

Rendimiento de 8 sistemas de recuperación de información del espacio web español

Por Carlos Benito Amat

Carlos Benito Amat, doctor en medicina y titulado superior del CSIC, participó en el mantenimiento de la base de datos Índice Médico Español. Ha sido responsable de la unidad de Documentación de Radiotelevisión Valenciana. Desde 2001 dirige la unidad de Documentación Científica del Instituto de Agroquímica y Tecnología de Alimentos (IATA-CSIC).



Resumen: Los sistemas AltaVista, EnlaWeb, Lycos, Olé/Terra, Ozú, Sol, Ya y Yahoo se evaluaron en términos de:

- 1) su ajuste al espacio web español y
- 2) su rendimiento en la recuperación de información en dicho espacio.

Para la determinación del ajuste se eligieron como indicadores su cobertura general, su aporte específico, la proporción de resultados erróneos (enlaces perdidos) y la accesibilidad de los documentos web españoles a través de los ocho sistemas. La determinación del rendimiento de la recuperación se basó en el cálculo de la precisión y la exhaustividad relativa de los resultados de búsquedas realizadas en el primer trimestre de 2003. La elección de los temas de búsqueda y el juicio de relevancia corrió a cargo de usuarios reales. El 12,4% de los resultados de búsqueda condujeron

a enlaces erróneos. El 76% de los resultados obtenidos se obtuvieron a partir de un único sistema. El rendimiento de los sistemas, expresado en términos de exhaustividad y precisión, es reducido, con valores promedio de exhaustividad situados entre el 7 (AltaVista) y el 14% (Ozú) y de precisión entre el 9 (Sol) y el 30% (Ozú). Sólo Yahoo muestra un comportamiento típico con relación inversa entre los valores a lo largo de los 20 resultados de búsqueda. Los sistemas restantes muestran invariablemente un ascenso en las cifras de precisión a partir del segundo o tercer resultado de búsqueda, lo que abunda en la idea de defectos en el algoritmo de ordenación de los resultados obtenidos.

Palabras clave: Recuperación en internet, Motores de búsqueda, Rendimiento de la recuperación.

Title: Performance of eight information retrieval systems in the Spanish web

Abstract: Eight search engines (AltaVista, EnlaWeb, Lycos, Olé/Terra, Ozú, Sol, Ya and Yahoo) were evaluated in terms of the Spanish web space and their performance in retrieving information from this space. Indicators chosen were their relative coverage, specific offering, proportion of dead links and accessibility of Spanish websites. Performance was determined by relative recall and precision in retrieval during the first quarter in 2003. Search topics and relevance of results were determined by the end users. 12,4% of the searches led to dead links and 76% of the pages were returned by only a single system. System performance, expressed in terms of recall ranged from 7% (AltaVista) to 14% (Ozú) and precision between 9% (Sol) and 30% (Ozú). Only Yahoo displayed typical inverse relationship between recall and precision figures. The rest of the systems invariably showed an increase in precision figures starting with the second or third search result, suggesting problems with the sorting algorithm.

Keywords: Web retrieval, Search engines, Retrieval performance.

Benito Amat, Carlos. "Rendimiento de 8 sistemas españoles de recuperación de información del espacio web español". En: *El profesional de la información*, 2005, septiembre-octubre, v. 14, n. 5, pp. 335-346.

Introducción

Las razones que justifican el análisis de los sistemas de recuperación de información distribuida en internet son variadas y de gran peso (Bellardo Hahn, 1998; Travis, 1998):

—Ofrecen la oportunidad de estudiar las posibilidades de consulta en texto libre y recuperación de in-

gentes colecciones de documentos (electrónicos), una situación que parece acercarse en el entorno de la información ligada al conocimiento y al que no han de ser ajenas las iniciativas de "colecciones digitales" ahora incipientes.

—El acceso de usuarios no cualificados sin intermediación alguna, los "casual users" frente a "end users", brinda la oportunidad de estudio de sus hábitos

Artículo recibido el 16-05-05
Aceptación definitiva: 20-06-05



de consulta y recuperación y la extracción de conclusiones de extrema importancia práctica.

—Las características de los documentos y textos a recuperar (inestabilidad, desaforado crecimiento, heterogeneidad máxima) plantean una serie de desafíos especialmente interesantes para el diseño de los sistemas, sus mecanismos de actualización y su arquitectura de almacenamiento.

—Todas estas motivaciones multiplican su valor cuando se plantea el establecimiento de mecanismos de recuperación de información distribuida en entornos académicos, de investigación o, en general, ligados al conocimiento. Esto incluye la selección o el diseño de sistemas de recuperación de los contenidos en intranets y redes corporativas especializadas en el marco de lo que se ha dado en llamar *ECM* (*enterprise content management*), y también en la aplicación de estas tecnologías a la recuperación en almacenes de documentos electrónicos (Prudlo, 2005).

«El empleo de la metodología tradicional en la evaluación de los sistemas de recuperación en internet ha suscitado algunas críticas»

Varios autores han tratado el tema en la literatura española. **López Alonso** y **Mares Marín** (1996) presentaron un trabajo original donde evaluaban el rendimiento de *Excite*, *Lycos*, *OpenText* y *Savvy Search*. **Marcos Mora** (1998b) mostraba de forma comparativa las funcionalidades de 6 de los motores más utilizados en aquel momento (*AltaVista*, *Excite*, *HotBot*, *InfoSeek*, *Lycos* y *OpenText Index*). **Sánchez Montero** (1997) trató el problema en el marco de su trabajo sobre diseño de contenidos para intranets corporativas. Más tarde **Baró** (1997) aportó una relación, con descripciones muy genéricas de algunos sistemas de búsqueda en el entorno world wide web. La revisión de **Senso** (1998) y la aportación al tema de **Marcos Mora** (1998a) resultan más descriptivas de los servicios asociados que de los propios sistemas. Otro trabajo destaca por el número de sistemas analizados y por incluir entre ellos, por vez primera, a 10 sistemas nacionales (**Maldonado Martínez**; **Fernández Sánchez**, 1998). En él, las autoras presentaron un estudio pormenorizado de las “características documentales” de 10 sistemas internacionales y 10 españoles. Por características documentales las autoras entendían el estudio del esquema de datos, las posibilidades de recuperación y la presentación de resultados de cada uno. Como conclusión emitían un juicio sobre la operatividad de todos y un análisis global de ambos grupos.

Nota:

El profesional de la información considera que este artículo es de interés metodológico a pesar de que los estudios en que se basa se realizaron a principios del año 2003 con los buscadores entonces existentes, algunos de los cuales ya han desaparecido.

Los trabajos españoles de tesis han abordado este tema de estudio de forma desigual. **Olvera Lobo** (1999) se centró en 10 servicios de búsqueda internacionales: *AltaVista*, *Excite*, *Hotbot*, *InfoSeek*, *Lycos*, *Magellan*, *OpenText*, *WebCrawler*, *WWWorm* y *Yahoo*. Por su parte, **Martínez Méndez** (2001) emplea *AltaVista*, *Alltheweb* (*FAST*), *Google*, *MSN*, *Terra* y *WISEnut*. Hasta la tesis de **Amat** (2004), los diseños no incluían a usuarios reales en la formulación de peticiones y el juicio de relevancia. **Alonso Berrocal** (2000) no realiza experimento de recuperación alguno. Y, sin embargo, “de entre los recursos de internet, el tipo más usado por las pequeñas empresas españolas (94%) corresponde a los buscadores” (*PricewaterhouseCoopers*, 2004).

Si a las motivaciones antes enumeradas se añade otra, basada en el valor de la información relevante y la consideración “patrimonialista” de los recursos distribuidos en un medio social y cultural específico (un argumento ampliamente esgrimido en lo tocante, por ejemplo, a la producción y el consumo de información científica “nacional”) se reconocerá el interés del estudio de los sistemas dedicados a la recuperación en el espacio web de España.

Este estudio comprende el análisis de algunas características de los sistemas analizados y la determinación de su rendimiento atendiendo a los indicadores más usuales: precisión y exhaustividad.

El empleo de la metodología tradicional a la evaluación de los sistemas de recuperación en internet ha suscitado algunas críticas (**Harter**; **Hert**, 1997). En su mayoría se centran en la dificultad de determinar el número total de documentos relevantes en las colecciones. Sin embargo, la aplicación del concepto de relevancia y sus indicadores asociados, exhaustividad y precisión, se considera obligada (**Oppenheim**; **Morris**, **Mcknight**; **Lowley**, 2000). Así, la práctica totalidad de los trabajos agrupados por **Martínez Méndez** y **Rodríguez Muñoz** (2003) en el apartado de “estudios experimentales” de la recuperación de información en internet emplean el concepto y sus indicadores en su evaluación.

Tan importante como la determinación de los indicadores y medidas a emplear es el ajuste de las condiciones experimentales. **Michael Gordon** y **Praveen Pathak** (1999), autores de uno de los más sólidos trabajos de evaluación, enumeran siete requisitos básicos:

—Las búsquedas se deben basar en necesidades de información genuinas. Que los experimentadores determinen los temas puede introducir errores que, por ejemplo, favorezcan a un sistema en perjuicio de otros. Además es necesario ajustar el experimento a la increíble diversidad de las necesidades informativas de los usuarios finales y casuales.

—Cuando se trata de identificar información relevante a una necesidad, es necesario que la persona que la formula exprese también cuanta información contextual sea posible: las listas de palabras clave o expresiones formales (por ejemplo booleanas) sólo pueden reflejar el tema de forma aproximada. Si acompañan a la expresión natural del tema de búsqueda, en cambio, pueden contribuir a despejar ambigüedades.

—Se debe realizar un número suficiente de búsquedas para obtener resultados significativos.

—La prueba debe incluir los sistemas principales.

—Se debe explotar cada sistema empleando sus características distintivas. No hay por qué emplear el mismo perfil en todos los sistemas analizados.

—El juicio de relevancia debe correr a cargo del propio usuario. Si lo realiza el experimentador puede originar errores debidos a su falta de familiaridad con el tema o a su ignorancia de las necesidades y conocimiento previo reales.

—El buen desarrollo experimental exige la obtención de medidas significativas de rendimiento mediante el empleo de un buen diseño (por ejemplo, la presentación de los resultados en orden aleatorio); mediante el ajuste a indicadores habituales (como los de exhaustividad y precisión) y a través del empleo de análisis estadístico que informe de la significación de las diferencias halladas entre los sistemas.

A lo largo del presente experimento se han seguido estos requisitos básicos y otros adicionales revisados por **Olvera Lobo** (2000). En el siguiente apartado se introducen los conceptos e indicadores empleados en este estudio. Posteriormente se detallan las condiciones del experimento: expresión de las necesidades informativas, sistemas seleccionados, desarrollo de las búsquedas y evaluación de los usuarios. A continuación se presentan los resultados.

Relevancia, exhaustividad y precisión como criterios de rendimiento

En una reciente revisión que destaca por su claridad, **Gómez Díaz** (2003) resume el procedimiento tradicional de evaluación del rendimiento de la recuperación, sus supuestos básicos, sus conceptos subyacentes, sus indicadores principales y su metodología (**Gómez Díaz**, 2003). El tratamiento de **Abad García** (1997) es anterior, pero sus grandes líneas son coincidentes. Este apartado se apoya en ambas aportaciones y matiza la aplicación de los indicadores al estudio de los sistemas. La imprecisión terminológica y la ambigüedad de algunos conceptos se han limitado atendiendo a la discusión de los mismos realizada por **Lancaster** y **Warner** (1993).

«Toda petición se sometió a cada sistema repetidamente, empleando diversos perfiles y formulaciones alternativas, hasta hallar resultados óptimos»

A grandes rasgos, la relevancia mide la proximidad entre un documento y la formulación de una petición o expresión de la necesidad informativa. Una valoración meramente mecánica de la relevancia se obtiene mediante comparación entre un perfil de búsqueda (la expresión sintáctica o lógica que refleja la demanda) y los términos que reflejan el contenido informativo de los documentos. Si embargo, “para evaluar un sistema de información real, con usuarios reales que formulan peticiones reales basadas en genuinas necesidades de información, es imperativo determinar la medida en que el servicio satisface las necesidades informativas de los usuarios” (**Lancaster; Warner**, 1993). La influencia de los modelos probabilísticos de recuperación se ha extendido hasta considerar la relevancia una variable continua que puede adoptar valores intermedios entre el 1 y el 0. Ello no obsta para que, en muchos casos, se emplee una escala discreta o se haga en otros un manejo binario del concepto.

Basándose en el cálculo de relevancia, es posible determinar el valor de los índices de precisión y exhaustividad, indicadores de rendimiento de la recuperación:

—La exhaustividad mide la proporción de documentos relevantes que son recuperados. Corresponde al cociente entre el número de documentos relevantes recuperados y el total relevantes existentes en la colección.

—La precisión mide la proporción de documentos recuperados que son relevantes. Refleja la eficacia de

las búsquedas y pone en relación el número de documentos relevantes recuperados con el número total de documentos recuperados.

El cálculo de la exhaustividad presenta algunas dificultades en los sistemas tradicionales, y más aún en los sistemas de recuperación en internet. En ambos casos, la determinación del denominador de la expresión (número total de documentos relevantes existentes en la colección) se estima de forma indirecta. Se habla entonces de exhaustividad relativa.

«En ningún caso un mismo resultado fue cubierto por todos los sistemas y sólo en una ocasión uno fue devuelto por siete de ellos»

Las más recientes críticas al empleo de indicadores tradicionales proponen ciertas modificaciones, además de añadir nuevas variables de análisis (Vaughan, 2004) relacionadas con la estabilidad de las colecciones y el procedimiento de ordenación de resultados.

El modelo tradicional de evaluación de la recuperación de información parte de los experimentos realizados en los años 60 en Cranfield y se prolonga a partir de 1992 en la serie de conferencias anuales sobre recuperación textual *TREC (Text Retrieval Conferencas)*. En la actualidad, el *National Institute of Standards and Technology* estadounidense proporciona colecciones documentales y temas de búsqueda que cada sistema experimental procesa. Luego los resultados se comparan con controles (colecciones de referencia), constituidos por listas de documentos ordenados por relevancia. No es posible emplear este modelo puesto que hasta hace muy poco (Bailey; Craswell; Hawking, 2003) no se ha podido contar con uno de los elementos esenciales: la colección de evaluación o colección de referencia que, por ende, contara con relaciones estructurales similares a las de los documentos hipertextuales del espacio web (Gurrin; Smeaton, 2004). Una reciente alternativa, que propone el análisis automático del rendimiento de los sistemas (Can; Nuray; Sevdik, 2004) parece demasiado incipiente, pero no está exenta de interés.

Otras medidas

Aunque exhaustividad y precisión son los indicadores fundamentales en la determinación del rendimiento de los sistemas de recuperación, existen otros que permiten compararlos. De ellos, interesa la cobertura, el grado en que los sistemas incorporan el espacio web y algunos indicadores relacionados: solapamiento y aporte específico (Abad García, 1997). Una medida complementaria es la accesibilidad de los re-

sultados, es decir, la presencia de los documentos hallados en todos los sistemas de búsqueda que, teóricamente, los incorporan. Otra, relacionada con el grado de actualización de los sistemas, es el número de resultados que cada uno devuelve de forma duplicada.

1. Método

a. Indicadores y medidas empleados

A lo largo del presente experimento se determinan el solapamiento relativo y el aporte específico de cada sistema. El índice de solapamiento relativo se calcula mediante la relación entre el número de documentos recuperados simultáneamente por dos o más sistemas y el total de documentos recuperados por esos mismos sistemas (Abad García, 1997). La expresión habitual es:

$$\text{Solapamiento entre A y B} = \frac{N(A \cap B)}{N(A \cup B)}$$

El aporte específico se obtiene mediante la relación entre el número de documentos recuperados exclusivamente por un sistema y el número total de documentos recuperados (Abad García, 1997). Ambas medidas se han obtenido para cada sistema y cada búsqueda y, con posterioridad, se han promediado los valores calculados en cada consulta.

Igualmente, la determinación del rendimiento en términos de exhaustividad (E) y precisión (P) se ha realizado para cada uno de los sistemas, búsquedas y niveles de resultados, siguiendo el método aplicable a conjuntos de resultados ordenados (Salton; McGill, 1983).

Se ha calculado la exhaustividad relativa tomando como denominador de la ecuación la suma de documentos juzgados relevantes en cada búsqueda por el total de sistemas.

Se ha empleado una determinación binaria de la relevancia, es decir, los documentos se han considerado relevantes o no. El juicio de relevancia ha sido subjetivo, expresado por los usuarios finales.

b. Expresión de las necesidades informativas

20 estudiantes de la asignatura "Sistemas de información en red" de la diplomatura en biblioteconomía y documentación de la Universitat de València cumplieron un formulario de búsqueda diseñado para que expresaran la información que solicitaban. Las peticiones eran muy variadas, por ejemplo "la aplicación del método Nordoff-Robins en dificultades motrices", "biografías de compositores clásicos", "homosexualidad en la antigua Grecia", "el mundo de la Tierra Media de Tolkien: mapas y planos", los "tipos de conexión a internet que se ofrecen en todo el mundo", etc. Además del tema de búsqueda expresado en lenguaje

natural, se pidió a los participantes la posible traducción a una formulación booleana equivalente de la combinación de términos, aquellos otros que se deberían eliminar para evitar resultados erróneos y, además, se requirió que enumeraran los sistemas que empleaban habitualmente en sus búsquedas en la Red.

c. Selección de los sistemas

Siete herramientas de recopilación automática y un directorio temático fueron objeto de estudio: AltaVista, HispaVista, Lycos, Olé/Terra, Ozú, Sol, Yahoo y el directorio EnlaWeb. Su elección se basó en que:

—fueran operativos entre 2002 y 2004

—arrojaran suficientes resultados en respuesta a una batería de búsquedas previa: se desecharon aquellos que ofrecieron menos de 5 resoluciones

—atendiendo a los resultados de análisis previos (Amat, 2002), incluyeran, entre otras cosas, pruebas de “popularidad”.

Acerca de Yahoo hay que indicar que cuenta con una base de datos de recopilación automática además del directorio clasificatorio que inicialmente constituía. En este caso las búsquedas se realizaron en el primero de los componentes.

d. Desarrollo de las búsquedas

Todas fueron realizadas por el autor en las últimas semanas de marzo de 2003. La expresión textual de la necesidad de información y el resto de las expresiones contenidas en los formularios podían resolverse de diversas formas. Se empleó la técnica de perfil mejorado (best search) de forma que toda petición se sometió a cada sistema repetidamente, empleando diversos perfiles y formulaciones alternativas, hasta hallar resultados óptimos.

En el momento en que no se pudieron mejorar los resultados de todas las búsquedas en cada sistema se registraron, en su correspondiente orden, las urls de los 20 primeros resultados, siguiendo el criterio de **Leighton y Srivastava** (1999). Posteriormente se trasladaron a hojas de cálculo y se compactaron. El objetivo era evitar que los usuarios se vieran obligados a examinar repetidamente un mismo resultado procedente de dos o más sistemas diferentes. Las urls de cada conjunto de resultados se ordenaron aleatoriamente y se ocultó el sistema o sistemas del que procedía cada resultado.

e. Evaluación por los usuarios

Cada alumno recibió un archivo de hoja de cálculo con la lista de resultados desordenados. El número de urls que contenía rondaba en todos los casos el número de 100. Los participantes fueron instruidos sobre la expresión de su juicio de relevancia. En concreto se hizo especial énfasis en que el juicio se ciñera a la página visualizada y que no se extendiera a páginas conexas o a la sede que las albergaba. Conectaron con cada url y tras su examen determinaron si su contenido era pertinente (1) o no (0) a su necesidad de información. La imposibilidad de visualizar un documento se anotó como error (E) y, a efectos de cálculo, su valor se consideró nulo. Así mismo se anotaron como valores nulos las repeticiones de resultados en un mismo sistema.

f. Resultados y discusión

Los cambios en los sistemas de recuperación de información en internet son continuos. Afectan a su interfaz, a sus mecanismos de cobertura, a la representación de los documentos y su indización, así como a sus procedimientos de recuperación y ordenación. Los ha-



	B14	B16	B04	B12	B01	B09	B19	Bem	B11	B08	B05	Total por sistema	%
<i>AltaVista</i>	0	4	2	4	1	7	5	3	2	5	4	37	16,97
<i>EnlaWeb</i>	1	2	3	2	3	0	3	7	0	0	0	21	9,63
<i>Lycos</i>	0	0	1	1	1	0	7	2	1	1	1	15	6,88
<i>Olé</i>	2	0	4	2	6	1	3	2	5	2	1	28	12,84
<i>Ozú</i>	0	1	3	0	3	4	0	2	1	2	4	20	9,17
<i>Sol</i>	1	2	9	5	3	10	2	2	6	9	0	49	22,48
<i>Ya</i>	1	2	1	2	1	2	7	3	2	1	1	23	10,55
<i>Yahoo</i>	1	0	2	1	4	4	4	1	3	1	4	25	11,47
Total por búsqueda	6	11	25	17	22	28	28	18	27	21	15	218	

Tabla 1. Distribución por sistemas y por búsquedas evaluadas de los errores de conexión hallados

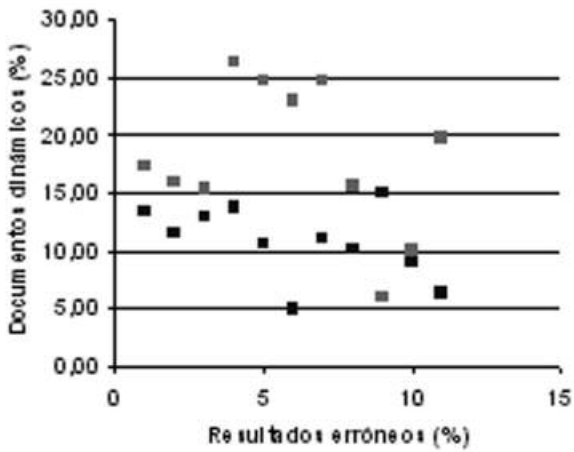


Figura 1. Correlación entre las proporciones de documentos dinámicos obtenidos en cada búsqueda y de errores devueltos

	Únicos	1	2	3	4	5	6	7
b01	127	100	21	6	0	0	0	0
b02	91	58	18	11	1	1	2	0
b03	113	90	13	4	4	1	1	0
b04	163	138	23	2	0	0	0	0
b05	57	25	15	7	6	3	1	0
b06	123	104	11	7	1	0	0	0
b07	98	79	9	8	2	0	0	0
b08	85	51	17	14	1	2	0	0
b09	122	94	18	10	0	0	0	0
b10	138	124	10	1	2	1	0	0
b11	109	83	13	7	2	4	0	0
b12	109	80	18	10	0	1	0	0
b13	97	60	29	5	2	1	0	0
b14	100	69	14	9	6	1	0	1
b15	109	83	13	7	2	4	0	0
b16	110	83	20	6	1	0	0	0
b17	102	70	16	9	4	3	0	0
b18	114	94	16	2	2	0	0	0
b19	142	109	24	8	1	0	0	0
b20	88	68	10	5	4	1	0	0
Total	2197	1662	328	138	41	23	4	1

Tabla 2. Número de sistemas de procedencia de los resultados de búsqueda

Ilazgos aquí expuestos son, en consecuencia, limitados y provisionales.

Aunque se realizaron las búsquedas correspondientes a los 20 temas propuestos en los 8 sistemas seleccionados (un total de 160), sólo se dispuso de los juicios de relevancia de 11 usuarios, lo que supuso 88 búsquedas evaluadas.

El experimento arrojó cinco series de resultados:

—El número de errores contabilizados al intentar acceder a las páginas resultantes.

—Una estimación del grado de cobertura y solapamiento de los sistemas.

—Una medición complementaria de la anterior, que determina la accesibilidad de los resultados.

—El rendimiento de los sistemas estudiados, en términos de exhaustividad y precisión.

—Algunos datos adicionales sobre el tipo de documentos recuperados y su relación con los errores.

2. Errores de conexión

La tabla 1 resume los errores de conexión hallados. Los resultados se distribuyen entre los sistemas analizados (filas) y las búsquedas evaluadas (columnas B14 a B05). Globalmente, el 12,4% de los resultados de búsqueda condujeron a páginas erróneas. Sólo en una ocasión el acceso se vio imposibilitado por la exigencia de identificación y contraseña de usuario. En el resto, los intentos de acceso devolvieron un error 404.

Para averiguar si existía relación entre el número de resultados erróneos y la naturaleza de los documentos recuperados, se compararon las proporciones respectivas. Las páginas dinámicas se identificaron por la extensión de los documentos (asp, php y otras) o por la inclusión en la url de una función CGI. Como término medio se halló una proporción del 10,8% en las búsquedas evaluadas. En este mismo conjunto de búsquedas (ver tabla 1) la proporción de errores superaba el 12%. La figura 1 ilustra la falta de correlación existente entre ambas proporciones.

Esta tasa de errores no se relaciona con el dinamismo de los documentos recuperados y, así, parece atribuirse más a un ritmo inadecuado de actualización de las bases de datos de cada sistema. La proporción de re-

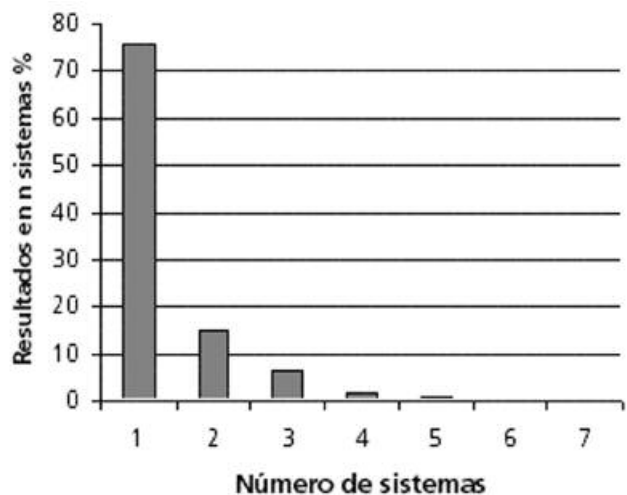


Figura 2. Accesibilidad de los documentos resultantes de las búsquedas

	<i>AltaVista</i>	<i>EnlaWeb</i>	<i>Lycos</i>	<i>Olé</i>	<i>Ozú</i>	<i>Sol</i>	<i>Ya</i>	<i>Yahoo</i>	Total b
b1	8	10	9	4	10	11	4	4	60
b2	0	0	0	0	0	0	0	0	0
b3	0	0	0	2	0	0	3	0	5
b4	0	0	0	0	0	0	0	0	0
b5	0	0	0	0	0	0	0	0	0
b6	0	0	0	0	0	0	0	0	0
b7	0	0	0	0	0	0	0	0	0
b8	0	0	0	0	0	0	0	0	0
b9	0	0	0	0	1	0	0	0	1
b10	0	0	0	0	0	0	0	0	0
b11	0	0	0	0	0	0	1	0	1
b12	0	0	0	0	0	0	0	0	0
b13	2	1	2	1	1	0	0	5	12
b14	0	0	0	2	0	4	0	6	12
b15	2	0	0	0	0	0	1	0	3
b16	0	0	0	0	0	0	0	0	0
b17	0	0	0	0	0	0	0	0	0
b18	0	0	0	0	0	0	0	0	0
b19	0	0	0	0	0	1	1	0	2
b20	0	0	0	0	0	0	0	1	1
Total s	12	11	11	9	12	16	10	16	

Tabla 3. Duplicidad de contenidos en los sistemas expresada por el número de resultados idénticos

sultados duplicados podría avalar esta conclusión: el sistema con más errores (Sol) también muestra la mayor proporción de duplicados (10%), junto con Yahoo. Sin embargo, no existe correlación significativa entre la proporción de errores y de resultados duplicados.

Accesibilidad de las páginas halladas

Entendemos accesibilidad como la presencia de los documentos hallados en todos los sistemas de búsqueda que, teóricamente, los incorporan. La tabla 2 ofrece algunas cifras absolutas en relación con la accesibilidad de los resultados de las búsquedas realizadas. Se han tenido en cuenta todas las consultas llevadas a cabo, no sólo las evaluadas. Para cada una de las búsquedas (b01 a b20) se ofrece el número de documentos resultantes y, en las columnas siguientes, la cantidad de sistemas (entre 1 y un máximo de 7) que devolvieron esos resultados. Puede apreciarse que sólo en una ocasión (búsqueda 14) un resultado fue devuelto por siete de los sistemas estudiados y en ningún caso todos los sistemas cubrían un mismo recurso.

La figura 2 representa gráficamente los datos, empleando una escala porcentual en el eje de valores. A pesar de que los sistemas analizados parten con el objetivo común de proporcionar acceso a los documentos del espacio web en España, casi el 76% de los resulta-

dos obtenidos se encuentran cubiertos de forma exclusiva por un único sistema entre ocho.

La tabla 3 refleja la duplicidad de contenidos de las bases de datos de los sistemas estudiados (un indicador relacionado con su grado de actualización). En ella se distribuyen las ocasiones en que los sistemas han devuelto más de una vez un resultado idéntico para determinada búsqueda. Las columnas de totales contienen la suma de resultados coincidentes (en un mismo sistema) para cada conjunto de resultados de búsqueda (total b) y para cada sistema a lo largo de las 20 búsquedas (total s).

Solapamiento e índice de aporte específico

El solapamiento de los sistemas se refiere al número de resultados coincidentes procedentes de los ocho sistemas estudiados. Se ha determinado búsqueda a búsqueda entre todos los sistemas tomados dos a dos y se ha promediado entre las 20 consultas realizadas.

Así, por ejemplo, la consulta número 2 contiene 16 coincidencias entre *Lycos* y *Olé/Terra*, lo que arroja un grado de solapamiento de $16/40 = 0,40$ ó 40%. Los valores promedios de solapamiento entre todos los siste-

	AltaVista	EnlaWeb	Lycos	Olé	Ozú	Sol	Ya	Yahoo
AltaVista		0,001 ± 0,001	0,03 ± 0,09	0,04 ± 0,04	0,04 ± 0,06	0,01 ± 0,03	0,13 ± 0,44	0,06 ± 0,07
EnlaWeb			0,01 ± 0,02	0,01 ± 0,02	0,01 ± 0,03	0,01 ± 0,02	0,01 ± 0,02	0,00 ± 0,01
Lycos				0,20 ± 0,18	0,03 ± 0,04	0,01 ± 0,02	0,20 ± 0,13	0,05 ± 0,06
Olé					0,06 ± 0,06	0,01 ± 0,02	0,21 ± 0,15	0,07 ± 0,07
Ozú						0,01 ± 0,02	0,05 ± 0,06	0,19 ± 0,13
Sol							0,01 ± 0,02	0,02 ± 0,05
Ya								0,05 ± 0,06
Yahoo								0,05 ± 0,06

Tabla 4. Solapamiento entre los sistemas estudiados (datos expresados con media ± desviación típica)

mas, acompañados de la desviación típica de cada distribución, se expresan en la tabla 4.

El valor máximo corresponde a los sistemas *Ya* y *Olé*, cuyos resultados aparecen solapados en un 21%. En general, es *Ya* quien presenta mayor grado de solapamiento con el resto, mientras que *EnlaWeb* (el único directorio evaluado) y *Sol* muestran los menores valores en conjunto.

El indicador de aporte específico, complementario al de solapamiento, se obtiene mediante la relación porcentual entre el número de documentos recuperados por un sistema exclusivamente y el total de documentos recuperados en una búsqueda determinada. Se ha calculado para cada uno de los ocho sistemas y para cada una de las 20 búsquedas. La figura 3 expresa gráficamente los promedios hallados, con valores máximo para *AltaVista* (10,65 ± 2,96) y mínimo para *Olé* (3,81 ± 3,31)

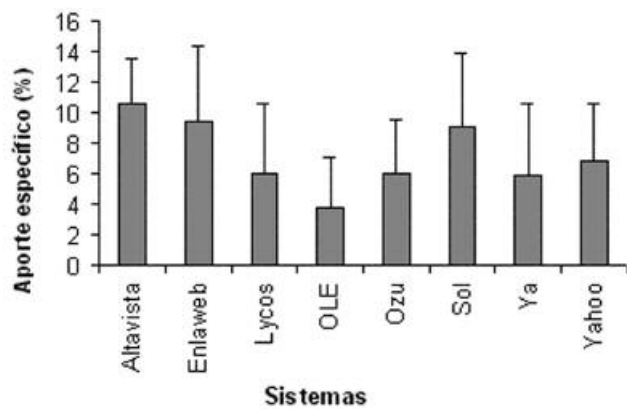


Figura 3. Aporte específico de cada sistema expresado en promedio y desviación típica

Del total de 3.200 resultados de búsqueda, el número de documentos resultantes únicos, descontadas las apariciones repetidas, fue de 2.197. De ellos, algo más del 75% fueron obtenidos por sólo un sistema, ob-

servación que, junto con las anteriores, permite calificar de deficientes los procedimientos de recopilación que utilizan y, en consecuencia, su cobertura. El hecho de que ningún resultado de búsqueda haya sido cubierto por los 8 sistemas estudiados ha de contrastarse con la especial naturaleza del sistema *EnlaWeb*: se trata de un directorio que incorpora sedes, no páginas de niveles inferiores, lo que dificulta que las urls ofrecidas coincidan con las de los sistemas de recopilación automática. No es de extrañar que su aporte específico —la proporción de documentos recuperados exclusivamente por este directorio— sea superior al de todos los sistemas excep-

AltaVista		EnlaWeb		Lycos		Olé		Ozú		Sol		Ya		Yahoo	
E	P	E	P	E	P	E	P	E	P	E	P	E	P	E	P
0,01	0,36	0,00	0,09	0,00	0,18	0,00	0,09	0,01	0,27	0,00	0,09	0,00	0,18	0,03	0,45
0,02	0,27	0,00	0,05	0,01	0,27	0,01	0,14	0,02	0,41	0,01	0,14	0,01	0,18	0,05	0,32
0,03	0,33	0,01	0,12	0,02	0,21	0,02	0,21	0,03	0,39	0,01	0,09	0,02	0,24	0,05	0,30
0,04	0,39	0,01	0,16	0,02	0,18	0,03	0,23	0,04	0,34	0,02	0,09	0,03	0,27	0,06	0,32
0,05	0,35	0,01	0,15	0,03	0,24	0,03	0,24	0,07	0,31	0,03	0,09	0,03	0,24	0,07	0,31
0,05	0,32	0,02	0,15	0,04	0,23	0,03	0,20	0,08	0,29	0,04	0,09	0,03	0,20	0,08	0,27
0,05	0,29	0,02	0,16	0,04	0,23	0,04	0,18	0,10	0,32	0,04	0,08	0,03	0,18	0,09	0,29
0,05	0,28	0,03	0,17	0,05	0,24	0,04	0,17	0,13	0,33	0,04	0,08	0,04	0,18	0,10	0,30
0,06	0,25	0,03	0,16	0,05	0,24	0,04	0,17	0,16	0,33	0,04	0,09	0,04	0,16	0,10	0,30
0,08	0,25	0,03	0,15	0,06	0,24	0,05	0,17	0,17	0,33	0,04	0,09	0,07	0,18	0,11	0,31
0,08	0,23	0,04	0,14	0,06	0,21	0,06	0,17	0,18	0,31	0,04	0,09	0,07	0,17	0,12	0,30
0,08	0,22	0,04	0,13	0,07	0,22	0,06	0,17	0,18	0,29	0,04	0,08	0,08	0,17	0,13	0,30
0,08	0,21	0,04	0,14	0,08	0,22	0,06	0,15	0,19	0,28	0,06	0,09	0,08	0,17	0,13	0,31
0,09	0,21	0,05	0,15	0,08	0,21	0,06	0,15	0,20	0,28	0,06	0,09	0,08	0,16	0,14	0,31
0,10	0,22	0,05	0,15	0,08	0,20	0,06	0,14	0,20	0,32	0,06	0,09	0,08	0,16	0,15	0,26
0,10	0,21	0,06	0,15	0,08	0,19	0,07	0,15	0,21	0,26	0,06	0,09	0,08	0,16	0,16	0,25
0,11	0,22	0,06	0,14	0,08	0,19	0,07	0,14	0,22	0,25	0,06	0,09	0,09	0,17	0,16	0,25
0,13	0,23	0,06	0,14	0,08	0,19	0,09	0,16	0,22	0,25	0,07	0,09	0,10	0,18	0,17	0,23
0,14	0,24	0,06	0,14	0,09	0,18	0,10	0,16	0,23	0,25	0,07	0,09	0,11	0,18	0,20	0,23
0,14	0,24	0,07	0,14	0,09	0,19	0,11	0,17	0,23	0,24	0,07	0,09	0,11	0,18	0,20	0,23

Tabla 5. Valores promedio de E-P de los sistemas analizados

tuando *AltaVista*. También su bajo nivel de solapamiento es indicativo.

Los datos sobre la cobertura de las sedes y la accesibilidad de los resultados aportados aquí son coherentes con hallazgos anteriores (Amat, 2002; 2003).

Exhaustividad y precisión de los sistemas

La tabla 5 muestra los valores promedio de exhaustividad (E) y precisión (P) de cada sistema en el conjunto de búsquedas evaluadas. Los resultados se detallan para cada nivel de respuesta, del 1 al 20. El cálculo de la exhaustividad toma como denominador el número total de documentos relevantes identificados por los usuarios en el contexto global de las búsquedas en los 7 sistemas. Las cifras son B14=22, B16=9, B04=28, B12=30, B01=33, B09=28, B19=4, Bem=46, B11=72, B08=3 y B05=49. La figura 4, por su parte, compara los trazados correspondientes a estas distribuciones.

El rendimiento de los sistemas, expresado en términos de exhaustividad y precisión, es reducido, con valores promedio de exhaustividad situados entre el 7% (*AltaVista*) y el 14% (*Ozú*) y de precisión entre el 9% (*Sol*) y el 30% (*Ozú*). Sólo *Yahoo* muestra un comportamiento típico con relación inversa entre los valores a lo largo de los 20 resultados de búsqueda. Los sistemas restantes muestran invariablemente un ascenso en las cifras de precisión a partir del segundo o tercer resultado de búsqueda, lo que abunda en la idea de defectos en el algoritmo de ordenación de las respuestas ofrecidas. La apariencia “contraída” del trazado correspondiente a *EnlaWeb* se justifica por su naturaleza de directorio y su consecuente mecanismo clasificatorio: a partir del ascenso inicial en los valores de precisión, se produce una estabilización del cociente. En este sentido, la horizontalidad de extensos tramos en algunos trazados (*Ozú* y *Yahoo*), con estancamiento de los valores de precisión y el posterior remonte, vuelven a ser indicativos del funcionamiento anómalo del algoritmo de ordenación de resultados.

Para avanzar en el conocimiento de estos resultados fue necesario analizar los componentes implicados y su interacción: el tema de búsqueda, aquí denominado indistintamente tema o búsqueda; el sistema y el orden de aparición de los resultados.

Se llevó a cabo un análisis de la varianza (*anova*) de estos tres factores. Los resultados indican que las diferencias de rendimiento dependen en gran medida del sistema empleado en las búsquedas. Por lo que respecta a la exhaustividad, el orden de presentación de los resultados también influye. El factor denominado “tema”, que en realidad se refiere tanto al tema de búsqueda

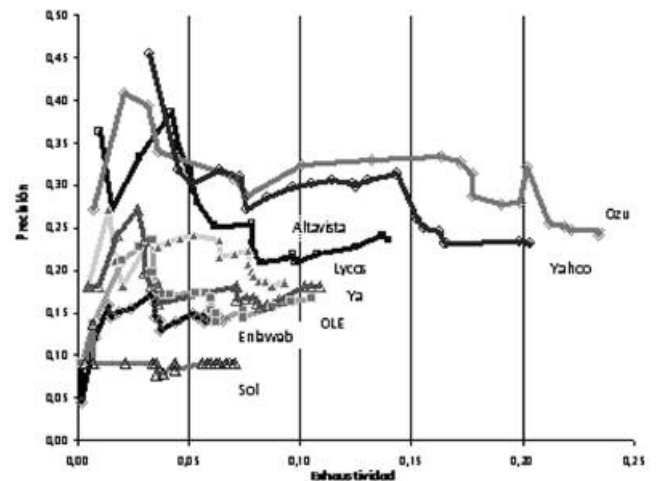


Figura 4: Diagrama general con los valores promedio de exhaustividad vs. precisión para los 8 sistemas analizados

queda como al usuario que lo ha propuesto y ha juzgado los resultados, tiene cierta importancia en relación con la exhaustividad, pero sobre todo es determinante de la precisión alcanzada. Así mismo, tiene relevancia la interacción entre sistema y tema. En otras palabras: el juicio de relevancia y el hecho de que determinados temas de búsqueda se ajusten más a la especial mecánica de la recuperación de un sistema son factores determinantes del rendimiento global.

La prueba de comparación múltiple de **Tukey (Wiener, 1962)** indica que existen diferencias significativas entre los sistemas. En relación con el tema de búsqueda se observa que la mejor precisión se obtiene en las consultas 9 y 11, mientras que las 4 y 7, con cifras muy bajas en precisión, arrojan mejores resultados en exhaustividad (figura 4). Los resultados de las búsquedas 2 y 10 son claramente inferiores tanto en precisión como en exhaustividad, aunque se aprecia que este último indicador se mueve en un reducido rango, entre el 5% y el 10%.

La interacción entre sistema y búsqueda es muy importante. *Ozú* y *Yahoo* muestran los mejores resultados de precisión en las búsquedas 3 y 11. *EnlaWeb* sólo destaca en la búsqueda 9, donde los resultados de todos los sistemas experimentan una mejoría en su precisión. Podría referirse este hallazgo como dependiente de un juicio de relevancia especialmente generoso. Algo similar sucede en la búsqueda 11, pero en este caso se deben exceptuar tanto este directorio como *Sol*.

Discusión, conclusiones y recomendaciones

El rendimiento de la recuperación de los sistemas analizados durante este estudio llevado a cabo en 2003 fue, en términos generales, reducido. Esto se refiere tanto a la recuperación como a la función de similitud en que se basó la ordenación de los resultados.

La primera conclusión que se obtuvo deriva de las cifras de exhaustividad y precisión alcanzadas por los sistemas, con valores promedio máximos del 23% y del 45% respectivamente, aunque se apreciaron variaciones entre ellos. El cálculo de un valor discreto para la relevancia, del que deriva la ordenación de resultados, se basó en una mera función estadística que pareció insuficiente. A este respecto, resulta revelador que sólo *Yahoo* mostrara un trazado típico y de tendencia continuada en el diagrama exhaustividad-precisión. En los restantes sistemas se produjeron repuntes en la precisión tras un arranque que indicaba resultados no relevantes encabezando las listas. Aunque los valores máximos de precisión se alcanzaron en los 5 primeros resultados y los valores de exhaustividad se estabilizaron en el tramo 1-15, estos hallazgos eran coherentes con una deficiente ordenación de los resultados de búsqueda.

La búsqueda de determinados temas ofreció mejores resultados en unos sistemas que en otros. Esta conclusión contradice la pretendida universalidad de los sistemas analizados. Por otra parte, este componente incorpora la mayor o menor generosidad en el juicio de relevancia de cada usuario individual, un factor que podría tener igual o mayor peso que la desigual orientación en la cobertura.

Un censo posterior (junio de 2004) de sistemas españoles de recuperación de información en internet ofrecía 36 servicios generales (no limitados regionalmente). Entre ellos no se incluía el directorio *EnlaWeb*, considerado inactivo desde el 22 de mayo de 2004 y ofrecían problemas de conexión *Avispanet* y *VenyBusca*. De los 34 servicios en operación en ese momento, 15 se analizaron en alguna de las fases de este estudio que presentamos. Dos de los restantes son sistemas especializados en información turística o en tecnologías de internet (*SearchIberia* y *Navegalis*, respectivamente). Se cuentan 7 sistemas de directorio. De los 10 restantes, sólo uno, la versión española de *Google* (introducida el 25 de septiembre de 2003) empleaba la recopilación automática y la recuperación contextual.

Si se tiene en cuenta el conjunto de variables estudiadas, *AltaVista* era el sistema que destacaba sobre los demás. Presentaba las mejores cifras de aporte específico (sobre un conjunto de sistemas del que cabe recordar que 9, con nulo aporte específico, se pueden considerar prescindibles en la búsqueda de información en el espacio web español). Además, presentaba la menor proporción de solapamiento si se descuenta *EnlaWeb* que, por su naturaleza de directorio, no incluía páginas sino sedes en su cobertura. *AltaVista* presentaba, por otra parte, una gran coherencia entre su nivel de representación de los documentos y las opcio-

nes de recuperación que ofrece a sus usuarios, así como buenas cifras de rendimiento.

Los sistemas dominantes en el cálculo del rendimiento, *Ozú* y *Yahoo*, empleaban tecnología de recuperación de *Google*. Sus bases de datos eran muy limitadas, según revelaban las cifras de aporte específico y de solapamiento, mostrando claras divergencias entre el nivel de representación de los documentos y las opciones ofrecidas a sus usuarios. Una mayor atención a la cobertura permitiría mejorar en mucho su utilidad en la recuperación de información de la web española. Por su parte, *Ya* estaba basado también en *Google* y presentaba datos de rendimiento ligeramente inferiores a los anteriores y una base de datos con gran solapamiento con los restantes sistemas.

«Si se tiene en cuenta el conjunto de variables estudiadas, *AltaVista* destaca sobre los demás»

Estos tres casos permitieron concluir que no bastaba con el empleo de programas de recuperación de reconocido prestigio y probado rendimiento para garantizar un nivel de recuperación de información adecuado en el espacio web.

El único directorio evaluado, *EnlaWeb*, ofreció datos coherentes con su propia naturaleza. Únicamente era destacable por su bajo nivel de solapamiento, aunque este dato resulta de un artificio: el hecho de que — como ya se ha dicho— sólo incluyera en su cobertura sedes, y no páginas individualizadas. Ni su rendimiento, ni su aporte específico ni su nivel de representación justificaron su empleo.

Del mismo modo, *Olé* y *Lycos* parecían prescindibles si se tiene en cuenta su alto grado de solapamiento. *Sol* añadía a su mínima aportación específica el peor rendimiento en términos de exhaustividad y precisión.

Las intenciones de flexibilización del registro de dominios españoles (*Ministerio de Ciencia y Tecnología*, 2003) pueden sentar las bases para facilitar la distribución online de información en el espacio web español. Pero para que esta información resulte realmente visible y accesible se necesitaría un acuerdo de mayor alcance: la sindicación de contenidos entre los 24 agentes registradores españoles y los propietarios de los sistemas de recuperación. La cobertura del espacio bajo dominio *.es* o alguna de sus combinaciones con subdominios genéricos se vería enormemente mejorada.

Un procedimiento similar permitiría a los sistemas de recuperación ofrecer información procedente de al-

Ahora renovar (o comenzar) la suscripción a “El profesional de la información” es mucho más ágil y sencillo.

Usted puede gestionar online su suscripción conectándose a esta página web:

<http://www.elprofesionaldelainformacion.com/suscripciones.html>

Si lo desea puede comunicar con nosotros dirigiéndose a esta dirección de correo electrónico:

suscripciones@elprofesionaldelainformacion.com

gunas bases de datos cuyo contenido se genera automáticamente en respuesta a transacciones. El protocolo *ICE* (*internet commerce enabled*), operativo desde 2000, y el desarrollo de funciones *CGI* y otras pasarelas y traductores posibilitarían el acceso (sin necesidad de innecesarias recargas de los índices).

Naturalmente, nada de ello ha de resultar en mejor cobertura si los módulos de recopilación no se programan al ritmo suficiente y si no se dota a los algoritmos de detección de duplicados de una mayor eficacia.

«Convendría la combinación con programas de visualización de la información que ofrezcan los resultados estructurados gráficamente, en lugar de ordenarlos en una mera lista unidimensional»

Se debería valorar la posibilidad de emplear agentes en lugar de almacenar en bases de datos locales la información procedente de páginas (obligatoriamente estáticas). Esta alternativa permitiría proporcionar una tasa de refresco inmediata a los sistemas. Acaso convendría la combinación con programas de visualización de la información que ofrecieran los resultados estructurados gráficamente, en lugar de ordenarlos en una mera lista unidimensional. El empleo de este tipo de programas o de soluciones basadas en la tecnología de mapas semánticos requiere, sin embargo, del cálculo de las relaciones entre unos y otros documentos recuperados, y conduciría, en su caso, a nuevos conceptos y procedimientos en la recuperación, muy próximos a los que emplean los sistemas de recuperación contextual, e incluso más avanzados.

Ya son apreciables, por otra parte, los intentos de indizar el contenido de documentos textuales o mixtos en formatos propietarios, sobre todo *MS Word* y *Adobe Acrobat*.

Por lo que respecta a las condiciones operativas de la indización y la recuperación, se debería obtener un mejor ajuste entre la estructura implícita de los documentos “planos”, es decir, aquellos basados únicamente en html y la correspondiente arquitectura de índices. De esta forma se ofrecería una gama mayor de puntos de acceso al usuario, al tiempo que se posibilitaría el refinamiento de búsqueda y la ganancia en relevancia.

La recuperación en los sistemas de información en internet no puede depender —ni en la selección ni en la ordenación de resultados— de funciones basadas en meros recuentos estadísticos de términos. Una mayor atención a aspectos estructurales implícitos en los documentos, con tratamiento diferenciado de los términos de los títulos, por ejemplo, podría añadir variables de utilidad al cálculo de la función de similitud y a la ordenación de resultados. Si es cierto que el número de enlaces externos de las sedes sigue modelos exponenciales de crecimiento, existe la posibilidad de que las funciones contextuales intervengan en la recuperación o, al menos, en la ordenación de los conjuntos de documentos recuperados. Los grupos de investigación nacionales comienzan a ofrecer avances de gran aplicación a la recuperación de información distribuida. La clasificación automática, la recuperación de pasajes y la visualización son de aplicación directa a la mejora del rendimiento en la recuperación.

El estudio de los sistemas de recuperación de información distribuida es deseable y necesario, tanto para su aplicación a los sistemas propietarios o a las intranets, en el contexto de lo que se ha dado en llamar *ECM* (*enterprise content management*), como a nivel general, sobre todo en el manejo de colecciones de documentos electrónicos. En este sentido, cabe finalizar con una apelación dirigida por **Manuel Castells** desde la perspectiva global de lo que viene en denominarse la “sociedad red”:

“Una vez que toda la información está en la red, una vez que el conocimiento está en la red, el conoci-

miento codificado, pero no el conocimiento que se necesita para lo que se quiere hacer, de lo que se trata es de saber dónde está la información, cómo buscarla, cómo procesarla, cómo transformarla en conocimiento específico para lo que se quiere hacer (...) Es ahí donde está, empíricamente hablando, la divisoria [brecha] digital en estos momentos” (Castells, 2001).

Bibliografía

Abad García, María Francisca. “Evaluación de los componentes de los sistemas de recuperación de la información”. En: *Investigación evaluativa en documentación: aplicación a la documentación médica*. Valencia: Universitat de València, 1997, pp. 125-163.

Abad García, María Francisca. “Evaluación de la eficacia de los SRI”. En: *Investigación evaluativa en documentación: aplicación a la documentación médica*. Valencia: Universitat de València, 1997, pp. 85-122.

Alonso Berrocal, José Luis. *Cibernetria: Análisis de los dominios web españoles*. Tesis doctoral. Universidad de Salamanca, 2000.

Amat, Carlos B. “Limitaciones en el nivel de representación, las posibilidades de recuperación y la cobertura relativa de los sistemas españoles de recuperación de información distribuida en internet”. En: *Revista española de documentación científica*, 2002, n. 25, pp. 329-338.

Amat, Carlos B. “Caracterización de una muestra de sedes web españolas bajo dominio .es”. En: *Boletín de RedTris*, 2003, n. 64, pp. 33-40.

Amat, Carlos B. “Evaluación de sistemas españoles de recuperación de información distribuida en Internet». Tesis doctoral. Universitat de València, 2004.

Bailey, P.; Craswell, N.; Hawking, D. “Engineering a multi-purpose test collection for web retrieval experiments”. En: *Information processing and management*, 2003, n. 39, pp. 853-871.

Baró i Queralt, Jaume. “Cerca i recuperació d’informació al world wide web: una aproximació a les eines disponibles”. En: *Jornades catalanes de documentació*, 1997, pp. 469-479.

Bellardo Hahn, Trudi. “Text retrieval online: historical perspective on web search engines”. En: *Bulletin of the American Society for Information Science*, 1998, n. 24, pp. 7-10.

Can, F.; Nuray, R.; Sevdik, A. B. “Automatic performance evaluation of web search engines”. En: *Information processing; management*, 2004, n. 40, pp. 495-514.

Castells, Manuel. *Internet y la sociedad red: lección inaugural del programa de doctorado sobre la sociedad de la información y el conocimiento*. Universitat Oberta de Catalunya, 2001. Consultado en: 07-04-04. <http://www.uoc.edu/web/esp/articles/castells/print.html>

Gordon, M.; Pathak, P. “Finding information on the world wide web: the retrieval effectiveness of search engines”. En: *Information processing and management*, 1999, n. 35, pp. 144-180.

Gómez Díaz, R. “La evaluación en recuperación de la información”. En: *Hipertext.net*, 2003. Consultado en: 16-05-05. <http://www.hipertext.net/web/pag188.htm>

Gurrin, C.; Smeaton, A. “Replicating web structure in small-scale test collections”. En: *Information retrieval*, 2004 n. 7, pp. 239-263.

Harter, S.; Hert, C. “Evaluation of information retrieval systems: approaches, issues and methods”. En: *Annual review of information science and technology*, 1997, n. 32, pp. 3-94.

Lancaster, Frederick; Warner, Amy. “Evaluation criteria and evaluation procedures”. En: *Information retrieval today*. Arlington: Information Resources Press, 1993, pp. 159-202.

Leighton, Vernon H.; Srivastava, J. “First 20 precision among world wide web search services (search engines)”. En: *Journal of the American Society for Information Science*, 1999, v. 50, n. 10, pp. 870-881.

López Alonso, M.; Mares Marín, J. “El futuro de la identificación de la información en internet”. En: *Fesabid*, 1996, pp. 513-518.

Maldonado Martínez, Ángeles; Fernández Sánchez, Elena. “Evaluación de los principales ‘buscadores’ desde un punto de vista documental: recogida, análisis y recuperación de recursos de información”. En: *Fesabid*, 1998, pp. 529-551.

Marcos Mora, Mari Carmen. “Motores de recuperación de información: un análisis comparativo (parte 1)”. En: *El profesional de la información*, 1998a, v. 7, n. 1-2, pp. 18-22.

Marcos Mora, Mari Carmen. “Motores de recuperación de información: un análisis comparativo (parte 2)”. En: *El profesional de la información*, 1998b, v. 7, n. 3, pp. 13-20.

Martínez Méndez, F. Javier. *Propuesta y desarrollo de una metodología para la evaluación de la recuperación de información en internet*. Universidad de Murcia, 2001.

Martínez Méndez, F. Javier; Rodríguez Muñoz, José Vicente. Síntesis y crítica de las evaluaciones de la efectividad de los motores de búsqueda en la web. En: *Information research*, 2003. Consultado en: 16-05-05. <http://informationr.net/ir/8-2/paper148.html>

Ministerio de Ciencia y Tecnología. *Orden CTE/662/2003, de 18 de marzo, por la que se aprueba el Plan Nacional de nombres de dominio de internet bajo el código de país correspondiente a España («.es»)*. En: *Boletín oficial del estado*, 2003, n. 73, pp. 11.917-11.924.

Olvera Lobo, María Dolores. *Evaluación de la recuperación de información en internet: un modelo experimental*. Facultad de Biblioteconomía y Documentación, Universidad de Granada, 1999.

Olvera Lobo, María Dolores. “Rendimiento de los sistemas de recuperación de información en la world wide web: revisión metodológica”. En: *Revista española de documentación científica*, 2000, n. 23, pp. 63-78.

Olvera Lobo, María Dolores. “Rendimiento de los sistemas de recuperación de información en la web: evaluación de servicios de búsqueda (search engines)”. En: *Revista española de documentación científica*, 2000, n. 23, pp. 302-316.

Oppenheim, Charles; Morris, A.; Mcknight, C.; Lowley, S. “The evaluation of www search engines”. En: *Journal of documentation*, 2000, n. 52, pp. 190-211.

PricewaterhouseCoopers. *Estudio de la industria de contenidos digitales en España*. Price Waterhouse Coopers España 2004. Consultado en: 07-04-05. http://www.pwc.com/es/esp/ins-sol/spec-int/ind_contenidos.html

Salton, Gerard; McGill, Michael. *Retrieval evaluation*. En: *Introduction to modern information retrieval*. New York: McGraw-Hill, 1983, pp. 157-197.

Sánchez Montero, J. “Hacia una optimización de los recursos de internet en la empresa”. En: *Revista española de documentación científica*, 1997, n. 20, pp. 52-60.

Senso, J. A. “Herramientas para realizar búsquedas en internet: una revisión”. En: *El profesional de la información*, 1998, n. 7, pp. 24-25.

Travis, Irene. “From ‘storage and retrieval systems’ to ‘search engines’: text retrieval in evolution”. En: *Bulletin of the American Society for Information Science*, 1998, n. 24.

Vaughan, Liwen. “New measurements for search engine evaluation proposed and tested”. En: *Information processing and management*, 2004, v. 40, n. 4, pp. 677-691.

Winer, B. “Design and analysis of single-factor experiments”. En: *Statistical principles in experimental design*, 2nd ed. New York: McGraw-Hill, 1962, pp. 149-260.

Carlos Benito Amat, Servicio de Documentación Científica y Biblioteca, Instituto de Agroquímica y Tecnología de Alimentos (IATA), Consejo Superior de Investigaciones Científicas (CSIC), Apdo. 73, 46.100 Burjassot (Valencia). carlos.benito@iata.csic.es