Consultivo de Sistemas de datos Espaciales¹ (*Consultative Committee for Space Data Systems-CCSDS*), siendo su texto idéntico a la norma ISO 14721:2003. El documento detalla cuatro funcionalidades básicas del modo siguiente:

- Volcado de Datos (*Data ingest*): se refiere al método de obtención de datos en un repositorio y pueden incluir los datos de carga (publicación o presentación de los datos de una fuente en el repositorio) y los datos de extracción (recolección de datos iniciada por un repositorio y actuaciones sobre las fuentes de datos).
- Gestión de datos (*Data management*): se refiere a las acciones que tienen lugar dentro del repositorio. A pesar de que no están accesibles directamente a los usuarios del repositorio, los resultados de estas acciones son a menudo visibles (por ejemplo, una imagen de alta resolución puede ser transformado en un formato de resolución más baja para la difusión web).
- Almacenamiento de archivo (*Archival storage*): detalla la infraestructura técnica utilizada para el almacenamiento y recuperación de los activos y los metadatos. Cuando se adoptan estándares abiertos que para la obtención de datos y se provee su acceso (búsqueda / recuperación), debe ser un proceso transparente para los usuarios del repositorio.
- Facilitar el acceso: Describe la forma en que el repositorio expone su contenido a los usuarios finales (por ejemplo, mediante búsquedas y/o interfaces de visualización).

La especificación de IMS DRI (*Digital Repositories Interoperability*) describe las funcionalidades del repositorio como un conjunto de funciones con dos etapas, la primera realizada por el usuario y la segunda por el repositorio: [ADL, 2002] propone un conjunto mínimo de funciones que un repositorio debe implementar para que los OA contenidos sean accesibles desde el exterior en un ambiente seguro. Este conjunto mínimo de funciones está compuesto de: buscar/exponer, reunir/exponer, enviar/almacenar, solicitar/entregar, y alertar/exponer.

Durante la última década se ha realizado un gran esfuerzo para desarrollar tecnologías que permitan catalogar y facilitar la localización de recursos de todo tipo. Los esfuerzos actuales para construir repositorios interoperables

¹ CCSDS <http://public.ccsds.org/>

están generando un elevado volumen de normativas, tecnologías de apoyo y estándares procedentes de variados sectores interesados en su desarrollo.

2.3.2 Soluciones Tecnológicas de los ROA

La tecnología web, en el sector de la educación, ha generado la aparición en los últimos años de muchos repositorios digitales de contenidos educativos, unos pertenecientes a organismos o entidades educativas públicas, otros tantos privados, etc. Por sí mismo, un único repositorio no proporcionaría la base adecuada de conocimientos si se desea contar con un conjunto general de conocimientos. Esto sólo puede hacerse a través de la unión de todos los repositorios digitales posibles para crear una red de conocimiento, generando diferentes infraestructuras de repositorios. Dichas relaciones entre los repositorios derivan en una topología variada de los repositorios distribuidos que va desde una descentralización total, hasta una centralización absoluta, pasando por las posibles situaciones intermedias.

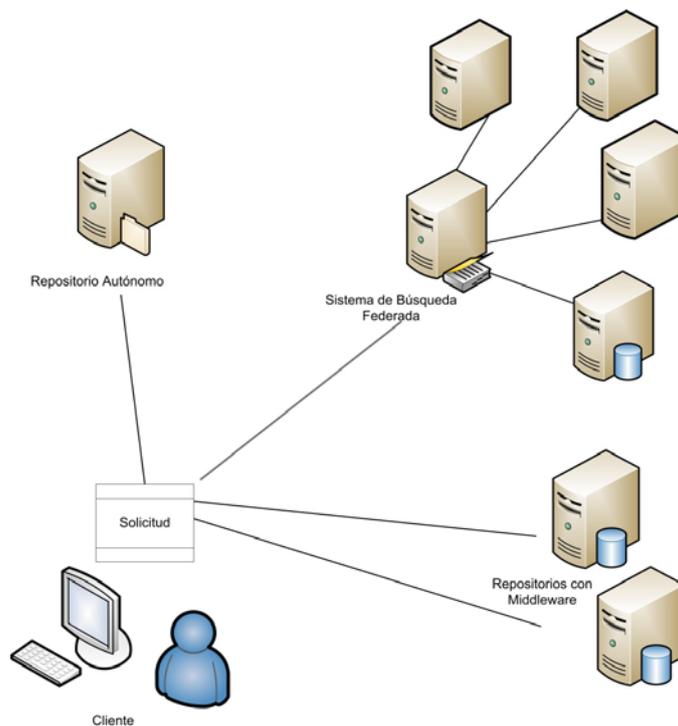


Figura 2-1. Taxonomía de los ROA en función de topología relacional

Muchos son ya los repositorios que utilizan una capa de abstracción, en base a la tecnología que sea, entre el ROA y el sistema cliente evitando que se tenga que tener en cuenta la infraestructura que adopta un repositorio de forma interna, sin embargo, resulta interesante conocer cómo se agrupan los repositorios que implementan dicha capa de abstracción. En este sentido se obtiene la siguiente clasificación (Figura 2-1):

- **Repositorios autónomos.** Aquellos que no tienen ningún sistema que permita una búsqueda externa, por lo tanto requieren búsquedas in situ.
- **Repositorios con *middleware*.** Son aquellos que tienen una interfaz de búsqueda externa y se pueden incluir en un sistema de búsqueda automatizada.
- **Repositorios con sistemas de búsqueda federada.** Son aquellos que además de realizar búsquedas internas también realizan búsquedas automatizadas en otros repositorios.
- **Sistemas de búsqueda federada.** Son sistemas que realizan búsquedas automáticas en otros repositorios.

Para promover este intercambio y la reutilización de objetos de aprendizaje son muchos los proyectos puestos en marcha a lo largo de los últimos años. Proyectos como SMETE², Merlot³, Edna⁴ o Ariadne⁵ han puesto en funcionamiento servicios para que las personas y otros repositorios puedan buscar y recuperar los metadatos de sus repositorios. Actualmente el proyecto GLOBE⁶ (*Global Learning Objects Brokering Exchange*) surgido en el año 2010, aún los mayores repositorios de contenidos educativos a nivel mundial. GLOBE Alliance pretende ser el proveedor único para la gestión de cada uno de ellos incorporando herramientas para la federación sus repositorios de objetos de aprendizaje. Esto implica tanto a las arquitecturas y componentes sobre los que residen los repositorios de objetos de aprendizaje como a los canales de comunicación y sus respectivos protocolos. Ambas partes habilitan la posibilidad de implementar clientes que realicen búsquedas y extraigan dichos contenidos educativos. El apartado siguiente detallará las arquitecturas y los protocolos para la interoperabilidad más importantes.

2.3.3 Estándares para la Interoperabilidad de ROAs

² SMETE <http://www.smete.org/smete/>

³ MERLOT <http://www.merlot.org>

⁴ EDNA <http://www.edna.edu.au/>

⁵ ARIADNE <http://www.ariadne-eu.org/>

⁶ GLOBE <http://www.globe-info.org/>

Existen diferentes perspectivas de arquitecturas y tecnologías que permiten la interoperabilidad de los diferentes ROAs. Entre los más destacados están CORDRA, IMS DRI o IMS *Learning Object Discovery & Exchange* (LODE).

El objetivo del proyecto CORDRA es definir un marco para la federación de repositorios de contenido a través de un registro de metadatos, conocido como Registro ADL (*ADL Registry*). Esto permite mostrar y proporcionar una implementación de referencia de la especificación CORDRA. La documentación del Registro ADL y CORDRA [CORDRA, 2005], consta de cinco volúmenes, parte de los cuales se puede descargar de la página del proyecto⁷. CORDRA está diseñado para ser un puente entre el mundo de los repositorios de contenidos educativos y las bibliotecas digitales. Tiene como objetivo identificar y especificar (no desarrollar) las tecnologías apropiadas y las normas de interoperabilidad que se pueden combinar en un modelo de referencia utilizado para habilitar una infraestructura de contenido de aprendizaje.

IMS DRI (*Digital Repository Interoperability*) [DRI, 2003] describe el modelo de información para la especificación de interoperabilidad de los repositorios digitales IMS. Proporciona recomendaciones para la interoperabilidad de las funciones más comunes de los repositorios. Es uno de los primeros estándares definidos y solamente especifica una estructura a seguir para el desarrollo de sistemas de búsqueda federada entre diferentes repositorios. Se basa en tecnologías de comunicación ya existentes como pueden ser SOAP o Z39.50; y también en especificaciones previamente definidas por este consorcio como IMS Metadata y *Content Packaging* [IMS Content Packaking, 2007]. La principal ventaja de esta arquitectura es que permite el acceso al contenido almacenado en los repositorios a través de aplicaciones externas como Sistemas de Gestión del Aprendizaje (*Learning Management Systems*, LMS), o de gestión de contenidos educativos (*Learning Content Management System*, LCMS) y buscadores específicos. Este tipo de aplicaciones permite realizar búsquedas en varios repositorios de forma simultánea, lo que a priori es un problema complejo, que además se agrava debido a la gran heterogeneidad de la que adolecen los repositorios digitales. Esta especificación introduce un componente intermedio opcional que contiene tres funciones orientadas a simplificar el problema:

- Una función de traducción, que transforma un lenguaje de consulta concreto en otro que sea válido en los repositorios existentes.

⁷

ADL
<http://www.adlnet.gov/Technologies/adlr/ADLRDocuments/ADL%20Registry%20Documentation/ADL%20Registry%20Documentation.aspx>

- Una función de agregación, que permite unir los metadatos de diferentes repositorios para que de esta forma pueda estar disponible para la búsqueda.
- Una función de federación que envía las consultas a diversos repositorios y que recoge las repuestas de estos.

Recientemente de nuevo el IMS ha sacado la especificación denominada LODE [LODE, 2010], del inglés *Learning Object Discovery & Exchange*, Descubrimiento de Objetos de Aprendizaje e Intercambio. Dicha especificación tiene por objeto facilitar el descubrimiento y recuperación de objetos de aprendizaje almacenados en más de una colección. En resumen LODE se basa en los siguientes supuestos:

- Los objetos de aprendizaje son descritos por metadatos tales como IEEE LOM o Dublin Core.
- Varias instancias de metadatos pueden ser necesarias con el fin de describir adecuadamente todos los aspectos de un objeto de aprendizaje (es decir, con el fin de proporcionar la información necesaria para apoyar los casos de uso LODE).
- Los metadatos pueden ser reunidos para crear catálogos de búsqueda de objetos de aprendizaje.
- Consultar dichos catálogos de metadatos es la principal manera de obtener la información necesaria para buscar objetos de aprendizaje, evaluar su utilidad, y recuperarlos.
- Los Catálogos de metadatos se almacenan en repositorios.
- En los repositorios se pueden realizar búsquedas mediante programación mediante el uso de APIs estándares tales como *Simple Query Interface (SQI)* o *Search/Retrieve with URL (SRU)*.
- Los grandes catálogos de grandes pueden ser creados por la recolección (es decir, copia de) los metadatos almacenados en los repositorios utilizando protocolos tales como la Iniciativa de Archivos Abiertos - Protocolo para la recolección de metadatos (OAI-PMH).

El modelo de Datos del Registro de IMS LODE [LODE, 2010] proporciona una forma de describir repositorios, sus colecciones de objetos de aprendizaje y metadatos, y los protocolos que soportan. Esto permite el registro de los repositorios en un registro central que puede ser accedido por los clientes LODE. Al consultar el Registro LODE, un cliente LODE obtendrá descripciones de repositorio conteniendo toda la información necesaria para conectarse automáticamente a los repositorios y obtener acceso a sus colecciones de metadatos. IMS LODE puede ser visto como una especificación resultado de la unión de los perfiles de otros protocolos de uso general. Su finalidad es extender su uso a fin de tener en cuenta los requisitos específicos para el ámbito educativo, en lugar de crear nuevos protocolos. La especificación LODE tiene por objeto facilitar el descubrimiento y recuperación de objetos de aprendizaje almacenados en más de una colección, señalando aspectos

interesantes que no hacen sino poner de relieve el grave problema de la interoperabilidad para los contenidos educativos.

La generación de estándares como los enunciados para la construcción de repositorios de objetos de aprendizaje permitirá en su evolución futura procesos automáticos de búsqueda y recolección de sus metadatos. Esto que a día de hoy es una meta por conseguir ha de permitir obtener los contenidos educativos de manera eficaz. El proyecto LODE pone de manifiesto las muchas lagunas existentes para completar la búsqueda automática de recursos educativos mediante la integración de todo tipo de tecnologías junto a la aparición de nuevos mecanismos de descripción de elementos como son los de los propios repositorios. El Anexo A, describe en más detalle estas arquitecturas para ROA interoperables.

2.4 PROTOCOLOS DE COMUNICACIÓN EN LA INTEROPERABILIDAD DE CONTENIDOS

Los objetos de aprendizaje son descritos por los metadatos almacenados en los repositorios. Los repositorios según hemos visto en el apartado anterior, están basados en servicios que permiten a las distintas partes el acceso a los elementos de las colecciones que albergan. Los usuarios, bien sean humanos o máquinas, interactúan con los servicios a través de blancos u objetivos, que son generalmente las direcciones de red de los servicios (los extremos de servicio). Las funcionalidades de búsqueda y/o recolección de los metadatos son provistas en cada repositorio a través de diferentes protocolos y están sujetos a las políticas de acceso de los repositorios y/o colecciones.

La interoperabilidad se basa en diferentes tipos de protocolos comunes, que controlan la interacción entre los sistemas y permiten que el intercambio de información sea eficiente. Los protocolos habilitados en los repositorios con las funcionalidades de búsqueda y/o recolección en el ámbito del e-learning son principalmente cinco: protocolo Z39.50, *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) [OAI, 2004], el Simple Query Interface (SQI) [SQI, 2005], la búsqueda de recuperación / a través de URL (SRU) [SRU, 2008], y la interfaz de publicación simple (SPI) [SPI, 2008]. Resulta de interés hacer una breve revisión de estos cinco protocolos de comunicación.

El protocolo Z39.50 se refiere al estándar internacional, ISO 23950: "*Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*", y a la norma ANSI Z39.50 / NISO-Z39.50-2003 [Z39.50, 2003]. La norma especifica un protocolo cliente/servidor para buscar y recuperar información de bases de datos remotas. La potencialidad de Z39.50 es que separa la interfaz de usuario en el cliente de los servidores de información, los

motores de búsqueda y bases de datos. Z39.50 ofrece una visión consistente de la información sobre una amplia variedad de fuentes, y ofrece a las implementaciones del cliente la capacidad de integrar la información de una serie de bases de datos y servidores.

SRU es un protocolo de búsqueda estándar basado en XML para realizar consultas y búsqueda en el ámbito de Internet, utilizando CQL (*Contextual Query Language*) [SRU, 2010]. Este protocolo de búsqueda y recuperación utiliza servicios de Internet y la Web para llevar los mensajes entre el usuario y el destino. Fue definido por la Biblioteca del Congreso de EE.UU. sobre la base del protocolo Z39.50 y disponible a través de su página web⁸. La iniciativa SRU busca llevar la funcionalidad Z39.50 a la Web y se basa fielmente en sus consultas y respuestas estructuradas. Por tanto, gran parte de la funcionalidad del protocolo SRU se deriva del protocolo anterior, sin embargo, sólo fueron tomadas las formas más útiles de este, y en una forma simplificada [Hammond, 2010].

Otro protocolo propuesto en el 2002 y actualmente en la versión 2.0 es el denominado *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) [OAI-PMH, 2008]. Este protocolo tiene como principal objetivo proveer un marco de trabajo independiente de aplicación para el “cosechado” o recolección (*harvesting*, en inglés) de documentos o recursos que permita la interoperabilidad entre sistemas. La comunicación de dicho protocolo se hace a través de HTTP y los metadatos a transmitir vía OAI-PMH habrán de estar etiquetados en Dublin Core.

⁸ Z39.50 <http://www.loc.gov/standards/sru/>

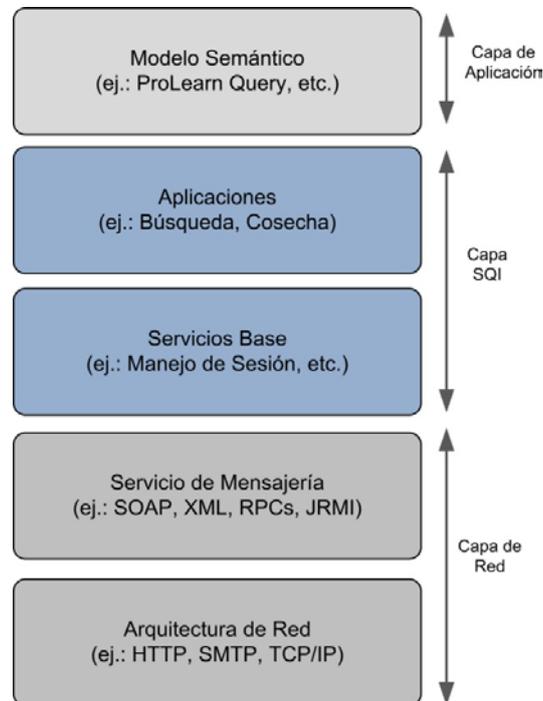


Figura 2-2. Capas definidas por SQI. Adaptada de [CWA 15454, 2005]

Finalmente, el SQI (*Simple Query Interface*), presenta una interfaz de programación de aplicaciones (API) para consultar repositorios de objetos de aprendizaje. La especificación fue aprobada el 6 de septiembre de 2005 por el Comité Europeo de Normalización [CWA 15454, 2005]. Su desarrollo ha reunido a una serie de colaboradores con diferentes áreas de interés, tanto de universidades, como de empresas privadas. El principal objetivo a la hora de diseñar SQI fue el de crear una especificación conceptualmente simple y que a la vez fuera fácil de implementar. El núcleo de SQI es un conjunto de funciones abstractas, conectadas mediante servicios web y para las que no se define ningún tipo de lenguaje de programación concreto.

SQI permite la interoperabilidad entre repositorios y sistemas a través de un servicio de consultas. Este servicio, pese a estar específicamente pensado para un contexto educativo, puede ser utilizado en una amplia variedad de escenarios, ya que se ha diseñado con un alto grado de abstracción, lo que facilita su adaptación. Según describe el estándar SQI, los repositorios deben implementar un conjunto de métodos, considerados el núcleo de la especificación. Estos métodos son la base de un protocolo de comunicación simple que permite el envío de consultas y la recepción de objetos de

aprendizaje. La topología de repositorios para SQI es *peer-to-peer*, es decir, cada uno de los miembros del sistema puede actuar como servidor (*target*) o cliente (*source*).

Para que la interoperabilidad sea posible, es necesario definir tanto un lenguaje de consulta, como un formato de resultados. Una de las características clave de SQI es que no pone ninguna restricción en este sentido, por lo que cualquier lenguaje de consulta o formato de resultados es válido, siempre que lo sea para los dos repositorios que están compartiendo información. Así mismo, tampoco limita otras características que dificultarían la interoperabilidad entre los repositorios como pueden ser: método de almacenaje, búsqueda interna, indización, etc. De modo que el foco del SQI, descrito con detalle en la sección A.4.4, se centra en el conjunto de servicios base para conectar los repositorios o realizar todo tipo de aplicaciones para realizar las búsquedas federadas.

2.5 CONCLUSIONES DEL CAPÍTULO

Diferentes sectores y organismos internacionales se encuentran desarrollando diversos esquemas y estándares para formalizar el uso de metadatos y establecer un conjunto de reglas semánticas, sintácticas y de contenido que pretenden describir homogéneamente a un OA. Su almacenamiento en los repositorios de contenidos educativos configura en la actualidad un campo abierto para posibilitar la interoperabilidad real de dichos contenidos en busca de la reutilización.

Como ejemplo de la importancia y el esfuerzo actual que se está realizando en las tecnologías asociadas, para hacer posible el acceso real a los contenidos digitales educativos, basta revisar el programa eContentplus⁹ financiado por la Comisión Europea (Programa plurianual comunitario de incremento de posibilidades de acceso, utilización y explotación de los contenidos digitales en Europa). Este programa está permitiendo a varias federaciones de repositorios enriquecer y fortalecer sus materiales educativos a través de diferentes proyectos resultantes tales como: ASPECT¹⁰ (*Adopting Standards and Specifications for Educational Content*), MELT¹¹ (*Metadata Ecology for Learning and Teaching*), MACE¹² (*Metadata for Architectural Contents in Europe*), ICOPER¹³ (*Interoperable Content for Performance in a Competency-driven Society*) o

⁹ EContentplus http://ec.europa.eu/information_society/activities/econtentplus/index_en.htm

¹⁰ ASPECT <http://www.aspect-project.org/>

¹¹ MELT <http://info.melt-project.eu/>

¹² MACE <http://portal.mace-project.eu/>

¹³ ICOPER <http://www.icoper.org/>

Share.TEC¹⁴ (*Sharing Digital Resources in the Teaching Education Community*). El eje fundamental de todos estos proyectos es la búsqueda de la interoperabilidad real entre los distintos componentes de los repositorios, un campo aún por conseguir plenamente.

La existencia de diferentes organismos trabajando en la materia tiene como ventaja el rápido avance en el desarrollo de especificaciones perfectamente válidas. Sin embargo esto genera una gran variedad de estándares y especificaciones existentes que no revierten en un estándar universal y por tanto contribuyen a la heterogeneidad de soluciones que se estandarizan.

Es evidente que la descripción de los recursos educativos mediante esquemas de metadatos facilita una gran cantidad de procesos como almacenamiento, búsqueda y recuperación. Aún así, pese a los avances en interoperabilidad y reutilización en los repositorios, se observan una serie de deficiencias:

- Los estándares actuales describen a los recursos mediante un elevado y variado conjunto de características. La completitud y calidad de los OA dependerá sin embargo del autor que anota la información de los mismos (experiencia, conocimiento del estándar, etc.).
- La información expresada por los estándares asociados a los OA, facilita el tratamiento automático de las metadatos. Pero no facilita el tratamiento de relaciones conceptuales entre OAs.
- Los estándares proporcionan facilidades de acceso a la información sobre los contenidos o el formato del OA, pero no incluyen información semántica explícita que pueda ser tratada de manera directa.

Por otro lado, los repositorios educativos contienen una gran cantidad de información sobre los recursos que almacenan, pero habitualmente de forma muy heterogénea debido a la gran variedad de recursos posibles, generando problemas de interoperabilidad entre sistemas. La interoperabilidad de los repositorios de contenidos educativos se basan en diferentes tipos de protocolos, que controlan la interacción entre los sistemas y permiten que el intercambio de información sea eficiente, sin embargo surgen gran cantidad de problemas asociados que dificultan la interoperabilidad real:

- Se necesitan mecanismos automáticos que permitan obtener una descripción clara del tipo de contenidos que almacena cada repositorio y los mecanismos de acceso a los mismos.
- Existen múltiples estándares y arquitecturas para buscar la interoperabilidad en la conexión con los repositorios, en aquellos que se preocupan por habilitar estas capas. Sin embargo los

¹⁴ ShareTEC <http://www.share-tec.eu/>

procedimientos automáticos de selección de repositorios no están aún solucionados al no existir mecanismos que los describan adecuadamente centralizando dicha información. Además aún son muchos los repositorios que ni siquiera permiten la abstracción del modelo de datos haciendo imposible automatizar el acceso a sus contenidos.

En resumen, no existe un estándar que cubra todas las necesidades del proceso búsqueda, recuperación y reutilización de los objetos de aprendizaje en los repositorios.

La tecnología actual no soluciona la búsqueda y recuperación automática o semiautomática de los OAs y por lo tanto, la reutilización de recursos de aprendizaje. Por este motivo se propone una arquitectura flexible y modular, que permita la recuperación inteligente de contenidos digitales educativos residentes en repositorios de objetos de aprendizaje que habiliten capas de interoperabilidad.

El contexto presentado hace necesario el uso de mecanismos de recuperación de contenidos educativos distribuidos en repositorios con distintas arquitecturas, distintos protocolos de comunicación, etc. El próximo capítulo analizará las tecnologías de recuperación de información web, así como su organización y servicios añadidos para facilitar la tarea de selección de contenidos que el usuario necesita. El fin es conseguir para el usuario una solución flexible en el complejo entorno educativo actual, de modo que se centralice de manera transparente el acceso y la recuperación de contenidos educativos dispersos. El próximo capítulo profundizará en los aspectos de la recuperación eficiente de contenido educativo que albergan los repositorios digitales.

3 RECUPERACIÓN EFICIENTE DE CONTENIDOS EDUCATIVOS

El gran crecimiento y desarrollo de la Web, colapsa de información a cualquier usuario interesado en acceder a información de calidad. A esto se suma la necesidad de extraer automáticamente información de distintos repositorios que albergan enormes volúmenes de datos potencialmente valiosos. Es por ello que en la actualidad una de las principales líneas de investigación en el campo de la computación está orientada al estudio y aplicación de técnicas y tecnologías de recuperación y organización de la información. La gran cantidad de contenidos al alcance del usuario en Internet no explota toda su utilidad cuando éste no cuenta con herramientas que recuperen información útil y la muestren organizada.

Gran parte de las necesidades de recuperación de información Web para los usuarios está resuelta con la implementación de algunos buscadores. Sin embargo, las carencias de los mismos plantean nuevas soluciones a la organización de información en cantidades a priori imposibles de manejar para los usuarios.

El término recuperación de información suele hacer referencia a la consulta y extracción de datos tanto estructurados como no estructurados. El énfasis de dichos sistemas de recuperación en el campo de los sistemas de información se centra en problemas como las consultas basadas en términos clave, la relevancia de los resultados para una consulta y el análisis y clasificación de los resultados. Para la recuperación de información y contenidos en la Web, se han desarrollado diferentes estrategias. Entre las más destacables está el uso de metadatos o la utilización de lenguajes semánticos basados en XML para indizar documentos Web y representar el conocimiento incluido en ellos, según hemos visto en el capítulo 2 en el caso de los contenidos educativos.

Algunos sistemas de recuperación de información incorporan técnicas de recuperación basada en la semejanza. En estos casos, el sistema puede valorar la idoneidad para un usuario de un ítem A y pedirle que recupere ítems "semejantes". En otros casos la semejanza se puede realizar en entre los propios usuarios de un sistemas, etc. Surgen así los denominados sistemas de recomendación. Un sistema de recomendación consiste en un tipo de tecnología encargada de filtrar la información de manera personalizada identificando conjuntos de ítems que pueden resultar interesantes a un cierto usuario.

El capítulo establecerá las bases de recuperación de contenido web en la denominada web oculta, identificando aspectos de evaluación de la efectividad de dicha recuperación. Se formalizarán los criterios la medida de eficiencia para la recuperación de contenidos educativos, concretándolos en los OA. Se introducirán los sistemas de recomendación de contenidos, profundizando en las tecnologías de aplicación para la propuesta de la solución planteada en capítulos siguientes junto con una revisión de trabajos relacionados con la recuperación de contenidos educativos.

3.1 RECUPERACIÓN DE CONTENIDOS EN WEB

Las soluciones desarrolladas por los motores de búsqueda o buscadores web convencionales son muy eficaces para recuperar los contenidos visibles de la Web. Hacen uso de rastreadores web, crawlers o arañas, que localizan y reúnen información web. Dichas arañas rastrean los sitios web y siguen de manera recursiva los hipervínculos presentes en los documentos. Los datos que las arañas extraen son tratados de forma diferente por los diversos motores de búsqueda. Dichos buscadores hacen uso del modelo de recuperación centralizada que copia la información recogida en una única base de datos centralizada, indexando los contenidos y ordenan los documentos clasificándolos en base a distintos atributos con sus pesos en la base de datos para las consultas de los usuarios. De modo que determinan el orden en el que la lista de sitios web debe aparecer en los resultados de una determinada búsqueda. Este método funciona bien cuando las fuentes de información dejan expuestos sus contenidos a los rastreadores web. Sin embargo, esto no es válido en la Web oculta, donde la información sólo puede ser accedida a través de mecanismos de búsqueda adaptados a fuentes específicas, como el caso de los repositorios con contenidos educativos.

Existe un gran volumen de contenidos educativos en la Web, que no son directamente accesibles a través de buscadores convencionales. Este tipo de información se dice que pertenece a la denominada Web oculta, profunda o invisible, frente a los contenidos de la web denominada visible.

La recuperación de información distribuida, o búsqueda federada, [Callan, 2000] pretende dar respuestas al problema de la recuperación de información en la Web oculta. El objetivo principal de la búsqueda federada es desarrollar modelos y estrategias para obtener el mayor beneficio de estas fuentes distribuidas. De esta manera, los usuarios perciben el sistema como un único punto de acceso a la información que requieren, independientemente del número de las fuentes que existan, de su localización o mecanismo de gestión. El proceso es totalmente transparente al usuario por lo tanto no percibe la complejidad del mismo.

La aportación fundamental de la búsqueda federada es que el proceso de búsqueda se realiza a través de mecanismos de búsqueda en fuentes de información individuales. Además, dicha búsqueda hace referencia a la ubicación de cada fuente distribuida y establece un control de la información relacionándola con las distintas fuentes de información oculta.

El mecanismo de búsqueda federada es más complejo y exhaustivo que el modelo de recuperación centralizada de los buscadores tradicionales. Lo cual le permite dar una solución real al problema de la web oculta. La búsqueda federada aplicada al ámbito de la recuperación de OA a través de sus repositorios se describe como la secuencia en la resolución de los tres subproblemas siguientes:

1. SELECCIÓN DE LOS REPOSITARIOS, es decir, analizar la descripción de los recursos en los repositorios y estudiar cómo se representa la información que se encuentra distribuida en dichos repositorios. Es muy importante conocer la temática, el área y el contexto a los que pertenece la información almacenada por los repositorios, así como otras tantas características clave de las distintas fuentes de información oculta (tiempos de respuesta, eficiencia, etc.).
2. SELECCIÓN DE RECURSOS, donde a partir de una necesidad de información y un conjunto de descripciones de éstos se recuperan los resultados y se decide cuáles serán los que tengan mayor probabilidad de satisfacer la consulta mediante un *algoritmo de recuperación*.
3. FUSIÓN DE LOS RESULTADOS, consiste en la integración y combinación de los resultados devueltos por las consultas sobre los n repositorios, formando una única lista que presenta al usuario una lista ordenada de resultados. El criterio de ordenación de dichos resultados será resultado de cierto *algoritmo de posicionamiento*, ranking o prioridad en la presentación de los contenidos recuperados.

En el contexto educativo actual existe un importante crecimiento de repositorios de objetos de aprendizaje formando parte de la web oculta en grandes bases de datos. Dichos objetos de aprendizaje pueden estar etiquetados en cualquier estándar (LOM, Dublin Core, etc). Se abre la necesidad de trabajar en el desarrollo de soluciones que permitan la búsqueda y recuperación eficaz de contenidos heterogéneos, haciendo frente al contexto distribuido donde se alojan.

3.2 MEDIDA DE LA EFECTIVIDAD DE LA RECUPERACIÓN

Tradicionalmente, la eficacia de un sistema de recuperación se mide en términos de precisión y exhaustividad. Estos dos parámetros caracterizan la capacidad del sistema para recuperar los documentos válidos y evitar los irrelevantes [Al-Maskari & Sanderson, 2010]. Otras medidas de eficacia se discuten en Korfhage (1997), [Baeza-Yates & Ribeiro-Neto, 1999] o [Järvelin & Kekäläinen, 2000].

La precisión, mide el porcentaje de los documentos recuperados que son verdaderamente importantes para la consulta. La exhaustividad, mide el porcentaje de los documentos importantes para la consulta que se han recuperado.

Otro problema más de la medida de la efectividad de la recuperación consiste el modo de definir los documentos que son realmente importantes y los que no lo son. Así la relevancia mide la proximidad entre un documento y la formulación de una petición o expresión de una necesidad informativa. Una evaluación mecánica de la relevancia se obtendría mediante comparación entre la expresión sintáctica o lógica de la consulta y los términos que reflejan el contenido de los documentos. Sin embargo para evaluar peticiones reales, es necesario determinar en qué medida el servicio satisface las necesidades de búsqueda de información de los usuarios finales [Barry & Chamber, 1998].

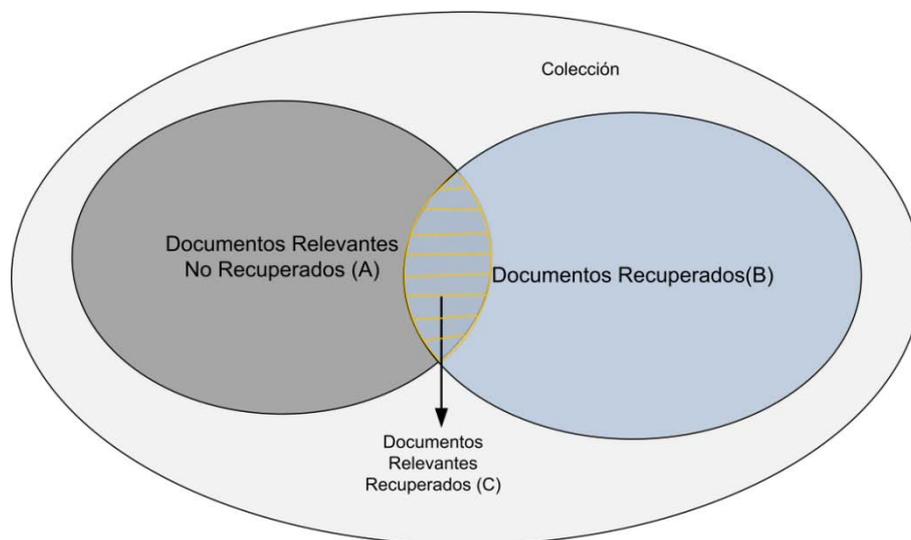


Figura 3-1. Conjuntos de estimación en la recuperación de documentos

Basándose en modelos probabilísticos de recuperación, se define la relevancia como una variable continua que puede adoptar valores intermedios entre 0 y 1. En muchos casos sin embargo, se emplea una escala discreta [Sampath& Prakash, 2009] y en otros se hace un manejo binario del concepto [Benito, 2005].

Dada una colección, fruto de una búsqueda obtendremos una serie de documentos recuperados, algunos de los cuales serán de interés o relevantes y otros muchos que podrían ser relevantes no serán devueltos por el sistema de búsqueda o recuperación, ver Figura 3-1. Basándose en el cálculo de la relevancia, es posible determinar el valor de los índices de exhaustividad (E) y precisión (P) que nos indicarán el rendimiento de la recuperación.

$$E = \frac{C}{C + A} \quad (3.1)$$

$$P = \frac{C}{C + B} \quad (3.2)$$

La medida de la exhaustividad (E) en la ecuación (3.1), calcula la proporción de documentos relevantes recuperados, determinando la bondad del sistema para encontrar documentos relevantes. Por otro lado el cálculo de la precisión (P), en la ecuación (3.2), mide la proporción de documentos recuperados que son relevantes y permite obtener una medida de la eficacia de la búsqueda.

Mientras que el cálculo de la precisión será un valor exacto, el cálculo de la exhaustividad tiene que ser sometido a aproximaciones. En los sistemas reales de recuperación en internet, la determinación del número total de documentos relevantes existentes en la colección se realiza mediante estimaciones aproximadas por distintos métodos, de modo que se habla entonces de exhaustividad relativa.

3.2.1 Medida del Rendimiento en la Recuperación de Contenidos Educativos

La determinación del rendimiento de la recuperación de OA en los distintos ROAs en este trabajo se basa en el cálculo de la precisión y la exhaustividad relativa así como de una nueva medida propuesta denominada ganancia y que tiene en cuenta el tiempo en la recuperación de contenidos educativos de los repositorios de objetos de aprendizaje. Estas medidas de rendimiento de las búsquedas realizadas se hacen en base a la formulación que se describe a continuación.

Dado un conjunto de n metadatos de OA recuperados para un determinado ROA J , el conjunto de dichos archivos de metadatos, vendrá determinado por $O_j = \{O_1, \dots, O_i\}$, con $i = 1, \dots, n$

La **relevancia** se encuentra relacionada con la utilidad o al uso potencial de los materiales recuperados en relación con alcanzar unos objetivos, interés o problemas intrínsecos al usuario [Schamber & Bateman, 1996]. En base a dicho planteamiento general del campo de la recuperación de contenidos, este apartado describe su adaptación en el contexto de este trabajo. Atendiendo a los criterios de los trabajos de [Benito, 2005], los metadatos de los OA recuperados han sido categorizados mediante criterios de relevancia basada en el manejo binario de la misma: $R = \{0,1\}$. Los criterios de selección de los valores binarios en la relevancia son:

- Si el OA no puede ser recuperado porque carece de la información etiquetada indicando la fuente del recurso (atributo <location> de la categoría <technical> de LOM), se califica como irrelevante y se le atribuye un 0
- En cualquier otro caso, se califica como relevante con valor 1

Dicho criterio binario en la relevancia permite elaborar una formulación propia para este estudio en el cálculo de la precisión de los sistemas de búsqueda de objetos de aprendizaje a través de las diferentes consultas en cada repositorio. Dicha formulación es detallada a continuación.

Se define la medida de precisión en la recuperación de OA de un determinado ROA j a través de la ecuación (3.3).

$$P_j = \frac{\sum_1^n R_i O_i}{n} \quad (3.3)$$

Generada una consulta Q en una serie de ROAs, el conjunto total de metadatos de OA recuperados será la unión de todos los metadatos recuperados en los m repositorios: $O = \cup_1^m O_p$. El cálculo de la **Exhaustividad Relativa** para un determinado repositorio J , en la ecuación (3.4), toma como denominador la suma de los OA juzgados relevantes en cada búsqueda comparándolos con los relevantes obtenidos en el total de repositorios.

$$E_j = \frac{\sum_1^n R_i O_i}{\sum_{j=1}^m \sum_{i=1}^n R_{ij} O_i} \quad (3.4)$$

La recuperación de metadatos en los repositorios de objetos de aprendizaje, contempla un escenario con una dinámica muy variable en el tiempo que debe de ser tenida en cuenta a la hora de estimar la medida de rendimientos en cualquier desarrollo aplicado. Para valorar estos aspectos dinámicos se ha realizado un análisis en el contexto de la recuperación de contenidos educativos en ROAs reales descrito en el Anexo B. Dicho estudio apunta cómo la recuperación de contenidos de repositorios distribuidos contempla una gran variabilidad relacionada con la cantidad de recursos con los que cuenta un repositorio y su índice de variación de contenidos en el tiempo, las

características de representación de los recursos, la actualización de contenidos, sus procedimientos de recuperación interna, etc.

Todos estos aspectos dinámicos han de ser contemplados en la formulación de la solución. Más aún cuando resulta especialmente importante la necesidad de atender las demandas de objetos de aprendizaje del usuario en un tiempo razonable. El proceso de recuperación de recursos los educativos, exige que el usuario tenga en un tiempo razonable el mayor número de OA que atiendan a sus requisitos. Esto es así en cualquier sistema de búsqueda en la web. Este requisito funcional hace necesario disponer de alguna métrica o formulación que nos permita evaluar esta característica. No se ha encontrado en la bibliografía ninguna medida que permita establecer este tipo de control sobre los contenidos por lo que se propone una nueva medida con tal fin. Por este motivo se define el término de ganancia temporal como sigue.

*DEF1. Denominaremos **Ganancia Temporal** de un repositorio J , $G_J(t)$, en la ecuación (3.5), a la estimación que relaciona el número de resultados para n consultas que devuelve un repositorio J respecto al tiempo en que éstos son devueltos.*

$$G_J(t) = \frac{\sum_1^n \frac{\text{resultados}_n}{\text{tiempo}_n}}{n} \quad (3.5)$$

Si se desea realizar una comparativa de dicha medida de la ganancia temporal entre el valor medio obtenido del conjunto de repositorios se calculará la medida de la **Ganancia Temporal media**, $\overline{G}(t)$, contemplando los m repositorios estudiados y utilizaremos la fórmula tradicional de la media, en la ecuación (3.6).

$$\overline{G}(t) = \frac{\sum_{j=1}^m G_J(t)}{m} \quad (3.6)$$

Quedan así expuestas las métricas con las que se va a cuantificar el rendimiento de la recuperación de objetos de aprendizaje en repositorios distribuidos a través de la definición de medidas adaptadas del campo de la recuperación de contenidos como son la precisión y la exhaustividad relativa. Se formula una nueva medida que cuantifica aspectos de tiempo en la extracción de contenidos y denominada Ganancia Temporal. A continuación se realiza una evaluación de los indicadores de de rendimiento planteados en la recuperación de contenidos digitales educativos.

3.2.2 *Evaluación del Rendimiento en Recuperación de Objetos de Aprendizaje*

Para estimar los aspectos de exhaustividad y precisión detallados en el apartado anterior así como de la medida propuesta en la Definición 1 se plantea el estudio de la recuperación de consultas en repositorios de objetos de aprendizaje. Para ellos, se han realizado 53 consultas referidas a los términos contenidos en la (Tabla 3-1) sobre los repositorios, LORNET y Merlot.

Tabla 3-1. Tópicos de Consultas realizadas

Nº de Consulta: Tópico	
Q1 Algorithm construction	Q28 Educational theories
Q2 Algorithmic languages	Q29 Fibre optics
Q3 Analog computing	Q30 Formal Systems
Q4 Architectural design	Q31 Formalized languages
Q5 Arithmetic and machine instructions	Q32 Heuristics
Q6 Artificial intelligence	Q33 Hybrid computing
Q7 Bioacoustics	Q34 Informatics
Q8 Biochemical genetics	Q35 Information systems design and components
Q9 Broadcasting, sound and television	Q36 Iterative methods
Q10 Cardiology	Q37 Linear Programming
Q11 Central processing units	Q38 Machine translation
Q12 Character recognition systems	Q39 Microwave links
Q13 Circuit design	Q40 Pediatrics
Q14 Codes and coding systems	Q41 Processing engineering
Q15 Computer architecture	Q42 Programming languages
Q16 Computer Sciences	Q43 Programming theory
Q17 Computer software	Q44 Pulmonary diseases
Q18 Computer technology	Q45 Real-time systems
Q19 Computer terminals graphic display devices and plotters	Q46 Recursive Functions
Q20 Computing for statistics	Q47 Sensor systems design
Q21 Computing systems design	Q48 Simulation
Q22 Control devices	Q49 Speech
Q23 Control Systems	Q50 Storage devices
Q24 Data transmission devices	Q51 Surgery
Q25 Dermatology	Q52 Telecommunications technology
Q26 Digital computing	Q53 Timing devices
Q27 Dynamic Programming	

Los expertos en la recuperación de información apuntan ciertas consideraciones sobre la creación de las consultas, coincidiendo en que deben de ser realizadas con ayuda de expertos en el área [Gordon and Pathak, 1999] [Hawking *et al.*, 2001]. Además según datos de openDOAR¹⁵, el idioma utilizado

¹⁵ OPENDOAR <http://www.opendoar.org/>

con gran diferencia en búsquedas de contenidos digitales es el inglés (77% de las consultas). Estos motivos hacen que en este estudio, los patrones de búsqueda o tópicos se han elegido al azar entre tópicos en inglés de Ciencia y la Tecnología de los Códigos UNESCO¹⁶, elaborados por expertos. Dichos tópicos están compuestos mayoritariamente por conjuntos de palabras (consultas largas) debido al carácter específico de dicha terminología, pero también por palabras únicas (consultas cortas) para evaluar diferentes comportamientos en los sistemas de búsqueda de los repositorios.

Las Figura 3-2 y Figura 3-3, muestran los valores de exhaustividad y precisión de cada uno de los dos sistemas evaluados. El cálculo de la exhaustividad toma como denominador el número total de documentos relevantes atendiendo al criterio de relevancia binaria establecida, identificados en el contexto global de la búsqueda. Las gráficas muestran cómo el repositorio Merlot tiene mayores valores tanto de precisión como de exhaustividad frente al repositorio Lornet. El repositorio Lornet contiene muchos elementos etiquetados, pero cuyos metadatos no permiten recuperar los contenidos educativos y por tanto los hacen carentes de todo valor en cuanto a reutilización automática. Este motivo hace que atendiendo a la formulación descrita en 3.2.1 las medidas de exhaustividad y precisión bajen de manera tan importante en dicho repositorio.

Con el fin de evaluar la variación en el tiempo del número de OA disponibles en los repositorios, así como los tiempos empleados en la extracción de las consultas se realiza una segunda prueba que permita evaluar el parámetro de ganancia temporal definido para el trabajo. En dos momentos de tiempo distantes en varios meses y manteniendo la misma batería de consultas se lanzan a un conjunto de cuatro repositorios relevantes en el sector. La Tabla 3-2, muestra los cálculos de las ganancias temporales para de las consultas realizadas en los dos momentos de tiempo definidos. Dicha tabla incluye el repositorio Agrega del que no se obtuvieron resultados en ninguna toma, así como los resultados del resto de los repositorios.

Tabla 3-2. Resultados de Ganancias Temporales en las dos muestras

Repositorio	M1	M2
Acknowledge	0,001202344	0
Agrega	0	0
Merlot	0,002349006	0,003986412
LORNET	0,00073739	0,002884153
Ganancia Media	0,001717641	0,00107218

¹⁶ UNESCO <http://www.unesco.org/new/en/unesco/>

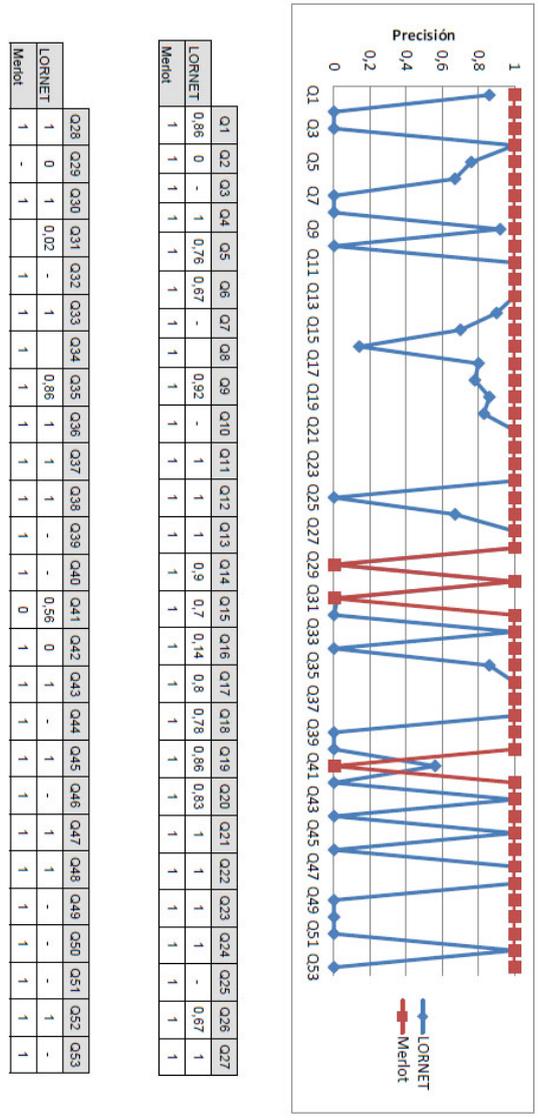
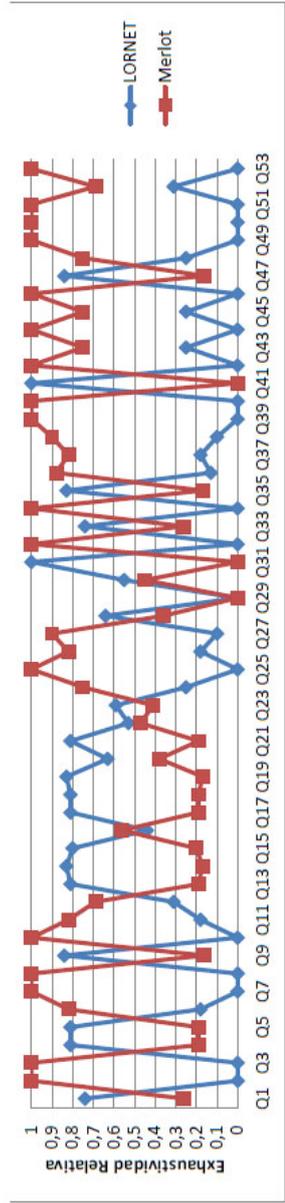


Figura 3-2. Valores de Precisión de los Repositorios Analizados



	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27
LORNET	0,74	0	0	0,81	0,81	0,18	0	0	0,64	0	0,18	0,31	0,81	0,83	0,8	0,44	0,81	0,81	0,63	0,63	0,81	0,53	0,59	0,25	0	0,18	0,1
Merlot	0,26	1	1	0,19	0,19	0,82	1	1	0,16	1	0,82	0,69	0,19	0,17	0,2	0,56	0,19	0,19	0,17	0,38	0,19	0,47	0,41	0,75	1	0,82	0,9

	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	Q41	Q42	Q43	Q44	Q45	Q46	Q47	Q48	Q49	Q50	Q51	Q52	Q53
LORNET	0,64		0,55	1	0	0,74	0	0,83	0,13	0,18	0,1	0	0	1	0	0,25	0	0,25	0	0,84	0,25	0	0	0	0,31	0
Merlot	0,36		0,45	0	1	0,26	1	0,17	0,88	0,82	0,9	1	1	0	1	0,75	1	0,75	1	0,16	0,75	1	1	1	0,69	1

Figura 3-3. Valores de la Exhaustividad Relativa de los Repositorios Analizados

Los resultados de la Tabla 3-2 muestran cómo la ganancia media baja en la recuperación de la segunda carga de consultas, debido a que dos de los repositorios no devuelven resultados en ese momento, sin embargo los otros dos repositorios que sí que devuelven metadatos a las consultas, mejoran el número de resultados devueltos respecto a la misma consulta realizada unos meses antes.

Con la introducción de estas medidas, ganancia temporal de un ROA y ganancia media del conjunto de ROAs contemplados para una consulta, podemos considerar un nuevo mecanismo de medida en la recuperación inteligente de contenidos educativos que permite optimizar en los planes de búsqueda y catalogación de objetos de aprendizaje mecanismos ponderables que permitan la mejora del proceso.

El objetivo de la recuperación está en obtener el mayor número posible de OA relevantes para el usuario que realiza la búsqueda. Este objetivo pasa por alcanzar la mayor ganancia media en cada momento. La ganancia media subirá su valor si se eliminan los accesos a aquellos repositorios que de manera temporal o repetidamente tienen ganancias temporales menores o nulas. Otra cuestión importante es la estimación de la idoneidad de los OA que se recuperan para un usuario. El mecanismo de organizar los objetos de aprendizaje extraídos del conjunto de repositorio deberá de incluirse como un aspecto más en la recuperación y filtrado de los OA relevantes para el usuario. El campo de los contenidos educativos impone nuevas perspectivas sobre el proceso de recomendación en cuanto a aspectos de personalización. El siguiente punto ahonda en los sistemas de recomendación como una solución para establecer una clasificación en los contenidos educativos recuperados que muestre al usuario de manera personalizada el conjunto recuperado.

3.3 SISTEMAS DE RECOMENDACIÓN

Los sistemas de recomendación seleccionan información para el usuario a partir de gustos y preferencias sobre un determinado tema. El área de la tienda virtual fue el primer campo de aplicación de estos sistemas. En la actualidad se encuentran sistemas de recomendación presentes en todos los ámbitos de oferta de servicios a través de Internet tales como música, páginas de noticias, bibliotecas virtuales. Su uso está teniendo igualmente una gran repercusión en todo tipo de sistemas de e-learning [Manouselis *et al.*, 2011]. El objetivo de cualquier sistema de recomendación es el de ofrecer, de forma personalizada, productos adecuados a cada usuario, que así tendrá menos dificultad para encontrar productos de su gusto. En este sentido, los sistemas de recomendación tienen el objetivo de adaptarse a cada usuario. Este trabajo

busca la obtención de una especie de repositorio personalizado para cada usuario a partir de la gestión eficaz de secuencias de búsqueda en diferentes repositorios.

La naturaleza de las recomendaciones permite diferenciar tres tipos de sistemas de recomendación: (i) los que se limitan a recordar al usuario los ítems que le han sido mostrados con anterioridad, (ii) los que les permiten explorar y expandir sus gustos y (iii) los sistemas híbridos derivados de los dos tipos anteriores. El tipo de recomendación empleado en el presente trabajo reside en un sistema que podríamos encuadrar en un tipo híbrido, puesto que recomienda al usuario activo contenidos educativos que han adquirido otros usuarios de su mismo ámbito de interés, pero teniendo siempre en cuenta el historial de los contenidos elegidos por el usuario con anterioridad. La forma de realizar estas recomendaciones combina gran cantidad de técnicas que permiten pronosticar las preferencias de los usuarios en lo que se denomina el mecanismo de personalización.

La personalización engloba una serie de procesos fundamentales e interdependientes:

- **Adquisición de datos del usuario.** Se trata de ampliar y utilizar la información contenida en los ficheros de *log* del sitio web para enriquecer los datos que se tienen acerca de la interacción que realiza el usuario con la plataforma.
- **Construcción de modelos.** Se trata de ampliar la información y las técnicas para la construcción de los modelos que sustentan las tareas de adaptación previstas para el sistema.
- **Identificación de las tareas de adaptación.** Relacionado con los modelos construidos y la definición de tareas adaptativas, trata de identificar para cada tarea de aprendizaje cooperativo que se plantee qué tipo de ayuda puede ser de utilidad para su realización.

Para lograr tal personalización, un sistema de recomendación debe contar con tres elementos principales.

- **Histórico de datos:** Se refiere a la información de que dispone el sistema antes de empezar el proceso de recomendación, la cual engloba datos del perfil de los usuarios, construido a través del comportamiento de los mismos en el sistema.
- **Datos de entrada:** se refieren a datos obtenidos a través de interacciones de acciones (comportamiento) de usuarios con el sistema, a los cuales se les va a realizar la recomendación.
- **Algoritmo:** utiliza algún método que se basa en los datos de entrada y en el histórico de datos para realizar las recomendaciones.

Los sistemas de recomendación se pueden dividir en dos grandes categorías [Lee *et al.*, 2001], [Cheung *et al.*, 2003] los denominados métodos basados en contenido y los basados en filtrado colaborativo.

En el enfoque **basado en contenido** se recomiendan productos por comparación entre sus contenidos y los perfiles de usuario mientras que el **filtrado colaborativo** (*Collaborative Filtering* CF) consiste en la predicción de preferencias de un usuario por determinados productos basadas en las opiniones o acciones de otros usuarios. Dentro de este último grupo podemos establecer a su vez otra clasificación en función de los algoritmos que atienden la colaboración: **Algoritmos basados en memoria** (o en el usuario) que tratan todos los artículos elegidos por los usuarios mediante técnicas estadísticas con el propósito de encontrar usuarios con preferencias similares (vecinos) y **los algoritmos basados en modelos** (o en ítems), los cuales desarrollan un modelo de valoraciones del usuario. Finalmente aparece una tercera categoría, **basada en conocimiento** (*Knowledge-based systems*), como un enfoque híbrido que extiende el filtrado colaborativo junto con el basado en contenido.

3.3.1 Recomendación Basada en Contenido

Para construir el perfil de un usuario, los métodos basados en contenido analizan las propiedades de ítems que el usuario ha visitado, comprado o evaluado anteriormente. Por lo tanto, en este caso los sistemas de recomendación no consideran información adquirida a través de otros usuarios. De esta forma el usuario recibirá recomendaciones de ítems similares a otros por los que él manifestó interés en el pasado. Para realizar una recomendación, dichos métodos, además de obtener datos relativos a propiedades de ítems, necesitan obtener información sobre el comportamiento de cada usuario sobre ellos, la cual se utiliza para construir su perfil. Para ello, se suelen utilizar técnicas basadas en reglas o en el cruce de palabras claves facilitadas por los usuarios, así como técnicas basadas en aprendizaje automático para elaborar los perfiles de forma automática.

Aparte de de construir los perfiles de usuario, es necesario también mantenerlos y actualizarlos, dado que en la actualidad constantemente se añaden nuevos ítems a los sistemas y además las preferencias de los usuarios pueden evolucionar con el paso del tiempo. Para las tareas de actualización también se utilizan técnicas de aprendizaje automático. Por esta razón algunos autores argumentan que los métodos basados en contenido tienen sus orígenes en el área de Recuperación de la Información (*Information Retrieval*), ya que múltiples técnicas que suelen ser utilizadas en tal área también lo son en muchos métodos basados en contenido. Para actualizar los perfiles de usuario, por ejemplo, en el área de la recuperación de la información también se utilizan valoraciones realizadas por los mismos usuarios acerca de ítems considerados (a priori) de su interés por el sistema. En la recuperación de la información se utiliza el término *retroalimentación relevante* (*relevance feedback*) para

referirse al dicho proceso. Las técnicas utilizadas por estos métodos suelen asignar pesos para diferenciar las propiedades de los ítems y de los perfiles de usuario.

3.3.2 Recomendación Basada en Filtrado colaborativo

El término filtrado colaborativo, fue utilizado por primera vez por Goldberg *et al.* [Goldberg *et al.*, 1992] y describe un tipo de algoritmos que, en lugar de buscar el contenido, utilizan datos sobre las preferencias de los usuarios sobre ítems para generar recomendaciones para el llamado usuario “activo”. Estos sistemas requieren que el usuario exprese sus valoraciones (*ratings*) sobre diferentes productos recomendables. A partir de dichas valoraciones se calculan coeficientes de correlación que muestran similitudes entre usuarios. Algunos sistemas populares son *Video Recommender* [Hill *et al.*, 1995], *GroupLens research system* [Konstan *et al.*, 1997] y *Ringo* [Sarwar *et al.*, 2001].

Para facilitar algún tipo de recomendación relativa a los ítems disponibles en un sistema, se necesita disponer de información sobre el interés de los usuarios en dichos ítems. Para ello, los sistemas de recomendación hacen uso de valoraciones (u opiniones) de usuarios. En la literatura se suele categorizar las valoraciones de acuerdo a como son obtenidas: de forma explícita o implícita. La forma explícita también puede denominarse realimentación activa (*active feedback*) y la forma implícita, realimentación pasiva (*passive feedback*). En la forma explícita las valoraciones se obtienen directamente de los usuarios. Dichas valoraciones se expresan mediante los llamados *ratings*, en los cuales el usuario asigna algún tipo de puntuación relativa a los ítems. En esta forma de valoración, el usuario expresa su deseo a través de un número discreto. Asimismo, el usuario puede añadir algún tipo comentario acerca de los ítems que fueron valorados, pues de esta manera se permite que otros usuarios tengan algún parámetro en relación a ítems de su interés. En la Figura 3-4 se puede observar cómo el repositorio MERLOT¹⁷ recoge un perfil de sus usuarios registrados para así recabar información sobre ellos y administrar servicios relacionados con recomendaciones a sus materiales.

Alternativamente, en la forma implícita las valoraciones son obtenidas de los usuarios de manera indirecta. Para ello, se suele analizar *timing logs*, registro de compras, registros de descargas, registros de navegación, etc. De esta forma, las valoraciones son obtenidas de manera subjetiva, pues se puede suponer, por ejemplo, que permanecer mucho tiempo leyendo la descripción de un ítem significa que el usuario tiene interés por tal ítem, si se descarga un contenido es evidente su interés, etc. Por lo tanto, se necesita definir métodos para

¹⁷ MERLOT <http://www.merlot.org>

interpretar el comportamiento del usuario adaptando dichos comportamientos a cada campo concreto.

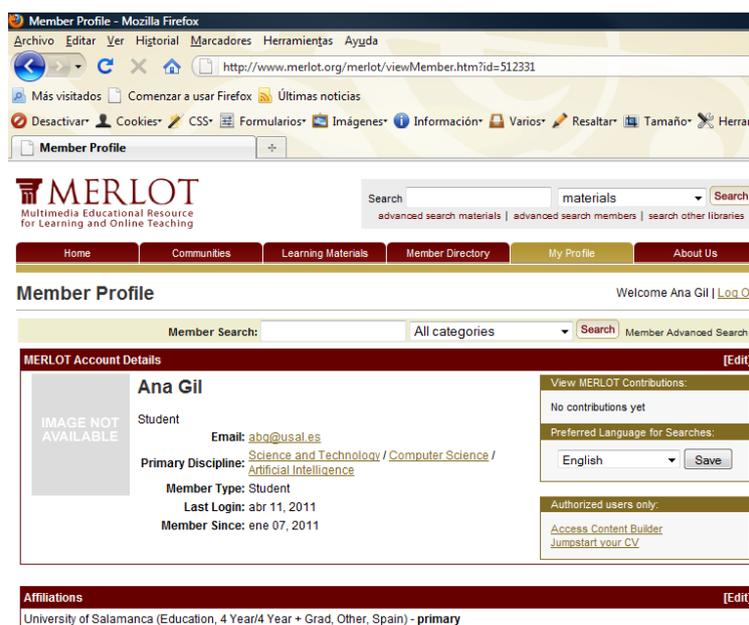


Figura 3-4. Toma de datos del usuario para recomendación en un sitio web

Las valoraciones implícitas tienen la ventaja de no encargar al usuario la labor de valorar ítems, sin embargo, la obtención de dichas valoraciones suele ser más costosa y éstas no siempre son eficaces debido a que son obtenidas de forma automática y subjetiva. Por lo tanto, en la mayoría de los sistemas de recomendación actuales todavía predominan las valoraciones explícitas.

Formalmente, el filtrado colaborativo en sistemas de recomendación puede ser definido a través de un escenario en donde haya un conjunto de m usuarios $U = \{u_1, u_2, \dots, u_m\}$ y un conjunto de n ítems $I = \{i_1, i_2, \dots, i_n\}$. Además, cada usuario u_i posee una lista de k valoraciones que ha realizado sobre un conjunto de ítems I_{ui} , donde $I_{ui} \subseteq I$. En este contexto, una recomendación realizada al usuario activo $u_a \in U$, consiste en un conjunto de N ítems $I_r \subset I$ de los cuales se predijo que el usuario activo tendría interés en ellos. Es importante decir que $I_r \cap I_{u_a} = \emptyset$, pues no se desea que sean recomendados ítems que el usuario ya haya adquirido. Para almacenar las valoraciones realizadas por los usuarios, los sistemas de recomendación suelen utilizar una matriz de valoraciones. En la Figura 3-5 se muestra una representación de tal estructura, donde se verifica si el usuario u_a podrá tener interés por el ítem i_j .

	i_1	i_2	...	i_j	...	i_n
u_1						
u_2						
...						
u_a						
...						
u_m						

Figura 3-5. Matriz de valoraciones

En los métodos de filtrado colaborativo las recomendaciones son realizadas por medio de predicciones del grado de interés del usuario activo en relación a determinados ítems. Dichas predicciones se basan en las preferencias del grupo de usuarios más semejante al usuario activo. En tal proceso se utiliza algún coeficiente de correlación, el cual expresa el grado de similitud entre dos usuarios. Como se ha comentado anteriormente, los métodos de filtrado colaborativo pueden clasificarse en dos tipos: los *basados en memoria* y los *basados en modelos*. Tal clasificación se realiza según la manera con la que se realizan las predicciones acerca del interés del usuario activo. En los métodos basados en memoria las predicciones son concebidas utilizando todas las valoraciones disponibles en el sistema, mientras que en los métodos basados en modelos se utilizan parte de ellas para construir un modelo de estimación de valoraciones.

Los *métodos basados en memoria* también son conocidos como métodos del vecino más cercano [Sarwar *et al.*, 2001], por ser este método el primero en ser empleado en el filtrado colaborativo. Además, debido a que inicialmente los métodos basados en memoria constituían la única metodología empleada en el filtrado colaborativo, muchas veces estos métodos son referidos como de filtrado colaborativo. Sin embargo, existen otros métodos basados en memoria más recientes y existe otra categoría de métodos que se emplea en el filtrado colaborativo.

Para realizar recomendaciones, estos métodos necesitan utilizar todos los registros presentes en la matriz de valoraciones, pues se necesita confrontar las valoraciones vinculadas al usuario activo con las valoraciones de tal matriz para buscar usuarios con preferencias semejantes. Por esta razón, los métodos basados en memoria también suelen ser llamados de métodos basados en el usuario.

Para proporcionar recomendaciones a un usuario, la primera tarea a desempeñar por algoritmos de esta categoría es la computación de la similitud entre los usuarios del sistema. El coeficiente de correlación de Pearson, es el método más utilizado para cumplir tal cálculo. La ecuación (3.6) expresa la forma de calcular tal coeficiente para dos usuarios (a, i) , donde $v_{a,j}$ expresa una valoración realizada por el usuario activo a sobre un producto j , \bar{v}_i el promedio de las valoraciones del usuario i y \bar{v}_a el promedio de las valoraciones del usuario activo.

$$w(a, i) = \frac{\sum_j (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{(v_{a,j} - \bar{v}_a)^2 \sum_j (v_{i,j} - \bar{v}_i)^2}} \quad (3.6)$$

En este contexto, el promedio de las valoraciones de un usuario representa un resumen de las preferencias del mismo. Tal promedio de valoración del usuario i sobre un producto j es calculado a través de la ecuación (3.7)

$$\bar{v}_i = \frac{1}{|I_i|} \sum_{j \in I_i} v_{i,j} \quad (3.7)$$

Después de disponer de los coeficientes de similitud, el algoritmo obtiene una lista de los n usuarios $U = \{u_1, u_2, \dots, u_n\}$ más similares son al usuario activo u_a , donde u_1 es el usuario más próximo a u_a , u_2 el segundo más próximo y así sucesivamente.

Una vez obtenidos los vecinos más cercanos al usuario activo, u_a , ya se puede realizar la predicción referente a un determinado ítem j , la cual define si tal ítem va a ser recomendado o no. Dicha predicción de recomendación se obtiene a partir de la suma ponderada de las valoraciones de otros usuarios mediante la ecuación (3.8), donde κ es un factor de normalización.

$$p_{a,j} = \bar{v}_a + \kappa \sum_{i=1}^n w(a, i)(v_{i,j} - \bar{v}_i) \quad (3.8)$$

La principal ventaja de los métodos basados en memoria es la rápida incorporación de información reciente, sin embargo, un grave problema inherente a estos métodos se refiere a la falta de disponibilidad de las valoraciones suficientes para realizar las predicciones, denominada en la bibliografía como problemas de arranque en frío (*cold start*). Además, la escalabilidad puede verse comprometida debido al volumen de datos que debe ser procesado cada vez que se tiene que realizar una recomendación a un usuario.

Los **métodos basados en modelos** fueron concebidos con el objetivo de solventar las limitaciones de los métodos basados en memoria comentadas anteriormente. Para ello, en dichos métodos se propone la construcción de un modelo de estimación de valoraciones. Tal modelo se crea utilizando parte de la base de datos de valoraciones, pues en los sistemas de recomendación actuales, en general, no se dispone de un número significativo de valoraciones frente a la cantidad de ítems que se ofrecen, por lo que difícilmente se dispone de una base de datos completa, con valoraciones de usuarios acerca de todos ítems.

Los métodos basados en modelos también son conocidos como métodos “basados en ítems”, pues éstos, a diferencia de los métodos basados en memoria, para predecir la valoración acerca de un determinado ítem también consideran la similitud de tal ítem con aquellos ítems que el usuario ya haya valorado. Para ello, en general se utilizan los mismos tipos de métrica para calcular similitudes.

Después de disponer de los datos relativos a la similitud de los ítems presentes en un sistema, se construye un modelo de estimación de valoraciones que también puede considerar atributos del usuario (edad, profesión, datos demográficos...). Dicho modelo se crea *off-line*, es decir, con anterioridad a la entrada del usuario activo en el sistema. Por lo tanto, se realiza el aprendizaje de un modelo de comportamiento de usuario con datos almacenados por el sistema de modo *off-line* y posteriormente tal modelo es aplicado para predecir las preferencias de un usuario cuando éste está usando el sistema. Para ello se suelen utilizar técnicas aprendizaje automático.

La principal ventaja de esta categoría de métodos es la simplicidad y velocidad con la que se realizan las recomendaciones, pues la mayor parte del procesamiento de datos se realiza *off-line*, por lo que el tiempo utilizado en la construcción del modelo no repercute en el tiempo de respuesta al usuario. Tienen, en cambio, la desventaja de que la información nueva no se incorpora de manera inmediata al modelo, sino sólo cuando se genera un nuevo modelo incluyendo la nueva información disponible. Por tanto, dichos métodos son más indicados para sistemas en los cuales las preferencias de sus usuarios cambian de forma lenta en relación al tiempo necesario para construir el modelo.

Las limitaciones inherentes a los métodos utilizados en los sistemas de recomendación repercuten en errores presentados a los usuarios, los cuales se clasifican en dos categorías: los falsos negativos y los falsos positivos. El primero se refiere a ítems que no fueron recomendados y que sí podrían interesar al usuario activo. El segundo se refiere a ítems recomendados de forma equivocada, donde uno o más ítems recomendados no son interesantes al usuario. Los falsos positivos son los errores más críticos, pues suelen llevar más fácilmente a insatisfacción del usuario.

Probablemente la limitación más significativa que se presenta actualmente en los sistemas de recomendación está relacionada con la dispersión (*sparsity*) de datos ocasionada debido a la inmensa cantidad de ítems disponibles en los sistemas de recomendación actuales. Los usuarios de sistemas de comercio electrónico, en general, pueden haber adquirido apenas el 1% del total de productos disponibles. Tal limitación es más comprometedor en los métodos de filtrado colaborativo basados en memoria, pues puede resultar inviable obtener una cantidad suficiente de valoraciones a partir de usuarios de un sistema. Además, las valoraciones solamente pueden ser aplicadas en contextos donde la información sea de un dominio homogéneo. Los métodos basados en modelos minimizan las dificultades debidas a la dispersión, sin embargo necesitan disponer de un cierto número, aunque pequeño, de valoraciones para posibilitar la construcción de un modelo de estimación de valoraciones.

Otra limitación ocasionada debido al excesivo número de ítems disponibles en los sistemas de recomendación se refiere a la escalabilidad (*scalability*), la cual puede volverse un factor comprometedor del rendimiento de un sistema, pues la búsqueda de los vecinos más cercanos, por ejemplo, puede ser inviable en sistemas que poseen bases de datos de gran volumen. El rendimiento es una propiedad primordial en los sistemas de recomendación, pues necesitan proporcionar respuestas a sus usuarios de forma rápida. Tal limitación no se presenta en métodos basados en modelos, pues en estos métodos, a diferencia de los demás, el procesamiento de los datos no se realiza en tiempo de ejecución.

Aunque las limitaciones mencionadas anteriormente puedan ser minimizadas a través del uso de métodos basados en modelos, existen limitaciones que también se presentan en estos métodos. El problema de la “primera valoración” (*first rater*) es un ejemplo de limitación común a todos los métodos de filtrado colaborativo. Tal problema se refiere a la imposibilidad de ofrecer recomendación acerca de un producto que recientemente fue incorporado a un sistema y que, consecuentemente, posee pocas evaluaciones de los usuarios. De forma análoga, el problema también ocurre con un usuario nuevo del sistema.

Otra limitación que ocurre solamente en los métodos de filtrado colaborativo es relativa al problema de la “oveja negra” (*gray sheep*). Dicha limitación se refiere a los usuarios cuyas preferencias no son compatibles con las preferencias de ningún grupo. Como consecuencia, dicho usuario no va a recibir recomendaciones. Dado que en los métodos basados en contenido para establecer las recomendaciones no se consideran las preferencias de otros usuarios del sistema, dicho problema no ocurre en esta categoría de métodos. Además, el problema de la primera valoración tampoco ocurre en dicha categoría de métodos, pues estos pueden facilitar recomendaciones basándose solamente en las propiedades de un ítem. Por otra parte, por no considerar el contexto social del usuario, las recomendaciones facilitadas por los métodos basados en contenido se limitan a ofrecer ítems similares a aquellos que el

usuario ha valorado positivamente, por lo que múltiples falsos negativos podrían ocurrir.

Una nueva aportación a este tipo de algoritmos proviene de la filosofía de colaboración y sus herramientas en la Web 2.0. a través de la información extraída de los gestores de referencias sociales o el uso de las llamadas folksomías. Estos son sistema de marcadores sociales en los que los usuarios guardan una lista de recursos de Internet que consideran útiles. La mayoría de estos servicios de marcadores sociales permiten que los usuarios añadan marcadores asociados a determinadas etiquetas o 'tags', lo que permite después hacer búsquedas y recomendaciones asociadas a dichas etiquetas. Entre ellos se encuentran servicios como CiteULike¹⁸, Conotea¹⁹, Bibsonomy²⁰, 2Collab²¹ para publicaciones científicas o, los más genéricos como delicious²² y flickr.com, etc.

Cada categoría de métodos referida anteriormente tiene sus ventajas y desventajas. Por lo tanto, los sistemas de recomendación actuales suelen combinar técnicas de más de una categoría de métodos, pues, de esta forma, se intenta que las limitaciones se eliminen mutuamente.

Con la tecnología disponible actualmente, en ciertos dominios todavía es muy costoso extraer propiedades relevantes de los ítems de manera automática o semiautomática. Además si se extrae de la web, ésta contiene más información que la que contienen sus documentos sin más. El uso de recomendadores en el ámbito académico tiene una clara aplicación: proporcionar acceso personalizado a los contenidos que albergan los muchos repositorios educativos. Es necesario estudiar el contexto de los contenidos en web y establecer las reglas de su consumo, sus accesos, sus relaciones con otros contenidos, etc. De los contenidos educativos se pueden extraer algunos rasgos que permiten su clasificación y tratamiento. Normalmente el gran número de contenidos educativos disponibles al usuario dificulta el acceso a aquellos que mejor se adaptan al conocimiento del estudiante y a los objetivos o preferencias de cada uno. Las técnicas de recuperación de la información actuales profundizan en aspectos multimedia como por ejemplo texto embebido en imágenes, los cuales pueden influenciar en el interés del usuario en un ítem,

¹⁸ CITEULIKE <http://www.citeulike.org/>

¹⁹ CONNOTEA www.connotea.org/

²⁰ BIBSONOMY <http://www.bibsonomy.org/>

²¹ 2COLLAB <http://www.2collab.com>

²² DELICIOUS <http://delicious.com/>

aspectos semánticos, estéticos, étnicos, etc. El punto siguiente profundizará en tecnologías relevantes de este tipo de técnicas utilizadas en el trabajo.

3.4 TECNOLOGÍAS DE SOPORTE A LOS SISTEMAS DE RECOMENDACIÓN

Existen diferentes metodologías que permiten habilitar mecanismos de filtrado de la información atendiendo a aspectos relevantes del usuario y al contexto. Los sistemas que incorporan esta capacidad de filtrado incrementan sus capacidades con aspectos de personalización permitiendo evaluar y seleccionar los contenidos disponibles a la medida del usuario. Para realizar esta tarea dichos sistemas deben de aprender de lo que el usuario considera útil, interesante o preferible, para lo cual necesitan estar en contacto con el usuario. Además el filtrado de información personalizada debe de ser flexible a los cambios en los intereses del usuario en el tiempo estableciendo criterios dinámicos en el tratamiento de los contenidos. Debido a esta naturaleza dinámica del contexto, este tipo de sistemas deben de aprender del entorno para el que se desarrolla y adaptarse. Los procesos de aprendizaje y adaptación de este tipo de sistemas se basan en el descubrimiento del patrón de los usuarios a través de sus transacciones o interacciones en línea con los contenidos. Los métodos de aprendizaje se implementan en los sistemas de filtrado de información a través de mecanismos de razonamiento automático y mecanismos de aprendizaje procedentes del campo de la inteligencia artificial mayoritariamente. El **aprendizaje** y la **adaptación** de los sistemas de filtrado de contenidos son aspectos fundamentales en la personalización de la información recuperada. Los sistemas de recomendación se basan en mecanismos de aprendizaje automático sobre experiencias anteriores del usuario en el contexto. Entre las diversas tecnologías para el aprendizaje automático, este trabajo hace uso del Razonamiento Basado en Casos (CBR) como un paradigma para el aprendizaje y el razonamiento a través de la experiencia. Los modelos CBR permiten la resolución de nuevas situaciones mediante la adaptación de soluciones y su personalización a través de razonamiento automático sobre soluciones dadas a problemas anteriores.

A su vez la minería de datos ofrece mecanismos que permiten el descubrimiento semi-automático de patrones, asociaciones, cambios, anomalías, reglas y estructuras significativas en los datos. De modo que técnicas minería de datos pueden ayudar a extraer el conocimiento de los datos y permiten obtener conocimiento off-line. Resulta adecuado valorar aspectos de minería de datos aplicada a la recuperación de información multietiquetada. Los siguientes apartados profundizan estos conceptos.

3.4.1 CBR aplicados a la Recomendación

El razonamiento basado en casos (CBR) es un tipo de mecanismo lógico, utilizado en el pensamiento humano, en el que se recurre a experiencias pasadas para resolver nuevos problemas [López de Mántaras y Plaza, 1997]. Así pues, si en un momento pasado se decidió resolver un problema utilizando una determinada solución (o grupo de soluciones) y, una vez aplicada esa solución se obtuvo un determinado resultado, entonces parece lógico que si se presenta un nuevo problema con características similares al resuelto en el pasado, se recurra a la experiencia adquirida en el pasado para dar solución al nuevo problema. Si el resultado obtenido tras aplicar la solución fue bueno, entonces parece lógico aplicar ahora una solución similar para resolver el problema actual. Por el contrario, si el resultado que se obtuvo no fue bueno, entonces la opción lógica debe de ser la de modificar la solución que aplicada en el pasado buscando obtener un resultado mejor. Se trata pues, de razonar a partir de recuerdos. Una vez finalizado el proceso, la persona ha realizado un aprendizaje de la nueva experiencia vivida. Hay que tener en cuenta dos aspectos del modelo de razonamiento basado en la experiencia. En primer lugar, el modelo se basa en la idea de que problemas similares tienen soluciones similares. Sin embargo, carecer de problemas similares no supone que el sistema no sea capaz de proponer buenos resultados, sino que la reutilización de memorias pasadas se convierte entonces en un proceso creativo. Sea cual sea el resultado de este proceso creativo, el individuo aprende de la nueva experiencia. En segundo lugar, en el proceso de razonamiento explicado se está emitiendo un juicio de valor que permite saber si la solución aplicada para resolver un determinado problema fue buena o no fue buena. Si este juicio es emitido por un experto en la materia del problema resuelto, mayores serán las posibilidades de incrementar la capacidad de aprendizaje [Schank, 1982].

El término CBR es muy general, y existen diferentes tipos, especialmente atendiendo a la representación, indexación o a los mecanismos de razonamiento aplicados sobre los casos. Típicamente un sistema CBR propiamente dicho viene caracterizado por el concepto de caso. Un caso debe contener cierto nivel de información y cierta complejidad en su estructura interna (con lo que, por ejemplo, un simple vector no sería considerado un caso). Además, este tipo de sistemas son capaces de adaptarse a distintos entornos o contextos, facilitando la generalización y cierto grado de independencia del entorno [López de Mántaras, 2001].

Los sistemas CBR diseñan su estructura sobre el concepto de caso. Para [Schank, 1982] un caso es un pedazo de conocimiento contextualizado que representa una experiencia, de tal forma que un caso contiene:

- ❖ el problema que describe el estado del mundo cuando ocurrió el caso,
- ❖ una descripción de la solución encontrada y/o

- ❖ un resultado que describe el estado del mundo después de que ocurrió el caso.

Un caso se puede ver como una lección aprendida cuando se ha resuelto un determinado problema. Los casos pueden representarse de diversas formas, como instancias, como vectores, como objetos, etc. Un caso se representa por medio de características llamadas índices. Los índices podrán ser restricciones, objetivos, solución, fracasos, etc. Según [Kolodner, 1993] un caso se compone de una descripción del problema, de la solución para ese problema y del resultado obtenido tras aplicar la solución. De esta forma se puede utilizar una notación matemática en la que un caso se represente por medio de una 3-tupla, $\langle P, S(P), R \rangle$, y es la que se muestra a continuación.

Case: \langle Problem, Solution, Result \rangle

Problem: initial_state

Solution: sequence of \langle action, [intermediate_state] \rangle

Result: final_state

Así pues, un caso queda definido a través de tres elementos:

- ❖ P o Problema: Representa una descripción del problema que se desea resolver.
- ❖ S(P) o Solución: Es la solución del problema P, y viene representada a partir del conjunto de operadores que se utilizan para construir la solución.
- ❖ R o Resultado: Es la eficiencia, que mide los recursos utilizados para alcanzar la solución.

3.4.1.1 Estructura de un CBR

Es necesario considerar una estructura que permita modelar un sistema CBR, de tal forma que se describan los componentes principales del sistema y las relaciones que se establecen entre ellos. [Riesbeck y Schank, 1989] proponen que el sistema debe estar formado por dos componentes fundamentales: la memoria de casos y el mecanismo de razonamiento (que a su vez puede necesitar de la inclusión de una base de conocimiento).

El **mecanismo de razonamiento** o modelo de actuación de un sistema CBR, se conoce como su ciclo de vida. En el ciclo de vida de un sistema CBR se especifican los pasos ordenados y relacionados por tiempo mediante los cuales se extrae y aprende información para resolver un problema específico. El ciclo de vida de un sistema CBR está formado por cuatro procesos secuenciales, conocidos como las cuatro "erres": "*Retrieve*" (Recuperación), "*Reuse*" (Adaptación), "*Revise*" (Revisión) y "*Retain*" (Retención o aprendizaje) [López de Mántaras *et al.*, 2005].

- i. Etapa "*Retrieve*" (Recuperación). En esta etapa primera se realiza la recuperación de casos, esto es, el acceso a los casos almacenados que cuentan con una descripción del problema más similar a la del actual. En esta etapa se llevan a cabo dos funciones distintas: acceder a los

casos almacenados y establecer la similitud entre casos (más concretamente entre descripciones de problema). De esta forma es necesario plantearse:

- Algoritmo de acceso a los casos almacenados: preferiblemente debe tratarse de un algoritmo que garantice un acceso rápido y eficiente.
 - Técnicas o métricas que permitan determinar la similitud entre casos: a aquellos casos a los que se accede mediante el algoritmo de acceso se les aplica una métrica de similitud que permita determinar cuál o cuáles de ellos son los mejores casos (los más similares al problema actual) [Golding y Rosenbloom, 1988].
- ii. Etapa “*Reuse*” (Reutilización o Adaptación). Una vez finalizada la etapa de recuperación, el sistema CBR pasa a ejecutar la etapa de reutilización. Esta etapa recibe como entradas los casos más similares recuperados durante la etapa anterior. La reutilización o adaptación consiste en trabajar con las soluciones correspondientes a los casos más similares recuperados en la etapa anterior para poder obtener una solución para el problema actual. Trabajar con las soluciones significa modificarlas y combinarlas, o simplemente decidir cuál de ellas es la más óptima y reutilizarla. Al igual que ocurría en la etapa anterior existen muchas posibilidades para realizar la adaptación. Lo ideal es encontrar un caso con una descripción de problema idéntica al actual. Pero, dado que en el mundo real es muy complicado que se presenten dos situaciones idénticas, se hace necesario crear una nueva solución basándose en las soluciones similares de las que se dispone. Además de los datos proporcionados por los casos similares recuperados se hace necesario utilizar algún tipo de conocimiento. El conocimiento puede venir dado a través de fórmulas o reglas.
- iii. Etapa “*Revise*” (Revisión). Al igual que ocurre con la etapa de recuperación, el proceso de evaluación de la bondad de la solución finalmente aplicada para resolver el problema actual puede ser problemático, puede ser necesario de un conocimiento considerable [Leake *et al.*, 2000]. Se comprueba si la solución propuesta en la etapa anterior es apropiada para el caso actual. Para ello se utiliza un sistema experto de conocimiento, o bien una persona experta. El resultado de esta etapa es un nuevo caso, para el que se haya obtenido una solución satisfactoria o una solución incorrecta y deba ser reparado. En ocasiones, en esta etapa se puede realizar una reparación de los fallos o errores detectados.
- iv. Etapa “*Retain*” (Retención y Aprendizaje). En esta última etapa se aprende a partir de la nueva experiencia adquirida. Para ello se almacena el caso actual y la solución aplicada para resolverlo. Además se tiene en cuenta el resultado obtenido en la etapa de revisión para asignar una eficiencia al caso. De esta forma el caso puede ser indexado

en la memoria de casos. Puede ser necesario reorganizar la memoria de casos. En caso de similitud con otros casos se puede aplicar generalización. En ocasiones durante esta etapa, o bien como una etapa adicional previa a la retención y aprendizaje se incluye una revisión de la base de conocimiento. En el caso de que se estén utilizando reglas u otro tipo de conocimiento experto, el conocimiento es revisado y actualizado en función de los resultados obtenidos para la experiencia que se acaba de vivir.

El segundo de los componentes del CBR, la **base de casos o memoria de casos**, es una de las partes más importantes de la estructura de un sistema CBR. De ella se extraen las soluciones anteriores y en ella se almacena lo aprendido. La memoria de casos es la encargada de mantener la representación y organización de estos. La base de casos debe tener en cuenta la estructura de los casos (representación de los casos), y tratar de facilitar en la mayor medida de lo posible cada una de las operaciones que se realizan en el ciclo CBR. Para facilitar dichas operaciones es necesario contemplar tanto la indexación y organización de los casos (asignación de índices para facilitar su recuperación) como el mantenimiento [Leake *et al.*, 2000]. Los modelos actuales tienden a incorporar tendencias como: (i) la representación de casos mediante objetos [Bergmann y Stahl, 1998]; para Bergmann [Bergmann *et al.*, 2005] se tiene representación orientada a objetos, utilizando los conceptos de clases y objetos; (ii) representación mediante XML (CBML), un estándar que permite la representación de casos utilizando XML; (iii) representación textual, necesaria cuando los datos con los que se trabaja son textuales, es decir, se trabaja con palabras o frases; (iv) representación jerárquica, la cual tiene en cuenta distintos niveles de abstracción para representar un caso; (v) representación en aplicaciones para procesamiento de imágenes, una representación especial para los casos en la que la definición del problema de la interpretación de imágenes incluye trabajar con informaciones indirectas (tipo y parámetros del sensor, luminosidad, información sobre objetos, etc.); y (vi), la representación de las soluciones propuestas como planes, la especialización del CBR se denomina planificación basada en casos (CBP) ([Bergmann *et al.*, 2005], [Hammond, 1989]) . La CBP es una especialización del CBR que requiere una representación más específica del concepto de caso [Cox *et al.*, 2006].

Adicionalmente a la memoria de casos puede hacerse necesaria la utilización de una **base de conocimiento**. La base de conocimiento contendrá informaciones tales como teorías, principios, reglas, etc. que permitan tomar decisiones para la solución de problemas y aprender de las experiencias. Tal y como se ha indicado, estos conocimientos suelen ser necesarios durante las etapas de adaptación y de revisión. Se trata de conocimientos que permiten realizar funciones tales como la generalización, la toma de decisiones, etc. Se trata de funciones complicadas en las que el mantenimiento también juega un papel muy importante. El conocimiento del que dispone el sistema puede ser un sistema experto. El sistema de conocimiento experto debe ser capaz de

aprender de las experiencias adquiridas por el sistema CBR. Probablemente se trata de la parte más complicada de un sistema CBR.

3.4.2 Minería de Datos y Sistemas de Recomendación

El área de la minería de datos aporta a través del desarrollo de sus técnicas importantes soluciones a los problemas que presentan los sistemas de recomendación personalizada de productos. El descubrimiento de reglas de asociación, basadas en el descubrimiento del patrón de los usuarios Web a través de sus transacciones o interacciones en línea con los contenidos constituyen la base de los mecanismos empleados por la Minería Web [Chang *et al.*, 2001]. Estas técnicas permiten habilitar mecanismos de personalización y recomendación en los sistemas. Tales técnicas en general, proporcionan una ventaja en el cómputo, así como una mejor eficacia en los procesos de recomendación que las técnicas tradicionales basadas en filtrados colaborativo, en particular en el contexto de los datos de *click-stream*. Hay que destacar que en todos los casos las reglas deben ser inducidas de manera *off-line* a través del procesamiento de los datos.

En lo que se refiere a técnicas predominantemente utilizadas en métodos de filtrado colaborativo basados en memoria, las máquinas de soporte vectorial (*Support Vector Machines*, SVMs) están entre las más populares. Tal técnica consiste en un método de aprendizaje supervisado en el cual se construye un clasificador lineal. Para ello, cada usuario es tratado como un vector compuesto por valoraciones sobre ítems. Dichos vectores son asociados a un espacio geométrico en el cual se construye un hiperplano de separación entre las posibles clases que, en este contexto, se refieren a grupos de usuarios con preferencias similares. A diferencia de otros métodos de aprendizaje, la precisión no está relacionada con el número de ítems, sino al margen de separación entre los datos.

Con respecto a los métodos de filtrado colaborativo basados en modelo, que construyen un modelo de estimación de valoraciones, las técnicas de aprendizaje automatizado son las más empleadas. Múltiples técnicas de este tipo, utilizadas para resolver otros problemas de minería de datos, también lo son para los sistemas de recomendación. Dichas técnicas, que en este ámbito reciben el nombre de técnicas de minería web, contribuyen al desarrollo de sistemas de recomendación personalizados y eficientes ya que posibilitan el descubrimiento de características de los usuarios que aumenten la probabilidad de realizar recomendaciones efectivas.

El agrupamiento (*clustering*) es un tipo de técnica utilizada en minería de datos que también se emplea en sistemas de recomendación. El agrupamiento se lleva a cabo con aprendizaje no supervisado para obtener grupos de usuarios que poseen preferencias similares. Las predicciones realizadas al usuario activo

estarán en conformidad con las opiniones de los miembros del grupo al que pertenece. Se puede aplicar la lógica difusa y de esta forma un determinado usuario puede pertenecer a varios grupos, aunque asignándole una medida de pertenencia del usuario a cada grupo. La utilización de técnicas de agrupamiento suele producir recomendaciones menos personales que otros métodos e, incluso en algunos casos, puede lograr menos precisión que el método del vecino más cercano. Por lo tanto, sólo conviene utilizar técnicas de agrupamiento de forma preliminar, cuando los datos disponibles para la construcción del modelo son analizados en una etapa inicial.

Otra técnica utilizada en minería de datos muy empleada en la construcción de modelos de recomendación son las redes Bayesianas. La técnica consiste en la utilización de un conjunto de entrenamiento para construir una estructura probabilística para realizar el filtrado colaborativo.

Además del agrupamiento y de las redes bayesianas, las reglas de asociación también pueden ser aplicadas para mejorar la personalización de sistemas de recomendación [Mican *et al.*, 2010]. En este contexto, las reglas pueden indicar relaciones entre ítems (en el caso de los métodos basados en contenido), entre usuarios (en el caso de los métodos basados en memoria) o entre usuarios e ítems (en el caso de los métodos basados en modelos).

Actualmente se suele combinar el uso de reglas de asociación con alguna otra técnica. Por ejemplo, un método que utiliza información proveniente de transacciones de grupos de usuarios con preferencias comunes (vecinos) para generar reglas de asociación sobre objetos Web. En otros casos, se propone la adaptación del proceso de inducción de reglas de asociación para sistemas de recomendación definiendo dos categorías de reglas: las que expresan relaciones entre ítems y las que expresan relaciones entre usuarios.

Especialmente interesante para el objetivo de este trabajo de investigación resulta un prometedor campo de la minería web aplicado a la clasificación de elementos etiquetados, tales como los objetos de aprendizaje. El siguiente apartado detallará más ampliamente estas técnicas.

3.4.2.1 *Técnicas de Aprendizaje Multi-etiqueta*

Las técnicas de aprendizaje Multi-etiqueta se originan como técnicas de la minería de datos para dar solución al problema de categorización de texto [Schapire and Singer, 2000] [McCallum, 1999] [Yang, 1999]. Un documento de texto puede pertenecer a varios temas predefinidos de forma simultánea. En el aprendizaje multi-etiqueta, el conjunto de entrenamiento se compone de casos. Cada caso está asociado a un conjunto de etiquetas. Una vez entrenado el sistema, la tarea consiste en dado un caso nuevo, predecir el conjunto de etiquetas de salida que lo describen y cuyo número se desconoce a priori. El conjunto de etiquetas que describen dicho nuevo caso se extraen a través del análisis de instancias de aprendizaje de casos vistos mediante conjuntos de

etiquetas conocidas. Estas técnicas ha atraído recientemente una considerable atención, para el reconocimiento de patrones en otros ámbitos en crecimiento tales como las redes sociales [Mika, 2005] [Xu *et al.*, 2006], la anotación semántica de imágenes [Zhang & Zhou, 2007] [Boutell *et al.*, 2004], la categorización de la música en emociones [Li & Ogihara, 2003] [Trohidis *et al.*, 2008] o la bioinformática [Diplaris *et al.*, 2005] [Roth & Fischer, 2006] [Zhang & Zhou, 2006]. Especialmente interesante para el objetivo de este trabajo de investigación resulta la aplicación de modelos generativos probabilísticos llamados modelos de mezcla paramétricos para texto multi-etiquetado, que han probado su eficacia para categorización multi-etiqueta de páginas web [Ueda & Saito, 2003]. Un compendio de estas técnicas y aplicaciones es recogido en la reciente publicación sobre *Data Mining an Knowledge Discovery* [Maimon and Rokach, 2010].

En [Katakis *et al.*, 2008] se indica cómo la clasificación tradicional a través de una sola etiqueta se vincula al aprendizaje de un conjunto de ejemplos que se asocian a una sola etiqueta λ de un conjunto de etiquetas disjuntas L , $|L| \leq 1$. Si $|L| = 2$, entonces el mecanismo de aprendizaje se denomina clasificación binaria (o de filtrado en el caso de datos de texto y web), mientras que si $|L| \geq 2$, entonces se llama clasificación multi-clase. En la clasificación multi-etiqueta, los ejemplos están asociados a un conjunto de etiquetas $Y \subseteq L$.

Existen dos tareas principales para el aprendizaje supervisado a partir de datos multi-etiqueta: la clasificación multi-etiqueta (MLC, del inglés *multilabel classification*) y el ranking de etiquetas (LR, del inglés *label ranking*). El MLC relaciona el modelo de aprendizaje con una bipartición del conjunto de etiquetas en relevantes o irrelevantes con respecto a la instancia. Poner en práctica métodos capaces de sugerir tanto un orden como una bipartición del conjunto de etiquetas de datos multi-etiqueta. Esta tarea es la denominada como ranking multi-etiquetado (MLR, del inglés *multi-label ranking*) [Brinker *et al.*, 2006] y constituye una generalización muy interesante y útil de la MLC y LR. Los métodos de clasificación multi-etiqueta se pueden clasificar en dos grupos diferentes: i) los métodos de la transformación del problema, y ii) los métodos de los algoritmos de adaptación [Tsoumakas & Katakis, 2007]. El primer grupo de métodos son algoritmos independientes y las transformaciones sólo afectan al espacio de las etiquetas. Transforman la tarea de clasificación multi-etiqueta en una o varias tareas de clasificación de etiqueta única, regresión o ranking de etiquetas. El segundo grupo de métodos especificados amplían algoritmos de aprendizaje para manejar directamente datos multi-etiquetados.

Existe una gran variedad de algoritmos de minería de datos multi-etiqueta, tales como: BR (*Binary Relevance*), LP (*Label Powerset*), RAKEL (*RANdom K-labELsets*) [Tsoumakas & Vlahavas, 2007] MLkNN (*Multi-Label K-Nearest Neighbor*) [Zhang & Zhou, 2007] [MLkNN, 2010] [Tsoumakas & Vlahavas, 2007].

A la luz de las aplicaciones prácticas de dichas técnicas revisadas [Maimon and Rokach, 2010], parece posible que la clasificación multi-etiqueta pueda ser aplicada como una implementación al problema de la clasificación automática de los objetos de aprendizaje recuperados. De manera que entrenando el sistema con un conjunto de objetos de aprendizaje tratados se puedan establecer ciertos patrones sobre la clasificación de los mismos respecto a sus etiquetas. Incluso añadiendo una modificación en los casos de entrenamiento añadiendo datos etiquetados sobre casos de uso de estos objetos de aprendizaje se puedan obtener patrones de usuario en el uso de dichos objetos, etc. Esta idea ha motivado que se haya sondeado esta posibilidad para lo cual se realiza una prueba con un conjunto de objetos de aprendizaje. La gran dimensión del espacio de etiquetas de los objetos de aprendizaje no hace sencilla su aplicación, exigiendo un alto número de casos para el entrenamiento de los algoritmos. Aunque se trata de una línea de trabajo activa en el momento de realizar esta memoria, estimo relevante dejarlo reflejado.

3.5 TRABAJOS RELACIONADOS

A continuación se enumeran trabajos relacionados, algunos centrados en su esfuerzo por generar arquitecturas que permitan realizar la recuperación de objetos de aprendizaje de un modo más o menos centralizado y otros, que partiendo de cierto conjunto de objetos de aprendizaje, se centran en posicionarlos de manera adecuada para ofrecer los más relevantes al usuario. En el objetivo de una solución integradora, el trabajo contempla dos escenarios relacionados, el de la recuperación automática de contenidos digitales por un lado al que se suma el filtrado personalizado de los contenidos recuperados.

3.5.1 Aproximaciones a la Recuperación Automática de Contenidos Educativos

Uno de los problemas que debe superarse para la normalización del uso de objetos de aprendizaje es la dificultad existente para encontrarlos [Bates, 2005]. Surgen por ello múltiples soluciones en la definición de arquitecturas y soluciones a su recuperación.

BILDU [Casquero *et al.*, 2008], propone una arquitectura basada en un conjunto de buscadores verticales integrados horizontalmente bajo una única interfaz, a la que una red social da soporte incorporando las características de la recomendación social y mecanismos de sindicación. Otros encapsulan la integración de contenidos y las búsquedas en las colecciones a través de mecanismos peer-to-peer, como en Edutella [Qu & Nejd, 2004] o LESSON [Zhou *et al.*, 2010] que utiliza un motor de metabúsqueda y sistema P2P para

compartirlos o [Santiago & Raabe, 2010] que proporciona a través de una red P2P entre instituciones, la posibilidad de búsqueda y recuperación de objetos de aprendizaje a través de un API dentro de los gestores de contenido.

Otras propuestas como [Kärger *et al.*, 2006], [Almaraz, 2008], [Santiago & Raabe, 2010] emplean una arquitectura basada en mediadores para integrar Repositorios de OA. Otro enfoque importante hace uso de arquitecturas de recuperación basadas en el uso de servicios web [Li *et al.*, 2009], [Dietze *et al.*, 2007] [Hanafy & Fakhry, 2008].

El uso de arquitecturas de recuperación de contenidos educativos basadas en tecnologías de agentes aparece en las propuestas de [Margain *et al.*, 2010] [Soto *et al.*, 2007], sin concretar ni detallar resultados ni herramientas en funcionamiento.

Por otro lado el éxito de la recuperación de contenidos educativos está en gran medida en expresar las características, restricciones o condiciones sobre el uso de los recursos de aprendizaje de forma correcta. El poder expresivo de las consultas permite seleccionar los repositorios o los OA de acuerdo a consultas más exactas de los usuarios y, permite presentar los resultados cumpliendo mejor con sus expectativas. Varios trabajos [Lee *et al.*, 2008], [Segura *et al.*, 2010], [Segura *et al.*, 2011] [Kärger *et al.*, 2006] ponen de manifiesto el poco poder de expresividad en la recuperación de los OA en sus respectivos ROAs aportando propuestas de mejora tales como la incorporación de aspectos semánticos en la expansión de los mecanismos de consulta. [Shih *et al.*, 2006] ha desarrollado una herramienta de edición de metadatos para apoyar la búsqueda y reutilización de OAs. En [Lee *et al.*, 2008] se propone una arquitectura con un enfoque ontológico para la recuperación de objetos de aprendizaje mediante la expansión del algoritmo de consulta con agregación de semántica, que denomina *Ontology Query Expansion Algorithm*. Para lo que define una ontología de contexto e infiere las intenciones del usuario mediante preguntas cortas que expanden la consulta en el ROA. Sin embargo realiza la simulación de esta propuesta, sobre el contexto del lenguaje de programación Java y realiza las pruebas sobre documentos de html en la Web visible, no haciendo uso de las características de los metadatos de los OA reales, ni de las funcionalidades de búsqueda de los ROA. El segundo de los trabajos [Segura *et al.*, 2010] hace nuevas aportaciones al mecanismo de la expansión semántica de las consultas, pero se queda en una propuesta no contrastada sobre ROA reales y sobre sus archivos de metadatos etiquetados. La expansión de las consultas en [Segura *et al.*, 2011] define un mecanismo dinámico de búsqueda y filtrado de los resultados que es validado en el repositorio MERLOT para la búsqueda de OA relacionados con el campo de la genética haciendo uso de la ontología Gene. El mecanismo de conexión con MERLOT utilizado está basado en un servicio Restful para la recuperación de los contenidos.

Haciendo uso de metodologías del área de la recuperación de información adaptadas aparecen propuestas interesantes incorporando cálculos de similitud con métricas sobre los OA [Yen *et al.*, 2010] o técnicas de clústering (*Formal Concept Analysis*, FCA) con objetivos pedagógicos para buscar en la web [Mayorga *et al.*, 2010].

Las soluciones propuestas son muchas y variadas, la gran mayoría no obtienen resultados sobre entornos reales y la mayoría no contempla los contenidos educativos en el caso particular de objetos de aprendizaje. Las propuestas analizadas buscan una solución basada en tecnologías con alta implantación donde la inclusión de algún aspecto que socialice la búsqueda mejorara los resultados. Tan sólo las propuestas [Margain *et al.*, 2010] [Soto *et al.*, 2007] hacen uso de tecnología de agentes para su desarrollo, pero en ninguno de los dos enfoques se lleva a cabo mediante la coordinación de tareas de agentes miembros de una organización. Del conjunto de soluciones estudiadas se considera un denominador común es que fomentan la interoperabilidad y la reutilización de los recursos incluyendo en sus desarrollos aspectos de relacionados con la adaptación, bien de los propios contenidos o bien de los mecanismos del entorno que permiten su mejor acceso. Este trabajo de *investigación* plantea la solución al problema de recuperación de contenidos digitales a través de una organización de agentes que permitirá a dichos agentes trabajar de forma coordinada, combinándose para resolver un problema común. Los agentes tendrán en cuenta una planificación de las tareas que se adaptará a los constantes cambios del entorno (usuarios, contenidos, repositorios, etc.).

3.5.2 Aproximaciones a la Recomendación de Objetos de Aprendizaje

Con tantos repositorios de objetos de aprendizaje, un gran desafío consiste en encontrar de manera eficiente los OA convenientes para los usuarios. Este objetivo ha despertado muchas investigaciones en el campo de la selección y recomendación de OA.

Los investigadores y desarrolladores de e-learning han comenzado a aplicar técnicas de recuperación de información junto con tecnologías aplicadas a recomendación, especialmente filtrado colaborativo [Bobadilla *et al.*, 2009], o minería web [Khribi *et al.*, 2009], para la recomendación contenidos educativos. Una reciente revisión de estas aplicaciones puede verse en [Manouselis *et al.*, 2010b]. Las características que manejan estas técnicas de filtrado de información en este contexto son la información de los atributos del artículo educativo (enfoque basado en contenido) y sobre el contexto del usuario (enfoque colaborativo).

De los primeros trabajos está el desarrollado mediante Altered Vista. Un sistema instruccional en el que se evalúan técnicas de recomendación basadas en filtrado colaborativo mediante algoritmos de vecinos próximos [Recker *et*

al., 2003] [Recker & Wiley, 2001]. Estos trabajos exploran cómo recoger del usuario evaluaciones de recursos de aprendizaje y propagarlos en forma de recomendaciones boca-a-boca (*Word-of-mouth*).

RACOFI (*Rule-Appling Collaborative Filtering*) propone un filtro colaborativo por reglas, con una arquitectura para la selección personalizada de contenidos educativos [Lemire *et al.*, 2005]. Su recomendación se reduce a combinar dos enfoques de recomendación mediante la integración del algoritmo de filtrado colaborativo que funciona con puntuaciones del usuario con un sistema de reglas de inferencia que realiza la asociación entre los contenidos y su índice de recomendación.

En [McCalla, 2004] se propone una mejora al filtrado colaborativo, denominado enfoque ecológico, para el diseño de sistemas e-learning. Los aspectos clave de dicha propuesta tienen en cuenta la acumulación gradual de información y se centra en los usuarios finales.

Manouselis *et al.* [Manouselis & Costopoulou, 2007] [Manouselis *et al.*, 2007] han realizado un caso de estudio con los datos recogidos de los usuarios del portal CELEBRATE para determinar un algoritmo de filtrado colaborativo adecuado.

CYCLADES [Avancini. & Straccia, 2005] propone un entorno donde los usuarios buscan, acceden y evalúan los recursos digitales disponibles en los repositorios a los que se tiene acceso a través de la Iniciativa de Archivos Abiertos (OAI). Dicho sistema ofrece recomendaciones sobre los recursos que se almacenan en repositorios diferentes a través de un sistema abierto. Las recomendaciones ofrecidas por CYCLADES se han evaluado a través de un estudio piloto con 60 usuarios, que se centró en probar el rendimiento (exactitud de predicción) de varios algoritmos de filtrado colaborativo.

En todos estos sistemas de recomendación colaborativa de objetos de aprendizaje, el problema clave de arranque en frío, descrito en la sección 3.3.2, no se ha abordado. Algunos trabajos recientes intentan solucionar dicho problema. En [Tang & McCalla, 2004] [Tang & McCalla, 2003] se incorpora un sistema de filtrado colaborativo para recomendación de artículos de investigación incorporando elementos de las características pedagógicas donde hacen uso de lo que denominan alumnos artificiales para solventarlo. El dominio de su sistema, sin embargo, se limita a trabajos de investigación por lo tanto los factores que influyen en la selección de los papers son distintos a los afectados en la selección de selección de objetos de aprendizaje.

Otras soluciones a la recomendación de contenidos educativos adoptan el enfoque híbrido. [Aijuan & Baoying, 2008] [Ghauth *et al.*, 2010] [Tsai *et al.*, 2006] [Wang *et al.*, 2007], hacen uso de algoritmos basados en las valoraciones de otros usuarios y basados en intereses extraídos a través de algoritmos de

vecinos próximos. El uso de estos algoritmos, basados en la correlación, permite calcular un índice de puntuación en la utilidad de los objetos de aprendizaje a través de los análisis de comentarios de estudiantes con perfiles similares. Dichos algoritmos basados en las preferencias mejoran la selección incorporando al algoritmo aspectos de preferencias de los alumnos. El patrón de preferencia de cada alumno se registra en su historial de preferencias, que se genera y actualiza de acuerdo a los comentarios de la preferencia del alumno. Si a un objeto de aprendizaje seleccionado se le ha dado una puntuación positiva, se realiza un incremento en las puntuaciones de preferencia de todas las características del objeto de aprendizaje. La combinación de las puntuaciones de un objeto de aprendizaje determinado por los dos algoritmos decide su posición en el resultado de la recomendación. Este algoritmo de preferencia ayuda a evitar el problema de arranque en frío, cuando no se tienen datos sobre el uso y las valoraciones de un determinado objeto de aprendizaje. Sin embargo, todas las características del objeto de aprendizaje seleccionadas son tratadas por igual, sin hacer ninguna distinción entre ellas lo que permitiría un tratamiento más preciso en los criterios de valoración del mismo por el usuario, afectando al propio patrón de preferencia de dicho usuario.

El uso de algoritmos basados en modelos biológicos, como el ACO (*Ant Colony Optimization*), es la base de [Yang&Wu, 2009], donde se propone un sistema de atributos basado en un sistema de colonia de hormigas (*attributes-based ant colony system*, AACS) para ayudar a los alumnos a encontrar a través de un modelo adaptativo los objetos de aprendizaje de un modo más eficaz. Este mecanismo está basado en el uso de actividades de aprendizaje y elementos educativos para predecir las trayectorias óptimas asociadas con el algoritmo ACO y recomendar la secuencia de objetos de aprendizaje. Este trabajo es interesante en cuanto a que genera una planificación en la trayectoria de aprendizaje del alumno. Sin embargo basa su recomendación en un itinerario de aprendizaje que afecta a un conjunto inicial de objetos de aprendizaje. El objetivo final es alcanzar ciertos conocimientos. La recomendación es la secuencia elaborada a través de la ruta óptima entre los distintos OA.

Los trabajos de [Kerkiri *et al.*, 2007] [Ochoa&Duval, 2006], [Wolpers *et al.*, 2007], plantean la necesidad de una selección de objetos de aprendizaje, atendiendo a contenidos educativos descritos por sus metadatos, en el sentido del presente trabajo de investigación. Proponen un mecanismo denominado, Atención Contextualizada de Metadatos (*Contextualized Attention Metadata*, CAM) para capturar información acerca de acciones a lo largo del ciclo de vida de objetos de aprendizaje, incluyendo la creación, el etiquetado, la oferta, su selección, uso y mantenimiento. Los trabajos estudiados proponen cuatro métricas con LOM y CAM para la clasificación y recomendación de dichos objetos de aprendizaje recuperados: Ranking del Análisis del Link, la Recomendación por Similitud, El Ranking Personalizado, y la Recomendación Contextual. Estas métricas permiten clasificar los objetos de aprendizaje

atendiendo a criterios tales como ranking de popularidad, la similitud de objetos basados en el número de descargas, etc. El cómo estas clasificaciones contribuyen a la selección de objetos de aprendizaje y cómo se combinan entre sí siguen siendo cuestiones abiertas y una materia de estudio altamente interesante.

Basándose en aspectos semánticos y considerando la información contextual procedente de actividades cognitivas del alumno y la estructura de contenidos objeto de aprendizaje, Qiyan *et al.* [Qiyan, 2010] proponen un framework de recomendación de objetos de aprendizaje que se adapta a las actividades cognitivas del alumno a través de un método basado en ontologías. El mismo planteamiento sigue el trabajo de Ruiz-Iniesta (2010), donde además se está desarrollando un framework que permita el desarrollo de recomendadores para OA de un modo sencillo.

Existen otras aproximaciones, con mecanismos o criterios de catalogación de contenidos educativos que requieren la intervención humana directa en su valoración, pero que hacen evidente la necesidad de proponer criterios de evaluación de la calidad de los contenidos. Entre ellos, destaca la evaluación de contenidos en el repositorio MERLOT o la herramienta LORI. El repositorio MERLOT (*Multimedia Educational Resource for Learning and Online Teaching*) ofrece el mejor ejemplo actual de aplicación masiva en la evaluación de contenidos educativos para la educación basada en Web [Vargo et al, 2003]. Las valoraciones de los contenidos se obtienen a través de comentarios y calificaciones en una escala de cinco puntos por parte de los usuarios de MERLOT y revisores nombrados (que evalúan mediante criterios de evaluación por pares). Dicha evaluación se realiza sobre tres propiedades generales: calidad de los contenidos, la eficacia potencial como herramienta de enseñanza-aprendizaje, y la facilidad de uso. La clasificación basada en la calidad de los resultados de búsqueda utiliza un promedio ponderado, de estas tres clasificaciones se evalúa por pares de expertos en la materia de trabajo de forma asincrónica. Por otra parte, devuelve las descripciones de los contenidos recuperados en una lista ordenada por el ranking que establece la evaluación de la calidad de dichos contenidos en orden descendente de valoración en donde los contenidos no evaluados aparecen los últimos. Y Finalmente LORI (*Learning Object Review Instrument*) es uno de los denominados “instrumentos” más conocidos para evaluar la calidad de recursos para la educación on-line. No se trata más que de un protocolo de evaluación de objetos de aprendizaje sobre nueve aspectos en una escala de de cinco puntos, que se puede implementar en línea mediante sistema de rúbricas, escalas de valoración y campos de comentarios. Como instrumento de evaluación se encuentra disponible en su web²³, y puede ser utilizada para la evaluación individual o por un panel de

²³ LORI <http://www.elera.net/eLera/Home/About%20%20LORI/>

expertos de un conjunto de OA, atendiendo a las indicaciones de [Vargo et al, 2003] [Nesbit et al, 2002]. Este instrumento evalúa la calidad del objeto, no atendiendo a aspectos masivos de uso del mismo o a criterios relacionados con aquellos que hicieron uso de los mismos.

3.6 CONCLUSIONES DEL CAPÍTULO

El número creciente de trabajos proponiendo sistemas para la recuperación automática de recursos de aprendizaje, evidencia la carencia de soluciones operativas según confirma el reciente trabajo de [Manouselis *et al.*, 2010a]. Del conjunto de las propuestas evaluadas se concluye que la incorporación de mecanismos que valoren atributos relacionados con los contenidos educativos (OA en nuestro trabajo), así como aspectos del contexto del usuario y su interacción con los contenidos, generan mecanismos de filtrado personalizado efectivo. Sin embargo una mirada más cercana a las propuestas revisadas pone de manifiesto la falta de aplicaciones sobre sistemas y contenidos educativos reales. La mayoría de los trabajos relacionados en este apartado, se basan en simulaciones o bien se han aplicado a un caso local de estudio o a un repositorio en particular. La mayoría de ellos realizan las pruebas con OA controlados a priori, para grupos reducidos de parámetros por lo general también locales. Son propuestas a la solución que no muestran resultados sobre un contexto real. Algunos de los contenidos educativos recomendados no son siquiera objetos de aprendizaje tal y como hemos definido en este trabajo y la gran mayoría por tanto no atiende a aspectos del etiquetado semántico de los recursos en su planteamiento.

A lo largo de este capítulo hemos formulado métricas que nos van a permitir la evaluación de la eficiencia en la recuperación contenido educativo. Hemos definido una nueva medida denominada ganancia temporal y realizando una evaluación satisfactoria sobre Repositorios reales.

La arquitectura propuesta en este trabajo facilitará evaluar las múltiples perspectivas de la recuperación de contenidos educativos desde un entorno real, abierto y escalable y a la vez será el soporte que permitirá implementar mecanismos de recomendación o ranking para los OA recuperados. La estrategia de recomendación utilizada en este trabajo es híbrida. Dicha recomendación se compone de un recomendador basado en contenido personalizado reactivo inicial cuyas propuestas son refinadas por un recomendador colaborativo incorporando aspectos de navegación social. El resultado será una lista ordenada de Los personalizada al usuario que realizó la búsqueda.

4 SISTEMAS MULTI-AGENTE BASADOS EN ORGANIZACIONES

Los retos a los que se enfrentan en nuestros días los desarrolladores de aplicaciones en los sistemas de información son cada vez más complejos. Los agentes inteligentes son objeto de gran interés por parte de múltiples campos de la informática y de la Inteligencia Artificial (IA) como paradigma para el desarrollo de aplicaciones [Jennings y Wooldridge, 1998].

La evolución del software, y concretamente el que incorpora elementos de la inteligencia artificial, tiende a la creación de entidades con comportamientos y conductas que emulan a las de los humanos. La teoría de agentes surge como una evolución de la inteligencia artificial distribuida y se sustenta sobre el concepto de agente [Russel y Norvig, 1995], como una entidad autónoma dotada con ciertas capacidades propias de los seres humanos. Una definición más formal sería la siguiente: un agente es una entidad física o abstracta que puede percibir su ambiente a través de sensores, es capaz de evaluar tales percepciones y tomar decisiones por medio de mecanismos de razonamiento sencillos o complejos, comunicarse con otros agentes para obtener información y actuar sobre el medio en el que se desenvuelve a través de ejecutores [Labidi y Lejouad, 1993]. En concreto, los agentes inteligentes se consideran entidades que tratan de imitar procesos de razonamiento o comportamientos humanos. Los agentes ya han sido empleados con éxito en diferentes ámbitos para facilitar tareas a los usuarios de recolección, supervisión y filtrado de información [Maes, 1994] [Cheong, 1996] [Ruqayya *et al.*, 2011] . Los agentes permiten el modelado y diseño de sistemas flexibles y personalizados para la gestión y el acceso a la información [Sycara, 2001], generando diferentes estrategias como la de recomendación sobre la diversidad de información disponible [De la Rosa *et al.*, 2003].

Un agente puede verse como una evolución del concepto de objeto software, perfeccionada gracias a la influencia de la inteligencia artificial, que permite incorporar características como la racionalidad, la inteligencia, la autonomía o el aprendizaje. Al igual que los humanos, los agentes deben tener habilidades sociales y ser capaces de realizar trabajos o resolver problemas de forma distribuida [D'Inverno y Luck, 2004]. Se habla entonces de un sistema multi-

agente, en el que los agentes cooperan e interactúan para conseguir los objetivos finales del sistema.

En la actualidad, pese a que el concepto de agente ya ha sido ampliamente estudiado, aparece una importante área de desarrollo en la consideración del agente como entidad que opera dentro de una sociedad. Las características básicas de un agente es que se encuentre situado en un entorno cooperando con otros agentes permiten dotarle de comportamientos sociales que emulando nuevamente a conductas humanas, generan en dichas sociedades de características más complejas.

El presente capítulo estudia el concepto de agente desde el punto de vista de su utilización en entornos cooperativos. La noción de agente evoluciona hacia la generación de sociedades de agentes y el desarrollo de Sistemas Multi-agente con capacidades de organizaciones. Comienza con una revisión de la de noción de agente inteligente. A continuación se expone una descripción de los sistemas multi-agente, con una especial consideración a aquellos que poseen propiedades organizacionales.

4.1 EL PARADIGMA DE AGENTE INTELIGENTE

El concepto de agente no es fácil de explicar debido a que este campo abarca una gran variedad de disciplinas de las ciencias de la computación (IA, Ingeniería del Software, Bases de Datos, Sistemas Distribuidos) e incluso de otras áreas de conocimiento como son la Biología, Psicología o la Sociología [D'Inverno y Luck, 2004]. Tener en cuenta tantos campos de conocimiento supone muchas ventajas, pero también muchos inconvenientes; el primero de ellos es establecer consenso en una definición. [Wooldridge y Jennings, 1995] propone la siguiente definición:

Un sistema computacional basado en hardware o (más habitualmente) basado en software que disfruta de las propiedades de autonomía, habilidades sociales, reactividad y proactividad.

Posteriormente, el mismo Wooldridge proporcionó una segunda definición [Wooldridge, 2002]:

Un agente es un sistema computacional que está situado en algún entorno, y que es capaz de actuar de forma autónoma en este entorno para poder alcanzar sus objetivos de diseño.

Aunque estas dos definiciones, son acertadas, existen otras tantas, en las que cada autor introduce su visión particular sobre el concepto de agente [Russell y

Norving, 2009] [Labidi y Lejouad, 1993], etc. Dado que no parece posible llegar a un acuerdo respecto a la definición, se opta por definir una serie de características que debe manifestar un agente. Para ello, se parte de las características propuestas por Wooldridge en su primera definición (autonomía, habilidades sociales, reactividad y proactividad), pero enfatizando el carácter inteligente de los agentes, es decir, su capacidad de iniciativa para actuar dirigidos por los objetivos a conseguir. Por tanto, a continuación se enumeran las características que se exigen en un agente:

- **Autonomía.** *Un agente es capaz de actuar sin la intervención de otros entes (personales o impersonales).*
- **Situación.** *Un agente es capaz de situarse dentro de un entorno (real o virtual).*
- **Reactividad.** *Un agente es capaz de percibir el entorno que le rodea y actuar en función las necesidades.*
- **Pro-Actividad o Racionalidad.** *Un agente es capaz de definir metas propias y planificar la secuencia de acciones que les permita alcanzar estas metas.*
- **Habilidad social.** *Un agente es capaz de interactuar con otros agentes (personales o impersonales).*
- **Inteligencia:** *Un agente es capaz de rodearse de conocimiento (creencias, deseos, intenciones y metas).*
- **Organización.** *Un agente es capaz de organizarse dentro de sociedades que siguen unas estructuras similares a las definidas en sociedades humanas o ecológicas.*
- **Movilidad.** *Un agente es capaz de moverse de unos sitios a otros, este movimiento supone el transporte del código y del estado.*
- **Aprendizaje.** *Un agente es capaz de adaptarse progresivamente a cambios en entornos dinámicos, mediante técnicas de aprendizaje.*

4.1.1 Clasificación de agentes

Como se ha observado en el apartado anterior, existe un número elevado de definiciones y características que identifican el concepto de agente, lo que provoca que la labor de clasificación de los diversos tipos de agente no resulte menos compleja que su definición. Existe un gran número de propuestas de clasificación de agentes [Russell y Norving, 2009] [Nwana, 1995] [Brenner *et al.*, 1998] [Franklin y Graesser, 1997] [Wooldridge, 2002] [Cetkovic y Parmee, 2002]. Prácticamente todas estas categorizaciones establecen los atributos de los agentes como el criterio principal de la clasificación así como los entornos de aplicación de éstos. Características comunes de los agentes, determinados por atributos, se toman como referencia fundamental, entre ellos encontramos a la autonomía, la pro-actividad, el aprendizaje o la cooperación.

Algunas de las principales propuestas, atendiendo a los criterios de clasificación que establecen los distintos autores, se muestran a continuación.

1. Según el *tipo de programa utilizado para implementar las funcionalidades del agente*, estableciendo un estado intermedio entre percepciones y acciones, [Russell y Norving, 2009] determinan cuatro tipos de agente:
 - **Agente reflejo simple.** agentes que reaccionan ante estímulos, proporcionando una respuesta y sin capacidad de memoria. Un ejemplo de agente reflejo simple puede encontrarse en [Mansury *et al.*, 2002] [Kleinsteins&Seiden, 2000].
 - **Agente reflejo con estado interno** (*Reflex Agent with State*). En este caso el agente incorpora, con respecto al agente de reflejo simple, un estado de memoria interna, que almacena experiencias pasadas, para elaborar las mejores respuestas en el futuro [Carrascosa *et al.*, 2008].
 - **Agente basado en metas** (*Goal Based Agent*). Se trata de un agente que persigue alcanzar objetivos concretos incorporando capacidades deliberativas, para lo cual suele utilizar técnicas de búsqueda y planificación [Pokahr *et al.*, 2003].
 - **Agente basado en utilidad** (*Utility Based Agents*). Un proceso deliberativo dirigido por metas simplemente, no permite que el agente adquiera características propias. La utilidad es una función que a cada estado le asocia su grado de utilidad. La completa especificación de la función de utilidad permite la toma de decisiones racionales en el momento en que es necesario satisfacer algunas metas que implica un conflicto entre ellas o cuando el agente puede desear obtener varias metas y no existe la certeza de lograr ninguna de ellas [Kephart y Walsh, 2004].
2. Atendiendo a la *utilización de las teorías de agentes en el desarrollo de productos software comerciales*, [Brenner *et al.*, 1998] proponen una nueva clasificación. Esta propuesta denomina agentes software a las entidades que presentan las características propias de los agentes y los clasifica atendiendo a su utilización en el desarrollo de software de carácter comercial: agentes de interfaz o asistentes personales, agentes de Internet, etc.
3. Atendiendo a *modelos biológicos naturales* [Franklin y Graesser, 1997] proponen una clasificación para los agentes que se asimila a la clasificación realizada en el mundo animal. La taxonomía de los agentes, concebidos como software al que se le incorporan capacidades propias de entidades de la naturaleza, concretamente de los humanos, se concibe por jeraquías.

4. Atendiendo al *diseño conceptual de los agentes* Cetkovic y Parmee [2002] identifican tres tipos de agentes: (i) Agentes interfaz, que ayudan al diseñador a interactuar con el sistema y a descubrir los detalles de bajo nivel que, en ocasiones, permanecen ocultos. (ii) Agentes de búsqueda, que se ocupan de gestionar los procesos de optimización, cooperación, monitorización de la población, restricciones, etc. (iii) Agentes de información, que trabajan con la información obtenida, buscando soluciones de interés. Dichos agentes desestiman las soluciones no validas y toman decisiones respecto a qué conflictos o problemas resolver y dónde resolverlos.

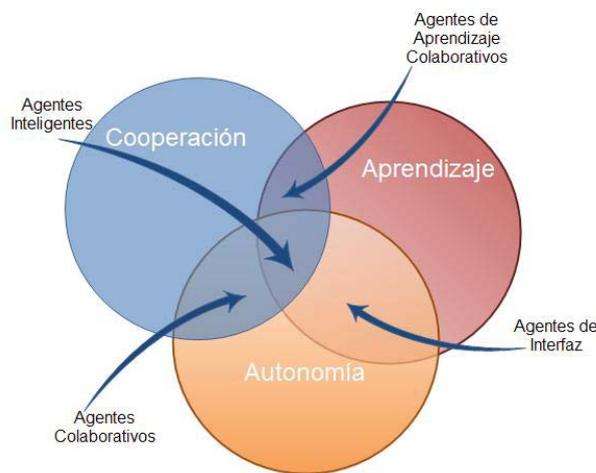


Figura 4-1. Clasificación de agentes según característica primaria [Nwana, 1995]

5. Teniendo en cuenta los *atributos propios de los agentes*, donde destacan la movilidad y el tipo de razonamiento utilizado por el agente y se consideran el resto de atributos como elementos que definen nuevos tipos. En base a estos atributos [Nwana, 1995] propone tres tipos de agentes en la siguiente clasificación:
- **Movilidad.** Atributo que describe la habilidad de los agentes para moverse a través de una red de comunicaciones. Atendiendo a este criterio distingue (i) agentes estáticos, que estando asignados dentro de una red, no pueden moverse de esta; y (ii) agentes móviles, son capaces de migrar de una plataforma de agentes a otra, o bien de un *host* a otro dentro de una red.
 - **Modelo de razonamiento.** Atributo que describe la representación del conocimiento del agente y la forma en que el

agente utiliza ese conocimiento para resolver problemas. Basándose en este atributo distingue entre (i) agentes deliberativos, que incorporan un modelo deliberativo para su razonamiento, estos agentes poseen un modelo simbólico interno de razonamiento y se ocupan de la planificación y la negociación con el objetivo de conseguir coordinarse con otros agentes; y (ii) agentes reactivos, que son agentes que reaccionan frente a estímulos solamente utilizando el conocimiento instantáneo del entorno.

- **Otras características.** Para el resto de agentes no clasificados según los dos atributos anteriores, Nwana considera otras tres características que denomina atributos primarios. Estos son la autonomía, la cooperación y el aprendizaje. Atendiendo a estos tres atributos distingue cuatro tipos de agentes (i) agentes colaborativos, que son agentes autónomos que cooperan con otros agentes; (ii) agentes interfaz, que son agentes autónomos con capacidad de aprendizaje; (iii) agentes de aprendizaje, que son agentes con capacidad de aprendizaje y que trabajan de forma coordinada; y (iv) agentes smart, que son agentes autónomos con capacidad de aprendizaje y que cooperan con otros agentes. Atendiendo a esta clasificación, se han obtenido tres grupos o clases de agentes, sin embargo, tal y como se aprecia en la Figura 4-1, Nwana indica que es posible combinar dos o las tres clases propuestas para un tipo de agente, definiendo así lo que llama *agente híbrido*.

Finalmente existen otros atributos secundarios, que también pueden ser tenidos en cuenta. Así pues, Nwana llega a la conclusión de que existen siete tipos de agentes, tal y como se puede apreciar en la Figura 4-2.



Figura 4-2. Clasificación de agentes software según Nwana [Nwana, 1995]

Concluyendo este apartado, se observa como criterio unificador de las clasificaciones enumeradas, que los atributos de los agentes, establecen el criterio principal en las taxonomías de los agentes propuestas. Todas ellas convergen en la necesidad fundamental de diferenciar aquellos agentes que son simplemente reactivos de los que incorporan algún mecanismo de razonamiento, haciendo muchas de ellas hincapié en la existencia de agentes híbridos, que flexibilizan las clasificaciones y permiten la integración de más de un criterio de clasificación.

4.1.2 Arquitecturas

En general, una arquitectura describe los componentes principales del sistema describiendo la forma en la que se relacionan e interactúan para alcanzar un objetivo final. Considerando el agente como un sistema más o menos complejo, su arquitectura describe la estructura interna de dicho agente, la forma en que se descompone en un conjunto de módulos y cómo estos módulos interactúan entre sí para alcanzar la funcionalidad requerida, determinando la arquitectura de agente [Mas, 2005].

Existen diferentes mecanismos de clasificación de las arquitecturas basadas fundamentalmente en el mecanismo de toma de decisiones de sus agentes. Dichos mecanismo derivan en la forma de razonamiento de los agentes, conformando fundamentalmente dos líneas importantes en la clasificación de dichas arquitecturas. Una primera propuesta por [Wooldridge, 1999], que tiene en cuenta el mecanismo de razonamiento derivado de los agentes que participan en su configuración y otra segunda donde el mecanismo de razonamiento se basa en la planificación dentro del contexto de la inteligencia artificial y que se reseñan a continuación.

Wooldridge propone una clasificación de arquitecturas atendiendo a los agentes participantes y enumentando cuatro propuestas según si participan en dicha arquitectura (i) *Agentes basados en lógica*, cuyo razonamiento y toma de decisiones se realiza a través de lógica y deducción [Genesereth y Nilsson, 1987] [Lésperance *et al.*, 1996] [Fischer, 1994]. (ii) *Agentes reactivos*, donde la toma de decisiones es llevada a cabo como un mapeo directo de una situación a una acción [Brooks, 1986] [Maes, 1989]. (iii) *Agentes BDI (Belief-Desire-Intention)*, donde la toma de decisiones depende de la manipulación de la representación de las creencias, deseos e intenciones del agente [Bratman *et al.*, 1988] [Rao y Georgeff, 1992]. Y Finalmente están las arquitecturas basadas en (iv) *Agentes basados en capa*, donde la toma de decisiones se realiza a través de varias capas de software, cada una llevando el razonamiento sobre el entorno a diferentes niveles de abstracción [Bratman *et al.*, 1988].

En el contexto de la inteligencia artificial, uno de los mecanismos de razonamiento se basa en lo que se denomina la planificación. La ejecución de

planes de acciones se basan en un alcanzar un objetivo final. De ese modo los sistemas de planificación utilizan modelos de representación del conocimiento y razonamiento de tipo simbólico diseñados en forma de búsquedas en un espacio de estados o en un espacio de planes y su modo de actuación está definido por la necesidad de satisfacer unos objetivos básicos (correctitud, completitud y optimalidad) para la elaboración de un plan dentro de una complejidad de espacio y tiempo. La planificación clásica presenta una desventaja en su aplicación en entornos de tiempo real, debido a que algunas veces estos algoritmos de planificación no pueden responder en el tiempo necesario que exige el sistema. Existen nuevas propuestas que buscan la implementación de otros modelos de representación del conocimiento así como de los mecanismos de planificación que sustentan el razonamiento de estas arquitecturas. La siguiente clasificación propone tres tipos de arquitecturas que se diferencian en el modelo de razonamiento que incorporan: (i) Las *arquitecturas deliberativas*, emplean modelos de representación simbólica del conocimiento, denominada su base de conocimiento y suelen estar basadas en la teoría clásica de planificación partiendo de un estado inicial en el que existen una serie de planes y un estado final al que se quiere llegar. Los agentes que la forman están dotados de un sistema de planificación que habitualmente se basa en la teoría clásica de planificación siguiendo el ciclo “Percepción → Deliberación → Acción”. Las arquitecturas intencionales se pueden implementar utilizando una aproximación deliberativa, entre esto tipos de arquitecturas la más ampliamente extendida y estudiada es la que se basa en el modelo BDI [Rao y Georgeff, 1992]. Por otro lado las (ii) *Arquitecturas Reactivas*, carecen de razonamiento simbólico y de representación del entorno, por lo que sus mecanismos de comunicación con otros agentes son muy básicos. Este tipo de agentes sigue el ciclo “Percepción → Acción” modificando sus comportamientos y el mismo entorno [Maes, 1990]. Finalmente las (iii) *Arquitecturas híbridas* resultan ser un compendio entre los dos modelos anteriores incluyendo comportamientos reactivos y deliberativos. Este tipo de agentes sigue el ciclo “Percepción→Decisión→Acción”.

De las arquitecturas descritas, este trabajo presta especial atención a la arquitectura basada en modelo BDI. El modelo BDI (creencias (*Beliefs*), deseos (*Desires*) e intenciones (*Intentions*)) es un modelo de agentes racionales, que reconoce la primacía de creencia, deseos, e intenciones en la acción racional. Dicho modelo suma tres fuerzas distintas: una filosofía subyacente basada en el razonamiento práctico humano, una arquitectura de software que es implementable en verdaderos sistemas, y una familia de lógicas que apoyan una teoría formal de agencia racional. Por un lado, la arquitectura basada en el modelo BDI se ha convertido en un estándar de facto para los modelos de agente y es aceptada por la FIPA²⁴, y, por otro lado, es suficientemente genérica

²⁴ FIPA (Foundation for Intelligent Physical Agents) <http://www.fipa.org/>

como para permitir el modelado de agentes, tanto naturales como artificiales. En el siguiente apartado se profundizará en dicho modelo.

4.2 AGENTES DELIBERATIVOS. EL MODELO BDI

Tal y como se ha podido apreciar a lo largo de este capítulo, existen muchas arquitecturas que permiten construir agentes deliberativos (agentes que utilizan un modelo simbólico para representar su conocimiento), donde la mayoría de ellas se basan en el modelo BDI [Bratman *et al.*, 1988]. La inteligencia artificial, ha venido definiendo la capacidad de razonamiento de un agente inteligente a través de lo que se han denominado *actitudes mentales*. El modelo BDI introducido por Bratman, limita estas actitudes mentales de los agentes a tres: creencias (*Beliefs*), deseos (*Desires*) e intenciones (*Intentions*).

- Las **Creencias** como el conjunto de proposiciones (conocimiento del entorno) que el agente acepta como verdaderas.
- Los **Deseos, metas u objetivos** que denotan una propiedad o conjunto de propiedades del entorno que el agente quiere que sean verdaderas. Es decir, representa algún estado final deseado.
- Las **Intenciones** como el conjunto de acciones planificadas por el agente que le permiten llegar a un estado deseado.

Además de las creencias, los deseos y las intenciones, es común definir un conjunto de planes, es decir, una secuencia de acciones [Corchado *et al.* 2008]. Si un agente tiene la intención de alcanzar un objetivo, debe tener la intención de ejecutar el plan que permita alcanzarlo, y además, debe creer que ese plan logrará dicho objetivo [Cavedon y Rao, 1996]. Por lo tanto, la arquitectura básica de un agente BDI, mostrada en la Figura 4-3 es un conjunto de creencias, planes, deseos e intenciones, procesadas por un intérprete que puede desempeñar diversas funciones, entre ellas el filtrado, el razonamiento y la deliberación.



Figura 4-3. Arquitectura básica de un agente BDI

Posiblemente el modelo BDI ha sido el más difundido y estudiado dentro de los modelos de razonamiento de agentes. Hay varias razones para que esto ocurriera pero quizás la más convincente es que el modelo BDI combina una serie de interesantes elementos. Primero, parte de un apreciable modelo filosófico de razonamiento humano fácil de comprender. Su planteamiento permite un número considerable de implementaciones [Georgeff y Lansky, 1987]. Incorpora finalmente, una semántica abstracta y elegante, la cual ha sido aceptada por la comunidad científica [Georgeff *et al.*, 1999] [Rao y Georgeff, 1998].

En un sistema de agentes BDI las entradas son una serie de posibles eventos que pueden ser tanto externos como internos. Dichos eventos se asumen como indivisibles y se recogen cuando se completan, tratándose a través de una cola de eventos. Por otro lado, la salida del sistema son acciones, igualmente indivisibles, realizadas por una función de ejecución que se basa en el estado actual del sistema así como en el conjunto de eventos presentes en la cola. La implementación de este modelo suele llevar los siguientes pasos, repetidos de forma continua:

- 1) Al inicio de cada ciclo, un generador de opciones lee la cola de eventos y genera una lista de opciones.
- 2) Un liberador de opciones selecciona un subconjunto de opciones y las agrega a la estructura de intenciones.
- 3) Si existe una intención que realiza una acción en este punto, el agente la ejecuta.

- 4) Si algún evento externo ha ocurrido durante el ciclo, es agregado como entrada a la cola de eventos.
- 5) Los eventos internos son agregados igualmente a la cola de eventos.
- 6) El agente modifica la estructura de deseos e intenciones eliminando aquellos alcanzados con éxito así como aquellos deseos imposibles y acciones no realizables.

Los agentes deliberativos BDI se modelan utilizando una estructura abstracta, basada en la lógica de situaciones (o mundos posibles), denominada árbol temporal y que contempla múltiples futuros y un único pasado [Rao y Georgeff, 1991]. En este árbol, cada nodo es una situación, y las ramas constituyen las diferentes opciones del agente en un momento dado. Para cada situación se definen a su vez otra serie de nuevas situaciones que el agente puede alcanzar desde el punto de vista de las creencias (situaciones que se consideran posibles), de los deseos (situaciones que se desean alcanzar) y de las intenciones (situaciones que se intentan alcanzar). Para generar el modelo, deben de existir ciertos axiomas de compatibilidad entre los elementos del modelo BDI, tales como:

- Compatibilidad entre creencias y objetivos. Si el agente desea alcanzar un objetivo, debe creer que dicho objetivo es cierto.
- Compatibilidad entre objetivos e intenciones. Antes de que el agente se enfoque en una intención, éste debe haberla formulado como deseo.
- Las intenciones conducen a acciones. Es necesario que el agente ejecute las intenciones a través de acciones simples.
- Relación entre creencias e intenciones. El agente conoce (cree en) sus propias intenciones.
- Relación entre creencias y objetivos. El agente debe conocer sus objetivos y deseos.
- No hay retrasos infinitos. Cuando un agente adopta una intención para alcanzar un objetivo, debe continuar hasta un determinado momento del futuro, sin posibilidad de posponer de manera infinita el alcanzar el objetivo.

Los agentes deliberativos BDI han sido exitosamente utilizados en múltiples sistemas desarrollados. Estos agentes proporcionan soluciones en entornos dinámicos e inciertos. Además, son capaces de enfrentarse a problemas del mundo real, incluso en situaciones donde sólo cuentan con una visión parcial

del problema y con un número limitado de recursos. Los agentes deliberativos BDI son capaces de proporcionar altos niveles de adaptabilidad a las condiciones del entorno de desarrollo. Esto resulta importante de cara a individualizar las acciones que puedan ser tomadas en cuenta para afrontar distintas situaciones de forma dinámica y autónoma.

El modelo BDI combina de manera satisfactoria modelos ontológicos del razonamiento humano y se mantiene en relación con otros modelos de agentes, sin embargo presenta un problema, la carencia de un modelo de aprendizaje.

La utilización de agentes deliberativos BDI será una herramienta fundamental en el desarrollo del modelo propuesto en capítulos posteriores. La variabilidad del entorno en el caso de estudio con situaciones como las impuestas por necesidades de los usuarios, de la variabilidad de la información, de las fuentes de dicha información, etc., comportan un entorno manifiestamente dinámico e incierto, sujeto a cambios bruscos y frecuentes.

4.3 MODELOS DE RAZONAMIENTO EN EL AGENTE

Existen múltiples razones por las que es deseable delegar actividades complejas en los ordenadores. El objetivo es la generación de un software que de manera autónoma pueda concebir buenas decisiones sobre acciones que nos beneficien y ejecutarlas de manera eficiente. Una de los más importantes logros de la inteligencia artificial es conseguir que los programas del ordenador puedan actuar como agentes autónomos y racionales. Los usos se extienden desde pequeños programas que inteligentemente buscan bienes de compra en la Web o generan ventas vía el comercio electrónico [Maes *et al.*, 1999], a sondas autónomas espaciales [Stranjak *et al.*, 2008].

En los trabajos [Corchado y Laza, 2003] y [Glez-Bedia *et al.*, 2002] se presenta una formulación que permite la incorporación de sistemas CBR como mecanismo deliberativo en agentes BDI, proporcionando un aprendizaje, adaptación y un mayor grado de autonomía que una arquitectura BDI pura [Bratman *et al.*, 1988]. Dicha formulación relaciona cadenas de creencias con planes de acción y deseos con funciones solución y eficacia de la solución. Esto permite relacionar los conceptos que definen la estructura interna y el comportamiento de un agente BDI con los conceptos manejados en sistemas de razonamiento basado en casos. Para ello establece un formalismo analítico en el que se proporciona una notación y una relación entre los componentes que caracterizan a un agente BDI.

A lo largo de este apartado describiremos cómo el mecanismo de razonamiento CBR (*Case Based Reasoning*), ya visto en el apartado 3.4.1, permite introducir un modelo que surte mecanismos de razonamiento a los agentes en los que se implementa. Se describirá cómo un sistema CBR puede incorporarse dentro de

un agente para proveerle de la capacidad de decidir y aprender autónomamente de la experiencia. Estos sistemas CBR proponen soluciones a un problema planteado basándose en la reutilización o la adaptación de soluciones almacenadas que fueron aplicadas de forma satisfactoria en similares problemas anteriores.

4.3.1 Agentes CBR- BDI, CBP-BDI

La relación entre sistemas CBR y agentes BDI puede ser establecida, asociando creencias, deseos e intenciones a casos. La ventaja de este enfoque es que un problema puede ser fácilmente conceptualizado en términos de agentes y, posteriormente, implementado en forma de un sistema CBR. En un agente BDI, se pone especial énfasis al razonamiento dirigido a la acción, es decir, a determinar qué debe hacer el agente [Wooldridge, 2002]. Este proceso involucra dos actividades fundamentales: a) determinar qué metas se desean alcanzar (deliberación) y b) decidir de qué manera alcanzar estas metas (planificación). Una de las principales carencias del modelo BDI que es el manejo de la memoria y de las experiencias pasadas. Ambos procesos, deliberación y planificación deben realizarse considerando las limitaciones de recursos de los agentes BDI, es decir, que los procesos de razonamiento deben tener en cuenta las restricciones impuestas por su falta de memoria y la efectividad de soluciones pasadas.

Para paliar la carencia en el manejo de situaciones pasadas y su repercusión en cuanto a efectividad y aprendizaje en el proceso e razonamiento de un modelo BDI, diversos autores [Glez-Bedia, 2004], [Bajo *et al.*, 2008], [Carrascosa *et al.*, 2008], [Corchado *et al.*, 2008], [Carneiro *et al.*, 2010] presentan un método de incorporación del motor CBR al modelo BDI. La idea principal de estos trabajos es la de utilizar los mecanismos que proporciona el modelo deliberativo BDI, como son creencias (*Beliefs*), deseos (*Desires*) e intenciones (*Intentions*), para poder obtener una representación de caso y poder ejecutar así el ciclo de razonamiento CBR.

Para poder integrar el sistema de razonamiento CBR dentro de la estructura de un agente deliberativo BDI, se plantea una formulación que permite relacionar el concepto de caso con los conceptos fundamentales BDI. La relación de identificación entre “estados intencionales” en el BDI y “estados CBR” permite plantear que agentes deliberativos, con arquitecturas BDI, puedan valerse de un ciclo de razonamiento tipo CBR, para generar soluciones S(P) a problemas P. Cuando el agente necesite resolver un problema, utilizará sus creencias, deseos e intenciones para construir su solución. Deseos previos, junto con creencias e intenciones, son almacenadas en forma de casos, y se convierten en material susceptible de recuperación para nuevos problemas. Los casos que puedan ser útiles al problema actual son seleccionados, y adaptados, generando el plan de

acción del agente que sería la nueva solución. Esta expresión es revisada y, si cumple con las propiedades requeridas, es almacenada, lo que constituye un mecanismo de aprendizaje asociado a cada proceso.

La formalización llevada a cabo para la integración de los dos modelos es la siguiente.

El entorno M y los cambios que se producen en dicho entorno se representan desde el punto de vista del agente. Así, el “mundo” del agente puede definirse como un conjunto de variables que influyen sobre los problemas que intenta resolver el agente y formalizarse de acuerdo a la ecuación (4.1).

$$M = \{\tau_1, \tau_2, \dots, \tau_s\} \text{ with } s < \infty \quad (4.1)$$

Las creencias del agente son objetos formados por algunos (o todos) de los atributos del mundo, y toman valores concretos, según la ecuación (4.2).

$$B = \{b_i / b_i = \{\tau_1^i, \tau_2^i, \dots, \tau_n^i\}, n \leq s \quad \forall i \in N\}_{i \in N} \subseteq M \quad (4.2)$$

Un estado del mundo $e_j \in E$ se representa como un conjunto de creencias que son ciertas en cierto instante de tiempo t , ecuación. Si $E = \{e_j\}_{j \in N}$ es el conjunto de estados del mundo en un instante concreto de tiempo t , entonces es posible obtener el conjunto de estados a lo largo del tiempo descrito por la ecuación (4.3).

$$e_j^t = \{b_1^{jt}, b_2^{jt}, \dots, b_r^{jt}\}_{r \in N} \subseteq B \quad \forall j, t \quad (4.3)$$

Los deseos u objetivos se imponen al comienzo de un problema, pero pueden cambiar a lo largo del tiempo. Se trata de aplicaciones que relacionan un estado del mundo actual y otro estado que se trata de alcanzar en un momento de tiempo posterior, definiendo dicho deseo de acuerdo a la ecuación (4.4).

$$d : \begin{matrix} E \\ e_0 \end{matrix} \rightarrow \begin{matrix} E \\ e^* \end{matrix} \quad (4.4)$$

Las intenciones son aplicaciones que indican la forma en la que el agente utiliza sus conocimientos para alcanzar sus objetivos. En el modelo CBR-BDI las intenciones deben determinar si existe compatibilidad entre los conocimientos que posee el agente y los requisitos necesarios para ser capaz de alcanzar sus deseos. De esta forma un deseo será alcanzable si se cumple la expresión presentadas en la ecuación (4.5)

$$i : \underset{(b_1, b_2, \dots, b_n, e_0)}{BxBx \cdots xBxE} \xrightarrow[n]{\quad} \underset{e}{E} \quad (4.5)$$

En el modelo CBR-BDI, las intenciones garantizan que se dispone de suficiente conocimiento en la base de creencias para ser capaz de alcanzar un deseo, proporcionando una solución.

Esta formulación permite relacionar cadenas de creencias con planes de acción y deseos con funciones solución y eficacia de la solución. Para integrar un mecanismo de razonamiento basado en casos en un agente BDI, el planteamiento que se sigue presenta un agente $A = \{\{b_i\}\{d_j\}\}$ que se enfrenta a un problema $P = (S_i, S_f)$ que debe solucionar para alcanzar sus objetivos. El agente necesita una intención $\{i_j\}$, esto es el uso adecuado de información combinada que posee en forma de creencias, para identificar la intención que constituye una solución $S(P)$. El agente consulta una memoria de casos tratando de encontrar una solución similar que ya hubiese sido utilizada en el pasado para resolver un problema de características similares a las de $S(P)$. El agente obtendrá la solución que considere más adecuada $S(P)$, y a partir de ella construye una intención o plan capaz de poner en marcha dicha solución. Es necesario tener en cuenta que tanto el problema como la descripción del entorno almacenados en un caso van a poder ser similares al actual, pero que al admitir dinamismo, nada nos dice que la evolución del entorno en el pasado que hizo de P una buena solución para $S(P)$ vaya a serlo en este caso.

Así pues, se dispone de un agente BDI con una serie de creencias que se encuentran almacenadas, junto con una serie de planes ejecutados en el pasado, en un motor de casos CBR. Formalmente se puede describir este sistema como:

1. Dada una base de casos o memoria de casos

$$C = (P, S(P), Eff)$$

Donde los problemas vienen definidos como $P = (S_i, \{\theta_j\}, S_f)$, siendo S_i el estado inicial del problema y S_f el estado final o estados objetivo que se desea alcanzar. $\{\theta_j\}$ con $j = 1, 2, \dots, m$ constituye el conjunto de operadores a partir de los que se construye la solución.

2. Las soluciones para los problemas estudiados, para las que se utiliza una representación matricial. $S(P) = \{S_k, \theta_h\}$, donde $k = 1, 2, \dots, n+1$ y $h = 1, 2, \dots, n$, con lo que $S_1 = S_i$ y $S_{n+1} = S_f$.

$$S(P) = \{S_1, \theta_1, S_2, \theta_2, \dots, \theta_n, S_{n+1}\}$$

que expresado en forma matricial:

$$S(P) = \left\{ \begin{matrix} (O(S_i)) \\ (R(S_i)) \end{matrix}, \begin{matrix} (O(S_f)) \\ (R(S_f)) \end{matrix} \right\}$$

donde los estados $O = O(O_1, O_2, \dots, O_n)$ forman el vector de objetivos y $R = R(R_1, R_2, \dots, R_n)$ el vector de recursos disponibles. Se utiliza esta notación teniendo en cuenta que los estados S_k y los operadores Θ_j se definen como:

$$S_k = \begin{pmatrix} \{O_r\}_{r=1,2,\dots,p} \\ \{R_s\}_{s=1,2,\dots,q} \end{pmatrix}$$

$$\Theta_j : S_k = \begin{pmatrix} \{O_r\} \\ \{R_s\} \end{pmatrix} \rightarrow \Theta_j(S_k) = \begin{pmatrix} \{O'_r\} \\ \{R'_s\} \end{pmatrix}$$

Siendo $\{O_r\}_{r=1,2,\dots,p}$ las variables con las que se representan los estados del problema.

3. La eficiencia o eficacia de una solución viene dada a través de restricciones sobre las relaciones que se establecen entre los objetivos a alcanzar y los recursos disponibles para alcanzar dichos objetivos.

$$\left. \begin{matrix} f(O_1, O_2, \dots, O_n) = 0 \\ g(R_1, R_2, \dots, R_m) = 0 \\ h(O_1, O_2, \dots, O_n, R_1, R_2, \dots, R_m) = 0 \end{matrix} \right\}$$

Se tiene entonces una relación que permite integrar los sistemas CBR en arquitecturas BDI. Dicha relación asocia creencias, deseos e intenciones con casos y su ventaja es que cualquier problema puede ser fácilmente conceptualizado en términos de agentes e implementado posteriormente en forma de CBR [Corchado y Laza, 2003].

Una especialización del agente CBR-BDI es el agente CBP-BDI [Corchado *et al.*, 2008]. Su estructura interna es similar, pero difiere del agente CBR-BDI en que la solución que lleva a cabo el agente no es un modelo, si no un plan, es decir, una secuencia de acciones. Un agente CBP-BDI debe incorporar un mecanismo de planificación que genere planes como soluciones a los problemas planteados. Muchos son los trabajos en el grupo de investigación BISITE que incorporan a un agente CBP-BDI diferentes mecanismos de planificación [Bajo *et al.*, 2007], [Bajo *et al.*, 2008], [Corchado *et al.*, 2007], [Corchado *et al.*, 2008].

En este sentido es reseñable el mecanismo de replanificación para los agentes CBP-BDI que presentado en [Corchado *et al.*, 2008] y cuya principal característica es la de ser capaz de replanificar en tiempo de ejecución. Un agente CBP-BDI calcula el plan o intención que es más fácilmente replanificable (MRPI). Esto es, aquel plan que puede ser más fácilmente sustituible por otro plan en caso de que se vea interrumpido (por ejemplo si un usuario cambia sus preferencias durante la ejecución del plan).

Las arquitecturas CBR-BDI y CBP-BDI expuestas resuelven uno de los problemas clásicos de arquitecturas deliberativas BDI como es la falta de capacidad de aprendizaje. El ciclo de razonamiento de un sistema CBR ayuda a los agentes a resolver problemas, facilita su adaptación a los cambios en el entorno e identifica nuevas posibles soluciones, mejorando y posibilitando importantes retos en las capacidades de razonamiento de los agentes. En el desarrollo del presente trabajo hacemos uso de agentes CBR-BDI. Dicha aplicación de agentes CBR-BDI permite resolver el problema de la recomendación de contenidos a partir de los mecanismos de aprendizaje incorporados por el CBR a través de la base de casos que relacionan el perfil de usuario con intereses afines en las búsquedas de contenidos digitales educativos.

4.4 SISTEMAS MULTI-AGENTE

Se habla de un sistema multi-agente (SMA), cuando dos o más agentes son capaces de trabajar de forma conjunta con el objetivo de resolver un problema [Mas, 2005]. Es necesario matizar esta afirmación, ya que en realidad, para que una asociación entre agentes que trabajan de forma conjunta sea considerada un SMA al menos uno de los agentes debe de ser autónomo y debe existir al menos una relación entre dos agentes en la que se cumpla que uno de los ellos satisface los objetivos del otro.

La característica principal de este tipo de sistemas es que no existe un sistema establecido de control global y los datos se encuentran organizados de forma distribuida (descentralizados), favoreciendo su computación asíncrona. Asimismo, cada agente puede decidir con libertad, dinámicamente, qué tareas debe efectuar y a quién asigna dichas tareas [Wooldridge, 2002].

En un SMA se hace necesario que los agentes trabajen de forma coordinada para alcanzar sus objetivos. El trabajo coordinado requiere mecanismos de comunicación, lo que permite establecer mecanismos de negociación entre los distintos agentes de un SMA [Ossowski y García Serrano, 1997] [Ossowski y Menezes, 2006].

Cuando hablamos de SMA, extendemos la idea de un agente en solitario completándola con una infraestructura para la interacción y comunicación. Los SMA idealmente tienen las siguientes características [Huhns y Stephens, 1999]:

- Son típicamente abiertos y tienen un diseño no centralizado.
- Contienen agentes autónomos, heterogéneos y distribuidos, con diferentes "personalidades" (cooperativo, egoísta, honesto, etc.).
- Proporcionan una infraestructura para especificar comunicación y protocolos de interacción.

En un SMA aparecen conceptos de gran interés: actividades conjuntas y cooperación, la aparición de conflictos y su resolución, aspectos de negociación, comunicación, compromisos y planificación de actividades, generación del modelo de conocimiento común, etc.

Un sistema multi-agente podría concebirse como una organización artificial formada por individuos dotados de ciertas habilidades. Los agentes así concebidos pueden compartir objetivos comunes o cada uno de ellos puede tener asignadas metas diferentes. Si la actitud es cooperativa pueden interactuar para brindar y recibir colaboración. La colaboración resulta particularmente útil cuando los agentes tienen habilidades diferentes que permiten resolver problemas que ninguno de los miembros de la organización podría resolver individualmente. Como contrapartida, exige solventar los conflictos que puedan surgir como consecuencia de la interacción entre dichos agentes dentro del sistema multi-agente.

4.4.1 *Interacción entre agentes en un SMA*

Cuando un agente forma parte de un sistema multi-agente, puede interactuar con otros miembros del sistema en el que participa y aprovechar sus creencias y sus habilidades. Dicha interacción aumenta las posibilidades de alcanzar las metas comprometidas, pero exige como contrapartida considerar el impacto que el comportamiento social provoca sobre la representación del conocimiento, el mecanismo deliberativo, la planificación y el modelo de comportamiento de cada agente individual en el conjunto del sistema. Es por ello que se hace necesario desarrollar tareas de comunicación, coordinación y negociación en el SMA. Para que los agentes puedan interactuar de forma coherente, deben compartir información sobre sus objetivos y tareas. Gracias al intercambio de esta información, los agentes coordinan el desarrollo de actividades, incorporando capacidades de negociación en caso de que aparezcan conflictos y de planificar sus acciones mediante la adaptación correspondiente de tareas para llevar a cabo un objetivo.

En las interacciones de un SMA hay que distinguir varios conceptos muy relacionados entre sí, pero que hacen referencia a características diferentes: comunicación, coordinación, cooperación, negociación y adaptación.

- **Comunicación.** Es la habilidad de los agentes para comunicarse entre sí, esto es, intercambiar información y conocimiento de forma comprensible. Permite a los agentes obtener el conocimiento necesario para decidir la secuencia de acciones que debe ejecutar en función de sus objetivos.
- **Coordinación.** La coordinación es una característica clave en el desarrollo de SMA. [Malone, 1988] [Malone y Crowston, 1994] describen la coordinación de acciones como un conjunto de acciones suplementarias que pueden realizarse en un entorno multi-agente para alcanzar un objetivo y que un agente, con los mismos objetivos, no podría alcanzar por sí solo.
- **Cooperación y negociación.** La cooperación es el mecanismo por el cual los agentes, que trabajan juntos para lograr un objetivo común, definen una estrategia (negociación) para alcanzar este objetivo.
- **Adaptación.** La adaptación en un SMA permite desarrollar sistemas con capacidad de reorganización y acomodo a los cambios en el entorno y que sean capaces de evolucionar en tiempo de ejecución. Los SMA permiten en las organizaciones dinámicas de agentes que los forman, mecanismos de auto-adaptación que les habilitan para operar en un entorno cambiante [Weyns & Georgeff, 2010]. Las claves de la construcción de este tipo de SMA son su capacidad de adaptación basada en el aprendizaje.

[Rodríguez *et al.*, 2009] propone una tipología con cuatro tipos de sistemas multi-agente adaptativos, en función de los mecanismos de interacción utilizados: mecanismos basados en interacciones directas, indirectas, basados en el refuerzo o basados en mecanismos de cooperación.

Los SMA se preocupan por las interacciones de los agentes que lo forman. Estos agentes forman parte de una colección y pueden coordinar su conocimiento, objetivos, habilidades y planes juntamente para tomar una acción o resolver una meta global. En todos los sistemas debe haber un proceso de racionalización para la coordinación del conjunto de agentes. Por lo general en estos sistemas, los agentes, con sus creencias, deseos e intenciones construyen el problema y el plan o secuencia de acciones para solucionarlo.

4.5 ORGANIZACIONES DE AGENTES

Los SMA deben permitir la participación de todo tipo de agentes, incluyendo distintas arquitecturas y lenguajes [Zambonelli *et al.*, 2003]. Por esta razón, no se puede confiar en el comportamiento de los agentes, siendo necesario establecer controles en base a normas o reglas sociales. Actualmente, las investigaciones centradas en el diseño de SMA desde el punto de vista

organizacional está ganando terreno [Rodríguez, 2010] [Andrade *et al.*, 2008]. Cada vez es más fuerte la idea de que no sólo se puede modelar la interacción entre un sistema de agentes a través de la interacción de sus miembros, sino que es necesario modelar también el comportamiento desde un punto de vista organizativo.

Los agentes en un sistema multi-agente basado en conceptos organizacionales se coordinan e intercambian servicios e información de forma conjunta, mediante procesos de negociación que dan lugar a acuerdos, e incluso pueden llevar a cabo otras acciones sociales más complejas. Los conceptos de reglas [Zambonelli, 2002], normas e instituciones [Esteva *et al.*, 2001] y estructuras sociales [Parunak y Odell, 2002] nacen de la idea de que es necesario un nivel de abstracción mayor, independiente del agente, y que defina explícitamente la organización en la que los agentes viven.

Para comprender este nuevo enfoque de interacción entre agentes resulta necesario aclarar un concepto clave, el de sociedad [Mauro, 2003]:

"Una sociedad artificial está definida como un conjunto de entidades artificiales interrelacionadas e interactuantes, que se rigen bajo determinadas reglas y condiciones"

La funcionalidad principal de una sociedad es permitir a sus miembros coexistir en un entorno compartido y llevar a cabo sus respectivos objetivos cooperando (o no) con el resto de miembros. La estructura está determinada por los roles, reglas de interacción y lenguaje de comunicación entre sus miembros. La ventaja principal del desarrollo de SMA desde el punto de vista social es que permite crear sistemas heterogéneos en cuanto a lenguajes, aplicaciones y características. Las arquitecturas que ayudan a modelar y construir SMA basados en organizaciones deben soportar marcos de coordinación entre agentes, además de ser capaces de adaptarse dinámicamente a cambios en su estructura, objetivos o interacciones [Dignum, 2004].

Las sociedades de agentes orientadas a la organización requieren una visión colectivista de la relación entre el agente y el entorno [Edmonds, 1999]. Un agente dentro de una sociedad necesita considerar no solamente su propio comportamiento sino también el comportamiento del sistema como un todo y cómo los agentes influyen unos en otros.

En [Davidsson, 2001] se propone una clasificación para las sociedades artificiales basándose en las siguientes características: (i) *Apertura*. Describe las posibilidades de cualquier agente de unirse a la sociedad. (ii) *Flexibilidad*. Indica el grado en el que el comportamiento del agente está restringido por las normas de la sociedad. (iii) *Estabilidad*. Define la previsibilidad de las

consecuencias de las acciones, y (iv) *Fidelidad o Confianza*. Especifica el grado en que los agentes pueden confiar en la sociedad.

La sociedad necesita soportar todas las anteriores características en diferentes grados atendiendo al propósito para el que sea creada. En función de dicho cumplimiento se pueden clasificar las sociedades en dos grandes grupos, las sociedades denominadas abiertas y las cerradas.

Las denominadas *sociedades abiertas*, no imponen restricciones a los agentes que forman la sociedad. Popper [Popper, 1982] define las sociedades abiertas, como sistemas en un estado, lejos de un equilibrio, que no muestran ninguna tendencia hacia un aumento en el desorden. De esta manera se exaltan características tales como la flexibilidad y el nivel de apertura, pero por otro lado se perjudican características como la estabilidad y la confiabilidad, debido a que no se tiene ningún control sobre qué entes entran en la sociedad y mucho menos sobre el rol que desarrollarán.

Por otro lado están las denominadas *sociedades cerradas* como contraposición a las abiertas. Se caracterizan por tener una gran confiabilidad y estabilidad, debido a sus políticas en las que ningún agente externo puede acceder; en este sentido, también las reglas que gobiernan el sistema están bien delimitadas con el objetivo de cumplir las metas especificadas. Todos los agentes de este tipo de sociedad persiguen el mismo objetivo común, aunque pueden tener metas locales.

[Davidsson, 2001] introduce dos nuevos tipos de sociedades agente, semi-abiertas y semi-cerradas, que combinan la flexibilidad de las sociedades abiertas con la estabilidad de las sociedades cerradas. Este equilibrio entre la flexibilidad y la estabilidad de resultados se logra mediante mecanismos que hacen cumplir la conducta ética entre los agentes. Así las *sociedades semi-abiertas* se determinan por tener cierto nivel de equilibrio respecto a las características anteriormente nombradas, debido a que para poder acceder a éstas es necesario, primero hacer contacto con un mecanismo de control de admisión, el cual evaluará el ente que solicita la entrada a la sociedad. De esta forma se permitirá el ingreso tan solo a entes que parezcan confiables y capaces ante el mecanismo de decisión. Las sociedades semi-abiertas limitan ligeramente la apertura y flexibilidad de las sociedades abiertas, pero son capaces de dar una mayor estabilidad y confianza.

Al igual que las sociedades cerradas las *sociedades semi-cerradas*, tampoco permiten el acceso a ningún nuevo ente a la sociedad, pero por medio de un mecanismo que comunica el exterior con el interior de la sociedad se puede solicitar la creación de un nuevo agente en el interior de la sociedad, el cual funcionará como un representante del agente externo. Este representante no es a imagen y semejanza del agente externo sino que es creado basado en los roles

existentes dentro de la sociedad. Esto amplía la flexibilidad y la apertura de la sociedad, sin perder la estabilidad y confiabilidad, ya que los agentes participantes están diseñados siguiendo los requerimientos de la sociedad y el propietario de la sociedad todavía controla la arquitectura general del sistema.

4.5.1 *Concepto de Organización Virtual*

Los actuales desarrollos informáticos con entornos de trabajo colaborativo y gestión distribuida del conocimiento a través de arquitecturas destinadas a compartir recursos y servicios han conducido a un mayor interés en lo que se ha denominado las *organizaciones virtuales*: un conjunto de personas e instituciones que necesitan coordinar los recursos y servicios a través de las diferentes fronteras institucionales [Foster *et al.*, 2001]. En la actualidad, existe un amplio abanico de investigaciones enfocadas a la construcción de organizaciones virtuales, a la simulación social basada en agentes y al estudio de su comportamiento [Castelfranchi, 2000] [Criado *et al.*, 2009] [Pavón *et al.*, 2008] [Iglesias, 2010]. La analogía entre sociedades de agentes y sociedades humanas es clara, por ello es necesario definir el término “organización” desde los dos puntos de vista.

Desde el punto de vista humano, una organización se define como *una formación o entidad social con un número de miembros que puede ser precisado y una diferenciación interna de las funciones que son desempeñadas por dichos miembros* [Peiro, 1990]. Por otro lado, un organización de agentes es *una colección de roles, que mantienen ciertas relaciones entre si, y que toman parte en patrones de interacción con otros roles de forma institucionalizada y sistemática* [Zambonelli *et al.* 2003].

Taylor en su teoría estructuralista [Taylor, 1911] modelaba las organizaciones como sistemas cerrados. Sin embargo en las organizaciones virtuales Spencer [Spencer, 1896] encuentra una serie de características como: (1) El crecimiento. (2) El hecho de volverse más complejas a medida que crecen. (3) El hecho de que haciéndose más complejas, sus partes exigen una creciente interdependencia. (4) Su vida tiene mayor extensión comparada con la vida de sus unidades componentes. (5) Una creciente integración es acompañada por creciente heterogeneidad. Todas ellas son características de los sistemas abiertos, por lo que podemos afirmar que las organizaciones son claramente sistemas abiertos.

Si nos centramos en el paradigma computacional, los sistemas pueden ser vistos, como entidades (comúnmente agentes) que proveen y consumen recursos. También es habitual que hayan sido diseñadas por distintos equipos de desarrollo así como que pueden entrar o salir de una organización en diferentes momentos por diferentes razones. Además pueden formar coaliciones u organizaciones entre ellas y tener unos mismos objetivos. Es por

ello que se define una Organización Virtual (OV) como [Foster *et al.*, 2001] [Boella *et al.*, 2005]:

"Una Organización Virtual es un conjunto de individuos e instituciones que necesitan coordinar sus recursos y servicios dentro de unos límites institucionales".

Por lo tanto, una OV es un sistema abierto formado por la agrupación y la colaboración de entidades heterogéneas y donde hay una separación entre la forma y la función que define su comportamiento.

Las Organizaciones Virtuales son un mecanismo adecuado para permitir la coordinación de agentes heterogéneos en entornos abiertos. Teniendo en cuenta muchos conceptos de la Teoría de Organización Humana, se desarrolla un modelo para Organizaciones Virtuales. Este modelo describe los aspectos estructurales, funcionales, normativos y ambientales del sistema. Está basado en cuatro conceptos principales: unidad de organización, servicio, norma y entorno.

El modelado de OV basadas en SMA abiertos es particularmente adecuado para el desarrollo de sistemas en entornos dinámicos y heterogéneos como el descrito en el entorno educativo para la recuperación de contenidos objeto de este trabajo. Los SMA modelados con estas características de OV no sólo posibilitan describir la composición estructural del sistema, sino también las normas y reglas para el control del comportamiento de los agentes, la entrada/salida dinámica de los componentes y la formación, también dinámica, de los grupos de agentes.

4.5.2 *Propiedades de una organización virtual*

Una OV es descrita por una estructura social y un conjunto de normas que regulan la interacción entre los agentes. Tal descripción identifica los componentes funcionales del sistema (agentes), sus responsabilidades (tareas) y sus recursos (conocimiento, software, hardware, etc.). Existen una serie de conceptos que definen y describen la noción de organización virtual de agentes tales como el concepto de Entidad Social, el tipo de estructura de la organización, su funcionalidad, normatividad, el entorno en el que se desarrollan, la dinamicidad del modelo, su adaptación y su aprendizaje social, su mecanismo de coordinación, etc. Estos diez elementos resultan fundamentales en la conceptualización de una organización virtual. A continuación se describirán brevemente.

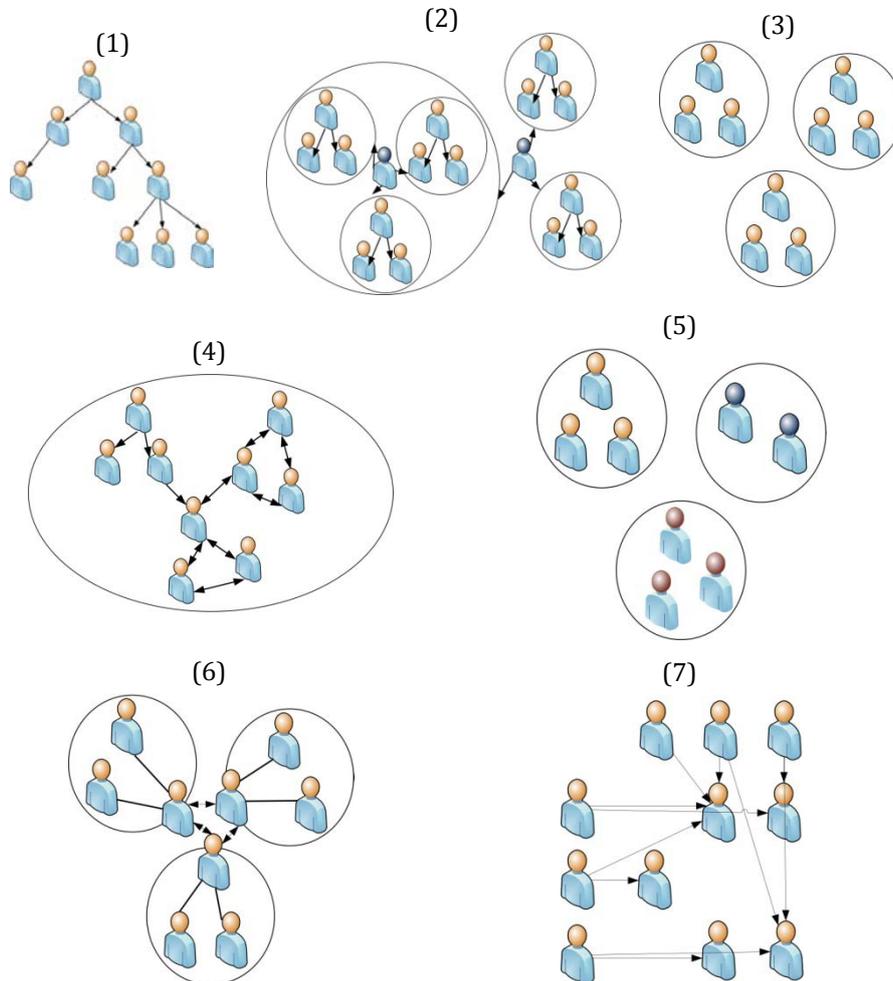


Figura 4-4. Topología de organización. De izquierda-derecha y de arriba-abajo (1) Jerarquía, (2) Holarquía, (3) Coaliciones, (4) Grupos, (5) Congregaciones. (6) Federaciones y (7) Organizaciones Matriciales

- I. Las organizaciones están formadas por componentes o **Entidades Sociales** que pueden estar compuestas a su vez por un número específico de miembros o agentes. Según [Pattison *et al.*, 1987] estas entidades tienen responsabilidades, tienen y consumen recursos, están estructuradas, tratan de alcanzar los objetivos globales y están regulados por normas y restricciones.
- II. Por otro lado las entidades de una organización no son independientes unas de otras debido a que interactúan entre sí. Estas relaciones no vienen dadas de manera individual dentro de una organización, sino que se requiere un conjunto de relaciones entre grupos de entidades,

generando una **Estructura**. La estructura puede definirse como la distribución, orden e interrelación de las diferentes partes que componen la organización. En dicha estructura, los agentes se ordenarán y comunicarán dependiendo de la topología en la que se encuentren estructurados. Existen diferentes tipos de tipologías siendo las siete siguientes las más relevantes:

- (i) *Jerarquías*. Los agentes organizand en una estructura de tipo árbol en la que los niveles inferiores tienen la funcionalidad básica y los niveles superiores la toma de decisiones.
 - (ii) *Holarquías*. Está formada por una anidación de holones. Cada uno de estos holones a su vez está formado por una jerarquía de entes.
 - (iii) *Coaliciones*. Son grupos temporales de agentes que se unen para la consecución de un objetivo concreto, que suele tener ciertos beneficios o reducir costes. Internamente suele representarse como una estructura plana o con un líder y externamente como una entidad única y atómica.
 - (iv) *Grupos*. Agrupaciones de agentes cooperativos que trabajan juntos en la consecución de un objetivo común, maximizando de esta forma la utilidad del equipo. Los grupos implican una mayor redundancia y flexibilidad para entornos inciertos, aunque también un aumento de las comunicaciones.
 - (v) *Congregaciones*. Agrupaciones de agentes con características similares. En este caso, no conlleva la consecución de un objetivo específico pero sí facilita la búsqueda de colaboradores adecuados para lograr alcanzar ese objetivo.
 - (vi) *Federaciones*. Agrupaciones de agentes con un representante. Los miembros interactúan sólo con el representante, que también actúa como intermediario entre el grupo y el exterior.
 - (vii) *Organizaciones matriciales*. Un agente puede ser controlado por más de un agente superior, por lo que es necesaria la utilización de mecanismos de evaluación de compromisos y resolución de conflictos locales.
- III. La **Funcionalidad** de una organización está determinada por su misión, que a su vez define la estrategia, los requerimientos funcionales y de interacción.
- IV. Las **Normatividad** o normas sociales, definen las consecuencias de las acciones de los agentes:
- Restricciones sobre la organización.
 - Obligaciones y sanciones a aplicar en el caso de no cumplirlas.
 - Control de acceso externo.
- V. La definición el **Entorno** de la organización (recursos, aplicaciones, supuestos, restricciones, *stakeholders*, etc.) establece la relación de los roles respecto al medio en el que se desarrolla dicha organización.

- VI. La **Dinámica** o dinamicidad de una organización está relacionada con la entrada y salida de agentes, así como con la adopción de roles por parte de los mismos, la creación de grupos y el control del comportamiento. Para definir la dinámica será necesario tener en cuenta:
- *La entrada de agentes.* Es necesario definir cuándo permitir a los agentes entrar en la OV, cuál será su posición y cuáles serán los procesos de expulsión de agentes con comportamientos anómalos.
 - *La adopción de roles.* Es necesario definir cómo los agentes adoptan un rol concreto y la asociación de agentes en roles.
 - *La creación dinámica de grupos.* Definir federaciones, coaliciones, congregaciones, etc.
 - *El control de comportamiento.* Es necesario definir como controlar la conformidad del comportamiento de los agentes a las normas de la sociedad.
- VII. La **Adaptación Social** consiste en la capacidad para involucrarse en el entorno y hacer una simbiosis que permita a ambos implicados el disponer el uno del otro.
- VIII. El **Aprendizaje social** resulta el proceso por el cual, gracias a un ambiente común, diferentes entes pueden interactuar y evaluar sus experiencias conjuntas con el objetivo de aprender a través de la experiencia pasadas de los entes de la sociedad. Un ente que aprende de otro ente, está ahorrando mucho esfuerzo en experimentación y tiempo, debido a que la información que captó es obtenida mediante el esfuerzo de otros. En función de los objetivos de aprendizaje propuestos se define la pertinencia de los datos encontrados y de ésta manera decidir que aprender y que no [Conte y Paolucci, 2001].
- Los dos tipos de aprendizaje en sociedad más relevantes son:
- IX. *Facilitación social.* El aprendizaje actualiza la base de conocimiento del agente que obtiene la nueva información por medio de la observación de lo que sucede a otro agente, mientras éste interactúa con el medio. Las repercusiones del entorno en respuesta a las acciones de un agente son la base de referencia para saber qué utilidad representan determinados comportamientos. Al observar las repercusiones del entorno en respuesta a las acciones de un agente, los demás tendrán un punto de referencia para saber qué utilidad representan determinados comportamientos [Mataric, 1997].
- X. *Aprendizaje imitativo.* El agente que aprende también tiene en cuenta la experiencia que recibe del medio después de interactuar con él. Es decir, después de que el agente percibe las acciones de otro agente que interactuó con el medio (facilitación social), éste no tomará la acción como totalmente cierta, sino que primero probará imitando la situación presentada, de ésta manera se verificará y se tendrá una mayor seguridad acerca de la validez de la información captada inicialmente. Es importante aclarar que el hecho de imitar a otro ente va ligado directamente al nivel de “confianza” que se tenga sobre éste

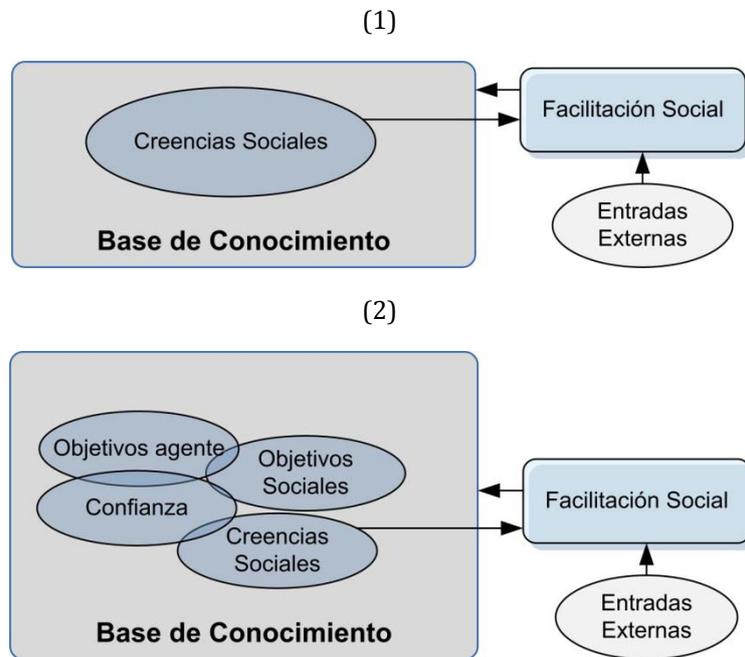


Figura 4-5. Aprendizaje por Facilitación social (1) y por Imitación (2)

El aprendizaje como vemos, está intrínsecamente relacionado con la adaptación. Tanto la facilitación social como la imitación son técnicas que nos permiten aprender de nuevas situaciones para poder actuar de cierta manera en el futuro. En definitiva, eso puede considerarse un proceso de adaptación a nuevas situaciones.

- XI. La **Coordinación** es un *mecanismo de coordinación* que determina el modo en el que uno o varios agentes realizan las tareas [Ossowski, 1998]. Cada organización necesita de un soporte de coordinación que determine explícitamente cómo deben organizarse y llevarse cabo las acciones y tareas dentro de la misma.

Los agentes coordinan sus conocimientos, objetivos, habilidades, tareas y planes conjuntamente para tomar una acción o resolver una meta global, así, de este modo se consideran y regulan todas las tareas a realizar y no se corre el peligro de ejecutar acciones no deseables. Desde una perspectiva práctica, es mejor, probablemente, concebir la coordinación como el esfuerzo de administrar el *espacio de interacción* [Wegner, 1997] [Busi *et al.*, 2001] de un SMA. Esta coordinación está relacionada con la planificación de acciones para la resolución de tareas, porque estos planes permiten:

- Conocer a alto nivel y predecir el comportamiento de otros agentes del sistema.
- Intercambiar resultados intermedios que lleven al progreso en la resolución de la tarea global.
- Evitar acciones redundantes, si ellas no son deseables.

Existen diferentes modelos de coordinación [Malone y Crowston, 1994] [Fox y Gruniger, 1998] [Cohen y Levesque, 1991] [Dunin-Keplicz, Verbrugge, 2002], la elección del modelo de coordinación dependerá del ámbito de actuación de la propia organización, pero siempre se debe intentar cubrir todas las posibilidades de actuación de los agentes.

- XII. **Adaptación.** Para que una organización pueda adaptarse rápidamente a los cambios en su entorno, los agentes deben poder coordinarse cuando sea necesario realizar cambios en sus metas o en los roles que tienen asignados. Es decir, es necesario que *la organización se adapte*.

Una OV puede verse por tanto como un sistema cooperativo, en el que la coordinación se basa en una planificación y distribución de tareas. La coordinación de tareas compartidas o tareas que se combinan para resolver un problema común requerirá o bien de una planificación centralizada, o bien de una planificación distribuida llevada a cabo por los propios agentes de forma individualizada. Esto constituye un problema abierto, ya que los sistemas clásicos de planificación [Sacerdoti, 1975] [McAllester y Rosenblitt, 1991] [Penberthy y Weld, 1992] no son apropiados por varios motivos: (i) Asumen que los agentes tienen un conocimiento completo de su entorno. (ii) Asumen que las acciones no van a fallar y finalmente (iii) Asumen que el entorno cambiará únicamente por la ejecución de las acciones del agente.

Ninguna de estas tres suposiciones puede realizarse de forma realista en un SMA en un entorno de desarrollo real, como el que plantea este trabajo, por lo tanto se requiere de una planificación que pueda adaptarse a cualquiera de las circunstancias del contexto en el que se desarrolla su funcionalidad. En este sentido se debe determinar cómo se debe adaptar la OV y de quién es la responsabilidad de la adaptación. Dichas cuestiones están determinadas por el objetivo que la OV deba o quiera cumplir, factor que se ve influenciado ampliamente por la convivencia en sociedad, el tipo de estructura social adoptada y la normativa y dinámica diseñada. De esta manera, las decisiones a tomar en el diseño de la organización estarán directamente relacionadas con ciertas funciones de utilidad o eficiencia a través de un beneficio individual o colectivo que afectará a la dinamicidad de la organización a través de procesos normativos que permitan su adaptación al entorno.

4.5.3 Entornos de desarrollo de Organizaciones Virtuales de Agentes

Las OV pueden considerarse sistemas abiertos formados por la agrupación y colaboración de entidades heterogéneas, dónde existe una clara separación entre estructura y funcionalidad [Foster *et al.*, 2001] [Boella *et al.*, 2005].

Podemos encontrar trabajos enfocados en el desarrollo de nuevas metodologías y procedimientos de diseño centrados en los aspectos de OV dentro los SMA como Gaia [Zambonelli *et al.*, 2003], Tropos [Bresciani *et al.*, 2004], OperA [Dignum, 2004], INGENIAS [Pavón *et al.*, 2005], CONOISE [Norman *et al.*, 2003] y otros [Ferber *et al.*, 2004] [Hubner, 2004] [Criado *et al.*, 2009] [Cosentino, 2005]. Actualmente, las abstracciones y herramientas disponibles todavía no son lo suficientemente potentes para desarrollar sistemas abiertos que traten con problemas dinámicos reales. La mayoría de los trabajos referenciados son, en cierto modo, incompletos, ya que no incluyen ni todas las etapas, ni todas las necesidades del desarrollo de sistemas de este tipo. Estas carencias se acentúan en las últimas fases del desarrollo donde se tienen que traducir las propiedades de la organización, las especificaciones del diseño y el análisis en código ejecutable.

La función primordial de una plataforma de agentes es ofrecer un entorno de ejecución para los agentes. El principal problema para la implementación de una OV es la falta de plataformas que den soporte a este tipo de sistemas. Ejemplos de plataformas de agentes son JADE²⁵, FIPA-OS²⁶, RETSINA [Giampapa y Sycara, 2002], una comparación de las mismas puede verse en [Argente *et al.*, 2004]. La mayoría de las plataformas nombradas tienen en cuenta aspectos organizacionales, pero en general únicamente en fases de análisis y diseño, por lo que no pueden utilizarse directamente para el desarrollo de SMA abiertos donde las estructuras organizacionales pueden emerger y cambiar dinámicamente en tiempo de ejecución.

Existe una arquitectura que sí incluye todas las características, se trata de THOMAS (*MeTHods, Techniques and Tools for Open Multi-Agent Systems*) [Carrascosa *et al.*, 2009] [Giret *et al.*, 2009]. THOMAS es una arquitectura abstracta para SMA abiertos a gran escala, basada en un enfoque orientado a servicios y direccionada principalmente al diseño de organizaciones virtuales [GTI-IA, 2009]. La arquitectura THOMAS, se fundamenta en varias áreas de investigación, entre ellas destacan la computación orientada a servicios, la estructura organizacional de las comunidades y las plataformas de agentes. Desde un punto de vista global, la arquitectura THOMAS, plantea una

²⁵ JADE <http://jade.tilab.com/>

²⁶ FIPA-OS <http://fipa-os.sourceforge.net/>

integración total de forma que los agentes puedan ofrecer e invocar servicios de forma transparente a otros agentes o entidades, así como que entidades externas puedan interactuar con los agentes de la arquitectura mediante el uso de los servicios ofertados. El apartado C.2, ofrece una descripción más detallada de los recursos de desarrollo que ofrece la arquitectura THOMAS.

THOMAS actualmente continúa su desarrollo hacia su nueva versión, denominada OVAMAH (*Multiagent-based Adaptive Virtual Organizations*) [Rodríguez *et al.*, 2009]. OVAMAH evoluciona a partir de su antecesora THOMAS realizando las modificaciones necesarias para que puedan ser implementadas organizaciones virtuales adaptativas. Estas propiedades son definidas desde un detallado estudio del estado del arte de métodos de desarrollo, trabajos y herramientas relacionadas con las organizaciones virtuales y su adaptación en [Rodríguez, 2010].

4.5.4 Metodologías de desarrollo de Organizaciones Virtuales

El desarrollo de sistemas multi-agente se ha planteado tradicionalmente a partir de un diseño centrado en el estudio de cada agente de forma aislada e independiente, orientado hacia su estructura y comportamiento. La integración con otros agentes se llevaba a cabo mediante plataformas de localización de agentes a través de servicios de direccionamiento y descripción de los mismos, tal y como propone el estándar FIPA (*Foundation for Intelligent Physical Agents*). La separación de los agentes del framework o herramienta de desarrollo y la inclusión de aspectos de organización, constituyen un avance reciente en los modelos de desarrollo de agentes.

En este trabajo se presenta un enfoque nuevo en el que el diseño se está dirigiendo hacia los aspectos organizativos de los agentes donde aparecen conceptos como grupo, roles, funciones, etc. estableciéndose dos niveles descriptivos: el de la organización y el del agente [Boissier, 2007]. Los agentes juegan papeles dentro de una organización; sin embargo, ellos no son la organización. La organización define el ajuste social en el cual el agente debe existir. Una organización incluye una estructura así como reglas, que obligan a un comportamiento válido del agente y a su interacción dentro de dicha organización. Este nuevo enfoque plantea dos metodologías de desarrollo diferentes: orientada a la organización y orientada al agente. Esto conlleva un impacto significativo sobre el desarrollo de metodologías de desarrollo de sistemas multi-agente previas como Gaia [Zambonelli *et al.*, 2003], MaSE [DeLoach, 1999] o MESSAGE [MESSAGE, 2000].

En las metodologías orientadas al agente, el diseñador se centra en las acciones individuales de los agentes y las estructuras sociales no se modelan, se forman con la interacción de los mismos. Ejemplos de estas metodologías son los inicios de la mayoría de las metodologías y herramientas de ingeniería orientadas a agentes. Así, GAIA, en su primera versión [Wooldridge *et al.*, 2000],

el inicio de MaSE [DeLoach, 1999], MESSAGE [Evans *et al.*, 2001], los inicios de ADELFE [Bernon *et al.*, 2003], INGENIAS [Pavón and Gómez-Sanz, 2003], TROPOS, en su primera versión [Giunchiglia *et al.*, 2002] y PROMETHEUS [Padgham and Winikoff, 2002] entre otras.

En las metodologías orientadas a la organización el diseñador desde un principio se centra en la organización del sistema, analizando el SMA desde una perspectiva global donde los objetivos de la organización representan una descripción a alto nivel de los propósitos de la sociedad.

La incorporación de los aspectos de organización en las metodologías de sistemas de multi-agente existentes se hace incorporando distintos aspectos. Por ejemplo la extensión de Gaia incorpora leyes sociales [Zambonelli *et al.*, 2001] y conceptos de organización [Zambonelli *et al.*, 2003] como también lo hace OPERA [Dignum *et al.*, 2005]. Mientras, otros proponen la organización como un agente institucional separado [Wagner, 2000]. MaSE se amplía hacia la denominada Organization-based MaSE (O-MaSE) mediante reglas de entidades ambientales [DeLoach, 2002a], [DeLoach, 2002b]. Otros ejemplos de estas metodologías son AGENT-GROUP-ROLE [Ferber *et al.*, 2003] [Ferber *et al.*, 2005], otra extensión de Gaia denominada ROADMAP [Juan y Sterling, 2002] y la extensión de TROPOS [Bryl *et al.*, 2009], todas ellas de tipo estructura organizativa.

Otra serie de propuestas establecen mecanismos para construir sistemas multi-agente altamente adaptativos como Adelfe, que incorpora interacciones sociales orientadas a la ética y sigue la teoría *Artificial Moral AgentS* (AMAS) [Bernon *et al.*, 2005], [Picard and Gleizes, 2004]. El objetivo de métodos como Adelfe es permitir a los diseñadores construir sistemas que con los mecanismos de organización implementados puedan producir alguna funcionalidad desconocida, no planificada previamente.

GORMAS (*Guidelines for Organization-based MultiAgent Systems*) [Argente *et al.*, 2008] es una guía metodológica para el análisis y el diseño de sistemas multi-agente abiertos desde la perspectiva de las organizaciones humanas. GORMAS consta de un conjunto de fases que cubren el análisis, el diseño de la estructura organizativa y el diseño de la dinámica de la organización. Con estas fases se especifica cuáles son los servicios que ofrece la organización, cuál es su estructura interna y qué normas rigen su comportamiento.

En el sistema desarrollado se ha tomado como ayuda la metodología de organizaciones GORMAS, lo que ha permitido definir los objetivos de la organización, su estructura, su dinámica y el entorno. Es decir, ha permitido diseñar un sistema abierto donde los agentes participantes tienen la capacidad de realizar búsquedas federadas sobre diferentes repositorios de objetos de

aprendizaje, permitiendo además la catalogación y ordenación de los resultados de las citadas búsquedas.

4.6 CONCLUSIONES DEL CAPÍTULO

La tendencia actual en el desarrollo de sistemas va enfocada hacia modelos no centralizados, distribuidos, con la implantación de paradigmas de computación de tipo *grid*, *clustering*, arquitecturas basadas en servicios, etc. Estos nuevos modelos conllevan la generación de software que contemple la autonomía, la robustez, la flexibilidad y la adaptabilidad de los sistemas a desarrollar. En este sentido, las complejas características de estos entornos abiertos y dinámicos dificultan la tarea de recuperación de contenidos. Por este motivo se hace necesario proponer nuevos enfoques que incluyan teorías, modelos, mecanismos y herramientas que permitan desarrollar sistemas con capacidad de reorganización y adaptación a los cambios en el entorno y que sean capaces de evolucionar en tiempo de ejecución.

El rápido crecimiento de internet es el más obvio signo del cambio en los modelos de interacción en sistemas distribuidos. El fenómeno de la interacción constituye una parte fundamental en la investigación del modelo de agentes. Resultan de particular interés aquellos sistemas que pueden cooperar y negociar unos con otros. En la evolución de internet a lo largo de las próximas décadas, la posibilidad de sistemas autónomos que puedan cooperar y negociar generará enormes oportunidades en áreas como el comercio electrónico, la gestión de información en la red, etc.

Las organizaciones virtuales permiten a los agentes que las forman, mecanismos de auto-adaptación que les habilitan para operar en un entorno cambiante. Estas organizaciones podrían aparecer en forma de sistemas de agentes dinámicos o emergentes, tales como las sugeridas por los dominios *grid*, ambientes inteligentes u otros en los cuales los agentes se agrupan de forma dinámica para ofrecer servicios compuestos y que pueden evolucionar en el tiempo.

Como hemos visto, cada vez es más frecuente modelar SMA no solamente desde el punto de vista individual del agente y sus capacidades de comunicación, si no que es necesario una ingeniería organizacional. La funcionalidad principal de una sociedad es permitir a sus miembros coexistir en un entorno compartido y llevar a cabo sus respectivos objetivos cooperando o no con el resto de miembros. El concepto de organización ha adquirido una gran importancia en el área de los sistemas multi-agente, al facilitar el análisis y diseño de mecanismos de coordinación y colaboración en sistemas abiertos. Las metodologías orientadas a aspectos organizativos mediante el análisis y diseño de la estructura organizativa y de sus normas sociales permiten controlar los

comportamientos de los agentes. En definitiva, en las metodologías orientadas a la organización se detallan cuales son los objetivos, su estructura, su dinámica y el entorno.

La inclusión de aspectos o factores sociales que permiten alcanzar acuerdos en las organizaciones de sistemas multi-agente constituyen una metodología importante para estructurar las interacciones entre mundos abiertos y dinámicos. El resultado de la inclusión de aspectos sociales en la organización de sistemas de agentes asegura agilidad (capacidad de adaptarse a circunstancias que cambian) y resistencia (capacidad de alcanzar los objetivos en un entorno dinámico e incierto) de dichos soluciones.

Las capacidades que proporciona el uso de agentes inteligentes (autonomía, situación, reactividad, racionalidad, inteligencia, organización, movilidad y aprendizaje) plantean la propuesta de diseñar la arquitectura para la recuperación de contenidos educativos en base a las posibilidades que ofrecen. La resolución de un problema como el que plantea este trabajo de investigación puede aplicar las características de dichos agentes inteligentes al desarrollo de un sistema estable, con la capacidad de reaccionar de forma inteligente a las necesidades del entorno dinámico en el que se plantea el trabajo. La idea de modelar la arquitectura como una organización virtual, en base a que la organización puede adaptar sus acciones frente a cambios para conseguir sus objetivos e interactuar con componentes heterogéneos, aparece como una solución teóricamente eficaz en el planteamiento de una solución escalable y flexible a la recuperación de contenidos heterogéneos distribuidos.

PARTE III. PROPUESTA

5 MODELO DE ARQUITECTURA BASADA EN ORGANIZACIONES VIRTUALES

Actualmente los usuarios tienen acceso a tecnologías que les permiten obtener información de una manera ubicua y en tiempo real. Esta apertura en la comunicación y la urgente necesidad de alcanzar una verdadera interoperabilidad entre entornos de aplicaciones educativas han creado la necesidad de investigar en la búsqueda e integración de contenidos educativos heterogéneos.

Los problemas de integración de las aplicaciones emergentes y los contenidos son causados generalmente por los diferentes tipos de datos y acceso que tienen que llevar a cabo los distintos tipos de usuarios. Además, las nuevas aplicaciones de la Web manejan información distribuida en distintos puntos geográficos haciendo que la integración de los datos y la seguridad de las aplicaciones sean más complicadas y, en muchas ocasiones, imposible de garantizar plenamente. Finalmente, hay que añadir los problemas relativos a los canales de comunicación por donde viaja la información, como es el caso de Internet. Tomando en cuenta estos y muchos otros riesgos, en este trabajo se presenta una arquitectura multi-agente para la búsqueda e integración de contenidos educativos heterogéneos a través de un modelo de recuperación mediante búsquedas federadas.

La arquitectura multi-agente utiliza un diseño organizacional. La finalidad de dicha arquitectura consiste en recuperar información que sea de utilidad para quien la ha solicitado, utilizando para ello, todas las ventajas proporcionadas por las organizaciones virtuales de agentes y el sistema de búsquedas.

Para el desarrollo de este sistema de búsqueda federada se decide el uso de los SMA y concretamente de aquellos que tienen en cuenta normas y restricciones sociales. Se opta por esta metodología dada la alta heterogeneidad del contexto educativo en el que nos movemos. Este tipo de arquitecturas ofrece la ventaja de poder describir la funcionalidad mediante un conjunto de roles, de tal forma que cada uno de los roles puede ser llevado a cabo por cualquier agente, siempre que posea las capacidades necesarias. En la práctica, esta nueva concepción resulta crucial ya que, al movernos en un entorno tan complejo y dinámico, los cambios que se produzcan en el contexto se podrán introducir también en el sistema de forma sencilla.

El capítulo está organizado de la siguiente manera. La sección 5.1 proporciona una introducción en la que se muestra la motivación que dio lugar al desarrollo de esta propuesta. La formalización de la arquitectura se presenta en la sección 5.2 y finalmente, la sección 5.3 describe líneas de trabajo futuro y presenta algunas de las conclusiones obtenidas.

5.1 MOTIVACIÓN

Para entender la motivación de este trabajo de investigación se hace necesario comprender el contexto actual en el que se centra la investigación llevada a cabo. A continuación se muestra el contexto, haciendo énfasis en el concepto clave de recuperación de información distribuida o búsqueda federada. Posteriormente, en la sección 5.1.2, se hace un resumen de su análisis y de los problemas encontrados, que darán lugar a la motivación de este trabajo. Encontramos tres clases diferentes de problemas, los relacionados con los ROA, los concernientes a la especificación de los OAs y los centrados en la problemática del filtrado, catalogación y recomendación de datos. Todos estos problemas nos llevan a la búsqueda de una solución, resumida en el apartado 5.1.3 y detallada posteriormente en la sección 5.2.

5.1.1 Contexto actual

Como se ha visto en las secciones primeras en el contexto educativo actual existe un gran volumen de contenidos en la Web, que no son directamente accesibles a través de buscadores convencionales. Este tipo de información se dice que pertenece a la denominada Web oculta, profunda o invisible, frente a los contenidos de la web denominada visible.

Las soluciones desarrolladas por los motores de búsqueda convencionales son muy eficaces para recuperar los contenidos visibles de la Web. Dichos motores hacen uso del modelo de recuperación centralizada que copia la información recogida en una única base de datos centralizada, indexan los contenidos y ordenan los documentos clasificados en la base de datos para las consultas de los usuarios. Este método funciona bien cuando las fuentes de información dejan expuestos sus contenidos a los rastreadores web. Sin embargo, esto no es válido en la Web oculta, donde la información sólo puede ser accedida a través de mecanismos de búsqueda adaptados a fuentes específicas, como el caso de los repositorios con contenidos educativos.

Uno de los enfoques más ampliamente aceptados en el contexto de la educación a distancia actual se basa en el uso del objeto de aprendizaje. Este paradigma se basa en fragmentar los contenidos en unidades modulares y autocontenidas para facilitar que puedan ser reutilizados en distintos contextos educacionales, así como por diferentes plataformas (como se ha descrito en el apartado 2.1).

Dichos objetos de aprendizaje están formados por un lado por el recurso en sí, destinado a la educación, y por otro por el etiquetado que a través de una serie de metadatos los describe completamente. La utilización sistemática de dicho etiquetado del OA permite automatizar procesos de búsqueda, catalogación y recuperación de dichos recursos.

The screenshot shows the MERLOT Learning Materials search interface. At the top, there is a search bar with the text 'materials' and a search button. Below the search bar is a navigation menu with links for Home, Communities, Learning Materials, Member Directory, My Profile, and About Us. The main content area displays search results for 'Pulmonary Disorders'. The results are sorted by Relevance and show 4 items. The first item is 'Breathing Abnormalities' by St. Petersburg College Web & Instructional Technology Services. The second item is 'Pulmonary Disorders' by Merck. The third item is 'Audio for Various Medical Conditions' by Merck & Company. The interface also includes a 'Browse Material Hitlist by Category' sidebar and a 'Contribute a Material' button.

Figura 5-1. Resultado de búsqueda en MERLOT a través de su interfaz Web

Habitualmente, los objetos de aprendizaje se almacenan en repositorios digitales. La mayoría de los repositorios poseen una interfaz web tradicional sobre un documento HTML de búsqueda que conecta directamente con la base de datos que alberga sus contenidos educativos. De entre ellos, la gran mayoría exige que el acceso del usuario mediante contraseña, de manera que puede ir guardando históricos de consultas, perfiles de usuario, etc. El usuario interactúa con la información que contiene dicho repositorio a través de su navegador web en el modo de interacción habitual: realiza las peticiones a través de la interfaz y como resultado el repositorio devuelve aquellos contenidos relacionados con su consulta. Existe una gran cantidad de repositorios de contenidos digitales que almacenan dichos contenidos mediante el formato de objeto de aprendizaje. Sin embargo a través de las interfaces web tradicionales, el usuario no tiene la posibilidad de manejar directamente toda la información etiquetada por dichos OA, ni siquiera puede saber si esa información está etiquetada o no conforme a algún estándar ya que

III. Propuesta

el usuario desconoce el modo en que dicho contenido educativo reside en el repositorio.

Como ejemplo, la Figura 5-1 muestra los resultados de la búsqueda “Pulmonary Disorders” a través de la interfaz web²⁷, disponible en del repositorio MERLOT (*Multimedia Educational Resource for Learning and Online Teaching*).

Los repositorios suelen ser altamente heterogéneos, con diferentes sistemas de almacenamiento, acceso a los objetos, métodos de consulta propios, etc. El problema de la heterogeneidad en los sistemas de bases de datos constituye un tema abierto que surge con las propias bases de datos y aunque solucionado en muchos de sus aspectos siguen apareciendo gran número de iniciativas para estandarizar los procesos y las tecnologías que lo conforman.

```
<lom>
  <general>
    <identifier>
      <catalog>MERLOT</catalog>

<entry>http://www.merlot.org/merlot/viewMaterial.htm?id=384820</entry>

    </identifier>

    <title>
      <string language="en">Pulmonary Disorders</string>
    </title>

    <description>
      <string>&lt;p&gt;This site is a section of the free Merck Online Manual that
      covers Pulmonary Disease.Â Some of the sub-topics include Asthma,
      Pneumonia,Â Â Diffuse Alveolar Hemorrhage and Pulmonary-Renal
      Syndromes, among others. Each of the sub-topics includes information
      about specific problems as well as symptoms and
      treatment.&lt;/p&gt;</string>
    </description>
  </general>
  <lifeCycle>
```

²⁷ MERLOT <http://www.merlot.org/>

```

<contribute>
  <role>
    <source>LOMv1.0</source>
    <value>author</value>
  </role>
  <entity><![CDATA[BEGIN:VCARD
FN:Merck
N:Merck
VERSION:3.0
END:VCARD]]></entity>
</contribute>
</lifeCycle>
<technical>
  <location>&lt;p&gt;This site is a section of the free Merck Online
Manual that covers Pulmonary Disease.Â Some of the sub-topics include
Asthma, Pneumonia,Â Â Diffuse Alveolar Hemorrhage and Pulmonary-Renal
Syndromes, among others. Each of the sub-topics includes information
about specific problems as well as symptoms and
treatment.&lt;/p&gt;</location>
</technical>
</lom>

```

Figura 5-2. Ejemplo de la descripción en LOM para un OA

Los repositorios de contenido educativo son cada vez más numerosos. A lo largo de los apartados 2.2, 2.3 y 2.4 de esta memoria hemos estudiado distintas iniciativas que buscan la interoperabilidad de sus contenidos. El contexto actual presenta una gran variedad de repositorios distribuidos con contenidos educativos de todo tipo. Muchos de estos repositorios no poseen ningún sistema en un nivel superior que permita la abstracción entre su funcionamiento interno y los usuarios finales de los datos almacenados. Otros se agrupan en redes haciendo uso de arquitecturas para facilitar su interoperabilidad en base a variados estándares que asignan una capa de abstracción que conecta las características internas de los mismos con el exterior, permitiendo una mayor automatización y tratamiento informático de

los OA que contienen según hemos revisado en el apartado 2.3. Siguiendo con el ejemplo del repositorio MERLOT, este implementa una arquitectura que permite la interoperabilidad de sus contenidos, proporcionando por ejemplo mecanismos de consulta offline a través búsquedas federadas mediante un servicio de cliente web SQI [SQI, 2005] (a través de la especificación WSDL [WSDL, 2001]) o a través de servicios web Restful [Fielding, 2000] [Fielding & Taylor, 2000]. A través de dichos servicios es posible acceder a la información etiquetada de los OA, como la que se visualiza en LOM en la Figura 5-2, resultado de la búsqueda “Pulmonary Disorders”.

En el contexto educativo actual los repositorios de objetos de aprendizaje forman parte de la web oculta en grandes bases de datos. Dichos objetos de aprendizaje están etiquetados en cualquier estándar (LOM, Dublin Core, etc). El desarrollo de soluciones que permitan la búsqueda eficaz de contenidos heterogéneos en contextos distribuidos se convierte en un reto y una necesidad. La recuperación de información distribuida, o búsqueda federada, [Callan, 2000] aporta una solución viable al problema de la recuperación de información en la Web oculta. La búsqueda federada desarrolla modelos y estrategias para obtener el mayor beneficio de estas fuentes distribuidas independientemente de su número, de su localización o mecanismos de gestión.

El mecanismo de búsqueda federada es más complejo pero más enriquecedor y exhaustivo que el modelo de recuperación centralizada de los buscadores tradicionales que no daban solución real al problema de la web oculta. La aportación fundamental de la búsqueda federada es que el proceso de búsqueda se realiza de forma simultánea en las fuentes de información individuales. Dicha búsqueda mantiene las referencias a la ubicación de cada fuente distribuida y establece el control de la información relacionándola con las distintas fuentes de información oculta.

La búsqueda federada en su aplicación al ámbito de la recuperación de OA de los repositorios se describe como la secuencia en la resolución de los tres subproblemas siguientes: (1) **Selección de Repositorios**, analizando la descripción de los recursos en los repositorios y evaluando cómo se representa la información que se encuentra distribuida en dichos repositorios. Es importante conocer la temática, el área y el contexto a los que pertenece la información almacenada por los repositorios, así como otras tantas características clave de las distintas fuentes de información oculta (dominio de los contenidos, sistemas de acceso, tiempos de respuesta, eficiencia, etc.). (2) **Recuperación de Contenidos**, donde a partir de una necesidad de información y el conjunto de descripciones de los OA de éstos se recuperan y se decide cuáles serán los que tengan mayor probabilidad de satisfacer la consulta mediante un *algoritmo de recuperación*. Y finalmente (3) **Fusión de Resultados**, mediante la integración y combinación de los resultados devueltos por las consultas sobre los n repositorios, formando una única lista que presenta al usuario una lista ordenada de resultados. El criterio de ordenación

de dichos resultados será resultado de cierto *algoritmo de posicionamiento*, ranking o prioridad en la presentación de los contenidos recuperados.

5.1.2 Funcionalidad de los ROA para Búsqueda Federada

Con el objetivo de conocer la situación real de los sistemas de educación a distancia se ha realizado un estudio preliminar de repositorios públicos que implementan capas que permiten búsqueda federada. Para ello se ha realizado el estudio de un conjunto representativo de ROAs que implementan la posibilidad de realizar búsquedas federadas a través del protocolo SQI. El conjunto de repositorios seleccionados para dicho estudio se muestra en la Tabla 5-1.

Tabla 5-1. Conjunto de ROA Públicos con soporte a SQI

Repositorio	URL
Acknowledge	www.acknowledge.be
Proyecto Agrega	http://contenidos.proyectoagrega.es
Ariadne	http://www.ariadne-eu.org/
AriadneNext	http://www.l3s.de
CGIAR	http://learning.cgiar.org/
EducaNext	http://www.educanext.org/
LACLO-FLOR	http://www.laclo.espol.edu.ec
LORNET	http://www.lornet.org/
MACE	http://www.mace-project.eu
MERLOT	http://www.merlot.org
NIME	http://nime-glad.nime.ac.jp/
OER Commons	http://www.oercommons.org
Edna Online	http://www.edna.edu.au/

El objetivo de este estudio, mostrado con detalle en el Anexo B, es clarificar las funcionalidades, así como detectar las carencias de repositorios públicos en funcionamiento con dicho protocolo y que albergan objetos educativos reales. Como consecuencia de dicho estudio se genera la necesidad de una solución centralizada para la búsqueda inteligente de contenidos educativos que subsane los problemas de heterogeneidad de cada una de las partes implicadas. Dicho estudio permite elaborar las pautas que guían y apoyan el desarrollo de la arquitectura de búsqueda federada propuesta en este trabajo de investigación.

El nacimiento de lo que ya se puede considerar como el paradigma de los OA [Richards, 2002] ha traído consigo una gran cantidad de ventajas a la hora de reutilizar contenido educativo gracias a las facilidades que un OA ofrece en cuanto a la especificación de contenidos (Figura 5-2), la búsqueda y la recuperación recursos educativos. Pero también, como no podría ser de otro modo, a lo largo del proceso de innovación se han producido diferentes retos que aún no han sido solventados. Aquellos repositorios que hoy en día todavía no implementan ninguna capa de abstracción que permita encapsular la lógica interna del repositorio, denominados autónomos, son monolíticos y difícilmente permiten automatizar los procesos de búsqueda y recuperación de recursos. En estos repositorios autónomos la consulta y extracción de OAs es un proceso lento en el que se hace necesaria la intervención manual. Evidentemente este tipo de repositorios están al margen de la solución integradora a través de la arquitectura propuesta por este trabajo.

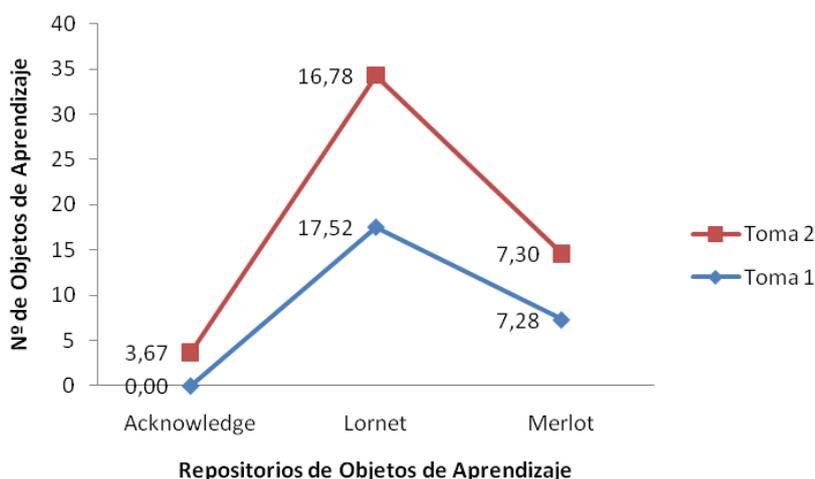


Figura 5-3. Promedio de resultados obtenidos para consultas en dos momentos de tiempo

Pero no sólo los problemas se derivan de no utilizar ninguna capa *middleware* de abstracción ya que los sistemas que sí la incluyen, como recoge el estudio del Anexo A, también adolecen de otras tantas complicaciones como la falta de confiabilidad o de disponibilidad, los elevados tiempos de acceso (Figura 5-4), resultados erróneos, escasez de resultados (Figura 5-3), etc.

Los problemas no superados obstaculizan la apuesta por la interoperabilidad de contenidos educativos y se pueden agrupar en torno a dos núcleos. El primer gran problema está relacionado con las dificultades que acompañan a la estructura monolítica de los Repositorios de objetos de aprendizaje. Entre estos problemas están el que no permiten su manejo externo con la flexibilidad y la

potencia necesaria que asegure una sencilla interoperabilidad de fuentes dispersas y heterogéneas. Los ROA son poco estables la mayoría de las veces y presentan un elevado número de errores. La presentación de la información a través de su interfaz web es generalmente deficiente. Los repositorios de objetos no se pueden añadir de forma dinámica, sólo se puede realizar consulta en los repositorios del registro. A través de sus interfaces de búsqueda no se visualiza la información referente al objeto de aprendizaje y en general únicamente se presenta como resultado el acceso al contenido. Además la gran mayoría de repositorios no permiten la gestión integral del usuario a través de sus intereses o respecto a otros usuarios operando en el mismo contexto.

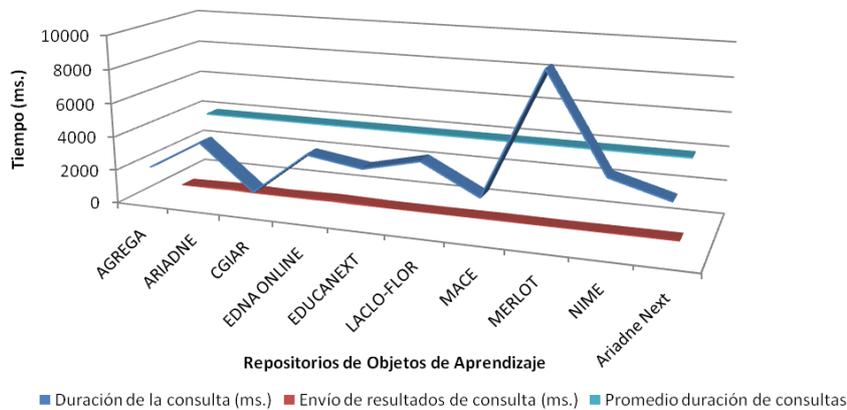


Figura 5-4. Tiempos de Acceso para evaluación de consultas en ROAs

Otros muchos aspectos pueden ser tenidos en cuenta como la carencia de perspectiva de la accesibilidad universal a los recursos web, presente en proyectos como Web-4-All²⁸, ya que los buscadores de los repositorios no cumplen en sus interfaces de acceso ni siquiera con el nivel mínimo del estándar de accesibilidad WAI²⁹(WAI-A).

El segundo grupo de problemas se relacionan directamente con la inexistencia de mecanismos automáticos que controlen la calidad técnica, semántica y sintáctica de los Objetos de Aprendizaje, asegurando una correcta especificación de dichos objetos de aprendizaje en cualquiera de los esquemas de metadatos que los describen.

²⁸ WEB-4-ALL <http://web4all.ca/>

²⁹ WAI <http://www.w3.org/WAI/>

Las carencias encontradas en los repositorios actuales tales como excesivo tiempo de respuesta o sus demasiados fallos en la conexión exigen soluciones que se adapten a la heterogeneidad de cada uno de dichos repositorios aislados y aseguren la interoperabilidad real y efectiva de los contenidos educativos de manera global. Una solución centralizada y global permitirá la búsqueda y la reutilización efectiva de los recursos al usuario final. Para ello es necesario subir el nivel de abstracción y fijarse en la clasificación de sistemas de almacenamiento y búsqueda de OAs. Como hemos visto, existen repositorios en los que en cada consulta desencadena a su vez nuevas búsquedas federadas en otros repositorios. Esto, aunque en principio se puede considerar como una ventaja porque se aumenta la posibilidad de incrementar el número de resultados, en la práctica tiene dos inconvenientes. El primero relacionado con el tiempo de respuesta, que aumenta considerablemente y el segundo derivado de la aparición de OAs repetidos en los resultados, debido las consultas repetida en el mismo repositorio de modo indirecto a través de las federaciones.

Más allá de las dificultades que se pueden encontrar en los accesos a los ROA y a la eficacia en la gestión que hacen de sus contenidos, también existen algunos problemas relacionados con la mala o incorrecta descripción de los OAs a través de los metadatos. Para evaluar la situación real del etiquetado de los objetos de aprendizaje disponibles en los repositorios activos se ha realizado un estudio disponible en el Anexo B. En dicho estudio se parte de que el IEEE LOM define perfectamente el recurso educativo, incluyendo una gran variedad de detalles como aspectos generales e identificación, objetivo educativo, derechos de uso y explotación, características técnicas o la relación con otros OAs. Sin embargo, como no existe una obligatoriedad en los atributos que hay que especificar para que un OA tenga una calidad mínima para ser utilizado dentro de un contexto educativo particular resulta ser un encuentro de obstáculos en la recuperación de contenidos educativos. En el estudio realizado se extraen dos problemas fundamentales: (i) la identificación unívoca de cada objeto de aprendizaje y (ii) la falta del elemento que define mediante URI o URL el lugar donde se almacena el recurso de aprendizaje.

- El primero relacionado con la identificación unívoca de cada objeto de aprendizaje. Un objeto a priori está perfectamente definido gracias a un identificador de catálogo (repositorio) y una entrada dentro de este repositorio. Pero como no es obligatorio completar estos campos, la identificación unívoca de cada OA se convierte en un proceso complejo. El mismo caso sucede en el caso de la identificación doble de un OA.
- El segundo está relacionado con la falta del elemento que define mediante URI o URL el lugar donde se almacena el recurso de aprendizaje que describe el archivo de metadatos. Y por lo tanto, aunque un OA esté perfectamente descrito, cuando no dispone de enlace al recurso o este es erróneo, dicho contenido no podrá ser accedido y por tanto utilizado.

Todos los problemas enumerados relativos a los ROA así como a los propios OA constituyen un auténtico reto para las aplicaciones que traten de hacer uso de dichos repositorios como fuentes de contenido educativo. El desarrollo de dichas aplicaciones deberá de trabajar con diseños que habiliten su uso con la inclusión de soluciones prácticas en el entorno altamente heterogéneo, dinámico y con elevado número de errores de los repositorios. Este contexto hace que sea necesario diseñar y construir sistemas con una alta tolerancia frente a fallos.

Además, hay que considerar un reto más en la gestión eficiente de los servicios y los elementos involucrados en estas plataformas mediante la inclusión de técnicas de gestión eficaz haciendo uso de las carencias en etiquetado de los contenidos educativos de manera que se facilite su almacenamiento, búsqueda, recuperación, etc. Es decir, mediante técnicas de filtrado, catalogación y recomendación.

Hay algunas propuestas para permitir al usuario encontrar los objetos de aprendizaje que le satisfacen. En la subsección 3.5.2 se enumeran algunos de estos trabajos, en los que algunos centran su esfuerzo en generar arquitecturas que permitan realizar dicha recuperación de un modo más o menos centralizado, otros esfuerzos se abstraen del problema de la recuperación y centran sus esfuerzos en posicionar el conjunto de OA de los que disponen de manera adecuada, para ofrecer los más relevantes para el usuario en primer lugar. De manera que una solución integral al problema de la recuperación de contenidos educativos va más allá de la simple recuperación, ya que necesita proveerse con mecanismos de catalogación de dichos objetos recuperados que posibiliten evaluar los más adecuados y generar su posicionamiento de acuerdo al usuario.

5.1.3 *Búsqueda de la solución*

Con la situación que se plantea en el contexto educativo, es evidente que se necesita un nuevo tipo de aplicaciones, que permitan la búsqueda de contenidos educativos en entornos distribuidos a través de diferentes formatos, servidores y redes. Este trabajo propone como solución una arquitectura de búsqueda federada de contenidos educativos, concretamente los denominados OA.

Las necesidades encontradas deben ser satisfechas con el propósito de proveer la creación de la federación y de esta manera la integración de los diversos componentes: los repositorios se encuentran distribuidos en lugares físicamente dispersos; el acceso a estos repositorios es de sólo lectura; los sistemas de información se encuentran funcionando bajo plataformas diferentes; el tipo de repositorios utilizados son distribuidos (metadatos almacenados en un sitio y OA en otro) y gestionados por sistemas son

autónomos, en algunos casos el acceso a los repositorios exige la creación de una sesión, en otros el acceso es directo, etc.

En este sentido, diseñar la arquitectura del sistema sobre la base que proporciona el uso de agentes inteligentes resulta idóneo, ya que se pueden aplicar las características de éstos (autonomía, situación, reactividad, racionalidad, inteligencia, organización, movilidad y aprendizaje) al desarrollo de un sistema estable, con la capacidad de reaccionar de forma inteligente a las necesidades del entorno. La idea de modelar la arquitectura como una organización virtual, en base a que una organización puede adaptar sus acciones frente a cambios para conseguir sus objetivos e interactuar con componentes heterogéneos, se plantea como una solución teóricamente eficaz.



Figura 5-5. Esquema general de distribución de tareas en la arquitectura

De la situación tecnológica heterogénea y cambiante, que acompaña al contexto educativo planteado, y en aras a la reutilización en un contexto operativo real, surge la motivación para diseñar un modelo de una arquitectura integradora, en la que una organización de agentes pueda ejecutar acciones de búsqueda y recuperación de dichos contenidos basándose en un modelo de búsqueda federada. La innovación de dicha arquitectura consistirá en dotar a una

organización de agentes de las capacidades auto-adaptativas necesarias para hacer frente a la problemática actual y en entornos altamente dinámicos como los analizados también adaptarse a sus cambios futuros. En la arquitectura se plantea una distribución de funciones que presenta la Figura 5-5.

De esta forma, quedan solucionados de una manera totalmente transparente al usuario los problemas de distribución e integración de los diferentes repositorios, de abstracción de la lógica interna de cada uno de ellos y de clasificación, almacenamiento y búsqueda de OAs. Además de añadir capacidades como una sencilla escalabilidad a posibles situaciones de uso de nuevos protocolos, lógicas internas de repositorios, catalogación o aplicaciones heterogéneas con servicios encaminados a cubrir funcionalidades relacionadas.

El pilar de este trabajo de *investigación* es la recuperación de objetos de aprendizaje en un entorno real utilizando la búsqueda federada en diferentes repositorios de Objetos de Aprendizaje. Es necesario proporcionar al usuario un marco que unifique la búsqueda y recuperación de objetos, facilitando así el proceso de aprendizaje, que filtre y clasifique adecuadamente los objetos de aprendizaje recuperados según ciertas reglas. La generación de dichas reglas para la organización de los objetos recuperados se realizará en base a los metadatos educativos y darán como resultado un contenido útil al usuario final. Se surtirán mecanismos que posibiliten evaluar los OAs recuperados y catalogar los más adecuados para un usuario. La arquitectura propuesta facilitará las múltiples perspectivas de la recuperación de contenidos educativos y su personalización. En las siguientes subsecciones se muestra la descripción detallada de la arquitectura propuesta, mostrando su formalización a través de una organización virtual de agentes.

5.2 FORMALIZACIÓN DE LA ARQUITECTURA PROPUESTA

Para el diseño y desarrollo de SMA se necesitan metodologías que den soporte a los diseñadores de una manera robusta y fiable. Muchas de las propuestas tradicionales detallan la estructura del SMA en términos de un modelo de roles, en el que se identifican los roles que los agentes juegan en el sistema, así como los protocolos de interacción en los que participan.

Estas metodologías, tipificadas como *orientadas al agente* asumen una perspectiva individualista, cuya entidad principal es el agente, el cual tiene definidas claramente sus tareas y habilidades que le ayudan a conseguir sus objetivos individuales. Se trata, pues, de sistemas cerrados, en los que no se permite la participación de agentes con comportamientos interesados, egoístas o no autorizados.

En los últimos años, los investigadores han llevado a cabo diversos estudios para ofrecer procedimientos y metodologías que permitan diseñar SMA abiertos que admitan la entrada y salida dinámica de sus entidades. Los SMA abiertos deberían permitir también la participación de agentes heterogéneos, con arquitecturas e incluso lenguajes distintos [Zambonelli *et al.*, 2003]. Por tanto, en ellos no se puede confiar en el comportamiento de los agentes, siendo necesario establecer controles en base a normas o reglas sociales.

Los investigadores y desarrolladores se han centrado en el desarrollo de un nuevo enfoque denominado *Metodologías Orientadas a la Organización*, que trata de describir en general, cuáles son los objetivos de la organización, su estructura organizativa (topología, jerarquía de roles e interacciones, su dinámica (cómo los agentes entran/salen de la organización, cómo adoptan roles, el ciclo de vida de los agentes y las normas sociales) y el entorno de la organización.

La Teoría de Organización determina cuáles son los conceptos básicos, relaciones y características intrínsecas de cada tipo de organización existente. Concretamente para definir una organización, con independencia de su tipo (humana o de agentes software) es necesario tener en cuenta los siguientes factores [Daft, 1998] [Fox, 1981]:

- **Estructura.** Comprende todos los elementos que persisten en la organización, independientemente de cuáles sean sus individuos finales en cada momento. Vendrá definida en base a los roles, sus agrupaciones, dependencias y sus patrones de interrelación.
- **Funcionalidad.** Especifica cuáles son los objetivos globales de la organización; los servicios y funcionalidades que ofrece; los objetivos que persiguen los distintos componentes de la organización y qué tareas y planes deben seguir para alcanzarlos.
- **Normalización.** Determina el conjunto de normas y acciones definidas para controlar los comportamientos de los miembros de la organización. Se incluyen aquí tanto las normas sobre la actuación (obligaciones, permisos y prohibiciones), como las sanciones y recompensas a efectuar sobre sus miembros.
- **Dinamicidad.** Especifica cómo evoluciona la organización a lo largo del tiempo, indicando el modo en que los agentes entran y salen de la organización de forma dinámica; adoptan determinados roles en función de sus capacidades y habilidades; y participan en aquellas unidades o agrupaciones de la organización en las que sean admitidos.
- **Entorno.** Comprende tanto los recursos de los que depende la organización, como los proveedores de dichos recursos y los clientes o beneficiarios de la existencia de la organización.

La arquitectura propuesta, denominada AIREH (*Architecture for Intelligent Recovery of Educational content in Heterogeneous environments*) se contempla

como un punto de comunicación intermedio entre los ROA, los OA que almacenan y los usuarios que hacen uso de ellos. El sistema proporciona un sistema de búsqueda federada, lo que permite realizar búsquedas en varios repositorios geográficamente dispersos de forma simultánea. Además, una vez que se han recibido los resultados de los diferentes repositorios se ha desarrollado una fase de catalogación y filtrado con el objetivo de adecuar los resultados obtenidos a las preferencias de los usuarios.

El modelo organizativo utilizado en la arquitectura hace uso de la plataforma THOMAS [Carrascosa *et al.*, 2009] [Giret *et al.*, 2009], actualmente en desarrollo hacia OVAMAH [Rodríguez *et al.*, 2009], para la definición de la estructura y las normas. La adaptación en este tipo de modelos se basa en la coordinación entre los participantes de la organización. Vamos a ver a continuación la organización propuesta con la que se pretende llevar a cabo las funcionalidades de la arquitectura.

5.2.1 *Diseño del Modelo de Organización Virtual*

Se ha utilizado GORMAS (*Guidelines for Organization-based MultiAgent Systems*) [Argente, 2008] como metodología de diseño y THOMAS (Métodos Técnicas y Herramientas para Sistemas Multi-agente Abiertos) [Carrascosa *et al.*, 2009] [Giret *et al.*, 2009] como plataforma final de desarrollo.

La metodología GORMAS define un Modelo de Organización Virtual (VOM) compuesto de cuatro dimensiones: (i) Dimensión estructural, describe cuáles son los elementos del sistema y cómo se relacionan entre sí; (ii) Dimensión funcional, detalla la funcionalidad específica del sistema, en base a servicios, tareas y objetivos; (iii) Dimensión del entorno, define cada elemento del entorno que puede ser accedido a través de puertos del entorno; y (iv) Dimensión normativa, describe cuáles son las normas de la organización, así como los objetivos normativos que los agentes deben seguir.

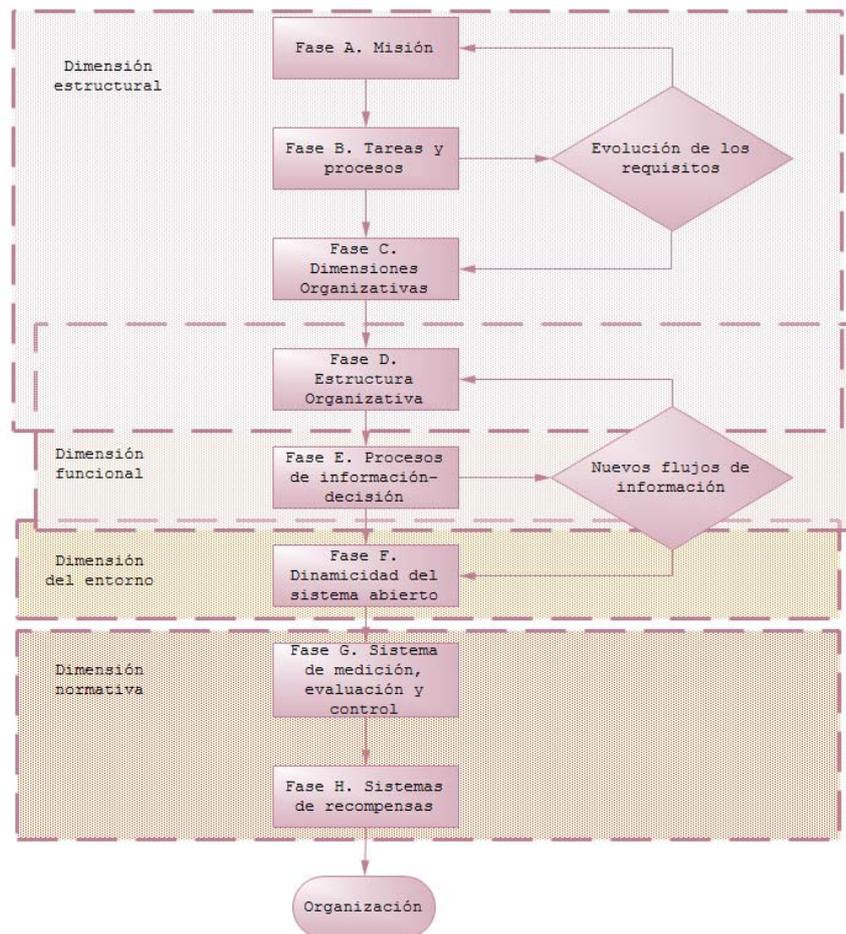


Figura 5-6. Fases de la metodología

Para definir las cuatro dimensiones que definen el VOM, el modelo GORMAS propone una secuencia-guía dividida en 8 fases:

- **Fase A: Misión.** Se realiza un análisis de la motivación que se persigue al definir la organización o sistema, es decir, del porqué de dicha organización o para qué se crea; de los resultados que, en conjunto, se esperan conseguir; y del entorno en el que existe, detallando los productos y/o servicios a ofrecer, los grupos de interés y su localización.

- **Fase B: Tareas y procesos.** Se analizan con mayor detalle los servicios a ofrecer en el sistema, sus requisitos y los procesos que conllevan. Se detallan también las tareas y objetivos asociados a dichos servicios.
- **Fase C. Dimensiones organizativas.** Se analizan las dimensiones de la organización (departamentalización, especialización, sistema decisor, formalización, coordinación), que imponen ciertos requisitos sobre los tipos de trabajo, así como sobre la diversidad e interdependencia de las tareas a realizar.
- **Fase D: Estructura organizativa.** Se determina y selecciona la estructura organizativa más adecuada para la organización, en función de sus dimensiones. Se hace uso de modelos organizativos para especificar los roles, interacciones y normas relacionados con la propia estructura.
- **Fase E: Procesos de información-decisión.** Para cada servicio identificado, se detallan las interacciones (flujos de información y de adopción de decisiones) necesarios para llevar a cabo el servicio. Además, se definen los contratos de calidad de servicio a los que se comprometen los proveedores y consumidores cuando éste sea llevado a cabo.
- **Fase F: Dinamicidad del sistema abierto.** Se establece la funcionalidad ofrecida como sistema abierto, que incluye los servicios que se deben publicitar y las políticas de adquisición y liberación de roles. Además, se diseñan los agentes propios del sistema.
- **Fase G: Sistemas de medición, evaluación y control.** Se cuantifican o evalúan las tareas y actividades y se establecen mecanismos para determinar si los objetivos del sistema se cumplen. Asimismo, se revisan las normas de la organización para especificar quiénes se encargan de ellas y las supervisan.
- **Fase H: Sistemas de recompensas.** Se determina el sistema de incentivos, para recompensar a los miembros que avancen en dirección de los intereses de la organización. También se analizan los sistemas de sanción para aquellos miembros que no cumplan con las normas dadas.

Para establecer el modelo de **interacción es necesario analizar** las necesidades y expectativas de los usuarios potenciales del sistema. A partir de este análisis es posible deducir los roles de los usuarios que participan en el sistema de reutilización de recursos educativos y la forma en la que van a intercambiar informaciones. Los roles iniciales identificados son los siguientes:

- **Usuario.** Representa al usuario o cliente del sistema:
 - Es el agente encargado de iniciar el proceso de búsqueda federada mediante el envío de un patrón de búsqueda al rol Query Manager.

- Recibe los resultados procedentes de la consulta y tiene la posibilidad de valorar tanto a los OA, como el orden en el que se le presentan estos resultados.
- Tiene acceso a información estadística tanto de los repositorios, como de los OA.
- **Query Manager.** Este agente se encarga de supervisar todo el proceso de búsqueda federada, como ya hemos visto, una búsqueda federada consiste en la búsqueda de OA en varios repositorios de forma simultánea.
 - Obtiene la consulta en lenguaje natural desde el agente Usuario y será el encargado de su formalización mediante lógica proposicional.
 - Solicita al agente Repositorio Manager la búsqueda federada de acuerdo al patrón que recibe.
 - Una vez que el agente Repositorio Manager notifica la finalización de la búsqueda federada, solicita al agente Catalogador que aplique las técnicas de catalogación y filtrado colaborativo sobre los resultados obtenidos, los cuales manejará el agente Resultados.
 - Una vez el agente Catalogador ha notificado la finalización, el proceso de ordenación de resultados de la búsqueda, ordena que éstos se transfieran al agente Usuario para que se consulten y valoren.
 - Envía al agente Estadísticas diversa información sobre la consulta federada una vez ha finalizado.
- **Repositorio Manager.** El agente Repositorio Manager tiene el control específico sobre las consultas que se realicen a los diferentes ROA, es decir, tiene el control sobre el proceso de búsqueda federada.
 - Este tipo de agente recibe la consulta formalizada desde del agente Query Manager.
 - Verifica los repositorios que se encuentran activos en el momento de la consulta y solicita al agente Estadísticas información estadística para elaborar un listado con aquellos repositorios cuyo rendimiento sea mayor. El rendimiento de un repositorio, se obtiene mediante la división del número de resultados y el tiempo necesario para obtenerlos.
 - Instancia a los agentes ROA necesarios para realizar la búsqueda federada, uno por cada repositorio incluido en el listado anterior. Una vez instanciados es el encargado de enviar a cada agente ROA la consulta en lenguaje formalizado.
 - Notifica al agente Query Manager que la búsqueda federada ha finalizado.
 - Tiene el deber de controlar el correcto funcionamiento de los agentes que adquieren el rol ROA y así evitar la pérdida de

rendimiento debido a que un repositorio deje de funcionar correctamente durante la consulta.

- **ROA.** Es el agente encargado de realizar la consulta a cada repositorio. Implementa el código cliente de cada una de las capas *middleware* posibles. Es decir, aunque únicamente se contempla un rol ROA, en la práctica existen distintos tipos de agentes que adquieren este rol, de forma que cada tipo de agente implementa una capa *middleware* diferente. Pudiera darse el caso de que para un mismo repositorio contenga distintas capas *middleware* implementadas, en ese caso se instanciaría el agente ROA que tuviese el mayor rendimiento de entre los posibles para dicho repositorio.
En una búsqueda federada existirán tantos agentes con el rol ROA como repositorios hacia los que va dirigida la consulta.
 - Este agente tiene la responsabilidad de realizar la consulta a cada ROA individual, diferentes instancias de este agente trabajarán de forma simultánea.
 - De forma previa a la realización de la consulta, debe instanciar al agente Traductor y enviarle la consulta en lenguaje formal para que sea traducida al lenguaje utilizado por el ROA específico hacia el que vaya dirigida la consulta.
 - El agente ROA realiza la consulta llevando a cabo todo el proceso necesario, definido en la especificación de la capa *middleware* que implemente dicho repositorio en concreto.
 - Es el encargado de enviar al agente Resultados los OA que hayan sido recibidos como respuesta a la consulta realizada.
 - Debe notificar al agente Repositorio Manager la finalización de la consulta.
 - Debe tomar datos estadísticos de la consulta que realiza y posteriormente enviarlos al agente Estadísticas.
 - Por último, también es el encargado de cancelar una consulta en el caso de que ésta no cumpla unos niveles QoS predefinidos. Si se da esta situación, en la que se tenga que cancelar una consulta, debe enviar el consiguiente error al agente Estadísticas.
- **Traductor.** Es el agente encargado de transformar la consulta del lenguaje formalizado al lenguaje requerido por el repositorio hacia el que se dirige la consulta.
 - Recibe la consulta formalizada mediante lógica proposicional desde el agente ROA.
 - Realiza la conversión de esta consulta desde un lenguaje formal hacia el lenguaje que sea capaz de tratar el ROA sobre el que se dirige la consulta. Este lenguaje es dependiente del ROA y/o de la capa *middleware* de interoperabilidad, por lo que existen diferentes instancias de este agente, una por cada uno de los

posibles lenguajes, aunque todas estas instancias tendrán el rol de Traductor.

- Una vez se realizada la transformación se enviará al agente ROA la consulta en el lenguaje con el que este opere.

- **Resultados.** El agente Resultados recibe los OA procedentes de cada uno de los agentes ROA, en cada una de las búsquedas federadas. De forma automática extrae la información del esquema de metadatos y elimina aquellos OA que no son válidos.

De la misma forma que en casos anteriores, aunque en teoría existe un único rol resultados, en la práctica existen diferentes tipos de agente resultados, de forma que cada uno de estos tipos de agentes implemente un esquema diferente de metadatos.

- Es el encargado de recibir los resultados de la búsqueda federada desde cada uno de los agentes ROA instanciados en cada búsqueda federada.
- Extrae la información útil procedente de cada uno los esquemas que describen a los OA, por lo tanto, será necesario que tenga acceso a los servicios de extracción de metadatos y a las estructuras de datos necesarias para almacenar la información suministrada por cada uno de los esquemas que definen a los OA.
- En un primer filtrado elimina aquellos OA defectuosos que impidan su correcto tratamiento y/o uso por parte de los usuarios del sistema.
- Este rol almacena un conjunto de datos mínimos convenientes para cada uno de los esquemas de descripción de los OA proporcionando niveles de abstracción en la gestión integral de recursos didácticos. Para cada uno de los OA de los resultados ejecutará un algoritmo de alineamiento para obtener el grado solapamiento del objeto determinado respecto al considerado como candidato ideal a la consulta del usuario.
- Toma datos estadísticos relevancia y exhaustividad de los ROA atendiendo a la formulación descrita en el apartado 3.2 y posteriormente se envían al agente Estadísticas.
- Los agentes con rol Catalogador tendrán acceso a los OA almacenados.

- **Catalogador.** Es el agente encargado de elaborar el ranking de OA que han sido devueltos como resultados de la búsqueda federada realizada por los agentes ROA y coordinada por el agente Repositorio Manager. Los resultados serán almacenados por el agente Resultados que a su vez aplica, de forma automática, un filtrado previo sobre OA, obteniendo la información útil de los mismos y eliminado aquellos que están incompletos.

El agente Catalogador implementa el sistema de razonamiento CBR que utiliza la información sobre búsquedas antiguas para elaborar un ranking con los objetos que más se adecuan a las necesidades (o

gustos) del usuario de la aplicación en función de la información obtenida previamente. Para ello, utiliza información del perfil del usuario, así como información sobre sus intereses educativos (filtrado basado en contenido). Posteriormente también utiliza información sobre la votación de los usuarios a los OA y sobre la conveniencia de los rankings de resultados creados previamente (filtrado colaborativo).

- El agente ordena los OA almacenados en el agente Resultados de acuerdo a las preferencias tanto generales (de perfil), como educativas del usuario.
- Para realizar este proceso, solicita al agente Estadísticas información sobre las votaciones de los OA y sobre ordenaciones previas, así como el *feedback* proporcionado por el usuario. Con toda esta información elabora el ranking de OA que más se adecue al usuario que ha realizado la consulta.
- Proporciona al agente Estadísticas el ranking creado para que sea almacenado y de este modo sea utilizado para la elaboración de ranking de resultados en el futuro. Este ranking almacenado será validado por el usuario durante el proceso de consultas de resultados.
- **Estadísticas.** Este agente es el encargado recoger los datos estadísticos del resto de los agentes. También suministra la información que contiene para aumentar el rendimiento del sistema, en el caso del Repositorio Manager; o para mejorar la calidad de los resultados, en el caso del agente Catalogador.
 - Recibe datos estadísticos de la consulta procedentes de los roles Repositorio Manager y ROA.
 - Recibe los datos de votación de los OA del agente Usuario.
 - Recibe el *feedback* sobre la ordenación de los resultados del agente Usuario.
 - Suministra datos estadísticos al agente Repositorio Manager para que elabore el listado de repositorios con un rendimiento mayor, que serán los que participarán en cada una de las búsquedas federadas.
 - Suministra datos estadísticos al agente Catalogador para que pueda elaborar el ranking de OA para cada una de las búsquedas federadas.
- **Supervisor.** Un agente que ejerza este rol tendrá un control general del sistema.
 - Analiza la estructura y sintaxis de todos los mensajes de entrada y salida del sistema.
 - Supervisa el correcto funcionamiento de los otros agentes del sistema (verifica periódicamente el estado de todos los agentes de la arquitectura, por ejemplo enviando un mensaje de *ping*).

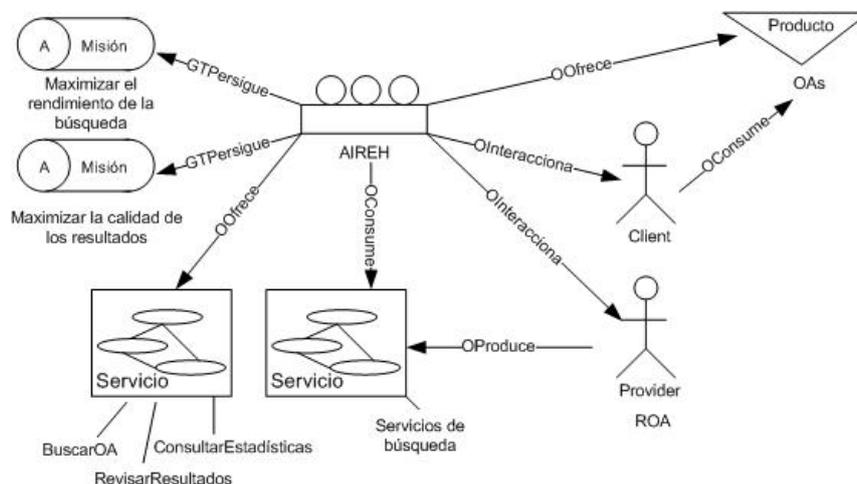


Figura 5-7. Diagrama del modelo de organización (vista funcional). Misión de la arquitectura

Siguiendo las pautas indicadas en la guía metodológica [Argente, 2008], una de las primeras tareas es instanciar la vista funcional (misión) del modelo de organización, ésta representa los productos y servicios que ofrece el sistema, el tipo de entorno, los objetivos globales que persigue (misión y justificación), los grupos de interés a los que afecta y la información que consume.

La Figura 5-7 detalla los elementos de dicho modelo de organización (Vista funcional). En ella se muestran los resultados (productos y servicios) que ofrece el sistema, el tipo de entorno sobre el que trabaja y los grupos de interés. También se justifica la razón de ser de la arquitectura, la cual pretende facilitar la búsqueda de recursos educativos mediante la puesta en marcha de un sistema de búsqueda federada.

Tabla 5-2. Misión Organizativa de la arquitectura

A.1. Misión organizativa	
Nombre.	AIREH
Dominio.	e-learning
Resultados.	<ul style="list-style-type: none"> • Producto: Objetos de aprendizaje. Finalidad: Descriptor de un recurso educativo asociado a la búsqueda realizada por el usuario. Tangibilidad: Lleva asociado un identificador y además se calcula el código HASH del esquema de metadatos. • Servicio: Buscar OA. Finalidad: Encontrar recursos educativos mediante la búsqueda de OA en diferentes ROA. • Servicio: Revisar Resultados. Finalidad: Comprobar la ordenación del sistema de archivos con el objetivo de mejorar el sistema en cada búsqueda.

<ul style="list-style-type: none"> • Servicio: Consultar estadísticas. Finalidad: Proporcionar al usuario del sistema información sobre patrones de uso de los OA.
Grupos de interés. <ul style="list-style-type: none"> • Cliente: Solicita OA de aprendizaje que se adecuen a una búsqueda realizada. • Proveedor: Facilita los OA que más se adecuan a la búsqueda realizada.
Tipo de entorno: Virtual y distribuido
Justificación. Sistema abierto que facilita un punto de comunicación común entre las entidades demandantes y entidades oferentes, de forma controlada sobre todos los servicios y productos que se ofrecen y generan. <ul style="list-style-type: none"> • Sistemas similares: Sistema de consulta propio de cada repositorio. • Ventajas: Acceso a un mayor número de repositorios de forma simultánea y centralizada y por lo tanto acceso a un mayor número de contenidos y mejor gestión de estos. • Desventajas: Mayor complejidad al ser un sistema abierto. • Singularidades: Potenciar la heterogeneidad de los miembros del sistema.

En la vista funcional de la unidad organizativa AIREH se muestran los servicios BuscarOA, RevisarResultados y ConsultarEstadísticas. Para proporcionar estos servicios, es necesario que los proveedores representados, en este caso los ROA, sean quienes ofrezcan los servicios de búsqueda que permiten la obtención de la información. En la Tabla 5-2 se muestra un análisis más detallado de los grupos de interés, en el que se describen las necesidades de los *stakeholders*, su relación concreta con la organización, así como qué influencia tienen las decisiones a tomar. Posteriormente en la Tabla 5-3 se muestra el análisis de las condiciones de entorno del sistema.

Tabla 5-3. Grupos de interés de la arquitectura

A.2. Grupos de Interés		
	Stakeholders	
	Demandante (Cliente)	Oferente (ROA)
Beneficiario	Primario	Secundario
Tipo	Cliente	Proveedor
Objetivos	Encontrar recursos de aprendizaje que se adecúen a sus necesidades	Maximizar el número de usuarios y la calidad de sus recursos.
Requiere	OA	--
Proporciona	Feedback sobre los resultados obtenidos	Recursos de aprendizaje

III. Propuesta

		encapsulados en OA.
Frecuencia	Frecuente	Frecuente
Beneficios	Obtener recursos de didácticos	Incrementar el número de usuarios de su ROA.
Poder de decisión	No	Sí (proporcionar nuevos servicios de búsqueda)
Influencia sobre sus intereses	Sí (ofrecerle un mayor número de contenidos educativos)	No

El sistema ofrece como producto final recursos educativos encapsulados en forma de OA. Éstos serán solicitados por los usuarios mediante un proceso de búsqueda. Por otro lado, también ofrece como producto información estadística sobre el rendimiento de los repositorios y sobre el uso de los OA, identificando aquellos que son más utilizados de acuerdo a los patrones de búsqueda. La misión de la organización será, por un lado maximizar el rendimiento del sistema de consultas mediante la disminución del tiempo y el aumento de resultados; y por otro lado, la maximización de la calidad de los resultados.

Tabla 5-4. Condiciones de Entorno de la arquitectura

A.3. Condiciones del entorno		
Condición	Valor	Justificación
Tasa de cambio	Dinámico	Podrán aparecer nuevos tipos de capas de middleware creándose nuevos protocolos de búsqueda
Complejidad	Simple	Existen pocos elementos distintos. Se tratará de un único tipo de proveedor y un único tipo de cliente. Y los productos a compartir tienen características similares.
Incertidumbre	Media	Entorno dinámico y simple.
Receptividad	Hostil	Se requiere información que proporcionan los proveedores, pero estos no tienen que estar siempre activos.
Diversidad	Uniforme	Los servicios proporcionados son independientes entre sí.

El siguiente modelo que se instancia, según la metodología GORMAS, es el modelo de organización, en la Figura 5-8 se puede apreciar la vista estructural, mientras que en la Figura 5-9 se aprecia la vista funcional. En este modelo se define una Unidad Organizativa que representa al sistema (AIREH). También, como se aprecia en la vista estructural, se representa al rol cliente como único ya que todos los posibles agentes que adquieran este rol tendrán acceso a la misma funcionalidad. De la misma forma se representa el proveedor como el

rol ROA, que será el rol encargado de acceder al servicio de búsqueda ofertado por el repositorio. Ambos (Cliente y Proveedor) se modelan mediante la relación *OcontieneA-Agente*, indicando así que ambos se considerarán como integrantes del sistema.

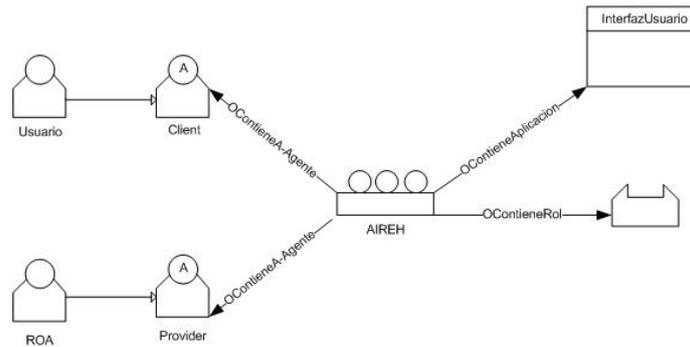


Figura 5-8. Diagrama del modelo de organización de la arquitectura (Vista estructural)

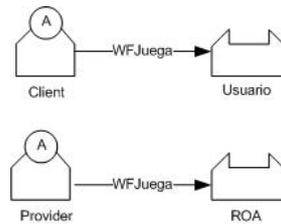


Figura 5-9. Diagrama del modelo de organización de la arquitectura (Vista funcional)

Siguiendo todas las pautas, obtenemos el modelado de la organización, desde su estructura interna hasta su funcionalidad. En la Figura 5-10 se muestra la vista funcional (funcionalidad externa) del modelo de organización asociado a AIREH, en el que los servicios conectados entre sí tienen asociados los roles del sistema las relaciones *WUtiliza* y *WFProporciona*.

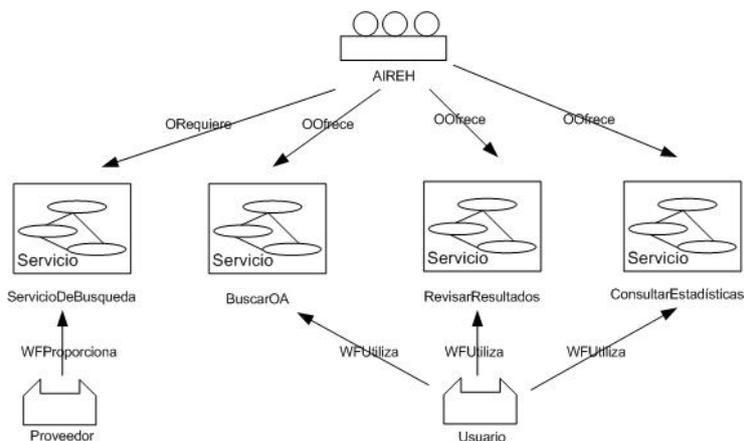


Figura 5-10. Diagrama del modelo de organización (Vista funcional, Funcionalidad externa)

De entre los tres servicios ofertados, el más complejo y representativo es el servicio de BuscarOA; a continuación en la Tabla 5-5 se muestra la descripción detallada de este servicio, mientras que en las Figura 5-11 y Figura 5-12 se muestran los diagramas del modelo de actividad.

Tabla 5-5. Tecnología de Unidad de Trabajo. Descripción de las características del servicio Buscar OA

B.2. Tecnología de Unidad de Trabajo	
Servicio	BuscarOA
Descripción	Realizar una búsqueda de OA sobre diferentes ROA, es decir, una búsqueda federada sobre diferentes repositorios.
Consumidor	Usuario.
Objetivo	Entontrar aquellos OA que se adecúan a las necesidades del usuario.
Beneficio	Facilitar el proceso de búsqueda de recursos educativos.
Perfil Servicio	
Entradas	Patrón <ul style="list-style-type: none"> Patrón. Patrón que identifique a los objetos de búsqueda que se desean recuperar
Precondiciones	Existe, al menos, un patrón de búsqueda.
Salidas	Descriptor de OA.
Postcondiciones	Si existen resultados que coincidan con el patrón dato, el usuario tiene acceso y puede consultar la información proporcionada por OA.

<p>Funcionalidad</p> <p>Tareas</p>	<p>Elaborar listado de repositorios + Transforma la consulta + Realizar búsquedas + Almacenar estadísticas + Filtrar resultados + Ordenar resultados</p> <ul style="list-style-type: none"> • Elaborar listado de repositorios: De forma previa a la búsqueda se elabora el listado con el conjunto de ROA sobre los que se realizará la búsqueda, haciendo uso de las estadísticas almacenadas. • Transformar la consulta: Se debe convertir la consulta dada por el usuario en lenguaje natural al lenguaje específico de cada uno de los ROA • Realizar búsquedas: Se realizarán una consulta a cada repositorio seleccionado, se recogerán los resultados y se guardarán las estadísticas que hayan recogido. • Almacenar estadísticas: Una vez realizadas las diferentes búsquedas en los repositorios se almacenarán diferentes estadísticas sobre las mismas. • Filtrar resultados: Se eliminarán los resultados que no cumplan la calidad mínima de los resultados obtenidos. • Ordenar resultados: Consiste en ordenar los resultados de acuerdo a las preferencias de los usuarios.
<p>Recursos</p> <p>Proveedor</p> <p>Producto</p>	<p>Valoraciones de resultados y de OA</p> <p>ROA</p> <p>OA</p>

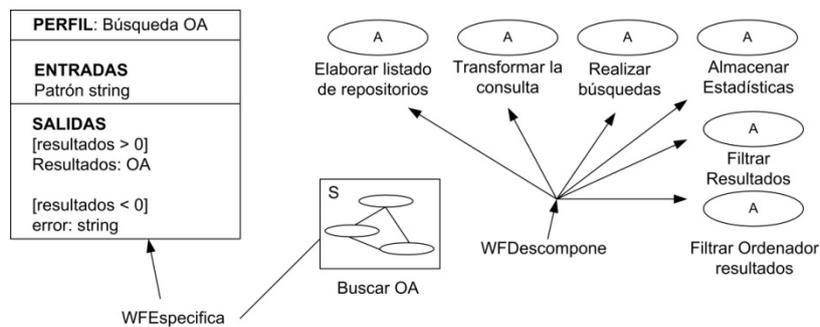


Figura 5-11. Diagrama del modelo de actividad del servicio Buscar OA (Perfil del servicio)

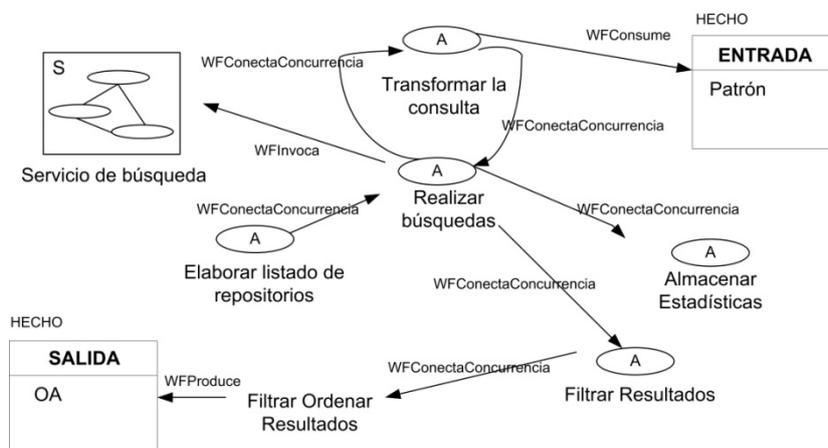


Figura 5-12. Diagrama del modelo de actividad del servicio BuscarOA (Relaciones entre A-Tareas)

Una vez que el rol Usuario recibe los resultados de la consulta que él mismo inicia, tiene la posibilidad de valorar tanto la calidad de los OA recibidos, como el orden en el que se le presentan los resultados. La importancia de este paso radica en que la realimentación obtenida desde el usuario se utilizará para mejorar la calidad en los futuros resultados. A continuación se presenta la descripción detallada del servicio Revisar resultados.

Tabla 5-6. Tecnología de Unidad de Trabajo. Descripción de las características del servicio RevisarResultados

B.2. Tecnología de Unidad de Trabajo	
Servicio	RevisarResultados
Descripción	Proporcionar realimentación sobre los resultados de una búsqueda previa
Consumidor	Cliente
Objetivo	Obtener toda la información posible sobre los resultados de la búsqueda
Beneficio	Mejorar la ordenación de resultados en futuras búsquedas
Perfil Servicio Entradas	Valoración OA + valoración orden OA <ul style="list-style-type: none"> • Valoración OA: El usuario valora si un OA es bueno o malo de acuerdo a sus intereses. • Valoración orden OA: El usuario valora si la ordenación en los resultados es buena o mala.

Precondiciones	Se tiene que haber realizado una consulta de forma previa
Salidas	--
Postcondiciones	La información proporcionada para el usuario se tiene que almacenar para futuras búsquedas.
Funcionalidad	Valorar OA + Valorar ordenación de resultados <ul style="list-style-type: none"> • Valorar OA: El usuario proporciona realimentación acerca de los OA incluidos en el listado de resultados. • Valorar ordenación de resultados: El usuario proporciona realimentación acerca de la ordenación de los resultados
Tareas	
Recursos	--
Proveedor	--
Producto	--

El rol Usuario también tiene acceso a la consulta de información estadística sobre el rendimiento de los repositorios y el histórico de resultados. En este caso no se muestra el detalle de este servicio ya que no tiene la complejidad suficiente.

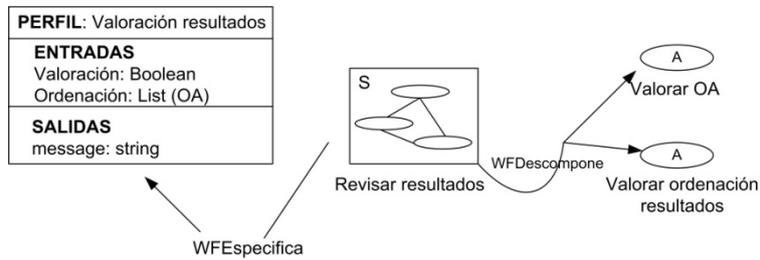


Figura 5-13. Diagrama del modelo de actividad del servicio RevisarResultados (Perfil del servicio)

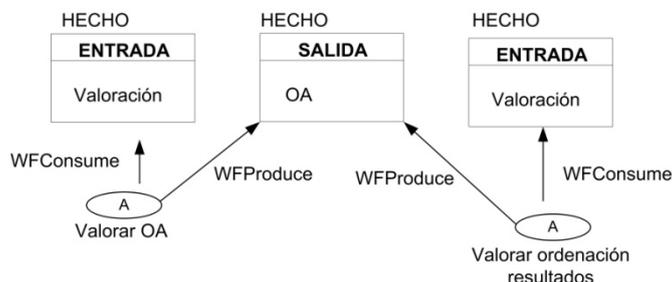


Figura 5-14. Diagrama del modelo de actividad del servicio RevisarResultados (Relaciones entre A-Tareas que lo componen)

La información que requiere el sistema, la proporcionan los proveedores a través del Servicio de Búsqueda. Por su parte, la información de los usuarios del sistema se obtiene a través de la aplicación Interfaz de Usuario, ya que fue identificada anteriormente. En la Figura 5-15 se muestra el diagrama del modelo de entorno de la unidad organizativa AIREH.

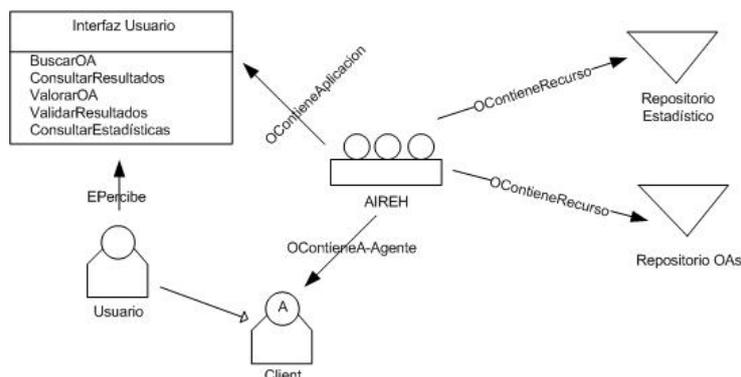


Figura 5-15. Diagrama del modelo de entorno de la arquitectura

El modelo GORMAS también establece la funcionalidad como sistema abierto, que incluye tanto los servicios que se deben publicitar como las políticas de adquisición y liberación de roles. Se determina qué funcionalidad debe ser implementada por agentes internos y cuál se publicita para poder ser suministrada por agentes externos. De este modo, se modela la dinámica de la organización como sistema abierto. Los roles ROA, Usuario, Resultados y Traductor serán accesibles por agentes externos, requiriéndose un proceso de adquisición de rol. El resto de roles (Query Manager, Repositorio Manager, Catalogador y Estadísticas) se asocian a agentes internos siendo no accesibles.

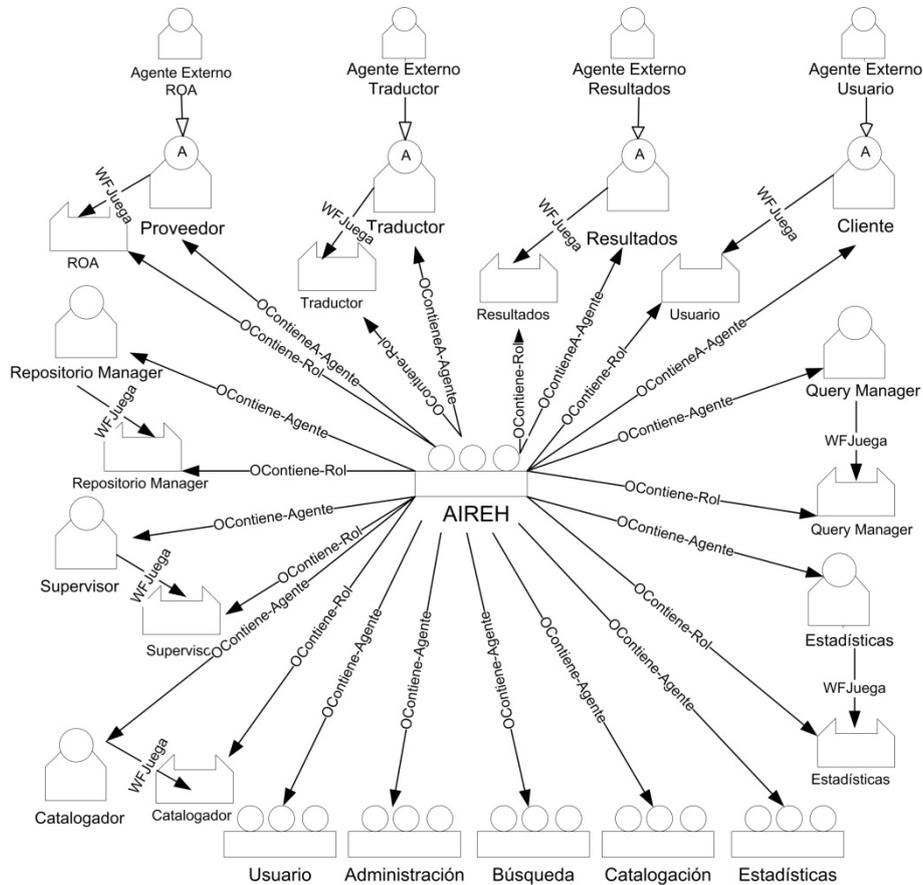


Figura 5-16. Diagrama actualizado del modelo de organización, en el que se muestran los agentes internos, externos y los roles que juegan

En la Figura 5-16 se muestra el modelo de organización, indicando los agentes externos (A-Agentes), los internos (Agentes) y los roles que juegan.

Los roles ROA, Usuario, Resultados y Traductor, como vimos en la figura anterior son accesibles por agentes externos, requiriéndose un proceso de adquisición del rol.

La dinámica del sistema, diseñado como una organización, puede quedar reflejada en diferentes escenarios: registro de nuevos agentes, de nuevos servicios, nuevos protocolos, peticiones de servicios, expulsiones del sistema, etc.

En la Figura 5-17 se muestra el escenario en que un nuevo ROA es registrado en la organización. ROA₁ se ha registrado como miembro de la plataforma THOMAS y pregunta al SF qué servicios ya definidos tienen un perfil similar al suyo. Esta petición es llevada a cabo usando el servicio *SearchService* (mensaje 1) del SF, donde *ROAServiceProfile* corresponde al perfil del servicio de búsqueda implementado por el guía ROA₁.

El SF devuelve los identificadores de servicios que satisfacen los requerimientos de búsqueda junto con un valor (*ranking value*) por cada servicio (mensaje 2). *Ranking value* indica el grado de adecuación entre el servicio y el servicio especificado propuesto. Entonces, ROA₁ ejecuta *GetProfile* (mensaje 3) para obtener información detallada sobre el servicio *ConsultROA*. Las salidas del servicio son "*service goal*" and "*profile*" (mensaje 4). El perfil del *ConsultROA* especifica que ese servicio debe ser proporcionado por agentes que juegue el rol *ROA* dentro de *ROAUnit*. Por este motivo, ROA₁ solicita al OMS el servicio *AcquireRole* para adquirir ese rol *ROA* (mensaje 5). El servicio *AcquireRole* es llevado a cabo satisfactoriamente (mensaje 6), ya que *ROAUnit* es accesible desde la organización virtual; finalmente ROA₁ es registrado como ROA.

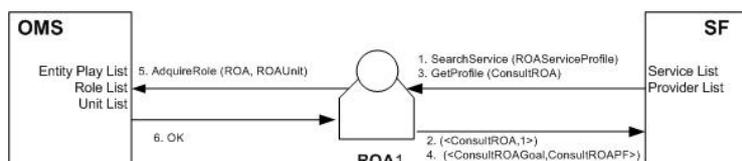


Figura 5-17. Ejemplo de adquisición de rol por parte de un agente externo

En nuestro caso, los roles de gestión y los servicios asociados a las unidades organizativas estarán disponibles gracias a la plataforma THOMAS [Carrascosa *et al.*, 2009] [Giret *et al.*, 2009]. Tal y como detallamos en el apartado C.3.1.2, el OMS (*Organization Management Systems*) ofrece los servicios necesarios para el adecuado funcionamiento de una organización. Facilita una serie de servicios para registrar o desregistrar componentes estructurales, en concreto, los roles, las normas y las unidades existentes en el sistema. También ofrece los servicios para informar sobre cuáles son dichos componentes.

Finalmente, se instancia el modelo de estructura que sigue el sistema. Cada uno de los componentes de la arquitectura se puede estructurar de diferentes formas. En este caso, según el árbol de decisión de la fase D [Argente *et al.*, 2008], se ha decidido emplear una estructura de tipo congregación.

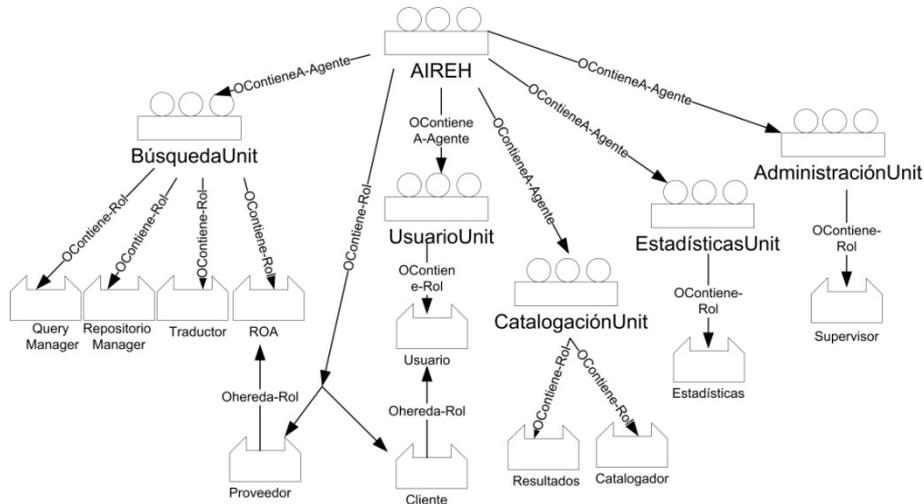


Figura 5-18. Diagrama actualizado del modelo de organización al adoptar una estructura de tipo congregación (Vista funcional)

En el diagrama de la vista estructural, adaptada conforme al patrón de congregación, la unidad AIREH contiene tantas unidades como funcionalidades (y productos), en este caso, UsuarioUnit, BúsquedaUnit, CatalogaciónUnit, EstadísticasUnit, AdministraciónUnit. Además, los roles proveedor y cliente se refinan dentro de estas nuevas unidades, para especializarse en la funcionalidad o bien en la utilización de los servicios concretos asociados a esos tipos de productos.

En el diagrama de la vista funcional, Figura 5-19, adaptada también según el patrón de congregación, se muestra el modelado interno de la unidad AIREH. A través del servicio RegistrarUnidad, los proveedores pueden solicitar la creación de nuevas unidades organizativas, para así atender la llegada de nuevas funcionalidades o productos.

Las unidades contienen los servicios de Adquirir Rol, Informar Unidad y Dejar Rol, además de los servicios propios dependientes del dominio, que ya habían sido identificados anteriormente. Así por ejemplo, un agente ROA externo que desee contactar con ellos, debe pasar previamente por un proceso de adquisición de rol correspondiente.

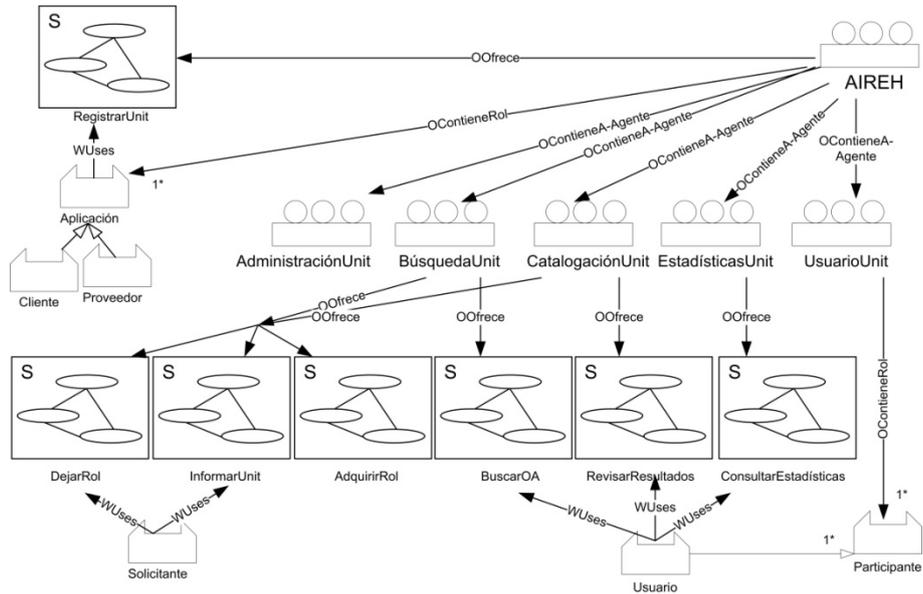


Figura 5-19. Diagrama actualizado del modelo de organización al adoptar una estructura de tipo congregación (Funcionalidad externa)

El siguiente paso es el diseño de la dinámica de la organización. Se analizan los procesos de información de la organización, detallando tanto su interacción con el entorno como las interacciones entre las entidades que la componen. Las interacciones producidas por la activación de los servicios se detallan en el modelo de interacción, donde se asocian los objetivos correspondientes del servicio con la interacción.

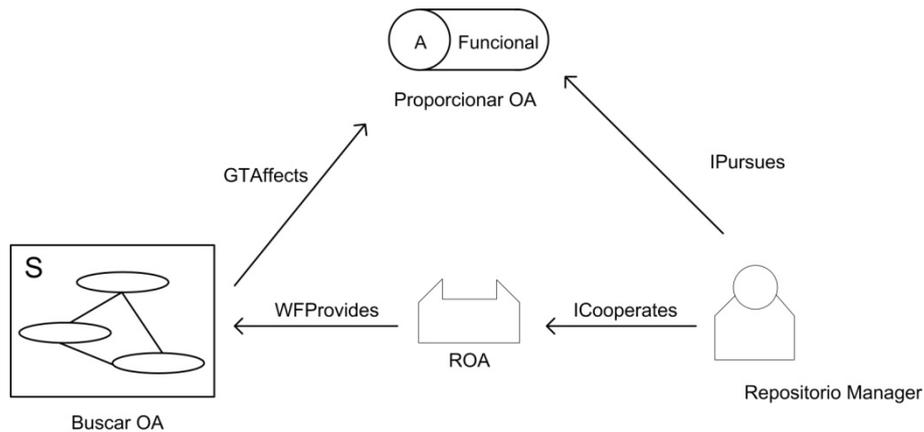


Figura 5-20. Diagrama del modelo de interacción para el servicio "Buscar OA"

Se genera también una primera especificación de la interacción, con diagramas de colaboración, donde se indican los mensajes a enviar. En la Figura 5-20 se muestra el diagrama del modelo de interacción inicial para el servicio "Buscar OA". En dicho diagrama se indican los roles que participan en la interacción y la ejecución del servicio que provoca su activación.

En la Figura 5-21 se muestra el diagrama de colaboración asociado a esta interacción, donde se detalla el modelo de proceso propuesto, indicando los mensajes que se envían las entidades participantes.

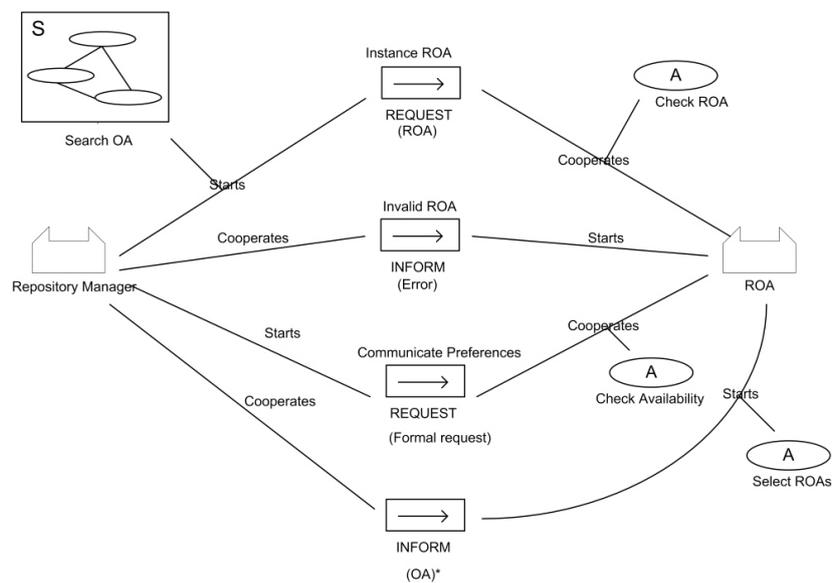


Figura 5-21. Diagrama de colaboración de la interacción BuscarOA

Una vez se ha realizado el análisis de los procesos de información y decisión, se procede a determinar la "apertura" del sistema, es decir, qué roles son accesibles por agentes externos y qué servicios necesitan ser publicitados para así atraer agentes al sistema.

En nuestro caso, los roles de gestión y los servicios asociados a las unidades organizativas estarán disponibles gracias a la plataforma THOMAS. El OMS ofrece los servicios necesarios para el adecuado funcionamiento de una organización. Dichos servicios se clasifican en: servicios estructurales, que modifican la especificación estructural y normativa propia de la organización; y servicios dinámicos, que permiten la gestión de la entrada y salida dinámica de agentes dentro de la organización, así como la adopción de roles.

El OMS facilita una serie de servicios para registrar o desregistrar componentes estructurales, en concreto, los roles, las normas y unidades existentes en el sistema. También ofrece servicios para informar sobre cuáles son dichos componentes. Los servicios dinámicos gestionan la creación de agentes nuevos en la organización, la entrada y salida dinámica de los agentes en las unidades del sistema, así como la adopción de roles.

Así, se actualiza el diagrama del modelo de entorno, incluyéndolo dentro del *framework* de ejecución de la plataforma THOMAS. Allí podrán generarse las unidades identificadas para el sistema, los agentes podrán llevar a cabo y publicitar sus servicios y será posible la entrada y salida de nuevas entidades gracias a las funcionalidades proporcionadas en THOMAS.

Respecto a las normas a definir, se implementarán internamente como parte de las acciones o tareas de los agentes. Se obliga a los proveedores de servicios a registrar una funcionalidad asociada a cada uno de los servicios que, como mínimo, deben ofrecer.

5.2.2 *Secuencia Funcional*

En esta sección se muestra el funcionamiento básico del SMA a través de las interacciones entre agentes que pueden estar instanciados en un momento de vida del sistema y la información que manejan. Se muestra a través de tres diagramas simples; el primero de ellos identifica el primer paso, la *Consulta*, a continuación se muestra el proceso de *Búsqueda* y posteriormente el comportamiento en la *Catalogación*.

Un usuario desea obtener contenidos educativos que atiendan a su interés respecto al tema concreto de los trastornos pulmonares (en inglés, “Pulmonary Disorders”). Si realiza la consulta directamente a través de la interfaz gráfica de un repositorio, por ejemplo MERLOT, sólo puede realizar la búsqueda manualmente apareciendo varias opciones en la Web de resultados, tal y como se muestra en la Figura 5-1. Esto mismo debería de ir haciendo de manera manual en todos y cada uno de los repositorios en los que quisiera extraer información anotando separadamente sus resultados. Este procedimiento es una labor inabordable y muy poco eficaz, según se ha detallado en las secciones anteriores.

Una solución adecuada incorpora, como propone la arquitectura descrita en apartados anteriores, mecanismos de búsqueda federada. La búsqueda federada permite desarrollar búsquedas más exhaustivas en un conjunto extenso de repositorios, cuando estos poseen una capa de abstracción que permite procesos de búsquedas off-line. Dicha búsqueda federada permite habilitar mecanismos automáticos de recuperación de contenidos en dichos repositorios a través de la secuencia en la resolución de tres subproblemas: la selección de los repositorios, la selección de los recursos y finalmente la fusión

de los resultados. Haciendo uso de los componentes de la arquitectura AIREH presentada, a continuación se muestra la secuencia de funcionamiento general del sistema mediante la representación de las interacciones de los agentes en el sistema para la resolución de las tareas.

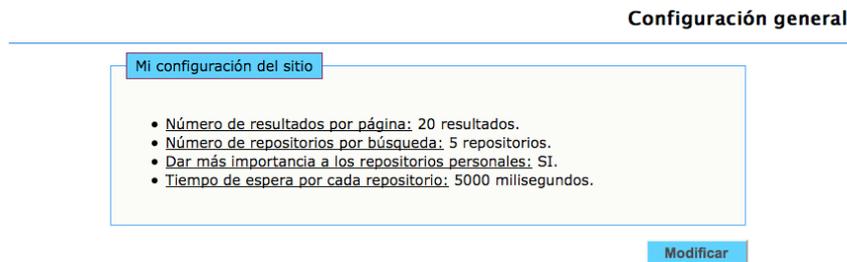


Figura 5-22. Parámetros generales configurables para la búsqueda

Una vez un usuario se da de alta en el sistema con datos personales (nombre, apellidos, nivel educativo y temas de interés), se genera un agente usuario con dicho perfil. El usuario puede además configurar parámetros generales sobre los procesos de búsqueda, Figura 5-22, limitando el número de repositorios y su prioridad en el proceso de selección de estos o el tiempo de espera, etc.

Cuando un usuario entra en la aplicación para solicitar contenidos educativos de su interés su Agente Usuario vincula su consulta al patrón de búsqueda e inicia el proceso enviando el patrón de búsqueda completo al rol Query Manager. La Figura 5-23, resume la secuencia funcional de la consulta desde que es emitida por el usuario hasta que se manda a los repositorios enumerando los agentes implicados en sus interacciones y que serán descritas en detalle.

III. Propuesta

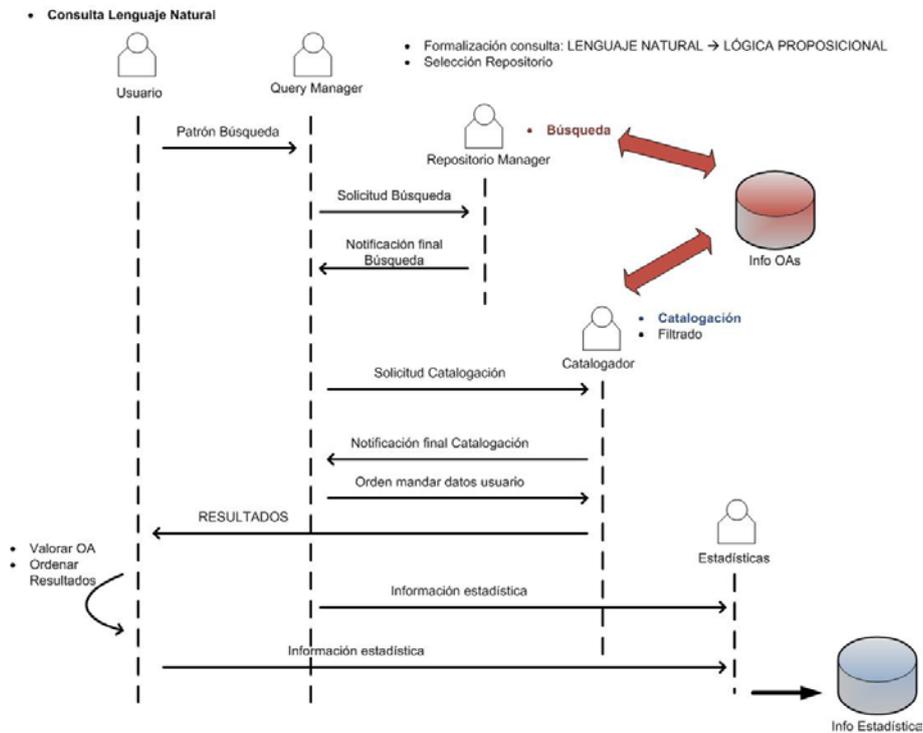


Figura 5-23. Diagrama de secuencia: Consulta

El sistema habilita dos modos de búsqueda, uno sencillo y otro más avanzado (Figura 5-24) que permite al usuario realizar la búsqueda seleccionando el término de la búsqueda y el idioma, lo que se almacena como la tupla de la consulta <tema, idioma>. Dicha tupla sería fácilmente ampliable con más características que refinasen la búsqueda, como formato del documento (pdf, word, etc.), autor, etc. A través de esta interfaz el usuario puede elegir repositorios sobre los que realizar las consultas. Existe más funcionalidades en la interfaz, tales como que el usuario lleve una gestión de su cartera de repositorios, incluso aquellos que exigen accesos con sesión (claves de acceso), incorpore repositorios nuevos, etc.

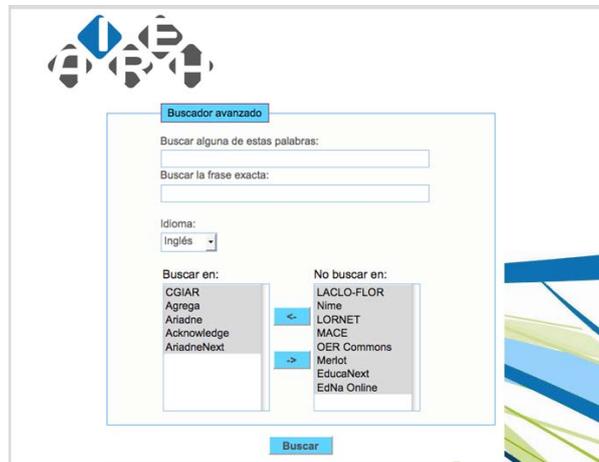


Figura 5-24. Interfaz de Usuario de Búsqueda de Contenidos en los Repositorios

Mediante lógica proposicional el Agente Query Manager genera la consulta planteada por el usuario a través de la interfaz, bien como la suma de sus términos individuales o como consulta en bloque:

Alguna de las palabras: Pulmonary Disorders \equiv Pulmonary \vee Disorders

La frase exacta: "Pulmonary Disorders " \equiv Pulmonary \wedge Disorders

El Agente Query Manager contacta con el Repositorio Manager enviándole la consulta formalizada. El Repositorio Manager analiza la disponibilidad de la información distribuida en dichos repositorios en base a tiempos de respuesta y eficiencia de los mismos entre los activos en ese momento y la vincula para su decisión final a la información estadística solicitada al agente Estadísticas junto con las condiciones establecidas por el usuario en el momento de la consulta (tiempos máximos de espera, número máximo de repositorios por búsqueda, etc). Es muy importante conocer la temática, el área y el contexto a los que pertenece la información almacenada por los repositorios, así como otras tantas características clave de las distintas fuentes de información oculta (tiempos de respuesta, eficiencia, etc.) y mantener una base de conocimiento sobre el funcionamiento operativo de dichos repositorios. Con esta información el Query Manager elabora un listado con los repositorios que optimizan el rendimiento en ese momento e instancia a los agentes ROA necesarios para que realicen la búsqueda federada enviéndoles la consulta en lenguaje formalizado.

La Figura 5-25, resume la secuencia funcional en la consulta y extracción de OA en cada repositorio una vez se instancia cada agente ROA. Cada uno de los agentes ROA incorpora todas las funcionalidades de la especificación de la capa

middleware del repositorio con el que realizan la comunicación para las consultas, incluyendo el tipo de lenguaje de consulta con el que opera el repositorio. La consulta es enviada en lenguaje formal al agente ROA en la instanciación del Repositorio Manager.

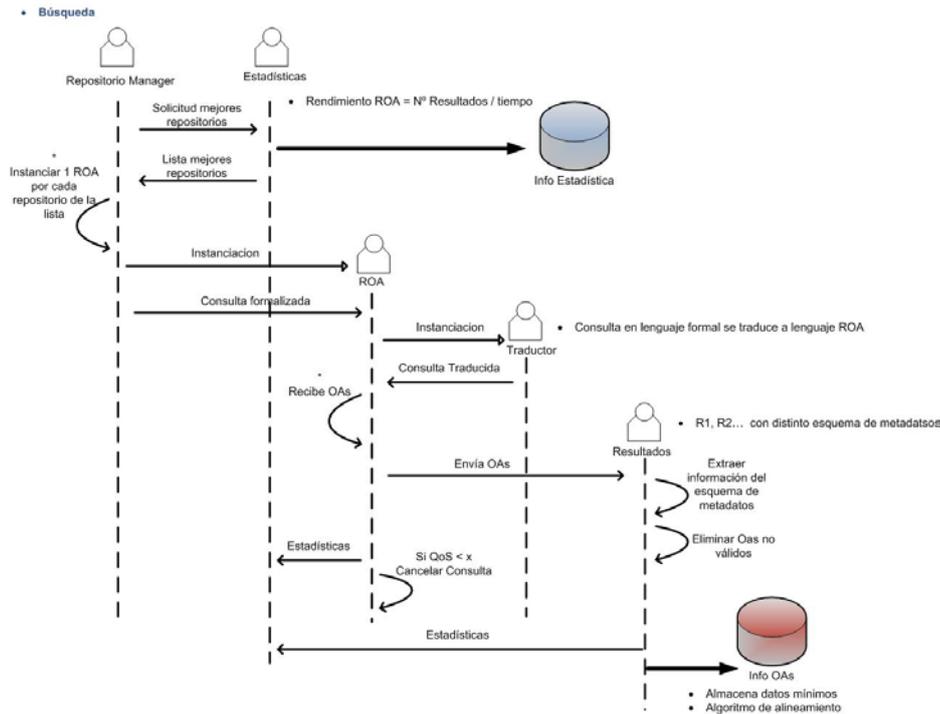


Figura 5-25. Esquema funcional Búsqueda de OA

Cada Agente ROA envía la consulta en lenguaje formal al agente traductor del lenguaje requerido por el repositorio. Existe un agente traductor por cada uno de los lenguajes requeridos por los ROA en esta consulta (*Very Simple Query Language* (VSQL), *ProLearn Query Language* (PLQL), *Lexicon Query Service* (LQS), etc). Cada agente ROA sólo instancia a un agente traductor, aunque cada uno de los agentes traductores puede recibir instancias de varios de los agentes ROA, que comparten el mismo lenguaje para su repositorio. El agente traductor les devolverá dicha consulta en el lenguaje requerido por cada ROA. La Tabla 5-7, muestra cómo la consulta emitida por el usuario a través de la interfaz en lenguaje natural es formalizada mediante álgebra relacional y una vez cada agente ROA se la envía al correspondiente agente traductor (en VSQL, ó en PLQL), éste se la devuelve traducida.

Tabla 5-7. Transformación de la consulta lenguaje natural a PLQL ó vsql

Lenguaje natural <term, lang>	<Pulmonary Disorders, english>	<"Pulmonary Disorders", english>
Álgebra relacional	$(\sigma_{term=Pulmonary}(R) \cup \sigma_{term=Disorders}(R)) \cap \sigma_{lang=en}(R)$	$\sigma_{term=Pulmonary\ Disorders}(R) \cap \sigma_{lang=en}(R)$
VSQ	<simpleQuery> <term> Pulmonary </term><term> Disorders </term> </simpleQuery>	<simpleQuery> <term> Pulmonary Disorders </term> </simpleQuery>
PLQL	lom.general.title = "Pulmonary" or lom.general.title = "Disorders" and lom.general.language = "en"	lom.general.title = "Pulmonary Disorders" and lom.general.language = "en"

Cada agente ROA, finalizada la extracción de OA de los repositorios, lo notificará al Repositorio Manager a la vez que envía los datos estadísticos de la consulta al agente Estadísticas. Finalmente cada ROA envía al Agente Resultados el conjunto de OA recuperados.

En una primera fase el agente resultados realiza una poda de entre todos los OA extraídos, filtrando aquellos OA no recuperables e incorporando aspectos de la petición del usuario que la falta de potencia expresiva en la consulta en el repositorio no hacen posible incorporar. Existe un rol de agente resultados por cada esquema de metadatos utilizado, por ejemplo uno para LOM, otro para Dublin-Core, etc. El agente resultados elimina los OA no recuperables analizando la cadena utilizada para acceder a este objeto educativo que depende de cada esquema de metadatos utilizado. Así la eliminación de los OA que no pueden ser recuperados depende del esquema y cada agente resultado

incorpora los servicios de extracción adecuados. Por ejemplo, el acceso al contenido educativo de un OA en el caso de un agente resultado que opere sobre LOM puede ser un localizador (URL) o un mecanismo que finalmente permite acceder a una localización (URI) y se representa mediante la etiqueta <location>. Además en LOM dentro de una instancia de metadatos es posible encontrar varias etiquetas <location>, ya que un OA podrá encontrarse replicado en diferentes sistemas con diferentes localizaciones. Lo mismo ocurre con el idioma en el que los usuarios quieren recuperar el OA, si el usuario desea contenidos en inglés, en ocasiones el poco poder de expresividad de los lenguajes de consulta del repositorio no permiten implementar en la consulta directa todos los requisitos del usuario necesarios para lograr la búsqueda de los contenidos, como en el caso del VSQL (ver Tabla 5-7). Por este motivo el Agente Resultados permite extraer en una segunda fase de entre todos los OA obtenidos y que son recuperables para un esquema de metadatos, aquellos que están en inglés. En el caso del VSQL para contenidos en LOM, el agente de resultado que opera con LOM, se quedará con todos aquellos que en la etiqueta <general> contiene el atributo <language="en">.

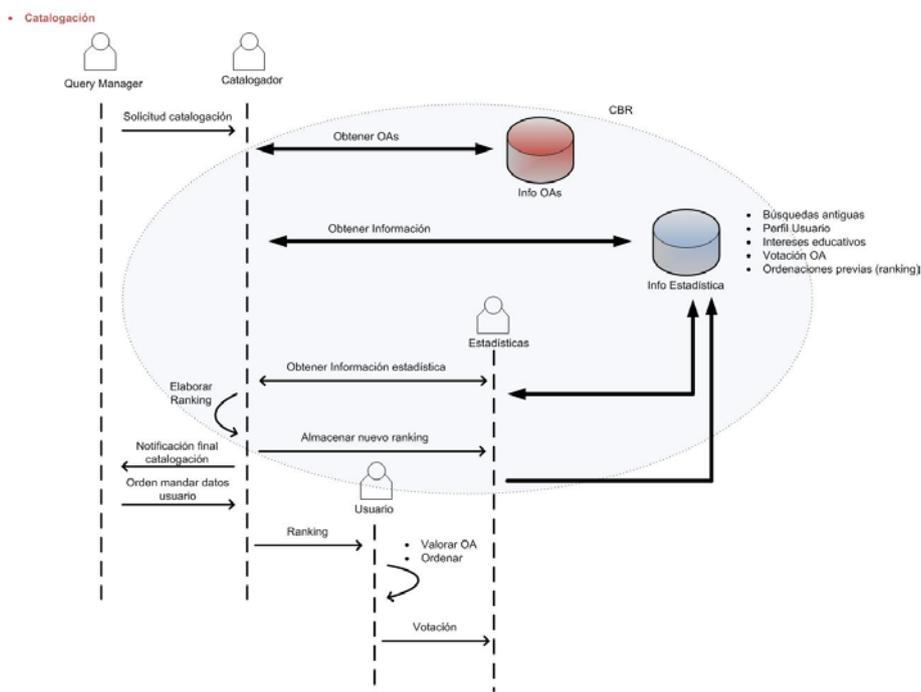


Figura 5-26. Diagrama funcional: catalogación de los OA recuperados

A partir de una necesidad de información y un conjunto de descripciones de éstos se recuperan los resultados y se decide cuáles serán los que tengan mayor probabilidad de satisfacer la consulta mediante el algoritmo de filtrado y

ordenación Q-Grams [Gravano and Ipeirotis, 2001] [Cheong and Tay, 2003] que permite la búsqueda de secuencias en las cadenas de caracteres de los objetos recuperados con respecto al idóneo para la consulta. Este método es utilizado con frecuencia como un método difuso ("fuzzy") de búsqueda en bases de datos ya que permite la búsqueda no exacta de consultas complejas. Genera además una cierta métrica determinada por el cociente de la distancia en q-gramas de cada secuencia. Este índice permite generar una valoración del emparejamiento de todos los OA en cada agente resultados respecto al idóneo, generando un primer acercamiento difuso a una fusión y ranking de resultados, que permite establecer un límite numérico inicial a los objetos recuperados que será mejorado a través del agente Catalogador.

El agente catalogador, Figura 5-26, es el encargado de realizar la fusión de los datos, asignándole un ranking a todos los OA pero formando una única lista que presenta al usuario una lista ordenada de resultados. El criterio de ordenación de dichos resultados se obtendrá incorporando el perfil de usuario, así como las valoraciones de otros usuarios en el contexto solicitando al agente estadísticas esta información para elaborar el ranking o prioridad en la presentación de los contenidos recuperados para un usuario determinado. La Figura 5-26 muestra la secuencia funcional para la catalogación de los OA recuperados. Para elaborar dicha ordenación de OA atendiendo a criterios de relevancia del usuario, el agente catalogador implementa un CBR.

5.2.3 Descripción del agente CBR-BDI para la recomendación

La elaboración de una única lista ordenada de OA que incorpore criterios de relevancia del usuario en el contexto es una de las tareas que realiza el agente catalogador que implementa un CBR.

En este sentido y como se detalló en la sección 4.3.1, los agentes CBR-BDI utilizan un sistema de razonamiento basado en casos. El sistema desarrollado viene caracterizado por trabajar con casos definidos con las características del contexto. Para facilitar la construcción de agentes CBR-BDI, distintos trabajos dentro del grupo BISITE [Corchado y Laza, 2003] [Bajo *et al.*, 2006] presentan un conjunto de equivalencias que funden las dos tecnologías a través de una biblioteca genérica. Un razonador basado en casos depende en gran medida la estructura y la representación y el contenido de su colección de casos. Por un lado el CBR, es un caso particular de técnica de búsqueda de vecinos próximos ampliamente utilizada en sistemas de recomendación y vista en la sección 3.3. Así los perfiles de usuario están representados por una colección de experiencias pasadas y, para recomendar un nuevo elemento, existe una amplia variedad de medidas de similitud que se pueden aplicar (Coeficiente de correlación de Pearson, función coseno, etc.). La similitud evalúa si un elemento se adapta a los intereses del usuario, de acuerdo al conocimiento que tiene el

sistema. Las técnicas de recuperación y su adaptación al CBR se han convertido en técnicas eficaces para el desarrollo de sistemas de recomendación [Cunningham, 2001], [Montaner *et al.*, 2002].

El uso de un razonador basado en casos tiene que ser fuertemente vinculado a la representación que del contexto se hace en la colección de casos. Cada caso en la base de casos del CBR se divide en dos partes. La primera de ellas describe un conjunto de atributos del elemento denominado objetivo (la definición del problema en la terminología CBR). El segundo de los elementos consiste en el conjunto de atributos que describe el interés del usuario en el tema dado (la solución al problema mediante un conjunto de metas en la terminología CBR).

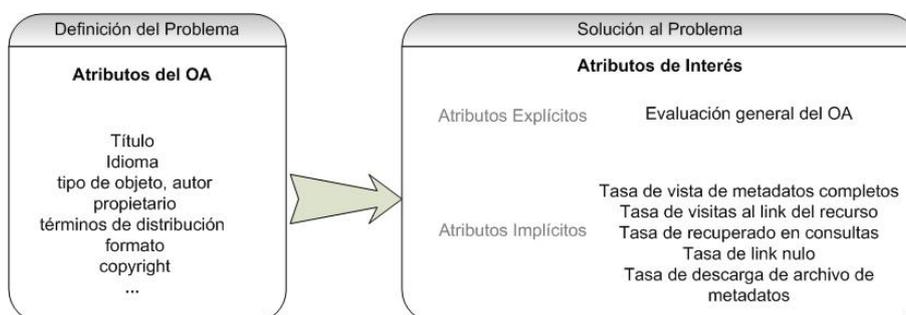


Figura 5-27. Ejemplo de Representaciones de casos en el dominio de los OA

La *definición de caso* que se proporciona en la biblioteca CBR-BDI dota de una gran potencia y flexibilidad al desarrollador, permitiéndole adaptar el caso a su problema concreto. En este trabajo el problema a solucionar consiste en posicionar ordenadamente, atendiendo a criterios de personalización, los contenidos educativos recuperados de distintos repositorios. Para ello se define el problema inicial en base a los elementos del contexto para los que es necesario proponer una lista ordenada de resultados relevantes de objetos de aprendizaje recuperados, según muestra la Figura 5-27.

Dado un conjunto de contenidos educativos $P = \{p_1, p_2, \dots, p_i\}$, cada uno de ellos, p_i , es caracterizado por una serie de atributos objetivos de entre el conjunto total T de posibles atributos. De manera que cada unidad de contenido educativo se representada por un vector $p_i = \{p_{i_1}, p_{i_2}, \dots, p_{i_n}\}$. En el caso concreto de los objetos de aprendizaje el conjunto total de atributos definido, $T = \{\text{título, idioma, keywords, formato, ...}\}$, es extraído fundamentalmente de la información descrita en su lenguaje de etiquetado en conformidad con el estándar de etiquetado utilizado (LOM, LOM-Es, DC, etc.). A la vez se introducen atributos que reflejan otros aspectos del entorno en el que ha sido recuperado, tales como el repositorio al que pertenece.

Una vez establecida la definición del problema en función de atributos, el objetivo del CBR es generar el ranking de dichos objetos de aprendizaje atendiendo a características particulares del usuario que se reflejen en las características de los objetos de aprendizaje disponibles, por ejemplo el nivel educativo del OA o el idioma del recurso, ver Figura 5-27.

La Tabla 5-8 muestra una descripción del problema formalizado mediante los atributos USER, QUERY, y PREF. El Campo QUERY se incluye para estimar otra propiedad objetiva pero sacándola del conjunto de parámetros objetivos favoreciendo un diseño modular del algoritmo de ranking que permita flexibilidad en la implementación. Mediante estos atributos, el sistema es capaz de obtener la descripción del problema a resolver. Además de estos atributos, la clase problemData proporciona los métodos que permiten actuar sobre el valor de dichos atributos. Los agentes CBR-BDI son agentes que trabajan con creencias, deseos e intenciones desde el punto de vista conceptual y con casos desde el punto de vista de implementación. Un caso está formado por una descripción del problema inicial con atributos como los descritos en la Figura 5-27, las soluciones propuestas y la eficiencia para la solución final.

Tabla 5-8. Atributos de los Casos

Campo del Caso	Tipo de Elemento
USER	Perfil del usuario (UserProfile)
QUERY	Consulta inicial del usuario (Query)
PREF	Preferencias del Usuario (Preference)
STIME	Sello de Tiempo (Time Stamp)

Cada usuario en el proceso tiene un grado de interés distinto en cada uno de los productos. Estos intereses pueden ser expresados por el usuario, formando parte de los denominados *atributos explícitos* o pueden ser capturados automáticamente por el sistema como consecuencia de la interactividad de dicho usuario, generando los llamados *atributos implícitos*. El interés explícito proporciona más confianza en el proceso de recomendación. Sin embargo, estos valores no siempre están disponibles. Los parámetros de estimación implícita del interés del usuario en el producto resultan de especial utilidad porque establecen una referencia directa al contexto en el que se realiza la evaluación con métodos menos invasivos para el usuario. En este trabajose distinguen los dos tipos de parámetros.

$$Int^e = \{voto\}$$

$$Int^i = \{visita\ link\ del\ recurso,\ vista\ metadatos\ completos,\ n^{\circ}\ veces\ recuperado,\ \dots\}$$

El parámetro explícito valora la votación de un usuario sobre un OA concreto, mientras que el implícito toma datos a través de la interacción del usuario con la interfaz (Figura 5-30) tales como si un OA ha sido visitado o ampliada la información sobre sus metadatos a través de la interfaz, etc.

Cada usuario tiene experiencia en varios artículos. Una experiencia guarda la información sobre las propiedades objetivas de un producto determinado Int^E , así como información subjetiva Int^I sobre el interés del usuario en ese tema.

$$E_i = \langle p_i, Int_i^I, Int_i^E \rangle$$

Por último, si un caso representa la experiencia del usuario en un determinado artículo, la base completa de casos para un usuario constituye la representación del perfil de usuario modelado en el contexto. Así, el sistema de recomendación mantiene una base de casos por cada usuario representando su perfil.

Formalmente, el sistema de recomendación se define a través de un escenario donde hay m usuarios $U = \{u_1, u_2, \dots, u_m\}$ y un conjunto de n OA, que denominamos $O = \{o_1, o_2, \dots, o_n\}$. Además, cada usuario u_i posee una lista de k valoraciones que ha realizado sobre un conjunto de OA I_{ui} , donde $I_{ui} \subseteq I$. En este contexto, una recomendación realizada por el usuario activo $u_a \in U$, consiste en un conjunto de N Objetos de aprendizaje $I_r \subset I$ de los cuales se evalúan de interés para dicho usuario en su contexto. Es importante señalar que en este ámbito de recomendación $I_r \cap I_{u_a} \neq \emptyset$, pues no ocurre como en comercio electrónico donde se desea que el usuario consuma productos nuevos, en este contexto para recomendar contenidos educativos no se plantea el perfil mercantilista de un negocio que busca diversificar las ofertas. En el contexto educativo un OA ya utilizado está en el modelo del usuario y puede ser nuevamente requerido y recomendado. El conjunto de valoraciones de los n usuarios sobre los m objetos de aprendizaje, constituye una matriz bidimensional.

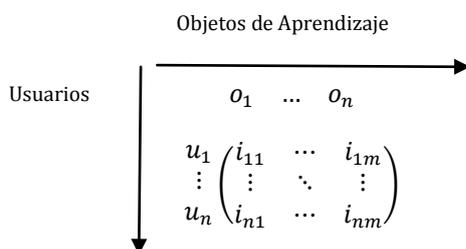


Figura 5-28. Matriz de valoración de OA por los votos de usuarios

Ampliando la experiencia de usuario con respecto cada OA añadiéndole las valoraciones sobre n elementos subjetivos, se obtiene una matriz tridimensional, que se puede visualizar mediante una superficie en el espacio tridimensional, según muestra la Figura 5-29.

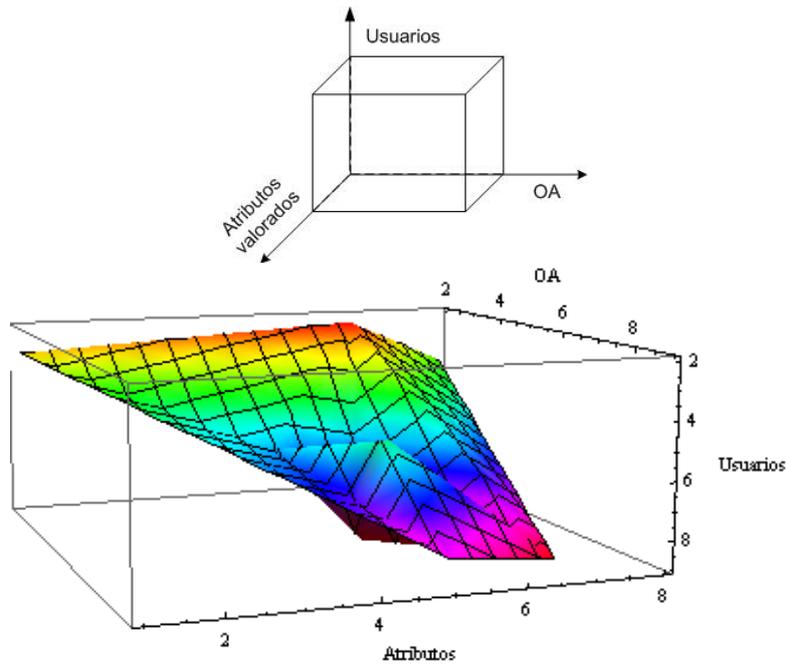


Figura 5-29. Representación del espacio \mathcal{R}^3 de experiencias de usuario

Cuando aplicamos CBR a la recomendación, en la *fase de recuperación* el sistema recupera elementos similares. Por lo tanto, la tarea de recuperación termina cuando el conjunto de los mejores elementos para una nueva búsqueda se ha encontrado. Definir el grado de similitud entre los elementos es el paso más importante de la fase de recuperación del CBR. El nivel de similitud entre dos ítems es computado mediante una función global de similitud, calculada mediante una ponderación en base a pesos de varias medidas de similitud. Para los atributos numéricos, funciones de similitud lineal y exponencial se han diseñado siguiendo [Falkman and Wilson, 1997]. Para los atributos etiquetados, se ha hecho uso del algoritmo de similitud semántica WordNet [Jiang and Conrath, 1997], ya que permite su uso para cualquier aplicación con consultas en lenguaje natural utilizando el diccionario relacional Wordnet®. Dicho algoritmo valora la relación de similitud entre dos términos en un valor $s \in [0,1]$ y calculado como una relación inversamente proporcional al número

de nodos a lo largo del camino más corto entre los términos. El camino más corto posible se produce cuando las dos entradas son iguales, en cuyo caso el valor de la similitud es de 1, generando el valor máximo de la relación de similitud. Aunque el Wordnet®³⁰ original está en inglés el proyecto crece hacia un consorcio de wordnets en todos los idiomas a través de The Global WordNet Association³¹, lo que permite la internacionalización del sistema que lo implementa.

Con una eficiente medida de similitud, dado un nuevo caso, podemos obtener una lista ordenada de casos similares. Haciendo uso de esta característica del CBR, cuando un usuario tiene interés en un tema bajo un contexto determinado, el sistema de recomendación puede generar una lista ordenada mediante un ranking de los elementos similares pero pesados en base al perfil del usuario.

La *fase de reutilización* consiste en adaptar las antiguas soluciones de los casos recuperados para el nuevo problema en base a las diferencias entre ellos. Una vez que el sistema ha recuperado un conjunto de elementos anteriores (los más similares), el sistema conoce el interés del usuario en los OA similares a través de los atributos de interés definidos (solución en la terminología CBR en la Figura 5-27). Suponiendo que el interés del usuario en un nuevo OA es similar a los intereses de otro usuario en OA similares, en la fase de reutilización, el sistema de recomendación calcula un valor que codifica el grado de interés para el nuevo elemento. Este valor se utiliza para decidir la posición en el ranking del conjunto de OA recuperados para el usuario en función del grado.

El valor del grado de interés se computa a partir de los valores de intereses de los OA similares seleccionados en la fase de recuperación. Calculamos esta en un proceso de dos pasos:

Primero se realiza la valoración del grado de interés para cada caso: $I_c = \mathcal{M}(f^e(Int_j^e), f^i(Int_j^i))$, donde f^e es la función que evalúa los intereses explícitos (las votaciones en nuestro sistema), f^i evalúa los intereses implícitos, \mathcal{M} es la función que combina los resultados numéricos de ambas funciones.

La función f^i , descrita por la ecuación (5.1), obtiene el resultado numérico de la agregación de los diferentes atributos implícitos, como una suma ponderada del conjunto establecido:

$$f^i = \sum_{j=1}^{|p_n^i|} \omega_j * Int_j^i \quad (5.1)$$

³⁰ WORDNET <http://wordnet.princeton.edu/>

³¹ GLOBAL WORDNET <http://www.globalwordnet.org/>

Donde $|p_n^i|$ es el número de atributos implícitos para el OA p_n , ω_j son los pesos asociados a los distintos atributos implícitos.

Finalmente la función de agregación de los atributos implícitos y explícitos computa la medida de evaluación general $\mathcal{M}(e, i) = \delta_e * f^e(Int_j^e) + \delta_i * f^i(Int_j^i)$.

El segundo paso consiste en calcular el grado de interés de cada OA recuperado r como la función de poderación del valor de interés sobre cada uno de los m OA similares j , extraídos en la fase de recuperación, ecuación (5.2). De manera que el grado de interés en cada OA r respecto a las preferencias del usuario es evaluado calculando los parámetros explícitos e implícitos respecto al perfil de usuario mediante la siguiente fórmula:

$$I_r = \frac{\sum_{j=1}^m (\text{Sim}(r, j) * \mathcal{M}_j)}{\sum_{j=1}^m \text{Sim}(r, j)} \quad (5.2)$$

Donde $\text{Sim}(r, j)$ es el valor de similitud obtenido en la fase de recuperación y \mathcal{M}_j es el valor del interés del OA i .

En la *fase de revisión*, el sistema evalúa el interés del usuario en la ordenación de los OA realizada. La idea es rastrear la interacción del usuario a través del sistema para conocer información relevante sobre el interés del usuario en los OA recomendados a través de la lista presentada, así como información explícita e implícita, a fin de conservar el nuevo caso.

III. Propuesta

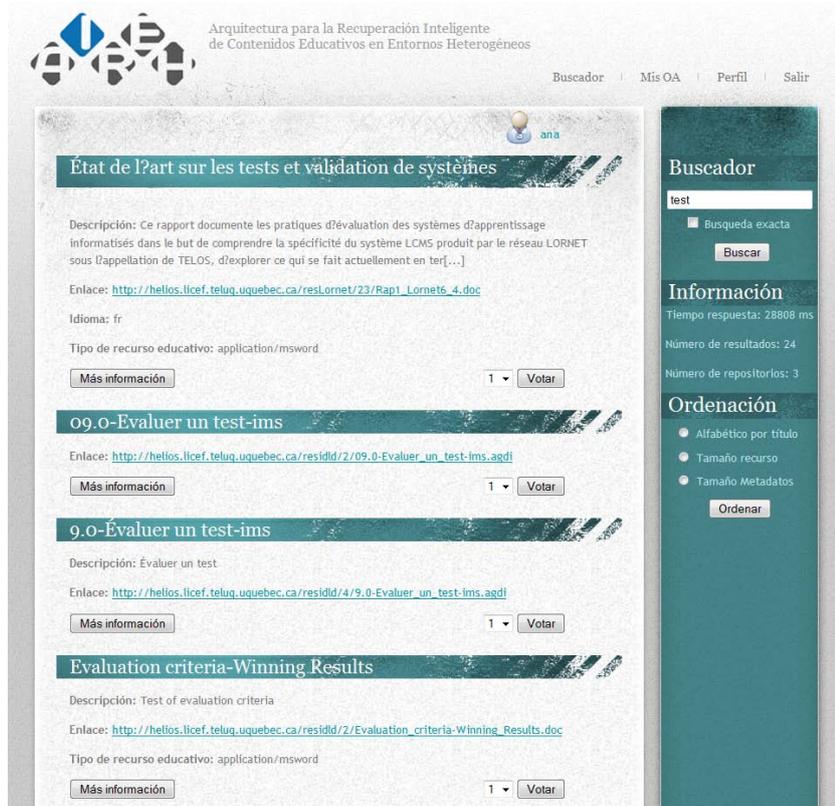


Figura 5-30. Presentación de resultados

Sobre la interfaz de presentación de los OA al usuario, Figura 5-30, éste puede interactuar visualizando la información etiquetada completa del objeto (Figura 5-31), descargándose, accediendo a la fuente del recurso educativo, valorándolo, etc.

Finalmente, en la *fase de retención*, el nuevo elemento se inserta en la base de casos con los atributos de interés que se agregaron en la fase de revisión.

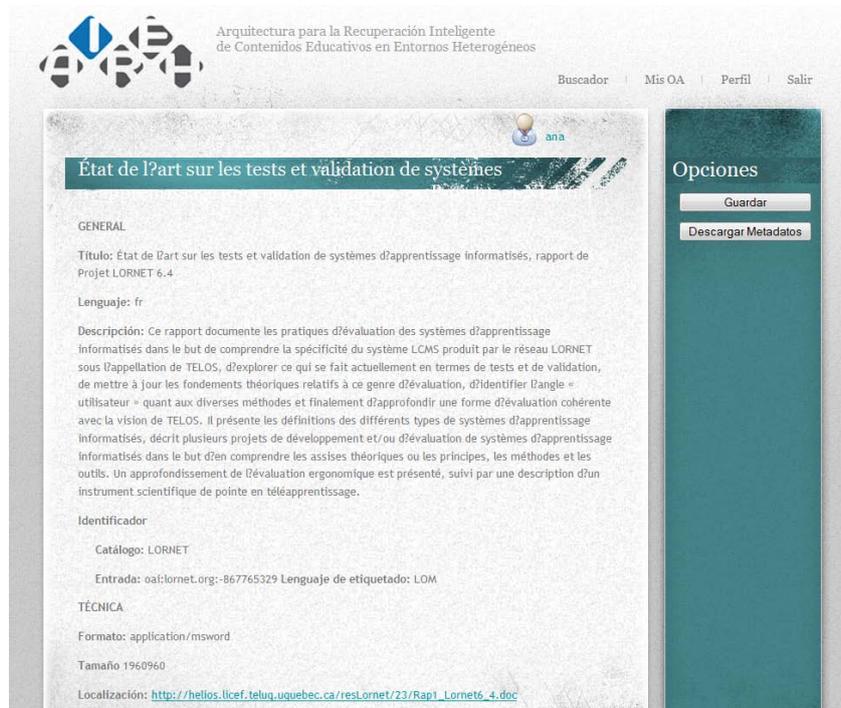


Figura 5-31. OA en la opción “Ver más” del usuario

5.3 CONCLUSIONES DEL CAPÍTULO

La solución propuesta pretende ser una arquitectura integradora, en la que una organización de agentes pueda ejecutar acciones de búsqueda y recuperación de contenidos educativos basándose en un modelo de búsqueda federada. La innovación de dicha arquitectura radica en utilizar un modelo organizativo para llevar a cabo sus objetivos, haciendo frente a entornos altamente dinámicos como es la Web profunda, donde se encuentran los contenidos. En la arquitectura se plantea una distribución de funciones analizada y diseñada a través de una guía metodológica, quedando abierta la posibilidad de añadir y modificar nuevas funcionalidades.

Se detalla el modelo de razonamiento del agente, mediante la incorporación de un modelo CBR aplicado a la recomendación de los OA. La definición de los casos y los parámetros implicados se abstraen del lenguaje de etiquetado, permitiendo la flexibilidad del proceso y su escalabilidad a otros contenidos.

III. Propuesta

El proceso de búsqueda queda integrado en los agentes de la organización de una manera totalmente transparente al usuario, haciendo frente a los problemas de distribución e integración de los diferentes repositorios, de abstracción de la lógica interna de cada uno de ellos y de clasificación, almacenamiento y búsqueda de OAs. Además de añadir capacidades como una sencilla escalabilidad a posibles situaciones de uso de nuevos protocolos, lógicas internas de repositorios, catalogación o aplicaciones heterogéneas con servicios encaminados a cubrir funcionalidades relacionadas.

PARTE IV. EVALUACIÓN Y
RESEARCH OVERVIEW

6 RESULTADOS EXPERIMENTALES Y EVALUACIÓN

Se ha presentado un nuevo enfoque para la búsqueda inteligente de contenidos educativos mediante la propuesta de una arquitectura basada en la aplicación de organizaciones. Dicha arquitectura se ha probado en el caso de estudio particular de búsqueda federada en repositorios de objetos de aprendizaje y posterior fusión de resultados obtenidos incluyendo aspectos de personalización. Las pruebas realizadas mediante el caso de estudio han permitido evaluar el modelo de arquitectura adoptada. A nivel teórico este nuevo modelo de sistemas multi-agente se adapta a la perfección a este contexto tan heterogéneo y dinámico. El principal objetivo de este trabajo de Tesis es demostrar que en la práctica, esta premisa razonable a nivel teórico, se cumple en la práctica en entornos un entorno real.

Para evaluar el rendimiento de la arquitectura propuesta en este trabajo en labores de recuperación de contenidos digitales educativos resulta preciso hacerlo conforme a un conjunto de métricas que permitan verificar su validez empíricamente. Siguiendo las consideraciones del estudio realizado en el Anexo B, se evalúa el comportamiento de la arquitectura mediante el uso de métricas empleadas en los procesos de recuperación de documentos relevantes distribuidos adaptadas al caso particular de repositorios de objetos de aprendizaje. Su adaptación junto a la definición de un parámetro nuevo, la ganancia temporal, al caso concreto de los objetos de aprendizaje establece como criterio de relevancia el permitir la reutilización de dichos objetos de aprendizaje como criterio de utilidad. La Exhaustividad y la Precisión adaptadas a la recuperación de los objetos de aprendizaje han sido definidas para este trabajo en la sección 3.2. y formuladas en las ecuaciones (3.1) y (3.2). La formulación de una nueva métrica, denominada Ganancia Temporal, es propuesta en este trabajo también en la sección 3.2, permitiendo calcular la eficacia de recuperación de contenidos en el tiempo.

Las siguientes subsecciones muestran resultados experimentales obtenidos para el caso de estudio, las conclusiones de la investigación, las contribuciones principales a la tecnología de agentes y SMA, y algunas líneas de trabajo futuro a corto, medio y largo plazo.

6.1 RESULTADOS

Los metadatos educativos constituyen un mecanismo sencillo para describir mediante etiquetas los contenidos a los que se se asocian. Todo esto queda ampliamente detallado y descrito a través de sus estándares en el capítulo 2 de esta memoria. Sin embargo el enfoque práctico de este trabajo de *investigación* exige conocer el uso real del etiquetado de los recursos educativos. Esta situación motiva el estudio empírico realizado sobre el uso del estándar más consolidado para los objetos de aprendizaje, el LOM. Partiendo de documentos reales de metadatos, recuperados de un conjunto de ROAs públicos, se elabora un informe de resultados que detalla la sección B.3 de los anexos. Dicho estudio evidencia cómo la dificultad del proceso de etiquetado, que es realizado fundamentalmente de forma manual, genera documentos con serias carencias. Dichas carencias afectan a problemas sintácticos en el etiquetado de los datos, ausencia de atributos fundamentales en la descripción, etc. Este hecho dificulta o imposibilita en ocasiones el uso directo de dichos metadatos en los procesos de búsqueda y recuperación de los contenidos a los que hacen referencia. La extensión de los niveles de especificación del estándar LOM revela asimismo atributos de difícil cumplimiento, otros irrelevantes o incluso algunos ambiguos. Como resultado de dicho estudio se establece una heurística que permite seleccionar unos cuantos elementos de entre todos los posibles para podar rápidamente en nuestro sistema la mayoría de los objetos irrelevantes recuperados. Una vez realizada la eliminación de los considerados irrelevantes se filtran los restantes teniendo en cuenta ciertos parámetros de calidad sobre los documentos de metadatos recuperados, integrando la información del contexto de usuario en el uso de dichos objetos de aprendizaje. El proceso es complejo y la asociación de más aspectos del mismo contribuye a mejorar la calidad de los contenidos recuperados. Un algoritmo de selección de contenidos eficaz para un usuario concreto debe de tener en consideración la semántica del objeto de aprendizaje y los aspectos técnicos utilizados para su búsqueda en los ROA. Dicho algoritmo de selección permitirá catalogar los resultados recuperados de forma automática mediante mecanismos inteligentes que impliquen las características y/o necesidades de dicho usuario en la búsqueda de OA. Para solventar estos requisitos en este trabajo el tratamiento de los documentos de metadatos recuperados atiende a tres aspectos, la completitud (*completeness*), confiabilidad (*reliability*) y la precisión de la ordenación de resultados (*ranking accuracy*) como criterios básicos en el desarrollo práctico del caso de estudio:

Completeness. La descripción del objeto de aprendizaje a través de sus metadatos debe de ser lo más completa posible para proporcionar resultados sobre los que planificar mecanismos de razonamiento automático que generen un aceptable nivel de confianza para su recomendación. Los atributos en los metadatos omitidos o mal etiquetados representan un agujero importante en la

precisión con la que han sido recuperados y pueden ser tratados. Algunos campos LOM pueden considerarse irrelevantes mientras que otros son absolutamente indispensables. Entre los que son elegidos como indispensables están el título, localización e idioma, para la automatización de los procesos de búsqueda y filtrado de OA relevantes para el usuario. La no existencia de título y localización es filtra a la entrada de cualquier OA en el sistema propuesto. Mientras que el completo etiquetado del contenido mejora su calidad en esta fase de preprocesado de los OA recuperados.

Confiabilidad. Una buena calidad en el etiquetado de los metadatos es fundamental para la recuperación de los OA correspondientes. Esto permite automatizar funcionalidades en la aplicación que vincula ambas partes del objeto (archivo de metadatos y contenido educativo) y dotan a los ROA que los albergan de un mayor índice de confianza en el sistema. Cualquier objeto de aprendizaje puede tener errores sintácticos o semánticos. Los errores sintácticos revelan una falta de conformidad con la especificación. En el caso de vocabularios controlados, se supone que uno debe utilizar alguno de los términos disponibles y no cualquier otro. Otros problemas son valores incompatibles con el atributo, faltas de ortografía, etc. Los errores semánticos sólo pueden ser detectados por la inspección heurística en el contexto. Algunos atributos pueden ser ambiguos o inadecuados de manera que el procesado de dicha información está sujeto a cierto grado de subjetividad que ha de limitarse lo más posible en los procesos automáticos con dichos metadatos. La propuesta presentada opta por incorporar los aspectos de confiabilidad mediante el uso de recomendación social.

Precisión en la ordenación de resultados. El conjunto de posibles objetos de aprendizaje para una petición del usuario puede ser muy grande. El algoritmo para la clasificación óptima de los resultados recuperados debe mostrar al usuario los objetos de aprendizaje recuperados mediante una lista ordenada de los mismos según un criterio de relevancia acorde a su petición y al contexto educativo en el que se realiza la consulta. La información mostrada al usuario sobre los objetos pertinentes localizados es importante para facilitar el proceso de elección último del usuario. El aprendizaje del sistema sobre los atributos de dominio relacionándolos con el usuario resulta un mecanismo clave en el proceso de ordenación de los OA implementado. Este trabajo incorpora como mecanismo de aprendizaje un agente CBR que permite ordenar los OA recuperados incorporando personalización.

Una vez descritos los ejes de aplicación del desarrollo práctico del caso de estudio veremos en puntos siguientes las pruebas realizadas y la evaluación de la recuperación de la arquitectura propuesta AIREH.

6.1.1 *Análisis del Caso de Estudio*

No cabe duda de que en los resultados de este trabajo tiene un peso importante el caso de estudio diseñado para la verificación de la arquitectura de recuperación. Sin embargo, el objetivo del mismo es la validación de dicha arquitectura de recuperación de contenidos educativos propuesta mediante un caso de estudio genérico. El caso de estudio se ha creado como un sistema de búsqueda federada en el que se dispone de 2 agentes ROA, un agente Usuario que realiza consultas y un agente de tipo Traductor. Los dos agentes ROA implementan el código cliente de la interfaz de acceso a repositorios SQL.

Tabla 6-1. Términos Seleccionados para la Batería de Consultas

Formalized languages, Formal Systems, Recursive Functions, Computer Sciences, Algorithmic languages, Analog computing, Artificial intelligence, Codes and coding Systems, Computer software, Digital computing, Heuristics, Hybrid computing, Simulation, Algorithm construction, Iterative methods, Control Systems, Cybernetics, Dynamic Programming, Linear Programming, Computing for statistics, Speech Sensor systems design, Storage devices, Circuit design, Processing engineering, Automation technology, Timing devices, Telecommunications, Radiocommunications, Machine translation, Architectural design, Bioacoustics, Educational theories, Cardiology, Allergies, Pulmonary diseases, Ethnolinguistics, Semiology, Dermatology, Pediatrics, Control devices, Data transmission devices, Real-time Systems, Central processing units, Computer architecture, Surgery, Fibre optics, Biochemical genetics, Microwave links, Computer technology, Informatics, Computing systems design technology, Arithmetic and machine instructions, Information systems design and components, Programming languages, Programming theory, Character recognition systems, Broadcasting, sound and television, Computer terminals, Graphie, Display devices and plotters

Se decide realizar una serie de pruebas de rendimiento de la recuperación sobre dos repositorios MERLOT³² y LORNET³³ (*stateless*). La elección de dichos dos repositorios está sujeta a varios aspectos. MERLOT es uno de los repositorios de objetos de aprendizaje a nivel de enseñanza superior más reconocidos disponibles en la actualidad. El repositorio LORNET para objetos aprendizaje es en realidad un gran proyecto que constituye una red de

³² MERLOT <http://www.merlot.org/>

³³ LORNET <http://www.siat.sfu.ca/research/projects/29/>

investigación de las universidades de Canadá. Por otro lado cada uno tiene un sistema de gestión de sesiones diferente, así MERLOT es de tipo *statefull* mientras que el repositorio LORNET es de tipo *stateless*. Por su parte en el caso de estudio el agente Traductor implementa para ambos el lenguaje de consulta VSQL. Ambos repositorios se ha hecho cargo de la tecnología desarrollada previamente en proyectos anteriores y constituyen los dos más relevantes en cuanto al número de objetos de aprendizaje que albergan. Son repositorios que se encuentran en continúa evolución tecnológica, lo que permite un banco de pruebas actualizado de la arquitectura presentada.

La arquitectura AIREH ha sido evaluada realizando una batería de pruebas para validar su eficiencia en entornos reales. Para realizar esta prueba se ha ejecutado una batería de 60 consultas en cada uno de los repositorios escogidos (Merlot y Lornet). Como punto de partida de dichas consultas se han tomado 60 códigos al azar, Tabla 6-1, dentro del campo de la ciencia y la tecnología dentro de los códigos UNESCO³⁴.

Se realiza la búsqueda sobre el mismo conjunto de consultas para poder tener una base de datos reales con los que comparar la solución desarrollada con respecto a estudios preliminares realizados a lo largo del trabajo y detallados en el Anexo B.

6.1.2 Rendimiento para la Arquitectura de Recuperación

Realizadas las 60 consultas detalladas en el apartado anterior se realiza una evaluación según las métricas ya descritas en la la sección 3.2.

El sistema es robusto frente a fallos, debido a que incorpora varios métodos en diferentes agentes en la organización durante todo el tiempo en el que transcurre la consulta. El agente Repositorio Manager, verifica los repositorios activos ante cualquier solicitud del agente Query Manager antes de elaborar la ordenación en la planificación de repositorios atendiendo a su rendimiento. Una vez son instanciados los agentes ROA que realizan la consulta en cada repositorio, cada uno de ellos está encargado de cancelar la consulta y notificar la situación al agente estadísticas si se producen situaciones que afectan a los niveles QoS establecidos, tales como tiempo de consulta, rendimiento del repositorio, etc. Aunque alguno de estos ROA cancele su operación de consulta, aún estarían operando el resto de agentes ROA instanciados, siempre monitorizados por el Repositorio Manager.

Debido a que el sistema filtra todos aquellos repositorios de los que no obtiene respuesta, el número de OA a los que tiene acceso el usuario consiste en la

³⁴ UNESCO <http://www.unesco.org/new/en/unesco/>

suma de los OA recuperados de los Repositorios activos en el momento de la consulta, según se muestra en la Figura 6-1.

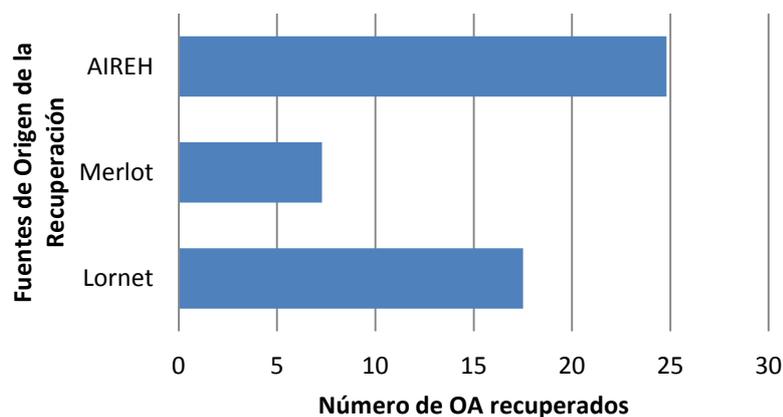


Figura 6-1. Documentos de Metadatos recuperados por el sistema

Este dato, resulta revelador en cuanto que eleva de manera importante el número de OA recuperados en el tiempo, haciendo crecer el valor de la ganancia temporal del sistema alrededor de un 15% de media. La Figura 6-2, muestra las ganancias temporales medias para la serie de 60 consultas realizadas.

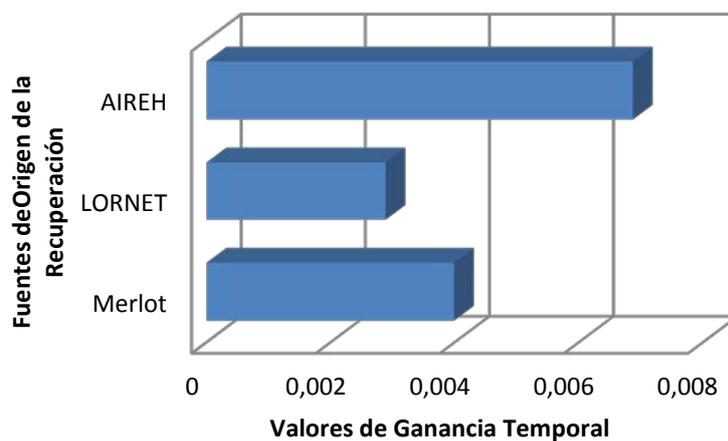


Figura 6-2. Valores de la Ganancia Temporal

En todas las consultas se obtienen resultados relevantes, con los criterios de relevancia que se han establecido en este trabajo. La recuperación de contenidos por la arquitectura propuesta depende de los OA devueltos por los repositorios aislados a los que accede, de manera que en el caso de que no exista ningún OA que atienda a la consulta requerida por el usuario, el sistema AIREH no puede solventar la falta de contenidos. Sin embargo la arquitectura asegura que de existir algún contenido para la petición del usuario, este será recuperado.

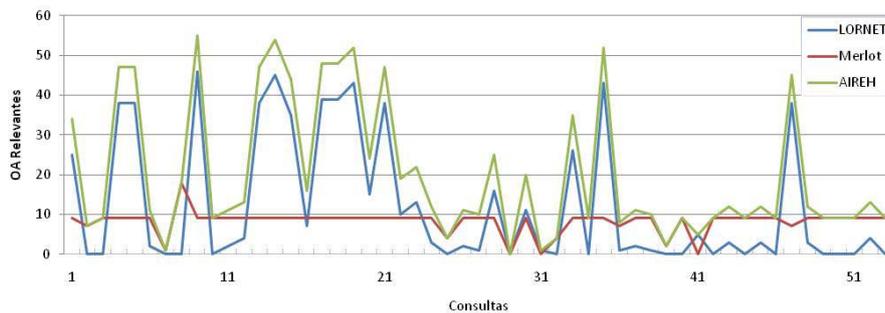


Figura 6-3. Comparativa de Documentos de Metadatos relevantes

Esta capacidad de obtener OA relevantes para el usuario se ve descrita en la obtención de las medidas de relevancia y exhaustividad evaluadas. La Figura 6-4 muestra la relevancia de AIREH, en su funcionamiento interno, algo inferior de media al repositorio Merlot, debido a la baja relevancia de contenidos en Lornet. Sin embargo la relevancia para el usuario, siempre según los criterios manejados en este trabajo sube a su valor máximo, 1, debido a que el sistema filtra todos aquellos OA recuperados que no tienen información sobre la ubicación del recurso en los metadatos (por ejemplo la Inexistencia de información para el atributo <location> de la categoría <technical> de LOM))

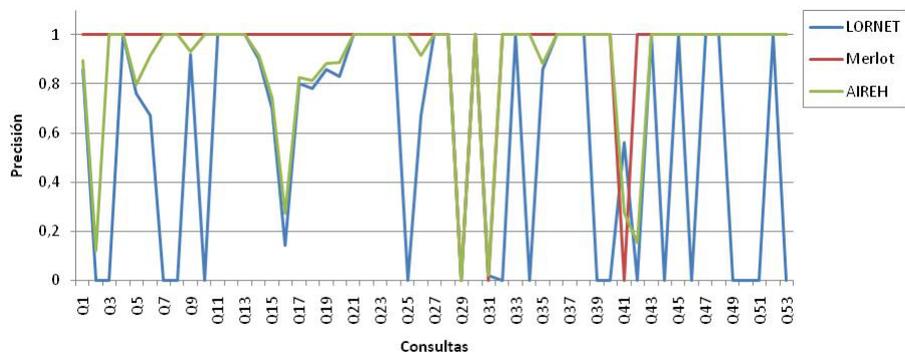


Figura 6-4. Comparativas de precisión a bajo nivel para el modelo AIREH

Como muestran los valores de la Figura 6-5 , el sistema propuesto optimiza la exhaustividad relativa, debido a que recupera todos los OA relevantes de los repositorios fuente de la búsqueda.

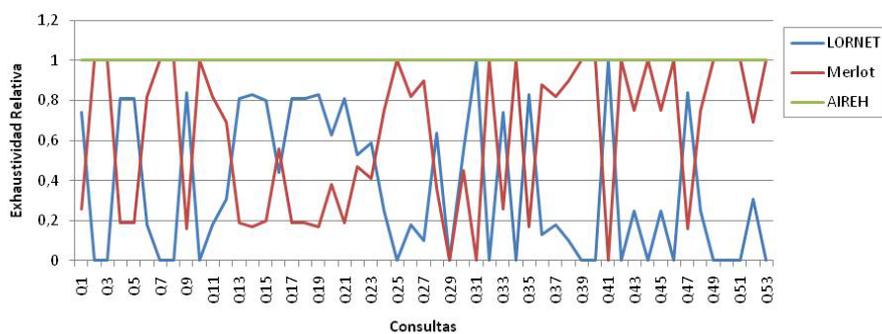


Figura 6-5. Comparativas de la Exhaustividad

Como se aprecia, tanto en los datos, como en las gráficas de evaluación el número de resultados ha mejorado considerablemente al utilizar una solución basada en la búsqueda federada, ver Figura 6-1.

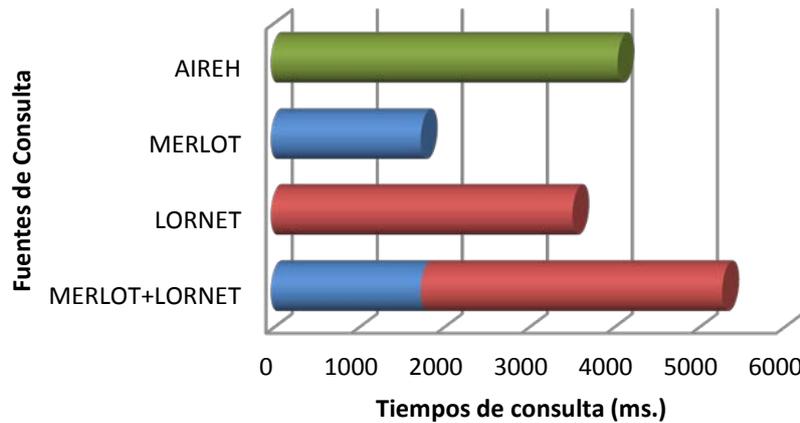


Figura 6-6. Comparativa de tiempos de consulta en ms. Entre las distintas fuentes

Se consigue dicho aumento en el número y la calidad de los resultados sin que además se incremente el tiempo de consulta. Esta optimización en los tiempos de consulta se debe a la paralelización realizada en la recuperación que consigue minimizar los tiempos entorno a un 15% de media sobre el repositorio que mayor tiempo requiere. Sin embargo este dato ha de ser puntualizado debido a que AIREH devuelve como solución a la consulta del usuario un conjunto de OA pertenecientes a los distintos repositorios, donde se han filtrado aquellos no reutilizables y los restantes se ordenan atendiendo a aspectos de personalización ya descritos. Ambos datos confieren a la arquitectura AIREH una importante mejora respecto al modo individual de búsquedas sobre este tipo de sistemas.

El mecanismo de búsqueda federada, se ha abordado como la resolución del problema de la recuperación de contenidos solventando tres fases: (1) la selección de los repositorios, (2) la recuperación de los contenidos y finalmente (3) la fusión de los Resultados. Además de la búsqueda centralizada de objetos de aprendizaje en distintos ROAs, el sistema realiza una clasificación de los objetos recuperados a través de un primer filtrado, mediante la eliminación de objetos de aprendizaje duplicados y de los no se detalla el acceso a la fuente del contenido. Esta etapa aborda la primera de las fases propuestas en la búsqueda federada mediante la selección de los ROA más idóneos de acuerdo a parámetros estadísticos y de rendimiento durante la sesión de búsqueda y la recuperación de los contenidos en base a la búsqueda del usuario a través de términos clave. Inmediatamente realiza un cribado inicial de los OA recuperados que facilita una primera organización homogénea del conjunto de documentos de metadatos recuperados. Este pretatamiento de los contenidos

recuperados fortalece la consecución de la segunda de las fases, la recuperación, en la búsqueda federada.

The screenshot shows a search results interface. At the top left is a logo with the letters 'A', 'R', 'E' and the text 'Arquitectura para la Recuperación Inteligente de Contenidos Educativos en Entornos Heterogéneos'. The main content area displays three search results:

- Accessible e-Learning**: Descripción: These training materials explore the theme of accessibility in relation to simple 'home grown' e-Learning, illustrating how simple techniques can greatly add to the learning experience of many learners. The inherent contradictions of 'universal access[...]
- ANG 1100 First level to begin learning english**: Descripción: Ce cours vous propose une approche fonctionnelle à l'apprentissage de la langue anglaise. Elle vise donc à développer la capacité d'utiliser cette langue dans les situations réelles où un besoin se fait sentir. La langue est donc considérée avant tout[...]
- Collaborative Learning Design: JigSaw**: Descripción: This is a version of a collaborative scenario where a group of learner explore a subject or solve a problem together. The group is subdivided in 4 groups labeled GROUP 1,2,3,4 of at least 4 learners.

On the right side, there is a search bar with the text 'learning object', a 'Buscador' section with a 'Buscar' button, an 'Información' section showing 'Tiempo respuesta: 46058 ms', 'Número de resultados: 50', and 'Número de repositorios: 3', and a 'Registro' section with fields for 'Usuario:' and 'Contraseña:' and an 'Entrar' button.

Figura 6-7. Presentación de resultados de una búsqueda al usuario

Una segunda etapa de filtrado incorpora los aspectos de calidad de los objetos recuperados en función con dos criterios (1) la valoración de calidad de los metadatos recuperados y (2) la evaluación del OA por los usuarios estimada a través de técnicas colaborativas. La valoración de la calidad de los metadatos recuperados incorpora parámetros del mismo tales como su tamaño (se estima que un mayor tamaño mejora la calidad informativa del contenido), la incorporación de las características principales (Título, autor, palabras clave, etc.), completitud, nivel educativo, etc. El segundo aspecto incorpora la valoración colectiva en el uso del OA a través de las evaluaciones de otros usuarios en la aplicación, mediante aspectos sociales implementados en la recomendación. La suma de dichos criterios sobre el ranking de objetos recuperados proporciona en el sistema una recomendación híbrida que comienza con una recomendación basada en contenidos y es refinada con

características colaborativas. Esta etapa concluye con la tercera de las fases planteadas en la búsqueda federada incluyendo características de recomendación en la fusión de los contenidos. La inclusión de aspectos de recomendación mejora la calidad de los contenidos recuperados para el usuario que generó la consulta debido a la personalización que de los contenidos se hace.

El mecanismo de recomendación empleado está basado en un método de filtrado híbrido que combina técnicas basadas en contenido con aspectos colaborativos. Este aspecto colaborativo se obtiene de las votaciones de los usuarios sobre los OA y se recogen a través de la interfaz, Figura 6-7 y pasan a formar parte de los atributos explícitos en la base de casos, descrita en la sección 5.2.3. El usuario puede emitir una valoración general de los objetos de aprendizaje recuperados a través de la interfaz, en una escala numérica de 1 a 10.

Para realizar la recomendación, se hace uso de modelo de razonamiento mediante un modelo CBR. El sistema accede a experiencias pasadas de los usuarios almacenadas en la base de casos descrita en el apartado 5.2.3. Se extrae de la base de casos un conjunto de estos que sean similares a la consulta introducida inicialmente. La base de casos está organizada de tal forma que se monitorizará la interacción del usuario en el contexto rellenando los atributos de interés para el usuario sobre los OA (caso).

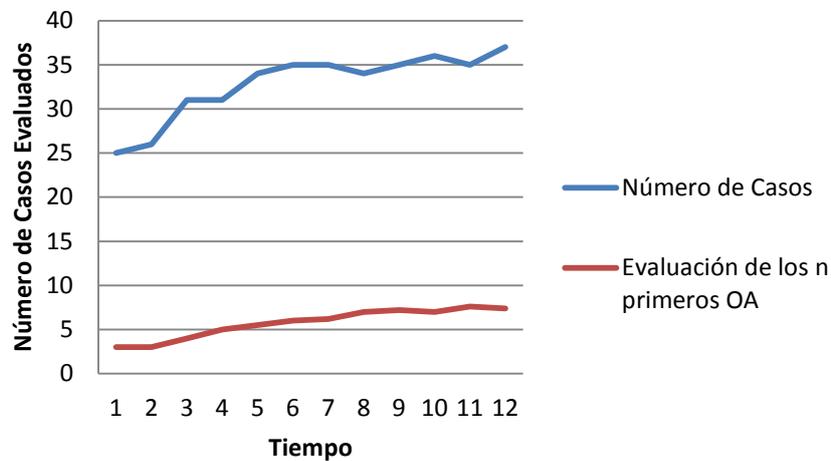


Figura 6-8. Evaluación de las recomendaciones del CBR

Con el grupo de casos recuperados se realiza la recomendación. Para validar la aplicación, se han analizado los resultados obtenidos con AIREH mediante la evaluación realizada durante 6 meses con 10 usuarios realizando la misma batería de consultas de una selección de palabras clave recogidas en la Tabla 6-1. La evolución del número de casos en la base de casos permite tener un mayor conocimiento y valoración de los posibles OA, así como de los usuarios. La Figura 6-8, muestra cómo de media los usuarios manifiestan un incremento en la valoración de los n primeros OA recomendados (n=5), a medida que el sistema resuelve nuevos casos. El éxito en el sistema se evalúa a través de la interacción del usuario con los OA recomendados, así como por la valoración que éste hace de los OA mostrados por el sistema a través de la interfaz. Esta mejora se debe a la capacidad del sistema para aprender y adaptarse a las experiencias adquiridas. De la misma forma, las experiencias obtenidas permiten una mejor adaptación al perfil del usuario. Es lógico dicha mejora en la recomendación, ya que al aumentar la variabilidad de los casos almacenados y su número, la posibilidad de encontrar casos parecidos al de la consulta aumenta y permite un mejor filtrado y ordenación de los OA recuperados. El mecanismo permite validar los OA recuperados con criterios que definen mejor los gustos y/o necesidades del usuario.

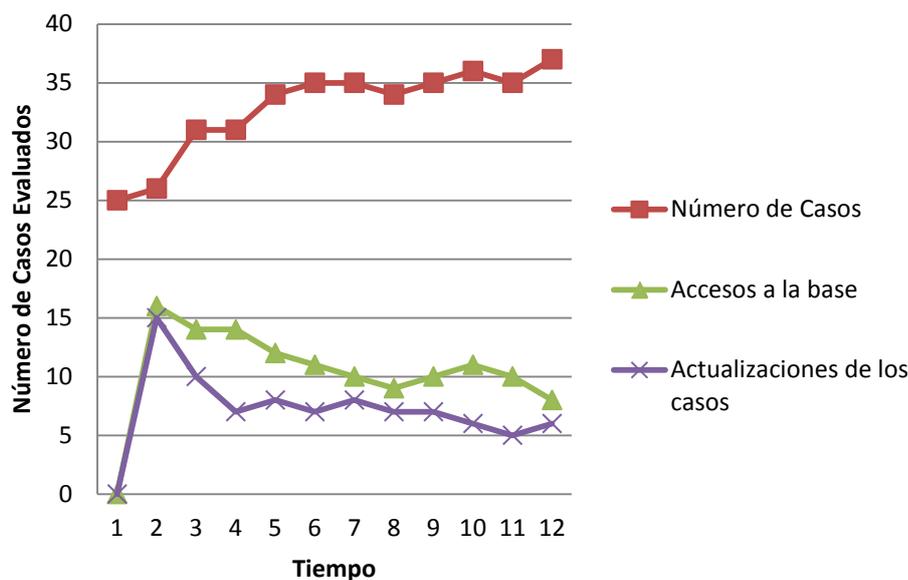


Figura 6-9. Evolución del CBR

La Figura 6-9 es posible apreciar cómo el número de actualizaciones de los casos desciende a medida que el sistema adquiere experiencias (el eje X representa el número de evaluaciones realizadas y el eje Y cuantifica el número

de casos referidos a los aspectos evaluados (número de casos, accesos a la base y actualizaciones de los casos).

Se ha realizado una limitación en el número de casos de la base en 50 casos para evitar el problema del crecimiento ilimitado de la base de casos. En situaciones en la que fuese necesario eliminar un caso, se hace uso del atributo *STIME* del caso, utilizado como marca de tiempo. Aquel caso con atributo menos próximo al tiempo de ejecución sería eliminado y sustituido por el nuevo caso.

AIREH propone una arquitectura que flexibiliza el desarrollo de soluciones mediante la incorporación de funcionalidades en la gestión de búsquedas de contenidos en repositorios de objetos de aprendizaje. Los recursos de referencia son los metadatos obtenidos mediante búsquedas federadas en distintos repositorios. La propuesta permite además la gestión de usuarios a los distintos repositorios y derechos de acceso, etc. Por este motivo parece razonable realizar una comparativa en cuanto a competencias en el desarrollo de AIREH con algunas herramientas que están apareciendo en la actualidad y pretenden dar solución al problema de la recuperación de contenido educativo disperso. Las más relevantes, estando aún fase de desarrollo y pruebas son la herramienta de búsqueda de GLOBE, utilizada para la búsqueda a través de un cliente web en los repositorios que forman parte de su alianza y que resulta ser la herramienta de ARIADNE³⁵ en estado beta ligada ahora al proyecto GLOBE³⁶. Dicho buscador a través de cliente web tiene la apariencia que muestra la Figura 6-10.



Figura 6-10 Buscador de OA en GLOBE

³⁵ ARIADNE <http://ariadne.cti.espol.edu.ec/>

³⁶ GLOBEFINDER <http://ariadne.cti.espol.edu.ec/globeFinder/>

También resultan relevante los trabajos [Paquette *et al.*, 2006] que han dado como resultado la actual herramienta PALOMAWeb³⁷, que permite la gestión de OA a través de sus metadatos y cuya interfaz aparece en la Figura 6-11.

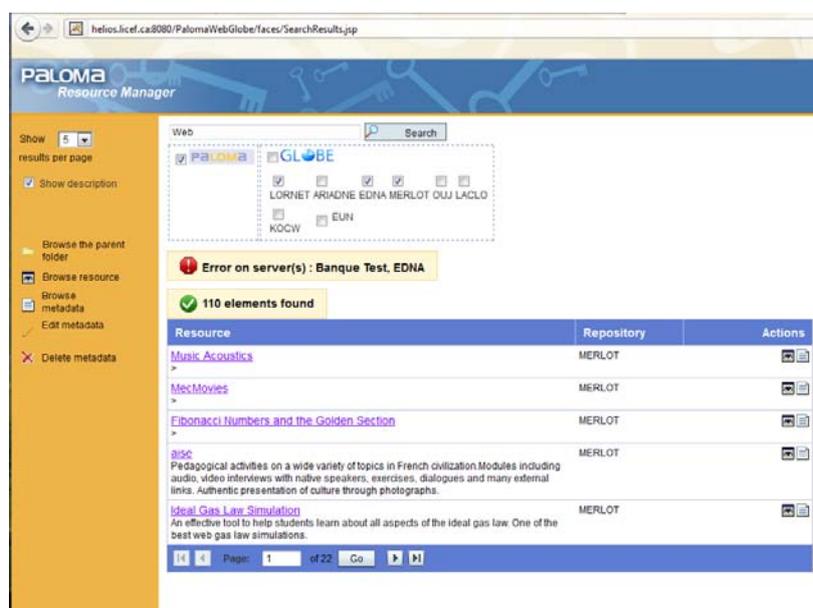


Figura 6-11. Buscador de OA PALOMA

Se han establecido una serie de características que muestran en la Tabla 6-2, la versatilidad en la búsqueda de contenidos digitales educativos por parte de las herramientas seleccionadas en comparación con la arquitectura propuesta. Sólo se han tomado aspectos perceptibles por el usuario debido a que no existe información relativa a las metodologías y tecnologías empleadas por estas herramientas.

Resulta evidente que la arquitectura propuesta en este trabajo para la búsqueda de contenidos educativos incorpora importantes características, tales como la personalización y la gestión de los metadatos que mejoran la labor del usuario y el tratamiento de los OA recuperados. La propuesta realizada, permite además un acceso flexible a los repositorios que el usuario establezca de interés, la incorporación de lenguajes de consulta de manera sencilla, siempre que los repositorios consultados cumplan con las características técnicas necesarias. Estas características de flexibilidad y adaptabilidad al entorno junto a la personalización de los resultados obtenidos presentan a AIREH como una solución eficaz al problema de contenido distribuido.

³⁷ PALOMAWEB <http://helios.lice.fca.8080/PalomaWebGlobe/>?

Tabla 6-2. Características Comparativas de buscadores Federados de OA

Soluciones a la Búsqueda de Recursos Educativos			
Características/Herramientas	PALOMA	GLOBE	AIREH
Acceso a Distintos Repositorios Distribuidos	SÍ	SÍ (sólo de la Alianza que lo conforma) Actualmente sólo a ARIADNE (En versión Beta)	SI
Inclusión de distintos lenguajes de consulta	NO	NO	SÍ
Personalización	NO	NO	SÍ
Incorporación de aspectos sociales	NO	NO	SÍ
Acceso a los Metadatos por parte del usuario	SÍ	NO	SÍ
Almacenamiento de histórico de Búsquedas	NO	NO	SÍ
Permite buscar con Criterios de Idioma	NO	No implementado	SÍ
Distintos estándares de etiquetado	No, sólo LOM	SÍ	SÍ
Cliente Web	SÍ	SÍ	SÍ

Hasta el momento para la recuperación de los contenidos en los ROA se ha utilizado la búsqueda por palabras clave en el contenido de los metadatos. El uso de clasificadores multietiqueta es en la actualidad una promesa a la construcción de potentes mecanismos para incrementar el poder de búsqueda en grandes volúmenes de datos etiquetados. El punto siguiente evalúa el trabajo realizado al respecto en este trabajo con la idea de sondear posibles mejoras futuras que refinen y amplíen las posibilidades de la arquitectura AIREH propuesta.

6.1.3 Técnicas de Clasificación Multietiquetado de Objetos de Aprendizaje

Los clasificadores multi-etiqueta, estudiados en el apartado 3.4.2.1, constituyen una técnica de la minería web aplicada al aprendizaje automático sobre colecciones de información. Dichos métodos permiten el aprendizaje off-line sobre conjuntos de datos etiquetados mediante diversos algoritmos. Parece razonable pensar que esta metodología se podría utilizar para la clasificación automatizada de OA en las colecciones extrayendo las etiquetas que clasifican los aspectos relevantes de cada OA. Esta nueva capacidad de consulta sería útil

para mejorar la eficacia en la selección de OA en distintos ROA. Permitiría extender el poder de expresividad en el filtrado de los objetos recuperados dentro del modelo de arquitectura presentado. Mediante la inclusión de esta nueva información como un campo adicional dentro de la base de casos del CBR, permitiría incrementar aspectos de similitud entre OA en los procesos de razonamiento. Igualmente podía hacerse uso de este tipo de técnicas para la clasificación y selección de los repositorios de objetos de aprendizaje. Para estudiar estas posibilidades se planificó un aprendizaje automático, en la construcción de un modelo multi-etiqueta con un conjunto de datos de entrenamiento de objetos de aprendizaje para posteriormente aplicarlo a un conjunto nuevo de datos (sin etiquetar), con el fin de obtener predicciones que puedan realizar una clasificación de casos nuevos.

Se utilizó un método de aprendizaje automático, similar a [Trohidis *et al.*, 2008], para realizar una evaluación empírica de los algoritmos MLkNN y RAKEL en un conjunto de datos multietiquetados de objetos de aprendizaje recuperados. El algoritmo RAKEL fue seleccionado, como un método que ha demostrado ser más eficaz que los dos primeros enumerados [Trohidis *et al.*, 2008]. Dicho método construye un conjunto formado por la unión de clasificadores LP. De esta manera RAKEL consigue tener en cuenta las correlaciones entre etiquetas, evitando los problemas de LP. El ranking de las etiquetas se obtiene por un promedio de las predicciones [0,1] de cada modelo establecido por etiqueta. A partir de los valores obtenidos se establece un valor umbral para generar la bipartición. MLkNN fue seleccionado, por ser representativo de los algoritmos de alto rendimiento en problemas con métodos de adaptación basándose en los k vecinos más cercanos (kNN) y denominados *lazy learning algorithm* [Wettschereck *et al.*, 2007].

El conjunto de datos para el entrenamiento se tomó de los mismos OA recuperados de las 60 consultas sobre los dos repositorios empleados para el caso de estudio. Se presentan los resultados experimentales del conjunto de datos que contiene 253 ejemplos de OA reales, anotados con una o más de las 38 etiquetas en inglés que corresponden a los tipos de consultas identificados por los profesores y alumnos como necesarias para su búsqueda, tales como "Programming languages", "Hybrid Computing", etc. Cada OA es descrito con 1.442 características extraídas de LOM. El conjunto de datos en formato (número de etiquetas, número de características, orden de los atributos, etc.) deben ajustarse al formato de los conjuntos de entrenamiento sobre la base del modelo construido.

De acuerdo con [Santos *et al.*, 2010], la evaluación de los métodos de aprendizaje de datos multi-etiqueta requieren medidas diferentes a las utilizadas en el caso de los datos de etiqueta única. Hay varias medidas que se han propuesto para la evaluación de biparticiones y la clasificación con respecto a los datos multi-etiqueta. Aquí, estamos interesados en evaluar la calidad de biparticiones y el ranking. Para ello se realizan experimentos sobre

los datos relatados de los OA para comparar los dos algoritmos RAKEL y MLkNN. La Tabla 6-3 muestra los resultados de predicción de los dos algoritmos competidores de clasificación utilizando una variedad de medidas. En el caso de RAKEL todos los indicadores superan significativamente a los del algoritmo MLkNN, especialmente si se tiene en cuenta la exactitud del subconjunto de medida (*Subset Accuracy*), que es igual a la pérdida de cero-uno para la tarea de clasificación de etiqueta única en la predicción de la etiqueta exacta.

Tabla 6-3. Resultados de Rendimiento

Class	RAKEL	MLkNN
Hamming Loss	0,078	0,0864
Subset Accuracy	0,1862	0,0948
Subset Recall	0,2195	0,1567
Example-Based Accuracy	0,2131	0,1538
Micro-averaged Precision	0,8677	0,7550
Micro-averaged Recall	0,2190	0,0717
Micro-averaged F-Measure	0,3397	0,2630
One-Error	0,1094	0,0888
Average Precision	0,3550	0,5581
Is-Error	0,7785	0,6962
Error Set Size	7,5411	4,3164
Coverage	22,0614	10,4075
Ranking Loss	0,4187	0,1766
Micro-averaged AUC	0,6127	0,8603

En relación con el tiempo, RAKEL es un algoritmo muy rápido, mientras MLkNN tiene un mayor consumo. Los resultados experimentales indican que no sólo RAKEL es más eficiente en la fase de entrenamiento y la de pruebas que MLkNN, sino que también mejora la precisión de la predicción. Los resultados demuestran que el algoritmo RAKEL se puede utilizar para mejorar la clasificación de los OA en los tipos de consultas basadas en el contenido de los metadatos. La Figura 6-12 muestra la clasificación de RAKEL para los OAs atendiendo a sus palabras clave en el etiquetado de los OAs entrenando el sistema según la consulta realizada para recuperar dicho OA. La gráfica de

dicha Figura 6-12 representa el número de ejemplos anotados con cada etiqueta. Basándose en la facilidad de las predicciones se puede realizar una clasificación las etiquetas en el siguiente orden descendente: L2, L3, L5, L6, L7, L8, L9, L11,..., L38.

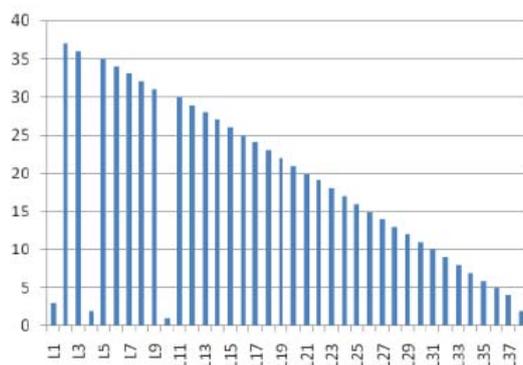


Figura 6-12. Número de ejemplos anotados con cada etiqueta

Mediante una instancia del algoritmo RAKEL se ha podido hacer tanto un ranking como una bipartición del conjunto de etiquetas de datos multi-etiquetados. De manera que es factible con este mecanismo realizar la tarea MLR. La Figura 6-13 muestra un ejemplo de ranking para la gramática empleada para etiquetado de OA. Faltaría establecer un criterio de umbral para condicionar las etiquetas almacenadas como relevantes para cada OA.

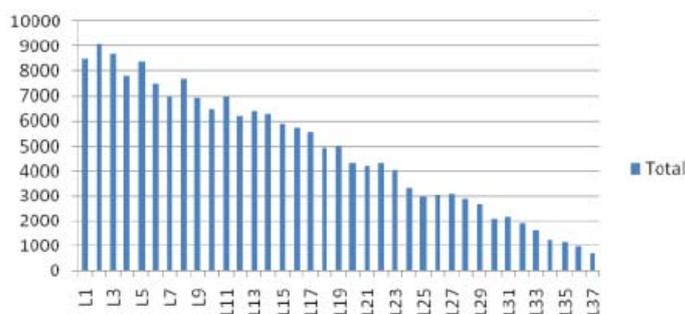


Figura 6-13. Ejemplo de la clasificación de Gramática LO.

Estos resultados son preliminares y en la actualidad es una línea en desarrollo. El objetivo será el apoyar la aplicación de los sistemas de recuperación de OA en los ROA de consulta a través de sus etiquetas. Esta capacidad de consulta

sería útil en la selección de OA a través de distintas aplicaciones de mejora del modelo propuesto.

Los clasificadores multi-etiqueta como RAKEL podrán ser utilizados para la anotación relevante automática de los OA. Esto a su vez, apoyaría la aplicación de técnicas más eficaces en la recuperación de información multi-etiqueta.

6.2 GENERALIZACIÓN DEL MODELO PROPUESTO

La innovación del modelo presentado, AIREH, es su arquitectura basada en un modelo organizativo de agentes. Esta propuesta permite lograr al usuario centralizar la búsqueda y ubicación de contenidos educativos relevantes, frente a ambientes altamente dinámicos, como en la Web profunda donde se localizan los repositorios de objetos de aprendizaje sobre la que se ha probado su utilidad. El proceso de búsqueda está integrado en la organización de los agentes de una manera totalmente transparente para el usuario, frente a los problemas de distribución y la integración de repositorios diferentes, la abstracción de la lógica interna de cada uno y la clasificación, el almacenamiento y la búsqueda de OAs. Teniendo en cuenta los resultados obtenidos a partir de la construcción de la aplicación AIREH, es posible concluir que la arquitectura basada en agentes es eficiente para resolver el problema de búsquedas federadas en repositorios heterogéneos. La solución propuesta permite la adaptación al entorno mediante el aprendizaje sobre los elementos participantes tales como usuarios, objetos de aprendizaje, repositorios, etc. El modelo de arquitectura presentado permite el ajuste del sistema en un entorno cambiante en función de la carga del sistema y de las condiciones de que se manifiestan en los repositorios en tiempo real en cuanto a rendimiento y disponibilidad de éstos. La arquitectura incluye varios aspectos relacionados con el aprendizaje en el contexto del objeto y las preferencias del usuario.

El análisis, estudio y definición de los mecanismos utilizados para dar solución a los problemas que involucran la búsqueda federada, la recuperación de recursos, y combinación de resultados pueden ser implementados en diferentes sistemas debido a que fueron creados en base a estándares de integración e interoperabilidad. El caso de estudio realizado, se centra en búsqueda federada sobre repositorios que implementan el protocolo SQL. De una manera sencilla puede hacerse sobre repositorios que basen su comunicación en cualquier tipo de protocolo. El diseño del Agente ROA del modelo de organización, ver Figura 5-17, permite la realización de la consulta a cada ROA individual y podría hacerlo por ejemplo para un repositorio OAI-PMH instanciado al agente Traductor para que sea traducida al lenguaje específico utilizado por el ROA. Como dicho protocolo permite la consulta y recuperación de metadatos en el estándar Dublin-Core, la integración de dichos metadatos para el proceso de

fusión la realizaría el agente Resultados. De manera que las características de la organización de agentes, permiten implementar de manera sencilla y escalable su generalización a cualquier ROA, atendiendo a su protocolo de comunicación, su lenguaje de consulta y la descripción de sus contenidos en cualquier estándar de metadatos.

El objetivo de este trabajo de *investigación* se centra en la recuperación de contenidos educativos, pero con el creciente auge de mecanismos de etiquetado para grandes repositorios de información, la solución planteada es fácilmente extrapolable a cualquier otro ámbito. La diferencia del campo de aplicación estaría en la semántica del contexto, afectando únicamente a los aspectos relacionados con el etiquetado de los recursos y la información relevante para la generación de la base de casos correspondiente. Un ejemplo de aplicación directa serían los BioBanks o Biorepositorios. Debido a que son colecciones digitales de metadatos que permiten el acceso a información sobre material biológico. Un ejemplo representativo de dichos BioBanks es el que soporta el Instituto Europeo de BioInformática con soluciones basadas en servicios web [Labarga *et al.*, 2007] sobre datos genómicos, son un campo de aplicación razonable de la arquitectura de recuperación propuesta.

Otro campo que se vislumbra como importante en cuanto a recuperación de información es el del documento electrónico en el ámbito de las administraciones públicas. En el caso del estado español, existe una norma técnica sobre documento electrónico perteneciente al ENI (Esquema Nacional de Interoperabilidad)³⁸ que ha sido regulado mediante el Real Decreto 4/2010, de 8 de enero. El Esquema Nacional de Interoperabilidad persigue la creación de las condiciones necesarias para garantizar el adecuado nivel de interoperabilidad técnica, semántica y organizativa de los sistemas y aplicaciones empleados por las Administraciones Públicas, que permita el ejercicio de derechos y el cumplimiento de deberes a través del acceso electrónico a los servicios públicos, a la vez que redunde en beneficio de la eficacia y la eficiencia. La norma referida en el RD 4/2010 regula el Esquema Nacional de Interoperabilidad en el ámbito de la Administración Electrónica y establece que las Administraciones Públicas han de adoptar las medidas necesarias para garantizar la interoperabilidad, recuperación y conservación de los documentos electrónicos a lo largo del tiempo. Resulta evidente que en dicha labor de conservación de documentos electrónicos las administraciones públicas generarán grandes volúmenes de documentos electrónicos que se almacenarán en repositorios. Esta situación configura un entorno donde el modelo propuesto, AIREH, podría tener un magnífico campo de desarrollo en la recuperación inteligente de documentos digitales.

³⁸ ENI <http://www.csi.map.es/csi/pg5e41.htm>

6.3 CONCLUSIONES

Un objetivo importante en el desarrollo de la tecnología de recuperación de contenidos en la Web profunda es mejorar la calidad, el alcance y precisión de los actuales motores de Web visible mediante la utilización de descripciones estructuradas de los recursos a través repositorios de metadatos semánticamente ricos. Esto sólo es posible haciendo que los metadatos de estos recursos sean accesibles. Por otra parte, debido a los derechos de autor restrictivos, el contenido digital real está libremente disponible sólo en número limitado de casos. El enfoque práctico propuesto en este trabajo ha consistido en recuperar los metadatos de una serie de repositorios de objetos de aprendizaje distribuidos geográficamente y centralizarlos. Una vez recuperados tratarlos adecuadamente y dejar que los usuarios realicen las búsquedas desde una única interfaz de usuario que centraliza su manejo, permitiendo implementar mejoras sustanciales en los procesos de ranking en los OA recuperados.

Cuando se construye una aplicación de búsqueda federada, es importante tener en cuenta el creciente número de repositorios disponibles, la necesidad de utilizar algoritmos avanzados, el alto grado intensivo en el cómputo para la recuperación de información, así como el creciente número de usuarios que necesitan perspectivas personalizadas sobre el contenido de dichos repositorios. Todos estos factores causan en la práctica problemas de escalabilidad y rendimiento.

Este capítulo describe cómo se han alcanzado los distintos objetivos definidos en esta investigación para validar y evaluar la hipótesis de partida: *“es posible modelar un sistema eficiente para la búsqueda, recuperación y gestión de contenido digital educativo en sistemas heterogéneos a partir de un modelo de organizaciones adaptativo que facilite una planificación distribuida y dinámica.”*

En este trabajo de investigación se ha presentado una arquitectura de recuperación de contenidos educativos basada en organizaciones de agentes. El modelo de organizaciones de agentes, permite dotar a una organización de capacidades auto-adaptativas en tiempo de ejecución para entornos altamente dinámicos. La principal novedad es que la arquitectura planteada basada en dicho modelo posee capacidad de planificación dinámica y adaptativa para llevar a cabo una distribución óptima de las tareas de los agentes miembros de la organización. Esta propuesta permite la recuperación inteligente de contenidos y su flexibilidad en el entorno altamente dinámico para el que se ha planteado. En resumen, la arquitectura presentada define las acciones que una organización de agentes debe llevar a cabo, considerando cambios producidos durante la ejecución de una consulta determinada mediante una planificación

adaptativa dentro de una organización de agentes y atendiendo a características del contexto (usuarios, perfiles, características de los contenidos, variabilidad de los ROA, etc.). Consideramos que los requerimientos más importantes establecidos durante el modelado han sido validados mediante el desarrollo de la arquitectura propuesta operativa en entornos reales. Para concretar, los siguientes objetivos han sido cubiertos a lo largo del trabajo:

- ❖ Se ha realizado un análisis de los problemas concretos de la recuperación de contenidos educativos de calidad para entornos heterogéneos detectando las carencias existentes que han establecido el conjunto de requisitos de partida en la elaboración de la solución.
- ❖ Se ha realizado un estudio de las tecnologías de almacenamiento y recuperación en entornos heterogéneos que ha permitido modelar la arquitectura propuesta atendiendo a múltiples perspectivas. Estos requisitos han formalizado la arquitectura dotándola de la flexibilidad y características necesarias para su adaptación a dichos entornos heterogéneos.
- ❖ Se ha hecho uso de mecanismos formales útiles para la resolución de la búsqueda automática de información en entornos heterogéneos y dinámicos en tiempo de ejecución. En este sentido, en entornos abiertos, cambiantes y heterogéneos, los SMA han sido dotados de capacidades organizacionales que permiten la generación de recomendaciones de forma autónoma en el contexto.
- ❖ En base a todos estos aspectos, se ha propuesto una arquitectura multi-agente para la recuperación y gestión de contenidos digitales para bases de datos heterogéneas sobre la base de organizaciones virtuales. Dicha arquitectura denominada AIREH ha demostrado ser flexible y adaptable a las necesidades del entorno de aplicación así como a los usuarios.
- ❖ Finalmente se ha hecho una implementación de la arquitectura propuesta con la teoría de agentes organizacionales como base. Dicha implementación se ha probado para tareas de búsqueda, filtrado y catalogación de objetos de aprendizaje. El prototipo del sistema se ha probado para un escenario de recuperación de contenido educativo específico, los repositorios de objetos de aprendizaje distribuidos, mediante una organización de agentes adaptada al problema definido.
- ❖ Se han evaluado empíricamente los resultados obtenidos en entornos reales de aplicación.

La arquitectura propuesta AIREH basada en un modelo de organizaciones de agentes proporciona al sistema la capacidad de obtener información del contexto (requisitos de usuario, rendimiento de repositorios en tiempo real, etc.) y una comunicación distribuida entre sus componentes que permite realizar una asignación dinámica de tareas entre los agentes. La arquitectura propuesta así modelada facilita las tareas de recuperación de contenidos en

repositorios heterogéneos. Permite desarrollar funcionalidades en recuperación de contenidos independientes de los lenguajes de consulta sobre los repositorios consultados, lo que resulta un logro significativo de la arquitectura propuesta, ya que posibilita desarrollos flexibles y adaptables a los cambios del entorno. De igual modo la arquitectura es flexible a la interoperabilidad de contenidos descritos en cualquier lenguaje de etiquetados (LOM, DC, etc.).

La implementación de las funcionalidades a través de la arquitectura propuesta AIREH, continúa en fase de desarrollo. Se trata de un modelo integrador que posibilita el desarrollo de múltiples funcionalidades. Sin embargo, los resultados obtenidos hasta el momento demuestran que el modelo de arquitectura AIREH basados en SMA flexibles y abiertos sustentado en organizaciones posibilita el desarrollo de recuperación de contenidos en entornos heterogéneos con funcionalidades complejas. Específicamente, el modelo propuesto proporciona a la arquitectura los servicios necesarios para llevar a cabo una planificación de tareas de manera dinámica. Cabe destacar que, tanto el modelo como las tecnologías presentadas en esta investigación, son tan solo un ejemplo del potencial para el desarrollo de sistemas de recuperación de contenidos en entornos heterogéneos debido a que el modelo se basa en el paradigma de las organizaciones virtuales.

El objetivo principal de este trabajo *conseguir modelar una arquitectura multi-agente basada en organizaciones como mecanismo eficaz para la gestión en la búsqueda, recuperación y catalogación de recursos digitales en entornos heterogéneos* se ha conseguido.

A todo esto se suma que la hipótesis de partida del proyecto OVAMAH³⁹ consiste en dotar de capacidades de autonomía a una organización virtual que permitan una respuesta dinámica frente a posibles situaciones cambiantes por medio de la adaptación y/o evolución de la propia organización, de tal manera que ésta sea capaz de detectar situaciones de interés, como por ejemplo fallos de funcionamiento, y pueda manejarlos maximizando la flexibilidad y la capacidad de adaptación. Por lo tanto, el trabajo presentado en esta memoria es una contribución importante en los objetivos a corto plazo para este proyecto, ya que *se ha modelado un sistema de recuperación de contenidos educativos capaz de adaptarse a los objetivos comunes de los agentes deliberativos que participen en una organización de agentes*.

³⁹ Proyecto OVAMAH: (Organizaciones Virtuales Adaptativas: Mecanismos, Arquitecturas y Herramientas) con referencia TIN 2009-13839-C03-03, otorgado por el Ministerio de Ciencia e Innovación. Proyectos de Investigación Fundamental No Orientada.

La investigación realizada en este trabajo ha contribuido a establecer las bases para futuros proyectos, no solo dentro del grupo de investigación BISITE, sino también en colaboración con otros grupos de investigación, principalmente GAIA (*Group for Artificial Intelligence Applications*) y <e-UCM> (*The e-Learning Research Group*) en la UCM .

En un intento de resumir las principales contribuciones a la investigación el siguiente punto hace una breve relación de la aportación de este trabajo al ámbito de recuperación de los contenidos en repositorios distribuidos.

6.3.1 Contribuciones a la investigación

El trabajo de investigación presentado en esta memoria aporta algunas nuevas contribuciones en los ámbitos de la Recuperación de Contenidos en Entornos Heterogéneos y sobre aplicaciones reales de la Teoría de Agentes y de las Organizaciones Virtuales.

- Desde el punto de vista de la ingeniería, se ha utilizado una metodología orientada al desarrollo de organizaciones que ha permitido cubrir el análisis, diseño de la estructura organizativa y de diseño de la dinámica de la organización de forma eficiente para la recuperación de contenidos en entornos heterogéneos.
- Desde el punto de vista del desarrollo, se ha podido comprobar en un entorno real la validez de este tipo de sistemas, el cual consiste en el desarrollo de una organización de agentes enfocada a centralizar los servicios de búsqueda, recuperación y catalogación de recursos educativos en entornos heterogéneos atendiendo a perfiles de usuario.
- Se han analizado y estudiado los resultados obtenidos para poder demostrar que dicha arquitectura es una solución viable.
- Se ha realizado un gran esfuerzo por obtener una retroalimentación de distintos investigadores y grupos de investigación en áreas relacionadas. El objetivo es robustecer esta investigación mediante el intercambio mutuo de ideas y conocimiento. Se ha puesto especial interés en difundir nuestras experiencias y los avances de esta investigación, desde sus etapas iniciales hasta su forma final, a través de diversas publicaciones y la asistencia a conferencias, congresos, *workshops*, citados en el Anexo D.

Como conclusión, decir que se han conseguido los objetivos con los que se inició el trabajo:

- (i) Se ha comprobado la adecuación de los modelos de organizaciones de agentes a la búsqueda y recuperación de recursos educativos;
- (ii) Se ha verificado, mediante datos empíricos su validez y que además eleva el rendimiento de manera ostensible; y por último

- (iii) Se ha propuesto un modelo recomendación híbrido basado en el filtrado colaborativo y en preferencias de usuario adaptado al contexto de la catalogación y selección de objetos educativos.

En este proyecto de investigación se prueba que es posible modelar un sistema eficiente para la gestión de sistemas abiertos a partir de un modelo de organizaciones adaptativo que facilite una planificación distribuida y dinámica. Específicamente, el modelo propuesto proporciona a la arquitectura a los servicios necesarios para llevar a cabo una planificación de tareas de manera dinámica. Cabe destacar que, tanto el modelo como las tecnologías presentadas en esta investigación, son un ejemplo del potencial para desarrollar sistemas de recuperación de contenidos basados en el paradigma de las organizaciones virtuales. Las ventajas de la arquitectura propuesta son la flexibilidad, la personalización, la uniformidad y la eficiencia.

La investigación presentada en esta memoria es relevante sobre todo desde el punto de vista científico, sentando las bases para el desarrollo de importantes proyectos; en especial el proyecto OVAMAH, de ámbito nacional.

6.3.2 Trabajo Futuro

La investigación presentada en esta memoria valida el modelo planteado para la recuperación inteligente de contenidos en repositorios distribuidos mediante el uso de una arquitectura multi-agente basada en organizaciones. Este resultado no es sino un punto de partida en un momento en el que es previsible una clara expansión de los almacenes de contenidos digitales y que deja abiertas gran cantidad de vías para un trabajo futuro. En este sentido la principal línea de trabajo futuro será la inclusión en la organización virtual de nuevos mecanismos de filtrado y su posterior validación mediante pruebas empíricas tal y como se ha realizado con el sistema de búsqueda federada. Junto con esta línea de trabajo principal, también se contempla el diseño y posterior desarrollo de algoritmos que permitan mejorar la búsqueda y recuperación de objetos de aprendizaje relevantes para el usuario. Es necesario incrementar el número de características que afecten al nivel de relevancia de los objetos de aprendizaje y amplíen las capacidades de evaluación de aspectos colaborativos en la recomendación por parte de los usuarios. Finalmente, como última línea de trabajo, se pretende una mejora en la interfaz, acorde a las necesidades de un sistema de este tipo, es decir, teniendo en cuenta además del caso de estudio, la organización virtual de agentes que se encuentra en niveles inferiores del sistema.

Aspectos como la inclusión de criterios que solucionen la internacionalización en la búsqueda de contenidos educativos deberán de ser tenidos en cuenta, solucionando los problemas para su recuperación multilingüe y la fusión de

dichos resultado atendiendo a criterios que incluyan elementos relacionados con el idioma, etc.

A partir de los resultados obtenidos en este trabajo posibles líneas de trabajo futuro pueden concretarse en los siguientes aspectos:

- ❖ Pruebas y validación.
Es necesario realizar pruebas mucho más exhaustivas con el objetivo de evaluar en detalle la arquitectura propuesta en términos de tiempo, aplicación de análisis y diseño, tiempo de cálculo de respuestas, calidad de los objetos de aprendizaje, etc. Los resultados obtenidos permitirán desarrollar modelos y sistemas más depurados y robustos.
- ❖ Resolución de nuevos problemas prácticos.
Para realizar una comprobación más exhaustiva de la validez del modelo propuesto, es necesario aplicarlo a nuevos problemas prácticos. De esta forma, sería posible comprobar si se adapta de forma adecuada a la resolución de problemas de características distintas, o bien si se ha construido un modelo muy restringido a los problemas concretos que se han estudiado durante la realización de este trabajo.
- ❖ Implementación de clasificadores multi-etiqueta.
En la sección 6.1.3 se relata un caso de estudio que muestra la posibilidad de incorporar técnicas de minería para clasificación multi-etiqueta en los objetos de aprendizaje. Incorporar estas técnicas en la extracción de perfiles de usuario o en la generación de clusters de objetos de aprendizaje puede aportar mejoras sustanciales en los procesos de recomendación de los contenidos.
- ❖ Incorporación de aspectos semánticos para la recuperación y catalogación de contenidos
Aún cuando los recursos educativos estén etiquetados de acuerdo a un estándar de metadatos, estos son principalmente descriptivos y no proveen información semántica directa asociada. Será necesario enriquecer semánticamente los procesos de búsqueda y tratamiento de los objetos de aprendizaje en base a modelos de conocimiento tales como ontologías de dominio. Esto permitirá incrementar las funcionalidades de la propuesta con funcionalidades basadas en semántica tanto en las consultas como para priorizar los resultados de dicha búsqueda atendiendo a su significado, estableciendo métricas entre los elementos del proceso atendiendo a métricas de distancias semánticas.

7 RESEARCH OVERVIEW

This chapter is a summary of the research presented in this work. The chapter is structured as follows: Section 8.1 is a brief introduction of the study. Section 8.2 describes related works for current problem. Section 8.3 presents the architecture for the proposed model. Section 8.4 demonstrates how the model can be used in a specific case study; to this end, the context and the metrics to validate the model will be explained, and an example within the developed system will be provided. Section 8.5 analyzes some experimental results obtained in a real environment case study, while Section 8.6 proposes a model generalization for the proposal. Finally, Section 8.5 will emphasize the conclusions obtained and Section 8.8 includes some contributions made by the present study. The chapter ends with an explanation in Section 8.9 of some future work.

7.1 INTRODUCTION

One of the most widely accepted approaches in the study of distance education today is based on the paradigm of the learning object (LO). This paradigm is based on the fragmented and self-contained content that enables modular units to be reused in different educational contexts and different platforms. These learning objects are formed both by the content itself, for education, and by labels that use a series of metadata to describe them completely. The aim of the systematic use of such labeling of learning objects is to automate the processes of searching, cataloging and retrieving these educational resources.

Usually, learning objects are stored in digital repositories. Most of these repositories have a traditional web interface on an HTML document search that accesses its educational content through queries to the database that hosts them. The vast majority of these repositories require user access by password which allows them to manage queries and develop user histories, profiles, etc. The user interacts with the information contained in the repository via their web browser in the usual mode of interaction: requests are made through the interface, and results are returned in the repository for the contents related to the query. There are many digital content repositories that store such content through the learning object format. But through the traditional web interfaces, it is not possible to manage all information directly labeled by these LO, nor to even know if that information is labeled or not according to any standard

because the user does not know how that educational content is in the repository.

There is a large volume of educational content on the Web, which is not directly accessible through conventional search engines. This information is said to belong to the so-called hidden, deep, or invisible Web, as opposed to the contents of the website which are referred to as visible. The solutions developed by conventional search engines are very efficient for retrieving the visible Web contents. These engines make use of a centralized recovery model that copies the information in a single centralized database, indexes and sorts the contents by classified documents in the database for user queries. This method works well when the information sources have left their content exposed to web crawlers. However, this does not apply in the deep Web where information can only be accessed via search mechanisms adapted to specific sources, such as the educational content repositories, which are the subject of this study.

As an example, Figura 5-1 shows the results of the search "Pulmonary Disorders" through the web⁴⁰ interface, available in the MERLOT (Multimedia Educational Resource for Learning and Online Teaching) repository.

Repositories are often highly heterogeneous, with different storage systems, access to objects with their own methods of consultation, etc. The problem of heterogeneity in database systems is an open issue that arises with each database and while it has been solved in many respects, it still remains an open problem with a large number of initiatives to standardize the processes and technologies that comprise it.

The current environment presents an increasing variety of distributed repositories of educational content. Chapter 2 of this report includes a study of various initiatives to ensure the interoperability of content. Many of these repositories do not have a system that allows a higher level of abstraction between the internal and end users of the stored data. Others are forming networks using architectures to facilitate interoperability based on various standards that assign an abstraction layer that connects the internal characteristics with the exterior, allowing for greater automation and computerization of the LO-containing process, as reviewed in section 2.4. Following this section is an example of the MERLOT repository, which implements an architecture that enables the interoperability of its contents by providing offline consultation mechanisms through federated searches using a web customer service SQI [SQI, 2005] (via WSDL specification [WSDL, 2001]) or the Restful Web services [Fielding, 2000] [Fielding & Taylor, 2000]. Through these services it is possible to access the information labeled for learning

⁴⁰ MERLOT <http://www.merlot.org/>

objects, displayed in LOM as in Figura 5-2 for the search result "Pulmonary Disorders".

In the current educational context there is a significant growth of learning object repositories as part of the hidden web in large databases. These learning objects can be labeled in any standard (LOM, Dublin Core, etc.). It is necessary to work towards and develop solutions that enable the efficient search of heterogeneous content with regards to the distributed context where they lodge. Distributed information retrieval or federated searches, [Callan, 2000] attempt to respond to the problem of information retrieval in the hidden Web. The main aim of federated search is to develop models and strategies to get the most benefit from these distributed sources. Thus, users perceive the system as a single point of access to information they require, regardless of the number of sources that exist, their location or management mechanism. The process is completely transparent to the user, who does not perceive the complexity involved, and gives a uniform treatment to the information retrieved.

The main contribution of federated search is that the search process is done through search mechanisms in individual information sources. In addition, the search refers to the location of each source and provides a distributed control of information related to the different sources of hidden information.

The federated search mechanism is thus a much more complex, rich and comprehensive centralized recovery model than the traditional search engines that do not provide real solutions to the problem of the hidden web. Federated search can be applied to the recovery of LO repositories such as the sequence described in the resolution of the following three subproblems:

1. SELECTION OF REPOSITORIES, ie, to analyze the description of resources in the repositories and to study how to represent information that is distributed in such repositories. It is very important to know the subject, area, and the context that the information stored in repositories belongs to, as well as many other key features of the various sources of hidden information (response times, efficiency, etc.).
2. SELECTION OF RESOURCES, given a need for information and a corresponding set of descriptions, a recovery algorithm is used to determine the results most likely to satisfy the query.
3. MERGER OF RESULTS is the integration and combination of results returned by queries on the n repositories, forming a single list that gives the user a ranked list of results. The sort order of the results will produce a ranking algorithm, or priority ranking in the presentation of the contents recovered.

7.1.1 *Analysis of the problem context*

In order to know the real situation of distance education systems, it was necessary to conduct a preliminary study of the repositories that implement layers that allow an offline search or federated search through SQI protocol, as shown in Anexo A. The aim of this study is to clarify the functionality and identify gaps in public repositories such as protocol operation and hosting real learning objects. The study created the need for a centralized solution for the intelligent pursuit of educational content to remedy the problems of heterogeneity for each of the parties involved. This study allowed us to develop guidelines that guide and support the development of federated search architecture proposed in this research.

The birth of what can already be considered as the LO paradigm has produced many advantages when it comes to reusing learning content. In addition to the content specification in LOs, the search and recovery of educational resources is evolving as a promising scenario. But the process of innovation has also produced different challenges that have not yet been solved. Overcoming the problems impeding the commitment to interoperability of educational content can be grouped around two core ideas. The first is related to the problems associated with the monolithic structure of learning object repositories, which does not allow external management with the flexibility and power necessary to ensure easy interoperability, and dispersed and heterogeneous sources. The LORs are unstable and have a high number of errors. The presentation of information through their web interface is really poor. The object repositories can not be added dynamically; it is only possible to query the registry repository. Additionally their search interface does not display the information on the learning object and generally only occurs as a result of access to the content. Nor do they allow comprehensive user management according to user interests in line with other users in the same context. Many other aspects can be considered a lack of perspective of universal accessibility to Web resources; this can be seen in projects such as Web-4-All⁴¹, as seekers of the repositories do not meet their interfaces to access even a standard minimum level of WAI⁴² (WAI-A).

The second group of problems is directly related to the absence of automatic mechanisms that control the technical quality, semantics and syntax of Learning Objects, ensuring the correct specification of such learning objects in any of the metadata schemas that describe them.

The main problem with those repositories is that they still do not implement any abstraction layer that can encapsulate the internal logic of the repository. This type of repository, known as autonomous, is monolithic and has great difficulty in automating the search and retrieval process of resources.

⁴¹WEB-4-ALL <http://web4all.ca/>

⁴² WAI <http://www.w3.org/WAI/>

Consequently, consultation and LO extraction is a slow process which requires the manual intervention of users who must reuse the learning resources. But the problems are not limited to those associated with not using any middleware layer of abstraction. The systems in which this layer is included, as seen in the study presented in Anexo A, also suffer from various problems such as lack of reliability or availability, high access times in some cases, erroneous results, poor results, etc.

The series of problems in the repositories require solutions that are adapted to the heterogeneity of each of these isolated repositories and can ensure real and effective interoperability of educational content globally. The solution will enable a centralized global search and the effective reuse of resources by the end user. This requires raising the level of abstraction and looking at the classification of systems storing and searching for LOs. As we have seen, there are repositories in which each query triggers new federated searches in other repositories. Although this can theoretically be seen as an advantage because it increases the number of results, in practice it has two drawbacks. The first relates to the response time, which increases considerably, and the second is a derivative of the repeated occurrence of LOs in the results, because repeated queries in the same repository are done indirectly through the federations.

Beyond the problems with the LOR, there are some problems with the specification of LOs. IEEE LOM defines the educational resource perfectly, including a variety of details and general aspects, the identification, educational purpose, use and rights of sale, and any technical or other relationship with LOs. But there is no obligation to specify the minimum attributes that a LO must have in order to be used within a particular educational context. It is derived from two fundamental problems:

- The first relates to the unique identification of each learning object. An object is well defined a priori through a catalog ID (repository) and an entry into this repository. But since it is not mandatory to complete these fields, the unique identification of each LO becomes a complex process. The same situation occurs in the case of the double identification of LO.
- The second relates to the absence of the element defined by the URI or URL that stores the learning resource described by the metadata file. Therefore, if a LO is perfectly described, that content can not be accessed and used by both either when there is no link to the resource or when it is wrong.

All the problems that have been observed in both the LORs and the LO are a real challenge for applications seeking to use these platforms as a source for educational content. The development of such applications should work with designs that enable use with the inclusion of practical solutions in

environments that are highly heterogeneous, dynamic, and contain a high number of errors in the repositories. This context makes it necessary to design and build systems with high fault tolerance.

It is also important to consider a further challenge in the efficient management of services and the elements involved in these platforms by including effective management techniques (e.g., filtering, classification and recommendation) that use labeled gaps in educational content to facilitate storage, search, recovery, etc. There are some proposals to allow users to find learning objects best suited to their needs. Section 7.2 lists some of these works, some of which focus their efforts on creating architectures that allow for a more or less centralized recovery; other efforts are abstracted from the problem of recovering and concentrating their efforts on the LO position. All provide an appropriate method of supplying the user with the most relevant content. Thus, a comprehensive solution to the problem of educational content recovering goes beyond simple recovery; it is necessary to have documentation with mechanisms of the objects retrieved that can be evaluated and generate the most suitable position according to the user.

7.1.2 *Over the Solution*

With the situation that arises from the context of education, it is clear that a new type of application is needed; one that allows the search for educational content in a distributed environment across different formats, servers and networks. This paper proposes a solution: the federated search architecture for educational content, particularly LOs.

Conflicting needs must be met in order to create the federation and thus integrate several components: the repositories must be distributed across physically dispersed locations; access to these repositories is read-only; information systems are operating under different platforms; the types used are distributed repositories (metadata stored in one place and LO in the other) and they are managed by autonomous systems; in some cases access to the repositories requires the establishment of a session, in others direct access, etc.

In this sense, the use of a system architecture based on the use of intelligent agents is ideal, since they can apply their characteristics (autonomy, status, reactivity, rationality, intelligence, coordination, mobility and learning) to develop a stable system with the ability to react intelligently to the needs of the environment. The idea to model the architecture as a virtual organization, on the assumption that an organization can adapt its actions to achieve its goals and interact with heterogeneous components, is proposed as a theoretically efficient solution.

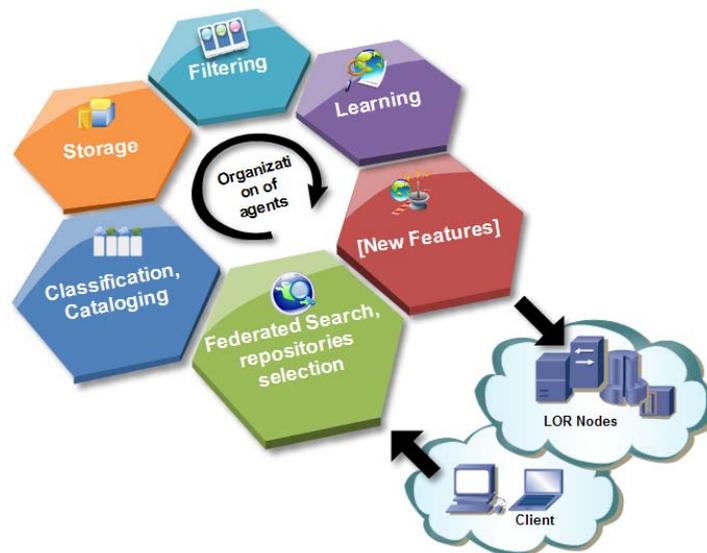


Figure 7-1. General scheme of distribution of tasks in architecture

From a heterogeneous and changing technological situation that accompanies the proposed educational context, and in order to reuse educational content in a real operational context, there is motive to design a model for an integrated architecture in which an organization of agents can execute actions for the search and retrieval of content based on a federated search model. The innovation of this architecture will be to provide an organization of agents with the self-adaptive capabilities required to address the current problems in highly dynamic environments such as those discussed, and to adapt to future changes. The features of the proposed architecture are presented in **Figure 7-1**.

This will solve the problems of distribution, the integration of different repositories, the abstraction of the internal logic of each repository, and the classification, storage and retrieval of LOs, in a completely transparent way. In addition to adding capacities, such as simple scalability, to possible situations involving the use of new protocols, it also adds internal logical repositories, cataloging or heterogeneous applications designed to cover services related features.

The cornerstone of this thesis is the recovery of learning objects in a real environment using federated search in different Learning Objects repositories. It is necessary to provide the user with a framework that unifies the search and retrieval of objects, thus facilitating the learning process that filters and properly classifies learning objects retrieved according to some rules. The

generation of such rules for the organization of the items recovered will be based on educational metadata and will provide useful content to end users. Mechanisms will be provided to document the objects recovered, which can be evaluated, and to generate the most suitable position according to the user. The architecture will provide multiple perspectives to assess the recovery of educational content.

When an architecture is defined, it must be applied to real problems in order to show its capabilities and possibilities for improvement. The following sections will show the results obtained using this architecture on real resources, with the subsequent study and conclusions. They also provide detailed descriptions, beginning with other related works that have provided different solutions in response to the same problems as those presented in this thesis.

7.2 PREVIOUS WORK

The following are related works, some focused on creating architectures that allow for the recovery of learning objects in a more or less centralized way, and others focused on obtaining a set of learning objects and then positioning them to provide the most relevance to the user.

7.2.1 Approaches to Automatic Recovery Educational Content

One of the problems that must be overcome in order to standardize the use of learning objects is the difficulty in actually finding them [Bates, 2005].

BILDU [Casquero *et al.*, 2008] proposes an architecture based on a set of horizontally integrated vertical search engines under one interface, which supports a network incorporating the features of social recommendation and mechanisms of association. Edutella [Qu & Nejd, 2004] proposes encapsulating other content integration and searching the collections through peer-to-peer mechanisms. LESSON [Zhou *et al.*, 2010] proposes the use of a metasearch engine and system P2P sharing, and [Santiago & Raabe, 2010] proposes a P2P network between institutions, with the possibility of searching and retrieving learning objects through an API within the content managers.

Some works [Karger *et al.*, 2006], [Almaraz, 2008], [Santiago & Raabe, 2010] employ an architecture based on mediators to integrate LO repositories. Other proposals for the recovery of educational content are based on methods adapted from the area of information retrieval [Yen *et al.*, 2010] with proposed similarity metric calculations on LO. [Mayorga *et al.*, 2010] include clustering techniques (Formal Concept Analysis. FCA) with pedagogical aims to search the web. Other approaches use recovery architectures based web services [Li *et al.*, 2009], [Dietze *et al.*, 2007] [Hanafy & Fakhry, 2008].

Proposals such as [Margain *et al.*, 2010] [Soto *et al.*, 2007] offer the solution of an architecture based on educational content retrieval using agent technology.

The successful recovery of educational content can provide in large part the characteristics, restrictions or conditions on the proper use of learning resources. The expressive power of queries to select LO repositories or queries according to user are more accurate and better present the results that meet user expectations. Various works [Lee *et al.*, 2008], [Segura *et al.*, 2010], [Segura *et al.*, 2011] [Karger *et al.*, 2006] show how a little power of expression in the recovery of LOs in their LORs can provide suggestions for improvement by incorporating many semantic aspects in the expansion of consultative mechanisms. [Shih *et al.*, 2006] has developed a metadata authoring tool to support the search and reuse of LOs. [Lee *et al.*, 2008] proposes an architecture with an ontological approach for learning object retrieval by expanding the query algorithm with semantic aggregation, which is called Ontology Query Expansion Algorithm. Lee (2008) defines a context ontology and infers user's intentions through short questions that expand the query in the LOR. But the simulation of this proposal, the context of the Java programming language, and performance tests on html documents on the visible Web, do not make use of the metadata characteristics of the real LO, or of the LOR search functionalities from the initial storage. The second study [Segura *et al.*, 2010] makes new contributions to the mechanism of semantic expansion of queries, but it remains a proposal that does not contrast real LOR or their metadata tagged files. Query expansion in [Segura *et al.*, 2011] defines a dynamic mechanism for searching and filtering the results that were validated in the MERLOT repository to search for LO related to genetics using Gene Ontology. The mechanism used is connected to MERLOT Restful service to recover the contents.

CYCLADES [Avancini. & Straccia, 2005] proposes an environment where users search, access and evaluate digital resources available in the repositories that are accessed through the Open Archives Initiative (OAI). This system makes it possible to offer recommendations on resources that are stored in different repositories and through an open system. The recommendations offered by CYCLADES were evaluated through a pilot study with 60 users, which focused on the test performance (prediction accuracy) of various collaborative filtering algorithms.

The proposed solutions are many and varied, most do not get results on real environments and the majority do not provide educational content in the particular case of learning objects. Search results have been improved by proposals seeking a solution based on technologies with high implementation and the inclusion of some social aspect. Only the proposals of [Margain *et al.*, 2010] and [Soto *et al.*, 2007] make use of agent technology for development, but neither of the two approaches is done through the coordination of member agent tasks in an organization. All of the solutions studied share a common denominator, which is that they promote interoperability and the reuse of

resources including the development aspects related to adaptation, or they use their own content or environmental mechanisms to allow for better access. This thesis presents the solution to the problem of the retrieval of digital content through an organization of agents, which allows those agents to work in a coordinated manner to solve a common problem, allowing the agents to take into account the planning tasks that they will adapt to the constantly changing environment (users, content repositories, etc.).

7.2.2 Approaches to Learning Object Recommendation

With so many Learning Object Repositories, a major challenge is to efficiently find the most suitable LOs for the users. This objective has attracted much research in the field of the selection and recommendation of LO.

Researchers and developers of e-learning have begun to apply information retrieval techniques with technologies for recommendation, especially collaborative filtering [Bobadilla *et al.*, 2009], or web mining [Khribi *et al.*, 2009], for recommending educational content. A recent review of these applications can be seen in [Manouselis *et al.*, 2010b]. The features that handle these information filtering techniques in this context are the attribute information of education article (content-based approach) and the user context (collaborative approach).

One of the first works in this context was developed by Altered Vista: a system in which instructional techniques are evaluated based on collaborative filtering recommendation algorithms with close neighbors [Recker *et al.*, 2003] [Recker & Wiley, 2001]. These works explore how to collect user reviews of learning resources and propagate them in the form of word-of-mouth recommendations.

RACOFI (Rule-Appling Collaborative Filtering) proposes a collaborative filtering by rules, with an architecture for the custom selection of educational content [Lemire *et al.*, 2005]. The author's recommendation is to combine both approaches to reduce recommendation by integrating a collaborative filtering algorithm that works with user ratings of a set of rules of inference, which creates an association between the content and rate of recommendation.

[McCalla, 2004] have proposed an improvement to collaborative filtering called the ecological approach to designing e-learning systems. Key aspects of this proposal take into account the gradual accumulation of information and focus on end users.

Manouselis *et al.* [Manouselis & Costopoulou, 2007] [Manouselis *et al.*, 2007] have conducted a case study with data collected from the CELEBRATE portal users to determine an appropriate collaborative filtering algorithm.

These recommendation systems for collaborative learning objects fail to address the key problem of cold start, as described in 3.3.2. Some recent works try to solve the problem. For example [Tang & McCalla, 2004] [Tang & McCalla, 2003] incorporate a collaborative filtering system for recommending research papers. The system incorporates elements of the pedagogical features, making use of what is called artificial pupils to resolve the problem of cold start. The system domain, however, is limited to research work, and the factors that influence the selection of the papers are different from those involved in the selection of learning objects selection.

Some solutions take a hybrid approach. [Aiju & Baoying, 2008] [Ghauth *et al.*, 2010] [Tsai *et al.*, 2006] and [Wang *et al.*, 2007] make use of algorithms based on reviews from other users according to interests which are extracted through nearest neighbor algorithms. These correlation-based algorithms are used to calculate an index score on the usefulness of learning objects through the analysis of comments from students with similar profiles. These algorithms improve preference-based selection algorithms by incorporating aspects of student preferences. The preference pattern of each student is recorded in a history of preferences that is generated and updated according to comments from the student's preference. If a selected learning object has been given a positive score, its preference score increases for all the features of the learning object. The combination of the scores for a learning object is determined by the two algorithms that decide the position of the learning object according to the outcome of the recommendation. This algorithm of choice helps avoid the problem of cold start, when there are no data on the use and value of a particular learning object. However, all of the selected learning objects are treated equally without any distinction between them, which would allow more precise assessment criteria of the user, affecting the very pattern of preference of the user.

The use of algorithms based on biological models, such as ACO (Ant Colony Optimization) is the basis of [Yang & Wu, 2009], which proposes a set of attributes based on a colony of ants (attributes- based ant colony system, AACs) to help students find their way through an adaptive model of learning objects more efficiently. This mechanism is based on the use of learning activities and educational elements to predict the optimal trajectories associated with the ACO algorithm, and recommend the sequence of learning objects. This work is interesting, but bases its recommendation on a path of learning through a different set of learning objects. The ultimate goal is to attain certain knowledge. The recommendation is the sequence produced by the optimal route between the different LOs.

The works by [Kerkiri *et al.*, 2007] [Ochoa & Duval, 2006], and [Wolpers *et al.*, 2007] suggest the need for selecting learning objects by taking into account the educational content described by their metadata, which falls in line with this

thesis. They propose a mechanism called Contextualized Attention Metadata (CAM) to capture information about the actions along the life cycle of learning objects, including their creation, labeling, supply, selection, use and maintenance. These studies proposed four metrics to LOM and CAM for classifying and recommending the learning objects retrieved: Link Analysis Ranking, Similarity Recommendation, Personalized Ranking and Contextual Recommendation. These metrics classify learning objects according to criteria such as popularity ranking, the similarity of objects based on the number of downloads, and more. How these rankings contribute to the selection of learning objects and how they combine with each other are still open questions and a highly interesting field of study.

Based on semantic aspects by considering contextual information from the student's cognitive activities and the LO content structure, Qiyan *et al.* [Qiyan, 2010] propose a framework for recommending learning objects to suit the student's cognitive activities through an approach based on ontologies. The same approach follows the work of Ruiz-Iniesta (2010), with a framework that simplifies the development of recommendations for LO.

There are other approaches, mechanisms or criteria for the categorization of educational content that require direct human intervention in their assessment, but list some criteria for assessing the quality of the content. Among these is the assessment contained in the MERLOT repository or LORI tool. The MERLOT repository (Multimedia Educational Resource for Learning and Online Teaching) offers the best current example of widespread application in the evaluation of educational content for Web-based education [Vargo *et al.*, 2003]. Content ratings are obtained through comments and ratings on a five point scale by users and reviewers appointed by MERLOT. This evaluation is based on three general properties: quality of content, potential effectiveness as a tool for teaching and learning, and ease of use. The classification is based on the quality of search results using a weighted average of these three classifications. The peer review process in MERLOT is carried out by two experts working asynchronously who return the descriptions of the contents recovered in a list sorted by the ranking, which has been established in turn by evaluating the quality of the content in descending order of assessment, where the contents are not evaluated at the end. LORI (Learning Object Review Instrument) is a tool known to assess the quality of education resources on-line. It is no more than an assessment protocol for learning objects in nine areas on a scale of five bridgesfive point scale that can be implemented in line by using rubrics, rating scales and comment fields. As an assessment tool, it is available at its website, which can be used to assess an individual or a panel of experts from a range of LOs, based on the advice of [Vargo *et al.*, 2003] [Nesbit *et al.*, 2002].

There are a growing number of papers proposing systems to recommend learning resources, evidenced by the lack of operational solutions as confirmed by recent work [Manouselis *et al.*, 2010]. All the evaluated proposals concluded

that the incorporation of mechanisms to assess attributes related to the educational content (Learning Objects in our work) as well as aspects of user context and their interaction with the content, create effective recommendation mechanisms.

However, a closer look at the revised proposals underscores the lack of applications on real systems and educational content. Most of the jobs listed in this section are based on simulations or have been applied to a local case study or a particular repository, with a priori control, for small groups of parameters that are usually local. The solution proposes that not all display results are from real context. Some of the recommended educational content objects are not learning objects, as defined in the present study, and the great majority does not therefore address aspects of semantic tagging of resources in their approach.

The architecture proposed in this work provides multiple perspectives to assess the recovery of educational content from a real, open and scalable environment, and will also will be a support mechanism to implement the recommendation or ranking for the recovered LOs

7.3 MODEL OVERVIEW

The design and development of SMA methodologies need to provide robust and reliable support to the designers. Many traditional approaches detail the structure of the SMA in terms of a role model, which identifies the roles that agents play in the system and the interaction protocols in which they participate.

These methodologies are classified as agent-oriented since they assume an individualistic perspective in which the lead agency includes agents with clearly defined tasks and skills that help each one to achieve its individual goals. It is, therefore, closed systems, which do not allow the participation of agents whose actions are in their own interest, selfish or unauthorized. In recent years, researchers have conducted several studies to provide procedures and methodologies that support open design SMA input and output dynamics of their bodies. Open MAS should also permit the participation of heterogeneous agents with different architectures and even different languages [Zambonelli *et al.*, 2003]. Therefore, they can not rely on the behavior of agents, which requires the need to establish controls on the basis of norms or social rules.

Researchers and developers have focused on developing a new approach called methodologies for the organization, which seeks to describe, in general terms, what the objectives of the organization are, its organizational structure

(topology, hierarchy of roles and interactions, its dynamics (how the agents enter / leave the organization, how they adopt roles, the life cycle of agents and social norms) and the environment of the organization.

Organization Theory sets out the basic concepts, relationships and intrinsic characteristics of each type of organization. In order to define an organization, regardless of type (human or software agents) it is necessary to consider the following factors [Daft, 1998] [Fox, 1981]:

- **Structure.** Includes all the elements that remain in the organization, regardless of what their individual goals are at any given time. Will be defined based on roles, their associations, dependencies and interaction patterns.
- **Functionality.** Specifies global goals of the organization, the services and features offered, the objectives pursued by the various components of the organization, and what tasks and plans must be followed to achieve them.
- **Standardization.** Determines the set of rules and actions defined to control the behavior of members of the organization. This includes performance standards (obligations, permissions and prohibitions), such as sanctions and rewards imposed or given to its members.
- **Dynamism.** Specifies how the organization evolves over time, indicating the manner in which agents enter and leave the organization in a dynamic way, adopt certain roles according to their abilities and skills, and participate in those units or groups of the organization into which they are admitted.
- **Environment.** Includes both the organizational resources, such as resource providers, and customers or beneficiaries of the existence of the organization.

The proposed architecture, called AIREH (Architecture for Intelligent Recovery of Educational content in Heterogeneous Environments), is seen as a communication point midway between the LOR, the LO that they store and users who use them. The system provides a federated search system that makes it possible to search multiple geographically dispersed repositories simultaneously. In addition, once the results of the different repositories have been received, an identification phase is developed and filtered in order to adapt the results to the user preferences.

The organizational model used in the architecture of the platform uses THOMAS [Carrascosa *et al.*, 2009] [Giret *et al.*, 2009], currently OVAMAH, to define the structure and rules. Adaptation in this type of model is based on coordination between the participants of the organization. The following sections describe how the proposed organization intends to perform the functions of architecture.

7.3.1 Design for the Virtual Organization Model

GORMAS (Guidelines for Organization-based Multiagent Systems) [Argente, 2008] was used as the design methodology and THOMAS (Technical Methods and Tools for Open Multi-agent Systems) [Carrascosa *et al.*, 2009] [Giret *et al.* 2009] as the final development platform.

The GORMAS methodology defines a Virtual Organization Model (VOM) composed of four dimensions: (i) structural dimension describes the elements of the system and how they relate to each other, (ii) functional dimension details the specific functionality of the system based on services, tasks and objectives, (iii) size of the environment defines each element of the environment that can be accessed through ports in the environment, and (iv) normative dimension describes the rules of the organization and the policy goals that the agents must follow. To define the four dimensions of the VOM, the model proposes a sequence of GORMAS divided into 8 phases:

- **Phase A: Mission.** An analysis of the motivation that seeks to define the organization or system, i.e., the reason for that organization or what is created; an analysis of the results that are expected to be achieved, and the environment that details the products and / or services to offer, the interest groups and their location.
- **Phase B: Tasks and processes.** The services to be offered in the system are discussed in more detail, including the requirements and processes involved. Also details the tasks and objectives related to those services.
- **Phase C: Organizational dimensions.** An analysis of the dimensions of the organization (departmentalization, specialization, decision-maker system, formalization, coordination), which impose certain requirements on the types of work and the diversity and interdependence of tasks.
- **Phase D: Organizational Structure.** Determines and selects the most appropriate organizational structure for the organization, depending on its size. Makes use of organizational models to specify the roles, interactions and norms related to the structure.
- **Phase E: information-decision processes.** Identifies detailed interactions for each service (information flows and decision making) needed to carry out the service. In addition, defines contracts as a service to those who engage providers and consumers after the fact.
- **Phase F: open system dynamics.** Provides the functionality offered as an open system, which includes the services that should be publicized, the procurement policies and release roles. In addition, designs the system for agents.
- **Phase G: Measurement, evaluation and control.** Quantifies or evaluates the tasks and activities and establishes mechanisms to determine

whether the system objectives are met. Also reviews the rules of the organization to specify who is in charge and supervises the process.

- **Phase H:** Reward systems. Determines the incentive system to reward members that move in line with the interests of the organization. Also discusses system sanctions for members who do not meet the guidelines given.

To establish the interaction model it is necessary to analyze the needs and expectations of potential system users. From this analysis it is possible to deduce the roles of users who reuse the educational resources, and how they are going to exchange information. The initial roles identified are:

- **User.** Represents the system user or customer:
 - Is the agent responsible for initiating the federated search process by sending a search pattern to the Query Manager role.
 - Gets the results from the query and is able to assess both the LO, and the order in which these results are presented.
 - Has access to statistical information from both the repositories and the LOs.
- **Query Manager.** This agent is responsible for overseeing the entire federated search process. As previously mentioned, a federated search is the simultaneous search of LO in multiple repositories.
 - Gets the natural language query from the user agent and is responsible for finalizing the query by propositional logic.
 - Queries the agent or federated search Repository Manager according to the pattern received.
 - Once the Repository Manager agent indicates the end of the federated search, the Query Manager asks the agent to apply the cataloguing techniques and collaborative filtering on the results, which will handle the agent results.
 - Once the agent has notified the Cataloguer that it has completed the management process of search results, it orders their transfer to the user agent for consultation and valuation.
 - Sends the official statistics or different information about federated query once it has finished.
- **Repository Manager.** Repository Manager agent has specific control over the queries that are made at different LOR; i.e., it has control over the federated search process.
 - This type of agent receives the query from the formalized Query Manager agent.
 - Checks the repositories that are active at the time of the query and asks the statistics agent for the statistics included in a list of those repositories whose performance is highest. The performance of a repository is obtained by dividing the number of results by the time needed to obtain them.

- Instantiates the LOR agents needed to perform the federated search, one for each repository included in the listing above. Once instantiated, it is responsible for sending each LOR agent a formalized query language.
- Is notified by the Query Manager when the federated search is complete.
- Is responsible for monitoring the proper functioning of the agents who take on the role of LOR and prevent yield loss if the repository stops working properly during the consultation.
- **LOR.** Is the agent responsible for conducting the consultation to each repository. Client code implements each of the possible middleware layers. Although only envisioned as a single role, in practice there are different types of agents who acquire the LOR role, so that each agent type implements a different middleware layer. There could be a case where a single repository containing middleware layers implemented would instantiate the LOR agent with the best performance among the potential agents for the same repository. In a federated search there will be many agents with LOR role as a repository that is directed towards the query.
 - This agent has the responsibility of making the request to each individual LOR, different instances of this agent will work simultaneously.
 - Prior to completing the consultation, the agent must instantiate the query translator and send formal language to be translated into the language used by the specific LOR to which the query is directed.
 - The agent performs the LOR query conducting all the necessary process, as defined in the specification of the middleware layer that implements the specific repository.
 - Is the agent responsible for sending the LO results that have been received in response to the query.
 - This Agent must notify the Repository Manager upon completion of the consultation
 - Should take statistical data from its query and send them to the Statistics agent.
 - Finally, it is responsible to cancel a query if it does not meet predefined QoS levels. If a query must be cancelled, the resulting error must send the official statistics
- **Translator.** Is the agent responsible for transforming the formalized language query language required by the repository to which the query is directed.
 - Receives the query expressed by propositional logic from the LOR agent.

- Performs the conversion of this query from a formal language to a language that is capable of dealing with the LOR on which the query is directed. This language is dependent on LOR and/or middleware layer for interoperability, so that there are different instances of this agent, one for each of the possible languages, but each one has the role of translator.
- Once the transformation had been carried out a query will be sent to the LOR agent in the language in which it operates.
- **Results.** The agent receives the LO results from each of the LOR agents in each federated search. Automatically extracts metadata schema information and eliminates those that are not valid LO. Although in theory there is only one results role, in practice there are different types of result agents so that each type of agent implements a different scheme of metadata.
 - Is responsible for receiving the federated search results from each one of the instantiated LOR agents in each federated search.
 - Extracts useful information from each diagram that describes the LO, therefore, will need to have access to metadata extraction and data structures needed to store the information provided by each of the schemes defines by the LO. An initial filter eliminates those defects that prevents the LO from receiving proper treatment and/or use by system users.
 - This role stores a minimum data set suitable for each of the schemes, and provides description of levels of abstraction in the comprehensive management resource. Runs an alignment algorithm for each of the LO of the results to determine the degree of overlap with respect to the particular object considered as the ideal candidate for the user's query.
 - Takes relevant statistical data as recall and precision of the LOR response to the formulation described in section 3.2 and then sends to the Statistics agent.

The Cataloguer role agents have access to the stored LOs.

- **Cataloguer.** This agent is responsible for preparing the ranking of LO that have been returned as federated search results by LOR staff and coordinated by the Repository Manager agent. The results will be stored by the agent who in turn automatically applies pre-filtering on LOs, obtaining useful information from them, and removes those that are incomplete. Cataloguer agent implements a CBR that uses previous search information to rank the items that best suit the needs (or likes) of the application user based on previously obtained information. It uses user profile information as well as their educational information (content-based filtering). It also subsequently uses information from the user's vote for LO and the appropriateness of the previously created ranking of the results (collaborative filtering).

- The agent orders the LO results to be stored in the agent according to the preferences of both general (profile) and user education.
- To carry out this process, it requests the agent statistics for the voting of the LO and previous ordinations, and the feedback provided by the user. With all this information it produces the ranking of LO that best matches the user who made the query.
- Provides the Statistics agent with the ranking information to be stored, which can then be used to produce future ranking. This stored ranking will be validated by the user during the results of the consultation process.
- **Statistics.** This agent is responsible for collecting statistical data from other agents. It also provides information it contains to increase system performance, in the case of the Repository Manager, or to improve the quality of the results or catalogue for the respective agents.
 - Receives the query statistics from the LOR Repository Manager roles and receives data from the LO rating of the user agent.
 - Receives feedback on the management of the results of the user agent.
 - Provides statistical data for the Repository Manager agent to prepare a list of repositories with the highest return, which will participate in each of the federated searches.
 - Provides statistical data for the cataloguer agent to produce the ranking of LOs for each of the federated searches.
- **Supervisor.** An agent carrying out this role will have overall control of the system.
 - Analyzes the structure and syntax of all messages entering and leaving the system.
 - Supervises the proper functioning of the other players in the system (sends a ping to periodically check the status of all agents of the architecture).

Following the guidelines in the methodological guide [Argente, 2008], one of the first tasks is to instantiate the functional vision (mission) of the organizational model, which includes the products and services offered by the system, the type of environment, goals, global aims (mission and justification), and the interest groups that are covered and the information they consume. This can be seen in Figura 5-6.

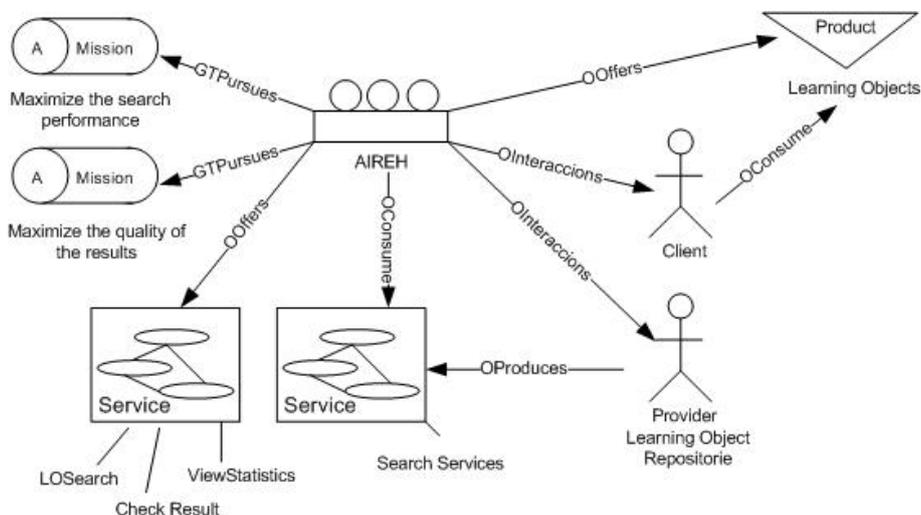


Figure 7-2. The organizational model Diagram (functional view). Mission architecture

Figure 7-2 details the elements of the organizational model (functional view), as it shows the results (products and services) offered by the system, the type of working environment, and the interest groups. Also justifies the rationale for the architecture, which aims to facilitate the search for educational resources through the implementation of a federated search system.

The functional view of the organizational unit shows AIREH SearchLO services, SeeStatistics ChechResults. In order to provide these services, the represented providers, in this case the LOR, must offer search services that enable the collection of information. Tabla 5-3 shows a more detailed analysis of the groups of interest in which the needs of the stakeholders are described, along with their specific relevance to the organization and the impact the decision making will have. Tabla 5-4 shows the analysis of the environmental conditions of the system.

The system offers educational resources as a final product encapsulated in the form of LO. These will be requested by users through a search process. It also provides statistical information such as product performance and use repositories of LO, identifying those that are used according to the search patterns. The mission of the organization will first maximize the system performance of queries by reducing time and increasing performance, and also, maximize the quality of results.

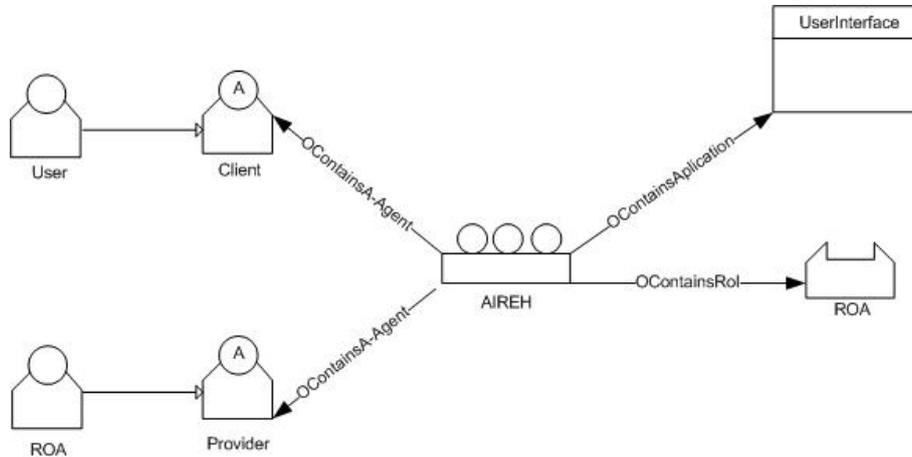


Figure 7-3. Organizational architecture model diagram (structural view)

In line with the GORMAS methodology, the next model in Figure 7-3 represents a structural view of the organization, while Figure 7-4 shows the functional view. This model defines an organizational unit that represents the AIREH system. Also, as seen in the structural view, the role represents the client as unique, since it is possible that all members who purchase this role will have access to the same functionality. In the same way the provider is represented as the LOR role, which will be the role in charge of accessing search service offered by the repository. Both customers and suppliers are modeled by the OcontainsAn-agent relationship, indicating that both are considered as part of the system.

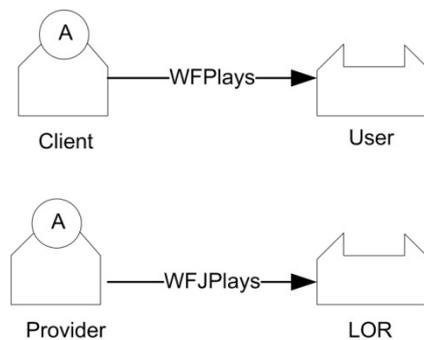


Figure 7-4. Organizational architecture model diagram (functional view)

Following all the guidelines, the next step is to get the model of the organization from its internal structure to its functionality. Figure 7-5 shows the functional

view (external function) AIREH associated organization model, whose services are connected to the associated roles and relationships of each system: WFPprovides WFUses.

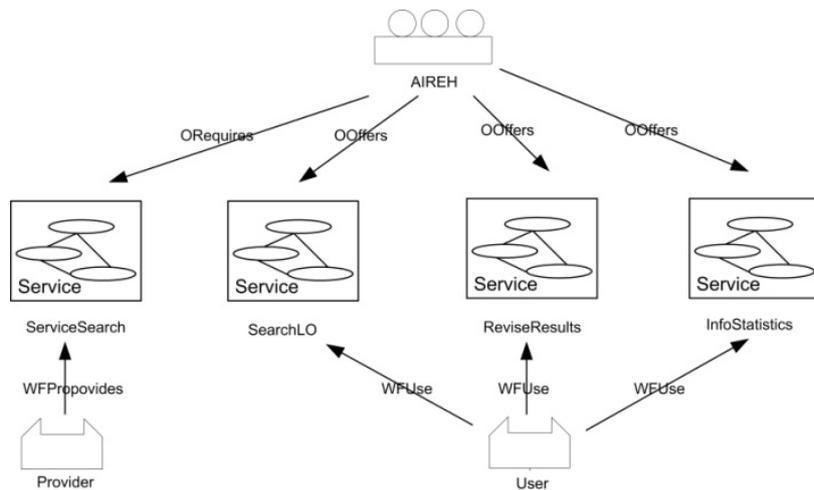


Figure 7-5. Organizational architecture model diagram (functional view, external functionality)

Among the three services offered, the most complex and representative is SearchLO service. Tabla 5-5 shows the detailed description of this service, while Figura 5-11 and Figura 5-12 show the pattern of activity diagrams.

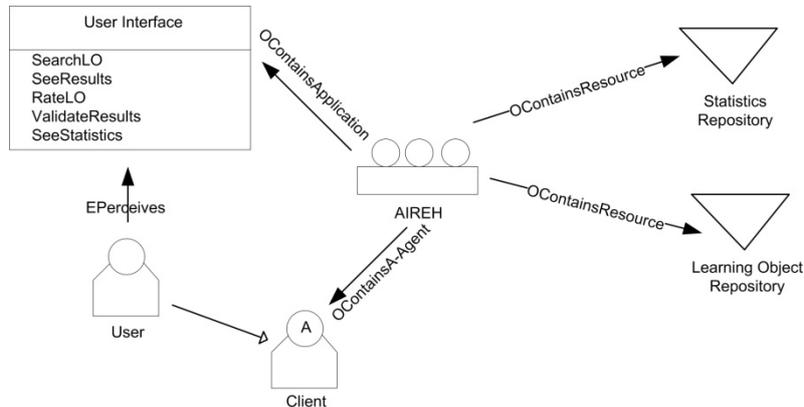


Figure 7-6. Model diagram of the architecture environment

Once the user role receives the results of the consultation that it initiated, it has the ability to assess both the quality of LO received, as the order in which results are presented. The importance of this step is that the feedback obtained

from the user will be used to improve the quality of future results. The detailed description of the Review Results Service is in Table 5-6.

The User role will also have access to statistical information regarding the performance of repositories and historical results. The service is presented in greater detail as it is quite complex.

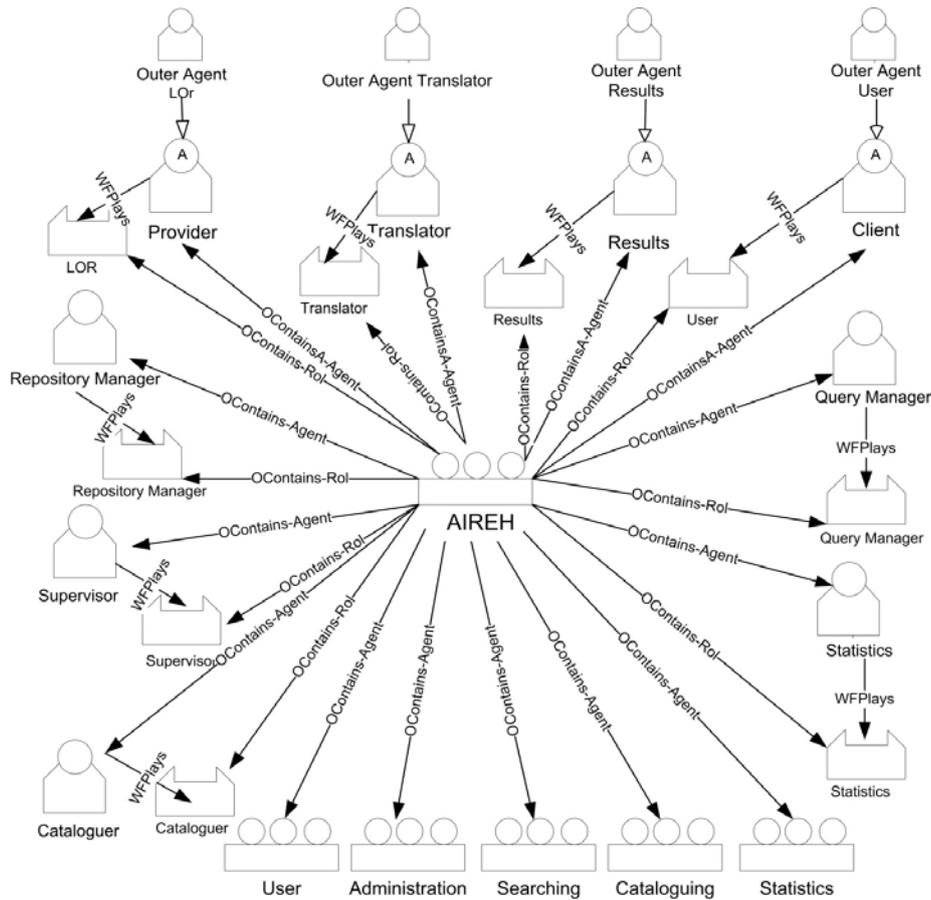


Figure 7-7. Updated diagram of the organizational model that shows both internal and external agents, and the roles they play

The information required by the system is provided through the Service Providers Search. The information of the system users is obtained through the UI application, which was identified earlier. Figure 7-6 shows the model diagram for the AIREH environment.

The GORMAS model also provides the functionality as an open system, meaning that both services should publicize the policies of acquisition and release roles, and determine what functionality must be implemented by internal staff and what is to be supplied by external agents. Thus, the dynamism of the organization is modeled as an open system. LOR roles, User, Results and translator will be accessible by outsiders that require a procurement role. The remaining roles (Query Manager, Repository Manager, cataloguer and Statistics) are associated with insiders being unreachable.

Figure 7-7 shows the organization model, indicating the external agent (A-Agents), internal (Agents) and the roles they play.

LOR roles, User, Results and Translator, as seen in the figure above, are accessible by external agents that require a procurement role.

The dynamism of the system, designed as an organization, can be reflected in different stages: registration of new players, new services, new protocols, service requests, expulsions from the system, etc.

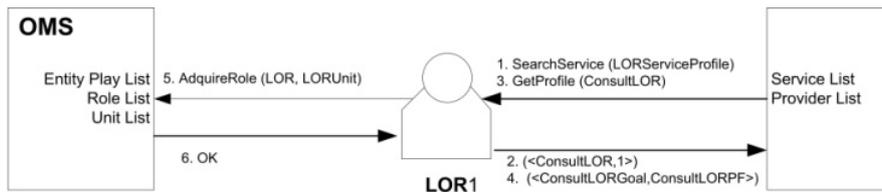


Figure 7-8. Example acquisition role by an external agent

Figure 7-8 shows a scenario where a new LOR is registered in the organization. LOR1 has been registered as a member of the platform and asks the Service Facilitator (SF) what services have already defined a similar profile. This request is carried out using the service SearchService (message 1) of the SF, which corresponds to the profile LORServiceProfile search service LOR1 implemented by the repository. SF returns the service identifiers to meet the search requirements with a value (rank value) for each service (message 2). The ranking value indicates the degree of alignment between the service and the specified service proposed. Then GetProfile LOR1 (message 3) obtains detailed information about the service ConsultLOR. The outputs of the service are "service goal" and "profile" (message 4). ConsultLOR profile specifies that the service should be provided by agents who assume the role in LORUnit LOR. For this reason, Organization Management Systems (OMS) asks the LOR1 service AcquireRole to acquire that LOR role (post 5). AcquireRole service is carried out successfully (message 6), as LORUnit is accessible from the virtual organization, and finally LOR1 is registered as LOR.

The units contain the Acquire Rol, Report Unit and Stop Rol services, in addition to the services dependent domain, which has been previously identified above. For example, if an external LOR agent wants to contact any of these services, it is first necessary to acquire the corresponding role.

In our case, the roles of management and services related to the organizational units will be available through the THOMAS platform [Carrascosa *et al.*, 2009] [Giret *et al.*, 2009]. As detailed in section C.3.1.2, OMS provides the services necessary for the proper functioning of an organization; it provides a range of services to register or unregister structural components, in particular, the roles, rules and existing units in the system. It also offers facilities to report on what those components are. Finally, it instantiates the structure model that follows the system. Each of the components of the architecture can be structured in different ways. In this case, the decision tree D phase [Argente *et al.*, 2008] has decided to use a congregation-like structure.

The structural diagram, see Figure 7-9, adapted according to the pattern of congregation AIREH unit, contains many features (and products), specifically the UserUnit, SearchUnit, CatalogingUnit, StatisticsUnit, and the AdministrationUnit. In addition, supplier and customer roles are refined into these new units to specialize in the functionality or the utilization of specific services associated with these types of products.

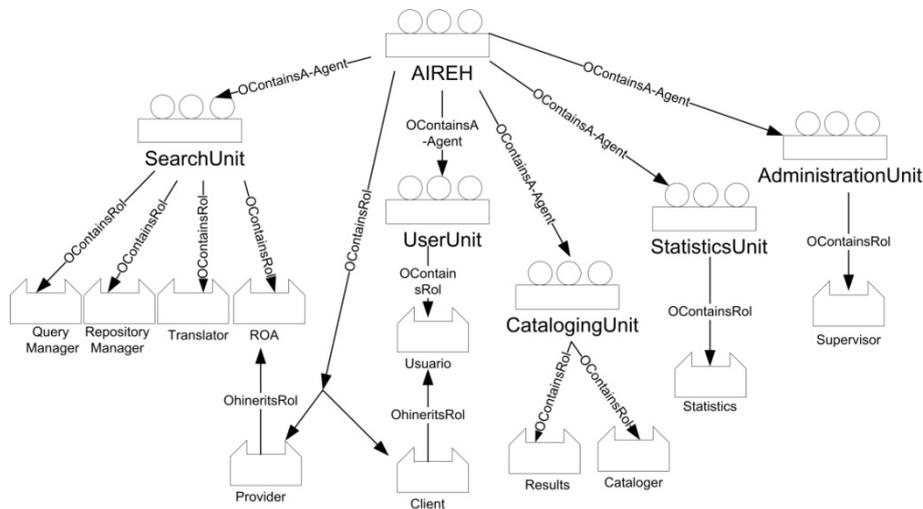


Figure 7-9. Adapted organizational model diagram according to a congregational structure (functional View)

The functionality diagram, see Figura 5-9, is also adapted to the pattern of congregation and displays the unit's internal modeling AIREH. Through the

Service Unit Registrar, providers can request the creation of new organizational units in order to meet the arrival of new features or products.

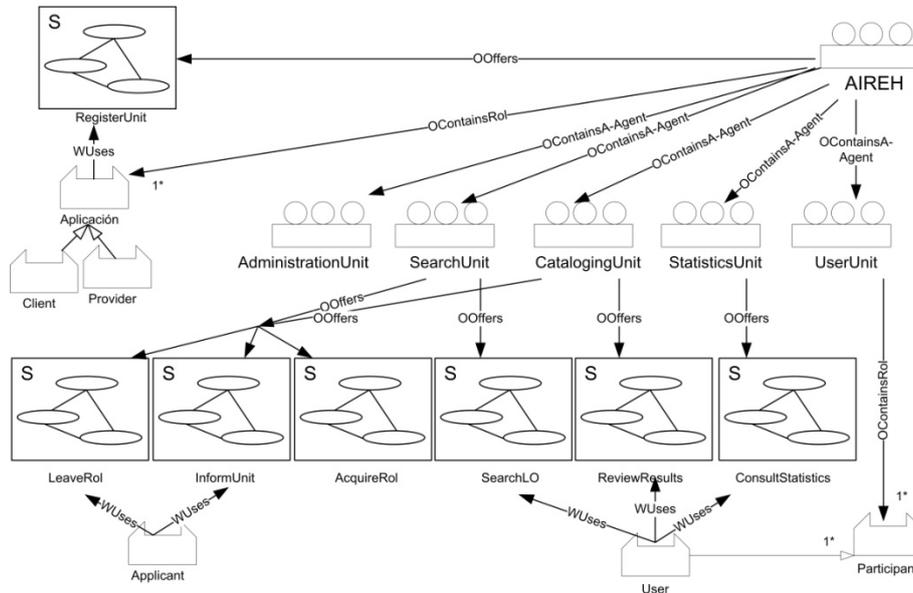


Figure 7-10. Adapted organizational model diagram according to a congregational structure (external functionality)

The units contain the Acquire Rol, Report Unit and Stop Rol services, in addition to the service dependent domain, which has been previously identified. For example, if an LOR external agent external wants to contact one of these services, it must first going through a process of acquiring the corresponding role.

The next step is the design of the dynamics of the organization. We analyze the information processes of the organization, detailing both its interaction with the environment, and the interactions between the component entities. The interactions produced by the activation of services are detailed in the interaction model, where the corresponding targets are associated with the interaction service.

7.4 EVALUATION FOR THE PROPOSAL

A new approach for intelligent search of educational content was introduced by proposing an architecture based on the application of virtual organizations of multi-agent systems. This architecture will be used in a particular case study

for federated search in repositories of learning objects and the subsequent fusion of results. These tests using the case study will assess the architectural model adopted. At a theoretical level this new heterogeneous and dynamic model of multi-agent systems fits perfectly into this context. The main objective of this thesis is to demonstrate that this assumption is reasonable at both a theoretical level and a practical level.

An evaluation of the architecture proposed in this research must be done under a set of metrics to verify its validity empirically. Following the considerations of the study in Annex B, the performance of the architecture will be evaluated based on the metrics used to measure the recovery of relevant documents by the Completeness and accuracy metrics described in section 3.2. These metrics from the field of distributed information retrieval have been adapted for the recovery of learning objects according to relevance criteria, which allows the reuse of learning objects such as the utility criterion. The recall and precision adapted to the recovery of learning objects are defined according to the formulation described in Equation (3.3) and Equation (3.4). The creation of a new performance measure referred to as temporary gain, which can calculate the efficiency of content recovery in time, is proposed in section 3.2 of this document. The following subsections show some experimental results, the conclusions of the investigation, the main contributions to the technology of agents and SMA, and some future lines of work in the short, medium and long term.

7.5 EXPERIMENTAL RESULTS

Educational metadata constitute a simple mechanism to describe content by using the tags to which they are associated. The practical aspect of this research requires knowledge of the actual use of the standards, i.e. LOM, for labeling educational resources, which is required for an empirical study. Based on actual documents for metadata, a set of public LOR are retrieved, and a report is produced detailing the results of Section B.3 of the annexes. This study shows the difficulty of the labeling process, which is basically done by hand and generates documents with serious shortcomings. These deficiencies affect syntactic problems in labeling data, they lack the key attributes in the description, and so on. This makes it difficult, or impossible in some cases, to directly use that metadata in the process of searching and retrieving the content to which they refer. The extension of the standard specification levels also reveals LOM attributes that are difficult to enforce, or are irrelevant or ambiguous. As a result of this study, a new pruning heuristic was established that can select some of the possible elements and quickly go through our system and filter out the majority of the irrelevant recovered objects. Once this process has taken place, the remaining elements are filtered according to

certain parameters related to the quality of the retrieved metadata documents, and the user context information is integrated into the use of the learning objects. The process is complex and the association of more aspects helps to improve the quality of the contents recovered. An algorithm for selecting effective content for a particular user should take into consideration the semantics of learning object use and the technical aspects for a search in the LOR, and should also catalog the results retrieved automatically through mechanisms involving smart features. The processing of metadata documents retrieved in this work addresses three aspects: completeness, reliability, and the accuracy of managing results (ranking accuracy):

Completeness. The description of the learning object through its metadata must be as complete as possible to provide results on which to plan automatic reasoning mechanisms that generate an acceptable level of confidence for their recommendation. The attributes in the missing or mislabeled metadata represent a major hole in the precision with which they have been recovered and can be treated. Some fields may be considered irrelevant LOM while others are absolutely essential (title, location, language, etc.) in order to automate the process of searching and filtering relevant LO to the user.

Reliability. A good quality metadata tagging is fundamental for LO recovery. This functionality automates the application that links all sides of the object (file, metadata and educational content) and equips the LOR hosting them with a higher rate of trust in the system. Any learning object can have syntactic or semantic errors. Syntax errors show a lack of conformity to specification. In the case of controlled vocabularies, it is assumed that one should use any of the terms available and not others. Other problems are incompatibility with the attribute values, misspellings, and so on. Semantic errors can only be detected by inspection heuristics in the context. Some attributes may be ambiguous or inadequate so that the processing of such information is subject to a certain degree of subjectivity and must therefore be limited as much as possible with that metadata.

Ranking accuracy. The set of possible learning objects that a user requests can be very large. The classification algorithm for optimal results should show the user retrieved learning objects by an ordered list according to a relevance criterion based on the request and the educational context in which the query is performed. The information displayed to the user about the relevant objects located is important to facilitate the user election process. The learning system attributes related to user domain is a key mechanism in the process of managing the implemented LO.

7.5.1 Performance Evaluation for Architecture Recovery

There is no doubt that the case study designed to verify the recovery architecture plays an important role in the results of this work. However, the

objective of the validation is only the recovery of the educational content of the proposed architecture using a generic case study. The case study was created as a federated search system with 2 LOR agents: an agent that queries user agent type, and a translator. Both LOR agents implement the SQL client code repository access interface.

Tests were performed on two repositories, each with a different session management. The tests were performed with the MERLOT⁴³ repository (stateful) and the LORNET⁴⁴ repository (stateless). Meanwhile the agent implements VSQL Translator query language.

The proposed architecture, AIREH, was evaluated by performing a battery of tests to validate their efficiency in real environments. The effectiveness of the system retrieval was measured in terms of precision (the fraction of retrieved documents retrieved by the system that are also relevant to the query) and recall (the fraction of the relevant documents present in the database that are retrieved by the system) as described in section 3.2. These parameters characterize the ability of the system to retrieve relevant documents and avoid irrelevant ones. To estimate these issues, 53 queries were analyzed, as shown in Tabla 6-1 in two repositories: LORNET and Merlot. Experts in information retrieval considerations suggest the creation of the queries and agree that this process should be done with the help of experts in the field [Gordon and Pathak, 1999] [Hawking *et al.*, 2001]. Also according to openDOAR⁴⁵, the most widely used language for searching digital content is English (77% of consultations). For this reason the search patterns or topics were chosen at random among topics in English for Science and Technology of UNESCO codes developed by experts. These topics are mainly composed of sets of words (longer queries) due to the specific nature of this terminology, but single words (short queries) were also used to assess different behaviors on the search engines repositories.

The system is robust against failure because it incorporates several methods within different agents in the organization during the time that the query is being processed. Prior to organizing the repositories according to performance, the Repository Manager agent checks the active repositories for any request made by the Query Manager. Once the LOR agents that perform the query in each repository have been instantiated, each agent is responsible for cancelling the query and reporting to the Statistics agent if there are any situations that affect the established QoS levels, such as query time, performance of the

⁴³ MERLOT <http://www.merlot.org/>

⁴⁴ LORNET <http://www.siat.sfu.ca/research/projects/29/>

⁴⁵ OPENDOAR <http://www.opendoar.org/>

repository, and so on. Although some of these LOR cancel the consultation process, there would still be other agents operating the instantiated LOR, which is always monitored by the Repository Manager. Because the system filters all the repositories that are not answered, the number of LO that the user has access to consists of the sum recovered from LO repositories at the time of consultation, as shown in Figure 7-11.

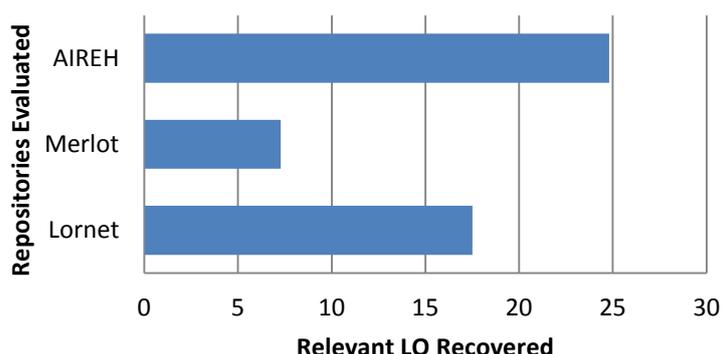


Figure 7-11. Metadata documents retrieved by the system in the testing phase

This data reveals that the proposed architecture significantly increases the number of LO to recover in time, and increases the temporary gain value of the system by about 15% on average as compared to isolated repositories such as Merlot or LORNET. Figure 7-12 shows the average temporary gains for the series of tested queries.

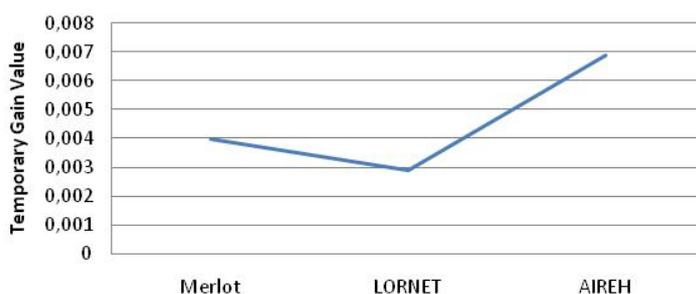


Figure 7-12. Temporary Gain Values

All the results are relevant consultations with the relevant criteria that have been established in this work. The recovery of content by the proposed architecture system depends on the LO returned by the isolated repositories, so

in the event that there is no LO to respond to the user's consultation, the AIREH system can not resolve the lack of content. However, the system ensures that any content related to the user's request will be recovered.

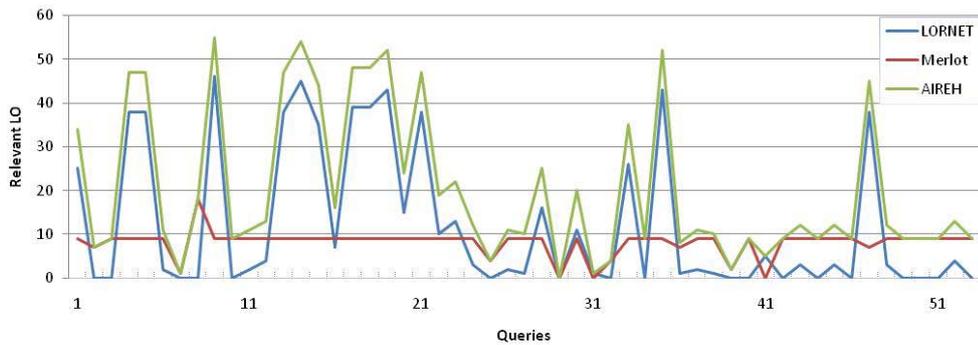


Figure 7-13. Comparison of relevant metadata documents (LO)

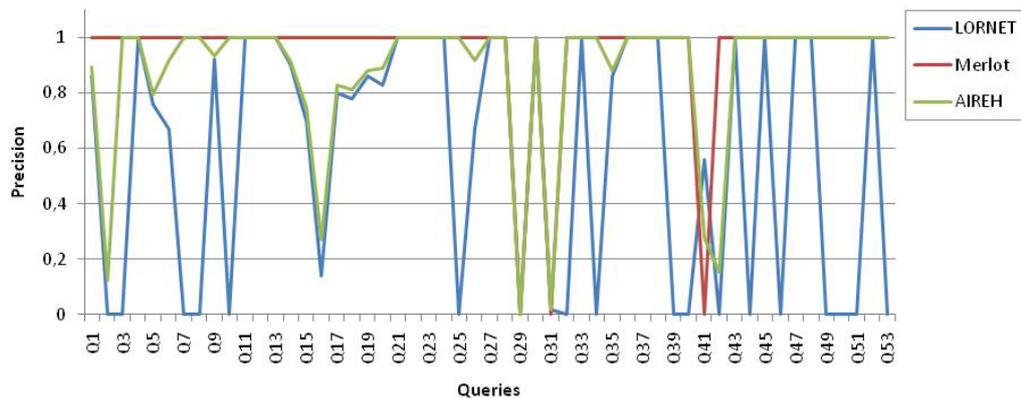


Figure 7-14. Comparative Precision Results for the Architecture Proposed

Related to measures of relevance of the retrieved content, Figure 7-13 shows that the relevance of AIREH in its internal operations is just under half the Merlot repository, due to the low relevance of content in Lornet. However, relative recall according to the criteria in Figure 7-15, increases to its maximum value, 1, because the system filters those LO that have no information about the location of the resource metadata (Lack of information for the attribute <location> under category <technical> in LOM).

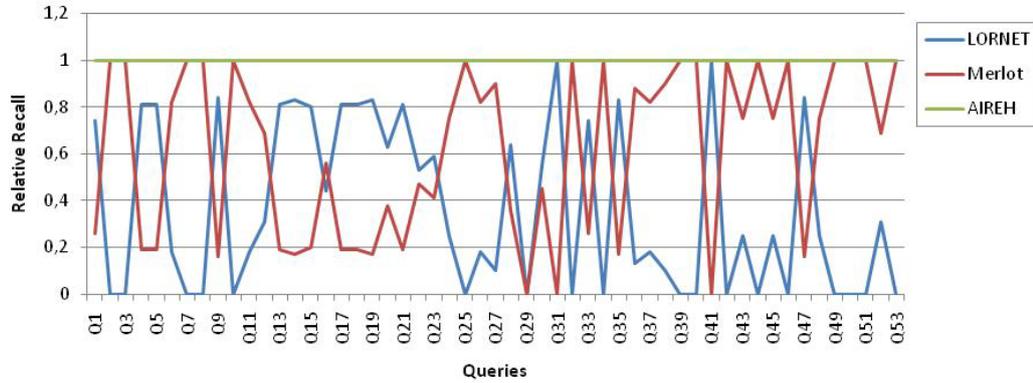


Figure 7-15. Relative Recall Comparative

As shown, both in the data as in the graphs, the assessment in the number of outcomes has improved considerably by using a solution based on federated search.

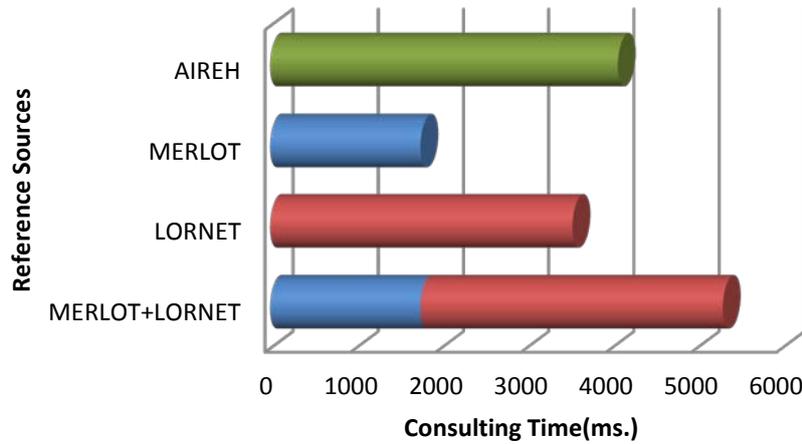


Figure 7-16. Average consulting time

It has achieved this increase in performance without increasing query time, Figure 7-16. This optimization in consulting time is due to the parallelization performed during the recovery, which minimizes the time involved by 15% on average compared to the repository that required the greatest amount of time. It is important to underscore this fact since the response that AIREH provides to the user query is a set of LO for the different repositories, having filtered out

those that are not reusable and ordering the remaining repositories according to the personalized aspects previously described. Both figures give the AIREH architecture a major improvement over the individual mode of research for this type of system.

The federated search mechanism has been addressed as the resolution of the issue of content retrieval by solving three phases: (1) the selection of repositories, (2) the recovery of content, and finally (3) the merger of results. In addition to the centralized search of learning objects in different repositories, the system performs a classification of objects retrieved through an initial filter, by removing duplicate learning objects but not detailed access to the source of content. This step addresses the first phase of the proposed federated search by selecting the most appropriate ROA according to statistical parameters and performance during the search and retrieval session of content, and is based on user research through key terms. An initial screening of recovered LO is immediately performed for the first organization that provides a homogeneous set of retrieved metadata documents. This pretreatment enhances the recovered achievement of the content for the second phase, the recovery in federated search.

A second stage of filtering incorporates aspects of quality of the retrieved objects according to two criteria (1) the assessment of the quality of metadata retrieved and (2) the assessment of LO estimated by users through collaborative techniques. The assessment of the quality of the retrieved metadata incorporates parameters such as size (a larger improvement in the quality of information content is estimated), main features (title, author, keywords, etc.) completeness, education, etc. The second incorporates the collective assessment on the use of LO through assessments from other users in the application, implemented by social aspects in the recommendation. The sum of these criteria on the ranking of retrieved objects in the system provides a hybrid recommendation that begins with a refined content-based recommendation and collaborative features. This stage ends with the third phase of the federated search, and includes recommendations on merging the content, which improves the quality of the retrieved content for the user that generated the query.

The recommendation mechanism used is based on a hybrid method that combines filtering techniques based on collaborative content aspects. This collaborative aspect comes from the feedback of users of the LO and is collected through the interface, as shown in Figura 5-30, and becomes part of the attributes explicit in the case base described in section 5.2.3 The user can issue a general assessment of learning objects retrieved via the interface, on a numerical scale of 1 to 10.

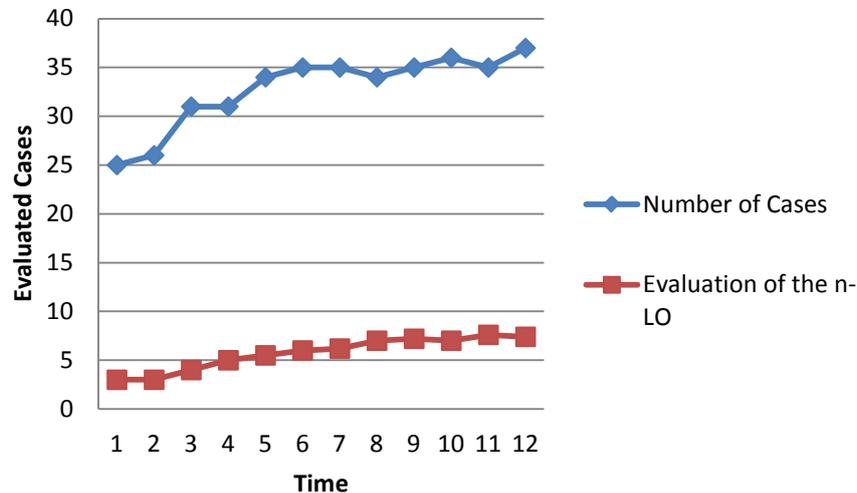


Figure 7-17. Evaluation of the recommendations of the CBR

To make the recommendation, the model applies the CBR model. The system accesses users' past experiences stored in the case, as described in paragraph 5.2.3. It is extracted from the base of a set of these cases that are similar to the query originally introduced. The case base is organized in such a way as to monitor user interaction in the context of filling out the attributes of interest to the user on the LO (case).

The recommendation is made according to the group of recovered cases. To validate the application, we compared the results obtained by the following evaluations of AIREH assessment by 6 months with 10 users. They perform the same battery of queries from a selection of keywords listed in Tabla 6-1. The evolution of the number of cases in the case base allows for greater knowledge and appreciation of potential LO, as well as users. The Figure 7-17 shows how the user will increase the value of the first n recommended LO ($n = 5$) as the system solves new cases. Success in the system is evaluated through user interaction with the recommended LO, as well as the assessment it makes of each. This improvement is due to the system's ability to learn and adapt to lessons learned. Likewise, the experiences allow a better adaptation to the user profile.

Figure 7-18 shows that the number of updated cases decreases as the system acquires experiences (the X axis represents the number evaluations during the time period and the y axis quantifies the number of cases concerning the aspects evaluated (number of cases, access to the base and updates of cases). It is logical, since by increasing the variability of the cases captured and the

number, the ability to find cases similar to the query that the user requires increases and may validate the recovered LO criteria defining user tastes and/or needs.

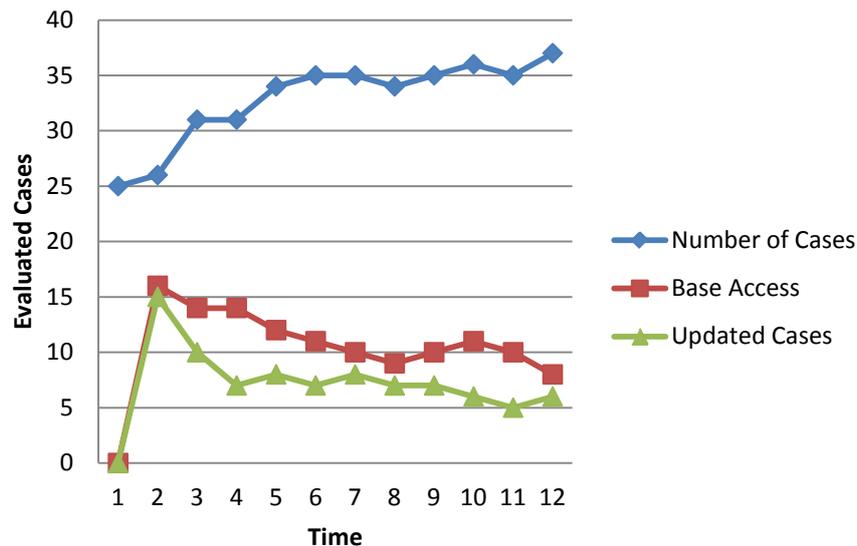


Figure 7-18. Evolution of the CBR

There has been a limitation on the number of cases in the base in 50 cases. This limitation avoids the problem of unlimited growth of the case base. In situations in which it was necessary to remove a case, it is made with the use of STIME attribute, used as a timestamp. That case with further time attribute compared with runtime would be eliminated and replaced by the new case.

AIREH proposes a flexible architecture that allows the development of a solution that provides functionality to manage content stored in repositories of learning objects. Reference resources are metadata through federated search mechanisms in various repositories accessed through federated searches. The architecture provides the user with management and access rights, etc. For this reason it seems reasonable to make a comparison in terms of competences between the development of AIREH and other tools that are emerging now and intend to solve the problem of recovering sparse educational content. Although still under development and being tested, the most relevant are the search tool GLOBE⁴⁶, which is used to search through a web client in the repositories that are part of the alliance, and the ARIADNE search tool, which is in beta testing

⁴⁶ GLOBE <http://ariadne.cti.espol.edu.ec/globeFinder/>

now and linked to the GLOBE project. This search through web client is shown in Figura 6-10.

Also of relevance is the work [Paquette *et al.*, 2006], which has resulted in the current PALOMAWeb tool⁴⁷, and allows the management of LO through its metadata; its interface is shown in Figura 6-11.

Table 7-1. Comparative characteristics of LO federated search engine

Solutions to the Educational Resources Search			
Features / Tools	PALOMA	GLOBE	AIREH
Access to distributed repositories	Yes	Yes (just what makes up the Alliance) Currently only ARIADNE (in beta)	Yes
Inclusion of different query languages	NO	NO	Yes
personalization	NO	NO	Yes
Incorporation of social aspects	NO	NO	Yes
Metadata access by the user	Yes	NO	Yes
Tracing historical storage	NO	NO	Yes
Search with Language criteria	NO	Not Implemented	Yes
Different standards of tagged languages	No, only LOM	Yes	Yes
Web Client	Yes	Yes	Yes

Table 7-1 shows a number of characteristics regarding the versatility of a digital educational content search by the selected tools as compared to the proposed architecture. Only the visible aspects have been taken by the user because there is no information on the methodologies and technologies used by these tools.

Clearly, the architecture proposed in this work to search for educational content incorporates important features such as personalization and metadata management that enhance the user's work and the treatment of the recovered LO. The proposal also allows flexible access to the repositories that the user sets of interest and the easy incorporation of query languages so long as the repositories surveyed meet the required specifications. These characteristics of flexibility and adaptability to the environment, along with the personalization

⁴⁷ PALOMAWEB <http://demo.liceftelug.quebec.ca/LomManager/?id=demo#B#>

of the results, present AIREH as an effective solution to the problem of delivered content.

So far for the recovery of those Learning Objects contained in ROAs has been used by search keywords in the metadata content. The use of multilabel classifiers guarantees the construction of powerful tools to increase the search power in large volumes of labeled data. The following section evaluates the work done in this paper about the idea of probing possible future improvements to refine and expand the possibilities of the AIREH architecture proposal.

7.5.2 *Multilabel Classification Techniques for Learning Object*

The multi-label classifiers studied in section 3.4.2.1, provide a web mining technique applied to machine learning on collecting information. These methods allow off-line learning on labeled data sets using various algorithms. It seems reasonable that this methodology could be used for the automated classification of LO in the collections by removing the labels that classify the relevant aspects of each LO. So far, search keywords in the content metadata have been used for the recovery of content in the LOR. This new ability to query would be useful to improve efficiency in the selection of LO in different LOR. It would make it possible to extend the expressive power in filtering the recovered objects within the architecture model presented in this study. By including this new information, such as a field within the database of CBR cases, it would be possible to increase LO aspects of similarity between the reasoning processes. By using these techniques it is possible to rank and select learning object repositories. To study these possibilities we planned a machine learning system in building a multi-label model with a training data set of learning objects for subsequent applications to a new set of data (untagged) in order to obtain predictions able to classify new cases.

We used a machine learning method, similar to [Trohidis *et al.*, 2008] for an empirical assessment of the MLkNN and RAKEL algorithms in a data set of multilabeled recovered learning objects. The RAKEL algorithm was selected as a method that has proved more effective than the first two listed [Trohidis *et al.*, 2008]. This method builds a set formed by the union of LP classifiers. RAKEL takes into account the correlations between labels, avoiding the problems of LP. The ranking of the labels is obtained by averaging the predictions [0,1] of each model established by label. From the values obtained, the threshold value is set to generate the bipartition. MLkNN was selected to be representative of high-performance algorithms in problems with adaptive methods based on the k nearest neighbors (kNN) and lazy learning algorithm [Wettschereck *et al.*, 2007].

The data set for training was taken from the 60 consultations recovered from the two repositories used for the case study. We present experimental results of the data set containing 253 examples of actual LO, annotated with one or more of the 38 labels in English corresponding to the query types identified by teachers and students as necessary for the search, such as “Programming languages”, “Hybrid Computing”, etc. Each LO is described with 1442 features extracted from LOM. The data set format (number of tags, number of features, order of attributes, etc.) must follow the format of the training sets based on the model built.

According to [Santos *et al.*, 2010], the evaluation of methods of learning multi-label data require different measures to those used in the case of single-label data. Several measures have been proposed for the evaluation of bipartition and classification with respect to multi-label data. Here, we are interested in evaluating the quality of bipartition and ranking. These experiments are performed on the data reported in the LO compared to the two algorithms RAKEL and MLkNN. Tabla 6-3 shows the prediction results of the two competing classification algorithms using a variety of measures. For all indicators, RAKEL significantly exceed those of the MLkNN algorithm, especially when taking into account the accuracy of measurement subset (Subset Accuracy), which is equal to the zero-one loss for the classification task tag only in predicting the exact tag.

In relation to time, RAKEL is a very fast algorithm, while MLkNN increased consumption. The experimental results indicate that RAKEL is not only more efficient in the training phase and the evidence than MLkNN, but also improves the accuracy of prediction. The results show that the RAKEL algorithm can be used to improve the classification of LO in the types of queries based on metadata content. Figura 6-12 shows the classification of LOs attending RAKEL for the keywords on the labels of the LOs training the system according to the survey conducted to recover the LO. The graph of Figura 6-12 represents the number of examples annotated with each label. Based on the ease of the predictions, classification labels can be made in the following descending order: L2, L3, L5, L6, L7, L8, L9, L11,..., L38.

Using an algorithm instance, RAKEL was able to do a ranking as a bipartition of the set of labels from multi-labeled data; making this mechanism feasible with the MLR task. Figura 6-13 shows an example of ranking for the grammar used for labeling LO. The only missing component would be a threshold criterion to determine the labels stored as relevant for each LO.

Multi-label classifiers such as RAKEL may be used for automatic annotation relevant to LO. This in turn would support implementation of effective techniques in multi-label information.

These results are preliminary and are currently being developed into an open researching branch. The objective is to support the implementation of the LO recovery systems in LOR consultation through their labels. This ability to query would be useful in the selection of LO across different applications to improve the proposed model.

7.6 PROPOSED MODEL GENERALIZATION

The innovation of the model presented and the implementation of AIREH is an architecture based on an organizational model for agents. This proposal allows the user to achieve a centralized search and location of relevant educational content, compared to highly dynamic environments, such as the Web where they are located in deep repositories of learning objects. The search process is integrated into the organization of the agents in a way that is totally transparent to the user, compared to the problems of distribution and integration of different repositories, abstracting from the internal logic of each and the classification, storage and the search for LOs. Taking into account the results obtained from the construction of the AIREH application, we can conclude that the agent-based architecture is efficient for solving the problem of federated search in heterogeneous repositories, due to the adaptation and learning of the parties involved. The architecture model presented allows the system to adjust in a changing environment in terms of workload and conditions that occur in real-time repositories with regards to their performance and availability. The architecture includes various aspects of the learning object context and user preferences.

The analysis, study and definition of the mechanisms used to solve problems involving federated search, resource recovery, and the combination of results can be implemented in different systems because they were created based on integration and interoperability standards. The case study focuses on federated search of repositories that implement the SQL protocol. This search could be very easily performed on repositories that base their communication on any type of protocol. The design of the LOR Agent in the organization model, as shown in Figure 7-8, allows each individual LOR to be consulted. It could do the same, for example, for a Translator agent instantiated in an OAI-PMH repository to translate to the specific language used by the ROA. As the protocol allows search and retrieval of metadata in the Dublin-Core standard, the integration of these metadata to the merger would make the agent process the results. So the characteristics of the organization of agents can implement a simple and scalable generalization to any LOR, based on their communication protocol, the query language and the description of its contents in any metadata standard.

The objective of this research focuses on the recovery of educational content, but with the growing popularity of labeling mechanisms for large repositories of information, the proposed solution is easily transferable to any other field. The difference in scope is the semantic context, affecting only the aspects related to labeling and information resources relevant to the generation of the corresponding base case.

Thus BioBanks, or biorepository, with digital collections of metadata that allows access to information on biological material, such as the European Bioinformatics Institute with solutions based on web services [Labarga *et al.*, 2007] on genomic data, is a reasonable field for the application of the proposed recovery architecture.

Another area that is seen as important in terms of information retrieval is the electronic document in the field of public administration. In the case of the Spanish state, there is a technical standard for electronic document belonging to ENI (National Interoperability Framework) that was regulated by Royal Decree 4 / 2010 of January 8. The National Interoperability Framework aims to create the conditions necessary to ensure an adequate level of the technical, semantic and organizational interoperability of systems and applications used by the government, which would allow the exercise of rights and the performance of duties via electronic access to electronic public services, while benefiting from their effectiveness and efficiency. The standard referred to in RD 4 / 2010 regulates the National Interoperability Framework in the field of e-Government and states that the government must take the necessary measures to ensure the interoperability, retrieval and preservation of electronic records over time. Clearly, an effort to preserve electronic government documents will generate large volumes of electronic documents that are stored in repositories. This sets up an environment where the proposed AIREH architecture could be significantly developed in the field of intelligent retrieval of digital documents.

7.7 CONCLUSIONS

An important objective in the development of recovery technology in the deep Web content is to improve the quality, scope and accuracy of existing visible Web engines through the use of structured descriptions of resources, ie, through semantic rich metadata. This is only possible if the metadata of these resources are accessible. Moreover, due to restrictive copyrights, the actual digital content is freely available only in limited cases. A practical approach is, therefore, to retrieve the metadata from a series of repositories of geographically distributed and centralized learning objects. After treating them properly and allowing users to perform searches from a single user interface, management becomes centralized and allows for the implementation of substantial improvements in the ranking process in recovered LO.

When building a federated search application, it is important to note the increasing number of available repositories, the need for advanced algorithms, the high intensive computation to information retrieval, as well as the growing number of users who need custom views on the content of these repositories. All these factors can cause scalability and performance problems when performing this type of application.

This research has presented a recovery architecture based on educational content from partner organizations. The model of partner organizations can provide an organization of self-adaptive capabilities in runtime for highly dynamic environments. The main novelty is that the proposed architecture is based on a model with dynamic and adaptive planning capabilities to carry out the optimal distribution of the tasks for the organization's member agents, thus enabling intelligent content retrieval and flexibility in a highly dynamic environment. In summary, the presented architecture can define the actions that an organization of agents must carry out, given the changes that may occur during the execution of a query, by using adaptive planning within an organization of agents according to context characteristics (users, profiles, features, content, variability of learning object repositories', etc.).

7.8 CONTRIBUTIONS

The research presented in this report provides some new contributions in the areas of content retrieval in heterogeneous environments, and real applications of the Theory of Agents and Virtual Organizations.

- The engineering methodology used has focused on the development of organizations, which has helped to cover the analysis, the design of the organizational structure, and the design of the dynamics of the organization in an efficient way for content retrieval in heterogeneous environments.
- The development demonstrated the validity of the system in a real environment, which included the development of an organization of agents focused on services that centralize the search, retrieval and cataloging of heterogeneous resources in educational environments while responding to user profiles.
- The results obtained were analyzed and studied to demonstrate that this architecture is a viable solution.
- A great effort was made to get feedback from other researchers and research groups in related areas. The aim is to strengthen this research through the exchange of ideas and knowledge. We are particularly interested in expanding our experiences and the progress of this

investigation, from its earliest stages to its final form, through various publications and attendance of conferences, workshops, etc.

In conclusion, it is possible to ascertain that the objectives for which this research began have been obtained:

- (i) It has shown the adequacy of organizational models for agents to search and retrieve educational resources;
- (ii) Empirical data has verified the validity and performance of the models, and finally
- (iii) A hybrid recommendation model was proposed based on collaborative filtering and user preferences adapted to the context of the documentation and selection of learning objects.

This research project proves that it is possible to model an efficient system for managing open systems from a model of adaptive organizations to provide distributed and dynamic planning. Specifically, the proposed model architecture provides the necessary services to carry out planning tasks dynamically. Notably, both the model and the technologies presented in this research are an example of the potential for developing content retrieval systems based on the paradigm of virtual organizations. The advantages of the proposed architecture are flexibility, customization, uniformity and efficiency.

The research presented in this report is relevant to everything from a scientific point of view, setting the stage for the major development of projects at a national level, in particular the proposed OVAMAH system:

- ❖ OVAMAH⁴⁸ Project (Adaptive Virtual Organizations: Mechanisms, Architectures and Tools) TIN 2009-13839-C03-03, awarded by the Ministry of Science and Innovation. Basic Research Projects not oriented.

The hypothesis for the OVAMAH project consists of providing autonomous capabilities to a virtual organization that allows for possible dynamic response to changing situations through the adaptation and/or evolution of the organization, so that it is capable of detecting situations of interest, such as malfunctions, and can manage to maximize its flexibility and adaptability. Therefore, the work presented in this thesis is an important contribution towards the short-term objectives for this project, as *it has modeled a retrieval system of educational content that can be adapted to the common goals of deliberative agents involved in an organization of agents.*

⁴⁸ BISITE <http://bisite.usal.es/webisite/?q=es/node/2>

7.9 FUTURE WORK

The research presented in this work is the starting point in an expanding field that is clearly leaving open many avenues for future work. In this sense, the main line of future work will be the inclusion in the virtual organization of new filtering mechanisms and the subsequent validation by empirical evidence, which has been done with the federated search system. In addition to this main line of work, there is also the design and further development of algorithms to improve the search and retrieval of learning objects relevant to the user, increasing the number of features that affect the level of relevance of learning objects. Finally, another line of work is to develop an interface tailored to the needs of a system of this type, which, as with the case study, includes the virtual organization agent found in lower levels of the system.

Issues such as the inclusion of criteria that address the internationalization of the search for educational content should be taken into account, in order to solve the problems of multilingual retrieval and the merging of results based on criteria that include elements related to language, etc. Possible areas of work that could be directly related to this work include:

- ❖ **Testing and Validation.** Much more extensive testing is needed in order to assess the proposed architecture in greater detail with regards to time, application analysis and design, calculation of response time, quality of learning objects, etc. The results could lead towards the development of more refined models and robust systems.
- ❖ **Resolution of new practical problems.** To perform a more thorough check of the validity of the proposed model, it is necessary to apply it to new practical problems. In this way, it would be possible to check if it can properly resolve problems of different characteristics, or if the model is limited to the specific problems that have been studied for this work.
- ❖ **Implementation of multi-label classifiers.** Section 8.5.2 below describes a case study that shows the possibility of including mining techniques for multi-label classification of learning objects. Incorporation of these techniques to extract user profiles or generate clusters of learning objects can provide substantial improvements in the process of recommending content.
- ❖ **Integration of semantic aspects to recover and catalogue content** Even when educational resources are labeled according to a metadata standard, they are mainly descriptive and do not provide semantic

information that is directly associated to it. They will be enriched with semantic search processes and the treatment of learning objects based on models of knowledge such as domain ontologies. This will increase the functionality of the proposal with semantic features based on the consultation, and prioritize the results of this search according to their meaning, establishing metrics between the elements of the basis of semantic distance metrics.

BIBLIOGRAFÍA

REFERENCIAS BIBLIOGRÁFICAS

- [ADL SCORM 2004, 2009] SCORM 2004 4th Edition Version 1.1 Documentation. Disponible on-line: <http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/2004%204th%20Edition/Documentation.aspx>
- [ADL SCORM, 2002] ADL SCORM (2002) Sharable Course Object Reference Model v1.2. Disponible on-line: http://adlnet.org/ADLDOCS/Other/SCORM_1.2_PDF.zip
- [ADL SCORM, 2006] Sharable Course Object Reference Model 2004 3rd Edition Documentation Suite Public Draft Disponible on-line: http://adlnet.org/ADLDOCS/Other/SCORM_1.2_PDF.zip.
- [Aijuan & Baoying, 2008] Aijuan, D., & Baoying, W. (2008). Domain-based recommendation and retrieval of relevant materials in e-learning. In IEEE international workshop on semantic computing and applications 2008 (IWSCA '08) (pp. 103–108)
- [Almaraz, 2008] Ángel Almaraz. Búsquedas Federadas Significativas en repositorios de Objetos de Aprendizaje. Disertación para MsC en Ciencias de la Computación. 2008.
- [Al-Maskari and Sanderson, 2010] Al-Maskari, A. and Sanderson, M. (2010), A review of factors influencing user satisfaction in information retrieval. *Journal of the American Society for Information Science and Technology*, 61: 859–868. doi: 10.1002/asi.21300
- [Andrade *et al.*, 2008] F. Andrade, P. Novais, J. Machado, J. Neves. (2008). Software Agents in Virtual Organizations: Good Faith and Trust. *Virtual Enterprises and Collaborative Networks* pp. 389-396.
- [Argente *et al.*, 2004] E. Argente, A. Giret, S. Valero, V. Julian, V. Botti. 2004. Survey of MAS Methods and Platforms focusing on organizational concepts. In: Vitria, J, Radeva, p and Aguilo, I (ed) *Recent Advances in Artificial Intelligence Research and Development, Frontiers in Artificial Intelligence and Applications*: 309–316.
- [Avancini. & Straccia, 2005] Avancini H. & Straccia U. (2005) User recommendation for collaborative and personalised digital archives. *International Journal of Web Based Communities* 1, 163–175.
- [Baeza-Yates & Ribeiro-Neto, 1999] Baeza-Yates, R., & Ribeiro-Neto, B. *Modern Information Retrieval*. Boston: Addison Wesley Longman.1999.
- [Bajo *et al.*, 2006] Bajo, J., De Paz, Y., De Paz, J. F., Martín, Q., & Corchado, J. M. (2006). SMas: A Shopping Mall Multiagent Systems. *Proceedings of IDEAL'06, Lecture Notes in Artificial Intelligence (LNAI)*. 4224, pp. 1166-1173. Springer-Verlag.
- [Bajo *et al.*, 2007] Bajo J., Corchado J.M. y Rodríguez S. Intelligent Guidance and Suggestions using Case-Based Planning. *Proceedings of ICCBR'07, LNAI*, vol 4626 pp. 389-403, Springer Verlag. (2007)
- [Bajo *et al.*, 2008] J. Bajo J, V. Julian, J.M. Corchado, C. Carrascosa, Y. de Paz, V. Botti, J.F. de Paz. An Execution Time Planner for the ARTIS Agent Architecture. *Engineering*

Applications of Artificial Intelligence Volume: 21 Issue: 5 Pages: 769-784 Published: AUG 2008

- [Barry & Schamber, 1998] Carol L. Barry, Linda Schamber, Users' criteria for relevance evaluation: A cross-situational comparison, *Information Processing & Management*, Volume 34, Issues 2-3, March-May 1998, Pages 219-236, ISSN 0306-4573, DOI: 10.1016/S0306-4573(97)00078-2.
(<http://www.sciencedirect.com/science/article/B6VC8-3TC6SNB-9/2/53eff96536226308c70141c551244623>)
- [Bates, 2005] Bates, T.: Questioning the value of re-usable learning objects in education: the need for a business case Tony Bates. II Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables. Barcelona, 2005.
- [Benito, 2005] Carlos Benito Amat. Rendimiento de 8 sistemas de recuperación de información del espacio web español. *El profesional de la información*, ISSN 1386-6710, Vol. 14, Nº 5, 2005, pags. 335-346
- [Bergman, 2001] Michael K. Bergman. White Paper: The Deep Web: Surfacing Hidden Value. *The Journal of Electronic Publishing*. Volume 7, Issue 1, August, 2001 DOI: 10.3998/3336451.0007.104
- [Bergmann *et al.*, 2005] Bergmann R., Kolodner J. y Plaza E. Representation in case-based reasoning. *The Knowledge Engineering Review*, Vol. 00:0, 1-4.c 2005, Cambridge University Press
- [Bergmann y Stahl, 1998] Bergmann R. y Stahl A. Similarity measures for object-oriented case representations. In B. Smyth and P. Cunningham, editors, *Advances in Case-Based Reasoning*, volume 1488 of *Lecture Notes in Computer Science*, pages 25-36. Springer, 1998.
- [Bernon *et al.*, 2003] Bernon, C., Gleizes, M., Peyruqueou, S., and Picard, G. 2003. ADELFE: a methodology for adaptive multi-agent systems engineering. In *Proceedings of the 3rd international Conference on Engineering Societies in the Agents World III* (Madrid, Spain, September 16 - 17, 2002). P. Petta, R. Tolksdorf, and F. Zambonelli, Eds. *Lecture Notes In Artificial Intelligence*. Springer-Verlag, Berlin, Heidelberg, 156-169.
- [Bernon *et al.*, 2005] Bernon, C., Camps, V., Gleizes M.P., Picard G. Engineering Adaptive Multi-Agent Systems: the ADELFE Methodology. In B. Henderson-Sellers and P. Giorgini (Eds.), *Agent-Oriented Methodologies*. Idea Group Pub, June 2005, pp.172-202.
- [Bobadilla *et al.*, 2009] Bobadilla, J., Serradilla, F., Hernando, A., & MovieLens. (2009). Collaborative filtering adapted to recommender systems of e-learning. *Knowledge-Based Systems*. doi:10.1016/j.knosys.2009.01.008.
- [Boella *et al.*, 2005] G. Boella, J. Hulstijn, and L. Van Der Torre. 2005. Virtual organizations as normative multiagent systems," in *HICSS IEEE Computer Society*.
- [Boissier, 2007] O. Boissier and B. Gateau. Normative multi-agent organizations: Modeling, support and control. In *Normative Multiagent Systems*, 2007
- [Boutell *et al.*, 2004] M., Boutell, J., Luo, X., Shen, C., Brown, Learning multi-label scene classification. *Pattern Recognition* 37 (2004) 1757-1771.

-
- [Bratman *et al.*, 1988] Bratman M.E., Israel D. and Pollack M.E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4. pages 349-355.
- [Brenner *et al.*, 1998] Brenner W., Zarnekow R., Wittig H. "Intelligent Software Agents : Foundations and Applications" Springer 98.
- [Bresciani *et al.*, 2004] P. Bresciani, A. Perini, P. Giorgini, F. Giunchiglia and J. Mylopoulos. 2004. Tropos: An Agent-Oriented Software Development Methodology, *Autonomous Agents and Multi-Agent Systems* 8 (3): 203-236.
- [Brinker *et al.*, 2006] K., Brinker, J., Fürkranz, E., Hüllermeier, A unified model for multi-label classification and ranking. In: *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI 06)*, Riva del Garda, Italy (2006) 489-493.
- [Bryl *et al.*, 2009] Volha Bryl, Paolo Giorgini, and John Mylopoulos. Designing socio-technical systems: From stakeholder goals to social networks. *Requirements Engineering*, 14(1):47-70, 2009.
- [Callan, 2000] Callan J. Distributed Information Retrieval. In W.B. Croft, editor, *Advances in information retrieval*, chapter 5, pages 127-150. Kluwer Academic Publishers, 2000.
- [Capera *et al.*, 2003] Capera, D., Georgé, J.-P., Gleizes, M.-P., & Glize, P. (2003). Emergence of organisations, emergence of functions. *Symposium on Adaptive Agents and Multi-Agent Systems*, (págs. 103-108).
- [Caplan, 2003] Caplan, P. *Metadata Fundamentals for All Librarians*. American Library Association. Chicaco. 2003.
- [Caplan, 2005] Caplan, P. You call it corn, we call it syntax-independent metadata for document like objects. *The Public Access Computer Systems Review*, v. 4, n. 6. 2005
- [Carneiro *et al.*, 2010] Carneiro, D., Novais, P., Andrade, F., Zeleznikow, J., Neves, J. (2010). Using Case-Based Reasoning to Support Alternative Dispute Resolution. In *Distributed Computing and Artificial Intelligence. Advances in Intelligent and Soft Computing*. Springer Berlin / Heidelberg Vo. 79, pp. 123-130.
- [Carrascosa *et al.*, 2008] Carrascosa C., Bajo J., Julián V., Corchado J.M. y Botti V.: Hybrid multi-agent architecture as a real-time problem-solving model. *Expert Systems with Applications*, vol. 34 (1) pp. 2-17. Pergamon-Elsevier Science LTD. doi:10.1016/j.eswa.2006.08.031. (2008)
- [Carrascosa *et al.*, 2009] C. Carrascosa, A. Giret, V. Julian, M. Rebollo, E. Argente, V. Botti .2009. Service Oriented MAS: An open architecture (Short Paper), *Proc. of 8th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, Decker, Sichman, Sierra and Castelfranchi (eds.), May, 10-15, 2009, Budapest, Hungary: 1291-1292
- [Casquero *et al.*, 2008] Oskar Casquero, Javier Portillo, Manuel Benito, and Jesús Romo. BILDU: Compile, Unify, Wrap, and Share Digital. *Learning Resources Interdisciplinary Journal of E-Learning and Learning Objects* Volume 4, 2008, pp.97-111.

- [Castelfranchi, 2000] Castelfranchi, C.: Engineering Social Order. In: Omicini, A., Tolksdorf, R., Zambonelli, F. (Eds.): Engineering Societies in the Agents World, LNAI 1972, Springer-Verlag, 2000, pp. 1 – 19.
- [Cavedon y Rao, 1996] Cavedon, L., & Rao, A. S. (1996). Bringing about rationality: Incorporating plans into a BDI agent architecture. In N. Y. Foo, & R. Goebel (Ed.), In Proceedings of the 4th Pacific Rim international Conference on Artificial intelligence: Topics in Artificial intelligence. Lecture Notes In Computer Science (LNCS). 1114. Springer-Verlag, London, UK.
- [Cetkovic y Parmee, 2002] Cvetkovic D. y Parmee I. Agent-based support within an interactive evolutionary design system. Artificial Intelligence for Engineering Design, Analysis and Manufacturing vol. 16 (5), pp. 331—342. (2002)
- [Chang *et al.*, 2001] Chang, G., Healey, M., McHugh, J.A.M., Wang, T.L.. Mining the World Wide Web: An Information Search Approach. Kluwer Academic Publishers, 2001
- [Cheong and Tay, 2003] Y.M. Cheong, J.C. Tay. Approximate String Matching for Multiple-Attribute, Large-Scale Customer Address Databases. Digital Libraries: Technology and Management of Indigenous Knowledge for Global Access. Lecture Notes in Computer Science, 2003, Volume 2911/2003, 168-172
- [Cheong, 1996] F. Cheong, *Internet Agents: Spiders, Wanderers, Brokers, and Bots*. Indianapolis, IN : New Riders, 1996.
- [Cheong *et al.*, 2003] K.W. Cheung, J.T. Kwok, M.H. Law, K.C. Tsui: Mining customer product ratings for personalized marketing. Decision Support Systems 35(2), 231-243, 2003.
- [Chiappe *et al.*, 2007] Chiappe, A., Segovia, Y., & Rincon, H. Y. (2007). Toward an instructional design model based on learning objects. Educational Technology Research and Development , 55, 671-681.
- [Conte y Paolucci, 2001] R. Conte and M. Paolucci. Intelligent social learning. Artificial Society and Social Simulation, 4(1):1-23, 2001.
- [Corchado *et al.* 2008] Corchado J. M. , Gonzalez-Bedia M. , De Paz Y. ,Bajo J. y De Paz J.F.: Replanning mechanism for deliberative agents in dynamic changing environments. Computational Intelligence ISSN: 0824-7935. Mayo 2008. Volumen 24, núm. 2, Pág. 77-107 DOI: 10.1111/j.1467-8640.2008.00323.x
- [Corchado *et al.*, 2007] Corchado J.M., Aiken J. and Bajo J.: A CBP Agent for Monitoring the CO2 Exchange Rate. Book Chapter in Case-Based Reasoning on Signals and Images. Petra Pernert (Ed.). Studies in Computational Intelligence vol. 73, ISBN: 978-3540731788, Springer Verlag. (2007)
- [Corchado y Laza, 2003] Corchado J. M. and Laza R. (2003). Constructing Deliberative Agents with Case-based Reasoning Technology, International Journal of Intelligent Systems. Vol 18, No. 12, December. pp.: 1227-1241.
- [CORDRA, 2005] Documentación del ADL CORDRA (*Content Object Repository Discovery and Registration/Resolution Architecture*) disponible en <http://www.adlnet.gov/Technologies/adlr/ADLRDocuments/ADL%20Registry%20Documentation/ADL%20Registry%20Documentation.aspx>

-
- [Cossentino, 2005] M. Cossentino. From requirements to code with the passi methodology, *Agent Oriented Methodologies IV*: 79–106. 2005
- [Cox *et al.*, 2006] Cox M. T., Muñoz-Avila H., y Bergmann R. (2006). Case-based planning. *Knowledge Engineering Review*. 20(3): 283-287.
- [CQL, 2008] CQL: Contextual Query Language (SRU Version 1.2 Specifications) 2008, Library of Congress. <http://www.loc.gov/standards/sru/specs/cql.html>
- [Criado *et al.*, 2009]. N. Criado, E. Argente, V. Julian and V. Botti. 2009. Designing Virtual Organizations. 7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS2009). Volume 55:440-449.
- [Cunningham, 2001] P. Cunningham, R. Bergmann, S. Schmitt, R. Traphoner, S. Breen and B. Smyth. WEBSELL: Intelligent Sales Assistants for the World Wide Web. In *e-Business and e-Work 2001* (e-2001), 2001
- [CWA 15454, 2005] CWA 15454:2005 - A Simple Query Interface Specification for Learning Repositories. Bernd Simon, Erik Duval, David Massart, Frans Van Assche, Stefaan Ternier. Version 1.0, November 2005
- [D’Inverno y Luck, 2004] D’Inverno M. y Luck M. *Understanding Agent Systems*. Springer Verlag. isbn 3540407006. 2004.
- [Davidsson, 2001] Davidsson, P.: Categories of Artificial Societies. In: A. Omicini, P. Petta, R. Tolksdorf (Eds.): *Engineering Societies in the Agents World II*, Springer Verlag LNAI 2203, 2001.
- [DCMI, 2008] DCMI Specifications 2008. Retrieved February 18, 2010 from <http://dublincore.org/specifications/>
- [DCMI, 2010] DCMI Specifications 2010. <http://dublincore.org/documents/dcmi-terms/>
- [De la Rosa *et al.*, 2003] Josep Lluís De La Rosa , Esteve Del Acebo , Beatriz López , Miquel Montaner, *From physical agents to recommender agents, Intelligent information agents: the agentlink perspective*, Springer-Verlag, Berlin, Heidelberg, 2003
- [DeLoach *et al.*, 2001] DeLoach, S. A., Wood, M. F. and Sparkman, C. H., “Multiagent Systems Engineering”. *The International Journal of Software Engineering and Knowledge Engineering*, 11(3), pp. 231-258, June 2001
- [DeLoach, 1999] DeLoach, Scott A. *Multiagent Systems Engineering: A Methodology and Language for Designing Agent Systems.. Agent Oriented Information systems 99 (AOIS'99)*.
- [DeLoach, 2002a] DeLoach, S. A. *Analysis and Design of Multiagent Systems Using Hybrid Coordination Media. Proceedings of Software Engineering in Multiagent Systems (SEMAS 2002)*. 2002.
- [DeLoach, 2002b] DeLoach, S. A. *Modeling Organizational Rules in the Multiagent Systems Engineering Methodology. Proc of the 15th Canadian Conference on Artificial Intelligence*. 2002.

- [DeLoach, 2006] DeLoach, Scott A. Engineering Organization-Based Multiagent Systems. A. Garcia *et al.* (Eds.): SELMAS 2005, LNCS 3914, pp. 109 – 125, 2006
- [Demsey y Patel, 2000] Demsey, L.; Heery, R. “A review of metadata: a survey of current resource description formats.” Disponible en: <http://www.ukoln.ac.uk/metadata/desire/overview/>
- [Di Marzo *et al.*, 2004] Di Marzo Serugendo, G., Gleizes, M.-P., & Karageorgos, A. (2004). AgentLink First Technical Forum Group Self-Organisation in Multi-Agent Systems. AgentLink Newsletter , Issue 16, 23-24.
- [Di Marzo *et al.*, 2006] Di Marzo Serugendo G., Gleizes M.-P., and Karageorgos A., Self-Organisation and Emergence in MAS: An Overview, *Informatica* 30 (2006) 45–54
- [Dietze *et al.*, 2007] Stefan Dietze, Alessio Gugliotta, John Domingue. A Semantic Web Service oriented Framework for adaptive Learning Environments. Lecture Notes in Computer Science, volume 4519/2007, pp: 701-715.
- [Dignum *et al.*, 2005] V. Dignum, J. Vazquez-Salceda, and F. Dignum. 2005. Omni: Introducing social structure, norms and ontologies into agent organizations. LNAI 3346.
- [Dignum, 2004] V. Dignum. A model for organizational interaction: based on agents, founded in logic, PhD. Thesis, 2004.
- [Diplaris *et al.*, 2005] S., Diplaris, G., Tsoumakas, P., Mitkas, I., Vlahavas, I., Protein classification with multiple algorithms. In: Proceedings of the 10th Panhellenic Conference on Informatics (PCI 2005), Volos, Greece (November 2005) 448-456.
- [DRI, 2003] IMS Digital Repositories Interoperability - Core Functions Information Model, K.Riley and M.McKell, Version 1.0, IMS, January 2003.
- [Duval & Hodgins, 2003] Duval, E. & Hodgins, W. (2003). A LOM Research Agenda. WWW2003 Conference, 20-24 May 2003, Budapest, Hungary.
- [Duval *et al.*, 2002] Duval, E.; Hodgins, W.; Sutton, S.; Weibel, S.. “Metadata Principles and Practicalities.” D-Lib Magazine, v.8, n.4, abril de 2002. Disponible en : <http://www.dlib.org/dlib/april02/weibel/04weibel.html>
- [Duval *et al.*, 2011] Erik Duval, Katrien Verbert and Joris Klerkx. Rainbow of Computer Science. Lecture Notes in Computer Science, 2011, Volume 6570/2011, 144-156
- [Edmonds, 1999] Edmonds, B.: Capturing Social Embeddedness: a Constructivist Approach. *Adaptive Behavior* 7, 1999, pp. 323-348.
- [EdNa Online, 2002] EdNa Online. EdNa Metadata Standard v.1.1. 2002 <http://www.edna.edu.au/edna/go/resources/metadata/cache/offonce/pid/261>
- [Esteva *et al.*, 2001] Esteva, M., Rodríguez, J., Sierra, C., Garcia, P., Arcos J.: On the formal specifications of electronic institutions, In: Dignum F., Sierra C. (Eds.) Agent-mediated Electronic commerce (The European AgentLink Perspective), LNAI 1991, Springer, 2001, pp. 126-147.
- [Evans *et al.*, 2001] R. Evans, P. Kearny, J. Stark, G. Caire, F. Garijo, J.J. Gomez-Sanz, F. Leal, P. Chainho, and P. Massonet. MESSAGE: Methodology for Engineering Systems of Software Agents. Technical Report P907, EURESCOM, 2001.

-
- [Ferber *et al.* 2003] Ferber J., Michel F., Gutknecht O., Ferber, J., O. Gutknecht, *et al.* (2003). "Agent/Group/Roles: Simulating with Organizations." Fourth International Workshop on Agent-Based Simulation (ABS03), Montpellier, France.
- [Ferber *et al.*, 2004] Ferber, O. Gutknecht, F. Michel. 2004. From Agents to Organizations: an Organizational View of Multi-Agent Systems, in: P. Giorgini, J. Muller, J. Odell (Eds.), Agent-Oriented Software Engineering VI, Vol. LNCS 2935 of Lecture Notes in Computer Science, Springer-Verlag: 214–230.
- [Ferber *et al.*, 2005] J. Ferber, F. Michel, and J. Baez. Agre: Integrating environments with organizations. In Danny Weyns, H. a. n. Parunak, and Fabien Michel, editors, Environments for Multi-Agent Systems, volume 3374 of LNCS, chapter 2, pages 48–56. Springer Berlin. Heidelberg, Berlin, Heidelberg, 2005.
- [Fielding & Taylor, 2000] Roy T. Fielding and Richard N. Taylor. 2000. Principled design of the modern Web architecture. In Proceedings of the 22nd international conference on Software engineering (ICSE '00). ACM, New York, NY, USA, 407-416.
- [Fielding, 2000] Roy T. Fielding. Architectural styles and the design of network-based software architectures. *PhD Thesis*, University of California, Irvine, 2000.
- [Fischer, 1994] Fischer, M.: A Survey of METATEM, the Language and its Applications. In: Gabbay, D., Ohlbach, H. (Eds.): Proceedings of the 1st. International Temporal Logic Conference, LNAI 827, Springer, 1994, pp. 480 – 505. [Brooks, 1986]
- [Foster *et al.*, 2001] I. Foster, C. Kesselman, and S. Tuecke. 2001. The anatomy of the grid: Enabling scalable virtual organizations, *Int. J. High Perform. Comput. Appl.*, vol. 15, n^o 3: 200-222
- [Franklin y Graesser, 1997] Franklin S. y Graesser A. "Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents," Proceedings of the Agent Theories, Architectures, and Languages Workshop, Berlin: Springer Verlag, 193-206. (1997)
- [Friesen *et al.*, 2002] Friesen, N. Roberts, A. y Fisher, S. CanCore: Metadata for Learning Objects. *Canadian Journal of Learning and Technology*. Volumen 28 (3) /Automne 2002.
- [Friesen, 2004] Friesen, N. Three Objections to Learning Objects. McGreal, R. (Ed). 2004. Online Education Using Learning Objects. Retrieved February 15, 2010, from the World Wide Web: <http://learningspaces.org/n/papers/objections.html>
- [Genesereth y Nilsson, 1987] Genesereth, M., Nilsson, N.: Logical Foundations of Artificial Intelligence, Morgan Kaufman, 1987.
- [Georgeff *et al.*, 1999] Georgeff M., Pollack M. y Tambe M. (1999). The Belief-Desire-Intention Model of Agency. Proceedings of the Fifth International Workshop on Agent Theories, Architectures and Languages (ATAL-98), Lecture Notes in Artificial Intelligence. Springer-Verlag, Heidelberg, 1999.
- [Georgeff y Lansky, 1987] Georgeff M. y Lansky A. L. (1987). Reactive reasoning and planning. Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87). Seattle, WA, 1987.

- [Georgeff y Lansky, 1987] Georgeff M. y Lansky A. L. (1987). Reactive reasoning and planning. Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87). Seattle, WA, 1987.
- [Ghauth *et al.*, 2010] Ghauth, Khairil, Abdullah, Nor. Learning materials recommendation using good learners' ratings and content-based filtering. Educational Technology Research and Development. 2010. Springer Boston. SN 1042-1629. <http://dx.doi.org/10.1007/s11423-010-9155-4>
- [Giampapa y Sycara, 2002] J. A. Giampapa and K. Sycara, "Team-Oriented Agent Coordination in the RETSINA Multi-Agent System," Tech. Report CMU-RI-TR-02-34, Robotics Institute, Carnegie Mellon University, December 2002. Presented at AAMAS 2002 Workshop on Teamwork and Coalition Formation.
- [Giret *et al.*, 2009] Giret, V. Julian, M. Rebollo, E. Argente, C. Carrascosa and V. Botti. 2009. An Open Architecture for Service-Oriented Virtual Organizations. Seventh international Workshop on Programming Multi-Agent Systems. PROMAS 2009: 23-33.
- [Giridhar *et al.*, 2006] Giridhar Manepalli, Henry Jerez, Michael Nelson "FeDCOR: An Institutional CORDRA Registry", D-Lib Magazine , February 2006 Volume 12 Number 2 ISSN 1082-9873
- [Giunchiglia *et al.*, 2002] Giunchiglia, F., Mylopoulos, J., and Perini, A. 2002. The tropos software development methodology: processes, models and diagrams. In Proceedings of the First international Joint Conference on Autonomous Agents and Multiagent Systems: Part 1 (Bologna, Italy, July 15 - 19, 2002). AAMAS '02. ACM, New York, NY, 35-36. DOI= <http://doi.acm.org/10.1145/544741.544748>
- [Glez-Bedia *et al.*, 2002] Glez-Bedia M., Corchado J. M., Corchado E. S. and Fyfe C. (2002) Analytical Model for Constructing Deliberative Agents, Engineering Intelligent Systems, Vol 3: pp. 173-185.
- [Glez-Bedia, 2004] Glez-Bedia M. (2004). Fundamentos cognitivos para el diseño de arquitecturas de agentes planificadores en contextos dinámicos de acción. Tesis Doctoral. Universidad de Salamanca.
- [Goldberg *et al.*, 1992] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, 35(12):61-70, 1992.
- [Golding y Rosenbloom, 1988] Golding A.R. y Rosenbloom P.S. Combining Analytical and Similarity-Based {CBR}, pp. 259-263.1988
- [Gordon and Pathak, 1999] M. Gordon and P. Pathak, Finding information on the World Wide Web: the retrieval effectiveness of search engines, *Information Processing and Management: an International Journal*, 35(2), p.141-180, March 1999
- [Gravano and Ipeirotis, 2001] L. Gravano and P. Ipeirotis. Using q-grams in a dbms for approximate string processing. In IEEE Data Engineering Bulletin 24(4), pages 28--34, 2001.
- [GTI-IA, 2009] GTI-IA. An Abstract Architecture for Virtual Organizations: The THOMAS project. <http://www.fipa.org/docs/THOMASarchitecture.pdf>

-
- [Hamel & Ryan-Jones, 2002] Hamel, C.J. & Ryan-Jones, D. Designing Instruction with Learning Objects. *International Journal of Educational Technology*, 3(1). Recuperado desde. <http://www.ed.uiuc.edu/ijet/v3n1/hamel/index.html>. 2002.
- [Hammond, 1989] Hammond K.J. Case-Based Planning: Viewing Planning as a Memory Task. Academic Press Inc., 1989.
- [Hammond, 2010] Tony Hammond. Case Study in OpenSearch and SRU Integration. D-Lib Magazine, July/August 2010 Volume 16, Number 7/8. Disponible en <http://dlib.org/dlib/july10/hammond/07hammond.html>.
- [Hanafy & Fakhry, 2008] Hanafy, M. Fakhry, M. A Proposed Architecture for Learning Object Selection and Discovery Based on WSMO. In the Proceedings of the 2008 4th International Conference on Next Generation Web Services Practices pp: 10-14.
- [Hawking *et al.*, 2001] Hawking, David and Craswell, Nick and Bailey, Peter and Griffiths, Kathleen. Measuring Search Engine Quality. *Information Retrieval*, Vol. 4 Nº 1, 2001, pp. 33-59.
- [Hill *et al.*, 1995] W. Hill, L. Stead, M. Rosenstein, G. Furnas: Recommending and evaluating choices in a virtual community of use. In: CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM Press/Addison-Wesley Publishing Co. 194-201, 1995.
- [Hillmann, 2007] Hillmann, D. Using Dublin Core. Disponible en <http://dublincore.org/documents/usageguide/>. 2007
- [Hubner, 2004] Hubner, J.F., Sichman, J.S., Boissier, O.: Using the Moise+ for a cooperative framework of mas reorganisation. In: LNAI Proc. of the 17th Brazilian Symposium on Artificial Intelligence (SBIA'04). Volume 3171, 506-515, Springer , 2004.
- [Huhns y Stephens, 1999] Huhns, M., Stephens, L.: Multiagent Systems and Societies of Agents. In: Weiss, G. (Ed.), Multi-agent Systems: a Modern Approach to Distributed Artificial Intelligence, MIT Press, 1999.
- [IEEE 1484.12.1-2002] IEEE 1484.12.1-2002, Draft Standard for Learning Object Metadata. The Institute of Electrical and Electronics Engineers, Inc. Retrieved March 23, 2010, from the World Wide Web: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf. Traducción al español disponible en <http://www.gist.uvigo.es/~lanido/LOMes/> Retrieved November 25, 20210.
- [Iglesias, 2010] José Antonio Iglesias Martínez. Modelado automático del comportamiento de agentes inteligentes. Phd Thesis. Departamento de Informática. Universidad Carlos III de Madrid. 2010.
- [IMS Content Packaking, 2007] IMS Digital Repositories Interoperability. Kraan, W., Poston Day, J., Ward, N. IMS Content Packaging Information Model. Versión 1.2 (Public Draft v2.0). 2007.
- [IMS DRI, 2003] IMS Digital Repositories Specification. V.1 2003. <http://www.imsglobal.org/digitalrepositories/>

- [IMS GWS, 2005] IMS General Web Service Services Primer. Colin Smythe. Version 1.0 . 19 December 2005
- [ISO/IEC 19788, 2010] ISO/IEC JTC1 SC36 Working Group 4: Management and Delivery of Learning, Education, and Training. Information technology -- Learning, education and training -- Metadata for learning resources -- Part 1: Framework 2010. http://www.iso.org/iso/catalogue_detail.htm?csnumber=50772
- [Järvelin & Kekäläinen, 2000] Järvelin, K., & Kekäläinen, J. (2000). IR evaluation methods for retrieving highly relevant documents. In Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 41–48). New York: ACM Press.
- [Jennings y Wooldridge, 1998] Jennings, N., Wooldridge, M. (eds.), Agent Technology: Foundations, Applications and Markets, Springer, 1998.
- [Jiang and Conrath, 1997] Jay J. Jiang, David W. Conrath. Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy [CoRR cmp-lg/9709008](http://www.corr.cmp.lg/9709008): (1997)
- [JORUM+ project, 2004] The JISC Online Repository for [learning and teaching] Materials. 2004
- [Juan y Sterling, 2002] Juan, T.P., A.; Sterling, L. Roadmap: Extending the Gaia methodology for complex open systems. in 1st Int. Joint Conference on Autonomous Agents and multiagent Systems (AAMAS 2002). 2002.
- [Kärger *et al.*, 2006] P. Kärger, C. Ullrich and E. Melis, Integrating learning object repositories using a mediator architecture, Proceedings of EC-TEL LNCS (4227) (2006), pp. 185–197.
- [Katakis *et al.*, 2008] I., Katakis, G., Tsoumakas, I., Vlahavas, Multilabel Text Classification for Automated Tag Suggestion, Proceedings of the ECML/PKDD 2008 Discovery Challenge, Antwerp, Belgium, (2008).
- [Kephart y Walsh, 2004] Kephart J.O. y Walsh W.E. An Artificial Intelligence Perspective on Autonomic Computing Policies. POLICY '04: Proceedings of the Fifth IEEE International Workshop on Policies for Distributed Systems and Networks, pp. 3. IEEE Computer Society. 2004
- [Kerkiri *et al.*, 2007] Kerkiri, T., Manitsaris, A., & Mavridou, A. (2007). Reputation metadata for recommending personalized e-learning resources. In Proceedings of the second international workshop on semantic media adaptation and personalization (pp. 110–115). Uxbridge.
- [Khribi *et al.*, 2009] Khribi, M. K., Jemni, M., & Nasraoui, O. (2009). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. Educational Technology & Society, 12(4), 30–42.
- [Kleinstein & Seiden, 2000] Kleinstein, S.H.; Seiden, P.E.; , "Simulating the immune system," Computing in Science & Engineering , vol.2, no.4, pp.69-77, Jul/Aug 2000
- [Kolodner, 1993] Kolodner J. (1993). Case-Based Reasoning. San Mateo, Morgan Kauffman.
- [Konstan *et al.*, 1997] J.A. Konstan, B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, J. Riedl: GroupLens: Applying Collaborative Filtering to Usenet News. 77-87, 1997.

-
- [Korfhage, 1997] Korfhage, R.R. (1997). *Information storage and retrieval*. New York: Wiley.
- [Kottler *et al.*, 2000] Kottler, H., Parsons, J., Wardengurg, S. & Vornbrock, F. Knowledge Objects: Definition, Development Initiatives, and Potencial Impact. New York: McGraw-Hill. 2000.
- [Labarga *et al.*, 2007] Alberto Labarga, Franck Valentin, Mikael Anderson, and Rodrigo López. Web Services at the European Bioinformatics Institute. *Nucleic Acids Res.* 2007 July; 35(Web Server issue): W6–W11. Published online 2007 July. doi: [10.1093/nar/gkm291](https://doi.org/10.1093/nar/gkm291).
- [Labidi y Lejouad, 1993] Labidi, S., Lejouad, W. (1993) Del'intelligence artificielle distribu'eeaux syst'emes multiagents
- [L'Allier, 1997] L'Allier, J. J. A Frame of Reference: NETg's Map to Its Products, Their Structures and Core Beliefs. 1997.
- [L'Allier, 1998] L'Allier, J. J. (1998). NETg's precision skilling: The linking of occupational skills descriptors to training interventions: <http://www.netg.com/research/pskillpaper.htm>
- [Leake *et al.*, 2000] Leake D.B., Bauer T., Maguitman A. y Wilson D.C. Capture, Storage and Reuse of Lessons about Information Resources: Supporting Task-Based Information Search. Proceedings of the AAAI-00 Workshop on Intelligent Lessons Learned Systems. Austin, Texas, pp. 33-37. 2000
- [Lee *et al.*, 2001] C.H. Lee, Y.H. Kim, P.K. Rhee: Web personalization expert with combining collaborative filtering and association rule mining technique. *Expert Systems Appl.* 21(3), 131-137, 2001.
- [Lee *et al.*, 2008] Ming-Che Lee, Kun Hua Tsai, Tzone I. Wang. A practical ontology query expansion algorithm for semantic-aware learning objects retrieval, *Computers & Education*, Volume 50, Issue 4, May 2008, Pages 1240-1257.
- [Lemire *et al.*, 2005] Lemire, D., Boley, H., McGrath, S., and Ball, M. Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation. *Technology and Smart Education*, 2(3). 2005. pp. 179-188.
- [Lésperance *et al.*, 1996] Lésperance, Y., Levesque, H., Lin, F., Marcu, D., Reiter, R., Scherl, R.: Foundations of a Logical Approach to Agent Programming. In: Wooldridge, M., Müller, J., Tambe, M. (Eds.): *Intelligent Agents II*, LNAI 1037, Springer, 1996, pp. 331 - 346.
- [Li & Ogihara, 2003] L., Li, M., Ogihara, M., Detecting emotion in music. In: Proceedings of the International Symposium on Music Information Retrieval, Washington D.C., USA (2003) 239-240.
- [Li *et al.*, 2009] Yushun Li; Zheng Chen; Ronghuai Huang; Xiaochun Cheng; , "An Integrated Learning Resource Management System with Web Services," *New Trends in Information and Service Science*, 2009. NISS '09. International Conference on New Trends in Information and Service Science , vol., no., pp.863-868, June 30 2009-July 2 2009 doi: 10.1109/NISS.2009.159 URL:

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5260804&isnumber=5260409>

- [LODE, 2010] David Massart, Nick Nicholas, Nigel Ward. IMS GLC Learning Object Discovery & Exchange (LODE). Version 1.0, (Base Document) Draft 14, 2 de Marzo de 2010.
- [López de Mántaras *et al.*, 2005] López De Mántaras R., McSherry D., Bridge D., Leake D., Smyth B., Craw S., Faltings B., Maher M.L., Cox M.T., Forbus K., Keane M., Aamodt A. y Watson I. Retrieval, reuse, revision and retention in case-based reasoning. *The Knowledge Engineering Review*, vol. 20 (3), pp.215-240. 2005
- [López de Mántaras y Plaza, 1997] López de Mántaras R., Plaza E. Case-Based Reasoning: An overview. *AI Communications Journal* 10(1) (1997) 21-29.
- [López de Mántaras, 2001] Ramón López de Mántaras. Case-Based Reasoning. *Machine Learning and Its Applications 2001*: 127-145
- [López-Batista *et al.*, 2011] López-Batista V.F., Prieta Pintado F., Gil A.B., Rodríguez S. and Moreno M.N.. A System for Multi-label Classification of Learning Objects. *Proceedings Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011. Series: Advances in Intelligent and Soft Computing, Vol. 87, 523-531. Corchado, E. et al.(Eds.). ISBN: 978-3-642-19643-0, 2011.*
- [Lujara *et al.*, 2007] Lujara, S.K., Kissaka, M.M., Bhalaluseca, E.P. and trojer L. Learning Objects: A new paradigm for e-learning resource development for secondary schools in Tanzania. *World Academy or Scienzie, Engineering and Technology*. Pp. 102-106, 2007
- [Maes *et al.*, 1999] Pattie Maes, Robert H. Guttman, and Alexandros G. Moukas. 1999. Agents that buy and sell. *Commun. ACM* 42, 3 (March 1999), 81-ff. DOI=10.1145/295685.295716 <http://doi.acm.org/10.1145/295685.295716>
- [Maes, 1989] Maes, P. How to Do the Right Thing. A.I. Memo No. 1180, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, December, 1989
- [Maes, 1990] Maes, P. Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back, In P. Maes (ed.), *Designing Autonomous Agents*, 1-2. Cambridge, MA: MIT Press. 1990.
- [Maes, 1994] Pattie Maes. 1994. Agents that reduce work and information overload. *Commun. ACM* 37, 7 (July 1994), 30-40. DOI=10.1145/176789.176792 <http://doi.acm.org/10.1145/176789.176792>
- [Maimon and Rokach, 2010] Oded Maimon and Lior Rokach (Ed.). *Data Mining and Knowledge Discovery Handbook. 2nd Edition*, Springer, 2010.
- [Malone y Crowston, 1994] Malone, T., Crowston, K.: *The Interdisciplinary Study of Coordination*. *ACM Computing Surveys* 26(1), March, 1994.
- [Malone, 1988] T. W. Malone. What is coordination theory? In *National Science Foundation Coordination Theory Workshop*. MIP, EE.UU., 1988.
- [Manouselis & Costopoulou, 2007] Manouselis, N. and Costopoulou, C. Experimental Analysis of Design Choices in Multi-Attribute Utility Collaborative Filtering.

-
- International Journal of Pattern Recognition and Artificial Intelligence, 21(2), 2007. pp. 311-331
- [Manouselis *et al.*, 2007] Manouselis, N., Vuorikari, R., and Van Assche, F. Simulated Analysis of Collaborative Filtering for Learning Object Recommendation. SIRTEL Workshop, EC-TEL 2007.
- [Manouselis *et al.*, 2010a] Manouselis N., Drachsler H., Vuorikari R., Hummel H. & Koper R. (2010) Recommender systems in technology enhanced learning. In Recommender Systems Handbook: A Complete Guide for Research Scientists & Practitioners (eds P.Kantor, F.Ricci, L.Rokach & B.Shapira), pp. 387-415. Springer, Berlin.
- [Manouselis *et al.*, 2010b] Manouselis, N., Vuorikari, R. and Van Assche, F. (2010), Collaborative recommendation of e-learning resources: an experimental investigation. Journal of Computer Assisted Learning, 26: 227-242
- [Manouselis *et al.*, 2011] Manouselis, N., Drachsler, H., Vuorikari, R., Hummel H. and Koper R. Recommender Systems in Technology Enhanced Learning. Recommender Systems Handbook, pages 387-415. Springer, 2011
- [Mansury *et al.*, 2002] Y. Mansury, M. Kimura M. J. Lobo, T.S. Deisboeck. Emerging Patterns in Tumor Systems Simulating the dynamics of multicelular clusters with an agent-based spatial agglomeration model. J Theor Biol. 219(3), 343-370, 2002.
- [Margain *et al.*, 2010] Ma. de Lourdes Margain Fuentes, Jaime Muñoz Arteaga, Francisco Álvarez Rodríguez, Jean Vanderdonk and Michael Orey. MIRROS: Intermediary Model to Recovery Learning Objects. Computación y Sistemas Vol. 13 No. 4, 2010, pp 373-384 ISSN 1405-5546
- [Mas, 2005] Mas, A. Agentes software y sistemas multiagente: conceptos, arquitecturas y aplicaciones. ISBN 84-205-4367-5, pp. 29-64. 2005
- [Massart *et al.*, 2010] Massart, D., Shulman E., Nicholas, N., Ward, N., Bergeron, F. (2010). Taming the Metadata Beast: ILOX. IMS Global Learning Consortium Series on Learning Impact. November 2010 from <http://www.imsglobal.org/articles/nov2010ilox.cfm>
- [Mauro, 2003] P.P. Mauro "The Emergence of Social Learning in Artificial Societies" in Applications of Evolutionary Computing: EvoWorkshops 2003: EvoBIO, EvoCOP, EvoIASP, EvoMUSART, EvoROB, and EvoSTIM, Essex, UK, pp. 467-478, 2003.
- [Mayorga *et al.*, 2010] José I. Mayorga, Juan Cigarrán, Miguel Rodríguez-Artacho. Retrieval and Clustering of Web Resources Based on Pedagogical Objectives. Actas de los Talleres de las Jornadas de Ingeniería del Software y Bases de Datos, Vol. 4, N° 3, pp. 124-130, 2010.
- [McCalla, 2004] McCalla, G. The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of Information about Learners. Journal of Interactive Media in Education, 2004 (7) Special Issue on the Educational Semantic Web. Volume 1, p. 18, 2004.
- [McCallum, 1999] A., McCallum. Multi-label text classification with a mixture model trained by em. In: Proceedings of the AAAI 99 Workshop on Text Learning. (1999).

- [Menendez, and Prieto, 2008] Menendez, Victor H. and Prieto, Manuel E. A Learning Object Composition Model. Information Systems and e-Business Technologies. Lecture Notes in Business Information Processing, pp. 469-474, 2008.
- [MESSAGE, 2000] MESSAGE: Methodology for Engineering Systems of Software Agents. Deliverable 1. Initial Methodology. July 2000. EURESCOM Project P907-GI
- [Mican *et al.*, 2010] Mican Daniel, Tomai Nicolae, Daniel Florian and Facca Federico. Association-Rules-Based Recommender System for Personalization in Adaptive Web-Based Applications Current Trends in Web Engineering. Lecture Notes in Computer Science Vol. 6385. Pages 85-90. 2010
- [Mika, 2005] P., Mika. Ontologies are us: A unified model of social networks and semantics. In Proc. Of ISWC05, 2005.
- [MLkNN, 2010] Algoritmo de clasificación MLkNN (Multi-Label K-Nearest Neighbor) <http://lamda.nju.edu.cn/datacode/MLkNN.htm>
- [Montaner *et al.*, 2002] Miquel Montaner, Beatriz López, and Josep Lluís de la Rosa. 2002. Opinion-Based Filtering through Trust. In Proceedings of the 6th International Workshop on Cooperative Information Agents VI (CIA '02), Matthias Klusch, Sascha Ossowski, and Onn Shehory (Eds.). Springer-Verlag, London, UK, 164-178.
- [Moreno & Bailly-Baillièrè, 2002] Moreno, F. , Bailly-Baillièrè M. Diseño instructivo de la formación on-line. Aproximación metodológica a la elaboración de contenidos. Barcelona: Editorial Ariel Educación. 2002.
- [MPEG-7, 2010] MPEG-7, Retrieved March 29 from MPEG Home Page <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>
- [Najjar *et al.*, 2003] Najjar, J., Duval, E., Ternier S., & Neven, F. (2003). Towards Interoperable Learning Object Repositories: the ARIADNE Experience. In *IADIS WWW/Internet 2003 Conference*, pp. 219-226.
- [Nesbit et al, 2002] Nesbit, J. C., Belfer, K., & Vargo, J. (2002). A convergent participation model for evaluation of learning objects. *Canadian Journal of Learning and Technology*, 28 (3), 105-120.
- [Norman *et al.*, 2003] Norman, T. J., Preece, A., Chalmers, S., Jennings, N. R., Luck, M., Dang, V. D., Nguyen, T. D., Deora, V., Shao, J., Gray, A. and Fiddian, N. (2003) CONOISE: Agent-based formation of virtual organisations. In: 23rd SGA International Conference on Innovative Techniques and Applications of AI, 2003, Cambridge, UK. pp. 353-366.
- [Norman *et al.*, 2004] Timothy J. Norman, Alun Preece, Stuart Chalmers, Nicholas R. Jennings, Michael Luck, Viet D. Dang, Thuc D. Nguyen, Vikas Deora, Jianhua Shao, W. Alex Gray, Nick J. Fiddian, Agent-based formation of virtual organisations, *Knowledge-Based Systems*, Volume 17, May 2004, Pages 103-111.
- [Nwana, 1995] Nwana H. S. Software Agents: An Overview. *Knowledge Engineering Review*, vol. 11 (2), pág 205-244. (1995)
- [OAI, 2004] *The open archives initiative protocol for metadata harvesting version 2.0*, Document Version 2004/10/12T15:31:00Z, C. Lagoze, H. Van de Sompel, M. Nelson,

- & S. Warner. Disponible en <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>, 2002.
- [OAI-PMH, 2008] The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14. Document Version 2008-12-07T20:42:00Z. Disponible on-line: <http://www.openarchives.org/OAI>
- [OAIS, 2002] Reference Model for an Open Archival Information System (OAIS). CCSDS 650.0-B-1, Blue Book, January 2002. Disponible en <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- [Ochoa&Duval, 2006] Ochoa, X. and Duval, E. Use of Contextualized Attention Metadata for Ranking and Recommending Learning Objects. Proceedings of 1st International Workshop on Contextualized Attention Metadata: Collecting, Managing and Exploiting of Rich Usage Information. 2006. pp. 9-16.
- [Ossowski y García Serrano, 1997] Ossowski S. and Garcia-Serrano A. (1997). Social coordination of autonomous problem-solving agents, *Agents and Multiagent Systems: Formalisms, Methodologies and Applications*. pp: 134-149, 1997 Springer Verlag, LNAI 1441 ISBN: 3-540-64769-4.
- [Ossowski y Menezes, 2006] Ossowski S. y Menezes R. On coordination and its significance to distributed and multi-agent systems. *Concurrency and Computation: Practice and Experience* 18(4), pp. 359-370. 2006
- [Padgham and Winikoff, 2002] Padgham, L. and Winikoff, M. 2002. Prometheus: a methodology for developing intelligent agents. In Proceedings of the First international Joint Conference on Autonomous Agents and Multiagent Systems: Part 1 (Bologna, Italy, July 15 - 19, 2002). AAMAS '02. ACM, New York, NY, 37-38. DOI=<http://doi.acm.org/10.1145/544741.544749>
- [Paquette *et al.*, 2006] Gilbert Paquette, Michel Léonard, Karin Lundgren-Cayrol, Stefan Mihaila, Denis Gareau. Learning Design based on Graphical Knowledge-Modelling. *Journal of Educational technology and Society ET&S*, 9 (1), 97-112, 2006.
- [Parunak y Odell, 2002] Parunak, H. V. D., Odell, J.: Representing Social Structures in UML. In: Wooldridge, M., Weiss, G., Ciancarini P. (Eds.): *Agent-Oriented Software Engineering II*, LNCS 2222, Springer, 2002.
- [Pattison *et al.*, 1987]. H. E. Pattison, D. D. Corkill, and V. R. Lesser. Distributed Artificial Intelligence, chapter Instantiating Descriptions of Organizational Structures, pages 59-96. Pitman Publishers, 1987.
- [Pavón and Gómez-Sanz, 2003] J. Pavón and J. Gómez-Sanz. Agent Oriented Software Engineering with. *INGENIAS. Multi-Agent Systems and Applications III*, 2691:394-403, 2003
- [Pavón *et al.*, 2005] J. Pavon, J. Gomez-Sanz, R. Fuentes. 2005. The *INGENIAS Methodology and Tools*, Idea Group Publishing, article IX: 236-276.
- [Pavón *et al.*, 2008] Juan Pavón, Candelaria Sansores, and Jorge J. Gómez-Sanz. 2008. Modelling and simulation of social systems with *INGENIAS*. *Int. J. Agent-Oriented Softw. Eng.* 2, 2 (February 2008)

- [Peiro, 1990] Peiró, J. M.: «Las Nuevas Tecnologías» en Organización. Nuevas perspectivas psicosociales, Barcelona, PPU S.A. 1990.
- [Picard and Gleizes, 2004] Picard, G. and Gleizes, M.-P. The ADELFE Methodology – Designing Adaptive Cooperative Multi-Agent Systems. In Bergenti, F. and Gleizes, M.-P. and Zambonelli, F., editor, Methodologies and Software Engineering for Agent Systems. Kluwer Publishing, 2004.
- [Pokahr *et al.*, 2003] Pokahr, A., Braubach, L. y Lamersdorf W. (2003) Jadex: Implementing a BDI-Infrastructure for JADE Agents, in: EXP - In Search of Innovation (Special Issue on JADE), Vol 3, Nr. 3, Telecom Italia Lab, Turin, Italy, September 2003, pp. 76-85.
- [Polsani, 2003] Polsani, P. Use and Abuse of Reusable Learning Objects. Journal of Digital Information, vol 3, iss 4, artº 164, 2003.
- [Popper, 1982] Popper, K.: The Open Universe: An Argument for Indeterminism, Hutchinson, London, 1982.
- [Porter *et al.*, 2004] Porter D., Curry J, Muirhead B. & Galan, N. A Report on Learning Object Repositories CANARIE Inc. 2002
- [Prieta & Corchado, 2010] F. de la Prieta, J.M. Corchado. Recuperación y Catalogación de Recursos Educativos Mediante Organizaciones Virtuales y Filtrado Colaborativo. Máster Sistemas Inteligentes Avances en Informática y Automática (4º Workshop), pp. 181-200, 2010.
- [Prieta & Gil, 2010] F. de la Prieta and A.B. Gil. *A Multi-agent System that Searches for Learning Objects in Heterogeneous repositories*. Advances in Intelligent and Soft Computing (Springer). Trends in Practical Applications of Agents and Multiagent systems: 8th International conference on Practical Applications of agents and multiagent systems (PAAMS 2010). Editors: Yves Demazeau *et al.*, pp. 355-362, Springer, Heidelberg (2010)
- [Qiyán, 2010] Qiyán Han; Feng Gao; Hu Wang, "Ontology-based learning object recommendation for cognitive considerations," 8th World Congress on Intelligent Control and Automation (WCICA), 2010, vol., no., pp.2746-2750, 7-9 July 2010
- [Qu & Nejdí, 2004] Changtao Qu, Wolfgang Nejdí. Interacting the Edutella/JXTA Peer-to-Peer Network with Web Services. SAINT 2004: 67-73.
- [Rao y Georgeff, 1991] Rao, A. S., & Georgeff, M. P. (1991). Modeling rational agents within a BDI-Architecture. In J. Allen, R. Fikes, & E. Sandewall (Ed.), Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning ({KR}'91) (pp. 473-484). San Mateo, CA, USA: Morgan Kaufmann publishers, Inc
- [Rao y Georgeff, 1992] Rao, A., Georgeff, M.: An Abstract Architecture for Rational Agent. In: C. Rich, W. Swartout, B. Nebel (Eds.): Proc. of KR'92, Morgan Kaufmann, 1992, pp. 439 - 449.
- [Rao y Georgeff, 1998] Rao A. S. and Georgeff M. P. (1998). Decision procedures of BDI. Logics. Journal of logic and computation 8(3). 1998.

-
- [Razavi *et al.*, 2005] Razavi, R., Perrot, J., & Guelfi, N. (2005). Adaptive modeling: an approach and a method for implementing adaptive agents. *Lecture Notes in Artificial Intelligence*, 3446, 136-148.
- [Recker & Wiley, 2001] Recker, M. and Wiley, D. (2001) A non-authoritative educational metadata ontology for filtering and recommending learning objects. *Journal of Interactive Learning Environments*. Special issue on metadata, pp. 1-17.
- [Recker *et al.*, 2003] Recker, M., Walker, A., and Lawless, K. What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. *Instructional Science*, 31(4-5), 2003, pp. 299-316.
- [[Reitbauer *et al.*, 2004] 57. Reitbauer, A., Battino, A., Karageorgos, A., Mehandjiev, P., Valckenaers, P., & Saint-Germain, B. (2004). The MaBE middleware: extending multi-agent systems to enable open business collaboration. *International Conference on Information Technology for Balanced Automation Systems in Manufacturing and Services - BASYS04*
- [Richards, 2002] Richards, G. The Challenges of the Learning Object Paradigm. *Canadian journal of learning and technology*. 2002, VOL 28; PART 3, pages 3-10.
- [Riesbeck y Schank, 1989] Reisbeck C., Schank R. *Inside Case-Based Reasoning*. Lawrence Erlbaum Ass. Hillsdale, New Jersey, 1989.
- [RLG, 2002] Research Libraries Group. *Trusted Digital Repositories: Attributes and Responsibilities*.
- [Robson, 1999] Robson, R. Object-oriented Instructional Design and Applications to the Web. In B. Collis & R. Oliver (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 1999* (pp. 698-702). Chesapeake, VA: AACE.
- [Rodríguez *et al.*, 2009] Rodríguez S., Pérez-Lancho B., De Paz J.F., Bajo J. and Corchado J.M.: *Ovamah: Multiagent-based Adaptive Virtual Organizations*. 12th International Conference on Information Fusion, Seattle, Washington, USA. Julio 2009
- [Rodríguez, 2010] Rodríguez S. *Modelo Adaptativo para Organizaciones Virtuales de Agentes*. Tesis Doctoral, Universidad de Salamanca. 2010.
- [Roth & Fischer, 2006] V., Roth, B., Fischer. Improved functional prediction of proteins by learning kernel combinations in multilabel settings. In: *Proceeding of 2006 Workshop on Probabilistic Modeling and Machine Learning in Structural and Systems Biology (PMSB 2006)*, Tuusula, Finland (2006).
- [Ruiz-Iniesta *et al.*, 2010] A. Ruiz-Iniesta, G. Jiménez-Díaz, M. Gómez-Albarrán. Personalización en Recomendadores Basados en Contenido y su Aplicación a Repositorios de Objetos de Aprendizaje. *IEEE-RITA Vol. 5*, pp. 31-38, Núm 1, Feb. 2010.
- [Ruqayya *et al.*, 2011] Abdulrahman, Ruqayya, Neagu, Daniel, Holton, D. Multi Agent System for Historical Information Retrieval from Online Social Networks. In *Agent and Multi-Agent Systems: Technologies and Applications Volume 6682*, pp. 54-63. *Lecture Notes in Computer Science*, 2011. Doi: 10.1007/978-3-642-22000-5_7

- [Russel y Norvig, 1995] Russell S. and Norvig P. (1995). *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
- [Russell y Norvig, 2009] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach* (Third edition). Prentice-Hall, Saddle River, NJ, 2009.
- [Sampath& Prakash, 2009] B.T. Sampath Kumar and J.N. Prakash. Precision and Relative Recall of Search Engines: A Comparative Study of Google and Yahoo. *Singapore Journal of Library & Information Management*. Volume 38, pp. 124-137, 2009
- [Santiago & Raabe, 2010] Rafael de Santiago, Andre L.A. Raabe, "Architecture for Learning Objects Sharing among Learning Institutions—LOP2P," *IEEE Transactions on Learning Technologies*, Vol. 3, No. 2. pp. 91-95, April-June, 2010
- [Santos *et al.*, 2010] Santos, A.M.; Canuto, A.M.P.; Neto, A.F.; , "Evaluating classification methods applied to multi-label tasks in different domains," *Hybrid Intelligent Systems (HIS), 2010 10th International Conference on* , vol., no., pp.61-66, 23-25 Aug. 2010.
- [Sarwar *et al.*, 2001] B. Sarwar, G. Karypis, J. Konstan, J. Riedl: Item-based Collaborative Filtering Recommendation Algorithms. *Proceedings International WWW Conference(10)*, Hong-Kong, 2001.
- [Schamber & Bateman, 1996] Schamber, L. and Bateman, J. (1996), "User criteria in relevance evaluation: toward development of a measurement scale", in Hardin, S. (Ed.), *Global Complexity: Information, Chaos and Control*, Proceedings of the 59th Annual Meeting of the American Society for Information Science, Baltimore, MD, October 21-24, Learned Information, Medford, NJ, pp. 218-25.
- [Schank, 1982] Roger Schank. *Dynamic Memory: A Theory of Learning in Computers and People* (New York: Cambridge University Press, 1982)
- [Schapire and Singer, 2000] R. E., Schapire and Y., Singer. Boostexter, A boosting-based system for text categorization. *Machine Learning*, 39(2/3):135 - 168, (2000)
- [Segura *et al.*, 2010] Segura, N.A.; Vidal, C.C.; Prieto, M.M. "Query expansion based on domain ontology for Learning Objects search," *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on* , vol.1, no., pp.446-451, 9-11 July 2010
- [Segura *et al.*, 2011] N. Alejandra Segura, Salvador-Sanchez, Elena Garcia-Barriocanal, Manuel Prieto, An empirical analysis of ontology-based query expansion for learning resource searches using MERLOT and the Gene ontology, *Knowledge-Based Systems*, Volume 24, Issue 1, February 2011, Pages 119-133
- [Sheth *et al.*, 1990] Amit P. Sheth and James A. Larson. Federated Database Systems for Managing Distributed, Heterogenous, and Autonomous Databases *ACM Computing Surveys*, 22(3):183- 236, September 1990.
- [Shih *et al.*, 2006] T.K. Shih, C.-C. Chang and H.W. Lin, Reusability on learning object repository. *Proceedings of ICWL LNCS (4181) (2006)*, pp. 203-214.
- [Sicilia *et al.*, 2005] Miguel-Ángel Sicilia, Elena García, Carmen Pagés, José-Javier Martínez and José-maría Gutiérrez. Complete Metadata Records in Learning Object Repositories: Some Evidence and Requirements. *International Journal of Learning Technology*, 2005. Volume 1, pp. 411—424.

-
- [Simon et al, 2005] Simon B., Massart D., van Assche F., Ternier S., Duval E., Brantner S., Olmedilla. D.A Simple Query Interface for Interoperable Learning Repositories. Workshop on Interoperability of Web-Based Educational Systems in conjunction with 14th International World Wide Web Conference (WWW'05). May, 2005, Chiba, Japan
- [SOAP, 2007] *Simple Object Access Protocol*, SOAP Version 1.2. W3C Recommendation 27 April 2007. Disponible en <http://www.w3.org/TR/soap/>
- [Sosteric et al., 2002] Sosteric, M. and Hesemeier, S. When is a Learning Object not an Object: A First Step towards a theory of learning objects. *International Review of Research in Open and Distance Learning*, vol. 3, no. 2, Oct. 2002.
- [Soto et al., 2007] Jesús Soto, Elisa García, Salvador Sánchez-Alonso. Semantic learning object repositories. *International Journal of Continuing Engineering Education and Life Long Learning* Issue: Volume 17, Number 6 / 2007 Pages: 432 – 446,
- [Spencer, 1896] Herbert Spencer. *The Study of Sociology*. New York. D. Appleton and Company. 1896
- [SPI, 2008] *A simple publishing interface for learning object repositories*, S. Ternier, D. Massart, F. Van Assche, N. Smith, B. Simon, & E. Duval. Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2008 (Vienna, Austria), AACE, June 2008, pp. 1840-1845. <http://www.editlib.org/p/28625>
- [SQI, 2005] *A simple query interface specification for learning repositories*, CEN Workshop Agreement (CWA 15454). B. Simon, D. Massart, F. Van Assche, S. Ternier, & E. Duval. Disponible en <ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/WS-LT/CWA15454-00-2005-Nov.pdf>, November 2005.
- [SRU, 2008] SRU Version 1.2 Specifications, Library of Congress. <http://www.loc.gov/standards/sru>
- [SRU, 2010] Search/Retrieve via URL. Draft SRU 2.0 December 9, 2010. Disponible en <http://www.loc.gov/standards/sru/>
- [Stockley y Laverty, 2003] Stockeley, D. y Laverty, C. The role of the librarian in learning object design. Proceedings of 2003 Intl. MERLOT Conf., Vancouver. Canada
- [Stranjak et al., 2008] Armin Stranjak, Partha Sarathi Dutta, Mark Ebden, Alex Rogers, and Perukrishnen Vytelingum. 2008. A multi-agent simulation system for prediction and scheduling of aero engine overhaul. In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: industrial track (AAMAS '08). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 81-88.
- [Sutton, 2004] Sutton, S.A. Gateway to Educational Material (GEM): metadata for networked information discovery and retrieval. <http://www.ra.ethz.ch/CDstore/www7/1897/com1897.htm>.
- [Sycara, 2001] Katia Sycara, *Multi-agent infrastructure, agent discovery, middle agents for Web services and interoperation*, Multi-agents systems and applications, Springer-Verlag New York, Inc., New York, NY, 2001

- [Tang&McCalla, 2003] Tang, T. Y. and McCalla, G. I. Smart Recommendation for an Evolving E-Learning System. In AIED 2003 Workshop on Technologies for Electronic Documents for Supporting Learning.
- [Tang&McCalla, 2004] Tang, T. Y. and McCalla, G. Utilizing Artificial Learner on the Cold-Start Pedagogical-Value based Paper Recommendation. In Proceedings of AH 2004: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems.
- [Taylor, 1911] Frederick Winslow Taylor. Principios y métodos de gestión científica, 1911.
- [Totschnig *et al.*, 2010] Totschnig, M., Derntl, M., Gutiérrez, I., Najjar, J., Klemke, R., Klerkx, J., Duval, E., & Müller, F. (2010). Repository Services for Outcome-based Learning. Fourth International Workshop on Search and Exchange of e-le@rning Materials (SE@M'10). September, 27-28, 2010, Barcelona, Spain.
- [Trohidis *et al.*, 2008] K., Trohidis, G., Tsoumakas, G., Kalliris, I., Vlahavas. Multilabel classification of music into emotions. In: Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR). (2008).
- [Tsai *et al.*, 2006] Tsai, K. H., Chiu, T. K., Lee, M. C., and Wang, T. I. A learning Object Recommendation Model based on the Preference and Ontological Approaches. Proceeding of the Sixth International Conference on Advanced Learning Technologies (ICALT'06) 2006.
- [Tsoumakas & Katakis, 2007] G., Tsoumakas, I., Katakis, I, Multi-label classification: An overview. International Journal of Data Warehousing and Mining 3 (2007) 1-13
- [Tsoumakas & Vlahavas, 2007] G., Tsoumakas, I., Vlahavas. Random k-labelsets: An ensemble method for multilabel classification. In: Proceedings of the 18th European Conference on Machine Learning (ECML 2007), Warsaw, Poland (2007) 406-417.
- [Ueda & Saito, 2003] N. Ueda, K. Saito, Parametric mixture models for multi-label text, in: S. Becker, S. Thrun, K. Obermayer (Eds.), Advances in Neural Information Processing Systems, vol. 15, MIT Press, Cambridge, MA, 2003, pp. 721-728.
- [Vargo *et al.*, 2003] Vargo, J., Nesbit, J. C., Belfer, K., & Archambault, A. (2003). Learning object evaluation: Computer mediated collaboration and inter-rater reliability. International Journal of Computers and Applications, 25 (3), 198-205.
- [Verbert y Duval, 2004] K. Verbert and E. Duval, Towards a global architecture for learning objects: a comparative analysis of learning object content models, in Proceedings of the ED-MEDIA 2004 World Conference on Educational Multimedia, Hypermedia and Telecommunications, pages 202-208, AACE, 2004
- [Verbert y Duval, 2007] K. Verbert and E. Duval. Evaluating the ALOCOM Approach for Scalable Content Repurposing. Second European Conference on Technology Enhanced Learning (ECTEL) – Creating new learning experiences on a global scale, 17-20 September 2007, Crete, Greece.
- [Wagner, 2000] Wagner, G. Agent-Oriented Analysis and Design of Organizational Information Systems. Proceedings of the 4th IEEE International Baltic Workshop on Databases and Information Systems, May 2000.

-
- [Wang *et al.*, 2007] Wang, T. I., Tsai, K. H., Lee, M. C., and Chiu, T. K. Personalized Learning Objects Recommendation based on the Semantic Aware Discovery and the Learner Preference Pattern. *Educational Technology and Society*, 10 (3), 2007. pp. 84-105.
- [Wettschereck *et al.*, 2007] Dietrich Wettschereck David W. Aha Takao Mohri. A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms. *Artificial Intelligence Review*, volume 11, pp. 273—314, 1997.
- [Weyns & Georgeff , 2010] Weyns, D.; Georgeff, M. . Self-Adaptation Using Multiagent Systems. *Software, IEEE Volume: 27, Issue: 1, 2010 , Page(s): 86 - 91*
- [Weyns *et al.*, 2004] Weyns, D., Schelfhout, K., Holvoet, T., & Glorieux, O. (2004). Role based model for adaptive agents. Forth Symposium on Adaptive Agents and Multiagent Systems at the AISB04 Convention
- [Wiley, 2000] Wiley, D. A.. Connecting learning objects to instructional design theory: A definition a metaphor, and a taxonomy <http://reusability.org/read/chapters/wiley.doc>. 2000.
- [Wiley, 2001] Wiley, D.A. Connecting learning objects to instructional design theory: A definition a metaphor, and a taxonomy. D. A. Wiley (Ed), *The Instructional Use of Learning Objects*. Bloomington, IN: Association for Educational Communications and Technology, USA, 2001. Retrieved March 23, 2010, from the World Wide Web: <http://www.reusability.org/read/>
- [Wolpers *et al.*, 2007] Wolpers, M., Najjar, J., and Duval, E. (2007). Tracking Actual Usage: the Attention Metadata Approach. *Educational Technology & Society*, 10(3), pp106-121.
- [Wooldridge *et al.*, 2000] Wooldridge, M., Jennings, N. R. and Kinny, D. (2000) *The Gaia Methodology for Agent-Oriented Analysis and Design*. *Journal of Autonomous Agents and Multi-Agent Systems*, 3 (3). pp. 285-312.
- [Wooldridge y Jennings, 1995] Wooldridge M. y Jennings N. R. (1995). *Intelligent Agents: Theory and Practice*. *The Knowledge Engineering Review*, vol. 10(2) pp. 115-152, 1995.
- [Wooldridge, 1999] Wooldridge, M.: *Intelligent Agents*. Weiss, G. (Ed): *Multi-agent Systems: a Modern Approach to Distributed Artificial Intelligence*, MIT Press, 1999
- [Wooldridge, 2002] Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*. Chichester, England, John Wiley & Sons, ISBN 047149691X.
- [WSDL, 2001] *Web Services Description Language (WSDL) 1.1*. W3C Note 15 March 2001. Disponible en la web <http://www.w3.org/TR/2001/NOTE-wsdl-20010315>
- [XQuery, 2010] *XQuery 1.0: An XML Query Language (Second Edition)*. W3C Recommendation 14 December 2010. Disponible en <http://www.w3.org/TR/xquery/>
- [Xu *et al.*, 2006] X. Wu, L. Zhang, and Y. Yu. Exploring Social Annotations for the Semantic Web. In *Proc. Of WWW06*, 2006.

- [Yang&Wu, 2009] Yang, Y. J., Wu, C. "An attribute-based ant colony system for adaptive learning object recommendation", *Expert Systems with Applications*, vol. 36: pp. 3034-3047, 2009.
- [Yang, 1999] Yang Y.. An evaluation of statistical approaches to text categorization. *Journal of Information Retrieval* 1 (1999) 67-88.
- [Yen *et al.*, 2010] Neil Y. Yen, Timothy K. Shih, and Qun Jin. 2010. A new paradigm of ranking & searching in learning object repository. In *Proceedings of the second ACM international workshop on Multimedia technologies for distance learning (MTDL '10)*. ACM, New York, NY, USA, 1-6. DOI=10.1145/1878052.1878054
- [Z39.50, 2003] *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*, 2003. Disponible en <http://www.loc.gov/z3950/agency/Z39-50-2003.pdf>
- [Zambonelli *et al.* 2004] Zambonelli, F., Gleizes, M.-P., Mamei, M., & Tolksdorf, R. (2004). Spray computers: frontiers of self-organisation for pervasive computing. *Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises - WETICE04* (págs. 397-402). IEEE Computer Society
- [Zambonelli *et al.*, 2001] Zambonelli, F., Jennings, N.R., and Wooldridge, M.J. Organisational Rules as an abstraction for the Analysis and Design of Multi-Agent Systems. *IJSEKE*. 11(3) pp. 303-328, June 2001
- [Zambonelli *et al.*, 2003] F. Zambonelli, N. R. Jennings, and M. Wooldridge. Developing multiagent systems: The Gaia methodology. *ACMTransactions on Software Engineering and Methodology*, 12(3):317-370, July 2003.
- [Zambonelli, 2002] Zambonelli, F.: Abstractions and Infrastructures for the Design and Development of Mobile Agent Organizations. In: Wooldridge, M., Weiss, G., Ciancarini P. (Eds.): *Agent-Oriented Software Engineering II*, LNCS 2222, Springer-Verlag, 2002, pp. 245 - 262.
- [Zhang & Zhou, 2006] M.L., Zhang, Z.H., Zhou. Multi-label neural networks with applications to functional genomics and text categorization. *IEEE Transactions on Knowledge and Data Engineering* 18(10) (2006) 1338-1351.
- [Zhang & Zhou, 2007] Min-Ling Zhang, Zhi-Hua Zhou, ML-KNN: A lazy learning approach to multi-label learning, *Pattern Recognition*, Volume 40, Issue 7, July 2007, Pages 2038-2048.
- [Zhou *et al.*, 2010] Shuigeng Zhou, Ming Xu, Jihong Guan, LESSON: A system for lecture notes searching and sharing over Internet, *Journal of Systems and Software*, Volume 83, Issue 10, October 2010, Pages 1851-1863.

Anexo A ESTÁNDARES PARA LA INTEROPERABILIDAD DE CONTENIDOS DIGITALES EDUCATIVOS

El presente anexo hace una revisión de las diferentes aproximaciones destinadas a la interoperabilidad de contenidos educativos. Este apartado pretende complementar el Los distintos apartados hacen referencia a los estándares más relevantes dirigidos al etiquetado de los contenidos mediante metadatos, a las arquitecturas de construcción de ROAs interoperables, a protocolos de búsqueda y recolección de metadatos en dichos repositorios.

A.1 ESTÁNDARES PARA EL MERCADO DE OBJETOS DE APRENDIZAJE

A.1.1 Dublin Core

Este conjunto de metadatos fue el pionero. Su origen está en el ámbito bibliotecario, por lo que no está específicamente pensando para ser utilizado en un contexto educativo. Fue elaborado por la DMCI (*Dublin Core Metadata Initiative*), una organización dedicada a fomentar la adaptación de estándares con el objetivo de lograr interoperabilidad entre sistemas y a desarrollar vocabularios especializados para la descripción de recursos; todo ello para poder crear sistemas de búsqueda de información más inteligentes. Al tratarse de un sistema de marcado muy genérico y que está muy extendido, es habitual utilizarlo también en cualquier otro contexto, incluido el educativo. De manera que aunque este estándar no surge específicamente para la descripción de los OA, puede ser aplicado a la mayoría de los recursos, siendo la base de desarrollo y llegando a estar integrado en un número importante de especificaciones de OA como IEEE LOM, EdNa, GEM, que describiremos en los siguientes apartados de esta memoria.

Las principales ventajas que ofrece este estándar son [Hillmann D., 2007]:

- **Simplicidad de creación y mantenimiento.** Esta especificación se creó teniendo en cuenta el punto de acuerdo entre simplicidad de uso y precisión en la descripción, por ello se diseñó un conjunto mínimo de

elementos, sencillo y que además no se especializara en ningún área o contexto concreto.

- **Semántica común, fácilmente entendible.** Gracias al uso de una semántica sencilla que describe los aspectos fundamentales de un recurso sin entrar en detalles innecesarios.
- **Alcance internacional.** Pese a que el estándar fue creado inicialmente en inglés, hoy en día, ya existe una traducción a diversos idiomas incluyendo finés, noruego, tailandés, japonés, francés, portugués, alemán, griego, indonesio y español.
- **Extensibilidad.** En la especificación se da la posibilidad de añadir nuevos elementos para satisfacer, de este modo, nuevas necesidades. Estos nuevos elementos podrían convivir sin mayor problema junto con los elementos actuales de Dublin Core.

La estructura de Dublin Core está basada en dos niveles denominados Simple (*Simple*) y Cualificado (*Qualified*). El estándar ISO 15836:2009, define el Conjunto clásico de Elementos Dublin Core, conocida como Dublin Core Simple (DC simple). Está formado por 15 elementos básicos que permiten describir cualquier objeto de información. Los elementos de Dublin Core son opcionales, repetibles y pueden aparecer en cualquier orden. Estos 15 elementos se presentan habitualmente divididos en tres grupos atendiendo tres aspectos fundamentales: elementos relacionados con el contenido, con la propiedad intelectual y finalmente con la instanciación.

Además de estos elementos básicos existen otros mecanismos que sirven para adaptar el Dublin Core a las necesidades concretas de información y que hacen que este modelo de metadatos sea aplicable a cualquier proyecto de sistema o servicio de información digital de cualquier ámbito. Mientras que el Nivel Simple está formado por 15 elementos, el Cualificado suma a estos 15 elementos otra serie de elementos cuantitativos (*qualifiers*) dedicados a la descripción más en detalle de los elementos simples.

Tabla A-1. Elementos de la especificación Dublin Core Simple

Elemento	Descripción	Ejemplo
Title	El nombre dado al recurso, habitualmente dado por el autor.	Title="The Sound of Music"
Subject	Los temas o tópicos del contenido del recurso descrito.	Subject="Aircraft leasing and renting"
Description	Breve resumen textual del contenido del recurso.	Description="Illustrated guide to airport markings and lighting signals[...]"
Type	Categoría del contenido del recurso descrito. Se recomienda seleccionar un valor dentro de un vocabulario	Type="Image" Type="Sound" Type="Text"

	controlado	
Source	Referencia al recurso del que se deriva el recurso descrito.	Source="Image from page 54 of the 1922 edition of Romeo and Juliet"
Relation	Relaciones del recurso descrito con otros recursos. Se recomienda el uso de un sistema de referencia formal. Existen diferentes tipos de relaciones: <i>isPartOf</i> , <i>isVersionOf</i> , <i>isFormatOf</i> , <i>hasFormat</i> , <i>isReferenceBy</i> , <i>isBasedFor</i> , <i>isBasedOn</i> y <i>required</i>	Relation="Engels' Origin of the Family, Private Property and the State"
Coverage	La extensión o ámbito del contenido del recurso. Puede ser espacial, temporal o cobertura jurídica.	Coverage="17th century" Coverage="1995-1996" Coverage="Boston, MA"
Creator	Persona, grupo de personas u organización encargadas de hacer el contenido del recurso.	Creator="Shakespeare, William"
Publisher	Entidad responsable de la publicación de los recursos.	Publisher="Funky Websites, Inc."
Contributor	Entidad responsable de realizar aportaciones al contenido del recurso.	Contributor="University of Salamanca"
Rights	Información acerca de la propiedad intelectual y distribución del recurso	Rights="Access limited to members"
Date	Fecha que debe estar relacionada con algún evento dentro del ciclo de vida del recurso.	Date="2009-03-12"
Format	Manifestación física o digital del recurso	Format="image/gif" Format="4 kB"
Identifier	Referencia inequívoca del recurso en un contexto dado.	Identifier="ISBN:0385424728"
Language	Idioma en el que está descrito el contenido intelectual del recurso	Language="fr"

Aunque Dublin Core (Metadata Element Set, Version 1.1.) es la especificación básica, actualmente forma parte de un nuevo esquema mucho más completo, que ha ido modificándose y ampliándose en diferentes versiones hasta llegar a la actual [DCMI, 2010] que reemplaza la del 2008 [DCMI, 2008]. La actual especificación contiene más elementos, algunos de los cuales se pueden utilizar con fines educativos, algunos de ellos son:

- **Audience.** Tipo de público hacia los que se dirige el recurso.
 - **educationLevel.** Nivel educativo hacia el que se dirige el recurso, que obviamente debe ser un recurso didáctico.
 - **instructionalMethod.** Incluye las técnicas docentes que se utilizarán junto con el recurso educativo que describe.
-

A.1.2 IEEE LOM

El estándar IEEE LOM (*IEEE Learning Object Metadata*), o simplemente LOM, es un modelo de datos utilizado para describir los OA. Resulta ser la única especificación reconocida internacionalmente desde el año 2002 a través del estándar de IEEE 1484.12.1:2002 [IEEE 1484.12.1-2002]. El propósito de este esquema es facilitar la búsqueda, evaluación, adquisición y uso de los OA, facilitando su intercambio y uso compartido, así como el desarrollo de catálogos e inventarios al tiempo que se toman en consideración la diversidad cultural y los diferentes contextos educativos en los que los objetos educativos y sus metadatos serán reutilizados.

Para este Estándar, un objeto educativo se define como cualquier entidad, digital o no, susceptible de ser usada en aprendizaje, educación o formación. En lo que respecta a este estándar, una instancia de metadatos para un objeto educativo describe las características relevantes del objeto educativo al que se aplica con la forma de un esquema conceptual de datos. Dichas características se pueden agrupar en las categorías general, ciclo de vida, meta-metadatos, técnica, uso educativo, derechos, relación, anotación y clasificación.

Se asume que el Estándar será referenciado por otros estándares que definirán descripciones de implementación del esquema de datos, de manera que una instancia de metadatos para un objeto educativo pueda ser usada por un sistema basado en tecnología educativa para gestionar, localizar, evaluar o intercambiar objetos educativos. El Estándar no define como un sistema basado en tecnología educativa representará o usará una instancia de metadatos de un objeto educativo.

Para concretar aspectos en profundidad del estándar LOM, se hace uso de las definiciones que se presentan a continuación:

- **Categoría (LTSC – Learning Technologies Standards Committee-Learning Object Metadata).** Grupo de elementos de datos relacionados.
- **Elemento de datos de LOM.** Un elemento de datos cuyo nombre, explicación, tamaño, orden, espacio de valores y tipo de datos están definidos en este estándar.
- **Tipo de datos.** Una variedad de valores determinada por sus características comunes y las operaciones sobre ellos.

- **Elemento de datos extendido.** Un elemento de una estructura de datos que está definido al margen de un estándar y permitido dentro de una instancia de la estructura de datos.
- **Espacio de valores.** El conjunto de los valores posibles de un determinado tipo de datos.

Tal y como se aprecia en la Figura A-1, la especificación se organiza en 9 categorías que se describen a continuación:

1. **General (General).** Agrupa la información general que describe un OA de manera global.
2. **Ciclo de vida (Lifecycle).** Agrupa las características relacionadas con la historia y el estado actual del OA, y aquellas que le han afectado durante su evolución.
3. **Meta-Metadatos (Meta-Metadatos).** Agrupa la información sobre la propia instancia de metadatos, en lugar del OA descrito por la instancia de metadatos.
4. **Técnica (Technical).** Agrupa los requerimientos y características técnicas del OA.
5. **Uso Educativo (Educational).** Agrupa las características educativas y pedagógicas del OA.
6. **Derechos (Rights).** Agrupa los derechos de propiedad intelectual y las condiciones para el uso del OA.
7. **Relación (Relation).** Agrupa las características que definen la relación entre este OA y otros OAs relacionados.
8. **Anotación (Annotation).** Permite incluir comentarios sobre el uso educativo del OA e información sobre cuándo y por quién fueron creados dichos comentarios.
9. **Clasificación (Classification).** Describe este objeto educativo en relación a un determinado sistema de clasificación

Cada una de estas nueve categorías en las que se organiza el esquema IEEE LOM está compuesta por los elementos de datos que describen el objeto educativo. Tanto categorías como elementos se estructuran también de forma jerárquica donde los elementos podrán ser datos agregados o simples. Cada uno de ellos se corresponde con un elemento dentro de un esquema de numeración.

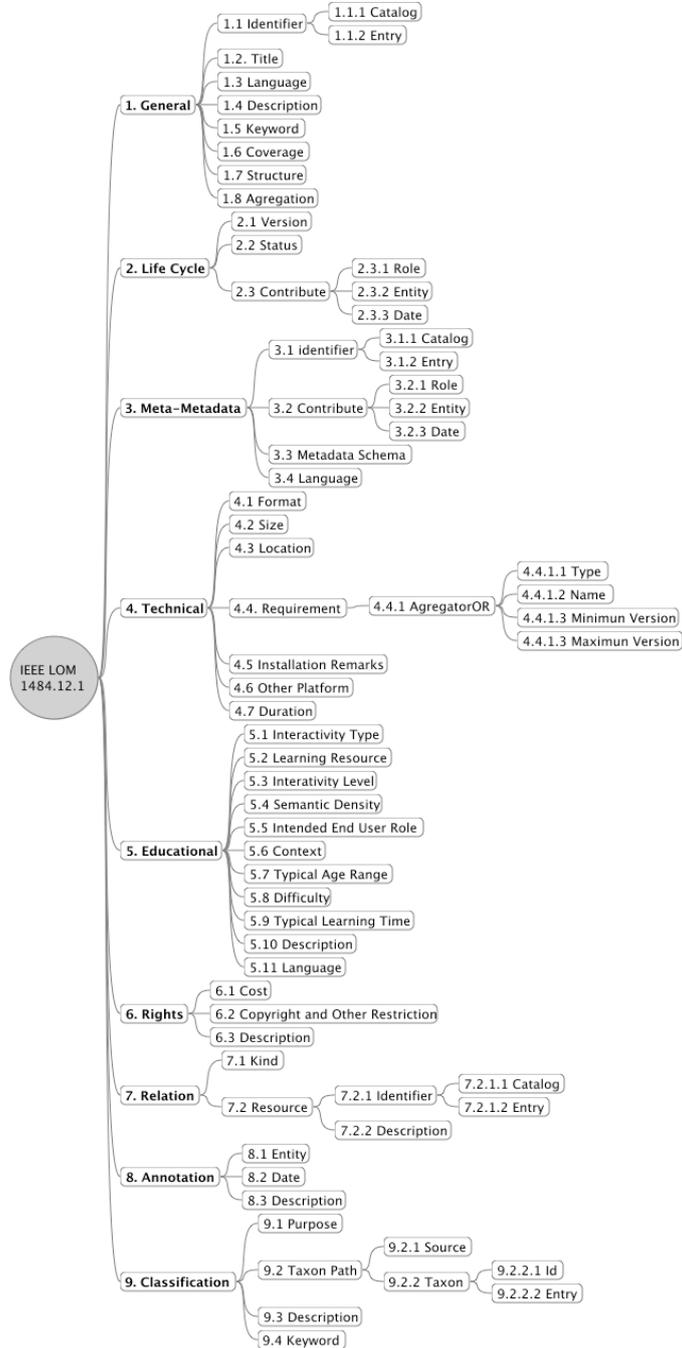


Figura A-1. El modelo de datos de IEEE LOM

Resulta necesario destacar que según la especificación del estándar ninguno de los elementos es obligatorio.

Tabla A-2. Descripción del modelo de datos de IEEE LOM

Núm.	Categoría	Descripción
1	General	
1.1	Identifier	Una etiqueta, única que identifica este OA.
1.1.1	Catalog	El nombre o denominación del esquema de identificación o catalogación para esta entrada. Un esquema de espacio de nombres.
1.1.2	Entry	El valor del identificador dentro del esquema de identificación o catalogación que designa o identifica este OA. Una cadena específica de un espacio de nombres.
1.2	Title	El nombre asignado a este OA.
1.3	Language	El idioma o idiomas humanos predominantes en este OA para la comunicación con el usuario.
1.4	Description	Una descripción textual del contenido de este OA.
1.5	Keyword	Una palabra clave o frase que describe el tema principal del OA.
1.6	Coverage	La época, cultura, zona geográfica o región a la que es aplicable este OA.
1.7	Structure	La estructura organizativa subyacente a este OA.
1.88	Agregation	La granularidad funcional de este OA.
2	Life Cycle	
2.1	Version	La edición de este OA.
2.2	Status	El estado de completitud o condición de este OA.
2.3	Contribute	Aquellas entidades (personas u organizaciones) que han contribuido al estado de este OA a lo largo de su ciclo de vida (Creación, edición, publicación, etc.).
2.3.1	Role	Tipo de contribución.
2.3.2	Entity	La identificación e información de las entidades (personas u organizaciones) que han contribuido a este OA ordenadas en función de la relevancia.
2.3.3	Date	La fecha de la contribución
3	Meta-Metadata	
3.1	Identifier	Una etiqueta única globalmente que menor identifica este registro de metadatos.
3.1.1	Catalog	El nombre o denominación del esquema de identificación o catalogación para esta entrada. Un esquema de espacio de nombres.
3.1.2	Entry	El valor del identificador dentro del esquema de identificación o catalogación que designa o identifica este registro de metadatos. Una cadena específica de un espacio de nombres.

3.2	Contribute	Aquellas entidades (personas u organizaciones) que han contribuido al estado de este OA a lo largo de su ciclo de vida (Creación, edición, publicación, etc.).
3.2.1.	Role	Tipo de contribución.
3.2.2	Entity	La identificación e información de las entidades (personas u organizaciones) que han contribuido a este OA ordenadas en función de la relevancia.
3.2.3	Date	La fecha de la contribución
3.3	MetadataSchema	El nombre y versión de la especificación utilizada para crear esta instancia de metadatos.
3.4	Language	El idioma de esta instancia de metadatos. Este es el idioma por defecto para todos los valores de tipo <i>LangString</i> en esta instancia de metadatos. Si no hay un valor para este elemento de datos, entonces no hay un valor por defecto para los elementos cuyo tipo de valor sea <i>LangString</i> .
4	Technical	
4.1	Format	El tipo de datos de (todos los componentes) este OA.
4.2	Size	El tamaño del OA digital expresado en octetos. El tamaño se representa como un valor decimal (base 10).
4.3	Location	Una cadena utilizada para acceder a este OA. Puede ser un localizador (URL), o un mecanismo que finalmente permite acceder a una localización (URI).
4.4	Requirement	Los requisitos técnicos para utilizar este OA.
4.4.1	AgregatorOR	Agrupación de múltiples requisitos. El requisito compuesto se satisface cuando al menos uno de los requisitos agregados se satisface.
4.4.1.1	Type	La tecnología requerida para usar este OA.
4.4.1.2	Name	El nombre de la tecnología requerida para utilizar este OA.
4.4.1.3	Minimum Version	La versión mínima posible de la tecnología necesaria para utilizar este OA.
4.4.1.4	Maximum Version	La versión máxima posible de la tecnología necesaria para utilizar este OA.
4.5	InstallationRemarks	Descripción de cómo debe ser instalado este OA.
4.6	Other Platform Requirements	Información sobre otros requisitos software o hardware.
4.7	Duration	El tiempo que dura un OA continuo cuando se reproduce a su velocidad normal.
5	Educational	
5.1	Interactivity Type	El tipo de aprendizaje predominante soportado por este OA. Aprendizaje activo, expositivo o combinado.
5.2	Learning	El tipo específico de recurso educativo. El tipo predominante debe

	Resource Type	aparecer en primer lugar.
5.3	Interactivity Level	El grado de interactividad que caracteriza a este OA.
5.4	Semantic Density	El grado de concisión de un OA.
5.5	Intended End User Role	El usuario principal para el que ha sido diseñado este OA. El predominante debe aparecer al principio.
5.6	Context	El entorno principal en el que se utilizará este OA.
5.7	Typical Age Range	Edad del destinatario típico.
5.8	Difficulty	Este elemento describe lo difícil que resulta, para los destinatarios típicos, trabajar utilizar este OA.
5.9	Typical Learning Time	Tiempo aproximado o típico que necesitan para asimilar el OA los destinatarios objetivo típicos.
5.10	Description	Comentarios sobre cómo debe utilizarse este OA.
5.11	Language	El idioma utilizado por el destinatario típico de este OA.
6	Rights	
6.1	Cost	Indicación de si este OA requiere pago
6.2	Copyright and Other Restrictions	Indicación de si existen derechos de autor u otras restricciones sobre este OA.
6.3	Description	Comentarios sobre las condiciones de utilización de este OA.
7	Relation	
7.1	Kind	Naturaleza de la relación entre este OA y el OA objetivo identificado por 7.2
7.2	Resource	El OA objetivo al que se refiere esta relación
7.2.1	Identifier	Una etiqueta, única global, que identifica el OA objetivo.
7.2.1.1	Catalog	El nombre o denominación del esquema de identificación o catalogación para esta entrada. Un esquema de espacio de nombres.
7.2.1.2	Entry	El valor del identificador dentro del esquema de identificación o catalogación que designa o identifica el OA objetivo. Una cadena específica de un espacio de nombres.
7.2.2	Description	Descripción del OA objetivo
8	Annotation	
8.1	Entity	La entidad (persona u organización) que creó esta anotación.
8.2	Date	La fecha en la que se creó esta anotación.
8.3	Description	El contenido de esta anotación
9	Classification	

9.1	Purpose	El propósito que se persigue al clasificar este OA.
9.2	Taxon Path	El camino taxonómico dentro de un sistema de clasificación específico. Cada nivel sucesivo representa un refinamiento sobre la definición dada en el nivel precedente.
9.2.1	Source	El nombre del sistema de clasificación.
9.2.2	Taxon	Un término concreto dentro de la taxonomía. Un taxón es un nodo que tiene definida una etiqueta o término. Un taxón puede poseer también una identificación o designación alfanumérica para ser utilizada como referencia estandarizada. Tanto la etiqueta como la entrada, o ambos, pueden ser utilizados para identificar un taxón particular.
9.2.2.1	Id	El identificador del taxón, tal como un número o una combinación de letras proporcionadas por la fuente de la taxonomía.
9.2.2.2	Entry	La etiqueta textual del taxón.
9.3	Description	Descripción del OA en relación con lo definido en 9.2
9.4	Keyword	Palabras clave y frases descriptivas del OA relacionadas con el elemento 9.1

Para algunos elementos de datos se definen vocabularios. Un vocabulario es una lista recomendada de valores apropiados. Se pueden usar también otros valores no incluidos en la lista. Sin embargo, los metadatos que se ajustan a los valores recomendados tendrán el máximo grado de interoperabilidad semántica

Tabla A-3. Correspondencia IEEE LOM y Dublin Core

Elemento	Descripción	Ejemplo
Title	El nombre dado al recurso, habitualmente dado por el autor.	Title="The Sound of Music"
Subject	Los temas o tópicos del contenido del recurso descrito.	Subject="Aircraft leasing and renting"
Description	Breve resumen textual del contenido del recurso.	Description="Illustrated guide to airport markings and lighting signals[...]"
Type	Categoría del contenido del recurso descrito. Se recomienda seleccionar un valor dentro de un vocabulario controlado	Type="Image" Type="Sound" Type="Text"
Source	Referencia al recurso del que se deriva el recurso descrito.	Source="Image from page 54 of the 1922 edition of Romeo and Juliet"
Relation	Relaciones del recurso descrito con otros recursos. Se recomienda el uso de un sistema de referencia formal.	Relation="Engels' Origin of the Family, Private Property and the State"

	Existen diferentes tipos de relaciones: <i>isPartOf</i> , <i>isVersionOf</i> , <i>isFormatOf</i> , <i>hasFormat</i> , <i>isReferenceBy</i> , <i>isBasedFor</i> , <i>isBasedOn</i> y <i>required</i>	
Coverage	La extensión o ámbito del contenido del recurso. Puede ser espacial, temporal o cobertura jurídica.	Coverage="17th century" Coverage="1995-1996" Coverage="Boston, MA"
Creator	Persona, grupo de personas u organización encargadas de hacer el contenido del recurso.	Creator="Shakespeare, William"
Publisher	Entidad responsable de la publicación de los recursos.	Publisher="Funky Websites, Inc."
Contributor	Entidad responsable de realizar aportaciones al contenido del recurso.	Contributor="University of Salamanca"
Rights	Información acerca de la propiedad intelectual y distribución del recurso	Rights="Access limited to members"
Date	Fecha que debe estar relacionada con algún evento dentro del ciclo de vida del recurso.	Date="2009-03-12"
Format	Manifestación física o digital del recurso	Format="image/gif" Format="4 kB"
Identifier	Referencia inequívoca del recurso en un contexto dado.	Identifier="ISBN:0385424728"
Language	Idioma en el que está descrito el contenido intelectual del recurso	Language="fr"

Por otro lado y como ya se comentó en el apartado A.1.1, el estándar LOM contiene a la especificación Dublin Core. La Tabla A-3 detalla la alineación de los elementos de un esquema que se corresponden con los del otro.

A.1.3 Perfiles de Aplicación

En muchas ocasiones la implementación de estándares o especificaciones en contextos educativos concretos exige considerar diferentes particularidades de cada comunidad, región o país. Por esta razón se desarrollan los denominados perfiles de aplicación (en inglés, *application profiles*) que combinan y utilizan uno o más estándares o especificaciones para crear nuevos esquemas dirigidos a situaciones específicas pero que, sin embargo, no agregan nuevos elementos y

se aseguran de mantener la interoperabilidad con los estándares o especificaciones originales. En la generación de perfiles de aplicación resulta fundamental mantener una separación entre la representación sintáctica de los metadatos, referido al lenguaje usado en su representación, y el significado, referenciado en la definición semántica de dichos metadatos. Un perfil de aplicación realiza un ensamblaje de metadatos escogidos de uno o más esquemas de metadatos, realizan inclusión de nuevos elementos de datos (elementos extendidos), nuevos vocabularios controlados para los mismos, así como modificaciones de los ya existentes en el estándar original. Como los perfiles de aplicación no introducen nuevos elementos [Duval, *et al.*, 2002]: cada elemento debe provenir de un `namespace` particular y si es necesario un nuevo término deberá crearse un nuevo `namespace` para él. El objetivo principal de los perfiles de aplicación es aumentar la interoperabilidad semántica del esquema resultante, sin que se comprometa la interoperabilidad básica que el estándar habilita en los límites de las distintas comunidades [Demsey y Patel, 2000].

Concretamente en España se desarrolla el Perfil de Aplicación (Norma UNE-71361:2010) denominado LOM-ES. El proyecto Agrega amparado por el gobierno de España ha modificado la versión oficial del estándar IEEE LOM para la comunidad educativa española. Ambos estándares, LOM y LOM-ES, son perfectamente compatibles, pero éste último contiene algunas modificaciones como es la agregación de nuevos elementos en los esquemas y nuevos términos en los vocabularios, modificación de cardinalidades, y además, también se han modificado restricciones que exigen la obligatoriedad de algunos términos en el etiquetado.

Otros perfiles de aplicación son el *United Kingdom Learning Object Metadata Core* o UK LOM Core⁴⁹. Un proyecto planteado por la comunidad de profesionales del Reino Unido para identificar la práctica común en dicho país sobre el mercado de contenidos de los OA. Aglutina la práctica común, que no la que sería deseable. Su objetivo no es ser preceptivo, sino reflejar lo que se está haciendo en el mercado de los contenidos. Hay tres tipos de elementos dentro de UK LOM: Obligatorio, Opcional y Opcional (recomendado).

Canadian Core Learning Resource Metadata (CanCore⁵⁰) es una iniciativa sostenida por TeleUniversite y la Universidad Athabasca Canadienses. Esta iniciativa presenta también un perfil de aplicación del estándar IEEE LOM, que además recomienda determinadas simplificaciones e interpretaciones sobre él, así como buenas prácticas a la hora de su implementación. A diferencia de muchos perfiles de aplicación, CanCore proporciona una guía detallada para la

⁴⁹ <http://zope.cetis.ac.uk/profiles/uklomcore> [Último acceso: 9 de Diciembre de 2010]

⁵⁰ <http://cancore.athabascau.ca/en/> [Último acceso: 10 de Diciembre de 2010]

interpretación y aplicación de cada elemento de datos en el estándar LOM. El detalle de esta especificación puede verse en [Friesen *et al.*, 2002] y a través de la página del proyecto.

A.1.4 Otros

En la actualidad, debido a la existencia de varios grupos de trabajo, se pueden encontrar diferentes especificaciones para la descripción de OA. Todas estas especificaciones son bastante similares, pero en cada una de ellas, cada grupo de trabajo presenta su visión particular de cómo debería ser un esquema de metadatos apropiado para la descripción de recursos educativos.

Además de las ya vistas en apartados anteriores, a continuación, se presentan algunas otras de las especificaciones más importantes:

ARIADNE Educational Metadata Recommendation. La fundación Ariadne⁵¹ (*Alliance for Remote Instructional Authoring and Distribution Networks for Europe*) fue creada al amparo de dos proyectos europeos que compartían el mismo objetivo, desarrollar herramientas y metodologías para la gestión eficiente y la reutilización de recursos educativos. Esta fundación desarrolló su propio conjunto de metadatos [Najjar *et al.*, 2003] de forma previa a IEEE LOM. La actual recomendación de metadatos para contenidos educativos, la versión 3.2 de *ARIADNE Educational Metadata Recommendation* de Febrero del 2002, conforma un perfil de aplicación de la especificación LOM, en el sentido de que esta recomendación es totalmente compatible con la especificación LOM, y es instanciada por la comunidad ARIADNE. Consta de un conjunto de 47 elementos, 27 de los cuales se puede asignar directamente a los elementos de LOM. Organizado en seis categorías: *General, Semantics, Pedagogical, Technical, Indexation, Annotation*. Tiene en cuenta las necesidades y exigencias específicas de una comunidad que es muy representativa en el ámbito de la Educación Superior en Europa y la Formación Profesional Continua.

GEM Metadata. El proyecto GEM (*Gateway to Educational Materials*)⁵² financiado por el departamento de Educación de U.S., surgió con el objetivo de definir un perfil de metadatos semánticamente ricos a través de vocabularios controlados. Su dominio específico surge de la necesidad de la descripción de materiales educativos en la Web. La especificación, de marzo de 2002, contiene los 15 elementos del estándar Dublin Core y otros siete elementos adicionales en aquel momento (*Audiencie, Cataloging, Duration, Essential Resources, Pedagogy, Quality y Standard*), algunos de los cuales ya forman parte de la

⁵¹ ARIADNE <http://www.ariadne-eu.org/>

⁵² GEM <http://www.thegateway.org/>

actual versión de Dublin Core [DCMI, 2010]. Una revisión del estándar se puede ver en [Sutton, 2004].

EdNa Metadata. La iniciativa amparada por el gobierno australiano y denominada Edna Online⁵³ también ha propuesto un esquema de metadatos basado en Dublin Core similar a la especificación GEM Metadata. Este esquema tiene como principales principios la simplicidad, la interoperabilidad, la extensibilidad y el refinamiento o abstracción. La especificación completa se puede consultar en [EdNa Online, 2002].

Metadata for Learning Resources (MLR). El grupo de trabajo ISO/IEC JTC1 SC36 *Working Group 4: Manage and Delivery of Learning, Education, and Training* (MDLET) se centra en el desarrollo de diferentes estándares y metadatos aplicados a recursos de aprendizaje. El estándar tiene como objetivo precisar un recurso de aprendizaje a través de normas, metadatos y atributos. Está dividido en partes estructurales que proporcionan niveles de abstracción en la gestión integral de recursos didácticos. Se puede acceder a la especificación completa en [ISO/IEC 19788, 2010].

A.2 MODELOS DE AGREGACIÓN DE CONTENIDO

Los OAs son las unidades mínimas de contenido reutilizable, que en la definición de una unidad de aprendizaje pueden agruparse o secuenciarse en el modo que se estime oportuno para conseguir el propósito educativo para el que se destine. En definitiva, los OAs son recursos digitales de cualquier tipo (páginas web, vídeos, animaciones, etc.). Sin embargo, la construcción de OAs hace referencia al uso de estándares para dotar a esos recursos de propiedades que permitan su acceso, agregación e intercambio sin problemas de interoperabilidad. La agregación de los OA para configurar cursos o unidades más complejas se realizará mediante una estructura que permita alcanzar las especificaciones de las diferentes situaciones de enseñanza. Con los estándares actuales existen diferentes modelos que permiten dicha agregación de contenidos a partir de los OAs. Los dos modelos principales son el desarrollado por AICC (*Aviation Industry CBT Committee*) y el que forma parte de SCORM (*Sharable Content Object Reference Model*).

A.2.1 AICC (*Aviation Industry CBT Committee*)

El comité CBT (*Computer-Based Training*) de la industria de la aviación (*Aviation Industry CBT Committee*, AICC⁵⁴) es una asociación internacional de

⁵³ EDNA <http://www.edna.edu.au/>

⁵⁴ AICC <http://www.aicc.org>

profesionales de la formación basada en tecnologías. La AICC desarrolla guías para la aviación en el desarrollo, intercambio y evaluación de CBT y tecnologías de formación relacionadas.

AICC establece una jerarquía de contenidos en base a tres estados, cada uno de los cuales incrementa el anterior. Estos son: (1) La Unidad Asignable (AU), como unidad mínima de aprendizaje. (2) El Bloque Instruccional como agrupaciones de AUs y (3) El Curso como nivel superior de la jerarquía.

AICC basa la comunicación entre los contenidos y el Sistema Gestor de Contenidos (LMS) en mensajes HTTP para conseguir que se produzca la correcta interpretación de los cursos por parte del LMS, llamándose a este protocolo mediante las siglas HACP (*HTTP AICC CMI Protocol*). AICC no exige que los componentes de un mismo curso se presenten alojados o empaquetados en el mismo fichero o lugar, permitiendo incluso que componentes de un mismo curso se encuentren alojados en diferentes servidores. Como sistema de codificación utiliza el ISO 8859, perfectamente válido para caracteres de lengua inglesa pero generando problemas para sistemas de codificación tipo UNICODE.

A.2.2 SCORM

Son muchas las organizaciones que promocionan la creación de especificaciones y estándares para las tecnologías relacionadas con el aprendizaje, como hemos visto en puntos anteriores. El Departamento de Defensa de los Estados Unidos y la Oficina de Políticas de Ciencia y Tecnología de la Casa Blanca decidieron crear la iniciativa *Advanced Distributed Learning* (ADL)⁵⁵ con el objetivo de impulsar, liderar y coordinar los diversos esfuerzos orientados al empleo de las tecnologías de la información en la modernización del aprendizaje. SCORM es el resultado del trabajo del comité CBT (*Computer-Based Training*) de la industria de la aviación (*Aviation Industry CBT Committee*, AICC), del Consorcio *IMS Global Learning, Inc.*, del IEEE, ARIADNE y otras iniciativas para crear un “modelo de referencia” unificado que proporcionara las especificaciones técnicas y las directrices del Departamento de Defensa (DoD) que cumplan con requisitos de alto nivel para contenido y sistemas educativos en la Web. La versión actual de SCORM es la denominada SCORM 2004 4ª Edición, Versión 1.1 (4th Ed.) [ADL SCORM 2004, 2009]. Su documentación completa puede encontrarse en la página web⁵⁶ de la iniciativa.

⁵⁵ ADL <http://www.adlnet.gov/>

⁵⁶ <http://www.adlnet.gov/Technologies/scorm/SCORMSDocuments/2004%204th%20Edition/Documentation.aspx> SCORM

Según ADL, la presencia de distintas especificaciones propuestas por diversos grupos no resultaba suficiente para garantizar la interoperabilidad entre sistemas, debido a la falta de implantación en algunos casos y los conflictos entre especificaciones en otros. La definición de SCORM tiene el objetivo de establecer un marco común para el aprendizaje a distancia, basándose en 6 principios esenciales:

- **Accesibilidad.** Posibilidad de localizar y acceder a OA desde una ubicación remota y su envío a otras muchas localizaciones.
- **Adaptabilidad.** Posibilidad de adaptar la enseñanza a distintas necesidades individuales u organizacionales.
- **Asequibilidad.** Posibilidad de aumentar la eficiencia y la productividad reduciendo el tiempo y el coste invertidos en la enseñanza.
- **Durabilidad.** Posibilidad de resistir la evolución de la tecnología y futuros cambios sin incurrir en rediseños, reconfiguraciones o recodificaciones excesivamente costosas.
- **Interoperabilidad.** Posibilidad de tomar componentes instruccionales desarrollados en una ubicación determinada y empleando unas herramientas y plataformas determinadas para su posterior aplicación en otra ubicación y otro conjunto de herramientas y plataformas.
- **Reusabilidad.** Flexibilidad para incorporar componentes instruccionales en múltiples contextos y aplicaciones

ADL reúne el trabajo de todas estas organizaciones dentro de un “Modelo de Referencia” común conocido como, el “Modelo de Referencia de Objetos de Contenido Compartibles” (*Sharable Content Object Reference Model*, SCORM) [ADL SCORM, 2002] [ADL SCORM, 2006]. SCORM proporciona un modelo de referencia que especifica cómo se deben definir los OA, sus metadatos, su empaquetamiento y su distribución. Para la definición de todos estos conceptos se emplean especificaciones ya existentes, en particular SCORM integra las siguientes especificaciones creadas por diferentes grupos:

- **IEEE Learning Object Meta-data 1484.12.1.** Empleado en el Modelo de Agregación de Contenido para definir los Metadatos de los OA.
- **IEEE ECMAScript API for Content to Runtime Services Communication 1484.11.2.** Empleado por el Entorno de Tiempo de Ejecución para definir el mecanismo de comunicación entre el contenido y el LMS (*Learning Management System*).
- **IEEE Data Model for Content Object Communication 1484.11.1.** Empleado por el Entorno de Tiempo de Ejecución para definir el modelo de datos empleado en la comunicación entre el contenido y el LMS.
- **AICC/Web-Based CMI Guidelines.** Empleado para definir la estructura del contenido en el Modelo de Agregación de Contenido.

- **IMS Content Packaging.** Empleado en el Modelo de Agregación de Contenido para agrupar objetos de contenido.
- **IMS Simple Sequencing.** Empleado para secuenciar las actividades en un curso.

SCORM se ha desarrollado para permitir a los desarrolladores de objetos de contenido su reusabilidad e interoperabilidad en diferentes LMSs. Con este fin centra su modelo en base a tres aspectos principales y para cada uno de ellos publica un documento técnico donde describe todos los detalles. Dichos temas principales y sus correspondientes documentos técnicos son:

- **Modelo de Agregación de contenidos (*Content Aggregation Model*, - CAM) [SCORM CAM, 2004].** SCORM describe a su Modelo de Agregación de Contenidos como una representación de medios neutrales de una taxonomía de aprendizaje, para que diseñadores e implementadores de instrucción agreguen recursos de aprendizaje con el propósito de compartir una experiencia de aprendizaje deseada. SCORM define un recurso de aprendizaje como cualquier representación de información que se usa en una experiencia de aprendizaje. Dichas experiencias de aprendizaje consisten en actividades que se apoyan en recursos de aprendizaje electrónicos o no electrónicos. Una actividad, en el proceso de crear y lanzar experiencias de aprendizaje, involucra la creación, descubrimiento y unión o agregación de simples *Assets* dentro de recursos de aprendizaje más complejos, que luego se deben organizar dentro de una secuencia predefinida de lanzamiento. El Modelo de Agregación de Contenidos da soporte a todo este proceso mediante la composición de cinco elementos: Modelo de contenido, Empaquetamiento del contenido, Metadatos y la Secuencia y Navegación. El manual SCORM CAM describe los distintos tipos de objetos de contenido permitidos dentro de la especificación y detalla los mecanismos que se deben seguir para su empaquetamiento, descubrimiento en repositorios y su distribución e interoperabilidad entre distintos LMS.
- **El entorno de Ejecución (*Run-Time Environment* - RTE) [SCORM RTE, 2004].** El requisito de adaptabilidad e intercambio de datos entre el contenido y el LMS da lugar a contenidos educativos más complejos de lo habitual y es necesario estandarizar el proceso de ejecución de estos contenidos para garantizar la interoperabilidad entre distintos LMS. Por ello, este manual define el proceso de ejecución y los mecanismos de comunicación que tanto el LMS como el propio contenido deben emplear.
- **Secuencia y Navegación (*Sequencing and Navigation* - SN) [SCORM SN, 2004].** Este manual define los mecanismos para que los LMS puedan concatenar las actividades educativas de modo consistente. El

manual recoge los eventos que pueden ser generados por los alumnos o por el sistema y que deben ser procesados por el LMS para decidir cuál es el recurso educativo que debe ser servido a continuación. También se recoge el modelo de datos para generar y procesar estos eventos.

En lo referente a metadatos, SCORM recomienda seguir el estándar IEEE LOM, imponiendo dos tipos de restricciones:

- **Elementos obligatorios.** Según LOM, todos los elementos son opcionales, pero SCORM en su perfil de aplicación recomienda que algunos elementos sean obligatorios para incrementar la interoperabilidad semántica y mejorar así la capacidad de búsqueda y reutilización.
- **Uso de vocabularios.** SCORM recomienda el uso de los vocabularios definidos por LOM, pero en el caso de que sea necesario crear nuevos vocabularios, se recomienda que éstos sean acordados por entidades verticales. Distingue entre (i) vocabularios *restringidos*, que indican la obligatoriedad; y (ii) vocabularios *mejores prácticas* que indican la conveniencia en el uso.

SCORM establece una lista de elementos obligatorios por categorías [SCORM CAM, 2004].

Los OA descritos mediante el estándar LOM pueden tener distintos grados de complejidad (obligatoriedad en los elementos). Según este grado de complejidad SCORM hace una división en niveles:

- **Elemento básico (*Asset*).** Representa al recurso de aprendizaje más básico.
- **Objeto reutilizable de aprendizaje (*Sharable Content Object* ó *SCO*).** Es un elemento básico o conjunto de ellos, dotados de información necesaria para poder ser gestionado por un LMS vía el entorno de ejecución de SCORM.
- **Actividad (*Activity*).** Una actividad es una instrucción, es decir, una acción a realizar sobre un SCO o sobre un *Asset*.
- **Organización de Contenido (*Content Organization*).** Representa el mapa o árbol de actividades, su secuencia de ejecución.
- **Agregación de contenido (*Content Aggregation*).** Representa el conjunto completo de todos los elementos anteriores.

A.2.3 Otros

Existen otros muchos modelos de agregación de contenidos digitales, algunos de ellos comparados en [Verbert y Duval, 2004] y que merecen ser al menos enumerados.

Wayne Hodgins, padre de los OAs y La Fundación para el aprendizaje (*Learnativity Foundation*), desarrollaron un modelo conceptual de contenido para OAs que provee una descripción granular más amplia [Duval & Hodgins, 2003]. Dicho modelo define en una estructura estándar Autodesk de cinco niveles la jerarquía de contenidos: Datos en crudo, objetos de información, objetos de aplicación (donde se incorporan los OA) y finalmente las colecciones.

NETg (*National Education Training Group Learning Object Model*). Contiene una jerarquía de cuatro niveles: curso, unidad, lección y tópico. En NETg un curso se ve como una matriz en tres dimensiones. Las unidades (elementos verticales), lecciones (elementos horizontales) y los tópicos (cada celda). Cada tópico representa un objeto de aprendizaje independiente que contiene un único objetivo de aprendizaje y tiene una actividad y una evaluación [L'Allier, 1997].

A finales de la década de 1990, Cisco Systems llevó a cabo un proyecto pionero para desarrollar una metodología que facilitase el paso de la creación y distribución de grandes e "inflexibles" cursos de formación para hacerlos reutilizables a la vez que permitiera sus búsquedas en las bases de datos. El modelo Cisco RLO/RIO⁵⁷, define piezas de información construidas alrededor de un objetivo de aprendizaje denominándolas *Reusable Information Objects* (RIOs). Para construir una experiencia de aprendizaje completa o lección se forma un paquete contenedor de una colección de RIOs, una presentación, un resumen y una evaluación que indicará si el estudiante ha conseguido el objetivo de aprendizaje de cada RIO.

El modelo de Contenido genérico para Objetos de Aprendizaje (ALOCOM) tiene como objetivo integrar la mayoría de los modelos de OA utilizando ontologías que definan los diferentes conceptos y tipos de componentes y sirva como base para una equivalencia entre los distintos modelos de contenido [Verbert y Duval, 2007].

A.3 ARQUITECTURAS PARA LA INTEROPERABILIDAD DE ROAS

Varias son las arquitecturas de configuración de los repositorios de objetos de aprendizaje con el objetivo de conseguir la interoperabilidad de sus contenidos. Entre los más destacables están CORDRA, IMS DRI e IMS LODS.

A.3.1 CORDRA

⁵⁷ CISCO Systems Reusable Information Object Strategy Version 3.0. Disponible en: http://www.cisco.com/warp/public/779/ibs/solutions/learning/whitepapers/el_cisco_rio.pdf

El proyecto CORDRA, Arquitectura para el Registro, Resolución y Descubrimiento de Repositorios de Objetos de Contenido (del inglés, *Content Object Repository Discovery and Registration/Resolution Architecture*) es un esfuerzo desarrollado en colaboración por la Corporación Nacional para Iniciativas de Investigación (CNRI, del inglés *Corporation for National Research Initiatives*), el Laboratorio de Arquitecturas para Sistemas de Aprendizaje (LSAL) y financiado por la iniciativa de Aprendizaje Avanzado Distribuido (ADL) del Departamento de Defensa de los Estados Unidos de América.

El objetivo de este proyecto es definir un marco para la federación de repositorios de contenido a través de un registro de metadatos, conocido como Registro ADL (*ADL Registry*), que permita mostrar y proporcionar una implementación de referencia de la especificación CORDRA. La documentación del Registro ADL y CORDRA consta de cinco volúmenes, parte de los cuales se puede descargar de la página del proyecto ⁵⁸[CORDRA, 2005].

El marco también conocido como "CORDRA especificación" pretende ser un modelo abierto, basado en estándares para el diseño e implementación de sistemas de información, incluyendo los registros y repositorios a los efectos del descubrimiento, el intercambio y la reutilización de la información. La especificación CORDRA describe cómo los dueños o administradores de la información digital distribuida pueden registrar la existencia de dicha información y así permitir que otras personas la recuperen y utilicen.

CORDRA está diseñado para ser un puente entre el mundo de los repositorios de contenidos educativos y las bibliotecas digitales. Su objetivo es el de identificar y especificar (no desarrollar) las tecnologías apropiadas y las normas de interoperabilidad que se pueden combinar en un modelo de referencia utilizado para habilitar una infraestructura de contenido de aprendizaje. Se trata de una arquitectura en la que una familia de servicios y herramientas integradas pueden apoyar el descubrimiento y la recuperación de los objetos digitales, tanto para contextos específicos como generales. Una aplicación de CORDRA en un dominio particular, implica el establecimiento del registro de servicios de CORDRA que implementan las políticas y prácticas de la comunidad. De ese modo los repositorios distribuidos, puede registrar los objetos y los metadatos asociados para que puedan ser descubiertos a través de los servicios de búsqueda.

El modelo CORDRA general se ilustra en la Figura A-2. Los componentes clave del modelo son:

⁵⁸ CORDRA

<http://www.adlnet.gov/Technologies/adlr/ADLRDocuments/ADL%20Registry%20Documentation/ADL%20Registry%20Documentation.aspx>

- Repositorios de Contenido: repositorios de datos locales para los contenidos educativos (OA, assets, etc.) y datos asociados (catálogos locales, metadatos, etc.)
- Repositorios de Sistema: repositorios de sistema CORDRA para los datos del sistema de gestión, modelos, registros, etc.
- Sistema de Identificación: Infraestructura para identificación del objeto, registro y resolución.
- Infraestructura de Servicios Comunes: Los servicios técnicos básicos y administrativos utilizados a través de una aplicación CORDRA (autenticación, gestión de derechos, reglas de procesamiento, etc.)
- Aplicaciones: sistemas de aplicaciones e interfaces (de búsqueda, descubrimiento, creación, personalización, entrega, etc.) para gestionar y entregar los contenidos educativos a los usuarios finales.

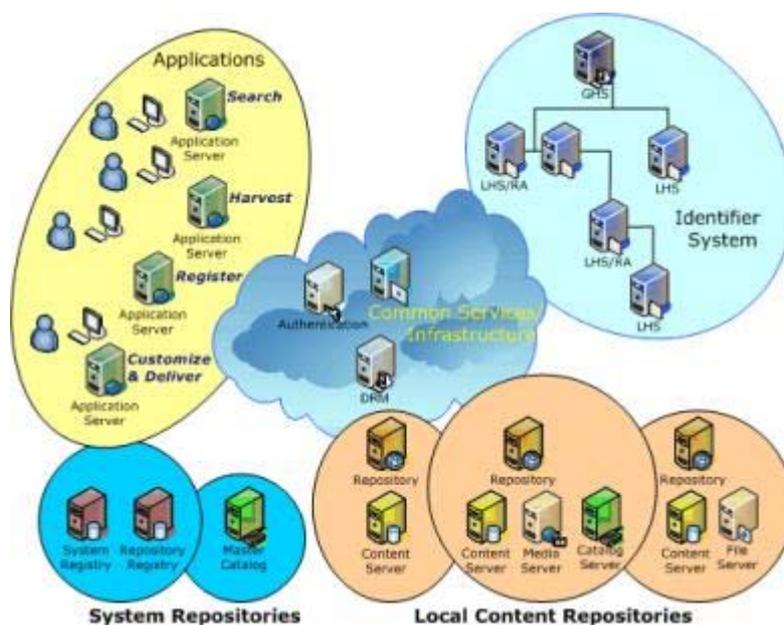


Figura A-2. Modelo CORDRA

La estructura del proyecto contempla la incorporación de grupos de repositorios organizados en federaciones que buscan integrar su contenido en un Registro central. Estas federaciones a su vez se registran en Registros de Registros (RoRs) conformando federaciones de federaciones. Estas federaciones pueden a su vez ser integradas en uno o más Registros maestros.

El nivel superior no se constituye en un área de control sino más bien un punto de entrada para acceder a alguna colección o contenido registrado en una de las diferentes federaciones [Giridhar *et al.*, 2006].

A.3.2 *IMS Digital Repositories Specification - IMS DRI (Digital Repository Interoperability)*

IMS DRI describe el Modelo de Información para la especificación de interoperabilidad de los repositorios digitales IMS. Proporciona recomendaciones para la interoperabilidad de las funciones más comunes de los repositorios. La especificación está compuesta por tres documentos disponibles a través de la página de consorcio ⁵⁹ [DRI, 2003]. Es uno de los primeros estándares definidos y solamente especifica una estructura a seguir para el desarrollo de sistemas de búsqueda federada entre diferentes repositorios. Se basa en tecnologías de comunicación ya existentes como pueden ser SOAP o Z39.50; y también en especificaciones previamente definidas por este consorcio como IMS Metadata y Content Packaging [IMS Content Packaging, 2007].

La propia especificación define a un repositorio digital como *una colección de recursos que son accesibles a través de Internet, sin la necesidad de tener conocimientos previos sobre su estructura. Los repositorios podrán tener activos (recursos) reales o metadatos que describan los activos (recursos). Ni los recursos, ni los metadatos, tendrán porqué estar almacenados en el propio repositorio.*

⁵⁹ DRI <http://www.imsglobal.org/digitalrepositories/index.html>

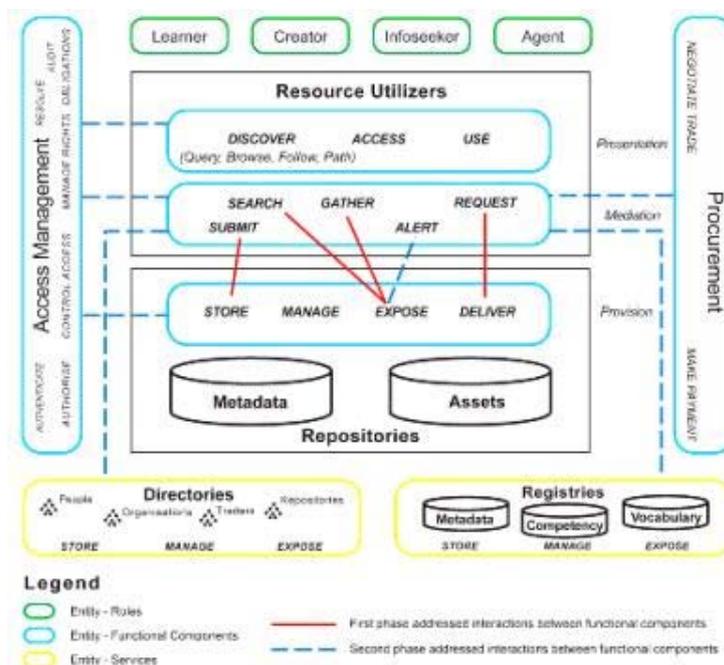


Figura A-3. Arquitectura funcional de la especificación DRI [DRI, 2003]

La especificación de DRI asume la alta heterogeneidad de los repositorios: formato de los recursos educativos, diferentes implementaciones y tecnologías, así como diferentes prácticas en su uso. Debido a ello, recomienda utilizar esta especificación en dos tipos de escenarios comunes, es decir, en dos tipos de repositorios habituales en el momento de la creación de la especificación:

- Repositorios que implementan el protocolo Z39.50 para su interoperabilidad.
- Repositorios que implementan el estándar XQuery [XQuery, 2010] y están basados en la especificación SOAP (siglas del inglés *Simple Object Access Protocol*) [SOAP, 2007].

La Figura A-3 muestra la arquitectura funcional de esta especificación, en ella, se pueden apreciar tres tipos de entidades bien diferenciadas:

- Los roles son todos los posibles clientes del sistema, juegan diferentes papeles y pueden ser tanto personas, como máquinas. Dentro de estos roles podemos encontrar los siguientes: *Learner*, *Creator*, *Infoseeker* y *Agent*.

- Los componentes funcionales son los módulos que debe tener un repositorio, externos e internos él, para permitir el acceso a los recursos que almacena y por lo tanto, para permitir la interoperabilidad. Dentro de estos componentes podemos encontrar, además del propio núcleo del repositorio: *Resource Utilizers*, *Access Management* y *Procurement Services*.
- Los servicios como Directorios o Registros, no definidos en la especificación de DRI.

En la Figura A-3 se aprecian un conjunto de líneas continuas rojas que se corresponden con el conjunto de funciones que permiten la interacción entre los manejadores de recursos (*resource utilizers*) y los repositorios. Este conjunto de funciones son la base de la interoperabilidad, tal y como, la define esta especificación. En el mismo esquema también existe una interacción en línea discontinua (función ALERT), ésta forma parte de la especificación DRI, pero aún no ha sido definida.

La principal ventaja de esta arquitectura es que permite el acceso al contenido almacenado en los repositorios a través de aplicaciones externas como Sistemas de Gestión del Aprendizaje (*Learning Management Systems*, LMS), o de gestión de contenidos educativos (*Learning Content Management System*, LCMS) y buscadores específicos. Este tipo de aplicaciones permite realizar búsquedas en varios repositorios de forma simultánea, lo que a priori es un problema complejo, que además se agrava debido a la gran heterogeneidad de la que adolecen los repositorios digitales. Esta especificación introduce un componente intermedio opcional que contiene tres funciones orientadas a simplificar el problema:

- Una función de traducción, que transforma un lenguaje de consulta concreto en otro que sea válido en los repositorios existentes.
- Una función de agregación, que permite unir los metadatos de diferentes repositorios para que de esta forma pueda estar disponible para la búsqueda.
- Una función de federación que envía las consultas a diversos repositorios y que recoge las repuestas de estos.

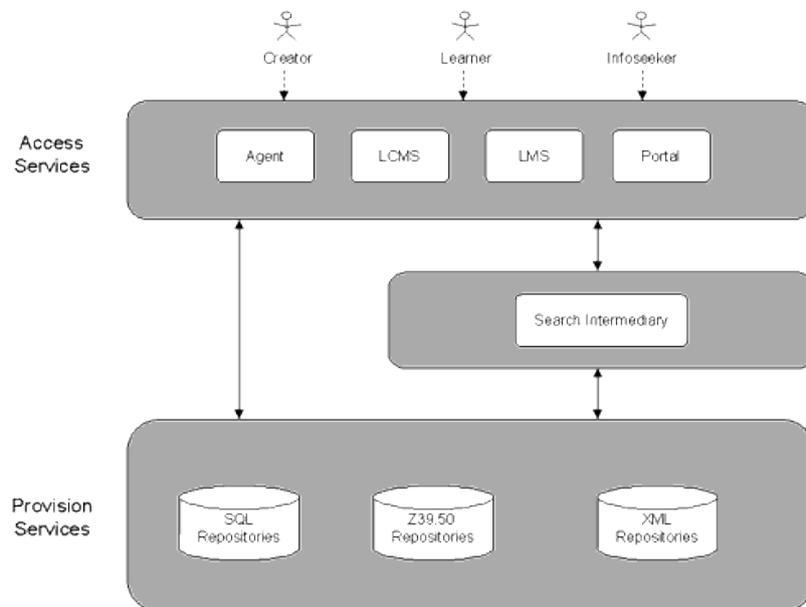


Figura A-4. Diagrama del Modelo General de Referencia diagrama con los distintos actores [DRI, 2003]

Tal y como ya hemos comentado, la primera fase y casi la única desarrollada, de la especificación IMS DRI define una serie de funciones del núcleo para permitir la interoperabilidad entre repositorios. Cada una de estas funcionalidades está compuesta a su vez por dos tareas o métodos, la primera en el cliente (*Resource Utilicers*) y el segundo, en el servidor (repositorio), la combinación de ambas funcionalidades, junto con la capa intermedia de búsqueda hace posible la interoperabilidad, según esta especificación. Las funciones que se proponen son las siguientes:

- **Search/Expose.** El modelo de búsqueda de referencia define cómo debe ser el proceso de búsqueda de metadatos asociados a los recursos de un repositorio. Las características principales de este modelo son dos, la primera de ellas es que es compatible con una amplia gama de configuraciones, y la segunda característica es que proporciona una capa intermedia de mediación entre sistemas heterogéneos. Como lenguaje de consulta define dos especificaciones ampliamente aceptadas en este contexto: XQuery y Z39.50.

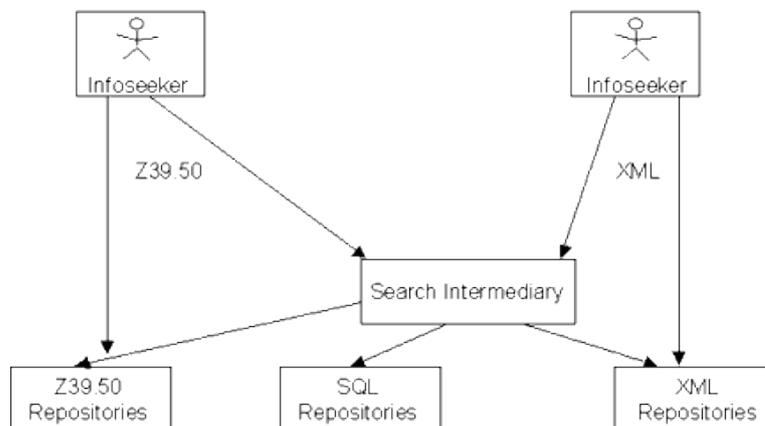


Figura A-5. Función de núcleo: Search/Expose [DRI, 2003]

- **Gather/Expose.** Muestra los metadatos de un repositorio derivados de la agregación de metadatos de diferentes consultas. Este tipo de funcionalidad puede interactuar con los repositorios de dos maneras:
 - Servicio *pull*, que implica una solicitud implícita de metadatos.
 - Servicio *push*, que implica la suscripción a un servicio externo.

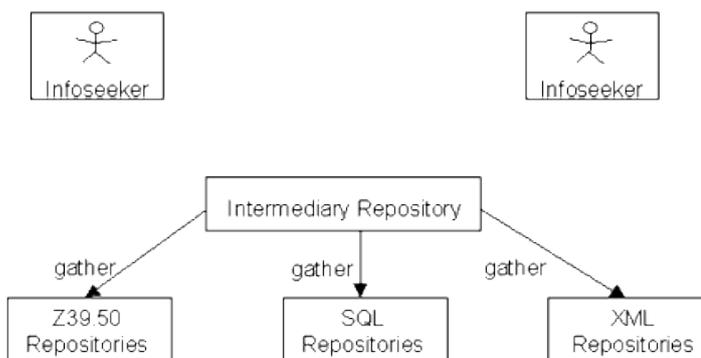


Figura A-6. Función de núcleo Gather/Expose [DRI, 2003]

- **Alert/Expose.** Esta función pese a formar parte del núcleo de la especificación, aún no está definida, pero cabe señalar que está orientada a ser un componente o capa intermedia que se puede implementar a través de correo electrónico o mediante el protocolo SMTP.

- **Submit/Store.** Esta función define como el objeto de aprendizaje se envía desde el repositorio hasta el punto de acceso a través de Internet.
- **Request/Deliver.** Solicitud de un recurso de aprendizaje que ha sido descubierto a través de una búsqueda o un servicio de Alerta.

La especificación define a través de casos de uso las posibles funcionalidades que puede realizar cada actor (agentes software, creador de contenidos, estudiantes, persona en busca de contenidos, etc.), las funcionales posibles son las siguientes:

- CU 1 – Creación de un curso y envío al repositorio.
- CU 2 – Búsqueda de un curso, posterior modificación del mismo y envío del curso modificado al repositorio.
- CU 3 – Búsqueda de metadatos y recuperación de los recursos mostrados en un repositorio por parte de algún actor humano.
- CU 4 – Búsqueda de metadatos y recuperación de los recursos mostrados en múltiples repositorios por parte de algún actor humano.
- CU 5 – Búsqueda de metadatos y recuperación de recursos por parte de agentes software en un único repositorio.
- CU 6 – Un repositorio de agregación recoge meta-datos de múltiples repositorios y rellena su propio repositorio.
- CU 7 – Servicio de suscripción que alertará al actor cuando algún cambio se produzca en los metadatos del recurso de un repositorio
- CU 8 – Servicio de suscripción que alertará al actor cuando algún cambio se produzca en los metadatos del recurso de múltiples repositorio

Resulta importante destacar cómo la especificación en el 2003 otorga un peso fundamental a la necesidad de realizar búsquedas sobre los repositorios, bien por parte de un actor humano o por un agente software. Aunque esta especificación sólo muestra una estructura a seguir para el desarrollo de sistemas de búsqueda federada entre diferentes repositorios posteriores especificaciones de IMS explican con más detalle alguno de diferentes aspectos que este documento simplemente enumera. Así en 2005 publican la especificación IMS General para los Servicios Web [IMS GWS, 2005].

Son muchas las cuestiones que planteaba esta especificación y que aún hoy en día permanecen como problemas abiertos. Tales como los mostrados en los casos de uso 3, 4 y 5, y que engloban la motivación de este trabajo de investigación.

Aunque el uso de plataformas de aprendizaje es cada vez más común en la mayoría de las organizaciones educativas, y el número de objetos de aprendizaje en línea, de forma gratuita o por suscripción, es enorme, la mayoría

de estos objetos de aprendizaje no son visibles directamente. Se encuentran en la denominada web invisible [Bergman, 2001], lo que dificulta su uso potencial y la reutilización. Mejorar el acceso a estos objetos de aprendizaje tendría un impacto significativo en el aprendizaje.

A.3.3 IMS Learning Object Discovery & Exchange (LODE)

En los últimos años, IMS ha producido especificaciones que hacen más fácil el intercambio y la reutilización de objetos de aprendizaje de propósito general tales como QTI, *Common Cartridge*, *Content Packaging*, *Simple Sequencing*, *Learning Design*. Todas estas especificaciones permiten a determinados tipos de objetos de aprendizaje facilitar el ser transferidos entre los sistemas y su reutilización, pero no abordan la cuestión de cómo se puede encontrar dicho contenido. IMS también ha producido las especificaciones, tales como IMS LRM y VDEX para ayudar al descubrimiento de dichos recursos mediante la unificación de las descripciones de este contenido. La pieza que falta en esta colección de especificaciones es un protocolo para apoyar el descubrimiento y el intercambio de todos estos contenidos interoperables.

El IMS ha sacado recientemente la especificación denominada LODE [LODE, 2010], del inglés *Learning Object Discovery & Exchange*, Descubrimiento de Objetos de Aprendizaje e Intercambio. Dicha especificación tiene por objeto facilitar el descubrimiento y recuperación de objetos de aprendizaje almacenados en más de una colección.

LODE se basa en los siguientes supuestos:

- Los objetos de aprendizaje son descritos por metadatos tales como IEEE LOM o Dublin Core.
- Varias instancias de metadatos pueden ser necesarias con el fin de describir adecuadamente todos los aspectos de un objeto de aprendizaje (es decir, con el fin de proporcionar la información necesaria para apoyar los casos de uso LODE).
- Los metadatos pueden ser reunidos para crear catálogos de búsqueda de objetos de aprendizaje.
- Consultar dichos catálogos de metadatos es la principal manera de obtener la información necesaria para buscar objetos de aprendizaje, evaluar su utilidad, y recuperarlos.
- Los Catálogos de metadatos se almacenan en repositorios.
- En los repositorios se pueden realizar búsquedas mediante programación mediante el uso de APIs estándares tales como *Simple Query Interface (SQI)* o *Search/Retrieve with URL (SRU)*.
- Los grandes catálogos pueden ser creados por la recolección (es decir, copia de) los metadatos almacenados en los repositorios utilizando protocolos tales como la Iniciativa de Archivos Abiertos - Protocolo para la recolección de metadatos (OAI-PMH).

IMS LODE puede ser visto como una especificación resultado de la unión de los perfiles de otros protocolos de uso general. Su finalidad es extender su uso a fin de tener en cuenta los requisitos específicos para el ámbito educativo, en lugar de crear nuevos protocolos. Se proponen tres modelos de datos principales:

- Un conjunto de contextos LODE para el Lenguaje de Consulta Contextual (*Contextual Query Language, CQL*) [CQL, 2008]: un modelo de datos con los atributos de los objetos de aprendizaje, que puedan ser utilizados para expresar la búsqueda mediante consultas significativas en el ámbito educativo.
- Un modelo de datos, denominado Información para el Intercambio de Objetos de Aprendizaje (*Information for Learning Object eXchange, ILOX*) [Massart *et al.*, 2010], que organiza conjuntos de metadatos en los objetos de aprendizaje para ser utilizados en el intercambio de datos, y
- Un modelo de datos, denominado Modelo de Datos del Registro del Repositorio de Objetos de Aprendizaje, del inglés *Learning Object Repository Registry Data Model*, para colecciones de objetos de aprendizaje, para ser utilizado en el descubrimiento y configuración de acceso a dichas colecciones.:

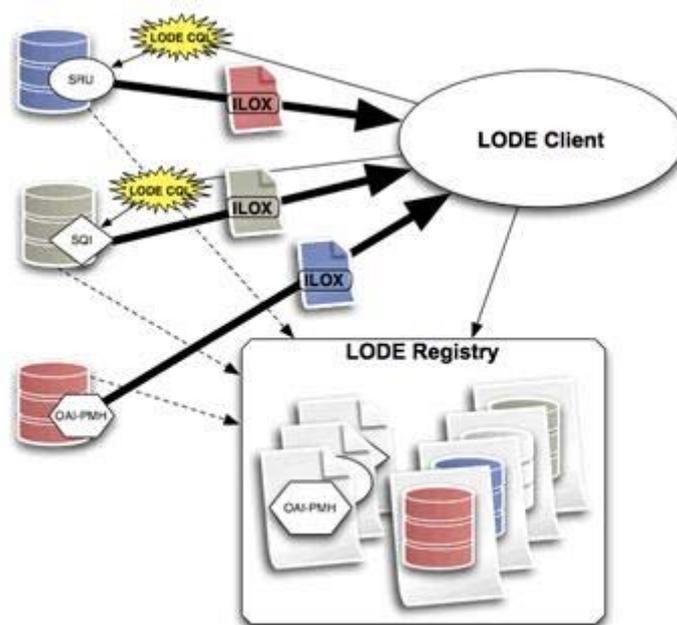


Figura A-7. Funcionamiento de IMS LODE [LODE, 2010]

La Figura A-7 muestra cómo la especificación IMS LODE se puede combinar con otras especificaciones para permitir a un cliente LODE (es decir, un sistema que se basa en la especificación IMS LODE descubrir y acceder a los objetos de aprendizaje) para obtener finalmente tales objetos de aprendizaje.

Los objetos de aprendizaje son descritos por los metadatos almacenados en los repositorios. Las funcionalidades de búsqueda y/o recolección de los metadatos son provistas en cada repositorio a través de diferentes protocolos, tales como ICS, SPI, OAI-PMH, etc. Con el fin de tener acceso a estos metadatos, el primer paso es descubrir los repositorios en los que se almacenan.

El modelo de Datos del Registro de IMS LODE proporciona una forma de describir repositorios, sus colecciones de objetos de aprendizaje y metadatos, y los protocolos que soportan. Esto permite el registro de los repositorios en un registro central que puede ser accedido por los clientes LODE. Al consultar el Registro LODE, un cliente LODE obtiene descripciones de repositorio conteniendo toda la información necesaria para conectarse automáticamente a los repositorios y obtener acceso a sus colecciones de metadatos.

Para los repositorios que implementan protocolos de búsqueda como SQI [SQL, 2005], o SRU [SRU, 2008], el LODE CQL (LOSW Context Set for CQL) permite a los clientes LODE expresar consultas en términos de atributos de los objetos de aprendizaje.

Por último, cualquiera que sea el protocolo utilizado por un cliente LODE para obtener metadatos (búsqueda y recolección), utilizando LODE ILOX para organizar las distintas instancias de metadatos devueltas por este protocolo, se garantiza que toda la información necesaria para acceder a los objetos de aprendizaje está presente y bien organizada, sin confundir el tipo de objeto de aprendizaje descrito, y se puede manejar fácilmente.

La especificación LODE tiene por objeto facilitar el descubrimiento y recuperación de objetos de aprendizaje almacenados en más de una colección, señalando aspectos interesantes que no hacen sino poner de relieve el grave problema de la interoperabilidad para los contenidos educativos. Entre otras cuestiones la propuesta presenta un marco genérico, (ILOX), para la organización de múltiples especificaciones de metadatos en un contenedor que puede ser manejado como un todo. Son los primeros pasos de una nueva propuesta, de momento reescribiendo arquitecturas basadas en servicios web para aplicarlas a los objetos de aprendizaje. El grupo de IMS LODE está trabajando en la actualidad en un perfil de IMS de ILOX para objetos de aprendizaje, que espera que esté listo en a lo largo del presente año 2011.

A continuación revisamos una serie de protocolos de comunicación para establecer la capa de comunicación entre un cliente en la extracción de contenidos educativos residentes en repositorios de objetos de aprendizaje.

A.4 PROTOCOLOS DE COMUNICACIÓN CLIENTE- ROA

Los metadatos que describen los contenidos educativos, son gestionados por los repositorios. Cualquiera que sea el mecanismo de gestión de dichos objetos de aprendizaje se necesita un protocolo de comunicación para la búsqueda de metadatos en dichos repositorios de objetos de aprendizaje. Entre los estándares más importantes están el Z39.50, el SRU, OAI-MPH o el SQL, que a continuación se detallan.

A.4.1 Protocolo Z39.50

El protocolo Z39.50 se refiere al estándar internacional, ISO 23950: "Information Retrieval (Z39.50): Application Service Definition and Protocol Specification", y a la norma ANSI Z39.50 / NISO-Z39.50-2003 [Z39.50, 2003]. La norma especifica un protocolo cliente/servidor para buscar y recuperar información de bases de datos remotas. Dicha asociación cliente-servidor, en el ámbito del protocolo se denomina como asociación Z. Los servicios Z39.50 son llevados a cabo mediante el intercambio de mensajes entre el cliente y el servidor. El objetivo de este estándar es facilitar la interconexión de clientes y servidores para aplicaciones donde los clientes busquen y recuperen información de servidores de Bases de Datos.

La búsqueda con este protocolo es realizada mediante una cadena en la que se especifican valores a ser comparados contra la BD. El subconjunto de resultados es llamado conjunto resultado y está disponible desde el servidor para referencia del cliente. Dicho conjunto es una lista ordenada de elementos, cada uno con un apuntador a un registro de la BD.

En el contexto del protocolo, el funcionamiento del mismo se efectúa mediante lo que denominan facilidades; es decir, las funciones conocidas para recuperar los registros de la base de datos. Dichas facilidades son descritas a continuación:

1. Inicialización. El cliente Z (origen) se conecta con el servidor y sugiere los parámetros básicos para la sesión. El servidor Z puede modificar estos parámetros, y si cliente y servidor llegan finalmente a un acuerdo comienza la asociación Z.
2. Búsqueda. La estrategia de búsqueda es una cadena de datos y parámetros asociados que definen los registros que se quieren recuperar como por ejemplo la base de datos a buscar, etc.
3. Recuperación. La recuperación comprende dos servicios: presentar y segmentar. Presentar es una petición al servidor Z para que envíe ciertos registros, y segmentar es el proceso de dividir un número

extenso de registros es grupos pequeños para que la transmisión sea más fácil.

4. Borrado de conjunto resultado. El servicio permite borrar un juego de resultados en el origen, el Servidor Z dará una respuesta acorde.
5. Control de acceso. El control del acceso construido a través del Z permite al servidor no autorizar virtualmente el acceso a operaciones pedidas por el cliente. Si el control de acceso no funciona, la asociación Z puede continuar con las operaciones que estén permitidas llevar a cabo o finalizar. El control de acceso puede realizarse sobre registros individuales si se requiere, así una biblioteca puede tener registros que no proveerá sin autorización, por razones de seguridad, pagos, etc.
6. Control de recursos/cuentas. Este servicio es enviado por el servidor como parte de una operación específica o de la asociación Z global.
7. Ordenar: Es llevada a cabo en el servidor, algunos clientes tienen implementada esa opción y pueden ordenar y filtrar la información después de recuperar los registros. Esta forma de trabajo es más fácil, más rápida y más flexible.
8. Visualizar. El servicio individual se denomina "scan" (consulta por índices o claves), y permite explorar u hojear una lista de autores o encabezamientos de materia. Esto contrasta con una búsqueda típica Z39.50, donde se recupera un conjunto de resultados en respuesta a una estrategia de búsqueda.
9. Servicios extendidos. Los servicios extendidos permiten al cliente Z aplicar paquetes de tareas sobre el servidor Z y controlar cómo van a operar.
10. Información sobre el servidor. El cliente Z puede consultar esta base de datos y descubrir los servicios concretos que ofrece el servidor así como sus características básicas.
11. Terminación. permite a un cliente Z o a un servidor Z interrumpir una asociación Z dando una razón para este cierre, por ejemplo, problemas en el sistema, límite de costes, violación de la seguridad, etc.

La potencialidad de Z39.50 es que separa la interfaz de usuario en el cliente de los servidores de información, los motores de búsqueda y bases de datos. Z39.50 ofrece una visión consistente de la información sobre una amplia variedad de fuentes, y ofrece a las implementaciones del cliente la capacidad de integrar la información de una serie de bases de datos y servidores. Son muchos los beneficios que presenta el protocolo, algunos de los más relevantes para este trabajo son:

- Los clientes modernos Z pueden enviar peticiones a varias bibliotecas simultáneamente ya sea la misma petición o diferentes. Esta característica permite un gran ahorro de tiempo al realizar búsquedas de elementos poco comunes o de un gran número de registros.
- El formato básico de intercambio es MARC. El cliente Z ofrece registros MARC para visualizar y realizar otros procesos posteriores. Todas las

bibliotecas "negocian" con registros bibliográficos. EL Z39.50 abre este mercado estandarizando las funciones básicas de búsqueda y recuperación de la información.

- Los servicios extendidos para la petición de documentos, actualización de bases de datos y almacenamiento de búsquedas pueden ser definidos y controlados mediante Z39.50.

A.4.2 *Search/Retrieve via URL (SRU)*

SRU es un protocolo de búsqueda estándar basado en XML para realizar consultas y búsqueda en el ámbito de Internet, utilizando CQL (*Contextual Query Language*) [SRU, 2010]. SRU es un protocolo de búsqueda y recuperación que utiliza servicios de Internet y la Web para llevar los mensajes entre el usuario y el destino. Fue definido por la Biblioteca del Congreso de EE.UU. sobre la base del protocolo Z39.50 y disponible a través de su página web⁶⁰. La iniciativa SRU busca llevar la funcionalidad Z39.50 a la Web y se basa fielmente en sus consultas y respuestas estructuradas. Por tanto, gran parte de la funcionalidad del protocolo SRU se deriva del protocolo anterior, sin embargo, sólo fueron tomadas las formas más útiles de este, y en una forma simplificada [Hammond, 2010]. En el momento en que comenzó el desarrollo de SRU, el uso de mecanismos de búsqueda similares era investigado en varias instituciones, en particular por el personal de la Biblioteca Real de los Países Bajos. Un grupo internacional de expertos de la comunidad Z39.50 colaboraron en el desarrollo de este nuevo protocolo en el ámbito de Internet/ Web / XML. Las especificaciones SRU se publicaron por primera vez en 2002 y ha sido popular por su uso en nuevas aplicaciones debido a la facilidad de aplicación.

SRU es muy flexible. Es basado en XML y la implementación más común es la SRU a través de URL, que utiliza el protocolo HTTP GET para la transferencia de mensajes. Otras versiones, sin embargo, se pueden ejecutar sobre el protocolo SOAP (SRU a través de SOAP), apoyado en las funcionalidades de servicios web, y a través de HTTP POST (SRU mediante POST), lo que evita algunas restricciones sobre la longitud y limitaciones sobre conjuntos de caracteres que están presentes con HTTP GET. Los registros devueltos en respuesta a una búsqueda pueden estar en cualquier formato XML bien definido. Lo que posibilita el recuperar metadatos de cualquier repositorio que implemente dicho protocolo.

A.4.3 *The Open Archives Initiative Protocol for Metadata Harvesting (OAI-MPH)*

⁶⁰Z39.50 <http://www.loc.gov/standards/sru/>

Este protocolo denominado *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-MPH), tiene como principal objetivo proveer un marco de trabajo independiente de aplicación para el “cosechado” o recolección (*harvesting*, en inglés) de documentos o recursos, lo que permite la interoperabilidad entre sistemas. Actualmente, la versión de este protocolo es la 2.0 [OAI-PMH, 2008] que fue propuesta el 14 de junio de 2002. La comunicación de dicho protocolo se hace a través de HTTP y los metadatos a transmitir vía OAI-PMH habrán de estar etiquetados en Dublin Core.

Dentro del protocolo OAI-MPH se definen dos tipos de participantes (1) El Proveedores de Datos, que implementa el protocolo como un medio para ofrecer los metadatos que describen sus recursos a otros sistemas y por otro lado (2) el Proveedores de Servicios, que realiza el servicio de búsqueda de metadatos a través de OAI-MPH.

Para comprender esta especificación es necesario definir algunos conceptos clave que veremos a continuación:

- Un recolector (*harvester*) es una aplicación cliente que utiliza el protocolo OAI-MPH como medio para realizar peticiones.
- Un repositorio es un servicio accesible a través de Internet que procesa peticiones en OAI-MPH. Dentro de los recursos almacenados por un repositorio se distinguen tres tipos diferentes de metadatos:
 - *Resource*. Un recurso es un objeto (físico y digital) que es almacenado en un repositorio o en otra base de datos y además es importante señalar que está fuera del alcance de OAI-MPH.
 - *Item*. Es un recurso u objeto que forma parte de un repositorio y cuyos metadatos pueden ser distribuidos o diseminados. Cada uno de los ítems debe tener un identificador único que identifique de forma unívoca al recurso, para ello se utiliza habitualmente la sintaxis proporcionada por las URI. Se recomienda, aunque no es obligatorio, que los repositorios organicen los ítems en conjuntos (*sets*), con el objetivo de permitir la recuperación selectiva (búsquedas selectivas).
 - *Record*. Esta entidad son metadatos en un formato específico que se encapsula mediante XML y que se envía como respuesta a una petición al repositorio mediante el protocolo OAI-MPH. Este fichero XML contiene tres elementos: *header*, *metadata*, *about*.

El protocolo OAI-MPH define dos tipos de búsquedas que además se pueden combinar para conseguir resultados más idóneos:

- Marcas de tiempo (*datestamps*). Se pueden utilizar marcas temporales para recuperar *records* que hayan sido creados, eliminados o modificados entre un determinado periodo de tiempo.

- Conjuntos (sets). Se podrá utilizar el conjunto como criterio de búsqueda de ítems.

Las peticiones a través de la especificación OAI-MPH se realizan a través del protocolo HTTP. Existen diferentes tipos de encapsulaciones dentro de este protocolo:

- HTTP *Request Format*. Puede utilizar tanto GET, como POST. Las peticiones se construyen a través de pares clave-valor (*key=value*), donde la clave es la palabra 'verb' de entre los cinco disponibles.
- HTTP *Response Format*. Respuesta HTTP a la petición anteriormente realizada, para ello, es necesario el correcto uso de los campos de la respuesta HTTP. Todas las respuesta HTTP deben contener un documento XML correctamente construido y utilizando la codificación de caracteres UTF-8.

OAI-PMH está compuesto por un conjunto de cinco servicios invocados a través de HTTP. Dichas peticiones de servicios son:

- `GetRecord`. Utilizado para recuperar una única entrada de un repositorio.

```
http://dominio.org/oai2?verb=GetRecord&identifier=oai:arXiv.org:cs/0112017&metadataPrefix=oai_dc
```

- `Identify`. Utilizado para recuperar información sobre el repositorio.

```
http://dominio.org/oai2?verb=GetRecord&identifier=oai:dominio.org:quant-ph/02131001&metadataPrefix=oai_dc
```

- `ListIdentifiers`. Es una forma corta de la petición `ListRecords`, que se verá a continuación, en la que sólo se devuelve el encabezado del conjunto de *records* recuperados.

```
http://dominio.org/OAI-script?verb=ListIdentifiers&from=1998-01-15&metadataPrefix=oldArXiv&set=physics:hep
```

- `ListMetadataFormats`. Utilizado para recuperar los posibles formatos de metadatos que maneja el repositorio.

```
http://www.dominio.edu/cgi-bin/pdataprov?verb=ListMetadataFormats&identifier=oai:dominio.edu:Dominio:text:1999.02.0119
```

- `ListRecords`. Utilizado para recuperar (*harvest*) un conjunto de records de un repositorio

`http://dominio.org/OAI-script?verb=ListRecords&from=1998-01-15&set=physics:hep&metadataPrefix=oai_rfc1807`

- `ListSets`. Utilizado para recuperar la estructura del repositorio.

`http://dominio.org/OAI-script?verb=ListSets`

Los trabajos iniciales para este protocolo se centraron en el desarrollo de marcos de interoperabilidad para la federación de archivos de *eprints*. Pero pronto quedó patente que la posibilidad de permitir el intercambio de múltiples formatos bibliográficos entre distintas máquinas utilizando un protocolo común, tenía aplicaciones más allá del ámbito bibliotecario. Este fue el motivo de su apertura al ámbito de cualquier material digital, incluyendo contenidos educativos.

A.4.4 CEN SQI (*Simple Query Interface*)

SQI, en inglés *Simple Query Interface*, presenta una interfaz de programación de aplicaciones (API) para consultar repositorios de objetos de aprendizaje. La especificación fue aprobada el 6 de septiembre de 2005 por el Comité Europeo de Normalización [CWA 15454, 2005]. Fue desarrollado por una serie de colaboradores con diferentes áreas de interés, tanto de universidades, como de empresas privadas.

El principal objetivo a la hora de diseñar SQI fue el de crear una especificación conceptualmente simple y que a la vez fuera fácil de implementar. El núcleo de SQI es un conjunto de funciones abstractas, conectadas mediante servicios web y para las que no se define ningún tipo de lenguaje de programación concreto, aunque se podría utilizarse tanto Java como PHP.

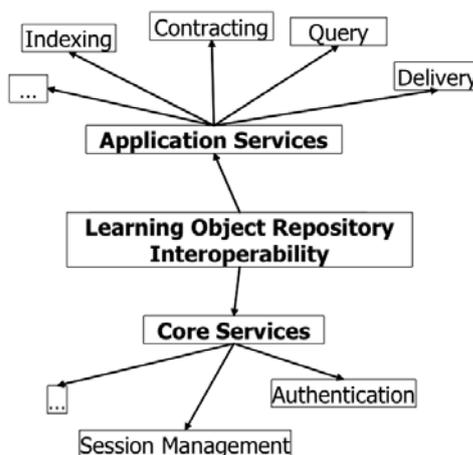


Figura A-8. Servicios de núcleo y de aplicación [CWA 15454, 2005]

La especificación de SQI, está compuesta, a su vez, por otros tres documentos:

- **Learning Object Repository Interoperability Framework** [Simon et. al, 2005]. Este documento proporciona una visión global del contexto que envuelve a la interfaz SQI.
- **Simple Query Interface Specification** [Simon et. al, 2005]. Contiene el núcleo de la especificación, es decir, las funciones que proporcionan la interoperabilidad entre repositorios.
- **Authentication and Session Management** [Simon et. al, 2005]. Documento centrado en la especificación de la gestión de sesiones y la autenticación de usuarios.

Antes de continuar, es importante señalar que la propia especificación de SQI define un repositorio de objetos de aprendizaje como *una colección de material didáctico, cursos, y objetos de aprendizaje a los que se le asocia una descripción (es decir, “metadatos”)*.

SQI diferencia entre dos tipos de servicios orientados a proporcionar interoperabilidad entre sistemas, como se aprecia en la Figura A-8.

- **Core services.** Son el núcleo del sistema, están orientados a proporcionar la funcionalidad interna del repositorio, habitualmente realizan tareas como la identificación unívoca de los objetos de aprendizaje, la autenticación de usuarios, la gestión de sesiones, etc.
- **Application services.** Utilizan los servicios de núcleo para proporcionar la interoperabilidad entre repositorios a través de los métodos propuestos por SQI (Gestión de consultas, indexación, etc.)

Tanto los servicios de aplicación, como los servicios del núcleo hacen uso de una infraestructura de red que permite que los repositorios puedan comunicarse entre sí, esta arquitectura típicamente se divide en 5 capas tal y como se muestra en la Figura 2-2.

SQI permite la interoperabilidad entre repositorios y sistemas a través de un servicio de consultas. Este servicio, pese a estar específicamente pensado para un contexto educativo, puede ser utilizado en una amplia variedad de escenarios, ya que se ha diseñado con un alto grado de abstracción, lo que facilita su adaptación.

Según describe el estándar SQI, los repositorios deben implementar un conjunto de métodos, considerados el núcleo de la especificación. Estos métodos son la base de un protocolo de comunicación simple que permite el envío de consultas y la recepción de objetos de aprendizaje.

La topología de repositorios para SQI es *peer-to-peer*, es decir, cada uno de los miembros del sistema puede actuar como servidor (*target*) o cliente (*source*).

Para que la interoperabilidad sea posible, es necesario definir tanto un lenguaje de consulta, como un formato de resultados. Una de las características clave de SQI es que no pone ninguna restricción en este sentido, por lo que cualquier lenguaje de consulta o formato de resultados es válido, siempre que lo sea para los dos repositorios que están compartiendo información. Así mismo, tampoco limita otras características que dificultarían la interoperabilidad entre los repositorios como pueden ser: método de almacenaje, búsqueda interna, indización, etc. De modo que el foco del SQI, se centra en el conjunto de servicios base para conectar los repositorios o realizar todo tipo de aplicaciones para realizar las búsquedas federadas.

La Figura A-9 muestra un proceso típico de intercambio, en el que un cliente/repositorio A realiza la consulta a un repositorio B. En ella, se puede apreciar como el cliente realiza una consulta al repositorio (en un lenguaje común), éste realiza la correspondiente consulta interna y devuelve los resultados (formato común).

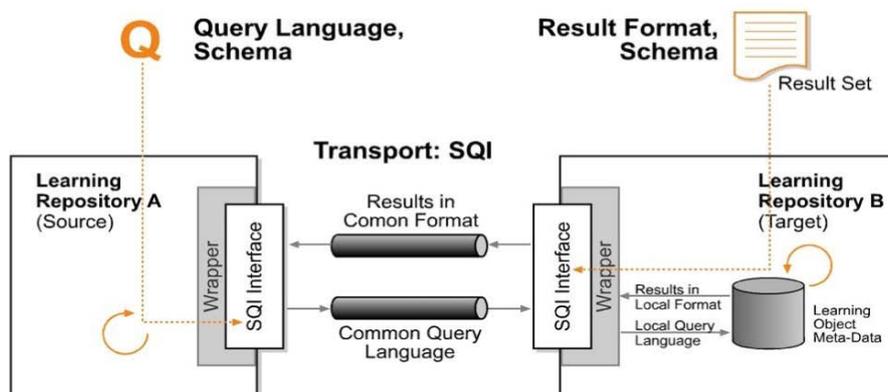


Figura A-9. Intercambio de información entre dos repositorios a través de SQI [CWA 15454, 2005]

El mayor objetivo seguido durante el diseño del conjunto de funciones que forman el estándar SQI era el de conseguir un estándar conceptualmente sencillo y a la vez fácil de implementar. Para ello, se adoptaron los siguientes principios de diseño:

- Como ya se ha visto de forma previa, SQI es neutral, tanto a nivel de resultados, como de lenguaje de consulta, debido a que los repositorios son altamente heterogéneos.

- Soporta tanto consultas síncronas, como asíncronas, con el objetivo de permitir la aplicación del estándar en diferentes situaciones.
- Soporta tanto acceso dinámico mediante sesión (*stateful*), como sin sesión (*stateless*).
- Está basado en un sistema de sesión, con el objetivo de separar la autenticación, de la propia gestión de sesiones.

El conjunto de funciones definidas puede ser fácilmente extensible, para ello se ha minimizado tanto el número de métodos, como el número de parámetros de éstos. Los cambios que se puedan producir en la interfaz SQI sólo requerirían la inclusión de algún método nuevo, de manera que no interferiría con los métodos ya implementados.

Por otro lado, también se sigue el principio de separar los métodos de comando o configuración de los métodos de consulta. Según este diseño, un método sólo devuelve información o datos de control, por lo que simplifica el desarrollo de interfaces.

Tal y como se ha comentado SQI puede implementarse para dos tipos de repositorios:

- Repositorios *stateful* (con sesión), es decir, repositorios que almacenan el histórico de consultas para un cliente.
- Repositorios *stateless* (sin sesión) que no tiene en cuenta la estados previos de un usuario, no guarda histórico de un usuario.

Dentro del núcleo de la especificación se pueden definir tres grandes conjuntos de métodos:

- Métodos de configuración.
- Métodos de consulta síncrona.
- Métodos de consulta asíncrona.

A.4.4.1 Métodos de configuración

- **setQueryLanguage.** Permite especificar el tipo de lenguaje que se utilizará en la consulta al repositorio. Para ello se utiliza el identificador del lenguaje, teniendo en cuenta que es sensible a mayúsculas y minúsculas.

Tabla A-4. Método `setQueryLanguage` - Especificación SQI

Método	<code>setQueryLanguage</code>		
Retorno	Void		
Parámetros	<code>targetSessionID</code>	String	Identificador de sesión

	queryLanguageID	String	Identificador de lenguaje.
Errores	NO_SUCH_SESSION QUERY_LANGUAGE_NOT_SUPPORTED METHOD_FAILURE		

- **setResultsFormat.** Permite especificar el tipo de formato en el que estarán formateados los resultados ofrecidos por el repositorio. El tipo de resultado se proporciona mediante dirección la URI que apunte al fichero que contenga la definición del esquema XML del esquema utilizado. También se puede utilizar identificadores predefinidos teniendo en cuenta que esta opción es sensible a mayúsculas y minúsculas.

Tabla A-5. Método setResultsFormat - Especificación SQI

Método	setResultsFormat		
Retorno	Void		
Parámetros	targetSessionID	String	Identificador de sesión.
	resultsFormat	String	Tipo de resultados.
Errores	NO_SUCH_SESSION RESULTS_FORMAT_NOT_SUPPORTED METHOD_FAILURE		

- **setMaxQueryResults.** Permite especificar el máximo número de resultados que se desean recibir. El número por defecto si no se utiliza este método es 100 y cuando se desea que no haya límite en el número de resultados se utiliza el valor 0.

Tabla A-6. Método setMaxQueryResults - Especificación SQI

Método	setMaxQueryResults		
Retorno	Void		
Parámetros	targetSessionID	String	Identificador de sesión
	maxQueryResults	Integer	Número máximo
Errores	NO_SUCH_SESSION INVALID_MAX_QUERY_RESULTS METHOD_FAILURE		

- **setMaxDuration.** Permite especificar el periodo de tiempo máximo (en milisegundos) que se desea que dure el proceso de búsqueda durante una consulta asíncrona. Si el número se fija a 0, entonces se entiende que no hay límite. El valor por defecto es 0 (sin límite de tiempo).

Tabla A-7. Método setMaxDuration - Especificación SQI

Método	setMaxDuration		
Retorno	Void		
Parámetros	targetSessionID	String	Identificador de sesión
	maxDuration	Integer	Máximo tiempo de espera
Errores	NO_SUCH_SESSION INVALID_MAX_DURATION METHOD_FAILURE		

A.4.4.2 Métodos de consulta síncrona

Tal y como se puede apreciar en la Figura A-10 en este escenario, el repositorio devuelve los resultados al cliente en la misma operación. Con el objetivo de facilitar la búsqueda federada, se pueden realizar varias consultas simultáneas utilizando el mismo identificador de sesión.

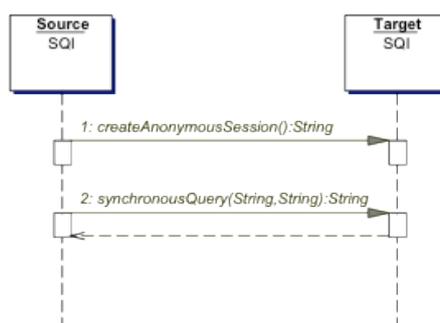


Figura A-10. Modo de consulta síncrono en la especificación SQI

- **setResultSetSize.** Permite especificar el número máximo de elementos que se desean obtener en una consulta simple. El número por defecto es 25 objetos de aprendizaje. Se utiliza el valor 0 para eliminar el límite.

Tabla A-8. Método setResultSetSize - Especificación SQI

Método	setResultSetSize		
Retorno	Void		
Parámetros	targetSessionID	String	Identificador de sesión
	resultSetSize	Integer	Número de resultados

Errores	NO_SUCH_SESSION INVALID_RESULTS_SET_SIZE QUERY_MODE_NOT_SUPPORTED METHOD_FAILURE
----------------	---

- **synchronousQuery.** Permite lanzar una consulta síncrona sobre un sistema. El método devuelve un conjunto de metadatos que concuerdan con la consulta. El índice empieza en 1, el número de resultados por consulta está determinado por `setResultsSetSize` y el número máximo de resultados está limitado por `setMaxQueryResults`.

Tabla A-9. Método `synchronousQuery` - Especificación SQI

Método	<code>synchronousQuery</code>		
Retorno	String		
Parámetros	<code>targetSessionID</code>	String	Identificador de sesión
	<code>queryStatement</code>	String	Consulta
	<code>startResult</code>	Integer	Resultado inicial
Errores	NO_SUCH_SESSION INVALID_QUERY_STATEMENT QUERY_MODE_NOT_SUPPORTED METHOD_FAILURE INVALID_START_RESULT NO_MORE_RESULTS METHOD_FAILURE		

- **getTotalResultsCount.** Permite obtener el número de resultados válidos que ha ofrecido una consulta. Sólo es válido para repositorios con sesión (*Stateful*).

Tabla A-10. Método `getTotalResultsCount` - Especificación SQI

Método	<code>getTotalResultsCount</code>		
Retorno	Integer		
Parámetros	<code>targetSessionID</code>	String	Identificador de sesión
	<code>queryStatement</code>	String	Consulta
Errores	NO_SUCH_SESSION QUERY_MODE_NOT_SUPPORTED INVALID_QUERY_STATEMENT METHOD_FAILURE		

A.4.4.3 Métodos de consulta asíncrona

La operación de consulta se inicia por el cliente hacia el repositorio. En este caso el cliente, también es necesario que implemente el servicio dónde se entregarán los resultados de la consulta. Por lo tanto, el cliente de forma previa a la consulta debería indicar el punto de acceso concreto al servicio de recepción de resultados.

Este método también permite realizar varias consultas simultáneas a través de un identificador de consulta que le tendrá que enviar el cliente al servidor y que éste le devolverá junto a la consulta en el servicio destinado a tal fin.

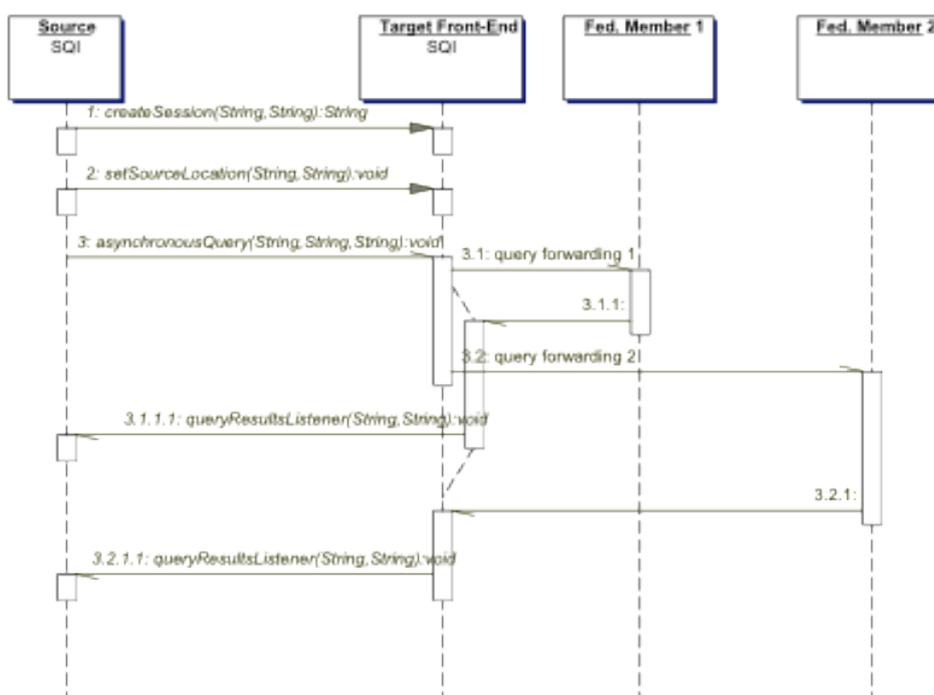


Figura A-11. SQI: Modo de consulta asíncrona [CWA 15454, 2005]

- **asynchronousQuery.** Permite lanzar consultas asíncronas de búsqueda, de forma que el sistema que realiza la llamada podrá realizar otras tareas mientras se procesa la solicitud. La localización de los resultados se realiza a través del método `setSourceLocation`. En este caso será importante el valor asignado por el método `setMaxDuration`, ya que será el máximo tiempo de demora entre la consulta y la obtención de

los resultados. El identificador de consulta (queryID) suministrado se devolverá junto con los resultados.A

Tabla A-11. Método asynchronousQuery - Especificación SQI

Método	asynchronousQuery		
Retorno	Void		
Parámetros	targetSessionID	String	Identificador de sesión
	queryString	String	Consulta
	queryID	String	Identificación del la consulta
Errores	NO_SUCH_SESSION QUERY_MOD_NOT_SUPPORTED NO_SOURCE_LOCATION INVALID_QUERY_STATEMENT METHOD_FAILURE		

- **setSourceLocation.** Este método deberá ser invocado antes de lanzar una consulta asíncrona. Será utilizado para especificar dónde está el sistema que lanza la consulta y que espera recibir los resultados. El punto de acceso se indica mediante la URL del servicio.

Tabla A-12. Método setSourceLocation - Especificación SQI

Método	setSourceLocation		
Retorno	Void		
Parámetros	targetSessionID	String	Identificador de sesión
	sourceLocation	String	Localización fuente mediante URL
Errores	NO_SUCH_SESSION QUERY_MODE_NOT_SUPPORTED METHOD_FAILURE		

- **queryResultsListener.** Método utilizado para recibir los resultados del repositorio en una consulta asíncrona.

Tabla A-13. Método queryResultsListener - Especificación SQI

Método	queryResultsListener		
Retorno	Void		
Parámetros	queryID	String	Identificador de la consulta

	queryResults	String	Resultado consulta
Errores	INVALID_QUERY_RESULTS NO_SUCH_QUERY METHOD_FAILURE		

A.4.4.4 Gestión de sesiones y autenticación

Para realizar una interacción entre un cliente y un repositorio es necesario estabilizar de forma previa una sesión. Se dice que la sesión está estabilizada cuando el cliente ha conectado correctamente con el repositorio.

Se puede obtener una sesión autenticada o anónima, en función de la privacidad que el repositorio requiera. Una vez estabilizada la sesión, el cliente recibirá un identificador de sesión (`sessionID`) que tendrá que utilizar en cada consulta con el repositorio.

Una sesión es válida hasta que se destruye, es decir, continúa abierta una vez realizada la consulta. Pero, si no se vuelve a realizar una consulta en 30 minutos, la sesión se pierde (auto-destruye). No obstante, como en cualquier sistema, se considera una mala práctica no cerrar sesiones si no se van a utilizar en el futuro. SQI, en su especificación, deja muy abierto este punto de gestión de la sesión, es por ello que muchos repositorios ni si quiera lo implementan este es el caso de los repositorios stateless cuyo identificador de sesión es único para todos los cliente.

En cuanto a los métodos para iniciar sesiones, se definen dos formas de crear sesiones:

- **Sesiones autenticadas.** Mediante un usuario y contraseña o cualquier otro tipo de credenciales.

Tabla A-14. Método `createSession` - Especificación SQI

Método	<code>createSession</code>		
Retorno	String		
Parámetros	userID	String	Identificador de usuario
	password	String	Contraseña de usuario
Errores	WRONG_CREDENTIALS METHOD_FAILURE		

- **Sesiones anónimas.** Obviamente no requieren ningún usuario ni contraseña.

Tabla A-15. Método createAnonymousSession - Especificación SQI

Método	createAnonymousSession
Retorno	String
Parámetros	--
Errores	METHOD_FAILURE

La sesión creada, ya sea identificada o anónima es válida hasta que se destruye, para destruir esta sesión la función utilizada es la siguiente:

Tabla A-16. Método destroySession - Especificación SQI

Método	destroySession		
Retorno	Void		
Parámetros	SessionID	String	Identificador de Sesión
Errores	NO_SUCH_SESSION METHOD_FAILURE		

A.4.4.5 Gestión de errores

A lo largo de los métodos que se han visto con anterioridad, cada uno de ellos estaba asociado con una serie de posibles errores. Intencionalmente, este conjunto de errores es pobre, con el objetivo de abarcar de esta forma el mayor número posible de repositorios y facilitar una vez más la máxima interoperabilidad.

Cada uno de estos errores (Tabla A-17) está compuesto por dos propiedades: (i) Mensaje, que es un texto libre que describe el error y (2) Código de error que identifica el problema.

Tabla A-17. Errores en la especificación SQI

Nombre	Código	Descripción
UNDEFINED	SQI-00000	Error no definido. Todos los métodos pueden lanzar este error.
METHOD_FAILURE	SQI-00001	El método falla aunque la llamada fuera correcta. Suele ser debido por un problema en el servidor. Todos los métodos pueden lanzar este error
NO_SOURCE_LOCATION	SQI-00002	En consultas asíncronas cuándo no se ha definido el lugar donde se enviarán los datos de forma previa a la consulta
INVALID_START_RESULT	SQI-00003	El número de resultados de comienzo es erróneo.

INVALID_QUERY_STATEMENT	SQI-00004	La consulta realizada es errónea atendiendo a la sintaxis del lenguaje.
INVALID_RESULT_SET_SIZE	SQI-00005	Número de resultados de la consulta erróneo.
INVALID_MAX_DURATION	SQI-00006	La duración máxima de la consulta es errónea.
INVALID_MAX_QUERY_RESULTS	SQI-00007	Número máximo de resultados inválido.
INVALID_QUERY_RESULTS	SQI-00008	Los resultados no pueden ser interpretados por el cliente en una consulta asíncrona.
QUERY_MODE_NOT_SUPPORTED	SQI-00009	Tipo de consulta no soportado (síncrona a asíncrona)
RESULTS_FORMAT_NOT_SUPPORTED	SQI-00010	Formato de resultados desconocido
QUERY_LANGUAGE_NOT_SUPPORTED	SQI-00011	Lenguaje de consulta no soportado
METHOD_NOT_SUPPORTED	SQI-00012	Método no soportado o implementado por el servidor (target)
NO_SUCH_SESSION	SQI-00013	Identificador de sesión erróneo
NO_SUCH_QUERY	SQI-00014	Consulta realizada inválida
WRONG_CREDENTIALS	SQI-00015	Autenticación errónea.

Anexo B ESTUDIO PRELIMINAR DE CONDICIONES

En los apartados 2.3 y 2.4 de esta memoria se presentaron varias arquitecturas contemplando soluciones a la interoperabilidad de repositorios de objetos educativos. Se detallaron distintas capas de implementación para la interoperabilidad (*middleware*) entre los ROA y sus clientes, tales como OAI-MPH, IMS DRI y SQI.

Por un lado, OAI-MPH es un estándar que permite la interoperabilidad entre sistemas fundamentalmente del ámbito de bibliotecas, pero que está extendiéndose a otros ámbitos en la actualidad pero no está directamente diseñado para intercambiar OA, aunque existan algunos ROA que lo implementen, pero son mínimos, por lo que no se incluirá en este estudio.

Por el otro lado, tenemos los estándares IMS DRI y SQI, específicamente diseñadas para ser utilizadas en un contexto educativo. Ambas especificaciones están formalmente aprobadas por organismos de reconocido prestigio (IMS y CEN, respectivamente). IMS DRI está en las primeras fases de la especificación y actualmente todavía no existe ningún repositorio que lo implemente. Por el contrario, SQI está totalmente definido y además existe un elevado número de ROA que sí lo implementan, de modo que sobre él se van realizar búsquedas masivas de objetos de aprendizaje y se analizará su funcionamiento así como las posibilidades en la recuperación de OA en los ROA que lo implementan.

Este estudio, que continúa el ya realizado en [Prieta & Corchado, 2010] añadiéndole medidas y estimaciones, se centrará en aquellos ROA que implementan la capa de interoperabilidad SQI. Se explorarán las características de cada ROA, realizando un análisis comparativo entre ellos.

Los repositorios enumerados en la Tabla 5-1, considerados los más relevantes en la actualidad, serán objeto de este estudio.

OpenDOAR⁶¹, a través de la web, proporciona una lista de calidad garantizada de repositorios de acceso abierto en todo el mundo, albergando todo tipo de información.

⁶¹ OPENDOAR <http://www.opendoar.org/>

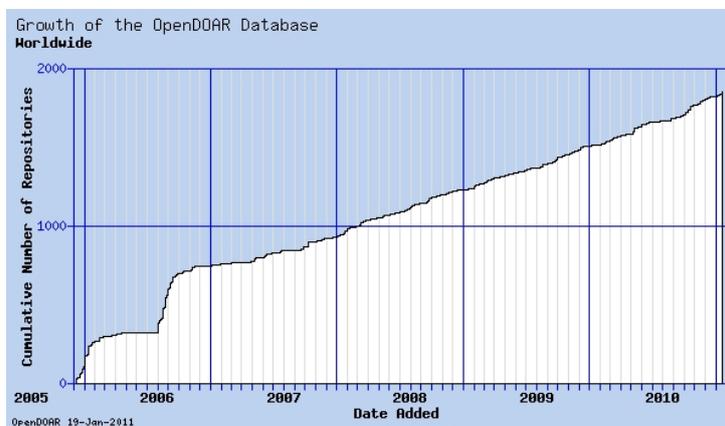


Figura B-12. Crecimiento de los repositorios según OpenDOAR

Sólo el 15% de los contenidos de los repositorios que indexa este buscador son Objetos de Aprendizaje. El buscador, tiene varias herramientas muy interesantes. “Policies Tool” indicando los criterios o políticas seguidas por el repositorio para el uso y depósito de ficheros, reutilización de metadatos y de preservación. La otra, “Opendoar Search”, es una herramienta para la búsqueda de texto completo en los contenidos de los repositorios que se encuentran en su directorio y está basada en la aplicación Google Custom Search Engine. Otra herramienta muy interesante por estar actualizándose en tiempo real es la de creación de gráficos, “OpenDOAr Charts”, que permiten analizar perfiles de repositorios por distintas variables (país, software, idioma, tipo de documento...), como la que muestra la Figura B-12. Tiene además un API que permite lanzar peticiones contra la base de datos de OpenDOAR con respuestas en ficheros XML. Resulta una fuente interesante de información actualizada de repositorios de acceso abierto.

El estudio se divide en tres fases que se detallan a continuación:

1. Análisis de los ROA, tanto a nivel individual como en su conjunto.
2. Realización de prueba de rendimiento de los ROA en la recuperación de metadatos
3. Análisis de los metadatos de los OA recuperados.

Finalmente, una vez se tienen los resultados del estudio se presentarán formalmente las carencias detectadas y se elaborarán las conclusiones.

B.1 FUNCIONAMIENTO DE LOS ROAS REALES

Resulta fundamental el conocer el funcionamiento real de los ROAs, indagando sobre las funcionalidades y características operativas. Según hemos visto en la sección A.4.4, El Comité Europeo de Normalización publicó en el 2005 la especificación Simple Query Interface (SQI), con el fin de facilitar la interoperabilidad entre los repositorios públicos y las aplicaciones que hicieran uso de ellos. Desde entonces, los principales repositorios de objetos de aprendizaje han ido adaptándose para cumplir con esta especificación. Este apartado presenta los resultados la evaluación del nivel de cumplimiento de la especificación SQI un conjunto de repositorios públicos disponibles, Tabla 5-1, así como otra serie de características:

- **Nombre.** Es el nombre por el que cada proyecto o grupo identifica a su repositorio.
- **Tipo de sesión.** Se distinguen dos tipos de sesiones diferentes:
 - *Stateful.* Cuando hay que solicitar un identificador de sesión para cada consulta (o consultas) de un mismo usuario.
 - *Stateless.* Cuando el identificador de sesión es el mismo para todas las consultas y usuarios.
- **Lenguaje de consulta.** Lenguaje mediante el cual se realiza la consulta al repositorio.
- **Formato de los resultados.** Esquema en el que se devuelven los resultados el ROA después de una consulta.
- **Tipo de conexión.** Indica si el repositorio permite conexiones síncronas y/o asíncronas.

Tabla B-18. Características básicas de enlace del ROA

Repositorio	Tipo sesión	Lenguaje	Formato resultados	Tipo conexión
Acknowledge	Stateless	vsqI	LOM	Síncrona
Agrega	Stateful	vsqI	LOM	Síncrona
Ariadne	Stateful	vsqI	LOM	Ambas
AriadneNext	Stateless	vsqI	LOM	Síncrona
CGIAR	Stateless	vsqI	LOM	Ambas
EducaNext	Stateful	vsqI	LOM	Síncrona
LACLO-FLOR	Stateful	vsqI	LOM	Ambas
LORNET	Stateful	vsqI	LOM	Síncrona
MACE	Stateless	vsqI	LOM	Síncrona
Merlot	Stateless	vsqI	LOM	Síncrona
Nime	Stateless	vsqI	LOM	Síncrona
OER Commons	Stateless	vsqI	LOM	Síncrona

Edna Online	Stateless	vsq1	LOM	Ambas
--------------------	------------------	-------------	------------	--------------

A tenor de estos primeros datos preliminares, se pueden extraer las siguientes conclusiones iniciales:

- **Respecto al esquema de los resultados.** En este sentido la uniformidad es la regla, todos los repositorios utilizan IEEE LOM como esquema para la descripción de los OA.
- **Respecto al lenguaje de consulta.** Del mismo modo, también todos utilizan el lenguaje VSQ1 [Simon *et al.*, 2005].
- **Respecto al tipo de sesión.** El 35% es de tipo de *statefull*, mientras que el resto no utiliza gestión de sesión individual para cada usuario (*stateless*).
- **Respecto al tipo de conexión.** Tal y como se aprecia en la **¡Error! No se encuentra el origen de la referencia.**, todos los repositorios implementan el tipo de conexión síncrona, y además existe un pequeño porcentaje (4 de 14) que implementa también conexión asíncrona. No se da la situación en que un repositorio implemente conexión asíncrona y no disponga de conexión síncrona.

Como se ha visto anteriormente, la interfaz SQI permite tanto peticiones síncronas como asíncronas, pero este estudio solo abarcará las conexiones de tipo síncronas, ya que no tiene sentido utilizar conexiones asíncronas en un sistema del que se esperan resultados en tiempo real. Para ello se hace uso de un cliente de servicio web SQI (a través de la especificación WSDL).

En el segundo paso del estudio se comprueba la corrección en el funcionamiento de los repositorios y la implementación de todas las funciones que especifica el estándar SQI. Las funciones que incluye este estudio son las siguientes:

- **Métodos de configuración:** [1] setQueryLanguage, [2] setResultFormat, [3] setMaxQueryResults

- **Gestión de sesiones.** (Sólo se comprobará en servidores *stateful*): [4] createSession, [5] createAnonymousSession, [6] destroySession

- **Consulta síncrona:** [7] setResultSetSize, [8] synchronousQuery, [9] getTotalResultsCount

Tabla B-2. Análisis de ROA - Implementación de métodos

Repositorio	Configuración			Gestión de Sesiones			Consulta síncrona		
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Acknowledge	Sesión inválida			Stateless			Ses. inv.	OK	OK

Agrega	OK	OK	OK	OK	OK	OK	OK	OK	OK
Ariadne	Repositorio no disponible								
AriadneNext	Repositorio no disponible								
CGIAR	OK	OK	OK	Stateless			OK	OK	OK
EducaNext	Repositorio no disponible								
LACLO-FLOR	No disponible			NO	OK	OK	No disponible		
LORNET	OK	No disp.		OK	OK	OK	NO	OK	OK
MACE	Sesión no válida			Stateless			Sesión no válida		
Merlot	OK	OK	OK	Stateless			OK	OK	OK
Nime	Repositorio no disponible								
OER Commons	Sesión no válida			Stateless			Sesión no válida		
EdNa Online	Repositorio no disponible								

Una vez realizada esta prueba de conexión el panorama cambia radicalmente, ya que 5 de los 14 repositorios que abarca el estudio no funcionan o no están disponibles en el momento concreto en el que se realiza el test, Junio de 2010, por lo que se eliminan del presente estudio inicial. Estos repositorios son: Ariadne, AriadneNext, EducaNext, Nime y EdNa Online. No obstante, en ocasiones anteriores y posteriores los repositorios Ariadne, Edna Online y EducaNext han funcionado correctamente.

Los repositorios que aparentemente funcionan sin ningún problema son todos *stateless*. De éstos, existen dos repositorios (MACE y LACLO-FLOR) que pese a realizar las correspondientes comprobaciones mediante un identificador de sesión válido, indican el error SQI_00013, que se corresponde con un identificador inválido. Dado que se está utilizando un identificador válido y aún así el repositorio no responde correctamente, se asume el funcionamiento erróneo del repositorio y se apartan del estudio.

Después del eliminar los repositorios con comportamiento anómalo, sólo quedan disponibles los repositorios Acknowledge, Agregas, LORNET y Merlot. Los dos últimos son perfectamente válidos, funcionan con corrección y además implementan todos los métodos de la interfaz SQI. Por el contrario, el repositorio Acknowledge sólo implementa los métodos `synchronousQuery` y `getTotalResoultCount`, que son los esenciales para poder realizar consultas a través de la especificación SQI; para el resto de métodos devuelve el error SQI_00013 al igual que pasaba con MACE y LACLO-FLOR.

B.2 RENDIMIENTO DE LOS ROA

Es importante manejar estadísticas sobre el número de OA, así como la facilidad o no de su recuperación. Para ello se desarrolla una primera aplicación, que conecta a través del uso del API SQI con diferentes ROAs. Esta primera

aplicación denominada BRENHET, permitía realizar la conexión con los repositorios a través de SQL, así como realizar estadísticas de conexión y volúmenes de recuperación [Prieta & Gil, 2010]. Como resultado de las estadísticas para 10 búsquedas durante varias consultas en el segundo trimestre de 2009 se obtiene la siguiente tabla que muestra los valores medios de tiempos necesitados para la recuperación de contenidos:

Tabla B-3. OA recuperados y tiempos medios

Repositorio	OA Recuperados	Tiempo (ms.)
Ariadne	10,9	2476,1
EducaNext	4,8	18255,9
Lornet	0	5006
Agrega	0	3731,8
LACLO-FLOR	0	3119,1
Acknowledge	0	326,9
Acknowledge News	0	298,1
Ariadne Next	0	526,5
CGIAR	3,7	297,9
EdNa Online	18,3	18294
Merlot	7,6	8477,6

Los resultados muestran elevados tiempo de respuesta para muchos repositorios que en más de la mitad de repositorios consultados ni siquiera se obtienen resultados.

En una segunda fase del estudio se analiza de modo más preciso, el proceso de recuperación de objetos de aprendizaje de trece repositorios operativos a lo largo del primer semestre de 2010. Se evalúa el rendimiento de dichos repositorios en base a parámetros relevantes en el ámbito de la recuperación de información y descritos en el apartado 3.2. Dicho estudio adapta diversas técnicas y metodologías de la recuperación de información mediante el ajuste de indicadores habituales como son la exhaustividad y precisión a través del empleo de análisis estadístico así como la cuantificación de errores en los sistemas. Se propone a su vez un nuevo indicador denominado ganancia temporal que establece una medida de los objetos de aprendizaje recuperados de un repositorio por unidad de tiempo.

B.2.1 Análisis de Errores

En algunos casos, como resultado de las consultas se producen errores. Un conjunto de dieciséis errores es detallado en la especificación SQL, resumidos en la Tabla B-4.

Tabla B-4. Códigos de error SQL

CODE	DESCRIPTION
SQL-00000	Undefined
SQL-00001	Method failure
SQL-00002	No source location
SQL-00003	Invalid start result
SQL-00004	Invalid query statement
SQL-00005	Invalid results set size
SQL-00006	Invalid max duration
SQL-00007	Invalid max query results
SQL-00008	Invalid query results
SQL-00009	Query mode not supported
SQL-00010	Results format not supported
SQL-00011	Query language not supported
SQL-00012	Method not supported
SQL-00013	No such session
SQL-00014	No such query
SQL-00015	Wrong credentials
SQL-00016	No more results

Evidentemente los errores en el transcurso de la búsqueda y recuperación implican a muchos procesos además del uso del SQL, que como hemos visto en la sección 0 en muchos casos ni siquiera está totalmente operativo en los repositorios. Realizando un estudio sistemático de todos los ROA en las diferentes consultas en las que se ha devuelto error y una vez analizadas, se han agrupado los errores en 4 tipos diferentes generales, etiquetados como:

1. Tipo 1. No se puede conectar con el servicio encargado de la gestión de sesiones.
2. Tipo 2. Se puede conectar al servicio encargado de la gestión de sesiones, pero no se puede obtener sesión.
3. Tipo 3. No se puede conectar con el servicio encargado de la gestión de consultas.
4. Tipo 4. Se puede conectar al servicio encargado de la gestión de consultas, pero no se puede realizar una consulta.

Para realizar esta prueba se ha realizado una batería de de 60 consultas en cada uno de los repositorios enumerados. Como punto de partida de dichas consultas se han tomado nuevamente 60 códigos al azar dentro del campo de la ciencia y la tecnología dentro de los códigos UNESCO⁶². Dichos términos son los siguientes:

⁶² UNESCO <http://www.unesco.org/new/en/unesco/>

Formalized languages, Formal Systems, Recursive Functions, Computer Sciences, Algorithmic languages, Analog computing, Artificial intelligence, Codes and coding Systems, Computer software, Digital computing, Heuristics, Hybrid computing, Simulation, Algorithm construction, Iterative methods, Control Systems, Cybernetics, Dynamic Programming, Linear Programming, Computing for statistics, Speech Sensor systems design, Storage devices, Circuit design, Processing engineering, Automation technology, Timing devices, Telecommunications, Radiocommunications, Machine translation, Architectural design, Bioacoustics, Educational theories, Cardiology, Allergies, Pulmonary diseases, Ethnolinguistics, Semiology, Dermatology, Pediatrics, Control devices, Data transmission devices, Real-time Systems, Central processing units, Computer architecture, Surgery, Fibre optics, Biochemical genetics, Microwave links, Computer technology, Informatics, Computing systems design technology, Arithmetic and machine instructions, Information systems design and components, Programming languages, Programming theory, Character recognition systems, Broadcasting, sound and television, Computer terminals, Graphie , Display devices and plotters

En cada una de las 60 consultas y para cada repositorio se recoge además de los metadatos de cada uno de los OA devueltos, diferente información que ayudará en la evaluación del rendimiento de dichos repositorios en cuanto al número de errores producidos, los tiempos de acceso al repositorio, tiempos de consulta, el número de resultados que devuelven, etc.

La Figura B-2 representa para los 13 repositorios evaluados, los errores tipificados en el estudio realizado en junio de 2010 así como la no existencia de ningún error. Se observa cómo 4 de los 13 repositorios funcionan correctamente, éstos son Acknowledge, Merlot, LORNET y Agrega; el resto de repositorios presenta un nivel intolerable de errores. Cabe destacar que en 8 de ellos el error es de tipo 4, es decir, se puede establecer conexión, pero el servicio de consultas no funciona con corrección, nos referimos a los repositorios Nime, AriadneNext, Edna Online, CGIAR, MACE, OER Commons y LACLO-FLOR.

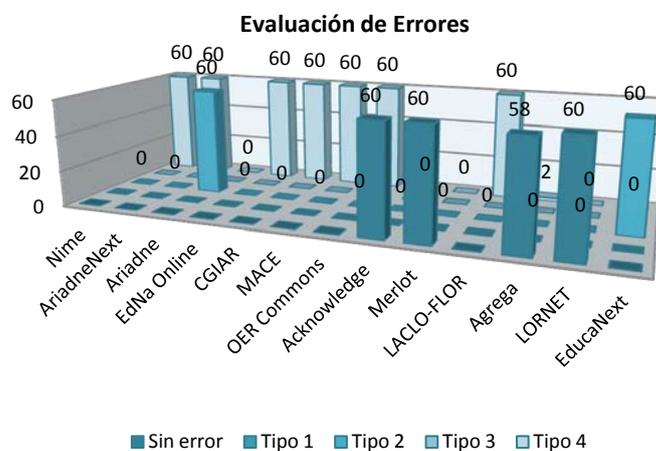


Figura B-2. Errores en los repositorios Junio de 2010

Parece que no hay término medio en cuanto al funcionamiento correcto de los repositorios se refiere, es decir, los repositorios por lo general funcionan, entendiendo que devuelven respuestas a las consultas o no funcionan. Otro asunto es el tipo de respuestas devueltas y la calidad de las mismas en relación a las consultas realizadas.

Resultados Sin Errores en los ROA

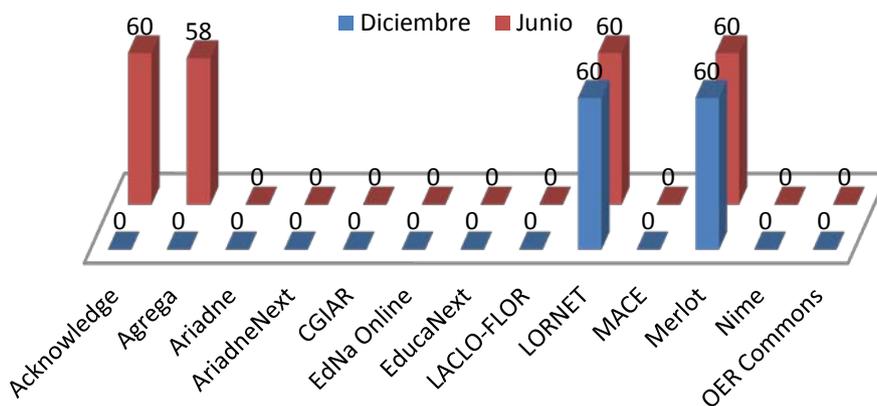


Figura B-3. Comparativa de recuperación de OA sin errores en las dos tomas

Realizada una segunda prueba de recuperación de datos para los mismos términos y los mismos repositorios en Diciembre de 2010. Se encuentra que en el momento de la recuperación, sólo dos (Lornet y Merlot) de los repositorios devuelven resultados en esta ocasión, manteniendo el comportamiento de funcionamiento total o de no funcionamiento absoluto registrado en la primera ocasión de toma de resultados, ver Figura B-.

B.2.2 *Tiempos vs. Número de resultados obtenidos*

Una vez realizada la primera toma de datos (junio de 2010), el trabajo de campo sigue exclusivamente en aquellos repositorios que funcionan correctamente, en ausencia de los que producen error para la muestra realizada. El siguiente paso, será analizar el tiempo necesario para realizar una serie de consultas en cada uno de los repositorios, para ello es necesario separar entre repositorios *stateful* y *stateless*.

Por un lado, en cuanto a repositorios *stateful* el tiempo de consulta total es la suma de los tiempos de obtención de sesión y la propia consulta. Los datos de los dos repositorios de tipo *stateful* son los siguientes:

- **Agrega.** La media del tiempo de la consulta acotada al 10% es de 25,504 segundos. Mientras que la media acotada al 10% del establecimiento de sesión es de 0,893 segundos y la media de la propia consulta es de 19,187 segundos. En la Figura B-4 se pueden apreciar los tiempos de duración de las 60 consultas realizadas a este ROA.
- **LORNET.** La media al 10% es de 12,797 segundos. El tiempo medio es de 9,522 segundos por consulta y de 2,967 segundos para el establecimiento de sesión.

Como se puede apreciar el tiempo del repositorio Agrega es ostensiblemente mayor que el de LORNET, aunque llama la atención que en el repositorio LORNET el tiempo de obtención del identificador de sesión sea más elevado que el de la consulta en sí, lo que puede inducir a pensar que el servicio de búsqueda de este repositorio es muy eficiente. En los siguientes gráficos (Figura B- y Figura B-) se puede apreciar los tiempos (milisegundos) de cada una de las consultas, el primero corresponde al repositorio Agrega y el siguiente al repositorio LORNET.

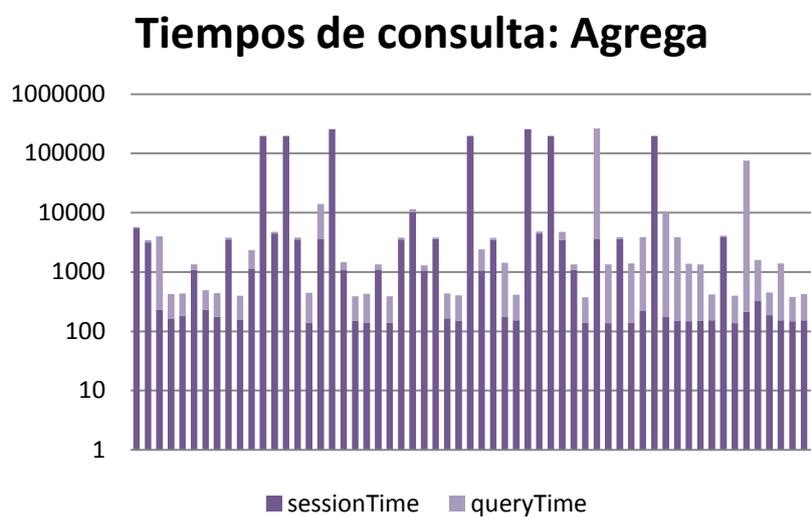


Figura B-4. Tiempos (ms.) de consulta para el repositorio Agrega

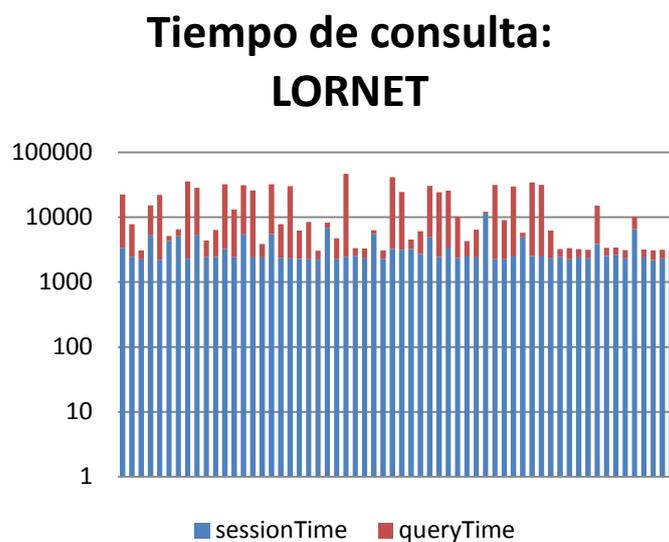


Figura B-5. Tiempos (ms.) de consulta par el repositorio LORNET

Tiempos de consulta: Acknowledge y Merlot

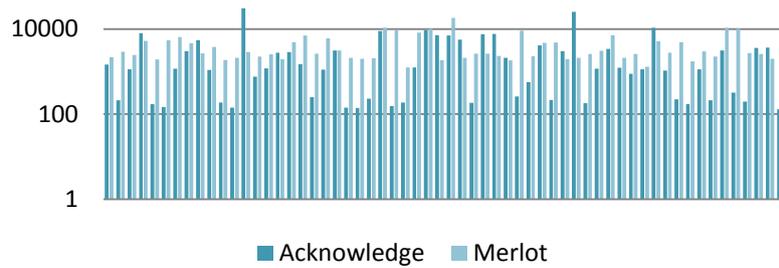


Figura B-6. Tiempos de consulta (ms.) de los repositorios Acknowledge y Merlot

Por otro lado, los repositorios *stateless* el tiempo de consulta total es igual al tiempo de la propia consulta ya que no es necesario solicitar un identificador de sesión. También existen dos repositorios de este tipo (Acknowledge y MERLOT), los resultados en milisegundos se aprecian en la Figura B-:

- **Acknowledge.** El tiempo medio de consulta al 10% es 2,282 segundos.
- **Merlot.** El tiempo medio al 10% es de 3,824 segundos.

En la Figura B- se visualiza la comparativa de los tiempos empleados en cada una de las 60 consultas en los dos repositorios. Se observa que los repositorios *stateless* tienen un tiempo menor pero en contrapartida ofrecen menos funcionalidades por no tener en consideración el histórico de los estados anteriores, respecto a los repositorios *stateful*.

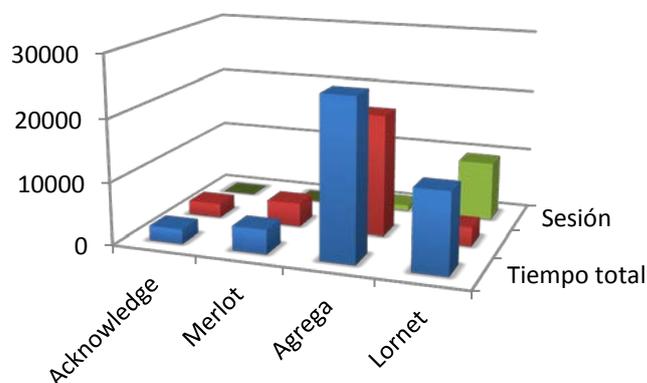


Figura B-7. Análisis de rendimiento ROA, Comparación de tiempos stateless y statefull

El último paso del presente estudio será comprobar el número de resultados que devuelve cada repositorio. Cabe destacar que pese a que el repositorio Agrega no devuelve ningún error, tampoco ningún resultado en la mayoría de las ocasiones. Se pudiera pensar que es debido a que todas las consultas han sido en lengua inglesa y el repositorio es español, poniendo de relieve el problema de la internacionalización en la recuperación de resultados multilingües y la falta de monitorización correcta en los protocolos de consulta con los repositorios. En la Figura B- se puede apreciar la media de resultados de cada uno de los tres repositorios que sí devuelven resultados en la primera ocasión, sin embargo en la recuperación de metadatos realizada en Diciembre Acknowledge no devuelve ningún valor.

La recuperación de metadatos en los repositorios de objetos de aprendizaje, entraña una dinámica variable en el tiempo con la que es necesario contar. Dicha dinámica depende de la cantidad de recursos con los que cuente el repositorio y lo que varíen en el tiempo, de si cambia las características de representación de los recursos, sus procedimientos de recuperación interna, etc.

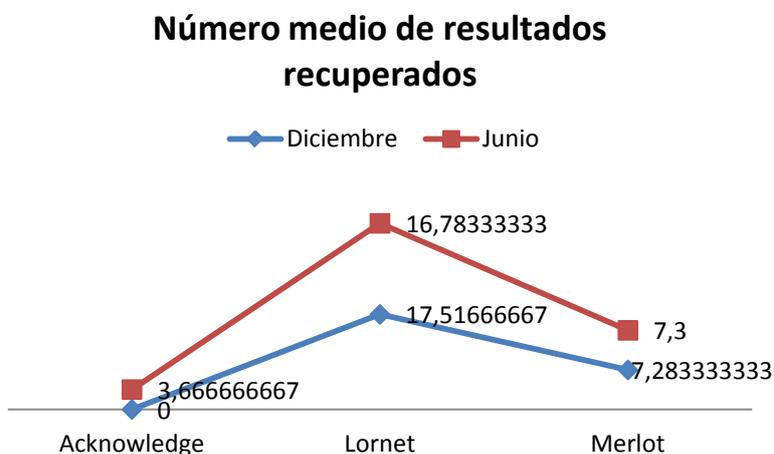


Figura B-8. Número medio de resultados

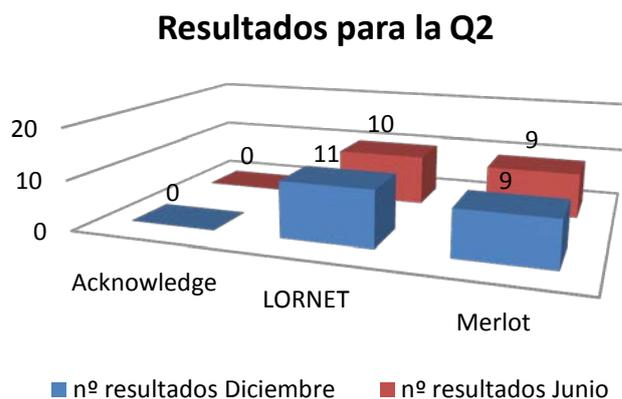


Figura B-9. Resultados recuperados para una misma consulta en dos tomas

La Figura B-, muestra el número de resultados recuperados para la misma consulta en tres repositorios en dos instantes de tiempo. Los resultados obtenidos reflejan que el grado de actualización de los repositorios es muy bajo ya que no son demasiado distintos. Sin embargo parece lógico pensar que una mejora sustancial de este tipo de sistemas con el tiempo aumente su uso y por tanto su grado de actualización.

Consultas a Lornet

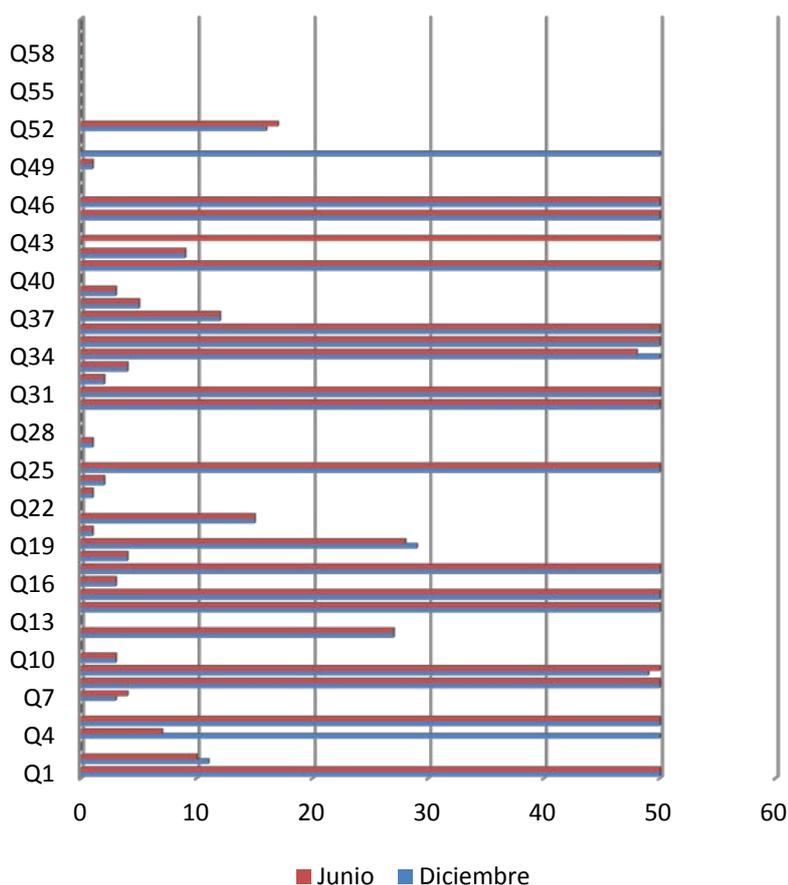


Figura B-10. Resultados recuperados para las 60 consultas en Lornet en las dos tomas

La Figura B- muestra la variación para un mismo repositorio de los resultados de las 60 consultas distintas a lo largo de los dos instantes en el tiempo.

La dinámica de lo ROA en cuanto a la recuperación de sus recursos, exige que el usuario tenga en el menor tiempo el mayor número de OA, por lo que fue necesario contar con una medida que evaluase dicha característica. Se propuso una nueva medida, denominada ganancia temporal de un repositorio J , $G_j(t)$, que

relaciona el número de resultados para n consultas que devuelve un repositorio J respecto al tiempo en que éstos son devueltos. Otra medida es la Ganancia Temporal media, $\overline{G}(t)$, de en m repositorios estudiados.

Con la introducción de estas medidas, ganancia temporal de un ROA y ganancia media del conjunto de ROAs contemplados para una consulta, podemos considerar un nuevo mecanismo de mejora en la recuperación inteligente de contenidos educativos. El objetivo de la recuperación está en obtener el mayor número posible de OA relevantes para el usuario que realiza la búsqueda. Este objetivo pasa por alcanzar la mayor ganancia media en cada momento. La ganancia media subirá su valor si se eliminan los accesos a aquellos repositorios que de manera temporal o repetidamente tienen ganancias temporales menores o nulas. Otra cuestión es la calidad de los OA que se recuperan, que deberá de incluirse como un aspecto más en la recuperación y filtrado de los OA relevantes para el usuario.

B.3 ESTUDIO SOBRE EL ETIQUETADO DE LOS OA EN IEEE LOM

Como hemos ido viendo en cada punto de la presente memoria, cada objeto de aprendizaje es una entidad digital que contiene un documento etiquetado con información descriptiva. Dicha información etiquetada de acuerdo a estándares, permite que dicho objeto de aprendizaje sea fácilmente detectado mediante mecanismos de búsqueda automáticos o semiautomáticos. Una buena descripción del recurso educativo facilitará en gran medida su reutilización mientras que una descripción pobre e incompleta dificultará, incluso hasta hacer imposible su uso, y por ende su reutilización. Este problema se agrava en el caso de que se utilicen esquemas como IEEE LOM, en el que no es obligatorio utilizar ninguno de sus campos, por lo que se deja al propio autor la responsabilidad de describir de forma correcta al recurso didáctico. Motivo por el cual es importante valorar las características reales del etiquetado de los OA, para lo cual se realizará un estudio sobre los metadatos recuperados según el mecanismo anterior. Dichos metadatos están en el formato IEE-LOM.

El primer análisis se realiza sobre las 9 categorías que define este estándar. Tal y como se aprecia en la Figura B-, la categoría más utilizada es *General* (100%), seguida de cerca por la categoría *Technical* (98,83%) y de la categoría *Life Cycle* (95,33%). Estas categorías son algunas de las que más información proporciona sobre el OA. Destacar, que el porcentaje de objetos que no aporta datos sobre la categoría *Technical* (1,17%) no podrían utilizarse ya que no contendrían la dirección al recurso educativo que describen.

Etiquetado Real en IEEE-LOM

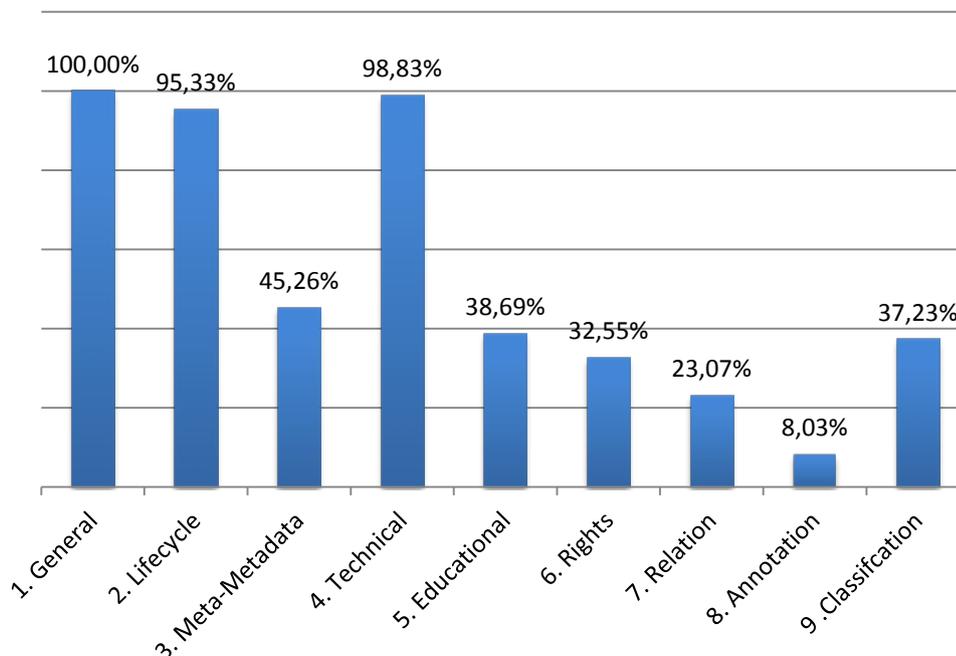


Figura B-11. Análisis de los metadatos en IEEE-LOM de los OA por Etiquetas

Por otro lado, sería necesario que la categoría *Educational* (38,69%) fuera más utilizada ya que es la que describe el contexto educativo sobre la que se debe o puede utilizar el recurso, lo que facilitaría en gran medida su tratamiento automatizado. También sería interesante que la categoría *Rights* (32,55%) tuviera una mayor aceptación, con lo que se conseguiría que no hubiera ningún uso incorrecto de cada OA concreto. Finalmente, destacar las pocas relaciones existentes entre OA, ya que la categoría *Relation* sólo es utilizada por el 23,07% de los OA analizados.

A continuación se estudiará cada categoría por separado, la primera de ellas es la categoría *General*. Se puede apreciar en la Figura B-, que tanto el campo que identifica al objeto y el título del mismo es utilizado en todos los OA analizados, le sigue muy de cerca el campo *Description* con un 94,60% de OA que lo utilizan. Resulta extraño comprobar que la identificación del idioma del OA sólo se utiliza por el 47,15%, este campo se considera fundamental, ya que una de los primeros filtros que se debe utilizar es el idioma del OA. Los campos que menos

se utilizan en esta categoría son *Agregation* (20,73%), *Structure* (5,96%) y *Coverage* (4,23%).

Categoría 1. General

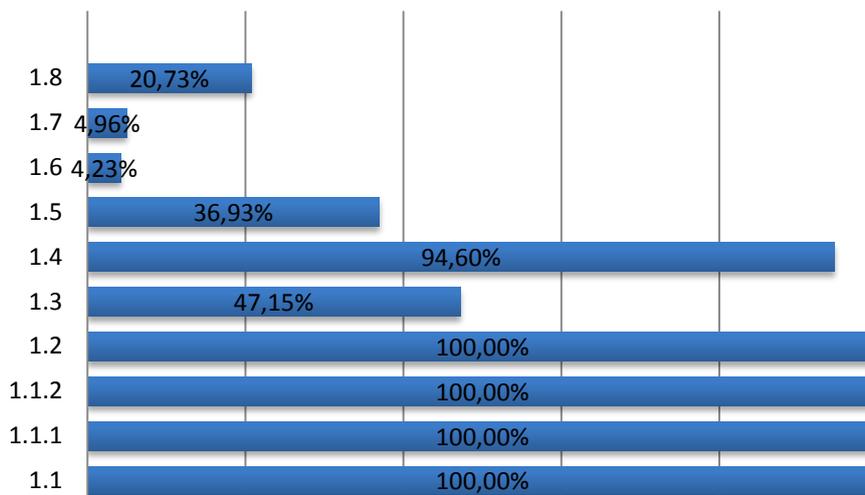


Figura B-12. Análisis campos OA, Categoría 1. General

De entre las etiquetas de la categoría *General* debería resultar de especial importancia el uso de las palabras clave que describen al OA (*keywords*), los resultados se aprecian en la Figura B-. Sólo existe un 47,15% de OA analizados, que utilizan este campo para describir el contenido del recurso educativo, lo cual es claramente una debilidad, ya que resulta un campo fundamental para ayudar a computar de forma automática si el OA es válido para un proceso de reutilización concreto. Del porcentaje de OA que utilizan el campo de palabras clave, la mayoría de ellos sólo con incluye una única palabra clave (57,31%) y con 2 y 3 palabras clave, 18,18% y 7,51%, respectivamente. La media de palabras clave es de 7,95 y existen 5 objetos de aprendizaje que tienen más de 100 palabras clave, lo cual se considera un error ya que no ayuda a analizar el contenido real del OA.

Uso de keywords

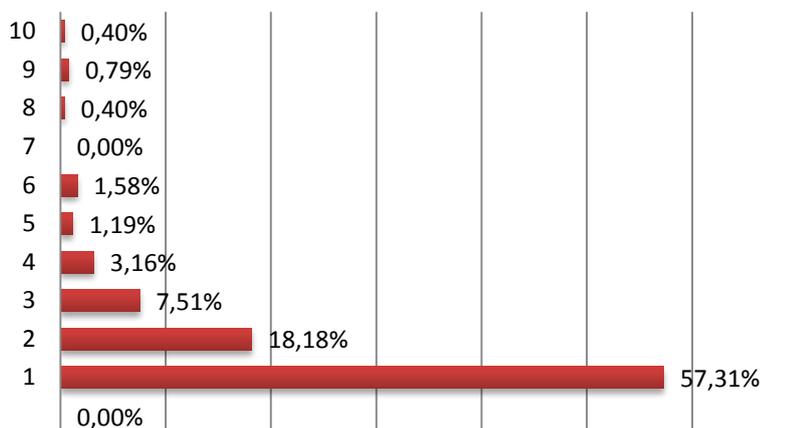


Figura B-13. Análisis campos OA, Campo 1.6 Keywords

En esta categoría, resulta interesante describir el identificador del OA (*Identifier*), que a su vez se compone de los elementos *Entry* y *Catalog*. Como ya hemos visto, todos los objetos analizados disponen de identificador, pero un porcentaje de 64,38% dispone de un único identificador, mientras que el 35,62% dispone de 2 identificadores, lo que dificulta la identificación unívoca del recurso. También se da la situación de que para un identificador de un OA no se completa alguno de los dos campos necesarios (*Entry* o *Catalog*) según el esquema IEEE LOM.

Analizando de forma conjunta las categorías *Life Cycle* y *Meta-Metadata*. Tal y como se aprecia en la Figura B-, la mayoría de campos de estas dos categorías no son utilizados, e incluso existe un buen número de campos (2.3.2, 2.3.3, 3.2.2, 3.3 y 3.4) que no son utilizados por ninguno de los 685 OA analizados. En cuanto a los campos más utilizados, cabe destacar que la mayoría de OA (94,31%) utiliza el campo 2.3.1 correspondiente al autor del recurso de aprendizaje.

Categorías 2. Life Cycle y 3. Meta-Metadata

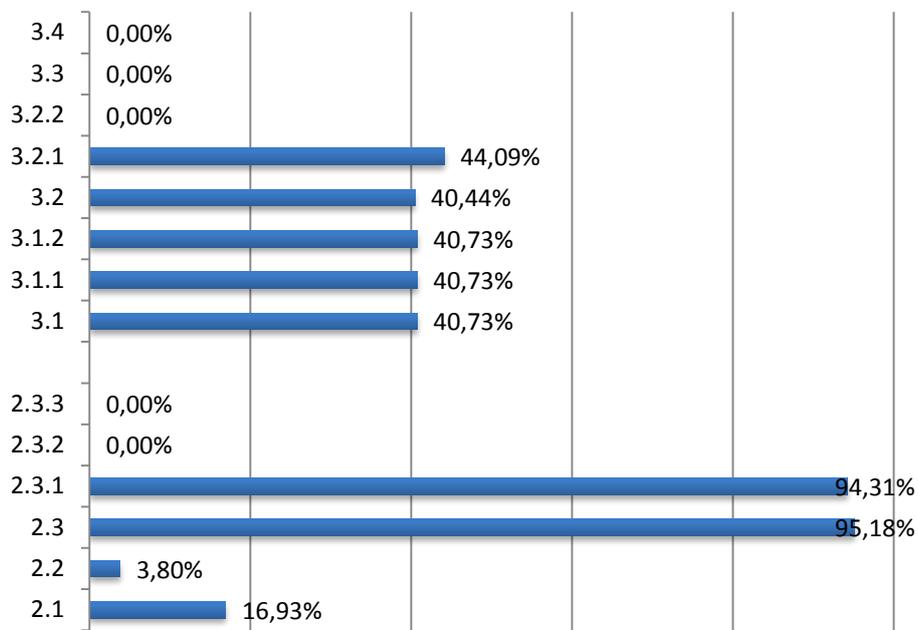


Figura B-14. Análisis campos OA, Categorías 2. Life Cycle y 3. Meta-Metadata

Seguimos con el análisis de la categoría *Technical* (Figura B-), en ella se especifican los detalles técnicos del OA, como tamaño, duración, requisitos, versión, sistema operativo, etc. y uno de los datos más importantes de todo el esquema, el enlace al recurso (*4.3 Location*) que se suele representar mediante una URI o URL. En general, se comprueba que se utilizan pocos campos de esta categoría, además del citado campo *Location* (93,43%), ya que se utiliza en aproximadamente un tercio de los objetos los descriptores *Format* (41,31%) y *Size* (24,82); y en menor medida se utiliza el descriptor 4.4 (17,37%) que da información sobre requerimientos técnicos del recurso de aprendizaje descrito por el OA.

Respecto a la localización del recurso vemos que este campo sólo lo utilizan un 93,43% de OA, lo cual implica que el porcentaje de OA que no lo utiliza (45 objetos) no podrán ser utilizados en un proceso de reutilización ya que no se puede acceder al recurso que describe. De los que sí lo utilizan la mayoría tiene una única localización (99,53%) y el porcentaje restante tiene el recurso duplicado en dos localizaciones diferentes.

Categoría 4. Technical

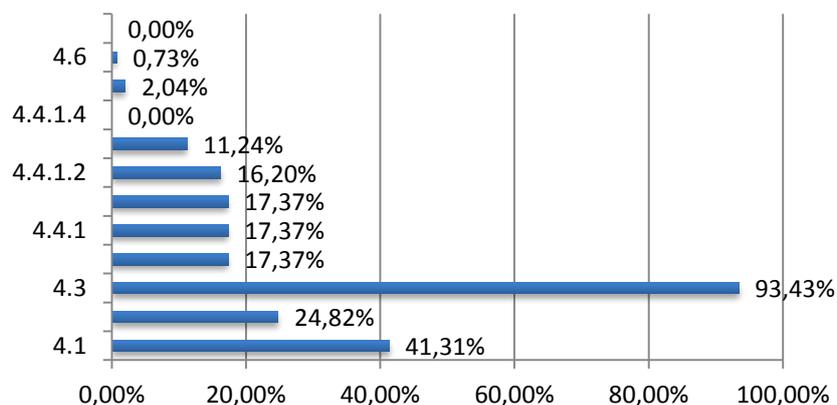


Figura B-15. Análisis campos OA, Campo 4.3 Location

La categoría *Educational*, que describe el entorno educativo dónde debe ser utilizado el recurso descrito, pudiera ser una de las categorías más importantes para automatizar la búsqueda y reutilización de OA. Sin embargo en la práctica, según se puede ver en la Figura B-, no es muy utilizado por los autores. Los campos de esta categoría se utilizan poco, ya que todos los porcentajes están por debajo del 32,55% y un buen número de elementos está por debajo del 10%.

Los campos más utilizados son *Intended End User Role* (24,82%), *Learning Resource* (28,03%) y *Context* (32,55%), que definen el tipo de recurso y el contexto en el que se debe utilizar. Mientras que se echa en falta el uso de elementos importantes como *Difficulty* o *Typical Learning Time*.

Categoría 5. Educational

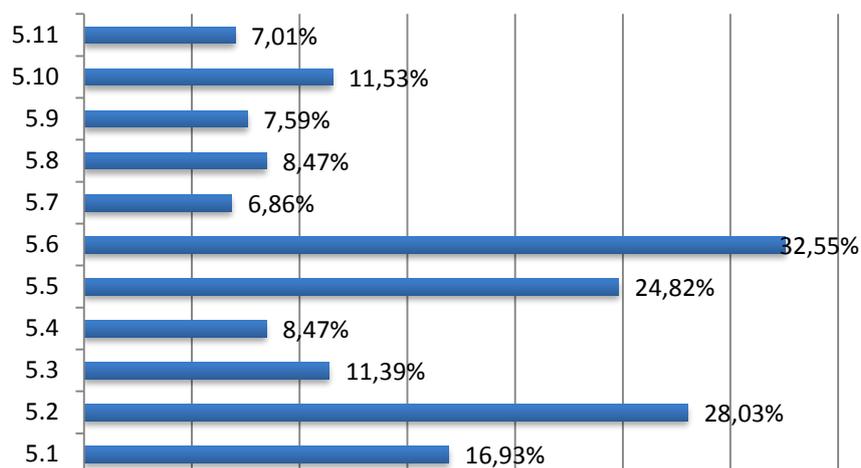


Figura B-16. Análisis campos OA, Categoría 5. Educational

La siguiente categoría analizada es *Rights*, que contiene únicamente 3 elementos relacionados con las licencias de uso del recurso educativo, estos tres elementos sitúan su porcentaje de uso entre el 20% y el 30%, tal y como se aprecia en el gráfico de la Figura B-

Categoría 6. Rights

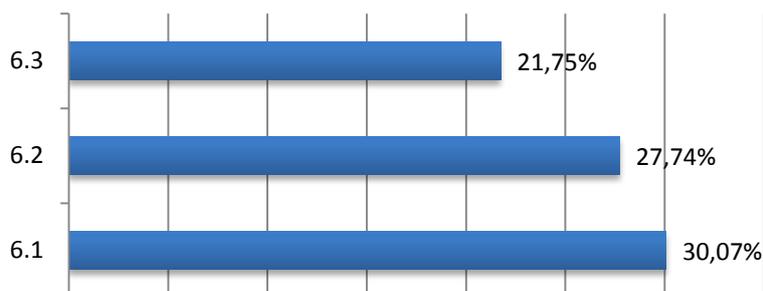


Figura B-17. Análisis campos OA, Categoría 6. Rights

El estudio sigue con el detalle de uso de la categoría *Relation* (Figura B-), la cual define las relaciones con otros OA. Existe un amplio porcentaje que no define ninguna relación con ningún OA (77,23), del porcentaje que sí define relaciones

la mayoría lo hace con un único OA (61,54%) o con dos (31,41%), mientras que el máximo número de relaciones que define alguno de los 685 OA analizados es de 7. También cabe destacar que existen algunos OA que completan el campo *Description* del OA relacionado pero no proporcionan los datos de identificador y catalogo para identificar correctamente al recurso referenciado, por lo que la relación no podría utilizarse aunque esté perfectamente definida.

Categoría 7. Relation

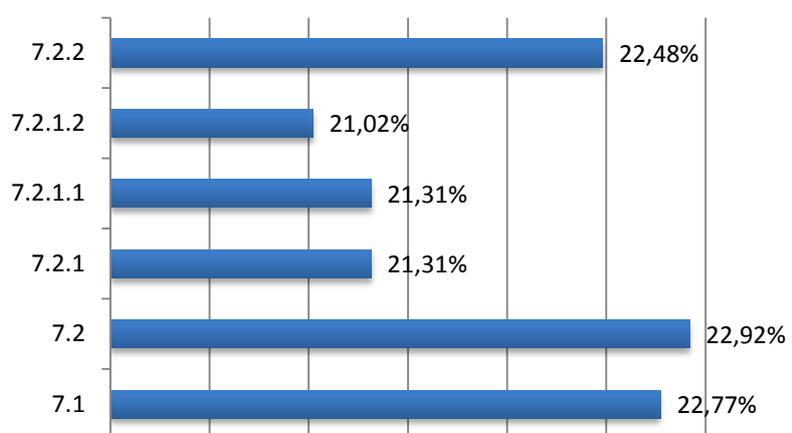


Figura B-18. Análisis campos OA, Categoría 7. Relation

La penúltima categoría a analizar es la categoría *Annotation* que se utiliza realmente poco, los tres campos que tiene están por debajo del 10%, los más utilizados son el de entidad que realiza la anotación (7,3%) y la fecha en que se realiza (6,86%), el campo *Description* a penas se utiliza, tal y como se observa en la Figura B- .

Categoría 8. Annotation

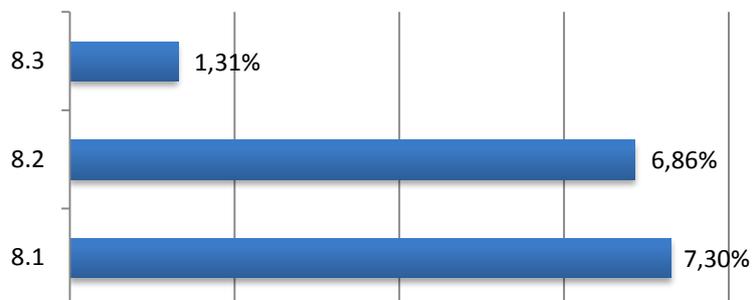


Figura B-19. Análisis campos OA, Categoría 8. Annotation

Para terminar se detallará el uso de la categoría *Classification*, la cual sitúa al OA dentro de una categoría arbitraria definida por el propio autor del recurso. Lo primero que destaca al comprobar la gráfica de resultados (Figura B-) es que los campos *Description* y *Keywords* no son utilizados por ninguno de los 685 OA analizados; y que el resto de categorías se utiliza por los autores, con porcentajes situados entorno al 35% de OA.

Categoría 9. Classification

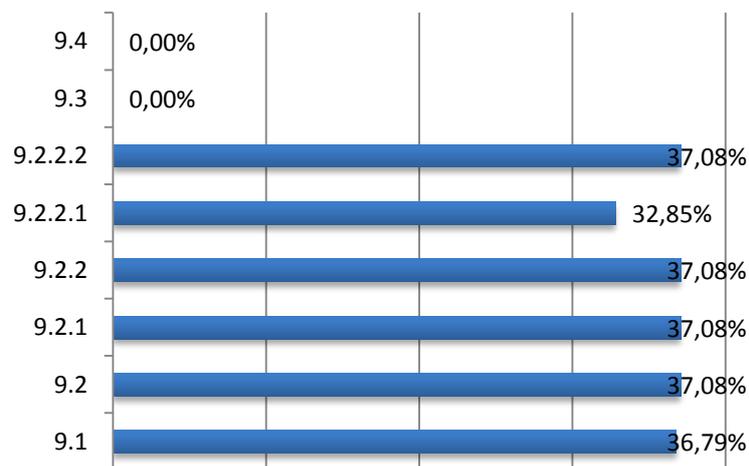


Figura B-20. Análisis campos OA, Categoría 9. Classification

Antes de finalizar el estudio, resulta importante detallar el grado de uso de los elementos del esquema Dublin Core incluidos dentro de los objetos descritos mediante IEEE LOM analizados, ya que como se comentó en el apartado A.1.2 había una equivalencia entre ambos esquemas, detallada en la Tabla A-3.

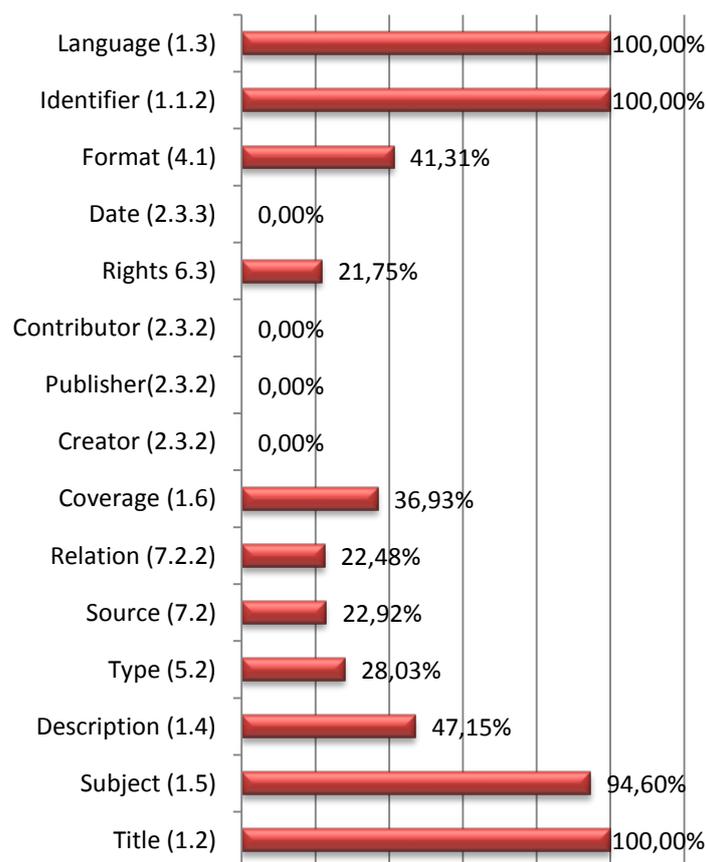


Figura B-21. Correspondencia IEEE LOM con Dublin Core

Tal y como se aprecia en Figura B-, existe un buen conjunto de elementos que se utiliza prácticamente en todos los OA, éstos son *Title*, *Subject*, *Identifier* y *Lenguaje*; pero también existe un buen número de elementos que apenas se utiliza como *Date*, *Publisher*, *Creator* y *Contributor*. Respecto a estos tres últimos elementos decir, que aunque la correspondencia entre Dublin Core y IEEE LOM es oficialmente mediante el elemento 2.3.2, en la práctica se utiliza el elemento 2.3.1 (94,31%) codificado mediante tarjetas VCARD, por lo que se

cumplirían los requisitos de estos tres elementos y tan sólo quedaría como elemento no utilizado el elemento *Date*.

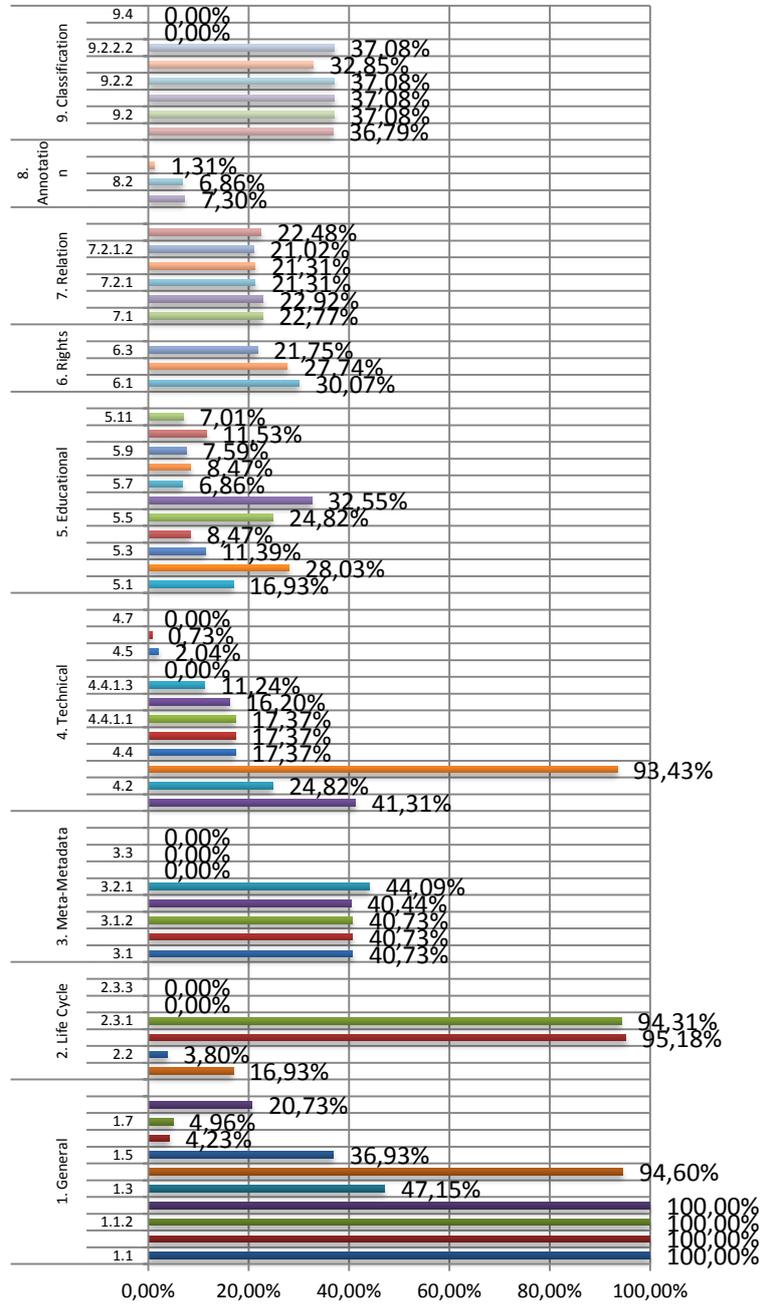


Figura B-22. Análisis campos OA, Porcentaje de uso de elementos

Respecto al resto de elementos, destacan el elemento *Description*, *Format* y *Coverage* con porcentajes alrededor del 40% y en menor medida también se utiliza *Rights*, *Source*, *Type* y *Relations* con un porcentaje en torno al 25%. Como resumen final del estudio la Figura B- da una visión global sobre el uso del estándar IEEE LOM en el etiquetado real de OA mediante el estadístico presentado.

Para concluir, en el estudio en la Figura B- se presenta el estándar IEEE LOM, eliminando aquellos elementos menos utilizados, es decir, aquellos elementos con un porcentaje de uso menor al 20%. Si nos atenemos a los resultados, el número de categorías se reduciría a 8, en lugar de las 9 iniciales y el número de elementos se reduciría de 64 a 35.

Es importante tener en cuenta esta reducción, ya que este subconjunto de elementos corresponde a los más utilizados tanto por los autores a la hora de describir el recurso (Título, descripción, palabras clave, etc.), como por los responsables del ROA a la hora de definir aspectos técnicos (Identificador, Localización, etc.).

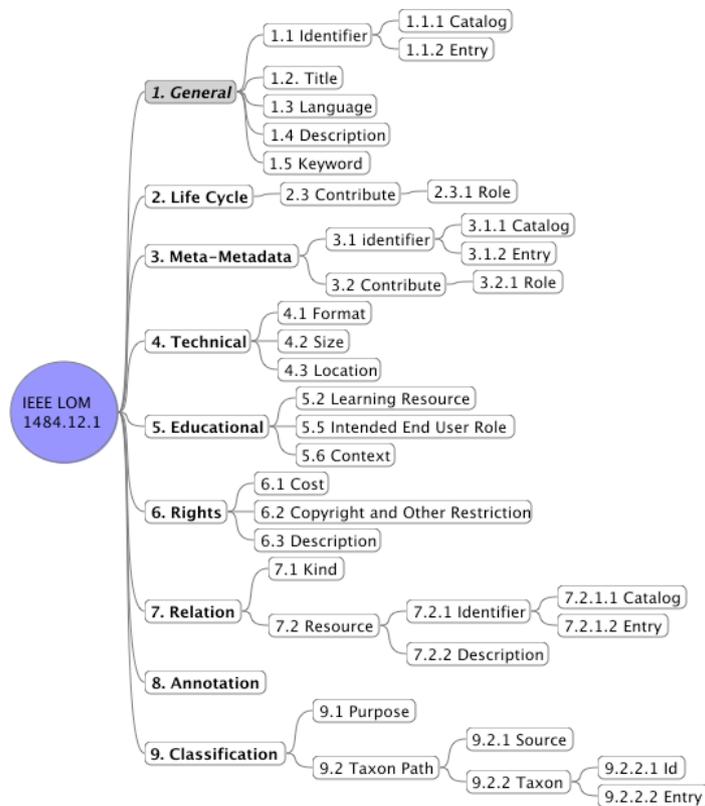


Figura B-23. Elementos de IEEE LOM más utilizados por los autores

Anexo C HERRAMIENTAS UTILIZADAS

C.1 EMFGORMAS

Como guía metodológica en la realización del diseño de la organización se ha utilizado GORMAS. Para la realización del proceso se ha utilizado la herramienta EMFGormas.

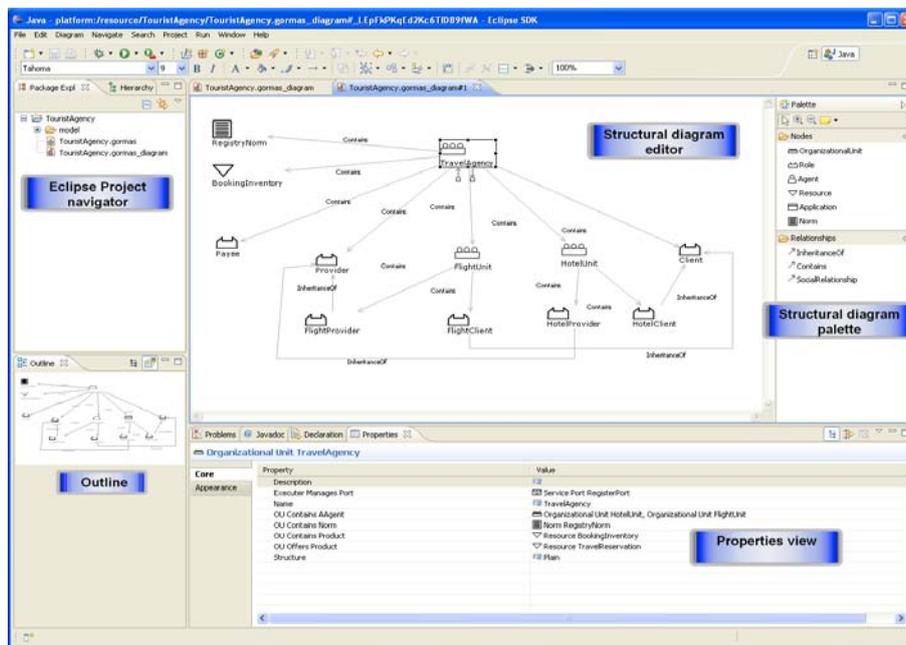


Figura C-1. Interfaz EMFGormas e Eclipse

EMFGormas⁶³ es un nuevo enfoque para el modelado de SMA abiertos orientados a servicios que utiliza la tecnología de Eclipse MDA (*Model Driven Architecture*)⁶⁴ [Soley, 2005]. Ofrece una herramienta CASE basada en un meta-

⁶³EMFGORMAS <http://users.dsic.upv.es/grupos/ia/sma/tools/EMFGormas/index.html>

⁶⁴ MDA <http://www.eclipse.org/modeling/>

modelo unificado para la ingeniería de sistemas abiertos en los que las entidades constituyentes interactúan entre ellas por medio de servicios.

La iniciativa MDA propone un enfoque de desarrollo de software basado en el modelado y en el mapeo automático de los modelos de origen a modelos destino. Los modelos que representan un sistema y su entorno puede ser visto como los modelos origen y el código fuente puede ser visto como un modelo de destino. Siguiendo los estándares de MDA, la plataforma Eclipse es una iniciativa de código abierto que ofrece un entorno de trabajo para crear este tipo de herramientas. La propia plataforma Eclipse está organizada como un conjunto de subsistemas (implementados en uno o más plugins). La siguiente figura muestra algunos de los plugins como EMF, GEF, GMF, Mofscript, Xpand2 que son ofrecidos por la plataforma para permitir la creación de herramientas basadas en la especificación de metamodelos.

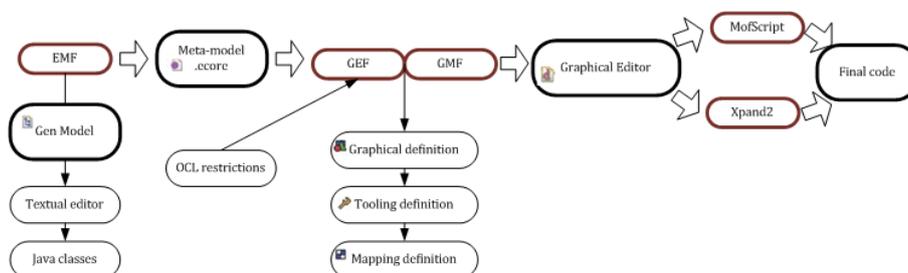


Figura C-2. Estructura de plugins de Eclipse

C.2 NETBEANS

Para la construcción de los servicios web implementados se hace uso de NetBeans⁶⁵ en la versión 7.0. Se trata de un entorno de desarrollo integrado libre, hecho principalmente para el lenguaje de programación Java. Existe además un número importante de módulos para extenderlo. NetBeans IDE es un producto libre y gratuito sin restricciones de uso. NetBeans maneja el uso de servicios Web. Permite probar, construir aplicaciones clientes que accedan a servicios web, y generar código para invocar a los servicios web existentes externos o crear otros nuevos (tanto RESTful como basado en SOAP).

Los servicios web son sistemas software diseñados para hacer posible la interoperabilidad entre equipos informáticos sobre una red. La interfaz de comunicación con el servicio es definida haciendo uso de un formato que una

⁶⁵ NETBEANS <http://netbeans.org/>

máquina puede procesar (por ejemplo el estándar para descripción de servicios web basados en SOA es WSDL, *Web Services Description Language*).

Los sistemas que deseen hacer uso de un servicio web deben, en primer lugar, localizarlo en un sistema de directorio que haga uso del estándar UDDI (*Universal Description, Discovery and Integration*). El directorio permite buscar los servicios web disponibles y obtener la descripción de los mismos en formato WSDL. Una vez se tiene la descripción del servicio, el sistema puede empezar a hacer uso de él accediendo al servidor que lo proporciona siguiendo la descripción proporcionada por el servicio UDDI. La comunicación se realizará a través del protocolo de comunicación SOAP (*Simple Object Access Protocol*).

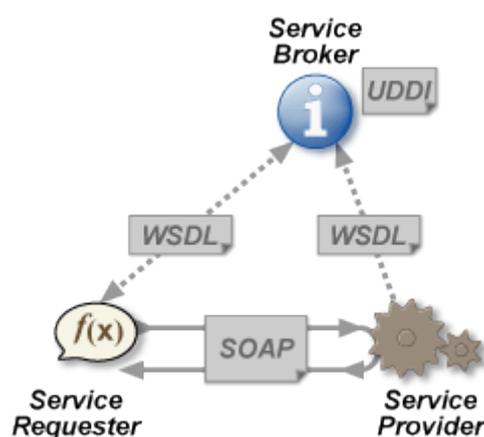


Figura C-3. Esquema de comunicación con servicios web

Debido a que en este trabajo se conoce el servicio web que se va a utilizar y la dirección del mismo en internet, no se hace uso de servicio de directorio. Por otro lado el servicio de directorio en estos servicios web no se emplea a día de hoy de manera general. Por ello que se describen a continuación únicamente los estándares SOAP y WSDL por un lado y RESTful como alternativa a SOAP.

SOAP es un protocolo para el intercambio de información estructurada orientado a servicios web en redes computacionales. La construcción del mensaje se realiza haciendo uso de *Extensible Markup Language (XML)*⁶⁶ y el intercambio del mismo se efectúa a través de protocolos como *Remote*

⁶⁶ Conjunto de reglas para el codificado de documentos en formatos que son procesables por máquinas.

*Procedure Call (RPC)*⁶⁷ o *Hypertext Transfer Protocol (HTTP)*⁶⁸. El protocolo es un estándar del *World Wide Web Consortium (W3C)* y se encuentra actualmente en su versión 1.2.

WSDL es un lenguaje basado en XML que proporciona un modelo para la descripción de servicios web. Para ello, describe la forma de comunicación (requisitos del protocolo y formatos de los mensajes) que posibilita la interacción con los servicios listados en su catálogo. Las operaciones y mensajes que soporta se describen en abstracto y se ligan después al protocolo concreto de red y al formato del mensaje.

La estructura de un fichero WSDL consta de las siguientes partes:

- Tipos de datos (etiqueta `<types>`): se indican los tipos de datos utilizados en los mensajes.
- Mensajes (etiqueta `<message>`): se definen los elementos de los mensajes, haciendo uso de los tipos antes definidos.
- Tipos de puerto (etiqueta `<portType>`): se definen las operaciones permitidas y los mensajes intercambiados.
- **Bindings** (etiqueta `<binding>`): se listan los protocolos de comunicación utilizados.
- Servicios (etiqueta `<services>`): conjunto de puertos y direcciones de los servicios. Se utiliza para ello la información definida con las etiquetas anteriores.

Otro modo de trabajar con servicios web es a través del uso de la arquitectura RESTful. Los servicios web RESTful son servicios construidos usando el estilo de arquitectura RESTful. *REpresentational State Transfer (REST)* es un estilo de arquitectura para sistemas hipermedia distribuidos, tales como la World Wide Web. El centro de la arquitectura RESTful es el concepto de los recursos identificados por los identificadores de recurso universal (*Universal Resource Identifiers URIs*). Estos recursos pueden ser manipulados usando una interfaz estándar, tales como el HTTP, y la información es intercambiada usando representaciones de estos recursos. Construir servicios web usando RESTful es una alternativa a SOAP para desplegar servicios en internet, debido a que son más ligeros y tienen la capacidad de transmitir datos directamente desde HTTP.

⁶⁷ Protocolo de comunicación entre procesos que permite a un programa informático ejecutar una subrutina o procedimiento que se encuentra en otro equipo a través de la red.

⁶⁸ Protocolo de red para la distribución de sistemas de información hipermedia (texto, imagen, vídeo, audio...). HTTP es usado en cada transacción en la World Wide Web.

C.3 THOMAS

En la actualidad, las arquitecturas y modelos computacionales para SMA abiertos a pequeña o gran escala son considerados una de las claves para el éxito de la tecnología de agentes en escenarios reales.

THOMAS (*MeTHods, Techniques and Tools for Open Multi-Agent Systems*) [GTI-IA, 2009] es el nombre dado a una arquitectura abstracta para SMA abiertos a gran escala basada en un enfoque orientado a servicios y direccionada principalmente al diseño de organizaciones virtuales. Podríamos definir varias áreas de investigación en las que se fundamenta la construcción de la arquitectura THOMAS, entre ellas destacan la computación orientada a servicios, la estructura organizacional de las comunidades y las plataformas de agentes.

C.3.1 Arquitectura

La arquitectura está formada básicamente por un conjunto de servicios estructurados modularmente. THOMAS toma como base la arquitectura FIPA⁶⁹, expandiendo sus capacidades respecto al diseño de organizaciones e impulsando las capacidades de los servicios. En THOMAS, existe un módulo dedicado al manejo de las organizaciones que hayan sido introducidas en la arquitectura. Dicha introducción se realiza mediante una redefinición del FIPA *Directory Facilitator* de forma que éste sea capaz de hacer frente a los servicios de un modo más elaborado, siguiendo las directrices Arquitecturas Orientadas a Servicio.

Los agentes acceden a la infraestructura ofertada por THOMAS por medio de una serie de servicios incluidos en lo que se denomina OMS (*Organization Manager Service*). Los componentes principales de THOMAS podemos verlos en la figura siguiente.

⁶⁹ FIPA <http://www.fipa.org/>

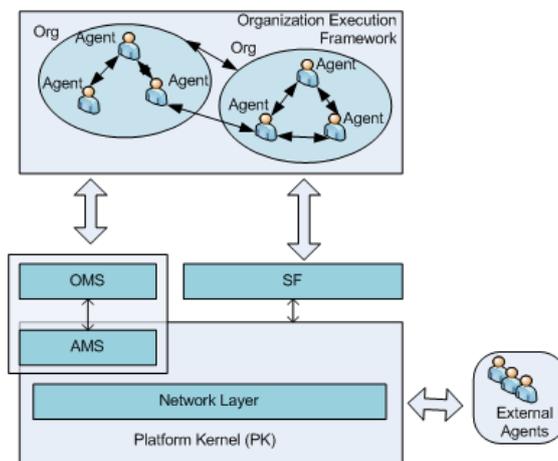


Figura C-4. Arquitectura THOMAS [GTIIA, 2009]

Como se observa, existen tres componentes principales dentro de la arquitectura de THOMAS:

- **Service Facilitator (SF).** Ofrece tanto servicios simples como complejos a los agentes y OV activas. Básicamente su funcionalidad se resume en ofrecer un directorio de páginas amarillas y verdes sobre los servicios disponibles.
- **Organization Manager Service (OMS).** Principalmente se encarga de la gestión de organizaciones y de las entidades incluidas en ellas. Permite, por lo tanto, la creación y gestión del ciclo de vida de una organización.
- **Platform Kernel (PK).** Permite mantener los servicios básicos de gestión de una plataforma de agentes.

Además de estos componentes, existe otro componente denominado **Execution Framework (EX)** que permite a cualquier agente crear su propia OV con una estructura y unas normas, junto con la oferta y la demanda de servicios necesarios.

C.3.1.1 Service Facilitator (SF)

Este componente ofrece el soporte necesario para que los agentes y las organizaciones ofrezcan y descubran servicios. Principalmente, el SF provee de un lugar donde las entidades autónomas pueden registrar la descripción de servicios como entradas en un directorio. Es decir, su función es actuar como un gestor de páginas amarillas, para que de esta manera se puedan hacer búsquedas que determinen qué entidades ofrecen un servicio concreto. Gracias a esto, se podrá localizar servicios de acuerdo a los perfiles u objetivos que se

desean satisfacer mediante la invocación del SF, incorporando además mecanismos de composición de servicios.

El SF será el que controle el acceso a la plataforma THOMAS, utilizando técnicas de seguridad y gestión de permisos.

Un servicio es caracterizado normalmente como la interacción entre dos entidades, y es modelado como comunicaciones entre procesos independientes. En general, un servicio ofrece unas capacidades (cada una de ellas encaminadas a cumplir un determinado objetivo), necesita que se verifiquen ciertas condiciones antes de su ejecución, consiste en el intercambio de uno o más mensajes y tras su ejecución (con éxito) se tienen una serie de efectos en el entorno. Además, existen parámetros adicionales, independientes de su funcionalidad, como la calidad del servicio, plazos o protocolos de seguridad, etc. En último lugar destacar que los resultados obtenidos por los servicios pueden ser mejorados usando mecanismos de composición automática de servicios [Brogi *et al.*, 2003] (como por ejemplo, *matchmaking* parcial). Para esto, el SF tiene una descripción de los procesos internos que tienen lugar cuando se ejecuta el servicio. En THOMAS, al disponer de los protocolos de FIPA existe un mecanismo bien establecido para estandarizar las interacciones, así cada servicio tiene asociado un protocolo.

Hay que considerar que estamos hablando de servicios semánticos, por lo que será importante que la ontología utilizada en el servicio permita, al leer la descripción de un servicio, que cualquier entidad obtenga toda la información necesaria para saber cómo interactuar con él o cómo construir una aplicación que sea capaz de utilizarlo.

Un servicio es definido como una tupla del tipo $(SID, goal, prof, proc, ground, ont)$, donde:

- **SID.** Es el identificador único del servicio.
- **goal.** Es el objetivo, el propósito que el servicio pretende conseguir y proporciona el primer nivel de abstracción para la composición del servicio.
- **prof.** Es el perfil del servicio y lo describe en términos de su IOPEs (*Inputs, Outputs, Preconditions y Effects*) y sus atributos no funcionales. Este tipo de representación incluye una descripción de lo que cumple el servicio, las limitaciones a su aplicabilidad y de su calidad, así como los requisitos que deben cumplir los clientes para poder utilizarlo.
- **proc.** Describe cómo debe usar el servicio un cliente. Se especifica el contenido semántico para el uso del servicio, las situaciones en las que se obtiene, y, cuando sea necesario, también se especifica el proceso paso a paso para obtener los resultados. En otras palabras, se

especifica cómo llamar a un servicio y lo que ocurre cuando el servicio se ejecuta.

- **ground.** Especifica en detalle cómo puede un agente acceder al servicio. El *grounding* especifica un protocolo de comunicación, los formatos de mensaje, el puerto de contacto y otros detalles del servicio. Se especifica mediante el estándar OWL-S ampliado con protocolos FIPA.
- **pnt.** Es la ontología que provee significado a todos los elementos del servicio. El lenguaje elegido es OWL-DL.

C.3.1.2 Organization Management System (OMS)

El componente OMS es el principal responsable del manejo de las organizaciones y sus entidades. Es el responsable de los ciclos de vida de las organizaciones, incluyendo la especificación y administración de sus componentes estructurales (roles, unidades y normas) y sus componentes de ejecución (agentes participantes y roles que juegan; unidades activas en cada momento).

- Un *rol* representa una posición dentro de la unidad donde se define. Lleva asociado unas reglas de interacción, impuestas por la estructura de su unidad y la posición concreta que ocupa en ella; así como unas reglas de comportamiento que especifican su funcionalidad (tipos de servicios que ofrece y requiere), limitan sus acciones (prohibiciones, obligaciones y permisos) y determinan las consecuencias de las mismas (sanciones y recompensas).
- Una *norma* indica las obligaciones, permisos y prohibiciones de los roles respecto al registro, solicitud y desempeño de servicios, su composición, o bien la calidad de sus resultados. Permite definir aquellas restricciones que no se pueden expresar en las precondiciones (o postcondiciones) de un servicio determinado.
- Una *unidad* representa a agrupaciones de agentes y admite la recursión (unidades dentro de otras). Establece así la estructura topológica del sistema. Por ejemplo, facilita la representación de estructuras de tipo jerárquico, matricial, coalición, etc.

Para la gestión de estos componentes, el OMS maneja las siguientes listas:

- **UnitList.** Guarda la relación de las unidades existentes, de sus unidades directamente superiores (SuperUnit), objetivos y tipo.
- **RoleList.** Guarda la relación de los roles existentes en cada unidad, de qué roles heredan y cuáles son sus atributos (accesibilidad, posición).
- **NormList.** Guarda la relación de las normas del sistema.
- **EntityPlayList.** Guarda la relación de las unidades en las que se ha registrado como miembro cada agente, así como el rol que juega en dicha unidad.

El OMS ofrece los servicios necesarios para el adecuado funcionamiento de una organización. Dichos servicios se clasifican en: (i) servicios estructurales, que modifican la especificación estructural y normativa propia de la organización; y (ii) servicios dinámicos, que permiten la gestión de la entrada y salida dinámica de agentes dentro de la organización, así como la adopción de roles.

Las organizaciones son estructuradas por medio de unidades, las cuales representan grupos de entidades (agentes u otras unidades), y que se relacionan porque tienen propósitos comunes. Las unidades tienen una topología interna (por ejemplo, jerárquica, de grupo o plana) que impone restricciones a las relaciones entre los agentes y al control (por ejemplo, supervisión o información entre las relaciones).

En THOMAS, una unidad “virtual” es definida para representar el “mundo” del sistema en el cual los agentes participan por defecto. El OMS crea organizaciones dentro de esta unidad “virtual” por medio del registro de unidades, que también pueden estar compuestas a su vez por más unidades.

A parte de esto, los roles se definen en cada unidad. Los roles representan toda la funcionalidad necesaria para lograr el objetivo de la unidad. También pueden tener asociadas normas para controlar las acciones propias de un rol (por ejemplo, qué servicios de agentes jugando ese rol pueden pedir, ofrecer o servir o qué permisos tiene para acceder a recursos, etc.). Como resultado, los agentes pueden dinámicamente adoptar roles dentro de las unidades, por lo que el OMS controla este proceso de adopción de roles y controla qué entidades juegan qué rol en cada momento.

C.3.1.3 Platform Kernel (PK)

El componente PK mantiene básicamente el manejo de los servicios básicos en una plataforma multi-agente. Por tanto, está encargado de gestionar el ciclo de vida de los agentes presentes en las distintas organizaciones y permite además disponer de un canal de comunicaciones (incorporando diferentes mecanismos de transporte de mensajes) que facilite la interacción entre las diversas entidades. Por otro lado, el PK ofrece una conectividad segura y los mecanismos necesarios que permiten la interconectabilidad multidispositivo para aplicaciones que así lo requieran. Los servicios ofrecidos son en la mayoría de los casos heredados de FIPA aunque con ciertos matices.

Los servicios de PK necesarios en la arquitectura THOMAS son clasificados en cuatro clases: (i) *Registration*, son servicios necesarios para añadir, modificar y eliminar agentes nativos de la plataforma; (ii) *Discovery*, son servicios que proporcionan la funcionalidad de obtención de información sobre los agentes nativos en la plataforma; (iii) *Management*, son servicios para controlar el estado de activación de los agentes nativos de la plataforma; (iv)

Communication: son servicios para la comunicación de agentes dentro de la plataforma y fuera de ella.

Desde un punto de vista global, la arquitectura THOMAS, plantea una integración total de forma que los agentes puedan ofrecer e invocar servicios de forma transparente a otros agentes o entidades, así como que entidades externas puedan interactuar con los agentes de la arquitectura mediante el uso de los servicios ofertados.

THOMAS actualmente está en continuo desarrollo y cabe destacar su nueva versión OVAMAH (*Multiagent-based Adaptive Virtual Organizations*). OVAMAH es la evolución de THOMAS, donde se están desarrollando las modificaciones necesarias para que puedan ser implementadas organizaciones virtuales adaptativas. Estas propiedades son definidas desde un detallado estudio del estado del arte de métodos de desarrollo, trabajos y herramientas relacionadas con las organizaciones virtuales y su adaptación en [Rodríguez, 2010].

Una vez se ha realizado la descripción de la arquitectura a continuación se revisarán los componentes técnicos de la misma.

C.3.2 Componentes Técnicos

La aplicación THOMAS⁷⁰ está formada por dos componentes fundamentales:

- ❖ Los Servicios Web, que implementan los servicios de la plataforma. Estos servicios están descritos en el documento de la arquitectura abstracta.
- ❖ Los agentes intermediarios (SF, OMS), que actúan de traductores entre los agentes FIPA de la plataforma y los Servicios Web. Este tipo de agentes reciben peticiones de servicios mediante el protocolo FIPA-request y se encargan de acceder a los servicios web correspondientes. Esta solicitud de los servicios web se realiza empleando el API de mindswap.

La siguiente figura muestra un esquema de la aplicación THOMAS.

⁷⁰ THOMAS <http://www.dsic.upv.es/users/ia/sma/tools/Thomas/index.html>

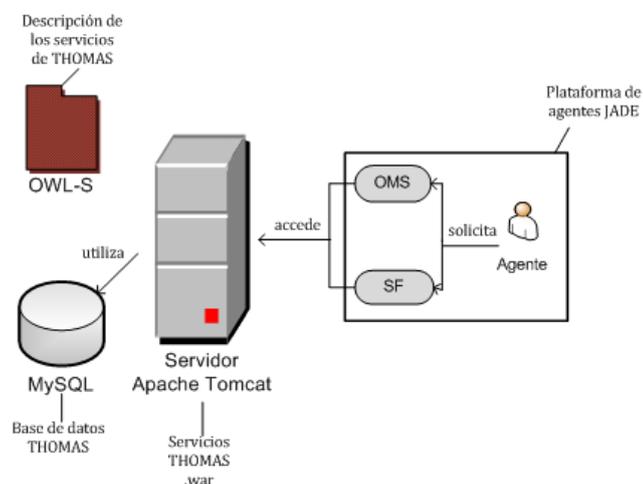


Figura C-5. Aplicación THOMAS

Los **servicios** proporcionados hacen uso de una base de datos MySQL⁷¹ que es necesario configurar previamente. Tanto en el ejemplo ejecutable como en el código suministrado asume que la base de datos se encuentra en la misma máquina en la que se proporcionan los servicios (localhost). En el caso de no ser así es necesario modificar los archivos *.java* de los servicios donde se configura el acceso a la base de datos.

La implementación de las **normas** implica que al registrar una nueva norma dentro de la plataforma se lleven a cabo las acciones necesarias para poder controlarla. De modo que en la implementación del servicio *RegisterNorm* cada vez que se solicita el registro de una nueva norma se analiza la norma y se llama al módulo de gestión de normas (*NormativeManager*) que se encargará de llevar a cabo las acciones necesarias en función del tipo de norma. El primer paso necesario para considerar la gestión de normas es desarrollar un analizador que nos permita determinar si una norma es válida con respecto a la especificación BNF de la gramática normativa. Este proceso de análisis Léxico-Sintáctico es llevado a cabo por la clase *NormativeLanguageParser*. Esta clase ha sido generada de forma automática por el generador de analizadores sintácticos *JavaCC (Java Compiler Compiler)* a partir de la gramática del lenguaje normativo. *JavaCC* genera un *parser* para una gramática presentada en notación BNF en código Java. La salida proporcionada por la función principal (llamada *Norm*) de la clase *NormativeLanguageParser* es el árbol de parseado correspondiente a una frase dada. Dicho árbol contiene una serie de

⁷¹ MYSQL <http://www.mysql.com/>

anotaciones que serán empleadas por el módulo analizador para determinar el tipo de norma.

De este modo, el módulo de análisis (*NormativeChecker*) se encarga de *parsear* una norma, haciendo uso del módulo anteriormente comentado, y analizar el árbol de análisis sintáctico para determinar si se trata de una norma de tipo conocido. La función principal de esta clase (*analyzeNorm*) devuelve un array que contiene el tipo de la norma (incompatibility, maxCardinality, o Unknown) y la información contenida en la misma.

Ejemplo: Normas Incompatibilidad

FORBIDDEN role1ID REQUEST AcquireRole

MESSAGE(CONTENT(role 'role2ID'))

La plataforma muestra una interfaz gráfica en la que es posible gestionar los componentes de la misma (SF; OMS, etc.):

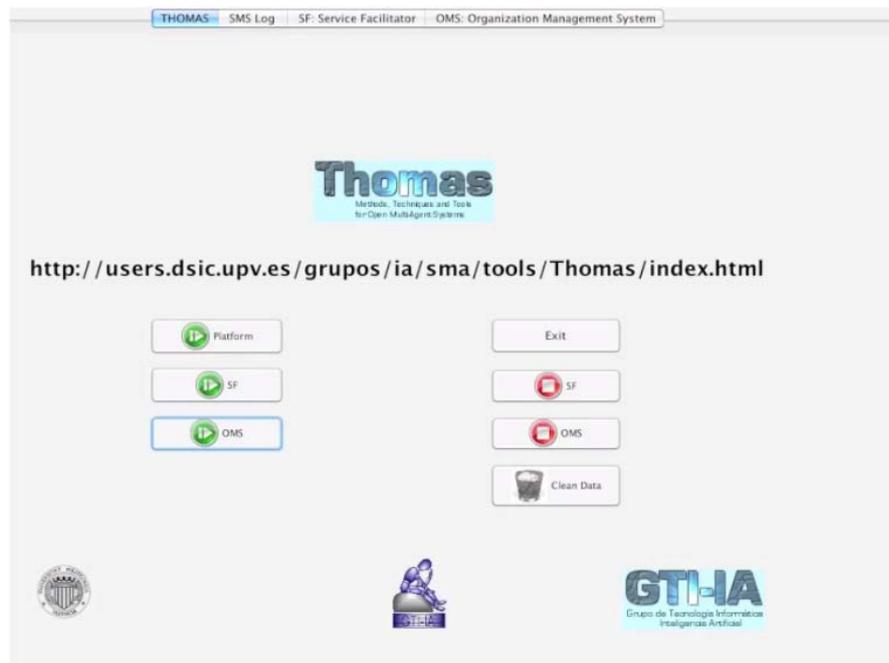


Figura C-6. Interfaz THOMAS

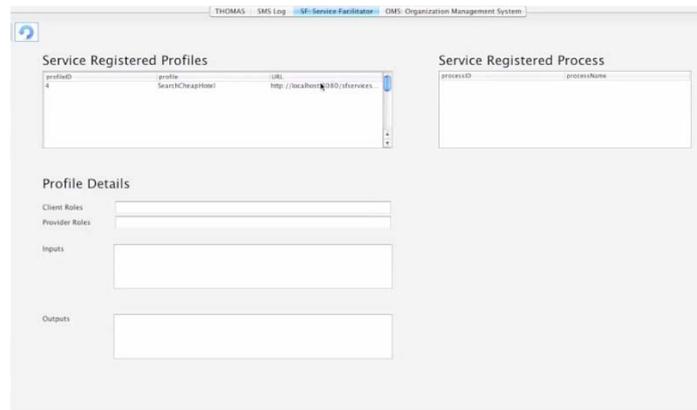


Figura C-7. Interfaz manejo de SF en THOMAS

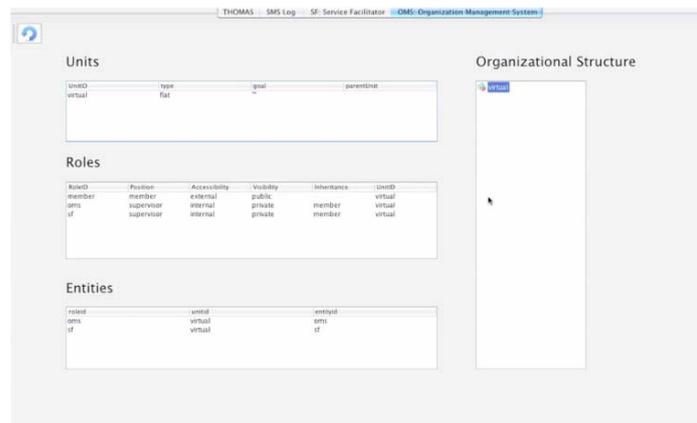


Figura C-8. Interfaz manejo de OMS en THOMAS

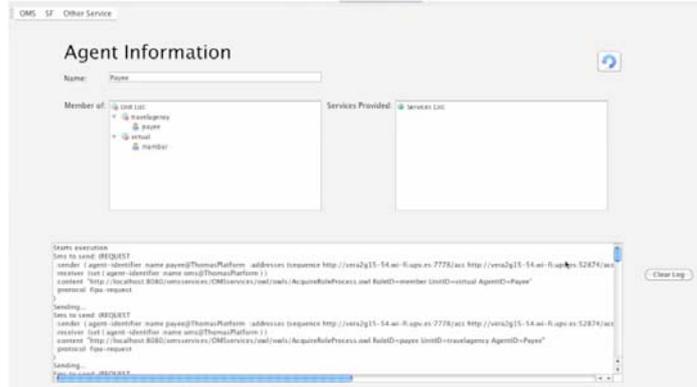


Figura C-9. Interfaz manejo de agentes e interacciones en THOMAS

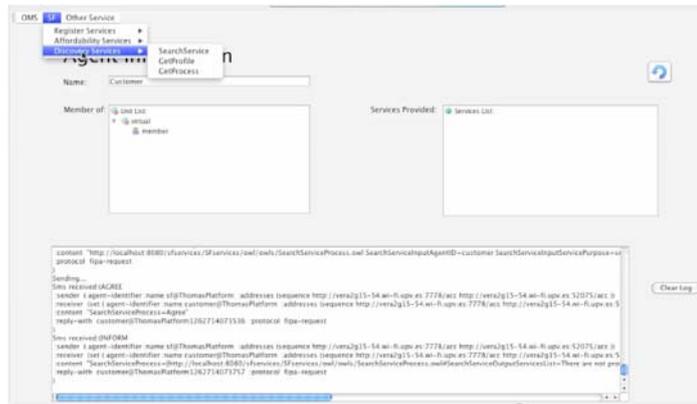


Figura C-10. Interfaz manejo de servicios en THOMAS

C.4 MULAN

Mulan es una biblioteca Java de código abierto dedicado a la minería de datos multi-etiqueta y disponible a través de su página web⁷². Incluye implementaciones de un gran número de algoritmos de aprendizaje, las capacidades básicas para la reducción de dimensionalidad y clasificación multi-etiqueta así como un marco de evaluación extensa.

Se ha utilizado el clasificador Rakel de la Biblioteca de Java para los mecanismos de aprendizaje multi-etiqueta [López-Batista *et al.*, 2011]. Para los

⁷² MULAN <http://mulan.sourceforge.net/index.html>

experimentos se siguió la directiva que está disponible en línea para el programa MULAN y que consta de tres partes:

- 1) Cargamos el conjunto de varias etiquetas de datos que se utilizará para entrenar el modelo. Los datos de entrenamiento de los objetos de aprendizaje se estructuran y se pasan a través de dos archivos a Mulan con la especificación de un conjunto de datos multi-etiqueta de la muestra: un archivo XML especifica los nombres de las etiquetas (LO.xml), y un archivo ARFF especificando los datos reales (LO.arff).
- 2) Creamos una instancia del algoritmo de aprendizaje que queremos entrenar, con el fin de construir un modelo y obtener predicciones. Vamos a crear una instancia del algoritmo Rakel. Rakel es en realidad un algoritmo de meta y puede aceptar cualquier caso con varias etiquetas como parámetro, pero se suele utilizar en conjunción con el algoritmo LP. Se acepta un clasificador de etiqueta única como parámetro. Vamos a utilizar el algoritmo C4.5 de Weka para este fin (J48). Rakel tiene tres parámetros que deben ser seleccionados antes del entrenamiento del algoritmo: a) el tamaño del subconjunto, b) el número de modelos y c) el umbral para el resultado final. Se utilizó un interno de 10 veces la validación cruzada en el conjunto de entrenamiento, La selección automática de estos parámetros es la siguiente: el tamaño del subconjunto se varió de 2 a 5, el número de modelos de 1 a 100 y el umbral de 0,1 a 0,9 con un paso de 0,1. Se llevaron a cabo 10 experimentos diferentes de validación cruzada para la evaluación. Los resultados que siguen son promedios de estas 100 simulaciones de los diferentes algoritmos. El número de vecinos en MLkNN se establece en 10.
- 3) Formamos el clasificador utilizando el conjunto de datos LO que cargamos de LOR diferentes. Dos diferentes experimentos de validación cruzada de 10 veces se llevaron a cabo para la evaluación. Los resultados de salida son promedios en estas 100 simulaciones de los algoritmos.

Para la evaluación empírica de los algoritmos se utiliza el método crossValidate de la clase Evaluador de Mulan. Esto devuelve un objeto MultipleEvaluation, que se imprimió al archivo para ver los resultados en términos de todas las medidas de evaluación aplicables disponibles en Mulan. Para obtener predicciones se llenaron las instancias de datos no etiquetados. El resultado de la predicción es devuelto como una instancia de la clase MultiLabelOutput. Los resultados de la salida contienen los datos de bipartición de etiquetas realizando una clasificación a los datos de entrada una vez entrenado el sistema.

Anexo D PROYECTOS, PUBLICACIONES Y TRABAJOS RELACIONADOS

El proyecto de investigación relacionado directamente con el desarrollo de esta memoria es:

- ❖ Proyecto OVAMAH⁷³: Organizaciones Virtuales Adaptativas: Mecanismos, Arquitecturas y Herramientas. TIN 2009-13839-C03-03, otorgado por el Ministerio de Ciencia e Innovación. Proyectos de Investigación Fundamental No Orientada.
El principal objetivo de este proyecto es el estudio y propuesta de nuevos métodos, herramientas y/o mecanismos que permitan la evolución y organización de organizaciones virtuales basadas en tecnología de sistemas multi-agente. De tal manera que permita la detección de situaciones no cooperativas o no deseadas mediante la evaluación dinámica del sistema así como su posterior tratamiento para que la organización evolucione de forma autónoma y sea capaz de adaptarse a la nueva situación.

Respecto a las publicaciones, el siguiente listado muestra algunas de las más relevantes relacionadas con el trabajo de investigación.

Artículos en revistas internacionales:

- ❖ A.B. Gil, F. de la Prieta and S. Rodríguez. Automatic Learning Object Extraction and Classification in Heterogeneous Environments. Special Issue on Web Engineering for Artificial Intelligence and Distributed Computing in the International Journal of Web Engineering and Technology (IJWET)(Preliminary acceptance).
- ❖ Ana B. Gil, Fernando de la Prieta, Sara Rodríguez and Juan M. Corchado. Smart Recommendation and Recovery of Learning Objects by Virtual Organizations of Agents. *Expert Systems with Applications* ISSN: 0957-4174. Oxford, England (United Kingdom). Factor Impacto (JRC-2010): 1.926. 34/108 Segundo Cuartil (Q2) en Comp. Sciences, Artificial Intelligence, 50/247 Primer Cuartil (Q1) en Engineering, Electrical & Electronic, 15/75 Primer Cuartil (Q1) en Operations Research & Management Science (In press)

⁷³ BISITE <http://bisite.usal.es/webisite/?q=es/node/2>

- ❖ Ana B. Gil, Sara Rodríguez and Juan M. Corchado. Architecture for Intelligent Recovery of Educational content in Heterogeneous Environments. *International Journal of Innovative Computing, Information and Control* (IJICIC), ISSN: 1349-4198. ICIC INT (Ed.). JAPAN. Factor Impacto (JRC-2010): 1.667. 12/60 Q1 en AUTOMATION & CONTROL SYSTEMS; 40/108 Q2 en COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE. (In press). Cristian I. Pinzón, Juan F. De Paz, Ana de Luis, Ana Gil and Javier Bajo. (2010) A Multi-Agent Based Solution to Detect and Block DoS Threats on Web Service. *Journal of Information Assurance and Security*. ISSN 1554-1010, Volume 5, 2010. pp. 455-463.
- ❖ A.-B. Gil and R. Rodríguez and F.J. García-Peñalvo and R. López (2008). SHARP Online: An Adaptive Hypermedia System Applied to Mathematical Problem Solving. *Journal of Universal Computer Science (JUCS)* ISSN: 0948-695X, Vol. 14, No. 19, pp. 3099-3113. Online: http://www.jucs.org/jucs_14_19/sharp_online_an_adaptive. Factor de impacto: Publicado (JCR-2008) 0.488 (JCR-2010): 0.587. Posición: (JCR 2008): 74/86 cuarto cuartil (Q4) en COMPUTER SCIENCE, SOFTWARE ENGINEERING; 71/84 cuarto cuartil (Q4) en COMPUTER SCIENCE, THEORY & METHODS.
- ❖ A.-B. Gil, F.-J. García-Peñalvo, F. (2008). Learner Course Recommendation in e-Learning Based on Swarm Intelligence. *Journal of Universal Computer Science (JUCS)* ISSN: 0948-695X, Vol. 14, No. 16, pp. 2737-2755. online: http://www.jucs.org/jucs_14_16/learner_course_recommendation_in. Factor de impacto: Publicado (JCR-2008) 0.488 (JCR-2010): 0.587. Posición: (JCR 2008): 74/86 cuarto cuartil (Q4) en COMPUTER SCIENCE, SOFTWARE ENGINEERING; 71/84 cuarto cuartil (Q4) en COMPUTER SCIENCE, THEORY & METHODS.
- ❖ Gil-González, A.-B. and García-Peñalvo, F. (2008) Multiagent system for learning objects recovery with context attributes. *Int. J. Computer Applications in Technology* ISSN 09528091, Vol. 33, No. 4, pp.320-326.
- ❖ Gil, A.B; García, F.J. (2003) E-Commerce Recommenders: Powerful Tools for the E-business. *Crossroads issue on Cognitive Science*, ACM. Winter 2003, Issue 10.2. pp. 1-6. ISSN 1528-4981. New York, NY, USA.

Artículos en actas de congresos internacionales y/o de reconocido prestigio:

- ❖ A.B. Gil, S. Rodríguez, F. de la Prieta. and B. Martín. (2011). *Educational Content Retrieval Based on Semantic Web Services*. Seventh International Conference on Next Generation Web Services Practices (NWeSP 2011) pp. 135-140. October 2011. (IEEE conference proceedings Catalog Number: CFP11388-CDR)
- ❖ A.B. Gil, F. de la Prieta and S. Rodríguez. (2011). *Automatic Learning Object Extraction and Classification in Heterogeneous Environments*. Advances in Intelligent and Soft Computing Volume 89. Highlights in Practical Applications of Agents and Multiagent Systems, pp. 109-116. 9th

-
- International Conference on Practical Applications of Agents and Multiagent Systems (PAAMS 2011). Factor De Impacto: (JCR-2007)Top 67 del Computer Science Conference Ranking AI (0.56/1) Posición: 50/701
- ❖ F. De la Prieta, A. Gil, S. Rodríguez, B. Martín. (2011). *BRENHET2, A MAS to Facilitate the Reutilization of LOs through Federated Search*. Trends in Practical Applications of Agents and Multiagent Systems, 177-184. Advances in Intelligent and Soft Computing, Vol. 90. 9th International Conference on Practical Applications of Agents and Multiagent Systems. Corchado, J.M. *et al.* (Eds.) Springer, Heidelberg. ISBN: 978-3-642-19930-1. Factor De Impacto: (JCR-2007) Top 67 del Computer Science Conference Ranking AI (0.56/1) Posición: 50/701
 - ❖ Vivian F. López Batista, Fernando Prieta Pintado, Ana Belén Gil, Sara Rodríguez and María N. Moreno. (2011). *A System for Multi-label Classification of Learning Objects*. Proceedings Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011. Series: Advances in Intelligent and Soft Computing, Vol. 87, 523-531. Corchado, E. *et al.* (Eds.). ISBN: 978-3-642-19643-0.
 - ❖ Ana Gil, Fernando de la Prieta and Vivian F. López. (2010). *Hibryd Multiagent System for Automatic Object Learning Classification*. 5th International Conference on Hybrid Artificial Intelligence Systems (HAIS 2010) E.S. Corchado Rodríguez *et al.* (Eds.): HAIS 2010, Part II, LNAI 6077, pp. 61-68. Springer, Heidelberg.
 - ❖ Francisco de Paz, Sara Rodríguez, Ana Gil, Juan M. Corchado, Pastora Vega. (2010). *Enhanced Self Organized Dynamic Tree Neural Network*. 5th International Conference on Hybrid Artificial Intelligence Systems (HAIS 2010) E.S. Corchado Rodríguez *et al.* (Eds.): HAIS 2010, Part II, LNAI 6077, pp. 85-92. Springer, Heidelberg (2010)
 - ❖ Juan F. De Paz, Ana B. Gil, Emilio Corchado (2010). *Automatic workflow during the reuse phase of a CBP system applied to microarray analysis*. 4th International Workshop on Practical Applications of Computational Biology and Bioinformatics 2010 (IWPACBB 2010) Series: Advances in Intelligent and Soft Computing, Vol. 74 Rocha, M.P.; Fernández Riverola, F.; Shatkay, H.; Corchado, J.M. (Eds.) 2010, ISBN 978-3-642-13213-1, pp. 17-24. 16th-18th June, 2010.
 - ❖ Fernando de la Prieta and Ana Belén Gil. (2010). *A Multi-agent System that Searches for Learning Objects in Heterogeneous repositories*. Advances in Intelligent and Soft Computing (Springer). Trends in Practical Applications of Agents and Multiagent systems: 8th International conference on Practical Applications of agents and multiagent systems (PAAMS 2010). Editors: Yves Demazeau *et al.*, pp. 355-362, Springer, Heidelberg (2010). Factor De Impacto: (JCR-2007) Top 67 del Computer Science Conference Ranking AI (0.56/1) Posición: 50/701.
-

- ❖ Gil, A.B., García F.J. (2007). *Course Recommendation in e-learning Based in Swarm Intelligence*". En las actas del IFIP TC 13 Seminar: Trends in HCI. Salamanca, 26 de Marzo de 2007. pp. 141-160, ISBN: 978-84-690-6531-0.
- ❖ Ana-Belén Gil, Francisco J. García. (2007). *Un Sistema Multiagente de Recuperación de Objetos de Aprendizaje con Atributos de Contexto*. Actas del Taller de Trabajo Zoco'07/CAEPIA. Integración de Aplicaciones Web para la Web Semántica. ISBN: 978-972-8969-04-2, pp. 1-11. 12 de Noviembre, Salamanca, 2007. Factor De Impacto: (JCR-2007) Top 67 del Computer Science Conference Ranking AI (0.55/1).
- ❖ Gil, A. B., and García, F. J. (2002). *Mediated E-Commerce Sites based on Adaptive Multiagent Systems*". En las actas de I International Workshop on Practical Applications of Agents and Multiagent Systems (PAAMS), Corchado, J. M., Siles, J. A., Marcos, D., Alonso, L., Molpeceres, F., García, M., Díaz, F., Fdez-Riverola, F. (Eds.), pp. 31-38, ISBN 84-932864-2-7. 23-25 de Octubre de 2002, Salamanca.

Artículos en actas de congresos:

- ❖ Ana Gil, Sara Rodríguez, Fernando de la Prieta, Beatriz Martín. (2011). *AIREH: Búsqueda y Recuperación de Objetos de Aprendizaje Mediante una Arquitectura de Sistema Multiagente*. 2º Taller sobre Ingeniería del Software en eLearning (ISELEAR'11). Actas del 2º Taller sobre Ingeniería del Software en eLearning. José Luis Sierra y Antonio Sarasa (Eds.) pp.207-220, 2011, ISBN: 978-84-694-7325-2
- ❖ Ana-Belén Gil, María Moreno, Francisco García. (2008). *Integration of Web Mining and Semantic Web*. Actas del "Web Mining and Semantic Web. I Workshop". Moreno M.N., Gil, A.B. y López V.F. (eds.), Universidad de Salamanca, 2008, pp. 45-52. (ISBN: 978-84-691-5945-3), Salamanca, October 27- November 9.
- ❖ Ana-Belén Gil, María Moreno, Francisco García (2008). *Introduction to Web Mining and Semantic Web*. Actas del "Web Mining and Semantic Web. I Workshop". Moreno M.N., Gil, A.B. y López V.F. (eds.), Universidad de Salamanca, 2008, pp. 53-70. (ISBN: 978-84-691-5945-3), Salamanca, October 27- November 9.
- ❖ Ana-Belén Gil, Erla Morales, Francisco J. García. (2007). *E-Learning Multi-agent Recommender for Learning Objects*. Actas del IX Simposio Internacional de Informática Educativa (SIIE'2007), ISBN: 978-972-8969-04-2, pp. 163-169. 14-16 de Noviembre, Porto, Portugal.
- ❖ Morales, E., Gil, A.B., García F.J. (2007). *Arquitectura para la recuperación de Objetos de Aprendizaje de Calidad en Repositorios Distribuidos*. En las actas del SHCA 2007, 11 de Septiembre de 2007, Zaragoza. Actas de Talleres JISBD Vol. 1 - Nº 1 pp. 31-38, ISSN: 1988-3455, 2007.
- ❖ Gil-González, A. (2006). *Semantic Web": Building the Logical Web*. En las actas de Second International Congress on Tools for Teaching Logic.

-
- Salamanca, 26 al 30 de Septiembre de 2006. pp. 45-48, ISBN: 84-690-0348-8.
- ❖ Morales, E., García, F. Barrón, A. Gil, A. (2006). *Sistema de gestión de Objetos de Aprendizaje de calidad*". Actas del III Simposio Pluridisciplinar sobre Objetos y Diseños de Aprendizaje Apoyados en la Tecnología 25, 26 y 27 de Septiembre Oviedo. ISBN: 84-611-5186-8.
 - ❖ Ana Gil, Zahia Guessoum, Francisco García. (2002). *Recomendadores en un Sistema Multiagente Adaptativo para el Comercio Electrónico*". Taller en Sistemas Hipermedia Colaborativos y Adaptativos dentro de las JISBD'2002. El Escorial, 18 al 22 de Noviembre del 2002.
 - ❖ Francisco García, Ana B. Gil. (2002). *Personalización de Sistemas de Recomendación* en el Workshop de Investigación sobre Nuevos Paradigmas de Interacción en Entornos Colaborativos Aplicados a la Gestión y Difusión del Patrimonio Cultural, COLINE'02. Granada, 11 y 12 de noviembre de 2002.
 - ❖ Ana Gil, Francisco García, Zahia Guessoum. (2002). *Dynamic and Adaptive E-Commerce architecture Based on Agent Technologies*. En las actas de 9º Workshop Internacional Design, Specification, And Verification Of Interactive Systems, DSV-IS'2002. Proceedings, Peter Forbrig, Quentin Limbourg, Bodo Urban, Jean Vanderdonck (Eds.), pp. 282-295. Rostock (Alemania), 12-14 de Junio 2002.
 - ❖ A. Gil, F. García, Z. Guessoum. (2002). *An Adaptive Agent Model for E-Commerce Architecture*. En las actas de AH'2002, *Adaptive Hipermedia and Adaptive Web-Based Systems*, Lecture Notes in Computer Science, Vol. LNCS 2347: Paul De Bra, Peter Brusilovsky, Ricardo Conejo (Eds.), pp. 505-509. ISBN: 3-540-43737-1 Springer Verlag. Málaga, 29-31 Mayo 2002.
 - ❖ F.García, F. Paternò, A. Gil. (2002). An adaptive e-commerce system definition. En las actas de AH'2002, *Adaptive Hipermedia and Adaptive Web-Based Systems*, Lecture Notes in Computer Science, Vol. LNCS 2347: Paul De Bra, Peter Brusilovsky, Ricardo Conejo (Eds.), pp. 592-597. ISBN:3-540-43737-1 Springer Verlag. Málaga, 29-31 Mayo 2002.
 - ❖ García, F. J., Gil, A. B., Moreno, M. N., Laguna, M. Á. (2002). An Agent-Based Architecture Definition for Mediated E-Commerce Site. En Proceedings del 2º International Workshop on Object-Oriented Business Solutions WOBS'02. Desarrollado dentro de 5º International Conference on Business Information Systems BIS'02 (April 24, 2002. Poznan, Polonia). R. Corchuelo, A. Ruiz-Cortés, and R. Wrembel (Eds.). Pages 23-30. 2002. Publicado además en BIS 2002: Proceedings of the 5 th International Conference on Business Information Systems. W. Abramowicz (Ed.). Pag. 413-420. ISBN 83-916842-0-2. 2002.
 - ❖ Gil, A. B., García, F. J., Guessoum, Z. (2002). *Adaptive Agents for E-commerce Applications*. In *New Methods and Tools Supporting E-Commerce – Proceedings of the ZOCO Meeting*, R. Corchuelo, A. Ruiz, J. A. Pérez (Eds.), pp 27-36. Editorial Catedral, Salamanca. ISBN 84-96086-02-X. 2002.

- ❖ Ana B. Gil, Francisco García, Zahia Guessoum. (2001). *Adaptive Agents for E-Commerce Applications*. ZOCO en JISBD'2001, pp. 43-52. Almagro, 20 de Noviembre de 2001.

También se ha realizado el registro de propiedad intelectual de la aplicación BRENHET. Dicha aplicación se corresponde a un trabajo preliminar a la propuesta que recoge esta memoria y presentado con los siguientes datos:

AUTORES (P.O. DE FIRMA): Fernando de la Prieta Pintado, Sara Rodríguez González, Javier Bajo Pérez, Juan Francisco de Paz Santana, Davinia Carolina Zato Domínguez, Ana Belén Gil González, Emilio Santiago Rodríguez y Juan Manuel Corchado Rodríguez.

TITULO: BRENHET: Búsqueda de Recursos Educativos en un entorno HETrogéneo.

OBJETO DE PROPIEDAD INTELECTUAL: Programa de ordenador

CLASE DE OBRA: Programa de ordenador

NÚMERO DE SOLICITUD: SA-196-10

NÚMERO DE ASIENTO REGISTRAL: 00/2011/2390

FECHA: 2010

En la segunda convocatoria del programa piloto de prototipos orientados al mercado en TCUE, desarrollado durante el curso 2010/11, fue seleccionado el trabajo titulado "Plataforma multiagente para la recuperación y catalogación de recursos educativos". Beatriz Martín García, estudiante de Ingeniería informática ha realizado un prototipo comercial bajo la tutela de Ana Belén Gil, que valida el presente trabajo de tesis ampliándolo con una perspectiva comercial y un plan de negocio.

Distintas Becas y Estancias han contribuido de manera relevante al desarrollo de este trabajo en distintos Centros entre los que cabe destacar por su duración las realizadas en el LIP6 de la Universidad Pierre et Marie Curie en París (Francia).