



**TESIS DOCTORAL**

***Gestión de Conocimiento Colaborativa  
Mediante Información Semántica***

**AUTOR**

***Jaime Moreno Llorena***

**DIRECTOR DE TESIS**

***Xavier Alamán Roldán***

**PROGRAMA DE DOCTORADO**

***Ingeniería Informática y Telecomunicaciones***

***Departamento de Ingeniería Informática***



***Septiembre de 2008***

*A María, Pablo y Lucía,  
por lo que ha quedado en este trabajo de lo que les pertenecía.*

## **i. Agradecimientos**

Naturalmente el mayor de los agradecimientos es para mi familia y mis amigos que me han soportado todo este tiempo. A María, Pablo y Lucía, por su comprensión y apoyo total, y por lo que ha quedado en este trabajo de lo que les pertenecía. A M<sup>a</sup> Luisa y Jaime con mis hermanos, y a Gloria, José María, Kika y Pedro por el ánimo y la cobertura durante todos estos años. A Manuel y Maca Sendón, y a Begoña Soto y Pablo Martín de la Hoz por el ejemplo y los buenos consejos. Y también a Olga Martín y a Paco Molina por su amistad y su apoyo incondicionales.

Asimismo quiero agradecer a los compañeros que han hecho posible la realización de este trabajo. A Vicente López por abrir la puerta. A Xavier Alamán por su acogida, su ayuda y todo lo que me ha enseñado, por sus revisiones meticulosas y su colaboración continuada en las experiencias con las asignaturas de Sistemas Operativos (SSOO) y Teoría de Autómatas y Lenguajes Formales II (TALF2). A Ruth Cobos por el apoyo y el contraste de impresiones que me ha brindado, por abrir la mina de KnowCat (KC) y no dejar que se cerrase, y por su colaboración en las experiencias realizadas, especialmente en la de Inteligencia Artificial (IA). A José Antonio Macias por la acogida que me dedicó cuando llegué al departamento y los comentarios y buenos consejos que me ha dado desde entonces. A Juana Calle por su paciencia, su buena disposición y su permanente ayuda.

Mi agradecimiento también se dirige a los becarios y colaboradores que han formado el equipo de KC, cuya contribución ha sido imprescindible para mantener el servicio del mismo, y la realización de las experiencias con los prototipos de Semantic KnowCat (SKC). Entre todos ellos agradezco especialmente la colaboración de M<sup>a</sup> Carmen Morillo y Javier Hidalgo, que además han participado en el desarrollo de piezas fundamentales de algunos prototipos de los módulos de SKC.

Tampoco puedo olvidar mi agradecimiento a otros compañeros del departamento que han colaborado en la investigación con sus alumnos. A Rosa Carro y Álvaro Ortigosa, con Xavier y los demás profesores de SSOO, que han participado a lo largo de los años en las experiencias llevadas a cabo con dicha asignatura. A Fernando Díez con Ruth y los demás profesores de IA por su participación en la experiencia realizada con esta asignatura. A Manuel Alfonseca y Antonio Martínez por su colaboración en las experiencias con Sistemas Informáticos 2 (SI2). A Juan de Lara y, de nuevo, a Manuel por su interés por nuestro sistema y los proyectos en los que lo han involucrado. A los cientos de estudiantes que han participado a lo largo de varios años en las actividades con KC en las asignaturas SSOO, IA, TALF2 y SI2, sin ellos no hubiese sido posible este trabajo.

También tienen mi agradecimiento Pablo Haya y Germán Montoro, y el resto de compañeros de ODISEA, por su disposición, tolerancia y apoyo, que han permitido mantener los servidores de KnowCat en funcionamiento durante tantos años para hacer posible este proyecto.

Por último, no quiero dejar de agradecer su colaboración a los compañeros de las asignaturas en las que imparto docencia, Informática General y Estructura de Datos y de la Información I, que me han facilitado las cosas cuando me ha hecho falta, entre ellos a Rosa Carro, Eduardo Serrano, José Dorronsoro, Germán Montoro, Iván Catador, David Vallet y especialmente Antonio Carlos Fernández.

## **ii. Resumen**

Uno de los efectos indeseables de la introducción de las tecnologías de la información y las comunicaciones, Internet y la Web ha sido la sobrecarga de información. Se están buscando soluciones a este problema desde distintos campos de investigación, Gestión del Conocimiento, Minería de Información, Web Semántica, etc. Una hipótesis interesante para buscar soluciones a este problema en el ámbito de los sistemas de gestión de conocimiento en red, es aprovechar la energía residual de la actividad de personas, servicios y otras entidades que interaccionan con ellos, y las características de los elementos implicados, comunidades de usuarios, conocimiento y red. El sistema de gestión de conocimiento en red KnowCat es un entorno basado en Web que permite la gestión del conocimiento gracias a la interacción de una comunidad de usuarios sin necesidad de supervisión humana. Semantic KnowCat es un nuevo sistema propuesto para comprobar esta hipótesis integrando la aproximación de KnowCat e ideas procedentes de distintos campos de investigación: Gestión del Conocimiento, Trabajo Colaborativo, Web, Hipermedia Adaptativa y la Minería de Información. Disciplinas estas que en su mayoría concurren en el campo de investigación emergente de la Web Semántica.

### **iii. Abstract**

One of the undesirable effects produced by the introduction of information and communication technologies, the Internet and the Web, has been the information overload. Solutions for this problem are being looked in different research fields, such as Knowledge Management, Information and Data Mining, Semantic Web, etc. One interesting hypothesis for finding a solution to this problem in the network knowledge management systems context is to use the characteristics of the elements involved in the knowledge management, the user's community, the knowledge and the network, and to use residual power from the activity of people, services and other entities that interact with them. KnowCat is an original system for knowledge management over the Web, which facilitates knowledge management as the result of the user's community interaction without need for has supervision. This approach is closely related to the proposed hypothesis and provides a good starting point for research it. Semantic KnowCat is the new system proposed for proving this hypothesis, which integrates the KnowCat approach and ideas from some other research fields: Knowledge Management, Human Computer Interaction, Computer-Supported Cooperative Work, Information and Data Mining and Semantic Web.

## iv. Índice

1	Introducción.....	1-1
1.1	Motivación.....	1-2
1.2	Contexto.....	1-3
1.3	Propuesta.....	1-9
1.4	Objetivos de la investigación.....	1-11
1.5	Principales contribuciones .....	1-11
1.6	Terminología.....	1-13
1.7	Plan de tesis .....	1-14
2	Estado de la Cuestión .....	2-1
2.1	Planteamiento del Problema .....	2-2
2.2	Campos de Investigación Relacionados.....	2-6
2.2.1	Gestión del Conocimiento.....	2-7
2.2.2	Interacción Persona-Ordenador.....	2-13
2.2.3	Web Semántica .....	2-48
2.2.4	Minería de Información y de Datos .....	2-68
2.3	Sistema KnowCat (KC) .....	2-81
2.4	Conclusiones.....	2-85
3	Aproximación Conceptual.....	3-1
3.1	Propuesta.....	3-2
3.2	El sistema Semantic KnowCat (SKC) .....	3-7
3.2.1	Aspectos Generales .....	3-7
3.2.2	Conocimiento .....	3-9
3.2.3	Comunidad de Usuarios y Fases del Sistema .....	3-14
3.2.4	Nodos y Redes de Nodos.....	3-16

3.2.5	Seguimiento de actividades .....	3-17
3.2.6	Cristalización del conocimiento .....	3-17
3.3	Propuesta de Arquitectura para SKC .....	3-19
3.3.1	Ontologías .....	3-23
3.3.2	Gestor de Base de Datos.....	3-28
3.3.3	Motor de Análisis .....	3-30
3.3.4	Motor de Cristalización .....	3-37
3.3.5	Módulo de Interacción.....	3-39
3.4	Conclusión .....	3-45
4	Aproximación experimental .....	4-1
4.1	Objetivos Experimentales .....	4-2
4.2	Prototipo de SKC sobre KC.....	4-3
4.2.1	Monitor de Cliente .....	4-4
4.2.2	Módulo de Análisis de Conocimiento.....	4-15
4.2.3	Módulo de Red de Nodos para soporte de la Interacción entre Instancias del Sistema.....	4-29
4.3	Campo Experimental: Proyectos de Innovación Docente con KC .....	4-39
4.3.1	Objetivos generales de los PIDs.....	4-40
4.3.2	Relación de asignaturas participantes en los PIDs .....	4-41
4.3.3	Metodología general de las actividades con el KC en los PIDs.....	4-43
4.3.4	Experiencias de los PIDs empleadas en este trabajo de investigación.....	4-45
4.4	Experimentos realizados y resultados obtenidos .....	4-51
4.4.1	Medida de la interacción de los usuarios en el lado del cliente: intensidad de la interacción de los usuarios, interés y aproximación a la selección del conocimiento sin manifestación explícita de opinión.....	4-52
4.4.2	Calificación semántica de ítems de conocimiento: Asociación automática entre ítems de conocimiento, recomendación y nuevas vistas del conocimiento.....	4-56
4.4.3	Interés de los usuarios: Descubrimiento y aplicación .....	4-62



4.4.4	Red de nodos SKC: Agentes de red, descubrimiento y vinculación de nodos, y establecimiento de red.....	4-71
5	Conclusiones y Trabajos Futuros .....	5-1
5.1	Recapitulación .....	5-2
5.2	Conclusiones y Principales Contribuciones.....	5-15
5.3	Difusión de los resultados.....	5-19
5.4	Cuestiones Abiertas y Trabajos Futuros .....	5-20

Referencias

Apéndices 1: Especificación Formal del Sistema SKC

## v. Figuras

### Estado de la Cuestión (Capítulo 2)

Figura 2-1.	<i>Evolución del volumen información digital de acuerdo con [IDC 2007].....</i>	2-5
Figura 2-2.	<i>Mapas de calor, CrazyEgg [CrazyEgg 2008] izquierda y ClickHeat [ ClickHeat 2008] derecha .....</i>	2-24
Figura 2-3.	<i>Sistema de Eye-tracker en sujeto [Prendinger 2007] y mapa de calor [EyetrackIII 2008] .....</i>	2-24
Figura 2-4.	<i>Cluster Map [Geroimenko 2002].....</i>	2-37
Figura 2-5.	<i>Topic Maps [Geroimenko 2002] .....</i>	2-37
Figura 2-6.	<i>Interfaz hiperbólica para ontologías [Fensel 2001] .....</i>	2-38
Figura 2-7.	<i>Ejemplo de fichero DOT de Graphviz [Gansner 2000] .....</i>	2-39
Figura 2-8.	<i>Ejemplo de fichero etiquetado por Freeling [Carreras 2004].....</i>	2-41
Figura 2-9.	<i>Documentos en un espacio de términos de dos dimensiones .....</i>	2-71
Figura 2-10.	<i>Descomposición de matriz con SVD de acuerdo con [JorgeBotana 2007].....</i>	2-74
Figura 2-11.	<i>Efecto reducción de dimensionas empleando SVD de acuerdo con [JorgeBotana 2007] .....</i>	2-75
Figura 2-12.	<i>Ejemplo de fichero de registro de actividad de servidor Web .....</i>	2-79

### Aproximación Conceptual (Capítulo 3)

Figura 3-1.	<i>Diagrama Simplificado de Casos de Uso de SKC .....</i>	3-19
Figura 3-2.	<i>Arquitectura general del sistema SKC propuesto .....</i>	3-21
Figura 3-3.	<i>Conceptos generales considerados en SKC .....</i>	3-23
Figura 3-4.	<i>Conceptos relacionados con el de Consideración .....</i>	3-24
Figura 3-5.	<i>Conceptos relacionados con el de Conocimiento del Nodo.....</i>	3-24
Figura 3-6.	<i>Conceptos relacionados con el de Grafo de Conocimiento de Nodo.....</i>	3-26
Figura 3-7.	<i>Conceptos relacionados con el de Documento .....</i>	3-26
Figura 3-8.	<i>Modelización de la Comunidad de Nodo .....</i>	3-27

<i>Figura 3-9. Gestor de Base de Datos de SKC</i> .....	3-29
<i>Figura 3-10. Módulo de Interactividad de SKC</i> .....	3-39
 <u><i>Aproximación Experimental (Capítulo 4)</i></u>	
<i>Figura 4-1. MC en el cliente (CW3) y el servidor (SW3)</i> .....	4-5
<i>Figura 4-2. Líneas típicas del LOG de un servidor Web</i> .....	4-8
<i>Figura 4-3. Línea del MC en LOG de un servidor Web</i> .....	4-8
<i>Figura 4-4. Fragmento de LOG de un servidor Web de un nodo SKC con MC</i> .....	4-9
<i>Figura 4-5. Ejemplo análisis de interfaz para establecer elementos objeto de la actividad</i> .....	4-12
<i>Figura 4-6. Proceso análisis actividad de usuarios en el cliente de SKC, a partir de LOG Web con registros del MC</i> .....	4-14
<i>Figura 4-7. Primeras líneas fichero resultado análisis</i> .....	4-14
<i>Figura 4-8. Proceso obtención vectores peso palabras para ítems conocimiento desde HTML</i> .....	4-19
<i>Figura 4-9. Vista interactiva del conocimiento como grafo de ítems relacionados (ventana de arriba) y servicio de recomendación sensible al contexto (ventanas de abajo, área inferior en la de la izquierda y toda la de la derecha)</i> .....	4-24
<i>Figura 4-10. Líneas de registro del MC extraídas del LOG de un nodo SKC</i> .....	4-25
<i>Figura 4-11. Proceso análisis actividad por usuario en el cliente de SKC, a partir de LOG Web con registros del MC</i> .....	4-26
<i>Figura 4-12. Bases de datos de los nodos SKC</i> .....	4-34
<i>Figura 4-13. Página de nodos SKC en el servidor para su descubrimiento en la Web</i> .....	4-36
<i>Figura 4-14. Mapa de red de nodos SKC del prototipo</i> .....	4-38
<i>Figura 4-15. Integración de nodos de red SKC en marco de recomendación del prototipo del sistema</i> .....	4-39
<i>Figura 4-16. Figura 1. Repositorio o área de conocimiento sobre Inteligencia Artificial</i> .....	4-49
<i>Figura 4-17. C frente a I en la experiencia de IA</i> .....	4-54
<i>Figura 4-18. C frente a I en la experiencia de SO</i> .....	4-54

<i>Figura 4-19.</i>	<i>Clasificación automática de documentos por temas en nodo KC sobre SSOO.....</i>	<i>4-57</i>
<i>Figura 4-20.</i>	<i>Agrupación automática de documentos por temática de área de conocimiento, en un nodo de SSOO .....</i>	<i>4-58</i>
<i>Figura 4-21.</i>	<i>Agrupación automática de documentos por temática de área de conocimiento en nodo de SI.....</i>	<i>4-59</i>
<i>Figura 4-22.</i>	<i>Mapeo entre temas de dos nodos KC de SSOO con distintos documentos pero árboles de conocimiento iguales.....</i>	<i>4-60</i>
<i>Figura 4-23.</i>	<i>Mapeo entre temas de dos nodos KC de TALF con distintos documentos y árboles de conocimiento .....</i>	<i>4-61</i>
<i>Figura 4-24.</i>	<i>Agrupación de nodos KC por temáticas .....</i>	<i>4-62</i>
<i>Figura 4-25.</i>	<i>Código cromático para interés relativo en experiencia SI. ....</i>	<i>4-64</i>
<i>Figura 4-26.</i>	<i>VIUs de los dos periodos de actividad de la experiencia en nodo de SI.....</i>	<i>4-65</i>
<i>Figura 4-27.</i>	<i>VIUs de los cinco periodos de actividad de la experiencia en nodo de TALF.....</i>	<i>4-67</i>
<i>Figura 4-28.</i>	<i>Agrupación de usuarios por interés mediante sus VPPIUs. ....</i>	<i>4-69</i>
<i>Figura 4-29.</i>	<i>Página de Nodos SKC Activos indexada por Google. ....</i>	<i>4-75</i>

## **vi. Tablas**

### Estado de la Cuestión (Capítulo 2)

Tabla 2-1.	<i>Cantidades de algunos dispositivos digitales y de abonados de telefonía móvil en 2006.....</i>	<i>2-3</i>
Tabla 2-2.	<i>Evolución del uso de la red propuesta por Baeza [Baeza 2006].....</i>	<i>2-4</i>
Tabla 2-3.	<i>Recursos para la gestión de conocimiento según Cobos en [Cobos 2002b].....</i>	<i>2-12</i>
Tabla 2-4.	<i>Clasificación de aplicaciones y servicios de groupware según la coincidencia de los participantes en el tiempo y el espacio.....</i>	<i>2-15</i>
Tabla 2-5.	<i>Clasificación de aplicaciones y servicios de groupware según la coincidencia de los participantes en el tiempo.....</i>	<i>2-15</i>
Tabla 2-6.	<i>Clasificación de los sistemas Hipermedia Adaptativos según Medina [Medina 2000].....</i>	<i>2-26</i>
Tabla 2-8.	<i>Características de algunas herramientas de mapeo entre ontologías [Choi 2006].....</i>	<i>2-61</i>
Tabla 2-9.	<i>Conceptos relacionados con el modelo vectorial en el campo de la minería de texto.....</i>	<i>2-72</i>

### Aproximación Experimental (Capítulo 4)

Tabla 4-1.	<i>Relación de asignaturas participantes en los TIDs.....</i>	<i>4-42</i>
Tabla 4-2.	<i>Nodos conocidos por cada nodo existente.....</i>	<i>4-72</i>
Tabla 4-3.	<i>Coeficientes de similitud entre nodos.....</i>	<i>4-73</i>

### Conclusiones y Trabajos Futuros (Capítulo 5)

Tabla 5-1.	<i>Resumen de objetivos y de componentes prototipadas.....</i>	<i>5-9</i>
Tabla 5-2.	<i>Resumen de grupos de experimentos.....</i>	<i>5-10</i>
Tabla 5-3.	<i>Resumen experimentos realizados.....</i>	<i>5-10</i>
Tabla 5-4.	<i>Resumen resultados obtenidos en los experimentos realizados por series.....</i>	<i>5-12</i>

## **vii. Fórmulas**

### Estado de la Cuestión (Capítulo 2)

<i>Fórmula 2-1. Peso de palabra en documento en agrupación y categorización.....</i>	<i>2-72</i>
<i>Fórmula 2-2. Peso de palabra en documento en recuperación de la información [Baeza 1999.....</i>	<i>2-72</i>
<i>Fórmula 2-3. Módulo de un vector.....</i>	<i>2-72</i>
<i>Fórmula 2-4. Distancia euclídea entre dos vectores.....</i>	<i>2-72</i>
<i>Fórmula 2-5. Coseno ángulo entre dos vectores.....</i>	<i>2-73</i>
<i>Fórmula 2-6. Producto escalar .....</i>	<i>2-73</i>
<i>Fórmula 2-7. Coseno ángulo vectores función componentes.....</i>	<i>2-73</i>

### Aproximación Experimental (Capítulo 4)

<i>Fórmula 4-1. Peso palabra en documento y frecuencia término documento inverso en colección.....</i>	<i>4-20</i>
<i>Fórmula 4-2. Similitud entre vectores en función del coseno del ángulo que forman.....</i>	<i>4-21</i>
<i>Fórmula 4-3. Peso de palabra en el interés de usuario .....</i>	<i>4-28</i>

# 1 Introducción

En este capítulo se esbozan las razones que han estimulado el trabajo de investigación que nos ocupa, así como los elementos que han servido de punto de partida y de referencia para el mismo. Además se avanza la propuesta planteada junto con los objetivos pretendidos y una previsión de las contribuciones esperadas. Después se dan unas reseñas terminológicas y se traza el plan de trabajo que se va a seguir. A lo largo de los siguientes capítulos que integran esta memoria se desarrollan con mayor profundidad las cuestiones apuntadas en esta introducción.

## **1.1 Motivación**

Uno de los efectos negativos de la extensión del uso de las tecnologías de la información y las comunicaciones (TIC) en los últimos años ha sido la sobrecarga de información, la Web es un caso paradigmático de este fenómeno. En un periodo de tiempo muy reducido se ha pasado de tener poca información digital disponible a tener tanta [IDC 2007] que resulta inabarcable, poco manejable, difícil de clasificar y de seleccionar por su calidad -utilidad, valor, precisión y relevancia-, por lo que acaba siendo menos útil de lo que cabría esperar. Por ello, desde hace algún tiempo se ha hecho imprescindible establecer mecanismos para afrontar tal sobrecarga de información.

El problema resulta especialmente grave cuando la información implicada constituye repositorios de conocimiento, de información que permite aprovechar la información, los cuales se van haciendo imprescindibles para los grupos a los que se dirigen y terminan siendo esenciales para el desarrollo de los organismos a las que pertenecen y, en ocasiones, para el progreso de las áreas de conocimiento a las que se refieren. En el caso de la Web, la cuestión adquiere una dimensión global, debido a que en ella tiene cabida cualquier campo de conocimiento y alcanza a comunidades de usuarios cada vez más amplias, que en muchas ocasiones no disponían previamente de acceso a repositorios de conocimiento semejantes y que ahora tienen la oportunidad de emplearlos para su desarrollo.

Las primeras soluciones propuestas pretendían atajar el problema con procedimientos manuales empleando equipos humanos destinados a la selección de la información y del conocimiento, ejemplos de esta estrategia son los primeros directorios Web como el de los años iniciales de Yahoo! [Yahoo 2008]. Actualmente no es posible manejar el volumen de información y de conocimiento existentes sin la ayuda de las tecnologías que propician su crecimiento y sin procedimientos más sofisticados de tratamiento y explotación. Desde distintos campos de investigación se buscan soluciones a este problema de gestión de información y conocimiento, campos como Minería de Información y de Datos [Chang 2001], Interacción Persona-Ordenador [SIGCHI 2008] o Web Semántica [Berners-Lee 2001], entre otros.

Para indagar el fenómeno de la sobrecarga de información y sus soluciones es imprescindible disponer de algún sistema que reúna las características más destacadas de los ámbitos donde se manifiesta el problema, pero con unos parámetros y una escala controlables. En nuestra opinión los sistemas de gestión de conocimiento en red reúnen estas características y la Web podría considerarse desde alguna perspectiva como un inmenso sistema de gestión de conocimiento, como se explicará después en el capítulo de Aproximación Conceptual.



La hipótesis que proponemos para buscar remedio al problema de la sobrecarga de información en los sistemas de gestión de conocimiento en red, es explotar las características de dichos sistemas y de los elementos que concurren en ellos. Esto creemos que se podría conseguir aprovechando la energía residual de la actividad de los propios sistemas y de las entidades que interactúan con estos, y sacando partido de las características de los elementos implicados en sus labores de gestión, que son los usuarios, el conocimiento y la propia red.

No parece que las bases necesarias para comprobar esta suposición se encuentren en un solo dominio de conocimiento, sino que haya que buscarlas en la integración de ideas procedentes de distintos campos. Esto es algo frecuente en las áreas de investigación donde concurren diversas disciplinas como ocurre con el tema de este trabajo, que se sitúa en la confluencia en la Web de la Gestión del Conocimiento, la Minería de Información y Datos, y la Interacción Persona-Ordenador, disciplinas todas ellas que convergen en el ámbito de la Web Semántica.

Por lo tanto, este trabajo indaga en el aprovechamiento de la energía excedente de la actividad de los sistemas de gestión de conocimiento en red y las entidades involucradas en ella, investiga la utilización de las características de todos los elementos implicados en su gestión, y busca las ideas y técnicas en varias áreas de conocimiento, todo ello con el objetivo de encontrar vías de solución para el problema de sobrecarga de información en los sistemas de gestión de conocimiento en red y en entornos similares, como la propia Web.

## **1.2 Contexto**

Como se ha dicho, la comprobación de la hipótesis propuesta requiere de un enfoque multidisciplinar en el que deben considerarse áreas de conocimiento diversas. De las más importantes se habla brevemente a continuación, señalando para cada una su relación con el trabajo de investigación del que trata este documento. Además, se presentan de forma concisa algunos sistemas existentes relacionados con el problema planteado y con las vías de solución que se proponen. Por último, se dedican unas líneas al sistema de gestión de conocimiento KnowCat [Alamán 1999][Cobos 2003], que es el antecedente de este trabajo y se emplea como punto de partida de él y como plataforma de prototipado del sistema propuesto en el mismo. En el siguiente capítulo se vuelve sobre estos temas para profundizar en los aspectos relevantes para la propuesta dentro de los campos de conocimiento mencionados y los sistemas a los que se hace referencia.

### *Gestión del Conocimiento*

El campo de estudio de la Gestión del Conocimiento es la colección, organización, clasificación y distribución del conocimiento [Malthotra 2000]. La mayoría de las definiciones de conocimiento se refieren a la información y al grupo de personas que interactúa con ella movido

por algún interés [McDermott 1999]. Se podría decir que el conocimiento se manifiesta en una comunidad cuando los individuos de la misma emplean la información de que disponen para resolver problemas o llevar a cabo proyectos. Para soportar la gestión del conocimiento no es suficiente que el sistema empleado proporcione la infraestructura para guardar y manejar la información involucrada, sino que además deberá atender las necesidades de interacción y colaboración del grupo [Churman 1971].

El modo en que se trata de la información en el ámbito de Gestión del Conocimiento sirve de base para la aproximación que se propondrá en este trabajo de dos formas. Por un lado, se considera el interés de los usuarios como la fuente de energía del mecanismo de selección de información planteado. Por otro lado, se comparte con dicho ámbito la idea de que los sistemas que soportan dichos mecanismos deben dar cobertura a las necesidades de interacción y colaboración de los usuarios, además de almacenar y facilitar la manipulación de la información que constituye el conocimiento que permiten gestionar, puesto que se considera que la actividad de grupo constituye el motor del mecanismo propuesto.

#### *Interacción Persona-Ordenador*

El campo de estudio de la Interacción Persona-Ordenador (IPO) [SIGCHI 2008] es el diseño, la implementación y la prueba de sistemas interactivos destinados al uso humano y de todos los fenómenos relacionados con tales sistemas. Vinculados con este campo están algunas áreas de especial interés para el trabajo que nos ocupa, entre las que cabe destacar Groupware, Awareness, Monitorización de Actividad, Hipermedia Adaptativa, Recomendación, Visualización de la Información, y Procesamiento del Lenguaje Natural, así como la aplicación de algunas de ellas en la llamada Web 2.0. El Groupware es una tecnología basada en las TIC diseñada para facilitar el trabajo en grupo, de la que se ocupa el área de conocimiento de CSCW -Computer\_Supported Cooperative Work- [ForakerDesign 2005]. En el ámbito del CSCW el Awareness [Dourish 1992] se refiere a la comprensión de la actividad de los demás participantes en un entorno colaborativo, que establece un contexto para la propia actividad en el entorno. La Monitorización de la Actividad del Usuario integra una serie de técnicas dirigidas a conseguir datos sobre la interacción de los usuarios [Hilbert 2000] que puedan ser aprovechadas por el sistema. La Hipermedia Adaptativa [Brusilovsky 1998] es una línea de investigación en el área de los sistemas adaptativos de usuario, y su objetivo es incrementar la funcionalidad de hipermedia acomodándola al usuario. La Recomendación [Adomavicius 2005] trata de filtrar la información para seleccionar los ítems de la misma que puedan resultar de más interés para cada usuario. La Visualización de la Información, que pretende producir representaciones gráficas de la estructura de información abstracta para usuarios humanos. El objetivo de la Visualización de la Información [Geroimenko 2002] es revelar patrones, tendencias y otros aspectos de un fenómeno, centrándose en la información abstracta, y representar la información no espacial y

no numérica en una forma visual efectiva. El Procesamiento del Lenguaje Natural forma parte del campo de la Inteligencia Artificial y se ocupa de la investigación de mecanismos computacionalmente eficaces para emplear el lenguaje natural en la comunicación entre personas y máquinas. Por último, la Web 2.0 [O'Reilly 2005] representa las tendencias dominantes en la Web en los últimos años, donde han cobrado gran importancia una nueva filosofía de las aplicaciones Web y la combinación en ellas de diversas tecnologías disponibles. Las aplicaciones características de la Web 2.0 aprovechan la estructura de la Web, las redes sociales, la inteligencia colectiva y la colaboración, son ejemplo de ellas los weblog [Weblog 2008], las folksonomías [Golder 2006] y las wikis [Leuf 2001]. La Web 2.0 integra tecnologías ya existentes para dotar a las interfaces Web de prestaciones parecidas a las aplicaciones interactivas del ordenador, a todas ellas se las hace referencia con el término AJAX [Woychowsky 2007] [Zakas 2007].

Del campo de Interacción Persona-Ordenador y de las áreas de conocimiento vinculadas con él se toman ideas y técnicas fundamentales para la aproximación que se planteará. El área de conocimiento de CSCW y las tecnologías de Groupware, incluido el Awareness, proporcionan las bases de los sistemas de trabajo colaborativo, como el que servirá para atender las necesidades de interacción y colaboración de los usuarios en el sistema que se propondrá [Churman 1971]. La Monitorización de la Actividad del Usuario permite obtener datos de su interacción, de forma transparente para él, que pretenden evidenciar su posible interés por algunos de los recursos accesibles a través del sistema que se sugerirá. La Hipermedia Adaptativa proporciona las bases para el modelado de los elementos del sistema que se propondrá, y para la anotación automática, el registro de actividad de dicho sistema y el empleo de elementos virtuales. La Recomendación se incorpora en el sistema que se planteará como medio de sacar provecho de la actividad excedente de los usuarios, de facilitar sus tareas, de evidenciar relaciones entre ellos y con otros elementos del sistema, y de mejorar la gestión de conocimiento. La Visualización de Información, aporta los fundamentos necesarios para presentar en dicho sistema la información disponible en cada momento de forma adecuada para que los usuarios puedan aprovecharla. El área de Tratamiento del Lenguaje Natural, contribuye con las técnicas necesarias para procesar de forma automática el conocimiento expresado de forma natural en el sistema que se propondrá, reduciendo de este modo los requerimientos de atención de los usuarios durante el proceso. Finalmente, la aproximación y el sistema que se presentarán encuentran en la Web 2.0 el marco adecuado para su planteamiento basado en el aprovechamiento de la red, de la inteligencia colectiva y de la colaboración, y su preocupación por conseguir que la interfaz de usuario se adecue a las necesidades de la actividad y a los datos que permite manejar.

### Minería de Información y de Datos

El campo de conocimiento de la Minería de Información y de Datos [Chang 2001] se ocupa de las técnicas para la extracción y tratamiento de contenidos disponibles en bases de datos. En los últimos años se han dado varias circunstancias que han propiciado el desarrollo de este campo. Por un lado, estaban disponibles varias técnicas sólidas y maduras aplicables al análisis de datos. Por otro lado, el abaratamiento de los dispositivos de almacenamiento y el actual desarrollo de la infraestructura de comunicaciones han propiciado la creación de grandes bases de datos y de mecanismos de acceso a ellas. Además, el abaratamiento de la capacidad de cálculo ha permitido disponer de la potencia de procesamiento necesaria.

La Minería de Texto es un campo de la Minería de Información que trabaja sobre información textual poco estructurada, a diferencia de la Minería de Datos que lo hace sobre información completamente estructurada. Su aplicación permite la detección de patrones en documentos y colecciones de documentos textuales. De la aplicación de la Minería de Información y Datos a la información y conocimiento accesibles a través de la Web y generados por la actividad de los servidores, ha surgido la Minería Web. Estas tecnologías incluyen procedimientos que soportan los sistemas de búsqueda en la Web, las aplicaciones de análisis de los registros de los servidores WWW, y otros sistemas y aplicaciones relacionadas con la Web. [Chang 2001]

Las técnicas de Minería de Información y Datos, en particular las propias del tratamiento de Textos y las utilizadas en el ámbito de la Web [Baeza 1999], se emplean de varias maneras en la aproximación que se propondrá. En primer lugar, se aplican en la implementación de instrumentos automáticos para la obtención de información sobre el conocimiento tratado, sobre los usuarios del mismo y sobre la interacción que se establece entre unos y otros, utilizando para ello el análisis de los registros de actividad del sistema y de los propios contenidos. Dicha información se utiliza para enriquecer el conocimiento tratado por el sistema y también para facilitar y mejorar su gestión. Además, se propone emplear tal información en los algoritmos de selección del conocimiento, junto con los datos obtenidos por otros procedimientos.

### Web Semántica

La Web Semántica [Berners-Lee 2001] es una propuesta de desarrollo de la Web actual hacia una red de recursos estructurados formalmente y enriquecidos con meta-información semántica explícita, que permita una mejor explotación del sistema empleando mecanismos automáticos y sin necesidad de participación humana directa. Las ontologías [Mahesh 1996] son expresiones formales del entendimiento común y compartido de un dominio. El uso de ontologías surge como respuesta a la necesidad de reutilizar y compartir el conocimiento. Este concepto constituye un elemento esencial de la Web Semántica, que utiliza las ontologías para dotar de significado explícito a los recursos que integra. La separación entre información y su modo de

presentación a través de la Web es otra característica importante de la Web Semántica. Esta idea da lugar al concepto de Documento Virtual [Milosavljevic 1999], que es un documento Web cuyos contenidos, enlaces, o ambas cosas, se generan en el momento de su solicitud. Dicha idea también proporciona el vínculo de la Web Semántica con los Sistemas Hipermedia Adaptativos [Brusilovsky 1998], que pueden considerarse como un antecedente de la aplicación de dicho concepto. Los Agentes Inteligentes [Dinverno 2001] están llamados a convertirse en una categoría de usuarios artificiales fundamentales de la Web Semántica. Se trata de entidades computacionales capaces de satisfacer las necesidades de otras entidades -artificiales o humanas-. Para conseguirlo se basan en sus propios conocimientos, se mueven por peticiones expresas o por su iniciativa, y son capaces de recurrir a otras entidades para conseguir sus objetivos. Los Agentes Inteligentes están dotados de autonomía, habilidades sociales, reactividad, y pro-actividad; cualidades y habilidades que les permiten actuar como entes independientes, interactuar con otras entidades, ser sensibles al entorno, actuar sobre él y tomar iniciativas; todo ello para alcanzar sus objetivos.

El interés del campo de investigación de la Web Semántica es muy cercano al de este trabajo, ambos indagan en cómo solucionar el problema de la sobrecarga de información en entornos de red. La Web Semántica pretende facilitar la explotación de la Web sin necesidad de intervención humana directa, dotando a la Web actual de la infraestructura necesaria para ello. La Web en su conjunto puede verse como un sistema de gestión de conocimiento global, aunque es mucho más que eso. En la propuesta de este trabajo se trata de conseguir unos objetivos afines en un ámbito más restringido, el de los sistemas de gestión de conocimiento en red. En este caso, se aspira a aprovechar todo lo posible la interacción humana, que se considera esencial para la selección del conocimiento, pero facilitando dicha interacción con recursos técnicos adecuados que resulten transparentes para el usuario. Las ideas y técnicas del área de la Web Semántica pueden aplicarse al ámbito de este trabajo y los resultados de esta investigación pueden resultar de utilidad en el terreno de dicha área. Además del marco general que proporciona la Web Semántica para la aproximación propuesta, como el empleo de ontologías, tienen especial importancia dos campos vinculados con ella. Uno es el área de conocimiento de la Hipermedia Adaptativa, ya mencionada en relación con IPO, cuyas ideas se aplican en las interfaces interactivas del sistema que se propondrá. El otro es el área de investigación de los Agentes Inteligentes, que se trata en la automatización de ciertas interacciones entre instancias del sistema que se presentará y se propone como soporte para dotar al entorno del carácter autónomo y pro-activo que se considera esencial.

#### *Sistemas Relacionados*

Actualmente existen varios sistemas que emplean técnicas consideradas en este trabajo. Los indexadores de la Web, como Google [Google 2008] o Yahoo! [Yahoo 2008], emplean

procedimientos de minería de texto para facilitar el acceso a la información publicada en la red. Algunos de estos sistemas utilizan, además, ciertas características de los documentos que manejan -sobre todo páginas Web- para realizar clasificaciones automáticas -por ejemplo, determinando el soporte, el idioma o la nacionalidad del objeto- y establecer la importancia de los mismos -atendiendo a circunstancias como el número y calidad de las páginas de la Web desde las que se puede acceder a ellos-. Por su parte, las wikis [Leuf 2001] son entornos de trabajo colaborativo en los que un grupo de usuarios puede participar en la elaboración y mantenimiento de artículos vivos que sintetizan el conocimiento colectivo sobre determinados temas y que se estructuran en forma de enciclopedia para su manipulación. También de forma colaborativa las folksonomías [Golder 2006] permiten organizar el acceso a repositorios de recursos digitales -como fotos, vídeos o textos- atendiendo a los criterios de clasificación que los propios usuarios van estableciendo sin considerar ningún esquema preestablecido. Existen otros servicios que se basan en la participación -involuntaria a veces- de los usuarios para organizar y facilitar el acceso a la información y el conocimiento disponible en la Web. Este es el caso de las recomendaciones de productos en sitios Web de comercio electrónico, como la de la librería en línea Amazon [Amazon 2008] donde se proponen libros relacionados con uno seleccionado en función de la coincidencia de ellos en las compras de los usuarios. Otro caso que goza de gran popularidad en el momento actual es el de los Weblogs [Weblog 2008], que proporcionan soporte para la creación y publicación personal o colectiva de contenidos en la red, y que frecuentemente sirven para mostrar selecciones subjetivas de elementos accesibles en la Web y establecer redes de recursos destacados en ella para los seguidores de cada Weblog.

### *KnowCat*

Relacionado con todos estos sistemas está el sistema de gestión de conocimiento en red KnowCat [Alamán 1999][Cobos 2003], que es un entorno basado en Web que permite la gestión del conocimiento gracias a la interacción de una comunidad de usuarios sin necesidad de supervisión humana. KnowCat proporciona un entorno de trabajo colaborativo para trabajar sobre un repositorio de conocimiento constituido por documentos que se organizan en una estructura de temas al que se llama árbol de conocimiento. El aspecto más característico del sistema es la capacidad para seleccionar el conocimiento que maneja atendiendo a su calidad mediante un proceso que se denomina cristalización. Este mecanismo emplea las opiniones explícitas emitidas por los usuarios sobre los documentos y los datos recogidos de su actividad con ellos para determinar la calidad de los mismos. La capacidad de cristalizar el conocimiento es la diferencia más significativa de este sistema respecto a otros que proporcionan entornos de elaboración colectiva de repositorios de documentos como las wikis [Leuf 2001]. El planteamiento de KnowCat tiene mucho que ver con las proposiciones enunciadas en la hipótesis del apartado anterior, y proporciona una base ideal para poner en práctica tales

supuestos y comprobar su validez. Por ello en este trabajo se parte de la aproximación a la gestión del conocimiento de dicho sistema y se emplean las ideas y técnicas de las áreas de investigación mencionadas con tal fin.

### **1.3 Propuesta**

Como ya se adelanto al principio del documento, la hipótesis de partida de esta tesis propone que podrían encontrarse vías de solución para el problema de la sobrecarga de información en los sistemas de gestión de conocimiento en red, empleando la energía residual de la actividad de personas, servicios y otras entidades que interaccionan con ellos, aprovechando las características de los elementos implicados en la gestión -comunidades de usuarios, conocimiento y red-, y aplicando ideas y técnicas de campos de investigación que se integran en la Web Semántica.

Con el objetivo final de comprobar la validez de la hipótesis propuesta se sigue el siguiente proceso. Para empezar, se revisan los campos de investigación mencionados en los apartados anteriores. A continuación, se diseña un entorno basado en los supuestos de la hipótesis. Más tarde, se realiza un prototipo operativo del entorno que permite realizar experimentos para probar los supuestos fundamentales. Después, se establecen las experiencias y pruebas de campo necesarias para comprobar los aspectos esenciales de la hipótesis. Finalmente, se realizaran las prácticas y los ensayos señalados y se estudiarán los resultados obtenidos para establecer conclusiones y líneas futuras de investigación.

El sistema propuesto se denomina Semantic KnowCat (SKC), se trata de un sistema de trabajo en grupo que permite la gestión de una base de conocimiento y la selección de los documentos que integran la misma por su calidad, mediante la interacción vía Web de una comunidad de usuarios. Para conseguir esto sin supervisión, el sistema emplea información sobre la actividad de los usuarios, los documentos que constituyen el repositorio de conocimiento y las opiniones de dichos usuarios por tales documentos. La arquitectura de SKC se construye alrededor de ontologías [Mahesh 1996], que se refieren tanto al conocimiento que el sistema gestiona como a las entidades que intervienen en su proceso. En el repositorio los documentos se organizan por temas que forman grafos de conocimiento. Cada instancia del sistema es un nodo SKC que se dedica a una determinada materia y tiene una comunidad de usuarios y unos grafos de conocimiento propios.

SKC sigue la filosofía de gestión del conocimiento de KnowCat (KC) [Alamán 1999][Cobos 2003], sistema que se presentó brevemente en el apartado anterior y del que se hablará más en detalle en el siguiente capítulo. Sin embargo, SKC procura conseguir la gestión de forma diferente: reduciendo la necesidad de que los usuarios manifiesten explícitamente su opinión sobre el conocimiento; aumentando el aprovechamiento de la información generada por el

sistema, los usuarios y otras entidades con las que aquél se relacione; y ampliando la explotación de las características de los elementos implicados en la actividad, como la comunidad, el conocimiento y la Web. Para ello SKC integra ideas y técnicas de los campos de conocimiento mencionados en el apartado anterior y que también se tratarán en profundidad en el capítulo siguiente.

Aunque los diseños de los sistemas SKC y KC son muy diferentes, como se verá después, esto no ha impedido que KC haya servido como plataforma de prototipado para SKC. De este modo se ha podido disponer de un modelo de SKC dotado de la funcionalidad necesaria para realizar varios experimentos destinados a probar los aspectos fundamentales de la hipótesis propuesta. En todos los casos se ha utilizado el sistema KC como núcleo del modelo operativo, al que se han añadido los módulos necesarios para dotarle de nuevas funcionalidades, respetando el funcionamiento básico del sistema anfitrión. En concreto se han prototipado tres módulos de SKC sobre KC: el primero es el Módulo de Cliente, que se encarga de registrar la interacción de cada usuario con el sistema; el segundo es el Módulo de Análisis de Conocimiento, que se ocupa de procesar el conocimiento y los datos recogidos por el entorno, para obtener información oculta y valiosa para enriquecer el conocimiento del repositorio o para facilitar su gestión; y el tercero es el Módulo de Red, que permite establecer vínculos automáticos entre distintas instancias del sistema repartidas por la Web, con el fin de implantar y mantener una red de conocimiento entre todas ellas. Además, se ha dotado al sistema de la infraestructura de base de datos necesaria para soportar las nuevas funcionalidades. Esta estrategia de prototipado ha permitido emplear el modelo de SKC en experimentos realizados en el marco de programas de actividades de KC y utilizar repositorios de conocimiento del KC original en ensayos con el prototipo.

Las experiencias realizadas con el prototipo de SKC se han diseñado para poner a prueba varios aspectos de la hipótesis propuesta relacionados con las funcionalidades de los prototipos de módulos implementados. El primer bloque de experimentos se ha dirigido a estudiar el aprovechamiento residual de la actividad de los usuarios con el sistema, atendiendo a lo que los usuarios demuestran de forma involuntaria cuando realizan operaciones conscientes con el entorno. La información obtenida de este modo, con requerimientos mínimos de atención por parte de los implicados, puede servir para mejorar la gestión del conocimiento. El segundo bloque de experiencias se ha orientado a analizar la explotación del conocimiento oculto en el repositorio -que no se aprecia a simple vista-, para enriquecer el evidente y mejorar con él la gestión realizada por el sistema. El tercer bloque de experimentos se ha encaminado a explorar la representación del interés del usuario, como resultado del análisis de su interacción con el conocimiento, y su posible aprovechamiento en el sistema. Por último, el cuarto bloque de



experimentos se ha dirigido a la indagación en la vinculación automática y autónoma entre nodos del sistema dispersos por la Web, para formar una red de nodos de conocimiento.

### **1.4 Objetivos de la investigación**

Como ya se ha anticipado, el objetivo fundamental de la tesis propuesta es probar que se pueden encontrar vías de solución para resolver el problema de la sobrecarga de información en el ámbito de los sistemas de gestión de conocimiento en red, entorno representativo de otros más generales donde se produce el mencionado problema, sobre la base de:

- Obtener beneficio de la energía residual de la actividad de personas, servicios y otras entidades que interaccionan con ellos.
- Aprovechar las características de los elementos implicados, comunidades de usuarios, conocimiento y red.

Estos objetivos generales se persiguen en este trabajo mediante los siguientes objetivos operativos experimentales que responden a casos de especial interés:

- Seleccionar el conocimiento por su calidad, minimizando los requerimientos de manifestaciones explícitas de opinión sobre el mismo por parte de los usuarios. Con ello se pretende reducir la necesidad de que los usuarios tengan que dedicar atención al acto de opinar sobre los ítems de conocimiento para conseguir seleccionarlos.
- Revelar conocimiento implícito del repositorio del sistema para enriquecer el explícito y mejorar con ello la gestión realizada. Se trata de analizar los ítems de conocimiento para descubrir información relevante para la actividad del sistema que no era evidente.
- Obtener conocimiento de la actividad de los usuarios con el entorno para incorporarlo en el repositorio y aprovecharlo en la gestión llevada a cabo por el sistema. Se trata de analizar la actividad de los usuarios para revelar información no evidente y mejorar la gestión efectuada.
- Establecer redes de conocimiento, integrando varios nodos de sistemas de gestión de conocimiento a través de la Web de forma automática y autónoma. Con ello se trata de abrir los nodos del sistema entre ellos.

### **1.5 Principales contribuciones**

Como se verá en el capítulo de Conclusiones, además de procurar demostrar la hipótesis propuesta, dirigida a identificar vías de solución para resolver el problema de la sobrecarga de información en el ámbito de los sistemas de gestión de conocimiento en red, con los resultados

de este trabajo se espera aportar varias contribuciones al ámbito del mismo y de otros semejantes, entre ellas pueden mencionarse cinco:

- La primera, un mecanismo de monitorización de interacción de los usuarios en el cliente y de registro de actividad de los mismos en el servidor para aplicaciones Web de gestión de conocimiento -y de acceso a contenidos, en general-, así como un proceso para determinar los elementos de conocimiento -contenido- objetos de la actividad observada.
- La segunda, una medida de la intensidad de interacción de la comunidad de usuarios en aplicaciones del tipo mencionado, a partir de la monitorización de los eventos de la interfaz del sistema y del registro de actividad del mismo.
- La tercera aportación, una aproximación a la selección automática y desasistida de conocimiento -contenidos en general- por su calidad, basada en el índice interés de la comunidad de usuarios por los ítems que lo constituyen en sistemas de las características indicadas.
- La cuarta, una representación del interés de los usuarios en las instancias de tales sistemas mediante unos vectores de interés de usuario, así como la proyección de dicho interés dentro y fuera de las mencionadas instancias mediante otros vectores descriptivos del mismo.
- La quinta contribución aportada sería la propuesta de convertir las instancias de tales sistemas en entidades artificiales activas en la Web y establecer ámbitos en el que dichas entidades colaboren con los usuarios humanos de forma activa.

Por otro lado, tal como se indicará en el capítulo de Aproximación Conceptual y luego en las Conclusiones:

- La sobrecarga de información se ha tratado desde distintas áreas de investigación de modo más o menos independiente. Sin embargo, en este trabajo se pretende integrar ideas y técnicas de un número considerable de áreas de conocimiento, este planteamiento multidisciplinar explícito supone una tendencia innovadora que se manifiesta en iniciativas distintas como las Web Semántica o las Web 2.0.
- La visión de la Web como un gran sistema de gestión de conocimiento reafirma la función como repositorio de información de la red, pero además reclama el papel esencial de la comunidad de usuarios en ella, que forma también una red activa, y dota de un propósito fundamental al conjunto, que es colaborar para sacar provecho de la información y el conocimiento. Este planteamiento proporciona nuevas oportunidades

de afrontar algunos problemas de la Web y de obtener beneficio de ella, lo que resulta también innovador.

- La aproximación propuesta pretende constituir un marco experimental para las áreas de investigación que involucra, en especial para los campos de gestión de conocimiento, de trabajo colaborativo, y de la Web. El sistema que se presentará establece un entorno con unas características que pueden resultar interesantes para investigar en dichas áreas, puesto que aporta un medio acotado de trabajo en grupo sobre la Web con una utilidad evidente, en el que participan usuarios humanos y automáticos con objetivos claros y concurrentes.
- La Web del futuro incluirá las aportaciones de la Web 2.0, de la Web Semántica y otras contribuciones que surjan con el tiempo. La aproximación planteada en este trabajo espera participar en este proceso, señalando un espacio de confluencia de ideas y tecnologías diversas para la gestión del conocimiento, y que resulten aplicables para el tratamiento del problema de la sobrecarga de información en ámbitos más generales.

## **1.6 Terminología**

Existe una serie de términos que se utilizan repetidamente a lo largo de este trabajo y requieren una aclaración previa:

- Estructura de conocimiento, es el árbol o el grafo de temas donde se organizan los documentos que constituyen el conocimiento que se gestiona en los sistemas KnowCat (KC) y Semantic KnowCat (SKC).
- Nodos de los sistemas KC y SKC, son instancia de los respectivos sistemas de gestión de conocimiento que se dedican a un determinado tema, incluyen un repositorio de conocimiento, una estructura que lo organiza y una comunidad virtual de usuarios.
- Cristalización de conocimiento, en el ámbito de KC y SKC, se refiere al proceso por el que se selecciona el conocimiento atendiendo a su calidad gracias a la opinión manifestada por la comunidad de usuarios hacia él.
- Conocimiento implícito, se refiere de forma figurada al conocimiento que, estando en el sistema, no es evidente.
- Ítem de conocimiento, es cada uno de los elementos básicos de los que se considera compuesto el conocimiento manejado por el sistema.
- Enriquecimiento del conocimiento, es el proceso por el que se mejora el conocimiento existente añadiendo información a los ítems de conocimiento.

- Energía residual de la actividad, de forma retórica, nos referimos con este concepto al esfuerzo empleado para realizar una tarea que puede aprovecharse para llevar a cabo otras.
- Monitorización de actividad, es el proceso por el que se registra la actividad de un usuario o sistema.

## **1.7 Plan de tesis**

Esta memoria comienza con unos apartados de preámbulo seguidos de cinco capítulos, de los que el primero es esta Introducción.

En el segundo capítulo, se revisa el Estado de la Cuestión, donde se profundiza en el planteamiento del problema esbozado en el capítulo anterior, se revisa el estado del arte de los campos de conocimientos considerados para el trabajo, se habla en detalle del sistema KnowCat, y se presentan algunas referencias interesantes de trabajos relacionados con el proyecto.

En el tercer capítulo, se presenta la Aproximación Teórica al tema de la tesis, en la que se especifica la Propuesta planteada, detallando los objetivos perseguidos, determinando sus fundamentos en el estado del arte, estableciendo el proceso hasta alcanzar las metas marcadas y previendo la trascendencia de los resultados esperados. Además, en este capítulo se presenta la especificación del sistema Semantic KnowCat (SKC) que se utiliza en forma de prototipo para la realización de los experimentos del proyecto.

En el cuarto capítulo, se expone la Aproximación Experimental realizada, en la que se presentan los detalles del Prototipo de SKC sobre KC realizado, el Proyecto de Innovación Docente de la universidad que sirve de marco para la realización las experiencias, así como los Experimentos Realizados y los Resultados Obtenidos.

En el último capítulo, se recogen unas Conclusiones finales que sirven de recapitulación de los puntos más destacados del proyecto y se presentan los Trabajos Futuros que resumen las líneas de investigación abiertas y las nuevas propuestas.

Por último, se enumeran las Referencias a las que se hace alusión a lo largo del trabajo y se incluye una sección de Anexos para documentación relacionada con el proyecto.

## **2 Estado de la Cuestión**

En este capítulo se describe con más detalle el problema planteado en la introducción del documento. Además, se revisa el estado del arte de los campos de conocimiento también presentado anteriormente, mostrando las líneas de investigación más destacados relacionadas con la investigación que nos ocupa y las referencias de algunos trabajos de especial interés para la misma. Por último se incluyen unas conclusiones, recapitulando las ideas y referencias más importantes presentadas, poniéndolas en relación con la tesis propuesta.

## **2.1 Planteamiento del Problema**

Como se afirmó en la introducción de este documento, uno de los efectos indeseables de la extensión del uso de las tecnologías de la información y las comunicaciones (TIC) en los últimos años ha sido la sobrecarga de información.

A la extensión de las TIC han contribuido varios factores que se han estimulado mutuamente de forma positiva en los últimos veinticinco años, cada vez con más intensidad, entre ellos han tenido especial importancia el progreso tecnológico digital, la popularización del empleo de dispositivos basados en esta tecnología, así como el desarrollo y la generalización de las redes de comunicación.

El avance de la tecnología digital ha propiciado el abaratamiento de los dispositivos que soportan las TIC y la diversificación de sus aplicaciones. Por un lado, las ventajas del empleo de la tecnología digital en la mayoría de los campos, ha estimulado la oferta de soluciones basadas en ella para los sectores profesionales implicados. Por otro lado, la presión comercial, mediática, e institucional, la evidencia de los beneficios y la creciente oferta de soluciones para aspectos diversos de cualquier actividad, han provocado una predisposición positiva hacia dicha tecnología, que ha propiciado el aumento de la demanda, la competencia y el abaratamiento de los productos. Además, la disponibilidad de dispositivos digitales asequibles ha facilitado la total adopción de los mismos en el entorno profesional y su posterior irrupción masiva en el mercado de consumo.

La popularización del empleo de dispositivos digitales en todos los ámbitos ha supuesto un aumento de la oferta y la demanda de información en formato digital, puesto que los dispositivos de uso generalizado abarcan todo el ciclo de vida del medio digital: captación de información, como las cámaras digital de fotografía y de vídeo, los dispositivos de digitalización de audio o los sistemas de obtención de texto a partir de soportes diversos - documentos impresos o voz-; reproducción de soportes digitales, como los reproductores de MP3 o MP4 o los sintetizadores de sonido y de voz; almacenamiento de material digital, como los DVDs, los discos duros o las memorias flash; procesamiento de material digital, como los ordenadores personales y portátiles, las PDAs o los navegadores GPS; y comunicación de dicho material, como los dispositivos de conexión y soporte de redes, los dispositivos de emisión y recepción de TV digital o los terminales de telefonía móvil, que son actualmente dispositivos multifunción capaces de intervenir en todo el ciclo digital. Una lista parcial de cantidad mundial de dispositivos y abonados generadores de información digital en 2006 puede verse en la Tabla 2-1 [IDC 2007].

El desarrollo de las redes de comunicación, sobre todo de Internet, y de los servicios y aplicaciones basados en ellas, especialmente de la Web y las aplicaciones P2P, ha permitido la publicación de la información digital disponible y el acceso a la misma para su intercambio y explotación. Esta oportunidad ha supuesto un estímulo fundamental para la proliferación de ese tipo de información, pero también ha sido esencial la generalización del uso de las redes de comunicación en muchos ámbitos de la actividad humana -los negocios, las instituciones, el trabajo, la educación y el hogar-. El número estimado de usuarios de Internet en 2006 era de 1.100 millones, más de la mitad con banda ancha, y en 2010 se prevé que será de 1.600 millones aproximadamente, de los que la mayoría, unos 1.300 millones, dispondrán de banda ancha para esas fechas [IDC 2007].

*Tabla 2-1. Cantidades de algunos dispositivos digitales y de abonados de telefonía móvil en 2006*

<b>Categoría</b>	<b>Millones en 2006</b>
Cámaras Digitales	400
Teléfonos con Cámara	600
Ordenadores Personales	900
Reproductores de Audio	550
Abonados de Telefonía Móvil	1600

El uso generalizado de las redes se ha hecho evidente de varias formas: con la extensión del empleo de servicios genéricos, como el de acceso, el correo electrónico, la Web o los sistemas de comunicación personal -mensajería instantánea, telefonía...-; con la multiplicación del uso de aplicaciones para explotar y compartir información y conocimiento, como los buscadores -Google [Google 2008] o Yahoo [Yahoo 2008], por ejemplo-, las aplicaciones colaborativas de respaldar e intercambiar recursos -eMule [eMule 2008], Flickr [Flickr 2008] y YouTube [YouTube 2008], entre otros- y de publicación sencilla de contenidos -como los weblog [Weblog 2008] o las wikis [Leuf 2001]- o de difusión -radio y TV -; o con la expansión de la utilización de servicios para obtener diversas prestaciones a través de la red, como servicios informativos, bancarias, administrativos o comerciales.

Además, la evolución del modo de uso de la red -la democratización del medio, ligada al éxito de las aplicaciones sociales características de la Web 2.0- ha acelerado el aumento desmesurado de proveedores de información, que han pasado de decenas de miles a cientos de millones en menos de veinte años, la Tabla 2-2 ilustra este proceso [Baeza 2006].

Como resultado del éxito de la expansión del uso de las TIC se ha provocado un aumento colosal de la cantidad de información disponible, tanto accesible a través de las redes de comunicación como en repositorios fuera de ellas. Dicho crecimiento no sólo se debe al

aumento imparable del número total de elementos de información existentes, sino también al tamaño cada vez mayor de los mismos -especialmente del material audiovisual-. [IDC 2007] En 2006 el volumen estimado de información existente en el universo digital -información creada o capturada en formato digital y replicada- era de 161 EB (161\*10<sup>9</sup> GB) y en 2010 se estima que será 6 veces mayor, unos 988 EB, una cuarta parte de esas cifras corresponde a registros visuales.

Tabla 2-2. Evolución del uso de la red propuesta por Baeza [Baeza 2006]

Fuentes	Ingenieros y Científicos	Webmasters		Comunidades
Tecnologías	Alpha Chip Surfers	Algoritmos Basados en Enlaces Escala de la Infraestructura		Tecnologías Facilitadoras Incentivos Correctos Masa Crítica
Magnitudes	Miles de Grupos de Noticias	Millones de Sitios Web	Decenas de miles de millones de Documentos Web	Millones de millones de Artefactos de Conocimiento
Productos	Usenet/Newsgroups	Directorio Yahoo/Altavista	Google/Inktomi	Flickr/Y!Answer

Al mismo tiempo, la facilidad para cualquiera de publicar información ha favorecido la diversidad de calidad de las aportaciones y la incertidumbre de los usuarios hacia ellas. Mientras que al principio de la Web, en la primera mitad de los noventa, eran decenas de miles de usuarios, en su mayoría expertos acreditados, los que publicaban información para unos pocos millones de usuarios, también mayoritariamente especializados; durante el desarrollo comercial de la red, en la segunda mitad de los noventa, eran algunos millones de usuarios, muchos especializados y con respaldos institucionales, los que publicaban información para cientos de millones de usuarios, sin conocimiento especiales la mayoría; y desde la popularización del uso de la red y del éxito de las aplicaciones sociales, a partir del cambio de siglo, cientos de millones de usuarios, muchos sin preparación especial, acreditación ni respaldo de ningún tipo y de forma voluntaria a menudo, publican información para más de mil millones de usuarios como ellos, que también son proveedores de información potenciales. De los 161 EB de información digital existente en 2006, el 75% fue creado por los consumidores, y en 2010 se espera que de los 988 EB de información digital existente para entonces, 692 EB sean generados por individuos, consumidores y trabajadores [IDC 2007]. La evolución de la información digital entorno al momento actual, según la misma fuente, se ilustra en la Figura 2-1.

La gran cantidad de información disponible en la red, la posible falta de criterio en la publicación, la virtual carencia de acreditación de los proveedores y la potencial ausencia de



critérios para la selección, componen un panorama en el que es muy difícil que los usuarios puedan discriminar la información accesible -distinguir la basura digital de la información valiosa-. Sin embargo, la dificultad de su explotación no disminuye el valor y el interés de toda la información existente, ni el origen incierto de parte de ella devalúa al resto, como tampoco la inexperiencia objetiva de los proveedores ni su falta de respaldo tienen que disminuir necesariamente la valía de sus aportaciones. Estas circunstancias no hacen sino agravar el problema y dificultar la solución.

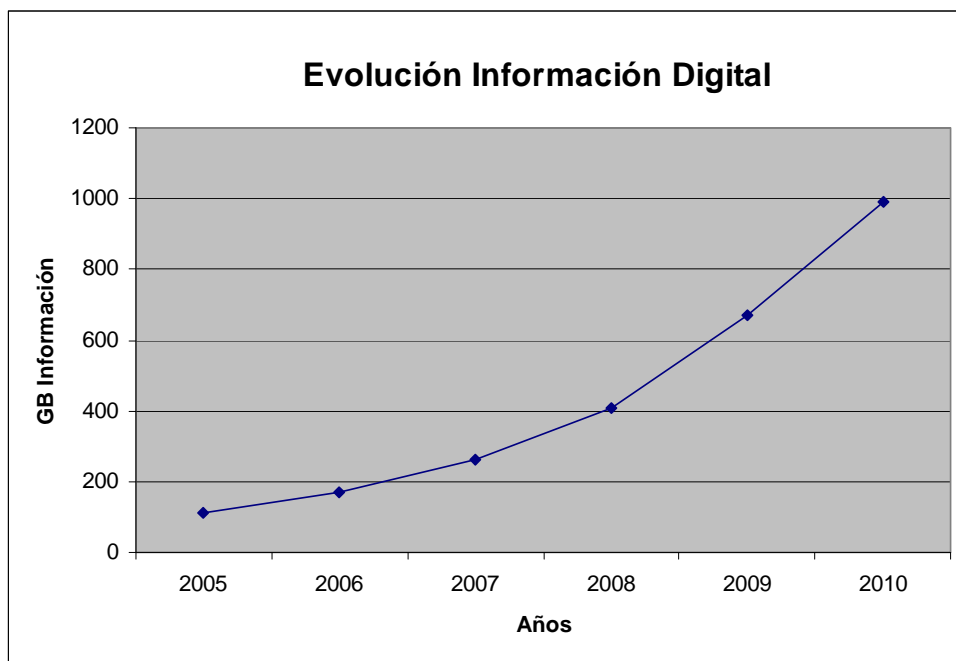


Figura 2-1. Evolución del volumen información digital de acuerdo con [IDC 2007]

La rapidez del incremento de la información digital y la complejidad del problema que plantea su aprovechamiento, no ha permitido el desarrollo paralelo de todas las tecnologías necesarias para su gestión y explotación, y ha provocado el efecto indeseable de sobrecarga de información. Este fenómeno se empezó a percibir de forma incipiente hace algo más de una década, con la popularización de Internet y de la Web, y se ha hecho más evidente y alarmante con el paso del tiempo, como resultado del proceso de extensión y asimilación de las TIC. El efecto de sobrecarga seguirá agravándose en el futuro, si los medios de gestión de información que se vayan proponiendo no se adaptan a los volúmenes de datos cada vez mayores y a su crecimiento acelerado, y si dichos medios no dan soluciones a las dificultades que el funcionamiento de la red provoca.

Si todo esto no fuera poco, el problema de sobrecarga se acentúa cuando la información a la que afecta resulta crítica para los grupos que la emplean. Aunque hay muchos casos en los que usuarios y organizaciones dependen de la información para alcanzar sus objetivos, el fenómeno

es especialmente grave cuando la información de la que se trata integra repositorios de conocimiento, de información que hace posible obtener y aprovechar otra información. Dicha gravedad se debe a que cuando el conocimiento se hace explícito en un repositorio, se suele convertir en un recurso imprescindible para los colectivos a los que se dirige, acaba siendo fundamental para el desarrollo de las organizaciones propietarias del mismo y, en algunas ocasiones, también esencial para el avance de los campos de conocimiento a los que se refiere.

Por otro lado, en el ámbito de la Web, la sobrecarga de información toma una dimensión global, puesto que la red es accesible por comunidades cada vez más amplias de usuarios y en ella hay espacio para cualquier área de interés y campo de conocimiento. De este modo, gran cantidad de información y conocimiento valiosos han empezado a estar al alcance de muchos colectivos que antes no podían acceder a ellos y que ahora tienen la oportunidad de aprovecharlos para su desarrollo. Este hecho estimula el interés de los beneficiarios y favorece la incorporación de nuevos grupos a la red, pero convierte a todos en dependientes de los recursos que se han puesto a su alcance.

Además, las TIC han transformado en poco tiempo el modo de trabajar con la información y el conocimiento, la forma de obtenerlos, almacenarlos, buscarlos, acceder a ellos y utilizarlos. Los sistemas basados en las TIC han sustituido en un periodo muy corto de tiempo a los sistemas que existían previamente, de tal manera que sería muy difícil prescindir ahora de ellos y volver a la situación anterior, después de haber abandonado, e incluso desmantelado, los sistemas alternativos preexistentes. Al mismo tiempo, los nuevos hábitos de consumo informativo han provocado un aumento de los requerimientos de información y exigencia hacia ella, que no se podrían atender sin el empleo de las propias TIC, pero tampoco simplemente con la oferta desmesurada de información en bruto sin seleccionar ni elaborar, que actualmente anega la red y otros sistemas fuera de ella.

Todo esto hace imprescindible encontrar soluciones para la sobrecarga de información, por supuesto sin renunciar a los recursos generados, aplicando para ello las propias TIC y tecnologías vinculadas con ellas, y aprovechando el interés y la dinámica de los grupos que participan en la red y en otros sistemas de creación e intercambio de información y de conocimiento.

## ***2.2 Campos de Investigación Relacionados***

En esta sección se revisan los campos de investigación relacionados con el tema de esta tesis. Cada campo se presenta en un apartado que incluye un resumen de sus fundamentos y de los aspectos vinculados con el proyecto que nos ocupa. Se tratan cuatro áreas principales, que son Gestión del Conocimiento, Interacción Persona-Ordenador, Web Semántica y Minería de Información y de Datos. Los dos temas centrales se refieren a campos muy amplios, por lo que

los apartados correspondientes son más extensos que los otros, a pesar de concentrarse en las cuestiones más relevantes para investigación que se presenta aquí.

### **2.2.1 Gestión del Conocimiento**

Este apartado se dedica al área de investigación de gestión del conocimiento. En él se incorpora un examen de sus bases, así como de los problemas y beneficios de su aplicación. Además se incluye un bloque dedicado a los sistemas que llevan a la práctica este tipo de gestión, en el que se muestran las dificultades de la implantación de tales sistemas en las organizaciones y se presentan algunas clasificaciones de los mismos.

#### *Datos, Información y Conocimiento*

Según Coakes [Coakes 2003] la relación entre datos, información y conocimiento se acepta de forma general en la literatura de gestión del conocimiento. Esta relación se considera habitualmente como una jerarquía, en la que los abundantes datos son la base sobre la que reposa la menos copiosa información que, a su vez, sirve de soporte para el reducido conocimiento. Sin embargo, la distinción entre estos términos resulta con frecuencia arbitraria, puesto que se utilizan, a menudo, de forma intercambiable, lo que no ayuda a desentrañar la relación emergente que existe entre ellos. Tal confusión ha arraigado en la literatura especializada, donde diversas aplicaciones reivindican ser consideradas sistemas de gestión de conocimiento.

Para el mismo autor, la relación entre datos, información y conocimiento se establece del siguiente modo. Los datos son invariantes acerca de un objeto o fenómeno al que se puede prestar atención, tienen un significado potencial para quien pueda interpretarlos, y son un posible punto de partida para un proceso mental. Relacionado con los datos está el concepto de *capta*, que es el resultado de seleccionar ciertos datos como consecuencia de prestarles atención. La información surge como resultado de la atribución de significado a determinados datos a los que se atiende, *captas*. Tal asignación de significado es particular del individuo que la realiza y del objeto concreto de su atención, por ello es un concepto personal, variable y efímero. Por último, el conocimiento aparece como resultado de la consideración de la información como objeto de observación y fuente de datos, y de asignar un significado al *capta* resultante, con el objetivo de utilizarlo en futuras asignaciones de significado, nuevos *captas*. En este sentido, la información que constituye el conocimiento representa conceptos, métodos, comportamientos, valores y normas, que constituyen un filtro a través del que se perciben eventos y observaciones, y que permite asignar significado a los nuevos *captas* inducidos. Por lo tanto, el conocimiento se puede ver como una transformación del significado de cierta información o *capta*, que puede servir para asignar significados a nuevos *captas* en el futuro. El conocimiento puede ser personal o colectivo, pero siempre resultará más estable que la información.

Por lo tanto, el conocimiento es el resultado de la elaboración de la información que se genera en la organización -grupo humano- a partir de los datos que se pueden observar de forma directa. Así que se puede establecer un proceso por el que algunos datos se transforman en información que luego puede llegar a convertirse en conocimiento. Los datos habitualmente son fáciles de capturar, estructurar y transferir, y muy susceptibles de tratar con las Tecnologías de la Información y las Comunicaciones (TIC). Estas tecnologías pueden ayudar en el proceso de la información que se consigue elaborando algunos datos y dando sentido a los resultados obtenidos, pero casi siempre la necesaria interpretación requiere de la intervención humana. Finalmente, la aplicación de las TIC al manejo del conocimiento es mucho menos provechosa que a los datos y la información de los que proviene, puesto que el conocimiento es difícil de identificar, capturar, estructurar y transformar automáticamente, y resulta imprescindible la participación de personas en su proceso. De este modo el requerimiento de participación humana aumenta y el aprovechamiento de las TIC disminuye, a medida que se avanza en el proceso de transformación de los datos en conocimiento [Davenport 1999].

### *Gestión del Conocimiento*

En el momento actual parece que la mayoría de la comunidad interesada en este campo está de acuerdo con definir el conocimiento como el resultado del tratamiento de la información y la interacción de un grupo de personas interesadas en esa información [McDermott 1999], y pensar que su gestión implica la colección, organización, clasificación y distribución de ese conocimiento [Malthotra 2000]. De estas consideraciones se puede interpretar que el fin último de un sistema de gestión del conocimiento no puede ser sólo el almacenamiento de la información, sino también cubrir las necesidades de distinto índole de la comunidad que lo utiliza [Churman 1971]. En definitiva, los medios informáticos implicados en un sistema de gestión del conocimiento tienen que responder a dos requerimientos esenciales, facilitar el trabajo colaborativo de los usuarios que intervienen en la gestión y proporcionar un soporte sólido para la administración de la información que constituye la base del conocimiento que se gestiona [Cobos 2002b].

El proceso de gestionar el conocimiento involucra varias acciones [Benjamins 1999]: en primer lugar, recolección del conocimiento que se está gestionando; en segundo lugar, organización y estructuración del conocimiento, imponiéndole una estructura con el fin de gestionarlo eficientemente; además, refinamiento del conocimiento, corrigiéndolo, actualizándolo, eliminándolo, en definitiva manteniéndolo; y por último, distribución del conocimiento, llevando el conocimiento a quien lo necesita.

La gestión del conocimiento examina cómo las organizaciones pueden gestionar, almacenar, recuperar y aumentar sus propiedades intelectuales de modo eficiente. El término gestión del

conocimiento apunta a una cuestión importante en las organizaciones actuales, que dependen cada vez más de desplegar los activos no tangibles, como el conocimiento práctico de cómo hacer las cosas (know-how) y resolver problemas tácticos en periodos de tiempo más cortos. [Ackerman 2003]

Para Nonaka [Nonaka 1995] la gestión de conocimiento en la organización es su capacidad para obtener nuevo conocimiento, diseminarlo en su seno y aplicarlo en la creación y mejora de productos, servicios y sistemas. La gestión del conocimiento en las entidades es un proceso que debe identificar la información que permitiría mejorar los procesos más importantes para conseguir los resultados pretendidos, determinar dónde se encuentra ese conocimiento y hacer que este esté disponible donde y cuando se necesite y para quien lo requiera.

Según Alavi [Alavi 2001] [Gottschalk 2004], el propósito de la gestión del conocimiento es ayudar a las compañías -y a las organizaciones en general- a crear, compartir y usar el conocimiento de forma más útil. La gestión del conocimiento efectiva provoca menos errores, menos trabajo, más independencia en tiempo y espacio para los que trabajan con el conocimiento, menos preguntas, mejores decisiones, menos reinversiones, mejores relaciones con los clientes -o destinatarios de la actividad-, mejor servicio y mayor rentabilidad, en definitiva, aumenta la innovación y el provecho. El reciente interés por el conocimiento de las organizaciones ha vinculado la cuestión de su gestión con el de su beneficio.

Para el mismo autor, los beneficios de la introducción de la gestión del conocimiento en las organizaciones abarcan distintos aspectos: aprovechamiento de la experiencia, transferencia del conocimiento, obtención de nuevo conocimiento, mejora de la comunicación en el seno de la organización, aumento de la capacidad analítica, diferenciación estratégica, aumento de la competitividad, aprovechamiento de recursos, aumento de la productividad, reducción de costes, eficacia organizativa, fomento de la innovación y la creatividad o reconocimiento del trabajo del personal, entre otros muchos.

#### *Sistemas para la Gestión del Conocimiento*

La implantación de un sistema de gestión del conocimiento no es sencilla, puesto que además de la infraestructura que requiere, conlleva un cambio profundo en la organización, que afecta a los procedimientos de trabajo, las atribuciones del personal, la cultura de la entidad, etc. Por ello, la implantación de un sistema así no puede considerarse sólo como una cuestión tecnológica, no debe abordarse sin involucrar a toda la organización en el proceso, ni realizarse de forma brusca sino progresivamente.

En opinión de Ackerman [Ackerman 2003] la mayoría de las aproximaciones que ofrecen soporte informático para la gestión del conocimiento demuestran un entendimiento limitado del tema. En general, hay dos enfoques del soporte de la gestión del conocimiento mediante

software. La primera explota la idea de externalizar el conocimiento y recomienda poner más y más información en repositorios compartidos. Estas bases de datos de información, o memorias de la organización, tienen la ventaja de emplear técnicas informáticas estándares y transmiten la esperanza de una fácil reutilización de la tal información. Estas aproximaciones tradicionales suelen estar enfocadas a recolectar, proporcionar y filtrar el conocimiento explícito disponible. Sin embargo, la visión de repositorio para la gestión del conocimiento tiene algunas limitaciones importantes y no puede ser aplicada en todas las situaciones. La información almacenada en un repositorio puede utilizarse y transferirse con facilidad, pero la información descontextualizada suele ser difícil de emplear. Los usuarios necesitan, a menudo, encontrar a otras personas informadas o gente que puedan ayudarles a aplicar la información a la situación o al problema que les ocupa. De forma parecida, cuando el conocimiento es tácito, el acceso a personas con experiencia es a menudo indispensable. Si hay un problema complejo o nuevo que resolver, acceder a expertos suele ser preferible que recurrir a documentos estáticos.

Esta circunstancia ha llevado a la consideración de un segundo tipo de gestión del conocimiento, la cual Ackerman [Ackerman 2003] denomina “experiencia compartida” (expertise sharing). Los campos de recursos humanos y comportamiento de la organización han hecho alusión a la importancia del personal en la vida de la organización. Ackerman defiende la importancia de aumentar lo que él llama redes de conocimiento y está de acuerdo con otros autores en considerar la memoria en las actividades laborales como un recurso de valor incalculable en las organizaciones.

La “experiencia compartida”, atiende a los componentes humanos -los aspectos cognitivos, sociales, culturales y organizativos del trabajo del conocimiento- además del almacenamiento y recuperación de la información. En comparación con la aproximación tradicional, que se preocupa fundamentalmente del intercambio de conocimiento en la organización, el enfoque de Ackerman [Ackerman 2003] se dirige a las actividades organizadas por los propios miembros de las organizaciones. Para hacer posible la puesta en común, las organizaciones tratan de conectar la gente entre si con el fin de reforzar la comunicación, el aprendizaje y el conocimiento de la propia organización. Ackerman cree que debe establecerse y cultivarse este segundo tipo de gestión del conocimiento.

Earl [Earl 2001] [Gottschalk 2004] desarrolló una taxonomía para la gestión del conocimiento estableciendo escuelas que representaban una perspectiva particular respecto al tema o una orientación propia de aproximarse a él. Cada escuela se propuso como tipos ideales que no eran mutuamente excluyentes. Según esta clasificación se pueden establecer tres escuelas principales: la económica, preocupada sobre todo por los ingresos y de la explotación de los activos de conocimiento; la organizativa, centrada en la red y cuyo propósito es reunir el

conocimiento; y la estratégica, preocupada principalmente por la ventaja competitiva y con los objetivos de identificar, explotar y sacar beneficio de las posibilidades del conocimiento.

La escuela económica, por tanto, se preocupa de proteger y explotar los activos intelectuales y el conocimiento para producir un flujo constante de ingresos. Se preocupa de gestionar el conocimiento como un activo, que incluye patentes, marcas comerciales, derechos y el conocimiento práctico de cómo hacer las cosas. Esta escuela está más preocupada de la explotación del conocimiento que de su exploración. Un factor crítico de éxito para esta escuela parece ser el desarrollo de un grupo especializado o función que gestione la propiedad del conocimiento de forma agresiva a través de la contabilidad del capital intelectual, de la gestión del mismo, y de la creación de mercados de conocimiento efectivos y eficientes.

Por su parte, la escuela organizativa describe el uso de estructuras organizativas o redes para compartir y reunir el conocimiento. A menudo llamadas comunidades de conocimiento, que de forma ideal son grupos de personas con intereses comunes, con los mismos problemas o experiencias. Estas comunidades se diseñan y mantienen por un propósito determinado, de negocio en el contexto empresarial, y pueden ser internas a la organización o de carácter interorganizativo.

Por último, la escuela estratégica considera la gestión del conocimiento como una dimensión de la estrategia competitiva o como la esencia de la misma. La perspectiva de gestión de la organización determina la aproximación a la gestión del conocimiento en esta escuela, la perspectiva pueden estar basada en: información, preocupada del acceso a la información disponible en la organización; tecnología, preocupada de las aplicaciones de tecnologías para la gestión del conocimiento; y cultura, preocupada de la puesta en común del conocimiento.

Desde otro punto de vista Benjamins [Benjamins 1999] propone una clasificación de los sistemas de gestión del conocimiento según dos dimensiones. La primera de ellas diferencia entre sistemas verticales frente a horizontales. Mientras los primeros se desarrollan para situaciones particulares de la actividad de la organización y tienen poca aplicación en otros contextos, los segundos se desarrollan de forma más general y pueden aplicarse en diversas situaciones. La segunda dimensión de la clasificación de Benjamins distingue entre sistemas centrados en procesos frente a centrados en productos. Los primeros consideran la gestión del conocimiento como un proceso de comunicación, que se puede mejorar con el soporte para trabajo en grupo. Parte de la consideración de los usuarios como la fuente de conocimiento más importante de la organización, entre los que se pueden resolver cualquier problema alcanzando un compromiso. Los segundos, se centran en la gestión del conocimiento contenido en los documentos, su creación, almacenamiento y reutilización en el sistema de información que soporta la memoria de la organización. Parte de la consideración del conocimiento como un

recurso tangible o en la idea de soportar el desarrollo de conocimiento individual de los usuarios y su uso presentando la información adecuada en el momento preciso.

Como ya se ha dicho antes, las herramientas y sistemas de gestión del conocimiento suelen responder a dos requerimientos, por un lado facilitar el trabajo colaborativo de los usuarios que los utilizan y, por otro, proporcionar el soporte adecuado para la administración de la base del conocimiento a la que se aplican. Sin embargo, no todos los recursos de gestión del conocimiento cubren ambos requerimientos del mismo modo. Entre ellos, unos se concentran en facilitar el trabajo colaborativo, otros en el establecimiento de estructuras de conocimiento robustas, y el resto procuran atender a ambas cuestiones con parecido interés. Cobos [Cobos 2002b] propone una clasificación de los recursos de gestión del conocimiento atendiendo a estos dos aspectos, con la que clasifica algunos sistemas de gestión del conocimiento:

Tabla 2-3. Recursos para la gestión de conocimiento según Cobos en [Cobos 2002b]

Recursos Gestión Conocimiento 2002		Soporte Gestión Estructura Conocimiento	
		Sí	No
Soporte Gestión Colaboración	Sí	Sistemas Integrados de Gestión del conocimiento:	Espacio Compartido:
		<ul style="list-style-type: none"> <li>• Dynasites</li> <li>• Herramienta KnowNet</li> <li>• KnowCat</li> <li>• Meta4 KnowNet©</li> <li>• Microsoft® SharePointTM</li> <li>• Plumtree Corporate Portal</li> <li>• Portal Server 2001</li> <li>• Sintagma</li> <li>• Zaplet Appmail Suite</li> </ul>	<ul style="list-style-type: none"> <li>• BSCW</li> <li>• Groove</li> </ul>
			Sistemas de Recomendación:
		<ul style="list-style-type: none"> <li>• NewKnow</li> <li>• Jasper II</li> <li>• Coins</li> </ul>	
	No		Aprendizaje Colaborativo:
		Mediadores de Información:	<ul style="list-style-type: none"> <li>• WISE</li> <li>• CSILE</li> <li>• GENTLE</li> <li>• DEGREE</li> </ul>
		Sistemas Basados en Ontologías:	
		<ul style="list-style-type: none"> <li>• C-Web</li> <li>• IBROW</li> <li>• Ontobroker</li> <li>• WebCadet</li> <li>• Planet-Onto</li> </ul>	
		Librerías Digitales:	
		<ul style="list-style-type: none"> <li>• COSPEX</li> </ul>	

En esta clasificación, la mayoría de las subcategorías incluidas necesitan una pequeña descripción. Los sistemas integrados de gestión de conocimiento son herramientas que combinan el conocimiento en un espacio común en forma de un repositorio que recoge la



memoria colectiva, incluyendo soporte para gestionar la colaboración y el conocimiento. Las herramientas que proporcionan espacios compartidos permiten interactuar a los usuarios un entorno virtual para compartir el conocimiento. Los sistemas de recomendación facilitan la selección de la información aprovechando la interacción de otros usuarios con esa misma información. Los sistemas de aprendizaje colaborativo facilitan el estudio de un grupo de personas que comparten su conocimiento para obtener nuevo conocimiento. Los sistemas mediadores de información proporcionan una interfaz para que los usuarios puedan realizar consultas en fuentes dispersas y, a veces, heterogéneas de conocimiento, obteniendo resultados de forma integrada y uniforme. Los sistemas basados en ontologías emplean una representación compartida del conocimiento para facilitar el trabajo sobre el mismo. Las librerías digitales son sistemas integran tecnologías de la información y comunicación para emular y ampliar el servicio de un biblioteca convencional.

### **2.2.2 Interacción Persona-Ordenador**

En este apartado se trata el campo de investigación de Interacción Persona-Ordenador, que abarca temas muy diversos relacionadas con esta tesis, entre los que están el área de conocimiento del CSCW -Computer\_Supported Cooperative Work-, los sistemas de trabajo colaborativo, el tratamiento de la conciencia en los grupos -awareness-, algunas técnicas de monitorización de la actividad de los usuarios, la hipermedia adaptativa, los sistemas de recomendación y las técnicas de visualización de la información. Además, se habla de otras áreas vinculadas con el campo de Interacción Persona-Ordenador, también importantes para este trabajo, como el procesamiento del lenguaje natural y como el fenómeno de la llamada Web 2.0 y las tecnologías que se emplean en ella.

Según el grupo de interés sobre Interacción Persona-Ordenador [SIGCHI 2008] -Special Interest Group on Computer-Human Interaction, SIGCHI- de ACM -Association for Computer Machinery- [ACM 2008], la disciplina de Interacción Persona-Ordenador (IPO) se ocupa del diseño, implantación y evaluación de sistemas informáticos interactivos, destinados al empleo del ser humano, y del estudio de los principales fenómenos relacionados con ellos. [Lorés 2001] Los objetivos más importantes de la IPO son desarrollar o mejorar la seguridad, utilidad, efectividad, eficiencia y usabilidad de los sistemas que incorporan ordenadores [Diaper 1989], considerando todos los elementos involucrados en ellos, que son los equipos y los programas, pero también las personas y el entorno de los que forman parte.

Para implementar sistemas interactivos es imprescindible entender los aspectos psicológicos, ergonómicos, organizativos y sociales, que establecen el modo en que las personas trabajan y emplean los ordenadores. Este entendimiento es fundamental para desarrollar los instrumentos y las técnicas necesarias para el diseño de sistemas informáticos adecuados para cada aplicación

concreta. Con ello se pretende conseguir una interacción eficiente, efectiva y segura, para los usuarios individuales o integrados en grupos. [Preece 1994]. La IPO constituye un punto de encuentro entre diversas disciplinas, entre las que están gráficos por ordenador, sistemas operativos, factores humanos, factores ergonómicos, ingeniería industrial, psicología cognitiva e ingeniería informática. [Lorés 2001]

### Trabajo Colaborativo

De acuerdo con Brinck [ForakerDesign 2005], el groupware es una tecnología diseñada para facilitar el trabajo en grupo. Esta tecnología puede utilizarse para comunicar, cooperar, coordinar, resolver problemas competir o negociar. A pesar de que tecnologías tradicionales como el teléfono se podrían calificar de groupware, el término se usa normalmente para referirse a tecnologías específicas dependientes de las redes de ordenadores modernas, como el correo electrónico, los foros de noticias, los videoteléfonos o el chat.

Según el mismo autor, el CSCW -Computer\_Supported Cooperative Work- designa el campo de estudio que examina el diseño, adopción y uso del groupware. A pesar del nombre, el campo de estudio no está restringido a los temas de la “cooperación” y el “trabajo”, sino que también se ocupa de cuestiones como la “competición”, la “socialización” y el “juego”. El campo típicamente atrae a personas interesadas en cuestiones como el diseño de software y el comportamiento social y organizacional, incluyendo personas de la empresa, informáticos, psicólogos de organización, investigadores de comunicaciones y antropólogos, entre otras especialidades.

Para Grudin [Grudin 1994a] en el universo del software, el groupware está situado entre los sistemas que soportan las organizaciones y las aplicaciones de usuario. Los primeros se comenzaron a desarrollar a mediados de la década de los sesenta y alcanzaron prominencia antes porque el coste de los primeros ordenadores impidió que se aplicaran a objetivos distintos de los principales de las organizaciones. Las segundas no surgen hasta comienzo de los ochenta con la aparición de los ordenadores personales y su rápida popularización. Después de esto, todavía fueron necesarios algunos años para que se acuñaran los términos Groupware y CSCW a mediados de los ochenta, cuando se dieron las condiciones necesarias para ello: el abaratamiento de los ordenadores personales, el desarrollo de la infraestructura de comunicaciones necesaria, la asimilación del empleo de los ordenadores por los usuarios y la madurez de las aplicaciones de uso individual.

Las tecnologías de groupware, según Johansen [Johansen 1988], se pueden clasificar de forma general considerando dos dimensiones. La primera atendiendo a si los usuarios de los grupos están trabajando juntos al mismo tiempo, modo “síncrono”, o lo hacen sin coincidir en el

tiempo, modo “asíncrono”. La segunda considerando si los usuarios están trabajando en grupo en el mismo lugar, “cara a cara”, o en diferentes lugares, “a distancia”.

Las aplicaciones existentes no siempre se ajustan exclusivamente a un valor de estas dimensiones, como se pone de manifiesto en la siguiente tabla. Por ello se han propuesto taxonomías que utilizan criterios distintos de clasificación [Grudin 1994b, Coleman 1997].

*Tabla 2-4. Clasificación de aplicaciones y servicios de groupware según la coincidencia de los participantes en el tiempo y el espacio*

	<b>Síncrono o “Mismo Tiempo”</b>	<b>Asíncrono o “Distinto Momento”</b>	
<b>Mismo Lugar o “Cara a Cara”</b>	Soporte Presentación Soporte Decisión	Juegos Multijugador Cuadro de Mando Soporte Votación Escritura Colaborativa Gestión Contactos Gestión Documental Gestión Conocimiento Control Flujo Trabajo Plataforma Educativa Ordenador Compartido	
<b>Distinto Lugar o “A Distancia”</b>		Conferencia Pizarra Charla Mensajería Instantánea	Gestión Conocimiento Seguimiento, Información y Recomendación Subasta Teleoperación

Algunas de las aplicaciones y servicios de groupware más importantes se resumen en la siguiente tabla, clasificadas atendiendo al primero de los criterios de clasificación de la tabla anterior.

*Tabla 2-5. Clasificación de aplicaciones y servicios de groupware según la coincidencia de los participantes en el tiempo*

<b>Síncronas</b>	
Soporte Presentación	Proporcionan el soporte necesario para presentar información a un grupo de usuarios.
Soporte Decisión (Decision Support)	Facilitan soporte para facilitar la tomar decisiones a un grupo de usuarios.
Conferencias (Conferences)	Permite la comunicación interactiva oral de un grupo a través de un sistema, como es el caso de las multiconferencias y las videoconferencias.
Pizarra Compartida (Shared Whiteboard)	Permite a un grupo de usuarios compartir una superficie de dibujo, que normalmente se usa como apoyo de un sistema de comunicación síncrona (conferencia, chat...).
Charla (Chat)	Permiten la interacción escrita entre un grupo de usuarios.
Mensajería instantánea	Permite el intercambio inmediato de mensajes entre usuarios.

<b>Síncronas o Asíncronas</b>	
Juego Multijugador (Multi-Player Games)	Facilitan la infraestructura necesaria para que un grupo de usuarios puedan jugar a través de un sistema.
Cuadro de Mando	Permite acceder a información clave actualizada sobre la actividad del grupo.
Soporte Votación	Proporcionan soporte para la emisión y proceso de votación.
Escritura Colaborativa (Collaborative Writing)	Permiten la colaboración de los miembros de un grupo para la preparación de un documento, facilitando la a los usuarios la anotación y modificación de los mismos.
Gestión de Contactos	Proporcionan el soporte para gestionar un repositorio compartido de referencias de contacto.
Gestión Documental	Facilitan la gestión de un repositorio documental compartido.
Gestión Conocimiento	Proporcionan soporte para gestión de una base de conocimiento.
Control Flujo Trabajo (workflow)	Permite controlar el movimiento de información en el seno de una organización siguiendo los procesos para ello establecidos.
Seguimiento, Información y Recomendación	Facilita el seguimiento de la actividad de los usuarios del grupo para que todos sean conscientes de ella y puedan aprovecharla.
Plataforma Educativa	Proporcionan una combinación de funcionalidades que son necesarias para utilizar un sistema basado en TICs como plataforma de formación.
Ordenador Compartido	Se podría considerar un procedimiento rudimentario de soporte para la colaboración, que es muy común, en el que se utiliza un mismo ordenador como plataforma de trabajo de un grupo de colaboradores y a través de él comparten información, se comunican, coordinan, etc.
Subasta	Proporcionan soporte para la oferta de bienes y su valoración y posterior intercambio mediante un procedimiento de puja.
Teleoperación	Permite el control de sistemas de forma remota y simultanea con otros usuarios.
<b>Asíncronas</b>	
Correo electrónico (e-mail)	La más común de las aplicaciones de groupware. Esencialmente es una tecnología que permite el envío de mensajes entre usuarios, aunque actualmente permite hacer operaciones más sofisticadas, como reenviar mensajes, adjuntar ficheros, crear grupos de correo, etc.
Listas de correo (mailing lists)	Hace posible enviar mensajes de correo electrónico a grupos de usuarios. La diferencia con los foros es que en estos los usuarios deben acceder a los mensajes solicitándola de forma explícita.
Foros de noticias (newsgroups)	Permite publicar mensajes de correo electrónico en una clasificados en temas para que los usuarios autorizados puedan acceder a ellos bajo demanda.
Weblog (blog)	Proporciona un espacio Web donde se pueden recopilar artículos de forma colaborativa, sobre una temática determinada.
Wiki	Permite la edición colaborativa de un sitio Web de forma interactiva a través de un navegador Web.
Calendario y Agenda (calendars)	Soporte para coordinación de recursos, planificación, gestión de proyectos, etc.

Entre las aplicaciones que incorporan tecnologías de groupware existen algunas de propósito general, los Sistemas de Trabajo Colaborativo, que pretenden proporcionar un entorno que fomente la comunicación y la coordinación entre un grupo de personas [OpenGroupware 2008]. Existen numerosos productos que podrían incluirse en esta categoría, la mayoría de ellos

incluyen algunas funcionalidades que pueden considerarse fundamentales, aunque en ocasiones incorporan otras utilidades extras.

Algunas de las funcionalidades que se pueden considerar comunes a los sistemas de trabajo colaborativo son las siguientes: soporte para la comunicación asíncrona, como el correo electrónico y las listas de distribución, los foros de discusión y los tableros de noticias; soporte para la comunicación síncrona, como pueden ser las charlas, la mensajería instantánea o la multiconferencia; soporte para la presentación de información y su debate, como la presentación de pantallas o las pizarras electrónicas; soporte para la coordinación y gestión de actividades y tareas, como el seguimiento de actividades, la gestión de proyectos, el control de flujo de trabajo, y los calendarios y agendas compartidas; soporte para la creación colectiva, como la escritura colaborativa; soporte para compartir información, como los contactos compartidos, transferencia de ficheros, gestión documental o vinculaciones hipertextuales entre elementos.

Otras funcionalidades adicionales que pueden incorporar los sistemas de trabajo colaborativo, pero que no son tan comunes y sólo algunos sistemas lo hacen, son entre otras: soporte para la decisión, como los cuadros de mando o la asistencia para votaciones; soporte para conferencia avanzada, como las videoconferencias o las presentaciones multimedia sincronizadas; soporte para juego, como los entornos multijugador o la gestión de torneos; soporte para la teleoperación, como pueden ser los escritorios remotos.

En la actualidad pueden encontrarse aplicaciones de trabajo colaborativo adaptadas a distintos entornos, que aunque comparten muchas funcionalidades, añaden otras que son específicas del entorno en cuestión o se han implementado de manera que se acomodan a las expectativas de los usuarios correspondientes. Entre las plataformas de trabajo colaborativo específicas pueden mencionarse las dedicadas a teletrabajo, teleformación, juego, gestión de proyectos, asistencia técnica (“help desk”), atención al cliente (“customer care”), gestión de conocimiento, redacciones digitales en medios de comunicación, etc.

Actualmente hay muchas oportunidades de colaborar mediante ordenadores fuera del entorno laboral a través de Internet, ejemplos de ello son aplicaciones con tanto éxito como las charlas, los juegos o las subastas, que en su mayoría son de participación voluntaria. Aunque en el entorno laboral existen cada vez más aplicaciones colaborativas, no suelen tener tan buena aceptación. En este contexto es más difícil que los usuarios se presten voluntariamente a utilizar dichas aplicaciones. [Palen 2002]

De acuerdo con Grudin [Grudin 1994a] las aplicaciones de trabajo colaborativo con éxito, como el correo electrónico, las bases de datos o los sistemas de gestión de código, comparten algunas características: los beneficiarios de ellas no son los directivos o las personas que toman las decisiones, sino gente que utiliza el ordenador de forma más rutinaria; son aplicaciones

orientadas a organizar y manejar la información sin incorporar nociones de rol, proceso e interacciones sociales; además, en ellas se suele adoptar una perspectiva del lugar de trabajo o uso más que de la tecnología empleada.

Brinck [ForakerDesign 2005] señala que con las aplicaciones de groupware es mucho más difícil acertar que con el software tradicional. Habitualmente, un sistema de groupware no puede dar buen resultado sin que la mayoría de los miembros del grupo o todo ellos estén dispuestos a adoptar el sistema. En contraste, un sistema de usuario individual puede tener un buen resultado aunque sólo una fracción de sus destinatarios lo adopta.

Grudin considera que existen ocho retos para los desarrolladores de aplicaciones de trabajo colaborativo [Grudin 1994a]:

1. Disparidad entre trabajo requerido y beneficios obtenidos, las aplicaciones de trabajo colaborativo a menudo requieren trabajo adicional de individuos que no reciben los beneficio directo del uso de la aplicación.
2. Masa crítica e interés general, las aplicaciones de trabajo colaborativo deben alcanzar una cantidad de usuarios mínima y evitar que los usuarios se muevan por sus intereses individuales para llegar a ser útiles.
3. Desbaratamiento de los procesos sociales, el groupware puede llevar a actividades que violen tabúes sociales, amenacen estructuras políticas existentes, o desmotiven a los usuarios claves para su éxito.
4. Manejo de excepciones, las aplicaciones de trabajo colaborativo pueden no responder al amplio rango de excepciones e improvisaciones que caracterizan muchas actividades de grupo.
5. Acceso infrecuente, las funcionalidades que soportan los procesos de grupo se utilizan con relativa poca frecuencia, por lo que requieren una accesibilidad discreta e integrada con otras características más utilizadas.
6. Dificultad de evaluación, los obstáculos casi insuperables para el análisis y la evaluación significativos y generalizables de las aplicaciones de trabajo colaborativo impiden aprender de la experiencia.
7. Falta de intuición, el empleo de la intuición en el desarrollo de aplicaciones multiusuario resulta totalmente inadecuado, por lo que su empleo es una mala decisión y un procedimiento de diseño propenso a errores.
8. La adopción de procesos, las aplicaciones de trabajo colaborativo requieren de una introducción más cuidadosa en entorno de trabajo que otros tipos de aplicaciones.

Para el diseño de aplicaciones de groupware es fundamental entender los grupos y su dinámica, pero también tener una buena comprensión de la tecnología de red y de cómo los aspectos de esta tecnología -por ejemplo, los retrasos en las vistas sincronizadas- afectan a la experiencia de

los usuarios. Además, todas las cuestiones relacionadas con el diseño de las interfaces de usuarios tradicionales mantienen su importancia, puesto que la tecnología involucra personas.

Según Brinck [ForakerDesign 2005], también otros aspectos específicos de los grupos requieren consideración. Por ejemplo, no sólo el comportamiento de un grupo de un millón de personas es distinto al de un grupo de cinco, si no que los requerimientos de prestaciones de las tecnologías que soporta ambos varía también. La facilidad de uso debe ser mejor para groupware que para sistemas de uso individual, porque el ritmo del uso de una aplicación está marcado a menudo por el ritmo de la conversación. La sensibilidad del sistema y su fiabilidad llega a ser una cuestión más significativa. Los diseñadores deben entender el grado de homogeneidad de los usuarios, de los posibles papeles que desempeñan en el trabajo cooperativo y de quién son las decisiones claves y qué influye en ellas.

Por último, algunas cuestiones de carácter coyuntural también deben considerarse en el diseño de las aplicaciones con soporte colaborativo, cuestiones como el cambio de actitud de la dirección respecto al uso del groupware, la madurez y mayor exigencia de los usuarios de estas tecnologías y la convivencia de las aplicaciones colaborativas con actividades tradicionales en grupo.

Grudin destaca en [Grudin 2003] el cambio de actitud de la dirección de las organizaciones respecto al groupware, aunque siempre se ha encargado tomar decisiones sobre su adquisición y frecuentemente han sido sus principales beneficiados, han pasado de delegar la interacción con ellas y su adopción personal, a convertirse en sus primeros y más entusiastas usuarios. Este fenómeno tiene implicaciones considerables en las características que las aplicaciones y sistemas deben incorporar para cubrir las expectativas de este poderoso colectivo, que es el que decide en las organizaciones, aunque sus necesidades no siempre coincidan con las del resto usuarios.

En el contexto del desarrollo de una aplicación colaborativa de anotación multimedia Bargerón [Bargerón 2004] observa que actualmente los usuarios se han hecho más sofisticados y demandan soporte a niveles muy detallados, de forma que su disposición para adoptar interfaces de propósito general parece reducirse.

Dentro de un trabajo sobre colaboración mediante anotación de documentos Cadiz [Cadiz 2000] plantea el interés de investigar el papel de actividades ajenas a las aplicaciones de trabajo colaborativo, como la interacción cara a cara o a través del teléfono, en los resultados obtenidos con estas tecnologías. De este modo, el diseño de aplicaciones que incorporen trabajo colaborativo requiere, además de la consideración de los problemas de interacción entre las personas y las TICs, del análisis de los procesos de comunicación, coordinación y acceso a la información que se producen dentro de los grupo, a través o no de esas tecnologías.

[Gutierrez 2005] En los procesos de grupo intervienen aspectos individuales, del grupo y de la organización, que es imprescindible considerar en su análisis. Los primeros vienen dados por las características de la comunicación humana, los patrones de la actividad individuales, como las prácticas y los estilos de trabajo, y el modo de la interacción con el grupo mediante interfaces de soporte para ello. Los aspectos del grupo están asociados a cuestiones como el entorno en el que se desarrolla la actividad, la implicación de los participantes en ella, la conciencia de equipo -*awareness*-, o las dinámicas del mismo, como el grado de colaboración o el rendimiento conseguido. Por último, los aspectos de la organización se relacionan con cuestiones como el modo de representar el conocimiento en la entidad, la manera de estructurarlo, y la forma en la que se realiza la gestión de actividades, personas y recursos.

También es necesario tener en cuenta para el análisis y diseño de sistemas CSCW algunos problemas de carácter general, como la separación entre trabajo individual y en grupo, el empleo de ordenadores, la adaptación a las aplicaciones de trabajo en grupo, el funcionamiento de las aplicaciones en Internet o quién se beneficia del uso de las tecnologías. Otros problemas tienen que ver con el diseño, como el establecimiento de interfaces para las actividades y el control del grupo, el soporte para procesos de grupo, el control de concurrencia, y el control de acceso a la información, entre otros.

Los objetivos generales para el análisis y el diseño de sistemas de trabajo en grupo son dos fundamentalmente. El primero de ellos es abordar el comportamiento del grupo para mejorar el balance de carga de trabajo entre los participantes, para obtener una representación de la organización en términos de los roles desempeñados por los actores implicados y de las relaciones que se establecen entre ellos, y para establecer el modo de soportar cuestiones como la conciencia de grupo o los protocolos sociales. El otro objetivo fundamental es conseguir establecer una conexión entre estos estudios y las metodologías y herramientas para desarrollo de aplicaciones.

Para conseguir dichas metas generales se deben alcanzar varios objetivos particulares. En primer lugar, establecer un marco conceptual y metodológico para estudiar y desarrollar sistemas cooperativos. En segundo lugar, establecer un modelo del sistema que capture las peculiaridades del comportamiento del grupo. Además, es deseable emplear una notación gráfica para la representación del sistema. Por último, determinar un procedimiento de validación y verificación de propiedades. Después de todo ello se estará en disposición de acometer el desarrollo de la aplicación.

La metodología AMENITIES [Gea 2002] proporciona un marco y un procedimiento adecuado para el análisis y el diseño de sistemas CSCW, alcanzando todos los objetivos enunciados. El marco metodológico considera la ingeniería de tres aspectos fundamentales: sistema, que se



ocupa del entorno, la interacción con él y los factores críticos relacionados que afectan al diseño, como aspectos sociales, humanos u organizativos; requisitos y software, que se refieren al análisis de requerimientos y taras, y al diseño basado en modelos de tareas para el modelado del sistema y el diseño de la participación de los involucrados -*stakeholders*-; y usabilidad, que analiza la organización, grupos, usuarios e interfaces para medir cómo las aplicaciones de trabajo en grupo soportan los procesos grupales.

El esquema general de la metodología AMENITIS [Gea 2002] muestra los cuatro modelos en los que se basa. El primero es el Modelo de Requerimientos, que se dedica a la obtención de requisitos, mediante la etnografía aplicada, y la descripción de los mismos, empleando casos de uso. El segundo es el Modelo de Cooperación, COMO, que constituye el núcleo central de la metodología y permite la modelización organizacional, cognitiva, de la interacción y de la información. El tercero es el Modelo Formal, que permite el análisis automático del sistema mediante el empleo de Redes de Petri Coloreadas. Por último, el cuarto es el Modelo de Desarrollo, que facilita el diseño de la arquitectura del sistema mediante UML.

La metodología AMENITIS se ha aplicado al análisis del entorno de gestión de conocimiento colaborativo KnowCat [Gea 2004], del que luego se hablará en este capítulo.

### Awareness

De acuerdo con Dourish y Bellotti [Dourish 1992], en el ámbito del CSCW - Computer\_Supported Cooperative Work- el *Awareness* se refiere al entendimiento de la actividad de los demás participantes en un entorno colaborativo, que proporciona un contexto para la propia actividad en dicho entorno. Este contexto se emplea para asegurar que las contribuciones individuales son relevantes para la actividad del grupo en conjunto, y para evaluar las acciones individuales respecto a los objetivos del grupo y sus progresos. La conciencia de la actividad es un medio para controlar el proceso de trabajo colaborativo.

Según estos autores, los sistemas de soporte para el trabajo colaborativo tienen distintos modos de proporcionar *awareness*. Uno de estos mecanismos, al que se refieren como *Informacional*, trata de proporcionar facilidades explícitas para que los colaboradores se informen de sus actividades -por ejemplo, mediante un fichero de registro o vía correo electrónico-. Otro mecanismo, al que denominan de *Rol Restringido*, consiste en proporcionar soporte manifiesto para distintos papeles en el sistema colaborativo, de forma que cada uno establezca un tipo de relación con los objetos del trabajo compartido y con los otros colaboradores, lo que normalmente va ligado a una serie de operaciones permitidas. Además, los autores proponen una tercera aproximación a la que llaman *Shared Feedback* -Retroalimentación Compartida-, que se basa en hacer evidente la información sobre la actividad individual de los componentes

del grupo en el entorno colaborativo, presentándola en un espacio compartido en lugar de en los espacios privados de cada uno.

Hudson y Smith [Hudson 1996] ponen de manifiesto dos problemas asociados al soporte del *awareness*, por un lado la posible violación de la intimidad de los usuarios del sistema colaborativo al monitorizar su actividad, y por otro lado la posible molestia que el exceso de información sobre la actividad del grupo podría provocar a sus integrantes. Para estos autores es imprescindible llegar a compromisos en la implantación de funcionalidades de *awareness* que eviten estos inconvenientes para aprovechar plenamente las ventajas de la conciencia en los sistemas de trabajo colaborativo.

En un trabajo sobre un entorno de oficina virtual que integra comunicación, colaboración y *awareness* [Sohlenkamp 1994] se establecen algunas características de funcionalidad de conciencia de actividad en un entorno de CSCW en modo síncrono y asíncrono. Cuando la colaboración se produce en el mismo momento el *awareness* se ocupa de mostrar lo que el reto del grupo está haciendo en ese instante. En este caso los objetivos de la conciencia de actividad del grupo se dirigen a proporcionar información sobre cuestiones como qué están haciendo los colaboradores, cuáles están disponibles para contactar con ellos, el nivel de disponibilidad propio, el grado de información propia proporcionada al resto del grupo, qué documentos están utilizando otros usuarios o lo que estos están haciendo exactamente durante una sesión de trabajo en grupo. Cuando la colaboración no se produce en el mismo momento, el *awareness* se encarga de mostrar que han hecho los demás componentes del grupo recientemente. En esta ocasión los objetivos de la conciencia de actividad grupal se orientan a suministrar datos sobre aspectos como qué objetos compartidos se han cambiado por otras personas, cómo se han cambiado tales elementos o qué colaboradores han dejado mensajes para otros y dónde están dichos mensajes. Los objetivos y las tareas encomendadas al *awareness* dependen del grado de acoplamiento temporal de la colaboración.

Otro tipo de conciencia, vinculada con la interacción persona-ordenador y los sistemas adaptativos, es la de contexto de la interacción, *context-awareness* [Byun 2001], que se refiere a la capacidad del sistema para adaptar su comportamiento al contexto personal del usuario o del entorno en el que se produce su interacción. Para que el sistema mantenga este tipo de conciencia debe tener un modelo del usuario y otro del contexto de su actividad con dicho sistema. Además, el sistema debe estar dotado de sensores para detectar el contexto combinados con las técnicas de modelado de usuario tradicionales.

### Monitorización Actividad Usuario

La implantación predominante de la arquitectura cliente/servidor en el ámbito de Internet y de la Web, ha limitado el desarrollo de técnicas de monitorización continua de la actividad de los

usuarios en ese contexto durante los últimos tiempos, debido a que dicha monitorización se ha centrado en el servidor fundamentalmente. Sin embargo, es evidente el interés de desplazar el análisis del lado del servidor Web al del cliente, puesto que en el servidor sólo hay información parcial y entrecortada, mientras que en el cliente es posible hacer un seguimiento continuo, así como considerar el uso de la Web en un entorno multiaplicación, y monitorizar las páginas Web completas, y no como un conjunto de elementos separados [Fenstermacher 2002a].

En este contexto tiene especial importancia el uso de los eventos que generan las interfaces de usuarios. Hilbert y Redmiles [Hilbert 2000] revisan diversas técnicas para extraer información de tales eventos y aplicarlas a la interacción persona ordenador. La estandarización de los eventos proporciona la base para llegar a la monitorización global más allá de los límites del navegador Web o de cualquier aplicación aislada [Fenstermacher 2002b].

En ámbito de la Web, la recomendación del modelo de objeto de documento -DOM, por sus siglas en inglés-[DOM 2008] realizada por el W3C [W3C 2004] establece una serie de eventos [DOMEvent 2008] que deberían implementar los clientes Web que soporten dicho modelo -que incluyen las últimas versiones de los navegadores más populares-. Entre los eventos relacionados con la interacción del usuario con las páginas Web en el cliente del sistema están: los de interfaz de usuario, entre los que se cuentan, *focus* y *blur*, para indicar cuándo ganan o pierden el foco ciertos elementos de la interfaz; algunos básicos de carácter general, que incorporan *load*, *unload*, *resize* y *scroll*, entre otros, que indican respectivamente la carga o descarga de las páginas, los cambios de tamaño de las ventanas y los desplazamientos de sus áreas visibles; los referidos al ratón, que incluyen *click*, *dblclick*, *mousedown*, *mousemove*, *mouseup*, *mouseover* y *mouseout* que corresponden al clic y doble clic del ratón, a la presión y liberación de sus botones, su movimiento sobre la página, así como su posición sobre algunos elemento de ella y su salida de ellos; y los relacionados con el teclado, que son *keydown* y *keyup*, e indican la presión de una tecla y su liberación, respectivamente.

El lenguaje interpretado JavaScript [JavaScript 1999] permite manejar los eventos de DOM desde las páginas en los clientes Web, aunque con ciertas peculiaridades dependiendo de la implementación particular de los módulos correspondientes en cada navegador, por lo que es imprescindible considerar la compatibilidad entre versiones para desarrollar *scripts* que funcionen en varios sistemas. Otros lenguajes de cliente también permiten el manejo de DOM y de sus eventos.

La monitorización de la interacción de los usuarios en el sistema puede vulnerar la privacidad de los mismos. Por esta razón los clientes Web y los lenguajes de programación suelen imponer algunas limitaciones sobre esta actividad. En particular, el uso de eventos y de accesos a páginas Web no solicitadas ha sido objeto de restricciones sucesivas en los últimos tiempos. Estas

limitaciones, por supuesto, deben ser respetadas, pero también obtener la aprobación de los implicados para su empleo no restringido. Estas consideraciones son importantes en la aplicación de estas técnicas, pero no reducen el interés de indagar en sus posibilidades.



Figura 2-2. Mapas de calor, CrazyEgg [CrazyEgg 2008] izquierda y ClickHeat [ClickHeat 2008] derecha

Una aplicación interesante del análisis de los eventos en los clientes Web son los mapas de calor de clics, que permiten visualizar las áreas de las páginas Web donde se han producido los clics durante un intervalo de tiempo, representando la frecuencia de los mismos en cada punto con distintos colores. Como resultado las zonas afectadas aparecen señaladas en los mapas con colores indicativos del grado de actividad que han tenido. Una de las aplicaciones de estos mapas en el ámbito de la interacción personas-ordenador es el análisis de las interfaces de usuario. Las aplicaciones que emplean este tipo de visualización deben recopilar los eventos de ratón correspondientes y mantener un registro de su número por cliente y posición de las páginas, para luego generar el mapa en cuestión. Aplicaciones como CrazyEgg [Crazyegg 2008] y ClickHeat [ClickHeat 2008] generan mapas de calor de clics (ver Figura 2-2).

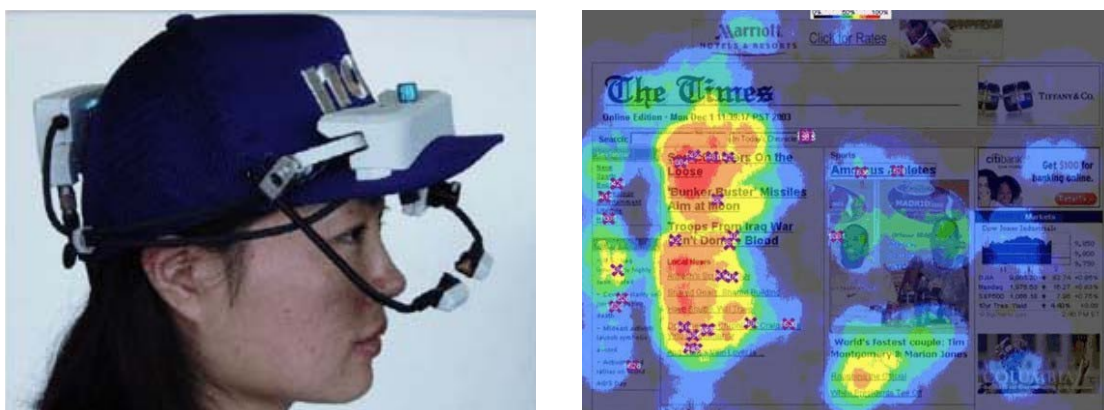


Figura 2-3. Sistema de Eye-tracker en sujeto [Prendinger 2007] y mapa de calor [EyetrackIII 2008]

Sin embargo, el seguimiento de los eventos de usuario no proporciona un registro completo de su actividad con el sistema, puesto que existen periodos de la misma durante los que no se generan eventos a pesar de que la interacción con el sistema continúe, por ejemplo cuando el

usuario dedica algún tiempo a leer un texto o a revisar un gráfico. Existen algunos procedimientos más sofisticados que el registro y análisis de eventos para obtener impresiones más completas de la actividad del usuario con el sistema. Se trata de técnicas como el *eyes tracking*, para el seguimiento de los ojos [Hassan 2007]. Los resultados de este tipo de análisis pueden visualizarse de forma gráfica sobre las interfaces analizadas, en ocasiones empleando mapas de calor análogos a los de clics, pero utilizando los datos sobre los movimientos de los apuntadores rastreados en lugar de los clics de ratón (ver Figura 2-3). Estos procedimientos se utilizan en diversos campos entre los que está la interacción persona-ordenador. El principal inconveniente de estos sistemas es el requerimiento de dispositivos y aplicaciones especiales, así como el coste de todo ello.

Por lo tanto, a pesar de que el seguimiento de eventos de usuario en el cliente no proporciona toda la información de la interactividad del mismo, facilita una información suficientemente detallada para algunos análisis sin infraestructuras especiales ni costosas, suministrando, a pesar de ello, datos más ricos que el seguimiento de la actividad observada solamente desde el lado del servidor en aplicaciones cliente/servidor como son las basadas en el servicio Web.

### *Hipermedia Adaptativa*

Brusilovsky presenta la Hipermedia Adaptativa como una línea de investigación en el área de los sistemas adaptativos de usuario. El objetivo perseguido es incrementar la funcionalidad de la hipermedia haciéndola personalizada. Los Sistemas Hipermedia Adaptativos (SHA) son sistemas que reflejan algunas características del usuario, como objetivos, preferencias o nivel de conocimientos, en un modelo de usuario y aplican este modelo para adaptar al usuario diversos aspectos del sistema, como contenidos, navegación o apariencia [Brusilovsky 1998].

Los Sistemas Hipermedia Adaptativos (SHA), por tanto, aparecen con el propósito de mejorar la usabilidad de los sistemas hipermedia tradicionales. Muchos de ellos facilitan la actividad de los usuarios, porque adaptan el sistema a las características de estos. El diseño de un SHA sugiere cuatro cuestiones sobre la adaptación: ¿qué?, ¿a quién?, ¿cómo?, y ¿cuándo? adaptar. Por otro lado, tres elementos están presentes explícita o implícitamente en la mayoría de los SHA: los modelos de dominio, usuario y adaptación. Dependiendo de las respuestas a las anteriores cuestiones se diseñarán los tres elementos. En la Tabla 2-6 se presentan las aproximaciones documentadas al tema desde distintos puntos de vista y se relaciona cada criterio con las cuestiones más importantes resueltas [Medina 2000].

Tabla 2-6. Clasificación de los sistemas Hipermedia Adaptativos según Medina [Medina 2000]

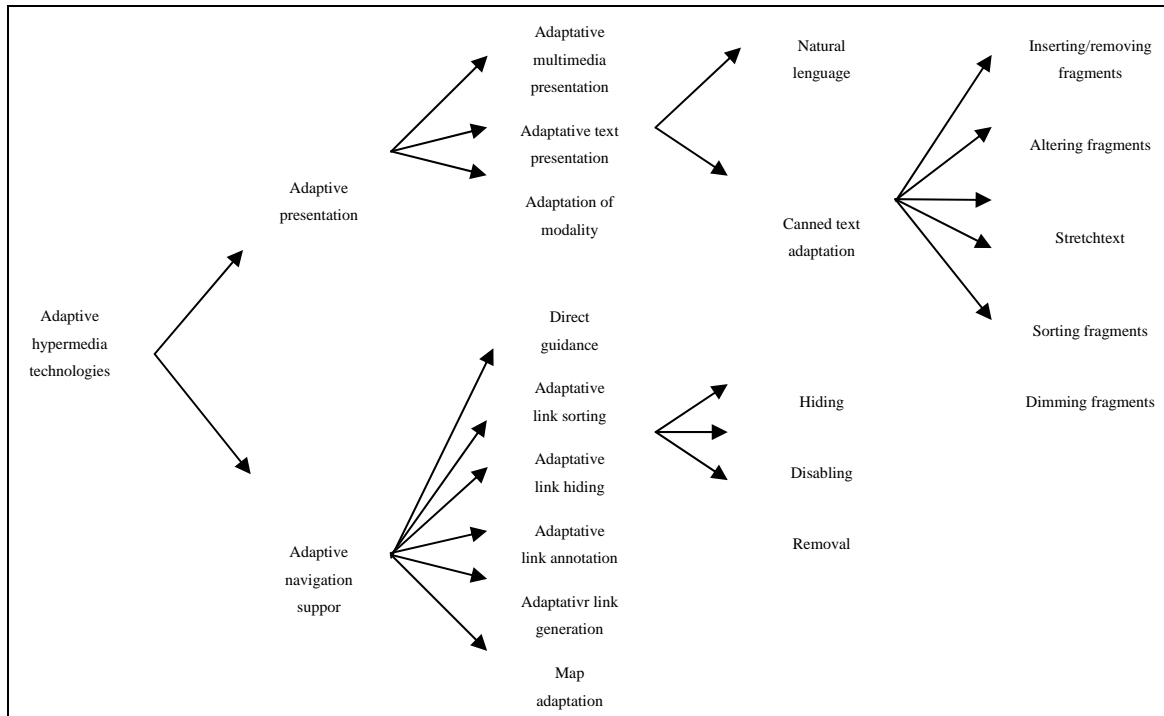
Criterio	Tipos				
<b>Dominio de Aplicación</b> <i>¿Qué?</i>	<b>Sistemas Hipermedia Adaptativos Generales:</b> Los documentos incluyen temas muy diferentes.				
	<b>Sistemas Hipermedia Adaptativos Específicos:</b> Todos los documentos explican conceptos de un dominio de información uniforme.				
<b>Adaptación a</b> <i>¿A quién?</i>	<b>Usuario</b>	<b>Representación del Modelo de Usuario</b>	Pares (atributo/valor)		
			Modelo Bayesiano		
			Aproximación probabilística		
			Redes de Petri		
	<b>Grupo de Usuarios</b>	Recomendaciones personalizadas			
Recomendaciones a los grupos de usuarios					
<b>Otras:</b> Por ejemplo, adaptación a distintos tipos de dispositivos en tecnología móvil					
<b>Métodos de Adaptación</b> <i>¿Qué? ¿Cómo?</i>	<b>Navegación Adaptativa</b>		Guía	Local o Global	
			Ayuda para la Orientación	Local o Global	
			Vistas Personalizadas		
	<b>Presentación Adaptativa</b>		Explicaciones Adicionales		
			Explicaciones con Prerrequisitos		
			Explicación Comparativa		
			Explicaciones con Variaciones		
Ordenación					
<b>Clase de Prerrequisitos</b> <i>¿Cómo?</i>	<b>Prerrequisitos Pedagógicos:</b> Establecer relaciones entre conceptos relacionados con el aprendizaje				
	<b>Prerrequisitos como Mecanismos de Ordenación:</b> Intenta establecer un orden parcial entre páginas				
<b>Integración de Información de Distintos Orígenes</b> <i>¿Qué?</i>	<b>Sistemas de Hipermedia Adaptativa Abiertos:</b> Sistemas que pueden integrar recursos de información localizados en cualquier sitio de la Web				
	<b>Sistemas de Hipermedia Adaptativa Cerrados</b>				

<b>Interacción con la Adaptabilidad</b> ¿Cuándo? ¿Cómo?	<b>Sistemas Hipermedia Adaptables:</b> El modelo de usuario sólo se cambia por petición expresa del usuario ( <i>Adaptability</i> )	
	<b>Sistemas Hipermedia Adaptativos:</b> El modelo de usuario se actualiza automáticamente como la navegación de usuario y el contenido y la estructura de enlaces se adapta a ello ( <i>Adaptivity</i> )	El usuario puede tener algún control sobre el comportamiento de adaptación del sistema
		El usuario no puede tener ningún control sobre la adaptación
	<b>Sistemas Hipermedia Adaptables/Adaptativos</b> ( <i>Adaptability</i> y <i>Adaptivity</i> simultáneamente)	
<b>Creación de Documentos Hipermedia</b> ¿Cuándo? ¿Cómo?	<b>Sistemas Hipermedia Adaptativos Dinámicos:</b> Los documentos se crean dinámicamente bajo demanda del usuario	
	<b>Sistemas Hipermedia Adaptativos No Dinámicos:</b> Los Documentos existen antes de su uso, aunque su presentación es adaptada a cada usuario	
	<b>Intermediación entre Hipermedia Adaptativa y Dinámica:</b> Los documentos se crean dinámicamente bajo demanda del usuario	
<b>Historia de Navegación</b> ¿Cómo?	<b>Representación explícita de la historia</b>	
	<b>La representación de la trayectoria de navegación del usuario no es explícita</b>	
<b>Información Contextual</b> ¿A quién?	<b>Contexto de Usuario:</b> Papel del usuario en un grupo, localización física, etc.	
	<b>Contexto de Texto:</b> Frase, párrafo o documento circundante	
	<b>Contexto Espacial:</b> Trayectoria de navegación del usuario a través del hiperespacio hasta llegar a la página actual	

Brusilovsky presenta en su revisión del estado del arte en el campo de la Hipermedia Adaptativa, desde su último artículo en el mismo sentido cinco años antes, la clasificación actualizada de tecnologías en esta área [Brusilovsky 2001] se presenta en la Tabla 2-7.

Es interesante resaltar la diferencia entre los términos *Adaptability* y *Adaptivity*, mientras el primero se refiere a la capacidad de un sistema de permitir cambiar el modelo de usuario por solicitud expresa del usuario, el segundo corresponde a la característica del sistema que le permite actualizar dicho modelo de forma automática como resultado de la consideración de ciertas condiciones, como la interacción del usuario con el entorno. Estas posibilidades se incorporan en el trabajo de Frasincar sobre el diseño de un sistema de información Web inteligente [Frasincar 2002], en el que se presenta una metodología, llamada Hera basada en Ingeniería Web [Murugesan 1999], que considera explícitamente estos aspectos. El proyecto emplea ontologías que se construyen con Protege-2000 y se representan en RDFS.

Tabla 2-7. Clasificación sistemas Hipermedia Adaptativos según Brusilovsky [Brusilovsky 2001]



En los modelos de usuario para la adaptación anteriores a la aparición de la Hipermedia Adaptativa ya se incluían varios aspectos que siguen teniendo importancia en las posteriores modelizaciones [Kobsa 1993]: conocimiento, planes y preferencias. Para manejar el primero de ellos se puede establecer una clasificación de los usuarios por grupos con un determinado nivel de conocimiento, considerando determinadas características que puedan evaluarse con facilidad y que constituyan el prototipo de cada clase. En el segundo aspecto se parte de la consideración de un plan como una secuencia de acciones dirigidas a alcanzar un objetivo. Para manejarlo se proponen dos alternativas: una es que el sistema mantenga una librería de todos los posibles planes de usuario preestablecidos, la otra que el entorno proporcione un mecanismo para la construcción de planes a partir de una librería de posibles acciones con sus efectos y prerequisites. Para el tercer aspecto se pueden establecer agrupaciones, como en el primero.

Según Brusilovsky en [Brusilovsky 1998], se pueden identificar tres etapas en el proceso de adaptación, la de recolección de datos del usuario, la de procesamiento de esos datos para construir o actualizar el modelo de usuario, y la de aplicar el modelo de usuario para proporcionar la adaptación. Además, existen varios problemas relacionados con el empleo automático de los modelos de usuarios en los sistemas adaptativos. Uno de ellos es que no son completamente fiables porque el sistema puede equivocarse al asignar o emplear el modelo, el otro que algunas componentes del modelo de usuario no se pueden deducir sin colaboración del usuario. Al mismo tiempo, algunas investigaciones en este campo demuestran la existencia de un cierto número de técnicas para el modelado de usuario automático. En particular el seguir las



acciones del usuario para determinar los conocimientos del usuario y también el utilizar el plan de conocimiento como base para deducir los objetivos del usuario.

Continúa diciendo que el área de la Hipermedia Adaptativa tiene una problemática especial, porque el seguimiento de la actividad del usuario en el entorno hipermedia suministra insuficiente información para modelar al usuario. La única información que el sistema puede registrar sobre la actividad del usuario es la trayectoria seguida por el hiperespacio y el tiempo empleado en cada nodo. La trayectoria y los patrones de navegación del usuario pueden ser una fuente de información interesante, pero resulta difícil actualizar el modelo de usuario utilizando sólo esta información. Alguna información acumulativa sobre la navegación del usuario, como el tiempo gastado en un nodo o el número de visitas realizadas a él, es fácil de utilizar. Sin embargo de estos datos no se puede deducir lo que el usuario ha hecho en el nodo. En general, esta clase de información no es de confianza y no puede utilizarse como única fuente para construir el modelo de usuario.

Siguiendo con la misma referencia de Brusilovsky, la solución del modelado de usuario automático en el área de Hipermedia Adaptativa depende del área de aplicación. Se identifican dos áreas que disponen de fuentes adicionales de información para alimentar el modelo: los sistemas de ayuda en línea y los de hipermedia educativa en sistemas de tutoría inteligente. Los primeros permiten obtener fácilmente el nivel y objetivos de los usuarios. Los segundos pueden obtener información para el modelo de usuario del modelo de estudiante, que se actualiza principalmente analizando las respuestas del estudiante a los cuestionarios, preguntas o problemas incorporados en el sistema.

En general queda la posibilidad de involucrar al usuario en el proceso de modelado para que colabore en la obtención de la información. Esto puede hacerse de tres formas correspondientes a las etapas del proceso de adaptación que es el modelado de usuario: permitiendo que los usuarios proporcionen la información requerida por el mecanismo del modelado de usuario, en vez de que este intente adivinarla; permitiendo que los usuarios hagan ellos mismos la adaptación deseada, en lugar de que la haga el sistema por ellos; y permitiendo que el usuario pueda actualizar directamente la información del modelo de usuario, editándola. En este caso, si los usuarios no tienen capacidad para modificar toda la información del modelo de usuario, el sistema puede proporcionar un procedimiento para que un administrador colabore en esta tarea.

Lo que es similar en estas aproximaciones es la idea de involucrar al usuario en el proceso de modelado del mismo, para obtener información adicional de él y hacer el proceso más sencillo y fiable. Lo que es diferente es la forma en que se implementa y usa. Proporcionar realimentación es el método más sencillo para los usuarios, mientras que obtener datos para el modelo de usuario directamente de este es el más sencillo para el sistema. El primero no requiere grandes

modificaciones en la interfaz ni una carga para el usuario, pero la cantidad de información que puede proporcionar es muy pequeña. El segundo requiere una interfaz específica que puede ser complicada y su uso puede requerir dedicación y preparación por parte del usuario, pero proporciona gran cantidad de información útil. Una alternativa intermedia es la adaptación dirigida por el usuario, esta posibilidad requiere más atención del usuario que la realimentación, pero no tanta dedicación ni preparación como la de recibir datos directamente para el modelo. Además, la cantidad de información que se puede conseguir con este procedimiento es grande.

Además del Modelado de los Elementos del Sistema, al que nos hemos referido ya, otro tema de Hipermedia Adaptativa con gran interés para este proyecto es el de la Virtualita, en particular el concepto de documento virtual y el modo en que este se emplea en la adaptación.

Según Milosavljevic en [Milosavljevic 1999], por Documento Virtual se entiende un documento Web cuyos contenidos, nodos o enlaces, o las tres cosas, se crean según se necesitan. Existen varias clases de documentos virtuales en la Web cuyo contenido se establece dinámicamente. En primer lugar, se puede utilizar una plantilla para que se sustituyan los contenidos del nodo en el momento de ejecución. En segundo lugar, se pueden usar aplicaciones para generar valores de un sólo uso. En tercer lugar, se pueden utilizar scripts CGI y motores de búsqueda para componer documentos virtuales a partir de fragmentos de otros documentos bajo demanda. En cuarto lugar, se pueden generar meta-datos para generar resúmenes para los usuarios, donde los usuarios hacen las extracciones y condensaciones en el mismo momento. Y finalmente, se pueden emplear técnicas de generación de lenguaje natural para construir dinámicamente documentos virtuales a partir de datos en bases de datos o de conocimiento.

Existen varios temas de investigación en torno a los documentos virtuales [Watters 1999]:

- Generación, respecto a en qué momento se genera el documento virtual o cómo se hace.
- Búsqueda, en relación con el modo de encontrar los documentos virtuales, en qué dominio y si el documento existirá en el momento que el usuario decida buscarlo.
- Revisita, en el sentido de que los usuarios tienen la expectativa de que un documento virtual encontrado una vez estará disponible en búsquedas posteriores.
- Versionado, en el sentido de que los documentos virtuales van cambiando con el tiempo y es necesario determinar el documento generado en un momento determinado.
- Autenticación, respecto a quién es el responsable de los contenidos de documentos virtuales que se forman a partir de diferentes fuentes.
- Referencia, en relación con el modo de citar documentos virtuales o versiones de estos.

- Anotación, en el sentido de atender a la necesidad de los usuarios de añadir información en los documentos virtuales, como comentarios, anotaciones, enlaces, etc.

Martins propone un procedimiento de generación de documentos hipermedia virtuales mediante adaptación de documentos en tres fases [Martins 2002]: filtrado, empaquetado y formateado. Para ello se consideran modelos de documento, usuario, empaquetado y formato. El filtrado se encarga de seleccionar los documentos y de establecer su orden, a partir de un modelo de documento y de usuario. El empaquetado se ocupa de determinar la parte de cada uno de ellos que se muestra, en base a un modelo de empaquetado. Y el formateado de fijar la manera en que, finalmente, se presentan al usuario, de acuerdo con un modelo de formateo.

### *Sistemas de Recomendación*

Los sistemas de recomendación son sistemas de filtrado de información destinados a seleccionar los ítems de la misma que puedan interesar más a cada usuario. En general, los sistemas de filtrado de información tratan de eliminar de forma automatizada los contenidos no deseados o redundantes de un flujo de información antes de presentarlos a los usuarios. El objetivo fundamental de este tipo de sistemas es reducir el impacto de la sobrecarga de información. Generalmente, los sistemas de filtrado comparan el perfil de un usuario con las características de otros usuarios o de los ítems de información para conseguir sus objetivos. Existen gran número de aplicaciones que incorporan la recomendación entre sus funcionalidades, especialmente las de carácter comerciales, entre ellas puede mencionarse Amazon [Amazon 2008] que sugiere libros y otros artículos y MovieLens [Movielens 2008] que propone películas. El problema de la recomendación consiste en estimar valoraciones de los ítems de información que no ha visto un usuario. Normalmente esta estimación se basa en la valoración realizada por ese usuario de otros ítems y en otras informaciones, según la aproximación empleada.

Los sistemas de recomendación suelen clasificarse en tres categorías [Adomavicius 2005]. La primera es la de recomendación basada en contenido, *content-based recommendations*, en la que se recomienda a los usuarios ítems parecidos a los que prefirieron en el pasado. La segunda categoría es la de recomendaciones colaborativas, *collaborative recommendations*, en la que se sugiere a los usuarios ítems que eligieron previamente otras personas con parecidos gustos y preferencias. Por último, la tercera categoría es híbrida de las dos anteriores, *hybrid approaches*, que combinan métodos colaborativos y basados en el contenido. Por otro lado, además de los procedimientos que establecen valores absolutos de valoración, existen otras aproximaciones que consideran valores de preferencia relativos para los usuarios, se trata del filtrado basado en preferencias, *preferente-based filtering*.

En los sistemas de recomendación basados en el contenido, la utilidad de un ítem para un usuario se estima a partir de los grados de utilidad asignados por el usuario a otros ítems que se

consideran similares al primero. Este tipo de recomendación está basado en el campo de recuperación de información y en el de filtrado de información. Debido a los avances recientes en estos dos campos y la importancia de varias aplicaciones basadas en texto, muchos sistemas basados en contenido se centran en la recuperación de ítems que contienen información textual, como documentos, sitios Web y mensajes de noticias Usenet. En estos sistemas, la mejora sobre los sistemas de recuperación de información está en el empleo del perfil del usuario como instrumento de filtrado, aprovechando la información sobre sus gustos, preferencias y necesidades.

El procedimiento básico que se sigue en los sistemas de recomendación basados en contenidos consiste en establecer un perfil de los ítems, que son un conjunto de atributos que los caracterizan [Adomavicius 2005]. Comparando estos perfiles, el sistema debe determinar si los ítems son adecuados o no para incluirlos en las recomendaciones. En el caso más habitual de contenidos asociados a textos, los ítems suelen expresarse en términos de palabras clave. La importancia de una palabra en la descripción de un ítem se indica mediante un peso que puede establecerse de distintas formas. El mecanismo más conocido para asignar pesos a las palabras en un documento dentro de una colección en el ámbito de Recuperación de Información, consiste en considerar la frecuencia de las palabras en el documento y en el conjunto de la colección, de modo que se asignen mayores peso a las palabras que son más frecuentes en el documento y menos en la colección. De este modo, como se hace también en Recuperación de Información, los ítems se representan por vectores de pesos de palabras, que pueden compararse entre sí para determinar su parecido. El procedimiento más habitual para establecer el grado de similitud es determinar la distancia entre los vectores, que viene dada por el coseno del ángulo que forman. Sin embargo, esta no es la única aproximación propuesta para determinar la similitud entre ítems, entre las alternativas formuladas pueden mencionarse los clasificadores Bayesianos y varias técnicas de aprendizaje automático.

Existen varios problemas asociados con los sistemas de recomendación basados en contenido [Adomavicius 2005]. El primero de ellos es un par de limitaciones: una es la necesidad de que los ítems estén en un formato que permita su comparación automática, o que tengan asociada información adecuada para ello, lo que complica el tratamiento de material multimedia; y la otra limitación es que no se puede distinguir dos ítems distintos que tienen exactamente la misma caracterización. El segundo problema es la demasiada especialización, por la que los sistemas no pueden recomendar ítems que no se parezcan a otros ítems ya valorados por los usuarios, este inconveniente se trata de manejar en ocasiones introduciendo cierto grado de aleatoriedad en la recomendación, utilizando algoritmos genéticos por ejemplo. De forma parecida, los sistemas tampoco pueden recomendar ítems que son demasiado parecidos a otros que el usuario

ha visto anteriormente. Por último, el tercero de los problemas se produce cuando hay un usuario nuevo en el sistema y no se tiene constancia de los documentos que le interesan.

En los sistemas de recomendación colaborativos, la utilidad de un ítem para un usuario se estima a partir de los grados de utilidad asignados a dicho ítem por otros usuarios que son considerados similares al primero. Según el algoritmo empleado para la recomendación los sistemas de este tipo pueden clasificarse en dos grupos [Adomavicius 2005]: por un lado los basados en memoria o también llamados heurísticos, *memory-based* o *heuristic-based*, que hacen predicciones sobre la utilidad de los ítems partiendo de la experiencia obtenida analizando las valoraciones previas realizadas por los usuarios; y por otro lado los basados en modelo, *model-based*, que emplean las valoraciones generadas por los usuarios previamente para aprender y establecer modelos con los que predicen las valoraciones. Sin embargo, existen aproximaciones que combinan ambas modalidades en algunos sistemas.

La aproximación heurística o basada en memoria para la recomendación colaborativa emplea algoritmos empíricos para realizar predicciones sobre la utilidad de los ítems para los usuarios, agregando las valoraciones realizadas anteriormente por otros usuarios parecidos a los primeros. Existen diversas formas de llevar a cabo dicha agregación, que puede ser un simple promedio, pero en la mayoría de los casos tiene en cuenta la similitud entre los usuarios de algún modo. Se han propuesto distintos procedimientos para establecer la similitud entre dos usuarios, en la mayoría de los casos se establece comparando la valoración realizada por ambos sobre los mismos ítems. Las dos aproximaciones más populares [Adomavicius 2005] son la basada en correlación, *correlation-based*, y en el coseno, *cosine-based*. En el primer caso la similitud suele establecerse mediante el coeficiente de correlación de Pearson. En el segundo caso, los usuarios suelen representarse como vectores en un espacio multidimensional y la similitud medirse por el coseno del ángulo que forman, aunque existen diversas modificaciones del procedimiento base dirigidas a conseguir mejoras en el rendimiento.

Existen diversas aproximaciones para los sistemas basados en modelo para la recomendación colaborativa [Adomavicius 2005]. Algunas aproximaciones proponen el empleo de técnicas probabilísticas para el filtrado colaborativo, otras emplean procedimientos encuadrados en el marco del aprendizaje automático, y otras plantean soluciones basadas, por ejemplo, en modelos Bayesianos, en regresión lineal o también en modelos de máxima entropía.

Los sistemas de recomendación colaborativos puros no tienen algunos de los inconvenientes de los sistemas basados en contenido, en particular los problemas derivados de la incapacidad de estos sistemas para recomendar contenidos que no son similares a los ya vistos por los usuarios. Sin embargo, los sistemas colaborativos siguen teniendo problemas con los nuevos usuarios, puesto que necesitan aprender sus preferencias a partir de sus valoraciones previas para poder

hacer predicciones. Se han propuesto distintas técnicas para tratar mejor este inconveniente, como integrar procedimientos basados en contenidos y colaboración, considerar la popularidad de los ítems, su entropía, la personalización de los usuarios o combinaciones de unas y otras.

Además, los sistemas colaborativos tienen dos nuevos problemas [Adomavicius 2005]. El primero se produce cuando se incorpora un nuevo ítem, puesto que hasta que el nuevo elemento es valorado por un número considerable de usuarios el sistema no puede recomendarlo. El segundo problema particular de la aproximación colaborativa es la producida por la escasez de usuarios, puesto que es necesaria una masa crítica de ellos para que el sistema pueda funcionar correctamente. El empleo de información sobre el perfil del usuario puede servir para superar este inconveniente, para el que se han propuesto otras soluciones. Los sistemas de recomendación híbridos [Adomavicius 2005] combinan los métodos de basados en contenido y basados en colaboración, con la intención de evitar algunas de las limitaciones de ambas aproximaciones. Se han planteado varios modos de combinación de dichos métodos en un sistema. Una posibilidad es implementar los métodos basados en colaboración y en contenido por separado y combinar sus predicciones. Otra alternativa consiste en incorporar algunas características del mecanismo basado en contenido en la aproximación colaborativa. También se puede hacer a la inversa e integrar características del mecanismo colaborativo en la aproximación basada en contenido. Por último, se ha propuesto la opción de construir un nuevo modelo unificado y único que incorpore ambas aproximaciones.

Una aplicación interesante de los sistemas de recomendación es la Navegación Colaborativa, que permite obtener información de los movimientos de una comunidad en el hiperespacio para facilitar la navegación de sus miembros en él. La búsqueda de información en la Web consume mucho tiempo y puede llegar a ser frustrante por la gran cantidad de información y recursos accesibles. La Navegación Colaborativa en la Web puede mejorar la manera de encontrar y manejar los contenidos disponibles, gracias a la información sobre ellos obtenida de forma automática o por otros usuarios dispuestos a compartirla.

Barra presenta en [Barra 2002] un proyecto en desarrollo, denominado GAS (Group Adaptive System) que permite la navegación colaborativa en la Web. El objetivo de dicho proyecto consiste en suministrar información que se adapte automáticamente a grupos dinámicos de usuarios. Para ello se tienen en cuenta: por un lado, la estructura de la Web, utilizando el algoritmo CLEVER [Chakrabarti 1998] y otros procedimientos [Kleinberg 1998, 1999]; y, por otro lado, las interacciones de los usuarios cuando navegan por ella. La información obtenida por ambos medios se reúne en un entorno colaborativo, que ofrece a los usuarios un área de información consistente adecuado a los intereses y actividades del grupo. En resumen, GAS es un sistema que permite a los usuarios realizar navegación colaborativa asíncrona, combinando

recursos descubiertos por un grupo de usuarios y sugerencias obtenidas aplicando el algoritmo CLEVER para proporcionar recomendaciones al grupo.

La ventaja obtenida de las actividades desarrolladas por otras personas del grupo son varias. Por un lado, permite el ahorro del tiempo necesario para encontrar en la Web información relevante gracias al aprovechamiento de las navegaciones y selecciones realizadas por otros, por otro, obtener recomendaciones de información relevante basada en otras opiniones y, también, utilizar la navegación por la Web como otro canal de comunicación.

La utilización del algoritmo CLEVER permite obtener información no sólo basada en las páginas Web recomendadas por el grupo, sino también en la estructura propia de las páginas. En este algoritmo una consulta -el resultado de una consulta a un motor de búsqueda basado en texto- determina un sub-grafo de nodos -páginas HTML- y vínculos -enlaces HTML- ampliado con varios nodos cercanos. En particular, el algoritmo comienza evaluando un primer grupo de páginas con ayuda de un motor de búsqueda convencional, incrementando el conjunto inicial con nuevas páginas apuntadas desde estas páginas o que apuntan a ellas. Este proceso se repite varias veces.

Kleinberg [Kleinberg 1998, 1999] ha desarrollado un grupo de herramientas algorítmicas para extraer información de estructuras enlazadas, demostrando su efectividad en una variedad de contextos dentro de la World Wide Web. En particular se enfocaron en la utilización de enlaces para analizar una colección de páginas relacionadas de una manera amplia con un tema, y descubrir la página más relevante en ese tema. La calidad de los métodos de búsqueda necesariamente requiere de una evaluación humana, debido a la subjetividad inherente en el concepto de relevancia. Mejorar la calidad de los métodos de búsqueda en la Web es un problema amplio e interesante que desde muchas perspectivas tiene que ver con la eficiencia y la facilidad de almacenaje del algoritmo empleado.

Kurhila presenta en [Kurhila 2002] una experiencia interesante en la que se implementa un sistema para seguir la trayectoria de los usuarios entre páginas Web, en un entorno colaborativo de enseñanza basado en navegación social que se llama EDUCO. El sistema presenta a cada usuario la ubicación de los demás en tiempo real y permite el intercambio de opiniones mediante charlas y comentarios que se pueden asociar a las páginas accedidas. Además muestra mediante gráficos la popularidad de los sitios visitados por toda la comunidad. Para ello el sistema se comporta como un PROXY, que intermedia entre la comunidad de usuarios y las fuentes Web, sustituyendo los enlaces de las páginas por vínculos que pasan a través de él, y los clientes utilizan unos módulos especiales que mantienen abierto una conexión permanentemente por donde el servidor envía las actualizaciones del estado del sistema sin retardos.

Cantador presenta en [Cantador 2007] un mecanismo para la identificación de comunidades de interés a partir de los gustos y preferencias manifestados mediante perfiles basados en ontologías. Las comunidades identificadas se emplean en un sistema de recomendación de elementos anotados para ilustrar su utilidad. Relacionado con este artículo está el trabajo de Castells [Castells 2005] que indaga en el empleo de técnicas de personalización basadas en ontologías, para mejorar la fiabilidad en sistemas de recuperación de la información procurando respetar la intimidad de los usuarios y mantener el control de los mismos sobre el proceso.

### Visualización de la Información

Según Chen en [Geroimenko 2002], el atractivo y potencial de la Visualización de la Información se reconoce cada vez más en un amplio rango de sistemas de información. La Visualización de la Información pretende producir representaciones gráficas de la estructura de información abstracta para usuarios humanos, del mismo modo que la Web Semántica intenta generar un marco de descripción de recursos (*Resource Description Framework*) universal que los agentes *software* puedan utilizar. Por un lado, la Visualización de la Información y la Web Semántica pueden complementarse mutuamente en cuestiones relacionadas con la organización de recursos de información y su acceso. Por otro, ambas difieren en algunos aspectos fundamentales de la forma de definir y representar la semántica. En este sentido, mientras la Web Semántica enfatiza en aproximaciones formales y orientadas a las máquinas, la Visualización de la Información lo hace en el significado que los modelos de visualización espacial pueden comunicar a los usuarios.

El objetivo de la Visualización de la Información es revelar patrones, tendencias y otros discernimientos de un fenómeno, centrándose en la información abstracta. El principal desafío de la Visualización de la Información es representar la información no espacial y no numérica en una forma visual efectiva. Una definición que recoge estas cuestiones es debida a Card [Geroimenko 2002]: “La Visualización de la Información es el empleo de representaciones de datos abstractos, basadas en ordenador, interactivas y visuales, para amplificar su entendimiento”.

Las interfaces de usuario, objeto de la Visualización de la Información, no suelen beneficiarse del enriquecimiento semántico realizado para las interfaces entre máquinas, materia de la Web Semántica. En los sistemas basados en ontologías se utilizan éstas para organizar y recuperar la información, pero no es frecuente emplearlas también para enriquecer su presentación al usuario.



En este sentido, el producto Spectacle, que forma parte de la aplicación de gestión de información y contenidos Administrator, proporciona el componente Cluster Map para la visualización de datos ontológicos [Geroimenko 2002] (ver Figura 2-4). Sirva como ilustración de su funcionamiento este ejemplo, para presentar una colección de ofertas de trabajo organizadas por una ontología muy sencilla, cada oferta –objeto- se representa

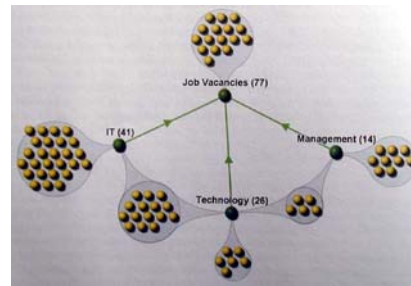


Figura 2-4. Cluster Map [Geroimenko 2002]

por una esfera amarilla, las clases por esferas grandes verdes etiquetadas con su nombre y cardinalidad, las relaciones entre estas mediante flechas y las asociaciones de ofertas –objetos- a sus más específicas clases mediante globos que representan agrupaciones, que pueden representar las intersecciones entre distintas clases de ofertas –objetos-.

<p>(a) Forma de grafo</p>	<p>(b) Geometría hiperbólica</p>	<p>(c) Tres dimensiones</p>
<p>(d) Lista de Texto</p>		

Figura 2-5. Topic Maps [Geroimenko 2002]

Como después se explicará en más detalle en el apartado de Web Semántica, el W3C ha desarrollado RDF para el enriquecimiento de los contenidos proporcionados a través de la Web.

Con el fin de enriquecer contenidos la Organización Internacional para la Estandarización (*International Organization for Standardization, ISO*) ha desarrollado los llamados *Topic Maps*, que recientemente se han adaptado para la Web. Ambas propuestas pretenden representar el conocimiento sobre recursos de información anotándolos. Le Grand y Soto presentan en [Geroimenko 2002] diversos modos de visualizar los *Topic Maps*, como puede verse en la ilustraciones que integran la Figura 2-5: en forma de grafo (a), utilizando geometría hiperbólica (b), en tres dimensiones (c), como de lista de texto (d) o en forma de mapas (e, f).

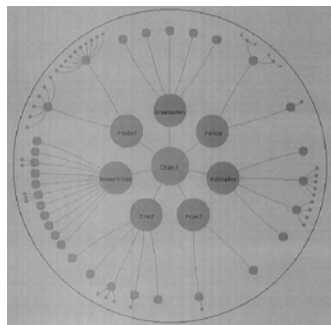


Figura 2-6. Interfaz hiperbólica para ontologías [Fensel 2001]

Los sistemas Ontobroker y On2broker [Fensel 2001] también utilizan un sistema de Visualización de Información hiperbólico para la interfaz de usuario con la que se puede navegar e interrogar al sistema. Las clases en el centro de la representación se muestran con círculos de mayor tamaño, mientras las de alrededor con círculos menores (ver Figura 2-6). Esta representación permite una navegación más rápida a clases lejanas desde el centro, así como de las clases cercanas y su vecindad.

Por su parte, Freire propone en [Freire 2004] una interfaz para visualizar y manipular estructuras hipermedias complejas en forma de grafos, que permite su manejo directo empleando múltiples técnicas contextuales para fijar el foco sobre parte de la información disponible y evitar el desorden y la sobrecarga de la misma. El sistema propuesto permite al usuario una manipulación de alto nivel de la estructura de información gracias a un mecanismo de ojo de pez basado en la agrupación jerárquica de la información. La navegación por la estructura facilita la orientación mediante animaciones entre los distintos estados de la presentación. Además, el sistema proporciona múltiples vistas sobre el mismo repositorio de datos.

Existen muchas más técnicas de Visualización de Información [Baeza 1999], desarrolladas para diversos aspectos de la Web, que van desde la estructura interna de los documentos, a la estructura global de la red. Muchas de estas representaciones pueden ser aplicadas en el ámbito de la Web Semántica [Geroimenko 2002].

Graphviz (Graphical Visualization Software) [Gansner 2000] es un paquete de software de código abierto que proporciona soporte para la visualización de gráficos con distintos diseños. El paquete facilita la descripción de los grafos mediante un lenguaje en formato de texto sencillo -ficheros DOT- y muy versátil y permite la generación de diagramas en diversos

formatos, tanto de imágenes de mapas de bits como de gráficos vectoriales (ver Figura 2-7). Existen muchos recursos basados en Graphviz, interfaces interactivos, herramientas auxiliares, librerías de aplicación y de enlace con varios lenguajes de programación, entre otros.

```
graph KC_SSO01SKC_2_3 {
dpi="72"; size="8,8";
node [ width="0.25", style="filled" ];
"T3" [ shape="circle", color="#9999FF", fillcolor="#CCCCFF", URL="inform.pl?b=KC_SSO01SKC&t=2" ];
"T7" [ shape="circle", color="#9999FF", fillcolor="#66c2a5", URL="inform.pl?b=KC_SSO01SKC&t=7" ];
"T3"--"T7"[ style="setlinewidth(8)", label = "0.23" ];
"T6" [ shape="circle", color="#9999FF", fillcolor="#66c2a5", URL="inform.pl?b=KC_SSO01SKC&t=6" ];
"T3"--"T6"[ style="setlinewidth(8)", label = "0.22" ];
"T27" [ shape="circle", color="#9999FF", fillcolor="#66c2a5", URL="inform.pl?b=KC_SSO01SKC&t=27" ];
"T3"--"T27"[ style="setlinewidth(8)", label = "0.21" ];
"T4" [ shape="circle", color="#9999FF", fillcolor="#66c2a5", URL="inform.pl?b=KC_SSO01SKC&t=4" ];
"T3"--"T4"[ style="setlinewidth(8)", label = "0.21" ];
"D2" [ shape="box", orientation="90.0", color="#9999FF", fillcolor="#f46d43",
URL="inter.pl?b=KC_SSO01SKC&t=3&d=2" ];
"T3"--"D2"[ style="dotted", label = "0.85" ];
"D1" [ shape="box", orientation="90.0", color="#9999FF", fillcolor="#fee08b",
URL="inter.pl?b=KC_SSO01SKC&t=3&d=1" ];
"T3"--"D1"[ style="dotted", label = "0.66" ];
"D8" [ shape="box", orientation="90.0", color="#9999FF", fillcolor="#66c2a5",
URL="inter.pl?b=KC_SSO01SKC&t=6&d=8" ];
"T3"--"D8"[ style="dotted", label = "0.21" ];
"D37" [ shape="box", orientation="90.0", color="#9999FF", fillcolor="#66c2a5",
URL="inter.pl?b=KC_SSO01SKC&t=27&d=37" ];
"T3"--"D37"[ style="dotted", label = "0.21" ];
{rank=max; "D37" "D8" "D1" "D2" };
{rank=min; "T7" "T6" "T27" "T4" }
}
```

Figura 2-7. Ejemplo de fichero DOT de Graphviz [Gansner 2000]

### Procesamiento del Lenguaje Natural

El lenguaje natural (LN) es el lenguaje hablado o escrito que utilizan las personas para comunicarse. El Procesamiento del Lenguaje Natural (PLN) es una disciplina de las áreas de conocimiento de la Inteligencia Artificial (IA) y de la Lingüística Computacional (LC). Como es sabido, la IA estudia los procesos para imitar la inteligencia de los seres vivos y su principal aplicación es la automatización de tareas que requieren de comportamientos inteligentes, como es el caso de entender el lenguaje de los seres humanos. Por su parte, el área de LC es un campo de conocimiento multidisciplinar que aplica la informática para el tratamiento del lenguaje humano, integrando lingüística e informática [Uszkoreit 2000], con psicología cognitiva y lógica. Entre las aplicaciones más destacadas de LC pueden citarse el soporte para corpus lingüístico, el diseño de análisis sintáctico del lenguaje natural y de etiquetados para su procesamiento, la definición de lógicas para el procesamiento de dicho lenguaje, la traducción automática y el estudio de las relaciones entre lenguajes naturales y artificiales.

Sobre estas bases, el PLN investiga y trata de formular mecanismos computacionalmente eficaces para permitir la comunicación entre hombres y máquinas empleando LN, para facilitar la comunicación entre personas con ayuda de máquinas o para facilitar el tratamiento automático de la información expresada en LN. Algunas de las aplicaciones más importantes de esta disciplina son el análisis del lenguaje, la comprensión del mismo, el reconocimiento del habla, la síntesis de voz, la traducción automática, la recuperación de información, la extracción de ésta o la síntesis de contenidos, entre muchas otras. En el ámbito de tratamiento automático

del lenguaje natural el término hace referencias a las propiedades computacionales y de otro tipo implicadas en la comprensión, producción y uso de las lenguas naturales.

Las tecnologías del PLN dependen de modelos formales, que son representaciones del conocimiento del lenguaje a distintos niveles, fonético, morfológico, sintáctico, semántico, pragmático y discursivo. Unos pocos modelos formales se utilizan para conseguir ese conocimiento, entre los que están máquinas de estados, sistemas de reglas formales, lógica y teoría probabilística. Los fundamentos de dichas tecnologías están en informática, lingüística, matemáticas, ingeniería electrónica y psicología. La vinculación esencial entre lenguaje y pensamiento ha puesto a las tecnologías del PLN en el centro del debate sobre máquinas inteligentes, y las investigaciones sobre cómo interaccionan las personas con medios complejos de comunicación parecen indicar que dichas tecnologías serán esenciales para el desarrollo de otras en el futuro. [Jurafsky 2000]

Doyle [Doyle 1997] destaca algunas de las ventajas de las interfaces de lenguaje natural. La primera es que proporcionan un vocabulario inmediato para referirse a los contenidos del ordenador. La segunda es que proporcionan un medio de acceder a información en el ordenador de forma independiente de su estructura y codificación. La tercera ventaja es que ese tipo de interfaces protege al usuario de los lenguajes formales de acceso a los sistemas subyacentes. Por último, el LN proporciona interfaces que se pueden utilizar con un mínimo de preparación.

El mismo autor indica algunas razones que dificultan la comprensión del lenguaje natural. Primero la complejidad de la representación del objeto a interpretar, que suele requerir el empleo de conocimiento adicional para extraer la información significativa. Segundo las clases de mapeos entre palabras, uno-a-uno, uno-a-muchos o muchos-a-muchos, que en casos requieren de gran cantidad de conocimiento de entorno no incluido en el elemento a interpretar. Tercero el grado de interacción entre los componentes de la representación del fuente, por el que un pequeño cambio en alguna palabra de una sentencia puede provocar un cambio de interpretación del conjunto. Después, la presencia de ruido en la entrada, que dificulta la percepción adecuada de la misma. Además, el problema del acoplamiento de los modificadores, que se pone de manifiesto porque las sentencias no son intrínsecamente jerárquicas y en ocasiones no está claro a qué se refieren sus modificadores. También, el problema del ámbito de los cuantificadores, palabras que pueden tener distintas lecturas según las circunstancias. Finalmente, las frases elípticas, en las que la interpretación de una sentencia puede depender de la interpretación de sentencias anteriores.

Actualmente, el empleo de técnicas de PLN en el ámbito de recuperación de la información (RI) es cada vez más frecuente [Baeza 2004]. Esto también ocurre en el campo de la Web Semántica, donde dichas técnicas están despertando gran interés, como se puso de manifiesto en el

Congreso Internacional de Web Semántica de 2007 [ISWC 2007], donde hubo una considerable presencia de herramientas de PLN para RI. Aunque la Web Semántica no se fundamenta directamente en técnicas de inteligencia artificial, el empleo de procedimientos de PLN y RI tienen mucho que aportar para la construcción de la nueva Web, especialmente en la transición hacia ella desde la actual, como instrumentos de anotación y mapeo automático entre anotaciones.

FreeLing [Carreras 2004] es una herramienta de procesamiento de lenguaje que permite analizar un texto para identificar las categorías gramaticales a las que pertenecen las palabras que lo forman y determinar el lema -entrada- al que corresponde cada una de dichas palabras en un diccionario de referencia. Con todo ello, la herramienta consigue establecer la interpretación morfológica más probable de las palabras del texto. Como resultado del análisis, FreeLing proporciona una versión etiquetada del texto, indicando para cada aparición de palabra su forma original junto con el lema y la correspondiente interpretación morfológica que se consideran más factibles (ver Figura 2-8).

```
...
Fundamentos fundamentos NP00000
de de SPS00
los el DA0MPO
archivos archivo NCMP000
de de SPS00
secuencias secuencia NCFP000
de de SPS00
el el DA0MS0
shell shell NP00000
Escritura escritura NP00000
de de SPS00
programas programa NCMP000
con con SPS00
el el DA0MS0
shell shell NP00000
Programación programación NP00000
controlada controlar VMP00SF
...
8 8 Z
Personalización personalización NP00000
de de SPS00
shells shells NP00000
de de SPS00
Linux_Exportación linux_exportación NP00000
de de SPS00
variables variable NCFP000
a a SPS00
el el DA0MS0
nuevo nuevo AQ0MS0
shell shell NP00000
Definición definición NP00000
de de SPS00
alias alias NCFP000
de de SPS00
comandos comando NCMP000
...
```

Figura 2-8. Ejemplo de fichero etiquetado por FreeLing [Carreras 2004]

FreeLing [FreeLing 2008] se distribuye en forma de paquete de código abierto de analizadores del lenguaje. El paquete incluye librerías para la integración de funcionalidades de análisis en el desarrollo de aplicaciones y varios diccionarios morfológicos para varios idiomas, entre los que están el inglés, el castellano, el italiano, el catalán y el gallego. La última versión incorpora en algunos de los diccionarios sentido basado en WordNet [WordNet 2006] [Fellbaum 1999], en los que las palabras se agrupan en conjuntos de sinónimos cognitivos que expresan diferentes conceptos. El paquete también incluye un analizador de demostración implementado con las propias librerías que permite utilizar la herramienta sin desarrollos complicados, de especial utilidad para prototipado de aplicaciones.

Además de FreeLing existen diversas herramientas de procesamiento del lenguaje disponibles en el ámbito académico, cómo GATE [GATE 2008] de la universidad de Sheffield, NLTK [NLTK

2008], OpenNLP [OpenNLP 2008], Stanford NLP Software [StanfordNLP 2008], WEKA [WEKA 2008] y UIMA [UIMA 2008], entre otros.

### Web 2.0

El concepto de la Web 2.0 [O'Reilly 2005] parte de una lluvia de ideas sobre el estado de la Web, realizada en 2004 por O'Reilly y MediaLive Internacional, después de la crisis de las compañías de la nueva economía vinculadas con Internet, que siguió a la llamada burbuja punto-com de finales de los noventa y que marcó los primeros años del nuevo siglo.

En los comienzos de las Web, a principios de los noventa, y antes del éxito de las empresas punto-com, la red estaba basada en páginas estáticas que no se actualizaban con mucha frecuencia y no había una idea clara de cómo hacer negocio con ella, más allá de proporcionar acceso a la misma, a esta etapa se la podría considerar como la Web 1.0. Durante la segunda mitad de los 90 y del desarrollo de la burbuja económica, la Web se llenó de páginas dinámicas que se generaban al vuelo a partir de bases de datos y en las que la estética era importante, etapa a la que algunos se refieren como Web 1.5. El modelo de negocio predominante en este periodo se basaba en facilitar servicios y contenidos que proporcionaran accesos a los sitios Web y posibles ingresos por publicidad. En el contexto de la Web 2.0 se suele hacer referencia a toda la Web anterior como 1.0.

Después de la crisis y hasta el momento actual, la Web ha tomado una nueva forma, donde ha cobrado especial valor la interacción y la colaboración, de forma que la Web se convierte en un punto de encuentro donde los contenidos los proporcionan los propios usuarios o se enriquecen con su participación, esta etapa es a la que se refiere la Web 2.0. Además, las empresas que han sobrevivido a la crisis y las que han surgido después tienen algunas características comunes en sus modelos de negocio, y en los servicios y aplicaciones que ofrecen. Para O'Reilly [O'Reilly 2005] los principios fundamentales que se deducen de la selección natural producida a lo largo de la historia de la Web son los siguientes:

1. La Web debe considerarse cómo una plataforma, donde los navegadores, los servidores y el resto de aplicaciones involucradas son fungibles, y el valor se encuentra en los servicios que se ofrece a través de ellos. De este modo, el valor de los programas es proporcional a la escala y el dinamismo de los datos que permiten gestionar. La forma en la que está constituida la Web también es fundamental, en ella es más importante aprovechar los recursos de los usuarios y encontrar el modo de alcanzar al mayor número de ellos, aunque sean pequeños y no formen parte de grupos destacados, porque el pelotón representa la masa fundamental de la red. Debido a esto, los servicios se deben plantear de tal modo que mejoren cuanto más gente los use, no al contrario.

2. Aprovechar la inteligencia colectiva, que se manifiesta en el éxito de sistemas que lo emplean: la propia Web se construye con hipervínculos que los usuarios establecen entre los recursos que en ella se publican; el directorio de enlaces de Yahoo [Yahoo 2008] es el resultado de la suma del trabajo de millones de usuarios de la Web; la innovación de Google [Google 2008] en la recuperación de información en la Web fue utilizar la propia estructura de enlaces de la red para mejorar los resultados de las búsquedas; el producto de eBay [eBay 2008] es la actividad colectiva de todos sus usuarios y su ventaja competitiva proviene de la masa crítica de compradores y vendedores global; y Amazon [Amazon 2008] fomenta la colaboración de los usuarios y utiliza la actividad de los mismos para producir mejores resultados de búsqueda y recomendación. Estos fundamentos los emplean nuevos servicios populares e innovadores de la Web 2.0 como: Wikipedia [Wikipedia 2008], que es una enciclopedia en línea elaborada entre los usuarios de la Web, empleando el sistema conocido como Wiki [Leuf 2001] que permite la edición colaborativa de un sitio Web de forma interactiva a través de un navegador convencional; Flickr [Flickr 2008] y del.icio.us [DelIcioUs 2008], que son sistemas para compartir fotografías y marcadores, respectivamente, empleando etiquetado colaborativo con palabras elegidas libremente, técnica conocida como folksonomias [Golder 2006]; Cloudmark [ClickHeat 2008], que permite el filtrado cooperativo de spam; el marketing viran, que permite la propagación de las recomendaciones de usuario a usuario; o gran parte de la infraestructura de la Web -Linux [GNU 2008], Apache [Apache 2008], MySQL [MySQL2008], etc.-, basadas en software abierto [OSI 2008], que confía en el desarrollo colectivo como soporte de proyectos de cualquier dimensión. Los weblogs [Weblog 2008] son un ejemplo especialmente significativo de aprovechamiento de la inteligencia colectiva. Se trata de páginas personales vivas con forma de diario en el que los usuarios van publicando sus opiniones. Además en ellas suele haber contenidos de otras páginas Web que se mantienen actualizados gracias la tecnología RSS [RSS 2008], si bien es posible hacer referencia permanentes, o permalinks, a partes concretas de otros Weblogs a las que se asignan URLs estáticas para poderlas apuntar siempre, aunque el sitio donde estén vaya cambiando.
3. Los datos son lo fundamental, hasta el momento todas las aplicaciones significativas de Internet han sido desplazadas por los repositorios de datos correspondientes: enlaces de Google [Google 2008], directorio de Yahoo [Yahoo 2008], productos de Amazon [Amazon 2008], vendedores y productos de eBay [eBay 2008], etc. Sin embargo, las bases de datos que se han hecho más valiosas son las que se han enriquecido con capas de valor añadidas por los usuarios. En este sentido, la introducción del modelo de

programación ligera de Google ha permitido la creación de servicios de valor añadido en forma de aplicaciones Web híbridas, o mashups, que enlazan varias fuentes de datos accesibles a través de la Web. La posesión de datos se ha convertido, por tanto, en algo esencial en este contexto, pero esto podría desembocar en el futuro en un movimiento de datos libres, parecido al de software libre actual.

4. Software como servicio, no como un producto, que ha marcado el fin del ciclo de actualización de versiones de las aplicaciones. Esto tiene importantes consecuencias en el modelo de negocio de las compañías. Por un lado, las operaciones se convierten en una competencia fundamental de la empresa, que debe preocuparse del mantenimiento diario de todo lo necesario para que el software siga funcionando. Por otro lado, los usuarios se deben tratar como parte del equipo de desarrollo, considerando las prácticas de desarrollo del software abierto, que involucra a aquellos en distintas fases del proceso, especialmente en la depuración y prueba.
5. Modelos de programación ligeros, en contraposición a los complicados servicios Web que las grandes compañías han creado para proporcionar un soporte fiable a las aplicaciones distribuidas. Los pragmáticos RSS [RSS 2008], ya mencionados, son los servicios Web más extendidos debido a su gran simplicidad. La simplicidad es lo que ha permitido que algunas interfaces simples basadas en AJAX [Woychowsky 2007] [Zakas 2007], que combina Javascript [JavaScript 1999] con XML [XML 2008], se hayan podido violar con facilidad para reutilizar sus datos en nuevas aplicaciones creativas, que han contribuido, paradójicamente, a su popularidad y difusión, como ha pasado con los mapas de Google en contraposición a los propuestos por otras compañías. De todo esto se desprende que hay que apoyar los modelos de programación ligera que permiten sistemas débilmente acoplados, que es preferible la sindicación que la coordinación entre servicios Web, y que no es mala idea programar para facilitar la reutilización de los datos en nuevos servicios remezclados.
6. Software no limitado a un solo dispositivo, es otra característica importante de las Web 2.0. Todas las aplicaciones Web cumplen esta característica de forma básica, puesto que implican al menos dos ordenadores: el cliente, donde reside el navegador; y el servidor. Sin embargo, en aplicaciones que integran servicios por varios ordenadores este esquema puede ser mucho más sofisticado, involucrando multitud de servidores y clientes de distintos tipos.
7. Experiencia enriquecedora del usuario, para superar las limitaciones de la interfaz tradicional de la Web. La Web 2.0 integra tecnologías ya existentes para dotar a las interfaces Web de prestaciones parecidas a las aplicaciones interactivas del ordenador.



El conjunto de tecnologías empleadas se le dio el nombre de AJAX [Woychowsky 2007] [Zakas 2007], que consta de: presentación basada en los estándares XHTML [XHTML 2008] y CSS [CSS 2008], visualización e interacción dinámica mediante DOM [DOM 2008], intercambio y manipulación de datos mediante XML [XML 2008] y XSLT [XSLT 2008], recuperación asíncrona de datos usando XMLHttpRequest [Woychowsky 2007] [Zakas 2007], e integración de todo ello con el lenguaje interpretado JavaScript [JavaScript 1999] que se ejecuta en el navegador Web del usuario.

Sobre las tecnologías basadas en XML -el propio XML, XHTML y XSLT- y sobre CSS se habla un poco más en el apartado que trata de la Web Semántica. El resto de tecnologías mencionadas, JavaScript, DOM, XMLHttpRequest y AJAX, se comentan brevemente a continuación, así como del formato RSS y de algunos de los sistemas mencionados de la Web 2.0 más interesantes para este trabajo, weblogs, folksonomías y wikis.

Javascript [JavaScript 1999] es un lenguaje de programación interpretado -scripting language-, que se emplea fundamentalmente en la implementación de aplicaciones Web, sobre todo en el lado del cliente, aunque también puede utilizarse en aplicaciones de servidor. Javascript tiene influencias de varios lenguajes, su sintaxis es parecida a la de C, como le ocurre al lenguaje Java del que toma parte de su nomenclatura. Sin embargo, Javascript no es un lenguaje orientado a objetos como Java y, a pesar de su nombre, no tiene una vinculación especial con él. Actualmente, la mayoría de los navegadores Web y todos los de uso más extendido interpretan el código Javascript.

Para interactuar con las páginas Web el lenguaje Javascript incorpora una implementación del modelo de objetos de documentos DOM -*Document Object Model*-[DOM 2008], promovido por el consorcio de las Web -W3C-, que permite representar los documentos estructurados, como las páginas HTML o los ficheros XML, mediante objetos que disponen de métodos y propiedades. Con ayuda de esta interfaz de programación -API, *Application Programming Interface*- para acceder a los documentos, Javascript puede cambiar los contenidos de las páginas Web cargadas en un navegador de forma dinámica. Aunque la implementación de DOM no es exactamente igual en todos los clientes Web, el soporte en los más extendidos permite funcionalidades similares en ellos. Otros lenguajes interpretados de cliente Web disponen también de interfaces de programación DOM.

XMLHttpRequest [Woychowsky 2007] [Zakas 2007] es otra API que existe para Javascript [JavaScript 1999] -y otros lenguajes interpretados de cliente Web-, que permite transferir datos en XML -y en otros formatos basados en texto-, entre clientes y servidores Web mediante protocolo HTTP de forma síncrona y asíncrona, aunque se suele emplear casi siempre de esta

segunda manera. La interfaz permite establecer múltiples conexiones simultáneas entre cliente y servidor, para intercambiar datos en cualquier momento, sin necesidad de recargar páginas completas en el navegador, lo que puede aprovecharse para crear aplicaciones Web más dinámicas y amigables.

Esto es precisamente lo que pretende AJAX -acrónimo de *Asynchronous Javascript And XML*- [Woychowsky 2007] [Zakas 2007] que es una técnica de desarrollo de aplicaciones Web interactivas, que se ejecutan en el cliente y que integran XML, XHTML, CSS, DOM, Javascript y XMLHttpRequest, para conseguir una experiencia de usuario similar a la de las aplicaciones de escritorio convencionales, sin los inconvenientes de la interfaz Web de los sitios tradicionales, caracterizada por las recargas de páginas completas y las esperas y el consumo de ancho de banda necesarios para ello. AJAX no constituye una tecnología en sí, sino que designa un modo de emplear todas las tecnologías mencionadas para conseguir un tipo de aplicaciones Web con una interfaz mejorada. AJAX se basa en estándares abiertos, es independiente de la plataforma y puede funcionar en diferentes sistemas operativos, arquitecturas de ordenadores y navegadores Web.

Un weblog [Weblog 2008] es un sitio Web que se actualiza de forma periódica donde las entradas -textos o artículos- se muestran de forma cronológica inversa y los propios autores deciden lo que se publica. Los weblogs se han hecho muy populares en los últimos tiempos en el ámbito de las aplicaciones de la Web 2.0 [O'Reilly 2005]. Los weblogs podrían verse como diarios digitales públicos, por lo que a veces se los denomina bitácoras -en referencia al cuaderno de los barcos que sirve para registrar los acontecimientos relevantes diarios-. Aunque hay weblogs que son auténticos diarios personales en línea, es muy frecuente que se dediquen a temas particulares de interés para los autores y recopilen comentarios y noticias sobre ellos. Además, es habitual que los weblogs incorporen funcionalidades interactivas para que los usuarios participen, poniendo comentarios por ejemplo, y haya vínculos con otros weblogs y sitios Web, mediante enlaces o incorporación de contenidos utilizando RSS.

RSS [RSS 2008] es una familia de formatos para la redifusión de contenidos que se actualizan frecuentemente, como titulares de noticias, entradas de Weblogs [Weblog 2008] o paquetes de archivos multimedia digitales -podcasts-, destinados a suscriptores de un sitio Web, para su uso particular o su redifusión -o publicación sindicada- en otros sitios Web. Aunque los contenidos se pueden obtener y utilizar sin necesidad de un navegador Web, mediante los llamados agregadores de contenidos o lectores de RSS, las últimas versiones de clientes Web más populares permiten utilizar estos formatos de datos directamente. Las aplicaciones que funcionan como agregadores acceden regularmente a los proveedores de contenidos -o canales- a los que están suscritos, en busca de nuevos contenidos para descargar todas las actualizaciones que encuentren. Los formatos RSS utilizan el formato XML e incluyen *Really Simple*

*Syndication* -RSS 2.0-, *RDF Site Summary* -que corresponde a las versiones 0.90 y 1.0 de RSS- y *Rich Site Summary* -RSS 0.91-. La redifusión Web basada en RSS se ha hecho muy popular por su empleo en los Weblogs, aunque no es exclusiva de este tipo de aplicaciones Web. Muchos proveedores de contenidos ofrecen servicios de RSS, puesto que la redifusión parece beneficiar a usuarios finales, intermediarios y proveedores.

Folksonomía -en inglés folksonomy- [Golder 2006] es el procedimiento colaborativo de creación y gestión de etiquetas -tags- para la anotación y clasificación de contenidos. A diferencia de los procedimientos tradicionales de indexación, en los que eran expertos los encargados de generar los descriptores y de asignarlos a los materiales, en este caso son los propios creadores y consumidores de los contenidos los que se ocupan de dichas tareas. Para ello los implicados eligen palabras claves libremente del lenguaje natural, en lugar de seleccionarlas de un vocabulario controlado. Como resultado de la anotación los contenidos quedan calificados con palabras claves que permiten clasificarlos en una taxonomía plana, que no está estructurada en niveles. Las folksonomías son un tipo de aplicaciones sociales que se han popularizado en los últimos años dentro del contexto de la Web 2.0 [O'Reilly 2005]. Esta modalidad colaborativa de gestión de contenidos se emplea con distintos fines, que van desde anotar y compartir fotografías [Flickr 2008] hasta clasificar y poner en común direcciones de Internet [DelIcioUs 2008], pero también lugares [Tagzalia 2008] e incluso deseos [43Things 2008].

Wiki [Leuf 2001] es el termino que designa a un tipo de aplicación Web, que permite a sus usuarios crear, editar, enlazar y publicar páginas Web de forma fácil, colaborativa y directa, a través de la propia Web empleado un navegador convencional. A los sitios Web construidos por este procedimiento se les llama wikis y a las páginas, enlazadas entre sí, que los constituyen páginas wiki. Las wikis están en permanente edición, puesto que sus usuarios pueden modificarlas en cualquier momento. Según Leuf y Cunningham las características esenciales de la wiki son las siguientes [Leuf 2001]: por un lado, las wikis invitan a sus usuarios a crear nuevas páginas wiki o editar las existentes en sus correspondientes sitios Web, empleando sólo un navegador Web estándar sin ningún tipo de complemento extra; por otro lado, las wikis promueven las asociaciones de temas significativos entre páginas wiki diferentes, haciendo la creación de enlaces entre ellas algo fácil e intuitivo, y mostrando si una página pretendida de destino existe o no; además, las wikis no son sitios Web cuidadosamente terminados para usuarios ocasionales, si no que pretenden involucrar al visitante en un proceso en curso de creación y de colaboración que cambia constantemente el correspondiente sitio Web. Por otro lado, las wikis permiten la escritura colaborativa de documentos mediante un lenguaje de marcado -etiquetas o tags- sencillo al que se conoce como wikitext. Desde otro punto de vista las wikis pueden verse como bases de datos, gestionadas colaborativamente, que permiten crear,

revisar, buscar y modificar información. Los sistemas wiki disponen habitualmente de un mecanismo de control de cambios, que permite a los usuarios seguir las modificaciones realizados a lo largo del tiempo, para facilitar las rectificaciones imprescindibles en un sistema tan abierto, que no suelen restringir las posibilidades de edición a los usuarios. Existen wikis dedicadas a temas diversos, como definiciones de palabras (Wikitionary), textos originales (Wikisource), libros (Wikibooks), citas (Wikiquote) o noticias (Wikinews), pero la más popular es la enciclopedia Wikipedia [Wikipedia 2008], aunque cualquiera puede crear una sobre un tema de su interés.

### **2.2.3 Web Semántica**

Este apartado está constituido por varios bloques dedicados a los aspectos más relevantes de la Web Semántica relacionados con la tesis. En primer lugar, se presentan los fundamentos de la misma, mostrando su vinculación con la Web actual. En segundo lugar, se introduce el concepto de Ontología, elemento esencial de la nueva Web. A continuación se describe la torre de lenguajes que aspiran a sustentar la Web Semántica sobre la Web actual. Después se tratan algunos temas relativos a las ontologías y que se refieren a su definición, implementación y explotación. Por último, se habla de los Agentes Software, dedicando atención a la definición de estos elementos, a los Sistemas Multi-Agente y a cuestiones relacionadas con su implementación.

La Web Semántica [Berners-Lee 2001] es una propuesta de evolución de la Web actual, constituida principalmente por documentos destinados al entendimiento humanos, hacia otra Web, que incorpore información adecuada para que los ordenadores puedan tratarla automáticamente. De esta forma la Web Semántica podría verse cómo una red global de información enlazada de modo que las máquinas puedan procesarla fácilmente, pero también cómo un modo eficiente de representar datos en la Web, o cómo una Base de Datos enlazada de forma global [Palmer 2001].

La idea de la Web Semántica parte de Beemers-Lee, el creador de la Web, las URIs, el protocolo HTTP (Hypertext Transfer Protocol) y el lenguaje HTML (Hypertext Markup Language). Cuenta con un equipo dedicado a ella en el Consorcio de la Web (W3C) [W3C 2008], que trabaja en mejorar, extender y estandarizar el sistema, propiciando el desarrollo de especificaciones, lenguajes y herramientas, así como en la publicación y divulgación de todo ello. [Needleman 2003].

La Web Semántica constituye un gran desafío tecnológico, aunque la finalidad del proyecto y su futuro parecen claros en general, no resulta siempre fácil lograr el consenso sobre el modo de afrontar los retos para alcanzar los objetivos marcados. Todavía no están maduras todas las tecnologías necesarias para soportar completamente la Web Semántica, ni tampoco está

disponible la mayoría de la información que será necesaria para su funcionamiento. Sin embargo, el camino se va aclarando en ambos sentidos, para conseguir llegar a la Web Semántica desde la Web actual, que se considera el punto de partida ineludible.

### *Hacia la Web Semántica*

Para trazar una trayectoria desde la Web actual hasta la Web Semántica es necesario considerar el punto de partida. Hoy en día la mayoría de los datos existentes en la Web están ocultos en ficheros HTML, adecuados para las personas pero no para explotarlos automáticamente de forma generalizada.

El lenguaje HTML está inspirado en la norma SGML (*Standard Generalized Markup Language*) que surge de las investigaciones de IBM, durante la década de los 60, sobre el modo de gestionar documentos de forma automática [Goldfarb 1999]. La idea del marcado consiste en incorporar información sobre los documentos dentro de ellos mismos. De esta forma se puede establecer el modo de presentación de un documento, pero también especificar otras muchas cuestiones sobre él, como su historia, su estructura o las palabras claves que lo describen. Por tanto, el lenguaje HTML es una aplicación de SGML para documentos de hipertexto. Aunque permite incorporar cierta información adicional sobre los documentos, su objetivo principal es describir, de modo básico, la estructura y el formato de los mismos. Una de las limitaciones más evidentes de las primeras versiones del lenguaje era la rigidez con que se manejaban los formatos.

El éxito de la Web y los requerimientos de presentaciones mejoradas hicieron que el W3C promoviera un nuevo lenguaje para adjuntar formatos flexibles a los documentos HTML: las hojas de estilos en cascada CSS (*Cascading Style Sheets*) [CSS 2008]. Para ello, se extendió el lenguaje HTML con un mecanismo sencillo que soportaba la incorporación de los nuevos estilos, definidos por los diseñadores de cada sitio Web, de manera compatible con los formatos fijos predefinidos en el lenguaje.

Sin embargo, esta solución no dio respuesta a todas las necesidades que la evolución de las Web originaba, no se trataba sólo de un problema de presentación de los datos, que es lo que HTML hace fundamentalmente, sino también de describirlos suficientemente. Fue entonces cuando el W3C desarrolló una versión simplificada de la norma SGML, que se conoce como XML (*Extensible Markup Language*) [XML 2008].

El nuevo lenguaje preservaba las principales ventajas del SGML a la vez que eliminaba su inconveniente más destacado, la complejidad. XML no pretende sustituir a HTML, puesto que en determinados contextos necesita a éste para presentar los datos que él describe. Sin embargo, XML aspira a ser el lenguaje de descripción de datos de la nueva Web, y esto conlleva el que todo en ella se escriba o rescriba en XML. De esta forma HTML se puede reformular en XML,

el resultado de ello es XHTML (*eXtensible HTML*) [XHTML 2008], que es el lenguaje llamado a remplazar a HTML en la Web que viene.

En XML es posible definir nuevas etiquetas y atributos, para ello se utiliza la definición de tipo de documentos o DTD (*Document Type Definition*). La DTD no está escrita en XML, sino en SGML lo que complica su creación, además de no seguir la filosofía de la nueva Web. Por esta razón se ha introducido más adelante la especificación XMLS (*XML Schema*), que permite definir los documentos XML utilizando la propia sintaxis XML.

### Ontologías

Como se anticipó en la introducción de este documento, la necesidad de reutilizar y compartir el conocimiento ha motivado la búsqueda de representaciones adecuadas del mismo. El uso de ontologías es una posibilidad con bastante aceptación en los últimos años. El concepto de ontología proviene de la metafísica y se refiere a la rama de ésta que se dedica al estudio sistemático de la existencia, el concepto del ser, sus modos, sus principios, sus propiedades, sus divisiones y sus causas.

Sin embargo, en el ámbito de sistemas basados en conocimiento, una ontología es una entidad computacional, un recurso que contiene el conocimiento sobre qué conceptos existen en el mundo y cómo se relacionan entre ellos [Mahesh 1996]. En este sentido una ontología es un recurso artificial que se crea para representar de forma simplificada la realidad y así poder trabajar con ella.

En este mismo contexto se enmarca la concisa y generalmente aceptada definición de Gruber, donde se dice que una ontología es una representación conceptual compartida del conocimiento, hecha de un modo explícito y formal [Gruber 1993]. En esta definición, el término conceptual se refiere a un modelo abstracto de cómo se conciben las cosas en el área de trabajo y el término explícito significa que los conceptos y relaciones del modelo abstracto se expresarán de forma clara y evidente mediante términos y definiciones [Gruninger 2002]. Considerando todo lo dicho, una ontología debe concebirse como un entendimiento común y compartido de un dominio, que puede comunicarse entre personas y sistemas informáticos [Pérez 2002].

Chandrasekaran aporta algunas consideraciones interesantes sobre las ontologías. Por un lado, afirma que el análisis ontológico clarifica la estructura del conocimiento de los dominios tratados, a la vez que apoya la idea de que las ontologías permiten que ese conocimiento se pueda compartir. Y por otro lado, observa que aunque los elementos que conforman lo que existe no dependen de los objetivos de los diseñadores de las ontologías, normalmente los aspectos que se eligen de esos elementos para configurar las ontologías suelen depender de tales objetivos y por lo tanto también lo hacen las ontologías resultantes. De este modo, opina que es poco probable que una ontología pueda servir para todos los posibles usos potenciales; es decir,

una ontología de un dominio y el conocimiento basado en ella, serán más apropiados para unos usos que para otros y no se podrán aplicar a tareas muy diferentes de aquellas que los motivaron [Chandrasekaran 1999].

Gruninger afirma en este sentido que aunque el uso de ontologías se debe originalmente a la necesidad de compartir y reutilizar el conocimiento, la posibilidad de hacer ambas cosas está limitada porque los usuarios de ontologías (y otros diseñadores) no siempre comparten las mismas presunciones que los diseñadores originales. También observa que es difícil para los usuarios identificar cuáles son las presunciones implícitas y entender las diferencias clave en la ontología, aunque reconoce que algunos desacuerdos entre las personas son sólo superficiales [Gruninger 2002].

Se pueden considerar cuatro tipos fundamentales de ontologías [Pérez 2002]:

- Ontologías de tarea [Guarino 1998], que son las que han sido creadas para una aplicación concreta y las diseñadas para una actividad o tarea específica, como la venta de productos o el diagnóstico de una enfermedad.
- Ontologías de un dominio [Steve 1998], en las que se representa el conocimiento especializado pertinente de un dominio o subdominio, como la medicina, las aplicaciones militares, la cardiología, etc.
- Ontologías genéricas [Steve 1998], en las que se representan conceptos generales y fundacionales del conocimiento como las estructuras parte/todo, la cuantificación, los procesos o los tipos de objetos.
- Ontologías de representación [Steve 1998], en las que se especifican las conceptualizaciones que subyacen a los formalismos de representación del conocimiento, por lo que también se denominan meta-ontologías.

Es posible distinguir varias aplicaciones de las ontologías [Gruninger 2002]:

- Para comunicación, entre sistemas informáticos, entre humanos o entre humanos y sistemas informáticos.
- Para inferencia computacional, para internamente representar y manipular planes y planificar información, y para analizar las estructuras internas, algoritmos, entradas y salidas de sistemas implementados en términos teóricos y conceptuales.
- Para reutilización (y organización) del conocimiento, para estructurar u organizar librerías o repositorios de planes y planificar la información de un dominio.

Existen distintas formas de aplicar las ontologías para la gestión del conocimiento en intranets e Internet, por ejemplo emplearlas como soporte para la memoria corporativa centralizada de una organización, para la interoperatividad entre bases de datos, para la automatización de la

generación de documentos con algún propósito particular o para construir, mantener e interrogar el conocimiento distribuido de una organización [Benjamins 1999].

Relacionado con este asunto está el desarrollo de vocabularios ontológicos, que se ha producido de forma paralela y gozan de un amplio apoyo. Se trata de juegos de términos que se emplean para referirse a los conceptos que manejan las ontologías. Algunos son de propósito general como Dublin Core (“<http://dublincore.org>”), y otros de carácter específico, como Chemdex (“<http://www.chemdex.com>”) para comercio electrónico entre compañías, WordNet [WordNet 2008] [Fellbaum 1999] para procesamiento del lenguaje natural o IMS (Instructional Metadata System, “<http://www.imsproject.org>”) y EML (Educational Modeling Language, “<http://eml.ou.nl>”) para temas de formación.

#### Separación de Presentación y Contenido

Una propuesta interesante de la Web Semántica es la de separar el contenido y la presentación en los documentos que maneja. De este planteamiento viene la idea de documento virtual [Milosavljevic 1999], que consiste en un documento que no existe como unidad integrada en ninguna URI (Uniform Resource Identification) y se compone dinámicamente combinando información que puede estar dispersa por la red. Este enfoque facilita la conceptualización, procesamiento y recuperación de la información que compone los documentos, que además pueden adaptarse a los requerimientos de las interfaces y condiciones de uso. Esta posibilidad permite integrar en la Web Semántica propuestas de la Hipermedia Adaptativa [Brusilovsky 1998], de la que se habló en el apartado de Interacción Persona-Ordenador, que constituyen un antecedente en el que las ontologías subyacentes suelen estar implícitas en el sistema y no se pueden modificar. Aunque la Web Semántica comparte con la Hipermedia Adaptativa algunas áreas de intersección, el ámbito de estudio de la primera se limita a los sistemas y aplicaciones que emplean la Web, mientras que campo de interés de la segunda tiene mayor recorrido y otras áreas de aplicación, que en los últimos años se ha orientado también a las aplicaciones y a los sistemas Web.

Para la generación dinámica de documentos en la Web existen actualmente multitud de alternativas –estándares, lenguajes y tecnologías-, todas ellas con ventajas e inconvenientes. Entre los que permiten la generación de los documentos en el servidor pueden citarse CGIs (Common Gateway Interface), C, Perl, ASP (Active Server Page) –con Visual Basic, Javascript-, Servlets, JSP (Java Server Page) –con Java-, PHP (inicialmente deriva de Personal Home Page Tools), Servidores de Aplicaciones, Bases de Datos. Y entre los que generan los documentos en el cliente: Javascript y Java. Fuera de estas dos categorías están por ejemplo XSL (eXtensible Stylesheet Language) y XML (Extensible Markup Language) [XML 2008].



### *Torre de Lenguajes para la Web Semántica*

Berners-Lee propuso una torre de lenguajes para sustentar la Web Semántica [Berners-Lee 2000], en ella los lenguajes de cada nivel están soportados por los de niveles inferiores para proporcionar nuevas capacidades:

- En el nivel más bajo, URIs (Uniform Resource Identifier) [Shadbolt 2006] y Unicode (Unicode Worldwide Character Standard) proporcionan una forma estándar de definir referencias a entidades y símbolos de intercambio.
- En el segundo nivel, XML (Extended Markup Language) [XML 2008] proporciona un modo estándar de representar árboles etiquetados y XML Schema la forma de definir gramáticas para crear documentos XML legales. También en este nivel, se facilita el mecanismo de espacio de nombres (NS) para combinar documentos con vocabularios heterogéneos. Hasta aquí se proporciona el soporte sintáctico para la Web Semántica.
- En el tercer nivel, que es realmente el primero de la Web Semántica, RDF (Resource Description Framework) [RDF 2008] aporta la base para el proceso de meta-datos, que proporciona la posibilidad de interacción entre aplicaciones sobre la Web, intercambiando información semántica comprensible para las máquinas [W3C 1999]. RDF Schema define un lenguaje de modelado minimalista sobre RDF, introduciendo clases, relaciones entre clases y propiedades, y restricciones de dominio y rango como primitivas de modelado.
- En el cuarto nivel, el de Ontología, el lenguaje OWL (de Web Ontology Language, a pesar del cambio de orden en las siglas) [OWL 2008] proporciona un modo común de procesar el contenido semántico de la información de la Web, para ello el lenguaje permite definir ontologías e instancias de las mismas. Dicho lenguaje ha conseguido la convergencia entre las distintas iniciativas preexistentes -OIL (Ontology Inference Layer) y DML (Data Manipulation Language)- y está construido sobre RDF y RDF Schema.
- El nivel Lógico es el quinto y pretende proporcionar un lenguaje universal para lógica monótona, superando las limitaciones de intercambio entre sistemas de reglas y la diversidad de motores existentes. Aún se encuentra en fase de desarrollo.
- El siguiente nivel es el de Prueba, que todavía se encuentra en una fase de definición muy preliminar. La idea es proporcionar un lenguaje que permita explicar por qué se ha llegado a una conclusión o se ha inferido un cierto conocimiento.
- Sobre todos los demás está el nivel de Confianza. Como el anterior se encuentra todavía en un estado embrionario, su intención es establecer una medida de la credibilidad del conocimiento que se maneja, por ejemplo, una afirmación sobre cualquier tema no

merece la misma fiabilidad si proviene de un experto que si lo hace de otra persona cualquiera.

Por lo tanto, la Web Semántica se construye sobre la Web actual utilizando XML y su familia de tecnologías. Esta familia se compone de tres grupos fundamentales: el de las tecnologías para diseñar y crear los lenguajes basados en XML y sus documentos; el de las que permiten el acceso, manipulación transformación y análisis de documentos XML; y el de las que permiten añadir una jerarquía de significados –metadatos- a los datos y recursos de la Web [Geroimenko 2002].

En el primer grupo, el de las tecnologías para diseñar y crear los lenguajes basados en XML [XML 2008], se incluyen, entre otros: XML, el metalenguaje que permite la creación de lenguajes de marcado para cualquier dominio y propósito; XML Namespace, con el que se previene la colisión de nombres en los documentos XML empleando nombres cualificados, que consisten en un espacio de nombres, que es una URI (Uniform Resource Identification), y un identificador local.; XML Schema, con el que se pueden definir documentos XML y poder crearlos y validarlos; XPath, que proporciona el modo de cómo apuntar a partes de un documento XML; XInclude, establece el modo de incluir en ficheros XML otros ficheros o partes de ellos; XPointer, que extiende XPath definiendo identificaciones de fragmentos para referencias URI; XLink, que facilita el establecimiento de enlaces entre documentos XML, incluyendo enlaces de doble sentido, múltiples y variantes más sofisticadas que las básicas existentes en HTML; XForms, que especifica el uso de técnicas Web en diversas plataformas, como televisiones o teléfonos móviles; XSignature, que proporciona sintaxis y reglas de procesamiento para firma digital; y XML Encryption, que define la sintaxis y las reglas de proceso para encriptar contenidos en XML.

En el segundo grupo, el de las que permiten el acceso, manipulación transformación y análisis de documentos XML, se encuentran: DOM (Document Object Model) [DOM 2008], que es una interfaz de programación de aplicaciones que describe un documento XML como un árbol de nodos, y define el modo en que los nodos se estructuran, acceden y manipulan; XSL (eXtensible Stylesheet Language), que es un lenguaje para transformar documentos XML en otros documentos XML; XQuery, que es un lenguaje para interrogar datos almacenados en formato XML como se hace con las bases de datos; CSS (Cascading Style Sheets) [CSS 2008], que es un lenguaje para especificar hojas de estilo que puede utilizarse para dar formato a documentos XML; XHTML (eXtensible HTML) [XHTML 2008], que es la reformulación de HTML en XML; SVG (Scalable Vector Graphics), que es un lenguaje para describir vectores de dos dimensiones y combinaciones de vectores y mapas de bits en XML; SMIL (Synchronized Multimedia Integration Language), que permite crear presentaciones multimedia en la Web, integrando, sincronizando y enlazando elementos multimedia independientes; X3D (eXtensible

3D), que es un lenguaje de marcado que permite expresar contenidos VRML (Virtual Reality Markup Language) en XML; WML (Wireless Markup Language), que es un lenguaje para presentar contenidos en teléfonos móviles y dispositivos asistentes personales al estilo de las páginas Web; y MathML (Mathematical Markup Language), que permite la representación de fórmulas matemáticas.

Y en el tercer grupo, el de las tecnologías que permiten añadir una jerarquía de significados a los datos y recursos de la Web, están: RDF (Resource Description Framework) [RDF 2008], que constituye el primer nivel de la nueva Web y proporciona el medio de definir modelos de datos de forma sencilla; RDF Schema, que es la tecnología clave para describir características básicas de un modelo RDF; OWL (Web Ontology Language), lenguaje para la expresión de ontologías en la Web que extienden RDF Schema y es heredero de OIL (Ontology Interface Layer) y DAML+OIL (DARPA Agent Markup Language + OIL); o Topic Maps, que es un estándar escrito en XML que permite definir redes semánticas sobre recursos de información, de forma parecida a como lo hacen RDF y RDF Schema.

Conviene señalar que RDF emplea un modelo de datos que incluye tres tipos de objetos [Broekstra 2002]: recursos, que pueden ser prácticamente cualquier cosa que se pueda referenciar con una URI (un conjunto de páginas Web, una de ellas o una parte de esta, pero también cualquier objeto aunque no sea accesible a través de la red); propiedades, que pueden ser un aspecto específico, característica, atributo o relación usada para describir un recurso; y sentencias, que permiten establecer el valor de una propiedad para un determinado recurso. Estos tres objetos constituyen respectivamente el sujeto, predicado y objeto de las sentencias RDF y son los componentes de las triplas que pueden utilizarse para representarlas. Se pueden utilizar distintas sintaxis para escribir RDF, como pueden ser XML o las simples triplas. Cada una tiene sus ventajas e inconvenientes y la elección de la más adecuada depende del uso, aunque la sintaxis empleada no suele ser lo más importante [Lassila 2000].

Desde 2004 se dispone de especificaciones sólidas de RDF -y RDF Schema-, con semántica formal bien definidas y sintaxis RDF/XML clara [Herman 2007]. Actualmente existe una amplia lista de recursos y herramientas disponibles para el empleo de RDF [Shadbolt 2006], como entornos de programación para la mayoría de los lenguajes de uso actual, sistemas de almacenamiento de triplas de RDF -algunos con capacidad para manejar conjuntos de datos muy grandes-, conversores entre RDF y distintos formatos o una gran cantidad de documentación, de material formativo y de datos. Además, existe una comunidad muy activa de desarrolladores.

De igual forma, desde 2004 se dispone de una especificación estable de OWL [Herman 2007]. Se han definido tres niveles distintos en el lenguaje [Shadbolt 2006], para regular las

capacidades y los requerimientos de la implementación según las necesidades de expresividad de las aplicaciones [OWL 2008]: OWL-Lite, es el nivel más sencillo de los tres que proporciona soporte para clasificación jerárquica y restricciones simples; OWL\_DL, es un nivel intermedio en el que se soporta la máxima expresividad posible para garantizar que todas las conclusiones sean computables en un tiempo limitado; y OWL\_Full, es el nivel que proporciona la máxima expresividad y libertad sintáctica de RDF, pero sin garantías computacionales. Cada lenguaje de la familia OWL es una extensión del de nivel anterior, de forma que las ontologías y las conclusiones válidas en los niveles inferiores son también válidas en los superiores. Además, en la actualidad existe un cierto volumen de herramientas y recursos disponibles para el soporte de OWL [Herman 2007], como motores de razonamiento OWL autónomos o incluidos en entornos de programación o editores de ontologías. Sin embargo, todavía es necesario que aumente el número de estos recursos para soportar su producción y aplicación [Shadbolt 2006].

Por otro lado, se están desarrollados grandes ontologías, directamente en OWL o a partir de otras definidas en diferentes formatos. Entre dichas ontologías pueden citarse [Herman 2007]: eClassOwl, ontología de productos y servicios de negocio electrónico; Gene Ontology, para describir genes y los atributos de los productos de genes en distintas organizaciones; BioPAX, para datos de senderos biológicos; o UniProt, para secuencias de proteínas y datos y terminología de anotación. Aunque normalmente se suele hablar de ontologías como las enumeradas, de carácter muy formal y serio, que acostumbran a ser complicadas y estar relacionadas con la ciencia y la tecnología, también son muy importantes las ontologías de tipo más ligero y superficial [Shadbolt 2006], que aparecen en otros contextos como los negocios, para organizar con esquemas simples grandes cantidades de datos.

Relacionado con la penetración de las ontologías en la Web está el fenómeno de las Folksonomias [Golder 2006], ya mencionadas antes, que son el resultado del etiquetado colaborativo de recursos accesibles a través de la Web, por grupos numerosos de personas de forma espontánea, sin etiquetas preestablecidas, y que forman parte de aplicaciones de software social encuadradas en el contexto de la Web 2.0 [O'Reilly 2005], a la que se dedica un apartado en este documento. Algunos ejemplos de este tipo de aplicaciones [Shadbolt 2006] pueden ser Flickr [Flickr 2008], para clasificar y compartir fotografías, o "del.icio.us" [DelIcioUs 2008], que permite publicar y compartir referencias de direcciones favoritas en la red.

Además, ya existen un gran número de vocabularios esenciales -no todos ellos basados en OWL-, entre ellos están [Herman 2007]: SKOS, sobre sistemas de conocimiento, tesauros y glosarios; Dublín Core, sobre recursos de información librerías digitales, con extensiones para derechos, permisos y tratamiento de derechos digitales; FOAF, sobre personas y organizaciones; DOAP, para la descripción de proyectos software; Music Ontology, para la descripción de CDs

y temas musicales; o SIOC, para la descripción de comunidades en línea enlazadas semánticamente.

Considerando todo lo dicho, puede concluirse que los primeros cuatro niveles de la Web Semántica se encuentran bastante consolidados en la actualidad. Los dos primeros ya habían ganado su respaldo en el ámbito de la Web actual. El tercero, RDF/RDFS, ha conseguido arraigarse, superando algunos detractores sobre la adecuación de las primeras propuestas [Patel 2002 a,b] [Broekstra 2002], gracias a la aceptación de los desarrolladores y a la creciente disponibilidad de recursos. El cuarto, el de Ontología, se encuentra en este momento en fase de consolidación, una vez se han hecho converger en OWL las propuestas iniciales, OIL [Fensel 2000] [Harmelen 2000] y DAML [Hendler 2000], y se van desarrollando herramienta para su aplicación, así como ontologías y vocabularios adecuados para empleo en diversos campos. El quinto nivel, Lógico, esta actualmente en la etapa de desarrollo. El resto de niveles están, como ya se ha dicho, en fases de definición.

No conviene olvidar, sin embargo, que desde la introducción de la idea de la Web Semántica [Berners-Lee 2001] se ha realizado una gran cantidad de investigaciones en el terreno de los lenguajes de marcado semántico. Desde distintos foros, el W3C, agencias en Europa y Estados Unidos, y diversas comunidades de investigación han efectuado numerosas propuestas de lenguajes para la Web Semántica que no siempre han trascendido. Algunos de estos lenguajes se han adoptado directamente, otros se han ampliado o integrado con terceros y el resto se han terminado desechando salvo, quizás, por algunas ideas claves [Gil 2002].

#### Ingeniería ontológica

Otro tema importante directamente relacionado con las ontologías y, por lo tanto con la Web Semántica, es el de la ingeniería ontológica. Pretende el soporte del desarrollo y uso de ontologías a través de su ciclo de vida, diseño, evaluación, validación, revisión, mapeo, integración, compartición y reutilización. Construir ontologías es difícil y costoso en recursos y tiempo, especialmente si el objetivo es el diseño de una ontología lo bastante formal como para soportar inferencia automática. Una de las razones de este fenómeno es que las ontologías requieren del consenso de una comunidad cuyos miembros pueden tener una visión radicalmente distinta del dominio considerado. La búsqueda del consenso se puede tratar de varias formas. En un extremo, las pequeñas ontologías ligeras pueden desarrollarse por un gran número de personas y luego integrarse. En el otro, las ontologías de formalismo riguroso pueden desarrollarse por consorcios y organizaciones de estandarización. En el primer caso habrá una gran necesidad de soporte para el mapeo y unión de ontologías, mientras en el segundo se requerirá mejor infraestructura para diseño y análisis colaborativos [Gruninger 2002].

Para proponer una ontología consensuada de algún dominio, es importante que la gente que tiene que usarla tenga una actitud positiva hacia ella. Imponer el uso de una ontología particular a gente que no ha contribuido en su definición no es probable que tenga éxito. Preferiblemente, una ontología se construirá con la colaboración de expertos en el dominio, representantes de los usuarios finales y especialistas en las tecnologías de la información utilizadas. Semejante esfuerzo de cooperación requiere del uso de una metodología que guíe el proceso de desarrollo de la ontología y de herramientas para acceder, examinar, codificar, modificar y descargar la ontología [Benjamins 1999].

Respecto al desarrollo de la ontología, Motta propone el empleo de una estrategia de análisis del-centro-hacia-afuera, en lugar de las alternativas de-arriba-a-abajo o de-abajo-a-arriba seguidas por otros autores. Considera que esta alternativa es más efectiva al permitir identificar primero los conceptos básicos del dominio para después generalizarlos o particularizarlos si se considera oportuno. También en relación con este asunto, se opta por que un equipo muy reducido de expertos se encargue del diseño de la ontología, en lugar de hacerlo de forma colaborativa con un grupo mayor y menos cualificado. Se llama la atención sobre el hecho de que los pequeños errores e inconsistencias en el diseño pueden hacer fracasar el sistema resultante [Motta 2000].

Stojanovic hace algunas consideraciones interesantes sobre la evolución de las ontologías en [Stojanovic 2002]. En su opinión las ontologías evolucionan por varias razones: los errores de diseño que incorporan, las modificaciones en el entorno donde operan que pueden invalidar asunciones hechas durante el diseño, y los cambios en las necesidades de los usuarios después de la definición inicial de la ontología. Por tanto las ontologías son entidades en evolución que además de su diseño inicial requieren de un permanente mantenimiento.

Dada la importancia que las ontologías van tomando en diferentes campos, Duineveld habla de la necesidad de disponer de herramientas que ayuden en la creación y mantenimiento de dichas ontologías [Duineveld 2000]. Esto se ha hecho más evidente con el auge de las ontologías en el ámbito de la Web Semántica. Los criterios que pueden guiar la elección de un editor de ontologías son muy diversos y numerosos [Denny 2004], pero pueden citarse tres criterios fundamentales [OntologyEd 2008]: el primero sería considerar el grado de abstracción que el editor proporciona respecto al lenguaje de representación ontológico empleado y las posibilidades de navegación visual en el modelo de conocimiento que brinda; el segundo criterio es disponer de motores de inferencia y de extracción de información incorporados en el sistema, así como de soporte para ontologías de alto nivel; por último, el tercer criterio sería la capacidad para importar y exportar lenguajes ajenos de representación de conocimiento para mapeo entre ontologías.

La lista de editores de ontologías es muy grande y es previsible que crezca a medida que el empleo de ontologías aumente, Denny [Denny 2004] proporciona una tabla muy amplia de herramientas de edición con detalles de sus características. Para la realización del trabajo que nos ocupa se ha empleado la herramienta Protégé [Protege 2008] de la universidad de Stanford, que es un editor ontológico y un gestor de bases de conocimiento.

#### Mapero entre Ontologías

La necesidad del mapeo entre ontologías surge del requerimiento de utilizar más de una ontología simultáneamente en entornos distribuidos y heterogéneos, cómo la Web y la Web Semántica, donde diferentes sistemas acceden a múltiples ontologías que consideran de formas distintas los dominios y sus descripciones. Noy presenta en [Noy 2002] las bases de un proyecto de evaluación de herramientas para el mapeo de ontologías. En este estudio preliminar se pone de manifiesto la problemática del proceso de mapeo y los diferentes modos en que se ha intentado resolver. Por un lado, del análisis de Noy se desprende la dificultad de los expertos humanos para llegar a un acuerdo sobre el modo de establecer mapeos entre ontologías, aunque se refieran exactamente a los mismos dominios, por la diferencia de criterios de unas y otras personas. Por otro lado destaca, la diversidad de técnicas empleadas por las herramientas de mapeo para intentar ayudar en el proceso de comparación de las ontologías.

Wiesman plantea en [Wiesman 2001] un procedimiento para mapeo automático de ontologías basado en asociación de nodos por identificación de palabras en la información clasificada en ellos. En realidad lo que Wiesman intenta es resolver un problema de interoperatividad entre agentes que utilizan ontologías para representar el significado de la información que pretenden comunicarse. El origen del problema es la falta de estandarización en la representación de los contenidos derivada de la dificultad para establecer una ontología global, si esto fuera posible.

Wiesman analiza el problema de la interoperatividad, que es un problema de mapeo entre ontologías. En primer lugar observa la posible existencia de estructura y semántica heterogéneas. La primera se refiere a la diferencia en la representación de la información, que es un problema entre bases de datos pero no entre agentes que enmascaran esta cuestión. La segunda se refiere a los distintos significados que se pueden asignar a la información descrita, que si es un problema para la comunicación entre agentes.

En este último caso, semántica heterogénea, Wiesman identifica las siguientes diferencias entre ontologías: estructurales, en la forma de organizar la semántica; nomenclatura, distintos nombres para referirse a la misma información, sinonimia, o el mismo nombre para informaciones distintas, polisemia; y representación, los mismos datos reflejados de formas distintas. Este último puede refinarse en diferencias en unidades, precisión y expresión. Para ilustrar esta clasificación presenta dos ontologías que describen datos de una persona. En ellas

se pueden observar problemas de estructura, mientras en una el campo *street* recoge número en la otra no, problemas de nomenclatura, mientras en una se refiere al nombre de pila como *first name* en la otra lo hace como *christian name*, etc.

Para solucionar el problema, cuando al comunicarse dos agentes se dan cuenta de que no están utilizando la misma ontología y no saben mapearlas, Wiesman propone establecer un procedimiento, valiéndose de las palabras clasificadas en las ontologías de cada agente y ciertas reglas de combinación dependientes del dominio y el contexto de trabajo, para intentar identificar información clasificada en una ontología en la otra y establecer así relaciones entre sus nodos.

Para establecer un mapeo entre repositorios XML de información Amann propone un procedimiento que emplea una ontología para representar el dominio de conocimiento al que la información de los repositorios se refiere [Amann 2002]. En este contexto se plantean dos posibles aproximaciones: una considera la estructura de la información de los repositorios como distintas vistas de un dominio global representado por una ontología (*local as a view*), mientras la otra considera la representación de ese dominio general como la unión de las estructuras de los distintos repositorios (*global as a view*). Esta segunda alternativa se emplea en algunos proyectos que pretenden mapear repositorios relacionales, sin embargo en la propuesta de Amann que considera repositorios XML se considera mejor la primera aproximación.

Según Choi [Choi 2006] el mapeo de ontologías se necesita en los tres procesos principales que se realizan con ellas. El primer proceso es la integración de ontologías, por el que se genera una nueva ontología en un dominio a partir de dos o más ontologías en dominios diferentes, pero relacionados. El segundo es la fusión de ontologías, que es el proceso de generar una nueva ontología coherente a partir de dos o más ontologías preexistentes con el mismo tema. El tercer proceso es la alineación de ontologías, que es el mecanismo de establecer enlaces entre dos ontologías originales.

Dicho autor clasifica el mapeo entre ontologías en tres categorías. La primera es la de mapeo entre una ontología global integrada y ontologías locales, que se emplea para identificar un concepto encontrado en una ontología en una vista, o para realizar una pregunta sobre otras ontologías. Esta clase de mapeo soporta la integración de ontologías en la Web Semántica, en los sistemas de gestión de conocimiento y de integración de datos e información. La segunda categoría es la de mapeo entre ontologías locales, que permite transformar las entidades de una ontología fuente en las entidades de una ontología destino en base a relaciones semánticas. La principal aplicación de este tipo de mapeo está en la Web y la Web Semántica, que comparten una naturaleza descentralizada donde no hay una ontología global y es preciso el mapeo entre las ontologías locales para hacer posible la interoperación con ellas. La tercera categoría es la de



mapeo en fusión y alineación de ontologías, que sirve para establecer correspondencias entre ontologías fuentes que se van a fusionar o alinear, y determinar el conjunto de conceptos solapados, sinónimos o conceptos únicos de las fuentes. Esta clase de mapeo se emplea en la resolución de solapamientos entre ontologías, que es cada vez más frecuente por la proliferación de ontologías dispersas en diferentes contextos.

Tabla 2-8. Características de algunas herramientas de mapeo entre ontologías [Choi 2006]

	Entrada	Salida	Interacción Usuario	Estrategia de Mapeo o Algoritmo	Conocimiento			
					Estructurado	Basado Instancia	Léxico	Dominio
<b>MOMIS</b>	Modelo de Datos	Una ontología global integrada (GVV, "Global Virtual View")	El diseñador involucrado en la anotación esquema y establecimiento umbral de clasificación para GVV	Equivalencia de nombres: Sinónimos, Subordinados, Encaje de agrupación	Sí	No	Sí	No
<b>LSD</b>	Esquemas fuente y sus instancias	Pares de términos relacionados entre un esquema global y local	El usuario proporciona mapeos para entrenar a la fuente y realimentación sobre los mapeos propuestos	Aproximación de aprendizaje multi-estrategia: técnicas de aprendizaje automático	No	Sí	Sí	Sí
<b>CTXMATCH</b>	Conceptos en jerarquía de conceptos	Relación semántica entre conceptos	No (CTXMATCH es algoritmo, no aplicación)	Deducción lógica	Sí	No	Sí	Sí
<b>GLUE</b>	Dos taxonomías con sus instancias de datos en las ontologías	Un conjunto de parejas de conceptos similares	Mapeos definidos por usuario para datos de entrenamiento, medidas similitud, establecimiento de peso de aprendizaje y analizar sugerencias de "system's match"	Aproximación de aprendizaje multi-estrategia: técnicas de aprendizaje automático	No	Sí	Sí	Sí
<b>MAFRA</b>	Dos ontologías	Mapeos de las dos ontologías mediante una ontología puente semántica	El experto del dominio interactúa con los módulos de enlace semántico y de semejanza, que tienen una interfaz gráfica de usuario (GUI)	Puente semántico	Sí	Sí	Sí	Sí
<b>LOM</b>	Dos listas de términos de dos ontologías	Una lista de parejas encajados de términos ordenadas por su grado de similitud	Requiere validación del usuario al final del proceso	Similitud léxica de términos completo, parte de palabra, conjunto de sinónimos y encaje de caracteres	No	No	Sí	No
<b>ONION</b>	Términos en dos ontologías	Conjunto de reglas de articulación	Un experto elige, borra o modifica las	Encaje lingüístico, estructura,	Sí	No	Sí	Sí

		entre las dos ontologías	asociaciones sugeridas mediante una GUI	heurísticas basadas en deducción				
<b>PROM PT</b>	Dos ontologías	Una ontología fusionada	El usuario acepta, rechaza o ajusta las sugerencias del sistema.	Analizador basado en heurística	Sí	No	Sí	No
<b>FCA-Merge</b>	Dos ontologías y un conjunto de doc. de conceptos en las ontologías	Una ontología fusionada	La generación de la ontología de fusión requiere la interacción del experto	Análisis lingüístico y algoritmo TITANIC para "computation for pruned concept lattice"	Sí	Sí	Sí	Sí

### *Conocimiento y Ontologías*

Hablando de la gestión del conocimiento en relación con Internet en el seno de las organizaciones, Dieng presenta dos alternativas para organizar la memoria corporativa basada en documentos (conocimiento no formalizado) [Dieng 2000]. En ambos casos se trata de asociar conocimiento formalizado a los documentos con el que se podrá razonar y recuperar partes de ellos. Una es añadir el conocimiento directamente a los documentos, la otra consiste en almacenar tal conocimiento en forma de meta-información fuera de los documentos. Cuando los formatos de los documentos son heterogéneos, se puede almacenar pequeñas referencias de ellos con enlaces a los originales, o envolver estos en documentos XML. Estas alternativas pueden aplicarse a la memoria documental de cualquier colectivo.

Motta propone una alternativa para el enriquecimiento de documentos mediante una aproximación basada en ontologías. El enriquecimiento consiste en la asociación de información a los documentos, pero sin incorporarla a estos como se hace en los sistemas basados en anotaciones. Para ello se parte de un acercamiento de-arriba-a-abajo inspirado en el enfoque ontológico de Gruber [Gruber 1993], en lugar de seguir una aproximación de-abajo-a-arriba característica de los sistemas de anotación de documentos [Motta 2000].

### *Ontologías y Bases de Datos*

Existen varios procedimientos para explotar las ontologías como bases de datos, desde la representación en memoria del grafo correspondiente, al empleo de una base de datos relacional que almacene la ontología en notación de triplas RDF, pasando por la utilización de bases de datos XML nativas [Emerick 2002]. La elección del sistema más adecuado depende de las características del proyecto. El procedimiento de representar en memoria del grafo puede ser adecuado, y más rápido, para sistemas que manejen ontologías de tamaño moderado puesto que los requerimientos de memoria son proporcionales a éste y al número de instancias tratadas. El almacenamiento en triplas podría escalar mejor, si el gestor de base de datos relacional es adecuado y el diseño establecido para el mapeo de la ontología también [Hjelm 2001].

Lo cierto es que, de acuerdo con Shadbolt [Shadbolt 2006], como RDF y RDF Schema han ganado terreno, la necesidad de almacenar contenidos RDF ha crecido. En lo que se refiere al almacenamiento de triplas, las capacidades de las soluciones son diferentes. Mientras unas se centran en proporcionar medios adecuados para razonar con las triplas [Jena 2008], otras lo hacen en almacenar grandes cantidades de datos [3store 2008] [Oracle 2008]. Al mismo tiempo, hay soluciones de almacenamiento de triplas que funcionan como extensiones de los navegadores Web actuales [PiggyBank 2008] y otras que lo hacen como sistemas que pueden funcionar con varias bases de datos de terceros [OpenRDF 2008]. Por otro lado, la necesidad de un sistema de acceso fiable y estándar a los contenidos RDF almacenados, ha llevado al W3C a la introducción del lenguaje SPARQL, que es una recomendación desde enero de 2008 [SPARQL 2008].

En relación con el diseño de bases de datos relacionales que implementen un mapeo de las triplas RDF directamente, sin módulos accesorios, se han propuesto distintas opciones. La alternativa más sencilla consistiría en almacenar todas las triplas en una única tabla con tres columnas, esta solución que podría tener sentido para aplicaciones básicas, puesto que tiene un claro problema de escalabilidad. Existen múltiples diseños que pueden adaptarse a los requerimientos de un sistema con mayores pretensiones. Una podría ser crear varias tablas para las declaraciones, el modelo de datos, los literales, los recursos y las claves de las declaraciones, respectivamente; otra crear tablas para almacenar los recursos, las declaraciones y los prefijos [Hjelm 2001]. Algunas alternativas como éstas pueden encontrarse en [Melnik 2001].

#### *Aplicaciones de la Web Semántica*

El modelo de la Web semántica se está empezando a utilizar en diversos ámbitos -comercial, corporativo, educativo...- para variadas aplicaciones como la formación, la gestión de recursos humanos, las comunidades virtuales o el comercio electrónico. Pueden servir como ejemplo de la aplicación de esta tecnología sistemas de propósito general como IBROW [Benjamins 1998b, Fensel 2001] que proporcionan modelos conceptuales para el manejo del conocimiento distribuido en áreas de trabajo donde la información sobre el dominio tiene una estructura conocida inicialmente, o del ámbito de los negocios como el sistema para soporte de decisiones WebCADET [Caldwell 2000] o el administrador de noticias en grupos de trabajo Planet-Onto [Domingue 2000]. También se están aplicando las ontologías como soporte de sistemas de búsqueda automática de conocimiento y gestores de consultas sobre la red, como hace Ontobroker [Fensel 2001] que propone un potente lenguaje para crear esquemas conceptuales del conocimiento en la Web y un motor para búsquedas estructuradas, funcionalidades que On2broker [Fensel 2001] amplía.

Benjamins propone un sistema basado en ontologías para implementar un servicio de intermediación inteligente para reutilizar componentes de conocimiento en la WWW (BROW), en particular métodos para la resolución de problemas (Problem-Solving Methods, PSM) de especial importancia para la construcción de sistemas basados en conocimiento. Emplea dos ontologías, una para representar el dominio de las tareas a resolver y otra el de las PSM, pero un mismo lenguaje en ambos dominios. La solución para algunas tareas es una combinación de varias PSMs. Además es necesario explicar cómo las PSMs se adaptan a las tareas [Benjamins 1998 a, b]. Relacionado con este proyecto está el trabajo de Fensel sobre el lenguaje UPML (*Unified Problem-solving Method description Language*) para la representación de la arquitectura y componentes de las PSMs y facilitar su reutilización y adaptación semiautomática [Fensel 1999 a, b].

Ontobroker [Fensel 2001] es un proyecto que aplica técnicas de inteligencia artificial para mejorar el acceso a fuentes heterogéneas de información, dispersas y semiestructuradas, publicadas en la Web o en una Intranet. Emplea ontologías para anotar las páginas Web, formular consultas y derivar respuestas. Lo esencial de la cuestión es definir una ontología y usarla para anotar, estructurar y envolver ciertos documentos Web, y que cualquier usuario puede emplear los servicios de interrogación avanzada e inferencia de Ontobroker para consultar el conocimiento incluido en esos documentos. Para alcanzar este objetivo el sistema proporciona tres lenguajes y dos herramientas. Su arquitectura consta de tres elementos principales, una interfaz de interrogación para formular consultas, un motor de inferencia para derivar respuestas, y un robot para recolectar el conocimiento de la Web. El sistema proporciona un lenguaje de representación para formular ontologías, un lenguaje para formular consultas, que es un subconjunto del anterior, y un lenguaje de anotación para que los proveedores enriquezcan los documentos con información ontológica.

El proyecto On2broker [Fensel 2001] amplía Ontobroker en varios sentidos, en él se separan claramente los motores de consulta e inferencia y se integran nuevos estándares como XML y RDF. La razón de estas innovaciones está en responder a dos de los problemas de complejidad significativa de Ontobroker: el esfuerzo requerido de inferencia computacional para un número grande de hechos, y el esfuerzo de anotación humana para añadir semántica a los documentos HTML.

La arquitectura general de On2broker incluye cuatro motores fundamentales: un motor de consultas (*Query Engine*), un agente de información (*Info Agent*), un motor de inferencia (*Inference Engine*) y un gestor de base de datos (*DB Manager*). Las Ontologías (*Ontologies*) son el elemento fundamental del sistema. El agente de información las usa para extraer hechos, el motor de inferencia para inferirlos, el gestor de base de datos para estructurar la base de datos y proporcionar hechos al motor de consulta.

Calvo propone en OSHI y WebOSHI [Calvo 2002] un sistema para compartir y reutilizar ontologías de fuentes documentales históricas, en el que se emplea una meta-ontología de las ontologías para la conceptualización de los distintos documentos que maneja la aplicación. Hezen aplica ontologías a un sistema de aprendizaje Hipermedia Adaptativo abierto, considerando de múltiples ontologías sobre un mismo dominio, fijándose especialmente en la valoración de los conocimientos del alumno [Hezen 2002].

### Agentes Software

Además de las Ontologías, otro tema relacionado con la Web Semántica y de especial interés para este proyecto es el de los Agentes Software y los sistemas basados en ellos. Se trata de unos conceptos ampliamente utilizados que paradójicamente no tienen todavía una definición precisa. Sobre esta cuestión trata este apartado.

La introducción del concepto de agente en inteligencia artificial se debe en parte a la dificultad para resolver problemas en situaciones donde no existía una completa definición del entorno de la cuestión o de las entidades involucradas en el proceso de resolución. En entornos complejos como estos, los agentes parecen responder de forma efectiva y eficiente para la solución de problemas. Para ello los agentes cuentan con un mecanismo que les permite ser sensibles al entorno y obtener datos de él, y otro para, después de procesar la información, actuar sobre el entorno como respuesta. Se trata de una idea sencilla pero poderosa que ha sido adoptada rápidamente en muchas áreas de informática por su utilidad y versatilidad.

Actualmente existe una gran variedad de entidades computacionales, que bajo la denominación de agentes, se aplican en diversos campos. El problema del éxito de la metáfora de agente es que no ha permitido llegar a un consenso sobre qué es realmente un agente. Distintos autores han propuesto definiciones que no siempre coinciden en todos sus términos [Franklin 1994] [Dinverno 2001]. Mientras se busca una definición adecuada se puede intentar encontrar los atributos que caracterizan a los pretendidos agentes para procurar identificar qué tienen en común.

En este sentido se puede establecer una noción débil de agente, que involucra: *autonomía* o habilidad para funcionar sin intervención; *habilidades sociales*, por medio de las cuales los agentes pueden interactuar unos con otros; *reactividad*, que permite a los agentes percibir y responder a los cambios en el entorno; y *pro-actividad*, gracias a la cual los agentes pueden establecer su comportamiento en pro de un objetivo. Estas características son aceptadas en general por muchos como representativas de las cualidades clave que pueden servir para evaluar la condición de ser agente. Al mismo tiempo, se puede establecer una noción fuerte de agente, dominante en inteligencia artificial, que además de los aspectos incluidos en la noción débil

incorpora otros de carácter mentales como creencia, deseo, intención, conocimiento, etc. [Wooldridge 1995] [Dinverno 2001].

Sin embargo, otros autores [Etzioni 1995] establecen como características deseables de los agentes cuestiones como: *autonomía y continuidad temporal*, que confiere a los agentes persistencia en el tiempo; *personalidad*, para facilitar la interacción efectiva; *habilidad de comunicación*, con otros agentes y personas; *adaptabilidad* a las preferencias de los usuarios; y *movilidad*, por medio de la cual los agentes se pueden trasladar entre diferentes máquinas y arquitecturas. Además, la autonomía requiere que los agentes estén *orientados por objetivos* y acepten peticiones de alto nivel, *colaborativos* en el sentido de que puedan modificar sus peticiones y clarificarlas, *flexibles*, para manejar acciones en forma de guión, *activarse por sí mismo* en el sentido que puedan sentir cambios en el entorno y decidir cuándo ponerse en acción. Otras características, como *veracidad, benevolencia y racionalidad*, se consideran también frecuentemente parte de la noción de agente.

Aunque todas estas últimas cualidades también conviene tenerlas presentes al pensar en la caracterización de los agentes, parece que entre todas las características mencionadas, las cuatro incluidas en la referida noción débil se pueden considerar como esenciales [Jennings 1998] [Dinverno 2001]. Las distintas clases de agentes tienen más acentuadas unas características que otras, o incorporan algunas nuevas que marcan su singularidad.

Los sistemas multi-agente son sistemas distribuidos que incluyen varios componentes, siendo cada uno de ellos un agente independiente capaz de resolver problemas, y que forman entre todos un conjunto coherente. Estos sistemas no tienen generalmente una arquitectura o configuración preestablecida para incorporar los agentes. Además no hay objetivos globales de sistema, los agentes son heterogéneos y tienen sus propios objetivos y capacidades. En consecuencia en este tipo de sistemas los agentes necesitan coordinar sus actividades y cooperar con los demás, para evitar la multiplicación de esfuerzo y aprovechar las capacidades de los otros. Por esta razón es necesario considerar cuestiones nuevas que no surgen cuando se consideran agentes individuales y aislados [Dinverno 2001]:

- Modelado de agentes, el agente necesita un modelo para representar el mundo con el que interactúa y, como en él existen otros agentes, un modelo de éstos también.
- Planificación multi-agente, en algunos casos los agentes deberán compartir planes para coordinar sus comportamientos o alcanzar un objetivo utilizando a otros.
- Relaciones sociales, los agentes deben mantener relaciones sociales con otros agentes, por ejemplo si un agente realiza un servicio para otro el segundo adquiere una obligación de reciprocidad con el primero.

- Interacción, los agentes deben interactuar y, como la interacción puede no estar predefinida en su mundo, para ello necesitan modelos de los demás para decidir cómo hacerlo y medir los resultados.
- Comunicación, los agentes necesitan comunicarse para explotar la interacción y poder coordinarse, en general cada agente debe persuadir a otros de adoptar sus objetivos y alterar sus planes convenientemente.

Los mismos problemas de indefinición de los que se ha hablado previamente para los agentes surgen con los sistemas multi-agente.

Entre los esfuerzos de estandarización de los sistemas de agentes, tiene especial relevancia el realizado por FIPA (Foundation for Intelligent Physical Agents) [FIPA 2008], que es una organización promovida por un consorcio de importantes empresas del sector de las TIC (Tecnologías de la Información y las Comunicaciones) interesadas en el campo de los agentes. FIPA ha propuesto una serie de recomendaciones con mucha aceptación, que tratan sobre la arquitectura de las plataformas para agentes y del mecanismo de comunicación entre estos.

La recomendación FIPA [FIPA 2008] basa la comunicación entre agentes en el intercambio de mensajes sobre HTTP -protocolo de comunicación de la Web- en un lenguaje que sigue las especificaciones de FIPA ACL (Agent Communication Language). Los mensajes pueden representarse en formato XML -formato popular de intercambio de datos-, y tienen la posibilidad de incluir contenidos, que sólo los agentes implicados deben entender, y utilizar ontologías -esquema conceptual formalizado de un dominio- para determinar el significado de dichos contenidos. Tales mensajes se utilizan en actos de comunicación, que establecen las secuencias de mensajes de los diálogos necesarios entre agentes, según protocolos de interacción predefinidos.

La arquitectura FIPA [FIPA 2008] establece un modelo lógico de referencia para la creación, registro, localización, comunicación, migración y baja de los agentes. El modelo establece una plataforma de agentes que incluye los agentes, un sistema de gestión de los mismos, un sistema de transporte de mensajes y un directorio. El sistema de gestión se encarga de controlar el acceso a la plataforma y el uso de la misma. El sistema de transporte de mensajes establece el procedimiento de comunicación entre agentes de diferentes plataformas. El directorio permite el registro de los servicios de los agentes y la localización de los agentes que proporcionan determinados servicios.

Actualmente, existen varias plataformas de agentes que están basadas en las especificaciones de la arquitectura de FIPA, como JADE [JADE 2008] o FIPA-OS [FIPA-OS 2008].

KQML (Knowledge Query Manipulation Language) [Finin 1994] es otro sistema de interacción entre agentes software que incluye un lenguaje y un protocolo de comunicación, que fue

desarrollado por DARPA Knowledge Sharing Effort Consortium (consorcio de la Defense Advanced Research Projects Agency) e implementado por varios grupos de investigación. KQML tiene unas características muy similares a FIPA ACL, cómo estar basados en la misma teoría del acto de hablar o tener una sintaxis inspirada en LISP, aunque son distintos en algunos aspectos, como detalles sintácticos, el tratamiento de instrucciones dirigidas a manejar algunas relaciones sociales o el modo de gestionar preguntas múltiples.

Entre las plataformas para la implementación de sistemas multi-agentes, preocupadas de proporcionar soporte para la comunicación y la coordinación entre agentes, pueden mencionarse tres [Bordini 2006]. La primera, TuCSoN (Tuple Centre Spread over the Network) [TuCSoN 2008] es un entorno para la coordinación de un sistema multi-agente, basado en un modelo y una infraestructura de propósito general que proporciona servicios programables para la comunicación y coordinación de los agentes. La segunda, DESIRE (DEsign and Specification of Interacting REasoning components) [DESIRE 2008] es un método de desarrollo compositivo para sistemas multi-agente, basado en noción de arquitectura compositiva, que fue desarrollado por el equipo de Treur en la Vrije Universiteit Amsterdam. La tercera, JADE (Java Agent DEvelopment Framework) [JADE 2008] es un entorno de trabajo Java para el desarrollo de aplicaciones multi-agente distribuido. El entorno JADE representa un agente middleware que proporciona un conjunto de servicios disponibles y fáciles de usar, así como varias herramientas gráficas para depurar y probar los desarrollos. Uno de los objetivos principales de la plataforma JADE es soportar la interoperatividad de los sistemas multi-agente siguiendo estrictamente las especificaciones FIPA respecto a la arquitectura de la plataforma y la infraestructura de comunicaciones.

En el ámbito de la plataforma KnowCat se han utilizado agentes en investigaciones realizadas en los últimos años [García 2006], algunas de ellas relacionadas con los prototipos desarrollados para este trabajo sobre dicha plataforma [Dumitrescu 2007], así como la plataforma JADE antes mencionada.

#### **2.2.4 Minería de Información y de Datos**

El presente apartado se ocupa del campo de investigación de Minería de Información y de Datos, que incluye dos áreas de gran interés para la aproximación propuesta en esta tesis, una es el área de Minería de Texto, que se ocupa en general del análisis de información textual poco estructurada, y el otro área es la de Minería Web, que se dedica al análisis de la información accesible a través de la Web y de los datos vinculados con los servicios ofrecidos a través de ella. Relacionados con la primera de dichas áreas están los fundamentos del Modelo Vectorial, empleado de forma generalizada para el análisis textual, y el Análisis de la Semántica Latente, que aporta una aproximación diferente al análisis de textos y tiene especial interés para este



trabajo. Vinculado con la segunda de las áreas mencionadas está la Minería de Uso Web, que también se aplica en la aproximación propuesta para el análisis de la actividad de los usuarios con sistemas basados en dicho servicio.

Como se avanzó en la introducción de este trabajo, a principios de los noventa se dieron dos circunstancias que permitieron la aparición de la minería de información y de datos (Data Mining) [Piatetsky-Shapiro 1991], a las que podemos añadir una más. De acuerdo con Piatetsky-Shapiro, por un lado, varias técnicas aplicables al análisis de datos que se habían venido desarrollando desde los años sesenta habían demostrado ya su madurez y solidez, como las técnicas estadísticas [Morgan 1963], los métodos de aprendizaje automático [Hunt 1966], las redes neuronales [Minsky 1969], y los algoritmos evolutivos [Holland 1975]. Por otro lado, empezaba a ser fácil disponer y compartir grandes repositorios de datos [Piatetsky-Shapiro 2000], como resultado del abaratamiento de los dispositivos de almacenamiento y el desarrollo de las redes de comunicación.[ARCADIA 2002] Pero además, el incremento y depreciación de la capacidad de cálculo permitió disponer de equipos con suficiente potencia. La Minería de Información y de Datos surgió de la aplicación de aquellas técnicas a esos repositorios utilizando tales capacidades de cálculo.

#### Minería de texto

La Minería de Texto es un campo de la Minería de Información, como la Minería de Datos, que se diferencia de ésta en que trabaja sobre información textual en lugar de hacerlo sobre información estructurada. La información textual tiene una cierta organización interna, pero está básicamente desestructurada. Sus aplicaciones principales son el Descubrimiento de Asociaciones, el Descubrimiento de Tendencias y la Detección de Sucesos. La Detección de Sucesos, en el ámbito de noticias en medios de información, permite la localización de historias en un flujo continuo de noticias, que corresponden a acontecimientos nuevos o no identificados previamente [Chang 2001].

En este contexto se ubica el método de Agrupación Incremental [Yang 1998] [Chang 2001], que se basa en un modelo vectorial [Salton 1989] [Chang 2001]. Consiste en procesar secuencialmente documentos de entrada y hacer crecer unas agrupaciones de forma incremental. En esta aproximación, las historias se representan por vectores cuyas dimensiones son el número de términos con raíces distintas en la colección, y cuyos elementos son el peso de los términos en las historias. El peso de los términos en las historias es una combinación de su frecuencia y de la frecuencia del documento inverso. La primera es el número de ocurrencias del término en la historia. La segunda, un valor asignado a las palabras que es alto para las raras y bajo para las comunes. Para establecer el grado de similitud entre historias y agrupaciones es

necesario establecer la correlación entre ellos, que puede cuantificarse por el coseno del ángulo que forman.

Los modelos vectoriales también se utilizan en el área de recuperación de información [Baeza 1999], que es un caso de Descubrimiento de Asociaciones, para buscar respuestas a preguntas en las que coincidencias parciales son también deseables. De forma general el procedimiento consiste en establecer un vector representativo de las raíces de los términos más importantes de la pregunta y de cada documento, y asignar a cada término un peso que es un valor representativo de la importancia de la palabra para caracterizar documentos y preguntas. Existen distintas técnicas para establecer tales pesos, en general se procura dar menos valor a las palabras más comunes y raras en el ámbito que se está tratando, mientras que se asigna más al resto, y luego considerar la frecuencia de ellas en los textos sobre los que se trabaja. Para establecer el grado de similitud entre dos textos es necesario, en primer lugar, reformular sus vectores en términos comunes, y luego establecer la correlación entre ellos, que puede cuantificarse, como en el caso anterior, por el coseno del ángulo que forman.

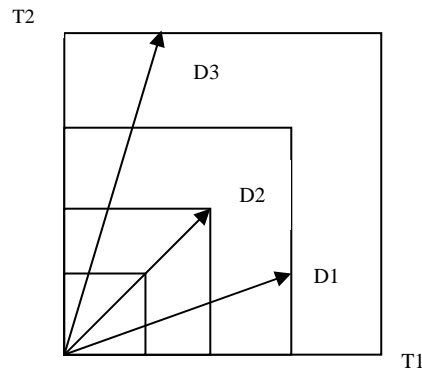
#### *Modelo Vectorial en Minería de Texto*

El procesamiento de recuperación de información textual y también el de agrupación y categorización de documentos textuales, comienza con un tratamiento inicial de los mismos, en el que se eliminan las palabras de parada -stopwords-, que son términos muy frecuentes en el idioma en el que está escrito el documento, por lo que pueden aparecer en escritos de cualquier tema y no aportan información específica sobre ellos. Se trata de palabras como los artículos, las conjunciones, las preposiciones o los verbos auxiliares.

El siguiente paso es eliminar las variantes morfológicas de las palabras, reduciendo todas las variaciones de cada una a una forma única. Este es el caso de los plurales, los tiempos y las personas de los verbales o las variaciones de género de las palabras. En inglés se dispone del algoritmo de Porter [Porter 2006] que permite realizar esta reducción, existen aproximaciones similares para otros idiomas y otras técnicas para conseguir resultados similares [Baeza 1999]. En el presente trabajo hemos utilizado el analizador de lenguaje Freeling [Carreras 2004] [Freeling 2008] para realizar estos dos pasos, trabajando con documentos en castellano.

La utilización de procedimientos de análisis del lenguaje natural como Freeling no es muy frecuente en este tipo de procesos, por el coste computacional que tiene y el poco beneficio que supone respecto a técnicas menos costosas. Este fenómeno se hace evidente en los sistemas de recuperación de información en la Web, entornos más populares en los que se emplean métodos similares a los que nos referimos, que se caracterizan por comparar textos pequeños -preguntas- con repositorios enormes de documentos. Sin embargo, parece que la situación podría ser diferente cuando se comparan textos más grandes, como documentos enteros, con repositorios

de tamaño moderado de documentos [Brants 2004]. Además, en la actualidad la potencia de los equipos informáticos permite emplear técnicas más costosas sin tanto impacto en el rendimiento de las aplicaciones.



*Figura 2-9. Documentos en un espacio de términos de dos dimensiones*

Una vez que se han convertido los documentos en conjuntos de términos significativos, se puede representar dichos documentos en un espacio vectorial de términos, lo que permitirá la comparación automática entre aquellos. En la Figura 2-9 se muestran tres documentos en un espacio de dos dimensiones, que corresponden a dos términos: en el documento D1 aparecen tres veces el término T1 y una el T2; en D2 está dos veces T1 y otras dos T2; y en D3 aparece una vez T1 y cuatro T2.

Aunque la frecuencia de aparición de las palabras en los documentos podría ser una medida de la importancia de aquellas en estos, no todos los términos frecuentes en el uso general del idioma se han eliminado de los documentos al quitar las palabras de parada. Se trata de términos que pueden aparecer en documentos de distintos temas y no son característicos de ellos. Por eso en lugar de considerar directamente la frecuencia de la palabra en los documentos se establece un peso en el documento para cada palabra que aparece en él. Para el cálculo de dicho peso se toma en consideración la frecuencia de aparición de la palabra en el documento junto como la frecuencia de la palabra en una colección de documentos representativa del uso general del idioma. La Tabla 2-9 muestra la representación de los diversos conceptos que se utilizan en las explicaciones posteriores.

La frecuencia normalizada de una palabra en un documento es el cociente de la frecuencia de la palabra en el documento entre la frecuencia de la palabra más popular en él, que es el máximo de las frecuencias de todas las palabras que este incluye. La normalización es un modo de reducir la magnitud del valor obtenido tomando en consideración el tamaño del documento.

Tabla 2-9. Conceptos relacionados con el modelo vectorial en el campo de la minería de texto

Concepto	Representación	Definición
Frecuencia de palabra en documento	$f_{k,i}$	Número de ocurrencias de la palabra k en el documento i
Frecuencia normalizada de palabra en documento	$fn_{k,i} = \frac{f_{k,i}}{\max_{l=1}^m (f_{l,i})}$	Número normalizado de ocurrencias de la palabra k en el documento i
Número de documentos donde aparece la palabra en la colección	$n_k$	Número de documentos en los que aparece la palabra k en la colección representativa del uso general del idioma
Número total de documentos en la colección	$N$	Número total de documentos en la colección representativa del uso general del idioma
Frecuencia de la palabra en el documento inverso	$fdi_k = \log \frac{N}{n_k}$	Frecuencia de documento inverso del término k en la colección utilizada de referencia

La frecuencia de la palabra en el documento inverso es el logaritmo decimal del cociente del número de documentos de la colección de referencia entre el número de documentos donde aparece la palabra en la colección. En esta fórmula el cociente expresa la relación inversa entre las dos magnitudes y el logaritmo permite, por un lado, trabajar con números menores, puesto que los valores del cociente pueden llegar a ser muy altos cuando se manejan muchos documentos, y por otro lado, que la frecuencia sea cero cuando una palabra aparece en todos los documentos.

$p_{k,i} =$	$(1 + \log(f_{k,i})) \cdot fdi_k$ si $f_{k,i} \geq 1$ $0$ si $f_{k,i} = 0$	$p_{k,i} = fn_{k,i} \cdot fdi_k$
<p>Fórmula 2-1. <i>Peso de palabra en documento en agrupación y categorización</i></p>		<p>Fórmula 2-2. <i>Peso de palabra en documento en recuperación de la información [Baeza 1999]</i></p>

La forma de combinar estos elementos en una fórmula matemática para obtener fácilmente valores adecuados depende de las aplicaciones y de preferencias personales. En el ámbito de agrupación y categorización algunas veces se emplea la Fórmula 2-1, en la que se considera la frecuencia de la palabra en el documento por su logaritmo decimal. Sin embargo, en contexto de la recuperación de la información es frecuente que se utilice la Fórmula 2-2, donde se emplea la frecuencia normalizada de las palabras en el documento.

$ \vec{v}  = \sqrt{\sum_{i=1}^n v_i^2}$	$ \vec{v}_1 - \vec{v}_2  = \sqrt{\sum_{i=1}^n (v_{1,i} - v_{2,i})^2}$
Fórmula 2-3. <i>Módulo de un vector</i>	Fórmula 2-4. <i>Distancia euclídea entre dos vectores</i>

Existen varias medidas de similitud entre vectores, como son la distancia euclídea que los separa o el coseno del ángulo que forman. Ambas medidas utilizan el módulo de los vectores en sus ecuaciones, que se calcula como la raíz cuadrada de la suma de los cuadrados de sus componentes. En la Fórmula 2-3 se representa el módulo  $|\vec{v}|$  del vector  $\vec{v} = (v_1, v_2, v_3 \dots v_n)$ , en función de sus n componentes  $v_i$ .

$\cos(\theta) = \frac{\vec{v}_1 \cdot \vec{v}_2}{ \vec{v}_1  \cdot  \vec{v}_2 }$	$\vec{v}_1 \cdot \vec{v}_2 = \sum_{i=1}^n v_{1,i} \cdot v_{2,i}$	$\cos(\theta) = \frac{\sum_{i=1}^n v_{1,i} \cdot v_{2,i}}{\sqrt{\sum_{i=1}^n v_{1,i}^2} \cdot \sqrt{\sum_{i=1}^n v_{2,i}^2}}$
<i>Fórmula 2-5. Coseno ángulo entre dos vectores</i>	<i>Fórmula 2-6. Producto escalar</i>	<i>Fórmula 2-7. Coseno ángulo vectores función componentes</i>

Por un lado, la distancia euclídea entre dos vectores se define como el módulo de la diferencia de los dos vectores, que es la raíz cuadrada de la suma de los cuadrados de las diferencias de las componentes correspondientes de ambos vectores. En la Fórmula 2-4 se obtiene la distancia  $|\vec{v}_1 - \vec{v}_2|$  entre los vectores  $\vec{v}_1 = (v_{1,1}, v_{1,2}, v_{1,3} \dots v_{1,n})$  y  $\vec{v}_2 = (v_{2,1}, v_{2,2}, v_{2,3} \dots v_{2,n})$ , en función de sus respectivas componentes  $v_{1,i}$  y  $v_{2,i}$ .

Por otro lado, el coseno del ángulo  $\theta$  entre dos vectores  $\vec{v}_1$  y  $\vec{v}_2$  viene dado por la expresión de la Fórmula 2-5, donde el numerador  $\vec{v}_1 \cdot \vec{v}_2$  es el producto escalar de los vectores y el denominador el producto de sus módulos.

Como es sabido el producto escalar de dos vectores es la suma de los productos de sus correspondientes componentes, esto es lo que indica la Fórmula 2-6. Por lo tanto el coseno del ángulo que forman dos vectores puede expresarse en función de sus componentes como se muestra en la Fórmula 2-7.

Castells presenta en [Castells 2007] una interesante adaptación del modelo de espacio vectorial para la recuperación de información empleando ontologías. Con ello se pretende mejorar los resultados de las búsquedas sobre repositorios grandes de documentos explotando una base de conocimiento basada en ontologías. Como resultado se consigue una mejora evidente al emplear esta técnica respecto a las búsquedas tradicionales basadas en palabras claves.

Análisis de la Semántica Latente

Según Landauer [Landauer 1998], el análisis de la semántica latente (Latent Semantic Análisis, LSA) es una teoría y un método para extraer y representar el significado de las palabras debido al uso contextual de las mismas, mediante el procesamiento estadístico de grandes corpus de textos. La idea subyacente es que la suma de todos los contextos donde una palabra dada

aparece y no lo hace, proporciona un conjunto de condiciones que determina en gran medida la similitud de significado de palabras y de conjuntos de ellas entre si. Además, el LSA parece reflejar algunos aspectos del tratamiento mental humano del significado, por lo que se ha especulado con la posibilidad de establecer una teoría psicológica del significado a partir de su funcionamiento [GutierrezR 2005].

El LSA no es un procesamiento tradicional de lenguaje natural o un programa de inteligencia artificial, no emplea diccionarios, bases de conocimiento, redes semánticas, gramáticas, análisis sintáctico ni morfológico, u otros recursos similares. El mecanismo parte de texto en bruto, que se considera compuesto de palabras, definidas como cadenas de caracteres únicas, que se agrupan en pasajes o muestras significativas, tales como sentencias, párrafos o documentos, que determinan contextos de uso. [Landauer 1998]

El primer paso del proceso consiste en transformar el texto de partida en una matriz en la que cada fila corresponde a una palabra y cada columna a un pasaje de texto o marco contextual, y en la que cada una de las celdas contiene la frecuencia de una palabra en un contexto. A continuación, las celdas de la matriz se someten a una transformación por la que se establece un peso de palabra por muestra, mediante alguna función que considera la importancia de las palabras en los pasajes y en el dominio en general. Por último, el procedimiento de LSA aplica una descomposición en valores singulares (singular value decomposition, SVD) a la matriz, tratamiento que es una forma generalizada de análisis factorial. [Landauer 1998]

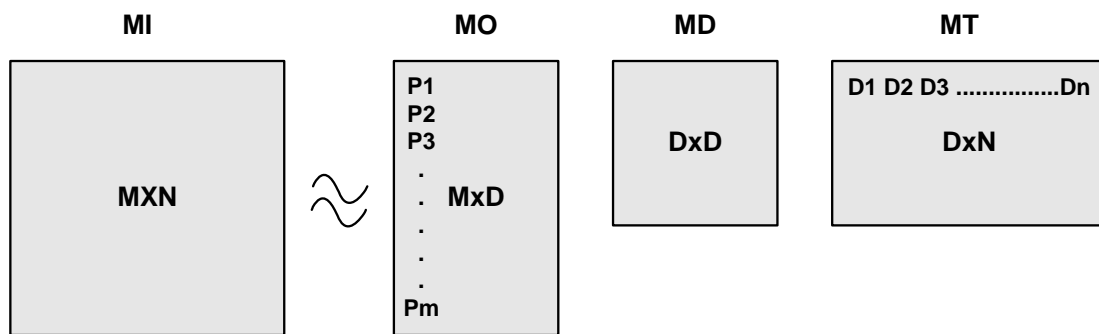


Figura 2-10. Descomposición de matriz con SVD de acuerdo con [JorgeBotana 2007]

De este modo, la matriz inicial (MI) se descompone en el producto de tres matrices [Landauer 1998][GutierrezR 2005] (ver Figura 2-10 ): la primera (MO), se calcula procesando linealmente el número de columnas de MI y describe las palabras originales cómo vectores de valores de factores ortogonales derivados, representando las palabras como vectores en espacios de palabras; la segunda matriz (MT), se obtiene a partir de las matriz transpuesta de MI procesada de forma análoga a MO y describe los pasajes originales de forma equivalente, representándolos como vectores en espacios de palabras; y la última, es una matriz diagonal (MD), con valores

distintos de 0 sólo en su diagonal, que se calcula a procesando linealmente el número de filas, el número de columnas y la cantidad de dimensiones de la matriz MI, y contiene valores escalares tales que cuando las tres componentes se multiplican matricialmente, se consigue una matriz semejante a la original. La SVD permite descomponer cualquier matriz, utilizando no más factores que la menor de las dimensiones de la matriz. La restauración de la matriz de partida puede ser más o menos precisa dependiendo del número de factores de la matriz diagonal considerados, lo que puede facilitar el proceso según las necesidades.

Sin embargo, lo interesante de esta reducción de dimensiones no es sólo mejorar el manejo de una matriz de palabras y pasajes, que es deseable porque suele ser muy grande, sino crear un espacio vectorial semántico en el que palabras y pasajes estén representados por medio de vectores que contengan la información esencial para establecer conceptos que los relacionen, que son los espacios de palabras. La representación vectorial del lenguaje conseguida permite realizar comparaciones entre los elementos considerados mediante alguna medida de similitud, como cósenos o distancias euclídeas. Además, es posible representar nuevos pasajes en el espacio vectorial para realizar categorizaciones de los mismos. [Jorge-Botana 2006]

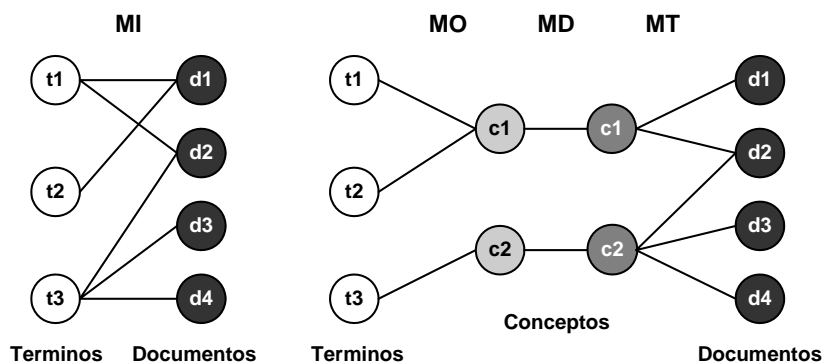


Figura 2-11. Efecto reducción de dimensiones empleando SVD de acuerdo con [JorgeBotana 2007]

La Figura 2-11 ilustra el efecto de la reducción de dimensiones conseguida mediante SVD [Jorge-Botana 2007]. En el lado izquierdo se muestra la relación entre palabras y pasajes mediante líneas que los unen. En el lado de la derecha se representan las mismas palabras y pasajes enlazados a través de los conceptos semánticos -dimensiones abstractas- fruto de la reducción. A cada palabra se le infiere una probabilidad de estar representada por cada uno de esos conceptos.

Como resultado del proceso algunas palabras pueden alcanzar cierta probabilidad de encontrarse en pasajes donde no están y otras obtener probabilidad baja de aparecer en pasajes donde realmente están. Esta situación puede parecer paradójica, pero se debe a que el significado es contextualmente dependiente, no fijo, y a que existen similitudes semánticas latentes entre palabras. Por ejemplo, la dependencia contextual del significado puede observarse en el uso de

la palabra “carta”, que dependiendo del ámbito donde se encuentre puede referirse a un mensaje escrito que se envía entre personas para comunicarse o a la lista de platos de un restaurante. La similitud semántica latente puede observarse también en el uso de esa palabra, que tiene el mismo significado que “menú” en algunos contextos. Debido a que los dos términos son sinónimos, no es frecuente que estén juntos en un texto. Sin embargo, es probable que ambos aparezcan al lado de algunas palabras relacionadas con su significado común, como “plato”, “sopa” o “postre”. La relación entre “carta” y “menú” se establece a través de sus vínculos con otras palabras.

La LSA se ha empleado en aplicaciones muy diversas, Gutiérrez [Gutiérrez 2005] menciona algunas cómo medir la coherencia de textos, comparar un texto con un resumen, establecer la adecuación de textos para posibles lectores, corregir cuestionarios de respuesta abierta o estudiar la comprensión de textos en el ámbito psicolingüístico. Además se pueden citar otras aplicaciones interesantes, como estimar de forma automática las interacciones de usuario en llamadas telefónicas -call routing- [Jorge-Botana 2007], evaluar la usabilidad de aplicaciones Web [Jorge-Botana 2006] o recuperación de información [Ozcan 2004] [Baeza 1999].

### Minería Web

Durante los últimos años se está desarrollando la Minería Web (Web Mining) como resultado de la aplicación de las técnicas de la Minería de Información y Datos a los grandes repositorios de información y conocimiento que están disponibles y accesibles desde cualquier sitio a través de la Web [ARCADIA 2002], así como a los datos generados por la actividad de los servidores Web. Estos procedimientos se están aplicando, por ejemplo, en sistemas de recuperación de información en Internet [Baeza 1999], en análisis de registros de los servidores Web [Chang 2001], en servicios públicos de información [Chau 1999] y en análisis de transacciones electrónicas [Kohavi 2001].

Desde la perspectiva de la minería de datos la Web es una red de páginas conectadas por enlaces, dichas páginas se agrupan en sitios Web soportados por servidores que proporcionan el acceso a ellas. Además, las páginas pueden estar formadas por varios componentes que se integran visualmente, como los marcos a los que se suele considerar como páginas independientes. Asimismo, los contenidos suelen incluirse en las páginas en forma de texto organizado en apartados que se componen de párrafos y palabras, y que a menudo van acompañadas de imágenes u otros recursos audiovisuales y enlaces vinculados a unos y otros.

La Minería Web se usa generalmente de tres formas distintas [Garofalakis 1999] [Srivastava 2000] [Chang 2001]: minería de contenidos Web, minería de estructura Web y minería de uso Web. A continuación se habla brevemente de cada una de ellas.



La minería de contenidos Web permite el descubrimiento automático de patrones en los documentos Web y la extracción de información útil de sus contenidos, sobre todo cuando son de tipo texto, pero también a veces cuando son imágenes u otros soportes. Los datos que requiere suelen ser fáciles de conseguir, puesto que casi siempre son accesibles a través de la red sin restricciones. Idealmente el contenido de la Web puede verse como un índice de cadenas, palabra, frases y párrafos que se integran en páginas y sitios. Este tipo de minería se emplea fundamentalmente para la recuperación de información y utiliza principalmente las técnicas de minería de texto tratadas anteriormente.

La minería de contenidos Web permite la clasificación automática de las páginas Web de distintas formas, como pueden ser el idioma, identificando la lengua en que están escritas, o la temática, mediante la asignación de palabras claves. En general se parte de un conjunto de categorías predefinidas y se asignan etiquetas a los documentos, tomando como referencia una serie de ejemplos previamente clasificados. En estos procesos se suelen emplear [Nasraoui 2004] procedimientos de clasificación basados en modelos vectoriales, que representan los documentos como vectores de peso de palabras, de los que ya se ha hablado, pero también algunas veces se emplean modelos estadísticos o algoritmos genéticos, entre otros.

La minería de estructura Web permite el descubrimiento automático de esquemas en la estructura de enlaces de la Web, que permiten obtener información de su topología [Nasraoui 2004]. De este modo es posible determinar distintas circunstancias interesantes en la estructura global de la Web, como la existencia de páginas que se comportan como puntos de interconexión *-hubs-* en las que hay muchos enlaces a otras páginas, o revelar páginas que gozan de gran consideración en la red *-authorities-* y están apuntadas por gran número de sitios, o también la existencia de grupos de páginas que se referencian mutuamente de forma cerrada, formando islas, que pueden denotar comunidades o especialidades. De forma parecida es posible analizar la estructura interna de los sitios Web, observando el modo en que sus páginas se enlazan entre sí, como la identificación de páginas de navegación *-slippery pages-* que proporcionan fundamentalmente acceso a otras páginas, o páginas de contenido *-sticky pages-* que son las que incorporan realmente la información publicados por el sitio en la Web.

Vinculado con la minería de estructura Web está el algoritmo HITS (acrónimo del inglés *Hypertext Induced Topic Selection*) de Kleinberg [Kleinberg 1998, 1999] para ordenar por su utilidad las páginas Web que se consideran relacionadas con una dada o con una cuestión planteada *-query-* en el ámbito de recuperación de información, teniendo en cuenta sus grados de *hub* y de *authority*. El procedimiento parte de la idea de que para establecer la utilidad de las páginas en un determinado contexto -establecido por la página de referencia o la *query-* hay que tener en cuenta la consideración de las mismas en la red, que se desprende del modo en que dichas páginas están vinculadas entre sí y con otras. El algoritmo consta de tres fases. La

primera es la creación de un conjunto de páginas base a partir del texto de la página de modelo o la *query*, empleando técnicas de minería de textos, de las que se habló anteriormente, que sólo analizan los contenidos. La segunda fase es la identificación de las páginas candidatas al conjunto de la respuesta final, a partir de las páginas a las que señalan las incluidas en el conjunto base y las páginas que apuntan a las de dicho conjunto. La tercera y última fase es la ordenación y selección de las páginas del conjunto de respuesta a partir de los candidatos, teniendo en cuenta que las relaciones entre *hubs* y *authorities* refuerzan sus respectivos valores, puesto que el valor de un *hub* aumenta si apunta a muchos *authorities* y el valor de un *authority* lo hace cuantos más *hubs* lo señalan.

Los datos que requiere la minería de estructura Web suelen ser fáciles de obtener, porque son accesibles a través de la red sin restricciones. Este tipo de minería permite organizar, agrupar y ordenar los sitios Web e identificar redes sociales, entre otras cosas. De forma ideal la estructura general de la Web puede representarse como un grafo donde las páginas son los nodos y los enlaces las aristas que los unen, y de forma análoga la estructura interna de los sitios Web puede verse como un grafo de páginas, compuestas de elementos de información, y vinculadas por enlaces.

La minería de uso Web facilita el descubrimiento automático de patrones en el acceso a los sitios Web. A pesar de que los datos tratados en todas las modalidades de minería Web cambian con el paso del tiempo, la información de uso es la más variable. Los datos necesarios para el análisis de uso suelen ser acceso restringido para las entidades propietarias de los sitios Web y resultan difíciles de conseguir para agentes no vinculado con ellas. Las aplicaciones más comunes de este tipo de minería son la caracterización de los usuarios, el seguimiento de su actividad en los sitios Web, la personalización, y cada vez más la recuperación de información. De modo ideal, el uso de la Web representa sus datos en forma de perfiles que pueden recoger información sobre los sitios visitados, las trayectorias de navegación, los documentos leídos, las preguntas realizadas y cualquier aspecto relevante de las actividades llevadas a cabo.

Los datos que se emplean en el análisis del uso de un sitio Web son fundamentalmente: las peticiones de recursos *-hits-*, que son los ficheros solicitados al servidor por los navegadores, entre los que están los documentos HTML y XML, y también las imágenes y los ficheros audiovisuales en distintos formatos; las páginas vistas, que representan las agregaciones de recursos que aparecen simultáneamente en una página Web; y las sesiones de usuario, que identifican secuencias de páginas o recursos visitados por un usuario durante un periodo de actividad continuada con el sitio Web. Sin embargo, existen otros datos que pueden tener interés para este tipo de estudios, aunque son menos empleados, como el registro de la interacción con las páginas Web en el navegador, que se emplea en una de sus modalidades en el presente

trabajo y del que se habla en la sección de interacción persona-ordenador de este mismo capítulo.

La minería de uso de la Web emplea dos fuentes de datos principales [Nasraoui 2004]. La primera es el registro de actividad de bajo nivel de los servidores Web -log del servidor o *Web log*-, que registra todas las peticiones de recursos solicitados al sistema por los navegadores Web. La segunda es el registro de actividad de alto nivel de las aplicaciones Web -*application log*-, que recoge las acciones relevantes para la aplicación realizadas por los usuarios, y que pueden dar sentido a los registros de bajo nivel en determinadas arquitecturas, por ejemplo en entornos donde hay páginas que se generan en tiempo de ejecución y pueden aparecer en el registro del servidor como accesos a los mismos recursos, cuando realmente incluyen contenidos diferentes.

Las arquitecturas de aplicaciones Web de varios niveles son muy frecuentes actualmente, en concreto: las de dos niveles, en las que el servidor Web genera las páginas dinámicamente al ser solicitadas ejecutando programas -genéricamente referenciados como CGIs- a partir de datos almacenados en una base de datos; y las arquitecturas de tres niveles, en las que el servidor Web recurre a un servidor de aplicaciones para generar las páginas bajo demanda, también a partir de una base de datos. Los servidores de aplicaciones permiten el seguimiento preciso de las sesiones de los usuarios, imprescindibles para los procesos transaccionales -característicos, por ejemplo, de las aplicaciones comerciales- que requieren un control de secuencias de acciones que los servidores Web esenciales no permiten.

```
62.36.67.34 - - [20/Apr/2004:20:02:35 +0200] "GET /img/version.gif HTTP/1.1" 304 -
62.36.67.34 - - [20/Apr/2004:20:02:35 +0200] "GET /img/note.gif HTTP/1.1" 304 -
213.37.225.150 - - [20/Apr/2004:20:02:35 +0200] "GET /5.0a/monitorKC.pl?est=1 HTTP/1.1" 200 330
62.151.101.192 - - [20/Apr/2004:20:02:47 +0200] "GET /5.0a/controlNT2.pl?b=KC_SOG2401&t=18 HTTP/1.1" 200 4531
62.151.101.192 - - [20/Apr/2004:20:02:47 +0200] "GET /5.0a/monitorKC.pl?est=1 HTTP/1.1" 200 330
```

*Figura 2-12. Ejemplo de fichero de registro de actividad de servidor Web*

Los ficheros de registro de actividad de los servidores Web [LogWeb 2008] suelen incluir ciertos datos de forma estándar, aunque es frecuente que se puedan configurar para guardar datos adicionales, atendiendo a las necesidades de su aplicación. Cada línea de los ficheros de *Web log* corresponde a la petición de un recurso solicitada por un cliente Web. Entre los datos que suelen aparecer en los ficheros log están: la dirección IP -que identifica un equipo conectado a la Web- desde la que se realizó la petición, el momento en que se recibió, la línea de petición del cliente incluyendo el modo en que se hizo la solicitud mediante el protocolo HTTP [Http 2008] (generalmente GET o POST) junto con la identificación del recurso y el protocolo utilizados, el código de resultado de operación enviado al cliente, así como el tamaño del objeto devuelto. Además, estos ficheros pueden incluir otros datos como: un identificador del cliente, un identificador del usuario, y las cabeceras de la petición HTTP que indican el

---

servidor desde el que se pide el recurso y el tipo de navegador utilizado para la solicitud. Los datos que no están disponibles en las líneas de registro se sustituyen por un guión en la posición correspondiente. En Figura 2-12 se puede ver un fragmento de un fichero de registro de actividad común de un servidor Web.

En la tercera línea, se puede apreciar la dirección IP del cliente (213.37.225.150), el identificador del cliente y del usuario (representados por guiones al no estar disponibles), la línea de petición (entre comillas) que incluye el protocolo (GET), la identificación del recurso (/5.0a/monitorKC.pl?est=1) y el protocolo empleado (HTTP/1.1), así como el código devuelto por el servidor (200) y el tamaño del elemento enviado (330).

El proceso de análisis de los ficheros de registro de actividad de los servidores Web comienza con un preproceso que consta de cinco pasos fundamentales [Nasraoui 2004]. El primero es eliminar las entradas de peticiones de recursos poco significativos para el estudio, que normalmente son solicitudes de ficheros de imágenes y multimedia. El segundo paso es eliminar las entradas generadas por agentes software *-spiders & crawlers-* utilizadas por los indexadores de la red y otros sistemas automáticos que acceden a los sitios Web, que tampoco suelen aportar información útil para el proceso. El tercer paso es reconocer las peticiones realizadas por cada usuario, en los ámbitos de las visitas aisladas y de las secuencias de visitas, mediante la identificación de los visitantes registrados en el sistema, el empleo de *cookies* -que son cadenas de texto que, con permiso de los usuarios, los servidores pueden depositar y obtener de los navegadores para mantener un seguimiento de la interacción con ellos-, o con la ayuda de métodos heurísticos para establecer las posibles identidades de los visitantes a partir de otros datos disponibles en el fichero de registro, como la dirección IP o el identificador del navegador. El cuarto paso del preproceso es determinar las sesiones de usuario, analizando las secuencias de peticiones de cada usuario con un máximo de tiempo entre peticiones consecutivas, intervalo que se establecen de forma empírica. Por último, el quinto paso es determinar las trayectorias seguidas por los usuarios a través de las páginas del sitio Web [Huysmans 2004], para ello hay que tener en cuenta que no todas las visualizaciones de las páginas provocan peticiones al servidor, debido a diversos mecanismos de almacenamiento intermedio *-caching-* que la Web proporciona -en navegadores y servidores PROXY- para reducir el tiempo de acceso a los recursos y el tráfico en la red. El establecimiento de estas trayectorias suele requerir de un conocimiento sobre la estructura interna de los sitios Web que, habitualmente, no es evidente en los ficheros log.

Después del preproceso, el fichero de registro está en un formato adecuado para aplicar las técnicas de minería para descubrir los patrones y las reglas que describen los comportamientos de los usuarios [Huysmans 2004]. Ejemplos de este tipo de patrones y reglas pueden ser: el porcentaje de los visitantes que van a una página después de pasar por otra, o el número de

usuarios registrados que acceden al sitio Web una vez a la semana por lo menos. A esta etapa se la conoce como la fase de descubrimiento de patrones.

Entre las aplicaciones del descubrimiento de patrones de uso en los registros Web, Chang menciona las siguientes [Chang 2001]: análisis estático, que permite obtener información para la monitorización del sistema; descubrimiento de reglas de asociación, para identificar páginas que se visitan juntas con una frecuencia superior a un valor establecido; realización de agrupaciones, que permite descubrir nuevas clases entre los registros correspondientes a perfiles de uso; realización de clasificaciones, que permite catalogar los registros en clases previamente establecidas; y descubrimiento patrones secuenciales, que permite identificar patrones que se repiten con una cierta frecuencia en los ficheros log.

Para Nasraoui [Nasraoui 2004] la minería de uso Web puede servir para mejorar la usabilidad de los sitios Web, por ejemplo: para descubrir la conveniencia de establecer enlaces entre páginas que no los tenían originalmente y hacer más navegables los sitios; para detectar modos de uso con los que dirigir los sitios Web al empleo que se les da y con los que intentar predecir lo que se hará con el sistema a partir de cada momento; o para identificar perfiles de usuario con los que realizar recomendaciones basadas en filtrado colaborativo -del que se habla en otro apartado de este capítulo-.

### **2.3 Sistema KnowCat (KC)**

En esta sección se habla del sistema KnowCat [Alamán 1999] [Cobos 2003], que es uno de los elementos fundamentales que han servido de estímulo, punto de partida, referencia y plataforma de prototipado para el trabajo presentado en esta tesis. La descripción del sistema KnowCat (KC) presentada en este apartado se ha realizado tomando como referencia [Cobos 2002a].

KnowCat es un entorno basado en Web que permite la gestión del conocimiento gracias a la interacción de una comunidad de usuarios sin necesidad de supervisión humana. El objetivo de esta gestión es provocar la selección del conocimiento disponible en el dominio para obtener su mejor representación o, dicho de otro modo, catalizar el proceso de cristalización del conocimiento en un dominio. Esta es la característica que el acrónimo KnowCat pretende resumir a partir de la expresión en inglés del término “catalizador de conocimiento”, “Knowledge Catalyser”. Pero también es fundamental el modo en que KC lo consigue: valiéndose de la interacción con el sistema de una comunidad de usuarios, que regulan su actividad sólo con las normas incorporadas en el entorno, pero sin ningún tipo de supervisión humana.

KC permite trabajar con conocimiento explícito, del que puede manejarse en forma de documentos de algún tipo, como textos, imágenes, secuencias de vídeo o esquemas

organizativos, por ejemplo. Además, asume que aunque el conocimiento tratado pueda estar en evolución debe ser propenso a estabilizarse en algún estado y perdurar en el tiempo, como ocurre con el conocimiento enciclopédico, el documental, el derivado de la investigación o el debido a la organización del propio conocimiento.

En el sistema, el dominio de conocimiento se representa por un árbol jerárquico: su raíz corresponde al dominio de conocimiento sobre el que se está trabajando, la primera ramificación refleja la división inicial del dominio en temas principales o áreas de conocimiento, la siguiente la subdivisión en temas de las áreas, y así sucesivamente. Este proceso se puede repetir tanto como los usuarios implicados en la definición lo consideren oportuno, el número de niveles establecidos determinará el grado de detalle alcanzado. Normalmente, la resolución más adecuada vendrá dada por el dominio, la comunidad y los objetivos perseguidos.

Cada rama del árbol puede tener dos tipos de elementos asociados: un conjunto de descripciones y un grupo de refinamientos. El primero, es una serie de documentos o unidades atómicas de conocimiento, alternativas entre sí, que compiten por describir de la mejor manera el tema correspondiente. El segundo, es un conjunto de alternativas candidatas a definir la descomposición en subtemas del tema en cuestión, dichos subtemas son, por tanto, nodos del árbol de conocimiento y subramas de la rama correspondiente.

El documento es la unidad atómica de conocimiento que el sistema reconoce. Se considera como tal cualquier paquete de información al que el autor confiere un sentido completo en el contexto del dominio en el que se está trabajando, y que, directa o indirectamente, mediante una representación, pueda ser tratado con procedimientos informáticos. Los documentos clasificados en cada uno de los nodos del árbol de conocimiento están en competencia con el objetivo de determinar la mejor descripción del tema representado por los vértices correspondientes. La opinión de los usuarios sobre los distintos documentos, expresada de forma explícita mediante votaciones e implícita por la interacción con ellos, al utilizar el sistema, se emplea para la selección, cristalización, de los documentos más populares para cada tema.

Para permitir añadir información a los documentos preservando su integridad, el sistema proporciona las anotaciones. Con ellas los usuarios pueden opinar o proponer modificaciones sobre los documentos sin alterarlos. Los autores pueden utilizar las anotaciones para añadir información en sus documentos ya publicados, como aclaraciones o pequeños cambios. De forma parecida, para el caso en que el propio autor de un documento quiere cambiarlo total o parcialmente de forma más profunda, el sistema proporciona las versiones. Con ellas los autores pueden proponer nuevos documentos candidatos a sustituir a otros anteriores. Generalmente la sustitución requiere la aceptación de la comunidad de usuarios. Con las anotaciones y versiones el sistema permite la evolución de los documentos en su seno.

Árboles de conocimiento, documentos, versiones y anotaciones son unidades de conocimiento. Todas ellas están bajo un proceso de cristalización específico, que se adapta a las características de cada tipo de unidad y a la clase de interacción que los usuarios pueden mantener con ella. En general el proceso de cristalización procura seleccionar un elemento dominante de entre todos los candidatos de una categoría que están en competición. Así, por ejemplo, ocurre en cada tema con los documentos que pretenden describirlo, por un lado, y los conjuntos de subtemas que intentan refinarlo, por otro. En el caso de las anotaciones, el proceso de cristalización no pretende la selección de sólo una de ellas, sino que tolera la persistencia en el sistema de las candidatas más votadas. En todos los casos los candidatos que no reciben el suficiente apoyo desaparecen del sistema.

El proceso de cristalización de las unidades atómicas de conocimiento, los documentos, se basa en el cálculo del llamado grado de cristalización. Se trata de una medida de la aceptación de los usuarios recibida por el documento. En este cálculo intervienen tanto las consideraciones explícitas, expresadas mediante votos y anotaciones, como las implícitas, debidas a la interacción con el documento de los usuarios. La cristalización es en definitiva la evolución de tal aceptación, de forma que un documento aceptado de forma mantenida durante un periodo de tiempo es candidato a cristalizar, y otro que no, aspirante a desaparecer del sistema. Para que posteriormente un documento cristalice es preciso que su grado de aceptación supere un cierto umbral establecido inicialmente. Sin embargo un documento cristalizado también puede dejar de estarlo, esto ocurre si después de haber alcanzado dicho estado, durante un periodo de tiempo determinado su grado de aceptación decae por debajo de otro umbral. En todos los casos lo que se valora es la evolución del grado de aceptación. Aunque la opinión de todos los usuarios se suele tomar en cuenta para evaluar la aceptación, la de algunos considerados expertos, porque han aportado al sistema algún documento que ha cristalizado, se le asigna más valor.

Los usuarios son el motor fundamental del funcionamiento de KC, sin ellos no habría conocimiento y sería imposible provocar su selección. Como en determinados momentos del desarrollo de los nodos no se puede garantizar la suficiente masa crítica de usuarios para que los procesos de cristalización funcionen adecuadamente, los nodos del sistema pasan por una serie de fases de maduración. En cada ocasión la organización de los usuarios depende de la fase en la que se encuentre el nodo correspondiente, que también determina su modo de funcionamiento. Actualmente en el sistema se consideran tres, la Fase Inicial, la Fase Activa y la Fase Estable.

En la Fase Inicial de los nodos, como no hay masa de usuarios ni conocimiento suficiente para el correcto funcionamiento de los mecanismos de cristalización, se establece un Grupo de Coordinación que se encarga de supervisar la actividad del nodo hasta que éste pueda empezar a funcionar de forma autónoma. Durante esta fase el Grupo de Coordinación se encarga de tomar

decisiones que en etapas posteriores se alcanzarán con la colaboración de toda la comunidad. El resto de usuarios del sistema constituyen, en esta fase, el Grupo de Colaboradores. Durante la etapa inicial los usuarios en esta categoría sólo pueden ingresar documentos en el sistema y revisar los documentos añadidos por otros usuarios, su actividad no se considerará, como después se hará, para la cristalización del conocimiento.

El Grupo de Coordinación se constituye cuando se crea un nuevo nodo de KC, aunque se pueden añadir nuevos miembros posteriormente. Una de las principales tareas del grupo es elaborar el árbol de conocimiento del nodo, haciendo propuestas y opinando sobre ellas, a través del sistema, hasta promover un consenso mediante un sistema votaciones. Durante esta etapa todos los usuarios del nodo de KC pueden ingresar documentos en el sistema, pero sólo los pertenecientes al Grupo de Coordinación pueden opinar sobre ellos e influir en la cristalización de los mismos. En este sentido y durante la etapa inicial el Grupo de Coordinación ejerce labores de Comité Editorial. Es el Grupo de Coordinación el encargado también de decidir el fin de la fase inicial, promover su propia disolución y hacer que el sistema empiece a funcionar en modo no supervisado, pasando a la denominada Fase Activa.

En la Fase Activa existe una masa de usuarios, y conocimiento, suficiente para el correcto funcionamiento de los mecanismos de cristalización. En este caso todos los usuarios de un nodo del sistema se integran en la llamada Comunidad Virtual del nodo. Los usuarios del nodo pueden ingresar documentos en el sistema utilizando el árbol de conocimiento definido en él. Los documentos están sometidos a un proceso de cristalización en cada nodo del árbol. Cuando un documento cristaliza en una rama, el usuario se convierte en Experto en el área de conocimiento correspondiente y en los temas más cercanos en el árbol de conocimiento. Los usuarios expertos reciben un determinado número de derechos de voto que pueden emplear en apoyar otros documentos ubicados en los nodos donde son considerados expertos. Este mecanismo está inspirado en la práctica de “*peer review*” que emplean la mayoría de las revistas científicas.

El grupo de usuarios Expertos en un tema se reúne en la llamada Comunidad Virtual de Expertos del tema. Pertenecen a ella los autores de documentos que han cristalizado en ese tema, en el tema del que desciende, en los temas que tiene como descendientes y en los temas que están a la misma altura en el árbol de conocimiento. De este modo la las Comunidades Virtuales de Expertos están estrechamente ligadas al árbol de conocimiento del nodo.

Estando el sistema en la Fase Activa, puede llegarse a una situación en la que ya no haya apenas cambios en el árbol de conocimiento, pocos documentos se publiquen en el sistema, la mayoría de la actividad en el nodo sea de consulta y el conocimiento esté muy cristalizado. En estas



condiciones se podría considerar que el sistema habría alcanzado una Fase Estable que duraría hasta que la situación cambiara y el nodo se reactivara para volver a la fase anterior.

## **2.4 Conclusiones**

En esta sección se sintetizan las ideas y referencias más destacadas de los campos de investigación revisados, Gestión del Conocimiento, Interacción Persona-Ordenador, Web Semántica, y Minería de Información y de Datos. También, se recapitulan algunos sistemas que se consideran de especial interés para el trabajo realizado. Además, todas las cuestiones recogidas se relacionan con la aproximación propuesta.

### *Gestión del Conocimiento*

Semantic KnowCat (SKC) es el entorno propuesto para comprobar los aspectos fundamentales de la aproximación planteada, de él se habló someramente en el capítulo anterior y nos ocuparemos en profundidad en el siguiente. SKC es un sistema de gestión de conocimiento y como tal, proporciona los recursos adecuados para las acciones de recolección, organización, refinamiento y distribución del conocimiento [Benjamins 1999] [Malthotra 2000], que permiten su administración, y también para cubrir las necesidades de interacción de sus usuarios [Churman 1971], que Churman sugiere para reforzar la gestión y facilitar la interacción entre los usuarios y de estos con el sistema.

Pero además, dicho entorno facilita la adquisición automática de conocimiento, empleando la información obtenida a partir de los datos disponibles. Para ello el sistema busca las relaciones existentes entre datos, información y conocimiento [Coakes 2003] en el ámbito de la aplicación. Los datos considerados son el propio conocimiento, la información utilizada para el funcionamiento del sistema, así como los registros de la actividad del mismo y de los usuarios que lo utilizan. Las ideas de Benjamins [Benjamins 1999] tienen especial importancia en la definición de los mecanismos de enriquecimiento del conocimiento, de registro de actividad del sistema y de adaptación de los contenidos para los interlocutores del entorno, que se incorporan en sistema SKC.

En SKC las Tecnologías de la Información y el Conocimiento (TIC) se emplean especialmente en la captación y tratamiento de datos, pero también en el proceso de la información y en la gestión y la explotación del conocimiento. A pesar de las consideraciones de Davenport [Davenport 1999] sobre el aprovechamiento de las TIC para procesar conocimiento, la contribución de estas tecnologías como mediadoras de la participación humana y el tratamiento de los datos generados por su actividad en relación con el conocimiento, resultan muy positivas para la gestión del mismo en todas las acciones que según Benjamins [Benjamins 1999] involucra el proceso. Además, la aproximación propuesta aplica las TIC para descubrir el

conocimiento y hacer que éste esté disponible donde y cuando se necesita y para quien lo requiera, siguiendo algunas ideas de Nonaka [Nonaka 1995]. De acuerdo con Alavi, todo esto contribuye a que sea más útil el proceso del conocimiento y su aprovechamiento para la comunidad de usuarios que lo mantiene y explota [Alavi 2001] [Gottschalk 2004].

La aproximación propuesta comparte con Ackerman [Ackerman 2003] la idea de que los sistemas de gestión del conocimiento no pueden ser meros repositorios de conocimiento explícito para poner en común, sino que es necesario considerar el papel fundamental de las personas con experiencia como depositarios de conocimiento tácito, que también hay que gestionar y compartir. En la aproximación propuesta los usuarios del sistema se consideran parte del conocimiento gestionado.

El sistema de gestión de conocimiento planteado se preocupa principalmente de la organización de dicho conocimiento para facilitar su reunión e intercambio, encuadrándose en el enfoque de la segunda de las escuelas de gestión consideradas por Earl [Earl 2001] [Gottschalk 2004]. Por otro lado, considerando la clasificación de Benjamins [Benjamins 1999], aunque los nodos -instancias- de SKC se dirigen a temas particulares, el sistema tiene una vocación generalista que permite su aplicación en múltiples contextos. Además, siguiendo con la taxonomía del mismo autor, el sistema se fundamenta en la colaboración entre los usuarios y pretende ocuparse de ella como fuente de conocimiento, a pesar de que se dirige a la gestión de documentos tangibles. Finalmente, desde la perspectiva de Cobos [Cobos 2002], SKC como KnowCat (KC) es un sistema que permite la gestión de conocimiento estructurado, facilitando soporte para su realización colaborativa, aunque a diferencia de KC el sistema propuesto emplea ontologías para representar el conocimiento compartido e ideas de los sistemas de recomendación para facilitar la interacción con grafos de conocimiento como estructuras menos rígidas.

### Interacción Persona-Ordenador

Como se adelanto en la introducción de este documento, del campo de Interacción Persona-Ordenador y de las áreas de conocimiento vinculadas se toman diversas ideas y técnicas que resultan esenciales para la aproximación planteada. Groupware, Awareness, Monitorización de Actividad, Hipermedia Adaptativa, Recomendación, Visualización de la Información, Procesamiento del Lenguaje Natual, y la Web 2.0, son los ámbitos más importantes de la Interacción Persona-Ordenador que se toman como base para el sistema SKC propuesto.

### **Groupware**

Además de un sistema de gestión de conocimiento, y precisamente por eso [Churman 1971], el entorno propuesto para comprobar los fundamentos de la aproximación planteada, SKC, es una aplicación de trabajo en grupo mediante ordenadores conectados en red -groupware-, que se

enmarca en el área del CSCW [ForakerDesign 2005] dentro del campo de la Interacción Persona-Ordenador [SGICHI 2008].

El sistema SKC permite la colaboración entre los usuarios tanto si coinciden en el espacio y en el tiempo, como si no lo hacen, por lo que su clasificación con estos criterios [Johansen 1988] no parece muy relevante, aunque pone de manifiesto la diversidad de contextos potenciales de utilización. Además, SKC se puede considerar un sistema de trabajo colaborativo, porque proporciona un entorno de uso general que fomenta la comunicación y la coordinación en el seno de una comunidad de usuarios [OpenGroupware 2008], debido a que la gestión de conocimiento que el sistema proporciona puede aplicarse a un amplio rango de actividades en diversos ámbitos.

Aunque las aplicaciones colaborativas de uso voluntario, como las de carácter lúdico, tienen más aceptación que las de uso obligatorio [Palen 2002], la aproximación propuesta permite que en condiciones impuestas, los usuarios obtengan beneficios de su participación en aplicaciones en las que prima una perspectiva de uso, se obvia la tecnología, y no se establecen grandes regulaciones. Condiciones que para Grudin [Grudin 1994a] son fundamentales para el éxito de las aplicaciones colaborativas.

Brinck [ForakerDesign 2005] señala que el número de usuarios es un factor esencial en las aplicaciones de trabajo en grupo, porque afecta al comportamiento del colectivo y a los requerimientos de los sistemas. La facilidad de uso, la sensibilidad del sistema y su fiabilidad, son también cuestiones fundamentales en este tipo de aplicaciones, así como la preferencia de los usuarios por interfaces menos generales, que apunta Bargerón [Bargerón 2004].

Por su parte, Grudin identifica varios aspectos problemáticos de las aplicaciones de trabajo colaborativo que los diseñadores deben considerar [Grudin 1994a]. Entre ellas están la compensación entre esfuerzo realizado y beneficio obtenido por los usuarios, la orientación de la aplicación al interés general más que particular, la masa crítica de usuarios requerida, la violación de límites sociales establecidos, el tratamiento de excepciones en las actividades grupales, el manejo del uso poco frecuente, la dificultad de evaluación, lo poco intuitivo del diseño, y la dificultad de integración en los entornos de uso. Las actividades ajenas a las aplicaciones de trabajo colaborativo son también fundamentales en los resultados obtenidos con ellas [Cadiz 2000].

Todos estos aspectos y muchos otros deben tenerse en cuenta en el análisis y diseño de las aplicaciones de groupware [Gea 2002], por lo que se ha procurado considerarlos en la aproximación planteada, y en los prototipos y experiencias realizadas.

### **Awareness**

En el ámbito del CSCW el *Awareness* hace referencia la conciencia de la actividad de los demás participantes en el entorno colaborativo [Dourish 1992], que sirve para poner en contexto la propia actividad y controlar el proceso de trabajo grupal. La información necesaria puede conseguirse directa y explícitamente, y mostrarse separada de los elementos compartidos, o recopilarse de modo implícito y mostrarse en el espacio de trabajo común. Existen varias formas de proporcionar *awareness* [Dourish 1992], como proporcionar roles que establecen las relación de los usuarios con el entorno, facilitar mecanismos para que los usuarios puedan informarse de su actividad, o suministrar procedimientos para evidenciar la actividad colaborativa individual en el grupo.

El *awareness* puede tener distinta función dependiendo de si el entorno CSCW es síncrono o asíncrono [Sohlenkamp 1994]. En entornos síncronos se ocupa de mostrar la situación de los usuarios en cada momento, cuestiones como qué hacen y con qué, su disponibilidad o la información propia proporcionada. En entornos asíncronos se encarga de informar sobre lo que han hecho los usuarios recientemente, sobre qué objetos se han cambiado, de qué manera y por quien, y acerca de los mensajes pendientes.

La conciencia del sistema sobre el contexto de la interacción con él es otro tipo de *awareness*, [Byun 2001] que proporciona a los sistemas la posibilidad de adaptarse a la situación de uso. Entre los posibles problemas que el *awareness* podría provocar están [Hudson 1996] la violación de la intimidad individual y la sobrecarga de información generada.

En nuestra aproximación se proporciona diversos procedimientos para incorporar el *awareness* en el proceso de gestión colaborativa de conocimiento, de acuerdo con [Dourish 1992]. En primer lugar, mediante la asignación de distintos roles a los usuarios participantes. En segundo lugar, gracias a la posibilidad de mostrar en el área compartida las observaciones de los usuarios sobre los elementos que constituyen el repositorio de conocimiento. En tercer lugar, a través de la publicación de los resultados de la opinión colectiva respecto a los elementos de de conocimiento, fruto del proceso de cristalización de los mismos que manifiesta la selección realizada por el grupo de forma explícita, mediante votaciones y anotaciones, pero también implícita, considerando el interés aparente por los elementos de conocimiento. En cuarto lugar, se contempla la generación de informes de actividad destinados a roles específicos y accesibles bajo demanda fuera del entorno compartido, así como la gestión del propio perfil en el sistema. Además, el entorno incorpora los mecanismos de identificación y seguimiento de actividad necesarios para soportar funcionalidades de carácter síncrono y asíncrono como las apuntadas por [Sohlenkamp 1994]. Por último, en todos los casos, no se pierden de vista las consideraciones de [Hudson 1996] sobre violación de intimidad y sobrecarga de información, y

se ha considerado en el diseño la posibilidad de adaptación la situación de uso, sugerida por [Byun 2001].

### ***Monitorización de la Actividad del Usuario***

La arquitectura cliente/servidor predominante en Internet y en la Web ha dificultado el desarrollo de técnicas de monitorización de actividad continua, contextualizada y completa de los usuarios en este entorno, debido a las dificultades para recopilar datos en el lado del cliente, donde está la mayoría de la información necesaria para ello [Fenstermacher 2002a]. El empleo de los eventos [Hilbert 2000] de las interfaces gráficas de usuario basadas en ventanas proporciona una alternativa para este seguimiento en el cliente. El modelo de objeto de documento, DOM [DOM 2008], establece una serie de eventos que incorporan los navegadores Web actuales más populares, con los que es posible llevar un control de algunas interacciones significativas de los usuarios con las páginas Web, como los objetos sobre los que se realizan las operaciones, los cambios en las áreas de visualización o las operaciones de entrada de datos mediante varios periféricos. Lenguajes como JavaScript [JavaScript 1999] permiten manejar dicho objeto de documento desde las páginas Web.

El objeto DOM [DOM 2008] y sus eventos [DOMEvent 2008] se han empleado en aplicaciones como los mapas de clics [Crazyeeg 2008] [ClickHeat 2008], que sirven para el análisis de las interfaces de usuario en el campo de la interacción persona-ordenador. Sin embargo, el control de los eventos de usuario no permite un seguimiento completo de la actividad, puesto que existen periodos de tiempo durante los que no hay eventos, aunque la interacción de los usuarios con las páginas Web continua, por ejemplo mientras se revisa un texto o un gráfico. Este tipo de actividades se puede monitorizar con técnicas más sofisticadas, como el seguimiento de los ojos [Hassan 2007]. El inconveniente de estos procedimientos es que requieren de equipamientos especiales para su realización.

En la aproximación planteada en este trabajo se ha empleado la monitorización de interacción del usuario con el cliente Web basadas en eventos del objeto de documento DOM [W3C 2004] [DOM 2008], que proporciona un seguimiento mejorado de la interacción del usuario respecto al análisis del registro de actividad en el servidor, pero sin la necesidad de los recursos especiales requeridos por otras técnicas más sofisticadas y completas como el seguimiento de ojos [Hassan 2007]. Con su empleo se ha dotado al sistema SKC propuesto de un medio de obtener datos sobre el interés aparente de los usuarios hacia los elementos de conocimiento objeto de gestión, como alternativa o complemento al requerimiento de manifestaciones explícitas de opinión de los usuarios del entorno.

### ***Hipermedia Adaptativa***

La Hipermedia Adaptativa es una línea de investigación dentro de los sistemas adaptativos de usuario, su objetivo es incrementar la funcionalidad de la hipermedia haciéndola personalizada. Los Sistemas Hipermedia Adaptativos reflejan algunas características del usuario en un modelo de usuario que aplican para adaptar diversos aspectos del sistema [Brusilovsky 1998]. Existen tres etapas en el proceso de adaptación, la de recolección de datos del usuario, la de procesamiento de esos datos para construir o actualizar el modelo de usuario, y la de aplicar el modelo. El seguimiento de la actividad del usuario en el entorno hipermedia suministra insuficiente información para modelar al usuario, la solución del modelado de usuario más o menos automático depende del área de aplicación y de la disponibilidad de fuentes adicionales de información. En general se puede involucrar al usuario en alguna de las tres etapas del proceso de modelado, según en cuál se haga la inversión y beneficios serán diferentes.

La incorporación de la adaptación en el sistema SKC propuesto está relacionada con el reforzamiento de la interfaz de usuario, para manejar la potencia proporcionada por las ontologías y la oferta de información adicional. Pero además sugiere la posibilidad de hacer que el sistema se adapte a otros elementos activos con los que interacciona, como es el caso de los agentes software.

Un Documento Virtual es un documento Web cuyos contenidos, enlaces, o ambas cosas, se crean según se necesitan [Milosavljevic 1999]. Existen varias clases, basados en plantillas, generados a partir de fragmentos de uno documento o varios documentos, como resultado de aplicaciones que proporcionan valores de un solo uso, o también a partir de repositorios de información. Martins propone un procedimiento de generación de documentos hipermedia virtuales mediante adaptación de documentos en tres fases [Martins 2002]: filtrado, que selecciona los documentos; empaquetado, que determina qué partes de ellos se presentarán; y formateado, que establece el aspecto final. Para ello se consideran modelos de documento, usuario, empaquetado y formato. Los Documentos Virtuales proporcionan el soporte para la adaptación del sistema propuesto a los usuarios y otros elementos del mismo, y el seguimiento de la actividad de todos ellos en el mismo.

### ***Sistemas de Recomendación***

Los sistemas de recomendación [Adomavicius 2005] pretenden filtrar la información disponible para identificar los ítems de la misma que pueden tener más interés para cada usuario, eliminando automáticamente los que no son deseados o resultan irrelevantes y reduciendo así el impacto de la sobrecarga de información. Para ello este tipo de sistemas suele comparar el perfil de un usuario con las características de otros o de los elementos de información considerados. Hay sistemas de recomendación [Adomavicius 2005] que se basan en los contenidos que los

mismos usuarios eligieron en el pasado, otros sistemas que se fundamentan en principios colaborativos, considerando las elecciones de otros usuarios con parecidas características, y los sistemas de recomendación que combinan ambas estrategias. Las dos primeras modalidades tienen algunas limitaciones que pueden reducirse al combinarlas. Todas las variedades de sistemas de recomendación requieren representar los elementos que comparan en un formato adecuado para ello y establecer una medida apropiada para hacerlo.

La navegación colaborativa es una aplicación sugestiva de sistema de recomendación, que puede mejorar la manera de encontrar y manejar información en la Web o en una red de información enlazada, ofreciendo a los usuarios datos relevantes obtenidos automáticamente o por otros usuarios dispuestos a compartirlos. Para ello se puede [Kurhila 2002] [Barra 2002]: convertir el sistema en un PROXY, que intermedie entre la comunidad de usuarios y las fuentes Web, sustituyendo los enlaces de las páginas por vínculos que pasan a través de él; incluir en los clientes módulos especiales con conexión abiertas permanentemente para actualizar el estado del sistema sin retardos; o extraer información de la estructura de enlaces de las páginas Web.

En la aproximación planteada en este trabajo, los sistemas de recomendación se incorporan como medio de aprovechar la actividad residual de la comunidad de usuarios, para facilitar la actividad de sus integrantes, para hacer evidentes las relaciones ocultas que se establecen entre usuarios y elementos de conocimiento, y para que todo ello redunde en la mejora de la gestión de conocimiento que el sistema realiza y en la reducción de la sobrecarga de información existente.

### ***Visualización de la Información***

La Visualización de la Información pretende producir representaciones gráficas de la estructura de información abstracta para usuarios humanos, del mismo modo que la Web Semántica intenta generar un marco de descripción de recursos universal adecuado para los agentes software. Ambas pueden complementarse en la organización de recursos de información y su acceso. Sin embargo, difieren en algunos aspectos fundamentales de la forma de definir y representar la semántica, mientras la Web Semántica se preocupa de las máquinas, la visualización lo hace de los usuarios humanos [Geroimenko 2002]. El objetivo de la Visualización de la Información es revelar patrones, tendencias y otros discernimientos de un fenómeno, centrándose en la información abstracta, y representar la información no espacial y no numérica en una forma visual efectiva. Las interfaces de usuario no suelen beneficiarse del enriquecimiento semántico realizado para las interfaces entre máquinas, aunque existen diversas técnicas de visualización que pueden aplicarse en la Web Semántica.

La Visualización de la Información proporciona un recurso imprescindible para la integración en un sistema de algunas de las ideas, técnicas y otros recursos propuestos en la aproximación

planteada en este trabajo. Sus técnicas son necesarias para dotar al sistema propuesto de la interfaz que permita a los usuarios tratar con el nuevo modelo de representación, visualizar los nuevos contenidos enriquecidos y las ontologías que los representan.

### ***Procesamiento del Lenguaje Natural***

El Procesamiento del Lenguaje Natural (PLN) es una especialidad de los campos de conocimiento de la Inteligencia Artificial y de la Lingüística Computacional, que se ocupa de los mecanismos computacionalmente eficaces para la comunicación utilizando el lenguaje natural entre personas con ayuda de máquinas, y entre personas y máquinas o para facilitar el tratamiento automático de la información en lenguaje natural. Entre las aplicaciones más importantes del PLN están el análisis del lenguaje y su comprensión, el reconocimiento y la síntesis de voz, la traducción automática, la recuperación y la extracción de información, y la síntesis de contenidos.

El empleo de las técnicas de PLN en el campo de recuperación de la información es más común cada vez [Baeza 2004]. Esto también ocurre en el ámbito de la Web Semántica, donde el PLN y la recuperación de información tienen un interesante papel en la construcción de la nueva red, sobre todo en el proceso de transición hacia ella desde la Web actual, como instrumento para facilitar los procesos automáticos o semiautomáticos de anotación, mapeo entre anotaciones de información, etc.

FreeLing [Carreras 2004] es una herramienta de procesamiento de lenguaje que permite analizar un texto para identificar las categorías gramaticales de las palabras que lo forman y determinar el lema al que corresponde cada una de ellas en un diccionario de referencia, con el fin de establecer la interpretación morfológica más probable de dichas palabras. Como se comenta en el apartado sobre minería de información y de datos, al hablar del modelo vectorial en minería de texto, es posible emplear analizadores como Freeling en el procesamiento de documentos textuales para eliminar las variaciones morfológicas de las palabras. Aunque esta práctica no es habitual en los sistemas más populares de recuperación de información en la Web, que comparan textos con una desproporción de tamaño grande -frases sueltas frente a páginas Web-, en el contexto del sistema propuesto en este trabajo, donde se comparan documentos completos con textos de tamaño moderadamente mayor, la aportación de este tipo de aproximación parece positiva [Brants 2004].

En la aproximación planteada, el PLN se considera en el procesamiento de la información textual que integra el conocimiento o que se asocia a los elementos que lo forman y que el sistema planteado maneja. En particular el PLN se aplica a la anotación automática de conocimiento, a su clasificación, y al mapeo entre las ontologías que lo organizan.



## **Web 2.0**

El concepto de Web 2.0 [O'Reilly 2005] hace referencia a las tendencias predominantes en la Web actual, después del desarrollo comercial de la Web durante los noventa, que culminó con la crisis de las compañías de la nueva economía al final de la década. Los principios fundamentales en los que se basa la Web 2.0 [O'Reilly 2005] son: (1) la Web debe considerarse como una plataforma sobre la que se proporcionan servicios, (2) debe aprovecharse la inteligencia colectiva que sustenta la propia Web y las aplicaciones colaborativas -buscadores, wikis, weblogs, etc.- que triunfan en ella, (3) los datos son lo fundamental no las aplicaciones que los manejan, (4) el software es un servicio no un producto, (5) los modelos de programación ligeros como RSS [RSS 2008] son preferibles a los complicados servicios Web, (6) las aplicaciones deben ser distribuidas no estar encerradas en un solo dispositivo, y (7) las interfaces deben estar enriquecidas mediante las tecnologías integradas en AJAX [Woychowsky 2007] [Zakas 2007] para mejorar la pobre experiencia de usuario de la Web tradicional.

AJAX [Woychowsky 2007] [Zakas 2007] es una técnica de desarrollo de aplicaciones Web interactivas, que se centra en la programación del cliente y que integra varias tecnologías preexistentes, entre las que están XML, XHTML, CSS, DOM, JavaScript y XMLHttpRequest, para conseguir una experiencia de usuario similar a la de las aplicaciones de escritorio convencionales, superando las limitaciones de la interfaz Web tradicional. Vinculado con las aplicaciones de la Web 2.0 está también RSS, que es una familia de formatos para la redifusión de contenidos que se actualizan frecuentemente, como titulares de noticias, entradas de Weblogs [Weblog 2008] o paquetes de archivos multimedia digitales, destinados a suscriptores de un sitio Web, para su uso particular o su redifusión en otros sitios Web.

Existen varias aplicaciones características de la Web 2.0, que tienen especial interés para este trabajo, entre las que están Weblogs, Folksonomías y Wikis. Un Weblog [Weblog 2008] es un sitio Web que se actualiza de forma periódica donde las entradas -textos o artículos- se muestran de forma cronológica inversa y los propios autores deciden lo que se publica. Una Folksonomía [Golder 2006] es un procedimiento colaborativo de creación y gestión de etiquetas para la anotación y clasificación de contenidos en una taxonomía plana popular y espontánea, que no requiere de la colaboración de expertos. Wiki [Leuf 2001] es el término que designa a un tipo de aplicación Web, que permite a sus usuarios crear, editar, enlazar y publicar páginas Web de forma fácil, colaborativa y directa, sin la intervención de editores expertos, a través de la propia Web, empleado un navegador convencional.

La aproximación propuesta y el sistema planteado para ponerla en práctica, se inscriben en la corriente que la Web 2.0 representa, especialmente el aprovechamiento de la red, de la inteligencia colectiva y de la colaboración, la visión de las aplicaciones como servicios, la

valoración de los datos por encima de los programas, y la preocupación por conseguir que la interfaz de usuario se adecue a las necesidades de la actividad y de los datos que permite manejar.

### Web Semántica

La Web Semántica [Berners-Lee 2001] es una red global de información enlazada de tal modo que las máquinas puedan procesarla con facilidad. Un concepto fundamental en la Web Semántica es el de ontología [Mahesh 1996], que puede definirse como una representación conceptual compartida del conocimiento, hecha de un modo explícito y formal [Gruber 1993]. La Web Semántica es uno de los pilares fundamentales de la aproximación presentada en este trabajo. Como se verá después, las ontologías constituyen un elemento esencial del sistema SKC propuesto, que las emplea para la descripción del propio sistema y para la representación del conocimiento que gestiona.

La Web Semántica parte de una idea del creador de la Web, Berners -Lee, que propuso una torre de lenguajes para sustentarla sobre la Web actual con siete niveles [Berners-Lee 2000, 2001]: Unicode y URIs, proporcionan una forma estándar de definir referencias a entidades y símbolos de intercambio; XML, XMLS y NS, proporcionan un modo estándar de representar árboles etiquetados, definir gramáticas para crear documentos XML y combinar documentos con vocabularios heterogéneos, respectivamente; RDF y RDFS, el primero aporta la base para el proceso de meta-datos y el segundo un lenguaje de modelado minimalista sobre RDF; Ontología, el lenguaje OWL proporciona un modo común de procesar el contenido semántico de la información de la Web; Lógico, proyecta aportar un lenguaje universal para lógica monótona; Prueba, pretende proporcionar un lenguaje para explicar por qué se ha llegado a cierta conclusión; y Confianza, aspira a establecer una medida de la credibilidad del conocimiento obtenido.

Actualmente los primeros tres niveles están muy afianzados desde hace algún tiempo. El de Ontología, se encuentra en fase de consolidación en este momento, una vez se han hecho converger en el lenguaje OWL las propuestas iniciales, OIL [Fensel 2000] [Harmelen 2000] y DAML [Hendler 2000], y se van desarrollando herramientas para su aplicación, así como ontologías y vocabularios adecuados para su empleo en diversos campos. El quinto nivel, Lógico, está actualmente en la etapa de desarrollo. El resto de niveles están todavía en fases preliminares de definición.

La especificación de la torre de lenguajes para soportar la Web Semántica y su estado de desarrollo, así como las observaciones realizadas por algunos autores sobre su diseño e implementación, han sido de especial importancia para determinar los lenguajes y estándares que se emplean en el sistema propuesto en este documento.

La construcción de las ontologías, elementos fundamentales de la Web Semántica, es un proceso difícil y costoso en recursos y tiempo, entre otras cosas porque requiere del consenso de una comunidad que normalmente tiene entendimiento distinto del conocimiento que se pretende representar. Otro problema importante en el campo de las ontologías es el mapeo entre ellas, donde el inconveniente fundamental sigue siendo la diversidad de entendimiento mencionado [Noy 2002]. Choi [Choi 2006] habla de tres procesos en los que se precisa del mapeo entre ontologías: la integración de varias ontologías de diferentes dominios para generar una nueva en otro dominio; la fusión de ontologías de un mismo dominio para generar otra coherente en él; y la alineación de ontologías estableciendo enlaces entre ellas. Para el mismo autor, hay tres categorías de mapeo entre ontologías: (1) el mapeo entre una ontología global integrada y otras locales, que permite identificar conceptos entre ellas; (2) el mapeo entre ontologías locales, que hace posible transformar las entidades de unas ontologías en las de otras; y (3) el mapeo en la fusión y alineación de ontologías, que sirve para establecer correspondencias entre distintas ontologías e identificar sus conceptos únicos, sinónimos o solapados. Entre los distintos procedimientos propuestos para realizar el mapeo Wiesman sugiere hacerlo de forma automática asociando los nodos de las ontologías comparando las palabras de la información que contienen [Wiesman 2001].

Los problemas derivados del manejo de las ontologías [Noy 2002] constituyen un aspecto fundamental para el diseño, el desarrollo y la explotación del sistema SKC propuesto en este trabajo. En particular, la vinculación entre instancias de SKC a través de la red para compartir e intercambiar conocimiento, que requiere de un mecanismo automático [Wiesman 2001] para identificar relaciones entre las ontologías [Choi 2006] que describen los respectivos repositorios de conocimiento.

Los Agentes Inteligentes están llamados a convertirse en una categoría de usuarios artificiales fundamentales de la Web Semántica. En este momento existe una gran variedad de entidades computacionales que, bajo la denominación de agentes, se aplican en diversos campos. Distintos autores han propuesto definiciones que no siempre coinciden en todos sus términos [Franklin 1994]. Sin embargo se puede establecer una noción débil de agente que involucra autonomía, habilidades sociales, reactividad, y pro-actividad. Estas características son aceptadas generalmente como cualidades clave para evaluar la condición como tales de los agentes [Dinverno 2001]. Por su parte, los Sistemas Multi-Agente son sistemas distribuidos que incluyen varios componentes, cada uno de ellos es un agente independiente capaz de resolver problemas, que colaboran para conseguir algo coherente. Los mismos problemas de indefinición de los agentes surgen con los Sistemas Multi-Agente.

FIPA [FIPA 2008] es una organización de empresas interesadas por el área de los agentes, que ha propuesto una serie de recomendaciones con mucha aceptación sobre el mecanismo de

comunicación entre agentes y sobre la arquitectura de las plataformas para ellos. La recomendación sobre comunicación establece el intercambio de mensajes entre agentes empleando el lenguaje FIPA ACL y HTTP como soporte. Los mensajes pueden escribirse en XML, incluir contenidos definidos libremente por los usuarios para sus agentes y utilizar ontologías para contextualizar tales contenidos. La arquitectura FIPA [FIPA 2008] define un modelo lógico de referencia para la creación, registro, localización, comunicación, migración y baja de los agentes. KQML [Finin 1994] es otro sistema de interacción entre agentes promovido por DARPA que incorpora un lenguaje y un protocolo de comunicación. KQML tiene unas características muy similares a FIPA ACL, aunque se diferencian algunos detalles.

En la actualidad, hay varias plataformas de agentes capaces de proporcionar soporte para la comunicación y la coordinación entre agentes, entre ellas pueden mencionarse tres [Bordini 2006] TuCSon [TuCSon 2008], DESIRE [DESIRE 2008], y JADE [JADE 2008]. Esta última es un entorno de trabajo Java para el desarrollo de aplicaciones multi-agente distribuidas basado en la arquitectura de FIPA, que se ha utilizado en varias investigaciones vinculadas con el trabajo que aquí se presenta.

Los Agentes Inteligentes se han incorporado en el sistema SKC propuesto para dotarle del carácter autónomo y pro-activo que requiere en algunas de sus funcionalidades, de las que luego se hablará, en particular en el módulo que se ocupa de la interacción entre instancia del sistema y en los que se ocupan de algunos procesos internos de metabolización de conocimiento.

### Minería de Información y Datos

Durante la última década se han dado las condiciones adecuadas para el desarrollo de la Minería de Información y Datos, como resultado de la disponibilidad de varias técnicas maduras y sólidas, de grandes repositorios de datos, y de capacidad de cálculo suficiente. La Minería de Información y de Datos surgió de la aplicación de aquellas técnicas a esos repositorios utilizando tales capacidades de cálculo. Últimamente se está desarrollando la Minería Web (Web Mining) como resultado de la aplicación de las técnicas de la Minería de Información y Datos a los grandes repositorios de información y conocimiento disponibles y accesibles a través de la Web, y a los registros de datos generados por la actividad de los servidores Web. El empleo de la Minería de Información y Datos es, junto a las técnicas de Interacción Persona-Ordenador y al uso de Ontologías y Agentes, otro de los elementos claves incorporados en la propuesta de arquitectura que se presenta más adelante.

La Minería de Texto es un campo de la Minería de Información que trabaja sobre información textual poco estructurada. Permite el descubrimiento de asociaciones, tendencias y la detección de sucesos. Esta última facilita la localización de historias en un flujo continuo de noticias en medios de información, con métodos como el de Agrupación Incremental basado en un modelo

vectorial [Yang 1998] [Salton 1989] [Chang 2001]. Los modelos vectoriales también se utilizan para recuperación de información [Baeza 1999], que es un caso de Descubrimiento de Asociaciones, para buscar respuestas a preguntas en las que coincidencias parciales son también deseables.

El empleo del Modelo Vectorial en Minería de Texto requiere un tratamiento del material textual parecido al siguiente. Primero se eliminan las palabras de parada, que son términos de uso muy frecuente en un idioma que resultan poco significativos. Segundo se eliminan las variantes morfológicas de las palabras, sustituyendo todas ellas por una forma única. Tercero se representan los textos como vectores en el espacio de términos únicos, estableciendo un vector por texto asignando un peso a cada término del espacio en función de su importancia en el texto, por ejemplo considerando en una expresión matemática la frecuencia de la palabra en el texto frente a su frecuencia en una colección de textos representativa del uso general del idioma. Por último, cuarto, se pueden comparar los vectores unos con otros para determinar su similitud empleando alguna medida de distancia, como la distancia euclídea o el coseno del ángulo que forman.

En particular la Minería de Texto se presenta como base para el mapeo automático entre ontologías, que es esencial para la integración entre instancias del sistema SKC propuesto, y que también sirve para el soporte de múltiples modos simultáneos para organizar, acceder y presentar el conocimiento. El Modelo Vectorial se utiliza en todos los casos como medio para establecer relaciones entre los elementos implicados mediante los textos asociados a ello, como se verá después en este documento.

El Análisis de la Semántica Latente [Landauer 1998] es una teoría y un método para extraer y representar el significado de las palabras debido al uso contextual de las mismas, mediante el procesamiento estadístico de grandes corpus de textos. El procedimiento [Gutiérrez 2005] se inicia creando una matriz de frecuencias de términos en textos, en la que cada columna corresponde a uno de los textos considerados y cada fila a un término que aparece en alguno de ellos. A continuación, la matriz se transforma en otra de pesos de palabras, sustituyendo las frecuencias por valores representativos de la importancia de los términos en cada texto y en el conjunto de textos considerado. Por último, se aplica una descomposición en valores singulares a la matriz, de modo que la matriz inicial se descompone en el producto de tres matrices, una de términos en espacios de términos, otra de los textos en espacios de términos y otra diagonal de valores tales que multiplicando matricialmente las tres se obtiene una matriz semejante a la original, con más o menos precisión dependiendo del número de factores considerados en la matriz diagonal. Con este procedimiento se consigue crear un espacio vectorial semántico, en el que términos y textos estén representados por vectores con información esencial para establecer conceptos que los relacionan, que son los espacios de términos. Los vectores obtenidos permiten

realizar comparaciones entre los elementos que representan mediante alguna medida de similitud, como cósenos o distancia euclídea [Jorge-Botana 2006].

Aunque el Análisis de la Semántica Latente no se ha empleado en los prototipos desarrollados para este trabajo, la vinculación con la aproximación propuesta es evidente, por lo que se ha considerado como una referencia fundamental para ella y para futuros trabajos que la desarrollen en el ámbito del proyecto que nos ocupa.

La Minería Web es el proceso de aplicar técnicas de Minería de Información y Datos al descubrimiento de patrones en la Web, en sus contenidos, estructuras o registro de uso [Garofalakis 1999] [Srivastava 2000] [Chang 2001]. La Minería de Contenidos Web permite el descubrimiento automático de patrones en los documentos Web y la extracción de información útil de sus contenidos, fundamentalmente para la recuperación de información, empleando técnicas de minería de texto de las que se habló anteriormente.

La Minería de Estructura Web permite el descubrimiento de esquemas en la estructura de enlaces de la Web, como páginas de interconexión *-hubs-*, páginas muy apuntadas por otras *-authorities-*, o grupos de páginas interrelacionadas de forma cerrada. Este tipo de minería también se ocupa de analizar la estructura interna de los sitios Web para observar el modo en que se relacionan las páginas en su interior, por ejemplo para identificar las páginas de navegación, cuya función principal es facilitar el acceso a otras, o reconocer las páginas de contenido, que son las que incorporan la información en los sitios Web. Vinculado con esta clase de minería está el algoritmo HITS [Kleinberg 1998, 1999], que permite ordenar un conjunto de páginas Web por su relevancia, considerando sus condiciones como *hubs* y *authorities*. Las Minerías de Contenidos y Estructura Web proporcionan la base de los populares sistemas de recuperación de información que indexan la Web, como Google y Yahoo.

La Minería de Uso de la Web permite descubrir patrones de uso [Chang 2001]: análisis estático, descubrimiento de reglas de asociación, realización de agrupaciones, realización de clasificaciones, y descubrimiento patrones secuenciales. Esta minería emplea dos fuentes de datos principales [Nasraoui 2004] que proporcionan datos cronológicos de uso muy numerosos. Una fuente son los registros de actividad de bajo nivel de los servidores Web, logs de los servidores, que registran las peticiones de recursos solicitados a los sistemas. La otra fuente son los registros de actividad de alto nivel de las aplicaciones Web, que recogen las acciones relevantes para ellas. Los datos de alto nivel obtenidos por las aplicaciones Web pueden dar sentido a los datos de bajo nivel registrados por los servidores.

El proceso de análisis de los registros de uso Web es muy laborioso, suele (1) comenzar con un tratamiento de los ficheros originales para eliminar entradas poco significativas -a veces en varios pasos-, (2) seguir con la identificación de las entradas asociadas a cada usuario, (3)

continuar con el reconocimiento de sesiones de usuario -secuencias de accesos separados entre sí menos de un tiempo máximo-, (4) proseguir determinando la trayectoria de los usuarios por las páginas Web, [Nasraoui 2004] y (5) terminar aplicando diversas técnicas de minería para descubrir patrones y reglas de comportamiento de los usuarios [Huysmans 2004]. Este análisis se complica por las limitaciones de los registros considerados, que a veces tiene datos incompletos, por las características del servicio y de la red, y que no siempre están en un modo adecuado para su tratamiento ideal, debido por ejemplo a su perspectiva de registro, ordenación o resolución.

La Minería Web se emplea de distintas formas en la aproximación planteada. La Minería de Uso Web proporciona el fundamento para el estudio de los registros de actividad de los usuarios y otros elementos activos del sistema SKC propuesto, en particular para el análisis de la interacción de los usuarios con el sistema y de sus manifestaciones implícitas de interés, donde también se utilizan ideas de Minería de Estructura Web, para establecer el significado de los registros atómicos de los ficheros Log utilizados. Además, en el módulo encargado de integrar las instancias de SKC formando una red de conocimiento, se utilizan buscadores convencionales de la Web, como Google o Yahoo, que utilizan técnicas de Minería de Contenidos y de Estructura Web. También, las técnicas Minería de Contenido Web, características del área de recuperación de información, sirven de base para los mecanismos de vinculación entre elementos de conocimiento propuestos en este trabajo.

#### *Sistemas Relacionados con la Propuesta*

Hay seis sistemas que tienen especial relevancia para el trabajo de investigación que aquí se presenta, de algunos de ellos se ha hablado ya en estas conclusiones, pero se considera conveniente presentarlos todos juntos en este subapartado para evidenciar el espacio que delimitan, donde se ubica la propuesta que se ha esbozado y se desarrollará después.

Los buscadores de la Web son sistemas informáticos que indexan automáticamente los recursos accesibles a través de la Web y responden a cuestiones planteadas con selecciones de los mismos. Estos buscadores suelen utilizar técnicas de recuperación de información [Baeza 1999] basadas en Minerías de Contenido Web, que emplea procedimientos de Minería de Texto con modelos vectoriales, y Minería de Estructura Web, que se sirve de algoritmo como HITS [Kleinberg 1998, 1999] para establecer el valor de los resultados considerando el modo en que se vinculan los recursos entre sí en la Web. Los buscadores se han convertido en instrumentos imprescindibles para encontrar información en la red. La demanda de sus servicios ha hecho de algunos de ellos sistemas fiables, autónomos y muy potentes, capaces descubrir cualquier recurso de Internet en poco tiempo, si es accesible a través de una URL al alcance de sus robots indexadores, que son los programas encargados del descubrimiento de recursos en la red. Los

buscadores Web convencionales, como Google [Google 2008] o Yahoo [Yahoo e008], disponen de una interfaz de formularios HTML para enunciar una pregunta, con la que se invoca un CGI al que se pasan los datos de la búsqueda. Como resultado los buscadores devuelven una página HTML con una lista de enlaces a páginas Web que responden a la cuestión realizada. A pesar de que estos resultados son casi siempre fáciles de tratar de forma automática, los buscadores convencionales actuales están dirigidos a usuarios humanos y los usuarios de otro tipo no suelen ser bien recibidos, aunque algunos ofrecen ciertos servicios para ellos. Cuando la Web Semántica esté implantada plenamente habrá servicios equivalentes a los buscadores actuales destinados a usuarios artificiales como los agentes software.

Los buscadores de la Web son aplicaciones que integran dos aspectos esenciales de la aproximación de gestión de conocimiento planteada en este trabajo. Por un lado, las técnicas de Minería de Información y de Datos, en particular la Minería de Texto a través de la Minería de Contenido Web, que permiten establecer relaciones entre los elementos de información que componen la Web y de estos con los intereses de los usuarios, aprovechando que todos ellos utilizan un formato de texto común. Por otro lado, las técnicas de Minería de Estructura, que sacan partido del trabajo realizado por la comunidad global de creadores de páginas Web, al enlazar páginas y recursos formando la red, para desentrañar el valor implícito de dichos elementos y utilizarlo para seleccionar y ordenar los conjuntos de elementos de información obtenidos previamente con las técnicas de Minería de Texto. Es interesante observar que el trabajo de enlace de los creadores de los sitios Web, que aprovechan los buscadores para calificar los recursos de la red, no consistía en emitir votos con ese fin y que las valoraciones obtenidas son el resultado de la actividad colectiva provocada por el interés particular de cada uno.

Los sistemas de recomendación permiten filtrar la información, para seleccionar la parte de ella que puede ser de más interés para cada usuario, eliminando de forma automática los contenidos inapropiados o redundantes antes de su presentación. Con ello se pretende reducir el impacto de la sobrecarga de información. Los sistemas de filtrado suelen comparar el perfil de un usuario con las características de otros usuarios o de los ítems de información para alcanzar su propósito. Entre las aplicaciones que incorporan mecanismos de recomendación pueden citarse Amazon [Amazon 2008] que hace sugerencias comerciales de libros y otros artículos y MovieLens [MovieLens 2008] que propone películas.

Los sistemas de recomendación aportan otro ingrediente interesante para la aproximación de esta tesis, se trata de aprovechar lo que se sabe de los contenidos que se gestionan y de los usuarios y la comunidad implicados en su tratamiento, para hacer predicciones que permitan mejorar el servicio, sin pedir, en el mejor de los casos, ninguna colaboración a los usuarios que obtienen un valor añadido de su actividad con el sistema. De nuevo la cuestión es sacar



provecho de información latente y del cruce de datos disponibles o que pueden obtenerse con una inversión económica y rentable.

Un weblog [Weblog 2008] es un sitio Web de actualización frecuente en el que las entradas se muestran de forma cronológica inversa y son los autores los que deciden qué se publica. A los weblogs se los llama a veces bitácoras, porque son diarios digitales de carácter público. Aunque los weblogs pueden ser simples diarios íntimos, es muy frecuente que se dediquen a cualquier tema que el propietario considera de interés para compartir. Además, los weblogs suelen facilitar la participación de los visitantes y estar vinculados con otros weblogs y sitios Web, utilizando enlaces o mecanismos de sindicación de contenidos.

Los weblogs son un ejemplo de cómo un sistema puede auto-incentivar a sus usuarios para que la labor voluntaria de un colectivo se canalice de forma productiva, y cómo este impulso se puede aprovechar para la difusión de contenidos, y la obtención de relevancia y oportunidades de negocio. Otra vez los vínculos entre contenidos de la Web, establecidos por sus creadores sin emitir votos explícitos, permiten una selección colectiva de los recursos movida sólo por intereses particulares.

Folksonomía [Golder 2006] es el procedimiento colaborativo de creación y gestión de etiquetas para la anotación y clasificación de contenidos, sin la participación de expertos. Las folksonomías son un tipo de aplicaciones sociales basadas en este procedimiento, donde los usuarios eligen las etiquetas libremente del lenguaje natural, para clasificar los contenidos que comparten en una estructura plana, sin niveles.

Las folksonomías aprovechan, como las otras aplicaciones características de las Web 2.0 - weblogs y wikis-, la potencia de las redes sociales. En este caso, la contribución colectiva a la organización del material gestionado es evidente, pero además es explícita, sin que por ello se merme, sino todo lo contrario, la fluidez de la colaboración y los resultados obtenidos. De nuevo el interés particular parece el motor de la disposición colaborativa, pero además parece esencial la naturalidad del acto de clasificación sin encorsetamientos ni supervisiones.

Wiki [Leuf 2001] es un tipo de aplicación Web que permite a los usuarios crear, editar, enlazar y publicar páginas Web de forma fácil, colaborativa y directa, empleado un navegador convencional. Las wikis son sitios Web contruidos con este tipo de aplicaciones y que permanecen abiertos a las posibles modificaciones de los usuarios. Las wikis tienen tres características fundaméntales [Leuf 2001]: (1) invitan a los usuarios a crear y editar sus páginas sólo con un navegador Web estándar sin añadidos; (2) promueven los enlaces entre páginas wiki, facilitando su creación y mantenimiento; (3) son sitios Web en permanente construcción, en los que se pretende involucrar a los usuarios de forma indefinida. Las wikis también pueden verse como bases de datos gestionadas colaborativamente.

Las wikis reafirman algunas de las observaciones realizadas en los sistemas anteriores, pero se hacen evidentes dos en particular. Por un lado, la facilidad de uso es esencial para estimular la participación voluntaria, sobre todo si es explícita y no un resultado residual de otra actividad principal. Por otro lado, el anonimato, la adscripción voluntaria a grupos y la supervisión relajada pueden provocar una desinhibición, que sirva para superar prejuicios, estimular la creatividad y mejoren los resultados grupales.

KnowCat es un sistema de trabajo en grupo que facilita la gestión de un repositorio de conocimiento mediante la interacción de una comunidad de usuarios a través de la Web [Alamán 1999][Cobos 2003]. Esto se puede hacer sin supervisión empleando información sobre la actividad de los usuarios y sus opiniones sobre los documentos que forman parte de la base de conocimiento. En este repositorio, los documentos se organizan en temas que forman un árbol de conocimiento. Cada instancia del sistema es un nodo KnowCat que se dedica a una temática concreta y tiene una comunidad de usuarios y un árbol de conocimiento propios.

El sistema SKC propuesto en este trabajo, sigue la filosofía de KnowCat en el modo de procurar la gestión del conocimiento, aunque intenta conseguirlo de modo distinto [Moreno 2005]: (1) reduciendo la necesidad de que los usuarios manifiesten explícitamente su opinión sobre el conocimiento; (2) aumentando el aprovechamiento de la información generada por el sistema, los usuarios y otras entidades con las que aquél se relacione; y (3) ampliando la explotación de las características de los elementos implicados en la actividad, como la comunidad, el conocimiento y la Web. Además, aunque SKC y KC siguen el mismo enfoque para la gestión del conocimiento, sus diseños son muy diferentes [Moreno 2005], puesto que la arquitectura de SKC se construye alrededor de ontologías, que se refieren tanto al conocimiento que el sistema gestiona como a las entidades que intervienen en su proceso. De este modo, los árboles de conocimiento de KC se tratan como ontologías en SKC. Estas diferencias no impiden que KC haya servido como plataforma de prototipado para SKC [Moreno 2007a, b, 2008].

# 3 Aproximación Conceptual

En el presente capítulo se hace un acercamiento conceptual a la hipótesis planteada al principio de este trabajo. Para ello se parte del estado de la cuestión desarrollado anteriormente, que se utiliza para justificar el problema considerado, para fundamentar la hipótesis propuesta, y para predecir las consecuencias de su confirmación. De este modo se presentan los antecedentes de la aproximación planteada y las innovaciones que ésta aporta al ámbito del problema y a los campos de conocimiento implicados, y se proyectan los resultados esperados buscando su generalización. A continuación, se propone un sistema ideal que reúne las características más importantes de los supuestos establecidos, empleando ideas y técnicas de los campos de conocimiento revisados. Finalmente, se describe un posible diseño de dicho sistema, que muestra una implementación del mismo con tecnologías disponibles y que servirá para el desarrollo de los prototipos necesarios para las experiencias que se presentan en el siguiente capítulo.

### **3.1 Propuesta**

Como se explicó al principio del capítulo anterior, la actual sobrecarga de información es un problema que involucra inmensos volúmenes de información de calidad diversa y muchas veces incierta, en el que están implicados normalmente grupos muy numerosos de usuarios, de los que casi siempre no se tiene referencias, y en el que las TIC proporcionan la infraestructura para la generación y tratamiento de la información y la interacción entre los distintos elementos. La Web es el ejemplo más general y representativo del entorno del problema.

Para investigar la sobrecarga de información sería deseable disponer de un sistema que reuniera algunas características significativas de los entornos en los que se manifiesta el problema, pero con una escala manejable y unos parámetros manipulables. En nuestra opinión, los sistemas de gestión de conocimiento en red pueden servir como modelos escalables para investigar el asunto y sus posibles soluciones, puesto que reúnen muchos de los elementos y características de entornos grandes y complejos, como la Web, pero en un ámbito mucho más reducido y controlable. De hecho, la Web podría verse desde algunas perspectivas exigentes como un gran sistema de gestión de conocimiento, puesto que puede proporcionar medios para recoger, organizar y distribuir el conocimiento [McDermott 1999], y para cubrir las necesidades de colaboración de la comunidad que trata con él [Churman 1971].

Nuestra hipótesis es que hay varios aspectos latentes en los sistemas afectados por la sobrecarga de información que pueden contribuir positivamente a la solución de tal problema. Por un lado, el aprovechamiento de la energía residual de la actividad de los elementos activos que intervienen en dichos sistemas, como son las personas, los servicios, las aplicaciones y otras entidades artificiales que interaccionen con los mismos. Por otro lado, la explotación de las características de los elementos y las actividades vinculados con los sistemas afectados, como pueden ser la red, las entidades activas ya mencionadas, la información y el conocimiento implicados, o los procesos y las interacciones de unos y de otros.

La energía residual de los elementos activos de los sistemas a la que nos referimos es, de forma figurada, el esfuerzo sobrante de la actividad de dichos elementos o, de otro modo, la potencia que puede aprovecharse para tareas adicionales, sin perjuicio de su aplicación original. Por ejemplo, si los usuarios de la Web establecen enlaces desde sus páginas a otras, podría considerarse como si emitieran votos de interés hacia las páginas enlazadas. Algo similar ocurre cuando los usuarios de un sistema de publicación de fotos clasifican sus imágenes con palabras claves elegidas libremente por ellos mismos, tal acción podría considerarse como manifestación de apoyo a la inclusión de determinados conceptos en una ontología del dominio de las imágenes. En ambos ejemplos los usuarios dedican energía a realizar las acciones que les

interesan, pero es posible aprovechar dicha energía para llevar a cabo otras tareas sin esfuerzo ni atenciones adicionales por su parte.

Las características a las que nos referimos de los elementos y las actividades vinculados a los sistemas, son las peculiaridades de aquellos a las que se puede sacar partido para el objetivo que se pretende. En el ejemplo de los enlaces del párrafo anterior, se están utilizando el interés propios de los usuarios al seleccionar enlaces y las propiedades de los hipervínculos de las páginas Web, entre otros; en el ejemplo de las fotografías, se están aprovechando características tales como la motivación personal de los usuarios para publicar sus obras o como la cultura compartida por el grupo. De nuevo en los dos ejemplos no se pide nada nuevo a los elementos implicados, que mantienen sus propiedades inalteradas todo el tiempo, aunque se saca provecho de sus características para conseguir algunos objetivos adicionales.

Consecuentemente, aplicando la hipótesis a los modelos de estudio propuestos, creemos que los sistemas de gestión del conocimiento en red se pueden beneficiar de aprovechar la energía residual de la actividad de las personas, los servicios y otras entidades artificiales que interactúan con ellos, así como de explotar las características de la red, de las entidades involucradas en sus actividades, de las interacciones entre las mismas y del conocimiento que gestionan dichos sistemas. Con ello pretendemos estudiar posibles vías paliativas para la sobrecarga de información en los sistemas de gestión de conocimiento empleados, pero con la ambición de que los resultados obtenidos sean de aplicación en ámbitos más amplios, como la propia Web.

El sistema de gestión de conocimiento Semantic KnowCat (SKC) se propone para ilustrar la aproximación planteada y comprobar los aspectos fundamentales de la misma. De dicho sistema se habló someramente en la introducción de este trabajo y se hablará en profundidad en el apartado siguiente.

El enfoque de la propuesta tiene sus antecedentes en varias referencias que se han revisado en el capítulo anterior. En primer lugar, el sistema de gestión de conocimiento KnowCat [Alamán 1999][Cobos 2003], que emplea la colaboración de una comunidad de usuarios para gestionar conocimiento sin supervisión. También, los procedimientos empleados en recuperación de información por algunos de los buscadores actuales de la Web, como Google [Google 2008], para mejorar sus resultados [Baeza 1999], son un antecedente de la propuesta realizada. En concreto la consideración de los enlaces recibidos por un sitio Web para establecer su posible interés y utilidad [Kleinberg 1998, 1999]. Además también son antecedentes de la propuesta las técnicas empleadas en algunos sistemas de recomendación [Adomavicius 2005], que tienen en cuenta la historia de la actividad del grupo y los individuos para mejorar el servicio que ofrecen

a los usuarios. Aplicaciones como Amazon [Amazon 2008] o MovieLens [MovieLens 2008] pueden ilustrar el empleo de este tipo de técnicas.

Al mismo tiempo, las tendencias de la Web en los últimos años -que la Web 2.0 representa [O'Reilly 2005]- han venido a apoyar de forma más general el planteamiento propuesto. En particular en lo que se refiere al aprovechamiento de la estructura de la Web, de las redes sociales, de la inteligencia colectiva y de la colaboración; aplicaciones como los weblog [Weblog 2008], las folksonomías [Golder 2006] y las wikis [Leuf 2001], pueden servir de ejemplo de estas tendencias.

Las ideas subyacentes a las referencias mencionadas y algunas de las técnicas empleadas en ellas se han aplicado en el trabajo que aquí se presenta, pero además se han considerado otras aportaciones de las áreas de conocimiento donde dichas referencias se ubican y de los campos de investigación relacionados con ellas y con el ámbito del trabajo: Gestión del Conocimiento, Interacción Persona-Ordenador, Web Semántica, y Minería de Información y de Datos.

Del primero de los campos de investigación considerados, la Gestión del Conocimiento, se extraen las ideas que hacen del sistema SKC propuesto un entorno de gestión del conocimiento preocupado tanto de estructurar adecuadamente el conocimiento [Malthotra 2000], como de facilitar el trabajo colaborativo de los usuarios que lo manejan [Churman 1971]. La arquitectura planteada pretende abundar en las prestaciones del sistema para el soporte del trabajo colaborativo y proporcionar una base más robusta para la administración de la información en el sistema [Cobos 2002b]. Éste también incorpora las consideraciones de Ackerman [Ackerman 2003] sobre la importancia de incluir los componentes humanos como parte del conocimiento que el sistema gestiona.

Del segundo de los campos de investigación mencionados, la Interacción Persona-Ordenador, se toman referencias de las áreas de Groupware, Awareness, Monitorización de Actividad, Hipermedia Adaptativa, Recomendación, Visualización de la Información, y Procesamiento del Lenguaje Natural, así como de la ya mencionada Web 2.0.

El sistema SKC propuesto además de un sistema de gestión de conocimiento, es un sistema de trabajo en grupo -Groupware- para poder desempeñar mejor su función [Churman 1971]. Además, el sistema incorpora diversas características de Awareness [Dourish 1992], para soportar la actividad en grupo y su seguimiento [Sohlenkamp 1994], que procuran respetar la intimidad de los participantes [Hudson 1996], y adaptarse a las situaciones de uso [Byun 2001]. Como parte del seguimiento de la actividad de los usuarios, SKC incluye un mecanismo de Monitorización de Actividad, que utiliza eventos de la interfaz de usuario [Hilbert 2000], a través de las páginas Web [DOM 2008], como medio de obtener datos sobre el interés aparente

de los usuarios hacia los elementos de conocimiento, que sirven para enriquecer el conocimiento del sistema.

De la Hipermedia Adaptativa [Brusilovsky 1998], se adoptan los fundamentos para el modelado de los elementos del sistema SKC, la anotación automática, el registro de actividad del sistema y el empleo de elementos virtuales. Tales modelos son fundamentales para el seguimiento y adaptación del entorno, y en general constituyen una parte fundamental del corazón del sistema. La anotación automática, junto con el registro, amplifican la capacidad del usuario para gestionar el conocimiento y el propio sistema. El empleo de documentos virtuales [Milosavljevic 1999] es uno de los recursos empleados por la adaptación.

La Recomendación [Adomavicius 2005] se incorpora en SKC como medio de aprovechar la actividad residual de los usuarios, para facilitar sus tareas, para evidenciar las relaciones implícitas que se establecen entre ellos y los elementos de conocimiento, y para mejorar con todo ello la gestión de conocimiento y la sobrecarga de información. La Visualización de la Información [Geroimenko 2002] sirve para dotar a SKC de la interfaz de usuario necesaria para tratar con su modelo de representación, visualizar los contenidos enriquecidos y las ontologías que los representan. El Procesamiento del Lenguaje Natural se utiliza en el tratamiento de la información textual [Baeza 2004] que integra el conocimiento o que se asocia a los elementos que lo forman y que SKC maneja. En concreto dicho procesamiento se emplea en la anotación automática de conocimiento, en su clasificación, y en el mapeo entre las ontologías que lo organizan.

Del tercero de los campos de investigación enumerados, la Web Semántica [Berners-Lee 2001], se toman las ideas de usar ontologías [Mahesh 1996], agentes [Dinverno 2001] y la generación dinámica de documentos [Milosavljevic 1999]. Con las ontologías se dota de un mecanismo de representación adecuado a los distintos dominios que se consideran en SKC: los documentos - unidades de conocimiento fundamentales- y las estructuras de conocimiento que los contienen; los usuarios y sus jerarquías; los nodos -instancias del sistema- y la red que los agrupa; etc. Los agentes proporcionan la base para la integración de las instancias del sistema formando redes. La generación dinámica de documentos permite la adaptación a los requerimientos de uso del sistema y del conocimiento que gestiona.

Finalmente, del campo de Minería de Información y de Datos [Chang 2001] se obtienen algunas aportaciones de Minería de Textos y Minería Web. Las ideas de Minería de Texto, que están vinculadas al área de recuperación de la información y de agrupación y categorización de textos [Baeza 1999], se aplican al análisis de los documentos textuales clasificados en las estructuras de conocimiento del sistema SKC con diversos fines, entre los que están: determinar vinculaciones en el interior de dichas estructuras, clasificar los documentos automáticamente

dentro de ellas, buscar relaciones entre los usuarios y de éstos con otros elementos del sistema, y establecer vínculos entre las instancias del sistema e integrarlas en una red de conocimiento. En estos procesos ha tenido especial relevancia el empleo de modelos vectoriales [Baeza 1999], y ha resultado fundamental la inspiración del Análisis de la Semántica Latente [Landauer 1998], a pesar de que no se ha aplicado directamente en los prototipos de SKC desarrollados. Las ideas de Minería Web [Chang 2001] se emplean en el análisis de los registros de actividad de SKC para descubrir patrones de uso con los que facilitar el desempeño del sistema y la interacción con él. En concreto su aplicación ha sido fundamental en el análisis del interés de los usuarios y la calificación automática de algunos elementos integrantes del conocimiento que el sistema gestiona.

Como se puede ver, el esfuerzo de integración de ideas y técnicas de diversas áreas de conocimiento es significativo. Aunque el problema de la sobrecarga de información se ha tratado desde distintos campos de investigación de forma independiente, este trabajo innova en la integración de técnicas e ideas de áreas diversas para buscar soluciones al mismo. Esta circunstancia no es caprichosa en el contexto de la propuesta, por el carácter interdisciplinario del ámbito del problema y la variedad de aspectos que concurren en él. La Web Semántica marca un espacio de convergencia de muchas de las disciplinas consideradas en este proyecto, para hacer que los datos accesibles en la Web sean más adecuados para su procesamiento automático. La Web 2.0 también establece un punto de encuentro entre algunas de dichas disciplinas, con el fin de mejorar el uso de la Web para los individuos y los grupos. La tendencia que estos dos ejemplos evidencian, la búsqueda de la integración de campos diversos para responder a los problemas que la Web plantea, apoya el enfoque de la aproximación propuesta en este trabajo.

Por otro lado, como ya se ha dicho antes, la Web podría verse como un gran sistema de gestión de conocimiento desde algunas perspectivas. Esta visión, además de reconocer el carácter de repositorio de datos de la red, reivindica el papel esencial de la comunidad de usuarios en ella, que también forman una red -pero activa- y dota de un objetivo fundamental al conjunto: colaborar para explotar la información y el conocimiento. Este planteamiento proporciona nuevas oportunidades de aproximación a algunos problemas de la Web y de obtener provecho de ella, algo que resulta también innovador. En particular, como se ha sugerido, el problema de sobrecarga de información en la Web podría mejorar aprovechando la energía residual de la actividad de los usuarios, de las aplicaciones y de las entidades artificiales que concurren en la red, y con la explotación de las características de la propia red, y de las entidades involucradas, como son los usuarios, la información y el conocimiento.

Además, la aproximación planteada aporta un marco experimental para los campos de investigación implicados, especialmente para las áreas de gestión de conocimiento, de trabajo



colaborativo, y de la Web. El sistema SKC propuesto proporciona un entorno con unas características de especial interés para la investigación en dichos campos: un medio acotado de gestión de conocimiento en grupo sobre la Web, en el que intervienen interlocutores humanos y automáticos con objetivos claros y concurrentes, y que tiene una utilidad práctica evidente.

Por último, la Web futura incorporará las aportaciones de la Web 2.0, de la Web Semántica y de otras enseñanzas aprendidas por el camino. La aproximación presentada en este trabajo contribuye en este proceso, mostrando el punto de convergencia de tecnologías diversas para la gestión del conocimiento y, a través de ellas, para el tratamiento de la sobrecarga de información en la Web y en ámbitos que compartan características similares con ella.

A continuación se presenta la descripción del sistema Semantic KnowCat (SKC), al que nos hemos referido ya varias veces a lo largo de este documento, que está destinado a ilustrar la aproximación planteada y a servir de base para probar algunos de sus supuestos principales. La propuesta recogida en este apartado consta de una descripción general del sistema y de una proposición de arquitectura para él. Además, como anexo de este trabajo se incluye una especificación más formal del sistema mediante la descripción de sus casos de uso fundamentales.

## **3.2 El sistema Semantic KnowCat (SKC)**

Semantic KnowCat (SKC) es un sistema que hereda las capacidades de KnowCat (KC) [Alamán 1999][Cobos 2003], e incorpora algunas mejoras y nuevas funcionalidades, con el objetivo final de probar los postulados presentados en este trabajo. Se ha preferido incluir de nuevo en esta sección los requerimientos coincidentes con las especificaciones de KC presentadas en el apartado correspondiente del capítulo anterior con el fin de proporcionar una visión completa del nuevo entorno.

### **3.2.1 Aspectos Generales**

Semantic KnowCat (SKC), como KnowCat (KC), es un entorno basado en Web que permite la gestión del conocimiento gracias a la interacción de una comunidad de usuarios sin necesidad de supervisión humana. También como en KC, en SKC el objetivo de esta gestión es provocar la selección del conocimiento disponible en el dominio para obtener su mejor representación. El proceso por el que se realiza esta selección a partir del conocimiento proporcionado explícitamente, y de los registros de interacción con el sistema y de actividad del mismo, se denomina cristalización. Una completa descripción de KC puede encontrarse en el capítulo de Estado de la Cuestión de este documento y en [Cobos 2002a].

Además, SKC es un sistema que aplica algunas ideas de la Interacción Persona-Ordenador y de la Web Semántica, junto con otras de la Minería de Información y Datos. Así, aunque SKC mantiene el espíritu y las funcionalidades de KC, la aplicación de las nuevas ideas conlleva cambios profundos en la arquitectura del sistema y, por tanto, en la implementación de las funcionalidades heredadas.

Asimismo, SKC incorpora una serie de nuevas funcionalidades que no estaban disponibles en KC. Estas incluyen infraestructura para redes de nodos, múltiples modos de organizar, acceder y presentar el conocimiento, y soporte para realizar el seguimiento particularizado de la actividad de los usuarios y de los demás elementos del sistema.

En relación con la primera de las funcionalidades añadidas, el nuevo sistema permite el establecimiento de redes de nodos preparados para interactuar automáticamente a través de la Web y para tomar en consideración sus repositorios de conocimiento y sus comunidades de usuarios. SKC incorpora algoritmos de cristalización preparados para aprovechar esta nueva posibilidad de interacción entre aplicaciones. Tal consideración entre nodos tiene sentido cuando todos los nodos implicados se dedican a similares áreas de conocimiento. En tal caso se pueden establecer distintas relaciones entre los grupos y repositorios consagrados a temas parecidos, que pueden ser de referencia, de colaboración, de subordinación, de competencia, etc. Dependiendo de la naturaleza de la relación, el efecto sobre cada nodo de la actividad de los otros es distinto. De la interacción entre nodos se encargan agentes, que son responsables de la gestión de la relación y el intercambio de información.

En lo que se refiere a la segunda de las funcionalidades incorporadas, el nuevo sistema amplía las capacidades del anterior en lo que al modo de acceso al conocimiento se refiere. Por un lado, SKC permite la convivencia de varias organizaciones del conocimiento sobre los repositorios de los nodos y durante todas las fases de evolución de éstos, con lo que es posible navegar por la información y acceder al conocimiento de múltiples formas. En este sentido, pueden coexistir varios árboles de conocimiento sobre la misma base del mismo y diversas organizaciones de usuarios del sistema a la vez. Por otro lado, el sistema permite que la presentación de la información se adapte al contexto y al usuario, lo que hace que la visualización de los elementos del sistema sea diversa. De esta forma, el sistema acomoda la presentación del conocimiento a las circunstancias de su solicitud, del usuario que lo requiera, de su posición en el grupo, del material que se trate, de la ubicación de este en el árbol de conocimiento, del uso que la comunidad haya hecho de ese material antes, etc. Aunque siempre se respeta la unidad del documento, cuando es posible éste se proporciona acompañado de la información disponible en el sistema que es relevante en el contexto de la interacción.

Respecto a la tercera de las nuevas funcionalidades, SKC lleva un control constante y personalizado de la interacción de los usuarios con el sistema. Los usuarios están en todos los momentos identificados y su actividad registrada y analizada. De esta forma se obtiene toda la información posible de la actividad de los usuarios para enriquecer el conocimiento aportado por ellos explícitamente a SKC. El seguimiento no se limita a los usuarios, sino que se generaliza a todas las entidades que interactúan con el sistema, puesto que en SKC además de los usuarios otros programas pueden también interaccionar con él -como instancias del mismo sistema o, en el futuro, otras aplicaciones de terceros-. Algunas de estas entidades, como ocurre con otras instancias de SKC, pueden tener tras de sí comunidades de usuarios que gocen de consideración a la hora de valorar su actividad con respecto a los nodos relacionados a través de los agentes encargados del intercambio de información. Algo parecido pasa con los repositorios de conocimiento que pudieran estar asociados a esas entidades, cuyo contenido podría también considerarse a través de los agentes.

### **3.2.2 Conocimiento**

SKC permite trabajar con conocimiento explícito, del que puede manejarse en forma de documentos de algún tipo, como textos, imágenes, secuencias de vídeo o esquemas organizativos, por ejemplo. Además, el sistema asume que, aunque el conocimiento tratado pueda estar en evolución, debe ser propenso a estabilizarse en algún estado y perdurar en el tiempo, como ocurre con el conocimiento enciclopédico, el documental, el derivado de la investigación o el debido a la organización del propio conocimiento. La aplicación de SKC es dudosa para gestionar conocimiento voluble, que nunca alcanza un estado estable, e inadecuada para el manejo de conocimiento que evoluciona tan rápidamente que la intensidad de la interacción de los usuarios no es suficiente para aportar la información necesaria al sistema para posibilitar su gestión. Determinar la adecuación de SKC para la gestión de una determinada clase de conocimiento no depende sólo de ésta, sino también de la forma en que el resto de los elementos activos del sistema interactúan con él. En lo que se refiere a la comunidad de usuarios dependerá de sus grados de motivación, dedicación y dinamismo, y en general, de la capacidad del propio sistema para manejar el conocimiento, interactuar con los integrantes de la comunidad y resto de agentes activos, y registrar y analizar las actividades de todos.

En SKC el conocimiento se organiza mediante grafos que representan las áreas en que se puede dividir el dominio que se está tratando y las relaciones que se establecen entre ellas. Como un mismo dominio puede organizarse de diversas formas, el sistema permite que varios grafos -organizaciones- coexistan. Todos ellos comparten un vértice que representa el dominio que cada uno a su manera pretende describir. El resto de vértices de los distintos grafos no tienen relaciones explícitas a priori, aunque un vértice en un grafo pueda representar la misma área de

conocimiento que otro en un grafo diferente, o referirse a una parte del área delimitada por aquél, o abarcar las áreas acotadas entre varios vértices en el otro grafo, etc. Desentrañar estas relaciones es parte de las tareas de análisis que el sistema realiza de forma invisible.

Cada vértice del grafo puede tener dos tipos de elementos asociados: un conjunto de descripciones del área o tema representado y un grupo de vértices relacionados. El primero, es una serie de documentos o unidades atómicas de conocimiento, alternativas entre sí, que compiten por describir de la forma mejor posible el tema correspondiente, de ellos se habla más adelante en este capítulo. El segundo, es un grupo de vértices representativos de áreas o temas relacionados con el actual. Las relaciones que se pueden establecer entre vértices son variadas. Pueden ser de inclusión, pero también podrían ser de otras clases, equivalencia, pertenencia, etc. Inicialmente, el conjunto de vértices incluido en otro está constituido por elementos del mismo grafo que éste; después como resultado de la actividad del sistema se pueden ir incorporando otros vértices al conjunto provenientes de diferentes grafos, que incluso podrían pertenecer a otros nodos SKC integrados en una misma red con el inicialmente considerado.

A diferencia de lo que ocurría en KC, en el nuevo sistema no se considera el concepto de refinamiento que establecía una competencia entre subdivisiones alternativas de los temas. Ahora las relaciones entre vértices no se consideran excluyentes entre sí y pueden coexistir mientras gocen de un nivel mínimo de aceptación, de forma parecida a como sucede con las anotaciones según se hablará después. Los usuarios pueden proponer nuevas relaciones entre vértices, su aceptación permitirá su persistencia o extinción.

Los grafos de conocimiento suelen ser árboles jerárquicos. Cuando es así, la raíz representa el dominio de conocimiento sobre el que se está trabajando, la primera ramificación refleja la división inicial del dominio en áreas de conocimiento, la siguiente la subdivisión en temas de las áreas, y así sucesivamente. Este proceso se puede repetir tantas veces como consideren oportuno los usuarios implicados en la definición; el número de niveles establecidos determinará el grado de detalle alcanzado. Normalmente, la resolución más adecuada vendrá dada por el dominio, la comunidad y los objetivos perseguidos. Esto ocurre de forma similar con otras clases de grafos, donde el número de vértices establecidos puede también ser indicativo del detalle de la representación.

Los usuarios del sistema pueden proponer nuevos grafos para organizar el conocimiento. Para ello es necesario que describan la estructura del grafo, vértices y aristas, y que determinen qué prototipos de documentos se clasificarán en cada nodo. El procedimiento para establecer tales prototipos depende del estado de desarrollo del nodo: cuando la masa de conocimiento en SKC es suficiente, el promotor puede utilizar algunos documentos del sistema para indicar el tipo de documentos que se clasificarían en los nodos del nuevo grafo. Si esto no fuera posible, puede

utilizar documentos que no estén en el sistema como prototipos, crear documentos de ejemplo o confiar en que los usuarios empiecen a clasificar material en el grafo sin la colaboración del sistema. El empleo de esta última alternativa no es recomendable salvo con el primer grafo definido en un nodo de SKC.

Como la masa crítica de usuarios y documentos necesaria para el adecuado funcionamiento del sistema será mayor a medida que se incluyan nuevos grafos en el nodo, el número de grafos que coexistan en el sistema estará regulado en determinados momentos de la evolución de los nodos.

Cuando los usuarios ingresan un documento en el sistema lo hacen a través de uno de los grafos disponibles. Como clasificar el documento en todos los grafos no es razonable y podría llegar a ser inabarcable, el sistema proporciona un mecanismo automático para liberar a los usuarios de esta tarea. Tal mecanismo utiliza los prototipos de documentos establecidos para los vértices de los grafos y los documentos previamente clasificados en ellos. En todo caso, el sistema prevé la posibilidad de que el autor de un documento pueda modificar manualmente la clasificación realizada por el sistema de forma automática. Además las clasificaciones automáticas entran en un régimen de selección similar a las realizadas por los usuarios, lo que garantiza la permanencia de las que gocen de aceptación y la extinción de las que no. Una vez que las clasificaciones automáticas han alcanzado un cierto grado de aceptación se las empieza a considerar en pie de igualdad con las propuestas realizadas por un usuario.

A lo largo del desarrollo de un nodo, puede suceder que alguien desee modificar alguno de los grafos, añadir un tema, eliminarlo o cambiarlo de posición en el grafo. De forma parecida a lo que se hace con los documentos, y que se explicará más adelante en estas especificaciones, el sistema facilita un procedimiento para poder proponer modificaciones y decidir si se van a llevar a cabo. Para ello, todos los usuarios de un nodo SKC pueden proponer cambios sobre los grafos, y algunos de ellos decidir, mediante votación, si las modificaciones se realizan. Los usuarios encargados de tomar tal decisión serán distintos dependiendo de la etapa en la que se encuentre el nodo. En las primeras fases serán los miembros del llamado Grupo de Coordinación los encargados de ello, en fases posteriores los llamados Expertos del grafo en cuestión. El sistema se encarga, en caso de necesidad, de consolidar de nuevo la base de conocimiento, reubicando documentos, restableciendo relaciones entre vértices, etc.

Todos los grafos propuestos están en competencia, de forma parecida a lo que ocurre con los documentos clasificados en sus vértices, pero -a diferencia de éstos- en esta ocasión no se trata de seleccionar la mejor organización del conocimiento en el nodo, sino de establecer la ordenación de las mismas según su aceptación y segregar las menos populares. La opinión de los usuarios sobre los distintos grafos, expresada de forma explícita mediante votaciones e

implícita por el uso de unos grafos en vez de otros en su interacción con el sistema, se emplea para la selección y cristalización de éstos en cada nodo de SKC.

El documento es la unidad atómica de conocimiento que el sistema reconoce. Se considera como tal cualquier paquete de información al que el autor confiere un sentido completo en el contexto del dominio en el que se está trabajando, y que, directa o indirectamente –mediante una representación-, pueda ser tratado con procedimientos informáticos. Se pueden considerar documentos un artículo de investigación, una receta de cocina o un corto cinematográfico; también pueden serlo algunas partes de otras unidades de información mayores, con tal de que su autor las considere provistas de sentido independiente, como el capítulo de un libro, la grabación en vídeo de una jugada de un partido de fútbol o un tema de música de un disco; o de forma inversa conjuntos de ellas como libros, series de TV o lotes de imágenes.

Además los documentos SKC deberán ser capaces de competir por la hegemonía en los vértices de los grafos en los que se les clasifique. En todo caso el sistema vela por mantener la unidad de los documentos procurando así preservar sus contextos y los sentidos que los autores intentaron conferirles. Sin embargo, no cabe duda que el propio sistema forma parte del contexto del documento desde el momento que éste se incluye en él y que los sentidos pretendidos originalmente por los autores pueden alterarse al enfrentar unos documentos con otros y someterlos a la interacción con los demás elementos del sistema –comentarios de otros usuarios, clasificaciones automáticas en vértices de otros grafos o publicación de qué determinados usuarios gustan de ellos-.

Precisamente para permitir añadir información a los documentos preservando su integridad, el sistema proporciona las anotaciones. Con ellas los usuarios pueden opinar o proponer modificaciones sobre los documentos sin alterarlos. Como en KC, las anotaciones se clasifican por tipos –comentario, aclaración, crítica...- y están disponibles para determinados grupos de usuarios dependiendo de los documentos. La asociación de los grupos de usuarios con los documentos que pueden anotar es configurable en SKC. Por ejemplo, los autores pueden utilizar las anotaciones para añadir información en sus documentos ya publicados, como aclaraciones o pequeños cambios.

De forma parecida, para el caso en que el propio autor de un documento quiere cambiarlo total o parcialmente de forma más profunda, añadiendo, modificando o quitando conocimiento, el sistema soporta las versiones. Con ellas los autores pueden proponer nuevos documentos candidatos a sustituir a otros anteriores. Normalmente se tratará de revisiones que intentarán mejorar los documentos previos considerando o no las anotaciones recibidas para estos, pero otras veces de completas transformaciones del documento anterior. En la mayoría de los casos una simple sustitución no es posible porque SKC debe preservar la integridad del conocimiento

que maneja, esto quiere decir que, por un lado, debe gestionar las consecuencias de la sustitución a la vista de los documentos implicados y la información en el sistema sobre los anteriores y, por otro, sólo permitir el cambio cuando los usuarios del entorno estén de acuerdo con él. Con las anotaciones y versiones el sistema permite la evolución de los documentos en su seno.

Desde la perspectiva de SKC el documento se considera de forma abstracta, puesto que es la reunión de sucesivas versiones, de las que sólo una es la representante del grupo en cada momento. En este sentido los documentos que ingresa el usuario en el sistema son versiones que integran un documento SKC del que sólo una es representante, y las anotaciones son consideraciones que los usuarios manifiestan hacia algunas de esas versiones. Sobre la conceptualización de documentos, versiones y anotaciones en SKC puede verse el apartado de Ontologías en este mismo capítulo.

De forma parecida a los grafos que estructuran el conocimiento en los nodos SKC, los documentos clasificados en cada uno de sus vértices están en competencia, aunque en este caso con el objetivo de determinar la mejor descripción del tema representado por los vértices correspondientes. La opinión de los usuarios sobre los distintos documentos, expresada de forma explícita mediante votaciones e implícita por la interacción con ellos, al utilizar el sistema, se emplea para la selección, cristalización, de los documentos más populares en todos los vértices de cada grafo.

SKC maneja Elementos de Contenido, como documentos, vértices, grafos o nodos, pero los presenta de manera combinada en forma de Contenido Virtual a sus Interlocutores, usuarios o agentes de red de los que luego se hablará. Para ello el sistema acomoda los contenidos, la navegación y la presentación, a las características y preferencias de los interlocutores y a las limitaciones de su medio de acceso al entorno, tomando en consideración las propiedades de los contenidos, las restricciones del sistema y el contexto de las interacciones en curso. Esto se consigue siguiendo un proceso de tres etapas para la confección dinámica de los Contenidos Virtuales. La primera sirve para determinar qué Elementos de Contenido se incluirán, la segunda para establecer qué partes de cada uno de ellos se mostrarán y, la última, para fijar con qué formato se presentarán. De este modo, cuando, por ejemplo, un usuario solicita un documento determinado, el sistema le devolverá el documento o alguna parte de él, dependiendo de las circunstancias, pero casi siempre junto con componentes de otros elementos de conocimiento que el entorno considera oportuno adjuntar, como pueden ser enlaces a otros documentos o la referencia de algún usuario que en ese preciso momento está viendo el mismo documento.

SKC se ocupa de la revisión constante del conocimiento y de la información generada por el sistema y por las entidades que interaccionan con él. Con ello se procura obtener nuevo conocimiento con el que enriquecer el disponible en cada momento y poner de manifiesto información relevante para el propio sistema y para las entidades que interactúan con él. En relación con el análisis de los Elementos de Contenido, el sistema se encarga de establecer automáticamente relaciones entre ellos y también con otros elementos del entorno, tales como los interlocutores. Esta labor se realiza en segundo plano y permite: determinar relaciones entre documentos y de éstos con vértices de grafos de conocimiento; establecer mapeos entre los grafos que pueden coexistir en un nodo o nodos relacionados en una red SKC; proponer relaciones entre nodos integrados en una red SKC; y determinar prototipos de documentos que los usuarios manejan para identificar vinculaciones con otros usuarios y distintos Elementos de Conocimiento. En general, como el documento es el elemento atómico de conocimiento en SKC, estableciendo vínculos automáticos el sistema consigue revelar muchas conexiones entre los elementos del entorno que están relacionados con ellos.

### **3.2.3 Comunidad de Usuarios y Fases del Sistema**

Los usuarios son el motor fundamental del mecanismo del sistema, sin ellos no habría conocimiento y sería imposible provocar su selección. En SKC la comunidad de usuarios se organiza de varias formas, como pasa con el conocimiento, según la actividad que se esté realizando. Inicialmente se definen dos grupos de actividades, aunque se podrían establecer más si fuera necesario, uno relacionado con las tareas administrativas y el otro relativo a la interacción con la base de conocimiento y sus organizaciones.

Respecto al primer grupo, en SKC existen labores administrativas a distintos niveles, se establecen Súper-administradores con capacidad de gestión absoluta en todo el sistema, en cada servidor de SKC se definen Administradores con capacidad de gestión en alguno de los servidores y los nodos definidos en él, y en cada nodo Responsables encargados de la gestión exclusivamente en ese ámbito. En KC se ha incorporado un módulo de administración que incorpora un planteamiento y funcionalidad similar al que este párrafo se refiere.

En relación con el segundo grupo, la organización depende de distintas variables. Según el estado de desarrollo de cada nodo, que se regula mediante fases, se establece una serie de grupos con cometidos especiales. Si en el sistema se definieran nuevos estados podrían aparecer nuevos grupos y cometidos.

En la Fase Inicial de los nodos, como no hay masa de usuarios ni conocimiento suficiente para el correcto funcionamiento de los mecanismos de selección automática del conocimiento, se establece un Grupo de Coordinación que se encarga de supervisar la actividad del nodo hasta



que éste pueda empezar a funcionar de forma autónoma. Durante esta fase el Grupo de Coordinación se encarga de tomar decisiones que en etapas posteriores se alcanzarán con la colaboración de toda la comunidad. El resto de usuarios del sistema constituyen, en esta fase, el Grupo de Colaboradores. Durante la etapa inicial los usuarios en esta categoría sólo pueden ingresar documentos en el sistema y revisar los documentos añadidos por otros usuarios. Su actividad no se considerará, como después se hará, para la cristalización del conocimiento.

El Grupo de Coordinación se constituye cuando se crea un nuevo nodo de SKC, aunque se pueden añadir nuevos miembros posteriormente. Una de las principales tareas del grupo es elaborar la primera estructura de conocimiento, haciendo propuestas, opinando sobre ellas a través del sistema, e interactuando con él hasta promover un consenso. Durante esta etapa todos los usuarios del nodo SKC pueden ingresar documentos en el sistema, pero sólo los pertenecientes al Grupo de Coordinación pueden opinar sobre ellos e influir en la cristalización de los mismos. En este sentido y durante la etapa inicial el Grupo de Coordinación ejerce labores de Comité Editorial. Es el Grupo de Coordinación el encargado también de decidir el fin de la Fase Inicial, promover su propia disolución y hacer que el sistema empiece a funcionar de modo no supervisado, pasando a la llamada Fase Activa.

En la Fase Activa existe una masa de usuarios, y conocimiento, suficiente para el correcto funcionamiento de los mecanismos de selección automática del conocimiento. En este caso todos los usuarios de un nodo del sistema se integran en las llamadas Comunidades Virtuales del nodo. Las dinámicas de estas comunidades están íntimamente ligadas a las estructuras de conocimiento que coexistan en el nodo. De hecho, hay tantas comunidades virtuales como grafos para organizar el conocimiento se hayan definido en él y todos los usuarios del mismo forman parte de cada una de ellas. Los usuarios del nodo pueden ingresar documentos en el sistema utilizando alguno de los grafos definidos en él, el sistema se encarga de clasificar los documentos en el resto de los grafos. Los documentos están sometidos a un proceso de cristalización en cada uno de los grafos del nodo. Cuando un documento cristaliza en un grafo, el usuario se convierte en experto en el área de conocimiento correspondiente al vértice donde se encuentra el documento y en los vértices conectados con el primero con determinadas relaciones, que pueden estar en otros grafos. Los usuarios expertos reciben un determinado número de derechos de voto por grafo, que pueden emplear en apoyar otros documentos ubicados en los nodos de los grafos correspondientes en los que son expertos. Este mecanismo está inspirado en la práctica de “peer review” que emplean la mayoría de las revistas científicas.

Estando el sistema en la Fase Activa, puede llegarse a una situación en la que ya no haya apenas cambios en los grafos de conocimiento, pocos documentos se publiquen en el sistema, la mayoría de la actividad en el nodo sea de consulta y el conocimiento esté muy cristalizado. En

estas condiciones se podría considerar que el sistema ha alcanzado una Fase Estable que duraría hasta que la situación cambiara y el nodo se reactivara para volver a la fase anterior.

### **3.2.4 Nodos y Redes de Nodos**

El nodo es el núcleo de SKC donde se reúnen los elementos fundamentales del sistema: el conocimiento sobre el dominio tratado y la comunidad de usuarios que interactúa con dicho conocimiento. El nodo es también el que soporta y constituye el nivel superior del sistema, la red SKC.

Los nodos disponen de unos Agentes de Red SKC para interactuar con otros nodos a través de la red. A partir del momento de su creación los agentes de los nodos buscan nuevos nodos permanentemente para integrarse con ellos en una red SKC. Además, intercambian información con los nodos ya conocidos para mantener la red común. Los agentes se encargan también de identificar los nodos que ya no están activos, y de controlar el acceso a la información y conocimiento de los nodos que cada uno representa.

Para poder actuar en el ámbito de red, los nodos SKC mantienen una representación de la red, que incluye modelos de todos los nodos que la integran, así como de la relación establecida con cada uno de ellos. La información que SKC maneja de cada nodo está formada por dos elementos: el Conocimiento de Red, que representa el conocimiento de todos los nodos que forman la Red SKC; y la Comunidad de Red, que refleja todas las comunidades de usuarios de cada nodo. Esta información se mantiene siempre actualizada gracias a la comunicación mantenida por cada nodo con los demás a través de los agentes.

Las relaciones entre nodos pueden ser de muchas clases: desde una básica soportada por un protocolo sencillo para el mantenimiento de la red, hasta una relación profunda con intercambio del conocimiento gestionado en los nodos implicados.

Las relaciones más profundas entre nodos sólo se producen entre nodos dedicados a temas afines. Las relaciones de esta clase pueden ser de distintos tipos, dependiendo de la consideración que la comunidad de un nodo tenga hacia el otro. Además, pueden ser asimétricas, porque las consideraciones de cada nodo por el otro no sean iguales. De manejar todas estas cuestiones se encargan los agentes.

Al iniciar la actividad de un nodo, SKC establece una página principal accesible desde la red IP con toda la meta-información necesaria para establecer comunicación con otros nodos SKC, procurando su registro en los buscadores más importantes. Es posible indicar al sistema dónde puede localizar a otros nodos SKC, lo que es imprescindible si el nodo no tiene acceso a Internet. Una vez iniciado, el sistema se encarga permanentemente de mantener la información

de la red, localizando nodos, poniéndose en contacto con ellos, y también detectando los nodos que han dejado de estar accesibles en la red.

### **3.2.5 Seguimiento de actividades**

En SKC todos los Interlocutores que interactúan con el sistema están siempre identificados y se mantiene un control de sus sesiones. Los datos recogidos se emplean en el proceso de cristalización del conocimiento y sirven para elaborar informes destinados a los usuarios del entorno. En SKC no sólo los usuarios interactúan con el sistema; se podría decir que la mayoría de los elementos que se consideran en el entorno son activos, interactuarán unos con otros y, por lo tanto, con el sistema. El análisis del trazo de sus actividades constituye una fuente fundamental de información, imprescindible para el funcionamiento del sistema.

Además, el seguimiento es particularizado, porque al tomar los datos de actividad de los distintos elementos considerados no sólo se registra de qué tipo de elemento se trata, sino también qué elemento es en particular. En el caso de los usuarios, esto supone que el sistema los debe identificar siempre que inicien una sesión de interacción con el entorno; algo parecido ocurrirá con los agentes encargados del intercambio de información con otros nodos o de los demonios encargados de la anotación automática de los documentos, por ejemplo.

Sin embargo, la aplicación de estos procedimientos no sólo permite obtener información sobre los usuarios y algunos parámetros del entorno, en el sistema todas las operaciones llevadas a cabo interactiva o automáticamente dejan registro. Con la información del registro se puede establecer un cronograma de la evolución de un documento con, por ejemplo, las visitas de sus versiones, o la variación del número de usuarios activos a lo largo del tiempo, o del uso de los enlaces propuestos por el sistema.

### **3.2.6 Cristalización del conocimiento**

En SKC los procedimientos de Cristalización se encargan de seleccionar el conocimiento gestionado por el sistema. Esta selección se realiza atendiendo a las consideraciones que sobre el conocimiento manifiestan los usuarios y a la información disponible en el sistema, generada por SKC o proveniente del resto de entidades que interactúan con él.

En el proceso de cristalización intervienen un objeto y un contexto. El objeto es el elemento que se somete a un proceso de selección como puede ser un documento, una anotación, un nodo, etc. El contexto son los parámetros que intervienen en dicha selección como son el conjunto de objetos implicados, el criterio de valoración establecido, el grupo de usuarios involucrados, la historia del proceso y el momento en el que se realiza la valoración. La cristalización de un objeto no depende sólo de él, sino también de los distintos contextos donde se sitúa. Por lo tanto

un mismo documento podría llegar a cristalizar de manera diferente en distintos grafos. Aunque esto no quiere decir que no se consideren los grados de cristalización que haya podido obtener el documento en otros ámbitos, sí indica que los procesos de cristalización deben considerarse independientes en SKC.

Nodos, grafos de conocimiento, documentos, versiones y anotaciones son Unidades de Conocimiento. Todas ellas están bajo un proceso de cristalización específico, que se adapta a las características de cada tipo de unidad y la clase de interacción que los usuarios pueden mantener con ella. En ocasiones el proceso de cristalización procura seleccionar un elemento dominante de entre todos los candidatos de una categoría que están en competición, así ocurre en cada tema con los documentos que pretenden describirlo, por ejemplo. En otras ocasiones no es así, en el caso de las anotaciones, el proceso de cristalización no pretende la selección de una sola de ellas, sino que tolera la persistencia en el sistema de las candidatas más votadas. En todos los casos, los candidatos que no reciben el suficiente apoyo desaparecen del sistema. En SKC algunas Unidades de Conocimiento emplean esquemas de cristalización diferentes a sus equivalentes en KC. Por ejemplo los grafos de conocimiento siguen uno más parecido al de las anotaciones que al de los documentos, puesto que el proceso no pretende la selección de uno sólo, al contrario de lo que ocurre con los árboles de conocimiento en KC.

SKC incorpora un procedimiento de cristalización en el ámbito de Red SKC en cada nodo para los nodos relacionados con el primero de determinada forma. En este caso se trata de ordenar los nodos con temática similar relacionados de cierta manera con el actual por su grado de aceptación.

En SKC, como en KC, el proceso de cristalización de las unidades atómicas de conocimiento - los documentos- se basa en el cálculo del llamado Grado de Cristalización [Cobos 2003]. Se trata de una medida de la aceptación recibida por un documento por parte de los usuarios. En este cálculo intervienen tanto las consideraciones explícitas, expresadas mediante votos y anotaciones, como las implícitas, debidas a la interacción con el documento de los usuarios, que en el caso de SKC se beneficia de la identificación permanente de los usuarios gracias al control de sesiones. La cristalización es en definitiva la evolución de tal aceptación, de forma que un documento aceptado de forma mantenida durante un periodo de tiempo es candidato a cristalizar, y otro que no, aspirante a desaparecer del sistema. Para que un documento cristalice es preciso que su grado de aceptación supere un cierto umbral establecido inicialmente. Sin embargo, un documento cristalizado también puede dejar de estarlo, esto ocurre si después de haber alcanzado dicho estado, durante un periodo de tiempo su grado de aceptación decae por debajo de otro valor establecido. En todos los casos lo que se valora es la evolución del grado de aceptación. Aunque la opinión de todos los usuarios se suele tomar en cuenta para evaluar la

aceptación -en algunas fases esto no es así-, a la opinión de los llamados Expertos se le asigna más valor que a la del resto de usuarios.

Como en SKC todos los elementos del sistema registran su actividad constantemente, los Interlocutores están siempre identificados y controladas sus sesiones de interacción con el sistema, es posible emplear toda esta información en los procesos de cristalización de las Unidades de Conocimiento, de la misma forma que en KC se consideraba el número de visitas para la cristalización de los documentos. Los datos considerados en cada caso dependen del tipo de Unidad.

### 3.3 Propuesta de Arquitectura para SKC

Con el fin de poner en práctica todo lo dicho se propone un diseño de arquitectura para el sistema Semantic KnowCat (SKC). Los requerimientos generales del mismo, desde el punto de vista de las entidades que interactúan con el núcleo del sistema, se recogen en los siguientes casos de uso básicos (ver Figura 3-1), análisis que se amplía como anexo de este documento.

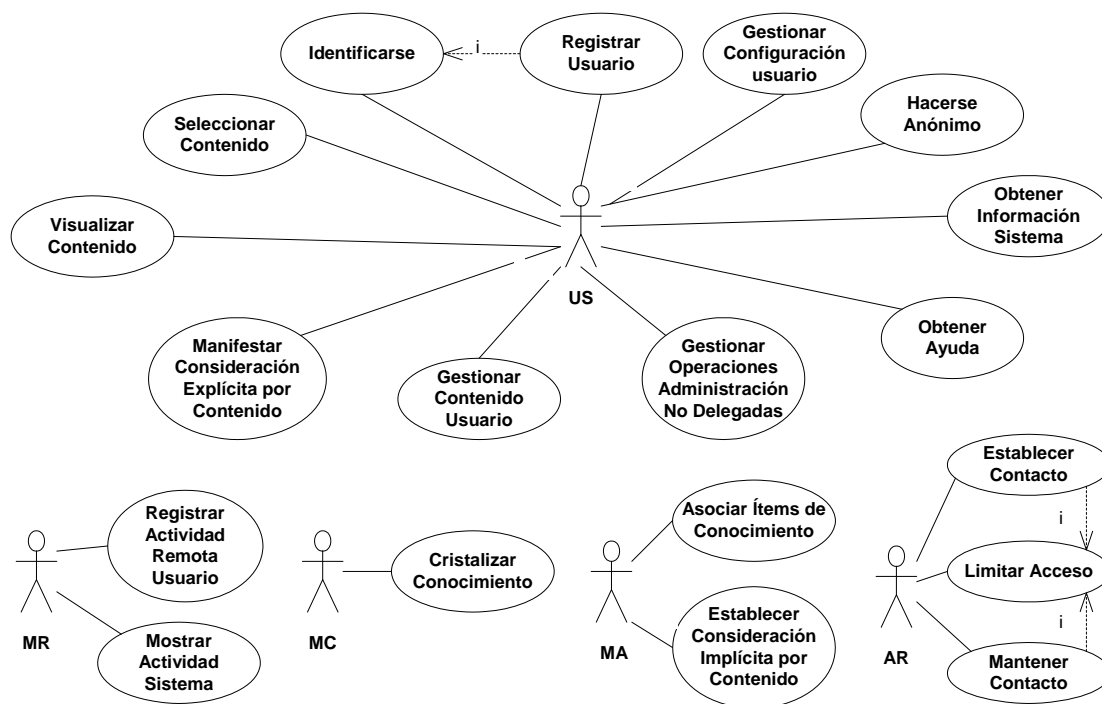


Figura 3-1. Diagrama Simplificado de Casos de Uso de SKC

Los actores considerados para el análisis son cuatro: US, representa a los usuarios humanos que interactúan con el sistema para trabajar con el conocimiento que este gestiona; AR, personifica a los agentes que interactúan con SKC en representación de otras instancias del sistema existentes en la red; MR, encarna al monitor del sistema, que se encarga de recopilar datos sobre la actividad del usuario e informar a estos sobre la actividad de SKC; MA, representa al módulo de

análisis, que interactúa con el núcleo del sistema para enriquecer el conocimiento que se maneja; y MC, al módulo de cristalización, que se encarga de interactuar con el núcleo de SKC para seleccionar el conocimiento gestionado.

El Monitor Remoto (MR) es un componente del propio sistema como el Módulo de Análisis (MA) y el Módulo de Cristalización (MC). Cada instancia del sistema tiene asimismo un Agente Remoto (AR), que además de interactuar con el propio sistema, lo hace también con los agentes de otras instancias de SKC a través de la red. Todas estas entidades se consideran actores, porque interactúan con el propio sistema de forma autónoma. Este enfoque permite capturar parte de los requerimientos que el metabolismo del sistema impone.

Los nombres de los casos de la mayoría de los usos representados en el esquema dan una idea de su finalidad, algunos sin embargo necesitan cierta explicación.

El Usuario del Sistema (US) necesita seleccionar y visualizar los contenidos que constituyen el conocimiento con el que puede trabajar y que SKC gestiona. Además, necesita gestionar los contenidos que él proporciona al sistema y su propia configuración. Sin embargo, no todas las operaciones de administración pueden ser realizadas por el interesado, se trata de las operaciones no delegadas que el usuario puede solicitar y seguir mediante el mecanismo que el sistema proporciona. Finalmente, el US puede manifestar de forma explícita la consideración que tiene hacia los contenidos del sistema.

El Monitor Remoto (MR) obtiene datos sobre la actividad del usuario en el sistema, más allá del servidor donde el núcleo del sistema reside, para que puedan emplearse como fuente de información para establecer la consideración implícita que los usuarios demuestran por los contenidos que manejan.

El Módulo de Análisis (MA) analiza los contenidos y los registros de actividad de los usuarios y el sistema para obtener información con la que enriquecer el conocimiento gestionado. Este módulo se encarga de dos tareas fundamentales: establecer asociaciones entre los elementos de conocimiento de forma automática, organizando el conocimiento de distintas formas, para facilitar la gestión del mismo; y establecer la consideración implícita que los usuarios demuestran por los contenidos que manejan, a partir de los registros de su actividad.

De forma general, la arquitectura propuesta se construye alrededor de las Ontologías, como en On2broker [Fensel 2001], que representan los elementos sobre los que trabaja SKC: por un lado el conocimiento que el sistema gestiona y por otro los modelos de todas las entidades que en él se necesitan para desarrollar tal actividad.

Para organizar el Conocimiento la comunidad de usuarios del nodo puede establecer tantas ontologías como considere oportuno; inicialmente no existe en el sistema ninguna organización

preestablecida. Siempre serán los usuarios los encargados de realizar propuestas y refinarlas, y la comunidad la responsable de aceptarlas al final.

Los Modelos de las Entidades del Sistema son también ontologías pero, a diferencia de las anteriores, no son los usuarios los encargados de especificarlas. Definen la representación de los elementos que el sistema maneja en el interior de los nodos y también la que se comparte entre ellos dentro de la red. De estas ontologías se hablará más adelante.

Además, la arquitectura del sistema incluye cuatro módulos activos fundamentales (ver Figura 3-2): el Gestor de Base de Datos, el Motor de Análisis, el Motor de Cristalización y el de Interacción.

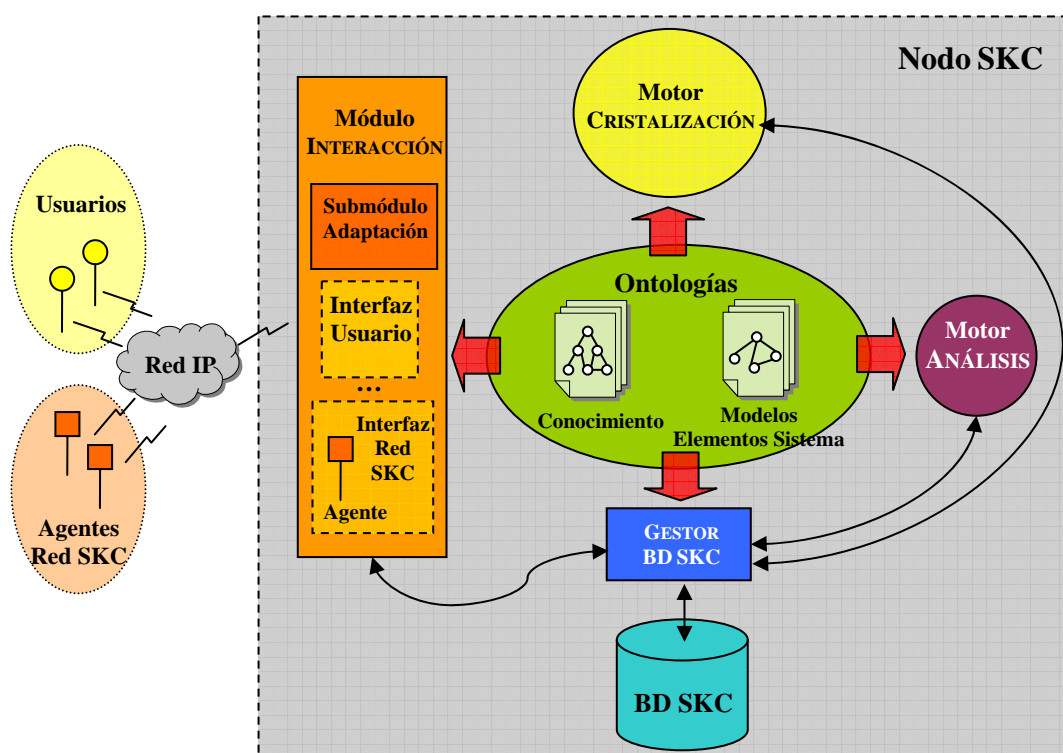


Figura 3-2. Arquitectura general del sistema SKC propuesto

El Módulo de Interacción soporta la comunicación con otras entidades a través de una Red IP, agrupando las Interfaces del sistema. En esta aproximación se consideran dos interfaces, la Interfaz de Usuario y la Interfaz de Red SKC. La primera permite que los usuarios accedan al nodo y la segunda que el sistema pueda interactuar automáticamente con otros nodos como él - mediante los Agentes de Red SKC- para formar y explotar una Red de Nodos SKC. Las interfaces utilizan un Submódulo de Adaptación que se encarga de acomodar los contenidos, la navegación y la presentación a las características y preferencias de los interlocutores y las limitaciones de su modo de acceso al sistema.

El Motor de Análisis se ocupa de la revisión constante del conocimiento y de los registros de actividad del sistema. El primero está constituido por los documentos y los grafos que lo organizan, y el segundo por la información generada por el sistema y las entidades que interactúan con él. Como resultado del análisis se obtiene información y conocimiento nuevos, que sirven para enriquecer el conocimiento disponible en el sistema y para revelar datos importantes destinados a los módulos de Interacción y de Cristalización.

El Motor de Cristalización se dedica a seleccionar el conocimiento que el sistema gestiona. Para ello toma en cuenta la consideración que los usuarios demuestran a través del módulo de Interacción, y la información disponible en el sistema. Esta última, es generada por el propio SKC o suministrada por el resto de entidades que interactúa con él. En todos los casos la información es procesada por el Motor de Análisis.

Ambos motores emplean aplicaciones que funcionan en segundo plano para llevar a cabo sus objetivos, son los Demonios de Análisis y Demonios de Cristalización. Se trata de aplicaciones que realizan operaciones repetitivas configurables y que dependen de un módulo de control que regula su actividad. A estas aplicaciones no se las puede considerar agentes, como a los Agentes de Red SKC, porque no exhiben las cualidades características de este tipo de entidades.

El Gestor de Base de Datos trabaja sobre una Base de Datos y articula todos los elementos anteriores, proporcionando soporte para: almacenar las Ontologías de forma adecuada para su explotación; atender a todas las peticiones de manejo de información requeridas por el Módulo de Interacción; y cubrir todas las necesidades de información y almacenamiento del Motor de Análisis y del de Cristalización.

Normalmente cada instancia del sistema SKC se configura sobre una única máquina. Sin embargo, la arquitectura se adapta a un despliegue distribuido del nodo, de forma que los distintos módulos se pueden ejecutar en diferentes equipos, con tal de que se establezcan las condiciones de accesibilidad adecuadas entre ellos.

La infraestructura sobre la que montar el sistema consiste en un diseño en tres niveles: un servidor Web, encargado de la interacción con los distintos interlocutores; un servidor de aplicaciones, que soporte los requerimientos de procesos y proporcione soporte para el manejo de sesiones y transacciones; y un servidor de base de datos, que responda a las necesidades de gestión de información. Tal arquitectura puede funcionar en un solo equipo, pero permite su escalado en distintas máquinas si el volumen o el peso del requerimiento de servicio sobre alguna de sus partes así lo aconsejan.

A continuación se describirán más en detalle los elementos más importantes de la arquitectura propuesta, a través de cinco apartados dedicados a: Ontologías, Gestor de Base de Datos, Motor de Análisis, Motor de Cristalización y Módulo de Interacción.



### 3.3.1 Ontologías

Como ya se ha dicho, en el centro de la arquitectura se encuentran las ontologías que modelizan los distintos elementos que concurren en SKC; unas describen el propio sistema mientras que otras representan el conocimiento con el que el sistema trabaja.

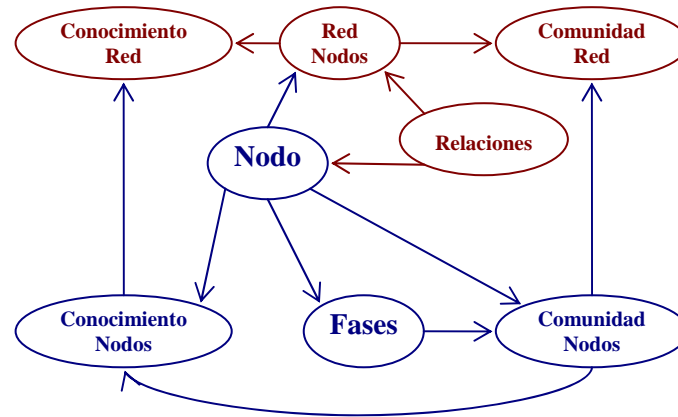


Figura 3-3. Conceptos generales considerados en SKC

Dentro del primer grupo, se encuentra la ontología que muestra el esquema de las entidades de ámbito general que el sistema considera (ver Figura 3-3). El Nodo es el epicentro de la conceptualización, es la base del nivel superior de Red y nexa de unión de los elementos esenciales del sistema: el Conocimiento de Nodo del dominio de conocimiento tratado y la Comunidad de Nodo de usuarios que interactúa con ese conocimiento.

En el nivel de red se identifican dos nuevos conceptos, el Conocimiento de Red, que integra los dominios de conocimiento tratados en todos los nodos de la red y el de Comunidad de Red, que como el anterior reúne todas las comunidades de usuarios de los nodos en el ámbito de la red. Además, se incorpora el concepto de Relación entre Nodos, que modeliza la vinculación que se establece entre los nodos de la red que se dedican a dominios de conocimiento semejantes.

Los nodos pasan por varias Fases a lo largo de su existencia; de estas fases dependen distintos aspectos del funcionamiento del sistema, que alcanzan desde la organización de la comunidad de usuarios, hasta el cálculo de los distintos grados de cristalización considerados en el sistema, pasando por las funcionalidades disponibles en cada caso.

La Consideración (ver Figura 3-4) que los usuarios pueden demostrar sobre algunas entidades del sistema se modeliza como la combinación de la consideración expresada de forma Explícita, como son los votos o las anotaciones, y de forma Implícita a través de la actividad desarrollada en relación con las entidades consideradas, como pueden ser el número de visitas realizadas o el tiempo invertido en cada una de ellas.



todos los grafos de conocimiento definidos en el nodo y Relaciones de Grafos Virtuales de Nodo establecidas entre las imágenes de vértices de distintos grafos, representados por los Vértices de Grafos Virtuales de Nodo. Los vértices de este grafo se incorporan a medida que se definen los grafos de conocimiento del nodo, y las relaciones entre sus vértices como resultado de la actividad del sistema, en ambos casos de forma automática supervisada por la comunidad de usuarios.

Un concepto parecido al anterior es el de Grafo Virtual de Conocimiento de Red. En este caso, se trata de un grafo que incluye una representación de los vértices de todos los grafos de conocimiento definidos en los nodos relacionados de la red, y Relaciones de Grafos Virtuales de Red establecidas entre las imágenes de vértices de los grafos de distintos nodos, representados por los Vértices de Grafos Virtuales de Red. Este concepto, igual que el previo, soporta parte del enriquecimiento del conocimiento en los distintos nodos. También en este caso, los vértices del nuevo grafo se incorporan a medida que se establecen las relaciones entre nodos en la red y se definen los grafos de conocimiento del nodo, mientras que las relaciones entre sus vértices se crean como resultado de la actividad del sistema. En todos los casos el proceso se realiza de forma automática supervisada por la comunidad de usuarios.

De obtener la información necesaria sobre los otros nodos de la red se encarga el Agente de Red, que se ocupa de conseguir la información necesaria para crear los Vértices de Grafos Virtuales de Red de otros nodos en la red relacionados con el que nos ocupa, y también de proporcionar la información necesaria sobre los vértices de los grafos del nodo para crear sus representaciones en los nodos de la red que corresponda.

Los Vértices son el lugar donde se ubican los Documentos (ver Figura 3-7) en los grafos, y constituyen -junto con estos- los contenedores fundamentales del conocimiento manejado en el sistema. El documento se considera de forma abstracta, puesto que es la reunión de sucesivas Versiones, de las que sólo una es la representante del grupo en cada momento. Las versiones gozan, como los grafos, de una determinada Consideración por parte de los usuarios del sistema, que puede manifestarse de forma Explícita mediante Votos y Anotaciones, o Implícita por la utilización que se haga de aquellas, esta consideración cumple distintas funciones en el sistema; de entre ellas destaca la de servir de fundamento para establecer el Grado de Cristalización de los Documentos.



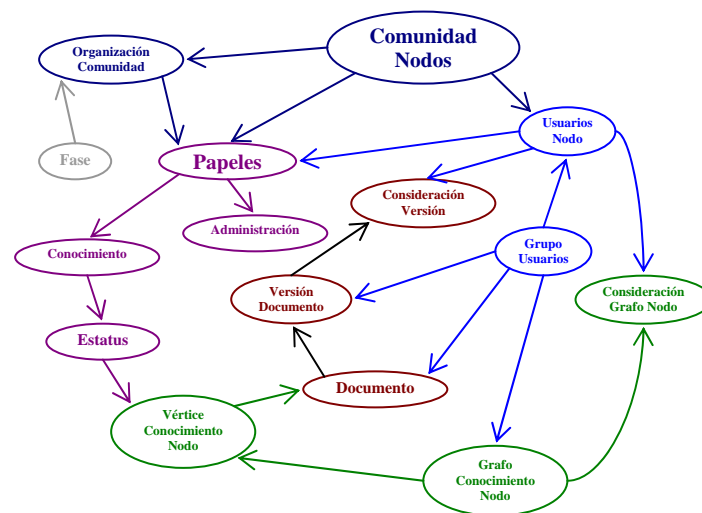


Figura 3-8. Modelización de la Comunidad de Nodo

Los usuarios del nodo desempeñan distintos Papeles en el sistema, en lo que se refiere a la Gestión del Conocimiento y la Administración del propio sistema. En el primer caso, además estos papeles pueden ir acompañados, en ciertas ocasiones, de unos determinados Estatus, que establecen la consideración de los usuarios en los vértices de los grafos que organizan el conocimiento y estipulan poderes concretos que al usuario se otorgan -como la consideración de experto y la posibilidad de opinar sobre la cristalización de los documentos de un tema en la fase activa de un nodo-.

Como ya se ha mencionado antes, los nodos pasan por varias fases que provocan, entre otras cosas, distintas Organizaciones de la Comunidad de Usuarios. Tales organizaciones determinan la dinámica de las comunidades y la distribución de papeles entre los usuarios –por ejemplo, en la fase inicial de los nodos se constituye una organización supervisada que establece la existencia de un grupo de coordinación con facultades de equipo editorial y capacidad de decisión, mientras que el resto de los usuarios sólo pueden ingresar conocimiento en el sistema y revisar el existente-.

El segundo grupo de ontologías que se mencionaba al comienzo de esta sección, es el de las que modelizan el conocimiento con el que el sistema trabaja. Dentro de este grupo se encuentran las ontologías específicas que describen los dominios de conocimiento tratados en cada nodo. Estas ontologías se encuentran representadas, como Grafos de Conocimiento de Nodo, en el esquema general especificado anteriormente y sirven para describir las distintas organizaciones que la comunidad de usuarios de cada nodo consensúan para manejar el dominio de conocimiento sobre el que en el nodo se trabaja.

Estas ontologías representan las áreas en que se puede dividir el dominio que se está tratando y las relaciones que se establecen entre dichas áreas. Como un mismo dominio puede organizarse de diversas formas, el sistema permite que varias ontologías coexistan en un mismo nodo. Todas ellas comparten un concepto que representa el dominio de conocimiento que todas describen, aunque cada una lo hace de una forma distinta. Los conceptos de cada ontología se encuentran relacionados en su interior de acuerdo a las vinculaciones establecidas en el momento de su definición, con la aprobación de la comunidad. Sin embargo no existen inicialmente relaciones entre conceptos de distintas ontologías: la actividad del sistema se encarga de establecer automáticamente tales vínculos entre los conceptos de ontologías distintas, con la supervisión de la comunidad de usuarios –un concepto de una ontología puede representar la misma área de conocimiento que otro en una ontología diferente, o referirse a una parte del área delimitada por aquella, o abarcar las áreas acotadas entre varios conceptos en la otra ontología, etc.-.

Algo parecido sucede con las ontologías de otros nodos integrados en la misma red SKC y que mantienen determinadas relaciones con el nodo que nos ocupa porque trabajan sobre dominios similares. Como en el caso anterior, la actividad del sistema se encarga de establecer automáticamente los vínculos entre los conceptos de las ontologías de los distintos nodos, con la supervisión de la comunidad de usuarios. La modelización de estas posibilidades se ha tratado anteriormente en este apartado, al hablar de los Grafos Virtuales de Conocimiento de Nodo y los Grafos Virtuales de Conocimiento de Red.

### **3.3.2 Gestor de Base de Datos**

Como se ha comentado al principio de este capítulo, el Gestor de Base de Datos de SKC (GBDSKC) trabaja sobre la Base de Datos de SKC (BDSKC) y es la espina dorsal que articula todos los elementos que constituyen el sistema: se encarga de almacenar las Ontologías de forma adecuada para su explotación, de atender a todas las peticiones de manejo de información requeridas por el Módulo de Interacción, y de soportar todas las necesidades de información y de almacenamiento del Motor de Análisis y del de Cristalización.

Sobre los procedimientos para explotar las ontologías como bases de datos, ya se habló en el capítulo de Estado del Cuestión. Se puede optar desde la representación en memoria del grafo correspondiente, al empleo de una base de datos relacional que almacene la ontología en notación de triplas RDF, pasando por la utilización de bases de datos XML nativas [Emerick 2002]. La elección del sistema más adecuado depende de las características del proyecto; el primer método puede ser adecuado, y más rápido, para sistemas que manejen ontologías de tamaño moderado, puesto que los requerimientos de memoria son proporcionales a éste y al número de instancias tratadas. En el segundo caso el sistema podría escalar mejor si el gestor de

base de datos relacional es adecuado y el diseño establecido para el mapeo de la ontología también [Hjelm 2001].

En esta aproximación se propone utilizar una representación de las ontologías en RDF en notación de triplas sobre una base de datos relacional, teniendo en cuenta que el sistema esta obligado a manejar varias ontologías simultáneamente, y que el tamaño de algunas de ellas y sus números de instancias puede llegar a ser grande. Un sistema en una red SKC que mantiene relaciones con varios nodos, además de las ontologías que representan a los elementos del sistema, podría tener varias ontologías en forma de grafos para organizar el conocimiento con decenas de vértices y relaciones, cientos de documentos, decenas de usuarios, así como los modelos de los nodos relacionados y de sus grafos de conocimiento, sumando cientos de vértices.

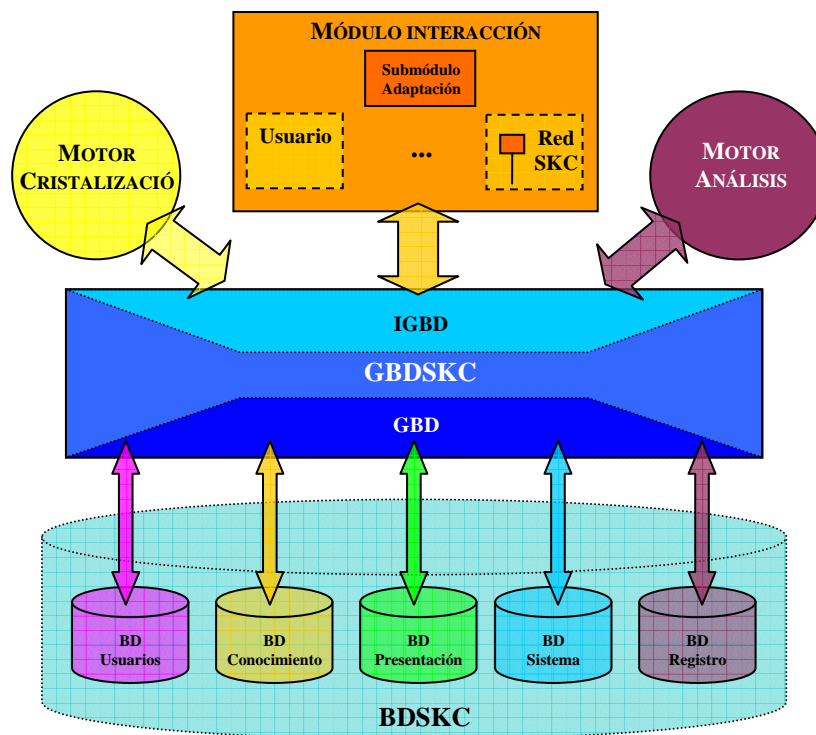


Figura 3-9. Gestor de Base de Datos de SKC

En cualquier caso, se considera oportuno dividir el sistema de Gestión de Base de Datos (ver Figura 3-9) en dos capas, la primera y más cercana a la conceptualización de las ontologías, sirve de interfaz del GBDSKC con el resto del sistema y permite enmascarar la base de datos que se encuentra por debajo. La segunda capa es el gestor de base de datos propiamente dicho. El resto de módulos del sistema utilizarán la primera capa para interactuar con el GBD, independizándoles de esta manera de las decisiones sobre el motor y modelo de base de datos elegido, y la representación de las ontologías en este.

Por otro lado, mediante las ontologías, lo que la base de datos almacena y permite manejar son los modelos con los que SKC representa el conocimiento y los distintos elementos del sistema. En este sentido, se puede considerar el repositorio formado por varias bases de datos:

- *Usuarios*, que reúne los modelos que representan a los usuarios individuales, sus agrupaciones y su integración en comunidades.
- *Conocimiento*, que reúne los modelos de documentos y grafos de conocimiento.
- *Presentación*, compuesta por las plantillas y elementos de presentación para usuarios y resto de interlocutores del sistema.
- *Sistema*, que incluye los modelos de sistema, nodos y red.
- *Registro*, que constituye el repositorio donde se registra toda la actividad del sistema.

La información y el conocimiento relativo a la red se reparten entre las anteriores bases de datos, puesto que la explotación de esta información se realiza en los distintos nodos en el contexto que cada una de ellas representa.

Esta aproximación basada en ontologías facilita la evolución de la estructura de datos del sistema sin necesidad de modificar el diseño de la base de datos. Así por ejemplo podría añadirse un nuevo atributo al modelo de usuario sin cambiar la base de datos subyacente. Esta posibilidad se potencia, como se explicará más adelante, con la aproximación propuesta para el motor de análisis.

El empleo de un gestor de base de datos como soporte del repositorio de información facilita, entre otras cosas, la resolución de problemas de concurrencia al acceder a los datos entre los distintos módulos y la posibilidad de que cada nodo del sistema pueda configurarse en una o más máquinas.

Todas las operaciones realizadas por el Gestor de Base de Datos de SKC van siendo anotadas en el registro del sistema.

### **3.3.3 Motor de Análisis**

Como ya se ha dicho antes, el Motor de Análisis se ocupa de la revisión constante del conocimiento y de la información generada por el sistema y las entidades que interactúan con él. Con ello se pretende obtener nuevo conocimiento con el que enriquecer el disponible en cada momento y poner de manifiesto información relevante para que la puedan utilizar los módulos de Interacción y el de Cristalización.

El conocimiento del sistema está en constante cambio debido a la interacción de usuarios y otras entidades a través de la red –en esta propuesta, los agentes de otros nodos SKC-. En cada nodo



los usuarios pueden interactuar de muchas formas: ingresando documentos y grafos para organizar el conocimiento; manifestando su consideración hacia ellos mediante votos, anotaciones o interaccionando con ellos; expresando de forma explícita sus preferencias sobre el modo de presentación de la información o de forma implícita a través de su perfil; estableciendo relaciones con otros nodos de la red SKC; cambiando el comportamiento del sistema al modificar la fase, y con ello la organización de la comunidad y de la forma de trabajo, etc. En la red SKC, los agentes deben interactuar manteniendo actualizados los datos básicos de todos los nodos incorporados, transmitiendo la información correspondiente a los cambios sufridos en cada nodo a los correspondientes relacionados, e intercambiando datos sobre las instancias del sistema que pueden ingresar o desaparecer de la red.

El Motor de Análisis está constituido por varios Demonios de Análisis, cada uno dedicado al examen de una parte del conocimiento o de los elementos del sistema. Cada demonio se centra en una de las bases de datos enunciadas en el apartado anterior, aunque no todas requieren el mismo grado de examen. De este modo existen demonios dedicados al análisis: de los usuarios, del conocimiento, de la presentación, del sistema y del registro. Aunque cada demonio trabaja sobre una base de datos, casi siempre requiere de datos almacenados en otras para llevar a cabo su labor.

Los demonios de análisis, como los de cristalización, son aplicaciones que funcionan en segundo plano realizando operaciones repetitivas, que son configurables. No son agentes, como los Agentes de Red SKC, porque no exhiben las cualidades características de tales entidades.

El Motor de Análisis utiliza una estrategia parecida a la de las arañas que emplean los buscadores para indexar la Web [Baeza 1999]. Tales sistemas suelen constar de uno o más procesos que se ejecutan en una máquina con acceso a Internet, desde allí recorren la Web visitando distintas páginas para recolectarlas –para que luego otro programa las pueda indexar-, siguiendo los enlaces que unen unas con otras, hasta agotar los vínculos o cambiar de trayectoria por un criterio establecido a priori. Normalmente parten de una serie de URLs seleccionadas por alguna razón y procuran alcanzar un determinado volumen de páginas a partir de ellas. Su trabajo nunca termina, porque la Web esta en permanente evolución, cuando se ha concluido un ciclo de recolección-indexación se vuelve a empezar según una planificación determinada.

En el caso que nos ocupa, cada uno de los demonios del motor de análisis utiliza las ontologías para recorrer las instancias de las clases almacenadas en las bases de datos, siguiendo algunas de las relaciones que enlazan sus conceptos. De este modo, por ejemplo, se recorrerán los distintos grafos que organizan el conocimiento, buscando los documentos clasificados en sus vértices para ir actualizando los valores de ciertos atributos, o se visitarán todos los usuarios registrados en un nodo para establecer el prototipo de su documento más visitado.

Para ello, el Motor de Análisis dispone de una unidad de control que regula la actividad de los demonios, establece la periodicidad o circunstancias que determinan su activación, el límite de recursos que pueden dedicar a desarrollar su labor, los periodos en los que pueden desempeñarla, los parámetros que fuerzan los cambios en la intensidad de su actividad y las condiciones que restringen el paralelismo en su desempeño.

A cada demonio se le asigna una serie variable de tareas, estas tareas vienen dadas por las clases y los atributos que debe tratar, las circunstancias en que se debe realizar y la forma en que tiene que llevarse a cabo. Este planteamiento permite que los demonios puedan adaptar su labor a la evolución del sistema, como lo permite la aproximación basada en ontologías que soporta el entorno.

Aunque las tareas encomendadas a cada demonio pueden ser más complicadas que la recolección de páginas Web para su indexación, las dimensiones del ámbito considerado -un nodo-, reduce los requerimientos de su aplicación. Además, las características de la arquitectura del sistema propuesto permiten la distribución de los módulos que constituyen un nodo en distintas máquinas convenientemente comunicadas. Tal posibilidad permite la escalabilidad del sistema, puesto que se pueden añadir más equipos en función de los requerimientos de la función de análisis o de otros módulos del sistema.

Como se puede deducir de lo dicho anteriormente, toda la actividad del Motor de Análisis se desarrolla en segundo plano. Por esta razón es fundamental que el módulo sea sensible a las condiciones del sistema, procurando que su actividad no afecte a la prestación del entorno. En la mayoría de los servicios cara al público en la red existen periodos de mayor actividad de los usuarios y otros donde ésta decae; se trata de aprovechar estos momentos para intensificar la actividad de análisis, como también se deberá hacer con los procesos automáticos relacionados con la red SKC. En otros momentos, habiendo una interacción moderada de usuarios, puede ser conveniente mantener una actividad moderada también de análisis. Además, en todos los casos el módulo tiene que adaptar su comportamiento a las condiciones cambiantes del entorno, reduciendo o aumentando la intensidad de su actividad.

En el Motor de Análisis tiene especial relevancia el tratamiento de los documentos y del registro de actividad. El primero tiene aplicación en la mayoría de los modelos de los elementos del sistema, y es fundamental para el establecimiento de vínculos entre ellos. El segundo tiene una utilidad parecida pero proporcionando una perspectiva temporal del desarrollo del sistema, de la interacción de sus usuarios, de la actividad de su red SKC, de la evolución de su conocimiento, etc. Ambas aproximaciones se basan en ideas del campo de Minería de Información y Datos [Baeza 1999] [Chang 2001]. A continuación tratamos con mayor detalle ambos aspectos.

### **3.3.3.1 Tratamiento de los Documentos**

El objetivo perseguido es establecer automáticamente relaciones entre los propios elementos que constituyen el conocimiento que maneja el sistema -documentos, grafos y nodos-, y también con otros elementos del entorno. Existen varias razones que justifican este interés, entre otras:

- Con el fin de que el sistema permita la coexistencia de varias ontologías para organizar el conocimiento en cada nodo, y que puedan utilizarse alternativamente para acceder a él, se necesita establecer mapeos entre las mismas.
- Algo parecido se requiere para soportar la posibilidad de compartir conocimiento entre distintos nodos relacionados en una red SKC, donde las ontologías que organizan el conocimiento serán casi siempre diferentes.
- Es interesante saber qué documentos se parecen a uno dado y qué vértices de los grafos de conocimiento tienen documentos similares. Con estos datos se pueden proponer accesos a documentos y temas relacionados con el documento o el tema que se está visitando en el propio nodo o en otros vinculados.
- Cuando un usuario de un nodo utiliza una determinada ontología para acceder al conocimiento e ingresar nuevos documentos, es útil que el sistema pueda proponer ubicaciones en otras ontologías que coexisten en el nodo, para facilitar la labor y evitar que la posible ventaja de disponer de varias organizaciones del conocimiento se convierta en un inconveniente.
- Cuando los nodos se integran automáticamente en una red SKC es conveniente que los sistemas puedan proponer de forma pro-activa posibles relaciones con otro nodo, que sólo son viables cuando los campos de conocimiento tratados en las parejas de nodos candidatos son similares.
- Cuando un usuario trabaja con el conocimiento de un nodo es útil saber qué tipo de documentos visita, cuáles inserta y en qué temas es experto; con esta información el sistema puede proponer el acceso a documentos que pueden interesar al usuario o también ponerle en contacto con otros usuarios del sistema con intereses parecidos.

En el sistema, el elemento atómico de conocimiento es el documento. Los grafos de conocimiento sirven para ordenar en sus vértices los documentos que se ingresan en los nodos, las relaciones de los usuarios a través del sistema se derivan, directa o indirectamente, de su interacción con los documentos, etc. En definitiva, estableciendo vínculos automáticos entre los documentos se pueden encontrar muchas relaciones entre distintos elementos del entorno.

Para este tratamiento se propone una aproximación basada en algunas ideas del área de Minería de Textos, de las que ya se habló en el capítulo de Estado del Cuestión. Basándose en ellas se pueden analizar los documentos textuales clasificados en las estructuras de conocimiento del sistema, que actualmente son la mayoría. Ideas de otras áreas de Minería de Información y Datos –minería de contenidos- podrían utilizarse para manejar, de forma parecida, otros tipos de documentos –audio, vídeo...-, o descubrir la estructura interna de unos y otros [Chang 2001], pero todo esto sobrepasa los objetivos de este trabajo.

De forma general el procedimiento propuesto consiste en: establecer un vector representativo de los términos más importantes, palabras claves, de cada documento; de forma parecida, establecer un vector para todos los vértices de los grafos de conocimiento, considerando los documentos ubicados en ellos; de manera análoga, establecer vectores de los nodos partiendo de los documentos que incluyen; y hacerlo también con los usuarios, considerando los documentos que manejan, etc. Comparando los vectores de los elementos se intentará desentrañar relaciones entre ellos, con procedimiento basado en [Chang 2001] y [Baeza 1999].

Para crear el vector de palabras claves de un documento es preciso identificar las palabras más características del mismo representadas por sus raíces. Su dimensión dependerá, en cada caso, del número de términos con la suficiente consideración de importancia, que se establece como un umbral inicial. La importancia de las palabras en los documentos es una combinación de su frecuencia y de la frecuencia del documento inverso. La primera es el número de ocurrencias del término en el documento. La segunda, un valor asignado a las palabras que es alto para las raras y bajo para las comunes.

Para establecer el grado de similitud entre dos documentos es necesario, en primer lugar, reformular sus vectores en términos comunes, y luego establecer la correlación entre ellos, que puede cuantificarse por el coseno del ángulo que forman. Este valor permite establecer relaciones entre ambos documentos de acuerdo a algún criterio.

Algo parecido puede hacerse con los vértices de los grafos de conocimiento. Si ya existen documentos ubicados en ellos se puede calcular el vector del documento prototipo del vértice, combinando los vectores de todos ellos. Si no es así, se puede proponer un ejemplo que sirva de modelo para calcular el vector correspondiente. Como los documentos ubicados en un vértice van cambiando, el vector en cuestión deberá recalcularse en función de los documentos que en cada momento se ubiquen en él. Comparando estos vectores se pueden establecer relaciones entre los propios vértices y también de éstos con todos los documentos.

De forma similar se puede asociar un vector prototipo a los nodos SKC, proporcionando inicialmente una serie de documentos representativos del dominio de conocimiento tratado, y

luego considerando los documentos que se vayan ingresando en el sistema. Con estos vectores se pueden identificar otros nodos dedicados a dominios de conocimientos similares.

Todo esto también es aplicable a los usuarios del entorno. Por ejemplo, como resultado de su interacción con el sistema, los usuarios visitarán ciertos vértices de los grafos de conocimiento más que otros. La combinación de los vectores prototipo de estos vértices, pueden representar el prototipo de temas que a un usuario interesan. Esto puede llevar a identificar usuarios con los mismos intereses. Un esquema parecido puede aplicarse a los temas en los que los usuarios son expertos o los documentos de los que son autores.

Entrando un poco más en detalle, algunos procedimientos de agrupación incremental – *incremental clustering*-, ideados para la clasificación de noticias en el contexto informativo, [Chang 2001] pueden servir en el ámbito de la propuesta para clasificar automáticamente los documentos en diferentes categorías, recalculando a la vez el nuevo vector prototipo como resultado de la clasificación. Sirva de ilustración el caso de la clasificación de un documento en un determinado vértice de un grafo de conocimiento; en tal caso, el vector del documento se compararía con el vector prototipo de los documentos ya clasificados en el vértice. Si se supera el umbral de similitud establecido, el nuevo documento se ubicaría en el vértice, y el vector prototipo de éste debería integrar, a partir de ese momento, también el vector del documento. En SKC, además, existe la posibilidad de que un documento desaparezca de un grafo, con los consiguientes cambios en los prototipos de los vértices donde este se encontrara. Todo esto es de aplicación también para los nuevos documentos ingresados en un nodo, visitados por un usuario, etc.

### **3.3.3.2 Tratamiento del Registro de Actividad**

El segundo aspecto a resaltar dentro del apartado sobre el Motor de Análisis es el tratamiento del registro de actividad. Para este proceso se propone una aproximación basada en algunas ideas del área de Minería Web, en particular las de Minería de Uso de la Web. Con ellas se analizan los registros de actividad del sistema para descubrir patrones de uso, que sirven para extraer información útil para el propio sistema, como hace de forma básica [Perkowitz 2000]. Ideas de otras áreas de la Minería Web –Minerías de Estructura Web- podrían utilizarse para intentar desentrañar los enlaces internos de los documentos incluidos en el sistema, pero su investigación va más allá de los objetivos de este trabajo.

En particular se propone emplear estas ideas para descubrir patrones de uso de la siguiente forma [Chang 2001]:

- Analizar de forma estática los ficheros de registro del sistema, para obtener la frecuencia de acceso a documentos, vértices, grafos, etc., determinando el número de

visitantes de los nodos, la frecuencia de errores y en general la información relevante para la monitorización del sistema. Con este tipo de datos se puede, por ejemplo, detectar algunos problemas del entorno, establecer los ciclos de actividad del sistema para planificar la intensidad y oportunidad de las labores de análisis, mantenimiento o interacción automática, etc.

- Descubrir reglas de asociación, también en los ficheros de registro del sistema, que procuran encontrar conjuntos de elementos accedidos conjuntamente con una frecuencia superior a un valor umbral preestablecido. Estas reglas proporcionan conocimiento sobre los comportamientos de acceso al sistema de los usuarios y vinculaciones entre elementos del sistema.
- Realizar agrupaciones sobre los ficheros de registro del sistema, para identificar la actividad de los usuarios por sesiones -interacciones sucesivas desde una misma ubicación en la red, entre las que no transcurre mucho tiempo y se las puede considerar provocadas por el mismo usuario o entidad- e identificar usuarios con patrones de navegación similares -porque visitan los mismos elementos del sistema-.
- Realizar clasificaciones para organizar los usuarios en categorías predefinidas según, por ejemplo, sus patrones de navegación -cómo acceden a muchos, pocos o ningún documento- o la intensidad de su actividad en el sistema -los que interactúan con el sistema casi todos los días, los que lo hacen casi todas las semanas o los que sólo acceden esporádicamente-.
- Descubrir patrones secuenciales en los ficheros de registro temporal del sistema, para identificar patrones que se repiten con frecuencia en sesiones de interacción con el sistema, como el acceso periódico a determinados elementos del mismo. Por ejemplo, un número significativo de usuarios acceden a sus propios documentos por lo menos una vez a la semana, pero no leen las anotaciones de éstos o proponen versiones de sus documentos durante un mes.

Todos los usuarios que interaccionan con el sistema están siempre identificados y se mantiene un control de sus sesiones, igual que ocurre con los agentes de red. Sin embargo, la aplicación de estos procedimientos no sólo permite obtener información sobre los usuarios y algunos parámetros del entorno; en el sistema todas las operaciones llevadas a cabo interactiva o automáticamente dejan registro. Con la información del registro se puede establecer un cronograma de la evolución de un documento con, por ejemplo, las visitas de sus versiones, o la variación del número de usuarios activos a lo largo del tiempo, o del uso de los enlaces propuestos por el sistema.

### 3.3.4 Motor de Cristalización

El siguiente módulo de la arquitectura propuesta es el Motor de Cristalización, que se dedica a seleccionar el conocimiento que el sistema gestiona, atendiendo a la consideración que los usuarios demuestran a través del Módulo de Interacción y a la información disponible en el sistema, generada por el propio SKC o suministrada por el resto de entidades que interactúan con él, información procesada en ambos casos por el Motor de Análisis.

La noción de cristalización es un concepto relativo que involucra un objeto y un contexto. El objeto puede ser un documento, un grafo de conocimiento, una anotación, un nodo o cualquier otro elemento que dentro de un nodo tenga sentido someter a un proceso de selección. El contexto viene dado por varios parámetros:

- *Criterio de agrupación*, tema de unos documentos, dominio de un grafo o de un nodo relacionado con otros en una red SKC, etc.
- Conjunto de *objetos en competencia*, el resto de documentos clasificados en el vértice de un grafo de conocimiento, los demás nodos relacionados en una red SKC, etc.
- *Criterio de valoración*, que depende del tipo de objeto y suele ser una combinación de los votos emitidos por los miembros de un determinado grupo de usuarios, los accesos de un documento, los usuarios que utilizan un grafo o visitan regularmente un nodo, etc.
- *Grupo de usuarios*, a pesar de que en cada momento el grupo de usuarios se puede considerar como un bloque y constituido con un criterio único, su representatividad y composición puede variar y esto afectar profundamente a la valoración del objeto en un momento dado.
- *Historia* del propio proceso de selección, puesto que el cálculo del grado de cristalización de un objeto considera los grados anteriores.
- *Momento* en el que se realiza la valoración, debido a que todos los parámetros enunciados son variables en el tiempo.

De esta forma, cuando un documento cristaliza en un vértice de un grafo de conocimiento de un nodo, es porque los usuarios que utilizan esa organización han demostrado una consideración hacia él superior a un umbral establecido y relativa al resto de documentos incluidos en el vértice, con la intención de seleccionar la mejor descripción del área de conocimiento que dicho vértice representa. De un modo parecido se podría describir en qué consiste la cristalización de un nodo o de un grafo, por ejemplo.

Aunque no siempre lo parezca, todos los parámetros que forman parte del contexto de cristalización son variables. Por ejemplo, en esta propuesta se considera que varios grafos de

conocimiento pueden coexistir en un nodo, esto significa que los documentos ingresados en el sistema se pueden ubicar en vértices de distintos grafos simultáneamente. En estas condiciones los documentos estarán sometidos a procesos de cristalización independientes en todos vértices donde se coloquen. Por lo tanto un mismo documento podría llegar a cristalizar en un grafo y desaparecer en los demás. La cristalización del documento no depende sólo de él, sino también de los distintos contextos donde se sitúa.

Todo lo anterior no quiere decir que en determinados casos los contextos correspondientes no consideren en sus criterios de valoración los grados de cristalización que el documento haya podido lograr en otros ámbitos. Sin embargo, en general, los procesos de cristalización deben considerarse independientes.

Bajo estas condiciones la masa crítica de usuarios y documentos necesaria para el adecuado funcionamiento del sistema será mayor a medida que se incluyan nuevos grafos en el nodo, si los usuarios se dividen en grupos que utilizan cada grafo y no participan en los procesos de cristalización de los demás, aunque lleguen a compartirse todos los documentos.

En su forma de trabajar el Motor de Cristalización es parecido al del Motor de Análisis. Como él está constituido por varios demonios, en este caso cada uno centrado en el proceso de determinados grupos de contextos de cristalización. De este modo existe un demonio por cada grafo de conocimiento para el proceso de los vértices correspondientes y otro para el proceso de los grafos de cada nodo junto con el de los nodos relacionados con el actual, si este está integrado en una red SKC. Aunque cada demonio trabaja sobre los elementos mencionados, dependiendo de los criterios de valoración empleados puede necesitar la consulta de información de otros elementos o acceder a diferentes bases de datos del sistema.

De nuevo en esta ocasión, cada uno de los demonios del Motor de Cristalización utiliza las ontologías para recorrer las instancias de las clases almacenadas en las bases de datos, siguiendo algunas de las relaciones que enlazan sus conceptos. De este modo, por ejemplo, se recorrerán los distintos grafos que organizan el conocimiento, buscando los documentos clasificados en sus vértices para ir evaluando y actualizando los valores de los atributos relacionados con el proceso, o se visitarán todos los grafos de un nodo para establecer su grado de cristalización.

Además, como el Motor de Análisis, el Motor de Cristalización dispone de una unidad de control con características parecidas y su actividad se desarrolla en segundo plano. También los demonios son configurables para adaptarse a las características de los contextos, en particular a los criterios de valoración, lo que contribuye a facilitar la evolución del sistema.



### 3.3.5 Módulo de Interacción

Como ya se ha comentado al principio de este capítulo, el Módulo de Interacción (ver Figura 3-10) de la arquitectura propuesta para el sistema SKC soporta la comunicación con otras entidades a través de una Red IP. En esta aproximación se consideran dos interfaces, aunque en el futuro podrían ser más. Una permite que los Usuarios accedan al nodo y la otra que el sistema pueda interactuar automáticamente con otros nodos como él, mediante los Agentes de Red SKC, para formar y explotar una Red de Nodos SKC. Las interfaces utilizan un submódulo que se encarga de adaptar los contenidos, la navegación y la presentación a las particularidades de los distintos interlocutores y su medio.

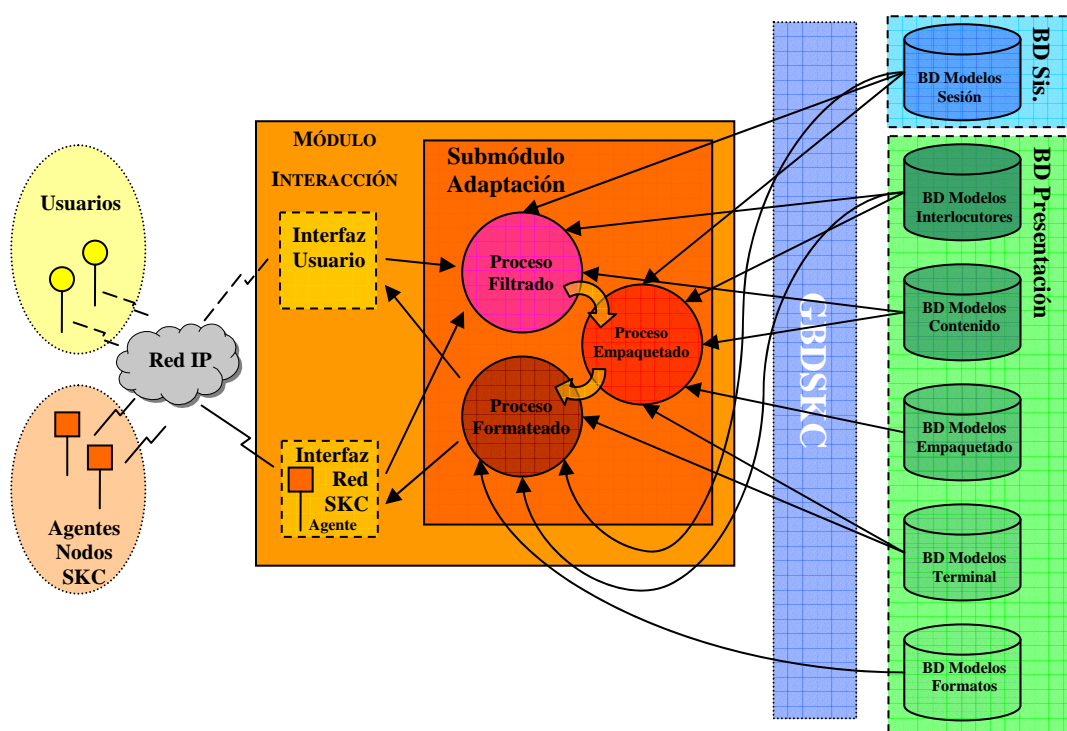


Figura 3-10. Módulo de Interactividad de SKC

La Interfaz de Usuario es la primera de las dos mencionadas, permite la interacción con el sistema a usuarios humanos a través de una interfaz Web convencional, aunque podría incorporar otras en el futuro como WAP, para teléfonos móviles, o PDA, para asistentes personales. Se encarga de recibir peticiones, procesarlas, responder a ellas y registrar toda su actividad.

La Interfaz de Red SKC es la segunda de las incluidas en el módulo. Su tarea fundamental es interactuar con otros nodos en el ámbito de la Red de Nodos SKC a la que el sistema puede estar adscrito, aunque también podría llegar a saber interactuar con otras aplicaciones en la

Web, como podrían ser algunos buscadores. En este caso la interacción con el entorno se hace a través de un Agente de Red SKC, que se comporta como cliente o servidor Web según las circunstancias.

La Red de Nodos SKC es el resultado de la interacción entre nodos, que permite el intercambio entre ellos de información y conocimiento. Los Agentes de Red SKC son los encargados de establecer, mantener y suspender las relaciones con otros nodos, así como del intercambio y salvaguarda de información y conocimiento compartido por los nodos.

El Submódulo de Adaptación es el encargado de acomodar contenidos, navegación y presentación, a las características y preferencias de los interlocutores y a las limitaciones de su medio de acceso al sistema. Para ello también se toman en consideración las propiedades de los contenidos, las restricciones del sistema y el contexto de la actual interacción. Todos estos elementos se representan mediante modelos en el entorno.

En el Módulo de Interacción tiene especial importancia el diseño y funcionamiento de la Red Nodos SKC y del Submódulo de Adaptación. En los siguientes subapartados trataremos con mayor detalle estos temas.

### **3.3.5.1 Red de Nodos SKC**

Para la integración de nodos SKC (NSKC) formando redes se propone un diseño basado en agentes [Dinverno 2001] que, cuando es posible, utiliza los recursos de búsqueda disponibles en la Web como medio de publicar la existencia de los nodos y de localizar a otros. Cualquier nodo aislado procura encontrar otros nodos en la red y asociarse con ellos formando una red SKC. El ansia mantenida de integración llevará a la incorporación de todos los nodos accesibles entre sí en redes SKC únicas por red IP, si las condiciones y recursos de ésta son los adecuados y la configuración de los nodos también. Además, un mecanismo parecido al de agregación permite la segregación de las redes SKC de los nodos que dejen de funcionar.

Todos los NSKC son autónomos, pero intentan integrarse con otros nodos formando redes SKC. Cada nodo tiene un agente de red encargado de sus relaciones con otros nodos, de establecer contacto con ellos, de intercambiar con ellos información, de conservar un registro del estatus de cada relación y de mantener un repositorio actualizado de la información obtenida de los nodos que son de interés para el NSKC que representa.

Las relaciones de cada NSKC no son iguales con todos los nodos de la red. Con la mayoría se mantiene un sencillo protocolo de reconocimiento mutuo y de intercambio de información básica sobre los nodos conocidos, que soporta la existencia de la propia red. Con algunos, sin embargo, las relaciones son más profundas: entonces además se intercambian datos relativos al

conocimiento gestionado en los nodos interlocutores, que se utilizará después para las gestiones del conocimiento de tales nodos.

En general la información intercambiada con otros nodos no es la que constituye el soporte fundamental de la base de conocimiento de cada uno -documentos, grafos o datos de usuarios- sino la meta-información empleada por el sistema que se aplica a esos elementos, es decir la obtenida del análisis de los registros del sistema y la elaborada por el sistema a partir de todas sus fuentes.

Este tipo de vínculos más profundos sólo tiene sentido que se establezca entre nodos dedicados al mismo dominio de conocimiento o áreas afines. Con ello se incorpora en el sistema el tipo de relaciones que se establecen en el mundo real entre repositorios de datos y grupos de trabajos dedicados a temas parecidos –considérense por ejemplo los grupos de investigación de distintas universidades o las diferentes publicaciones especializadas dedicadas a un tema-.

Además tales relaciones, también como en el mundo real, no son todas iguales; pueden clasificarse en varios tipos dependiendo de la consideración que la comunidad de usuarios de los nodos tengan sobre ellas. En unas ocasiones simplemente se pretende mantener un seguimiento del otro nodo, una referencia; en otras se trata de establecer una auténtica colaboración que, respetando los ámbitos delimitados por la independencia de ambos nodos, permita el trabajo independiente pero comprometido; en algunas otras se plantea una auténtica subordinación entre los nodos, puesto que uno considera al otro como referencia ineludible que no puede perderse de vista asumiendo su autoridad absoluta en la materia; o también en otras se considera al nodo interlocutor como un auténtico competidor que se pretende mantener siempre bajo vigilancia para no ser sorprendido, etc.

Al mismo tiempo, las relaciones establecidas entre los nodos SKC no tienen porqué ser simétricas, en realidad las vinculaciones entre dos NSKC tienen dos sentidos, la relación establecida entre un nodo hacia el otro y la de éste hacia el primero. Podría darse el caso de que un nodo considere a otro competidor, mientras que el segundo se considere subordinado al primero o no tenga ningún interés en él. De la gestión adecuada de esta asimetría se encargan los agentes de red implicados por parte de cada nodo.

También los agentes de red se dedican a limitar el acceso a la información suministrada a otros nodos y de no revelar a terceros los datos proporcionados por estos, al tiempo que los propios nodos se encargan de garantizar que toda la información obtenida en la red SKC se emplea sólo para los fines establecidos en el marco del tipo de relación implantada entre los nodos. Salvo en casos excepcionales la mayoría de los nodos encontrarán más beneficios en el intercambio de datos que en su reserva.

En general, todos los nodos que permiten a personas el acceso libre para revisar sus contenidos, no deberían tener ningún problema en que entidades automáticas como los agentes de red de otros nodos también lo hagan, teniendo en cuenta que los propios agentes de red de cada uno de ellos se encargaran de limitar el acceso a la información de acuerdo a condiciones predeterminadas. En este sentido, los nodos pueden restringir el suministro de datos sobre su base de conocimiento y su actividad, pero no el intercambio básico de información sobre los nodos que el sistema utiliza para el soporte de la red SKC.

Siempre que se crea un nuevo nodo se establece una página principal accesible desde la red IP donde este se sitúa. En esta página se incluye toda la meta-información necesaria para el establecimiento de comunicación entre nodos SKC. Si la ubicación no es visible desde toda Internet es preciso indicar al nuevo nodo dónde puede encontrar otros NSKC accesibles. En caso de que el nuevo nodo sea visible libremente desde Internet, además de poderse indicar explícitamente las ubicaciones de otros nodos, se procura el registro de su página principal en los buscadores más importantes. Todos los nodos SKC con acceso a esos buscadores realizarán periódicamente búsquedas específicas para intentar localizar nuevos NSKC. Cuando un nodo identifica uno nuevo, registra automáticamente su ubicación en su base de datos.

Independientemente del modo en que se haya identificado un nuevo nodo SKC, por registro explícito, a través de un buscador o mediante información de otro NSKC, el nodo descubridor intenta ponerse en contacto con el nuevo. En este primer contacto se intercambia información básica sobre los dominios de conocimientos sobre los que ambos trabajan y datos sobre todos los nodos registrados por cada lado. Por ambas partes, si se producen descubrimientos de nuevos nodos, se establece contacto con ellos. En todas las comunicaciones entre nodos se actualizan las respectivas bases de datos sobre otros nodos existentes.

Paralelamente a este proceso de identificación de nuevos nodos, cada nodo mantiene una política de segregación de los nodos que no dan señales de vida. Los agentes de red de cada NSKC establecen comunicaciones periódicas con todos nodos que tienen registrados: cuando uno no es accesible se anota la incidencia, si en varias ocasiones sucesivas se produce el mismo fenómeno se considera que el nodo en cuestión ha desaparecido. Cuando se identifica un nodo sospechoso de haber desaparecido se comunica al resto de los nodos el caso para que inicien sus procesos correspondientes.

Todos los datos intercambiados entre agentes van anotados con información temporal. Para evitar los problemas de sincronización, dicha información siempre se expresa en términos relativos referidos al momento de su intercambio. De esta forma, cuando, por ejemplo, un nodo informa a otro de que un tercero no ha respondido a sus requerimientos, indicará además que no

lo hace desde un cierto número de minutos, horas, días, etc. Cada nodo interpretará esta información respecto a su referencia de tiempo.

Para todo lo anterior, los nodos necesitan un modelo de sí mismos, de los otros nodos, de la red, de las relaciones establecidas entre ellos, de su propia base de conocimiento y comunidad de usuarios, y de ambas en los demás [Dinverno 2001]. Este entendimiento de las cosas compartido por todas las instancias del sistema, que permite el intercambio de información entre ellas, se basa en algunas de las ontologías que incluye el sistema, de las que se ha hablado en otros apartados de esta sección.

### **3.3.5.2 Submódulo de Adaptación**

El segundo tema a resaltar dentro del apartado sobre el Módulo de Interacción es el diseño y funcionamiento del Submódulo de Adaptación. La arquitectura que se propone es un diseño en tres capas inspirado en [Martins 2002], del que se habló en el capítulo de Estado de la Cuestión. El submódulo emplea una secuencia de tres procesos para realizar la adaptación, que utiliza una serie de modelos de los elementos considerados.

Las principales diferencias respecto al planteamiento de Martins están en los elementos – representados por modelos- considerados para la adaptación en las tres capas, la introducción del Modelo de Conocimiento Virtual para integrar contenidos diversos, el empleo del Modelo de Sesión para mantener memoria de condiciones no permanentes y el empleo de ontologías como soporte de todo ello. Además el diseño articula la coexistencia de varias interfaces mediante el Modelo de Interlocutor.

Los modelos que el módulo considera en el proceso de adaptación son los siguientes:

- El Modelo de Interlocutor recoge la imagen que el sistema tiene de la entidad, en general Interlocutor, que interactúa con él a través de las interfaces. Se apoya en los modelos de Usuario y Agente de Red SKC que corresponden a las dos interfaces incluidas en el sistema, si hubiera más, utilizaría los modelos adecuados de las entidades asociadas.
- El Modelo de Contenido Virtual permite, como se comentó antes, la presentación de diversos contenidos, entre los que maneja el sistema, integrados en una sola respuesta. Por ejemplo, cuando un usuario solicita un documento determinado, el sistema le devolverá el documento o parte de él, pero casi siempre junto con otros elementos que el entorno considera oportuno adjuntar, como enlaces a otros documentos o la referencia de otro usuario que en ese preciso momento está viendo el mismo documento. En realidad el sistema maneja Elementos de Contenido, como documentos, vértices o grafos, pero suele presentarlos de manera combinada en forma de Contenido Virtual.

- El Modelo de Empaquetado, establece qué parte de los Elementos de Contenido seleccionados para formar parte del Contenido Virtual se incluirán en la respuesta. Siguiendo con el ejemplo anterior, cuando en respuesta a una petición el sistema devuelve un documento a un usuario junto con enlaces a otros, mientras que del primero se muestra todo su texto, de los segundos sólo la URL para acceder a ellos, de esto se encarga el modelo que nos ocupa.
- El Modelo de Terminal representa las características del medio que el usuario está utilizando para acceder al sistema, las posibilidades gráficas de los equipos o de la conexión a la red pueden tener importancia a la hora de decidir qué parte de la información se va a presentar y cómo. Considérese por ejemplo que el sistema dispone de documento audiovisual con distintas versiones, cada una preparada para un ancho de banda determinado.
- El Modelo de Formato de Presentación, establece el modo en que se va a presentar el Contenido Virtual. En el caso de la Interfaz de Usuario normalmente será una página Web con una determinada apariencia, pero podría ser un fichero de hoja de cálculo o para imprimir, en el caso de la Interfaz de Red SKC normalmente será un documento de intercambio de datos como un XML.
- El Modelo de Sesión recoge la información relevante sobre interacciones realizadas desde que se inicio la sesión, por ejemplo puede recoger preferencias vigentes durante su duración, pero que no forman parte del Modelo de Interlocutor.

Los tres procesos incluidos en la secuencia que el Submódulo de Adaptación emplea para realizar su labor son los siguientes, presentados por orden de intervención:

- El Proceso de Filtrado determina qué Elementos de Contenido de los que maneja el sistema, como documentos, vértices o grafos, se van a incluir en el Contenido Virtual, combinación de documentos que el sistema utiliza para responder a las peticiones de contenidos de los interlocutores.
- El Proceso de Empaquetado establece qué secciones de los Elementos de Contenidos determinados por el Proceso de Filtrado deben aparecer en el Contenido Virtual, que servirá de respuesta a la demanda de contenido que inició el proceso.
- El Proceso de Formateado fija cómo se va a presentar la respuesta a solicitud emitida. Para ello establece el formato con el que se va a presentar el Contenido Virtual: una página HTML, un fichero XML, un fichero PS, etc.

Combinando todo lo anterior, a grandes rasgos, el módulo de interacción funciona de la siguiente forma cuando se solicita un contenido al sistema:

1. Cuando un interlocutor interactúa con una de las interfaces del Módulo de Interacción para hacer la solicitud, la interfaz afectada atiende la petición pidiendo al Submódulo de Adaptación que prepare el contenido solicitado.
2. La solicitud pasa en primer lugar por el Proceso de Filtrado, donde considerando el Modelo de Interlocutor, el de Contenido Virtual y el de Sesión, se determinan los Elementos de Contenido adecuados para componer el Contenido Virtual que constituirá la respuesta para el interlocutor.
3. La información conseguida, las listas de las referencias de los Elementos de Contenidos ordenados según algún criterio, se pasa al Proceso de Empaquetado, que considerando el Modelo de Interlocutor, el Modelo de Contenido Virtual, el Modelo de Empaquetado, el Modelo de Terminal y el Modelo de Sesión, determina las partes de esos Elementos que se incluirán en el Contenido Virtual respuesta a la solicitud realizada.
4. Toda la información acumulada se pasa al siguiente proceso, el de Formateo, que considerando el Modelo de Interlocutor, el de Terminal, el de Formato de Presentación y el de Sesión confeccionará el Contenido Virtual que entregará a la interfaz solicitante.
5. La interfaz presenta al interlocutor el contenido solicitado, en forma de Contenido Virtual.

Como siempre, todas las operaciones realizadas por las interfaces y el submódulo van siendo anotadas en el registro del sistema.

### **3.4 Conclusión**

En este trabajo se presenta una propuesta para manejar el problema de la saturación de información en un entorno concreto, paradigma de muchos otros sistemas de gestión de conocimiento en red. Para ello se parte de la experiencia con el sistema KnowCat (KC) y algunas ideas de diversos campos de conocimiento, Gestión del Conocimiento, Interacción Persona-Ordenador, Web Semántica, y Minería de Información y de Datos.

En particular se presenta el diseño de un nuevo sistema de gestión de conocimiento desasistido, Semantic KnowCat (SKC), que integra todo ello y presenta varias características novedosas:

- Convivencia de varias ontologías para la representación de un mismo dominio de conocimiento. Esto con una doble finalidad, facilitar la adaptabilidad dentro de cada nodo, y garantizar la interactividad entre instancias del sistema y otras aplicaciones de forma flexible y respetando las singularidades.
- Modelado de las entidades del sistema, registro de su actividad y análisis del mismo. En el sistema se consideran muchas entidades que interactúan entre sí: documentos,

estructuras de conocimiento, usuarios, grupos de usuarios, comunidades, nodos, redes de nodos, relaciones entre nodos de redes, documentos virtuales, etc. Pero se podrían considerar más, tales como partes de un documento o de un documento virtual, o colecciones de documentos. De todas ellas es preciso registrar información y procesarla para extraer conclusiones con las que alimentar el propio sistema.

- Interacción automática entre entidades del sistema, entre nodos de una red, entre documentos de un nodo, entre usuarios y documentos, entre estructuras de conocimiento, entre documentos y dichas estructuras, etc. Anotación automática de todos los elementos.
- Adaptación del acceso al conocimiento y la presentación del mismo, integrando toda la información disponible en el sistema proveniente del material ingresado en cada nodo, de la interacción con el entorno explícita e implícita de la comunidad de usuario, otros nodos y sistemas, y del análisis de materiales y registros de actividad.

La aproximación presentada aporta varias innovaciones en su ámbito de investigación. Por un lado, la integración de técnicas e ideas de áreas diversas para buscar soluciones al problema de la sobrecarga de información, que se venían tratando desde distintos campos de investigación de forma independiente. Por otro lado, la visión de la Web desde algunas perspectivas como un gran sistema de gestión de conocimiento, proporciona nuevas oportunidades de aproximación a algunos problemas de la Web y de obtener provecho de ella. Además, SKC proporciona a los campos de investigación implicados un entorno de prueba de especial interés, por las características que integra: medio acotado de gestión de conocimiento sobre la Web, interlocutores humanos y automáticos con objetivos claros y concurrentes, y una utilidad práctica evidente.



# 4 Aproximación Experimental

En el capítulo anterior se ha establecido un marco conceptual a partir de las hipótesis propuestas al principio del mismo. Por un lado, se ha elegido un entorno asequible con las características esenciales de ámbitos más generales donde se produce la sobrecarga de información, pero más reducido y controlable, que es el contexto de los sistemas colaborativos de gestión de conocimiento en red. Por otro lado, se ha propuesto un sistema hipotético que responde a los requerimientos fundamentales de la aproximación propuesta: facilita la colaboración de usuarios humanos y artificiales sobre un repositorio de información, procurando aprovechar las características de todos los elementos implicados y su actividad normal con el sistema, sin que los usuarios tengan que realizar acciones nuevas significativas, para intentar mejorar la consecución del objetivo del sistema, que es gestionar el conocimiento.

En este capítulo se pretende comprobar los supuestos principales planteados en el capítulo anterior, realizando experimentos factibles sobre el entorno delimitado con el sistema propuesto. Para ello se fijan unos objetivos experimentales, se prepara un prototipo del sistema diseñado con las funcionalidades requeridas, se establece el campo experimental necesario, y se realizan las experiencias y un primer análisis de sus resultados. En el siguiente capítulo se completa dicho análisis, se extraen las correspondientes conclusiones, y se establecen las líneas de investigación y los trabajos futuros a la vista de tales conclusiones.

#### **4.1 Objetivos Experimentales**

Como se adelantó en la introducción de este trabajo, el objetivo general de la tesis es probar que se pueden encontrar vías de solución para el problema de la sobrecarga de información en el contexto de los sistemas de gestión de conocimiento en red, entorno representativo de otros más generales donde se produce dicho problema, aprovechando la energía de la acción de las entidades activas de estos sistemas y las características de todos los elementos implicados en ellos. Para eso se proponen cuatro objetivos operativos, dirigidos a indagar en algunos de los aspectos esenciales de la actividad de dichos sistemas [Benjamins 1999]: la recolección de conocimiento; la organización y estructuración del mismo; su refinamiento y mantenimiento; y la distribución del conocimiento entre destinatarios necesitados de él.

Dichas metas operativas constituyen los objetivos experimentales perseguidos en este capítulo. El primero de ellos es seleccionar el contenido por su calidad, minimizando los requerimientos de manifestaciones explícitas de los usuarios sobre él. Con ello se pretende reducir la necesidad de que los usuarios dediquen atención a la selección del conocimiento. El segundo objetivo experimental es hacer evidente conocimiento que está latente en el repositorio, para enriquecer con él el conocimiento ya explícito y mejorar así la gestión del conjunto. Se trata de analizar los elementos de conocimiento existentes, para descubrir información oculta que pueda resultar de ayuda para la actividad del sistema. El tercero de los objetivos experimentales establecido es conseguir nuevo conocimiento de la actividad de los usuarios con el sistema, para incorporarlo también en el repositorio del sistema y aprovecharlo en la gestión del conocimiento que se lleva a cabo en el mismo. En este caso se trata de analizar la actividad de los usuarios, para evidenciar información útil que mejore el desempeño del sistema. Por último, el cuarto de los objetivos experimentales fijado es establecer redes de conocimiento entre los nodos del sistema de forma automática y autónoma a través de la Web. Con ello se pretende abrir los nodos del sistema y proyectarlos más allá de sus límites, implantando vinculaciones entre sus contenidos y entre sus comunidades de usuarios.

## **4.2 Prototipo de SKC sobre KC**

Como ya se ha dicho varias veces, el sistema Semantic KnowCat (SKC) descrito en el capítulo anterior emplea la aproximación a la gestión colaborativa del conocimiento sin supervisión del sistema KnowCat (KC) [Alamán 1999][Cobos 2003], integrando ideas y técnicas de las diversas áreas de investigación que se revisaron en el capítulo de Estado de la Cuestión.

En el apartado que se ocupa de la Arquitectura del SKC, también en el capítulo anterior, se explicó que aunque el sistema sigue el enfoque de gestión de conocimiento de KC, la estrategia que emplea y su diseño son bastante diferentes. Por un lado, SKC pretende realizar la gestión del conocimiento aprovechando más las características de los elementos que concurren en el sistema, y reduciendo los requerimientos de las manifestaciones explícitas de opinión de los usuarios. Por otro lado, la arquitectura de SKC se construye alrededor de ontologías, que se refieren tanto al conocimiento que el sistema gestiona, como a las entidades que intervienen en su proceso. De este modo, los árboles de conocimiento de KC deben tratarse como ontologías en SKC, y es necesario dotar al nuevo sistema de capacidades sensoriales, analíticas y expresivas superiores a las de KC. Estas diferencias no han impedido que KC se emplee como plataforma de prototipado para SKC.

Con ello se ha conseguido disponer de un modelo funcional de SKC dotado de la funcionalidad necesaria para alcanzar los objetivos experimentales propuestos, utilizando el sistema KC como núcleo operativo, al que se han añadido los módulos imprescindibles para conferirle las nuevas funcionalidades, pero respetando el funcionamiento básico del sistema anfitrión. Esta estrategia de prototipado ha permitido emplear el modelo de SKC en experimentos realizados en el marco de programas de actividades de KC y utilizar repositorios de conocimiento del KC original en ensayos con el prototipo, además de posibilitar el desarrollo y la prueba de las funcionalidades pretendidas de SKC con los recursos disponibles.

En esta sección se presenta cómo se han prototipado tres componentes de SKC sobre KC utilizando las técnicas e ideas de los campos de conocimiento considerados. En primer lugar un Monitor de Cliente (MC), que se encarga de registrar la interacción de cada usuario con el sistema de forma más detallada que los registros convencionales de los sistemas Web. En segundo lugar, un Módulo de Análisis de Conocimiento (MAC), que se ocupa de procesar el conocimiento y los datos recogidos por el sistema, para obtener información oculta y valiosa con la que enriquecer el conocimiento del repositorio o facilitar su gestión. Finalmente, un Módulo de Red (MR), que permite establecer vínculos automáticos entre distintas instancias del sistema repartidas por la Web, con el fin de implantar y mantener una red de conocimiento entre todas ellas. Además, se ha dotado al sistema de la infraestructura de base de datos necesaria para soportar estas nuevas funcionalidades.

### **4.2.1 Monitor de Cliente**

Este apartado trata de la integración de algunas de ideas y técnicas presentadas en el capítulo de Estado de la Cuestión para el desarrollo del Monitor de Cliente (MC) del sistema SKC, que como se ha visto es una aplicación Web cliente/servidor. Dicho monitor es un componente singular de Módulo de Análisis de SKC. El MC se encarga de registrar la actividad del usuario en el lado del cliente cuando se trabaja con el sistema, de hacer llegar los datos obtenidos hasta el servidor y de procesarlos allí. En el lado del cliente, el MC se dedica a tomar nota de la actividad del usuario mientras trabaja con el sistema, entre interacciones sucesivas con el servidor, al tiempo que se ocupa de informar al servidor de ello. En el lado del servidor, el MC se encarga de registrar los datos de actividad enviados, así como de ordenarlos y analizarlos con el fin de obtener nueva información que resulte aprovechable para la gestión del conocimiento. Todo ello de la forma más transparente posible para el usuario. Una primera aplicación del MC es obtener un indicador del grado de actividad de los usuarios con los elementos que constituyen el conocimiento gestionado por el sistema -documentos, temas, nodos y autores -, coeficiente que parece representativo del Índice de Interés de la comunidad por cada uno de dichos elementos.

Como es sabido, en los sistemas cliente/servidor estrictos, como es el caso de los sistemas basados en servidores Web, sólo queda constancia de las solicitudes de servicio recibidas por el servidor. Es por esto que, habitualmente, nada queda registrado sobre la actividad del usuario en el cliente entre solicitudes sucesivas de servicio al servidor. Los clientes Web convencionales tampoco toman nota de la actividad del usuario con los contenidos que se manejan a través de ellos. Para llenar este vacío, la aplicación del cliente tendría que encargarse de registrar la actividad de los usuarios en su lado y transferir periódicamente los datos obtenidos al servidor. Esto es precisamente lo que hace la parte del cliente del MC de SKC. Como resultado, en el servidor quedarán almacenados los datos de la actividad de los usuarios en el cliente y en el servidor.

Para la implementación del MC se emplean ideas y técnicas especialmente de los campos de conocimiento de Interacción Persona-Ordenador y de Minería de Información y Datos. Del primero de ellos, se utiliza la referencia de los sistemas de monitorización, como el control de eventos de la interfaz de usuario, en particular los del objeto DOM de los clientes Web, o la aplicación de técnicas de integración de aplicaciones de cliente en páginas HTML, típicas de la Web 2.0. Del segundo de los campos mencionados, tienen especial importancia para el desarrollo del módulo realizado las técnicas de Minería de Uso Web, que se han empleado en el análisis de los registros LOG del servidor Web del sistema, y las ideas de Minería de Estructura Web, aplicadas al análisis de la estructura interna de la interfaz del sistema, que también se emplea para el proceso del registro Web.

El tipo de información obtenida por un módulo similar al MC de SKC podría ser de interés en aplicaciones de campos diversos, como e-learning, marketing, groupware o gestión de información y conocimiento, donde se pretenda monitorizar, predecir, recomendar o adaptar, entre otros objetivos. Naturalmente, no es una novedad interesarse por la actividad de los usuarios con el sistema ni intentar aprovechar el registro de la misma para algún fin. Sin embargo, MC propone un modo asequible de conseguir esta información sobre un navegador convencional sin grandes modificaciones en la aplicación Web en el lado del cliente y en el servidor. Con todo, donde realmente propone algo nuevo es en el modo de aprovechar la información obtenida para seleccionar el conocimiento sin necesidad de que los usuarios manifiesten de forma explícita su opinión sobre el contenido, como se hacía por ejemplo en el mencionado sistema KC.

#### 4.2.1.1 Monitor de Cliente de SKC

El Monitor de Cliente (MC) del sistema Semantic KnowCat (SKC) está ideado con el propósito de obtener información de la actividad del usuario con el sistema en el lado del cliente, para utilizarla en la selección del conocimiento. El MC sólo monitoriza la actividad del usuario cuando interactúa con la interfaz Web del sistema a través de los navegadores Explorer y Mozilla.

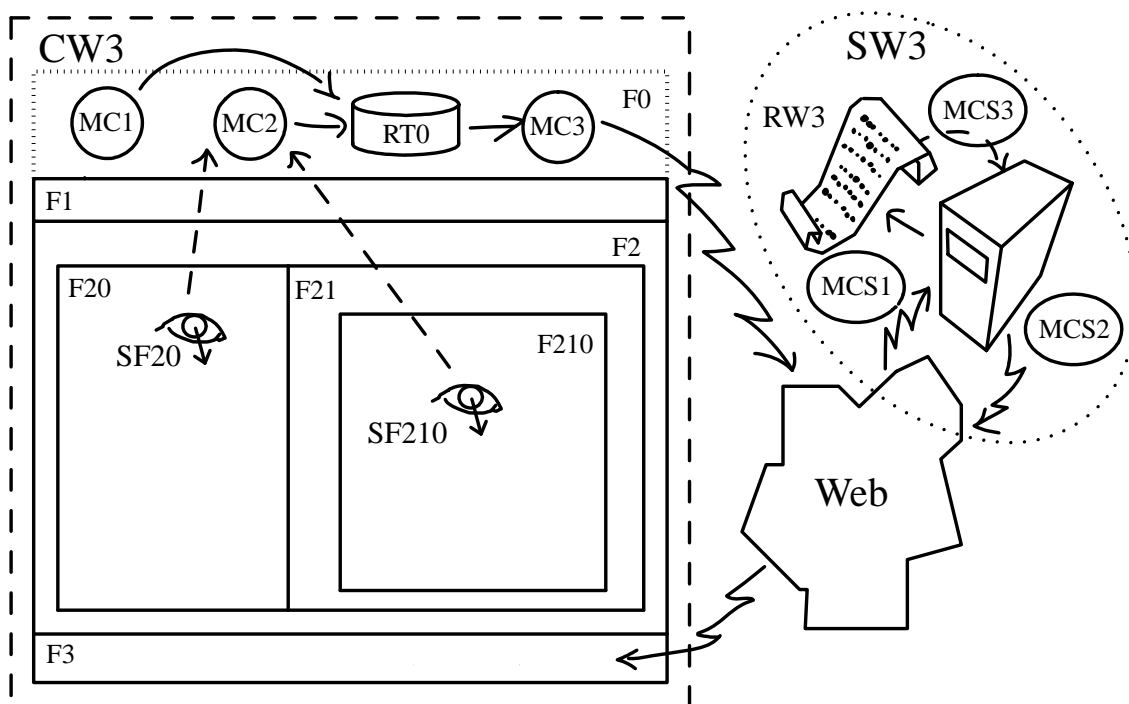


Figura 4-1. MC en el cliente (CW3) y el servidor (SW3)

El MC monitoriza la actividad del usuario a bajo nivel, atendiendo a ciertos eventos que se consideran indicadores de actividad, como la obtención de foco en la aplicación, los desplazamientos y las pulsaciones de ratón, los movimientos de la barra de desplazamiento o las

pulsaciones de teclado. Mientras se mantiene la actividad sobre la interfaz en el cliente, el monitor registra el número de ocurrencias de cada uno de esos eventos, pero deja de hacerlo cuando la actividad desaparece o se traslada a otras aplicaciones ajenas a SKC. Cada cierto tiempo, si ha habido actividad en el último intervalo, el MC notifica al servidor los datos registrados y pone a cero los contadores. En el servidor, el monitor toma nota de los datos enviados cada vez que llegan y los analiza cuando se le solicita.

El MC establece sesiones de cliente, que determinan una secuencia continuada de actividad con interrupciones inferiores a un límite establecido por configuración del sistema. Su planteamiento es similar al de las sesiones de servidor que se utilizan en algunas aplicaciones Web, pero considerando la actividad en el cliente y no en el servidor. El monitor consta de dos partes bien diferenciadas, una incrustada en la interfaz del sistema en el cliente, y la otra en el servidor del sistema, ésta última repartida entre las aplicaciones de administración del mismo y el servidor Web que lo soporta.

#### ***4.2.1.2 El MC en el Lado del Cliente***

En el lado del cliente el sistema consta de una aplicación de cliente (CW3 en la Figura 4-1) que trabaja sobre un navegador Web, Explorer o Mozilla. Dicha aplicación utiliza una interfaz basada en marcos (F0, F1... en la Figura 4-1) construida de tal forma que el código del marco base se carga cuando se inicia una sesión de interacción con el sistema y no se vuelve a cargar mientras dure la misma. Con ello se consigue dotar de memoria a la aplicación y proporcionar la lógica de monitorización necesaria para toda su interfaz.

En particular, el marco base carga una librería JavaScript que incluye variables y funciones (RT0, MC1, MC2.. en la Figura 4-1). Las variables permiten contabilizar eventos y mantener el control del proceso de monitorización. Las funciones se encargan de todas las tareas necesarias para la monitorización: identificar al usuario, identificar la sesión de interacción en el cliente de forma unívoca, establecer sensores (SF20 y SF210 en la Figura 4-1) en marcos y ventanas, implantar ciclos de monitorización, registrar localmente los eventos producidos durante dichos ciclos e informar al servidor de la actividad registrada al final de cada uno de ellos.

Sobre los espacios establecidos por el marco base, se cargan los distintos marcos que van a ir cambiando y constituyen la interfaz del sistema en el cliente. Cada vez que uno de estos marcos cambia, se inicializan los sensores necesarios y se establece el estado adecuado para continuar la monitorización a partir de la situación previa.

Los eventos que se consideran son cinco: movimiento del ratón, presión en los botones del mismo, activación de la barra de desplazamiento, cambio del foco sobre los elementos de la

interfaz y presión en las teclas del teclado. Los sucesos de cada uno de ellos se van contabilizando en variables independientes cada vez que los sensores detectan su aparición.

Los ciclos de monitorización establecen los momentos de informar al servidor sobre lo sucedido en el cliente en el último periodo. Para ello se utilizan temporizadores que controlan el envío de información mediante la recarga de un marco especial de la interfaz, que además muestra la actividad del entorno en ese sentido. Sin embargo, las notificaciones no se producen si en el ciclo no se ha producido actividad. La duración de los ciclos es configurable y debe elegirse de manera que informe al servidor frecuentemente, sin llegar a hacerse molesto y a sobrecargar el equipo y la línea. En realidad, los datos enviados y las respuestas correspondientes del servidor son muy ligeros, aunque la interacción proactiva de las aplicaciones Web no es habitual y puede disgustar a los usuarios, tanto como la observación de su actividad.

En este sentido, el objetivo y el comportamiento de la aplicación puede hacerla sospechosa de violar la privacidad de los usuarios, aunque en su implementación se hayan respetado las restricciones establecidas sobre el tema en los navegadores y en el lenguaje de programación. Estas consideraciones son importantes en la aplicación de estas técnicas, pero no reducen el interés de indagar en sus posibilidades. En particular, el uso de eventos y de accesos a páginas no solicitadas ha sido objeto de restricciones sucesivas en los últimos tiempos.

La posibilidad de mantener una memoria en el cliente es la base de la monitorización. Sin ella no sería posible seguir el proceso, acumular la cuenta de los eventos ni dosificar el envío de información al servidor. La aproximación utilizada no es la única posible, por ejemplo podrían haberse utilizado “cookies” para guardar datos y otros lenguajes de programación para implementar la funcionalidad. Sin embargo, el uso de JavaScript con marcos resuelve todos los requerimientos del problema, empleando elementos comunes y aceptados generalmente en las aplicaciones Web comerciales.

#### ***4.2.1.3 El MC en el Lado del Servidor***

En el lado del servidor (SW3 en la Figura 4-1) se realizan tres funciones fundamentales del monitor: se atiende a los informes de actividad de los ciclos de monitorización en el cliente, se registra la actividad correspondiente y se proporciona el soporte para el análisis de los datos almacenados.

La información de actividad de los ciclos se recibe mediante una llamada a un CGI (Common Gateway Interface) en el servidor Web (MCS1 en la Figura 4-1). El programa simplemente responde redireccionado la llamada a la dirección encargada de componer la respuesta adecuada (MCS2 en la Figura 4-1) que se dirige a un marco especial en el cliente. Aunque el mencionado CGI no ha hecho aparentemente nada, en realidad ha dejado una línea en el fichero de registro

del servidor Web (RW3 en la Figura 4-1) con la información recibida a través de la URL (Uniform Resource Locator) que le invocaba.

El fichero de actividad del servidor Web, su fichero de LOG, es el lugar elegido como repositorio de la actividad observada en el cliente. Esta elección no es caprichosa, puesto que en este archivo es donde se registra la actividad del sistema observada en el servidor Web de forma estándar. De este modo, se ha transformado la actividad observada en el cliente en parte de la actividad convencional apreciada en el servidor. Esta circunstancia facilita el tratamiento combinado y la aplicación de técnicas convencionales de procesamiento de este tipo de ficheros [Garofalakis 1999].

En la Figura 4-2 se muestra un par de líneas de un fichero típico de LOG de un servidor Web, que se presentan partidas para adaptarse al formato de este documento, donde el texto correspondiente a las líneas pares aparece sombreado para facilitar su identificación. Ambas líneas comienzan con la dirección IP de Internet desde donde se realizó la petición y contienen una indicación del momento del registro, entre corchetes, y los datos del recurso solicitado, entre comillas.

```
150.244.57.62 - - [01/Mar/2006:10:33:26 +0100] "GET /5.1a/descriptions.pl? b=KC_BDD763&t=24
HTTP/1.1" 200 1488
150.244.57.62 - - [01/Mar/2006:10:33:26 +0100] "GET /5.1a/control.pl? b=KC_BDD763&t=24&nav_space=1
HTTP/1.1" 200 10288.
```

Figura 4-2. Líneas típicas del LOG de un servidor Web

La Figura 4-3 muestra una línea del LOG de un registro de datos de actividad del MC, que se ha adaptado al formato del artículo como en el caso anterior. Como puede apreciarse, se trata de una línea de registro como las demás, pero se refiere a un recurso característico “infoSituacion.pl” e incluye datos codificados en la URL correspondiente, como el identificador de usuario que lo ha generado, precedido de “UsrID”, los momentos de inicio y fin del ciclo de monitorización “CntIni “ y “NtfTmp” en tiempo del cliente, o los números de eventos de bajo nivel registrados en el periodo para cada tipo. Los eventos considerados son: “MseM”, movimientos de ratón; “MseD”, pulsaciones de ratón; “Scr”, movimientos de scroll; “Fcs”, ganancias de foco; “Blr”, pérdidas de foco; y “KeyD”, pulsaciones de teclas.

```
150.244.57.62 - - [01/Mar/2006:10:33:29 +0100] "GET /5.1a/informSituacion.pl?
urlrsp=.../5.1a/monitorKC.pl%3Fest%3D1&UsrID=Monitor%20de%20BDD&CntIni=1141205148142&NtfTmp=114120552
9249&MseM=414&MseD=11&Scr=0&Fcs=19&Blr=19& KeyD=0 HTTP/1.1" 302 296
```

Figura 4-3. Línea del MC en LOG de un servidor Web

En la práctica, los nuevos registros de actividad del cliente (RAC) quedan intercalados con los habituales de actividad en el servidor Web (RAS). Unos y otros recogen datos comunes, como la dirección IP, el momento de anotación y la URL de llamada que los originó con todos sus parámetros. En el caso de los RAC, los parámetros de dicha URL incluyen los contadores de los distintos eventos, el identificador unívoco de la sesión en el cliente y las referencias de tiempo utilizadas en el mismo.



La Figura 4-4 muestra una sección de fichero de LOG, donde las líneas pares se han sombreado como en el caso anterior. En el fragmento reproducido pueden apreciarse las líneas RAC debidas al MC, que aparecen subrayadas e intercaladas entre las RAS generadas por el resto de solicitudes realizadas al sistema.

```

...
150.244.57.62 - - [01/Mar/2006:10:33:21 +0100] "GET /5.1a/descriptions.pl?b=KC_BDD763&t=25 HTTP/1.1"
200 1486
150.244.57.62 - - [01/Mar/2006:10:33:21 +0100] "GET /5.1a/control.pl?b=KC_BDD763&t=25&nav_space=1
HTTP/1.1" 200 10288
150.244.57.62 - - [01/Mar/2006:10:33:24 +0100] "GET /5.1a/descriptions.pl?b=KC_BDD763&t=48 HTTP/1.1"
200 1546
150.244.57.62 - - [01/Mar/2006:10:33:25 +0100] "GET /5.1a/control.pl? b=KC_BDD763&t=48&nav_space=1
HTTP/1.1" 200 10288
150.244.57.62 - - [01/Mar/2006:10:33:25 +0100] "GET /5.1a/descriptions.pl?b=KC_BDD763&t=25 HTTP/1.1"
200 1486
150.244.57.62 - - [01/Mar/2006:10:33:26 +0100] "GET /5.1a/control.pl?b=KC_BDD763&t=25&nav_space=1
HTTP/1.1" 200 10288
150.244.57.62 - - [01/Mar/2006:10:33:26 +0100] "GET /5.1a/descriptions.pl?b=KC_BDD763&t=24 HTTP/1.1"
200 1488
150.244.57.62 - - [01/Mar/2006:10:33:26 +0100] "GET /5.1a/control.pl? b=KC_BDD763&t=24&nav_space=1
HTTP/1.1" 200 10288
217.127.197.187 - - [01/Mar/2006:10:33:28 +0100] "GET /SI2/info.html HTTP/1.1" 200 14054
217.127.197.187 - - [01/Mar/2006:10:33:28 +0100] "GET /SI2/info_archivos/interrogante.jpg HTTP/1.1"
304 -
217.127.197.187 - - [01/Mar/2006:10:33:28 +0100] "GET /SI2/info_archivos/kc.gif HTTP/1.1" 304 -
217.127.197.187 - - [01/Mar/2006:10:33:28 +0100] "GET /SI2/info_archivos/ExpKcSi2_0506.gif HTTP/1.1"
304 -
217.127.197.187 - - [01/Mar/2006:10:33:28 +0100] "GET /SI2/info_archivos/image001.gif HTTP/1.1" 304
-
217.127.197.187 - - [01/Mar/2006:10:33:29 +0100] "GET /img/img_toolbar/nav_info.gif HTTP/1.1" 304 -
150.244.57.62 - - [01/Mar/2006:10:33:29 +0100] "GET /5.1a/informSituacion.pl?
urlrsp=../5.1a/monitorKC.pl%3Fest%3D1&UsrID=Monitor%20de%20BDD&CntIni=1141205148142&NtfTmp=114120552
9249&MseM=414&MseD=11&Scr=0&Fcs=19&Blr=19&KeyD=0 HTTP/1.1" 302 296
150.244.57.62 - - [01/Mar/2006:10:33:29 +0100] "GET /5.1a/monitorKC.pl?est=1 HTTP/1.1" 200 318
150.244.57.62 - - [01/Mar/2006:10:33:31 +0100] "GET /5.1a/descriptions.pl?b=KC_BDD763&t=23 HTTP/1.1"
200 2918
150.244.57.62 - - [01/Mar/2006:10:33:31 +0100] "GET /5.1a/control.pl?b=KC_BDD763&t=23&nav_space=1
HTTP/1.1" 200 10288
150.244.57.62 - - [01/Mar/2006:10:33:32 +0100] "GET /5.1a/descriptions.pl?b=KC_BDD763&t=81 HTTP/1.1"
200 1493
217.127.197.187 - - [01/Mar/2006:10:33:32 +0100] "GET /5.1a/informSituacion.pl?
urlrsp=../5.1a/monitorKC.pl%3Fest%3D1&UsrID=antonio%20martinez&CntIni=1141203026247&NtfTmp=114120553
8460&MseM=50&MseD=0&Scr=0&Fcs=1&Blr=1&KeyD=0 HTTP/1.1" 302 296
150.244.57.62 - - [01/Mar/2006:10:33:32 +0100] "GET /5.1a/control.pl?b=KC_BDD763&t=81&nav_space=1
HTTP/1.1" 200 10288
217.127.197.187 - - [01/Mar/2006:10:33:33 +0100] "GET /5.1a/monitorKC.pl?est=1 HTTP/1.1" 200 318
...

```

Figura 4-4. Fragmento de LOG de un servidor Web de un nodo SKC con MC

Entre el primer RAC de una sesión de cliente cualquiera y el último registro de esa sesión en el fichero de LOG del servidor Web, se encuentran la mayoría de los RAS correspondientes a dicha sesión de cliente. Determinar cuáles son resultaría muy interesante, porque permitiría relacionar la actividad en el cliente con los contenidos servidos por el sistema, que es el objetivo final de todo este mecanismo.

Encontramos algunas dificultades para realizar esta labor. Por un lado, los RAC se registran al final de cada ciclo, es decir hacen referencia a los RAS que les preceden en el LOG. Esto tiene varios inconvenientes derivados de la necesidad de considerar registros anteriores cuando se analiza el fichero secuencialmente en aparente orden cronológico. Por otro lado, ambos tipos de registros, RAC y RAS, comparten información que no siempre es suficiente para asociarlos entre sí de forma inequívoca. Por ejemplo, dos registros consecutivos del mismo momento y dirección IP pueden corresponder a las máquinas de usuarios distintas conectadas a la red mediante el mismo PROXY (servidor intermediario con Internet). Además, el servidor Web que

soporta el prototipo de SKC puede emplear un único fichero LOG para varios nodos que residen en el mismo sistema. Esto sucede en el ejemplo de la Figura 4-4, donde puede apreciarse que hay mezcladas líneas correspondientes a dos usuarios ubicados en equipos con direcciones IP diferentes accediendo cada uno a un nodo distinto.

Algunos de estos problemas pueden subsanarse complicando un poco los programas de análisis. Otros, considerando peculiaridades de las aplicaciones con las que se está trabajando, como puede ser el caso de sistemas que mantienen sesiones de servidor (seguimiento de las secuencias continuas de interacción de los clientes) o de sistemas que utilizan algún tipo de identificador que se incluye en todas sus URLs. Por último, hay ocasiones en que, simplemente, se puede obviar el asunto, por la poca incidencia que una asignación de actividad equivocada pudiera tener al principio o final de la sesión de cliente o entre usuarios concurrentes utilizando el mismo PROXY.

Todas estas cuestiones se manejan en la última de las funciones, la de análisis, y son tratadas por el MC en el lado del servidor (MCS3 en la Figura 4-1), tal como se enumera al principio de este apartado. A diferencia de las otras dos actividades, la atención a los informes de actividad y el registro de los mismos (que se van realizando como parte del servicio), la actividad de análisis se ejecuta bajo demanda en determinados momentos, como es habitual en el procesamiento de los ficheros LOG de los servidores Web.

Para realizar el análisis, el valor asignado a los sucesos de los distintos eventos considerados en el MC (de ratón, barras, foco y teclas), no debe ser el mismo, puesto que unos son muy frecuentes y pueden ser accidentales, como los movimientos de ratón, y otros infrecuentes y casi siempre voluntarios, como el movimiento de la barra de desplazamiento o la pulsación de botones y teclas.

En nuestro caso, la consideración de los eventos se ha establecido de forma empírica, comprobando la proporción de los distintos sucesos en actividades de distinta intensidad sobre documentos a través del sistema. En particular, se han considerado tres niveles de actividad para cada ciclo: alto (10), cuando se ha producido algún evento de barra de desplazamiento, pulsación de botón o tecla, o un número superior a treinta de movimientos de ratón; moderado (5), cuando se ha ganado el foco más veces que las que se ha perdido o se han producido entre quince y treinta eventos de movimiento de ratón, pero ningún evento de otros tipos; y nula o despreciable (0), en cualquier otro caso.

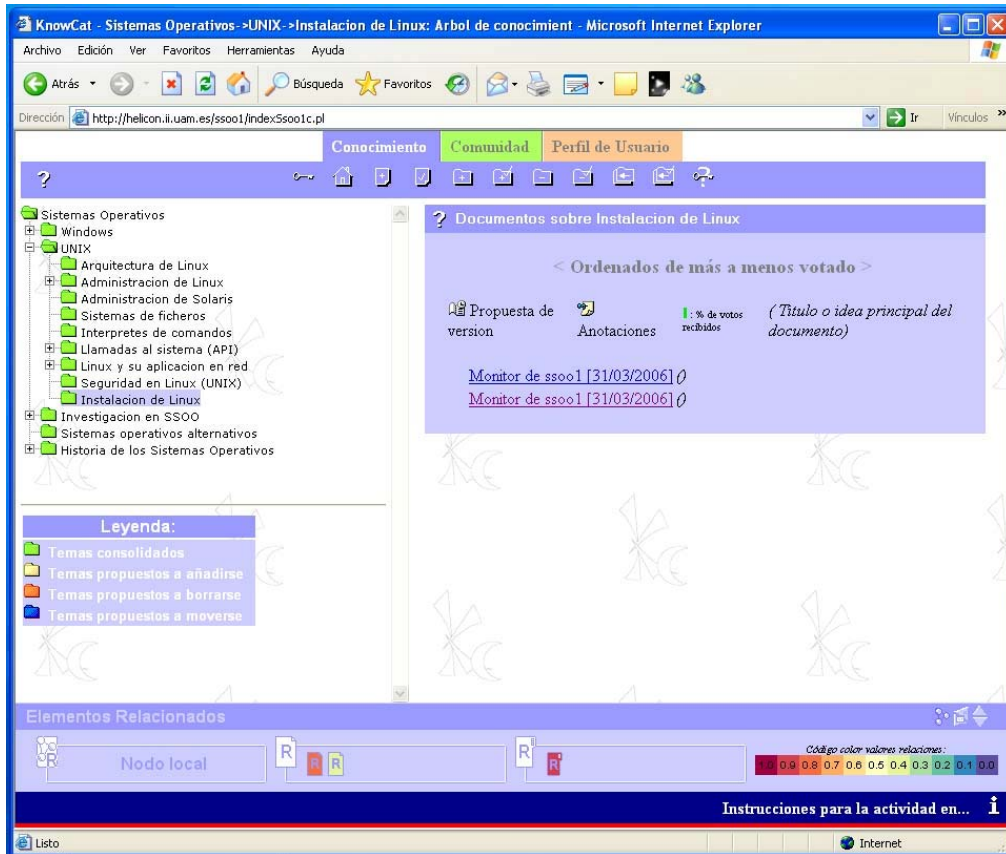
Aunque en este apartado sólo nos centramos en los documentos, el análisis de la actividad en el cliente es aplicable a otros tipos de elementos que forman parte del sistema SKC. En concreto: a los documentos, unidades de conocimiento; a los temas, agrupaciones de documentos; a los autores, creadores de documentos; y a los nodos, instancia del sistema, que agrupan a

documentos, temas y autores. En base a la intensidad de la interacción establecida sobre cada tipo de elemento se pueden establecer comparaciones entre ellos.

Para establecer los tipos de elementos sobre los que recae la actividad en el cliente en cada momento, es necesario analizar la organización de interfaz del sistema. En este caso particular, las páginas (o el CGI invocado desde cada marco), determinan sobre qué elemento particular recae la acción. El nodo siempre debe sumar actividad si se observan eventos en cualquier elemento que forma parte de él. Los temas deben sumar puntos cuando se trabaja con su estructura de datos o alguno de los documentos que incluyen. Los documentos puntúan cuando un marco que los muestra registra actividad o cuando lo hacen los marcos que presentan información relacionada con ellos. Con los autores se hace lo mismo. Siguiendo este criterio se ha establecido una clasificación de las páginas Web y los CGIs que presentan información relativa a cada tipo de elemento. Además, se ha determinado la incompatibilidad de páginas y CGIs que comparten marcos, y por lo tanto no pueden convivir en la interfaz simultáneamente.

Por ejemplo, en la Figura 4-5 se muestran dos momentos de la interacción con un nodo del prototipo de SKC. En la imagen (a), el sistema presenta a la derecha los documentos incluidos en el tema seleccionado en el árbol de conocimiento de la izquierda. En la imagen (b) el sistema muestra uno de los documentos enumerados en la imagen anterior. En ambos casos la actividad sobre la interfaz se refiere al mismo tema, puesto que el contenido de una pantalla es el general del tema y el de la otra el de uno de sus documentos. Sin embargo, la página (a) atañe concretamente al documento que presenta y a su autor, lo que no ocurre en la (b). Al acceder a la información del tema en la imagen (a), se invocan unos CGIs distintos que al hacerlo al contenido de la (b), a pesar de que en ambas ocasiones se comparte el mismo espacio central de la interfaz. Los registros correspondientes en el fichero LOG del servidor Web hacen referencia a unos CGIs u otros según el caso, de forma que puede establecerse cuándo la actividad observada debe recaer sólo sobre el tema implicado o también sobre cierto documento y su correspondiente autor. Además, las llamadas a los mismos CGIs son diferentes según los contenidos a los que se refieran, puesto que en ellas se especifica la información particular, que siempre incluye el identificador de nodo y según el caso el de tema, el de documento o el de autor asociados. De este modo, cuando se visualiza en la interfaz un documento distinto del mismo tema, la llamada registrada en el LOG puede distinguirse de la de otro documento.

(a)



(b)

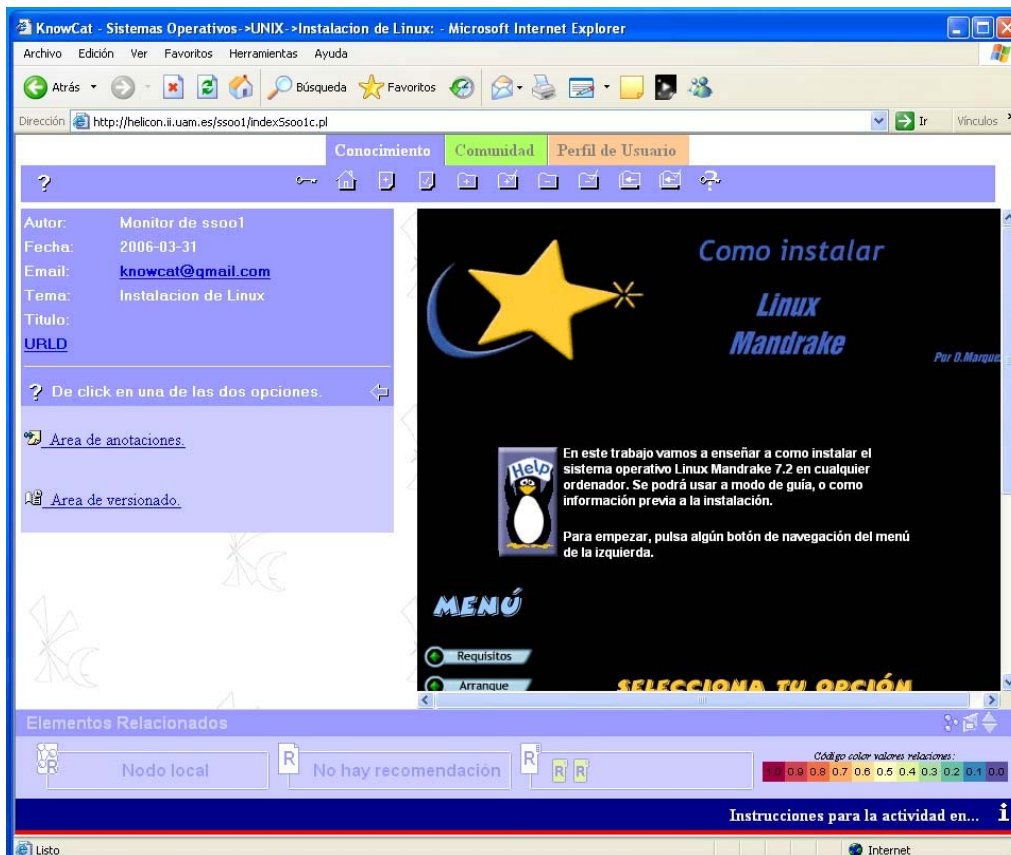


Figura 4-5. Ejemplo análisis de interfaz para establecer elementos objeto de la actividad

Por otro lado, las salidas de los CGI están siempre asociadas a determinadas zonas de la interfaz en el sistema, en el ejemplo anterior al área central. De esta forma, cuando en el registro aparece la referencia a unos CGI que utilizan la misma zona de la interfaz que otros registrados inmediatamente anteriores, significa que se ha producido un cambio en el contexto de la aplicación y que hay que revisar los objetos de la actividad que se observe a partir de ese momento. Por ejemplo, cuando después de un acceso a un documento se vuelve a la página general del correspondiente tema, la actividad debe dejarse de asignar a dicho documento y a su autor, porque los CGI correspondientes a la nueva página impiden que la interfaz muestre la anterior en la zona central de la misma. Este estudio de la interfaz y de su relación con los CGI involucrados debe hacerse para los distintos tipos de pantallas que constituyen el entorno de la aplicación.

Además, no todos los registros correspondientes a un nodo que aparecen en el LOG del servidor Web son relevantes para el análisis. Por ejemplo, las peticiones de las imágenes de los iconos de la interfaz, también las de los documentos físicos que se presentan en las páginas generadas mediante los CGI considerados, o las de páginas y recursos asociados a elementos genéricos del sistema como son los asociados a las páginas de ayuda.

Tomando en cuenta todo lo dicho, durante el análisis del fichero LOG se van estableciendo los elementos de cada categoría que deben contabilizar la actividad que se produzca sobre el sistema. Dichos elementos se van sustituyendo a medida que aparecen referencias de otros elementos excluyentes en el registro de actividad. Los elementos particulares, como un documento concreto, se reconocen por los identificadores que incluyen las URLs que los referencian.

El análisis de la actividad de los usuarios en el cliente, con los elementos de conocimiento que gestiona SKC mediante el fichero LOG del servidor Web con los registros del MC, requiere tres pasos (ver Figura 4-6). Los dos primeros constituyen una etapa básica del tratamiento que puede servir para otros procesos del LOG, como se verá más adelante en esta sección. El tercero es específico para el proceso propuesto en este apartado, y sirve para establecer un Índice de Interés de la comunidad de usuarios por los distintos elementos que integran el repositorio de conocimiento, en concreto para los documentos, los temas, los nodos y los autores.

En el primer paso de la etapa básica se divide el fichero LOG del servidor Web en ficheros LOG de sesiones de cliente clasificados por días. Como resultado aparecerán tantos directorios como días hubiera en el periodo considerado y tantos ficheros de sesiones por día como hubiera en el mismo. En cada fichero se incluirán las líneas de actividad del fichero de LOG general que se consideren generadas por una sesión de usuario en el cliente. Estas sesiones vienen determinadas por las trazas de actividad producidas sin que entre ellas transcurra un periodo de

tiempo máximo establecido. En el segundo paso de la etapa básica, se procesan los ficheros obtenidos de sesiones de cliente para eliminar registros irrelevantes para el análisis y para establecer tramos de actividad periódicos, consiguiendo unos LOGs refinados por usuarios y días.

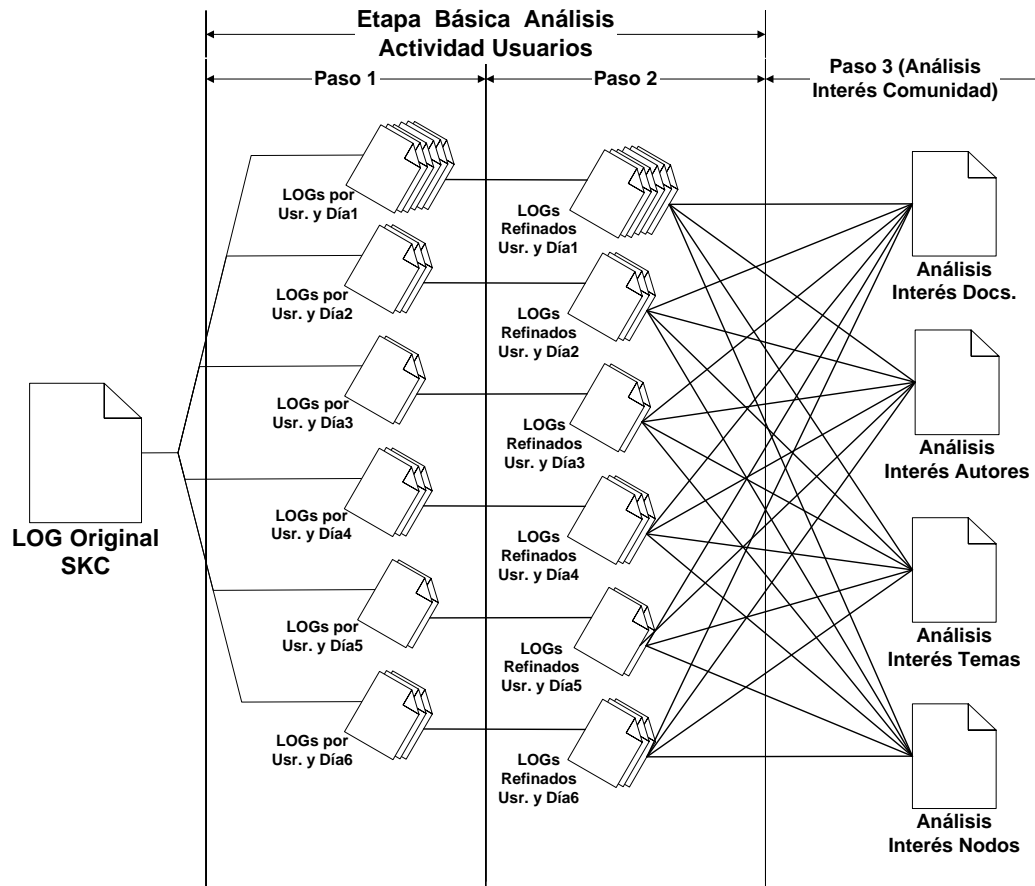


Figura 4-6. Proceso análisis actividad de usuarios en el cliente de SKC, a partir de LOG Web con registros del MC

En el paso específico de análisis se procesan los ficheros de sesión refinados obtenidos al final de la etapa básica para cuantificar los tramos de actividad por sesiones, y evaluar la actividad por nodos, temas, documentos y autores en función de los tramos de sesiones en los que aparecen. En este caso, en el análisis sólo se consideran las identidades de los usuarios para establecer las sesiones de su actividad, pero no para hacer un seguimiento personalizado de ellos, por lo que las conclusiones obtenidas son generales a la comunidad y no particulares para cada usuario.

```

Identif|SumaEval|Tramos|Media|MediaNormal
KC_TALF2N270:T11|3875|630|6.15079365079365|0.237467826939576
KC_TALF2N270:T12|3035|380|7.98684210526316|0.185990930261061
...
    
```

Figura 4-7. Primeras líneas fichero resultado análisis

Como resultado del proceso descrito se obtiene un fichero por cada uno de los tipos de elementos de conocimiento considerados, que incluye varios datos: identificación del elemento,

valoración de eventos recibido, número de tramos de actividad contabilizados, valoración media de eventos recibida por tramo y valoración media normalizada respecto al máximo de las medias obtenidas para todos los elementos del tipo considerados. Este último valor es el coeficiente propuesto para representar el Índice de Interés colectivo asignado a cada elemento en un periodo de uso del sistema, se trata de un número entre cero y uno, que es más cercano a la unidad cuanta más actividad se ha observado asociada al correspondiente elemento de conocimiento de SKC. En la Figura 4-7 pueden verse las primeras líneas del fichero de análisis de actividad de la comunidad de usuarios sobre los temas de un nodo.

## **4.2.2 Módulo de Análisis de Conocimiento**

En este apartado se presenta cómo se integran algunas de las técnicas e ideas expuestas en el capítulo de Estado de la Cuestión para implementar un prototipo de Módulo de Análisis de Conocimiento (MAC) para SKC sobre un sistema KC. Dicho módulo es parte del Módulo de Análisis de SKC del que se habló en la presentación de la arquitectura del sistema en el capítulo anterior. El MAC se encarga de procesar el conocimiento explícito en el seno de SKC con el fin de revelar otro latente y reincorporarlo al sistema de forma que se pueda aprovechar. Para ello, el módulo emplea textos asociados al conocimiento, el modo en que éste se estructura y la forma en que sus componentes interaccionan. Como resultado, el conocimiento revelado proporciona nuevas oportunidades de acceso e interacción con el sistema y el conocimiento.

Como ya se ha dicho antes, los árboles de conocimiento de KC se tratan como ontologías en SKC. Esto es posible porque el árbol de conocimiento de un nodo KC representa el entendimiento común y compartido de la comunidad correspondiente sobre el dominio al que se dedica el nodo [Gruber 1993]. Así que dicho árbol puede considerarse como representación de una ontología subyacente de ese dominio. De este modo, la asignación de documentos a los temas en el árbol de conocimiento supone la anotación semántica de esos documentos en el ámbito ontológico. Esta es la visión que el MAC tiene del árbol de conocimiento del nodo donde funciona. En este contexto, tiene sentido interesarse por la anotación automática de documentos -asignación automática de documentos a temas- [Kiryakov 2004] o por el mapeo entre ontologías -árboles- de distintos nodos [Noy 2002].

En el MAC se emplean técnicas de minería de texto que permiten el procesamiento de información textual poco estructurada mediante el uso de modelos vectoriales [Baeza 1999][Chang 2001]. Estas técnicas gozan de gran popularidad en la actualidad, sobre todo por su empleo en los indexadores automáticos de contenidos de la Web.

Además, el MAC utiliza procedimientos de análisis del lenguaje [Carreras 2004], propios del procesamiento del lenguaje natural. La aplicación de estas técnicas en el campo de la recuperación de información no está muy extendida, porque el esfuerzo computacional de su

empleo no se justifica con los beneficios que aporta en los casos más comunes, donde alguno de los textos comparados son pequeños y el repositorio a tratar grande -como ocurre al utilizar buscadores convencionales en la Web-. Sin embargo, parece que la situación puede ser diferente al comparar textos mayores en repositorios de tamaño moderado, que es el caso que nos ocupa, por lo que nos ha parecido interesante emplear dicha técnica [Brants 2004]. Existen otras alternativas a esta aproximación [Baeza 1999] que el prototipo podría incorporar para contrastar sus resultados.

Por último, el objetivo del módulo es convertir el resultado obtenido en algo útil para interactuar con el sistema y sus contenidos. Para ello es imprescindible resolver problemas relacionados con el filtrado de la información a mostrar, típicos de los sistemas de recomendación [Adomavicius 2005], y con la visualización de dicha información [Geroimenko 2002].

#### ***4.2.2.1 Módulo de Análisis de Conocimiento de SKC***

El Módulo de Análisis de Conocimiento (MAC) del sistema Semantic KnowCat (SKC) está concebido con el propósito de procesar el conocimiento existente en el entorno para revelar conocimiento implícito, que no es evidente en su estado original, y obtener nuevo conocimiento que pueda resultar aprovechable por el propio sistema y la comunidad que lo utiliza. A este proceso del conocimiento se le puede ver como una digestión porque pretende extraer del conocimiento existente algo nuevo y asimilable por el sistema.

El MAC trabaja sobre bases de conocimiento del sistema, depositando el resultado de su actividad en los mismos repositorios, de manera que el sistema y los usuarios puedan utilizar sus aportaciones de forma transparente. El MAC considera que el conocimiento está constituido por ítems. Los ítems pueden ser documentos, temas, árboles de conocimiento, nodos, usuarios, etc.

Con el nuevo conocimiento el sistema puede mejorar de varias formas la gestión que realiza, por ejemplo, proporcionando diferentes vistas del repositorio y nuevos servicios de acceso; simplificando a los usuarios la clasificación de los ítems de conocimiento aportados al sistema; o informando al usuario de relaciones implícitas entre ítems, dado el contexto de la interacción.

Todos los ítems de conocimiento considerados por el MAC deben tener asociados textos de descripción, que pueden asignarse de forma manual o automática; de esta segunda alternativa puede encargarse el propio módulo en algunas ocasiones. Los textos descriptivos asociados a los documentos que SKC maneja actualmente son los propios documentos, puesto que contienen información textual. En el caso de los temas -que no son sino colecciones de documentos-, vienen dados por los textos descriptivos de los documentos que se clasifican en su



seno o de los subtemas que contienen; aunque inicialmente también pueden usarse textos de modelo, que no tienen por qué formar parte del repositorio de conocimiento del sistema. Con los nodos pasa algo parecido, puesto que se les puede considerar los temas raíz de los árboles de conocimiento constituidos por los temas y por los documentos incluidos en ellos. En el caso de los usuarios, se les pueden asociar varios textos de descripción, considerando los documentos o temas que, por ejemplo, aportan o frecuentan.

El MAC realiza dos tareas fundamentales: por un lado, revela conocimiento que se encuentra latente en el sistema; y por otro, lo incorpora de forma explícita en el propio sistema para permitir su explotación. El conocimiento implícito se encuentra en las relaciones que se establecen entre los distintos ítems de conocimiento, por ejemplo en los contenidos que incluyen o en las interacciones que establecen unos con otros. El conocimiento explicitado se incorpora al sistema en su nuevo estado manifiesto, calificando a los ítems de conocimiento ya existentes, o en forma de nuevos ítems de conocimiento que se añaden al repositorio.

La vinculación mediante los contenidos se establece, en esta aproximación, obteniendo descriptores vectoriales de peso de palabras a partir de los documentos de texto asociados a los ítems. Con tales descriptores se pueden comparar los ítems, determinar la distancia que les separa y formar agrupaciones entre ellos.

Las asociaciones basadas en la interacción entre ítems de conocimiento se determinan analizando cómo los ítems se relacionan entre sí. De este modo, se considera cómo los temas agrupan documentos y otros temas en el árbol de conocimiento de los nodos, y cómo los usuarios aportan documentos al sistema.

El conocimiento incorporado en el sistema como resultado del análisis proporciona nuevas oportunidades de explotación del repositorio. Por un lado, los ítems de conocimiento enriquecidos pueden presentarse desde nuevas perspectivas gracias a los nuevos atributos. Por otro lado, los ítems incorporados al sistema por el conocimiento asimilado por este, permiten ofrecer a los usuarios distintas vistas del repositorio y nuevos servicios.

#### ***4.2.2.2 Vinculación por contenido***

En nuestra aproximación hemos considerado, inicialmente, cuatro clases de ítems de conocimiento: nodos que son instancias del sistema dedicadas a la gestión del conocimiento sobre un área con la ayuda de una comunidad de usuarios; temas estructurados en forma de árbol de conocimiento que desarrollan los distintos aspectos del tema principal del nodo; usuarios que constituyen la comunidad que participa en el nodo; y los documentos que describen los distintos temas y son aportados por los usuarios, consultados por ellos y objeto de su consideración.

A todos los ítems de conocimiento que se consideran en el sistema es posible asociarles documentos de texto que los describen. Estas asignaciones pueden tener diversos orígenes. En primer lugar, las asociaciones de textos a los ítems de conocimiento pueden provenir de la naturaleza de tales ítems; por ejemplo los documentos empleados en los experimentos son de tipo textual. En segundo lugar, tales asociaciones pueden derivarse de las relaciones explícitas de los ítems de conocimiento con otros ítems que tienen ya textos asociados. Esto ocurre con los temas que organizan los documentos, los usuarios que aportan documentos al sistema o el nodo que contiene unos y otros. En tercer lugar, las asociaciones de textos descriptivos a los ítems pueden desprenderse de relaciones más dinámicas, como la que se establece entre los usuarios y los documentos que visitan con mayor frecuencia o sobre los que manifiestan su opinión, o como las que se ponen de manifiesto entre documentos referenciados entre sí. Por último, siempre es posible asociar textos descriptivos a los ítems que incidan sobre algún aspecto de utilidad, como son los currículos de los usuarios, sus temas de interés, las palabras claves asociadas a los documentos o las descripciones de los temas. Este caso es completamente general y puede aplicarse a documentos de naturaleza no textual como podrían ser imágenes, sonido, etc.

Una vez que se ha asociado un texto descriptivo a uno de los ítems considerados, es necesario ponerlo de manera que pueda utilizarse como instrumento de comparación. Esto se consigue convirtiendo el texto en un descriptor, que estará ligado al aspecto al que se refiera. Por ejemplo, si el texto asociado a un usuario describe los temas que le interesan, el correspondiente descriptor se referirá a las preferencias del usuario; pero si el texto describe los documentos que él mismo ha elaborado, el correspondiente descriptor se referirá a su labor creativa. De este modo, los ítems podrán tener tantos descriptores como aspectos de ellos se tengan en consideración.

Como ya se ha dicho, en nuestra aproximación los descriptores son vectores de peso de palabras, que pueden utilizarse para determinar similitudes con otros vectores del mismo tipo y así relacionar los ítems de conocimiento correspondientes [Baeza 1999][Chang 2001]. El proceso de obtención de tales vectores se ilustra en la Figura 4-8 y parte de los textos asociados a los ítems. Como los textos pueden estar en distintos formatos, es preciso tratarlos para obtener sus contenidos desnudos en forma de texto plano. En nuestra aproximación inicial se han considerado ficheros de texto en formato PDF -convertidos previamente a HTML- y HTML, aunque ambos se transforman en ficheros de texto plano, antes de iniciar el proceso.

Después de que se ha eliminado el formato de los textos -generando ficheros de texto plano (HTX representados en la Figura 4-8)- es preciso identificar los lemas a los que se refieren las palabras (obviando las formas gramaticales en las que se presentan) y determinar las categorías

gramaticales a las que pertenecen. Con ello se unifican las referencias a conceptos, se reduce el número de palabras distintas consideradas y se identifican los términos que no tienen utilidad.

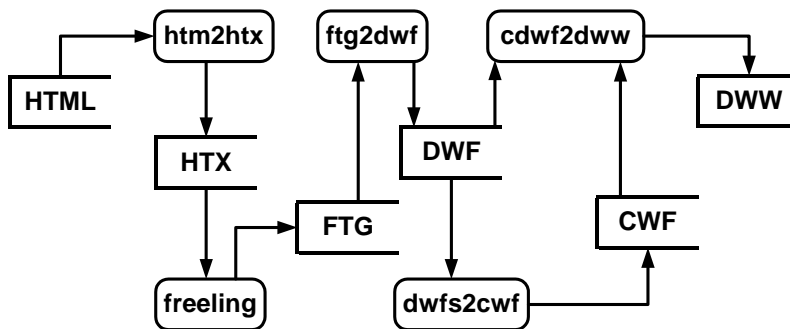


Figura 4-8. Proceso obtención vectores peso palabras para ítems conocimiento desde HTML

En nuestra aproximación hemos utilizado la herramienta de análisis de lenguaje FreeLing [Carreras 2004], que facilita la obtención de toda la información necesaria para alcanzar los anteriores objetivos. FreeLing permite analizar un texto para identificar las categorías gramaticales a las que pertenecen las palabras que lo forman y determinar los lemas a los que corresponden dichas palabras en un diccionario de referencia. Cuando FreeLing no encuentra en el diccionario un lema adecuado para alguna palabra, la considera como un nuevo lema. Con todo ello, la herramienta consigue establecer la interpretación morfológica más probable de cada una de las palabras que integran el texto, que servirá para determinar una aproximación semántica del mismo. Como resultado del análisis, FreeLing proporciona una versión etiquetada del texto (ficheros FTG a los que se refiere la Figura 4-8), indicando para cada aparición de palabra su forma original junto con el lema y la correspondiente interpretación morfológica que se consideran más factibles.

El texto etiquetado obtenido con FreeLing se procesa atendiendo a sus categorías gramaticales, para descartar completamente las entradas de palabras que se consideran poco relevantes para la comparación de textos, como son los determinantes, las conjunciones o las preposiciones. También se eliminan las etiquetas y las formas originales de las demás entradas. De este modo el texto original queda convertido en una secuencia de lemas, que existen en el diccionario de referencia utilizado, o que se han acuñado a partir de términos singulares que no aparecen en él. En esta secuencia las ocurrencias de distintas formas de las mismas palabras en el texto original aparecen como repeticiones de los mismos lemas. A todos los lemas incluidos en dicha secuencia se les puede atribuir un interés semántico para contribuir a la creación del descriptor, que es el objetivo del proceso.

Contando las apariciones de cada término en la secuencia de lemas se puede establecer la frecuencia de cada uno de ellos. De esta forma se generan los ficheros de frecuencia de palabras para cada texto asociado a un ítem de conocimiento (ficheros DWF en la Figura 4-8). En los ficheros DWF hay sólo una entrada por cada lema, que contiene el correspondiente identificador

y su frecuencia, normalizada respecto al máximo de apariciones de las demás palabras consideradas en el documento.

Siguiendo un proceso parecido al descrito -pero trabajando sobre una colección de textos representativos del uso general del idioma en el que se está trabajando- se genera un fichero de referencia con las frecuencias de palabras en dicha colección (fichero CWF referenciados en la Figura 4-8), que representa las frecuencias de las palabras en el uso común del idioma [Baeza 1999]. La colección de documentos se procesa como si se tratara del texto asociado a un ítem de conocimiento. Para que las palabras encontradas y su frecuencia sean representativas del uso general del idioma, la colección debe ser suficientemente amplia y abarcar una temática de tipo general. En nuestra aproximación se han utilizado los 748 artículos incluidos en los anuarios del periódico El País de cuatro años distintos, que tratan de los acontecimientos más destacados acaecidos en esos períodos en los principales ámbitos informativos, tales como sociedad, cultura, deportes, etc.

Los ficheros CWF son parecidos a los DWF, hay una entrada por cada lema, con el identificador en cuestión y su coeficiente de frecuencia de documento inverso. Esta frecuencia es el logaritmo en base diez del cociente del total de documentos en la colección  $N$ , entre el número  $n_k$  de documentos donde aparece el término (ver Fórmula 4-1). Dicho coeficiente es un indicador de la frecuencia de empleo del término en el uso general del idioma que representa la colección y denota la rareza de aquel.

$$p_{k,i} = f_{k,i} \times fdi_k = f_{k,i} \times \log \frac{N}{n_k}$$

*Fórmula 4-1. Peso palabra en documento y frecuencia término documento inverso en colección*

Partiendo de los ficheros de frecuencia de palabras de cada ítem de conocimiento (DWF), y utilizando el fichero de frecuencia de palabras en la colección de referencia (CWF), se establece un peso para cada palabra en el texto asociado al ítem. El peso de una palabra en un texto representa la relevancia de la palabra en él. Una palabra es más característica de un texto cuanto más frecuente sea en el correspondiente texto y menos en el uso general del idioma en que está escrito. En concreto, el peso  $p_{k,i}$  de una palabra  $k$  en un documento  $i$  es el producto de la frecuencia normalizada  $f_{k,i}$  de la palabra  $k$  en el texto  $i$ , por la frecuencia de documento inverso del término en la colección utilizada de referencia  $fdi_k$  (ver Fórmula 4-1).

El vector formado por las palabras que aparecen en el texto asociado a un ítem de conocimiento y sus respectivos pesos constituye el descriptor resultante del proceso, que se guarda en forma de un fichero (ficheros DWW resultado del proceso ilustrado en la Figura 4-8). Los ficheros DWW incluyen una entrada por cada lema que aparece en el texto, con el identificador de éste y el peso que se le atribuye. Los DWW se utilizarán para comparar los ítems entre sí, calculando

el grado de similitud entre los vectores de peso de palabras que representan. La similitud entre dos vectores se puede establecer de varias formas, a partir de alguna distancia como la euclídea o del coseno del ángulo que forman, que es el procedimiento empleado en esta aproximación.

$$\text{sim}(v_i, v_j) = \frac{\vec{v}_i \bullet \vec{v}_j}{|\vec{v}_i| \times |\vec{v}_j|} = \frac{\sum_{k=1}^n p_{k,i} \times p_{k,j}}{\sqrt{\sum_{k=1}^n p_{k,i}^2} \times \sqrt{\sum_{k=1}^n p_{k,j}^2}}$$

Fórmula 4-2. Similitud entre vectores en función del coseno del ángulo que forman

Por lo tanto, la similitud entre dos vectores  $v_i$  y  $v_j$  es el producto escalar de los dos vectores, partido por el producto de los respectivos módulos. El producto escalar de los vectores se calcula como el sumatorio de los productos de sus componentes  $p_k$  en cada una de sus  $t$  dimensiones. El módulo de un vector se computa como el sumatorio de los cuadrados de las componentes del vector (ver Fórmula 4-2).

El grado de similitud entre dos vectores es un coeficiente entre cero y uno. Cuanto más cercano esté el valor a la unidad, más parecidos resultarán los vectores, y cuanto más próximo a cero, menos semejantes serán. Las relaciones de similitud establecidas entre ítems de conocimiento están calificadas con dicho coeficiente. En nuestra aproximación se consideran vinculados los ítems de conocimiento que superan un determinado umbral del coeficiente de similitud de la relación entre ambos. Desafortunadamente, este umbral no puede establecerse de forma fija ni general para todos los casos, puesto que dependiendo de circunstancias tales como la temática de los nodos o la naturaleza de los documentos considerados, puede variar mucho la elección de su valor.

#### 4.2.2.3 Vinculación por interacción

Este tipo de vinculación entre ítems de conocimiento se establece analizando cómo éstos se relacionan entre sí. En nuestra aproximación hemos considerado la forma en que se organizan los documentos y los temas dentro del árbol de conocimiento, y cómo los usuarios se relacionan con los documentos que aportan al sistema. El análisis sigue un proceso que pasa por tres etapas.

En primer lugar, el MAC establece los ítems de conocimiento incluidos en el árbol que requieren ser tratados. La primera vez que se lleva a cabo el proceso en un nodo, se deben procesar todos los ítems de conocimiento del árbol, pero en procesos sucesivos sólo es necesario tratar los ítems que han cambiado, o los vinculados con éstos. En nuestra aproximación, por ejemplo, los cambios en los documentos afectan a los temas de la rama del árbol de conocimiento donde se ubican, al nodo y a los ítems relacionados con unos y otros de alguna manera, pero no a la totalidad de los elementos del repositorio. Este proceso selectivo de los ítems parece es esencial en sistemas con bases de conocimiento grandes o con actividad intensa.

En segundo lugar, el MAC identifica los usuarios responsables de los ítems de conocimiento válidos en el árbol de conocimiento. En esta aproximación sólo se consideran las vinculaciones entre los usuarios y los documentos válidos que han aportado al sistema. El texto asociado a cada usuario en función de sus aportaciones de documentos es la concatenación de los textos descriptivos de todos ellos. Otros tipos de vinculaciones parecidas a éstas se podrían tratar de forma similar.

Por último, el MAC recupera los componentes textuales que constituyen los textos asociados a los ítems de conocimientos a través de la Web. En nuestra aproximación, hemos partido de los textos de los documentos, que están vinculados de forma consustancial con ellos, para establecer los textos asociados con los demás ítems en función de las relaciones consideradas entre ellos, de las que se ha hablado antes. En nuestra aproximación hemos utilizado el programa GNU wget [GNU 2008b] para recuperar los ficheros con la información textual correspondientes a los distintos ítems de conocimiento e integrarlos -cuando son más de uno- para formar los textos descriptivos asociados. Estos textos suelen ser la concatenación de varios ficheros; por ejemplo el texto asociado a un tema estará formado por los textos asociados a todos los documentos y subtemas que incluye.

#### ***4.2.2.4 Enriquecimiento de conocimiento y explotación del mismo***

Como ya se adelantó, en nuestra aproximación el conocimiento revelado como resultado del proceso de análisis se incorpora al sistema de forma explícita; bien como descriptores que califican a los ítems de conocimiento preexistentes o bien en forma de nuevos ítems de conocimiento.

Los descriptores añadidos a los ítems de conocimiento proporcionan nuevos datos para mostrar facetas ocultas de los elementos a los que califican. Por ejemplo, el interés que despierta un determinado ítem puede servir para destacarlo entre los demás ítems o para ordenar todos ellos. También las palabras más características que incluye un ítem pueden resultar una referencia interesante para buscar información relacionada con él en otros repositorios de información.

En nuestra aproximación, el conocimiento explicitado por el proceso de análisis se incorpora al sistema en forma de una nueva categoría de elementos de conocimiento que representan las relaciones entre ítems de todas las clases previamente consideradas (documentos, temas, usuarios y nodos). Las relaciones así incorporadas en el repositorio proporcionan la base para ofrecer a los usuarios nuevas vistas multidimensionales del conocimiento y nuevos servicios para facilitar su explotación. En concreto, para demostrar esta propuesta hemos implementado una vista interactiva del grafo de algunas relaciones entre ítems de conocimiento del sistema (ver Figura 4-9 arriba), así como un servicio de recomendación sensible al contexto que

proporciona las referencias de ítems -de distintas clases- relacionados con el que el usuario está usando en cada momento (ver Figura 4-9 abajo).

La vista en forma de grafo integra las relaciones estáticas establecidas en el sistema con otras dinámicas que evolucionan a lo largo del tiempo. Entre las primeras están los vínculos jerárquicos que ligan los temas del árbol de conocimiento o los vínculos de autoría que unen a los usuarios con los documentos que aportan a la base de conocimiento. Entre las relaciones dinámicas pueden mencionarse las derivadas del carácter de los ítems presentes en el repositorio en cada momento y las debidas a las interacciones que se establecen entre ellos como consecuencia de la actividad del sistema. En la ilustración (ver Figura 4-9 arriba), puede verse un ejemplo de este tipo de vista, donde los temas están representados por círculos de color naranja, los documentos, como cuadrados claros, las relaciones estáticas están encarnadas por líneas negras y las relaciones dinámicas entre los ítems considerados están representadas en forma de líneas de color según el grado de similitud entre los vectores de los ítems correspondientes, que se incluye como etiqueta.

Por otro lado, el servicio de recomendación ilustra cómo se puede sacar beneficio del nuevo conocimiento para facilitar el uso del sistema y hacer más dinámica y atractiva la interacción con él. En la ilustración (ver Figura 4-9 abajo), puede verse un ejemplo de ventana del sistema que presenta un documento y que incorpora en el área inferior un panel de recomendación, en el que aparecen iconos representativos de distintos tipos de ítems de conocimiento en colores más cálidos cuanto mayor es el grado de similitud entre los vectores asociados a los correspondientes ítems. Además, en la ventana que aparece superpuesta se presenta una representación en forma de grafo de las relaciones más importantes que el documento con el que se está trabajando mantiene con otros ítems del sistema. En este grafo, como en el ejemplo de la vista, los temas se representan por círculos y los documentos por paralelogramos. Sin embargo en este caso los colores de las figuras representan el coeficiente de similitud de las relaciones que las ligan con el ítem de conocimiento central. Otros servicios serían posibles con un planteamiento parecido, como un asistente para ubicar documentos en el tema más adecuado dentro del árbol de conocimiento, o uno para localizar expertos sobre algún tema u otros usuarios interesados por las mismas cuestiones.

Tanto la vista en forma de grafo como el servicio de recomendación implementados permiten la navegación por el conocimiento de un modo diferente al que el sistema permitía cuando no se aprovechaba el conocimiento latente en el sistema. En ambos casos, los grafos se han generado con el paquete Graphviz [Gansner 2000].

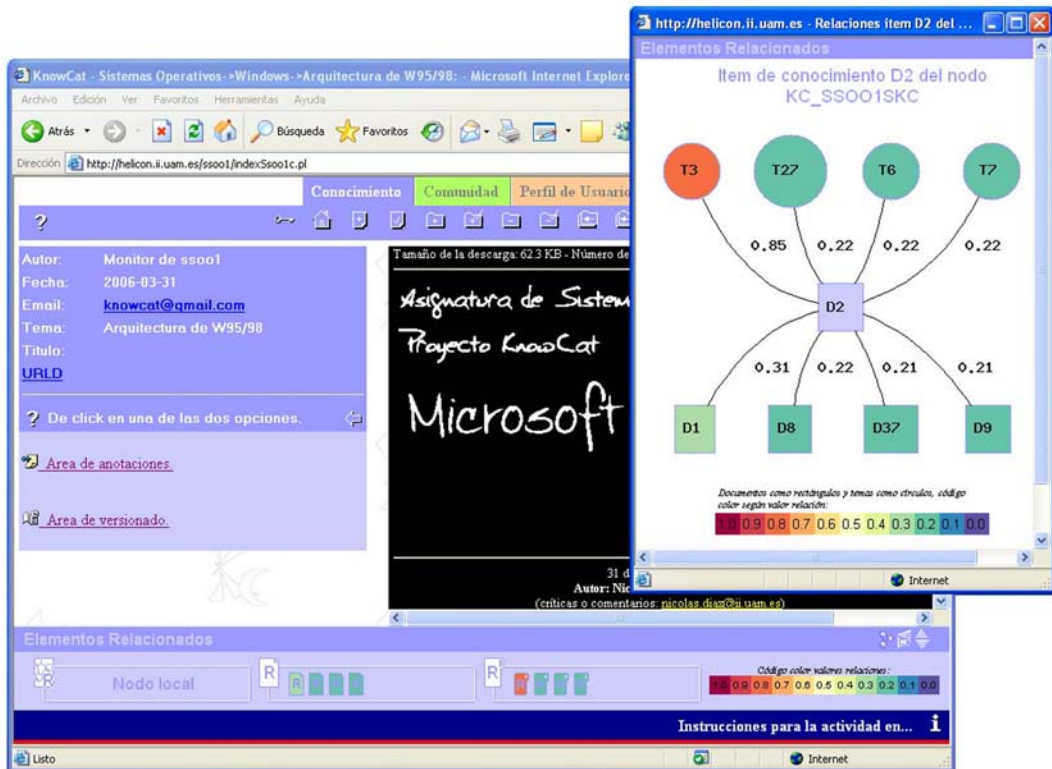
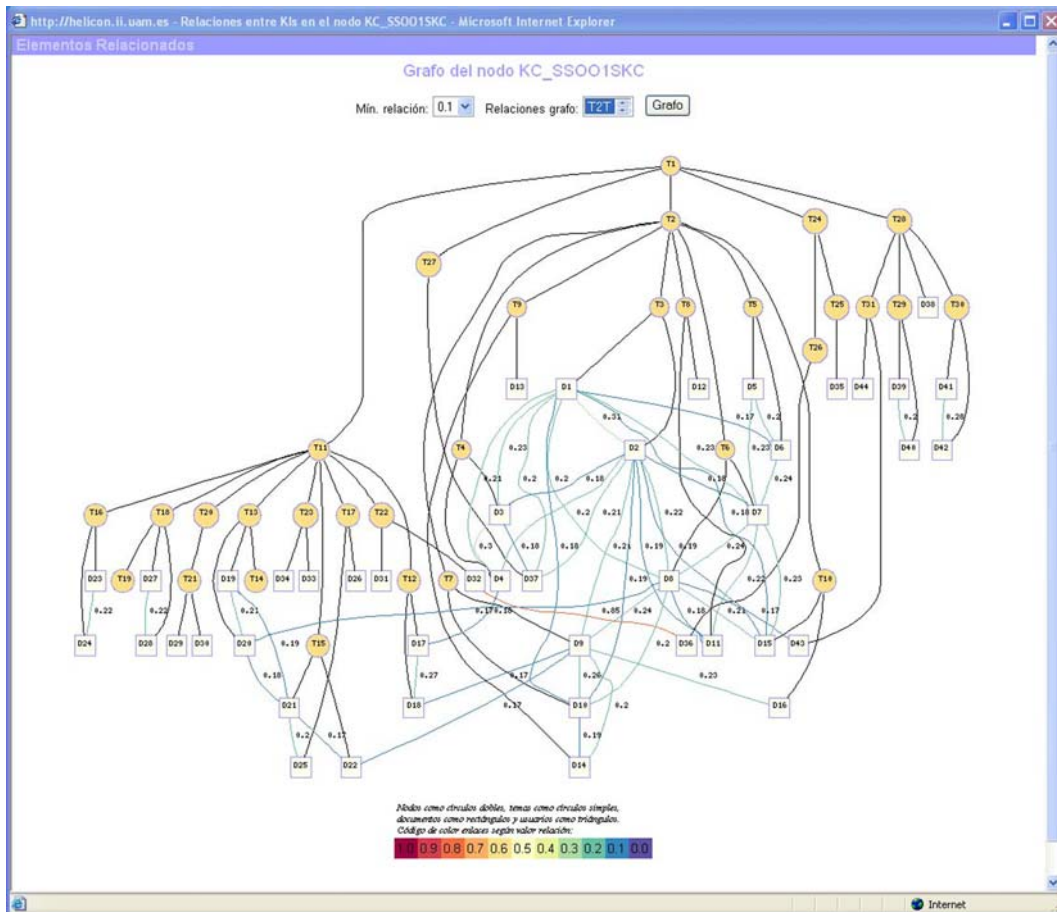


Figura 4-9. Vista interactiva del conocimiento como grafo de ítems relacionados (ventana de arriba) y servicio de recomendación sensible al contexto (ventanas de abajo, área inferior en la de la izquierda y toda la de la derecha)



#### 4.2.2.5 Vector de interés de usuario

En este subapartado se muestra cómo se integra el análisis del interés de los usuarios individuales, basado en los datos y los procedimientos del Monitor de Cliente (MC), con el procesamiento del conocimiento, realizado por el Módulo de Análisis de Conocimiento (MAC), en el ámbito del Módulo de Análisis del prototipo de SKC. Para ello se parte de los Vectores de Pesos de Palabras (VPP), asociados a los ítems de conocimiento vinculados al MAC, y de los Índices de Interés de la comunidad de usuarios por dichos ítems, asociados al MC.

Conviene recordar que, como se ha explicado antes en este documento, el MAC es un elemento del Módulo de Análisis de SKC, encargado de procesar el conocimiento y la información disponibles en el sistema para facilitar la gestión de dicho conocimiento. El MAC obtiene descriptores con los que calificar los ítems de ese conocimiento -documentos, temas, usuarios, etc.-, este es el caso de los Vectores de Pesos de Palabras (VPP) -de los que se habla antes en esta sección-, empleados para relacionar entre sí los ítems de conocimiento que tienen asociados textos descriptivos. Tampoco hay que olvidar que el Monitor de Cliente (MC), también descrito previamente en la sección, forma parte asimismo del mencionado Módulo de Análisis de SKC y permite monitorizar la actividad de los usuarios con el sistema en el lado del cliente -puesto que SKC es un sistema Web cliente/servidor-, con lo que se consigue mayor precisión que considerando sólo la información disponible habitualmente en el lado del servidor. El MC permite determinar un Índice del Interés de la comunidad de usuarios por los componentes del conocimiento del sistema con los que interactúan, de este índice se habló también anteriormente.

```
...
150.244.57.62 - - [01/Mar/2006:10:33:29 +0100] "GET /5.1a/informSituacion.pl?
urlrsp=../5.1a/monitorKC.pl%3Fest%3D1&UsrID=Monitor%20de%20BDD&CntIni=1141205148142&NtfTmp=114120552
9249&MseM=414&MseD=11&Scr=0&Fcs=19&Blr=19&KeyD=0 HTTP/1.1" 302 296
...
217.127.197.187 - - [01/Mar/2006:10:33:32 +0100] "GET /5.1a/informSituacion.pl?
urlrsp=../5.1a/monitorKC.pl%3Fest%3D1&UsrID=antonio%20martinez&CntIni=1141203026247&NtfTmp=114120553
8460&MseM=50&MseD=0&Scr=0&Fcs=1&Blr=1&KeyD=0 HTTP/1.1" 302 296
...
```

Figura 4-10. Líneas de registro del MC extraídas del LOG de un nodo SKC

Sobre la base de tales indicadores se propone un nuevo descriptor de interés para calificar a los usuarios, el Vector de Interés de Usuario (VIU), que permite relacionar los usuarios dentro del sistema por sus intereses. Además, se propone otro descriptor más general y potente, basado en dichos VIUs y los VPPs, el Vector de Pesos de Palabras de Interés de Usuario (VPPIU), con el que es posible relacionar los usuarios de SKC con otros ítems de conocimiento del sistemas o de fuera de él, con tal de que se les asocien VPPs.

Anteriormente se ha mostrado cómo se puede obtener una impresión sobre el interés de la comunidad de usuarios hacia los ítems del conocimiento de un nodo de SKC. El procedimiento correspondiente se basa en el análisis del registro de actividad -LOG- del servidor Web que

soporta el sistema. El proceso básico de análisis se describe de forma detallada en un apartado anterior, pero a grandes rasgos consiste en determinar la actividad por días y sesiones de usuario, y regularizar los registros correspondientes para facilitar su tratamiento posterior.

Cuando los usuarios deben identificarse para utilizar el sistema, las líneas de registro del Monitor de Cliente (MC) incluyen las referencias de quiénes las han provocado. Esto puede apreciarse en las líneas de registro del MC extraídas del LOG de un nodo del prototipo de SKC presentadas en la Figura 4-10, donde el identificador de usuario -indicado por el campo UsrId- de la primera es “Monitor de BDD” y en la segunda “antonio martinez”. Con esta información es posible hacer un seguimiento individual de la actividad de cada usuario a lo largo del tiempo, incluso a través de varias sesiones. Esto es lo que hemos hecho para establecer el Índice de Interés de Usuario (IIU) y los Vectores de Interés de los mismos (VIU).

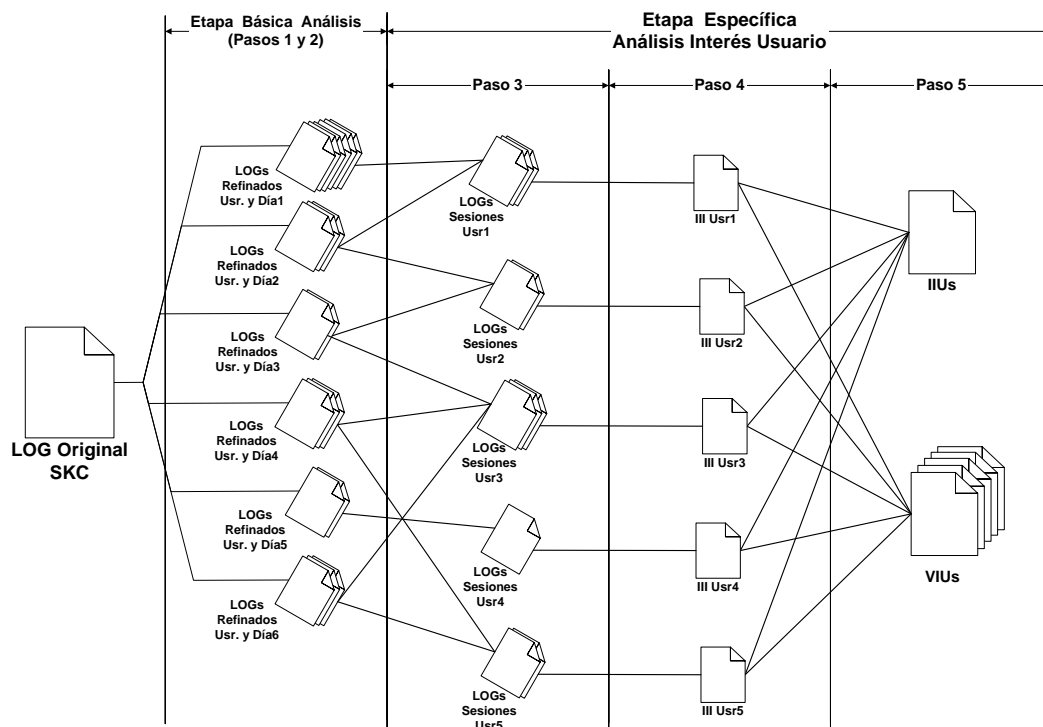


Figura 4-11. Proceso análisis actividad por usuario en el cliente de SKC, a partir de LOG Web con registros del MC

En concreto, (ver Figura 4-11)partiendo de los resultados obtenidos de la etapa básica del análisis descrito en la sección sobre el Monitor de Cliente, que son los ficheros refinados de LOG por sesiones de usuario y días, se acumulan por persona los valores asignados en los ciclos de monitorización a los ítems de conocimiento registrados. Como consecuencia, se obtiene para cada usuario una lista de todos los ítems de conocimiento accedidos por él, calificados con el valor Índice de Intensidad de Interacción mantenida (III) con cada uno de ellos. De esta lista se pueden extraer un par de objetos interesantes.

Por un lado, sumando por usuario los IIIs, puede asignarse a cada uno de ellos un valor representativo de la Intensidad de Interacción de Usuario (IIU) mantenida con el sistema, que puede utilizarse como índice del interés demostrado por el sujeto en su actividad con el nodo y servir para compararlo con los grados de interés evidenciados por otros miembros de la comunidad.

Por otro lado, dividiendo todos los IIIs de cada usuario por la suma de ellos, se obtiene para cada sujeto un vector que representa la contribución relativa de cada ítem al interés manifestado por el usuario. Estos son los Vectores de Interés de Usuario (VIUs), con los que es posible establecer comparaciones entre los intereses demostrados por los integrantes de la comunidad que usa el sistema.

Los IIUs pueden utilizarse como indicadores de la calidad de los VIUs, puesto que para valores de interés demasiado bajos los vectores de interés no resultarán significativos. Tampoco serán buenos los VIUs obtenidos a partir de un conjunto de IIIs en la que todos resulten demasiado reducidos, aunque los correspondientes IIUs no lo sean. Estos límites se pueden establecer a partir de las distribuciones de los IIUs y de los IIIs, descartando los valores extremos inferiores.

#### ***4.2.2.6 Vector de pesos de palabras de interés de usuario***

En un apartado anterior de esta sección se ha explicado el procedimiento empleado en el prototipo de SKC para crear descriptores mediante Vectores de Pesos de Palabras (VPPs) [Yang 1998] [Salton 1989] [Chang 2001], que permitían comparar ítems de conocimiento que tienen textos descriptivos asociados, y se mostró el modo de utilizar esos descriptores para establecer relaciones entre dichos ítems. Los VPPs son listas de lemas -palabras- a los que se ha dado un peso -puntuación-. Los lemas se obtienen a partir de las palabras significativas del texto de partida, agrupando las distintas formas en que puedan aparecer -número, tiempo, etc.-. Los pesos se establecen considerando la frecuencia de los lemas en el texto de origen y la frecuencia de los mismos en el uso general del idioma, de modo que las palabras más frecuentes en el texto y menos corrientes en el idioma tengan pesos mayores que las más comunes en el uso ordinario de la lengua, especialmente si son poco numerosas en el texto [Baeza 1999]. Para comparar los ítems unos con otros se determina la similitud entre los VPPs, determinado el coseno del ángulo que forman dichos vectores. El grado de similitud entre dos VPPs es un valor entre cero y uno, más cercano a la unidad cuanto más parecidos son los vectores, y más próximo a cero cuanto menos lo son.

Hasta ahora, en el prototipado de SKC los VPPs asociados a los ítems de conocimiento se basaban en textos que se consideraban vinculados totalmente con los respectivos ítems: los documentos textuales eran sus propias descripciones, a los temas que forman los árboles de conocimiento se les asociaba un texto descriptivo compuesto por la unión de las descripciones

de todos sus documentos y subtemas, las descripciones de los autores se obtenían de la concatenación de todos los documentos creados por ellos, etc.

Esta estrategia no parece adecuada para el tratamiento del interés de los usuarios, puesto que el interés de cada uno se reparte entre varios ítems de conocimiento de forma desigual. El vector de interés de usuario, VIU, de cada sujeto muestra cómo se distribuye su interés entre varios ítems de conocimiento, indicando para todos ellos un índice de interés, que es el tanto por uno de su participación en el total. Tomando como partida cada VIU y los VPPs de los ítems a los que el primero se refiere, es posible crear un descriptor por usuario representativo del contenido del interés demostrado por el correspondiente usuario en el sistema.

El nuevo descriptor es el vector de pesos de palabras de interés de usuario, VPPIU. Se trata de un VPP en el que el peso de las palabras se establece a partir del peso de las mismas en los descriptores de pesos de palabras de los ítems de conocimiento correspondientes y de la proporción de dichos ítems en el interés del usuario. De este modo, el peso  $p_{k,u}$  de una palabra  $k$  en el interés de un usuario  $u$  es el sumatorio del producto del índice de interés de ítem  $i_{i,u}$  del usuario por el peso de la palabra en el correspondiente ítem  $p_{k,i}$ , para cada uno de los  $n$  ítems de interés del VIU de dicho usuario (ver Formula 4.3).

$$p_{k,u} = \sum_{i=1}^n i_{i,u} \times p_{k,i}$$

*Fórmula 4-3. Peso de palabra en el interés de usuario*

Los VPPIUs, como los VIUs y los VPPs, dependen del periodo considerado para su cálculo y van evolucionando con el tiempo según se manifieste el interés de los usuarios por los ítems de conocimiento del sistema o varíen las descripciones de los ítems que involucran -por ejemplo, cuando los documentos se actualizan o se incorporan nuevos elementos en los temas-.

Los VPPIUs son como los demás VPPs y pueden compararse entre ellos y servir para establecer relaciones entre los usuarios y cualquier elemento que tenga asociado un descriptor de este tipo. En concreto, dentro del ámbito de SKC, tiene sentido determinar la similitud de intereses entre usuarios de una o varias instancias -nodos- del sistema, y también establecer el posible interés de los usuarios por la temática de un nodo o de algún tema de su árbol de conocimiento -temario-.

Además, como es fácil establecer VPPs para elementos externos al sistema y accesibles a través de Internet, con tal de que sean textuales o tengan asociada alguna descripción de este tipo, los VPPIUs pueden emplearse para establecer vínculos con ellos. En este sentido, podría resultar interesante descubrir recursos disponibles en otros repositorios de conocimiento -distintos nodos del sistema o repositorios basados en Wiki por ejemplo- que estuvieran en el área de interés de los usuarios dentro del campo de conocimiento del nodo al que estos pertenecen.

Pueden diseñarse diversos servicios y utilidades basadas en el interés de los usuarios dentro de SKC. Por ejemplo, es posible incluir el interés como una variable de los servicios de recomendación soportados con VPPs, de los que se ha hablado antes este documento, también establecer mapas de interés para la comunidad de usuarios, mostrando las vinculaciones entre sus integrantes debidas al interés manifestado por ellos, o crear mapas de interés que muestren las relaciones entre los usuarios y los demás ítems de conocimiento del sistema, según el interés demostrado por los primeros hacia los segundos -a los usuarios también se les considera ítems de conocimiento en SKC-.

### **4.2.3 Módulo de Red de Nodos para soporte de la Interacción entre Instancias del Sistema**

Como ya se ha visto, aunque SKC y KC siguen el mismo enfoque para la gestión del conocimiento, sus diseños son muy diferentes: la arquitectura de SKC se construye alrededor de ontologías [Gruber 1993], que se refieren tanto al conocimiento que el sistema gestiona como a las entidades que intervienen en su proceso. De este modo, los árboles de conocimiento de KC se tratan como ontologías en SKC, que representan las ontologías subyacentes de los dominios de conocimiento tratados, como se vio en el capítulo anterior. La asignación de documentos a los temas en el árbol de conocimiento supone la anotación semántica de esos documentos en el ámbito ontológico. Esto no ha impedido que KC sirva como plataforma de prototipado para SKC, como se en el capítulo que nos ocupa.

El Módulo de Análisis de Conocimiento (MAC) de SKC, presentado anteriormente, se encarga esencialmente de la anotación automática de documentos -asignación automática de documentos a temas- y proporciona las bases para el mapeo entre ontologías -árboles- de distintos nodos [Choi 2006], que es de lo que se encarga de forma conceptual el Módulo de Red (MR), del que trata este apartado. En la práctica, este módulo se ocupa de encontrar otros nodos de sistema existentes en la Web, establecer contacto y colaborar con ellos, con el objetivo de crear y mantener una red que integre los nodos activos conocidos en cada momento, y permita determinar y conservar relaciones particulares de intercambio de conocimiento entre aquellos que comparten intereses. Para lograrlo, el módulo emplea agentes software [Dinverno 2001] que actúan en representación de los nodos del sistema de forma descentralizada y que interaccionan entre si y con otras aplicaciones de la Web para conseguir los objetivos mencionados. Como resultado, en todas las instancias del sistema se dispone de una vista de los demás nodos accesibles a través de la Web y de vinculaciones entre sus repositorios de conocimiento cuando los nodos son afines.

La anotación automática y el mapeo entre ontologías son temas todavía abiertos y fundamentales en la transición de la Web actual a la Web Semántica (WS) [Noy 2002]. En este

momento, las tendencias de la Web, que la Web 2.0 sintetiza [O'Reilly 2005], establecen un marco en el que las redes sociales, la inteligencia colectiva y la colaboración son elementos esenciales, y en el que la actividad de los grupos genera cantidades ingentes de contenidos, que ellos mismos etiquetan a veces empleando ontologías ligeras propias. En este contexto, el mapeo automático entre estas ontologías, concebidas de forma espontánea e independiente por diferentes grupos de usuarios, tiene un gran interés para la integración de los contenidos anotados en ellas. SKC sirve de modelo experimental para indagar en dicha cuestión.

El problema de mapeo entre ontologías, se ha tratado con diversos procedimientos [Choi 2006], que siempre parecen requerir de validación humana. Esta circunstancia puede no ser un inconveniente en entornos colaborativos, aunque en ellos sí es fundamental que los usuarios no se aparten de sus objetivos principales, que no suele ser el proceso explícito el mapeo ontológico. Los sistemas y algoritmos que emplean los contenidos clasificados en las ontologías para realizar el mapeo, parecen los más autónomos, entre ellos están FCA-Merge, LOM, CTXMATCH [Choi 2006]. Esta estrategia creemos que es la más adecuada para trabajar con ontologías ligeras de forma desasistida, en entornos que puedan beneficiarse de su mapeo, pero que no tienen esta labor como su principal objetivo, como es el caso de SKC.

Por otro lado y como es sabido, la Web Semántica (WS) [Berners-Lee 2001] aspira a ser una red global de información vinculada de tal modo que resulte fácil de procesar automáticamente. La WS se basa en la Web actual, que esta orientada a los usuarios humanos. Para construir la WS a partir de la Web actual hay que dotar a ésta de la infraestructura necesaria para que usuarios artificiales puedan utilizarla también. Los agentes software son entidades computacionales autónomas, dotadas de habilidades sociales, capaces de reaccionar a los cambios de su entorno y de actuar sobre él de forma pro-activa [Dinverno 2001], que están llamados a ser usuarios de la WS.

Al mismo tiempo, el sistema SKC es un sistema colaborativo que explota redes de usuarios para llevar a cabo sus objetivos. Esta misma estrategia es la que las instancias del sistema pretenden explotar al integrarse en redes “colaborativas” de nodos de conocimiento, encarnándose en agentes software que cooperan con objetivos concurrentes con los de los usuarios humanos, pero a otro nivel. Como se ha dicho, los agentes son usuarios potenciales de la WS y su capacidad gregaria puede servir en ella como la tendencia humana de agruparse lo hace en la Web 2.0.

De este modo, aunque la funcionalidad central de la integración de nodos SKC podría lograrse con el uso de agentes y con el empleo de Servicios Web [WebServices 2008], por ejemplo, una combinación así desvirtuaría el modelo ligero, flexible, “colaborativo” y reticular, del que se pretende dotar al sistema. Alternativamente, las Web 2.0 prefieren procedimientos de

comunicación entre aplicaciones sencillos y más abiertos, como los que soportan las funcionalidades de RSS [O'Reilly 2005]. Este planteamiento está más en la línea del procedimiento de comunicación propuesto para los agentes del MR, basado en lenguaje de la recomendación FIPA ACL [FIPA 2008], del que se habló en el capítulo de Estado de la Cuestión. Esta filosofía es la que nos ha llevado también a pensar en los buscadores Web como medio de descubrimiento de instancias del sistema, en lugar de opciones más convencionales y rígidas, cómo las propias de las plataformas multiagente, -como el directorio de la arquitectura FIPA [FIPA 2008], descrita en un capítulo anterior- o las características de los Servicios Web -mediante el uso del catálogo de negocios en la red UDDI, la plataforma Universal Description, Discovery and Integration-.

Como es conocido, los buscadores Web se han convertido en instrumentos imprescindibles para encontrar información en la red. Algunos de ellos son sistemas fiables, autónomos y muy potentes, capaces descubrir cualquier recurso accesible en Internet en poco tiempo. Con ellos se podrían encontrar las referencias de otras plataformas de agentes en la red, si cada una de ellas publicara la información necesaria en una página Web. Los buscadores Web convencionales como Google [Google 2008] o Yahoo [Yahoo 2008] disponen de una interfaz HTML para realizar las búsquedas, con la que se invoca un programa al que se pasan las preguntas. Como resultado los buscadores devuelven una página HTML con una lista de enlaces a páginas Web que responden a la cuestión. Si las páginas requeridas contienen alguna cadena de caracteres muy peculiar y se utiliza para su búsqueda, es muy probable que todas las páginas esperadas estén en el resultado y que no aparezcan muchas adicionales. Es fácil tratar automáticamente las respuestas para obtener las URLs de las páginas que apuntan. Sin embargo, actualmente los buscadores convencionales están dirigidos a usuarios humanos y otros tipos de usuarios no suelen ser bien recibidos [Google 2008]. Esta circunstancia no resta interés experimental al mecanismo. Cuando la WS esté implantada plenamente habrá servicios de uso general equivalentes a los buscadores actuales, que permitan trabajar de forma efectiva con los grafos RDF [Shadbolt 2006], destinados a usuarios artificiales como son los agentes software.

#### ***4.2.3.1 El Módulo de Interacción entre Nodos de SKC***

En definitiva, el propósito del Módulo de Red (MR) de SKC es ampliar la proyección de los nodos del sistema, enriquecer los repositorios de conocimiento de los mismos y estimular a las comunidades de usuarios correspondientes. Todo esto se pretende conseguir mediante la divulgación del conocimiento disponible entre los colectivos interesados, en el momento y el lugar que puede resultar relevante, de forma accesible y poco intrusiva, y sin necesidad de mucha atención por parte de los destinatarios.

Como resultado del establecimiento de la red de nodos, SKC se convierte en un sistema de gestión de conocimiento descentralizado con ambición global, en el que los grupos de nodos dedicados a temas similares aparecen vinculados de forma flexible, pero manteniendo su independencia. El modelo propuesto se parece al que se establece en el mundo real entre los especialistas, los grupos de expertos, los artículos y las publicaciones, que se vinculan entre sí de forma abierta. Dicho modelo parece fundamental en el proceso del reconocimiento público del trabajo realizado en los ámbitos de investigación y motor del avance en los mismos.

#### **4.2.3.2 Agentes de nodos**

Como ya se ha mencionado, el MR se basa en agentes que colaboran en la creación y el mantenimiento de una red formada por todos los nodos SKC accesibles, y se ocupan del establecimiento y soporte de las relaciones de intercambio de conocimiento entre los nodos dedicados a los mismos temas.

Para ello, cada agente desarrolla varias tareas de forma autónoma y proactiva: localizar otros nodos en la Web, contactar con ellos para comprobar su actividad y obtener la información básica sobre los mismos, determinar el tipo de vinculación adecuada para mantener con ellos, obtener la información necesaria para establecer los correspondientes vínculos, visitar periódicamente a los nodos identificados para comprobar la continuidad de su actividad, desactivar nodos registrados que dejan de dar señales de vida, activar nodos inactivos que se vuelven a localizar en funcionamiento y revisar de forma reactiva la vinculación con los nodos activos según la evolución del propio nodo y de los otros. Para la localización de otros nodos en la Web los agentes interactúan entre ellos, pero también con otras aplicaciones Web, que en el prototipo desarrollado son buscadores de uso general como Google [Google 2008] o Yahoo [Yahoo 2008].

Además, cada agente atiende las demandas de los demás y aprovecha las interacciones de los desconocidos para identificarlos y establecer relaciones con ellos. Toda esta actividad requiere del empleo de diversas habilidades sociales basadas en la comunicación, en la colaboración y en el empleo de protocolos adecuados para las distintas situaciones.

Internamente los agentes disponen de una interfaz para interactuar con agentes como ellos (IIA) y otra interfaz para hacerlo con otras aplicaciones Web (IW3). Ambas interfaces emplean una red TCP/IP como infraestructura de comunicación. La IIA establece comunicaciones de bajo nivel entre los agentes, utilizando directamente sockets -no HTTP, como propone FIPA-, para soportar la conversación en lenguaje FIPA ACL [FIPA 2008], que encapsula los mensajes propios de los agentes en mensajes y secuencias de interacción característicos de dicho lenguaje. Cada agente establece su IIA en un puerto IP libre de la máquina donde reside. La IW3 se comporta como un navegador Web controlado por el agente al que pertenece, por lo que utiliza



el protocolo HTTP para acceder a los servidores Web considerados de interés, tal como lo haría cualquier usuario humano. La IW3 permite enviar peticiones a los servidores y recibir respuestas de ellos en formato HTML.

En el prototipo desarrollado sólo se han implementado los mensajes y las secuencias de interacción FIPA que son imprescindibles para conseguir una funcionalidad básica. Con la misma intención, los mensajes propios de los agentes son los imprescindibles para permitir las pruebas establecidas. Sin embargo, el protocolo FIPA permite conversaciones más elaboradas y proporciona soporte para mensajes de agente mucho más sofisticados, que serán necesarios en la implementación final del sistema. En este sentido, tiene especial interés para la implementación final del sistema, la posibilidad de que los mensajes de agente utilicen las ontologías que describen el propio sistema para hacer referencias inequívocas a los datos intercambiados.

Durante todo el tiempo los agentes permanecen atentos a las comunicaciones que otros agentes les dirigen. En el prototipo sólo se atiende una visita a la vez y las visitas concurrentes deben esperar hasta que las anteriores terminen, aunque el tiempo de espera está limitado en el visitante y las visitas se consideran fallidas si se sobrepasa ese tiempo. Además, cada agente sigue una política de visitas para mantener contacto con todos los agentes conocidos que se consideran activos. En el prototipo, se crean activos los agentes que no han dejado de atender un cierto número de visitas sucesivas, y la política de visitas consiste en tratar de acceder a los agentes activos con los que no se haya intentado contactar desde hace más tiempo. De forma parecida, los agentes mantienen un régimen periódico de visitas a los servidores Web que se utilizan para descubrir nuevos nodos en Internet.

Para poder interactuar con las aplicaciones Web mencionadas, los agentes necesitan información sobre ellas. En el caso de los buscadores Web considerados en el prototipo, la información necesaria para cada uno es: la URL del buscador en la red, el contenido y el formato de los datos que se le deben enviar, el modo de hacerlo -modo GET o POST de HTTP- y el formato de la respuesta. En el prototipo se consideran buscadores que devuelven páginas HTML con enlaces a los elementos buscados. Los agentes procesan dichas páginas para obtener la información requerida.

#### **4.2.3.3 Repositorio de conocimiento**

La Figura 4-12 muestra tres nodos SKC (N1, N2 y N3), accesibles a través de la Web, con sus respectivos agentes (AN1, AN2 y AN3) y repositorios de información (BDs).

La base de datos de cada nodo de SKC incluye la información completa de todos los ítems propios que constituyen el correspondiente repositorio de conocimiento fundamental (ver N1,

N2 y N3 en las BDs de la Figura 4-12). En los prototipos realizados sobre KC presentados hasta ahora los ítems de conocimiento considerados eran locales al nodo: documentos, autores, temas, árboles de conocimiento, relaciones que ligaban a unos con otros, y el propio nodo.

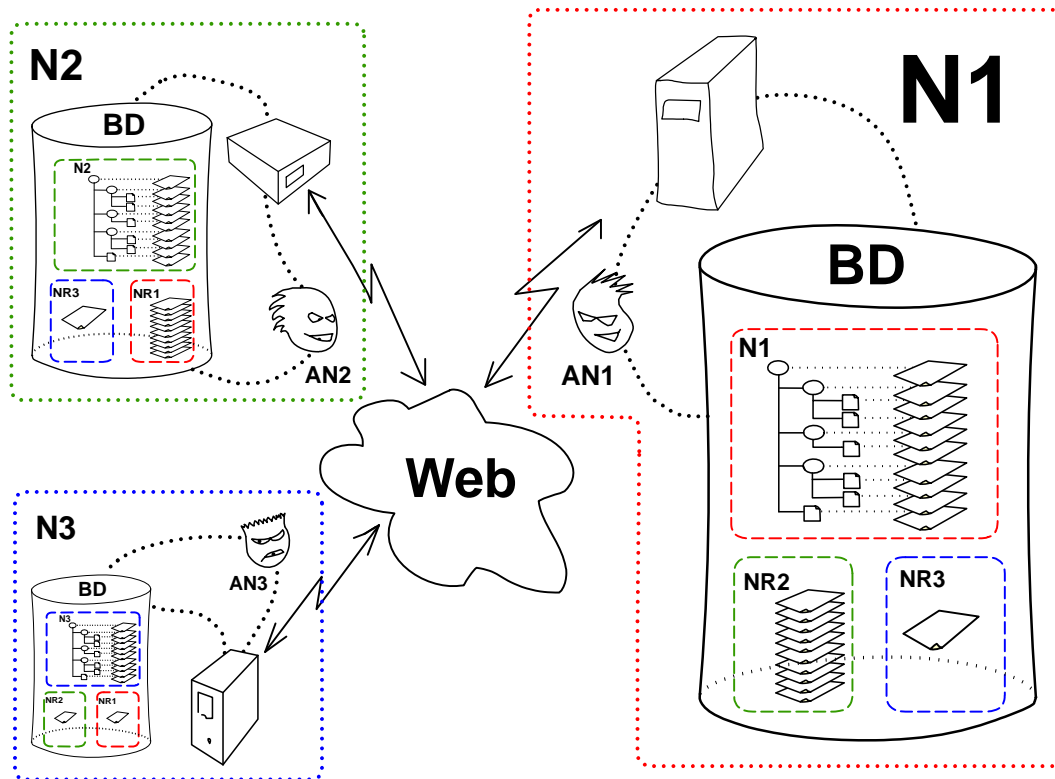


Figura 4-12. Bases de datos de los nodos SKC.

Además, para el soporte del MR, el sistema necesita guardar en la base de datos información acerca de los ítems de conocimientos de otros nodos (ver NR1, NR2 y NR3 en las BDs de la Figura 4-12): nodos, temas, autores y documentos, todos ellos remotos. La información sobre los ítems de conocimiento ajenos no tiene que ser tan exhaustiva como la de los propios, y su composición depende del tipo de ítems del que se trate, de la vinculación existente con el nodo al que pertenecen y la aplicación que se les vaya a dar. En todos los casos, los ítems de conocimiento ajenos residen en sus respectivos nodos y en las bases de datos de los otros nodos sólo se guardan referencias y los datos imprescindibles para su manipulación.

A pesar de que el sistema considera la condición remota de los nuevos ítems de conocimiento para su proceso, como distingue entre las distintas categorías de los demás ítems -documentos, temas, usuarios...-, los principios generales de manejo de todos son los mismos: los ítems de conocimiento están clasificados por tipos -documentos, documentos remotos, temas, temas remotos...-, cada tipo de ítems tiene una representación particular en la base de datos, a los ítems se les pueden asociar descriptores múltiples para calificarlos -en el prototipo se emplean Vectores de Peso de Palabras (VPP), introducidos anteriormente en este documento-, y se pueden relacionar unos ítems con otros en base a sus descriptores. De este modo, es posible

analizar los registros de datos y de actividad de los nuevos ítems cómo se hacía con los registros de los ya existentes, incorporar dichos ítems en funcionalidades y servicios preexistentes, descritos antes en este capítulo, e integrarlos con todos los demás en nuevas utilidades diseñadas específicamente para la explotación de la red de nodos SKC.

#### **4.2.3.4 Descubrimiento de nodos en la Web**

Para localizar nuevos nodos en la Web los agentes emplean tres estrategias distintas: usar referencias suministradas durante la instalación, emplear algunos buscadores Web de uso general conocidos y utilizar las referencias proporcionadas por otros nodos con los que se mantiene contacto.

El primer caso es el más sencillo, pero también el menos potente, simplemente parte de la suposición de que es posible que los encargados de la instalación de los nuevos nodos conozcan las referencias de otros nodos en Internet y que puedan proporcionarlas durante la configuración inicial de los sistemas para facilitar la integración de los nodos en la red SKC.

La segunda opción es más poderosa que la primera, porque permite explotar los recursos de uso general y dominio público disponibles en la Web para recuperar información, como Google o Yahoo. Estos recursos resultan relativamente persistentes y son autónomos y completamente independientes de SKC. Para utilizarlos es necesario suministrar al sistema información sobre cada uno de ellos, fundamentalmente, las correspondientes URLs, los modos de invocarlas y los formatos y los datos que requieren.

La última posibilidad complementa a las dos anteriores y sirve para dinamizar éstas, puesto que permite renovar las referencias de nodos y de buscadores conocidos, aprovechando la comunicación entre nodos y la colaboración entre ellos, que son dos aspectos fundamentales de la red SKC.

Como se mencionó antes, para que los buscadores puedan encontrar las referencias de los nodos en la Web, éstas deben publicarse de forma adecuada: en una ubicación rastreada por los indexadores de los buscadores de la Web, con acceso autorizado para los robots encargados de ello y con información que pueda tener interés para un usuario humano. Aunque este último no es nuestro objetivo principal, es compatible con nuestros intereses experimentales. En el prototipo, a cada servidor SKC se le dota de una página especial (ver Figura 4-13) con reseñas de todos los nodos activos que soporta, incluyendo una descripción de los mismos, un enlace para acceder a ellos, y otro enlace, fácil de identificar automáticamente, para acceder a la información que los agentes necesitan. Con ello, dichas páginas combinan datos de interés para los usuarios humanos con información destinada a los agentes del sistema.

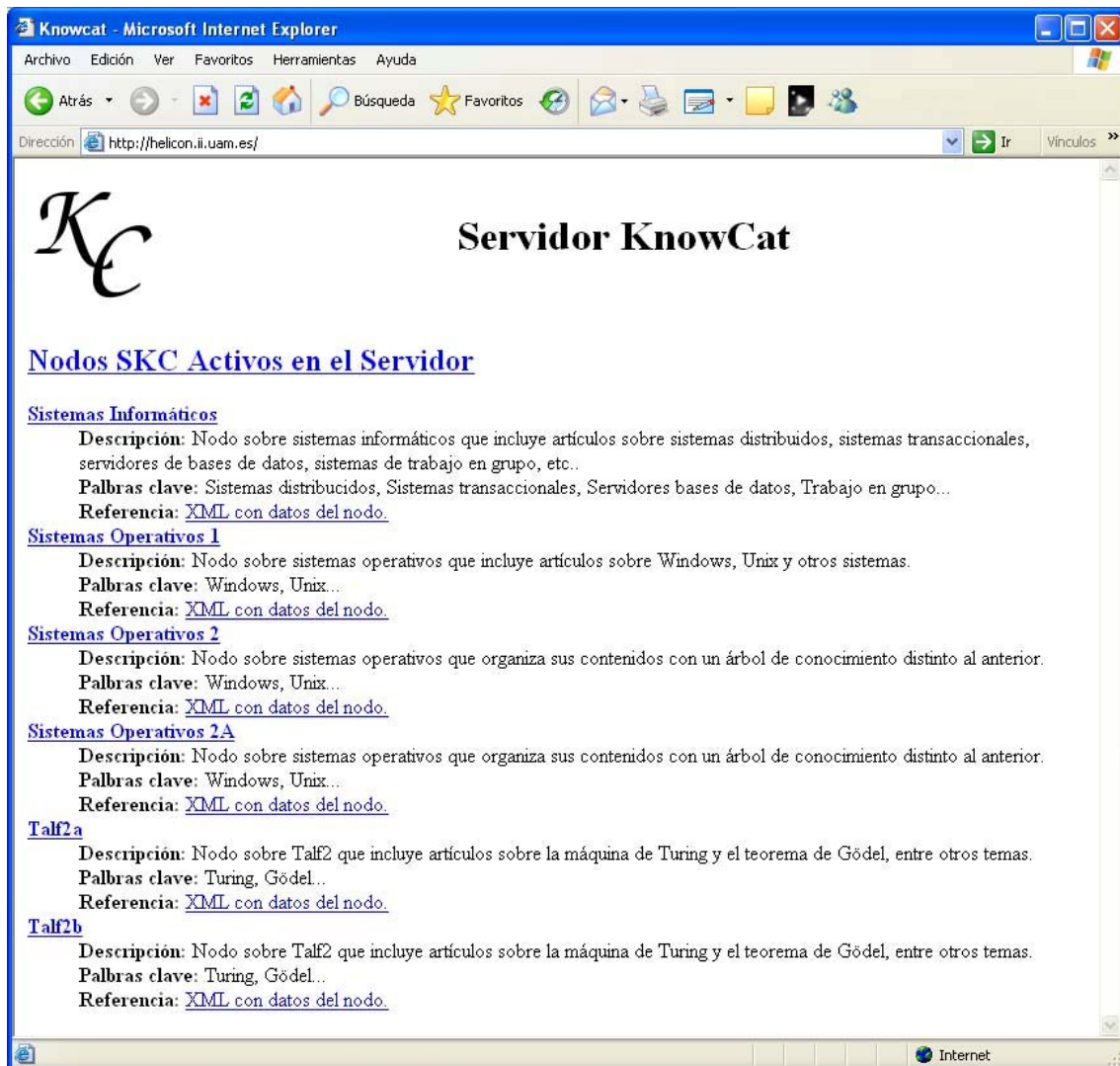


Figura 4-13. Página de nodos SKC en el servidor para su descubrimiento en la Web.

#### 4.2.3.5 Vinculación entre nodos

Mediante el MR los nodos SKC pueden vincularse de dos formas. En ambos casos, los nodos deben mantener contacto para intercambiar información sobre los nodos activos que conocen y que constituyen en cada momento la red SKC. El conjunto de nodos activos puede cambiar con el paso del tiempo, a medida que se crean nuevos nodos y otros desaparecen. Con la información intercambiada de este modo, todos los nodos de la red SKC tienen una imagen relativamente actualizada de los nodos activos que la integran en cada momento. En la vinculación entre nodos más básica sólo se intercambia información de este tipo. Cuando los nodos tratan de temas parecidos, además de la información anterior, intercambian datos sobre los contenidos de sus bases de conocimiento. Con esta información los nodos vinculados pueden establecer relaciones entre los ítems de sus respectivos repositorios. Esta es la

modalidad de vinculación más fuerte que se considera en el prototipo desarrollado. Como los nodos son sistemas en continua evolución, en los que los contenidos van cambiando, es posible que los tipos de vinculaciones entre nodo se alteren con el paso del tiempo.

Durante el proceso de vinculación, desde el momento que un nodo descubre a otro hasta que decide el tipo de vinculación que tendrá con él, el nuevo nodo pasa por una serie de estados intermedios desde la perspectiva del nodo descubridor. Cuando la identificación se produce a través de un tercero o como resultado de una primera visita del nuevo nodo, sólo se cuenta con una información de contacto mínima sobre él. Tras la primera visita al nuevo nodo, se obtiene el resto de los datos de identificación. Sin embargo, cuando se localiza un nuevo nodo a través de un buscador de la Web, se obtienen todos los datos de identificación directamente.

Una vez que se dispone de toda la información de identificación del nodo, en la siguiente interacción se obtiene la información que sirve para determinar la similitud entre temáticas y establecer el tipo de vinculación entre los nodos. Si las temáticas no se parecen, los nodos establecen una vinculación básica en la que, periódicamente, se intercambia información sobre los nodos activos que conocen y también la información necesaria para revisar la similitud entre ellos y la correspondiente vinculación. Por último, si la vinculación es de contenido, se precisa una interacción adicional, que también debe ser periódica, en la que se obtiene la información sobre los ítems de conocimiento del otro nodo, con el fin de establecer las relaciones entre los ítems propios y ajenos.

Aunque en el prototipo implementado las vinculaciones entre nodos y las relaciones entre ítems tienden a ser simétricas, es posible que en la implementación final esto no sea así, porque los criterios de similitud sean distintos en cada nodo.

#### ***4.2.3.6 Intercambio de conocimiento***

Como se ha visto, para establecer el tipo de vinculación entre los nodos y las relaciones entre sus ítems de conocimiento es necesario comparar los repositorios de conocimiento correspondientes. Sin embargo, no es conveniente que los nodos intercambien sus repositorios enteros para poder hacerlo, debido al gran tamaño que estos suelen tener, los recursos y el tiempo necesarios para procesarlos y los tratamientos reiterativos que podrían provocarse.

En el prototipo desarrollado se utilizan descriptores que representan a los ítems de conocimiento de los repositorios, dichos descriptores están concebidos especialmente como instrumentos de comparación y suelen tener un tamaño menor que el ítem al que se refieren. En concreto, para los experimentos realizados se han utilizado descriptores Vectoriales de Peso de Palabras, obtenidos a partir de textos asociados a los ítems de conocimiento que representan. Estos descriptores se crean previamente en los nodos que contiene los ítems a los que se

refieren, pueden enviarse con facilidad por la red, y permiten la comparación de los elementos que califican con un esfuerzo razonable. De sus características, elaboración y empleo en el sistema SKC se habla en el capítulo de Estado de la Cuestión y en otros apartados de este capítulo.

#### 4.2.3.7 Aprovechamiento de la red de nodos

Como resultado de la integración de los nodos en la red SKC, cada uno de ellos dispone de un registro de los nodos que hay en la misma. Con los datos registrados los nodos pueden establecer la similitud entre las temáticas de los distintos nodos y determinar aquellas que son afines con las suyas. Con esta información es posible trazar un mapa de la red SKC en el que se muestren los nodos y los vínculos que los ligan. En el prototipo realizado se proporciona una vista de la red en forma de grafo interactivo, donde aparecen todos los nodos registrados con una indicación visual de sus vinculaciones temáticas. Pulsando sobre las representaciones de los nodos en el grafo se puede navegar por la red SKC, accediendo a los correspondientes nodos.

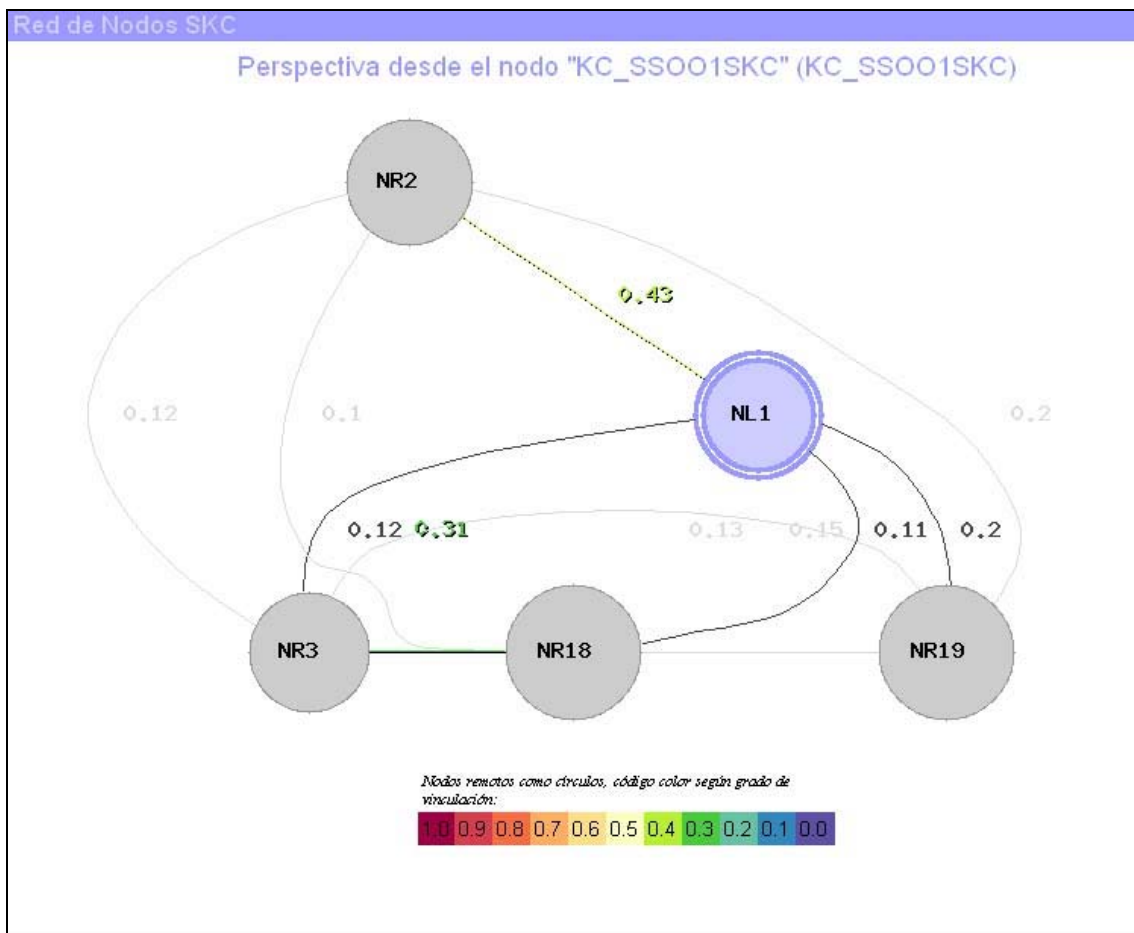


Figura 4-14. Mapa de red de nodos SKC del prototipo.

Además, cada nodo cuenta con registros de los ítems de conocimientos de los nodos con temáticas similares a la suya. Con tales datos los nodos pueden establecer el grado de similitud

entre sus ítems de conocimiento y los ítems de los nodos afines. En el prototipo preparado, esta información sirve para integrar los nodos de la red en el visor de relaciones entre ítems de conocimiento y en el servicio de recomendación, que se prototiparon anteriormente y se presentó en un apartado anterior de este capítulo.

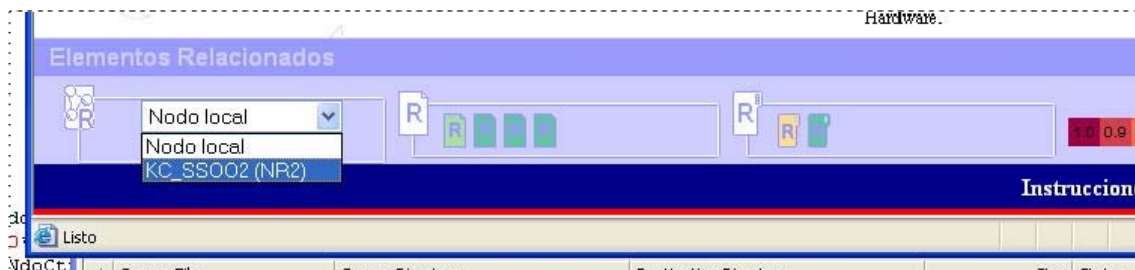


Figura 4-15. Integración de nodos de red SKC en marco de recomendación del prototipo del sistema.

Esta integración pone de manifiesto cómo las referencias de los ítems de los nodos afines a un dado se incorporan a su repositorio de conocimiento como ítems de las nuevas categorías “remotas” -documentos, temas, nodos y autores remotos-, que se manejan de forma similar y compatible con los ítems locales ya considerados previamente.

### **4.3 Campo Experimental: Proyectos de Innovación Docente con KC**

En la sección anterior, Prototipo de SKC sobre KC, se han presentado los tres módulos de Semantic KnowCat (SKC) que se han prototipado sobre la plataforma KnowCat (KC). En la siguiente sección se muestran varios experimentos diseñados para probar dichos módulos y las propuestas subyacentes. Tales experiencias se han incorporado en diversas actividades docentes realizadas con KC en el ámbito de dos Proyectos de Innovación Docente (PIDs) financiados por la Universidad Autónoma de Madrid (UAM) durante cinco años académicos, entre los cursos 2002/2003 y 2006/2007. En esta sección se habla de dichos PIDs que han supuesto un marco idóneo para el trabajo de campo necesario para este proyecto de investigación y que han permitido obtener datos esenciales para la realización del mismo.

Los PIDs de la UAM nacieron con la pretensión de promover una actualización de los procedimientos docentes tradicionales y han servido para fomentar la adecuación de los mismos a los requerimientos del proceso de convergencia hacia el Espacio Europeo de Educación Superior (EEES) en el que está inmersa la universidad europea desde hace algunos años. KC ha demostrado su aptitud como instrumento para dar protagonismo a los alumnos en el proceso formativo, facilitar el aprendizaje activo, el trabajo en grupo, así como el seguimiento y la tutela de tal proceso.

KC se ha utilizando de forma experimental en la docencia de varias asignaturas de diversas titulaciones desde 1998, cuando se puso a disposición de la comunidad docente de la UAM a través del Centro de Aprendizaje e Instrucción (CAI) de dicha universidad. A partir del curso 2002/2003 el sistema se empieza a emplear en el marco de los PDIs, primero durante un año en uno titulado “Plataforma KnowCat para la gestión colaborativa de materiales docentes en red” y después durante el resto del tiempo en otro titulado “Aprendizaje activo y tutelado en la generación colaborativa de materiales docentes en la Web con la asistencia del sistema KnowCat”. Durante todo este tiempo se han ido incorporando asignaturas a los proyectos, que luego han seguido utilizando el sistema año tras año.

El empleo continuado de KC, además de servir para proporcionar un valioso servicio a la comunidad académica, ha permitido la consolidación del sistema, su desarrollo y la investigación en nuevas propuestas basadas en sus postulados, como los trabajos relacionados con SKC, que han permitido elaborar los trabajos presentados aquí y han servido para mejorar el propio sistema KC en varios aspectos. Todo esto no hubiese sido posible sin la financiación de los PDIs, que han hecho posible disponer del equipamiento imprescindible y de los recursos humanos necesarios para mantener el sistema y el servicio.

### **4.3.1 Objetivos generales de los PDIs**

Los objetivos generales de los PDIs han ido evolucionando a lo largo del tiempo, pero podrían enunciarse de la siguiente forma:

1. Crear material docente de calidad, accesible a través de la Web, que sirva de apoyo para la docencia presencial. El material elaborado puede ser apuntes basados en los contenidos impartidos en clase, ampliaciones de tales contenidos, aplicaciones prácticas de los mismos u otros materiales relacionados con ellos.
2. Fomentar el trabajo en grupo de los estudiantes. El sistema proporciona un soporte para el trabajo en equipo y para la colaboración entre los alumnos con el fin de obtener conocimiento de calidad y útil para las asignaturas donde se emplea.
3. Potenciar el aprendizaje activo de los estudiantes mediante del empleo de Tecnologías de la Información y las Comunicaciones (TIC). Con el sistema los alumnos no son meros receptores de conocimiento, si no que participan activamente en su creación y selección. El empleo de las TIC proporciona un soporte muy adecuado para tal proceso.
4. Facilitar la evaluación continua y el seguimiento de la actividad de los alumnos. El sistema permite el registro de la actividad de los alumnos de forma sencilla y la obtención datos útiles para supervisarlos y evaluarlos.



5. Proporcionar nuevos procedimientos de evaluación de los conocimientos y las competencias de los alumnos. El sistema permite que los estudiantes evalúen el trabajo de sus compañeros mediante votaciones, comentarios o sugerencias. La información obtenida permite la evaluación de los alumnos y de la actividad del grupo desde nuevas perspectivas.
6. Mejorar el proceso de enseñanza y aprendizaje. El hecho de que el sistema se utilice por profesores y alumnos, permite la retroalimentación en el proceso de transmisión de conocimiento y abre la posibilidad de aplicarla en la mejora del mismo.

Estos objetivos están en línea con algunos de los requerimientos de la convergencia hacia el EEES. En primer lugar, el sistema favorece la participación activa de los alumnos en el proceso de aprendizaje, puesto que facilita el trabajo en grupo y la realización de actividades y trabajos supervisados y de seminarios. En segundo lugar, el entorno permite el seguimiento de la actividad de los estudiantes y la interacción directa con ellos, que constituyen dos aspectos fundamentales de la actividad de tutoría. En tercer lugar, el sistema proporciona nuevos mecanismos de evaluación, más allá de la valoración de los conocimientos adquiridos, que toman en consideración otras competencias conseguidas por los alumnos: instrumentales, como las capacidades de análisis, síntesis u organización; interpersonales, como las capacidades de crítica, de trabajo en grupo o de interacción; y sistémicas, como la capacidad de aprender, la autonomía o la preocupación por la calidad. Además, el entorno permite la generación de material didáctico de buena calidad accesible a través de la Web. Por último, como ya se ha dicho, el sistema está basado en las tecnologías de la información y las telecomunicaciones (TIC).

Como resultado, el entorno proporciona un soporte, que resulta apropiado para incorporar el Sistema Europeo de Transferencias de Créditos (ECTS son sus siglas en inglés) a cualquier asignatura en la que se pueda emplear, puesto que permite completar o sustituir clases presenciales, supervisar y tutelar la actividad de los estudiantes, disponer de nuevos criterios de evaluación y aprovechar la actividad de los alumnos para facilitar las labores del profesor requeridas en el nuevo contexto. A lo largo del tiempo de su desarrollo, los PIDs con KC han involucrado a varios centros de la UAM, a un número significativo de asignaturas diversas, a un grupo importante de docentes y a más de mil alumnos cada año.

### **4.3.2 Relación de asignaturas participantes en los PIDs**

En la Tabla 4-1 se resumen los datos más importantes de las asignaturas implicadas en los PIDs con KC a los que se dedica esta sección. Algunas de dichas asignaturas participaron en experiencias con el sistema antes de la vigencia de los proyectos y siguieron participando en el ámbito de los mismos.

Tabla 4-1. Relación de asignaturas participantes en los TIDs

ASIGNATURA	DEPARTAMENTO	FACULTAD O ESCUELA	PROFESORES QUE LA HAN IMPARTIDO
Inteligencia Artificial (Cursos: 2004/2005, 2005/2006, 2006/2007)	Ingeniería Informática	Escuela Politécnica Superior	D. Alberto Suarez González D. Fernando Diez Rubio Dña. Ruth Cobos Pérez
Razonamiento bajo Incertidumbre (Cursos: 1998/1999*, 1999/2000*, 2000/2001*, 2001/2002* y 2002/2003)	Ingeniería Informática	Escuela Politécnica Superior	D. Xavier Alamán
Teoría de Autómatas y Lenguajes Formales II (Cursos: 2002/2003, 2003/2004, 2004/2005, 2005/2006, 2006/2007)	Ingeniería Informática	Escuela Politécnica Superior	D. Xavier Alamán
Matemáticas para la Educación Infantil (Cursos: 2000/2001*, 2001/2002* y 2002/2003)	Matemáticas y Tecnología de la Comunicación	Centro Superior de Estudios Universitarios La Salle	D. Melchor Gómez García
Nuevas Tecnologías aplicadas a la Educación (Cursos: 2005/2006, 2006/2007)	Didáctica y Teoría de la Educación	Facultad de formación del Profesorado y Educación	D. Melchor Gómez García
Sistemas Informáticos II (Cursos: 2005/2006, 2006/2007)	Ingeniería Informática	Escuela Politécnica Superior	D. Antonio Martínez Martínez D. Manuel Alfonso Moreno

ASIGNATURA	DEPARTAMENTO	FACULTAD O ESCUELA	PROFESORES QUE LA HAN IMPARTIDO
Sistemas Operativos I (Cursos: 2004/2005, 2005/2006, 2006/2007)	Ingeniería Informática	Escuela Politécnica Superior	Dña Rosa Maria Carro Salas D. Alvaro Ortigosa Dña. Ana María Gonzalez Marcos
Bases del pensamiento matemático y su didáctica (Curso: 2004/2005)	Didáctica y Teoría de la Educación	Facultad de formación del Profesorado y Educación	D. Melchor Gómez García
Biología del desarrollo (Cursos: 2002/2003, 2003/2004, 2004/2005, 2005/2006, 2006/2007)	Bioquímica	Facultad de Ciencias	D. Roberto Marco D. Miguel Manzanares Fourcade (Cursos: 2004/2005, 2005/2006) D. Juan Jimenez (curso 2006/2007)
Atención a la diversidad Sociocultural (Curso: 2002/2003)	Psicología Evolutiva y de la Educación	Facultad de formación del Profesorado y Educación	Dña. Paloma González Aguado
Educación y desarrollo en diversidad (Curso: 2002/2003)	Psicología Evolutiva y de la Educación	Facultad de formación del Profesorado y Educación	Dña. Paloma González Aguado

Nota: \* Cursos académicos previos al inicio de los PIDs

### 4.3.3 Metodología general de las actividades con el KC en los PIDs

Aunque todas las experiencias realizadas en el marco de los PIDs son distintas, muchas de ellas comparten algunos rasgos que se intentan resumir a continuación. Cada experiencia tiene asociados uno o más nodos KC sobre los que se realizan las actividades, según un calendario establecido al principio de las mismas, trabajando en grupos de estudiantes, que a veces cambian a lo largo de la experiencia. Cada nodo cuenta con un repositorio de conocimiento estructurado jerárquicamente en forma de árbol. La raíz de ese árbol es un tema y las ramas que lo forman temas también. En algunas de esas ramas se colocan los documentos sobre los que se trabaja, normalmente en las ramas finales u hojas del árbol, que corresponden a temas sin subdivisiones en subtemas. Los alumnos suelen asignarse por temas hojas del árbol para trabajar en grupos pequeños, de forma que resulte más fácil la labor en equipo y la interacción entre ellos.

Los repositorios pueden incluir documentos de distintos tipos para las actividades, normalmente pueden clasificarse en las siguientes categorías:

- a) Apuntes o resúmenes de los contenidos impartidos en las clases presenciales. Este tipo de material es de utilidad para alumnos y profesores, los primeros para recordar el contenido de las clases y los segundos para supervisar el seguimiento de las mismas.
- b) Ampliaciones de la información proporcionada en las clases con otra obtenida de diferentes fuentes como libros, revistas o Internet. La información suministrada en las clases presenciales suele requerir una elaboración por parte de los estudiantes para completarla y asimilarla.
- c) Aplicaciones prácticas de los contenidos teóricos dados en clase. La realización de ejercicios prácticos permite a los alumnos asimilar los contenidos teóricos aportados y facilita a los profesores un modo de evaluar la comprensión de la materia por los estudiantes.
- d) Documentos interesantes relacionados con los contenidos vistos en las clases, por ejemplo sobre temas de actualidad, de investigación o de opinión. La principal utilidad de este tipo de material es interesar a los estudiantes por la materia en la que se está trabajando.

Las actividades que los alumnos realizan con los nodos suelen consistir en una combinación de las siguientes tareas: aportar documentos de alguna de las categorías anteriores al repositorio de conocimiento, normalmente referentes a algún tema asignado o elegido; revisar algunos de los documentos proporcionados por los compañeros, generalmente los relativos a ciertos temas establecidos por alumno; y manifestar opiniones sobre los documentos revisados mediante votaciones o calificaciones numéricas, o mediante comentarios textuales, que se recogen en el sistema por medio de anotaciones adjuntas a los documentos.

El prototipo de actividad con el sistema consiste en establecer un árbol de conocimiento, asignar grupos de alumnos a sus temas hojas, hacer que los alumnos elaboren documentos individualmente sobre los temas asignados para luego publicarlos en el sistema, hacer que los estudiantes revisen y manifiesten su opinión sobre los documentos publicados por sus compañeros en los temas asignados, y seleccionar, con ayuda del sistema, los documentos que han tenido más aceptación en cada tema.

Sin embargo, el sistema está abierto a la configuración de actividades diversas. En ocasiones el profesor encargado de la experiencia proporciona documentos de referencia a través del sistema sobre los que los estudiantes tienen que trabajar. Otras veces se asignan diferentes tareas a los distintos grupos de estudiantes que comparten un mismo repositorio de conocimiento. En algunas experiencias los comentarios sobre los documentos elaborados sirven a los autores para

refinar los documentos originales atendiendo a las opiniones ajenas y luego publicar en el sistema nuevas versiones de los mismos. En otros casos los comentarios se utilizan como medio de comunicación entre los miembros de los grupos para elaborar documentos de forma colegiada.

La evaluación de los alumnos participantes en las experiencias suele tomar en consideración la calidad de los documentos aportados. Para ello el profesor encargado de la experiencia acostumbra a considerar la selección realizada por el sistema a partir de la opinión manifestada por los estudiantes y su propia impresión sobre el material publicado. También es frecuente que la evaluación considere la adecuación de las opiniones emitidas por los alumnos sobre los trabajos de sus compañeros, para lo que se suelen analizar los votos emitidos y las anotaciones hechas a los documentos. Asimismo es habitual que la evaluación tenga en cuenta el grado de cumplimiento de las tareas encargadas y el modo en que éstas se han realizado, para lo que se suelen analizar los resultados obtenidos por los alumnos y los registros de actividad de los mismos.

El sistema proporciona informes para facilitar la evaluación de los estudiantes, en ellos se incluyen los parámetros monitorizados de la actividad del usuario y muchos de los datos que el sistema genera para la selección de los documentos. Además, es posible generar informes especiales, más elaborados, analizando los datos disponibles con ayuda del equipo humano de soporte del sistema. En algunos casos se han realizado estudios para dilucidar el intercambio de votos entre los alumnos o investigaciones para identificar copias entre los participantes en las experiencias. Todo ello resulta de gran ayuda para la evaluación de los conocimientos y de las competencias adquiridas por los alumnos en las actividades.

Con la finalidad de que las experiencias se desarrollen de forma adecuada, estimulando la intervención y manteniendo la motivación de los participantes, es habitual realizar reuniones con ellos durante las actividades. Todas las experiencias suelen empezar con una presentación de las mismas y un seminario sobre el funcionamiento básico del sistema. Dependiendo del plan de actividades y de las necesidades que se detecten durante el desarrollo de las experiencias, se establecen otras reuniones destinadas al soporte requerido.

#### **4.3.4 Experiencias de los PIDs empleadas en este trabajo de investigación**

Algunas de las experiencias desarrolladas en el ámbito de los PIDs con KC que nos ocupan, han servido para probar distintos módulos del sistema SKC prototipados sobre KC y las propuestas subyacentes. Las actividades correspondientes se diseñaron con objetivos pedagógicos acordes

con los requerimientos de los PIDs, pero de modo que sirvieran para las pruebas pretendidas. En este apartado se recoge una reseña de las experiencias implicadas.

#### **4.3.4.1 SSOO (2002/2003-2004/2005)**

Sistemas Operativos (SSOO) es una asignatura troncal de primer ciclo de Ingeniería Informática de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid, que se imparte durante el segundo semestre del segundo año. Suele tener una matrícula anual cercana a 400 alumnos, que se reparten en tres grupos. La carga docente semanal de los alumnos es 3 horas de clases teóricas y 2 de clases prácticas. El número total de créditos ECTS estimados para la asignatura es de 6. En el periodo considerado, el trabajo de los alumnos se evaluó al final de cada curso por la calificación obtenida en el examen final de la asignatura, ponderada con una nota de evaluación continua, que consideraba la participación en estas experiencias, la calificación obtenida en las prácticas de la asignatura y la nota alcanzada en un examen parcial que se realizaba a la mitad del periodo docente.

Durante los tres cursos considerados las actividades se han desarrollado sobre distintos nodos KC cada curso, aunque tomando como base los documentos de un nodo creado con el sistema por alumnos de la asignatura en años anteriores. Los documentos del nodo de partida se seleccionaron por su aceptación con ayuda del sistema y se utilizaron junto con los documentos generados cada año para alimentar los nodos de años sucesivos al principio de la actividad.

La actividad propuesta en el curso 2003/2004 puede servir de modelo de todas las demás. Para empezar, se establecieron dos nodos KC sobre SSOO con distintas estructuras de árboles de conocimiento y suficientes temas hojas como para asociar uno a cada grupo de cuatro participantes, y los alumnos de la asignatura se repartieron entre ambos nodos. Después, los estudiantes interesados en participar se tuvieron que dar de alta en su correspondiente nodo, utilizando un identificador que permitiera reconocerlos de forma sencilla, y elegir un tema hoja para trabajar sobre él, respetando el límite de cuatro personas por tema. Durante el siguiente mes, el trabajo los alumnos fue leer cada semana los artículos, de unas cuatro páginas, publicados por el profesor en los temas que habían elegido y en realizar una anotación al documento utilizando el sistema. El contenido de las anotaciones era diferente para cada alumno y semana, según el orden en que se eligió el tema, y podía consistir en la elaboración de una lista de palabras clave, un resumen o una lista de referencias para el documento. A continuación, cada alumno debía confeccionar un nuevo artículo, de no más de cuatro páginas, sobre el tema elegido y entregarlo quince días antes de fin de la experiencia. Durante las últimas dos semanas de la actividad, los alumnos tuvieron que revisar los artículos de sus compañeros de tema, anotarlos dando sus opiniones sobre ellos, y votar a los dos mejores artículos de sus temas,

incluidos los aportados por el profesor y excluidos los propios. Al final de la experiencia, el sistema seleccionó los documentos de cada tema hoja según la aceptación obtenida.

Para la elaboración de los nuevos artículos se pudo utilizar como referencia el material disponible en ambos nodos y el que se pudiera encontrar en otras fuentes, aunque no se admitió la copia literal de documentos o partes de ellos. De forma parecida, para efectuar las nuevas anotaciones no se permitió copiar las ya realizadas.

La evaluación del trabajo de los alumnos se realizó atendiendo a la calidad de los documentos aportados, al grado de participación en la actividad mediante interacción con el sistema, al interés demostrado y a la oportunidad de las anotaciones realizadas, y al grado de participación en las votaciones y otras manifestaciones solicitadas.

Estas experiencias, además de facilitar los objetivos generales de los PIDs, han permitido obtener un importante repositorio de documentos sobre la misma materia clasificados en árboles de conocimiento con distinta estructura, para algunos experimentos con los prototipos de los módulos de SKC de los que después se hablará.

#### **4.3.4.2 TALF2 (2002/2003-2006/2007)**

Teoría de Autómatas y Lenguajes Formales II (TALF2) es una asignatura optativa de segundo ciclo de Ingeniería Informática de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid, que se imparte durante el primer semestre del cuarto curso. Suele tener un solo grupo de entre 80 y 90 alumnos al año. La carga docente semanal de los alumnos es 3 horas de clases teóricas y 2 de clases prácticas. El número total de créditos ECTS estimados para la asignatura es de 5,6. La calificación final del curso es obtenida en un examen final de teoría, ponderada con una nota de evaluación continua, y la calificación de prácticas. Esta calificación de evaluación continua, implica la participación en la actividad con KC.

Durante los cinco cursos considerados, como pasó en las experiencias de SSOO, en cada curso las actividades se han desarrollado sobre distintos nodos del sistema, aunque en todos ellos se han tomado como base los documentos de años anteriores, menos en el primero que se partió de cero.

Las experiencias sobre TALF2 se iniciaron en el curso 2002/2003 creando tres nodos sobre la misma materia, pero con árboles de conocimiento distintos en cada uno de ellos. Se dividió a los alumnos de la signatura entre los tres nodos, y se les propuso que eligieran un tema del árbol de conocimiento correspondiente para realizar un artículo sobre él. Para ello se impuso la condición de que no hubiera más de cuatro trabajos por cada tema, los que consiguió articulando un procedimiento que resolviera los conflictos. Cuando todos los alumnos entregaron el documento sobre el tema asignado, se les pidió que anotaran tres trabajos de los otros dos

nodos, que trataran de temas los más parecidos posible al elegido por cada uno inicialmente, y que votarán el que les pareciera mejor de ellos.

Durante los cursos 2003/2004 y 2004/2005 la actividad se desarrolló en dos fases, utilizando un nodo distinto cada una, pero con árboles de conocimiento iguales. En la primera cada alumno tuvo que elegir un tema para anotar los documentos que el profesor publicó en cinco semanas sucesivas. En cada tema hubo entre siete y diez alumnos. Como en las experiencias de SSOO, el contenido de las anotaciones fue diferente para cada alumno y semana, según el orden en que se eligieron los temas, y pudo consistir en la elaboración de listas de palabras claves, de resúmenes o de listas de referencias para los documentos. Después los estudiantes tuvieron que votar los mejores artículos de cada tema elegido. En la segunda fase se asignaron temas distintos a los alumnos y se repitió la actividad con el otro nodo. Como resultado se obtuvieron manifestaciones de opinión de los usuarios sobre los temas asignados en ambas fases.

En la experiencia del curso 2005/2006 y en la del 2006/2007 se crearon dos nodos para la asignatura con árboles de conocimiento iguales, como el año anterior, que se utilizaron sucesivamente durante el curso. La actividad se inicio en uno de los nodos con el registro de los alumnos participantes y la elección de los temas de trabajo. Durante las cinco semanas siguientes los alumnos tuvieron que trabajar semanalmente leyendo los artículos que el profesor añadía cada semana a los temas elegidos y realizando ciertas anotaciones sobre ellos. De forma parecida a otras veces, el contenido de las anotaciones consistía en la elaboración de una lista de palabras clave, un resumen, una lista de referencias del documento o una opinión personal sobre el mismo, que se asignó semanalmente a cada alumno. Para terminar la actividad con el primer nodo, cada alumno tuvo que elaborar un documento original sobre el tema elegido y ponerlo en el sistema. Durante el resto del curso el trabajo se desarrolló sobre el otro nodo. En este caso, se incluyeron en el árbol de conocimiento los documentos aportados por los alumnos al final de la actividad con el primer nodo, pero no los aportados por el profesor. Además, todos los alumnos que se dieron de alta en el primer nodo aparecieron registrados en el nuevo y se asignó a cada uno un tema diferente al elegido en la primera ocasión para trabajar sobre él. La actividad de cada alumno consistió en leer los artículos incluidos en el tema asignado durante las primeras semanas de trabajo con el nuevo nodo. Durante las últimas semanas de la actividad, cada alumno tuvo que votar los dos mejores artículos del tema asignado, justificando su voto con una anotación al correspondiente documento.

La evaluación del trabajo de los alumnos se realizó considerando calidad de las aportaciones realizadas y el grado de cumplimiento en las tareas asignadas. Con esta experiencia, además de los objetivos generales de los PIDs, se ha conseguido obtener un repositorio de documentos sobre TALEF2 clasificados en árboles de conocimiento con distinta estructura y registros de



actividad de los usuarios, que han servido para algunos de los experimentos realizados con los prototipos de los módulos de SKC que se presentan más adelante.

#### 4.3.4.3 IA (2004/2005)

La asignatura de Inteligencia Artificial es una asignatura troncal de segundo ciclo de Ingeniería Informática de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid. La asignatura tiene una matrícula anual de unos 300 alumnos divididos en tres grupos. La docencia de esta asignatura se compone de clases teóricas y clases prácticas. La carga docente semanal de esta asignatura se compone de 3 horas teóricas y de 2 prácticas. El número total de créditos ECTS estimados para ella es de 6. La calificación final de los alumnos en la asignatura se obtuvo ponderando la nota obtenida en las partes práctica y teórica. La evaluación del trabajo realizado en la experiencia supuso el 15% de la nota de teoría.

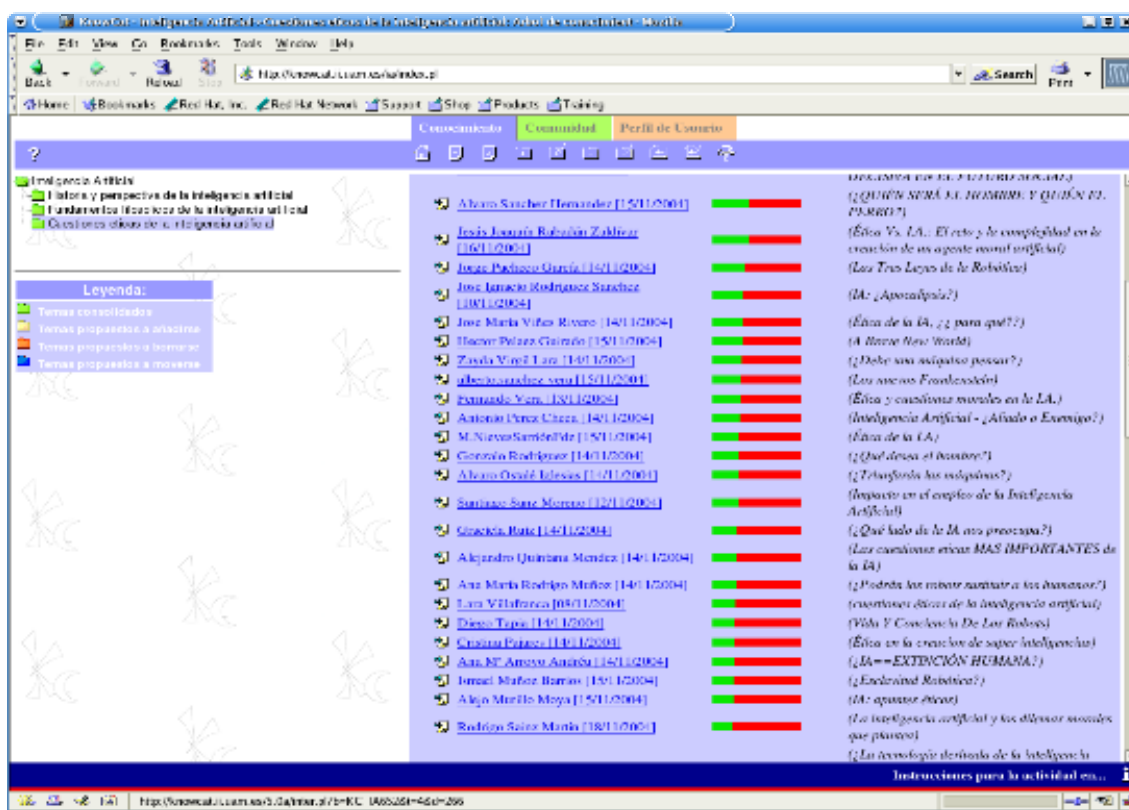


Figura 4-16. Figura 1. Repositorio o área de conocimiento sobre Inteligencia Artificial.

El objetivo de la experiencia [Cobos 2005] fue generar un repositorio de trabajos realizados por los alumnos sobre temas relacionados con la asignatura, fuera del temario y que supusieran una ampliación de los contenidos vistos en clase. El repositorio obtenido se generó y evaluó completamente por los alumnos, aunque bajo la vigilancia de los profesores de la asignatura, que se ocuparon de supervisar la forma de realizar la actividad y comprobar la corrección del

contenido de los trabajos. La labor de clasificación y evaluación de cada uno de los trabajos fue realizada entre los propios alumnos mediante el uso de votaciones y anotaciones.

Esta experiencia ha servido para realizar alguno de los experimentos con los prototipos de SKC sobre KC de la siguiente sección, gracias al repositorio de documentos sobre la asignatura y el registro de actividad obtenido de los usuarios, además de alcanzar los objetivos generales de los PIDs.

#### **4.3.4.4 SI2 (2004/2005,2005/2006)**

Sistemas Informáticos II (SI2) es una asignatura troncal de segundo ciclo de Ingeniería Informática de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid, que se imparte durante el segundo semestre de cuarto curso y que suele tener una matrícula anual de unos 300 alumnos. La carga docente semanal de los alumnos es 3 horas de clases teóricas y 2 de clases prácticas. El número total de créditos ECTS estimados para la asignatura es de 6.

Durante los dos cursos considerados la actividad ha sido esencialmente la misma. Cada año se propuso el desarrollo de 45 temas distintos de los 115 que constituyen la asignatura, partiendo de la información proporcionada en clase en forma de esquemas y de las referencias suministradas. Al comienzo de los cursos, se asignaron dos temas a cada alumno de forma aleatoria y de modo que no hubiera menos de cuatro alumnos por tema. Durante cuatro semanas cada alumno tuvo que elaborar y publicar un artículo sobre los temas establecidos. A continuación, se asignaron otros tres temas por alumno del mismo modo. Los estudiantes dispusieron de seis semanas para revisar los artículos incluidos en los temas establecidos, realizar anotaciones de ampliación o rectificación sobre ellos y valorar los mismos con una calificación entre cero y diez. Finalmente, los documentos se ordenaron en cada tema por calidad según las calificaciones obtenidas y quedaron a disposición de todo el grupo para la preparación del examen de la asignatura.

Como en otras ocasiones, la evaluación del trabajo de los alumnos consideró las contribuciones efectuadas y el cumplimiento en las tareas encomendadas. Para ello se partió de la valoración realizada por los alumnos de los artículos de sus compañeros a través del sistema y de la opinión de los profesores sobre el trabajo realizado. Como resultado los participantes recibieron hasta un punto adicional en la nota final de la asignatura, junto con el material de estudio mencionado.

Con las dos experiencias se han obtenido, además de los objetivos generales de los PIDs, un repositorio de artículos sobre la práctica totalidad del temario de SI2 y registros de la actividad de los alumnos, que han permitido realizar varios de los experimentos con los prototipos de los módulos de SKC presentados a continuación.

#### **4.4 Experimentos realizados y resultados obtenidos**

Los tres módulos de SKC prototipados sobre KC, de los que se hablo en otra sección de este capítulo, han servido para realizar varios experimentos diseñados para probar dichos componentes y las propuestas subyacentes, en el ámbito de los proyectos de innovación docente presentados en la sección anterior. Los experimentos llevados a cabo se han dividido en tres grupos.

El primer grupo de experimentos emplea el prototipo del Monitor de Cliente (MC) y trata de la medida de la interacción de los usuarios en el lado del cliente, dado que SKC es una aplicación cliente/servidor y que disponer de información en ese lado supone una ventaja sobre el registro tradicional de este tipo de sistemas. Con ello se indaga en aspectos como la intensidad de la interacción de los usuarios, que permite tener una impresión sobre el interés de los usuarios, sin necesidad de manifestaciones explícitas sobre él, para conseguir una selección tentativa del conocimiento sin requerimientos de expresiones directas de opinión sobre el mismo.

El segundo grupo de experimentos realizados utiliza el prototipo del Módulo de Análisis de Contenido (MAC) e investiga la calificación semántica de ítems de conocimiento, con el fin de establecer asociaciones automáticas entre dichos ítems, realizar recomendación sobre los mismos y proporcionar nuevas vistas del conocimiento disponible en el sistema. En concreto se han realizado experimentos de agrupación y clasificación automáticas de ítems de conocimiento, otros de mapeo entre árboles de conocimiento -representativos de ontologías subyacentes- y otros sobre asociación automática entre nodos de conocimiento.

El tercer grupo de experimentos llevados a cabo usa de forma combinada el prototipo del MC y el del MAC y se ocupan, más en detalle, del interés de los usuarios, en concreto de su descubrimiento, de su representación y de su aprovechamiento. En particular se han realizado experimentos de identificación del interés de los usuarios y de identificación de contenidos interesantes para los usuarios dentro y fuera del sistema.

Finalmente, el cuarto grupo de experimentos efectuados emplean el prototipo del Módulo de Red y el del MAC, y se dedica al establecimiento y mantenimiento de una red de instancias del sistema SKC para ampliar su proyección, enriquecer su conocimiento y estimular a sus usuarios. Para ello se indaga en el uso de agentes para el descubrimiento y la vinculación entre instancias, y para establecimiento y el mantenimiento de una red entre ellas. En concreto se han realizado experimentos de vinculación automática entre nodos y de descubrimiento de nodos SKC en la Web

#### **4.4.1 Medida de la interacción de los usuarios en el lado del cliente: intensidad de la interacción de los usuarios, interés y aproximación a la selección del conocimiento sin manifestación explícita de opinión**

Como se ha visto al principio de este capítulo, para paliar la falta de información sobre la actividad de los usuarios en el lado cliente, característica de los sistemas Web, Semantic KnowCat (SKC) dispone de un Monitor de Cliente (MC) encargado de tomar nota de la actividad de los usuarios, más allá de la observada habitualmente desde el lado servidor. El MC se ocupa también de analizar el registro de actividad obtenido, para conseguir nueva información que facilite la selección del conocimiento y que reduzca la necesidad de manifestaciones explícitas de opinión sobre el mismo por parte de los usuarios.

Para probar la utilidad de esta aproximación y la viabilidad de su implementación, se ha desarrollado el prototipo de MC de SKC sobre un sistema KnowCat (KC), descrito en la primera sección de este capítulo, para utilizarlo en dos experiencias reales de carácter docente en la Escuela Politécnica Superior de la Universidad Autónoma de Madrid durante el curso 2004-2005, en el marco de los Proyectos de Innovación Docente presentados en la sección anterior. En los dos casos, se han establecido actividades diferentes a través del sistema, intentando provocar distintos modos de interacción con el conocimiento incorporado en el sistema. Todos los alumnos eran conscientes de que su actividad en el cliente estaba siendo monitorizada, aunque en ningún caso se les pidió que se identificaran, sino que se les asignaron identificadores únicos automáticos en cada sesión.

Como se explica más detalladamente en [Cobos 2002a] y en [Cobos 2003], el sistema KC incluye un mecanismo selección del conocimiento, al que se denomina cristalización, que permite clasificar un conjunto de documentos por orden de relevancia, en base a información de accesos y votaciones explícitas de los usuarios. Este mecanismo ha sido suficientemente validado [Gómez 2001][Cobos 2003] [Cobos 2004] [Díez 2007], por lo que ahora lo hemos empleado para comprobar el desempeño del nuevo mecanismo de clasificación que proponemos. Así, una vez realizada la cristalización por los mecanismos de KC, se han procesado los ficheros LOG del servidor Web para establecer una clasificación de los documentos en función de la intensidad de interacción que se había mantenido con ellos en el cliente. Para el cálculo de la relevancia de cada documento se consideró: el número de ciclos activos registrados por el MC, los periodos en los que se había detectado actividad sobre cada documento, y la intensidad de la misma, en función de la cantidad de eventos de cada tipo detectados durante esos periodos. Todo ello, según se ha explicado en el apartado anterior.

Finalmente, se ha hecho una comparación de la clasificación obtenida por este procedimiento frente a la cristalización clásica del sistema KC.

#### **4.4.1.1 Experiencia 1: “Inteligencia artificial”**

La primera experiencia (E11) se ha desarrollado con casi 300 alumnos de la asignatura “Inteligencia Artificial” (IA), troncal del tercer curso de Ingeniería Informática. Se trató de un trabajo voluntario que podía suponer hasta un 15% de la nota de teoría de la asignatura. El resultado puede verse en la URL “<http://knowcat.ii.uam.es/ia/index.pl>” y más en detalle en [Cobos 2005].

Los alumnos se repartieron en tres grupos, cada uno dedicado a uno de los tres temas propuestos por los profesores. Cada alumno contribuyó con un documento sobre el tema de su grupo antes de una fecha límite. Posteriormente tuvieron un mes para emitir valoraciones personales sobre el trabajo de los compañeros del grupo. Después de esto se realizó una cristalización para seleccionar el mejor documento de cada tema. Sobre los contenidos de cada uno de estos documentos se incluyó una pregunta en el examen final de la asignatura. Los resultados obtenidos en esta experiencia fueron muy satisfactorios, puesto que la mayoría de los alumnos obtuvo buenos resultados en las cuestiones sobre los documentos seleccionados mediante el sistema.

#### **4.4.1.2 Experiencia 2: “Sistemas operativos”**

Esta segunda experiencia (E12) consiste en una actividad con 110 alumnos de la asignatura “Sistemas Operativos” (SO), que involucra los nodos “[http://knowcat.ii.uam.es/so1\\_05\\_21/index.pl](http://knowcat.ii.uam.es/so1_05_21/index.pl)”, “[http://knowcat.ii.uam.es/so1\\_05\\_22/index.pl](http://knowcat.ii.uam.es/so1_05_22/index.pl)” y “[http://knowcat.ii.uam.es/so1\\_05\\_26/index.pl](http://knowcat.ii.uam.es/so1_05_26/index.pl)”. En ella los alumnos se han repartido en tres grupos dedicados a dos o tres temas, según el número de estudiantes asignados. Cada alumno ha elegido uno de los temas de su grupo, sobre el que ha tenido que preparar un trabajo para poner en el sistema. A continuación todos los alumnos han tenido que valorar los demás trabajos aportados a sus respectivos temas, mediante votaciones y anotaciones. Después de todo esto se ha realizado una cristalización del conocimiento, que ha producido una clasificación de los documentos. Esta clasificación permitirá establecer los trabajos que se expondrán en clase durante el curso.

#### **4.4.1.3 Resultados obtenidos**

La primera de las experiencias, la de IA, es la más interesante de las dos por su dimensión en número de usuarios y documentos, y por la motivación que se ha conseguido con su

planteamiento, que se ha manifestado en la calidad de la cristalización conseguida y en los resultados obtenidos por los alumnos en base al material seleccionado.

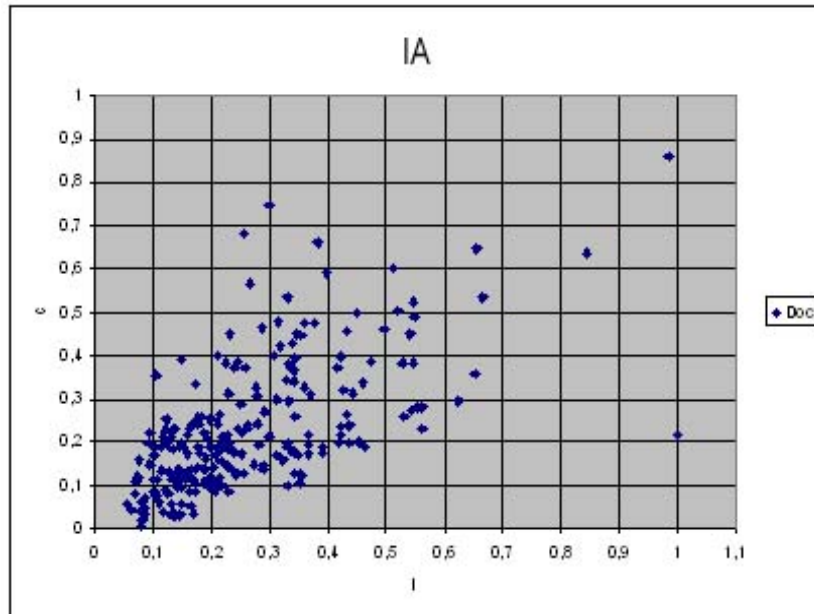


Figura 4-17. C frente a I en la experiencia de IA

El siguiente gráfico presenta en el eje de ordenadas el grado de cristalización (C) de los documentos que aparecen representados como puntos en el gráfico. En el eje de abscisas se representa la intensidad de la interacción (I) en el cliente para cada documento.

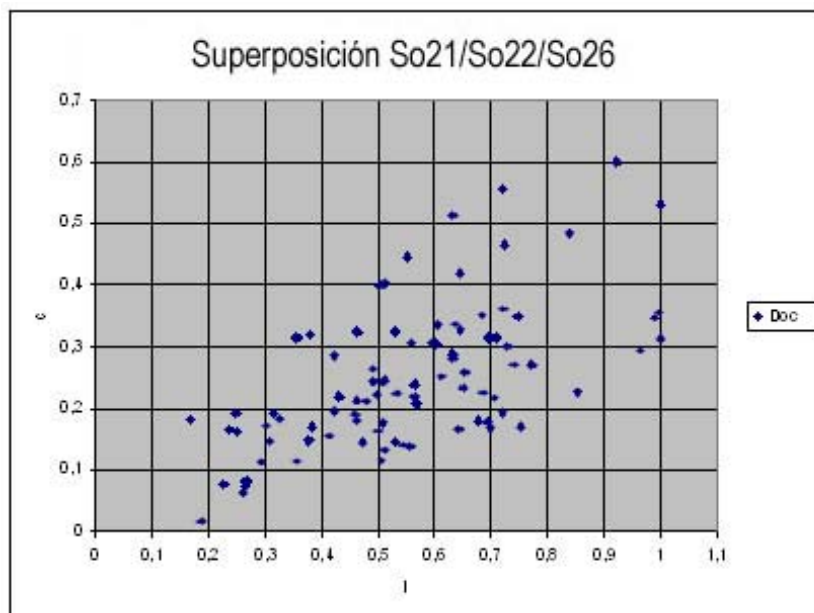


Figura 4-18. C frente a I en la experiencia de SO

Es evidente la correlación entre los dos parámetros, que se pone de manifiesto en la concentración de los puntos siguiendo la diagonal. Es significativa la coincidencia de documentos con altos grados C que manifiestan también altos valores de I, puntos en el

cuadrante superior derecho. Del mismo modo, es interesante comprobar que los documentos con bajos valores de ambos parámetros son muchos, situados en el cuadrante inferior izquierdo.

Algunos documentos, sin embargo, mantienen índices C bajos a pesar de haber sido utilizados activamente, se encuentran en el cuadrante inferior derecho. Se trata de una posibilidad razonable en documentos que, a pesar de suscitar interés, no son apreciados finalmente. Los casos que no parecen tener una explicación tan lógica son los que, manteniendo índices de intensidad de interacción bajos, sin embargo tienen grados de cristalización elevados, y están en el cuadrante superior izquierdo. Se trata de documentos que, a pesar de no haber sido suficientemente consultados y trabajados a través del sistema, han recibido muestras de aprecio suficientes para cristalizar. Llevado a ciertos límites puede poner de manifiesto votaciones arbitrarias o revisiones del material al margen de la interfaz del sistema. Para ilustrar la experiencia de SO, el siguiente gráfico superpone los resultados de los tres nodos utilizados. De nuevo, como en el caso de IA, se manifiesta la correlación entre los dos índices. En esta ocasión, además, la concentración de documentos en el cuadrante superior derecho y en el inferior izquierdo es muy significativa, lo que indica que la mayoría de los documentos tienen ambos índices altos o bajos a la vez. También es interesante observar que no aparece ningún documento en el cuadrante superior izquierdo.

#### **4.4.1.4 Conclusiones**

Un índice basado en la intensidad de la interacción de los usuarios en el cliente parece mostrar el interés de los usuarios por los elementos de conocimiento gestionados por un sistema como KnowCat (KC) o Semantic KnowCat (SKC). Un índice de estas características puede servir para matizar y detectar anomalías en un mecanismo de cristalización al estilo del de KC. Al mismo tiempo, un índice así podría servir para seleccionar el contenido sin necesidad de manifestaciones explícitas de opinión sobre el mismo o reduciendo significativamente la necesidad de éstas, como se pretende en SKC. Sin embargo, la monitorización en el cliente sólo es posible en condiciones determinadas y bajo restricciones importantes que no siempre se pueden asumir, superar o imponer. Además, la motivación de los usuarios parece una condición imprescindible para la manifestación adecuada del interés, como también lo es para la realización de la cristalización en KC.

#### **4.4.2 Calificación semántica de ítems de conocimiento: Asociación automática entre ítems de conocimiento, recomendación y nuevas vistas del conocimiento**

Como se ha visto en la primera sección de este capítulo, para enriquecer el conocimiento que maneja y facilitar su gestión, SKC dispone de un Módulo de Análisis de Conocimiento (MAC) encargado de procesar el conocimiento explícito que existe en el sistema para asimilarlo, con el objetivo de revelar conocimiento implícito que no se aprovechaba y que podría resultar de utilidad para mejorar la interacción de los usuarios con el sistema y el conocimiento. Con el fin de comprobar la viabilidad de las aproximaciones propuestas, hemos desarrollado un prototipo de los tres elementos del sistema SKC: un MAC, un Visor de Conocimiento (VC) para visualizar los grafos de relaciones entre ítems de conocimiento, y un Servicio de Recomendación (SR) contextual. Todos se han incorporado a un sistema KnowCat (KC).

El prototipo ha permitido realizar varios experimentos con nodos KC, preparados para ello durante varios cursos, en actividades docentes realizadas en la Escuela Politécnica Superior de la Universidad Autónoma de Madrid en el marco de los Proyectos de Innovación Docente presentados en la sección anterior. En concreto, se han utilizado cuatro nodos KC: uno sobre Sistemas Operativos (SSOO); dos sobre Teoría de Autómatas y Lenguajes Formales II (TALF2); y otro sobre Sistemas Informáticos (SI).

El nodo sobre SSOO es el resultado de la elaboración de un temario sobre dicha materia realizado por los alumnos durante cuatro años consecutivos, y consiste en un árbol de conocimiento de dos niveles de profundidad con más de 20 temas y 350 documentos.

Los nodos de TALF2 organizan en dos árboles de conocimiento distintos documentos aportados por alumnos durante un curso. Los dos nodos tratan sobre la misma materia, pero en cada uno los documentos y la estructuración del temario son diferentes. Ambos árboles tienen dos niveles, un nodo con 6 temas y 24 documentos y el otro con 12 temas y 50 documentos.

Por último, en el nodo de SI se ha desarrollado un temario sobre el correspondiente área de conocimiento, de manera que durante un curso los alumnos de una asignatura han aportando más 180 documentos sobre unos 40 temas diferentes estructurados en un árbol de conocimiento.

Los experimentos realizados se han dirigido a comprobar la viabilidad de la asociación automática de ítems de conocimiento empleando los vectores de peso de palabras asignados por los procedimientos propuestos. Para ello se han realizado tres grupos de experimentos que se describen en los siguientes apartados.



#### 4.4.2.1 Experimentos de agrupación y clasificación automática de ítems de conocimiento

Para probar la agrupación y clasificación automática de ítems de conocimiento se han realizado tres experimentos.

El primero de ellos (E211) toma como partida el nodo KC sobre SSOO, donde se han utilizado los dos documentos con más aceptación de cada tema para establecer el vector de peso de palabras (VPP) de cada uno de ellos. A continuación se han establecido los grados de similitud entre los VPP de los documentos restantes y los VPP de los temas mencionados. Con estos datos se ha obtenido un gráfico (ver Figura 4-19), donde cada fila de puntos corresponde a un tema y cada columna a un documento. Los grados de similitud entre documentos y temas se han representado por puntos de color, más claros cuanto más alto es el valor del coeficiente. Los documentos se han ordenado por los temas en que los alumnos los habían clasificado inicialmente de forma manual.

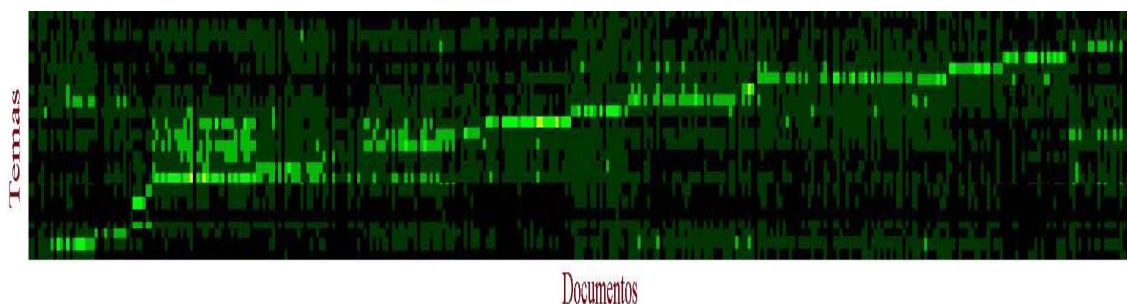


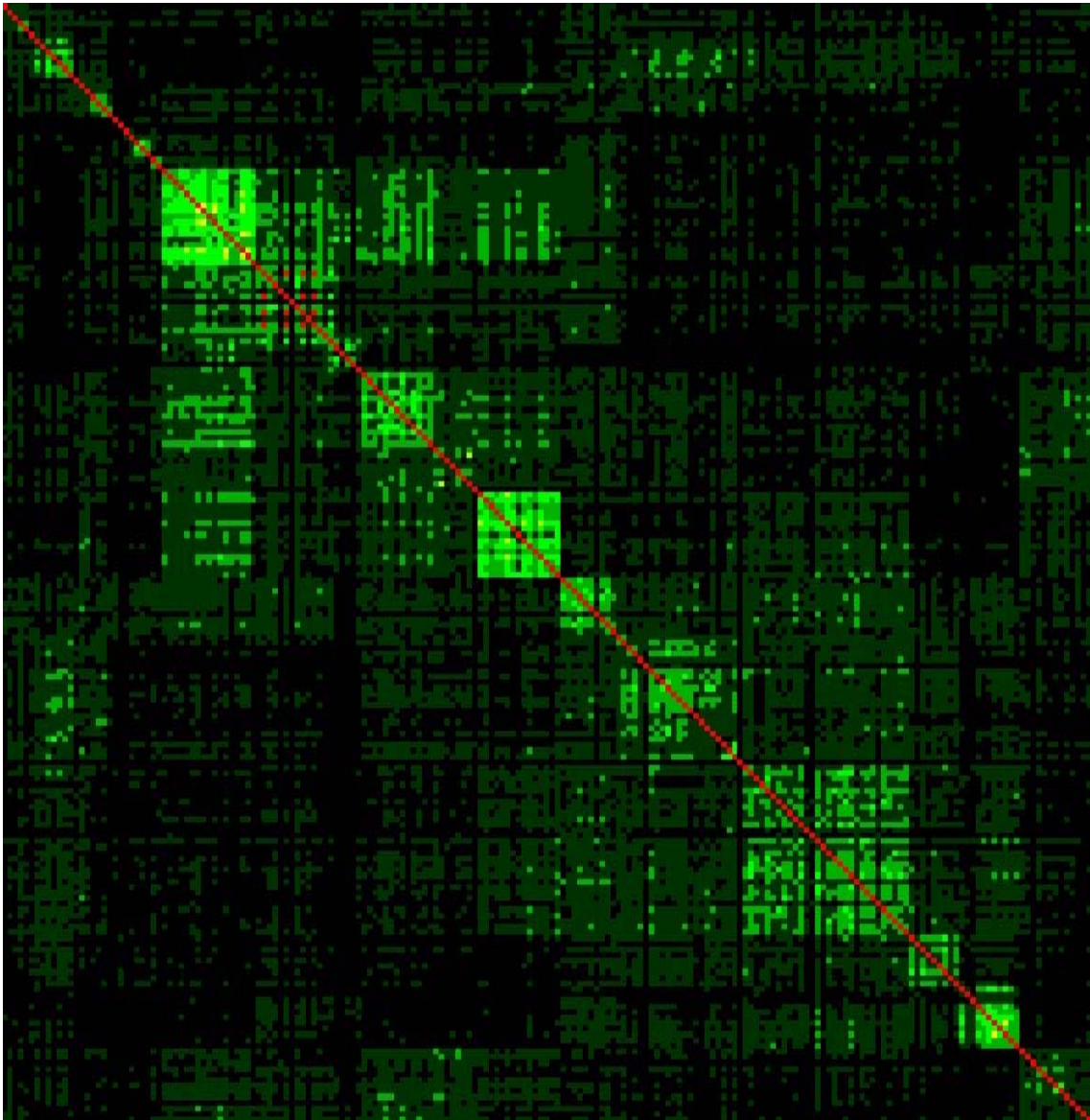
Figura 4-19. Clasificación automática de documentos por temas en nodo KC sobre SSOO

El segundo experimento (E212) compara los VPP de los documentos del nodo original de SSOO entre sí. En el gráfico (ver Figura 4-20) cada línea de puntos -vertical y horizontal- corresponde a un documento y estos aparecen ordenados por los temas en los que fueron clasificados manualmente. Como en el caso anterior, los grados de similitud entre elementos se han representado por puntos de color, más claros cuanto más alto es el coeficiente.

Como resultado del primer experimento, en el gráfico se puede ver cómo los valores más altos de similitud -puntos más claros- se alinean mayoritariamente en las filas que corresponden a los temas en los que se les clasificó manualmente. Esto indica que la clasificación automática coincide en la mayoría de los casos con la manual. Analizando a posteriori las anomalías se observa que corresponden a temas ambiguos en los que se habían clasificado documentos heterogéneos.

Como se puede observar en el gráfico de este segundo experimento, los mayores grados de similitud aparecen agrupados en torno a la diagonal formando bloques. En las condiciones expuestas, esto indica que la mayoría de los documentos son más parecidos entre sí cuando

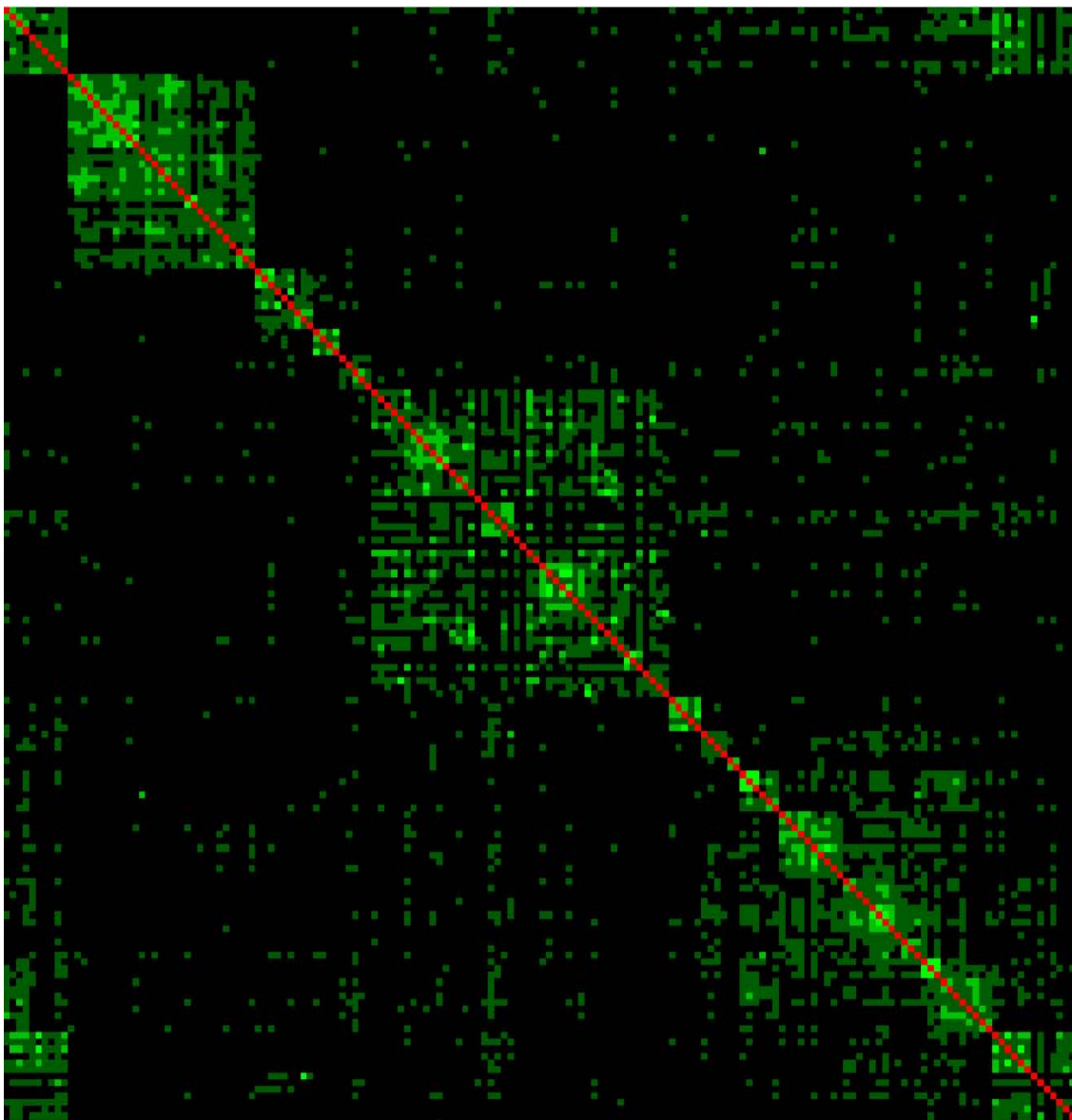
tratan de los mismos temas. Sin embargo, en este caso es interesante observar cómo los puntos claros fuera de los grupos de la diagonal aparecen en bandas, que muestran cómo se establecen relaciones entre documentos de temas similares.



Las filas y columnas corresponden a los mismos documentos ordenados por los temas de la clasificación manual inicial

Figura 4-20. Agrupación automática de documentos por temática de área de conocimiento, en un nodo de SSOO

En el tercer experimento realizado (E213), los resultados son similares al anterior cuando se comparan, en las mismas condiciones, los VPP de los documentos de un nodo KC sobre SI, como se puede apreciar en el correspondiente gráfico (ver Figura 4-21). En este caso, al ser mayor el número de temas y menor el de documentos por tema, las agrupaciones son bloques claros más pequeños.



Las filas y columnas corresponden a los mismos documentos ordenados por los temas de la clasificación manual inicial

Figura 4-21. Agrupación automática de documentos por temática de área de conocimiento en nodo de SI

#### 4.4.2.2 Experimentos de mapeo entre árboles de conocimiento

Para probar el mapeo entre árboles de conocimiento, que representan las ontologías que se consideran en cada nodo del sistema, se han realizado dos experimentos.

En el primer experimento (E221) de esta serie, de nuevo, se toma como partida el nodo KC sobre SSOO, aunque en este caso se han creado a partir de él dos nuevos nodos con los árboles de conocimiento iguales al original. Se han repartido los documentos del nodo inicial entre los recién creados, de manera que cada par de temas homólogos de los nuevos quede con un número similar de documentos diferentes, pero de parecida relevancia. Después se han calculado los VPP de los temas de los nodos nuevos y se han comparado los de uno con los del

otro. Con los valores de similitud obtenidos se ha elaborado un gráfico (ver Figura 4-22), en él que las líneas de bloques de color corresponden a los temas de un nodo y las columnas a los del otro. Los temas se han ordenado en ambas dimensiones de forma que los homólogos estén en las mismas posiciones en las entradas correspondientes de la tabla. Como en gráficos anteriores, los valores más altos de similitud se representan con colores más claros.

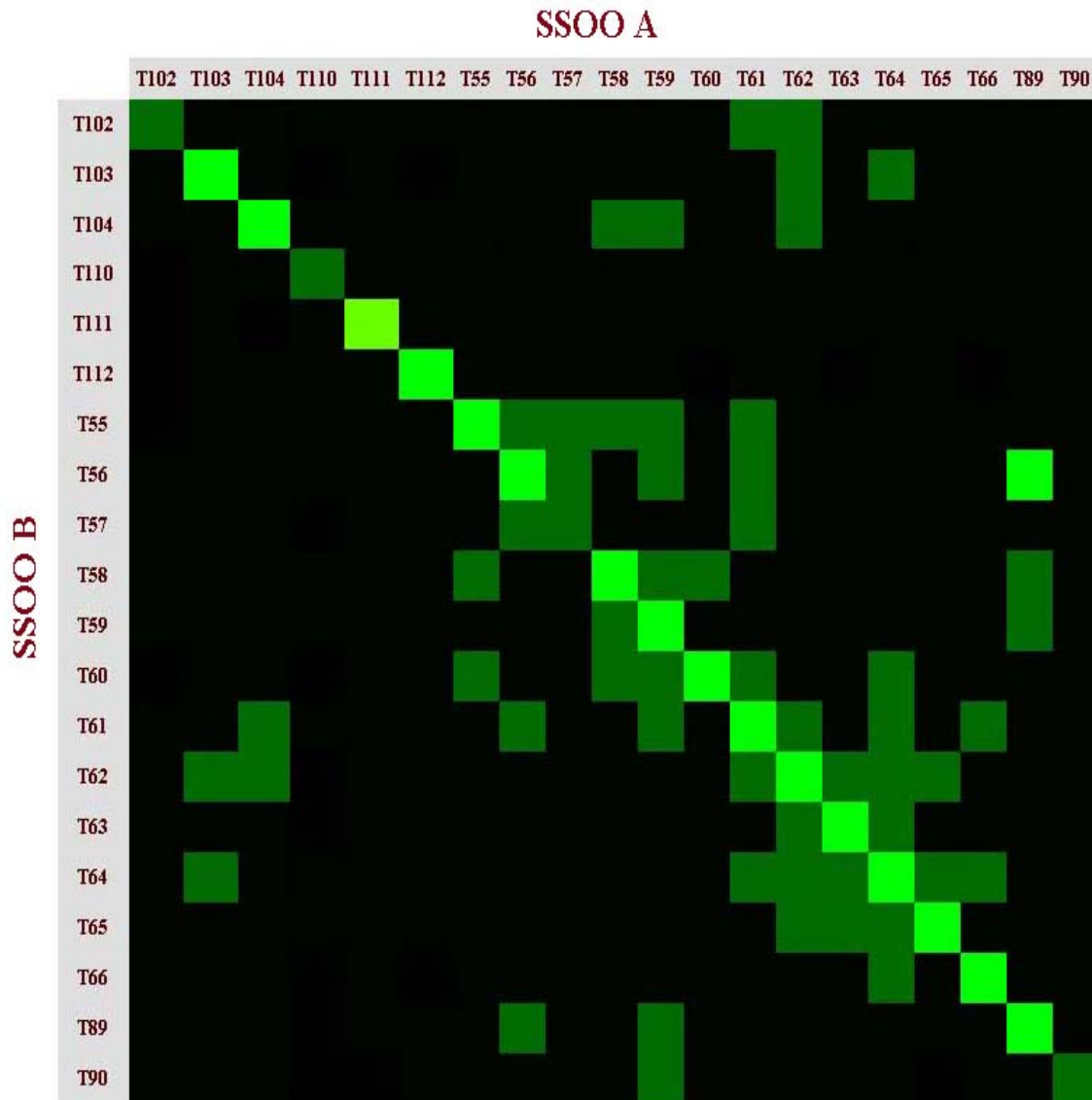


Figura 4-22. Mapeo entre temas de dos nodos KC de SSOO con distintos documentos pero árboles de conocimiento iguales

Como puede verse en la imagen de este primer experimento, los grados de similitud más elevados -bloques de color claro- se encuentran sobre la diagonal en casi todos los casos. Con el planteamiento propuesto, esto significa que es posible identificar las ramas de los árboles de conocimiento que contienen documentos que tratan de los mismos temas.

Para el segundo experimento de la serie (E222) se han utilizado dos nodos KC sobre TALF que tienen árboles distintos para organizar el conocimiento. De nuevo se han calculado los VPP de



los temas a partir de los documentos incluidos en ellos y se han calculado los grados de similitud entre los temas de distintos nodos comparando sus correspondientes vectores. El resultado se ha presentado en un gráfico (ver Figura 4-23) donde los temas de un nodo están en el eje de abscisas y los del otro en el de ordenadas. Como en otras ocasiones, los valores del grado de similitud se muestran con bloques de color, que son más claros cuanto mayores son los coeficientes representados. En este caso, se han marcado con una cruz las parejas de temas que se consideran vinculadas por su contenido mediante un análisis manual por un experto en la materia.

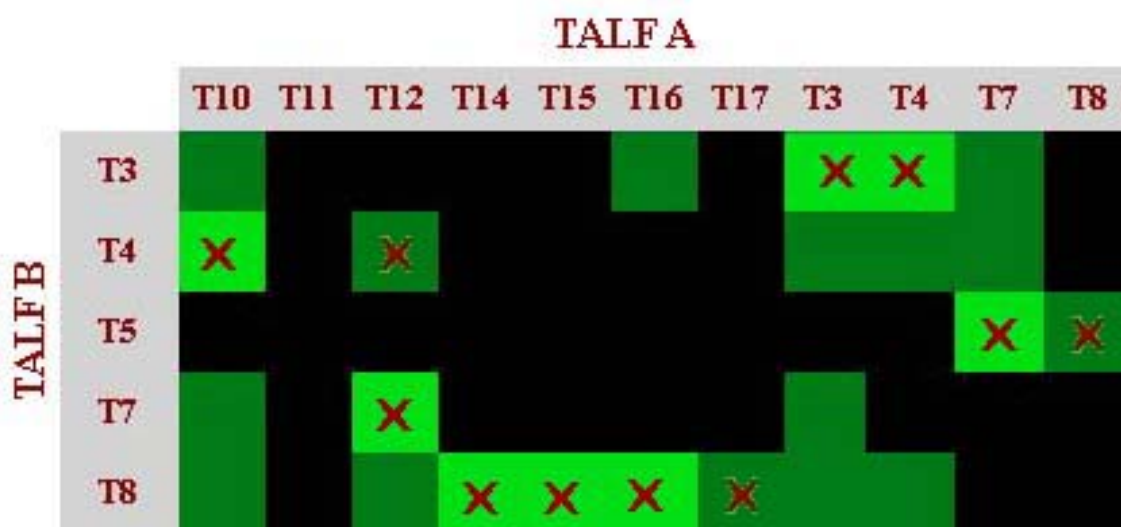


Figura 4-23. Mapeo entre temas de dos nodos KC de TALF con distintos documentos y árboles de conocimiento

Como resultado de este segundo experimento, se puede observar que la mayoría de las asociaciones hechas por el experto coinciden sobre bloques de color claro y que todos los bloques claros están en parejas de temas asociados por el experto. Por lo tanto, es posible identificar por el procedimiento propuesto y de forma automática los temas que tratan asuntos relacionados en árboles de conocimiento distintos.

#### 4.4.2.3 Experimento de asociación automática entre nodos de conocimiento

Partiendo de los documentos incluidos en cinco nodos KC, el de SI utilizado el primer grupo de experimentos, los dos de SSOO preparados para el anterior grupo y los dos de TALF utilizados en el mismo grupo, se ha establecido un VPP para cada uno de ellos. En todos los casos, los documentos contenidos en los nodos son distintos. En este nuevo experimento (E231) comparando dichos VPP se ha obtenido un gráfico (ver Figura 4-24), en el que cada línea de bloques, vertical y horizontal, corresponde a un nodo. Como en los diagramas anteriores, los colores más claros representan similitud mayor.

En el gráfico se puede ver que los índices de similitud entre los VPP de los nodos que tratan sobre los mismos temas son altos en comparación con los obtenidos al comparar los vectores de

nodos sobre temas distintos. Esto significa que -utilizando esta técnica- es posible identificar nodos que tratan sobre contenidos parecidos y diferenciarlos de otros sobre materias distintas.

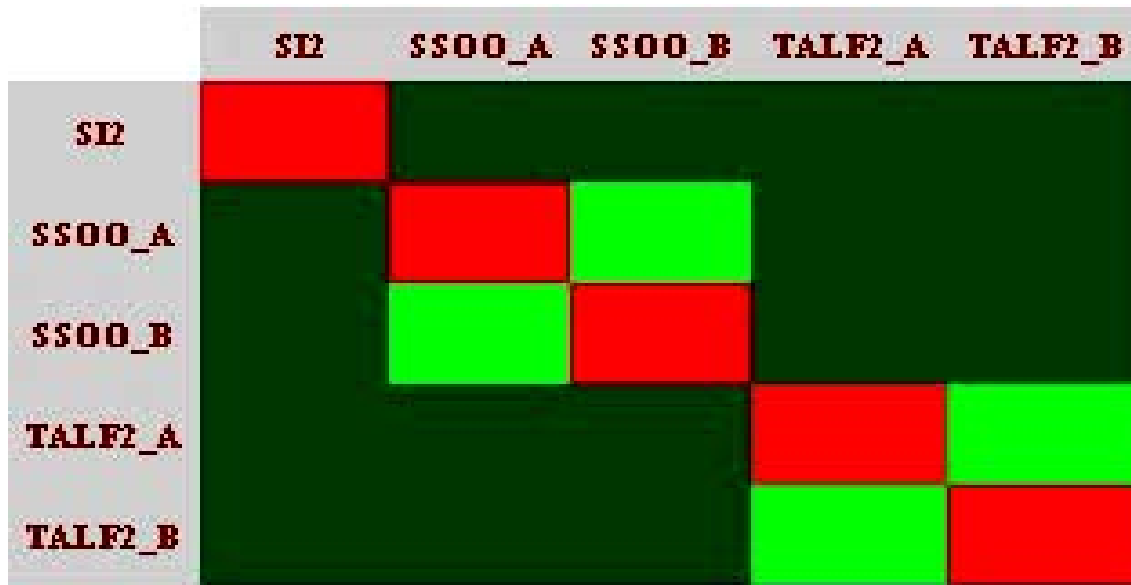


Figura 4-24. Agrupación de nodos KC por temáticas

#### 4.4.2.4 Conclusiones

El proceso de digestión de conocimiento propuesto parece capaz de explicitar conocimiento latente en un entorno de gestión de conocimiento, que puede resultar útil para facilitar la labor de gestión desempeñada por el sistema, la interacción entre las entidades del mismo y el acceso de los usuarios a los contenidos gestionados, entre otras aplicaciones interesantes. El enriquecimiento del contenido propuesto parece proporcionar un soporte muy potente para el intercambio automático del conocimiento entre sistemas de gestión de conocimiento abriendo una vía de desarrollo de los mismos en el ámbito de la Web Semántica.

Sin embargo, el umbral encontrado en los grados de similitud para considerar los ítems de conocimiento parecidos es bajo y es raro que aparezcan valores mucho más altos. En casi todos los casos considerados la mayoría de los ítems con similitud superior a 0,3 están relacionados por su contenido y todos los que no lo están tienen índices menores, aunque algunos objetivamente relacionados no alcanzan dicho valor. En ciertos casos el umbral es aún menor y esta entre 0,2 y 0,3. Sería deseable que el índice de similitud marcara con más claridad la separación entre ítems con distinto contenido y matizara el parecido entre los que lo tienen similar.

#### 4.4.3 Interés de los usuarios: Descubrimiento y aplicación

Como se ha mostrado al principio del capítulo, el Monitor de Cliente (MC) de SKC se ocupa de registrar la actividad de los usuarios en el lado del cliente y analizarla para facilitar la selección

del conocimiento, aprovechando el interés demostrado por la comunidad de usuarios hacia él, y el Módulo de Análisis de Conocimiento (MAC) de SKC se encarga de enriquecer el conocimiento que maneja el sistema y facilitar su gestión, explotando el conocimiento latente que el módulo descubre en el propio sistema. Combinando el análisis de la actividad de los usuarios del MC, aplicado a los usuarios individuales para evaluar sus intereses particulares, con el análisis de conocimiento del MAC, se pretende facilitar la clasificación de los usuarios y la identificación de conocimiento con utilidad potencial para ellos, dentro y fuera de SKC.

Con el objetivo de probar la viabilidad de las propuestas anteriores, se ha incorporado soporte para el interés de los usuarios en el MAC de SKC, que se había prototipo ya sobre una instancia del sistema KC junto con otros elementos fundamentales del entorno. Como otras veces, se han llevado a cabo experimentos con nodos KC dirigidos a probar los prototipos de SKC en actividades docentes de la Escuela Politécnica Superior de la Universidad Autónoma de Madrid, en el marco de los Proyectos de Innovación Docente presentados antes.

En esta ocasión, se han empleado dos nodos KC, uno sobre Sistemas Informáticos (SI) y otro sobre Teoría de Autómatas y Lenguajes Formales (TALF). El primero es una continuación de un nodo utilizado en experimentos anteriores, presentados en esta misma sección, mientras que el segundo es nuevo, a pesar de tratar del mismo tema que otros utilizados en experiencias previas, también descritas antes. En ambos casos, se han diseñado actividades con aspectos específicos para las pruebas del MAC con soporte para el interés de los usuarios, que se han realizado a lo largo del curso 2006-2007.

El nodo de SI continúa con el desarrollo del temario de la asignatura sobre esa materia iniciado un año académico antes. Durante el nuevo curso unos 90 alumnos han aportado entorno a 160 documentos repartidos entre unos 40 temas nuevos, que se añaden a otros tantos ya existentes y más de 180 documentos elaborados anteriormente.

El nodo de TALF recoge los documentos confeccionados por los alumnos a partir de otros aportados por un monitor a lo largo de un año académico en un árbol de conocimiento sobre temas de la correspondiente asignatura. En total unos 90 alumnos han trabajado sobre 6 temas y casi 450 documentos.

#### ***4.4.3.1 Experimentos de identificación del interés de los usuarios***

Estos experimentos están dirigidos a probar el proceso de identificación del interés de los usuarios. Para ello se han realizado dos experiencias, cada uno con uno de los nodos descritos anteriormente.

La primera de dichas experiencias (E311) utiliza el nodo de SI para que los alumnos desarrollen el temario de la asignatura a partir de los apuntes y las referencias proporcionados durante el

curso. Se trata de elaborar artículos que sirvan para preparar el desarrollo de uno de los temas durante el examen de la asignatura. La experiencia tiene dos fases: en la primera se asigna a cada alumno un par de temas para que elaboren los correspondientes artículos y los publiquen en el sistema; en la segunda, se asignan otros tres temas a cada alumno para que revisen, comenten y califiquen los trabajos realizados por sus compañeros en la fase anterior. Después de este proceso, el sistema ordena los artículos por su calidad dentro del correspondiente árbol de conocimiento. Como resultado, los alumnos disponen de 40 nuevos temas desarrollados para la preparación del examen y pueden conseguir hasta un punto adicional en la calificación final de la asignatura, según su participación en la actividad. Para hacer el seguimiento posible, los alumnos tienen que identificarse al principio de cada sesión con el sistema.

Al final de la experiencia se ha procesado el fichero de LOG registrado durante su realización. A partir de él se han determinados los Vectores de Interés de los Usuarios (VIUs) para dos intervalos de la actividad: el primero, desde el inicio de la misma hasta el final de la primera fase, etapa de elaboración de los artículos sobre los temas asignados; y el segundo, desde el comienzo de la segunda fase hasta el final de esta, intervalo de revisión de los artículos incluidos en los nuevos temas asignados.

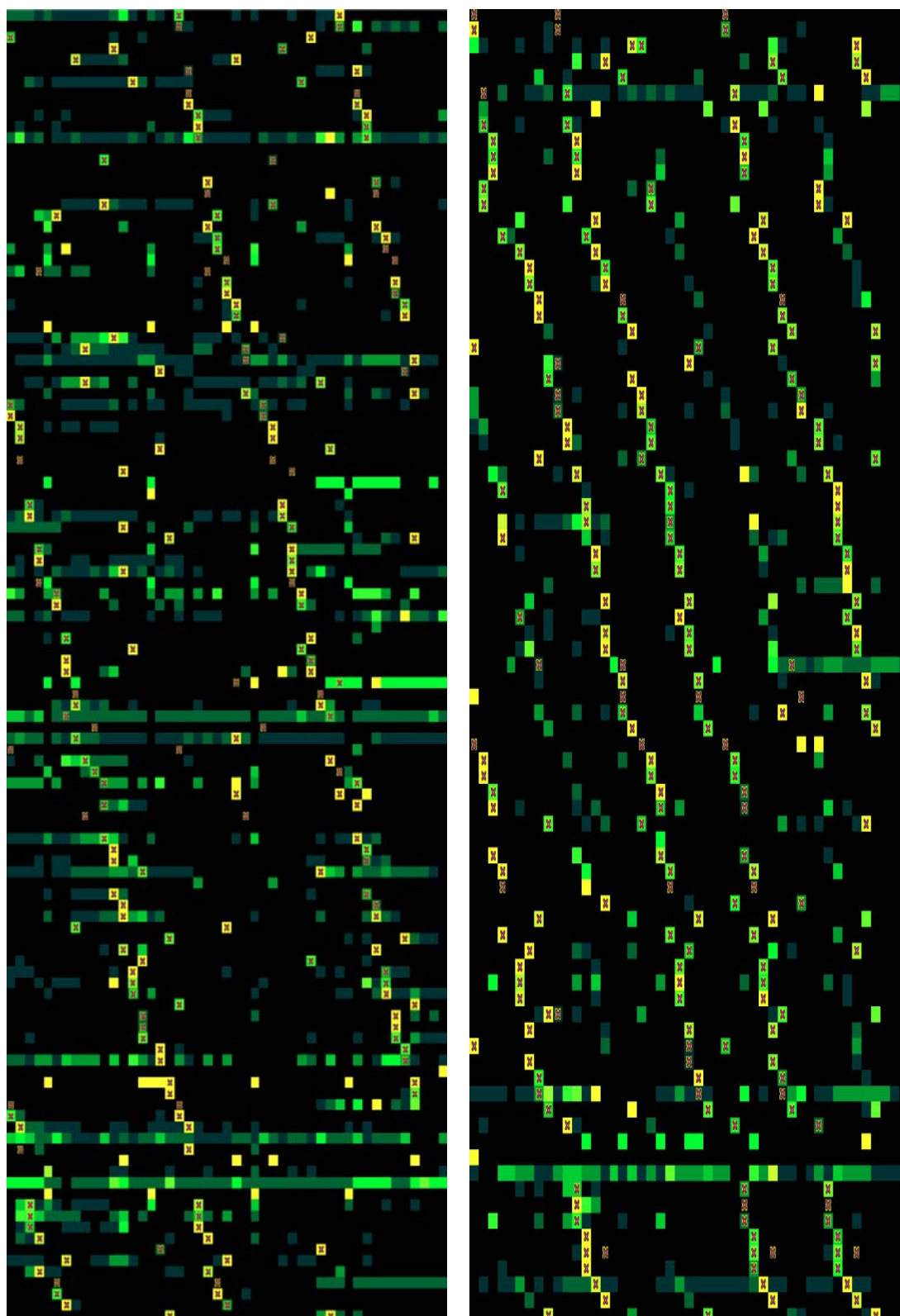
El VIU de cada usuario se representa mediante una fila de bloques de color. Cada bloque muestra el interés relativo del usuario por un tema, según la codificación cromática de la Figura 4-25. Los colores más claros indican un grado de interés mayor. El más claro representa el mayor interés relativo y el negro la ausencia del mismo.



Figura 4-25. Código cromático para interés relativo en experiencia SI.

Los VIUs de los intervalos mencionados se muestran en la Figura 4-26 en dos imágenes, a la izquierda el del primer periodo (a) y a la derecha el del segundo (b). En ambos casos, cada fila corresponde a un usuario y cada columna a un tema. Los temas considerados en los dos gráficos son los mismos, pero en el primero (a) hay más usuarios que en el segundo (b), debido a que en el segundo intervalo no hay registro de actividad de algunos usuarios que interactuaron con el sistema en el primero. Los temas asignados a los alumnos en cada intervalo aparecen marcados con una cruz. Algunos usuarios no tienen asignados temas en ningún caso, porque no son alumnos. En la primera imagen (a), las cruces están sobre los dos temas donde los alumnos debían elaborar los documentos. En la segunda imagen (b), las marcas están sobre los tres temas que debían revisar. En algún caso aparecen más asignaciones de temas de lo previsto, se trata de errores en el proceso que se corrigieron durante el desarrollo del experimento, pero quedaron registrados en el LOG.





(a) VIUs periodo de elaboración de documentos propio.

(b) VIUs periodo de revisión de documentos ajeno.

En ambas imágenes cada fila corresponde a un usuario y cada columna a un tema, y las cruces indican los temas asignados

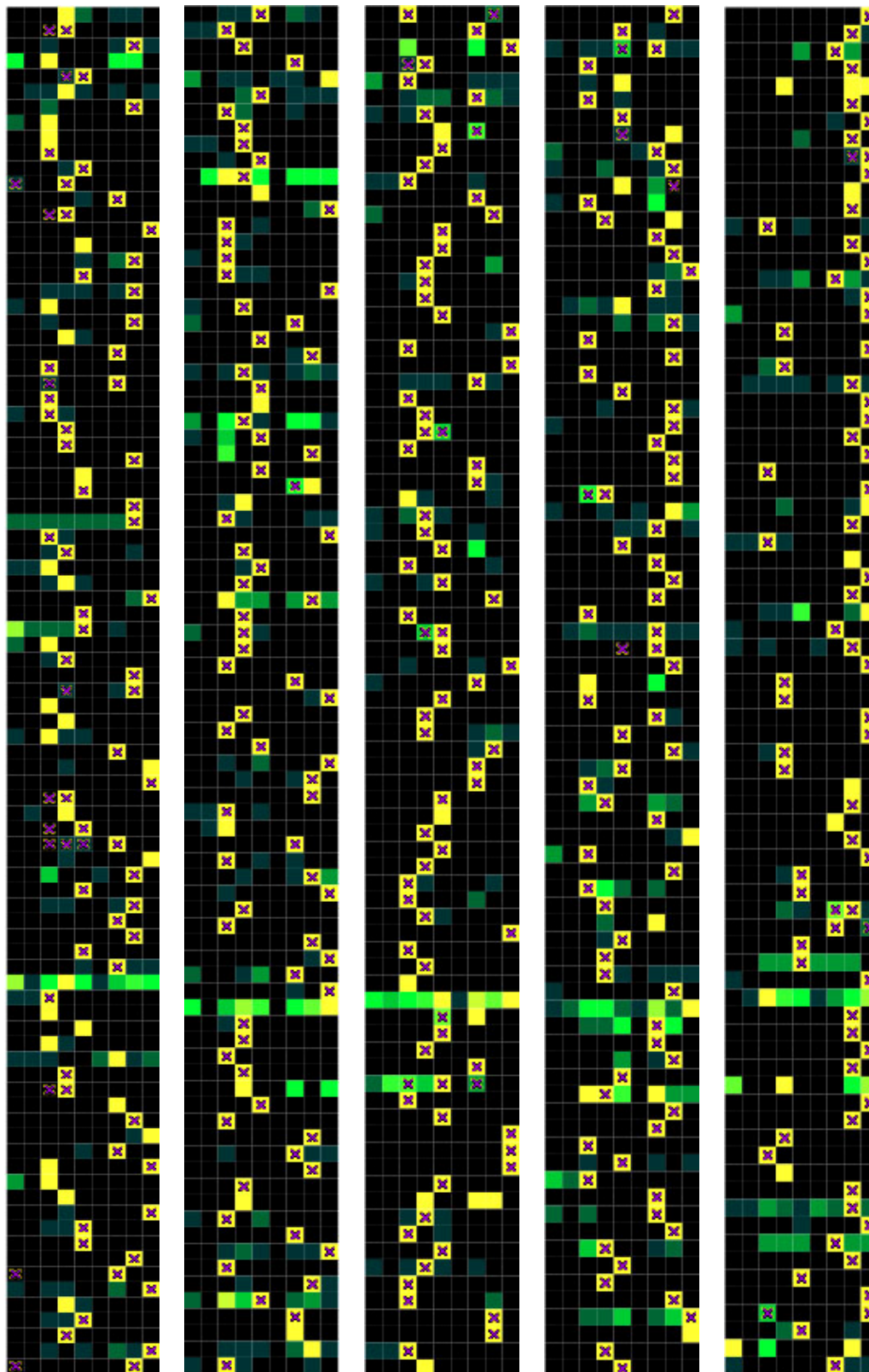
Figura 4-26. VIUs de los dos periodos de actividad de la experiencia en nodo de SI.

Como se puede ver en las imágenes de la Figura 4-26, los temas asignados a los usuarios en ambos intervalos, imágenes (a) y (b), se encuentran entre los que más interesan a la mayoría de ellos, puesto que las marcas de asignación suelen estar sobre alguno de los bloques de colores más claros de cada fila. El fenómeno es más evidente en la imagen del segundo intervalo (b), puesto que la revisión de material requeriría más interacción con el entorno que la elaboración del mismo. Esto se debe a que la experiencia no requiere consultar información en el sistema para preparar los documentos y a que el entorno no proporciona soporte para la redacción en línea de los mismos.

Otro fenómeno que llama la atención en las imágenes es el poco interés aparente de los alumnos por los temas que no se les han asignado. La justificación puede buscarse en las condiciones del experimento, los periodos de actividad ilustrados y las características del colectivo implicado. El experimento se ha diseñado para provocar en los usuarios un interés artificial por temas determinados en ciertos momentos y hacer que la actividad necesaria sea fácil de monitorizar. La motivación de los estudiantes en los periodos ilustrados es conseguir la nota adicional por participar correctamente en la actividad, pero no utilizar el material elaborado para estudiar -los alumnos suelen preparar los exámenes al final-.

Los VIUs del periodo de estudio del examen, desde el final del de revisión hasta la realización de la prueba, no difieren mucho de los dos intervalos mostrados en la Figura 4-26. Mirando el registro del sistema en dicho periodo, parece que cada alumno sigue más interesado en los temas que se le han asignado que en el resto del temario. El motivo puede estar en la forma en que se accede al material para estudiar, puesto que el entorno sólo puede monitorizar la actividad en línea y es muy probable que los documentos se impriman o copien para preparar el examen sin conexión con el entorno.

El segundo de los experimentos realizados (E312) emplea el nodo de TALF mencionado anteriormente. La actividad consiste en elaborar cinco pequeños resúmenes sobre otros tantos temas de la asignatura, utilizando los artículos de referencia proporcionados por el profesor a través del sistema. Para la preparación de cada artículo se dispone de un tiempo de dos semanas. Al principio de cada periodo se tiene que elegir un tema de trabajo por alumno, de forma que los estudiantes no repitan tema a lo largo de la actividad y no haya más de un número máximo de alumnos por tema e intervalo. Los resúmenes entregados en cada periodo quedan depositados en el sistema con el acceso restringido, sólo habilitado para el profesor encargado de su evaluación. Según el trabajo realizado, cada estudiante puede conseguir hasta un punto adicional en la nota de la asignatura. Como en el caso anterior, los participantes en la experiencia deben identificarse al principio de cada sesión para permitir el seguimiento de su actividad.



(a) Periodo entrega 1. (b) Periodo entrega 2. (c) Periodo entrega 3. (d) Periodo entrega 4. (e) Periodo entrega 5.

En las imágenes cada fila corresponde a un usuario y cada columna a un tema, y las cruces indican los temas asignados

Figura 4-27. VIUs de los cinco periodos de actividad de la experiencia en nodo de TALF.

Después de la experiencia se ha procesado el LOG del sistema para determinar los vectores de interés de los usuarios (VIUs) de los cinco periodos de entrega de documentos. Los VIUs correspondientes a cada intervalo se muestran en la Figura 4-27 en cinco imágenes, ordenados por el número de periodo, empezando por el primero a la izquierda y terminando por el quinto a la derecha (imágenes a, b, c, d y e de la Figura 4-27). En todos los casos, cada fila corresponde a un usuario y cada columna a un tema. Los temas son los mismos en las cinco imágenes, pero algunos usuarios no aparecen en todas ellas, porque abandonaron la actividad y dejaron de aparecer en el LOG. Como en el caso anterior, los temas asociados a los alumnos en cada intervalo aparecen marcados con una cruz y algunos usuarios no tienen temas señalados, porque no son alumnos. Además, en ciertas ocasiones, aparecen más asociaciones de temas por usuarios de las esperadas, debido a errores en el proceso de asignación que se corrigieron inmediatamente, aunque dejaron rastro en el registro. La representación de los VIUs es parecida a la del experimento anterior, de nuevo se utiliza la codificación cromática de la Figura 4-25.

De forma análoga a lo que pasó en el anterior experimento, los temas que más interesan a los alumnos están claramente entre los asociados a los usuarios en cada periodo. En esta ocasión, el fenómeno es más evidente que en el caso previo, por las características de la experiencia. Primero, el árbol de conocimiento se ha preparado para la actividad, presentando un número de temas ajustado a la misma. Segundo, para la preparación de los resúmenes se deben revisar documentos publicados en los correspondientes temas del propio árbol de conocimiento. Tercero, no se permite que los alumnos accedan a los documentos entregados. Además, el único estímulo para la participación es la nota adicional por el buen cumplimiento de las tareas asignadas. Por último, debido a todo lo dicho, los alumnos apenas se interesen por los temas que no tienen asociados en cada intervalo.

#### ***4.4.3.2 Experimentos de identificación de contenidos interesante para los usuarios dentro y fuera del sistema***

En los experimentos anteriores se demostró la utilidad de los vectores de interés de usuario (VIUs) para representar cómo se reparte la atención de los miembros de una comunidad entre los elementos de conocimiento del nodo al que pertenecen. Los VIUs no tienen sentido fuera del contexto donde se han definido, porque se refieren de forma explícita a elementos concretos del mismo y no proporcionan directamente información que pueda servir fuera de ese contexto. Ahora vamos a comprobar el beneficio de los vectores de pesos de palabras de interés de usuario (VPPIU) para aprovechar la información recogida en los VIUs de forma más general, y con aplicación dentro y fuera del nodo donde se produjeron.

### Agrupación Usuarios por Interés

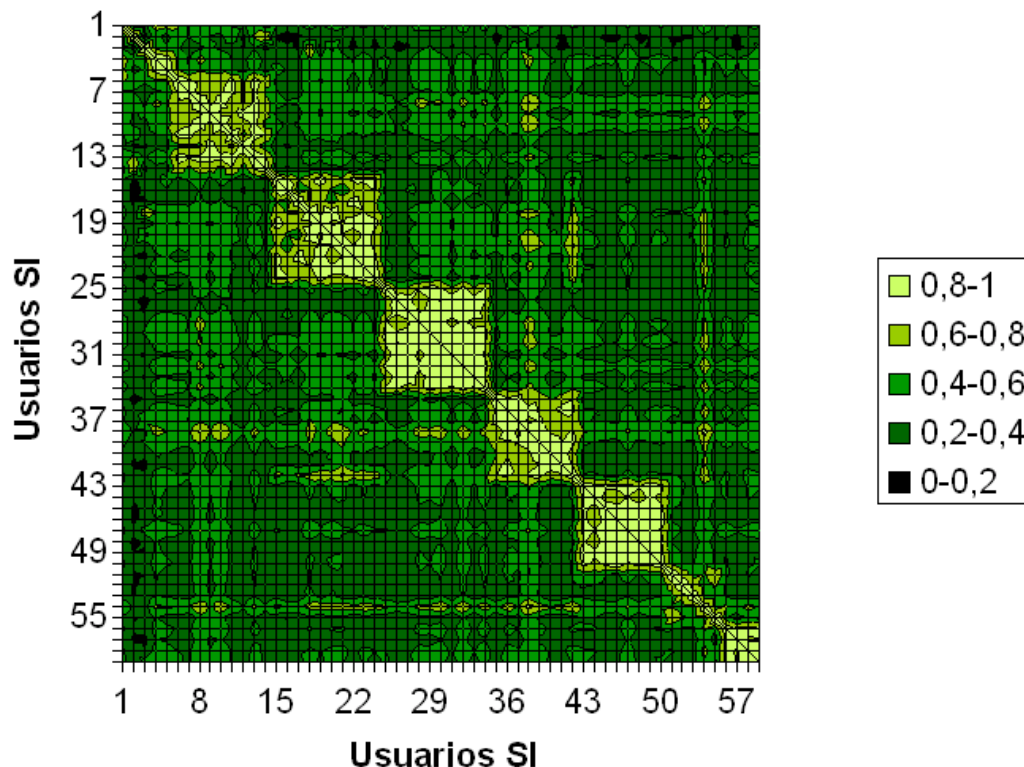


Figura 4-28. Agrupación de usuarios por interés mediante sus VPPIUs.

El experimento realizado en esta ocasión (E321) consiste en comparar entre si los VPPIUs generados a partir de los VIUs obtenidos en la experiencia E311, que se llevó a cabo con un nodo de SI y que se presentó en el apartado 4.4.3.1. Dicha actividad consistía en proponer a los alumnos primero dos temas, para desarrollarlos en sendos artículos, y después otros tres temas, para que evaluaran los trabajos aportados por diferentes usuarios. Como resultado, durante la actividad, los estudiantes se han mostrado más interesados por los temas asignados que por el resto de contenidos del árbol de conocimiento del nodo utilizado, que es lo que se pretendía. Los temas se han dividido en ocho paquetes para asignarlos a los alumnos de forma sistemática, así que se puede saber el grupo de temas establecido para cada uno. Los paquetes incluían selecciones variadas de los cuarenta temas de SI considerados en la actividad. Los VPPIUs correspondientes se han calculado en función de la distribución del interés -representada por el VIU de cada usuario- y de los VPPs de los artículos implicados, como se explico en el apartado 4.2.2.6.

Los VPPIUs de los usuarios se han comparado entre si, para comprobar si evidenciaban los intereses comunes de aquellos, de forma análoga a como se confrontaron los VPPs de otros elementos de conocimiento en experimentos anteriores. El resultado se ha presentado en un

gráfico (ver Figura 4-28) donde se representan todos los usuarios participantes en ambos ejes de coordenadas, ordenados por los grupos de temas asignados.

Como en otras ocasiones, los valores del grado de similitud se muestran con teselas de color, que en este caso son más claras cuanto mayores son los coeficientes representados. Como se puede apreciar en el gráfico, los grados más altos aparecen agrupados alrededor de la diagonal formando bloques. En las condiciones propuestas, esto indica que los VPPIUs permiten identificar los usuarios interesados por los mismos grupos de temas, empleando descriptores genéricos que pueden servir para compararlos con cualquier objeto que tenga asociado un VPP. Sin embargo, sólo se aprecian claramente cinco bloques en la diagonal, en lugar de los ocho que cabría esperar con el número establecido de grupos de temas. Esto puede deberse a la cantidad de usuarios asignada a cada paquete que han participado de forma efectiva en la actividad, y a la pérdida de rasgos significativos en los VPPIUs de algunos de esos paquetes, como consecuencia de las combinaciones particulares de temas diversos.

En estas condiciones los VPPIUs son VPPs representativos de las temáticas de interés particulares de cada usuario, y resultan análogos a los VPPs de los temas que forman los árboles de conocimiento de los nodos de SKC. Por lo tanto, dichos vectores pueden servir para comparar las temáticas que representan con los VPPs de documentos, temas o usuario del propio nodo, pero también de otros nodos SKC, como se demostró con otros VPPs, o incluso de fuentes externas cualesquiera.

#### **4.4.3.3 Conclusiones**

Los experimentos de este grupo se han dedicado al tratamiento del interés de los usuarios, en lo que se refiere al descubrimiento del mismo y su aplicación. Los primeros experimentos se han dirigido a determinar y describir el interés de los usuarios, y los últimos a la identificación de contenidos interesantes para los usuarios dentro y fuera del sistema.

Como resultado de los primeros experimentos se puede concluir que el mecanismo propuesto para representar el interés de los usuarios hacia los elementos de conocimiento del nodo SKC en el que trabajan, mediante Vectores de Interés de Usuario (VIU) que reflejan la distribución del interés de cada uno entre tales elementos, proporciona descriptores adecuados para identificar los focos de interés de los usuarios en el árbol de conocimiento de dicho nodo.

De los resultados de los últimos experimentos del grupo se puede afirmar que el procedimiento establecido para calificar a los usuarios de SKC con Vectores de Peso de Palabra de Interés de Usuario (VPPIU), considerando los VIUs de cada usuario y los Vectores de Peso de Palabra (VPP) de los elementos de conocimiento del sistema, dan lugar a descriptores adecuados para identificar usuarios con intereses parecidos. Además, los vectores VPPIU permiten identificar

elementos de información relacionados con los intereses de los usuarios dentro y fuera del sistema, si se establecen vectores VPP para tales elementos. Sin embargo, se ponen de manifiesto las limitaciones de los umbrales de similitud en la comparación de VPPs, como en experimentos anteriores, y se evidencia que la efectividad de los VPPIUs disminuye cuando los usuarios no tienen intereses muy concretos, como ocurre cuando se trabaja con VPPs de documentos de temática poco definidos.

#### **4.4.4 Red de nodos SKC: Agentes de red, descubrimiento y vinculación de nodos, y establecimiento de red**

Como se ha mostrado al principio de este capítulo, SKC dispone de un Módulo de Red, que se ocupa de descubrir otras instancias del sistema en la Web y de establecer contacto con ellas con el fin de crear una red sobre la que establecer relaciones de intercambio de conocimiento. Para ello, cada instancia del sistema dispone de un agente software que le representa y se encarga de interactuar con aplicaciones de la Web y de colaborar con los agentes de otras instancias. Como resultado SKC consigue construir una red de nodos de conocimiento, determinar los tipos de vinculación entre ellos y establecer las comunicaciones necesarias para mantener y explotar dicha organización. Para comprobar la viabilidad de las aproximaciones propuestas, hemos desarrollado un prototipo del Módulo de Red de nodos (MR) para SKC, que se ha incorporado a un sistema KnowCat (KC) junto con otros módulos prototipados anteriormente.

El prototipo ha permitido realizar varios experimentos con nodos KC, preparados para probar los prototipos de SKC durante varios cursos, en actividades docentes realizadas en la Escuela Politécnica Superior de la Universidad Autónoma de Madrid en el marco de los Proyectos de Innovación Docente descritos previamente en este documento. En particular, se han utilizado nodos KC, ya emplearon en experimentos presentados en los apartados anteriores: uno sobre Sistemas Operativos (SSOO); dos sobre Teoría de Autómatas y Lenguajes Formales (TALF); y otro sobre Sistemas Informáticos (SI).

El nodo de SSOO es el fruto del trabajo realizado durante cuatro cursos sucesivos por los alumnos de una asignatura sobre ese tema, se trata de un árbol de conocimiento que contiene 350 documentos en 20 temas organizados en dos niveles.

Los nodos de TALF contienen los documentos elaborados por alumnos durante un año académico en dos árboles de conocimiento distintos. Aunque los nodos tratan sobre el mismo tema, en cada uno los temas y los documentos son diferentes. En los dos nodos el árbol de conocimiento tiene dos niveles, pero en uno hay 50 documentos repartidos en 12 temas y en el otro hay 24 documentos en 6 temas.

Finalmente, el nodo de SI desarrolla un temario 40 temas sobre la correspondiente materia mediante 180 documentos aportados por alumnos durante un curso.

Los experimentos realizados se han dirigido a comprobar la viabilidad de la aproximación propuesta para la creación, mantenimiento y explotación de una red de nodos de conocimiento. Para ello se han realizado dos experimentos que se describen en los siguientes apartados.

#### 4.4.4.1 Experimentos de vinculación automática entre nodos

Para probar el proceso de establecimiento y mantenimiento de red de nodos de conocimiento mediante vinculación automática se ha realizado dos experimentos en las que se han utilizado cinco nodos. Tres de los cuales son los nodos originales de TALF y SI, los otros dos se han creado a partida del nodo de SSOO, manteniendo árboles de conocimiento iguales al de partida en los dos y repartiendo los documentos del nodo original de modo que en cada par de temas homólogos de los nuevos nodos queden con un número similar de documentos diferentes y de parecida relevancia.

Tabla 4-2. Nodos conocidos por cada nodo existente

	SSOOa	SSOOb	TALFa	TALFb	SI
SSOOa	-	1	1	2	3
SSOOb	2	-	2	1	1
TALFa	2	1	-	1	2
TALFb	1	2	2	-	3
SI	3	1	1	3	-

El primer experimento (E411) se ha iniciado registrado manualmente en cada nodo la información básica de identificación de algunos de los otros nodos, de tal forma que ninguno conozca a más de dos de los demás. En la Tabla 4-2 se indica con un “1” los nodos conocidos inicialmente por cada nodo de la primera columna de la izquierda.

A continuación se han puesto en funcionamiento los agentes de todos los nodos con la función de búsqueda en la Web desactivada. Los agentes han comenzado a visitarse, primero para completar la información de identificación de los nodos que conocían. Como resultado han obtenido la información de identificación de los nodos visitados y han descubierto nodos desconocidos que les conocían a ellos y les habían visitado (indicados con “2” en la Tabla 4-2).

Cuando se ha agotado el descubrimiento por contacto directo, los agentes han establecido sus vinculaciones con los nodos que conocen y están en disposición de visitarlos para obtener las referencias de los nodos registrados por estos. Así los agentes que no se habían descubierto



directamente consiguen las referencias de los otros o son visitados por los que han encontrado la suya (señalados con “3” en la Tabla 4-2). Después de un tiempo de actividad de los agentes, todos los nodos conocían al resto de los nodos activos existentes. El tiempo necesario para completar el proceso depende del orden de visitas seguido por cada agente, de la disponibilidad de los otros agentes al intentar interactuar con ellos y de los periodos establecidos entre interacciones sucesivas. En el mejor de los casos el proceso puede completarse en tres ciclos de interacción.

Si el experimento se repite con uno de los agentes desactivado, todos los agentes que conocen al inactivo terminan por considerarlo como tal. Si después se vuelve a activar dicho agente y tiene referencias de agentes activos, todos los nodos acaban teniendo la referencia del nodo reactivado como activo. En este caso, las visitas iniciadas por este nodo, provocan el cambio de consideración en los nodos visitados que le creían inactivo.

Tabla 4-3. Coeficientes de similitud entre nodos

	SSOOa	SSOOb	TALFa	TALFb	SI
SSOOa	-	<b>0,43</b>	0,12	0,11	0,20
SSOOb	<b>0,43</b>	-	0,12	0,10	0,20
TALFa	0,12	0,12	-	<b>0,31</b>	0,15
TALFb	0,11	0,10	<b>0,31</b>	-	0,13
SI	0,20	0,20	0,15	0,13	-

El segundo experimento (E412) parte de un escenario en el que cada nodo conoce a los demás de la red y ha completado sus registros. En estas condiciones, todos los nodos disponen de los descriptores de los otros y están en disposición de establecer el tipo de vinculación con ellos, comparando sus descriptores con los demás. Como se mostró en un experimento anterior los grados de similitud entre los descriptores de los nodos permiten identificar los nodos que tienen temáticas similares. Esto puede apreciarse en la Tabla 4-3, donde aparecen los coeficientes de similitud entre los nodos y destacadas en negativo los valores superiores a 0,3, que corresponden a parejas que se consideran afines.

Al activar los agentes en esta situación los agentes afines establecen una vinculación de contenido, mientras que el resto establecen una básica. A partir de este momento, las interacciones entre los nodos con distinta vinculación son diferentes. En el caso de los nodos con vinculación básica, la interacción se reduce a intercambiar sus descriptores y las referencias de nodos activos que tiene cada uno. En el caso de los nodos con vinculación de contenido, además se intercambian los descriptores de todos los ítems de conocimiento.

Si se desactiva uno de los nodos vinculados, los demás nodos lo considerarán inactivo al cabo de un tiempo. Cuando el nodo se reactiva y los demás nodos lo detectan, vuelven a marcarlo como activo, pero retiran la vinculación establecida previamente, y tienen que recuperar el descriptor correspondiente y determinar el tipo de vinculación otra vez. Al quedarse inactivo un nodo con vinculación de contenido, se desechan todos los ítems de conocimiento y los descriptores correspondientes, que luego es preciso volver a obtener en caso de reactivación y restauración de la vinculación perdida. Los agentes se encargan de determinar la vinculación entre los nodos y de desechar los ítems de conocimiento asociados a un nodo desactivado, pero del alta de los ítems en la base de datos y del establecimiento de relaciones entre ítems se ocupan otros módulos de SKC.

Si durante la actividad de los agentes se cambia el descriptor de cualquiera de los nodos, de forma que alguno de los valores de similitud pase de ser mayor de 0,3 a ser menor o viceversa, al cabo de un tiempo el tipo de vinculación entre los nodos afectados cambia y el repositorio de conocimiento se adapta a la nueva situación.

Como resultado de estos experimentos se puede concluir que los agentes son capaces de reconocer una red de nodos y adaptarse a los cambios de la misma colaborando, actuando sobre el entorno y reaccionando a sus cambios de forma reactiva. Por lo tanto, es posible establecer y mantener una red de nodos SKC con intercambio selectivo de conocimiento empleando el procedimiento propuesto.

#### ***4.4.4.2 Experimento de descubrimiento de nodos SKC en la Web***

Este experimento (E421) está dirigido a comprobar la viabilidad del procedimiento propuesto con el fin de descubrir nodos desconocidos para una red SKC en la Web. Para ello se utilizan los nodos considerados en los experimentos anteriores. Dichos nodos se implantan en un servidor SKC de modo que ninguno tenga conocimiento de los otros. Además, se prepara una página Web en el servidor con los datos de los nodos incluidos (ver la Figura 4-13). Después se hace que la página de nodos sea accesible para el indexador del buscador considerado, en este caso Google [Google 2008], y se la registra en él, para facilitar su hallazgo. De los detalles del proceso de descubrimiento de nodos en la Web se habla en el apartado sobre el Módulo de Red en la sección de Prototipo de SKC sobre KC al principio de este capítulo.

A continuación, se activan los agentes de los nodos y se espera hasta que el indexador registra en el buscador la página de nodos del servidor publicada. Desde el momento de la activación, los agentes comienzan a buscar páginas de nodos activos en los buscadores configurados. En un plazo de unos días, dependiendo de la dinámica de la indexación, alguno de los agentes encuentra la página buscada en Google (ver la Figura 4-29). En ese momento el agente descubridor accede a la página de nodos del servidor y obtiene los datos de los nodos incluidos.

Seguidamente se pone en contacto con todos ellos, y se construye la red de nodos SKC como ocurrió en los experimentos anteriores.

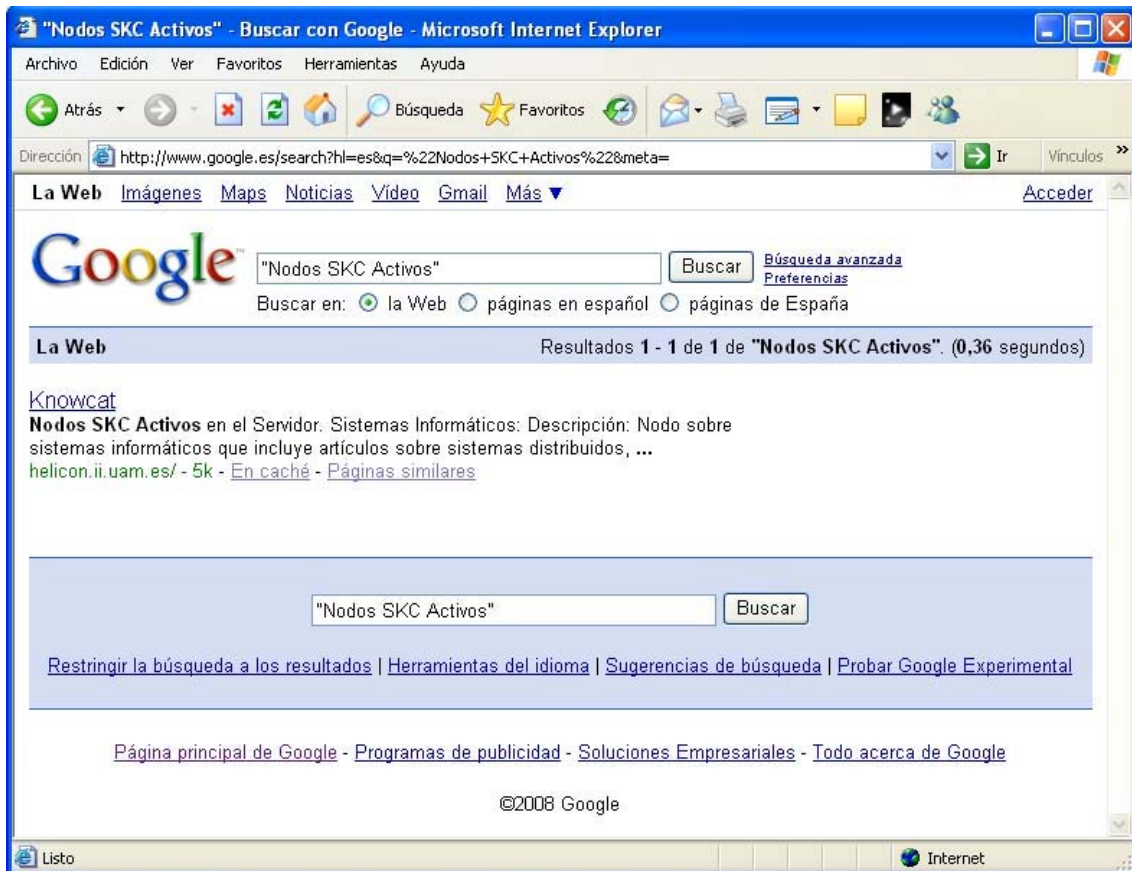


Figura 4-29. Página de Nodos SKC Activos indexada por Google.

Como conclusión de este experimento se puede afirmar que el procedimiento propuesto para descubrir nodos desconocidos en la Web es posible y resulta ser un potente mecanismo de integración de redes de nodos. Por lo tanto, es viable establecer una red de nodos SKC descubriendo nodos dispersos por la Web, sin conocer nada sobre ellos, empleando el procedimiento propuesto y buscadores Web convencionales.

#### 4.4.4.3 Conclusiones

Los experimentos que integran este grupo se han dirigido a la constitución y mantenimiento de una red de nodos SKC, mediante agentes software que representan a cada nodo y que se encargan de descubrir otros nodos y de establecer y mantener relaciones con ellos. Los primeros experimentos se han ocupado del proceso de vinculación automática entre los nodos, así como de la creación y mantenimiento de la red. El último experimento se ha ocupado específicamente del proceso de descubrimiento automático de nodos SKC en la Web, mediante el empleo de un buscador Web convencional.

Como resultado de los primeros experimentos se puede afirmar que el mecanismo sugerido para establecer y mantener una red de nodos SKC con intercambio selectivo de conocimiento entre agentes software es adecuado para el fin perseguido. Los agentes consiguen constituir la red de nodos y mantenerla operativa adaptándola a las circunstancias de las comunicaciones y la evolución de los nodos.

A partir de los resultados del último experimento se puede concluir que el procedimiento establecido para descubrir nodos SKC desconocidos en la Web, utilizando buscadores genéricos, constituye un mecanismo flexible de integración de redes de conocimiento entre instancias de SKC.

# 5 Conclusiones y Trabajos Futuros

Este capítulo comienza con una recapitulación de los puntos más destacados del proyecto, para dar una visión global del mismo. A continuación se recogen las conclusiones y principales contribuciones de la investigación, donde se hacen algunas consideraciones sobre los resultados obtenidos, se sintetiza su significado y se da una impresión de la trascendencia del trabajo realizado. Finalmente, se tratan las cuestiones abiertas y los trabajos futuros, donde se revisan las líneas de investigación sin agotar, otras que el proyecto ha motivado y algunas actividades futuras relacionadas con el trabajo.

## 5.1 Recapitulación

Como se ha dicho varias veces a lo largo de este trabajo, uno de los efectos nocivos de la extensión del uso de las tecnologías de la información y las comunicaciones (TIC) desde mediados de los noventa ha sido la sobrecarga de información. El progreso tecnológico digital, la popularización del empleo de dispositivos basados en esta tecnología, y también el desarrollo y la generalización del uso de las redes de comunicación, han contribuido a dicha extensión y a la proliferación de la información disponible. La democratización de la red, que convierte a cualquier usuario en proveedor de contenidos, está jugando un papel fundamental como catalizador del proceso en los últimos tiempos. Este fenómeno también provoca diversidad en la calidad de la información disponible e incertidumbre sobre aquella. La Web es el ejemplo más general y significativo del entorno del fenómeno.

El problema de sobrecarga se acentúa cuando la información a la que afecta es crítica para los grupos que la emplean, que es el caso de los repositorios de conocimiento. En el ámbito de la Web, la sobrecarga de información toma dimensión global, puesto que la red es accesible por comunidades cada vez más amplias y en ella hay espacio para cualquier área de interés. Además, las TIC han transformado en poco tiempo el modo de trabajar con la información y el conocimiento, de forma que actualmente sería muy difícil prescindir de ellas. También, los nuevos hábitos de consumo informativo han provocado un aumento de los requerimientos de información y exigencia hacia ella.

Para investigar la sobrecarga de información es necesario disponer de un sistema que aglutine algunas de las propiedades más relevantes de los entornos en los que se produce el problema, pero con unos parámetros y una escala controlables. Creemos que los sistemas de gestión de conocimiento en red reúnen estas características. Efectivamente, como se señaló en el capítulo de Aproximación Conceptual, la Web podría verse desde algún punto de vista exigente como un sistema de gestión de conocimiento global, puesto que es un entorno en el que comunidades interesadas se esfuerzan en el tratamiento de la información compartida, para hacerla útil y transformarla, eventualmente, en conocimiento [McDermott 1999]. Además, la Web puede proporcionar medios para recoger, organizar y distribuir dicho conocimiento [Malthotra 2000], y también para cubrir las necesidades de colaboración de la comunidad que trata con él [Churman 1971].

Como se indicó en la Propuesta recogida en dicho capítulo, nuestra hipótesis es que hay varios aspectos ocultos en los sistemas afectados por la sobrecarga de información que pueden contribuir positivamente a la solución de tal problema. Por un lado, el sacar partido de la energía excedente de la actividad de los elementos activos que intervienen en dichos sistemas, como son los usuarios, los servicios, las aplicaciones y otras entidades que se relacionan con ellos. Por

otro lado, la utilización de las propiedades de los elementos y las actividades vinculados con los sistemas afectados, como pueden ser la red, las entidades activas ya mencionadas, la información y el conocimiento implicados, o los procesos y las interacciones de unos y de otros.

Tal como se planteó en la mencionada Propuesta, aplicando dicha hipótesis al modelo experimental propuesto, en nuestra opinión los sistemas de gestión del conocimiento en red se pueden beneficiar de aprovechar la energía residual de la actividad de las personas, los servicios y otras entidades artificiales que interaccionan con ellos, así como de explotar las características de la red, de las entidades involucradas en sus actividades, de las interacciones de las mismas y del conocimiento que gestionan dichos sistemas. Con ello pretendemos obtener posibles vías atenuantes de la sobrecarga de información en los sistemas de gestión de conocimiento, pero con la ambición de que sean de aplicación en ámbitos más amplios, como la Web.

El enfoque planteado tiene sus antecedentes en diversas referencias, entre las que están: el sistema de gestión de conocimiento KnowCat [Alamán 1999][Cobos 2003]; los procedimientos empleados en los buscadores actuales de la Web, como Google [Google 2008], para mejorar sus resultados [Baeza 1999] considerando los enlaces recibidos por los sitio Web para establecer su relevancia [Kleinberg 1998, 1999]; las técnicas empleadas en algunos sistemas de recomendación [Adomavicius 2005], que tienen en cuenta la historia de los usuarios para mejorar el servicio, como es el caso de Amazon [Amazon 2008]. Además, las tendencias recientes de la Web, que la Web 2.0 encarna [O'Reilly 2005], que aprovechan la estructura de la Web, las redes sociales, la inteligencia colectiva y la colaboración, en aplicaciones como las wikis [Leuf 2001] o las folksonomías [Golder 2006].

Como se vio en la Aproximación Conceptual, las ideas que están detrás de las referencias indicadas y algunas de las técnicas utilizadas en ellas se han empleado en la aproximación propuesta, pero además se han tenido en cuenta otras contribuciones de las áreas de conocimiento donde dichas referencias se sitúan y de los campos de investigación vinculados con las mismas y con el entorno del trabajo: Gestión del Conocimiento, Interacción Persona-Ordenador, Web Semántica, y Minería de Información y de Datos. Entre las ideas y técnicas consideradas, algunas tienen especial relevancia.

Del primer campo de conocimiento mencionado, la Gestión de Conocimiento: los fundamentos de la los sistemas de gestión de conocimiento [Malthotra 2000] preocupados por facilitar el trabajo colaborativo [Churman 1971] de los usuarios y la consideración de las personas como parte del conocimiento gestionado [Ackerman 2003].

Del segundo de los campos mencionados, la Interacción Persona-Ordenador (IPO): las bases de los sistemas de Groupware [Churman 1971] y del Awareness [Dourish 1992], para soportar la actividad en grupo y su seguimiento [Sohlenkamp 1994] y para adaptarse a las situaciones de

uso [Byun 2001]; los mecanismos de Monitorización de Actividad, mediante eventos de la interfaz de usuario [Hilbert 2000] de las páginas Web [DOM 2008], para obtener información de la actividad de los usuarios; los fundamentos de Hipermedia Adaptativa [Brusilovsky 1998], para el modelado de los elementos del sistema, y el seguimiento y la adaptación del entorno mediante documentos virtuales [Milosavljevic 1999]; las técnicas de los sistemas de recomendación [Adomavicius 2005], ya citados entre los antecedentes, como medio de aprovechar la actividad residual de los usuarios; la Visualización de la Información [Geroimenko 2002], para facilitar la interacción con estructuras complejas de datos; el Procesamiento del Lenguaje Natural, para el tratamiento de la información textual [Baeza 2004]; y la filosofía de la Web 2.0 [O'Reilly 2005], ya mencionada como antecedente.

Del tercero de los campos enumerados, la Web Semántica [Berners-Lee 2001]: el uso de ontologías [Mahesh 1996], para representar el conocimiento y el propio sistema; el empleo de agentes [Dinverno 2001], para dotar al sistema de pro-actividad y capacidad gregaria; y la generación dinámica de documentos [Milosavljevic 1999], vinculada también con IPO, para la adaptación de los contenidos a los requerimientos de su uso.

Finalmente, del cuarto de los campos referidos, la Minería de Información y de Datos [Chang 2001]: las técnicas de Minería de Texto, del área de recuperación de la información y de agrupación y categorización de textos, mediante el empleo de modelos vectoriales [Baeza 1999], para el análisis del conocimiento; las ideas del Análisis de la Semántica Latente [Landauer 1998], como referencia para el estudio de las vinculaciones ocultas en el seno del conocimiento; y las técnicas de Minería Web [Chang 2001], para analizar la interfaz y los registros de actividad del sistema.

Tomando en consideración todo lo dicho, el objetivo general de este trabajo de investigación es probar que existen vías de solución para la sobrecarga de información en el ámbito de los sistemas de gestión de conocimiento en red, entorno representativo de otros más generales donde se produce el mencionado problema, sacando partido para ello de la energía de la actividad de las entidades activas de dichos sistemas y las propiedades de todos los elementos involucrados en ellos.

Para conseguirlo se proponen cuatro objetivos operativos experimentales, dirigidos a indagar en algunos de los aspectos esenciales de la actividad de los sistemas de gestión de conocimiento [Benjamins 1999]: la recolección de conocimiento, la organización y estructuración del mismo, su refinamiento y mantenimiento, y la distribución de dicho conocimiento entre destinatarios necesitados de él. El primer objetivo experimental (OE1) consiste en seleccionar el contenido por su calidad, minimizando los requerimientos de manifestaciones explícitas de opinión de los usuarios sobre él. El segundo de dichos objetivos (OE2) radica en evidenciar el conocimiento



que está latente en el repositorio, para enriquecer con él el conocimiento ya explícito y mejorar así la gestión del conjunto. El tercer objetivo experimental (OE3) es conseguir nuevo conocimiento de la actividad de los usuarios con el sistema, para incorporarlo también en el repositorio del sistema y aprovecharlo en la gestión del conocimiento que se lleva a cabo en el mismo. Por último, el cuarto de los objetivos experimentales (OE4) consiste en establecer redes de conocimiento entre los nodos del sistema de forma automática y autónoma a través de la Web.

El sistema de gestión de conocimiento Semantic KnowCat (SKC) se propone para aplicar la aproximación planteada y comprobar los aspectos fundamentales de la misma, condensados en los anteriores objetivos operativos experimentales. Para ello, SKC sigue la aproximación a la gestión colaborativa del conocimiento sin supervisión del sistema KnowCat (KC) incorporando las ideas y las técnicas principales mencionadas.

KC es un sistema de trabajo en grupo que facilita la gestión de un repositorio de conocimiento mediante la interacción de una comunidad de usuarios a través de la Web [Alamán 1999][Cobos 2003]. Esto se puede hacer sin supervisión empleando información sobre la actividad de los usuarios y sus opiniones sobre los documentos que forman parte de la base de conocimiento.

SKC sigue la filosofía de KC en el modo de procurar la gestión del conocimiento, aunque intenta conseguirlo: reduciendo la necesidad de que los usuarios manifiesten explícitamente su opinión sobre el conocimiento; aumentando el aprovechamiento de la información generada por el sistema, los usuarios y otras entidades con las que aquél se relaciona; y ampliando la explotación de las características de los elementos implicados en la actividad, como la comunidad, el conocimiento y la Web. Todo esto no sería posible sin la aplicación combinada en SKC de las ideas y técnicas de los diversos campos de investigación indicados. Aunque los diseños de ambos sistemas son muy distintos y SKC requiere de capacidades sensoriales, analíticas y expresivas más sofisticadas que las de KC, estas diferencias no han impedido que KC se haya empleado como plataforma de prototipado de SKC.

Con ello se ha podido disponer de un modelo funcional de SKC dotado de las características necesaria para alcanzar los objetivos experimentales propuestos, utilizado el KC como núcleo operativo, al que se han añadido los módulos imprescindibles para conferirle las nuevas funcionalidades, pero respetando el funcionamiento básico del sistema huésped. Esta estrategia de prototipado ha permitido emplear el modelo de SKC en experimentos realizados en el marco de programas de actividades de KC y utilizar repositorios de conocimiento del KC original en ensayos con el prototipo, además de posibilitar el desarrollo y la prueba de las funcionalidades pretendidas de SKC con los recursos existentes. De esta forma, se han prototipado tres

componentes de SKC sobre KC, el Monitor de Cliente (MC), el Módulo de Análisis de Conocimiento (MAC) y el Módulo de Red (MR).

El primero de ellos es el Monitor de Cliente (MC), que se encarga de registrar la interacción de cada usuario con el sistema de forma más detallada que los registros convencionales de los sistemas Web. MC registra los eventos de bajo nivel que la interacción del usuario con el sistema provoca en la interfaz del cliente, hace llegar sus percepciones al fichero de registro del servidor Web que soporta SKC y analiza dicho LOG para establecer la relación entre los rastros de la interacción y los elementos de conocimiento del sistema. Como consecuencia el MC obtiene dos resultados, un índice de interacción (I) por elementos de conocimiento, que parece representar el interés que cada elemento despierta en la comunidad de usuarios, y un vector de interés por cada usuario (VIU), que parece encarnar la distribución del interés de cada usuario entre los elementos de conocimiento del sistema.

El segundo de los componentes prototipados es el Módulo de Análisis de Conocimiento (MAC), que se ocupa de procesar el conocimiento y los datos recogidos por el sistema, para obtener información oculta y valiosa con la que enriquecer el conocimiento del repositorio o facilitar su gestión. Como resultado el MAC elabora unos Vectores de Peso de Palabra (VPP) para cada elemento -ítem- de conocimiento del sistema, a partir de los textos asociados con los mismos, que permiten comparar unos elementos con otros y establecer vínculos entre ellos. La asociación de textos con los elementos de conocimiento no es siempre directa, como sí pasa con los documentos que son textuales o los elementos que se les asignan descripciones directamente, y el MAC debe considerar las relaciones entre los distintos tipos de elementos para conseguir textos representativos de los mismos. Por ejemplo, los temas que agrupan elementos descritos textualmente se consideran definidos por la unión de los mismos, o los autores que han creado una serie de documentos, por la unión de las correspondientes descripciones. De este modo, cada ítem de conocimiento puede estar calificados por varios VPPs, relativos a distintos aspectos de su naturaleza o actividad. Así, cada usuario puede tener asociado un VPP en relación con su labor como autor y otro distinto vinculado con sus intereses personales respecto a los temas que constituyen el conocimiento del sistema.

Este último caso nos ha interesado especialmente en el prototipado del MAC, porque la vinculación que liga a los usuarios con los elementos de conocimiento, a través del interés, no es tan directo como ocurre en otras ocasiones -donde cada usuario era o no autor de un documento, o cada documento estaba o no en un cierto tema-. Los vectores de interés de usuario, VIU, generados con el MC, establecen una relación entre los usuarios y los ítems de conocimiento en forma de proporción -donde unos documentos interesan más que otros en el total del interés de cada usuario-. Los VIU se utilizan en el MAC para generar descripciones de los usuarios, proporcionales a sus intereses, en base a las descripciones de los elementos objetos

de su atención. Los descriptores correspondientes son los Vectores de Peso de Palabra de Interés de Usuario (VPPIU), que son VPPs que permiten relacionar los usuarios de SKC con otros ítems de conocimiento del sistema o de fuera de él, que tengan asociados VPPs.

Por último, el tercer componente de SKC prototipado es el Módulo de Red (MR), que permite establecer vínculos automáticos entre distintas instancias del sistema repartidas por la Web, con el fin de implantar y mantener una red de conocimiento entre todas ellas. Para conseguirlo, cada nodo del sistema tiene un agente que desarrolla varias tareas de forma autónoma y proactiva: localizar nodos en la Web, contactar con ellos, comprobar su actividad, obtener la información básica sobre los mismos, determinar el tipo de vinculación con ellos atendiendo a sus contenidos, obtener la información necesaria para hacerlo, desactivar nodos inactivos y revisar de forma reactiva las vinculaciones con todos los nodos de la red. Para la localización de otros nodos en la Web los agentes interaccionan entre ellos, y también con buscadores Web. Además, cada agente atiende las demandas de los demás y aprovecha las interacciones de los desconocidos para identificarlos y establecer relaciones con ellos. La vinculación entre nodos es distinta dependiendo de la similitud entre los temas de los que se ocupan. Cuando los nodos tratan de temas parecidos, los nodos pueden establecer relaciones entre los ítems de conocimiento de sus respectivos repositorios. Para todo ello se emplean los VPPs generados por los MACs de los distintos nodos.

Los módulos de SKC prototipados han servido para realizar cuatro grupos de experimentos diseñados para probar dichos componentes y las propuestas subyacentes, a través de los objetivos operativos planteados.

El primer grupo (GE1) emplea el prototipo del MC y trata de la medida de la interacción de los usuarios en el lado del cliente del sistema, con ello se indaga en aspectos como la intensidad de la interacción de los usuarios, que permite obtener una impresión sobre el interés de los mismos, sin necesidad de manifestaciones explícitas sobre aquel, para conseguir una selección tentativa del conocimiento sin expresiones directas de opinión. Los resultados de estos experimentos contribuyen fundamentalmente a tres de los cuatro objetivos experimentales presentados antes, OE1, OE2 y OE3.

El segundo grupo de experimentos (GE2) realizados utiliza el prototipo del MAC e investiga la calificación semántica de ítems de conocimiento, con el fin de establecer asociaciones automáticas entre ellos, realizar recomendación sobre los mismos y proporcionar nuevas vistas del conocimiento disponible en el sistema. Los resultados de este grupo apoyan básicamente los objetivos experimentales OE2 y OE4.

El tercer grupo de experimentos (GE3) llevados a cabo utiliza de forma combinada el prototipo del MAC y el del MC, y se ocupan, más en detalle, del interés de los usuarios, en concreto de su

descubrimiento, de su representación y de su aprovechamiento. Los resultados de los experimentos de este grupo se dirigen esencialmente a los objetivos OE2 y OE3.

Finalmente, el cuarto grupo de experimentos (GE4) efectuados emplea los prototipos del MR y del MAC, y se dedica al establecimiento y mantenimiento de una red de instancias del sistema SKC en la Web, con el objetivo de abrir el sistema al exterior, enriquecer su conocimiento y estimular a sus usuarios. Los resultados de tales experimentos contribuyen directamente al objetivo OE4, pero también apoyan el OE2.

Todos los experimentos se han realizado con datos obtenidos de experiencias que se han incorporado en diversas actividades docentes realizadas con KC y con los prototipos de SKC en la Escuela Politécnica Superior (EPS) de la Universidad Autónoma de Madrid (UAM). Las experiencias se han efectuado en el ámbito de dos Proyectos de Innovación Docente (PIDs), financiados por dicha universidad durante cinco años académicos, entre los cursos 2002/2003 y 2006/2007. Los PIDs de la UAM promueven la actualización de los procedimientos docentes tradicionales y han servido para fomentar la adecuación de los mismos a los requerimientos del proceso de convergencia hacia el Espacio Europeo de Educación Superior. Las experiencias se han diseñado cuidadosamente con un doble propósito: por un lado, contribuir al proceso de enseñanza/aprendizaje de forma efectiva; y por otro lado, crear las condiciones adecuadas para obtener datos significativos para esta investigación. En particular se han empleado en este trabajo las experiencias realizadas con las asignaturas de la EPS: Sistemas Operativos, con 400 alumnos anuales, entre los cursos 2002/2003 y 2004/2005; Teoría de Autómatas y Lenguajes Formales II, con una matrícula de entre 80 y 90 estudiantes por año, entre los cursos 2002/2003 y 2006/2007; Inteligencia Artificial el curso 2004/2005, con unos 300 alumnos ese año; y Sistemas Informáticos II, con 300 estudiantes al año, los cursos 2004/2005 y 2005/2006. En total las experiencias realizadas han involucrado cuatro asignaturas de la EPS, con más de 2000 alumnos en total, a lo largo de cinco años.

En la Tabla 5-1 se resumen el objetivo general del trabajo, los cuatro objetivos experimentales planteados y los tres componentes prototipados del sistema SKC. Los cuatro grupos de experimentos realizados a lo largo de la investigación, indicando sus vinculaciones con los objetivos y los componentes, se sintetizan en la Tabla 5-2.

Tabla 5-1. Resumen de objetivos y de componentes prototipadas

<b>Objetivo General</b>	
Probar que existen vías de solución para la sobrecarga de información en el ámbito de los sistemas de gestión de conocimiento en red, entorno representativo de otros más generales donde se produce el mencionado problema, sacando partido para ello de la energía de la actividad de las entidades activas de dichos sistemas y las propiedades de todos los elementos involucrados en ellos.	
<b>Objetivo Experimental</b>	<b>Descripción</b>
OE1	Seleccionar el contenido por su calidad, minimizando los requerimientos de manifestaciones explícitas de opinión de los usuarios sobre él.
OE2	Evidenciar el conocimiento que está latente en el repositorio, para enriquecer con él el conocimiento ya explícito y mejorar así la gestión del conjunto.
OE3	Conseguir nuevo conocimiento de la actividad de los usuarios con el sistema, para incorporarlo también en el repositorio del sistema y aprovecharlo en la gestión del conocimiento que se lleva a cabo en el mismo.
OE4	Establecer redes de conocimiento entre los nodos del sistema de forma automática y autónoma a través de la Web.
<b>Componente SKC Prototip.</b>	<b>Descripción</b>
MC	Monitor de Cliente, que se encarga de registrar la interacción de cada usuario con el sistema de forma más detallada que los registros convencionales de los sistemas Web.
MAC	Módulo de Análisis de Conocimiento, que se ocupa de procesar el conocimiento y los datos recogidos por el sistema, para obtener información oculta y valiosa con la que enriquecer el conocimiento del repositorio o facilitar su gestión.
MR	Módulo de Red, que permite establecer vínculos automáticos entre distintas instancias del sistema repartidas por la Web, con el fin de implantar y mantener una red de conocimiento entre todas ellas.

Tabla 5-2. Resumen de grupos de experimentos

Grupo Experimentos	Objetivos Experimentales	Componentes SKC Prototip.	Descripción
GE1	OE1, OE2 y OE3	MC	Medir la intensidad de la interacción de los usuarios observando la interacción de los mismos en el cliente del sistema, para investigar como conseguir una selección del conocimiento sin expresiones directas de opinión.
GE2	OE2 y OE4	MAC	Indagar en la calificación semántica de ítems de conocimiento, con el fin de establecer asociaciones automáticas entre ellos, realizar recomendación sobre los mismos y proporcionar nuevas vistas del conocimiento disponible en el sistema.
GE3	OE2 y OE3	MAC y MC	Investigar en la manifestación implícita del interés de los usuarios, en su representación y en su aprovechamiento
GE4	OE2 y OE4	MR y MC	Probar el establecimiento y el mantenimiento de una red de instancias del sistema SKC, para abrir el sistema al exterior, enriquecer su conocimiento y estimular a sus usuarios.

En la Tabla 5-3 se resumen los catorce experimentos realizados con indicación de sus adscripciones a los grupos presentados anteriormente y sus correspondientes descripciones.

Tabla 5-3. Resumen experimentos realizados

Grupo Experimentos	Experimento	Descripción
GE1	<b>E11</b>	Selección automática de conocimiento sin expresión directa de opinión a partir de la actividad de los usuarios con dicho contenidos a través SKC con actividad docente en una asignatura de Inteligencia Artificial.
	<b>E12</b>	Selección automática de conocimiento sin expresión directa de

		opinión a partir de la actividad de los usuarios con dichos contenidos a través SKC con actividades docentes en una asignatura de Sistemas Operativos.
GE2	<b>E211</b>	Clasificación automática de ítems de conocimiento en nodo SKC.
	<b>E212</b>	Agrupación automática de ítems de conocimiento por temas en nodo SKC en un nodo sobre Sistemas Operativos.
	<b>E213</b>	Agrupación automática de ítems de conocimiento por temas en nodo SKC en un nodo sobre Sistemas Informáticos.
	<b>E221</b>	Mapeo automático entre temas de nodos SKC sobre Sistemas Informáticos.
	<b>E222</b>	Mapeo automático entre temas de nodos SKC sobre Teoría de Autómatas y Lenguajes Formales.
	<b>E231</b>	Agrupación de nodos SKC de Sistemas Informáticos, Sistemas Operativos, y Teoría de Autómatas y Lenguajes Formales.
GE3	<b>E311</b>	Identificación del interés de los usuarios en un nodo de SKC sobre Sistemas Informáticos.
	<b>E312</b>	Evolución del interés de los usuarios en un nodo de SKC sobre Teoría de Autómatas y Lenguajes Formales.
	<b>E321</b>	Descubrimiento de contenidos interesantes para los usuarios de un nodo SKC.
GE4	<b>E411</b>	Establecimiento y mantenimiento automático y autónomo de una red de nodos SKC mediante agentes software.
	<b>E412</b>	Establecimiento y mantenimiento automático y autónomo de vínculos entre nodos de una red SKC mediante agentes software.
	<b>E421</b>	Descubrimiento automático y autónomo de nodos SKC desconocidos en la Web para su integración en una red SKC mediante agentes software y buscadores convencionales de la Web.

El grupo de experimentos GE1 consta de dos experiencias E11 y E12, dirigidas a selección automática de conocimiento sin expresión directa de opinión, a partir de la actividad de los usuarios con dicho contenidos en SKC. El grupo de experimentos GE2 consiste en dos series: la

primera incluye las experiencias E211, E212 y E213, orientadas a clasificación y agrupación automática de ítems de conocimiento en nodos SKC; la segunda serie contiene las experiencias E221 y E222, ocupadas del mapeo automática entre temas de nodos SKC. El grupo de experimentos GE3 está constituido por dos series también: la primera de ellas incluye las experiencias E311 y E312, dirigidas a identificación del interés de los usuarios y de la evolución del mismo en nodos de SKC; la segunda tiene sólo la experiencia E321, orientadas al descubrimiento de contenidos interesantes para los usuarios de un nodo SKC. El grupo de experimentos GE4 está formado por dos series como las anteriores: la primera de ellas incluye las experiencias E411 y E412, que se ocupan del establecimiento y del mantenimiento automático y autónomo de una red de nodos SKC y de sus vínculos mediante agentes software; la segunda sólo contiene la experiencia E421, dirigida al descubrimiento automático y autónomo de nodos SKC desconocidos en la Web para su integración en una red SKC mediante agentes software y buscadores convencionales de la Web.

En la Tabla 5-4 se resumen los resultados obtenidos en los experimentos realizados a lo largo de la investigación presentada en este documento. Los experimentos aparecen agrupados en series, cuando contribuyen a resultados conjuntos. Después de la tabla hay una valoración de los resultados por grupos de experimentos en relación con los objetivos establecidos anteriormente.

Tabla 5-4. Resumen resultados obtenidos en los experimentos realizados por series

Experimento	Resultado
<p><b>E11</b></p> <p><b>E12</b></p>	<p>Como resultado de los experimentos de selección automática de conocimiento sin expresión directa de opinión, a partir de la actividad de los usuarios con dicho contenidos en SKC:</p> <ul style="list-style-type: none"> <li>• Los coeficientes de intensidad de interacción (I), que se calculan para cada ítem de conocimiento del sistema analizando la actividad de los usuarios con ellos, pero sin considerar manifestaciones explícitas de opinión sobre los mismos, parecen ser indicadores del interés de los usuarios por dichos elementos.</li> <li>• Existe una correlación entre los grados de cristalización (C) de los ítems de conocimiento considerados en los experimentos -documentos- y los coeficientes I obtenidos para ellos. En KC el índice C permite ordenar los documentos por su calidad, teniendo en cuenta la actividad de los usuarios en el sistema y las manifestaciones explícitas de opiniones emitidas por los usuarios.</li> <li>• Sin embargo, hay algunas discrepancias entre C e I en documentos puntuales. Por un lado, hay algunos documentos poco valorados explícitamente -C bajo-, que son objeto de mucha actividad -I alta-. Esto parece indicar que dichos documentos despiertan interés, pero luego no cubren las expectativas</li> </ul>



	<p>provocadas. Por otro lado, hay documentos muy valorados explícitamente -C alto-, que son objeto de poca interacción -I baja-. Esto puede significar, llevado al límite, que se han valorado positivamente documentos sin suficiente interacción con ellos, lo que parece indicar una evaluación inadecuada de los mismos.</p> <ul style="list-style-type: none"> <li>• En definitiva, el coeficiente I, indicador de interés implícito por los elementos de conocimiento, parece comportarse como una aproximación de C, indicador de opinión explícita de calidad de dichos ítems, en condiciones de gran motivación -inducida en los experimentos-, puesto que los usuarios parece que tienden a interactuar más con los contenidos que consideran mejores y menos con el resto, en tales situaciones. Sin embargo, la motivación de los usuarios es fundamental para que ambos coeficientes resulten significativos.</li> <li>• Sería deseable contar con un indicador de la motivación de los usuarios.</li> </ul>
<p><b>E211</b> <b>E212</b> <b>E213</b></p>	<p>Como resultado de los experimentos de clasificación y agrupación automática de ítems de conocimiento en nodos SKC:</p> <ul style="list-style-type: none"> <li>• El procedimiento propuesto de calificar los elementos de conocimiento mediante Vectores de Peso de Palabra (VPP), elaborados a partir de los textos asociados a tales ítems, genera descriptores adecuados para la clasificación de documentos en temas de forma automática.</li> <li>• Así mismo, dicho procedimiento basado en los VPPs crea descriptores adecuados para la agrupación automática de los documentos por temas.</li> <li>• Sin embargo, el umbral encontrado que separa los ítems similares de los que no lo son resulta bajo -entorno a 0,3-, siendo raro que aparezcan valores mucho mayores, y no hay un contraste muy marcado entre los índices de similitud de los elementos parecidos y los que no los son.</li> <li>• El valor del umbral de similitud no parece fijo, sino dependiente de cada caso.</li> <li>• Sería interesante contar con un procedimiento automático para establecer el umbral de similitud.</li> <li>• También sería deseable disponer de una medida de similitud que marcara mejor el contraste entre ítems similares y diferentes.</li> </ul>
<p><b>E221</b> <b>E222</b></p>	<p>Como resultado de los experimentos de mapeo automática entre temas de nodos SKC:</p> <ul style="list-style-type: none"> <li>• El procedimiento empleado de calificar temas mediante VPPs, elaborados a partir de la unión de los textos asociados a los ítems de conocimiento incluidos en cada tema, genera descriptores propicios para el mapeo automático entre temas equivalentes con distintos documentos clasificados en ellos.</li> </ul>

	<ul style="list-style-type: none"> <li>• Así mismo, el procedimiento propuesto para calificar los temas con VPPs, resulta adecuado para el mapeo automático de árboles de conocimiento sobre la misma temática pero con estructuras diferentes.</li> <li>• La medida de similitud y el umbral correspondiente manifiestas los mismos inconvenientes que en los experimentos de la anterior serie.</li> </ul>
<b>E231</b>	<p>Como resultado del experimento de agrupación de nodos SKC:</p> <ul style="list-style-type: none"> <li>• El procedimiento empleado para asignar descriptores VPP a los nodos, a partir de la unión de los textos asociados a los elementos integrantes de su árbol de conocimiento, agrupa nodos con temáticas similares y permite distinguir los dedicados a temas diferentes.</li> <li>• La medida de similitud y el umbral correspondiente manifiestan los mismos problemas que en las anteriores series de experimentos.</li> </ul>
<b>E311</b> <b>E312</b>	<p>Como resultado de los experimentos de identificación del interés de los usuarios y de la evolución del mismo en nodos de SKC:</p> <ul style="list-style-type: none"> <li>• El procedimiento propuesto de calificar a los usuarios del sistema mediante Vectores de Interés de Usuario (VIU), elaborados a partir del análisis de la intensidad de la interacción de cada uno por los elementos de conocimiento del nodo, crea descriptores adecuados para identificar los focos de interés de cada usuario en el árbol de conocimiento del nodo SKC donde se desarrolla la actividad.</li> </ul>
<b>E321</b>	<p>Como resultado del experimento de descubrimiento de contenidos interesantes para los usuarios de un nodo SKC:</p> <ul style="list-style-type: none"> <li>• El procedimiento propuesto de calificar a los usuarios del sistema mediante Vectores de Peso de Palabra de Interés de Usuario (VPPIU), elaborados para cada usuario a partir de los VIUs y de las descripciones de los elementos de conocimiento del sistema, genera descriptores adecuados para identificar agrupar usuarios con intereses afines.</li> <li>• Además, el procedimiento propuesto para calificar los usuarios con VPPIUs, resulta adecuado para identificar elementos de información relacionados con los temas de interés de los usuarios dentro y fuera del sistema.</li> <li>• La medida de similitud y los umbrales de similitud están afectados de los mismos inconvenientes que en el anterior grupo de experimentos.</li> </ul>
<b>E411</b>	<p>Como resultado de los experimentos de establecimiento y mantenimiento automático y autónomo de una red de nodos SKC y de sus vínculos mediante agentes software:</p>

<b>E412</b>	<ul style="list-style-type: none"><li>• El procedimiento propuesto permite establecer y mantener una red de nodos SKC con intercambio selectivo de conocimiento empleando agentes software que representan a los distintos nodos.</li><li>• Los agentes son capaces de establecer una red de nodos y adaptarse a los cambios de la misma colaborando, actuando sobre el entorno y reaccionando a sus cambios de forma reactiva.</li></ul>
<b>E421</b>	<p>Como resultado del experimento de descubrimiento automático y autónomo de nodos SKC desconocidos en la Web para su integración en una red SKC mediante agentes software y buscadores convencionales de la Web:</p> <ul style="list-style-type: none"><li>• El procedimiento propuesto para descubrir nodos desconocidos en la Web cumple su cometido y parece ser un potente mecanismo de integración de redes de conocimiento entre instancias de SKC.</li></ul>

## 5.2 Conclusiones y Principales Contribuciones

En general, en todas las experiencias realizadas se han obtenido los resultados esperados, con lo que se pueden considerar cubiertos los objetivos experimentales propuestos. Como se ha visto estos objetivos inciden en algunos de los aspectos esenciales de la actividad de los sistemas de gestión de conocimiento [Benjamins 1999] -recolección, organización, estructuración, refinamiento, mantenimiento y distribución del conocimiento-, por lo que a través de ellos se alcanza también el objetivo general planteado.

El primer grupo de experimentos (GE1) ha puesto de manifiesto las posibilidades de un procedimiento desatendido, sin grandes requerimientos técnicos, para aproximar la selección del conocimiento por su calidad, analizando la actividad normal de los usuarios con el sistema, pero prescindiendo de manifestaciones explícitas de opinión de este colectivo sobre el contenido. Dichos experimentos han recurrido al conocimiento latente que existía en el SKC para obtener nuevo conocimiento con el que enriquecer el existente y facilitar la gestión realizada por el sistema. En este caso se han aprovechado la estructura jerárquica del árbol de conocimiento y la organización de áreas excluyentes de la interfaz de la aplicación, para establecer vínculos entre los eventos percibidos y los elementos de conocimiento afectados por la interacción. En definitiva, con los experimentos se ha mostrado cómo conseguir conocimiento nuevo a partir del existente y de la actividad de los usuarios, y cómo aplicarlo en la mejora de la gestión del conocimiento realizada por el sistema. Sin embargo, la monitorización en el cliente sólo se puede llevar a cabo en determinadas condiciones y bajo restricciones importantes que no se pueden aceptar, superar o imponer en todos los casos. Además, la motivación de los usuarios

parece imprescindible para que el interés se manifieste de forma adecuada, como también ocurre para la cristalización de conocimiento en KC.

Una selección de la información sin expresiones directas de opinión, en base a la interacción de una comunidad de usuarios interesados por ella, supone una aproximación muy interesante para intentar reducir el impacto de la sobrecarga de información en sistemas saturados, que reúnan algunas de las características del sistema de gestión de conocimiento propuesto y se utilicen en un contexto propicio. Entre las propiedades requeridas por el sistema están: utilizar una interfaz Web; trabajar con usuarios identificados dispuestos a que se monitorice su actividad; disponer de una interfaz estructurada, que permita el seguimiento de los elementos de información objetivos de la selección. El contexto debe de cumplir condiciones como: número de usuarios mínimo -adecuado a la cantidad de ítems de información en el repositorio-; proporción de usuarios por elemento de contenido variable según el tipo de información tratada, la dinámica del grupo y las actividades establecidas con el sistema; y comunidad con gran motivación y disponibilidad para trabajar; y objetivos y procedimientos bien establecidos.

El segundo grupo de experimentos (GE2) ha evidenciado las oportunidades que aporta un análisis sencillo de los contenidos, para revelar conocimiento oculto en el sistema. Para ello se han aprovechado la naturaleza textual de los elementos fundamentales de conocimiento - documentos-, las relaciones existentes entre estos y otros tipos de ítems de conocimiento dentro del correspondiente árbol, y los vínculos establecidos entre todos ellos con los usuarios - resultado de la actividad de éstos con los elementos de conocimiento del sistema-. Los descriptores correspondientes -basados en VPPs- han proporcionado un medio para automatizar la organización -clasificación, estructuración, ordenación...- del conocimiento y el descubrimiento de relaciones ocultas entre los elementos que lo constituyen, así como han aportado nuevas oportunidades de acceso a ese conocimiento y de enriquecimiento del mismo. Además, dichos descriptores han facilitado el soporte para algunos de los requerimientos esenciales de la implantación de la red de conocimiento entre nodos de SKC, proporcionando un medio básico para la identificación automática de nodos con temáticas similares y el mapeo entre árboles de conocimiento -ontologías en la versión definitiva de SKC-. Sin embargo, el umbral de similitud, que determina cuándo se empiezan a considerar parecidos o no los elementos confrontados, varía según la temática de los contenidos que se comparan, suele ser un valor relativamente bajo en el rango del coeficiente correspondiente -que está entre 0 y 1-, carece de mucho contraste -las cosas parecidas no se diferencian de forma demasiado evidente de las que no lo son-, y debido a todo esto, su valor tiene que fijarse empíricamente, lo que dificulta la automatización de su empleo.

El tercer grupo de experiencias (GE3) ha mostrado cómo se puede sacar provecho de la manifestación implícita del interés particular de los usuarios -los descriptores VIU- por los

elementos de conocimiento del sistema, para enriquecer el conocimiento vinculado a los usuarios -que también son elementos de conocimiento en SKC- y brindar nuevas oportunidades de mejorar la interacción con la aplicación y también la gestión realizada por la misma. Además, la combinación del VIU de cada usuario con los VPPs de los elementos de conocimiento del sistema, proporciona unos nuevos descriptores para los primeros -VPPIUs-. Estos descriptores representan el interés de los usuarios de forma que puede compararse con los VPPs de otros ítems de conocimiento del sistema -incluidos los de otros usuarios-, o asociados a cualquier contenido externo, para identificar información adaptada a los intereses de cada individuo. La agrupación de usuarios por intereses o la construcción de mapas de interés de los repositorios de conocimiento son aplicaciones interesantes de estos descriptores. Sin embargo, los umbrales de similitud vuelven a ser un inconveniente como en casos anteriores donde se emplean VPPs y la efectividad de los VPPIUs disminuye cuando los usuarios no tienen intereses muy focalizados.

Para terminar, el cuarto grupo de experiencias (GE4) ha demostrado cómo se puede aprovechar el conocimiento ajeno, latente en el entorno del sistema, más allá de los límites de los nodos, para enriquecer el conocimiento propio disponible, e intentar mejorar el desempeño del sistema procurando estimular a los usuarios del mismo con referencias documentales foráneas relevantes, que puedan marcar contrastes con los elementos de conocimiento autóctonos, mostrar puntos de vista alternativos y romper posibles prácticas de “endogamia” intelectual a las que son propensas los nodos aislados. Para ello ha resultado fundamental el conocimiento revelado con los descriptores que se han probado en experimentos anteriores, así como el aprovechamiento de las características del sistema, derivadas de sus condiciones de aplicación Web y de entidad activa. Sin embargo, en este caso también se manifiestan los problemas con los umbrales de similitud, de los que se ha hablado anteriormente en este apartado. Además, resulta un inconveniente, fuera del ámbito experimental, las restricciones de uso de los buscadores Web genéricos para los usuarios no humanos.

En definitiva, los experimentos realizados parecen evidenciar que existen vías para paliar la sobrecarga de información en los sistemas de gestión de conocimiento en red aprovechando la energía residual de la actividad de las entidades activas de estos sistemas, y las propiedades de todos los elementos involucrados en ellos. Además, esta conclusión es extensible a otros entornos de gestión de información con características parecidas a los sistemas objeto de esta investigación.

Entre las contribuciones que se desprenden de los resultados de los experimentos realizados destacan cinco especialmente. Primero, el mecanismo de monitorización de interacción de usuarios y de registro de actividad en el servidor, y el proceso para determinar los elementos de conocimiento objetos de la actividad percibida. Segundo, la medida de la intensidad de

interacción de la comunidad de usuarios (I) a partir de la monitorización de los eventos de la interfaz del sistema y el registro de actividad del mismo. Tercero, la aproximación a la selección automática y desasistida de conocimiento por su calidad, basada en el índice interés de la comunidad de usuarios por los ítems de conocimiento del sistema. Cuarto, la representación del interés de los usuarios en el nodo mediante VIUs y la proyección del mismo dentro y fuera de dicho nodo mediante los VPPIUs. Quinto, la propuesta de convertir las instancias del sistema en entidades artificiales activas en la Web y establecer un ámbito en el que dichas entidades colaboren con los usuarios humanos de forma activa.

Conjuntamente, como se indicó en el capítulo de Aproximación Conceptual, es evidente que en este trabajo se ha hecho un gran esfuerzo en integrar ideas y técnicas de diversos campos de conocimiento. El problema de la sobrecarga de información se ha tratado desde distintas áreas de investigación de forma más o menos independiente, así que este trabajo innova en el planteamiento multidisciplinar de la aproximación propuesta. Este enfoque está justificado por el carácter interdisciplinario del ámbito del problema y la variedad de aspectos que concurren en él. La Web Semántica establece un área de convergencia de gran parte de campos de conocimiento considerados en este proyecto, con el objetivo de hacer los datos de la Web apropiados para el procesamiento automático. La Web 2.0 también constituye un punto de encuentro de algunas de dichas disciplinas, en este caso con la intención de facilitar el empleo de la Web para individuos y grupos. La tendencia que se manifiesta en estos dos ejemplos, de buscar en la integración de campos diversos soluciones para los problemas que la Web plantea, apoyan la aproximación propuesta.

Al mismo tiempo, como se ha propuesto a lo largo del trabajo la Web podría verse como un gran sistema de gestión de conocimiento desde algunos puntos de vista. Este enfoque, tal como se dijo en el capítulo tres, al mismo tiempo que afirma la función como repositorio de datos de la red, reivindica el rol fundamental de la comunidad de usuarios en ella, que también forma una red, pero activa, y establece un propósito fundamental para el conjunto, que es colaborar para sacar provecho de la información y el conocimiento. Este planteamiento aporta nuevas oportunidades de encarar algunos problemas que la Web plantea y de sacar beneficio de ella, algo que también resulta innovador. En concreto, como se defiende en este trabajo, el problema de sobrecarga de información en la Web podría mejorar aprovechando la energía excedente de la actividad de los usuarios, de las aplicaciones y de las entidades artificiales que concurren en la red, y con la explotación de las características de la propia red, y de las entidades involucradas, como los usuarios, la información y el conocimiento.

Además, como también se señaló en ese capítulo, la aproximación presentada proporciona un marco experimental para los campos de investigación que involucra, en especial para el área de gestión de conocimiento, de trabajo colaborativo, y de la Web. El sistema SKC propuesto

proporciona un entorno con unas características que pueden resultar de interés para investigar en los campos mencionados. Por un lado, se trata de un medio acotado de trabajo en grupo sobre la Web. Por otro lado, es un entorno en el que participan interlocutores humanos y automáticos con objetivos claros y concurrentes. Además, el sistema tiene una utilidad práctica evidente. Este entorno de prueba constituye también una contribución interesante para las áreas de investigación implicadas.

Finalmente, como concluyó la Propuesta de la Aproximación Conceptual, la Web futura incorporará las aportaciones de la Web 2.0, de la Web Semántica y de otras referencias que aparezcan con el tiempo. La aproximación propuesta en este trabajo espera contribuir en este proceso, mostrando un punto de confluencia de ideas y tecnologías diversas para la gestión del conocimiento y, a través de ello, para el tratamiento de la sobrecarga de información en la Web y otros ámbitos que compartan características parecidas con ella.

### **5.3 Difusión de los Resultados**

Los resultados obtenidos en esta tesis se han difundido en varios foros y publicaciones a lo largo de los últimos años. Los trabajos realizados han dado lugar a varias contribuciones a workshops y congresos: una en el Workshop de Investigación sobre nuevos paradigmas de interacción en entornos colaborativos aplicados a la gestión y difusión del Patrimonio cultural COLINE'02 [Alamán 2002]; varias comunicaciones en los Congresos de Interacción Persona-Ordenador, AIPO, de los años 2004 [Moreno 2004], 2005 [Moreno 2005b], 2007 [Moreno 2007b] [Dumitrescu 2007] y 2008 [Moreno 2008b], que se ha celebrado en el marco del Congreso Español de Informática CEDI los años 2005 y 2007; una comunicación a las II Jornadas Nacionales de Metodologías ECTS [Cobos 2007]; y una ponencia en la III Jornada Pedagógica del Proyecto ADA-Madrid [Moreno 2008a]. Además, algunos de los resultados obtenidos en esta investigación se han publicado como capítulos en dos libros de la editorial Springer [Moreno 2005][Moreno 2007a] y las actividades docentes que han servido como base de los experimentos se han descrito en un capítulo editado por la Universidad Autónoma de Madrid (UAM) [Cobos 2008].

Por otra parte, la concepción del sistema Semantic KnowCat (SKC), el desarrollo de los prototipos presentados y la difusión de los trabajos investigación realizados con ellos han sido financiados parcialmente por el Ministerio de Ciencia y Tecnología, a través de los proyectos ARCADIA (TIC2002-01948), UCAT (TIN2004-03/40) y Mosaic (TSI2005-08225-C07-06).

Además, entre 2003 y 2007, el sistema KnowCat (KC), con algunos de los módulos prototipados del sistema SKC presentados en este trabajo, ha sido explotado en el marco de los Proyectos de Innovación Docente financiados por la UAM, que ha permitido realizar los experimentos descritos a lo largo del documento: "Plataforma KnowCat para la gestión

colaborativa de materiales docentes en red” (curso 2003-2004) y “Aprendizaje activo y tutelado en la generación colaborativa de materiales docentes en la Web con la asistencia del sistema KnowCat” (cursos 2004-2005, 2005-2006 y 2006-2007).

#### **5.4 Cuestiones Abiertas y Trabajos Futuros**

Entre las cuestiones que quedan abiertas en esta investigación pueden destacarse cuatro. La primera de ellas es conseguir un indicador de la motivación de los usuarios que sirva para evaluar de forma automática la bondad de coeficientes que dependen de dicha actitud, como la cristalización de conocimiento (C) o la intensidad de la interacción, como indicador de interés. La segunda cuestión abierta es encontrar una medida de la similitud entre elementos, que muestre de forma más intensa la diferencia entre los elementos parecidos y los que no lo son. La tercera cuestión, que está ligada con la anterior, consiste en disponer de un procedimiento automático para identificar el umbral de similitud entre parejas de temas. Por último, la cuarta cuestión abierta, encontrar alternativas al empleo de los buscadores Web genéricos para el descubrimiento automático de repositorios de conocimiento en la red.

Como resultado de este trabajo se han identificado seis líneas de investigación interesantes en el ámbito del proyecto. La primera de ella es el empleo del Análisis de la Semántica Latente para establecer los vínculos entre los elementos de conocimiento, como se hace ahora con las técnicas basadas en espacios vectoriales. La segunda línea de investigación es profundizar en la representación del interés de los usuarios, especialmente en lo que se refiere al tratamiento de la diversidad, donde el modelo actual parece tener limitaciones. La tercera línea identificada es la de visualización del conocimiento y la interacción con él, que constituyen aspectos críticos de su aprovechamiento. La cuarta línea considerada es profundizar en el empleo de los mecanismos de recomendación como medio de colaboración y de aprovechamiento de la información y el conocimiento implícitos en la actividad del grupo. La quinta línea identificada es la del empleo de la conciencia de grupo -awareness- como instrumento de fomentar la vinculación social de los usuarios en contexto del entorno colaborativo. Finalmente, la sexta línea de investigación es la de simulación de sistemas colaborativos, en los que es muy difícil realizar determinados experimentos por los requerimientos de colaboración de usuarios que resultan imprescindibles.

Para terminar, como continuación del trabajo presentado en este documento se están desarrollando varios proyectos vinculados con las tres últimas líneas de investigación enunciadas. El primero de ellos es el desarrollo de un simulador de la actividad de usuarios humanos en un sistema KnowCat, que emplea el Monitor de Cliente de SKC descrito en este trabajo para el registro de la interacción con el sistema de los usuarios virtuales. Este prototipo se está llevando a cabo en colaboración con profesores y alumnos de un curso de postgrado sobre “Simulación de Sistemas Complejos” de las Escuela Politécnica Superior (EPS) de la



Universidad Autónoma de Madrid (UAM). El segundo es un proyecto financiado por la Agencia Española de Cooperación Internacional (AECI) en su convocatoria de 2007 (A/7954/07), que se está realizando este año, y lleva por título “Entorno basado en tecnologías de la información y las comunicaciones (TICs) para monitorizar y analizar los procesos colaborativos en actividades con contenidos educativos”. El proyecto se ha incorporado funcionalidades de awareness en un sistema KnowCat que incorpora algunos de los resultados de este trabajo. En este proyecto participan la Universidad Nacional de Colombia y la Universidad del Cauca, por parte de Colombia, y la Universidad de Lleida y la UAM, por parte de España. El tercero es un proyecto de la Comisión Interministerial de Ciencia y Tecnología (CICYT) que ha sido aprobado recientemente (TIN2008-02081/TIN), titulado “METEORIC: Meta-Tool Environments for Model-Oriented Collaborative Web Applications” y que se llevará a cabo durante los próximos tres años. En dicho proyecto se desarrollará una herramienta de composición y generación de aplicaciones colaborativas, a partir de un repositorio de componentes que incluirán elementos basados en los resultados de este trabajo.

# **R**referencias

- [3store 2008] **3store**. <http://sourceforge.net/projects/threestore>. (Consultado enero 2008)
- [43Things 2008] **43 Things**. <http://www.43things.com/>. (Consultado febrero 2008)
- [Ackerman 2003] M. Ackerman, V. Pipek, V. Wulf. **Sharing Expertise, Beyond Knowledge Management**. MIT Press. 2003
- [ACM 2008] **Association for Computing Machinery**. <http://www.acm.org/>. (Consultado abril 2008)
- [Adomavicius 2005] G. Adomavicius, A. Tuzhilin. **Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions**. IEEE Transactions on Knowledge and Data Engineering, 2005, No.17, pp. 734-749.
- [Alamán 1999] X. Alamán, R. Cobos. **KnowCat: a Web Application for Knowledge Organization**. Proceedings of the World-Wide Web and Conceptual Modeling (WWWCM'99). París, Francia. Noviembre, 1999. In Lecture Notes in Computer Science 1727, P.P. Chen, et.al. (eds). Springer, 1999, pp. 348-359.
- [Alamán 2002] X. Alamán, R. Cobos, J. Moreno. **Una propuesta para la gestión colaborativa del conocimiento**. Proceedings of the Workshop de Investigación sobre nuevos paradigmas de interacción en entornos colaborativos aplicados a la gestión y difusión del Patrimonio cultural, COLINE'02. Granada, Spain, November 11-12, 2002.
- [Alavi 2001] Alavi, M. & Leidner, D.E. **Knowledge management and knowledge management systems: Conceptual foundations and research issues**. MIS Quarterly, 25(1), 2001, pp. 107-136.
- [Amann 2002] B. Amann, C. Beerli, I. Fundulaki, M. Scholl. **Ontology-Based Integration of XML Web Resources**. Proceedings of the ISWC2002, 2002.
- [Amazon 2008] **Amazon.com**. <http://www.amazon.com/>. (Consultado abril 2008)
- [Apache 2008] **The Apache Software Foundation**. <http://www.apache.org/>. (Consultado abril 2008)
- [ARCADIA 2002] M. Alfonso. **Memoria Científico-Técnica del Proyecto ARCADIA**. Documento interno del Dpto. Informática de la EPS de la UAM. Memoria proyecto sin publicar, 2002.
- [Baeza 1999] R. Baeza-Yates, B. Ribeiro. **Modern Information Retrieval**. Addison Wesley, 1999.
- [Baeza 2004] R. Baeza-Yates. **Challenges in the interaction of information retrieval and natural language processing**. Proceedings of the 5th international conference on

Computational Linguistics and Intelligent Text Processing (CICLing), volume 2945 of Lecture Notes in Computer Science. Springer, 2004, pp. 445-456.

[Baeza 2006] R. Baeza-Yates. **The Challenges and the Sciences of the Social Web**. [http://www.cibersociedad.net/public/k3\\_multimedia/ricardo\\_baeza.pdf](http://www.cibersociedad.net/public/k3_multimedia/ricardo_baeza.pdf). 2006. (Consultado abril 2008)

[Bargerón 2004] D. Bargerón, J. Grudin. **As Users Grow More Savvy: Experiences with an Asynchronous Distance Learning Tool**. Proceedings of the HICSS 2004, 2004.

[Barra 2002] M. Barra, P. Maglio, A. Negro, V. Scarano. **GAS: Group Adaptive System**. In Proceedings of the 2nd Adaptive Hypermedia and Adaptive Web-Based Systems (AH2002), Málaga, Spain, 2002, pp. 47-57.

[Benjamins 1998 a] V. R. Benjamins, D. Fensel, **Problem-Solving Methods**. International Journal of Human-Computer Studies as guest editorial of a special issue on Problem-Solving Methods, vol. 49, no. 4, 1998, pp. 305-313.

[Benjamins 1998 b] V. R. Benjamins, E. Plaza, E. Motta, D. Fensel, R. Studer, B. Wielinga, G. Schreiber, Z. Zdrahal, and S. Decker. **Ibrow3: An intelligent brokering service for knowledge-component reuse on the world-wide web**. Proceedings of the 11th Banff Knowledge Acquisition for knowledge-Based System Workshop (KAW98), Banff, Canada, 1998.

[Benjamins 1999] V. R. Benjamins, D. Fensel, S. Decker, A. Gómez. **(KA)2: building ontologies for the Internet: a mid-term report**. International Journal of Human-Computer Studies, vol. 51, iss. 3, September 1999, pp. 687-712.

[Berners-Lee 2000] T. Berners-Lee. **Semantic Web - XML2000**. <http://www.w3.org/2000/Talks/1206-xml2k-tbl>. 2000. (Consultado agosto 2008)

[Berners-Lee 2001] T. Berners-Lee, J. Hendler, and O Lassila. **The Semantic Web**. Scientific American, May 2001.

[Bordini 2006] R. Bordini, L. Braubach, M. Dastani, A. El Fallah Seghrouchni, J. Gomez-Sanz, J. Leite, G. O'Hare, A. Pokahr, A. Ricci. **A Survey of Programming Languages and Platforms for Multi-Agent Systems**. Informatica 30, 2006, pp. 33-44

[Brants 2004] T. Brants. **Natural Language Processing in Information Retrieval**. Proceedings of CLIN 2004 Antwerp, Belgium, 2004, pp. 1-13.

[Broekstra 2002] J. Broekstra, M. Klein, S. Decker, D. Fensel, F. van Harmelen, I. Horrocks. **Enabling knowledge representation on the Web by extending RDF Schema**. Computer Networks, vol. 39, iss. 5, 5 august 2002, pp. 609-634.

- [Brusilovsky 1998] P. Brusilovsky. **Methods and Techniques of Adaptive Hypermedia**. In P. Brusilovsky, A. Kobsa, J. Vassileva (eds.). Adaptive Hypertext and Hypermedia. Kluwer Academic Publishers, 1998, pp. 1-43.
- [Brusilovsky 2001] P. Brusilovsky. **Adaptive Hypermedia**. User Modeling and User-Adapted Interaction, vol. 11, no. 1/2, Kluwer Academic Publishers, 2001, pp. 87-110.
- [Byun 2001] H. E. Byun, K. Cheverst. **Exploiting User Models and Context-Awareness to Support Personal Daily Activities**. Proceedings of the Workshop on User Modeling for Context-Aware Applications at the 8th International Conference on User Modeling (UM2001), Sonthofen, Germany, 2001.
- [Caldwell 2000] N. Caldwell, J. Clarkson. **Web-Based Knowledge Management for Distributed Design**. IEEE Intelligent Systems, 2000, pp. 40 – 47.
- [Cadiz 2000] J.J. Cadiz, A. Gupta, J. Grudin. **Using Web Annotation for Asynchronous Collaboration Around Documents**. Proceedings of CSCW 2000, 2000, pp. 309-318.
- [Cantador 2007] I. Cantador, P. Castells, and A. Bellogín. **Modelling Ontology-based Multilayered Communities of Interest for Hybrid Recommendations**. 1st Workshop on Adaptation and Personalisation in Social Systems: Groups, Teams, Communities, at the 11th International Conference on User Modeling (UM 2007). Corfu, Greece, 2007.
- [Calvo 2002] A. Calvo, M. A. Serrano-Tenllado, A. Maroto, J. A. Romero del Castillo. **OSHI y WebOSHI: Sistema para compartir y reusar ontologías de fuentes documentales históricas**. Proceedings of the III Jornadas de Bibliotecas Digitales, El Escorial (España), 2002, pp. 67-76.
- [Carreras 2004] X. Carreras, I. Chao, L. Padró, M. Padró. **FreeLing: An Open-Source Suite of Language Analyzers**. Proceedings of LREC 2004 Lisbon, Portugal. 2004.
- [Castells 2005] P. Castells, M. Fernández, D. Vallet, P. Mylonas, and Y. Avrithis. **Self-Tuning Personalized Information Retrieval in an Ontology-Based Framework**. 1st IFIP WG 2.12 & WG 12.4 International Workshop on Web Semantics (SWWS 2005), November 2005. R. Meersman, Z. Tari, and P. Herrero (Eds.), Springer Verlag Lecture Notes in Computer Science, vol. 3762, 2005, pp. 977-986.
- [Castells 2007] P. Castells, M. Fernández, and D. Vallet. **An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval**. IEEE Transactions on Knowledge and Data Engineering 19(2), special issue on Knowledge and Data Engineering in the Semantic Web Era, 2007, pp. 261-272.
- [Chakrabarti 1998] S. Chakrabarti. **Automatic Resource Compilation by Analyzing Hyperlink Structure and Associate Text**. Proceedings of the 7th WWW Conference, 1998.

- [Chandrasekaran 1999] B. Chandrasekaran, J.R. Josephson, V.R. Benjamins. **What are ontologies, and why do we need them.** Intelligent Systems, IEEE, pp 20-26.
- [Chang 2001] G. Chang, M. Healey, J. McHugh, J. Wang. **Mining the World Wide Web: An introduction search approach.** Kluwer Academic Publishers, 2001.
- [Chau 1999] M. Y. Chau. **Web Mining Technology and Academic Librarianship: Human-Machine Connections for the Twenty-First Century.** First Monday, vol. 4, no. 6, 1999.
- [Choi 2006] N. Choi, I. Song, H. Han. **A survey on ontology mapping.** ACM SIGMOD Record archive, vol. 35, no. 3, 2006, pp. 34 - 41.
- [Churman 1971] C.W. Churman. **The Desing of Inquiring Systems.** Basic Books, New York, NY, 1971.
- [ClickHeat 2008] **Labsmedia: ClickHeat.** <http://www.labsmedia.com/clickheat/index.html> (Consultado abril 2008)
- [ClickHeat 2008] **Cloudmark.** <http://www.cloudmark.com/> (Consultado abril 2008)
- [Coakes 2003] Coakes, E. **Knowledge Management: Current Issues and Challenges.** Idea Group Publishing, 2003.
- [Cobos 2002a] R. Cobos, X. Alamán. **Cristalización del conocimiento de una comunidad de usuarios.** Proccidings of the IV Congreso Internacional de Interacción Persona-Ordenador, Madrid, Spain, 2002, pp. 128-135.
- [Cobos 2002b] R. Cobos, J. Esquivel, X. Alamán. **IT Tools for Knowledge Management: A Study of the Current Situation.** Journal of Novática and Informatik/Informatique, special issue on Knowledge Management, vol. III, no. 1, 2002.
- [Cobos 2003] R. Cobos. **Mecanismos para la cristalización del conocimiento, una propuesta mediante un sistema de trabajo colaborativo.** Tesis doctoral. Escuela Politécnica Superior, Universidad Autónoma de Madrid. 2003.
- [Cobos 2004] R. Cobos, M. Pifarré. **Learning among Equals in the Net: Analysis of KnowCat supporting group work.** In R. Navarro-Prieto, et. al. (eds.). HCI related papers of Interacción 2004, Springer, Dordrecht, The Netherlands, 2005, pp. 277-287
- [Cobos 2005] R. Cobos, F. Díez. **Una Experiencia Docente de Trabajo Colaborativo en Inteligencia Artificial.** Proceedings of SINTICE 2005 (Workshop on Information and Communication Technology in Education), Granada, Spain, 2005, pp. 11-18.
- [Cobos 2007] R. Cobos, X. Alamán, J. Moreno. **Aprendizaje activo y tutelado en la generación colaborativa de materiales docentes en la Web con la asistencia del sistema**

**KnowCat**. Proccidings of the II Jornadas Nacionales de Metodologías ECTS, Extremadura, September, 2007.

[Cobos 2008] R. Cobos, X. Alamán, J. Moreno. **Memoria del Proyecto Aprendizaje Activo y Tutelado en la Generación Colaborativa de Materiales Docentes en la Web con la Asistencia del Sistema KnowCat**. In A. Álvarez-Ossorio. El Proceso de Aprendizaje en la Universidad Autónoma de Madrid: La Implantación del Crédito Europeo. UAM Ediciones, 2008, pp. 271-292.

[Coleman 1997] D. Coleman. **Groupware: Collaborative Strategies for Corporate LANs and Intranets**. Prentice Hall, Upper Saddle River, NJ, 1997.

[Crazyeeg 2008] **Crazyeeg: Visualize your visitors**. <http://crazyegg.com/>. (Consultado abril 2008)

[CSS 2008] **Cascading Style Sheets Home Page**. <http://www.w3.org/Style/CSS/>. (Consultado abril 2008)

[Davenport 1999] T. Davenport. **Ecología de la Información**. Oxford University Press, 1999.

[DelIcioUs 2008] **Del.icio.us**. <http://del.icio.us/>. (Consultado enero 2008)

[Denny 2004] M. Denny. **Ontology Tools Survey, Revisited**. O'Reilly xml.com. <http://www.xml.com/pub/a/2004/07/14/onto.html>. (Consultado enero 2008)

[DESIRE 2008] **DESIRE**. <http://www.few.vu.nl/~wai/demas/tools2.html>. (Consultado enero 2008)

[Diaper 1989] D. Diaper. **The discipline of human-computer interaction**. In *Interacting with computers*, no. 1, vol. 1, Butterworth-Heinemann Ltd., Guildford, Reino Unido, 1989.

[Dieng 2000] R. Dieng. **Knowledge management and the internet**. *Intelligent Systems*, IEEE, vol 15, iss. 3, 2000, pp. 14-17.

[Díez 2007] F. Díez, R. Cobos. **A Case Study of a Cooperative Learning Experience in Artificial Intelligence**. In *International Journal Computer Applications in Engineering Education* 15(4), 2007, pp. 308-316.

[Dinverno 2001] M. d'Inverno, M. Luck. **Understanding Agent Systemas**. Springer, 2001.

[DOM 2008] **Document Object Model (DOM) Level 3 Core Specification**. <http://www.w3.org/DOM/>. (Consultado abril 2008)

[DOMEvent 2008] **Document Object Model (DOM) Level 2 Events Specification**. <http://www.w3.org/TR/DOM-Level-2-Events/>. (Consultado abril 2008)

[Domingue 2000] J. Domingue, E. Motta. **PlanetOnto : From News Publishing to Integrated Knowledge Management Support**. IEEE Intelligent Systems, 2000, pp. 26-31.

[Doyle 1997] P. Doyle. **Natural Language**.  
<http://www.cs.dartmouth.edu/~7Ebrd/Teaching/AI/Lectures/Summaries/natlang.html>. 1997.  
(Consultado Enero 2008)

[Dourish 1992] P. Dourish, V. Bellotti. **Awareness and coordination in shared workspaces**. In Proceedings of the ACM CSCW'92 Conference on ComputerSupported Cooperative Work, Toronto, Canada, November 1992, pp. 107-114.

[Duineveld 2000] A. J. Duineveld, R. Stoter, M. R. Weiden, B. Kenepa, V. R. Benjamins. **WonderTools? A comparative study of ontological engineering tools**. International Journal of Human-Computer Studies, vol. 52, iss. 6, 2000, pp. 1111-1133

[Dumitrescu 2007] D.A. Dumitrescu., R. Cobos, J. Moreno. **Sistema multiagente para la extracción y análisis de la interacción de los usuarios de un sistema colaborativo**. Proceedings of the VIII Congreso Internacional de Interacción Persona-Ordenador (INTERACCION 2007), Zaragoza, Spain, September 2007, pp. 127-129.

[Earl 2001] M.J. Earl. **Knowledge management strategies: Toward a taxonomy**. Journal of Management Information Systems, vol. 18, no. 1, 2001, p.p. 215-233.

[eBay 2008] **eBay**. <http://www.ebay.com/> (Consultado abril 2008)

[eMule 2008] **eMule**. <http://www.emule.com/> (Consultado agosto 2008)

[Etzioni 1995] O. Etzioni, D. Weld. **Intelligent agents on the Internet: Fact, fiction and forecast**. IEEE Expert, vol. 10, no. 4, 1995, pp. 44-49.

[Emerick 2002] J. Emerick. **Managing XML data storage**. ACM Crossroads, vol 8, Iss. 4, 2002, pp. 6 - 11.

[EyetrackIII 2008] **Eyetrack III: Online news consumer behavior in the age of multimedia**. <http://www.poynterextra.org/eyetrack2004/index.htm>. (Consultado abril 2008)

[Fellbaum 1999] C. Fellbaum (ed.). **WordNet: An Electronic Lexical Database**. MIT Press, 1999.

[Fensel 1999 a] D. Fensel , V.R. Benjamins, S. Decker, M. Gaspari, R. Groenboom, W. Grosso, M. Musen, E. Motta, E. Plaza, G. Schreiber, R. Studer, B. Wielinga. **The Component Model of UPML ina a Nutshell**. Proceedings of the First Working IFIP Conference on Software Architecture (WICSA1), San Antonio, Texas (USA), 1999.



[Fensel 1999 b] D. Fensel, V.R. Benjamins, E. Motta, B. Wielinga. **UPML: A Framework for knowledge system reuse**. Proceedings of the International Joint Conference on AI (IJCAI-99), Stockholm (Sweden), 1999.

[Fensel 2000] D. Fensel, I. Horrocks, F. Van Harmelen, S. Decker, M. Erdmann, M. Klein. **OIL in a nutshell: Knowledge Acquisition, Modeling, and Management**. Proceedings of the European Knowledge Acquisition Conference (EKAW-2000). R. Dieng et al. (eds.), Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, 2000.

[Fensel 2001] D. Fensel. **Ontologies: A silver bullet for knowledge Management and electronic commerce**. Springe, 2001.

[Fenstermacher 2002a] K. Fenstermacher, M. Ginsburg. **Mining Client-Side Activity for Personalization**. Presented at the Fourth Workshop on Advanced Issues in Electronic Commerce and Web Information Systems (WECWIS) 2002. Newport Beach, California. 2002.

[Fenstermacher 2002b] K. Fenstermacher, M. Ginsburg. **A Lightweight Framework for C/S Monitoring**. IEEE Computer. vol. 35, no. 3, march 2002.

[Finin 1994] T. Finin, R. Fritzson, D. McKay, R. McEntire. **KQML as an Agent Communication Language**. Proceedings of the Third International Conference on Information and Knowledge Management (CIKM94), ACM Press, 1994.

[FIPA 2008] **FIPA**. <http://www.fipa.org>. (Consultado enero 2008)

[FIPA-OS 2008] **FIPA-OS Agent Toolkit**. <http://sourceforge.net/projects/fipa-os/>. (Consultado enero 2008)

[Flickr 2008] **Flickr**. <http://www.flickr.com/>. (Consultado enero 2008)

[ForakerDesign 2005] Foraker Design. **Usability First: Your Online Guide to Usability Resources**. <http://www.usabilityfirst.com/index.txt> (Consultado abril 2008)

[Franklin 1994] S. Franklin, A. Graesser. **Is it an agent, or just a program?: A taxonomy for autonomous agents**. In J.P. Müller, M.J. Wooldridge, and N.R. Jennings (eds.). Intelligent Agents III-Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages. Lecture Notes in Artificial Intelligence, Springer-Verlag, vol. 1193, 1997, pp. 21-35.

[Frasincar 2002] F. Frasincar, G.J. Houbert. **Hypermedia Presentation Adaptation on Semantic Web**. Proceedings of the The 2nd Adaptive Hypermedia and Adaptive Web-Based Systems (AH2002), Málaga, Spain, 2002, pp. 133-142.

[Freeling 2008] **Freeling**. <http://garraf.epsevg.upc.es/freeling/>. (Consultado enero 2008)

- [Freire 2004] M. Freire, P. Rodríguez. **A graph-based interface to complex hypermedia structure visualization**. Proceedings of AVI '04, Gallipoli, Italy, 2004, pp. 163-166.
- [Gansner 2000] E. R. Gansner, S. North. **An Open Graph Visualization system and its Applications to Software Engineering**. Software - Practice Experience 30, 1203-1233, 2000.
- [Garofalakis 1999] M. N. Garofalakis, R. Rastogi, S. Seshadri, K. Shim. **Data mining and the Web: Past, present and future**. In Proceedings of the WIDM99 Conference, 1999, Kansas City, Missouri (USA), pp. 43-47.
- [GATE 2008] Gate. <http://gate.ac.uk/>. (Consultado enero 2008)
- [García 2006] G. García, R. Cobos. **ESMAP: A Multi-Agent Platform for Extending a Knowledge Management System**. In T. Nishida, et. al. (eds.). Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence. Hong Kong, Diciembre 2006. IEEE Computer Society. 2006. pp. 59-65.
- [Gea 2002] M. Gea, F.L. Gutiérrez, J.L. Garrido, J.J. Cañas. **AMENITIES: Metodología de Modelado de Sistemas Cooperativos**. Proceedings of the COLINE'02 : Investigación En Entornos De Interacción Colectiva, 2002.
- [Gea 2004] M. Gea, J.L. Garrido, F.L. Gutierrez, R. Cobos, X. Alamán. **Representación del comportamiento dinámico en modelos colaborativos: aplicación a la gestión del conocimiento compartido**. Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. no. 24, 2004, pp. 87-95.
- [Geroimenko 2002] V. Geroimenko, C. Chen (Eds.). **Visualizing the Semantic Web**. Springer, 2002.
- [Gil 2002] Y. Gil, V. Ratnakar. **A Comparison of (Semantic) Markup Languages**. Proceedings of the International FLAIRS Conference, Pensacola Beach, Florida, USA, 2002.
- [GNU 2008] **GNU Operating System**. <http://www.gnu.org/>. (Consultado abril 2008)
- [GNU 2008b] **GNU Wget Reference**. <http://www.gnu.org/software/wget/>. (Consultado abril 2008)
- [Gómez 2001] M. Gómez, A. Gutiérrez, R. Cobos, X. Alamán. **El Aprendizaje Colaborativo con Soporte Informático en el Diseño de Material para Desarrollo del Pensamiento Abstracto en Educación Infantil. Una Experiencia en Didáctica de las Matemáticas**. Tercer Simposio Internacional de Informática Educativa (SIIE). Viseu, Portugal, 2001.
- [Google 2008] **Google**. <http://www.google.com>. (Consultado Abril 2008)
- [Golder 2006] S. Golder and BA Huberman. **The structure of collaborative tagging systems**. Journal of Information Science, vol. 32, no. 2, 2006, pp. 198-208.

- [Goldfarb 1999] C. F. Goldfarb, P. Prescod. **Manual de XML**. Prentice, 1999.
- [Gottschalk 2004] P. Gottschalk. **Strategic Knowledge Management Technology**. Idea Group Publishing, 2004.
- [Gruber 1993] T. R. Gruber. **A Translation Approach to Portable Ontology Specifications**. Knowledge Acquisition, vol. 5, iss. 2, 1993, pp. 199-220.
- [Grudin 1994a] J. Grudin. **Groupware and social dynamics: eight challenges for developers**. Communications of the ACM archive, vol. 37, 1994, pp. 92-105.
- [Grudin 1994b] J. Grudin. **CSCW: History and Focus**. IEEE Computer, vol. 27, no.5, Mayo, 1994, pp. 19-26.
- [Grudin 2003] J. Grudin. **Leaders leading? A shift in technology adoption**. CHI Extended Abstracts 2003, pp. 930-931.
- [Gruninger 2002] M. Gruninger, J. Lee. **Ontology applications and design: Introduction**. Communications of the ACM, February, 2002.
- [Guarino 1998] N. Guarino. **Formal Ontologies and Information Systems**. Amsterdam: IOS Press, 1998.
- [Gutierrez 2005] F.L. Gutiérrez, J.L. Garrido. **Diseño de Sistemas Interactivos**. In In tutorials of the V Congreso Internacional Interacción Persona-Ordenador (INTERACCIÓN 2005), Granada, Spain, 2005.
- [GutierrezR 2005] R. M. Gutiérrez. **Análisis Semántico Latente: ¿Teoría psicológica del significado?**. Signos, vol. 38, no. 59, 2005, pp. 303-323.
- [Harmelen 2000] F. Harmelen, I. Horrocks. **FAQs on OIL: The Ontology Inference Layer**. IEEE Intelligent Systems, vol. 15, no. 6, nov./dic. 2000, pp. 67-73.
- [Hassan 2007] Y. Hassan-Montero, V. Herrero-Solana. **Eye-Tracking en Interacción Persona-Ordenador**. No Solo Usabilidad Journal, no. 6, oct. 2007.
- [Hendler 2000] J. Hendler, D. L. McGuinness. **The DARPA Agent Markup Language**. IEEE Intelligent Systems, vol. 15, no. 6, nov./dec. 2000, pp. 67-73.
- [Herman 2007] I. Herman. **State of the Semantic Web**. Keynote given at the International Conference on Semantic Web & Digital Libraries. Bangalore, India. 2007. <http://www.w3.org/2007/Talks/0223-Bangalore-IH/Slides.pdf>. (Consultado enero 2008)
- [Hezen 2002] N. Hezen, W. Nejdl. **Knowledge Modeling for Open Adaptive Hypermedia**. Proceedings of the 2nd Adaptive Hypermedia and Adaptive Web-Based Systems (AH2002), Málaga, Spain, 2002, pp. 174-183.

- [Hilbert 2000] D. Hilbert, D. Redmiles. **Extracting Usability Information from User Interface Events**. ACM Comput.Surv, vol. 32, no. 4, dec. 2000, pp. 384-421.
- [Hjelm 2001] J. Hjelm. **Creating the Semantic Web with RDF**. Wiley, 2001.
- [Holland 1975] J. H. Holland. **Adaptation in Natural and Artificial Systems**. The University of Michigan Press, Ann Harbor, MI, 1975.
- [Http 2008] **HTTP - Hypertext Transfer Protocol**. <http://www.w3.org/Protocols/>. (Consultado febrero 2008)
- [Hunt 1966] E. B. Hunt, J. Marin, P. J. Stone. **Experiments in Induction**, Academic Press New York, 1966.
- [Hudson 1996] S. E. Hudson, I. E. Smith. **Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems**. Proceedings of the 1996 ACM conference on Computer Supported Cooperative Work (CSCW 1996), Boston, Massachusetts, USA, 1996, pp. 248-257.
- [Huysmans 2004] J. Huysmans, B. Baesens, J. Vanthienen. **Web usage mining: a practical study**. Proceedings of the Twelfth Conference on Knowledge Acquisition and Management (KAM2004), Kule, Poland ,2004.
- [IDC 2007] IDC White Paper sponsored by EMC. **The Expanding Digital Universe: A Forecast of Worldwide Information Growth through 2010**. <http://www.emc.com/leadership/digital-universe/expanding-digital-universe.htm?hpid=1>. 2007. (Consultado abril 2008)
- [ISWC 2007] **ISWC 2007**. <http://www.iswc07.org/main/default.asp>. (Consultado enero 2008)
- [JADE 2008] **JADE, Java Agent Development Framework**. <http://jade.tilab.com>. (Consultado enero 2008)
- [JavaScript 1999] **Client-Side JavaScript Reference**. <http://docs.sun.com/source/816-6408-10/contents.htm>. (Consultado abril 2008)
- [Jena 2008] **Jena: A Semantic Web Framework for Java**. <http://jena.sourceforge.net/>. (Consultado enero 2008)
- [Jennings 1998] N.R. Jennings, K. Sycara, M.J. Wooldridge. **A roadmap of agent research and development**. Autonomous Agents and Multi-Agent Systems, vol. 1, no.1, 1998, pp. 7-38.
- [Johansen 1998] R. Johansen. **Groupware: Computer Support for Business Teams**. New York: The Free Press, 1988.

- [Jorge-Botana 2006] G. Jorge-Botana. **El Análisis de la Semántica Latente y su aportación a los estudios de Usabilidad**. No Solo Usabilidad journal, no. 5, 2006.
- [Jorge-Botana 2007] G. Jorge-Botana, R. Olmos, J. A. León. **Análisis de la Semántica Latente (LSA) y estimación automática de las intenciones del usuario en diálogos de telefonía (call routing)**. FAZ, no. 1, 2007.
- [Jurafsky 2000] D. Jurafsky, J. H. Martin. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. Prentice-Hall, 2000.
- [Kiryakov 2004] A. Kiryakov, B. Popov, I. Terziev, D. Manov, D. Ognyanoff, **Semantic Annotation, Indexing, and Retrieval**. Journal of Web Semantics 2, iss. 1, Elsevier, 2004, pp. 49-79.
- [Kleinberg 1998] J. Kleinberg. **Authoritative sources in hyperlinked environment**. In Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms. San Francisco, California, USA, 1998, pp. 25-27.
- [Kleinberg 1999] J. Kleinberg. **Authoritative sources in hyperlinked environment**. Extended version in Journal of ACM, no. 46, 1999.
- [Kobsa 1993] A. Kobsa. **User Modeling: Recent Work, Prospects and Hazards**. In M. Schneider-Hufschmidt, T. Kühme and U. Malinowski (eds.). Adaptive User Interfaces: Principles and Practice. North-Holland, Amsterdam, 1993.
- [Kohavi 2001] R. Kohavi, F. Provost. **Applications of Data Mining to Electronic Commerce**. Data Mining and Knowledge Discovery, vol. 1, 2001, pp. 1-7.
- [Kuhnila 2002] J. Kuhnila, M. Miettinen, P. Nokelainen, H. Tirri. **EDUCO: A Collaborative Learning Environment Based on Social Navigation**. Proceedings of the 2nd Adaptive Hypermedia and Adaptive Web-Based Systems (AH2002), Málaga, Spain, 2002, pp. 242-252.
- [Landauer 1998] T. K. Landauer, P. W. Foltz, D. Laham. **An introduction to Latent Semantic Analysis**. Discourse Processes, no. 25, 1998, pp. 259-284.
- [Lassila 2000] O. Lassila. **The Resource Description Framework**. IEEE Intelligent Systems, vol. 15, no. 6, nov./dec. 2000, pp. 67-73.
- [Leuf 2001] B. Leuf, W. Cunningham. **The Wiki Way. Quick collaboration on the Web**. Addison-Wesley, 2001
- [LogWeb 2008] **Logging Control In W3C httpd**. <http://www.w3.org/Daemon/User/Config/Logging.html>. (Consultado febrero 2008)

[Lorés 2001] J. Lorés, J. Abascal, J.J. Cañas, M. Gea, A.B. Gil, A.B. Martínez Prieto, M. Ortega, P. Valero, M. Vélez. **La interacción persona-ordenador**. AIPO, Lleida, Spain, 2001. (Electronic book: <http://griho.udl.es/ipo/ipo/libroe.html>)

[Malthotra 2000] Y. Malthotra. **From Information Management to Knowledge Management: Beyond the 'Hi-Tech Hidebound' Systems**. In K. Srikantaiah, M.E.D Koenig (Eds.). Knowledge Management for the Information Professional. Medford, N.J.:Information today Inc. pp.37-61.

[Martins 2002] L. C. Martins, T. A. S. Coelho, S. D. J. Barbosa, M. A. Casanova, C. J. P. de Lucena. **A Framework for Filtering and Packaging Hypermedia Documents**. Proceedings of the 2nd Adaptive Hypermedia and Adaptive Web-Based Systems (AH2002), Málaga, Spain, 2002. pp. 275-283.

[McDermott 1999] R. McDermott. **Why Information Technology Inspired but Cannot Deliver Knowledge Management**. California Management Review, vol. 41, no. 4, 1999, pp.103-117.

[Melnik 2001] S. Melnik. **Storing RDF in a relational database**. <http://infolab.stanford.edu/~melnik/rdf/db.html>. (Consultado enero 2008)

[Milosavljevic 1999] M. Milosavljevic, F. Vitali, C. Watters. **Workshop on Virtual Documents, Hypertext Functionality and the Web**. At the Eighth International World Wide Web Conference. Toronto, Canada, May 1999.

[Mahesh 1996] K. Mahesh. **Ontology Development for Machine Translation: Ideology and Methodology**. NMSU. Computing Research Laboratory. Technical Report MCCS-96-292. New Mexico, USA, 1996.

[Medina 2000] N. Medina-Medina, L. García-Cabrera, M. J. Rodríguez-Fortiz, J. Parets-Llorca. **Adaptation in an Evolutionary Hypermedia System: Using Semantic and Petri Nets**. Proceedings of the 2nd Adaptive Hypermedia and Adaptive Web-Based Systems (AH2002), Málaga, Spain, 2002, pp. 284-295.

[Minsky 1969] M. Minsky & S. Papert. **Perceptrons**. The MIT Press, Cambridge, MA, 1969.

[Moreno 2004] J. Moreno, X. Alamán, R. Cobos. **Propuesta de Diseño para la Gestión Colaborativa del Conocimiento Mediante Información Semántica**. Proceedings of V Congreso Internacional de Interacción Persona-Ordenador. Lleida, Spain, may 3-7, 2004, pp. 332-340.

[Moreno 2005] J. Moreno, X. Alamán. **A Proposal of Design for a Collaborative Knowledge Management System by means of Semantic Information**. In R. Navarro-Prieto, et. al. (eds.).

HCI related papers of Interacción 2004, Springer, Dordrecht, The Netherlands, 2005, pp. 307-319.

[Moreno 2005b] J. Moreno, X. Alamán, R. Cobos. **SKC: Midiendo la Intensidad de la Interacción de los usuarios**. Proceedings of the VI Congreso Internacional de Interacción Persona-Ordenador, Granada, Spain, sep. 2005, pp. 89-98.

[Moreno 2007a] J. Moreno, X. Alamán. **SKC: Measuring the user's interaction intensity**. In B. Fernández-Manjón, et. al. Computers and Education: E-learning, From Theory to Practice. Springer, 2008, pp. 123-132.

[Moreno 2007b] J. Moreno., X. Alamán. **SKC: Digestión de Conocimiento**. Proceedings of the VIII Congreso Internacional Interacción Persona-Ordenador (INTERACCION 2007), Zaragoza, Spain, 2007, pp. 281-290.

[Moreno 2008a] J. Moreno, R. Cobos, X. Alamán. **La asistencia del sistema Knowcat en la generación colaborativa de material docente en la Web**. Online proceedings of the III Jornada Pedagógica del Proyecto ADA-Madrid, 2008. [http://moodle.upm.es/adamadrid/file.php/1/web\\_III\\_jornadas\\_ADA/Comunicaciones/Moreno%20Llorena.pdf](http://moodle.upm.es/adamadrid/file.php/1/web_III_jornadas_ADA/Comunicaciones/Moreno%20Llorena.pdf). (Consultado agosto 2008)

[Moreno 2008b] J. Moreno, X. Alamán. **Red de Nodos de Conocimiento SKC**. Proceedings of the IX Congreso Internacional Interacción Persona-Ordenador (INTERACCION 2008), Abacete, Spain, 2008, pp. 147-150.

[Morgan 1963] J. A. Morgan, J. N. Sonquist. **Problems in the Analysis of Survey Data, and a Proposal**. Journal of the American Statistical Association, 1963, vol. 58, pp. 415-434.

[Motta 2000] E. Motta, S. Buckingham, J. Domingue. **Ontology-driven document enrichment: principles, tools and applications**. International Journal of Human-Computer Studies, vol. 52, iss. 6, june 2000, pp. 1071-1109.

[Movielens 2008] **Movielens: Helping you find the right movies**. <http://www.movielens.org>. (Consultado abril 2008)

[Murugesan 1999] S. Murugesan , et al. **Web Engineering: A New Discipline for Web\_Based System Development**. Proceedings of the First International Conference of Software Engineering (ICSE) Workshop on Web Engineering, ACM Press,1999, pp. 1-9.

[MySQL 2008] **MySQL**. <http://www.mysql.com/>. (Consultado abril 2008)

[Nasraoui 2004] O. Nasraoui. **Approaches to Mining the Web**. <http://webmining.spd.louisville.edu/Websites/tutorials/Chapter2-approaches-mining-web.pdf>. (Consultado febrero 2008)

[Needleman 2003] M. Needleman. **The W3C Semantic Web Activity**. Serials Review, vol. 29, iss. 1, spring 2003, pp. 63-64.

[NLTK 2008] **NLTK**. [http://nltk.sourceforge.net/index.php/Main\\_Page](http://nltk.sourceforge.net/index.php/Main_Page). (Consultado enero 2008)

[Nonaka 1995] I. Nonaka, H. Takeuchi. **The Knowledge Creating Company**. Oxford University Press. 1995.

[Noy 2002] N. F. Noy, M. A. Musen. **Evaluating Ontology-Mapping Tools: Requirements and Experience**. Proceedings of the EKAW02 Workshop (WS1), Sep 2002.

[OntologyEd 2008] **Ontology editor**. [http://en.wikipedia.org/wiki/Ontology\\_editor](http://en.wikipedia.org/wiki/Ontology_editor). (Consultado enero 2008)

[OpenGroupware 2008] **OpenGroupware.org**. <http://www.opengroupware.org/>. (Consultado abril 2008)

[OpenNLP 2008] **OpenNLP**. <http://opennlp.sourceforge.net/>. (Consultado enero 2008)

[OpenRDF 2008] **OpenRDF**. <http://www.openrdf.org/>. (Consultado enero 2008)

[Oracle 2008] **Semantic Technologies Center**. [http://www.oracle.com/technology/tech/semantic\\_technologies/index.html](http://www.oracle.com/technology/tech/semantic_technologies/index.html). (Consultado enero 2008)

[O'Reilly 2005] T. O'Reilly. **What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software**. O'Reilly (online publishing), 2005. <http://www.oreillynnet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>. (Consultado agosto 2008)

[OSI 2008] **Open Source Initiative**. <http://www.opensource.org/>. (Consultado abril 2008)

[OWL 2008] **OWL Web Ontology Language Overview**. <http://www.w3.org/TR/owl-features/>. (Consultado abril 2008)

[Ozcan 2004] R. Ozcan, Y. A. Aslangogun. **Concept Based Information Access Using Ontologies and Latent Semantic Analysis**. In Technical Report CSE-2004-8 University of Texas at Arlington, 2004.

[Palen 2002] L. Palen, J. Grudin. **Discretionary adoption of group support software: Lessons from calendar applications**. B.E. Munkvold (Ed.). Organizational implementation of collaboration technology. Springer, 2002.

[Palmer 2001] S. B. Palmer. **The Semantic Web: An Introduction**. <http://infomesh.net/2001/swintro/>. 2001. (Consultado enero 2008)



- [Patel 2002a] P.F. Patel-Schneider, D. Fensel. **Layering the Semantic Web: Problems and Directions**. Proceedings of the ISWC2002, 2002.
- [Patel 2002b] P.F. Patel-Schneider, J. Simeon. **Building de Semantic Web on XML**. In Proceedings of the ISWC2002, 2002.
- [Pérez 2002] C. Pérez. **Estudios de Lingüística Española (ELiEs)**. Vol. 18, 2002.
- [Perkowitz 2000] M. Perkowitz, O. Etzioni. **Towards Adaptive Web Sites: Conceptual Framework and Case Study**. Artificial Intelligence, vol. 118, april 2000, pp. 24-275.
- [Piatetsky-Shapiro, 1991] G. Piatetsky-Shapiro. **Knowledge Discovery in Real Databases**. A Report on the IJCAI-89 Workshop. AI Magazine, vol. 11, no. 5, 1991, pp. 68-70.
- [Piatetsky-Shapiro, 2000] G. Piatetsky-Shapiro. **Knowledge Discovery in Databases: 10 years after**. SIGKDD Explorations. ACM SIGKDD, 2000, vol. 1, no. 2, pp. 59-61.
- [PiggyBank 2008] **Piggy Bank**. [http://simile.mit.edu/wiki/Piggy\\_Bank](http://simile.mit.edu/wiki/Piggy_Bank). (Consultado enero 2008)
- [Porter 2006] **The Porter Stemming Algorithm**. <http://tartarus.org/martin/PorterStemmer/>. (Consultado febrero 2008)
- [Preece 1994] J. Preece. **Human-computer interaction**. Addison-Wesley, Reading, MA, 1994
- [Prendinger 2007] H. Prendinger, C. Ma, M. Ishizukab. **Eye movements as indices for the utility of life-like interface agents: a pilot study**. Interacting with Computers vol. 19, iss. 2, 2007, pp. 281-292.
- [Protege 2008] **Protégé**. <http://protege.stanford.edu/>. (Consultado enero 2008)
- [RDF 2008] **Resource Description Framework (RDF)**. <http://www.w3.org/RDF/>. (Consultado abril 2008)
- [RSS 2008] **RSS Advisory Board**. <http://www.rssboard.org/>. (Consultado abril 2008)
- [Salton 1989] G. Salton. **Automatic Text Processing : The Transformation, Analysis, and Retrieval of Information by Computer**. Addison-Wesley, Reading, Massachusetts, 1989.
- [Shadbolt 2006] N. Shadbolt, T. Berners-Lee, W. Hall. **The Semantic Web Revisited**. IEEE Intelligent Systems, vol. 21, no. 3, 2006, pp. 96-101.
- [SIGCHI 2008] **SIGCHI: Interactions Bulletin TOCHI CHI Letters**. <http://sigchi.org/>. (Consultado abril 2008)
- [Sohlenkamp 1994] M. Sohlenkamp, G. Chwelos. **Integrating Communication, Cooperation, and Awareness: The DIVA Virtual Office Environment**. Proceedings of the 1994 ACM

Conference on Computer Supported Cooperative Work (CSCW 1994). Chapel Hill, NC, USA, 1994, pp. 331-343.

[SPARQL 2008] **SPARQL Query Language for RDF**. <http://www.w3.org/TR/rdf-sparql-query/>. (Consultado enero 2008)

[StanfordNLP 2008] **Stanford NLP Software**. <http://nlp.stanford.edu/software/index.shtml>. (Consultado enero 2008)

[Steve 1998] G. Steve, A. Gangemi, D. Pisanelli. **Integrating Medical Terminologies with ONIONS Methodology**. H. Kangassalo (ed.). Information Model and Knowledge Bases VIII. Amsterdam, IOS-Press, 1998.

[Stojanovic 2002] L. Stojanovic, A. Maedche, B. Motik, and N. Stojanovic. **User-driven Ontology Evolution Management**. In Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management 2002 (EKAW02), Madrid, Spain, LNCS/LNAI 2473, 2002, pp. 285-300.

[Srivastava 2000] J. Srivastava, R. Cooley, M. Deshpande, P. N. Tan. **Web usage mining: Discovery and applications of usage pattern from Web data**. SIGKDD Explorations, vol. 1, no. 2, 2000, pp.1-12.

[Tagzalia 2008] **Tagzalia**. <http://www.tagzania.com/>. (Consultado febrero 2008)

[TuCSoN 2008] **TuCSoN**. <http://tucson.sourceforge.net>. (Consultado enero 2008)

[UIMA 2008] **UIMA**. [www.research.ibm.com/UIMA](http://www.research.ibm.com/UIMA). (Consultado enero 2008)

[Uszkoreit 2000] H. USZKOREIT. **What is Computational Linguistics?** [http://www.coli.uni-saarland.de/~hansu/what\\_is\\_cl.html](http://www.coli.uni-saarland.de/~hansu/what_is_cl.html). 2000. (Consultado enero 2008)

[Watters 1999] C. Watters, M Shepherd. **Research Issues for Virtual Documents**. In Proceedings of the Workshop on Virtual Documents, Hypertext Functionality and the Web at the Eighth International World Wide Web Conference, Toronto, Canada, 1999.

[W3C 2008] **W3C: World Wide Web Consortium**. <http://www.w3.org/>. (Consultado abril 2008)

[WEKA 2008] **WEKA**. <http://www.cs.waikato.ac.nz/~ml/weka/>. (Consultado enero 2008)

[Weblog 2008] **Blog**. <http://en.wikipedia.org/wiki/Weblog>. (Consultado febrero 2008)

[WebServices 2008] **Web Services Activity**. <http://www.w3.org/2002/ws/>. (Consultado mayo 2008)

[Wiesman 2001] F. Wiesman, N. Roos, P. Vogt. **Automatic ontology mapping for agent communication**. Informe técnicot. 2001

- [Wikipedia 2008] **Wikipedia**. <http://es.wikipedia.org/wiki/Wikipedia>. (Consultado abril 2008)
- [Wooldridge 1995] M.J. Wooldridge and N.R. Jennings. **Intelligent agents: Theory and practice**. Knowledge Engineering Review, vol. 10, no. 2, 1995.
- [WordNet 2006] **WordNet**. <http://wordnet.princeton.edu/>. 2006. (Consultado Enero 2008)
- [Woychowsky 2007] E. Woychowsky. **AJAX: Creating Web Pages with Asynchronous JavaScript and XML**. Prentice Hall, 2007.
- [XHTML 2008] **W3C Recommendation: XHTML™ 1.0 The Extensible HyperText Markup Language (Second Edition)**. <http://www.w3.org/TR/xhtml1/>. (Consultado abril 2008)
- [XML 2008] **Extensible Markup Language (XML)**. <http://www.w3.org/XML/>. (Consultado abril 2008)
- [XSLT 2008] **XSL Transformations (XSLT)**. <http://www.w3.org/TR/xslt>. (Consultado abril 2008)
- [Yahoo 2008] **Yahoo!**. <http://www.yahoo.com>. (Consultado abril 2008)
- [Yang 1998] Y. Yang, T. Pierce, J. Carbonell. **A study on retrospective and online event detection**. In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia, 1998, pp. 28-36.
- [YouTube 2008] **YouTube. Broadcast Yourself**. <http://www.youtube.com> (Consultado agosto 2008)
- [Zakas 2007] N. C. Zakas, J. McPeak, J. Fawcett. **Profesional Ajax**. Anaya Multimedia, Madrid, Spain, 2007.

# **A**péndice 1: *Especificación Formal del Sistema SKC*

## Índice

1	Introducción .....	3
2	Glosario .....	4
3	Actores .....	5
4	Casos de Uso .....	7
4.1	ECU: Registrar Usuario .....	9
4.2	ECU: Identificarse .....	12
4.3	ECU: Hacerse Anónimo .....	14
4.4	ECU: Gestionar Configuración Usuario .....	16
4.5	ECU: Obtener Ayuda Usuario .....	19
4.6	ECU: Seleccionar Contenido .....	21
4.7	ECU: Visualizar Contenido .....	26
4.8	ECU: Gestionar Contenido UI.....	31
4.9	ECU: Obtener Información Sistema .....	35
4.10	ECU: Manifestar Consideración Explícita por Contenido Ajeno UI.....	38
4.11	ECU: Gestionar Configuración Nodo RN .....	42
4.12	ECU: Gestionar Operaciones de Administración No Delegadas.....	44
4.13	ECU: Suplantar Identidad UI.....	48
4.14	ECU: Gestionar Configuración Usuarios.....	53
4.15	ECU: Gestionar Contenido Usuarios .....	56
4.16	ECU: Registrar Actividad Remota Usuario.....	59
4.17	ECU: Mostrar Actividad Sistema .....	61
4.18	ECU: Asociar Ítems de Conocimiento.....	63
4.19	ECU: Establecer Consideración Implícita por Contenido Ajeno de UI.....	66
4.20	ECU: Cristalizar Conocimiento .....	68

## **1 Introducción**

Este documento recoge los requerimientos principales del sistema SKC, a través de la especificación esquemática de los casos de uso esenciales. Su objetivo es dar una idea inicial de cómo se comporta el sistema desde el punto de vista de los distintos actores que interactúan con él. El documento sirve como base para el resto de especificaciones detalladas necesarias para su desarrollo completo.

Algunos de los actores mencionados son personal, mientras que otros sistemas o subsistema que interactúan con SKC mediante una interfaz.

## 2 Glosario

En este apartado se definen los términos que se utilizan para la descripción de los casos de uso a lo largo del documento.

- **Grafo de Conocimiento (Vista):** Organización del conocimiento que se maneja en el sistema en forma de red semántica de temas del área de conocimiento representada. Cada grafo de conocimiento determina una forma de ver el conocimiento. Los grafos están constituidos por vértices unidos por relaciones.
- **Vértice del Grafo de Conocimiento (Vértice):** Los nodos que forman el grafo de conocimiento, cada uno corresponde a un tema del área de conocimiento representada por el grafo al que pertenecen.
- **Relación de Grafo de Conocimiento (Relación):** Vínculo entre vértices (temas) dentro de un grafo de conocimiento.
- **Documento:** Unidad mínima de conocimiento considerada en el sistema. Los documentos están asociados a vértices de los grafos de conocimiento.
- **Consideración por un Elemento de Conocimiento:** Relación que vincula un usuario del sistema con un elemento de conocimiento.
- **Anotación:** Información asociada a un documento para aportar información o criticar el contenido o algún aspecto del documento. En una manifestación implícita de consideración.
- **Voto:** Manifestación explícita y cuantificable de la opinión de un usuario hacia un elemento de contenido.
- **Elemento de Contenido (Contenido):** vista, grafo, vértice, relación, documento y anotación.
- **Documento Virtual (DV):** Documento enriquecido con diverso conocimiento disponible en el entorno, y que se emplea para mejorar la interactividad con el sistema.

### 3 Actores

A continuación se definen los actores que intervienen en los casos de uso especificados en esta documentación.

- **Usuario Anónimo (UA):** Usuario que utiliza el sistema sin estar identificado, sólo puede ver la parte del conocimiento de acceso libre.
- **Usuario Identificado (UI):** Usuarios que utiliza el sistema estando identificado, puede acceder al conocimiento establecido para realizar las operaciones que se le permitan de las disponibles (ver, añadir, quitar, modificar...)
- **Responsable de Nodo (RN):** Usuario identificado del sistema que asume la responsabilidad de las decisiones sobre un nodo.
- **Administrador de Nodo (AN):** Usuario identificado del sistema encargado de la gestión de uno o más nodos. Se trata de una persona con conocimientos suficientes sobre el sistema, que asesora a los RN de los nodos que administra, realiza las operaciones solicitadas por ellos considerando las implicaciones de las mismas, y se encarga de gestiones del sistema independientes de la competencia de los RN. En ocasiones puede asumir las funciones de RN por ausencia o delegación.
- **Motor de Análisis (MA):** Módulo del sistema encargado de obtener nueva información de utilidad a partir de la información que se genera por la actividad del propio sistema. Fundamentalmente se encarga de explicitar la consideración implícita de los usuarios por los elementos de conocimiento.
- **Motor de Cristalización (MC):** Módulo del sistema encargado de analizar el conocimiento gestionado con el sistema para seleccionarlo en función de la consideración manifestada hacia él por los usuarios.
- **Monitor Remoto (MR):** Módulo del sistema que obtiene datos de la actividad de los usuarios y muestra, a estos, información sobre la actividad del sistema.

Entre algunos de estos actores se establecen relaciones:

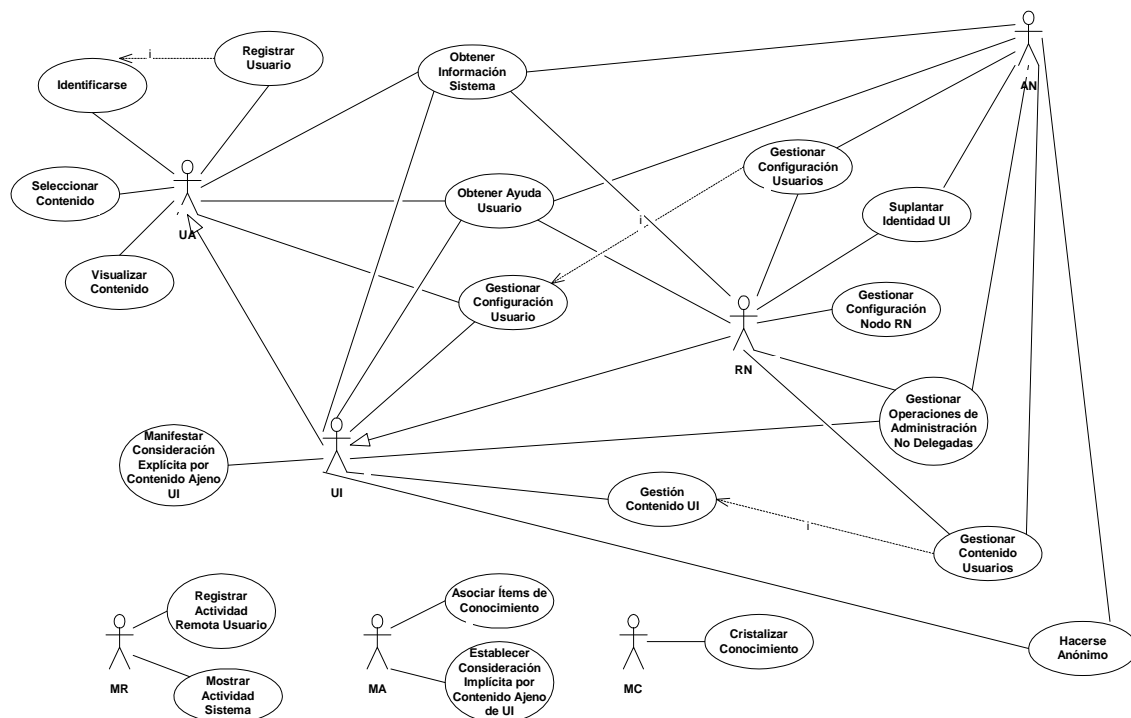
1. Cuando un Usuario Anónimo (UA) se identifica, se convierte en un Usuario Identificado (UI) o en Administrador de Nodo (AN).
2. Todos los Responsables de Nodo (RN) son también Usuarios Identificados (UI) del sistema, a además pueden realizar operaciones delegadas de administración de su nodo y suplantar a otros UI.



3. Los Administradores de Nodo (AN) están identificados, aunque no son UI. Pueden realizar operaciones reservadas para ellos sobre los nodos que gestionan. Entre ellas, pueden asumir la identidad de cualquier Usuario Identificado (UI) o Responsable de Nodo (RN), pero no comportarse realizando operaciones reservadas para estos sin ponerse en lugar de alguno de ellos.
4. Usuarios Identificados (UI) y los Administradores de Nodo (AN) vuelven a ser Usuarios Anónimos (UA) cuando así lo deciden.

## 4 Casos de Uso

El esquema general de los casos de uso es el siguiente:



• Figura1: Gráfico general de casos de uso del sistema.

En este gráfico se incluye la representación de los casos de uso que se especifican después en los apartados de esta sección. Esta es la enumeración de los casos de uso considerados indicando los actores que los usan:

1. Registrar Usuario
2. Identificarse
3. Hacerse Anónimo
4. Gestionar Configuración Usuario
5. Obtener Ayuda Usuario
6. Seleccionar Contenido
7. Visualizar Contenido
8. Gestionar Configuración UI
9. Obtener Información Sistema

10. Manifiestar Consideración Explícita por Contenido Ajeno UI
11. Gestionar Configuración Nodo RN
12. Gestionar Operación Administración No Delegadas
13. Suplantar Identidad UI
14. Gestionar Configuración Usuarios
15. Gestionar Contenido Usuarios
16. Registrar Actividad Remota Usuario
17. Mostrar Actividad Sistema
18. Asociar Ítems Conocimiento
19. Establecer Consideración Implícita por Contenido Ajeno de UI
20. Cristalizar Conocimiento

A continuación, en el siguiente apartado, se incluyen la Especificación de los Casos de Uso (ECU) aquí enumerados.

## **4.1 ECU: Registrar Usuario**

### **4.1.1 Descripción**

Los UA (y los actores que los generalizan) pueden darse de alta en el sistema como UI, si la configuración del nodo lo permite. Los RN y AN siempre pueden registrar usuarios en el sistema en su ámbito de gestión. Este caso de uso describe el registro de usuario que unos y otros realizan, los UA directamente y los RN y AN a través del caso “Gestionar Operaciones de Administración No Delegadas” que se presenta más adelante.

### **4.1.2 Flujo de Eventos**

#### **4.1.2.1 Flujo Básico**

1. Se inicia cuando el usuario manifiesta su intención de dar de alta un usuario en el sistema. El solicitante proporciona los datos necesarios para ello: identificador, contraseña, dirección de correo, etc.
2. El usuario intenta enviar los datos.
3. El sistema comprueba los datos localmente y solicita correcciones si lo considera oportuno.
4. El sistema intenta registrar el usuario con los datos proporcionados.
5. El sistema presenta un mensaje indicando el resultado de la operación.

#### **4.1.2.2 Flujos Alternativos**

##### ***4.1.2.2.1 Datos con Formato Inválidos***

1. Cuando el solicitante intenta enviar los datos, el sistema comprueba el formato de los mismos.
2. Si existen datos con formato inadecuado, se presenta el formulario de nuevo al solicitante, pero indicando los datos erróneos, para que este los corrija.
3. El solicitante corrige los datos introducidos y vuelve a intentar enviarlos.

#### **4.1.2.2.2 Registro Frustrado**

1. Cuando los datos introducidos por el solicitante tienen un formato adecuado y se ya han enviado al servidor, pueden encontrarse problemas al procesarlos allí que impidan la consumación del registro.
2. En tal caso, el sistema informa al solicitante de la circunstancia y da por finalizado el trámite, dejando al solicitante en la situación inmediatamente anterior al inicio del proceso.

#### **4.1.2.2.3 Cancelación de Proceso de Registro**

1. En cualquiera momento de la secuencia de registro el solicitante puede cancelar el proceso.
2. En tal caso, el sistema solicitará confirmación de la pretensión.
3. Si el solicitante no la confirma, se volverá al proceso de registro, en la posición previa a la solicitud de cancelación.
4. Si el solicitante confirma la cancelación, el sistema le informará de la circunstancia y se finalizará el proceso. El solicitante quedará en la situación inmediatamente anterior al inicio del proceso.

### **4.1.3 Requerimientos Especiales**

Algunos de los datos solicitados son obligatorios, como el identificador de usuario y la contraseña. El identificador de usuario debe ser único en el nodo y la contraseña de 6 caracteres, contener letras y números, y diferente al identificador. La dirección de correo electrónico deberá contener un signo "@" y el resto de campos cumplir los requerimientos propios de su clase.

### **4.1.4 Condiciones Previas**

#### **4.1.4.1 Disponibilidad de Autoregistro**

Dependiendo de la configuración del sistema puede estar habilitada o no la posibilidad de que los UA se puedan registrar directamente. El AN siempre puede registrar otros usuarios.

## **4.1.5 Condiciones Posteriores**

### **4.1.5.1 Disposición para Identificación**

El UA podría identificarse como el UI registrado inmediatamente después del haber sido registrado.

### **4.1.6 Notas**

Debería considerarse la posibilidad de que al terminar el proceso de auto-registro el UA quedara identificado como UI, si no se hace así debe informarse al usuario de ello en el mensaje final.

## **4.2 ECU: Identificarse**

### **4.2.1 Descripción**

Los UA pueden identificarse y convertirse en UI (como también son los RN) o en AN. Con ellos los UI pueden cambiar de identidad (que no es lo mismo que suplantar).

### **4.2.2 Flujo de Eventos**

#### **4.2.2.1 Flujo Básico**

1. Se inicia cuando el UA decide Identificarse.
2. El sistema solicita un usuario y una contraseña.
3. El UA debe suministrar estos datos y enviarlos.
4. El sistema intenta autenticar al usuario.
5. El sistema contabiliza las veces que se ha intentado la autenticación en la sesión actual.
6. Si tiene éxito, asigna el papel adecuado al usuario como UI o AN, sino se sigue por los flujos alternativos según el caso.
7. El sistema muestra el resultado de la operación al usuario y se pierde la cuenta de intentos asociada a la sesión.

#### **4.2.2.2 Flujos Alternativos**

##### ***4.2.2.2.1 Usuario y Contraseña Inválidos***

1. Si quedan intentos se informa al usuario del problema al tiempo que vuelve a pedir al usuario que se identifique, sino vuelve al flujo básico para informar al usuario del resultado.

##### ***4.2.2.2.2 Cuenta de Usuario Bloqueada***

1. Si la cuenta de usuario está bloqueada, porque un usuario autorizado está suplantando al usuario, entonces se informa al usuario sobre el problema y el tiempo que el usuario deberá esperar para volver a intentar usar la cuenta, una vez desbloqueada.

#### **4.2.2.2.3 Cancelación de Identificación**

1. El UA puede cancelar la operación de identificación en cualquier momento. En tal caso se salta al último paso del flujo básico para informa al usuario del resultado de la operación y se termina.

### **4.2.3 Requerimientos Especiales**

El número de intentos de identificación se indican en la fase de despliegue y configuración del nodo.

### **4.2.4 Condiciones Previas**

#### **4.2.4.1 Registro de Usuario**

Para que los usuarios puedan identificarse deben estar registrados en el sistema previamente. Se trata de una operación que, como ya se ha explicado, dependiendo de la configuración del nodo, podrá realizarla el propio UA o deberá encargarse de ella un RN o un AN.

### **4.2.5 Condiciones Posteriores**

#### **4.2.5.1 Identificación Positiva**

El UA pasa a ser un UI.

#### **4.2.5.2 Identificación Negativa**

El UA sigue como tal y vuelve al estado en que estaba antes de iniciar el intento de identificación.

### **4.2.6 Notas**

Debería considerarse la posibilidad de bloquear la cuenta sobre la que se han repetido intentos de identificación que han resultado negativos. Esta requeriría nuevos casos de uso para cubrir funcionalidades como la de desbloquear cuentas o solicitar el desbloqueo de las mismas.



## **4.3 ECU: Hacerse Anónimo**

### **4.3.1 Descripción**

Los usuarios reconocidos (UI, RN y AN) pueden dejar de estarlo en cualquier momento para volviendo a ser UA.

### **4.3.2 Flujo de Eventos**

#### **4.3.2.1 Flujo Básico**

1. Se inicia cuando el usuario identificado decide volver a ser anónimo.
2. El sistema solicita confirmación de la operación.
3. Si el usuario la confirma la sesión termina y el entorno aparece como si se hubiera acabado de iniciar para una UA.
4. Si no la confirma, el sistema se queda en el estado inmediatamente anterior a la solicitud de cierre de sesión, por lo que el usuario continúa con la misma identidad.

#### **4.3.2.2 Flujos Alternativos**

Ningunos.

### **4.3.3 Requerimientos Especiales**

Ningunos.

### **4.3.4 Condiciones Previas**

#### **4.3.4.1 Usuario Identificado**

Para que los usuarios puedan cerrar una sesión de usuario identificado, se han tenido que identificarse previamente como tales.

### **4.3.5 Condiciones Posteriores**

#### **4.3.5.1 Usuario Anónimo**

El usuario se convierte en UA.

### **4.3.6 Notas**

Ningunas.

## **4.4 ECU: Gestionar Configuración Usuario**

### **4.4.1 Descripción**

Parte de los elementos configurables del sistema puede gestionarlo el usuario, dependiendo del tipo de usuario.

El UA es el usuario más restringido del sistema y sólo puede gestionar: apariencia gráfica (colores, tipografías, iconos...), entorno interactivo (idioma, herramientas de navegación, ubicación de las mismas...) o visualización contenidos (vista preferida, composición de documentos virtuales...).

El UI tiene menos restricciones que el UA, además de las configuraciones disponibles para este, puede gestionar: sus datos personales y la persistencia de sus opciones de configuración.

### **4.4.2 Flujo de Eventos**

#### **4.4.2.1 Flujo Básico**

1. El usuario indica que desea gestionar su configuración.
2. El sistema presenta los aspectos configurables del sistema para él.
3. El usuario selecciona el aspecto que quiere configurar.
4. El sistema presenta una interfaz de configuración adecuada para cada caso.
5. El usuario indica sus preferencias e intenta enviar los datos.
6. El sistema comprueba los datos localmente y solicita correcciones si lo considera oportuno.
7. El sistema intenta realizar la configuración indicada.
8. El sistema presenta un mensaje mostrando el resultado de la operación.

## **4.4.2.2 Flujos Alternativos**

### ***4.4.2.2.1 Configuración Usuario Frustrada***

1. Cuando los datos introducidos por el usuario tienen un formato adecuado y se han enviado al servidor, pueden encontrarse problemas al procesarlos que impidan la consumación del proceso.
2. En tal caso, el sistema informa al usuario de la circunstancia y da por finalizado el trámite, dejando al usuario en la situación inmediatamente anterior al inicio de la operación.

### ***4.4.2.2.2 Cancelación de Configuración Usuario***

1. En cualquiera momento de la secuencia de configuración el usuario puede cancelar el proceso.
2. En tal caso, el sistema solicitará confirmación de la pretensión.
3. Si no se confirma se volverá al proceso de configuración, en la posición previa a la solicitud de cancelación.
4. Si se confirma la cancelación, el sistema informará al usuario de la circunstancia y se finalizará el proceso. El usuario quedará en la situación inmediatamente anterior al inicio del proceso.

### ***4.4.2.2.3 Aplicar Configuración Usuario por Defecto***

1. En cualquiera momento el usuario puede indicar que quiere aplicar la configuración establecida por defecto en el sistema.
2. En tal caso, el sistema solicitará confirmación de la pretensión.
3. Si no se confirma se volverá al proceso de configuración, en la posición previa a la solicitud de cancelación.
4. Si se confirma la operación, el sistema informará al usuario de la circunstancia, se aplicará la configuración y se finalizará el proceso. El usuario podrá seguir trabajando con la nueva configuración.

### ***4.4.2.2.4 Memorizar Configuración UI***

1. El sistema va memorizando la última configuración del UI y la utiliza por defecto durante la sesión de trabajo.

2. Al final de la misma el sistema pregunta al usuario si desea guardarla.
  - a. Si la respuesta es positiva, la configuración se utilizará al iniciar una nueva sesión del usuario.
  - b. Si la respuesta es negativa, en la siguiente sesión se utilizará la última configuración guardada, si existe, o la establecida por defecto en el sistema.

### **4.4.3 Requerimientos Especiales**

Ningunos.

### **4.4.4 Condiciones Previas**

#### **4.4.4.1 Configuración Usuario por Defecto**

El usuario siempre dispondrá de una configuración por defecto que el sistema aplicará cuando el usuario no está identificado o no ha manifestado sus preferencias antes.

### **4.4.5 Condiciones Posteriores**

#### **4.4.5.1 Caducidad Configuración Usuario**

La configuración de un usuario dura sólo hasta el fin de la sesión de servidor (durante una conexión continuada con el servidor) correspondiente a la conexión en la que se han especificado las preferencias.

#### **4.4.5.2 Caducidad Configuración UI**

La configuración del UI dura sólo durante la sesión, a no ser que el usuario decida guardarla. Si la configuración no se ha salvado en una sesión, en la siguiente se intenta utilizar la última registrada, y si no existe ninguna, la establecida por defecto durante el despliegue del sistema.

### **4.4.6 Notas**

Se podría considerar la persistencia de configuraciones de usuario asociada a navegador mediante el uso de cookies, permitiendo al usuario que pueda guardar en el navegador de su equipo la información necesaria para restaurar la configuración establecida cuando se vuelva a conectar al sistema, con ese mismo navegador desde esa misma máquina.

## **4.5 ECU: Obtener Ayuda Usuario**

### **4.5.1 Descripción**

Los usuarios pueden obtener ayuda sobre la funcionalidad del sistema que les incumbe según el papel que desempeñan en él (UA, UI, RN o AN).

### **4.5.2 Flujo de Eventos**

#### **4.5.2.1 Flujo Básico**

1. El usuario solicita ayuda mediante el elemento del entorno correspondiente.
2. El sistema abre una ventana o utiliza la ya abierta para tal fin, y presenta el la información correspondiente a contexto desde el que se le ha invocado.
3. El usuario puede cerrar la ventana de ayuda o dejarla abierta, y seguir trabajando con el sistema.

#### **4.5.2.2 Flujos Alternativos**

##### ***4.5.2.2.1 Ayuda sobre Tema Distinto al del Contexto Actual***

1. Si el usuario quiere información sobre un tema distinto al establecido por el contexto, tiene que seleccionar la opción correspondiente en la ventana de ayuda.
2. El sistema presentará los temas disponibles.
3. El usuario seleccionará el tema de su interés.
4. El sistema presentará el tema seleccionado en sustitución del anterior.

### **4.5.3 Requerimientos Especiales**

1. La ayuda debe ser contextual, presentando en primer lugar la información correspondiente al contexto desde el que se la ha invocado, aunque permita acceder a todos los contenidos que abarque, en el ámbito del usuario, mediante un índice.
2. La ayuda debería estar siempre disponible a través de la navegación general del entorno.

3. Además debería haber enlaces directos a temas de la ayuda desde los sitios del contenido o del entorno que se consideren oportunos. En tales casos, los enlaces llevarán artículo o sección adecuados de de la ayuda.
4. Los usuarios sólo deben tener acceso a la parte de la ayuda establecida para ellos.
5. La ayuda debe presentarse en una ventana separada de la aplicación, que siempre sea la misma y que pase a primer plano cada vez que se solicite ayuda.
6. La ventana de ayuda debe tener un elemento activo para que el usuario pueda cerrarla (además de los elementos que pueda tener la propia ventana ya incorporados).
7. Si un usuario cambia de papel en el sistema, la ayuda disponible debe acomodarse a la nueva situación.

#### **4.5.4 Condiciones Previas**

Ningunas.

#### **4.5.5 Condiciones Posteriores**

##### **4.5.5.1 Cierre Ventana Ayuda**

La ventana de ayuda puede permanecer abierta, si el usuario lo prefiere, mientras se interactúa con el nodo, pero debe cerrarse al terminar de trabajar con él.

#### **4.5.6 Notas**

Debería considerarse la posibilidad de incorporar algún mecanismo de búsqueda en la ayuda para facilitar la identificación de los temas que explican algún concepto.

También debería plantearse la posibilidad de incorporar algún mecanismo que facilite un acabado adecuado para impresión de los artículos incluidos en la ayuda.

## **4.6 ECU: Seleccionar Contenido**

### **4.6.1 Descripción**

El sistema permite a los usuarios acceder al conocimiento que gestiona, para ello proporciona un mecanismo para seleccionar los contenidos de su interés.

La interfaz de selección de contenido consta de varios módulos independientes que funcionan integrados para facilitar la navegación por los contenidos a varios niveles.

### **4.6.2 Flujo de Eventos**

#### **4.6.2.1 Flujo Básico**

1. El usuario accede al área del entorno para seleccionar contenido.
2. El sistema presenta las referencias de los contenidos disponibles en el contexto que el usuario se encuentra, utilizando la interfaz establecida para el correspondiente contexto en función de la situación, la disponibilidad, las preferencia manifestadas por el usuario y la configuración del entorno. Se pueden distinguir los siguientes contextos:
  - a. perspectiva general, para seleccionar vistas o grafos;
  - b. dentro de una vista, para seleccionar vértices y relaciones;
  - c. en el interior de un vértice, para seleccionar elementos vinculados, como documentos, usuarios, relaciones y otros vértices;
  - d. dentro de un documento, para seleccionar elementos vinculados, como versiones, anotaciones, usuarios y otros documentos y vértices.
3. El usuario selecciona el elemento de información deseado.
4. El sistema muestra el contenido correspondiente al elemento seleccionado en el área del entorno para destinado a tal fin.



## 4.6.2.2 Flujos Alternativos

### 4.6.2.2.1 *Cambiar Configurar Selección de Contenido*

En cualquier momento el usuario puede cambiar la configuración de la interfaz de selección de contenido. Para ello la interfaz proporciona un selector para elegir entre las distintas alternativas disponibles.

1. El usuario solicita cambiar la configuración de un determinado módulo de selección de contenido.
2. El sistema presenta las alternativas disponibles.
3. El usuario elige una.
4. El sistema sustituye la interfaz actual por la nueva.

### 4.6.2.2.2 *Cancelación Cambiar Configurar Selección de Contenido*

En cualquier momento el usuario que está cambiando la configuración de la interfaz de selección de contenido puede cancelar la operación

1. En medio del proceso de, el usuario decide cancelarlo y lo indica utilizando el elemento de la interfaz adecuada para ello.
2. El sistema pide confirmación.
3. Si el usuario confirma la cancelación, el sistema sigue con la misma configuración que antes de iniciar el proceso.
4. Si el usuario no confirma la cancelación, el sistema vuelve a la situación previa a la solicitud de cancelación, en el momento que se dejó el proceso de cambio de configuración.

### 4.6.2.2.3 *Volver Atrás*

En cualquier momento el usuario puede decidir volver atrás sobre sus pasos. Para ello el sistema habrá ido guardando la trayectoria seguida por el usuario.

1. El usuario solicita la trayectoria de su navegación.
2. El sistema presenta la trayectoria seguida.
3. El usuario selecciona punto de la trayectoria al que quiere volver.
4. El sistema se sitúa en el contexto de navegación del punto seleccionado, si es posible porque persisten las condiciones necesarias (se pretende ir a un documento

que no se ha borrado). Si no, el sistema indica el percance y permanece en la situación previa a la solicitud de vuelta atrás.

#### ***4.6.2.2.4 Cancelar Vuelta Atrás***

1. En cualquiera momento, el usuario puede cancelar el proceso de salto atrás seleccionando el elemento activo de la interfaz para este fin.
2. El sistema vuelve al contexto de navegación previo a la solicitud de vuelta atrás.
3. Si por alguna razón la acción no es posible, el sistema informa del problema y establece el contexto de navegación general.

#### ***4.6.2.2.5 Memorizar las Opciones de Selección de Contenidos en Uso***

1. El sistema va memorizando la última configuración de selección de contenidos y la utiliza por defecto durante la sesión de trabajo.
2. En cualquier momento el usuario puede solicitar que se guarde la configuración actual.
3. El sistema confirma la solicitud:
  - a. Si la respuesta es positiva, la configuración se guarda en la sesión actual, y si el sistema tiene medio para recordar dicha configuración entre sesiones del mismo usuario, se almacena también para utilizarse al iniciar una nueva sesión del usuario.
  - b. Si la respuesta es negativa, se cancelará la operación y todo quedará como antes de su inicio.

#### ***4.6.2.2.6 Memorizar las Opciones de Selección de Contenidos en Uso al Final***

1. Como se ha dicho antes, el sistema va memorizando la última configuración de selección de contenidos y la utiliza por defecto durante la sesión de trabajo.
2. Al final de la sesión, si el sistema tiene medio para recordar dicha configuración entre sesiones del mismo usuario, pregunta al usuario si desea guardarla.
  - a. Si la respuesta es positiva, la configuración se utilizará al iniciar una nueva sesión del usuario.
  - b. Si la respuesta es negativa, en la siguiente sesión se utilizará la última configuración guardada o la establecida por defecto en el sistema.

#### **4.6.2.2.7 *Restaura las Opciones de la Selección de Contenidos***

1. En cualquiera momento, el usuario puede restaurar la opción de selección de contenidos a su valor por defecto establecido en la configuración del sistema o a la última preferencia memorizada por el usuario.

#### **4.6.3 Requerimientos Especiales**

1. En cada caso sólo se muestran activas las categorías de elementos que existen.
2. Durante toda la navegación sobre el contenido se guarda un histórico de la trayectoria seguida sobre el mismo, que sirve para deshacer el camino en caso de necesidad.
3. El número máximo de pasos que se guardan en trayectoria histórica es un parámetro que se establece en la configuración del sistema durante su al despliegue.
4. La interfaz básica de selección de contenidos es independiente de la de visualización de los mismos, aunque esta última pueda incluir funcionalidades de navegación, como parte de los documentos virtuales utilizados para presentar el contenido.
5. La interfaz de selección de contenidos debe permitir la visión de los contenidos a varios niveles, como por ejemplo cambiar la vista utilizada permaneciendo en un documento o al pasar de un vértice a otro con el vinculado.
6. La interfaz de navegación sobre los contenidos tiene que ser modular, con alternativas para facilitar la selección los contenidos a distintos niveles (por ejemplo, disponibilidad de interfaces gráficas y textuales de los grafos, vértices o documentos).
7. Las opciones elegidas por los usuarios sobre la configuración de la interfaz de navegación deben conservarse durante la sesión de trabajo y poderse guardar como parte de las preferencias de usuario cuando el sistema disponga de un modo de recordarlas entre sesiones.

## **4.6.4 Condiciones Previas**

### **4.6.4.1 Preferencias Preestablecidas**

El sistema recuerda la última configuración utilizada para el selector de contenidos de cada nivel, y la utiliza por defecto. Si no, el sistema utiliza una configuración por defecto cuando se desplegó el sistema.

## **4.6.5 Condiciones Posteriores**

### **4.6.5.1 Opciones de Selección Activas**

Al final del procedimiento de selección de contenidos, siempre deben quedar unas opciones de selección activas, si el usuario no ha manifestado nada, se utilizará la configuración inicial establecida por defecto en el entorno, que se especifico al desplegar el sistema.

## **4.6.6 Notas**

Considerar el uso de cookies para la memorización de las opciones de selección de contenido, cuando no se dispone de otra memoria permanente entre secciones.

## **4.7 ECU: Visualizar Contenido**

### **4.7.1 Descripción**

Como ya se ha dicho en el caso anterior, el sistema permite a los usuarios acceder al conocimiento que gestiona, para ello es imprescindible un mecanismo para visualizar los contenidos correspondientes.

### **4.7.2 Flujo de Eventos**

#### **4.7.2.1 Flujo Básico**

1. Cuando el usuario selecciona la referencia de un contenido, el sistema lo presenta para su visualización en forma de un documento virtual, que integra el elemento solicitado junto con otra información relevante, en el área del entorno reservada para tal fin.
2. El usuario revisa el contenido, interactuando con él cuando el contenido es interactivo o navegable.
3. El usuario termina la visualización del contenido cuando:
  - a. Al interactuar con él, selecciona la referencia de otro contenido.
  - b. Cambia a otra actividad incluida en el sistema, que es incompatible con la visualización del documento.
  - c. Cierra explícitamente el contenido.

#### **4.7.2.2 Flujos Alternativos**

##### ***4.7.2.2.1 Cambiar Configurar Visualización Contenido***

En cualquier momento el usuario puede cambiar la configuración de la interfaz de visualización de contenido. Para ello la interfaz proporciona un selector para elegir entre las distintas alternativas disponibles.

1. El usuario solicita cambiar la configuración de un determinado módulo de selección de contenido.
2. El sistema presenta las alternativas disponibles.
3. El usuario elige una.

5. El sistema sustituye la interfaz actual por la nueva.

#### **4.7.2.2.2 *Cancelar Cambiar Configurar Visualización Contenido***

En cualquier momento el usuario que está cambiando la configuración de la interfaz de visualización de contenido puede cancelar la operación:

1. En medio del proceso de configuración, el usuario decide cancelarlo y lo indica utilizando el elemento de la interfaz adecuada para ello.
2. El sistema pide confirmación.
3. Si el usuario confirma la cancelación, el sistema sigue con la misma configuración que antes de iniciar el proceso.
4. Si el usuario no confirma la cancelación, el sistema vuelve a la situación previa a la solicitud de cancelación, en el momento que se dejó el proceso de cambio de configuración.

#### **4.7.2.2.3 *Memorizar Configuración de Visualización***

1. El sistema va memorizando la última configuración de visualización de contenidos y la utiliza por defecto durante la sesión de trabajo.
2. Al final de la misma, si el sistema tiene medio para recordar dicha configuración entre sesiones del mismo usuario, pregunta al usuario si desea guardarla.
  - a. Si la respuesta es positiva, la configuración se utilizará al iniciar una nueva sesión del usuario.
  - b. Si es negativa la respuesta, en la siguiente sesión se utilizará la última configuración guardada o la establecida por defecto en el sistema.

#### **4.7.2.2.4 *Marcar Contenido***

En cualquier momento el usuario puede decidir marcar el contenido que está visualizando para poder volver a él más adelante con facilidad. Para ello el sistema guarda una lista de todos los contenidos marcados.

1. El usuario solicita la operación mediante el elemento proporcionado por la interfaz.
2. El sistema confirma la solicitud:
  - a. Si la respuesta es positiva, la referencia del contenido se incluye en la lista y se indica como marcado en todas las interfaces que lo referencian.

- b. Si la respuesta es negativa, se vuelve a la situación inmediatamente anterior a la solicitud de marcado.

#### **4.7.2.2.5 Desmarcar Contenido**

Cuando un usuario está visualizando un contenido marcado, puede desmarcarlo si así lo quiere, para ello:

1. El usuario indica su intención utilizando el elemento de la interfaz adecuada para ello.
2. El sistema pide confirmación.
3. Si el usuario confirma la operación, el sistema quita el contenido de la lista correspondiente, que deja de aparecer señalado como marcado en todas las interfaces que lo referencian.
4. Si el usuario no confirma la operación, el sistema vuelve a la situación previa a la solicitud de la misma.

#### **4.7.2.2.6 Memorizar Contenidos Marcados**

1. El sistema va memorizando la lista de contenidos marcados durante la sesión de trabajo.
2. Al final de la misma, si el sistema tiene medio para recordar dicha lista entre sesiones del mismo usuario, pregunta al usuario si desea guardarla.
  - a. Si la respuesta es positiva, la lista se utilizará al iniciar una nueva sesión del usuario.
  - b. Si es negativa la respuesta, en la siguiente sesión se partirá de una lista vacía.

#### **4.7.2.2.7 Navegar A Través de un Documento Virtual**

En ocasiones los documentos virtuales presentan referencia a otros elementos de contenido que pueden seleccionarse para acceder a ellos. En tales casos los documentos virtuales se comportan como otro mecanismo de selección de contenidos.

1. El usuario selecciona una referencia en el documento virtual que esta visualizando.
2. El sistema presenta el documento referenciado:

- a. Si el usuario no ha indicado lo contrario el elemento referenciado se presenta sustituyendo al actual. En este caso, el acceso al nuevo elemento se reflejará en todas las interfaces de selección de contenido.
- a. Si el usuario lo indica al seleccionar la referencia, el elemento se presenta en una nueva ventana y las interfaces de selección no se hacen eco del acceso, pero sí lo hace el registro de trayectoria.

#### **4.7.2.2.8 Navegar por Contenidos Marcados**

En cualquier momento el usuario puede decidir ir a un contenido marcado.

1. El usuario puede elegir la referencia del elemento de conocimiento marcado al que quiere ir:
  - a. Seleccionando el elemento adecuado del entorno para acceder a la lista correspondiente. A lo que el sistema responderá mostrando dicha lista.
  - b. Seleccionando el elemento de conocimiento directamente, en alguna de las interfaces de selección en donde aparezca indicada su condición.
2. El sistema presenta el documento referenciado:
  - a. Si el usuario no ha indicado lo contrario el elemento referenciado se presenta sustituyendo al actual. En este caso, el acceso al nuevo elemento se reflejará en todas las interfaces de selección de contenido.
  - b. Si el usuario lo indica al seleccionar la referencia, el elemento se presenta en una nueva ventana y las interfaces de selección no se hacen eco del acceso, pero sí lo hace el registro de trayectoria.

#### **4.7.3 Requerimientos Especiales**

1. El número máximo de elementos de conocimiento que se guardan en lista de marcados es un parámetro que se establece al desplegar el sistema. Si en algún momento se pretende sobrepasar dicho límite, se avisará al usuario sobre la incidencia y se dará al usuario a elegir entre desmarcar algún elemento o cancelar la operación.
2. La interfaz básica de selección de contenidos es independiente de la de visualización de los mismos, aunque esta última pueda incluir funcionalidades de navegación, como parte de los documentos virtuales utilizados para presentar el contenido.



3. Las opciones elegidas por los usuarios sobre la configuración de la interfaz de visualización deben conservarse durante la sesión de trabajo y poderse guardar como parte de las preferencias de usuario cuando el sistema disponga de un modo de recordarlas entre sesiones.
4. El sistema debe mantener en la lista sólo los elementos que siguen siendo válidos (no han desaparecido del sistema). Durante una sesión interactiva, la comprobación de validez se realizará al intentar acceder al elemento, en caso de detectarse la invalidez se informará al usuario, se eliminara el elemento de la lista y se volverá al estado previo a la solicitud del elemento en cuestión.

#### **4.7.4 Condiciones Previas**

##### **4.7.4.1 Lista Elementos Marcados Previos**

El sistema recuerda la última lista guardada de elementos marcados. Si no, el sistema parte de una lista vacía. Es preciso comprobar la validez de los elementos marcados y limpiar la lista en caso necesario.

#### **4.7.5 Condiciones Posteriores**

##### **4.7.5.1 Opciones de Visualización Activas**

Al final del proceso de visualización de contenidos, siempre deben quedar unas opciones de visualización activas, si el usuario no ha manifestado nada, se utilizará la configuración inicial establecida por defecto en el entorno, que se especifico al desplegar el sistema.

#### **4.7.6 Notas**

Considerar el uso de cookies para la memorización de la lista de elementos marcados, cuando no se dispone de otro modo de memoria permanente entre secciones.

También debería considerarse la posibilidad de incorporar algún mecanismo que facilite un acabado adecuado para impresión de los contenidos visualizados.

## **4.8 ECU: Gestionar Contenido UI**

### **4.8.1 Descripción**

El entorno proporciona a los UI un mecanismo para gestionar los contenidos que aportan al sistema. Los RN, como UI que son, pueden gestionar sus propios contenidos también.

### **4.8.2 Flujo de Eventos**

#### **4.8.2.1 Flujo Básico**

1. El sistema incluye en su interfaz para UI un elemento activo para gestionar contenidos del usuario. Utilizando dicho elemento, el usuario indica que desea gestionar su contenido.
2. El sistema presenta los contenidos que el usuario puede gestionar, organizados según la configuración establecida, y distintas alternativas para organizarlos. El usuario puede elegir en este punto entre las alternativas de presentación, que en caso de ser seleccionadas se convertirán en la configuración establecida para futuras ocasiones:
  - a. Clasificar por categorías (vistas, temas, documentos, anotaciones...) y dentro de ellas ordenar por antigüedad de forma descendente.
  - b. Agrupar por tiempo en orden creciente o descendente y dentro de cada grupo por tipos.
  - c. Agrupar por vinculación (anotaciones de un mismo documento, documentos de un mismo tema de una vista, vistas del nodo...)
3. El usuario selecciona el contenido que quiere gestionar.
4. El sistema presenta el contenido y las gestiones disponibles para el tipo de documento y la coyuntura del momento, por ejemplo:
  - a. Anotación: modificar y borrar, modificar configuración...
  - b. Documento: anotar, versionar, borrar, vincular y desvincular otros documentos, modificar configuración documento...
  - c. Temas: vincular y desvincular documento, modificar configuración tema...

- d. Vistas: proponer añadir, modificar o quitar tema, modificar configuración vista...
  - e. Nodo: proponer añadir o quitar vista
5. El usuario selecciona la gestión deseada.
  6. El sistema presenta una interfaz de gestión adecuada para el caso.
  7. El usuario elige las opciones de su preferencia, proporciona los datos necesarios e intenta enviarlos al sistema (algunas gestiones pueden requerir de secuencias de interacción con el servidor).
  8. El sistema comprueba los datos localmente, solicita correcciones si lo considera oportuno y solicita confirmación si la gestión lo requiere.
  9. El sistema intenta realizar la gestión indicada.
  10. El sistema presenta un mensaje mostrando el resultado de la operación, se adecua la presentación del contenido a la gestión realizada (si el contenido ha dejado de existir se ubica al usuario en el contenido más cercano y general al eliminado), y se muestran las opciones de seguir gestionando contenidos o terminar.
  11. El usuario elige entre las opciones disponible:
    - a. Si se decide terminar se pasa a una situación de visualización del contenido que hubiera en pantalla, que será el que se estuviera gestionando si no se eliminó.
    - b. Si se decide seguir gestionando documentos, se pasa a la selección de contenidos para gestionar.

## **4.8.2.2 Flujos Alternativos**

### ***4.8.2.2.1 Gestión Sobre Contenido Actual***

1. Cuando el usuario está en un contenido propio, el sistema le ofrece la posibilidad de gestionarlo.
2. El usuario indica que desea gestionar el contenido actual.
3. El sistema presenta las gestiones disponibles para el tipo de documento y la coyuntura del momento.
4. El usuario selecciona la gestión deseada.

5. El resto de la secuencia de este flujo es como en el básico después de la selección de la gestión a realizar.

#### **4.8.2.2.2 *Gestión de Contenido Frustrada***

1. Cuando los datos introducidos por el usuario tienen un formato adecuado y se han enviado al servidor, pueden encontrarse problemas al procesarlos que impidan la consumación del proceso.
2. En tal caso, el sistema informa al usuario de la circunstancia y da por finalizado el trámite, dejando al usuario en la situación inmediatamente anterior al inicio de la operación.

#### **4.8.2.2.3 *Cancelación de Gestión en Curso***

1. En cualquier momento del proceso el usuario puede decidir cancelar la gestión en curso. Para ello el usuario debe indicarlo activando el elemento incluido la interfaz para tal fin.
2. El sistema pedirá confirmación de la operación:
3. Si se confirma la cancelación, el proceso seguirá por el punto en que estuviera.
4. Si no se confirma la cancelación, el sistema informará al usuario de la circunstancia y se finalizará el proceso. El usuario quedará en la situación inmediatamente anterior al inicio del proceso. En las gestiones que requieran de secuencias de interacción con el servidor, el sistema tiene que deshacer las operaciones intermedias que se hubieran realizado, para dejar el sistema en el estado previo.

### **4.8.3 Requerimientos Especiales**

1. Algunas gestiones pueden requerir, para su realización, de secuencias de interacción con el servidor. Si ello conlleva modificaciones de datos parciales en el servidor, el sistema tiene que ser capaz de:
  - a. Bloquear otras operaciones con el sistema que pudieran interferir con la operación en curso.
  - b. Restaurar el estado a la situación previa al inicio del proceso en caso de cancelación del mismo.
2. Cuando se seleccionan los contenidos a gestionar, se puede optar por varias modalidades de organización de los mismos. La elección realizada pasara a forma

parte de las preferencias de configuración del usuario, se recordarán durante la sesión en curso y se podrán guardar para mantenerlas en sesiones posteriores, junto con el resto de las preferencias de configuración.

3. Para que los RN y los AN puedan gestionar contenido de un UI deben asumir la identidad del UI correspondiente o utilizar los mecanismos de gestión de contenidos de grupos de usuarios.

## **4.8.4 Condiciones Previas**

### **4.8.4.1 Configuración Usuarios por Defecto**

El usuario siempre dispondrá de una configuración por defecto que el sistema aplicará cuando el usuario no ha manifestado sus preferencias antes. Esto incluye las preferencias de organización de los contenidos para su selección.

## **4.8.5 Condiciones Posteriores**

### **4.8.5.1 Caducidad Configuración UI**

Como ocurre con el resto de la configuración del UI, esta dura sólo durante la sesión en curso, a no ser que el usuario decida guardarla. Si la configuración no se ha salvado en una sesión, en la siguiente se intenta utilizar la última registrada, y si no existe ninguna, la establecida por defecto durante el despliegue del sistema.

## **4.8.6 Notas**

Ninguna en particular.

## **4.9 ECU: Obtener Información Sistema**

### **4.9.1 Descripción**

Los usuarios del entorno tienen acceso a información sobre el sistema, aunque dependiendo de su categoría (UA, UI, RN o AN) dicha información puede variar cualitativa y cuantitativamente.

### **4.9.2 Flujo de Eventos**

#### **4.9.2.1 Flujo Básico**

1. En cualquier momento el usuario puede acceder a la información del sistema disponible para su categoría mediante informes. Para ello el entorno dispone de un elemento activo para solicitar tal servicio.
2. El usuario solicita información accionando el elemento activo.
3. El sistema presenta los tipos de información disponibles.
4. El usuario selecciona el tipo de información deseado.
5. El sistema pide al usuario los parámetros necesarios para configurar el informe correspondiente a su gusto.
6. El usuario indica sus preferencias e intenta enviar los datos.
7. El sistema comprueba los datos localmente y solicita correcciones si lo considera oportuno.
8. El sistema intenta obtener el informe solicitado.
  - a. Si lo consigue, el sistema presenta el informe.
  - b. Si no lo consigue, el sistema presenta un mensaje explicando lo sucedido y un elemento activo para cancelar la operación.
9. El usuario dispone de la información solicitada para revisarla y proporciona:
  - a. Un elemento activo para imprimir el informe, que al ser activado prepara el documento en un formato adecuado y lo imprime.
  - b. Un elemento activo para salir del informe, que al ser activado interroga sobre si se desea otro informe o cerrar la ventana de información.

## **4.9.2.2 Flujos Alternativos**

### ***4.9.2.2.1 Formato Inválido en Datos Solicitud Información***

1. Cuando el usuario intenta enviar los datos, el sistema comprueba el formato de los mismos.
2. Si existen datos con formato inadecuado, se presenta el formulario de nuevo al usuario, pero indicando los datos erróneos, para que este los corrija.
3. El usuario corrige los datos introducidos y vuelve a intentar enviarlos.

### ***4.9.2.2.2 Solicitud Información Frustrada***

1. Cuando los datos introducidos por el usuario tienen un formato adecuado y se han enviado al servidor, pueden encontrarse problemas al procesarlos que impidan la obtención del informe solicitado.
2. En tal caso, el sistema comunica al usuario la circunstancia y da por finalizado la tarea, dejando al usuario en la situación inmediatamente anterior al inicio de la operación.

### ***4.9.2.2.3 Cancelación de Solicitud de Información***

1. En cualquiera momento del proceso el usuario puede cancelarlo.
2. En tal caso, el sistema solicitará confirmación de la intención.
3. Si no se confirma, se volverá al proceso en la posición previa a la solicitud de cancelación.
4. Si se confirma la cancelación, el sistema informará al usuario de la circunstancia y se finalizará el proceso. El usuario quedará en la situación inmediatamente anterior al inicio del proceso.

## **4.9.3 Requerimientos Especiales**

1. La información sobre el sistema debería estar siempre disponible a través de la navegación general del entorno.
2. La información debe presentarse en una ventana separada de la aplicación, que siempre sea la misma, que pueda mantenerse abierta con un informe mientras se sigue trabajando con el entorno, y que pase a primer plano cada vez que se solicite el servicio en ocasiones sucesivas.

3. La ventana de información debe tener un elemento activo para que el usuario pueda cerrarla (además de los elementos que pueda tener la propia ventana ya incorporados).
4. La ventana de información debe tener un elemento activo para que el usuario pueda imprimir el informe correspondiente (además de los elementos que pueda tener la propia ventana ya incorporados).
5. Si un usuario cambia de papel en el sistema, la información disponible debe acomodarse a la nueva situación.

#### **4.9.4 Condiciones Previas**

##### **4.9.4.1 Identificación Previa de Usuario**

Para poder obtener los informes disponibles para un determinado perfil de usuario que necesite identificación, el usuario deberá haberse identificado como tal antes de hacer la correspondiente solicitud.

#### **4.9.5 Condiciones Posteriores**

##### **4.9.5.1 Caducidad de Acceso a Informes**

El usuario mantiene el acceso a un informe solicitado con un determinado perfil mientras se mantiene como tal. Los informes obtenidos con un determinado perfil se pueden seguir utilizando a pesar de haber cambiado de estatus. Sin embargo, los informes nuevos que se soliciten siempre corresponderán al perfil actual del usuario.

#### **4.9.6 Notas**

Considerar la posibilidad de incluir opciones para generar informes en formatos para hojas de cálculo o bases de datos, como se hace para impresión.



## **4.10 ECU: Manifestar Consideración Explícita por Contenido Ajeno UI**

### **4.10.1 Descripción**

Durante toda su interacción con el sistema el usuario está manifestando implícitamente su consideración sobre el conocimiento ajeno que maneja. Además, los usuarios identificados (UI, RN y AN como algún UI) pueden expresar de forma explícita la consideración sobre determinados elementos de conocimiento. Al visualizar dichos elementos se proporciona al usuario la posibilidad de manifestar tal consideración.

### **4.10.2 Flujo de Eventos**

#### **4.10.2.1 Flujo Básico**

1. Al visualizar determinados elementos de contenido (documentos, anotaciones...) el sistema presenta al usuario la posibilidad de expresar su consideración por ellas mediante un elemento activo.
2. El usuario indica su deseo de expresar su consideración por el contenido que está visualizando accionando el elemento activo correspondiente.
3. El sistema muestra los tipos de consideración disponibles para el elemento de conocimiento en cuestión (por ejemplo, para anotación apoyo, y para documento, apoyo a adscripción a tema, anotación o voto) sobre los que el usuario puede manifestarse en unas circunstancias determinadas (una votación puede tener un determinado momento o el apoyo a una adscripción a tema cuando se ingresa un nuevo documento que el sistema intenta clasificar automáticamente en algún tema distinto del elegido por el autor).
4. El usuario selecciona el tipo de manifestación que quiere hacer.
5. El sistema solicita la información necesaria para ello, que depende del tipo de manifestación (anotación, apoyo, voto...).
6. El sistema comprueba los datos localmente y solicita correcciones si lo considera oportuno.
7. El sistema intenta realizar la operación solicitada:

- a. Si la operación se ha realizado con éxito, el sistema presenta un mensaje indicando el resultado.
- b. Si no lo consigue, el sistema presenta un mensaje explicando lo sucedido y un elemento activo para cancelar la operación.

#### **4.10.2.2 Flujos Alternativos**

##### ***4.10.2.2.1 Formato Inválido en Manifestación de Consideración***

1. Cuando el usuario intenta enviar los datos, el sistema comprueba el formato de los mismos.
2. Si existen datos con formato inadecuado, se presenta el formulario de nuevo al usuario, pero indicando los datos erróneos, para que este los corrija.
3. El usuario corrige los datos introducidos y vuelve a intentar enviarlos.

##### ***4.10.2.2.2 Manifestación de Consideración Frustrada***

1. Cuando los datos introducidos por el usuario tienen un formato adecuado y se han enviado al servidor, pueden encontrarse problemas al procesarlos que impidan la consumación de la manifestación.
2. En tal caso, el sistema comunica al usuario la circunstancia y da por finalizado la tarea, dejando al usuario en la situación inmediatamente anterior al inicio de la operación.

##### ***4.10.2.2.3 Cancelación de la Manifestación de Consideración***

1. En cualquiera momento del proceso el usuario puede cancelarlo.
2. En tal caso, el sistema solicitará confirmación de tal intención.
  - a. Si no se confirma, se volverá al proceso en la posición previa a la solicitud de cancelación.
  - b. Si se confirma la cancelación, el sistema informará al usuario de la circunstancia y se finalizará el proceso. El usuario quedará en la situación inmediatamente anterior al inicio del proceso

#### **4.10.2.2.4 Restauración de la Visualización del Elemento de Contenido Correspondiente a la Manifestación de Consideración**

1. En cualquiera momento del proceso el usuario puede abrir el elemento de contenido correspondiente a la manifestación en curso, para ello el sistema proporciona un elemento activo en la interfaz particular.
2. El usuario indica su deseo activando el mencionado elemento.
3. El sistema presenta el elemento de conocimiento sustituyendo al que se estuviera visualizando en el entorno.

### **4.10.3 Requerimientos Especiales**

1. Los usuarios identificados (UI, RN y AN, asumiendo la identidad de algún UI) son los que pueden manifestar su consideración por algunos elementos de conocimiento incluidos en el sistema bajo determinadas circunstancias.
2. Los tipos de consideración que los usuarios identificados pueden manifestar por los elementos de conocimientos dependen de:
  - a. El usuario del que se trate, algunas manifestaciones se reservan a un tipo de usuarios, a usuarios que realizan una determinada tarea en relación con ciertos elementos de contenido (expertos del tema en el que se incluye un documento) o mantienen una relación concreta con alguno de ellos (como ser el autor de un documento).
  - b. Los periodos establecidos para determinadas manifestaciones, como el intervalo de tiempo establecido para la realización de una anotación o votación.
  - c. Las circunstancias que concurran en un momento determinado, como el reciente ingreso de un documento que ha desencadenado su adscripción automática a algún tema distinto del elegido por el autor.
3. La manifestación de consideración debe hacerse mediante una ventana separada de la aplicación, que pueda mantenerse abierta simultáneamente con el elemento de contenido sobre el que se trate.
4. Si se deja de ver el elemento de conocimiento al que corresponde una ventana de manifestación de consideración abierta, el sistema debe avisar de ello.

5. La ventana de manifestación de consideración debe tener un elemento activo para abrir el elemento de conocimiento al que se refiere, en caso de que este se hubiera cerrado.
6. Sólo puede estar abierta un proceso de manifestación de consideración a la vez.
7. Si un usuario pretende cambiar de papel en el sistema, las manifestaciones de consideración en proceso deben cancelarse antes.

#### **4.10.4 Condiciones Previas**

##### **4.10.4.1 Identificación Previa de Usuario**

Sólo los usuarios identificados pueden manifestar su consideración elementos de conocimiento incluidos en el sistema. Son usuarios identificados los UI, que incluyen a los RN, y los AN cuando asumiendo la identidad de algún UI.

#### **4.10.5 Condiciones Posteriores**

##### **4.10.5.1 Caducidad de la Manifestación de Consideración en Curso**

Para que un usuario que tiene una manifestación de consideración en curso pueda terminar o cambiar de usuario tiene que terminar el proceso correspondiente previamente.

#### **4.10.6 Notas**

Considerar la posibilidad de tener abiertas varios procesos de manifestaciones de consideración simultáneamente.

Quizás lo más útil, en este sentido, sea poder ver manifestaciones de consideración ya completadas como referencias para otras nuevas, lo que podría hacerse mediante informes.

Puede que el acceso a los informes sobre consideraciones anteriores pueda estar incorporado en la interfaz del proceso de manifestación de nuevas consideraciones.

## **4.11 ECU: Gestionar Configuración Nodo RN**

### **4.11.1 Descripción**

Los usuarios RN de cada nodo, y los AN asumiendo su identidad, pueden gestionar la configuración de los nodos que tienen bajo su responsabilidad.

### **4.11.2 Flujo de Eventos**

#### **4.11.2.1 Flujo Básico**

1. El usuario indica que desea gestionar la configuración del nodo.
2. El sistema presenta los aspectos configurables del nodo.
3. El usuario selecciona el aspecto que quiere configurar.
4. El sistema presenta una interfaz de configuración adecuada para cada caso.
5. El usuario indica las preferencias e intenta enviar los datos.
6. El sistema comprueba los datos localmente y solicita correcciones si lo considera oportuno.
7. El sistema intenta realizar la configuración indicada.
8. El sistema presenta un mensaje mostrando el resultado de la operación.

#### **4.11.2.2 Flujos Alternativos**

##### ***4.11.2.2.1 Configuración de Nodo Frustrada***

1. Cuando los datos introducidos por el usuario tienen un formato adecuado y se han enviado al servidor, pueden encontrarse problemas al procesarlos que impidan la consumación del proceso.
2. En tal caso, el sistema informa al usuario de la circunstancia y da por finalizado el trámite, dejando al usuario en la situación inmediatamente anterior al inicio de la operación.

##### ***4.11.2.2.2 Cancelación de Configuración de Nodo***

1. En cualquiera momento de la secuencia de configuración el usuario puede cancelar el proceso.

2. En tal caso, el sistema solicitará confirmación de la pretensión.
3. Si no se confirma se volverá al proceso de configuración, en la posición previa a la solicitud de cancelación.
4. Si se confirma la cancelación, el sistema informará al usuario de la circunstancia y se finalizará el proceso. El usuario quedará en la situación inmediatamente anterior al inicio del proceso.

#### ***4.11.2.2.3 Aplicar Configuración de Nodo por Defecto***

1. En cualquiera momento el usuario puede indicar que quiere aplicar la configuración establecida por defecto en el sistema.
2. En tal caso, el sistema solicitará confirmación de la pretensión.
3. Si no se confirma se volverá al proceso de configuración, en la posición previa a la solicitud de cancelación.
4. Si se confirma la operación, el sistema informará al usuario de la circunstancia, se aplicará la configuración y se finalizará el proceso. El nodo dispondrá a partir de ese momento de la nueva configuración.

### **4.11.3 Requerimientos Especiales**

Ningunos en particular.

### **4.11.4 Condiciones Previas**

#### **4.11.4.1 Configuración de Nodo por Defecto**

Los nodos siempre disponen de una configuración por defecto, que el sistema aplicará cuando el usuario autorizado no la ha modificado.

### **4.11.5 Condiciones Posteriores**

#### **4.11.5.1 Caducidad Configuración de Nodo**

La configuración de un nodo se guarda en el servidor y se recuerda siempre entre sesiones.

### **4.11.6 Notas**

Ningunas.

## **4.12 ECU: Gestionar Operaciones de Administración No Delegadas**

### **4.12.1 Descripción**

Algunas categorías de usuarios no pueden realizar ciertas operaciones de administración por sí mismos. Esto puede deberse a que se requiere de la autorización o supervisión de otros usuarios, o a que el sistema no dispone del soporte imprescindible para que los interesados realicen directamente las operaciones o que estas se puedan hacer a través del entorno.

Para la gestión de todo ello el sistema proporciona un mecanismo que facilita el intercambio de solicitudes y notificaciones de operaciones entre las categorías de usuarios implicadas en su trámite: los usuarios de cada categoría, pueden enviar solicitudes de las operaciones que ellos no pueden realizar y recibir los resultados de las mismas; y los usuarios encargados de llevarlas a cabo, recibir dichas solicitudes y responder a las mismas con los resultados de sus tramitaciones.

Además, el mecanismo proporciona a los usuarios encargados de realizar las operaciones integradas en el sistema de forma efectiva, el modo de lanzarlas desde el entorno.

### **4.12.2 Flujo de Eventos**

#### **4.12.2.1 Flujo Básico**

1. En cualquier momento el usuario puede gestionar las operaciones de administración no delegadas. Para ello el entorno dispone de un elemento activo que le permite acceder a la interfaz correspondiente.
2. El usuario solicita la realización de gestiones administrativas no delegadas accionando el elemento activo mencionado.
3. El sistema presenta los tipos de gestiones disponibles.
4. El usuario selecciona el tipo de gestión que desea.
5. El sistema pide la información necesaria para la gestión dependiendo del tipo de la misma:
  - a. Si requiere de la selección de ítems, el sistema los presenta para que el usuario los seleccione.

- b. Si requiere del suministro de ciertos parámetros para llevar a cabo la gestión, el sistema los solicita para que el usuario los proporcione.
  - c. Si requiere una combinación de información de varios tipos, el sistema la solicita de forma adecuada.
6. El usuario proporciona la información que se le demanda e intenta enviarla.
7. El sistema comprueba los datos localmente y solicita correcciones si lo considera oportuno. El usuario atiende la petición.
8. El sistema intenta llevar a cabo la operación requerida (puede consistir en enviar una solicitud al usuario encargado de su realización, en consultar el resultado de las operaciones solicitadas ya realizadas, o en ejecutar efectivamente la operación cuando el usuario encargado de ello).
  - a. Si lo consigue, el sistema presenta un mensaje indicando el resultado positivo.
  - b. Si fracasa, el sistema presenta un mensaje explicando el problema y un elemento activo para cancelar la operación.
9. En ambos casos el sistema pregunta si se quiere realizar otra gestión:
  - a. Si la respuesta del usuario es afirmativa, se vuelve al punto donde se muestran los tipos de operaciones disponibles.
  - b. Si la respuesta es negativa, se vuelve a la situación inmediatamente anterior a la solicitud del servicio.

#### **4.12.2.2 Flujos Alternativos**

##### ***4.12.2.2.1 Solicitud Frustrada de Operación de Administración No Delegada (OAND)***

1. Cuando los datos introducidos por el usuario tienen un formato adecuado y se han enviado al servidor, pueden encontrarse problemas al procesarlos que impidan la consumación del proceso.
2. En tal caso, el sistema informa al usuario de la circunstancia y da por finalizado el trámite, dejando al usuario en la situación inmediatamente anterior al inicio de la operación.



#### **4.12.2.2.2 Cancelación de Solicitud de OAND**

1. En cualquiera momento de la secuencia de solicitud el usuario puede cancelar el proceso.
2. En tal caso, el sistema solicitará confirmación de la pretensión.
3. Si no se confirma se volverá al proceso de solicitud, en la posición previa a la petición de cancelación.
4. Si se confirma la cancelación, el sistema informará al usuario de la circunstancia y se finalizará el proceso. El usuario quedará en la situación inmediatamente anterior al inicio de la solicitud.

#### **4.12.2.2.3 Cambio de Identidad con Gestión de OAND en Proceso**

1. Si el usuario decide cambiar de identidad cuando una gestión de OAND en proceso. El sistema informa al usuario de la circunstancia antes del cambio, ofreciendo la posibilidad de cancelar la gestión o abortar el cambio de identidad para poderla completar.
2. Si el usuario decide cancelar la operación, el sistema lo hace y cierra la interfaz de gestión.
3. Si el usuario opta por cancelar el cambio de identidad, el sistema permanece en la misma situación inmediatamente anterior a la solicitud de dicho cambio.

### **4.12.3 Requerimientos Especiales**

1. El proceso de solicitud debe utilizar ventanas específicas y distintas a las de visualización del contenido, para que el usuario pueda recopilar información sobre el conocimiento implicado en la operación a la vez que se cumplimenta el formulario correspondiente.
2. Si un usuario pretende cambiar su identidad en el sistema y tiene una gestión de OAND en proceso, antes debe terminar la gestión o cancelarla.

#### **4.12.4 Condiciones Previas**

##### **4.12.4.1 Identificación Previa de Usuario**

Sólo los usuarios identificados pueden utilizar esta funcionalidad. Los RN son UI, y los AN pueden asumir la identidad de cualquier usuario. Unos y otros deben identificarse para acreditar su condición y que el sistema los reconozca.

#### **4.12.5 Condiciones Posteriores**

##### **4.12.5.1 Respuesta del Usuario Responsable de una Operación**

El usuario responsable de una operación, destinatarios de las solicitudes enviadas por el usuario interesado, tiene que intentar resolver las operaciones administrativas solicitadas y siempre responder al solicitante indicando el trámite realizado y el resultado obtenido.

#### **4.12.6 Notas**

Considerar la posibilidad de que el sistema responda automáticamente, en caso de que los usuarios responsables de una operación no lo hagan en un tiempo establecido por configuración.

## **4.13 ECU: Suplantar Identidad UI**

### **4.13.1 Descripción**

En cualquier momento los usuarios responsables de la gestión de los nodos (URGN), que son los RN y los AN, pueden asumir la identidad de otros usuarios de su ámbito de gestión: los RN dentro del nodo de su responsabilidad y los AN en cualquiera de los nodos bajos su control.

### **4.13.2 Flujo de Eventos**

#### **4.13.2.1 Flujo Básico**

1. En cualquier momento un URGN puede asumir la identidad de los UI adscritos a los nodos que gobiernan. Para ello, el entorno dispone de un elemento activo que permite acceder a la interfaz correspondiente.
2. El URGN solicita la asunción de la identidad de un UI accionando el elemento activo mencionado.
3. Si el URGN es un AN, el sistema presenta los nodos disponibles y el usuario selecciona en el que está el UI que pretende suplantar.
4. El sistema presenta la lista de UI disponibles en el nodo sobre el que se va a trabajar, clasificados por categorías y ordenados por identificadores dentro de cada una de ellas, si no se ha establecido otro criterio de ordenación durante la sesión (para lo que la interfaz incluye un con elemento activo).
5. El URGN selecciona el UI que va a suplantar.
6. El sistema pide confirmación de la operación, presentado toda la información relevante sobre el UI seleccionado. Además, si el usuario suplantado tiene una sesión abierta en ese momento, el sistema indicará la circunstancia y avisará que si se sigue con el proceso, se expulsará al usuario del sistema y no se le dejará volver a entrar en él hasta que haya terminado la suplantación.
7. Si el URGN no confirma la operación, se sigue en la misma situación previa a la selección del UI.

8. Si el URGN confirma la selección, el sistema aparecerá con las mismas funcionalidades y aspecto que el UI encontraría al acceder a él, pero incluyendo además:
  - a. Un área informativa que recuerda el usuario suplantado y el suplantador.
  - b. Un elemento activo para finalizar la suplantación.
  - c. Otro elemento activo para restaurar la configuración previa a la suplantación.
9. El URGN interactúa con el sistema como si fuera el usuario suplantado hasta que decide finalizar la suplantación utilizando el elemento activo correspondiente.
10. El sistema solicita confirmación de la operación.
11. Si el URGN no lo hace volverá a la situación en que estaba antes de la solicitud.
12. Si el URGN la confirma el sistema preguntará si se desea mantener la situación de bloqueo previa a la suplantación, indicando cual era esta y la razón de ella. El sistema terminará la operación de acuerdo con las preferencias del usuario dejando al URGN en la situación previa al cambio de temporal de identidad.

#### **4.13.2.2 Flujos Alternativos**

##### ***4.13.2.2.1 Solicitud de Suplantación Frustrada***

1. Cuando los datos introducidos por el URGN se han enviado al servidor, pueden encontrarse problemas al tratarlos que impidan la consumación del proceso.
2. En tal caso, el sistema informa al URGN de la circunstancia y da por finalizado el trámite, dejando al URGN en la situación inmediatamente anterior al envío de la solicitud (es probable que el URGN pretenda volver a intentarlo o suplantar a otro usuario del mismo nodo).

##### ***4.13.2.2.2 Cancelación de la Solicitud de Suplantación***

1. En cualquiera momento de la secuencia de solicitud el URGN puede cancelar el proceso.
2. En tal caso, el sistema solicitará confirmación de la acción.
3. Si no se confirma se volverá al proceso de solicitud, en la posición donde se produjo la petición de cancelación.

4. Si se confirma la cancelación, el sistema informará al URGN de la circunstancia y se finalizará el proceso. El usuario quedará en la situación inmediatamente anterior al inicio de la solicitud.

#### **4.13.2.2.3 Selección de Ordenación para Selección de UI**

1. Cuando se presenta la lista de UI en el nodo, el URGN puede seleccionar entre varias ordenaciones dentro de las categorías. Para ello el sistema presenta un elemento activo.
2. Si el URGN quiere cambiar la ordenación selecciona el elemento mencionado.
3. El sistema presenta una lista de las ordenaciones disponibles:
  - a. Por identificadores (elección por defecto del sistema).
  - b. Por Apellidos y Nombres.
4. El URGN selecciona la ordenación preferida, esta se considerará el valor por defecto durante el resto del proceso actual.
5. El sistema vuelve al punto del proceso en curso, pero utilizando la nueva ordenación.

#### **4.13.2.2.4 Suplantación de Usuario con Sesión Activa**

1. Si el URGN decide asumir la identidad de un usuario que tiene una sesión activa en ese momento, el sistema informará al URGN de la circunstancia y le dará a elegir entre echar al usuario del sistema o cancelar el proceso de suplantación.
2. Si el URGN decide echar del sistema al usuario suplantado, el sistema terminará la sesión abierta por dicho usuario informándole de la circunstancia, abrirá una nueva sesión controlada por el URGN y bloqueará la cuenta correspondiente hasta que la nueva sesión termine.
3. Si el URGN decide cancelar la operación, el sistema deja al URGN en la situación inmediatamente anterior al envío de la solicitud (es probable que el URGN pretenda volver a intentarlo o suplantar a otro usuario del mismo nodo).

#### **4.13.2.2.5 Restauración de Configuración Usuario Suplantado**

1. En cualquier momento durante la suplantación, el suplantador puede restaurar la configuración que el usuario suplantado tenía antes de que asumieran su identidad. Para ello el sistema incluye en la interfaz de suplantación un elemento activo.

2. El URGN solicita la restauración activando el elemento mencionado.
3. El sistema solicita confirmación de la operación.
4. Si el URGN no lo hace volverá a la situación en que estaba antes de la solicitud.
5. Si el URGN la confirma el sistema realiza la restauración.

#### ***4.13.2.2.6 Finalización de la Suplantación***

1. En cualquier momento durante la suplantación, el suplantador puede dar por terminada la misma. Para ello el sistema incluye en la interfaz de suplantación un elemento activo.
2. El URGN solicita la finalización activando el elemento mencionado.
3. El sistema solicita confirmación de la operación.
4. Si el URGN no lo hace volverá a la situación en que estaba antes de la solicitud.
5. Si el URGN la confirma el sistema preguntará si se desea mantener la situación de bloqueo previa a la suplantación, indicando cual era esta y la razón de ella. El sistema terminará la operación de acuerdo con las preferencias del usuario dejando al URGN en la situación previa al cambio de temporal de identidad.

### **4.13.3 Requerimientos Especiales**

1. En todo momento debe quedar claro el usuario suplantado y el suplantador.
2. El sistema no permite suplantaciones anidadas. Por eso cuando un URGN suplanta a otro URGN, en la interfaz del suplantado debe estar bloqueada la opción de suplantación.
3. La interfaz del usuario suplantado debe ser igual que la que este encontraría al acceder al sistema por si mismo, salvo en el caso anterior.
4. La interfaz del sistema debe incorporar los elementos necesarios para:
  - a. Informar de la identidad del suplantador.
  - b. Informar de la identidad del suplantado.
  - c. Dar por terminada la suplantación.
  - d. Incluir utilidades para facilitar la suplantación, como restaurar la última configuración vigente al iniciar la suplantación.
5. Un usuario no puede tener una sesión activa en el sistema cuando se le suplanta.

6. Si se pretende suplantar a un usuario que tiene una sesión activa en el sistema, el sistema debe notificar el hecho al URGN para que este decida si prefiere echar al usuario para seguir con la operación o desistir de ella.
7. Durante la sesión de suplantación no están vigentes las restricciones impuestas al suplantado, sino las asignadas al suplantador. Esto quiere decir que ciertas operaciones limitadas para el UI suplantado podrán realizarse cuando un URGN esté actuando en su nombre. En tal caso, dichas operaciones aparecerán en los menús cuando el usuario sea apropiado.

#### **4.13.4 Condiciones Previas**

##### **4.13.4.1 Identificación Previa de Usuario**

Sólo los usuarios responsables de la gestión de los nodos (URGN), que son los RN y los AN, pueden utilizar esta funcionalidad. Los URGN deben identificarse para acreditar su condición y que el sistema los reconozca.

#### **4.13.5 Condiciones Posteriores**

##### **4.13.5.1 Desbloqueo Automático de Cuenta de Usuario**

La cuenta bloqueada por una suplantación quedará liberada automáticamente al terminar dicha suplantación, a no ser que el URGN indique lo contrario en ese momento.

#### **4.13.6 Notas**

Cuando se produce una suplantación de un usuario con alguna sesión activa, el sistema puede terminar unilateralmente la sesión, pero el usuario no se dará cuenta hasta que se establezca alguna interacción con el entorno. Si estuviera activo el módulo de registro de actividad de cliente, la notificación podría llegar en el retorno del último registro. En cualquier caso, hay que estudiar el procedimiento de notificación general y el modo de manejar las distintas situaciones que pudieran producirse.

Los usuarios suplantados no pueden tener sesiones propias durante la suplantación, pero pueden tener sesiones paralelas cuando nadie está asumiendo su identidad. El uso de sesiones paralelas puede ser útil en determinadas situaciones si el usuario mantiene el control de lo que en ellas pasa. Sin embargo, cuando otro usuario controla alguna de las sesiones en nombre del propietario pero sin su control, la experiencia puede resultar desconcertante para ambos.

## **4.14 ECU: Gestionar Configuración Usuarios**

### **4.14.1 Descripción**

En cualquier momento, los usuarios responsables de la gestión de los nodos (URGN), que son los RN y los AN, pueden acceder y modificar la configuración de los UI de su ámbito de gestión: los RN dentro del nodo de su responsabilidad y los AN en cualquiera de los nodos bajos su control.

Aunque parte de esta labor la pueden realizar asumiendo la identidad de cada UI afectado secuencialmente, existen aspectos de la configuración de estos usuarios que podrían no están delegadas para ellos, y no siempre los procedimientos interactivos establecidos para la autogestión resultan los más adecuados para tratar la configuración de grupos numerosos de usuarios.

### **4.14.2 Flujo de Eventos**

#### **4.14.2.1 Flujo Básico**

1. En cualquier momento un URGN puede acceder y modificar la configuración de los UI adscritos a los nodos que gobiernan. Para ello, el entorno dispone de un elemento activo que permite acceder a la interfaz correspondiente.
2. El URGN solicita la edición de la configuración de UI accionando el elemento activo mencionado.
3. Si el URGN es un AN, el sistema presenta los nodos disponibles y el usuario selecciona en él que están los UI sobre cuyas configuraciones se pretende actuar.
4. El sistema presenta la lista de UI disponibles en el nodo sobre el que se va a trabajar, clasificados por categorías y ordenados por identificadores dentro de cada una de ellas, si es que no se ha establecido otro criterio de ordenación durante la sesión (para lo que la interfaz incluye un elemento activo como en otros casos).
5. El URGN selecciona los UI sobre cuyas configuraciones se pretende actuar.
6. El sistema presenta los aspectos configurables para ellos.
7. El URGN selecciona el aspecto que quiere configurar.
8. El sistema presenta una interfaz de configuración adecuada para cada caso.



9. El URGN indica sus preferencias e intenta enviar los datos.
10. El sistema comprueba los datos localmente y solicita correcciones si lo considera oportuno.
11. El sistema intenta realizar la configuración indicada.
12. El sistema presenta un mensaje mostrando el resultado de la operación y preguntado si el URGN quiere seguir gestionando la configuración de los usuarios.
13. Si el URGN indica que ha terminado, el sistema lo deja en la situación previa al inicio de la operación.
14. Si el URGN quiere seguir trabajando con la configuración de los usuarios, el sistema lo deja en el punto del proceso donde se seleccionan los UI.

#### **4.14.2.2 Flujos Alternativos**

Como en el caso Gestionar Configuración Usuario, pero aplicadas a conjuntos de usuarios:

- Configuración Usuario Frustrada
- Cancelación de Configuración Usuario
- Aplicar Configuración Usuario por Defecto.

Además, como en el caso Suplantar Identidad UI:

- Selección de Ordenación para Selección de UI.

#### **4.14.3 Requerimientos Especiales**

Como resultado de la gestión de configuración de un grupo, puede que el trámite sólo llegue a buen fin para algunos usuarios. En tal caso el sistema debe informar del resultado indicando los usuarios para los que el trámite no ha sido posible, pero dando como bueno la parte terminada sin problemas.

#### **4.14.4 Condiciones Previas**

##### **4.14.4.1 Identificación Previa de Usuario**

Sólo los usuarios responsables de la gestión de los nodos (URGN), que son los RN y los AN, pueden utilizar esta funcionalidad. Los URGN deben identificarse para acreditar su condición y que el sistema los reconozca.

#### **4.14.4.2 Configuración Usuario por Defecto**

Los usuarios siempre dispondrán de una configuración por defecto que el sistema aplicará cuando los usuarios no están identificados o no se hayan establecido valores distintos para los parámetros configurables.

#### **4.14.5 Condiciones Posteriores**

##### **4.14.5.1 Caducidad Configuración**

Las configuraciones de UI establecidas por este procedimiento se convierten en las iniciales de las siguientes sesiones de los usuarios.

#### **4.14.6 Notas**

Pensar en la conveniencia de considerar las gestiones sólo como acciones completas, de un modo menos tolerante que el actual.

## **4.15 ECU: Gestionar Contenido Usuarios**

### **4.15.1 Descripción**

En cualquier momento, los usuarios responsables de la gestión de los nodos (URGN), que son los RN y los AN, pueden acceder y modificar los contenidos de los UI de su ámbito de gestión: los RN dentro del nodo de su responsabilidad y los AN en cualquiera de los nodos bajos su control.

Aunque parte de esta labor la pueden realizar asumiendo la identidad de cada UI afectado secuencialmente, existen aspectos de la gestión de los contenidos de estos usuarios que podrían no están delegadas para ellos, y no siempre los procedimientos interactivos establecidos para la autogestión resultan los más adecuados para tratar los contenidos de grupos numerosos de usuarios.

### **4.15.2 Flujo de Eventos**

#### **4.15.2.1 Flujo Básico**

1. En cualquier momento un URGN puede acceder y gestionar los contenidos de los UI adscritos a los nodos que gobiernan. Para ello, el entorno dispone de un elemento activo que permite acceder a la interfaz correspondiente.
2. El URGN solicita la gestión de los contenidos de UI accionando el elemento activo mencionado.
3. Si el URGN es un AN, el sistema presenta los nodos disponibles y el usuario selecciona en él que están los UI cuyos contenidos se pretenden gestionar.
4. El sistema presenta la lista de UI disponibles en el nodo sobre el que se va a trabajar, clasificados por categorías y ordenados por identificadores dentro de cada una de ellas, si es que no se ha establecido otro criterio de ordenación durante la sesión (para lo que la interfaz incluye un elemento activo como en otros casos).
5. El URGN selecciona los UI cuyos contenidos se pretenden gestionar.
6. El sistema presenta los aspectos gestionables para ellos.
7. El URGN selecciona el aspecto que quiere gestionar.
8. El sistema presenta una interfaz adecuada para cada caso.

9. El URGN indica sus preferencias e intenta enviar los datos.
10. El sistema comprueba los datos localmente y solicita correcciones si lo considera oportuno.
11. El sistema intenta realizar la gestión indicada.
12. El sistema presenta un mensaje mostrando el resultado de la operación y preguntado si el URGN quiere seguir gestionando los contenidos de los usuarios.
13. Si el URGN indica que ha terminado, el sistema lo deja en la situación previa al inicio de la operación.
14. Si el URGN quiere seguir trabajando con los contenidos de los usuarios, el sistema lo deja en el punto del proceso donde se seleccionan los UI.

#### **4.15.2.2 Flujos Alternativos**

Como en el caso “Gestionar Contenido UI”, pero aplicadas a conjuntos de usuarios:

- Gestión de Contenido Frustrada
- Cancelación de Gestión de Contenido en Curso

Además, como en el caso “Suplantar Identidad UI”:

- Selección de Ordenación para Selección de UI.

#### **4.15.3 Requerimientos Especiales**

Como resultado de la gestión de configuración de un grupo, puede que el trámite sólo llegue a buen fin para algunos usuarios. En tal caso el sistema debe informar del resultado indicando los usuarios para los que el trámite no ha sido posible, pero dando como bueno la parte terminada sin problemas.

#### **4.15.4 Condiciones Previas**

##### **4.15.4.1 Identificación Previa de Usuario**

Sólo los usuarios responsables de la gestión de los nodos (URGN), que son los RN y los AN, pueden utilizar esta funcionalidad. Los URGN deben identificarse para acreditar su condición y que el sistema los reconozca.

#### **4.15.4.2 Configuración Usuario por Defecto**

Los usuarios siempre dispondrán de una configuración por defecto que el sistema aplicará cuando los usuarios no están identificados o no se hayan establecido valores distintos para los parámetros configurables.

#### **4.15.5 Condiciones Posteriores**

##### **4.15.5.1 Caducidad Configuración**

Las configuraciones de UI establecidas por este procedimiento se convierten en las iniciales de las siguientes sesiones de los usuarios.

#### **4.15.6 Notas**

Pensar en la conveniencia de considerar las gestiones sólo como acciones completas, de un modo menos tolerante que el actual.

## **4.16 ECU: Registrar Actividad Remota Usuario**

### **4.16.1 Descripción**

El MR recopila datos de la actividad del usuario en el cliente del sistema cuando accede al conocimiento a través del entorno. EL MR registra la información así obtenida en el sistema, para que este la utilice como indicador de intensidad de la actividad del usuario con los elementos de conocimiento.

### **4.16.2 Flujo de Eventos**

#### **4.16.2.1 Flujo Básico**

1. En cualquier momento, el MR puede solicitar registrar información de actividad en el cliente de un usuario proporcionando todos los datos necesarios.
2. El sistema intenta realizar la operación solicitada.
3. El sistema indica el resultado de la operación.

#### **4.16.2.2 Flujos Alternativos**

##### ***4.16.2.2.1 Identificación Usuario Registrado***

1. En cualquier momento, el MR puede solicitar la identificación de un usuario, proporcionando los datos correspondientes.
2. El sistema intenta realizar la operación solicitada.
3. El sistema indica el resultado de la operación.

### **4.16.3 Requerimientos Especiales**

1. El MR no es un usuario humano, esta circunstancia debe considerarse en la interacción del sistema con el módulo.
2. El MR tiene que tener su interfaz y parte de su lógica empotrada en el cliente del sistema.
3. El MR registra actividad solamente cuando esta existe, por lo que tiene un comportamiento asíncrono.

4. El MR asigna un identificador de usuarios unívoco a los usuarios que no se han identificado previamente. La determinación del identificador de usuario es, por tanto, previa al inicio del registro de actividad, sea mediante la identificación del usuarios como UI del sistema o por asignación directa de identificador.
5. La configuración del sistema determina si el MR registrará la actividad del sistema y si se requerirá la identificación previa del usuario.

#### **4.16.4 Condiciones Previas**

##### **4.16.4.1 Identificación Unívoca de Usuario**

Para poder realizar la tarea pretendida es preciso identificar unívocamente al usuario. Esto es posible si el usuario se identifica como UI del sistema o si se le asigna un identificador único vinculado a su sesión de cliente de interacción con el sistema, en caso contrario.

#### **4.16.5 Condiciones Posteriores**

##### **4.16.5.1 Reversibilidad del Establecimiento de Consideración**

En establecimiento de consideración debe ser reversible mientras los datos establecidos no se hayan utilizado para la cristalización.

#### **4.16.6 Notas**

Debe considerarse la implementación de este caso en combinación con el de “Mostrar Actividad del Sistema”, que se especifica también en este documento.

## **4.17 ECU: Mostrar Actividad Sistema**

### **4.17.1 Descripción**

Además de registrar la intensidad de la actividad de los usuarios con los elementos de conocimiento, el MR proporciona al usuario información sobre el sistema y la actividad del mismo, mediante la interfaz del entorno en el cliente.

### **4.17.2 Flujo de Eventos**

#### **4.17.2.1 Flujo Básico**

1. El MR solicita información sobre la actividad del sistema proporcionando todos los datos requeridos.
2. El sistema intenta obtener la información solicitada.
  - a. Si no la consigue, indica el problema encontrado.
  - b. Si consigue la información, la entrega al MR.

#### **4.17.2.2 Flujos Alternativos**

Como en el caso “Registro Actividad Remota Usuario”.

### **4.17.3 Requerimientos Especiales**

De forma parecida al caso “Registro Actividad Remota Usuario”:

1. El MR no es un usuario humano, esta circunstancia debe considerarse en la interacción del sistema con el módulo.
2. El MR tiene que tener su interfaz y parte de su lógica empotrada en el cliente del sistema.
3. El MR solicita información sobre el sistema solamente cuando existe actividad, por lo que tiene un comportamiento asíncrono.
4. Es imprescindible que el usuario esté identificado para poder determinar la información sobre la actividad del sistema que se le va a mostrar.
5. El MR asigna un identificador de usuarios unívoco a los usuarios que no se han identificado previamente. La determinación del identificador de usuario es, por



tanto, previa al inicio de la presentación de información sobre la actividad del sistema, sea mediante la identificación del usuarios como UI del sistema o por asignación directa de identificador.

6. La configuración del sistema determina si el MR mostrará información sobre la actividad del sistema y si para ello se requerirá la identificación previa del usuario.

#### **4.17.4 Condiciones Previas**

Como en el caso “Registro Actividad Remota Usuario”.

#### **4.17.5 Condiciones Posteriores**

Como en el caso “Registro Actividad Remota Usuario”.

#### **4.17.6 Notas**

Debe considerarse la implementación de este caso en combinación con el de “Registro de Actividad Remota de Usuario”, que se especifica también en este documento.

## **4.18 ECU: Asociar Ítems de Conocimiento**

### **4.18.1 Descripción**

El MA estudia los elementos de conocimiento para establecer y explicitar relaciones implícitas entre ellos. Asociar automáticamente elementos asignados explícitamente a otro con temáticas parecidas, o asociar usuarios con intereses similares, son ejemplos de ello.

### **4.18.2 Flujo de Eventos**

#### **4.18.2.1 Flujo Básico**

1. En cualquier momento, el MA puede proporcionar los datos necesarios para el establecimiento de una asociación entre elementos de conocimiento.
2. El sistema intenta realizar la operación solicitada.
3. El sistema indica el resultado de la operación.

#### **4.18.2.2 Flujos Alternativos**

##### ***4.18.2.2.1 Obtención de Información sobre Estado de Carga del Sistema***

1. En cualquier momento, el módulo puede solicitar información sobre el estado de carga del sistema (para determinar la oportunidad de interactuar con el entorno en un determinado momento).
2. El sistema intenta realizar la operación solicitada.
3. El sistema indica el resultado de la operación.

##### ***4.18.2.2.2 Bloqueo Funcionalidades Incompatibles con Proceso***

1. En cualquier momento, el módulo puede solicitar el bloqueo de las funcionalidades del sistema incompatibles con el proceso. Este es un trámite previo e imprescindible para realizar el proceso.
2. El sistema intenta realizar la operación solicitada.
3. El sistema indica el resultado de la operación.

#### **4.18.2.2.3 Desbloqueo Funcionalidades Incompatibles con Proceso**

1. En cualquier momento, el módulo puede solicitar el desbloqueo de las funcionalidades del sistema incompatibles con el proceso. Esta operación es imprescindible y sólo tiene sentido después del bloqueo correspondiente.
2. El sistema intenta realizar la operación solicitada.
3. El sistema indica el resultado de la operación.

#### **4.18.3 Requerimientos Especiales**

1. El MA no es un usuario humano, esta circunstancia debe considerarse en la interacción del sistema con el módulo.
2. El MA es un módulo que funciona en el lado del servidor de forma autónoma del resto del sistema. Para ello, sigue una planificación establecida que se puede variar de acuerdo a una política sensible a las circunstancias del desarrollo de la actividad (como el estado de carga del sistema en cada momento).
3. El MA analiza las asociaciones o las establece, solamente cuando su planificación y las circunstancias lo permiten, por lo que en sus interacción con el sistema tiene un comportamiento asíncrono.
4. El MA analiza y establece las asociaciones entre elementos de conocimiento, a medida que se producen cambios en el conocimiento. Por ejemplo, cuando se incorporan nuevos documentos al sistema o cuando cambia la consideración de un usuario por cierto elemento de conocimiento.
5. El MA debe analizar y establecer las asociaciones entre elementos de conocimiento, según el plan establecido y procurando que los datos estén lo antes posible disponibles, para facilitar la interacción del usuario y la realización de otros procesos que los pudieran requerir (como el establecimiento de consideración implícita o la cristalización).
6. Esta labor de análisis es especialmente pesada, debido a que puede requerirse el cruce constante de datos de muchos de los elementos de conocimiento incluidos en el sistema.

#### **4.18.4 Condiciones Previas**

##### **4.18.4.1 Disponibilidad de Datos de Elementos de Contenido**

Para la realización de esta actividad es imprescindible la existencia de datos actualizados sobre los elementos de conocimientos incluidos en el sistema (como vectores descriptivos, consideración, etc.).

#### **4.18.5 Condiciones Posteriores**

Ningunas.

#### **4.18.6 Notas**

Deben considerarse las dependencias del caso que en esta sección nos ocupa con el de “Manifestar Consideración Explícita...”, “Establecer Consideración Implícita...” y el de “Cristalizar Conocimiento”, que se especifican también en este documento.

## **4.19 ECU: Establecer Consideración Implícita por Contenido Ajeno de UI**

### **4.19.1 Descripción**

El MA analiza la actividad de los usuarios para establecer la consideración sobre los elementos de conocimiento que de ella se deriva. Comportamientos como el trabajo prolongado y repetido de un usuario con determinado elemento, indica el interés del usuario por este, aunque no haya explicitado dicho inclinación.

Para ello, el MA examina periódicamente los registros del sistema y determinar la consideración de los usuarios por los elementos de conocimiento ajenos con los que han trabajado. Los resultados de dicho análisis se incorporan al sistema mediante este caso de uso.

### **4.19.2 Flujo de Eventos**

#### **4.19.2.1 Flujo Básico**

1. En cualquier momento, el MA puede proporcionar los datos necesarios para el establecimiento de consideración implícita de un elemento de conocimiento.
2. El sistema intenta realizar la operación solicitada.
3. El sistema indica el resultado de la operación.

#### **4.19.2.2 Flujos Alternativos**

Como en el caso “Asociar Ítems de Conocimiento”.

### **4.19.3 Requerimientos Especiales**

1. El MA no es un usuario humano, esta circunstancia debe considerarse en la interacción del sistema con el módulo.
2. El MA es un módulo que funciona en el lado del servidor de forma autónoma del resto del sistema. Para ello, sigue una planificación establecida que se puede variar de acuerdo a una política sensible a las circunstancias del desarrollo de la actividad (como el estado de carga del sistema en cada momento).

3. El MA establece la consideración o analiza la actividad, solamente cuando su planificación y las circunstancias lo permiten, por lo que en sus interacción con el sistema tiene un comportamiento asíncrono.
4. El MA analiza la actividad de todos los usuarios por orden temporal creciente en periodos configurables consecutivos.
5. El MA debe mantener el establecimiento de la consideración correspondiente según el plan establecido y procurando que los datos estén lo antes posible disponibles, para facilitar la realización de otros procesos basados en ellos (como la cristalización).

#### **4.19.4 Condiciones Previas**

##### **4.19.4.1 Disponibilidad de Registros de Actividad Remota de Usuarios**

Para la realización de esta actividad es imprescindible la existencia de registros de actividad de usuario hasta un momento previo configurable.

#### **4.19.5 Condiciones Posteriores**

Ninguna.

#### **4.19.6 Notas**

Deben considerarse las dependencias del caso “Cristalizar Conocimiento”, que se especifica también en este documento, con el que en esta sección nos ocupa.

## **4.20 ECU: Cristalizar Conocimiento**

### **4.20.1 Descripción**

El MC selecciona el conocimiento que se gestiona en el sistema mediante el proceso de cristalización. Para ello, establece el grado de cristalización de cada elemento de contenido a partir de las demostraciones de consideración hacia él manifestadas por los usuarios de forma explícita e implícita.

#### **4.20.1.1 Flujo Básico**

1. En cualquier momento, el MC puede establecer el grado de cristalización de un elemento de conocimiento.
2. El sistema intenta realizar la operación solicitada.
3. El sistema indica el resultado de la operación.

#### **4.20.1.2 Flujos Alternativos**

Como en el caso “Asociar Ítems de Conocimiento”.

### **4.20.2 Requerimientos Especiales**

1. El MC no es un usuario humano, esta circunstancia debe considerarse en la interacción del sistema con el módulo.
2. El MC es un módulo que funciona en el lado del servidor de forma autónoma del resto del sistema, aunque compartiendo con él datos. Para desarrollar su labor sigue una planificación establecida que se puede variar de acuerdo a una política sensible a las circunstancias del desarrollo de la actividad (como el estado de carga del sistema en cada momento).
3. El MC analiza la consideración manifestada sobre determinados elementos (como documentos y grafos) de contenido y establece los correspondientes grados de cristalización, solamente cuando su planificación y las circunstancias lo permiten, por lo que en sus interacción con el sistema tiene un comportamiento asíncrono.
4. El MC analiza la consideración manifestada por los elementos de conocimiento de forma sistemática según la organización establecida para los mismos, por ejemplo:

los documentos dentro de un vértice de un grafo o los grafos (vistas) dentro del nodo.

5. El proceso de cristalización puede requerir el bloqueo de ciertas funcionalidades durante su realización, con el objetivo de garantizar la consistencia de los resultados.

### **4.20.3 Condiciones Previas**

#### **4.20.3.1 Disponibilidad de Datos de Consideración Suficientes**

Para la realización de esta actividad es imprescindible la existencia de datos de consideración suficientes obtenidos a partir de manifestaciones explícitas u obtenidos por el MA a partir de los registros de actividad de los usuarios de forma implícita.

### **4.20.4 Condiciones Posteriores**

Ninguna.

### **4.20.5 Notas**

Deben considerarse las dependencias del caso que en esta sección nos ocupa con el de “Manifestar Consideración Explícita...”, “Asociar Ítems de Conocimiento” y el de “Establecer Consideración Implícita...”, que se especifican también en este documento.