

Conferencia.
Facultad de Documentación.
Universidad de Murcia,
4 de mayo de 1999

AGENTES INTELIGENTES APLICADOS A LA GESTIÓN DE INFORMACIÓN PARA LA INVESTIGACIÓN

José Angel Martínez Usero

Objetivos

- Analizar las diferentes herramientas existentes para la búsqueda y recuperación de contenidos en Internet.
- Estudiar las diferencias y semejanzas, ventajas e inconvenientes de los índices temáticos/directorios, motores de búsqueda y agentes inteligentes.
- Destacar las características propias de los agentes inteligentes y sus posibilidades en la gestión de información científica, técnica y comercial.
- Enunciar las tendencias futuras en el ámbito de las herramientas para la búsqueda y recuperación de información en Internet.

1. EL DESARROLLO DE LOS ACTUALES MOTORES DE BÚSQUEDA: UNA INTRODUCCIÓN.

La aparición y posterior expansión de Internet y las tecnologías asociadas (sobre todo WWW) nos ha ofrecido la posibilidad de acceder a una ingente cantidad de información. Esta situación ha obligado a los profesionales de la información al diseño, creación y utilización de herramientas para la búsqueda y recuperación de la información pertinente de una forma rápida y sencilla, entre estas herramientas destacan los índices temáticos, los motores de búsqueda (“buscadores” = search engines) y los agentes inteligentes.

La localización y recuperación de la información en el WWW es uno de los principales problemas que afrontan los documentalistas, las herramientas existentes (índices temáticos y motores de búsqueda) no han resuelto adecuadamente los problemas de ruido documental y escasa exhaustividad en los resultados. Los metabuscadores pueden ayudar a solucionar estos problemas. Un paso más allá, los agentes de información ofrecen un conjunto de utilidades para la búsqueda y recuperación de contenidos en la web que supera a los anteriores y se imponen como una solución viable a los problemas de gestión de recursos web.

1.1. Problemas terminológicos

En España existen diferentes imprecisiones terminológicas a este respecto, Por un lado, a los motores de búsqueda se les ha denominado con otros términos “sinónimos”, tales como: buscadores, rastreadores, webcrawlers, agents, índices, directorios (VAQUERO, JR, 1997, p. 31). Por otro, durante cierto tiempo, se han confundido tres tecnologías que ahora tienen autonomía propia: los índices temáticos/directorios, los motores de búsqueda y los agentes inteligentes.

En principio, la diferencia entre motor de búsqueda e índice temático o directorio parece clara. Un índice temático es una página web (sitio web) en donde, las distintas materias se encuentran organizadas en torno a un conjunto de epígrafes. Esta diferencia se tambalea cuando nos encontramos con índices temáticos como Yahoo (www.yahoo.com), que presenta un interfaz similar a los motores e incluso permite realizar búsquedas sobre los recursos que tiene sistematizados. En general, la diferencia radica en el hecho de que los índices temáticos contienen direcciones que son recopiladas, organizadas y clasificadas manualmente y la búsqueda se lleva a cabo exclusivamente sobre los recursos indexados del directorio.

Los agentes inteligentes pueden realizar una serie de tareas sin que los humanos u otros agentes les tengan que decir qué hacer a cada paso que dan en su camino. Se diferencian de los motores de búsqueda en que éstos albergan contenidos estáticos (aunque se actualizan con cierta frecuencia) y responden directamente a las peticiones de los usuarios.

Si un motor de búsqueda pudiera almacenar peticiones de los usuarios y notificarles la llegada de información útil, entonces el motor de búsqueda sería un agente. Aunque hemos de decir que según los comentarios de Isidro F. Aguillo, la diferenciación no es radicalmente clara, puesto que se denominan agentes inteligentes a softwares que realmente no lo son (Aguillo, I.F., 1999, p. 241).

A continuación se presenta una definición estándar de cada una de las herramientas de búsqueda y recuperación de información mencionadas anteriormente.

Indice temático/Directorio: página web donde de forma manual se organizan jerárquicamente un conjunto de recursos web que han sido recopilados a partir de búsquedas (también manuales) en la Red.

Motor de búsqueda: programa que tiene por objetivo la estructura hipertextual de la web para recoger e indexar de forma automática los recursos disponibles en la Red, formando una gran base de datos integrada por documentos html (ahora también otros DTDs de SGML como xml, vml, dhtml etc.)

Agente inteligente: un agente inteligente es un software que asiste al cliente y actúa en su nombre.

1.2. ¿Cómo funcionan los motores de búsqueda?

Un motor de búsqueda está formado por cuatro elementos básicos (Codina, 1997, p. 22-23):

1. Un **programa** (también denominado robot, rastreador o webcrawler) que recorre el WWW buscando recursos de información y sus respectivas URLs.
2. Un **sistema automático** de análisis de contenidos e indexación de los documentos localizados por el robot.
3. Un **sistema de interrogación**, generalmente basado en la lógica booleana, que permite al usuario expresar su necesidad de información.
4. Un **programa que actúa de pasarela** entre el servidor de documentos html y la base de datos.

Funcionamiento: el motor de búsqueda recibe la consulta del usuario (query), formada por uno o más términos, realiza una consulta interna en la base de datos que contiene los recursos web indexados y ofrece una lista de aquellos recursos que cumplen una parte o el total de los requisitos establecidos en la consulta. Generalmente, los resultados aparecen ordenados según una puntuación (score) que el programa asocia automáticamente a cada recurso.

Para realizar una consulta es necesario tener en cuenta un conjunto de variables:

1. Lenguaje de interrogación, que debe ofrecer diferentes tipos de operadores: lógicos, de comparación, de truncamiento, de proximidad, de especificación de campo.
2. Posibilidad de refinar (refine) una búsqueda inicial.
3. Campos limitadores que nos permitan reducir la búsqueda: dominios, lenguas, países, fecha de creación del recurso.
4. Búsquedas alternativas: búsqueda simple, búsqueda avanzada, búsquedas combinando operadores e índices temáticos, etc.
5. Opciones avanzadas: buscar diferentes recursos (texto, sonido, imagen), guardar y reutilizar búsquedas, diferentes formatos en los resultados de búsqueda (estándar, detallado, compacto, etc.), búsqueda de conceptos

relacionados (related topics), consulta directa en bases de datos (infranets), etc.

1.3. Los metabuscadores: un paso más allá

La gran cantidad de información y el notable aumento de motores de búsqueda accesibles desemboca en la necesidad de realizar consultas simultáneas en diferentes motores de búsqueda y con una sola estrategia (query). De esta necesidad surgen los denominados “metabuscadores”, que ofrecen nuevas prestaciones y mejores y más exhaustivos resultados de búsqueda.

Los metabuscadores permiten automatizar el proceso de realizar una misma consulta en diversos motores de búsqueda, lo cual no significa que sea totalmente exhaustivo, puesto que el metabuscador envía la consulta solamente a aquellos motores de búsqueda con los que ha establecido un acuerdo previo.

En el funcionamiento de los metabuscadores cabe destacar algunas variables interesantes. Por una lado, la exhaustividad no está garantizada (desde el momento en el que sabemos que un motor de búsqueda es capaz de indexar a lo sumo un 30% de los recursos web disponibles). Por otro, los tiempos de respuesta pueden ser mucho más largos dada la necesidad de realizar múltiples búsquedas simultáneas (Metacrawler permite delimitar el tiempo máximo de espera de 1 a 10 minutos), además, la recuperación de recursos duplicados suele ser muy elevada, por ello, algunos metabuscadores ya han implementado la utilidad que permite eliminar los duplicados.

Algunos de los metabuscadores más prestigiosos del mercado son:

Metacrawler <http://www.metacrawler.com>

All-in-One Search Page <http://www.albany.net/allinone>

Search.Com <http://www.search.com>

Mundo Latino <http://www.mundolatino.org/123/>

1.4. Últimas tendencias

Partiendo de los problemas actuales que presentan los motores de búsqueda en cuanto a su funcionamiento y los resultados ofrecidos se pueden adelantar algunas vías de solución futura que marcan la tendencia en la evolución de los motores de búsqueda.

Los **resultados de la búsqueda** pueden ser satisfactorios o no tanto. Los motores de búsqueda ofrecen resultados muy diferentes ante una misma cuestión inicial; este hecho demuestra la poca exhaustividad de los motores en la indexación de los recursos web y pone de manifiesto los problemas derivados de la escasez de control lingüístico.

Actualmente se aboga por la incorporación de herramientas de análisis lingüístico y control terminológico en los motores de búsqueda, de forma que sea posible efectuar una recuperación menos ligada a la comparación de cadenas de caracteres y más vinculada a la comparación de conceptos.

La escasa **calidad de la información recuperada** es otro inconveniente de los actuales motores de búsqueda. Los mecanismos para aumentar la precisión en la búsqueda (refinamientos, búsquedas avanzadas, acotación por dominios, etc.), a veces, no funcionan como cabría esperar. A ello, hay que añadir el mínimo valor de algunos de los sitios web recuperados, el porcentaje de recursos repetidos y el porcentaje de recursos inactivos (que ya no existen físicamente en la red aunque continúan indexados).

En este sentido se empieza a hablar de una Internet para el gran público y una Internet de los recursos culturales, científicos y técnicos. La especialización de los motores de búsqueda es una buena vía para conseguir mejores servicios de información, la especialización conduce a la concentración del conocimiento en ciertos lugares donde los usuarios pueden encontrar fácilmente los recursos relacionados con su ámbito de conocimiento (KOSMYNIN, 1997, p. 148).

Los mecanismos de análisis e indexación de recursos.

En 1997 parecía que el uso de los metadatos (Metadata Workshop, 1998), (etiqueta "metadata" en la cabecera de un documento HTML) para la indexación de los recursos web constituiría la solución futura a los problemas de representación de contenidos en Internet y, a pesar de proyectos tan ambiciosos como el Dublin Core, los metadatos ya no constituyen una solución viable, puesto que sería ingenuo esperar que todos los particulares y empresas autocatalogarán con metadatos todos los recursos que publiquen en la web.

Más adelante los DOIs (Digital Object Identifier) como sistema para nombrar cualquier tipo de objeto (documento, imagen, sonido, etc.) estableciendo una asociación entre el nombre del objeto y la dirección de Internet donde se encuentra situado, han supuesto una alternativa, aunque orientada sobre todo al control de la propiedad intelectual de los recursos electrónicos (DOI Workshop, 1998).

Actualmente se opta por un DTD de SGML, como XML (Santamaría, F., 1998), que permita tratar los recursos web de una forma conceptual y trabajar con sus contenidos como si se tratara de registros de una base de datos documental.

Los motores de búsqueda en las Intranets.

La misma tecnología usada para la localización de información en Internet puede ser adaptada para la búsqueda de contenidos en una intranet corporativa e incluso superar estas prestaciones, puesto que son capaces de

mantener un informe de las revisiones de los documentos y otras utilidades de gestión. Por ejemplo, pueden mantener una base de datos de los currícula corporativos, de forma que cuando la empresa necesite formar un equipo para abordar un proyecto, el motor de búsqueda selecciona los candidatos atendiendo a la información contenida en la intranet.

Un ejemplo claro de la utilidad de implementar motores de búsqueda sobre una intranet corporativa lo representa el FBI que utiliza Altavista Search Intranet para gestionar la base de datos de historiales criminales, lo cual ha supuesto una reducción del tiempo de búsqueda de varios días a unos 2.8 segundos (Matthew, 1998, p. 122).

2. LOS AGENTES INTELIGENTES: DEFINICIÓN, CARACTERÍSTICAS Y TIPOLOGÍAS.

2.1. Una aproximación al concepto de agente inteligente.

El concepto de agente inteligente es tan amplio y posee tantas aplicaciones que no es sencillo aportar una definición exacta. Bradshaw (Bradshaw, J., 1997, p. 5) señala que la definición de agente depende del punto de vista del investigador o de los atributos propios del agente. En la misma línea, Nwana (Nwana, J. 1996) afirma que el concepto de agente ya se puede encontrar en la investigación en Inteligencia Artificial de la década de los 70s, pero que continúa siendo un término difuso, un meta-término o paraguas que da cobertura a diferentes enfoques. Finalmente, Tramullas (Tramullas, J., 1999, p. 247) comenta que la complejidad que rodea al ámbito de los agentes, donde intervienen la Inteligencia Artificial, la Sociología, la Lógica, las Telecomunicaciones y otras disciplinas, hace necesario el estudio de las situaciones en las que puede encontrarse un agente para definirlo de forma adecuada.

Stenmark ofrece una definición genérica: un agente inteligente es un software que asiste al cliente y actúa en su nombre (Stenmark, D., 1998-B)

Hipola y Vargas-Quesada lo definen como una entidad software que, basándose en su propio conocimiento, realiza un conjunto de operaciones destinadas a satisfacer las necesidades de un usuario o de otro programa, bien por iniciativa propia o porque alguno de éstos se lo requiere (Hipola, P., 1999, p. 13).

Maes lo define como: Programas de ordenador capaces de efectuar una tarea o actividad sin la manipulación directa de un usuario humano (Maes, P., 1994). Los agentes inteligentes han cambiado sustancialmente la forma de interacción hombre-máquina. El usuario delega diferentes tareas a los agentes que son capaces de actuar en su nombre. Además, los agentes tienen la característica esencial de aprender de diferentes formas:

- Observando e imitando el comportamiento del usuario
- Recibiendo un feedback positivo o negativo del usuario
- Recibiendo instrucciones explícitas del usuario

- Pidiendo consejo a otros agentes

2.2. Un agente se caracteriza por...

Las características que un programa debe poseer para ser considerado un agente inteligente, en opinión de los expertos, son:

- **Autonomía:** el agente debe tener control sobre sus propias acciones y ser capaz de lanzar acciones independientemente del usuario.
- **Capacidad de reacción:** los agentes pueden detectar cambios en su entorno y reaccionar en función de éstos.
- **Comunicatividad:** el agente es capaz de interactuar con los usuarios y otros agentes.
- **Consecución de metas:** los agentes tienen un propósito determinado y actúan en consecuencia hasta conseguirlo.

Otras características de los agentes inteligentes reseñadas en la mayoría de la literatura en este ámbito son: dinamismo (los agentes deberían ser capaces de funcionar independientemente del espacio y el tiempo), adaptabilidad (los agentes aprenden y cambian su conducta basándose en las experiencias previas), continuidad temporal (los agentes no deberían parar o reanudar su actividad para ciertas tareas, más bien su funcionamiento debería ser un proceso continuo) y movilidad (los agentes se pueden transportar de una máquina a otra e, incluso, entre diferentes arquitecturas y plataformas)

2.3. Una propuesta de clasificación para los agentes inteligentes.

Como consecuencia de una definición diáfana de agente inteligente, diferentes autores han propuesto una gran variedad de taxonomías o clasificaciones. A continuación se presenta una doble clasificación que pretende aclarar la tipología de agentes inteligentes basándonos tanto en el ámbito en el que actúan como en las tareas que llevan a cabo, para finalmente demarcar nuestro ámbito de actuación en los agentes de Internet que realizan tareas de recuperación de información.

En cuanto a su ámbito de actuación (Caglayan, 1997):

- Agentes de escritorio (agentes de sistema operativo, agentes de aplicaciones, etc.)
- Agentes de Internet (agentes de búsqueda, filtrado, recuperación de información, agentes de notificación, agentes móviles, etc.)
- Agentes de Intranet (agentes de customización cooperativa, agentes de bases de datos, agentes de automatización de procesos, etc.)

En cuanto a su función (Stenmark, D., 1998-A):

Stenmark clasifica los agentes en las siguientes tipologías: Interface agents, System agents, Advisory agents, Filtering agents, Retrieval agents, Navigation agents, Monitoring agents, Recommender agents, Profiling agents y otros que están surgiendo continuamente.

En el ámbito de la gestión eficiente del conocimiento, podemos destacar tres tipos:

1. **Filtering agents:** agentes que se usan para reducir la sobreabundancia de información mediante el borrado de los datos no deseados (por ejemplo, los datos que no satisfacen completamente el perfil de usuario). Muchos clientes de e-mail, así como los productos Agentware e InfoMagnet proporcionan prestaciones básicas de filtering agents.
2. **Retrieval agents:** agentes que buscan, recuperan y proporcionan la información como si fueran auténticos gestores de información y documentación (“information brokers”). Muchos productos se autoproclaman como retrieval agents, tanto aplicaciones cliente: AT1, BullsEye, Go-Get-It, Got-It, Surfbot y WebCompass; como aplicaciones servidor: Agentware e InfoMagnet.
3. **Monitoring agents:** proporcionan al usuario la información cuando sucede un determinado acontecimiento; por ejemplo cuando la información ha sido actualizada, trasladada de lugar o borrada. Algunos productos ejemplo son: WBI de IBM, BullsEye y SmartBookmarks.

En nuestro caso nos vamos a centrar en los denominados “retrieval agents”, esto es, agentes para la recuperación de información. Otros autores denominan a este tipo de software como agentes de información. Al fin y al cabo su función no se basa en la mera recuperación de información sino que disponen de un conjunto de utilidades conexas que nos permitirían denominarlos agentes para gestión del conocimiento.

3. AGENTES INTELIGENTES PARA GESTIÓN DE INFORMACIÓN.

En la última década los métodos para recuperar y gestionar la información han variado notablemente. A medida que Internet ha ido creciendo se ha complicado la forma de localizar la información disponible en la propia red. La mayoría de herramientas disponibles en la actualidad para la recuperación de la información en Internet se centran más en la cantidad que en la calidad. Resulta evidente, por tanto, la necesidad de una cierta automatización del proceso de búsqueda y que a su vez permita seleccionar el nivel de calidad de la información desde un punto de vista subjetivo (generalmente aprendido del comportamiento del propio usuario). En este aspecto, los agentes inteligentes pueden asistir al usuario en la búsqueda de información útil y relevante, a la vez que recuperar, analizar, manipular e integrar informaciones y datos de carácter heterogéneo.

Para un estudio detallado de las características relevantes de los agentes de información, éstos pueden ser clasificados en relación con su comportamiento, (Klusch, M., 1999, Preface).

En primer lugar, los agentes de información pueden ser cooperativos o no cooperativos, dependiendo de su habilidad para cooperar con otros agentes durante la ejecución de sus tareas.

3.1. La colaboración entre agentes de información.

El principal reto de colaboración entre agentes autónomos y heterogéneos se basa en el entendimiento mutuo. Ya existen marcos básicos en uso que proporcionan un lenguaje interfaz para la definición y localización de objetos, como CORBA, DCOM y, menos extendido, Java RMI.

Los agentes necesitan entenderse mutuamente cuando realizan una tarea común, para ello, se han desarrollado diversas herramientas y lenguajes para la representación y compartición del conocimiento de cada agente. De ellos destacan las ontologías comunes y los lenguajes de comunicación entre agentes (ACL), tales como KQML o FIPA ACL.

De igual forma, para facilitar los servicios de brokering y matchmaking en Internet se hace necesario un lenguaje común para la descripción y el procesamiento automáticos de las capacidades de los agentes. En este sentido se vienen utilizando lenguajes de codificación, como SGML o XML, así como lenguajes específicos para la descripción de las capacidades de los agentes, como LARKS.

La colaboración entre agentes también implica un conjunto de obligaciones sociales derivadas de la delegación de tareas y responsabilidades, el diseño de los roles de agente, el análisis de las relaciones entre roles, los tipos de conversación necesarios para conseguir acuerdos entre agentes, etc., que dado su carácter social, humano y subjetivo son los más difíciles de adquirir por parte de los agentes.

Independientemente de si son cooperativos o no, los agentes de información pueden ser ulteriormente clasificados en tres grandes grupos: adaptativos, móviles y racionales.

3.2. Los agentes de información adaptativos.

Aquellos agentes que son capaces de adaptarse a los cambios en la red. Un ejemplo claro son los asistentes personales en la web.

La adaptación de un agente a su entorno puede llevarse a cabo de una forma aislada o en colaboración con otros agentes, usando, respectivamente, métodos de aprendizaje simple o multi-agente.

Los agentes de información adaptativos simples son aquellos que no colaboran con otros agentes y cuyo principal objetivo es la búsqueda de información relevante a partir de las necesidades de información y preferencias de los usuarios del sistema. Para que la búsqueda sea pertinente existen diferentes técnicas para establecer el perfil y preferencias del usuario, entre otras: observar las actividades online del usuario, búsqueda heurística en web, programación de conceptos y principios de la consulta en bases de datos, la recuperación de información, el “data mining” y el “knowledge discovery”.

Los agentes de información adaptativos multi-agente son los más eficientes para llevar a cabo las tareas de filtrado cooperativo (collaborative filtering). Por ejemplo, un documento o producto anunciado en web es recomendado por un agente si este producto ya ha sido recomendado por otros agentes a usuarios con un perfil similar.

3.3. Los agentes de información móviles.

Se refiere a los agentes de información que son capaces de viajar autónomamente a través de Internet de un sitio a otro para llevar a cabo tareas o consultas en diferentes servidores. Un ejemplo son los agentes que permiten una reducción de la transferencia de datos entre servidores y aplicaciones.

Aunque actualmente existe una discusión constante sobre las características que debe cumplir un agente inteligente para considerarse móvil, ya se puede decir que la movilidad de agentes ha dejado de ser ciencia-ficción. Aunque su uso no se generalizará hasta que existan normas para su desarrollo y la seguridad en Internet deje de ser un impedimento (técnicas criptográficas).

Los problemas de seguridad afectan en un doble sentido. Cómo pueden los servidores estar protegidos ante las acciones maliciosas de los agentes móviles y, a la inversa, cómo puede un agente de información, conteniendo datos de carácter privado, estar protegido de los servidores y otros agentes mientras viaja por el ciberespacio.

La principal aplicación actual de los agentes móviles es en el área de las telecomunicaciones, donde se usan como una parte de la arquitectura descentralizada de la próxima generación de redes, como TINA.

3.4. Los agentes de información racionales.

Esta tipología de agentes tiene un marcado carácter comercial y su principal aplicación es en el ámbito del comercio electrónico. Ejemplos de este tipo de agentes son los ShopBots (robots de compra) o los agentes mediadores en subastas.

Los agentes de información racionales que actúan en una economía cambiante deben estar equipados con métodos sofisticados que les permitan tomar decisiones económicamente racionales.

En la actualidad un agente de información racional y personalizado puede realizar compras para su usuario hasta un límite preautorizado, filtrar información de los vendedores, comerciar con diferentes artículos en un mercado digital e incrementar gradualmente el nivel de fiabilidad en sus acciones.

Para la expansión de este tipo de agentes son necesarios métodos para la detección de fraudes, así como fiabilidad y seguridad para facilitar las transacciones que tienen lugar en el comercio electrónico. Además, es necesario desarrollar métodos que permitan a los agentes reaccionar ante los rápidos cambios que los vendedores realizan en sus productos y servicios.

4. LAS UTILIDADES DE LOS AGENTES DE INFORMACIÓN PARA EL DESARROLLO DE SERVICIOS Y PRODUCTOS DE INFORMACIÓN CIENTÍFICA Y TÉCNICA.

Las aplicaciones de los agentes inteligentes en el ámbito de la recuperación y difusión de la información han venido a solucionar (o, más bien, a mitigar) uno de los mayores problemas de la actualidad: la carencia de herramientas para la efectiva gestión de la información, que ha conducido a lo que se denomina "explosión de la información", "sobreabundancia de información" (information overload) o, más correctamente, "infoxicación" (Cornellà, A. 1998-A).

A continuación se presentan unos ejemplos que pretenden ilustrar la tecnología de agentes inteligentes como una herramienta para explotar de forma eficiente las posibilidades de Internet en la oferta de servicios de información técnica y comercial.

Actualmente, la tecnología de agentes inteligentes puede ser utilizada para la prestación de diversos servicios, por ejemplo se puede utilizar el agente CiteSeer (Giles, C.L., 1998) para establecer redes de citas en un sector específico de conocimiento, conocer los autores que más investigan, los centros de investigación que más actividad presentan, cual es la evolución en una tendencia de investigación, etc. Podemos utilizar BullsEye para la realización de informes de recursos electrónicos, para conocer los webs más representativos en relación con un ámbito de conocimiento hiperespecializado, para establecer que empresas, organismos oficiales o centros de enseñanza están ligados a un tecnología, etc. Además, Webwatch reúne un conjunto de prestaciones para la oferta de servicios de vigilancia tecnológica en un entorno cooperativo.

4.1. Agentes inteligentes para índices de citas: CiteSeer

Las últimas investigaciones (Snyder, H., 1999,) sobre la posibilidad de utilizar motores de búsqueda para el análisis de enlaces (links) y el

establecimiento de una estructura de redes de citas han puesto de manifiesto una serie de problemas.

En el momento del estudio (julio de 1998), sólo dos motores de búsqueda comerciales –Altavista y Hotbot- permitían contar las páginas con enlaces a otras páginas.

De las conclusiones del estudio se puede extraer que la fiabilidad de los resultados obtenidos es muy cuestionable. Los motores de búsqueda no poseen las prestaciones para realizar búsquedas que permitan un análisis de enlaces fiable. Hasta el momento el único análisis de enlaces posible es al nivel de los dominios (edu, com, org, etc.). Por tanto, y debido a la escasez de fiabilidad y utilidad de los motores de búsqueda para el análisis de enlaces, se puede afirmar que este tipo de análisis no es riguroso y no se puede comparar al análisis de citas tradicional.

Como una alternativa a la tentativa anterior, aparece un agente que puede analizar las estructuras de cita de los recursos web y realizar un análisis exhaustivo de las citas.

CiteSeer es un agente inteligente para creación de índices de citas de la literatura científica en soporte electrónico.

Funcionamiento básico del agente:

- Recorre la web y descarga artículos en diferentes formatos que convierte a texto.
- Analiza los artículos para extraer las citas (incluye un algoritmo para identificar los diferentes formatos de cita).
- De cada artículo se extrae la siguiente información: URL del fichero, cabecera, resumen, introducción si existe, citas, contexto de las citas y texto completo.
- Se extrae cada cita y se localiza el título, autor, año, páginas y formato de cita.
- Crea una base de datos con las citas.

Ventajas respecto a los índices tradicionales:

- Incluye los artículos más recientes, desde el momento en el que están disponibles en web.
- Posibilidad de crear bases de datos actualizadas sin limitarse a un conjunto restringido de publicaciones.
- Funciona de forma totalmente autónoma: reduce los costes de indización manual.
- Refleja el impacto de un artículo y su contexto (browsing), esto es, cómo un artículo dado es citado en las publicaciones posteriores.

Inconvenientes respecto a los índices tradicionales:

- No cubre publicaciones significativas (primary journals), sólo las disponibles en web.
- No distingue subcampos de forma precisa (por ejemplo, dos autores con el mismo nombre). Aunque utilizan bases de datos de nombres de autores y revistas (ficheros de autoridades automatizados) para ayudar a identificar los subcampos.

Utilidades en investigación científica y técnica:

- Localización de artículos por palabras clave, por la información contenida en las citas o mediante un vector de similitud de palabras.
- Recuperación de artículos sobre un ámbito específico de conocimiento siguiendo redes de citas.
- Evaluación y clasificación de artículos, autores y revistas según el número de citas recopiladas.
- Identificación de las tendencias en investigación.

4.2. Agentes inteligentes para informes técnicos: BullsEye

BullsEye es una aplicación que permite a los usuarios recuperar, gestionar y permanecer alertado sobre determinada información en Internet. El agente utiliza una colección de motores de búsqueda que recuperan los resultados relevantes del web visible e invisible (puesto que es capaz de lanzar consultas sobre bases de datos).

Posibilidades de búsqueda de información técnica:

- Búsqueda de recursos en web (toda la web o nuevos sitios)
- Búsqueda de noticias
- Búsqueda de negocios (patentes y marcas)

Prestaciones avanzadas:

- Búsqueda booleana avanzada
- Eliminación de links inútiles
- Análisis de las páginas
- Previsualización y marcado de sitios web
- Visualización por categorías (ranking de adecuación, tipo de sitio [com, edu, net, org, Europa, Asia, Norteamérica, etc.], palabras clave, etc.)
- Posibilidad de subscripción a sitios web (Track)
- Posibilidad de generar informes de forma automática.

4.3. Agentes inteligentes para vigilancia tecnológica: Webwatch

La nueva tecnología basada en la utilización de software de agentes inteligentes para recuperación de la información y gestión del conocimiento puede, en el marco de una empresa, ayudar a:

- Crear canales de información especializada con unos niveles de exactitud desconocidos en la industria.
- Establecer nuevos productos y servicios de información una vez han sido claramente identificadas las necesidades del cliente.
- Encontrar información relevante sobre unas líneas de trabajo determinadas sin necesidad de tener que realizar una búsqueda activa.
- Crear una vía de actualización constante sobre las últimas novedades en un determinado ámbito de conocimiento.
- Establecer contacto con otros usuarios que posean perfiles de trabajo e intereses similares.

Las aplicaciones de agentes inteligentes en el ámbito de la recuperación y difusión de la información han venido a solucionar uno de los mayores problemas de la actualidad: la carencia de herramientas para conseguir información filtrada, estructurada y de máximo interés para favorecer los procesos de innovación tecnológica.

Por ello, los agentes inteligentes constituyen la herramienta idónea para la oferta de servicios de vigilancia tecnológica en un futuro próximo. De hecho, se basa en anticipar las innovaciones tecnológicas e informar sobre la situación del mercado de la forma más adecuada posible con el objetivo que las empresas puedan adaptar sus medios, sistemas, materiales y tecnologías a la dinámica evolución del sector.

Este sistema ofrece ventajas con relación a los tradicionales Informes de Vigilancia Tecnológica:

- Tipologías de información: los contenidos abarcan todo tipo de documentos electrónicos (patentes, normas, legislación, artículos técnicos, todo tipo de literatura gris, programas de software, noticias, etc.)
- Medio de difusión: Internet supone el medio ideal para recibir la información de las últimas innovaciones en un ámbito de conocimiento.
- Actualización: la actualización de la información que llega al puesto de trabajo puede ser tan frecuente como la empresa defina (incluso diario).

5. LOS AGENTES INTELIGENTES EN EL MARCO DEL COMERCIO ELECTRÓNICO: UTILIDADES Y SERVICIOS.

En los últimos años el comercio electrónico ha experimentado un gran crecimiento debido a la convergencia de diferentes desarrollos tecnológicos en las telecomunicaciones, la informática y la gestión de información; a lo que ha contribuido un clima económico favorable.

Por otro lado, el ingente crecimiento cuantitativo y cualitativo de Internet y su sucesivo aumento de popularidad constituyen las bases para el desarrollo del comercio electrónico. Un valor añadido que ofrece la red Internet es la

posibilidad de garantizar una representación apropiada de las PYMEs y fomentar su participación en un mercado global (Adam, N.R., 1998, p.2).

En realidad, el comercio electrónico no ha surgido de forma espontánea o como un avance de las tecnologías EDI, sino que responde a una necesidad empresarial de realizar transacciones comerciales de una forma eficiente y a bajo costo.

El éxito del comercio electrónico dependerá, dicen (Cornellà, A., 1998 B), de que se aprovechen al máximo las características diferenciales de esta nueva tecnología, esto es: que aporte soluciones que no son posibles en el mundo real, que se den servicios que el cliente precise, que no pueda vivir sin ellos.

Hasta el momento las diferentes formas de estar presente en la red para realizar actividades de comercio electrónico están íntimamente relacionadas con el éxito o fracaso de los sitios web. Una tipología básica de e-catálogos, esto es, empresas que venden productos en Internet, es la siguiente:

- Stand alone: sitio web independiente utilizado por un comerciante (ej. Amazon.com para libros)
- E-mall: conjunto de catálogos de diferentes comerciantes (ej. Internet Shopping Network).
- Embedded: forma parte de un sitio web corporativo inicialmente utilizado para propósitos no comerciales (marketing, apoyo a clientes, etc.) (ej. Microsoft).

Todo ello supone una sobreabundancia de información y un problema en la localización de la información, productos y servicios que el cliente necesita.

Hoy en día, la búsqueda de un producto concreto en un buscador puede ofrecer varios cientos e incluso miles de resultados, entre los que se pueden encontrar unas pocas tiendas que efectivamente venden este producto y entonces proceder a la comparación de precios, modelos, etc., todo de forma manual.

Para solucionar estos problemas se han desarrollado varios agentes inteligentes para utilidades en el ámbito del comercio electrónico. De ellos destacamos los siguientes:

Buscador de productos: se trata de un agente personal que recoge información sobre el perfil de interés del usuario (profiling) y, utilizando un algoritmo para medir el grado en el que los productos responden a este perfil de usuario (matching), filtra los productos presentes en web y recupera un conjunto ordenado de tales productos. Algunos ejemplos son: NetSage Sage (<http://www.netsage.com>), Extempo Imp (<http://www.extempo.com>) y PersonalLogic (<http://www.personallogic.com>).

Filtrado de información cooperativa (collaborative filtering): se trata de agentes capaces de anticipar las necesidades individuales de un usuario a partir del contexto con otros usuarios. Esto es, el agente ofrece servicios y productos que han sido recomendados a otros usuarios con un perfil similar. Los agentes de información más populares que implementan esta tecnología son, entre otros: Firefly (<http://www.firefly.com>), LikeMinds (<http://www.likeminds.com>) y WiseWire (<http://www.wisewire.com>).

Robots de compra (shopbots): son sistemas que pueden obtener información sobre el precio y/o especificaciones técnicas de un producto en los diferentes sitios web y comparar la oferta de los distintos competidores, normalmente ordenada de acuerdo con un criterio elegido por el usuario. Algunos ejemplos de agentes que permiten la compra comparativa son: CompareNet (<http://www.compare.net>), Jango/Excite (la tienda de Excite), <http://jango.excite.com>, Jungle/Yahoo! (<http://www.jungle.com>), RoboShopper (<http://www.roboshopper.com>) o Fido (<http://www.shopfido.com>).

Mercados basados en agentes (agent-based marketplaces): aplicación de la tecnología de agentes donde múltiples usuarios y proveedores pueden reunirse para negociar e intercambiar información, productos y servicios. Algunos de los mercados basados en agentes y mercados de subastas ya disponibles en web son: Kasbah (<http://kasbah.media.mit.edu>), OnSale Exchange (<http://www.onsale.com/exchange.htm>), y FairMarket (<http://www.fairmarket.com>).

Formación de coaliciones: permite la colaboración entre agentes de información racionales, que pueden negociar y formar coaliciones para incrementar su beneficio individual o el beneficio del sistema en sí mismo. Aunque las técnicas de formación de coaliciones y negociación automática entre agentes son conocidas, todavía no existen sistemas multi-agente que puedan utilizarlas para utilidades de comercio electrónico en web.

De lo expuesto anteriormente, podemos deducir que Internet ha evolucionado de un espacio de información hacia un espacio de mercado, donde existen miles de tiendas electrónicas (storefronts), subastas y diferentes servicios comerciales.

Los primeros shopbots simplificaban la ardua tarea de buscar y comparar productos y precios, pero no ofrecían servicios de valor añadido, tales como la personalización de los servicios. En los últimos tiempos están surgiendo un conjunto de proyectos que pretenden ir más allá de los meros agentes para la búsqueda, recuperación, filtrado y comparación de productos y ofertas en el marco del comercio electrónico, sino que suponen un nuevo mercado basado en agentes.

Anthony Chaves, Pattie Maes y otros, (Chaves, A., 1997), del MIT Media Lab., han desarrollado un mercado basado en agentes denominado Kasbah, en el que los usuarios pueden asignar la tarea de comprar o vender a un

agente, el cual realiza la negociación y el establecimiento de acuerdos de forma automática.

Los proyectos de biblioteca digital de Stanford (<http://www-diglib.stanford.edu>), y la Universidad de Michigan (<http://http2.sils.umich.edu/UMDL/>) utilizan arquitecturas basadas en agentes. Puesto que como afirman Nabil Adam y Yelena Yesha (Adam, N., 1996), la evolución del comercio electrónico y las bibliotecas digitales tienen aspectos en común tales como el requerimiento de arquitecturas robustas y la creación de un mercado abierto para información, bienes y servicios. La tecnología de agentes en el ámbito del comercio electrónico y las bibliotecas digitales es útil para que los proveedores tengan conocimiento de las necesidades de los consumidores y alertar a los consumidores de las ofertas de los proveedores. En ambos sistemas el futuro se fundamenta en el uso de agentes que negocien los términos de una transacción.

En estos momentos no es fácil anticipar cual va a ser el futuro en los próximos años, pero quizá, a partir de lo expuesto, es fácil adivinar que el profesional de la información tendrá que desarrollar su trabajo en contacto con la tecnología de agentes, con el comercio electrónico, con las bibliotecas digitales, con las nuevas formas de ofrecer servicios y productos de información y, todo ello, en un contexto económico globalizado.

Práctica 1. Evaluación de motores de búsqueda

El alumno/a deberá evaluar 5 motores de búsqueda de ámbito regional, nacional o internacional, para ello usará una estrategia de búsqueda que aplicará en los 5 motores preseleccionados, de forma que puedan apreciarse claramente las diferencias entre éstos.

Las características que pueden presentar los diferentes motores de búsqueda se van a agrupar en tres apartados: recogida de la información, búsqueda y recuperación de la información y presentación de los resultados (Maldonado Martínez, 1998, p. 529-551).

Recogida de información

En este apartado hay que determinar si el robot es capaz de identificar las etiquetas "META" de los documentos HTML y extraer información de las mismas para ser usada en la búsqueda o presentación de resultados.

Búsqueda

Las características básicas que un motor de búsqueda debe cumplir desde el punto de vista de la recuperación de la información son las siguientes:

- Formularios de búsqueda: posibilidad de elegir entre simple o avanzado.
- Herramientas de búsqueda: posibilidad de utilizar operadores booleanos (AND, OR, NOT), paréntesis, comillas para los términos compuestos o frases y, finalmente, posibilidad de truncado en palabras derivadas.
- Clasificación temática: existencia de un índice general para aquellos que no saben concretar su tema de búsqueda.
- Campos de búsqueda: posibilidad de limitar la búsqueda a campos determinados, tales como: Título, URL, Descripción, Palabras Clave, Localización e Idioma.
- Control del vocabulario: descubrir si el motor dispone de alguna herramienta para eliminar sinonimias y polisemias.
- Detección de novedades: posibilidad de diferenciar los nuevos recursos incorporados.

Resultados

Hay que tener en cuenta si el motor de búsqueda permite elegir entre diferentes formatos de presentación de los resultados así como diversos criterios de ordenación.

Una vez se han analizado las características y se ha cumplimentado el formulario de la página siguiente, se realizará un informe (con gráficos de los datos más representativos) donde se expondrán los resultados derivados del análisis de los datos.

CUADRO PARA LA EVALUACIÓN DE MOTORES DE BÚSQUEDA

MOTORES	Recogida de información		Búsqueda y Recuperación de Información																Resultados			
	Metadata	No Metadata	?	Formularios		Herramientas de búsqueda						Clasif.	Campos de búsqueda						Control Vocab.	Nov.	Format.	Orden
				Simple	Avanz.	OR	AND	NOT	()	“ “	*		Tít.	URL	Desc.	Keyw.	Loc.	Leng.				

Práctica 2. Creación de informes técnicos con BullsEye (Intelliseek)

Para realizar esta práctica es necesario descargar (download) el agente de información BullsEye distribuido por Intelliseek, para ello, hay que conectar con el sitio web <http://www.intelliseek.com> y descargar la demo del programa que tiene una duración de 30 días.

Una vez instalado se diseñará una estrategia de búsqueda hiperespecializada, algunos ejemplos:

1. Uso de plásticos reciclados para la alimentación. Posibles descriptores: plastic*, polymer*, recicl*, "food contact", food, beverage*
2. Sistemas de seguridad en el comercio electrónico. Posibles descriptores: "electronic commerce", securit*, device*, system*, retailer* storefront*
3. Envases y embalajes biodegradables. Posibles descriptores: pack*, bio* biodegrad*, environment*

Estudiar las posibles combinaciones de descriptores para desarrollar una estrategia de búsqueda coherente.

Elegir una de las opciones de consulta (Web, News, Business, etc.) y lanzar la búsqueda.

Revisar los resultados de búsqueda, marcar los más interesantes para la posterior realización del informe.

Entregar un informe de la situación actual en el campo de conocimiento seleccionado a partir de los diferentes recursos de información recuperados. Al informe global ("estado de la técnica") habrá que anexar, entre otros, los siguientes informes (reports) generados automáticamente por BullsEye:

- Informe de los recursos web de interés.
- Informe de las últimas noticias publicadas en este ámbito del conocimiento.
- Informe de las patentes publicadas sobre el sector científico y/o técnico estudiado.

El objetivo de esta práctica ha sido que el alumno/a realice un informe sobre el estado de la técnica basado en la información de recursos web.

El alumno podrá realizar la práctica con otro agente de información semejante. En las siguientes direcciones se pueden encontrar ejemplos y demos.

<http://www.csee.umbc.edu/agents/related/ir/>
<http://w3.informatik.gu.se/~dixi/agent/product.htm>
<http://www.botspots.com>

Bibliografía

ADAM, N. R... et al. *Electronic Commerce: technical , business and legal issues*. New Jersey: Prentice Hall PTR, 1999.

ADAM, N. Y YESHA, Y. Electronic Commerce and Digital Libraries: towards a digital agora. *ACM Computing Survey*, vol. 4, nº 28, diciembre de 1996.

AGUILLO, I. F. Del multibuscador al metabuscador: las agentes trazadores de Internet. En: *Congreso ISKO (IV. Granada. 1999)*. La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de información. Granada: Isko: Universidad de Granada, 1999, p.239-245

BARÓ I QUERALT, J. Cercar i col.locar informació en el World Wide Web. *Quaderns de Comunicació*, 7, 1998.

BRASHAW, J. An introduction to software agents. En: Brashaw, J. *Software agents*, AAAI Press, 1997, p. 4-7

BullsEye.Com. Toda la información disponible sobre este agente inteligente para la recuperación de información (retrieval agent) puede ser localizada en <http://www.intelliseek.com/be/bullseye.htm> y <http://www.bullseye.com> (consultados en abril de 1999).

CAGLAYAN, A.; HARRISON, C. *Agent Sourcebook*. New York, etc.: Jonh Wiley & Sons, 1997.

CHAVES, A... et al. A Real-Life Experiment in creating an agent marketplace. En: *Proceeding of PAAM'97*. Practical Applications Company, 1997.

CODINA, L. Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información. Part I. *Information World en Español*, vol. 6, nº 5, 1997, p. 22-26

CORNELLÀ, A. Afectará la crisis económica al comercio electrónico?. *Extra!Net. Revista de Infonomía: La información en las organizaciones*. Mensaje 378. Recibido en IWETEL el 5/11/1998 y La Red es, o será, el mercado, según y cómo... <http://www.extra-net.net> octubre de 1998. (A)

CORNELLÀ, A. Nuevos instrumentos contra la infoxicación. *Extra!Net. Revista de Infonomía: La información en las organizaciones*. Recibido en IWETEL el 3/11/1998. (B)

DOI Workshop. Telematics for Libraries. *European Commission*, DG XIII-E4, August 1998.

ERIKSSON, J.; FINNE, N. y JANSON, S. To each and everyone an agent: augmenting web-based commerce with agents. *Intelligent Systems Laboratory. Final Report*, 1999.

GILES, C.L.; BOLLACKER, K.D.; LAWRENCE, S. CiteSeer: an automatic citation indexing system. In: *Digital Libraries '98 – The Third ACM Conference On Digital Libraries*, ACM Press, June, 1998. p. 89-98

GILES, C.L.; BOLLACKER, K.D.; LAWRENCE, S. CITESEER: an autonomus web agent for automatic retrieval and identification of interesting publications. In: *2nd International ACM Conference on Autonomus Agents*, ACM Press, May, 1998. p. 116-123.

HÍPOLA, P.; VARGAS-QUESADA, B. Agentes inteligentes: definición u tipología. Los agentes de información. *El profesional de la información*, vol. 8, nº 4, abril de 1999, p. 13-21

KLUSCH, M. (Ed.). *Intelligent information agents: agent-based information discovery and management on the Internet*. Berlin: Springer, 1999.

KOSMYNIN, A. From bookmark managers to distributed indexing: an evolutionary way to the next generation of search engines. *IEEE Communications Magazine*, junio 1997, p. 146-151

LELOUP, C. *Motores de búsqueda e indexación: entornos cliente servidor, Internet e Intranet*. Barcelona: Gestión 2000, 1998.

MAES, P. Agents that reduce work and information overload. *Communications of the ACM*, vol. 7, nº 37, 1994, p. 31-44. Disponible en <http://pattie.www.media.mit.edu/people/pattie> (visitado en mayo de 1999).

MALDONADO MARTÍNEZ, A.; FERNÁNDEZ SÁNCHEZ, E. Evaluación de los principales “buscadores” desde un punto de vista documental: recogida, análisis y recuperación de recursos de información. *Actas VI Jornadas Españolas de Documentación*, 1998, p. 529-551

MATTHEW, C. Bridging intranet profit and value. *Datamation*, diciembre/enero 1998, p. 120-124

METADATA Workshop. Telematics for Libraries. *European Commission*, DG XIII-E4, August 1998.

NWANA, J. Software agents: an overview. *Knowledge Engineering Review*, 11(3), 1996, p. 205-244

SANTAMARIA, F. XML (eXtensible Markup Language): nuevo estándar para la descripción de documentos en la World Wide Web. *Actas VI Jornadas Españolas de Documentación*, 1998, p. 819-824

SNYDER, H.; ROSENBAUM, H. Can search engines be used as tools for web-link analysis? A critical review. *Journal of Documentation*, vol. 55, nº 4, 1999, p. 375-384

STENMARK, D. *Intelligent Software Agents: a attempt to do a classification*, 1998 . <http://w3.informatik.gu.se/~dixi/agent/agent.htm> (visitado en enero de 1999). (A)

STENMARK, D. *Intelligent Software Agents: a study of what is available*, 1998. <http://w3.informatik.gu.se/~dixi/agent/agent.htm> (visitado en enero de 1999). (B)

TRAMULLAS, J. Agentes y ontologías para el tratamiento de la información: clasificación y recuperación en Internet. En: *Congreso ISKO (IV. Granada. 1999)*. La representación y la organización del conocimiento en sus distintas perspectivas: su influencia en la recuperación de información. Granada: Isko: Universidad de Granada, 1999, p. 247-248

VAQUERO, J.R. Motores de búsqueda. *Information World en Español*, vol. 6, nº 7-8, 1997, p. 31-32

Web Watch Revolution. <http://webwatch.krdl.org.sg/technology.html> (consultado en septiembre de 1999).

Preguntas de autoevaluación (V/F)

Un directorio, un índice temático y un motor de búsqueda tienen un funcionamiento similar.

Los metabuscadores pueden considerarse un tipo especial de agente inteligente.

El desbordamiento cognitivo es característico de la indexación realizada por los motores de búsqueda.

Los motores de búsqueda son aplicables a la gestión de contenidos en una intranet corporativa.

Dublin Core es un proyecto para la aplicación de los metadatos a la descripción de recursos web.

La autonomía es una característica de los agentes inteligentes.

Los agentes de Internet y los agentes de información son la misma cosa.

KQML es un lenguaje de comunicación entre agentes.

Los agentes de información adaptativos sólo pueden tener un aprendizaje multi-agente.

La principal aplicación de los agentes móviles es en el área del control del tráfico aéreo.

BullsEye es un agente de información para la oferta de servicios de vigilancia tecnológica.

CiteSeer permite la identificación de las tendencias en investigación mediante el análisis de citas.

Los ShopBots sirven para comparar las ofertas de distintos proveedores.

Kasbah es el nombre de la biblioteca digital del MIT Media Lab.

Las bibliotecas digitales y el comercio electrónico tienen aspectos en común.