

UNA TAXONOMÍA PARA LOS AGENTES DE INFORMACIÓN

- Autores:** Carlos Carrascosa, Vicente J. Julián y Miguel Rebollo
{carrasco, vinglada, mrebollo}@dsic.upv.es
Universidad Politécnica de Valencia
- Resumen:** En la última década Internet se ha convertido en el gran vehículo o contenedor de información. Es habitual en cualquier proceso de recuperación que un usuario se sienta desbordado por el elevado volumen de documentos o de referencias que se le suministra. En este punto, el empleo de técnicas inteligentes no sólo puede resultar adecuado para la mejora de los resultados obtenidos, sino que constituye actualmente un campo de investigación abierto. Una de las técnicas de inteligencia artificial más utilizadas en los últimos tiempos es la del paradigma de agente / sistema multi-agente. Así, en los últimos años diferentes aplicaciones basadas en el paradigma de agente han ido apareciendo intentando buscar fundamentalmente una solución al problema de la explosión de la información. En este artículo se presenta una panorámica de estas aplicaciones junto con una clasificación de las mismas de acuerdo a la función que realizan.
- Palabras clave:** agentes; taxonomía; internet; filtrado; búsqueda
- Abstract:** Over the last decade, Internet has become a huge information vehicle or container. A user is usually overflowed by the great quantity of documents or references he is supplied with. So, the use of artificial intelligence techniques is not only suitable for improving the obtained results, but now it is an open research field. One of the most lately used artificial intelligence techniques is the agent / multi-agent system paradigm. In this way, different agent-based applications have appeared in the last few years trying to look mainly for an answer to the information explosion. In this paper, a panoramic of these applications is presented along with their classification according to their function.

Introducción

En la última década los métodos para gestionar y organizar la información han variado notablemente. Hoy en día, Internet se ha convertido en el gran vehículo o contenedor de información. A medida que ésta ha ido evolucionando y creciendo se ha complicado la forma de poder localizar y consultar la información disponible en la propia red. La mayoría de herramientas disponibles actualmente para la localización y recuperación de información en Internet se centran más en la cantidad que en la calidad de la información. Es habitual en cualquier proceso de recuperación que un usuario

se sienta desbordado por el elevado volumen de documentos o de referencias que se le suministra. Resulta, en este caso, evidente la necesidad de una cierta automatización en el proceso que facilite la labor de localización de información de calidad y que a su vez permita diferenciar entre ella desde el punto de vista subjetivo del propio usuario. En este punto, el empleo de técnicas inteligentes resulta adecuado para la mejora de los resultados obtenidos.

En los últimos años diferentes aplicaciones basadas en el paradigma de agente han ido apareciendo intentando buscar fundamentalmente una solución al problema de la explosión de la información.

Definición

Actualmente, podemos encontrar un gran número de propuestas en la literatura acerca de lo que es y lo que no es un agente, sin que ninguna de ellas haya sido plenamente consensuada por la comunidad científica. Una de las definiciones más aceptadas es la propuesta en [Wooldridge 1995], según la cual un agente viene definido por su flexibilidad, entendiendo por flexible que un agente sea:

Reactivo, responda al entorno en que se encuentra.

Proactivo, que sea capaz de intentar cumplir sus propios objetivos.

Social, sea capaz de comunicarse con otros agentes mediante algún tipo de lenguaje.

En los trabajos más recientes relacionados con el tema se tiende a etiquetar a un agente en función del papel que desempeña. Así, por ejemplo, un agente con acceso a diversas fuentes de información y que es capaz de agrupar y manipular la información obtenida de dichas fuentes para ponerla a disposición del usuario se conoce como *agente de información*.

Más formalmente, los agentes inteligentes de información o agentes de Internet pueden definirse como aquellos sistemas software de computación que tienen acceso a múltiples y heterogéneas fuentes de información que están distribuidas geográficamente [Klusck 1999]. Esta clase de agentes suelen ayudar a sus usuarios en la búsqueda de información útil y relevante. Los agentes de información, además de proveer acceso transparente a las diferentes fuentes de información existentes en Internet, deben ser capaces de recuperar, analizar, manipular e integrar información heterogénea cuando ésta sea solicitada. En resumen, este tipo de agentes trata de resolver el problema de la obtención de la información correcta en el tiempo preciso.

Tipología

A continuación se presentan diferentes clasificaciones que pretenden aclarar la tipología de agentes inteligentes de información que se basan tanto en el ámbito en el que actúan como en las tareas que llevan a cabo. Al consultar la bibliografía referente al área de los agentes de información surgen términos como *Interface agents*, *System agents*, *Advisory agents*, *Filtering*

agents, Retrieval agents, Navigation agents, Monitoring agents, Recommender agents, Profiling agents.

La primera clasificación toma como criterio el ámbito de actuación de los agentes [Caglayan 1997]. De esta forma se puede hablar de:

Agentes de escritorio (agentes de sistema operativo, agentes de aplicaciones, agentes de interfaz, etc...)

Agentes Internet (agentes de búsqueda, filtrado, recuperación de información, agentes de notificación, agentes móviles, etc...)

Agentes Intranet (agentes de ayuda colaborativa, agentes de bases de datos, agentes de automatización de procesos, etc...)

Otra posible clasificación divide los distintos tipos de agentes de información basándose en el entorno donde actúa cada agente, y posteriormente divide cada categoría obtenida según la tarea realizada por cada agente. De esta forma podemos hablar de:

Agentes de Búsqueda en la Web: proporcionan servicios de búsqueda al usuario.

Agentes de Servicio en la Web: residen en lugares específicos de la red proporcionando servicios diversos.

Agentes de Filtrado de Información: filtran información electrónica basándose en las preferencias y gustos específicos del usuario.

Agentes de Recuperación de Información: proporcionan un conjunto de información personalizada al usuario de acuerdo con las preferencias de éste.

Agentes de Notificación: avisan al usuario cuando ocurre un evento de interés para éste.

Agentes de Servicio: proporcionan servicios especializados a los usuarios.

Agentes Móviles: viajan de un lugar de la red a otro para llevar a cabo tareas específicas de usuario.

La clasificación que se plantea en este artículo divide los agentes de información según su función y se puede considerar como una clasificación más generalista. Así, se puede hablar de:

Agentes de Búsqueda (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 *agentes de búsqueda*, tanto aplicaciones cliente como aplicaciones servidor.

Agentes de Filtrado (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 correo electrónico, así como otro tipo de productos más específicos proporcionan prestaciones básicas de *agentes de filtrado*.

Agentes de Monitorización (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.

Esta última clasificación se va a ver con más detalle en los siguientes puntos. Esta clasificación no es excluyente, pudiendo un agente concreto pertenecer a más de una categoría.

Agentes de búsqueda

Los agentes de búsqueda son el tipo de agente de información más conocido y con más ejemplos en la literatura. Este tipo de agente permite al usuario la búsqueda de información en servidores de la red (el agente en sí es independiente de si nos encontramos en Internet o en una Intranet). Actualmente existen en la red infinidad de robots de búsqueda con los que el usuario interactúa directamente con mayor o menor eficacia.

La forma de medir el grado de éxito de una consulta se basa en dos parámetros:

precisión, indica la relación existente entre los documentos relevantes para la consulta y el total de documentos devueltos por la misma.

recall, indica la relación entre el total de documentos relevantes devueltos y el total de documentos.

Ambos parámetros son inversamente proporcionales, por tanto si intentamos mejorar uno perjudicaremos el otro. De esta forma se debería intentar conseguir un equilibrio entre los dos, lo cual es bastante complejo en los actuales motores de búsqueda. Los agentes de búsqueda proporcionan un punto de inicio en la exploración en Internet para los usuarios. Las ventajas que podemos encontrar o que deberían proporcionar son:

Fácil de usar.

Incremento de la productividad si el usuario sabe lo que quiere.

Incremento de la precisión en la búsqueda sin aumentar notablemente el número de documentos devueltos.

Reducir la sobrecarga que generan los procesos de búsqueda en la Web y en el propio sistema.

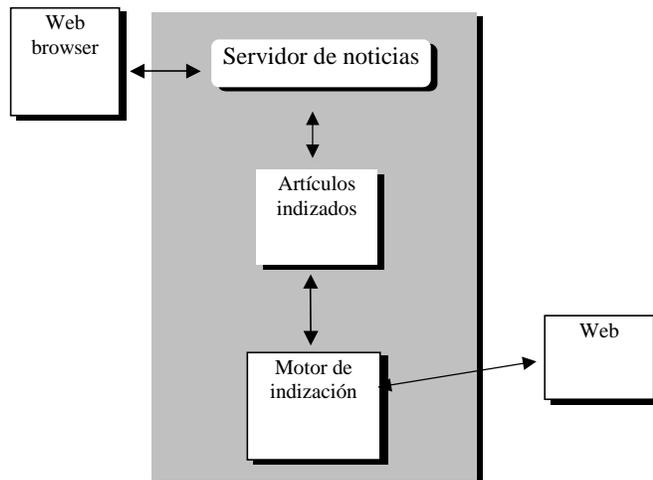


Ilustración 1: Esquema de un agente de búsqueda.

Este tipo de agentes podrían realizar las siguientes funciones en una corporación [Caglayan 1997]:

Indexar información sobre servidores de la organización para búsquedas a realizar por los empleados.

Desarrollar servicios especializados de búsqueda empleando para ello recursos externos si es relevante para la organización.

Promocionar mediante servicios específicos el uso de los recursos disponibles internos de una organización.

Permitir compartir resultados de búsquedas entre diferentes comunidades de usuarios dentro de una organización.

Agentes de filtrado

Para que un agente de información pueda servir eficientemente a sus distintos usuarios, necesita ser capaz de almacenar, aprender y manipular las preferencias y gustos de cada usuario, así como sus cambios. Este perfil de usuario puede servir como un indicador mucho más fiable de la información a recuperar que las palabras clave proporcionadas por el usuario, método que es utilizado por los distintos y múltiples buscadores de información que existen en la actualidad.

La tarea de filtrado de un agente de Internet consiste en determinar si un artículo es relevante o irrelevante para el usuario basándose en el perfil de usuario [Caglayan 1997], en un entorno donde la probabilidad de encontrar un documento relevante es muy baja en comparación con la alta probabilidad de encontrar un documento irrelevante.

Existen diferentes técnicas de aprendizaje que pueden ser utilizadas para aprender el perfil de usuario. Entre ellas se pueden destacar los clasificadores Bayesianos, el vecino más cercano, los árboles de decisión, TF-IDF, y las redes neuronales, entre otras. Después de haber logrado aprender el perfil, éste puede ser utilizado para el filtrado de información, pudiendo además anticiparse a la obtención de la información solicitada por el usuario. La arquitectura de un agente de filtrado de información es la siguiente:

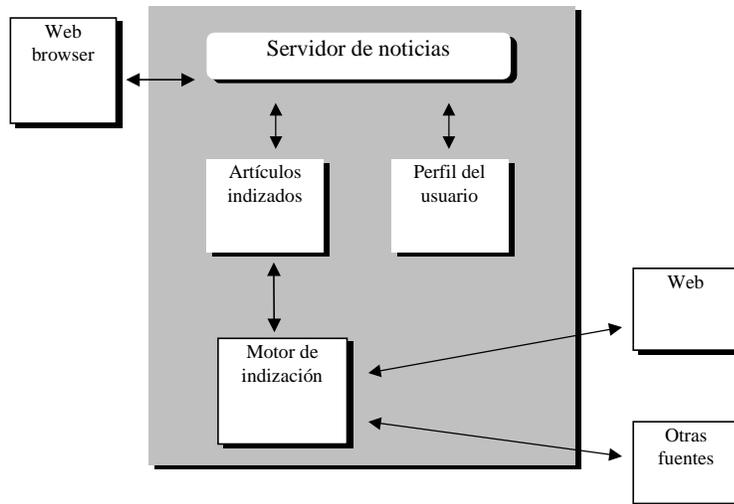


Ilustración 2: Esquema de un agente de filtrado.

El funcionamiento que sigue un agente de filtrado de información es el siguiente:

Establecer el perfil del usuario. Esto se puede llevar a cabo de dos formas diferentes:

Directa: Cuando el usuario determina su perfil a través de una interfaz en el que especifica, por medio de palabras clave, los temas de su interés.

Indirecta: Por el control efectuado por el propio agente de los enlaces seguidos por el usuario en una sesión concreta.

El agente comienza la búsqueda en las diferentes fuentes de información, indizando, mediante el motor de indización, todos los artículos que encuentre de acuerdo con el perfil de usuario establecido previamente.

Una vez filtrados aquellos artículos que se ajustan a la especificación del usuario, el agente elabora el informe que presenta los resultados obtenidos los cuales serán enviados a la dirección correspondiente.

El motor de indización asigna una palabra clave a cada artículo. La mayoría de sistemas de recuperación de información también generan Tesoros de sinónimos con la intención de indizar palabras por su raíz.

Para conocer la representatividad de un término dentro de un documento determinado y, por tanto para establecer la pertinencia del mismo al perfil del usuario, el motor de indización emplea técnicas de frecuencia de términos como la siguiente fórmula:

$$W_{ik} = f_{ik} [\log n - \log d_k + 1]$$

Donde f_{ik} es la frecuencia de ocurrencias de términos K en un documento i, d_k es el número de documentos en la colección con términos K, y n es la colección total.

Esta ponderación significa que el término K en un documento i es directamente proporcional al número de veces que él aparece en el documento, y la importancia del término K es inversamente proporcional al número de documentos de la colección donde el término aparece.

Algunas de las aplicaciones de este tipo de agentes son [Caglayan 1997]:

- Entrega de noticias de cualquier tipo (financieras, políticas, etc.) que puedan tener repercusión al usuario.

- Seguimiento de las noticias de un determinado perfil de usuario.

- Actualización automática de configuraciones hardware según un perfil, incluso proporcionando precios de posibles actualizaciones.

- Proporcionar noticias de un determinado perfil a diferentes comunidades de una organización.

Agentes de monitorización

Este tipo de agentes permiten tener alerta a un usuario frente a eventos que le puedan resultar de interés. Los eventos frente a los cuales estos agentes lanzan algún tipo de notificación pueden ser:

- Cambios en una página Web lo bastante importantes para un usuario.

- Comunicación de eventos de tipo personal como lanzar un mensaje el día del cumpleaños de alguien.

- Adiciones a los motores de búsqueda para determinadas palabras clave.

Los agentes de monitorización se pueden considerar como un software localizado en un determinado servidor que se encarga de descubrir y notificar eventos interesantes especificados previamente por el usuario.

La forma en que este tipo de agentes sirven la información a su usuario puede ser variada:

- Indicar únicamente qué página o páginas han cambiado y desde cuando ha sucedido esto.

- Descargar el texto de las páginas actualizadas y que son de interés para el usuario, filtrando en este caso imágenes, gráficos y demás.

- Construir páginas HTML que embeben las modificaciones encontradas por la Web por el agente y que son presentadas al usuario.

Aplicaciones

A continuación se puede ver un ejemplo de clasificación de agentes de información de acuerdo a la división propuesta:

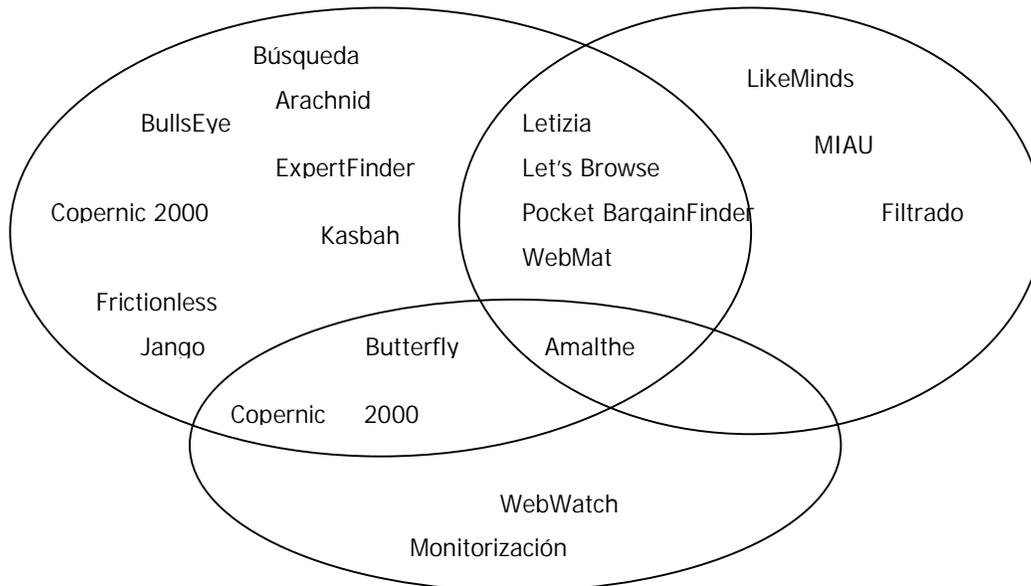


Ilustración 3: Ejemplo de clasificación de agentes de información.

Conclusiones

Hoy en día el término *agente* está sufriendo un mal uso al aplicarse para denominar casi cualquier tipo de aplicación informática. Se hace por ello necesario el determinar qué es y qué no es un agente, y en especial cuáles de estos

agentes se pueden considerar agentes de información, y qué se puede esperar uno frente a uno de estos agentes de información.

Con esta clasificación se pretenden caracterizar los agentes de información para permitir evaluar las herramientas que se autodenominan agentes, concluyendo la veracidad de dicha afirmación.

Referencias

[Caglayan 1997] Caglayan, Alper; Harrison, Colin (1997). Agent sourcebook. New York: Wiley Computer Publishing, cop.

[Klusck 1999] Klusck, Matthias (1999). Intelligent Information Agents. Springer-Verlag.

[Wooldridge 1995] Wooldridge, M.; Jennings, N. R. (1995). Intelligent agents: Theory and practice. The Knowledge Engineering Review, 10(2):115–152.

Direcciones Internet

- Amalthea [En línea]. Disponible en

- <<http://moux.www.media.mit.edu/people/moux/papers/PAAM96/node5.html>>
[Consulta: 13.02.2001]
- Arachnid [En línea]. Disponible en
<<http://dollar.biz.uiowa.edu/~fil/Papers/ICML.ps>> [Consulta: 13.02.2001]
 - BullsEye. [En línea]. Disponible en
<<http://www.intelliseek.com/prod/bullseye.htm>> [Consulta: 13.02.2001]
 - ButterFly [En línea]. Disponible en
<<http://www.neilvandyke.org/butterfly/>> [Consulta: 13.02.2001]
 - Copernic [En línea]. Disponible en
<<http://www.copernic.com>> [Consulta: 09.02.2001]
 - ExpertFinder [En línea]. Disponible en
<<http://www.media.mit.edu/~adriana/projects/EF/>> [Consulta:
13.02.2001]
 - FishMarket [En línea]. Disponible en
<<http://www.iii.csic.es/Projects/fishmarket>> [Consulta: 09.02.2001]
 - Frictionless Commerce [En línea]. Disponible en
<<http://www.frictionless.com/main.html>> [Consulta: 14.02.2001]
 - Jango – Excite . [En línea]. Disponible en
<<http://www.jango.excite.com>> [Consulta: 09.02.2001]
 - Kasbah – Market Maker. [En línea]. Disponible en
<<http://ecommerce.media.mit.edu/maker/maker.htm>> [Consulta:
09.02.2001]
 - Let's Browse [En línea]. Disponible en
<<http://lieber.www.media.mit.edu/people/lieber/Lieberary/Lets-Browse/Lets-Browse-Intro.html>> [Consulta: 09.02.2001]
 - Letizia. [En línea]. Disponible en
<<http://lcs.www.media.mit.edu/people/lieber/Lieberary/Letizia/Letizia-Intro.html>> [Consulta: 09.02.2001]
 - LikeMinds [En línea]. Disponible en
<<http://www.macromedia.com/software/likeminds/>> [Consulta:
09.02.2001]
 - MIAU project (Multiple Internet Agents for User-adaptive Decision)
[En línea]. Disponible en
<<http://www.dfki.de/imedia/miau/>> [Consulta: 09.02.2001]
 - Pocket BargainFinder [En línea]. Disponible en
<<http://www.accenture.com/xd/xd.asp?it=enWeb&xd=services\cstar\projects\PocketBargainFinder.xml>> [Consulta:
13.02.2001]
 - WebMate [En línea]. Disponible en
<<http://www.cs.cmu.edu/~softagents/webmate/Introduction.html>>
[Consulta: 09.02.2001]
 - WebWatch [En línea]. Disponible en
<<http://pharadox.nu/software/pww.html>> [Consulta: 14.02.2001]