

Búsqueda en Internet, una aguja en un pajar

RAMON NOGALES MORAN

Tradicionalmente las empresas han manejado volúmenes importantes de datos para la toma de decisiones, de hecho uno de los objetivos de la evolución de las arquitecturas de los sistemas de información hacia las redes de área local y las aplicaciones cliente servidor ha sido el conseguir mayores productividades en la manipulación de la información, cuya máxima expresión ha sido la introducción de los DATAWAREHOUSE en estas arquitecturas. Sin embargo, las herramientas tradicionales de búsqueda en estos sistemas están limitadas por los algoritmos utilizados, que no permiten gestionar más de 50.000 objetos. Además estaban ligados a un entorno concreto, el cual tenía una organización de datos y unos formatos básicamente propietarios.

La explosión de Internet ha hecho que nos encontremos en un escenario inundado de información, no de 50.000 objetos sino de millones. Hoy día determinar la información relevante de todos los datos disponibles para la toma de decisiones es prácticamente imposible, produciendo a los responsables una ansiedad añadida, aproximadamente el 33% de altos directivos según un trabajo realizado por Reuters dice vivir esta situa-

ción y es más, el 45% manifiesta que retrasa su toma de decisiones.

Por tanto uno de los problemas más importantes del gran crecimiento de la publicación de información en Internet, es el poder acceder a la misma, de forma productiva y esto es válido tanto para particulares como para las empresas.

INFORMACION DESORGANIZADA

Los usuarios de Internet se enfrentan a una ingente cantidad de información muy desorganizada y aunque solamente un tercio de la información disponible en Internet es de acceso libre, todavía es un volumen muy considerable y la búsqueda de la página Web o documento de interés es realmente un reto.

El caos del cyberspacio es solamente un aspecto del problema de la búsqueda. La competencia entre los diferentes Web's, hace que se establezca un clima de batalla de impactos de información entre ellos para captar la atención de los usuarios, lo cual hace más difícil, si cabe encontrar información relevante y fiable.

Cuando un usuario a través de un buscador generalista o a través de

servicios de un directorio pregunta por una determinada información el resultado es una cantidad enorme de direcciones que le llevaría mucho tiempo el desmenuzar. Para encontrar información relevante en esta gran cantidad de posibilidades hay que ser prácticamente un especialista en la búsqueda de información.

Desde el punto de vista de los usuarios empresariales, navegar es una pérdida de tiempo, es más productivo buscar, sin embargo para que la búsqueda sea productiva, hay que ser especialista de la materia.

La solución es poder utilizar herramientas de búsqueda que realicen el trabajo de buscar la información.

Cuando tenemos que seleccionar una herramienta de búsqueda o recuperación de información tenemos que tener en cuenta dos requisitos: que encuentre información relevante y que lo haga en el menor tiempo posible. La velocidad de respuesta no depende exclusivamente de la herramienta sino de el volumen de tráfico en la redes corporativas (RAL o Intranet), de la popularidad de los Web a los que





tiene que acceder y el ancho de banda de los enlaces de transmisión así como del tipo de contenido a ser transmitido.

Cuando desde nuestra empresa entramos no en la Intranet corporativa sino a Internet, el ancho de banda no es predecible y la respuesta a nuestra petición de búsqueda puede ser extremadamente baja en periodos de pico. El desarrollo de agentes inteligentes que realicen las funciones de búsqueda y recuperación de la información así como su clasificación en función de los perfiles de búsqueda que han sido introducidos como requisitos para el agente inteligente, permite el que no empleemos nuestro tiempo en esta tarea. Este desarrollo de los agentes inteligentes se realiza en estrecha colaboración con las empresas que lo van a utilizar, ya que son personalizaciones de los productos genéricos.

Las dos herramientas básicas de búsqueda en Internet son los motores de búsqueda y los directorios. Los directorios agregan información en bases de datos especiales de acuerdo

a unos esquemas de clasificación que combinan diferentes factores: alfabéticos, geográficos, tópicos. Este tipo de herramienta de búsqueda requiere un esfuerzo humano significativo para coleccionar, clasificar, codificar, anotar y editar los datos.

Cuando entramos en Internet, el ancho de banda no es predecible y la respuesta a nuestra petición de búsqueda puede ser extremadamente baja.

Los motores de búsqueda son sistemas que automáticamente coleccionan e indexan mediante robots la información sin intervención humana. Sin embargo, estos robots no son capaces todavía de sustituir a la inteligencia humana, por lo que la informa-

ción que hay en los directorios es más fiable que la que dan los motores de búsqueda generales.

FUTURO DE LOS MOTORES DE BÚSQUEDA

El negocio de los motores de búsqueda está cambiando con rapidez debido a que es obvio que no pueden vivir exclusivamente de la publicidad. Algunas firmas líderes incluso publicitan y crean valor para sus compañías esperando ser adquiridas por una entidad mayor, por ejemplo una gran compañía de ordenadores.

Los motores generales de búsqueda tienen un limitado futuro en la forma que trabajan actualmente.

Los directorios son herramientas excelentes de búsqueda para navegar pero requieren la interactividad humana para la operación y explotación. Como los web's en Internet aumentan muy rápidamente los directorios son presionados para ser actualizados y estos procesos son caros.

Hay un gran número de directorios tanto de tipo general como específicos, pero es creciente la dificultad de su existencia como directorios independientes. Solo aquellos directorios que son soportados por grupos de interés como: directorios telefónicos, asociaciones profesionales, grupos de industriales y líneas de productos son los que son más actualizados y visitados. Son los que tienen futuro.

Algunos de los directorios han sido adquiridos por empresas de servicios de búsqueda, para aumentar el automatismo de determinadas operaciones de evaluación y recopilación de datos. Los resultados de esta decisión no son tan seguros como los directorios que son mantenidos por personas expertas pero en muchos casos ha sido la manera de sobrevivir.

Los principales factores que determi-

nan las características y aceptabilidad de las herramientas de búsqueda incluyen el tamaño, los contenidos y la capacidad de concurrencia de las base de datos de indexación. Otro factor importante es la velocidad del mecanismo de búsqueda, la facilidad de uso, la selección de las características de la búsqueda y el diseño de los interfaces del usuario final.

INFORMACION ESPECIFICA

El éxito de la búsqueda en Internet depende del grado de efectividad del plan de búsqueda. La premisa básica de cualquier búsqueda es apuntar directamente a una información específica en vez de navegar.

Cuando un motor de búsqueda arranca es importante la selección de las palabras de búsqueda, ya que la búsqueda se deteriora sino elegimos bien las palabras. Los motores de búsqueda difieren unos de otros en como interpretan las mismas palabras o frases. Los usuarios deberán probar sinónimos plurales así como los operadores lógicos disponibles en la mayoría de los motores. Trabajar con la lógica booleana es básico para tener una respuesta más ajustada.

El proceso de búsqueda aumenta mucho más en su efectividad con la introducción de los metabuscadores los cuales lanzan simultáneamente a varios motores de búsqueda.

Una de las técnicas más utilizadas en las funciones de búsqueda de los metabuscadores es el almacenamiento intermedio de la información en una base de datos, lo cual permite refinar mejor los parámetros de las búsquedas y facilitar las consultas posteriores en base a esta información.

La mayoría de los buscadores utilizan técnicas derivadas de las tecnologías de los agentes inteligentes, sin embargo los motores denominados

inteligentes son aquellos que se basan en las tecnologías del lenguaje natural, las técnicas de aprendizaje y las redes neuronales. Este tipo de motores tratan de recrear lo que sería un dialogo humano. Sin embargo, el problema de estos agentes, es que son perseguidos por los propietarios de los Web's ya que se llevan la información y les quitan publicidad.

Los motores inteligentes también se emplean en capturar y almacenar preferencias de los usuarios. El objetivo es saber con precisión sus requerimientos y mejorar el plan de búsqueda.

El filtrado colaborativo es otra técnica empleada por los motores inteligentes, se basa, en la comparación

Algunos de los directorios han sido adquiridos por empresas de servicios de búsqueda, para aumentar el automatismo de determinadas operaciones.

de las preferencias de diferentes usuarios sobre un mismo dominio. Hay numerosos proyectos en el campo de los motores inteligentes ya que es la tendencia tecnológica más interesante de cara al futuro.

Otra de las líneas de investigación actual es la capacidad de la búsqueda visual, ya que el usuario puede definir con mayor precisión lo que quiere ver de una imagen. Los atributos básicos que actualmente se están considerando son: color, forma y tex-

tura, también se está investigando sobre rasgos faciales y huellas y en el futuro sonido.

LOCALIZACIONES ESPACIALES

No está cerca el tener un buscador de este tipo debido a su gran complejidad. Cuando las bases de datos multimedia son grandes y los videos complejos un buscador exhaustivo multimedia puede ser muy poco práctico. Además hay que tener en cuenta también las necesidades de considerar localizaciones espaciales y relaciones entre objetos y regiones de una imagen, lo cual complica la situación. La tendencia irá a sistemas de filtrado y personalización previas a la búsqueda concreta.

Actualmente los motores de búsqueda se están incorporando a los navegadores de Internet, con el objetivo de estructurar la información en canales temáticos y así poder facilitar al usuario el uso de Internet. Sin embargo, el problema de estas herramientas de búsqueda es que cada día son más perseguidas por los propietarios de los Web's ya que se llevan la información y les quitan publicidad.

El mercado de los motores de búsqueda y del software de recuperación de textos era de 80 millones de dólares en el año 1.990, según Delphi Consulting Group el mercado en el 1.996 ha sido de 700 millones de dólares. Más de 60 pymes están trabajando en este mercado. Ninguna de ellas lo domina y no es capaz de conseguir estándares de facto. ■

*RAMON NOGALES MORAN

Ramón Nogales Morán es Director de
Desarrollo Tecnológico de Telefónica
Sistemas.