

Herramientas avanzadas para la búsqueda de información médica en el web

I.F. Agullo

Introducción

Los contenidos depositados en la red Internet son ya, por volumen, accesibilidad, calidad y variedad e incluso coste, el recurso de información en medicina más importante. El tremendo potencial de la red no es ya sólo valioso para los profesionales, sino que en los últimos años pacientes y público en general han encontrado respuesta, generalmente rigurosa y fidedigna, a sus dudas y preguntas.

Ello no impide que exista cierta preocupación tanto por el explosivo crecimiento de la información ofrecida, a veces a costa de su calidad, como por la dificultad que plantea la recuperación de recursos críticos entre la maraña de datos que se proporcionan. La recuperación de información médica en la red requiere herramientas, técnicas y especialistas que permitan incrementar y profundizar en los resultados básicos obtenidos desde los portales de gran público, que se han generalizado como fuentes primarias en los últimos meses. Aun reconociendo la indudable valía de muchos de ellos, cierta ligereza en su uso contrasta con la necesaria rigurosidad que un profesional médico o un intermediario especialista deben exigir en las tareas que recaban información de calidad.

El portal, aunque resulta ser una vía de entrada cómoda y potente, no deja de ser un conjunto de gran diversidad y heterogeneidad en cuanto a la forma y fondo de los contenidos. A menudo se olvida este carácter y se asume equivalente nivel de calidad a todos los recursos presentes, lo cual puede distar considerablemente de la realidad. Por ello es conveniente recurrir directamente a las herramientas individuales, conocer sus métodos de selección y producción, su estructura y contenidos y los mecanismos de localización y recuperación de información. Profundizar en las características específicas de estas herramientas y dominar el manejo de varias de ellas es un reto cuyo cumplimiento es cada vez más deseable entre los profesionales de la salud.

Estas herramientas son servicios muy conocidos entre los usuarios de Internet e incluyen tanto los motores de búsqueda como los índices y directorios. Los más populares entre los primeros suelen ser los generales (Altavista, FAST, Google), con una gran cobertura de contenidos, pero que ofrecen mucho ruido documental en sus respuestas por la falta de precisión de sus mecanismos de recuperación. Todo lo contrario ocurre con los segundos, donde se prefieren los especializados, dada su mayor calidad en las respuestas y el valor añadido de una descripción realizada

por profesionales, es decir, no de forma automática. Aunque haya índices generales muy utilizados (Yahoo, Lycos, Terra), los servicios especializados, auténticas supersedes, tanto por la calidad de la información como por la fiabilidad de las fuentes, suelen ser un referente básico en la recuperación de información médica de la red. Aunque frecuentemente tienen carácter internacional o cobertura geográfica amplia, no pueden ser exhaustivos en la recopilación de los recursos disponibles en la red. Por ello, a la hora de elegir herramientas para trabajar en la red (no necesariamente aquellas «impuestas» en un determinado portal), además de evaluar criterios tales como el seguimiento de normas o estándares, la frecuente actualización y la incorporación de uno o varios mecanismos de evaluación de la pertinencia, debe asumirse la necesidad de utilizar varias de las citadas herramientas de forma simultánea o consecutiva para solventar las carencias individuales. Por tanto, antes de proceder a la resolución de una necesidad concreta, es importante evaluar y seleccionar entre las diferentes herramientas disponibles aquella o aquel grupo que mejor se adapten a nuestra petición. En la tabla 1 se ofrece un esquema que puede ayudar a realizar dicha selección. Se trata de un modelo general, que puede servir de guía, aunque la selección final requiere un análisis más específico del caso concreto a resolver.

La inspección de la tabla 1 muestra que resulta muy importante el conocimiento de los mecanismos de recuperación de información en los motores de búsqueda generales, puesto que su cobertura de una determinada especialidad es superior incluso a la de herramientas más especializadas. El filtrado de la información mediante una adecuada estrategia de búsqueda resulta en estos casos clave para reducir el ruido documental que producen los buscadores. En la tabla 2 hemos presentado una guía práctica

Laboratorio de Internet. CINDOC-CSIC. Madrid. España.

Correspondencia:
I.F. Agullo.
CINDOC-CSIC.
C/ Joaquín Costa, 22.
28002 Madrid. España.

Correo electrónico:
isidro@cindoc.csic.es

TABLA
1

Herramientas de recuperación de primera generación basadas en búsqueda (motores) y navegación (índices)

Búsqueda de un término genérico o disciplina

	Focalizando la navegación	Primer criterio	Segundo criterio	Otros criterios
Elija un índice o directorio	El mejor índice mundial	El mayor índice mundial: 1. Open directory. 2. Yahoo	Indización por especialistas: 1. Open directory. Ordenación según criterios de impacto: 1. Google directory (mirror of od)	
	El mejor índice regional (p. ej., español)	El mayor índice español: 1. Terra (Olé) 2. Yahoo España		Otros índices (nativos) «Mirrors» regionales en español
	El mejor índice especializado (p. ej., médico)	El mayor índice especializado		Índices muy especializados
Problemas con los índices	Difícil localización de índices especializados	Consulta de directorios de índices: 1. Virtual library. 2. Argus clearinghouse 3. Yahoo	Directarios analíticos. Vortales	
	Categorías confusas o de difícil navegación	Utilización del buscador interno. Visualización del árbol de navegación		Sinónimos
Recursos de calidad	Acceda a la Internet invisible			
Búsqueda de un término concreto				
Elija un motor de búsqueda	El mejor buscador mundial	El mayor buscador mundial: 1. Google. 2. Fast. 3 Altavista. 4. Northernlight	Operadores avanzados: 1. Altavista. 2. Fast	Recursos relacionados o similares: 1. Google. 2. Altavista
	El mejor buscador regional (p. ej., español)		Delimitadores por dominios e idiomas: 1. Google 2. Altavista Ordenación y agrupación de resultados: 1. Google 2. Northernlight	Operadores y delimitadores combinados: 1. Altavista 2. Fast El mejor buscador especializado (p. ej., médico) Actualidad: selector por fechas: 1. Altavista. 2. Northernlight
Problemas con buscadores	Ruido documental	Buscadores muy especializados: directorios de buscadores		
	Silencio documental	Multibuscadores y metabuscadores		
Archipiélagos de calidad		Internet invisible: fuentes de referencia		

de estrategias genéricas que resultan especialmente útiles en la recuperación de información en un *corpus* documental de texto completo, tan amplio y heterogéneo como es Internet.

Asimismo la tabla 1 recoge una importante, muy conocida y utilizada, parte de Internet, la constituida por las grandes bases de datos bibliográficas de medicina, los catálogos de bibliotecas, el creciente número de revistas electrónicas y los nuevos depósitos de documentos y separatas, una tendencia emergente en los últimos años. Llamamos a este heterogéneo conjunto «Internet invisible», dado que su in-

formación, al estar accesible sólo a través de pasarelas, resulta imposible de indizar por los motores de búsqueda y es por tanto invisible a éstos. Es una información de elevada calidad y a menudo de difícil localización, por lo que en la tabla 3 se describe no sólo su estructura y clasificación, sino las herramientas de recuperación más adecuadas en cada caso.

Herramientas de segunda generación

Sin embargo, las herramientas citadas anteriormente no son capaces, y posiblemente demuestren más esta debili-

TABLA
2

Reducción de ruido documental en buscadores

Búsqueda en texto libre	Formule secuencias de términos factibles en documentos a texto completo mediante el uso de la adyacencia estricta (encerrando el texto entre comillas)
Posiciones privilegiadas	Utilice los delimitadores, bien individualmente (título) bien en conjunción con otros (dominio, fecha)
Uso efectivo de operadores	Haga uso de los operadores booleanos, aunque dando prioridad al operador de adyacencia facultativa (near) sobre el de co-ocurrencia (and)
Focalizar la búsqueda	Utilice el mayor número de términos posibles
Lenguaje natural	Aproveche la formulación de la pregunta en lenguaje natural cuando esté disponible
Opciones superavanzadas	Ante silencios documentales utilice operadores de expansión, relación o similitud (like, related)

TABLA
3Internet
invisible

Áreas	Componentes	Metaherramientas	Otras herramientas
Pasarelas web	Registros bibliográficos	Pasarelas Z39.50	Clientes Z39.50
	Obras de referencia	Directorios	
	Bases de datos	Multibuscadores: 1. Profusión «Deep Web»: 1. LexiBot. 2. BullsEye. 3. Copernic	Directorios
	Documentos formatos especiales	Buscadores: 1. PDFSearch 2. Google	Depósitos
	Revistas electrónicas	Directorios de títulos	Buscadores de artículos: 1. Comerciales. 2. Bibliográficos
Depósitos de software	Archivos multimedia	Buscadores: 1. Altavista. 2. Hotbot	Depósitos
	Shareware, freeware y similar	Buscadores, directorios y metabuscadores: 1. Shareware.com. 2. Tucows. 3. Softseek	
	Otros	Buscadores generales, índices especializados	
Usenet/Listas	Mensajes de correo	Buscadores de Usenet news. Directorios de listas: 1. RedIRIS. 2. Liszt	

dad en el futuro, de enfrentarse con el enorme crecimiento del volumen de información en la red. La tabla 4 indica que en el primer trimestre de 2001 están disponibles más de 3.500 millones de páginas, con un crecimiento estimado superior al 100% anual.

Los mayores motores búsqueda (Google, FAST o Altavista) apenas cubren un 20-25% del web, mientras que los principales índices es dudoso que lleguen a un 5%. Otras desventajas de este ingente volumen de información es que, por razones de celeridad en la respuesta, obliga a limitar las prestaciones de búsqueda de forma que suelen faltar ciertas capacidades avanzadas. Otros problemas importantes derivan de la diferente cobertura de la red (las sedes comerciales y de los países desarrollados están mejor indizadas), el elevado porcentaje de enlaces no activos y la desactualización de los recursos debido a frecuencia de revisión muy baja o inadecuada.

Las citadas herramientas están instaladas en el ordenador remoto (*server side*) y por tanto limitadas por restricciones generalmente ajena al usuario final. Una nueva generación de herramientas y la adopción de nuevas estrategias pueden ayudar significativamente, así como el reconocimiento de nuevas realidades y el descubrimiento de fuentes ocultas de datos relevantes hasta la fecha frecuentemente infráutilizados.

Las herramientas de segunda generación, instaladas en el ordenador cliente (*client side*) son capaces de tratar con grandes volúmenes de información, automatizando tareas que incrementan la productividad final de los recursos recuperados. Comprobamos las prestaciones ofrecidas por este tipo de programas en la resolución de problemas de información relacionadas con salud y medicina. Nuestra clasificación reconoce 6 grupos de acuerdo con las tareas específicamente documentales que abordan: clientes Z39.50, volcadores, multibuscadores, agentes trazadores, indizadores y mapeadores (tabla 5).

TABLA
4Tamaño y distribución según grandes áreas del WWW
(datos elaborados de diferentes fuentes)

Wespacio	Sedes	Contenidos
Comercial	85%	68%
Académico	6%	20%
Gubernamental	1%	5%
Salud	1%	2%
Otros	7%	5%
Total	> 60 millones	-5.000 millones de páginas web

TABLA
5Tareas especializadas: herramientas
de segunda generación

Elaboración de bibliografías	Clientes Z39.50
Archivo automático de documentos	Volcadores
Elaboración de directorios	Multibuscadores, trazadores
Elaboración de bases de datos de recursos	Indizadores
Descripción cuantitativa de sedes web	Mapeadores

Clientes Z39.50

Hemos comentado anteriormente la importancia de la llamada «Internet invisible», una fracción muy voluminosa de información que aunque accesible a través de la red por distintas razones no es indizada por los motores de búsqueda. Desde hace algunos años se ha ido generalizando el uso del protocolo Z39.50, que permite interrogar un elevado número de servidores de información simultáneamente y con el mismo y único interfaz y lenguaje de interrogación. Así, puesto que, por ejemplo, la Biblioteca Nacional de la Medicina (creadora de Medline) tiene instalado un servidor Z39.50, resulta posible recuperar registros bibliográficos médicos tanto de artículos como de mo-

TABLA 6 Algunos clientes
Z39.50

Bookwhere 2000 4.0	Sea Change (www.bookwhere.com)
EndNote 4.0	Research Information (www.endnote.com)
ZNavigator 1.0h	EnWare (www.enware.es)
ZSearcher 1.1	Integral Concepts (www.integralconcepts.com)
ZPista 1.07	Ifigenia Plus (www.ifigenia.es/zeta)

nografías de la citada Medline y varias docenas de los mejores catálogos de bibliotecas médicas del mundo.

Los programas que permiten realizar esta proeza son estrictamente clientes (tabla 6) a los que se puede configurar para atender simultáneamente cientos de peticiones. Bookwhere 2000 (fig. 1) es una herramienta excelente, que resulta potente tanto para búsquedas concretas como para la elaboración de listados y bibliografías.

Volcadores

Uno de los más importantes problemas desde el punto de vista documental de Internet es la gran volatilidad de los recursos, lo que unido a un inexistente sistema de archivo hace que información muy relevante pueda perderse o ser inaccesible en períodos de tiempo muy cortos. Se ha desarrollado una serie de programas que permiten no sólo almacenar automáticamente sedes con un gran volumen de información, sino que mantienen la estructura y elementos dinámicos de las páginas, respetando así su integridad y reflejan fielmente su objetivo original (tabla 7).

El sistema es muy eficaz y, dado el discreto tamaño físico de la mayoría de las sedes, permite mantener un archivo «offline» de un elevado número de páginas con mecanismos elaborados de actualización de la información y programación de la labor de volcado que alcanza elevados niveles de automatismo y sofisticación. Así, resulta posible el archivo

periódico de aquellas sedes médicas que se actualizan con mucha frecuencia sin necesidad de la intervención directa del usuario, que puede fijar la hora de volcado a períodos de poco tráfico y bajo coste (horario nocturno), tarea que es realizada y controlada automáticamente por el programa. Nuestra recomendación es Teleport Pro (fig. 2), que permite ejecutar el volcado utilizando hasta 10 agentes simultáneamente, lo que lo convierte en una de las herramientas más rápidas del mercado.

Multibuscadores

Este grupo de programas supone uno de los intentos más logrados de mejorar el rendimiento de los motores de búsqueda, mediante el sencillo procedimiento de realizar la recuperación en varios de dichos motores de forma simultánea. Aunque tal procedimiento ya está disponible en forma de multibuscadores (tabla 8) «tradicionales», la segunda generación proporciona la ventaja de la perspectiva cliente: los programas multibuscadores disponen de opciones de automatización de tareas y son extremadamente flexibles en su configuración, lo que les permite resolver labores específicas. Esta posibilidad es particularmente interesante en temas médicos, pues permite seleccionar las fuentes contra las que realizar la interrogación formando grupos de motores especialmente perfilados según temáticas o cualquier otro criterio que se estime conveniente. Uno de los programas más interesantes es Copernic (fig. 3), que permite no sólo diseñar e interrogar grupos de motores a voluntad, sino que elimina duplicados y direcciones erróneas, filtra con criterios definibles, vuelca contenidos y exporta los resultados en formatos reutilizables.

Trazadores (tabla 9)

Este grupo explora la naturaleza hipertextual del web, por lo que al contrario que en el caso anterior los resultados se obtienen por rastreo o exploración del árbol hipertextual a

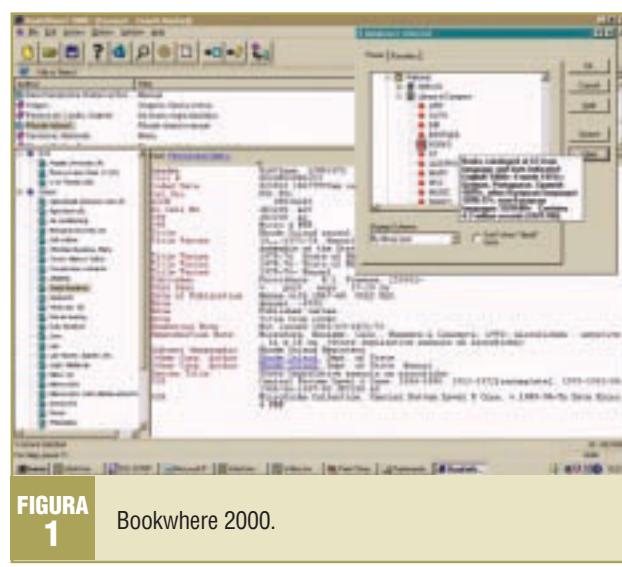
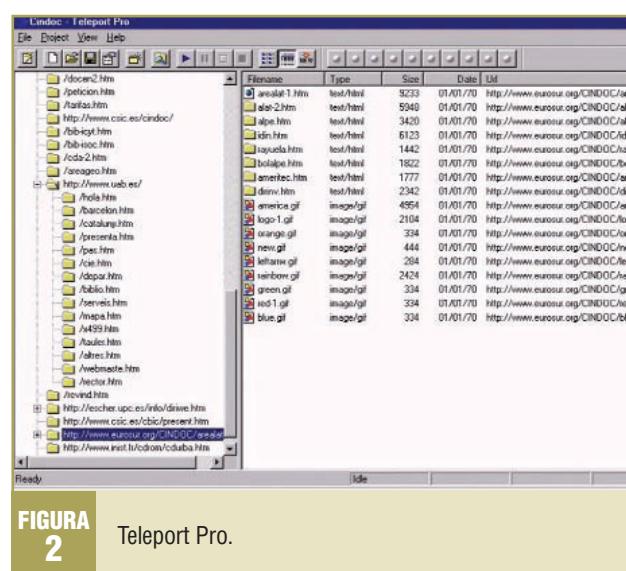

FIGURA 1 Bookwhere 2000.

FIGURA 2 Teleport Pro.

TABLA
7Programas para el volcado automático
de sedes web

BackStreet 2.7	Soft. Resolutions (softwaresolutions.net/backstreet)
BlackWidow 4.14	Softbyte Labs (www.softbytelabs.com)
Ecatch 3.0	LaMine (www.ecatch.com)
Httpweazel 1.10	(www.httpweazel.com)
HTTrack Website Copier 3.15	Xavier Roche (www.httrack.com)
Internet Angel 5.0	Tassilo Rinder (ineta.de)
Internet Search Agent 2.0	Renegade (www.renegade-software.com/ISA)
JOC WebSpider 2.90	JOC Software (www.jocsoft.com)
Leech 3.2.4	Aeria (www.aeria.com)
MemoWeb 4.0	GoTo Software (www.memoweb.com)
Offline Explorer Pro 2.1	Metaproducts (www.metaproducts.com)
SuperBot 2.60	Elytesys (www.sparkleware.com/superbot)
Surf Saver 2.2	askSam Systems (www.surfsaver.com)
Teleport Pro 1.29	Tennyson Maxwell (www.tenmax.com)
WebCopier 3.0	(home.columbus.rr.com/mklimov)
Webleech 2.4	Voodoo Software (www.voodoo-software.com)
WebMirror 1.40	MaccaSoft. (www.maccasoft.com/webmirror)
WebReaper 9.8	Mark Otway (www.webreaper.net)
Website Extractor 7.8	Asona (www.asona.org)
WebStripper 2.02	Solent Soft (www.solentsoftware.com/webstripper)
WebVCR 2.1	Net Results (www.netresultscorp.com)
WebWhacker 2000 4.0	Blue Squirrel (www.bluesquirrel.com)
WebZip 4.1	Spidersoft (www.spidersoft.com)

partir de una sede determinada. Dicha sede, que recibe el nombre de «semilla», puede ser cualquier página web especializada con un elevado número de enlaces (una supersede médica) o un buscador en el que se ha introducido una estrategia de búsqueda. A partir de ahí se comprueban

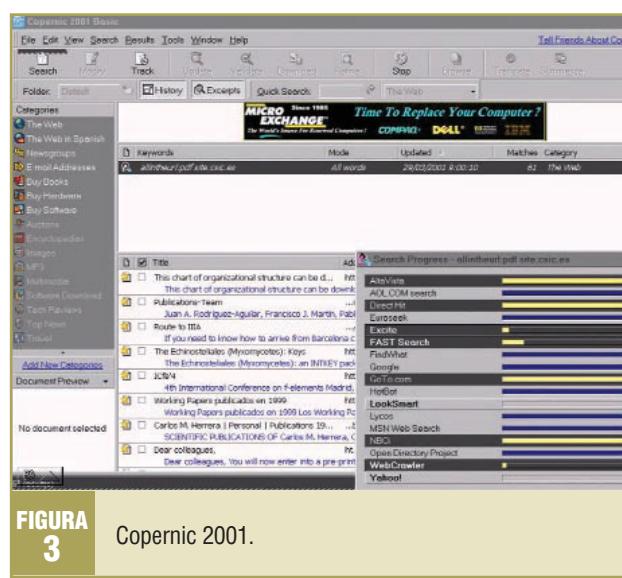
todos las páginas enlazadas con la primera (hijas) de acuerdo a una serie de criterios de pertinencia que pueden estar ponderados. Aquellas seleccionadas se utilizan como fuente para una exploración hipertextual ulterior (nietas) y así sucesivamente hasta un nivel prefijado o hasta agotar el tiempo programado.

Esta técnica indirecta es lenta y proclive a bucles y ruido documental, pero resulta muy valiosa para encontrar respuestas inesperadas. Hay que indicar que el rendimiento mejora considerablemente con una adecuada retroalimentación del sistema (aprendizaje).

Este grupo es el más heterogéneo de los presentados, ya que no todos los programas ofrecen las mismas posibilidades. Las semillas suelen ser los grandes buscadores, mientras que los mecanismos de evaluación de la pertinencia pueden ser muy pobres o extremadamente sofisticados sistemas multilingües.

Indizadores

Se trata de programas desarrollados a partir de los multibuscadores y que a partir de los resultados obtenidos, mediante un módulo de indización, ofrecen una serie de palabras clave descriptivas de los contenidos de las sedes seleccionadas o incluso un pequeño «resumen»

FIGURA
3

Copernic 2001.

TABLA
8

Un breve listado de multibuscadores

Answwwer 2.1	(www.answwwer.net)
Bingooo 1.09	Bingooo (www.bingooo.com)
CiteLine 3.5	Caredata (www.citeline.com)
Copernic 5.0	Agents Technologies (www.copernic.com)
Delta Dark 2.1	(www.scorpsoft.com)
Direct Seek 2.3	Direct Seek (www.directseek.com)
EZSearch 3.0	American Systems (www.americansys.com)
FirstStop 3.0	Brush Systems (www.firststopwebsearch.com)
Geowhere 1.22 (www.geowhere.com)	
Glacino 1.1	Glacino (www.glacino.com)
Golden Eye 007 2.0	(www.6bytes.com)
InfraNET 2000 1.75	Infradev (infradev.freeshell.org)
Internetowy Poszukiwacz 2000 3.0	(www.poszukiwacz.pl)
JOC WebFinder 2.60	JOC (www.jocsoft.com)
MetaEureka 1.0	4C Software (www.metaeureka.com)
NetPuller 2.52	(www.wutils.com)
Patrol Search 1.3 (www.patrolsearch.com)	
SearchLight 2.11	SpyCatcher (www.spycatcher.com)
SearchPad 1.61	LeisureSoft (searchenginesoftwaretools.com)
SearchRocket 2.0	SearchCube (www.searchrocket.com)
SearchWolf 2.03	Trellian (www.msw.com.au/searchwolf)
Simple Find 2.02	ACD Soft (www.adcsoft.com/simplefind.htm)
SurfPilot 1.01	Invention Office (www.surfpilot.net)
WebFerret 4.0	Ferretsoft (www.ferretsoft.com)
WebFinder 2	AAA.soft (www.trionex.net)
WebSeeker 5.0	Blue Squirrel (www.bluesquirrel.com)

construido mediante extracción de frases significativas del texto (tabla 10).

Estos extractos no pueden utilizarse directamente, pero ayudan en la preparación de resúmenes por parte de indizadores humanos. Algo parecido ocurre con los descriptores, aunque en general la calidad de éstos es suficiente para que una revisión de frecuencias baste para corregir la indización. Un importante valor añadido de estos programas es su ca-

pacidad para exportar los resultados en formatos reutilizables por paquetes ofimáticos o editores web.

Es muy probable que estos programas evolucionen muy rápidamente en un futuro cercano, ya que los algoritmos de indexación son cada vez más potentes. Ello permitirá mantener grandes directorios de recursos realizados de forma semiautomática y actualizados periódicamente sin apenas intervención humana.

TABLA
9

Agentes trazadores

CISpider 1.21	Artificial Intelligent Lab (ai.bpa.arizona.edu/go)
Cmore 2.2	CSS Group (www.cmoreweb.com)
DigOut4U 1.5	ARISEM (www.arisem.com)
Personal Search Agent 0.8	DataVillage (www.datavillage.com/psa.htm)
StrategicFinder 2.2	Digimind (www.digimind.com)
URL Spider Pro 2.0	Innerprise (www.innerprise.net)
WebBandit 4.45	JW Software Gems (www.1billionsites.com/webbandit.htm)
WebStorm 3	SharpeWare (www.sharpeware.com)
WebWolf 2.03	Trellian (www.msw.com.au/webwolf)

TABLA
10Programas que seleccionan palabras clave
o realizan «resúmenes»

AnswerChase Professional 4.0	AnswerChase (www.answerchase.com)
BullsEye Pro 3.0	Intelliseek (www.intelliseek.com)
Copernic Summarizer 2.0	Copernic (www.copernic.com)
iCrossReader 2.8	Insight (www.insight.com.ru)
Lexibot 2.0	BrightPlanet (www.lexibot.com)
Subject Search Spider 2.30	Kryloff Technol (www.kryltech.com)

TABLA
11Programas para la descripción cuantitativa
de sedes

Astra Site Manager 2.0	Mercury Interactive (www.merc-int.com)
Blueprint 1.2.6	Exit 0 (www.exit0.com)
Content Analyzer 3.0	Microsoft Co. (www.microsoft.com)
Custo 1.9	(www.netwu.com)
Funnel web Profile 2.0	(www.quest.com)
Intellitamper 2.08	(www.interllitamper.com)
Lighthouse (Java)	Anton Leuski (toowoomba.cs.umass.edu/~leuski/lighthouse)
LinkBot Pro 6	Watchfire (www.watchfire.com)
LinkViewer 3.0	GradeTools (www.gradetools.com)
Power Mapper 2.0	Electrum Multimedia (www.electrum.co.uk)
Tree Studio 1.5	Inxight Soft (www.inxight.com)

Mapeadores

Aunque originalmente diseñados para asistir a los editores de páginas web, los programas mapeadores proporcionan descripciones gráficas («mapas») y cuantitativas muy detalladas de las sedes y sus contenidos (tabla 11). Los datos numéricos que se obtienen pueden ser utilizados para evaluar dichos contenidos y constituyen una importante ayuda para discernir la calidad hipertextual de una sede. Las relaciones que se ilustran a través de los enlaces y la calidad de su mantenimiento son fuentes adicionales de información sobre la pertinencia y utilidad del recurso analizado.

Los datos extraídos se pueden utilizar con fines comparativos, ya que los valores absolutos tienen relativo valor, pero aun así podemos obtener valores del tamaño, densidad hipertextual y multimedia de la sede, su estructura de niveles, la diversidad de los enlaces y en estudios temporales la evolución del conjunto.

Conclusiones

El problema de la recuperación de información médica en Internet se agudizará en el futuro debido a la explosión de la información en el web, fenómeno que no alcanzará una asymptota en el crecimiento ni a corto ni a medio plazo. Dicho problema se agravará con un incremento asimismo incontrolado de la heterogeneidad de la calidad de dicha información. Sólo es posible solventar ambos retos con un acercamiento profesional que implica conocer mejor las herramientas

disponibles y aplicar criterios más rigurosos y exhaustivos. La revisión de métodos presentada se inicia proponiendo una mejor evaluación y selección de herramientas, que no son equivalentes ni intercambiables, como a menudo se ha sugerido. Cada índice o buscador tiene sus ventajas e inconvenientes y algunos son más apropiados que otros para determinadas tareas. Incluso muy a menudo será necesario el recurso simultáneo a varias herramientas para cubrir las carencias de cobertura que presentan todas ellas.

Dado el carácter crítico de la información sobre medicina y salud, el valor añadido de la indización humana por parte de expertos debe ser tenido muy en cuenta. Aunque los motores generalistas resultan paradójicamente más apropiados para búsquedas muy específicas, son los índices y directorios las herramientas recomendadas para identificar recursos especializados. Sin embargo, incluso una selección cuidadosa deja en evidencia las carencias de muchas de las herramientas disponibles en los servidores remotos, por lo que la disponibilidad de clientes especializados, las llamadas herramientas de segunda generación, ha permitido afrontar con ciertas garantías de éxito el control individualizado del ciberespacio de la salud. Estas herramientas, entre las que se incluyen robots y agentes, permiten automatizar tareas, profundizar en la cobertura de ciertas labores, incrementar el rendimiento de la recuperación y, en fin, hacer viables proyectos de localización y recuperación de datos que de otra manera no serían posibles realizar desde el escritorio de un usuario final que no domine la programación avanzada.

Importante consecuencia de esta nueva potencia y flexibilidad es la libertad obtenida para una evaluación más rigurosa y estricta de la información recabada, labor esta última para la que también estas herramientas ayudan y que resulta clave en el universo web en el que a partir de ahora debemos desenvolver nuestra actividad profesional.

Bibliografía general

- Aguillo IF. Herramientas de segunda generación. Anuario SOCA-DI. Barcelona, 1998; p. 85-112.
- Aguillo IF. Del multibuscador al metabuscador: los agentes trazadores en Internet. IV Congreso ISKO España, Granada, 1999; p. 239-45.
- Aguillo IF. Internet invisible o infranet: definición, clasificación y evaluación. VI Jornadas Españolas de Documentación, Bilbao, 19-21 de octubre de 2000; p. 249-69.
- Aguillo IF. A new generation of tools for search, recovery, and quality evaluation of world wide web medical resources. J Manag Med 2000;14:240-8.
- Eysenbach G, Diepgen T. Towards quality management of medical information on the internet: evaluation, labelling, and filtering of information. BMJ 1999;317:1496-502.
- Hipola P, Montes A. Descripción y evaluación de agentes multibuscadores. El Profesional de la Información 2000;8:15-26.
- Maldonado A, Fernández E. Análisis comparativo de buscadores en Internet. El Profesional de la Información 2000;9:40-6.
- Tramullas J, Olvera MD. Recuperación de la información en Internet. Madrid: RA-MA, 2000.