

Agentes de información

Alfonso López Yepes
Rodrigo Sánchez Jiménez
José Ramón Pérez Agüera *

Artículo recibido:
26 de noviembre de 2004.

Artículo aceptado:
12 de enero de 2005.

RESUMEN:

Este artículo realiza un repaso sobre las tipologías de agentes de información y describe aspectos como movilidad, racionalidad y adaptatividad, y el ajuste final de estos conceptos a entornos distribuidos como Internet, donde este tipo de agentes tienen un amplio grado de aplicación. Asimismo, se propone una arquitectura de agentes para un sistema multiagente de recuperación de información donde se aplica un paradigma documental basado en el concepto de ciclo documental.

Palabras Clave: Agentes; Recuperación de información; Ingeniería del software.

* Los tres autores son integrantes del Departamento de Biblioteconomía y Documentación de la Universidad Complutense de Madrid, España.
(Alfonso: alopez@ccinf.ucm.es); (José: jose.aguera@ccinf.ucm.es);
(Rodrigo: rsanchezj@ccinf.ucm.es).

ABSTRACT

Information Agents

*Alfonso López-Yepes; Rodrigo Sánchez-Jiménez
and José Ramón Pérez-Agüera*

This article summarizes the main information agent types reflecting on issues such as mobility, rationality, adaptability and the final adjustment of these concepts to distributed environments such as the Internet, where this kind of agents has wide range application. Likewise, an information agent architecture is proposed to create a multi-agent information retrieval system in which a documentary paradigm based on the documentary cycle is developed.

Keywords: Agents; Information retrieval; Software engineering.

INTRODUCCIÓN

Las primeras aplicaciones de agentes de información se encuentran en sistemas federados de bases de datos en los cuales los agentes de información servían como elementos de comunicación que ponían a disposición de los usuarios los recursos federados del sistema de forma unificada. Pero es sobre todo a partir de la aparición de Internet cuando nos encontramos con un uso masivo de este tipo de software en entornos de recuperación y gestión de la información.

La razón fundamental por la cual Internet supone un cambio tan radical en el uso de las tecnologías basadas en agentes reside en que al tratarse de un entorno de red, las técnicas de computación distribuida son, en muchas ocasiones, más que necesarias. Por otro lado, el aumento exponencial de la cantidad de información que contiene la Red ha hecho patente la necesidad de software cada vez más inteligente que pueda afrontar con garantías la gestión y recuperación de la información contenida en Internet. El paradigma de agentes se adapta perfectamente a estas necesidades, y es por esta razón por la cual se ha producido un enorme auge de este tipo de software en los últimos años.

Los denominados agentes de información se han ido definiendo en función de su uso a través de Internet, y es destacable que ya desde las primeras

referencias bibliográficas con las que hemos trabajado¹ se defina a Internet como entorno de trabajo ideal para el cual desarrollar agentes de información, además de un entorno de experimentación sin precedentes.

Vemos de esta forma que el desarrollo de las investigaciones sobre agentes de información está indisolublemente unido al desarrollo de Internet como sistema de información. Es más, se podría decir que, tras una primera observación, la presencia de agentes de información en Internet ha aumentado en los últimos años y pensamos que esta presencia va a seguir aumentando próximamente debido a que la automatización de muchas de las tareas que se realizan hoy en día en Internet pasa por la implementación de agentes o sistemas de agentes, debido a la capacidad de adaptación de este tipo de software a las características de Internet, ya que cuentan con bastante autonomía y permiten un alto grado de personalización.

LOS AGENTES DE INFORMACIÓN

Los agentes de información son agentes software que tienen acceso a múltiples fuentes de información heterogéneas geográficamente distribuidas. Como ya hemos mencionado, el auge de Internet ha supuesto la proliferación de este tipo de agentes, ya que mediante su uso se intentan resolver los problemas asociados al manejo de información distribuida tal y como se presenta en la Red.

Este tipo de agentes, como veremos a continuación, tiene funciones muy variadas, ya que pueden asistir al usuario en la búsqueda y filtrado de información relevante, informar cuándo nuevos datos de interés están disponibles, negociar la compra o venta de productos, participar en subastas electrónicas, etcétera. En cada uno de estos casos los agentes de información ayudan al usuario en la ejecución de tareas, y para llevar a cabo este objetivo tienen que ser capaces de capturar y almacenar las preferencias del usuario. Además, deben ser capaces de actuar adecuadamente ante nuevas situaciones no previstas; es decir, como agentes software deben tener capacidad de aprendizaje y autonomía, siendo ésta una de las características más difíciles de conseguir debido al carácter heterogéneo tanto de los usuarios como de las fuentes de información a las que acceden.

Los agentes de información pueden llevar a cabo sus tareas de manera independiente o trabajar en coordinación con otros agentes. Por esa razón, y en función de la habilidad para cooperar con otros agentes en la ejecución

1 C. A. Knoblock; Arens, Y y Hsu, C., "Cooperating Agents for Information Retrieval", en *Proceedings of the Second International Conference on Cooperative Information Systems*, University of Toronto Press, Toronto, Ontario, Canada, 1994

de tareas, los agentes de información se pueden clasificar en agentes no cooperativos, que se corresponden con agentes individuales, y agentes cooperativos, que se corresponden con sistemas multiagente. A su vez, tanto los agentes cooperativos como los no cooperativos se pueden subdividir en tres tipos fundamentales:

- Agentes de información racionales, que son agentes que se utilizan en el comercio electrónico y que median por su usuario en compras o subastas. Por ejemplo, agentes que buscan el mejor precio de un producto (*shopbots*), agentes que participan automáticamente en mercados y subastas, etcétera.
- Agentes de información adaptativos que son capaces de adaptarse por sí mismos a cambios en su entorno. En el desarrollo de este tipo de agentes hay que contar con dos aspectos importantes. El primero es que en Internet los agentes tienen que construirse para tratar de una manera fiable y segura con información incierta e incompleta. El segundo es la personalización, ya que interesan sistemas que se adapten a cada usuario y aprendan de su comportamiento.
- Agentes de información móviles que son capaces de viajar autónomamente a través de Internet de un sitio a otro para la ejecución de sus tareas en diferentes servidores (p.e. obtener información)

Pese a que aquí realizamos una categorización muy definida de cada uno de los tipos de agentes de información que se pueden encontrar es necesario resaltar el hecho de que en muchos casos encontraremos agentes que implementan características de diferentes modelos, siendo bastante difícil su clasificación unívoca en cualquiera de los tipos que hemos definido aquí.

Agentes de información no cooperativos

Como ya hemos visto se trata de agentes individuales que dan servicio a un usuario concreto cuyos hábitos son aprendidos por el agente con el objetivo de ajustarse a sus necesidades de información. Este tipo de agentes ha proliferado mucho en los últimos años debido a que suelen ser la base de agentes de interfaz destinados a asistir de forma personalizada a usuarios concretos en tareas de recuperación de información. Pese a esto, las tareas que pueden desempeñar este tipo de agentes son muy variadas, a continuación podemos ver algunos ejemplos:

- *Búsqueda*: ayudan al usuario en la tarea de recuperar información sobre la web.

–Bullseye: búsqueda sobre más de 800 buscadores. Permite almacenar búsquedas, añadir comentarios a las páginas, etcétera.

–CiteSeer: especializado en documentos científicos (publicaciones de investigación). Citas a cada artículo, documentos similares, bibliografía, usuarios que han visto el artículo y también han visto otros artículos similares, etcétera.

- *Monitorización*: vigilan cambios en una página indicada por el usuario, aparición de páginas en buscadores, etcétera. Se comunica con el usuario a través de e-mail.
 - TracerLock
- *Filtrado*: seleccionan información en función de las preferencias del usuario.
 - BotBox
- *Navegación*: agentes de interfaz que ayudan al usuario a navegar por la Web.
 - Leticia
- *Comercio Electrónico*: recomiendan productos, comparan precios, etcétera.
 - MySimon

Últimamente se está produciendo un auge de este tipo de agentes en Internet debido a su aplicación a Servicios Web.²

El objetivo es conseguir un acceso a la información en Internet más flexible y personalizado, que responda no sólo a necesidades de información, sino a necesidades relacionadas con las acciones que se realizan con la información recuperada.

Se abre aquí todo un mundo de servicios, íntimamente relacionados con el desarrollo de la Web Semántica y muy a tener en cuenta, ya que en los próximos años viviremos un auge espectacular de este tipo de servicios, los cuales estarán centrados en agentes que sirvan como asistentes para realizar este tipo de tareas.

La complejidad de las tareas que debe realizar un agente es cada vez mayor. Esto está provocando que los diseñadores de agentes software tiendan cada vez más a desarrollar sistemas multiagente donde los agentes colaboren en la resolución de problemas de forma distribuida.

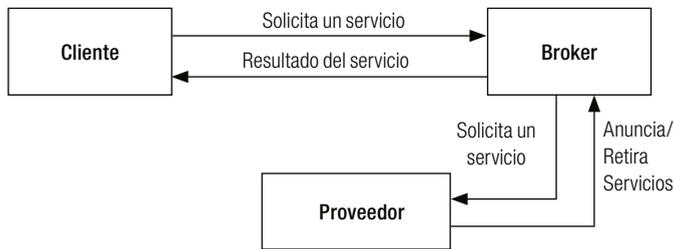
Pese a esta tendencia hacia los sistemas multiagente, los agentes de usuario no cooperativos seguirán sin duda estando muy presentes en los próximos años.

2 A. Craig y C.A. Knoblock. "Deploying information agents on the web", en *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-2003)*, Acapulco, Mexico.

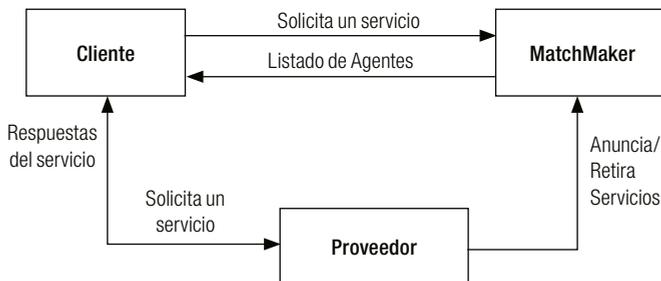
Agentes de información cooperativos

Frente a la necesidad de resolver problemas complejos es una buena solución que los agentes puedan cooperar entre sí haciéndose cargo cada uno de una parte del problema. De aquí deriva la idea de los sistemas multiagente que ya hemos mencionado anteriormente. En la definición dada para los sistemas multiagente dentro del campo de los agentes de información, se deben cuidar muy especialmente dos aspectos clave. Por un lado el mecanismo de intermediación entre los agentes, y por otro la forma de resolver las heterogeneidades de la información.

Existen dos modelos básicos de intermediación entre agentes de información, el modelo *service brokering* y el modelo *matchmaking*. En el modelo de *service brokering* no hay comunicación directa entre el proveedor y el solicitante, el broker contacta con el proveedor con el objetivo de negociar, controlar la transacción y devolverle los resultados al solicitante.³



Por el contrario, en el modelo de *matchmaking* el resultado es una lista de proveedores que pueden proporcionar el servicio. En este caso es el propio solicitante el encargado de contactar y negociar con el proveedor del servicio.



3 A. López-López y L. Sandoval Tafolla. "Un Sistema Multiagente para Recuperación de Información Distribuida", en *Proceedings National Meetings of Computer Science ENC 99, Workshop of Distributed and Parallel Systems*, pp. 496-501, Pachuca Hgo. México, Septiembre 1999.

Como se puede ver por las figuras existen varios tipos de agentes involucrados en este proceso y que pueden ser tipificados en función de estos modelos.

Los tipos de agentes que podemos encontrar según estos modelos de interacción entre agentes de información son:

- Agentes proveedores (*provider*), que constituyen la base de la cadena de "consumo" de información y servicios. Son agentes productores que proporcionan capacidades, como por ejemplo servicios de búsqueda de información o venta de productos para sus usuarios y a otros agentes.
- Agentes solicitantes (*requester*), consumen información y servicios ofrecidos por agentes proveedores en el sistema. Serían equivalentes a los consumidores en el mundo real. Estos agentes pueden tener un doble papel, ya que pueden preguntarle a un agente intermediario cuál de los posibles proveedores puede llevar a cabo un determinado servicio, o pedir directamente la intermediación del mediador para la realización del servicio.
- Agentes intermediarios (*middle*) cuya misión es mediar para que pueda tener lugar una correcta comunicación entre solicitantes y proveedores. Para que esto sea posible los proveedores tienen que registrar sus capacidades ante uno o varios agentes mediadores.

Adaptación, racionalidad y movilidad

Como ya se ha señalado en la introducción de esta sección, independientemente de su grado de cooperación, podemos diferenciar entre agentes de información adaptativos, racionales y móviles. Aunque estas características se pueden encontrar de forma conjunta también pueden servirnos para diferenciar entre las distintas funcionalidades de un agente de información, ya que suelen ir relacionadas con las tareas desempeñadas por el agente.

Adaptación a los cambios en el entorno

El carácter abierto de los principales entornos de información, como Internet, provoca que los cambios se produzcan con gran rapidez y la información contenida en el entorno se amplía y modifica a una velocidad que hace difícil conservar información actualizada sobre el estado del propio entorno. Por esta razón los agentes deben tratar con la incertidumbre y con información incompleta y vaga, lo que hace necesario dotarlos de capacidad de adaptación para que puedan tomar sus decisiones de forma inteligente, rápidamente y en función de unas condiciones difícilmente predecibles.

A los cambios en el entorno hemos de añadir la necesidad de que los agentes se adapten a los usuarios que los utilizan, aprendiendo de sus particularidades y configurando con el tiempo un modelo de usuario que le permita solucionar las necesidades de información de la persona que usa el agente. Vemos, pues, que la adaptabilidad de los agentes de información viene definida por dos elementos: el entorno donde se encuentra la información y el usuario que utiliza el sistema para recuperar esa información.

Una posible solución a este problema basada en agentes simples reside en la creación de equipos de agentes que colaboren en un entorno abierto, y que luego todos éstos aporten la información necesaria para realizar una adaptación gradual a los cambios producidos en el entorno. Entre estos agentes que colaboran para mantenerse informados sobre el estado del entorno para poder luego adaptarse a él, encontramos desde agentes personales de asistencia para recuperar información, hasta arañas encargadas de recorrer la Web en busca de nueva información para incluirla en sus bases de datos. Asimismo la colaboración entre este tipo de agentes permite definir perfiles de usuario, poniendo en contacto a individuos con intereses comunes y aprendiendo de estos intereses para definir de la forma más precisa posible las necesidades de información de los usuarios.

El abanico que se abre entre los agentes personales y las arañas es muy amplio, y se amplía cada día más con el desarrollo de Servicios Web cada vez más complejos. El camino hacia una web personalizada pasa por el mejoramiento de los mecanismos de adaptación de los agentes de información, y esta capacidad de adaptación es uno de los principales campos de investigación en el área de agentes de información.

Agentes de información racionales

Entre las distintas funcionalidades que se les están asignando a los agentes de información también encontramos cada vez más a menudo agentes que no sólo reúnen información, sino que también ayudan a tomar decisiones y aprenden con el tiempo a poder tomar mejores decisiones. Por esta razón la utilización de agentes de información en el área del comercio electrónico es otra de las vías de extensión de este tipo de tecnologías. La idea fundamental radica en dotar a este tipo de agentes con un grado de autonomía cada vez mayor para que realicen actividades de tipo económico en nombre de sus propietarios. Este tipo de agentes requieren también un alto grado de adaptabilidad, pero sin duda su punto fuerte reside en su capacidad para tomar las decisiones correctas en entornos donde tienen que manejar una gran cantidad de datos de tipo concreto, y actuar de forma correcta en función de esos datos.

Los principales componentes para diseñar agentes de información racionales son la gestión dinámica de cadenas de suministro, la negociación descentralizada y la contratación multilateral flexible, lo que permite la comunicación simultánea con varios agentes en tiempo real. En los últimos años se vienen realizando numerosos esfuerzos en este sentido, de cara al desarrollo de métodos que permitan la formación de coaliciones entre agentes autónomos, agentes mediadores de subastas y agentes basados en espacios de mercado.

Los espacios de mercado en Internet del tipo tiendas virtuales como Amazon, y las subastas tipo E-bay, están proporcionando el entorno indispensable para desarrollar este tipo de sistemas, y facilitar el intercambio de datos relevantes de utilidad para los clientes que más tarde se dirigen a estos sitios.

La seguridad es otro de los aspectos importantes en este tipo de agentes, ya que se tiene que prevenir la posibilidad de suplantación de los clientes, o la estafa mediante diferentes métodos de engaño. El desarrollo de protocolos criptográficos y su integración en la lógica de estos agentes racionales es una de las áreas de trabajo fundamentales para la investigación y el desarrollo de este tipo de agentes. Por esta razón la mayoría de los investigadores y desarrolladores de este tipo de sistemas insisten en la importancia de establecer una serie de estándares para protocolos que garanticen la seguridad y privacidad de las transacciones.

Movilidad de los agentes de información

La movilidad de los agentes en un sistema distribuido como Internet tiene muchas ventajas que van desde el ahorro de ancho de banda hasta el aprovechamiento de potencia de cálculo de forma distribuida. Su gran problema es la seguridad ya que se debe permitir la ejecución de código proveniente de otras máquinas, con los problemas de desconfianza que esto supone.

La utilización de agentes de información móviles ya se ha aplicado de forma más o menos experimental en entornos de intercambio de información. Destaca en este terreno Gossip, un sistema que utilizaba agentes móviles para la búsqueda y el intercambio de archivos. Aunque la idea de Gossip no llegó a la etapa de producción, sí ha inspirado otras iniciativas similares adaptadas a redes p2p. La más significativa hasta el momento es el proyecto Galileo⁴ del cliente de Bittorrent⁵ Azureus, donde se está intentando desarrollar un sistema de agentes móviles que permitan la búsqueda de ficheros intercambiables entre los distintos clientes de la plataforma.

4 <http://sourceforge.net/projects/sondagalileo/>

5 Bittorrent es un protocolo de transmisión de ficheros en redes p2p que está teniendo una gran aceptación.

Si bien aún es pronto para hablar de un uso extendido de agentes de información móviles, al menos en Internet, sí es necesario tenerlos en cuenta en el cercano futuro.

AGENTES INTELIGENTES Y RECUPERACIÓN DE INFORMACIÓN TRADICIONAL

Para implementar muchos de los tipos de agentes aquí descritos se utilizan las técnicas básicas de la Recuperación de Información (*Information Retrieval*). Ahora bien, existen varias excepciones relacionadas con el trabajo de los agentes de información, que necesitan una readaptación de las metodologías tradicionales aportadas por la RI, incluso en algunos casos esto ha provocado el replanteamiento de algunos de los enfoques clásicos de la RI.

La RI ha trabajado tradicionalmente sobre colecciones de datos estáticas, concentradas en un solo lugar y organizadas en registros de acuerdo con alguna forma de esquema o patrón. En Internet los datos se encuentran enlazados a través de enlaces hipertextuales, están distribuidos geográficamente, aparecen bastante desestructurados y pueden contener información no textual. Estas diferencias con las fuentes tradicionales con las que trabajaba la RI tradicional han supuesto planteamientos nuevos adaptados al nuevo entorno, entre los que sin duda destaca el concepto de conectividad, donde se analizan los patrones de enlace para analizar la estructura de la Web y posibilitar así una recuperación efectiva y eficiente de la información que ésta contiene. A pesar de esto todavía se usan muchas técnicas de RI clásicas en sistemas que cuentan con agentes de recuperación de información.⁶

Por otro lado también es importante destacar que las tareas de los agentes de información están empezando a diferir bastante del concepto clásico de RI, ya que dentro de este campo se asume un paradigma conversacional donde en cada interacción el usuario emite una consulta y el sistema elige una respuesta adecuada extraída del espacio de información, la cual devuelve como respuesta. Una vez hecho esto el usuario recibe esta respuesta y puede elegir entre reformularla o hacer una nueva. Este paradigma no se puede aplicar directamente a los agentes de información, ya que los usuarios pueden no saber cómo expresar de forma precisa la manera en la que se debe realizar una consulta, de tal forma que son los agentes de información los encargados de indicarle al usuario qué tipos de consulta son los adecuados. De hecho el esquema conversacional puede incluso no aplicarse al estar continuamente

6 TF-IDF Term Frequency-Inverse Document Frequency es un conocido algoritmo para la asignación de pesos a palabras extraídas de un documento, que sirve como técnica básica de indexación automática.

activos los agentes de información, de forma que podrían intentar acceder a la información de forma proactiva; es decir sin la interacción del usuario. De esta forma un agente puede estar operando durante un largo espacio de tiempo con una consulta, refinándola por sí mismo y lanzándola contra distintos sistemas para comparar resultados de forma autónoma y devolver aquellos que crea más apropiados. Esto evidencia la capacidad de aprendizaje de los agentes y da como resultado que la misma consulta en momentos distintos o para usuarios distintos, pueda tener resultados diferentes.

Por estas razones las nociones técnicas que aparecen en muchos textos clásicos de recuperación de información están siendo replanteadas. Algunas técnicas, como el TF-IDF, dependen de la definición de una colección finita de datos, lo que supone un modelo muy próximo al existente en las bases de datos tradicionales. Esto en la Web es impracticable desde el momento en que no tenemos una colección definida de datos sobre la cual trabajar de forma controlada.

Las medidas tradicionales de la RI, precisión y cobertura, también se ven afectadas por estos cambios, ya que ambas medidas pierden mucho de su sentido si tenemos en cuenta que en Internet la colección sobre la que se obtienen los datos es abierta y difícilmente mensurable en términos absolutos.

Por lo que hemos podido ver en este breve repaso Internet se adapta muy bien a modelos de computación distribuida como los que ofrecen los agentes, y es aquí donde éstos tienen su mayor potencial tanto en el desarrollo de nuevos sistemas de agentes como en la evolución de una disciplina clásica como es la RI.

PROPUESTA DE UNA ARQUITECTURA DE SISTEMA MULTIAGENTE (ASM)⁷ DE RECUPERACIÓN DE INFORMACIÓN

Una vez que hemos visto los enfoques más desarrollados en el área de agentes de información vamos a proponer una posible arquitectura de un sistema multiagente de recuperación de información que difiere de las que hemos visto hasta el momento, y que puede resultar interesante frente a la adaptación de la RI tradicional al campo de los sistemas multiagente.

La idea fundamental se basa en reconocer los roles de los distintos actores implicados en el proceso de recuperación de información. Una vez reconocidos estos roles podemos asignarles las tareas correspondientes a cada uno de ellos a los agentes, de tal forma que podamos automatizar el proceso y resolver el problema de la recuperación de información de forma distribuida.

El proceso de recuperación de información queda definido a partir de una serie de etapas que comienzan con la definición de políticas de adquisición de documentos, y finalizan con la recuperación por parte de un usuario de la información contenida en el sistema. A estas etapas se les debe añadir un proceso de retroalimentación en el que la descripción de los documentos se va refinando en función de la capacidad de recuperación que ofrecen sus descriptores y los lenguajes documentales designados para normalizarlos.

A partir de aquí podemos diferenciar entre cuatro etapas diferentes dentro del proceso de recuperación de información:

- La adquisición de los documentos que pasan a formar parte de la colección, para lo cual es necesario definir las políticas de adquisición documental.
- La indización de los documentos seleccionados en la fase de adquisición, mediante la asignación de descriptores normalizados a partir de lenguajes documentales diseñados para esta tarea.
- La clasificación de los documentos a partir de una estructura temática jerarquizada que describa los temas tratados en la colección y permita diferenciar entre estos temas.
- La consulta de usuarios, donde el usuario final realiza una consulta al sistema para recuperar documentos que se adapten a sus necesidades de información.

A estas cuatro etapas hemos de añadir la gestión y actualización de los lenguajes documentales que se lleva a cabo mediante el proceso de retroalimentación realizado a partir de las consultas efectuadas por los usuarios finales.

ROLES DE LOS AGENTES INVOLUCRADOS EN EL SISTEMA

A partir de las etapas descritas anteriormente podemos definir una serie de roles encargados de realizar las tareas relacionadas con cada una de las etapas citadas. De esta forma tendríamos los siguientes roles asignables a agentes:

- Agente encargado de adquirir y seleccionar los documentos que pasan a formar parte del sistema.
- Agente encargado de indizar los documentos que entran a formar parte del sistema.
- Agente encargado de clasificar los documentos que pasan a formar parte del sistema.

- Agente encargado de facilitarle al usuario final la recuperación de los documentos contenidos en el sistema.
- Agente encargado de normalizar los conceptos y de gestionar y actualizar los lenguajes documentales utilizados en el sistema.

A continuación haremos una breve descripción de las tipologías a las que pertenecerían estos agentes.

Agente de adquisición

Este agente se correspondería con el rol encargado de incorporar documentos a la colección que contiene el sistema en función de unos parámetros definidos a través de políticas de adquisición de documentos. Estas políticas se definen siempre en función de los intereses de los usuarios del sistema, lo cual quiere decir que no son estáticas, ya que se modifican en función de la interacción que tienen los usuarios con el sistema y el grado de satisfacción que se observa por parte de éstos, definido mediante el grado de acceso de los usuarios a los documentos que recuperan. Este aspecto toma una gran importancia, sobre todo en colecciones referentes a dominios específicos donde existe una masa crítica de usuarios especializados con necesidades muy concretas de información, pero es aplicable a todo tipo de sistemas de RI ya que es una de las bases de su éxito.

En el caso de nuestro agente de adquisición de documentos aplicado a un entorno como Internet, donde la incorporación de nuevos documentos a la colección puede hacerse de forma automatizada, el diseño del agente se corresponde muy estrechamente con el de los *crawlers* o arañas web que utilizan los buscadores para recorrer la web en busca de nuevos documentos.

De forma concreta contaríamos con un agente encargado de crear y gestionar pequeños *crawlers* que son enviados a través de la WWW en busca de un tipo específico de documentos.

Las políticas de adquisición se pueden definir previamente de forma manual por parte de los administradores del sistema y a partir de ahí evolucionar en función de los documentos que vayan formando parte del sistema y de las peticiones que vayan haciendo los usuarios en sus consultas, las cuales son siempre una guía bastante exacta de lo que esperan encontrar los usuarios en la colección. La forma de definir las políticas se puede basar además en la utilización del análisis de conectividad, entendiendo que las páginas de una temática dada están enlazadas entre sí y por lo tanto podemos recuperar documentos que sean enlazados a partir de los que tenemos en la colección y que a su vez son recuperados con más frecuencia por los usuarios del sistema. De

esta forma las políticas de adquisición pueden ir modificándose en función de las peticiones de los usuarios y así el sistema puede aprender qué temáticas son las más solicitadas y adquirir documentos relacionados con ella.

Este agente de adquisición se comunicaría con el agente de usuario con el objeto de redefinir y actualizar sus políticas de adquisición en función de las consultas. También debería poder comunicarse con el agente de indización para que éste comience a indizar los documentos que entran a formar parte del sistema.

Agente de indización

Este agente se encarga de indizar automáticamente los documentos y de extraer los conceptos que son más representativos de su contenido. Para realizar esta tarea debe comunicarse con el agente de gestión de tesauro, normalizar términos y extraer conceptos a partir de los términos que aparecen en los documentos.

Este agente contiene un motor de inferencia basado en lógica difusa y está encargado de asignarle relevancia a cada uno de los descriptores que se le han asignado previamente al texto mediante la comparación con el tesauro o tesauros contenidos en el sistema. El índice de relevancia de un concepto determinado es calculado agregando los índices de relevancia parciales que se obtienen para cada una de las marcas o etiquetas definidas en la configuración del agente, en función de la frecuencia relativa del concepto y de la importancia que aparece en la configuración del agente para la marca en cuestión.

Una vez que se obtiene un índice de relevancia previo para un concepto en función de las etiquetas en las que aparece y de su frecuencia de aparición, el agregador de relevancia obtiene el índice de relevancia final en función del cálculo de la primera aparición del concepto en el texto y de su papel en el tesauro (si es que se ha podido normalizar) como término general o término específico, para medir así el grado de precisión que nos proporciona cada descriptor.

De esta forma el agente de indización le asigna una relevancia o peso total a cada concepto aparecido en cada documento, de tal forma que permite la generación de un vector de pesos para cada documento útil para su posterior recuperación. La diferencia con los cálculos de asignación de pesos tradicionales basados en TF-IDF estriba en que en este caso no se utilizan cálculos sobre el conjunto de la colección, lo que permite indizar el documento independientemente de la colección, sin que esto suponga una pérdida de precisión en la asignación de pesos.

Agente de clasificación

Este agente es el encargado de asignarles los documentos a las categorías de una clasificación temática previamente establecida con el objeto de mejorar las posibilidades de recuperación del documento. También permite establecer la similitud entre un documento en concreto y otros relacionados temáticamente para ampliar la exhaustividad de la recuperación.

Para realizar esta labor debe comunicarse con el agente de indización y utilizar los términos de indización como guía para clasificar temáticamente los documentos, y con el agente de gestión del tesoro que será el encargado de gestionar también la clasificación que utiliza este agente para realizar su tarea.

Para implementar este agente se pueden utilizar los algoritmos de *clustering* que agrupan los documentos que contienen una mayor cantidad de conceptos comunes.

Agente de gestión de tesauros

Éste es el agente encargado de la gestión del tesoro que se utiliza como base de conocimiento del dominio frente al reconocimiento y la normalización de conceptos.

Además el agente debe implementar algoritmos de evaluación del tesoro para estudiar si las distintas actualizaciones han redundado en una mayor eficiencia de la base de conocimiento en función de la capacidad de recuperación que tienen los descriptores en el momento en que un usuario intenta recuperar un documento perteneciente a la colección. Por esta razón este agente debe comunicarse con el agente de interfaz con la intención de actualizar el tesoro para que siempre utilice como descriptores los términos referidos a un concepto con mayor capacidad de recuperación o, lo que es lo mismo, aquellos que son más utilizados por los usuarios a la hora de recuperar un documento que contenga un determinado concepto.

Ahora bien, como ya se ha mencionado, la propuesta del agente de gestión de tesauros va más allá de permitir la mera consulta del lenguaje, y pasa por realizar tareas de mantenimiento interno y actualización del tesoro. La necesidad de actualización de los tesauros es un problema que viene siendo abordado desde hace tiempo en la literatura del área de documentación. La doctora Blanca Gil Urdiciain dedica una sección del capítulo referente a tesauros de su *Manual de Lenguajes Documentales precisamente a este tema*.⁸

8 B. Gil Urdiciain. *Manual de Lenguajes Documentales*. Madrid Ed. Noesis, 1996. pp. 215-220.

En este texto la doctora Gil resalta el hecho de que la extensibilidad que facilita la estructura de los tesauros documentales permite ampliar y modificar el vocabulario en función de las necesidades que surgan a lo largo de su uso continuado. Las necesidades de modificación se hacen más patentes sobre todo en los tesauros de nueva creación, ya que se requiere un tiempo de adaptación del lenguaje documental al entorno en el que se utiliza. La actualización del tesoro ha de hacerse tanto para incorporar la terminología derivada del desarrollo de la ciencia o materia a la que se dedica, como para cubrir lagunas o fallos detectados durante su utilización.

Este segundo proceso de corrección puede automatizarse a partir de que el agente de descripción de tesauros detecte estos fallos o lagunas. Uno de los ejemplos más claros de la automatización de este proceso se da a partir del estudio de las consultas realizadas por los usuarios al recuperar documentos del sistema. Estas consultas pueden ser procesadas de forma estadística para establecer la capacidad de recuperación de los descriptores utilizados en el tesoro. A partir de la medida de la capacidad de recuperación de cada descriptor y no-descriptor contenidos en el tesoro podemos establecer un índice de relevancia que mantenga siempre como descriptor al término con mayor capacidad de recuperación. Podríamos decir, a partir de la práctica de los usuarios, que el agente de gestión de tesauros debe ser capaz de aprender qué términos son los más utilizados para la recuperación de determinados documentos, de tal forma que les asigne siempre a estos documentos los términos con mayor capacidad de recuperación. A esta labor de aprendizaje podemos añadir distintas metodologías de evaluación y mantenimiento de tesauros que permitan mantener su consistencia en todo momento.

Como se puede advertir, las tareas realizadas por el agente de gestión de tesauros son lo suficientemente complejas como para justificar su separación como agente autónomo.

Agente de recuperación

Este agente es el encargado de gestionar el trato directo con el usuario final del sistema, el cual puede ser un usuario humano u otro agente de usuario que quiere consultar la colección. Se trata de un agente de interfaz cuyo propósito fundamental es traducir las consultas del lenguaje natural a los parámetros que utiliza el sistema para representar los documentos.

El proceso de búsqueda se realiza mediante la implementación del modelo de espacio vectorial, comparando un vector generado a partir de la consulta del usuario con los vectores que definen cada uno de los documentos, y que han sido generados por el agente de indexación anteriormente descrito.

Una de las características fundamentales de este agente es su capacidad de adaptación a las necesidades del usuario. Tal adaptación se realiza analizando las consultas de los usuarios para poder establecer un perfil de usuario con el propósito de mejorar la capacidad de recuperación del sistema. Este perfil se puede definir a partir de un vector generado mediante el análisis de todas las consultas hechas por cada usuario; es decir, cada vez que entra un usuario al sistema se le identifica y se registran sus consultas para componer un vector global correspondiente a ese usuario. De esta forma podemos conocer el perfil de búsqueda de cada usuario y adaptar las capacidades de recuperación del sistema a este perfil mediante la comparación de vectores.

Por otro lado la comunicación con el resto de agentes se basa en la información recopilada por el agente de interfaz. De esta forma los términos introducidos por los usuarios para la recuperación le son comunicados al agente de gestión de tesauros para actualizar el tesoro contenido en el sistema. De este modo se garantiza que el agente de gestión de tesauros normalice siempre con base en los términos que utilizan los usuarios para recuperar los documentos; es decir, por aquellos que tienen mayor capacidad de recuperación.

Comparación de la propuesta con los modelos tradicionales

La gran diferencia entre esta propuesta y la del sistema de arquitectura por multiagentes (SMA) que se han estudiado para la realización de este trabajo radica en que en este último se contempla la recuperación de información como un conjunto amplio de operaciones, mientras que en la mayoría de los sistemas estudiados el agente se reduce a la fase final de recuperación de la información, pero no interviene en el resto del proceso. Esto limita las posibilidades del agente y su capacidad de aprendizaje, e imposibilita una verdadera cooperación de los agentes para la resolución distribuida que supone el problema de la recuperación de la información.

CONCLUSIÓN

Una vez que hemos analizado los tipos de agentes de información que nos podemos encontrar llegamos a la conclusión de que, pese a excepciones, el gran éxito alcanzado por los agentes de información en Internet está limitando hasta cierto punto su desarrollo. Creemos que esta afirmación que podría parecer contradictoria no lo es, ya que se basa en la idea de que la necesidad de desarrollar agentes funcionales que sean capaces de trabajar en un entorno como Internet está limitando hasta cierto punto la investigación a largo

plazo en este campo. De esta forma encontramos numerosos artículos destinados a describir las diferentes aplicaciones de los agentes de información que se pueden realizar en campos muy concretos de aplicación.

Se echan de menos en la bibliografía consultada para la elaboración de este trabajo, artículos que reflexionen de forma general sobre el diseño de agentes de información que vayan más allá de sus aplicaciones concretas.

BIBLIOGRAFÍA

- [BL01] Travis Bauer and David B. Leake. Real time user context modeling for information retrieval agents. In *Tenth international conference on Information and knowledge management*, pages 568 – 570. ACM Press New York, NY, USA, 2001.
<http://portal.acm.org/citation.cfm?id=502693&dl=ACM&coll=portal>.
- [Ble98] M Bleyer. Multi-agent systems for information retrieval on the world wide web. Master's thesis, Universidad de Ulm, Facultad de Informática, 1998.
<http://www.informatik.uni-ulm.de/ki/Students/mb/cema/Diplomarbeit.ps.gz>.
- [BS95] M. Balabanovi'c and Y. Shoham. Learning information retrieval agents: Experiments with automated web browsing. In *AAAI-95 Spring Symposium on Information Gathering from Heterogenous, Distributed Environments*, 1995.
<http://citeseer.ist.psu.edu/context/202852/0>.
- [DR04] Juan L. Dinos-Rojas. Arquitectura de un sistema basado en agentes para la recuperación de metadatos rdf con base en una ontología de documentos. Master's thesis, Universidad de Puerto Rico, Departamento de Ingeniería Eléctrica y Computadoras, 2004.
<http://grad.uprm.edu/tesis/dinosrojas.pdf>.
- [ea99] Brewington, Brian et al. *Mobile agents in distributed information retrieval*, chapter 15, pages 355–395. Springer, 1999.
<http://agent.cs.dartmouth.edu/papers/brewington:ir.ps>.
- [Gil99] Gil Urdiciain, Blanca. *Manual de lenguajes documentales*. Noesis, 1999.
- [JYW98] Les Miller Jihoon Yang, Vasant Honavar and Johnny Wong. Intelligent mobile agents for information retrieval and knowledge discovery from distributed data and knowledge sources. In *IEEE Transactions on Power Systems*, 1998.
<http://citeseer.ist.psu.edu/yang98intelligent.html>.
- [KH94] Arens, Yigal Knoblock, Craig A. and Hsu, Chun-Nan. Cooperating agents for information retrieval. In *Second International Conference on Cooperative Information, Ontario, Canada*, octubre 1994.
<http://www.isi.edu/info-agents/papers/knoblock94-coopis.pdf>.

- [Klu99] Klusch, Matthias. *Intelligent Information Agents. Agent-Based Information Discovery and Management on the Internet*. Springer, 1999.
- [Kno04] Tuchinda, Rattapoom Knoblock, Craig A. Agent wizard: Building information agents by answering questions. In *Proceedings of Intelligent User Interfaces, Island of Madeira, Portugal, 2004*, febrero 2004.
<http://www.isi.edu/info-agents/papers/tuchinda04-iui.pdf>.
- [LRJ94] Brian Logan, Steven Reece, and Karen Sparck Jones. Modeling information retrieval agents with belief revision. In W. Bruce Croft and C. J. van Rijsbergen, editors, *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. Dublin, Ireland, 3-6 July 1994 (Special Issue of the SIGIR Forum)*, pages 91–100. ACM/Springer, 1994.
- [LS99] López-López, Aurelio and Sandoval Tafolla, Luis. Un sistema multiagente para recuperación de información distribuida. In *National Meeting of Computer Science (ENC'99)*, septiembre 1999.
<http://cseg.inaoep.mx/allopez/psmrid.ps>.

