

# 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

## Introducción

Las razones que justifican el análisis de los sistemas de recuperación de información distribuida en Internet (SIDi, vulgo *buscadores*) son variadas y de gran peso. A pesar de esto, se cuentan muy pocos trabajos originales que aborden dicho estudio en España. Por fortuna, tanto la aportación de Angeles Maldonado y Elena Fernández (1) como la tesis doctoral y posteriores publicaciones de María Dolores Olvera (2, 3), son serias en sus planteamientos y su elaboración, aunque la naturaleza de su objeto de estudio limite en mucho el alcance de sus conclusiones.

En el editorial de presentación del número del ASIS Bulletin dedicado a la evolución de la recuperación de textos, Irene Travis hace énfasis en los motivos que apoyan el análisis de esos sistemas (4). De su enfoque, y de otros (5), incluidos en el mismo monográfico, se pueden enumerar las siguientes:

- 1) Los SIDi ofrecen la oportunidad de estudiar las posibilidades de consulta en texto libre y recuperación de ingentes colecciones de documentos (electrónicos), una situación que parece avecinarse en el entorno de la información ligada al conocimiento y al que no han de ser ajenas las iniciativas de «colecciones digitales» ahora incipientes.
- 2) El acceso de usuarios no cualificados sin intermediación alguna, los «casual users» frente a «end users», brinda la oportunidad de estudio de sus hábitos de consulta y recuperación y la extracción de conclusiones de extrema importancia práctica.
- 3) Las características de los documentos y textos a recuperar (inestabilidad, desahogado crecimiento, heterogeneidad máxima) plantean una serie de desafíos especialmente interesantes al diseño de los sistemas, a sus mecanismos de actualización y a su arquitectura de almacenamiento.
- 4) 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.

A las motivaciones anteriores se puede añadir una quinta: ¿Hasta qué punto los SIDi pueden hacer frente al panorama aquí sólo esbozado? ¿Son los sistemas españoles capaces de recuperar de forma eficiente la información española? Maldonado y Fernández presentaron un estudio pormenorizado de las «características documentales» de 10 SIDi 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 sistema. Como resultado, emitían un juicio sobre la operatividad de todos y un análisis global de ambos grupos. Aunque su trabajo es de un alto valor descriptivo, no se ajusta en modo alguno a los procedimientos de evaluación esquematizados por Oppenheim en su revisión metodológica (6). Por el contrario, Olvera realiza un gran trabajo «tipo Cranfield», pero concentra su esfuerzo en sistemas internacionales. Desde cierto punto de vista, ambos originales son complementarios.

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 a la producción y el consumo de información científica «nacional») se reconocerá el interés del estudio de los SIDI desarrollados en España. Para llevarlo a cabo, es rigurosamente necesario analizar los tres componentes principales de cualquier sistema de recuperación de información: 1) los documentos; 2) las características operativas y el esquema conceptual de los sistemas y, naturalmente, 3) la interacción usuario-sistema y la satisfacción de las demandas planteadas.

Por estudio de los documentos ha de entenderse una profundización en el manido comentario (camino de convertirse en frase hecha) que une heterogeneidad, ingente volumen y falta de estabilidad como características del espacio Web. El análisis del esquema conceptual y la operatividad de los sistemas exige la comparación entre estos sistemas y otros, tradicionales o no. La línea de los usuarios y la satisfacción de sus demandas se ha de plantear teniendo en cuenta no sólo la mecánica de ajuste entre perfiles y documentos. La cultura, el estado de conocimiento previo y otros factores se han de tener en cuenta en este tercer componente de la evaluación.

En el marco de un proyecto de evaluación de los SIDI españoles, se ha trazado ya una panorámica del campo (7) y se ha procedido a la caracterización de los documentos del espacio Web español (8). El presente trabajo avanza en la línea trazada. Su objetivo general es la investigación de dos elementos clave del esquema conceptual de los sistemas españoles de recuperación de información distribuida y un tercer elemento crucial en la valoración de esos mismos sistemas. Los dos primeros se refieren al esquema de datos y a la dinámica de recuperación. El tercero a la cobertura.

El esquema de datos no sólo refleja el nivel de representación de los documentos en cada sistema sino que, junto con los datos relativos a su cobertura, ilustra el funcionamiento del principal módulo de cualquier buscador, su «crawler» o módulo de recopilación. Por otra parte, las funcionalidades del «indexer» (módulo de indización) y el «searcher» (de recuperación) quedan reveladas (aunque sea parcialmente) por el análisis de la mecánica de recuperación. Véase si no la excelente descripción que Risvik y Michelsen realizan del sistema FAST (9).

## **Fuentes y método**

Los principales apartados metodológicos del presente estudio son 1) la selección de los sistemas objeto de estudio; 2) el examen de cada esquema de datos y su comparación con algún conjunto autorizado y normalizado; 3) la caracterización de la mecánica o de las opciones de recuperación, y 4) la estimación de su cobertura.

## Selección de los sistemas

Aunque se dispone de un sinfín de «listas de buscadores recomendados», la selección tuvo como punto de partida la lista distribuida en Red IRIS hasta mediados de 2001 y la recopilación de Buscopio (10). En ambos casos se incluyeron inicialmente en el estudio los sistemas españoles de ámbito y temática general operativos a finales del año 2000. Siguiendo el procedimiento de Maldonado y Martínez, se realizó una «prueba de popularidad» en dos sistemas que permitían cuantificar el número de enlaces dirigidos a la dirección de cada servicio. Para minimizar el efecto de la integración de los sistemas de búsqueda en portales, la URL empleada correspondió a la dirección específica de las páginas de búsqueda (cuando fue posible). Así, por ejemplo, se sustituyó la expresión «link:www.telepolis.com» por la más exacta «link:buscador.telepolis.com». La tabla 1 muestra las cifras correspondientes a esta estimación (30 de Agosto de 2002) junto con los mismos datos obtenidos por Maldonado y Martínez en 1998.

**Tabla 1**  
**«Popularidad» de los sistemas seleccionados**

	<i>Google</i>	<i>Altavista</i>
Altavista	580	58
Apali	440	229
BIWE	2110	635
Buscopio	1630	664
Elcano	2320	24
Elindice	1830	6175
Enlaweb	444	118
Eureka	406	382
Hispavista		
Lycos	8980	5
Ole (Terra)	268	3
Ozú	3120	1865
Salman	260	28
Sol	1070	547
Telepolis	362	70
Trovator	1390	18
Ya	1330	164
Yahoo	42500	1366

Número de enlaces recibidos por los diversos SIdI según dos sistemas que permiten su determinación. La cifra de Yahoo corresponde al portal (yahoo.es).

La lista inicial incluía la porción española del sistema Excite, y los sistemas llamados El buscador y Ozu.com. Desafortunadamente, la inestabilidad del espacio Web ha pasado factura a los propios sistemas: Excite.es cesó en su actividad el 11 de Junio de 2001 (11), Ozu.com se fusionó con Ozú.es y la URL de El buscador (<http://www.elbuscador.com/>) está ocupada actualmente por un proveedor de contenidos eróticos para Internet y servicios 906. Además, los intentos de conexión con el índice ofrecen un error desde finales de agosto de 2002. Naturalmente, sólo se seleccionaron las porciones españolas de los sistemas Altavista, Lycos y Yahoo.

En Marzo y Abril de 2001 se realizaron envíos de una encuesta electrónica dirigida a todos los servicios. La encuesta se basaba en el modelo de formulario de DESIRE (Fase 1) (12) y tenía por objeto la determinación de los mecanismos de recopilación e indización empleados por cada servicio, además de otras características. Sólo uno de los servicios cumplimentó la encuesta, aunque sin ajustarse a la estructura requerida. Se impuso el examen directo de los servicios.

### **Determinación de los esquemas de datos**

Las fuentes para la determinación de los datos recopilados por cada sistema fueron las páginas de búsqueda avanzada y/o las páginas de ayuda tanto a la inscripción como a la recuperación, los formularios de solicitud de incorporación al sistema, las listas de preguntas frecuentes (FAQ) disponibles y determinadas comprobaciones en la interrogación a los sistemas.

Como base de comparación se eligió la segunda versión del conjunto de elementos del Dublin Core Metadata (13). La correspondencia entre unos y otros esquemas se estableció de modo que, además de la relación obvia entre los elementos *title*, *date*, *type-format*, *identifier* y *language* (respectivamente título, fecha, formato, URL e idioma) se emparejaron *coverage* y ámbito geográfico, *subject* con palabras-clave y clasificación, *creator* con responsable, *publisher* con organismo editor y *description* con descripción. Se asignó a cada sistema una puntuación para cuantificar el nivel de representación de los documentos Web recopilados. Así, la correspondencia entre un elemento del esquema y su correspondiente elemento del Dublin Core equivalía a un punto, a excepción de *type-format*, valorado con puntos cuando se incluía, además del formato, su tamaño y de *subject*, puntuado doble también cuando, además de los epígrafes clasificatorios, se incluían palabras clave. Como los esquemas clasificatorios identificados dependían de cada sistema, no hubo oportunidad de asignar 3 puntos a esta variable, lo que hubiera significado que el esquema clasificatorio elegido estaba apoyado por alguna autoridad (LC, CDU, o cualquier otra). Igualmente, la combinación *type-format* hubiera totalizado 3 puntos si se hubiera incorporado en algún caso el subelemento *version*.

### **La mecánica (las opciones) de recuperación**

Las páginas de búsqueda simple y avanzada, la ayuda a la búsqueda y las fuentes empleadas en la determinación del esquema de datos se utilizaron para caracterizar los mecanismos de recuperación de cada sistema. Además de determinar las posibilidades y sintaxis de la expresión de perfiles, las posibilidades de combinación de criterios y la limitación por elementos (campos) se investigaron la existencia de mecanismos de refinamiento de búsqueda (*relevance feedback*) y búsqueda por similitud (*query by example*). Se procedió a la construcción de una escala de mecanismos de recuperación, de forma que el análisis pudiera traducirse en puntuaciones. Desde el nivel inferior, los tramos de esta escala son los siguientes: 1) empleo de operadores lógicos (búsqueda booleana); 2) empleo de expresiones compuestas (búsqueda por frase); 3) búsqueda por proximidad; 4) anidamiento de expresiones; 5) búsqueda en elementos

definidos (limitación por campo); 6) truncación; 7) enmascaramiento; 8) refinamiento de resultados; 9) búsqueda por similitud, y 10) posibilidad de seguimiento y alerta.

A grandes rasgos se puede decir que las diferentes fases de esta secuencia corresponden a niveles progresivos de procesamiento. En su valoración, sin embargo, se ha de tener en cuenta el hecho de que varios servicios españoles han licenciado la arquitectura de construcción de índices y de recuperación de otros sistemas. Por otra parte, y al igual que en el sistema de puntuación propuesto para los esquemas de datos, existe una variable que puede alcanzar valores superiores a la unidad: la limitación por campos. En función del número de elementos que permitan precisar así la expresión de búsqueda, los sistemas obtendrán mayor o menor puntuación.

## Estudio de la cobertura

Dos series de operaciones se realizaron para determinar la cobertura de los servicios. En primer lugar, se procedió a la extracción de una muestra aleatoria de dominios españoles de segundo nivel (sedes). La muestra se extrajo el 4 de enero de 2001 y consistió en un conjunto de 500 sedes bajo dominio .es del total de 29.858 registrados en esa fecha por ES-NIC. *En primer lugar*, se emplearon los indicadores de cobertura resumizados por Abad (14), utilizando el mismo método que Castillo, Martínez y Server (15). En concreto, se determinaron la tasa de solapamiento, la tasa de cobertura y el índice de aporte específico de cada servicio en relación con la muestra de sedes. Tal y como se pone de manifiesto en el trabajo de caracterización de la muestra (8), no todos los dominios contenían información o eran accesibles.

*En segundo lugar*, se determinó si un dominio era incluido en un sistema, mediante un conjunto de búsquedas (marzo de 2001). Los perfiles empleados correspondieron a elementos de la URL de los dominios en aquellos sistemas que permitían tal modalidad de búsqueda. En los restantes, se utilizaron palabras del título de la portada (*home page*) de cada dominio en combinación booleana o como frase. Los resultados se examinaron para garantizar que cada sistema devolvía exactamente el dominio en cuestión.

## Resultados y discusión

Atendiendo a la escala propuesta, los SIDI españoles presentan esquemas de datos insuficientes. Sobre un total de 19 puntos, que corresponderían a un ajuste total a la lista de elementos del Dublin Core y a la interpretación cuantitativa que se ha propuesto, 1 sistema alcanzan una puntuación de 9 (ver tabla número 2) le siguen otro con 8 puntos y 3 puntuados con 7. La puntuación mínima adjudicada es de 2. La tabla 2 permite apreciar que, salvo en el caso de Altavista, los elementos *Subject* y *Type-Format* en combinación, son los criterios que en mayor medida contribuyen al aumento de las puntuaciones. El servicio con el esquema de datos más valorado ([www.altavista.es](http://www.altavista.es) o [es-es.altavista.com](http://es-es.altavista.com)) distribuye sus méritos relativos entre todos los elementos. La existencia de la combinación palabras-clave y esquema clasificatorio rinde las mejores puntuaciones en lo que al elemento *Subject* respecta. También contribuye a ellas la inclusión de valores relativos al tipo de fichero y a su tamaño en la

**Tabla 2**  
**Comparación entre el Dublín Core Element Set y los esquemas de datos de los servicios analizados**

	<i>Tit- le</i>	<i>Sub- ject</i>	<i>Des- cription</i>	<i>Publis- her</i>	<i>Da- te</i>	<i>Type- Format</i>	<i>Iden- tifier</i>	<i>Sour- ce</i>	<i>Lan- guage</i>	<i>Rela- tion</i>	<i>Cove- rage</i>	<i>Pun- tos</i>
Altavista	1	1	1	0	1	1	1	1	1	1	1	9
Apali	1	1	1	0	1	0	1	0	1	0	0	5
BIWE	1	2	1	0	0	1	1	0	0	0	1	6
Buscopio	1	0	0	0	0	0	1	1	0	0	0	2
Elcano	1	1	1	1	2	2	1	0	0	0	0	8
Elindice	1	1	1	0	1	0	1	0	0	0	0	4
Enlaweb	1	0	1	0	0	0	1	0	0	0	0	2
Eureka	1	2	1	0	1	0	1	0	1	0	1	7
Hispavista	1	1	1	0	0	0	1	0	1	0	1	5
Lycos	1	1	0	0	1	2	1	0	1	0	1	7
Ole (Terra)	1	1	0	0	1	2	1		1	0	0	6
Ozú	1	0	0	0	0	2		1	0	1	0	5
Salman	1	1	1	0	1	0	1		1	0	1	6
Sol	1	1	1	0	0	2	1	1	0	1	0	7
Telepolis	1	2	1	0	0	0	1	0	1	0	1	6
Trovator	1	1	1	0	1	0	1	0	0	0	1	5
Ya	1	1	1	0	0	0	1		1	1	1	6
Yahoo	1	1	1	0	1	0	1	0	1	0	1	6

variable *type-format*. Ninguno de los sistemas analizados alcanza puntuaciones máximas en estos dos elementos y sólo uno obtiene una puntuación doble en la variable *date*: incluye no sólo la fecha de creación de los documentos recopilados, también la de modificación.

Los diez «tramos» en la escala de valoración de las opciones o mecanismos de recuperación son cubiertos de forma irregular por los sistemas. Las puntuaciones máximas (ver tabla número 3) corresponden a los fragmentos españoles de sistemas internacionales o a aquellos que han licenciado sus programas: Hispavista (8 puntos) y Ya (también 8) emplean tecnología FAST, Ozú (9) y Sol (6) han licenciado los programas de Google y Ole-Terra (8) los de Lycos. Las mayores aportaciones a la puntuación final proceden, naturalmente, de la variable «limitación por campos», donde las puntuaciones más elevadas son, precisamente, las que corresponden a estos sistemas. Sólo el sistema Apali, a través de su mecanismo «*personal search*», que requiere registro previo, ofrece la posibilidad de almacenamiento y reejecución diferida de perfiles de búsqueda. Las puntuaciones se ha otorgado a los servicios atendiendo a las instrucciones de búsqueda, tal y como se ha especificado en el apartado metodológico. Se debe precisar, no obstante, que el sistema Altavista (puntuación total de 8 y de 2 en el componente de limitación) posee muchas más opciones de limitación por elemento (por campo) de las especificadas en sus instrucciones de búsqueda. Finalmente, y en relación con la truncación, merece la pena que se anote que dos sistemas (Hispavista y Salman) realizan tanto truncación como enmascaramiento automáticos.

De las 500 sedes seleccionadas aleatoriamente, sólo 166 eran accesibles en el momento del estudio. Las razones de esta limitación se detallan y comentan en el trabajo antes citado (8). Partiendo, en consecuencia, de un global de 166 sedes, los datos relativos a la cobertura de los sistemas estudiados se ofrecen en la tabla 4. Puede apre-

**Tabla 3**  
**Valoración de las opciones y mecanismos de recuperación de los sistemas analizados**

	<i>Bool</i>	<i>Fras</i>	<i>Prox</i>	<i>Anid</i>	<i>Limi</i>	<i>Trun</i>	<i>Masc</i>	<i>Refi</i>	<i>Simi</i>	<i>Aler</i>	<i>Puntos</i>
Altavista	1	1	1	1	2	1	1	0	0	0	8
Apali	1	1	0	0	0	0	0	0	0	1	3
BIWE	1	1	0	0	1	1	0	1	0	0	5
Buscopio	1	1	0	0	0	0	0	1	0	0	3
Elcano	1	1	0	0	1	1	1	0	0	0	5
Elindice	Error 404										
Enlaweb	1	1	0	0	0	1	1	0	0	0	4
Eureka	1	1	0	0	1	1	0	0	0	0	4
Hispavista	1	1	0	0	4	1	1	0	0	0	8
Lycos	1	1	0	1	5	0	0	0	0	0	8
Ole (Terra)	1	1	0	1	5	0	0	0	0	0	8
Ozú	1	1	0	0	4	1		1	1	0	9
Salman	0	1	0	0	2	1	1	0	0	0	5
Sol	1	1	0	1	1	1	1	0	0	0	6
Telepolis	1	1	0	1	0	1	1	0	0	0	5
Trovator	1	0	0	1	1	0	0	0	0	0	3
Ya	1	1	0	0	6	0	0	0	0	0	8
Yahoo	1	1	0	1	3	1	0	0	1	0	8

Puntuaciones parciales y total obtenidas por los sistemas analizados en razón a las opciones de recuperación que ofrecen: 1) empleo de operadores lógicos (búsqueda booleana); 2) empleo de expresiones compuestas (búsqueda por frase); 3) búsqueda por proximidad; 4) anidamiento de expresiones; 5) búsqueda en elementos definidos (limitación por campo); 6) truncación; 7) enmascaramiento; 8) refinamiento de resultados; 9) búsqueda por similitud, y 10) posibilidad de seguimiento y alerta.

ciarse en ella que 2 sistemas presentaban nula cobertura y que los valores máximos en relación con el aporte global de cada sistema (el porcentaje de dominios cubiertos) rebasa un tercio del conjunto de dominios sólo en dos casos (Buscopio, 36,75% y Altavista 34,94%). En cuanto al aporte específico, es decir, al número de dominios que cada buscador ha incluido en su cobertura de forma específica, Altavista e Hispavista rebasan el 3% y 9 de los sistemas en cuestión ofrecen una aportación exclusiva nula: la búsqueda de contenidos en sedes Web españolas podría prescindir de estos sistemas.

Los resultados expresados están sujetos a las mismas limitaciones que los obtenidos por trabajos anteriores (1, 3): las determinaciones se han realizado en fechas concretas y no es posible que la repetición de los cálculos ofrezca resultados idénticos. Por otra parte, el presente trabajo forma parte de una serie que, sólo a su conclusión, permitirá la formulación de conclusiones sólidas.

No obstante, algunos hallazgos merecen ser destacados: el esquema de datos de la mayoría de los sistemas es rudimentario. Teniendo en cuenta que la mayor parte de los elementos del esquema de comparación son descriptivos (en oposición a analíticos, como sería el caso de las palabras-clave, la clasificación y la descripción) y dependen sólo de la configuración de los programas de recopilación, no se explica la falta de eficiencia que se ha hallado. Otro tanto sucede con las propiedades de los respectivos módulos de indización. A excepción de las variables situadas en los tra-

**Tabla 4**  
**Cobertura relativa de los sistemas analizados**

	$S(a)$	<i>Global</i>	$S(x)$	<i>Específico</i>
Altavista	58	34,94	6	3,61
Apali	0	0	0	0
BIWE	24	14,46	0	0
Buscopio	61	36,75	1	0,6
Elcano	5	3,12	0	0
Elindice	30	18,07	2	1,2
Enlaweb	0	0	0	0
Eureka	4	2,41	1	0,6
Hispavista	44	26,5	5	3,01
Lycos	54	32,53	1	0,6
Ole (Terra)	23	13,86	0	0
Ozú	21	12,65	1	0,6
Salman	7	4,22	0	0
Sol	25	15,06	0	0
Telepolis	9	5,42	0	0
Trovator	19	11,45	0	0
Ya	34	20,48	3	1,81
Yahoo	24	14,46	1	0,6
teórico global = 166				

Sobre la muestra de 166 sedes,  $S(a)$  es el número absoluto de sedes cubiertas por cada sistema,  $S(x)$  el número de sedes incluidas exclusivamente en cada sistema y las columnas restantes, los aportes global y específico de cada SIdI.

mos finales de la escala empleada (*query by example*, *relevance feed-back* y *alerta*) y de aquellos que requieren una especial arquitectura de índices (truncación y enmascaramiento sobre todos), es difícil justificar las bajas puntuaciones de gran parte de los sistemas analizados en la limitación por campos o la búsqueda de términos próximos, precisamente las opciones de interrogación que permiten un mejor ajuste entre necesidad de información y resultados obtenidos.

No existe correlación alguna entre las puntuaciones asignadas a los esquemas de datos y las obtenidas por las respectivas opciones de recuperación, hallazgo que habla a las claras de la mencionada falta de eficiencia de los sistemas. En otras palabras: de nada sirve un esquema de datos completo si la recopilación no se traduce en

Si hubiera que atenerse a los datos de cobertura resultantes del presente trabajo, se concluiría que las dos terceras partes de los contenidos de las sedes Web españolas resultan inaccesibles a los usuarios de los sistemas analizados. Por fortuna, y además de la existencia de SIdI internacionales con otros presupuestos, la muestra empleada se limitaba a las sedes bajo dominio de primer nivel .es, y no tenía en cuenta la plétora de sedes registradas en otros dominios más liberalizados (sobre todos .com). En todo caso, cabe anotar que, aunque existe una correlación positiva entre la aportación global y la específica de los sistemas estudiados ( $p < 0,01$ ) los valores son relativamente débiles (entre 0,53 y 0,66, dependiendo del estadístico empleado).



En términos generales, los SIDl españoles no se pueden considerar válidas «islas de información filtrada», por emplear otra de tantas metáforas náuticas al uso, acuñada en el contexto de la edición electrónica (16). Los sistemas estudiados, en las condiciones observadas y con los métodos aquí empleados, ofrecen escaso acceso a poca información insuficientemente representada.

## Reconocimientos

Ricardo Fornás (buscopio.com) ha criticado con extraordinario acierto el enfoque de éste y otros trabajos en la misma serie. Víctor Castelo y Diego R. López (RedIRIS) facilitaron la muestra aleatoria de sedes bajo dominio .es. Alejandro de la Cueva y Josep Lluís Canet siguen los progresos de este proyecto con especial atención.

## Bibliografía

1. MALDONADO MARTÍNEZ, A.; FERNÁNDEZ SÁNCHEZ, E. Evaluación de los principales «buscadores» desde un punto de vista documental: recogida, análisis y recuperación de recursos de información. *Sextas Jornadas Españolas de Documentación*, 1998, 29-21 de octubre, Valencia, p. 529-552. [http://esabid98.florida-uni.es/Comunicaciones/a\\_maldonado/A\\_Maldonado.htm](http://esabid98.florida-uni.es/Comunicaciones/a_maldonado/A_Maldonado.htm) [19 de septiembre de 2002].
2. OLVERA LOBO, M. D. Rendimiento de los sistemas de recuperación de información en la world wide web: revisión metodológica. *Revista Española de Documentación Científica*, 2000, 23 (1): 63-77.
3. OLVERA LOBO, M. D. Rendimiento de los sistemas de recuperación de información en la web: evaluación de servicios de búsqueda (search engines). *Revista Española de Documentación Científica*, 2000, 23 (3): 302-316.
4. TRAVIS, I. From «Storage and Retrieval Systems» to «Search Engines»: Text Retrieval in Evolution. *Bulletin of the American Society for Information Science*, 1998, 24 (4): 1. <http://www.asis.org/Bulletin/Apr-98/travis.html> [17 de septiembre de 2002].
5. BELLARDO HAHN, T. Text Retrieval Online: Historical Perspective on Web Search Engines. *Bulletin of the American Society for Information Science*, 1998, 24(4): 7-10. <http://www.asis.org/Bulletin/Apr-98/hahn.html> [17 de septiembre de 2002].
6. OPPENHEIM, C.; MORRIS, A.; McNIGHT, C.; LOWLEY, S. Progress in documentation: The evaluation of WWW search engines. *Journal of Documentation*, 2000, 56 (2): 190-211.
7. AMAT, C. B. Recuperación en Internet: Cuatro modelos complementarios y una agenda para su integración. *Boletín de Red IRIS*, 1999, 48. Accesible en <http://www.rediris.es/rediris/boletin/48/enfoque2.html> [18 de septiembre de 2002].
8. AMAT, C. B. Caracterización de una muestra de sedes Web españolas bajo dominio .es. *Boletín de Red IRIS*, 2002 (en prensa).
9. RISVIK, K. M.; MICHELSEN, R. Search Engines and Web Dynamics. *Fast Search and Transfer ASA*. Accesible en [www.idi.ntnu.no/~algkon/generelt/se-dynamicweb1.pdf](http://www.idi.ntnu.no/~algkon/generelt/se-dynamicweb1.pdf) [30 de agosto de 2002].
10. Buscador de buscadores. Prisa.com, SA. Accesible en <http://buscadores.buscopio.com> [10 de septiembre de 2002].
11. ENOS, L. Excite@Home is raising funds to improve its bottom line while at the same time taking steps to cut costs. *E-Commerce Times*, 11 de junio de 2001.

12. KOCH, T.; ARDÖ, A.; BRÜMMER, S. y LUNDBERG, S. *The building and maintenance of robot based internet search services: A review of current indexing and data collection methods*. Lund University Library, Net Lab, 1996. Accesible en <http://www.lub.lu.se/desire/radar/reports/D3.11/> [22 de septiembre de 2002].
13. Dublin Core Metadata Element Set, Versión 1.1: Reference Description. Dublin Core Metadata Initiative, 1999. Accesible en <http://dublincore.org/documents/dces> [17 de septiembre de 2002].
14. ABAD GARCÍA, M. F. Evaluación de los componentes de los sistemas de recuperación de la información en *Investigación Evaluativa en Documentación*. Valencia, Universitat de Valencia, 1997, pags. 125-163.
15. Castillo Blasco, L ; Martínez de Pablos, MJ ; Server, G: Evaluación de la información contenida en seis sedes web de las Escuelas Universitarias y Facultades de Biblioteconomía y Documentación españolas. *Revista Española de Documentación Científica*, 1999, 22 (3): 325-332.
16. BUTLER, D: The writing is on the Web for Science journals in print. *Nature*, 1999, 397 (6716): 195-200.

CB Amat

Instituto de Agroquímica y Tecnología de Alimentos. CSIC

Apdo 73 46100 Burjassot (Valencia). Dir-el: [carbea@iata.csic.es](mailto:carbea@iata.csic.es)