

CLASIFICACIÓN DE RECURSOS Y FUENTES DE INFORMACIÓN DE CIENCIAS DE LA SALUD EN INTERNET

M.^a LUISA MAQUEDANO MARTÍNEZ

Doctora en Ciencias de la Información
Hospital Puerta de Hierro de Madrid

Resumen: En el presente artículo se ofrece una clasificación de recursos y fuentes de información de los servidores de INTERNET en el área de Ciencias de la Salud, que es el resultado parcial, revisado y actualizado de la tesis doctoral «Biblioteca Virtual de Ciencias de la Salud». Asimismo se explican los criterios utilizados para la creación de los grupos que integran la clasificación, y se comentan las dos grandes dificultades encontradas al revisar las páginas Web: el creciente número de servidores y la carencia de un criterio único para indizar el contenido. El objetivo principal es proporcionar al profesional sanitario una herramienta útil para la búsqueda y recuperación de la información en los servidores de INTERNET, y el URL para su acceso es: <http://web.jet.es/marisa>.

Palabras clave: Ciencias de la Salud, Búsqueda y recuperación de la información, INTERNET.

Abstract: This article presents a classification of resources and information sources of Health Sciences in the INTERNET Web servers. It forms part of the revised and updated version of the dissertation "Health Sciences Virtual Library". It also deals with the two most important problems concerning information search and retrieval in the Internet: the increasing number of servers and the lack of and indexing vocabulary. The main aim of this article is to provide health professionals with a useful tool for information search and retrieval in INTERNET Web servers. It is available from the URL: <http://web.jet.es/marisa>.

Key words: Health Sciences, Seeking and information retrieval, INTERNET.

1. INTRODUCCIÓN

Las razones de la elaboración esta clasificación surgen de la observación de los problemas reales a los que se enfrentan los profesionales de la salud a la hora de localizar la información sanitaria, obtener el documento y editar sus trabajos de investigación en el soporte tradicional en papel. Todo ello se debe por un lado, al incesante crecimiento de la documentación científica, especialmente de los artículos publicados en el área de Ciencias de la Salud, puesto que la obsolescencia de las investigaciones está íntimamente ligada a su carácter de ciencia empírica. Y por otro lado, a la creación de las nuevas herramientas destinadas a recuperar esa información. En la actualidad, el factor del desarrollo industrial aplicado a las nuevas tecnologías de la informática y las redes de comunicaciones suscita un nuevo enfoque en la forma de acceder a la literatura científica, y en consecuencia, una nueva forma de abordar el tema del acceso a las fuentes del saber científico.

La presente clasificación¹ en la página Web del proveedor de INTERNET, Jet Internet, pretende proporcionar al profesional sanitario una herramienta útil para el acceso online a la información del área de Ciencias de la Salud, y facilitarle la comunicación con la comunidad científica. Su principal objetivo es evitar el desánimo, la pérdida de tiempo y el alto coste económico de las comunicaciones que procura toda búsqueda ineficaz, especialmente en la red INTERNET que ha experimentado un crecimiento inusitado en la última década. La posibilidad de transferir todo tipo de datos binarios que ofrece la aplicación WWW a través de cualquier «navegador», permite al usuario copiar la presente clasificación en el disco duro de su ordenador personal, o seleccionar los URL de determinados grupos para organizar y ampliar la información en la categoría más adecuada del área de su interés.

2. METODOLOGÍA

En primer lugar se han realizado búsquedas en servidores WWW de reconocido prestigio, tanto en los de carácter general como en los especializados en Ciencias de la Salud. Los servidores utilizados en la fase de inicio se relacionan más abajo. El contenido de la información proporcionada

¹ MAQUEDANO, M. L.: "Clasificación de recursos y fuentes de información de Ciencias de la Salud". *Jet Internet*, 1997-98. Disponible en: URL: <http://web.jet.es/marisa>.

por cada uno de ellos, se ha analizado exhaustivamente y se ha accedido a todos y cada uno de los hiperenlaces, incorporando los campos URL, Home Page y Copyright a una base de datos relacional con una clasificación preestablecida. Esta operación se ha repetido sucesivamente con todos los hiperenlaces de aquellos servidores que ofrecían información relacionada con distintos recursos y fuentes de información sanitaria en INTERNET.

Debido a las frecuentes modificaciones que experimentan los URL de los servidores de INTERNET, se ha revisado y actualizado la siguiente lista a fin de que resulte operativa para el usuario:

- Home page: Centro de Comunicaciones CSIC-RedIRIS²
URL: <http://www.rediris.es/>
- Home page: NetCenter³
URL: <http://home.netscape.com/index.html>
- Home page: W3 Search Engines⁴
URL: <http://cuiwww.unige.ch/meta-index.html>
- Home page: U.S. National Library of Medicine⁵
URL: <http://www.nlm.nih.gov/guides.html>
- Home page: US National Network of Libraries of Medicine Home page⁶
URL: <http://www.nlm.nih.gov/>

² RedIRIS. Centro de Comunicaciones CSIC-RedIRIS. RedIRIS 1998. Disponible en: URL: <http://www.rediris.es/>.

³ Netscape. NetCenter. Netscape 1998. Available from: URL: <http://home.netscape.com/index.html>.

⁴ CUI. W3 Search Engines. CUI 1997. Available from: URL: <http://cuiwww.unige.ch/meta-index.html>.

⁵ NLM. U.S. National Librry of Medicine. National Library of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/nlmhome.html>.

⁶ NN/LM. US National Network of Libraries of Medicine Home page. National Network Libraries of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/>.

- Home Page: NN/LM's Guide to Internet Discovery Tools⁷
URL: <http://www.nlm.nih.gov/tools.html>
- Home page: Health Information Resources Guides⁸
URL: <http://www.nlm.nih.gov/guides.html>
- Home page: Document Delivery Suppliers⁹
URL: <http://www.nlm.nih.gov/pnr/docsup/>
- Home page: McGill Health Sciences and Osler Libraries: Health Sciences and Internet Guides¹⁰
URL: <http://www.health.library.mcgill.ca/resource/hsintgui.htm>
- Home page: Health Sciences Resources on the Internet¹¹
URL: <http://www.health.library.mcgill.ca/resource/page.htm>
- Home page: Yahoo - Health¹²
URL: <http://www.yahoo.com/Health/>
- Home page: WWW VL: Biosciences: Medicine: Other Medical Indices¹³
URL: http://www.ohsu.edu/clinweb/wwwvl/other_med.html

⁷ NN/LM. NN/LM's Guide to Internet Discovery Tools. National Network Libraries of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/tools.html>.

⁸ NN/LM. Health Information Resources Guides. National Network Libraries of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/guides.html>.

⁹ Shipman J. Document Delivery Suppliers. NN/LM's Document Delivery Service Information. National Network Libraries of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/pnr/docsup/>.

¹⁰ Mc Gill University Health Sciences & Osler (History of Medicine) Libraries. McGill Health Sciences and Osler Libraries: Health Sciences and Internet Guides. Mc Gill University Health Sciences & Osler (History of Medicine) Libraries. Available from: URL: <http://www.health.library.mcgill.ca/resource/hsintgui.htm>.

¹¹ GAROUFALIS, A.: *Health Sciences Resources on the Internet*. Mc Gill University Health Sciences & Osler (History of Medicine) Libraries 1998. Available from: URL: <http://www.health.library.mcgill.ca/resource/page.htm>.

¹² Yahoo! Inc. Yahoo - Health. Yahoo! Inc. 1994-98. Available from: URL: <http://dir.yahoo.com/Health/>.

¹³ APPELYARD, R.: *WWW VL: Biosciences: Medicine: Other Medical Indices*. Biomedical Information Communication Center, Oregon Health Sciences University, 1996.

- Home page: Library and information science, librarianship finding internet resources in this subject area¹⁴

URL: <http://www.ub2.lu.se/lisres.html>

Los criterios aplicados a esta clasificación de recursos y fuentes de información de los servidores de INTERNET en el área de Ciencias de la Salud, son criterios documentales modificados por los nuevos tipos de soporte telemático: las bases de datos en soporte electrónico que la tecnología informática ha incorporado para el almacenamiento y recuperación del documento; y por los nuevos medios de acceso a la información: las aplicaciones y herramientas de las redes de telecomunicaciones, especialmente de la red INTERNET.

En esta clasificación las definiciones de cada grupo se han realizado a posteriori, es decir, después de analizar el tipo de información ofrecido en las páginas Web a través de los hiperenlaces de los servidores mencionados y observar qué características tenían en común para tratar de unificarlas e incluirlas en un grupo o subgrupo determinado, con el fin de facilitar la búsqueda. La mayoría de los servidores titulan sus páginas con sinónimos del tipo, guías, índices, directorios, etc., indistintamente, pero mientras unos presentan listas de 'sitios', otros ofrecen acceso a bases de datos, y los más una clasificación arbitraria de aquellos recursos propios o ajenos, que resultan interesantes al autor de la misma. Sin olvidar, que la opción del hiperenlace dificulta aún más la sistematización.

Conviene aclarar, puesto que esta clasificación parte de estos conceptos, que el término «fuentes de información» hace referencia al tipo de documentos según el nivel informativo en soporte electrónico, y el término «recursos» a los distintos sistemas telemáticos: aplicaciones y herramientas de INTERNET, teniendo en cuenta que los «recursos» son en ocasiones, recurso y fuente de información a la vez, como es el caso de algunos «buscadores», que además de ejecutar la búsqueda sistematizan la información.

Debido a la continua evolución de los diferentes tipos de información que aparecen en los servidores de Ciencias de la Salud, o a los intereses de los usuarios, es probable que en futuras revisiones esta clasificación sufra modificaciones, ya sea añadiendo o eliminando grupos; o bien, modificando el nombre de algunos de sus componentes.

¹⁴ TRAUGOTT KOCH: Library and information science, librarianship finding internet resources in this subject area 1998. LUND UNIV LIBRARY, SWEDEN. Available from: URL: <http://www.ub2.lu.se/lisres.html>.

3. CLASIFICACIÓN

A) SEGÚN EL TIPO DE RECURSOS (*RESOURCES*)

1. *Aplicaciones*

Sistemas telemáticos que se utilizan para el acceso online a los diferentes tipos de recursos y fuentes de información de la red INTERNET. La mayoría de las aplicaciones funcionan con el modelo «cliente-servidor», en el que existe un «programa-cliente» en el ordenador del usuario que interactúa con el «programa servidor» del equipo remoto. Esta forma de conexión está regulada por protocolos normalizados que permiten la comunicación entre distintos tipos de ordenadores y sistemas operativos.

Según el tipo de protocolo estas aplicaciones reciben el nombre de:

- Servidores WWW
- Servidores Gopher
- Servidores FTP
- Servidores Telnet
- Servidores Usenet
- Servidores E-Mail

2. *Herramientas*

Son los diferentes tipos de sistemas de gestión de bases de datos online para la búsqueda de información en los servidores de INTERNET.

- *Buscadores*. Motores de búsqueda de todo tipo de recursos y fuentes de información de los servidores de INTERNET, que contienen su propia base de datos. Realizan dos tipos de búsqueda: por categorías y por palabras clave.
- *Metabuscadors*. Motores de búsqueda que realizan la búsqueda en diferentes buscadores, simultáneamente.
- *Gateways*. «Pasarelas» o interfaces para el acceso online a los sistemas de gestión de bases de datos (SGBD).

B) SEGÚN EL TIPO DE FUENTES DE INFORMACIÓN (*INFORMATION SOURCES*)

1. *Bases de datos*

Fuentes de información primaria o secundaria, clasificadas por tipo y contenido (*).

2. *Directorios*

En este apartado se han incluido aquellos servidores que contienen una clasificación tipológica de recursos y/o fuentes de información, según el concepto mencionado anteriormente. Se incluyen también algunos «buscadores», en razón de la importancia de la clasificación por categorías de su base de datos.

- Biblioteconomía
- Ciencias médicas
- Enfermería
- Farmacia & Farmacología
- Recursos generales

3. *Índices*

Estos servidores ofrecen listas de 'sitios' de un determinado tipo de recurso y/o fuente de información.

- Asociaciones y organizaciones académicas
- Bases de datos
- Editoriales
- Enfermedades
- Especialidades médicas
- Foros de debate

(*) La clasificación de las bases de datos no está incluida en la página Web.

- Historia de la ciencia, tecnología & medicina
- Hospitales
- Normas & Terminología
- Publicaciones electrónicas
- Telemedicina
- Web's & Opac's bibliotecarios

4. *Guías*

Recursos y/o fuentes de información específica de un host, que se ha incorporado a cada subgrupo por la relevancia informativa de la entidad en sí misma.

- Asociaciones bibliotecarias
- Asociaciones médicas
- Control de calidad
- Glosarios
- Hospitales virtuales
- Instituciones bibliotecarias
- Instituciones sanitarias
- Instituciones universitarias
- Investigación
- Política & Gestión sanitaria
- Productos & Servicios sanitarios
- Proveedores documento *online*
- Recursos información
- Revistas *online*
- *Software*
- Telemática bibliotecaria

4. INTERNET Y LA COMUNIDAD CIENTÍFICA

Todo parece indicar que el fulgurante avance tecnológico en el área de la informática y de las telecomunicaciones está produciendo un cambio fundamental en la vida científica. No es posible predecir el impacto de las nuevas tecnologías en la comunidad científica¹⁵, pero sí parece claro que inciden tanto en las nuevas formas del flujo de la información como en las relaciones entre los individuos de la comunidad científica, las instituciones y otros miembros de la sociedad. Esta “revolución” es connatural a la “juventud digitalizada” actual que oye la música en discos compactos, ocupa su tiempo de ocio con juegos informáticos y “navega” por INTERNET desde cualquier terminal u ordenador personal. Este convencimiento es la causa de que un gran número de expertos en estas nuevas tecnologías desarrollen herramientas cada vez más potentes y capaces de producir importantes cambios en el orden económico y social. La expansión de la industria de los ordenadores, de los discos compactos que permiten almacenar grandes volúmenes de información, de los nuevos productos informáticos y la creciente utilización de las redes con conexiones vía satélite o fibra óptica está transformando las relaciones sociales, comerciales y científicas entre los individuos y revolucionando el campo de la Medicina con estos nuevos ingenios electrónicos.

La Era de la Información como se conoce ya el siglo XXI se caracteriza por el desarrollo de las nuevas tecnologías de la informática y de las telecomunicaciones, que arrinconan el concepto de mausoleo de la biblioteca del siglo XIX y el de fuente enciclopédica del conocimiento de la biblioteca del siglo XX para dar paso al concepto de “biblioteca virtual”, como auténtica fuente global del conocimiento, y de ámbito universal en la cobertura y en el acceso. Ciertamente, la “biblioteca virtual” facilita, potencialmente, a cualquier usuario final el acceso *online* al conocimiento universal, pero si bien soluciona gran parte de las dificultades de que adolecen las bibliotecas tradicionales, no es menos cierto que el usuario se enfrenta a un medio hostil, el de las nuevas tecnologías, que es desconocido y evoluciona vertiginosamente. Su uso exige el aprendizaje del funcionamiento del *hardware* y el *software* teleinformático, de los sistemas de gestión de bases de datos, de los lenguajes de recuperación de la información en los diferentes tipos de bases de datos, y de las aplicaciones y herramientas de INTERNET. Este hecho, desmoralizante para la mayoría de los

¹⁵ GREENWOOD, M. R. C.: “Twenty-first century science”, *Bull Med Libr Assoc*, 1995, 83 (1), 8-13.

usuarios de la red, especialmente para los neófitos, que se enfrentan a un medio poco conocido, lo es aún más para los profesionales de la medicina, que necesitan localizar la información de una forma inmediata y precisa para el desarrollo de sus labores asistenciales y de investigación.

A este escollo hay que añadir, como ya se ha dicho reiteradamente en todos los medios de comunicación, el crecimiento desmesurado de los servidores de la red INTERNET desde finales de los ochenta, y de la consiguiente repetición de las fuentes, añadiría yo; pero el principal problema del usuario final para el acceso a la información tanto de carácter general como especializada, no radica principalmente en el creciente número de servidores, ni siquiera en la abundante información de todo tipo de productos y servicios que se producen diariamente, sino en la falta de un criterio unívoco a la hora de introducir los datos para su posterior recuperación. En términos documentales, es la carencia de un lenguaje documental o de indización de los documentos que se incorporan a las bases de datos.

Son muchas las utilidades que la aplicación de la Telemedicina en el campo de la Documentación Científica¹⁶ proporciona a los profesionales sanitarios, especialmente a través de la aplicación WWW, que no sólo sirven para buscar información sino también para crearla:

- Información y acceso a todo tipo de fuentes y recursos (bibliotecas, catálogos, hospitales virtuales, bases de datos, etc.).
- Comunicación con el resto de la comunidad científica mundial mediante el correo electrónico, listas de distribución y grupos de noticias.
- Intercambio de información mediante protocolos normalizados para recogida de datos de las actividades hospitalarias (sesiones clínicas, conferencias, estadísticas, etc.).
- Apoyo a la docencia mediante cursos de formación continuada.
- Ayuda a la investigación.
- Edición de las publicaciones en revistas electrónicas.
- Desarrollo de la informática médica y de los sistemas expertos para mejorar la asistencia clínica y quirúrgica.

¹⁶ ZUNDEL, K. M.: "Telemedicine: history, applications, and impact on librarianship". *Bull Med Assoc*, 1996, 84, 71-9.

5. EL PROBLEMA DE LA BÚSQUEDA Y RECUPERACIÓN DE LA INFORMACIÓN

El problema de la búsqueda y recuperación de la información surge ya en el siglo XVII en la llamada Revolución Científica¹⁷ con la constitución de un nuevo concepto de ciencia moderna que dinamiza el carácter estático del saber tradicional encapsulado hasta entonces en bibliotecas y monasterios como templos del conocimiento absoluto. El saber científico se caracterizó a partir de entonces, por el aspecto relativo y asociativo de los fenómenos observados en las hipótesis de trabajo, condicionadas a su vez por los factores ideológicos y culturales contemporáneos. Este aspecto provisional e inacabado de la ciencia da lugar a un nuevo instrumento de comunicación, la revista científica, que a partir del siglo XIX se convierte en el vehículo por excelencia de la difusión de las publicaciones científicas. La Revolución Industrial y la aparición de las nuevas tecnologías aplicadas al desarrollo de las máquinas y las telecomunicaciones ofrecen la posibilidad de generar y procesar masivamente la información y difundirla a todos los ámbitos de la sociedad, lo que provoca un crecimiento excesivo de publicaciones.

El crecimiento de la información científica no sólo se caracteriza por el envejecimiento y obsolescencia del contenido debido al dinamismo y relatividad de las explicaciones científicas, sino también porque su ritmo es mucho mayor que el de la mayoría de los fenómenos sociales. Así, en el siglo XIX se produce una auténtica explosión de la bibliografía médica en Europa, especialmente en Alemania a causa del aumento de las publicaciones en el formato de artículo de revista, que plantea ya un auténtico problema para la búsqueda y recuperación de la información.

La constitución de la Documentación Médica como disciplina coincide con las nuevas corrientes sobre la teoría de la Ciencia de la Documentación que inician Paul Otlet y Henri La Fontaine con la fundación del Instituto Internacional de Bibliografía en 1895 y la elaboración de un repertorio bibliográfico universal clasificado por materias y autores al que se aplicó el Sistema de Clasificación Decimal Universal inventado por el norteamericano Melvin Dewey.

¹⁷ LÓPEZ PIÑERO, J. M. y Terrada, M. L.: *La información científica en medicina y sus fuentes*. Valencia, Cuadernos de Documentación e Informática Biomédica, IX. Instituto de Estudios Documentales e Históricos sobre la Ciencia (Universitat de València-C.S.I.C.), 1993.

La mayor aportación para evitar la duplicación de trabajos mediante la aplicación de un sistema de clasificación a los documentos médicos la realizó John Shaw Billings, director de la National Library of Medicine¹⁸ (NLM) desde 1864, cargo que abandona treinta años más tarde dejándola convertida en una de las bibliotecas de medicina más importante del mundo. Diseña el Index Catalogue, que en 1950 había indizado más de medio millón de libros y folletos y dos millones y medio de artículos de revista. A partir de esa fecha el repertorio se conoce con el nombre de *National Library of Medicine Catalog*. Pero la gran contribución de Billings a la bibliografía médica fue la creación en 1879, de un índice de «bibliografía en curso», el Index Medicus, que después de diversas interrupciones en su publicación por factores económicos, se constituyó como el mejor repertorio bibliográfico de la literatura médica publicado hasta ahora.

En 1963 y después de tres años de diseño, la NLM crea uno de los primeros sistemas de gestión de bases de datos (SGBD) en línea, MEDLARS (Medical Literature Analysis and Retrieval System), que permite el acceso a toda la información biomédica almacenada en esta institución. En este ingenio provisto de la más alta tecnología informática se preparan y fotocomponen las publicaciones bibliográficas, del tipo Index Medicus. MEDLARS contiene más de 40 bases de datos de las que MEDLINE¹⁹ (Index Medicus online) es la más representativa. Las otras bases de datos dan información sobre catálogos, series, datos químicos y toxicológicos, materiales audiovisuales, cáncer y otras áreas especializadas de la medicina.

El sistema MEDLARS responde a un tipo de bases de datos, las llamadas documentales que permiten el almacenamiento y recuperación de grandes volúmenes de texto de longitud variable como puede ser la información bibliográfica. Pero el gran logro de NLM en el área de la clasificación de los documentos, es el proyecto de investigación Unified Medical Language System²⁰ (UMLS), que se inició en 1986 y cuyo propósito es ayudar a profesionales sanitarios e investigadores en el desarrollo de aplicaciones que recuperen e integren la información biomédica electrónica de una variedad de fuentes. El fin último es facilitar al usuario el acceso a diferentes sistemas de información, registros de pacientes en soporte informático, ba-

¹⁸ NLM. The National Library of Medicine. National Library of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/pubs/factsheets/nlm.html>.

¹⁹ NLM.MEDLINE. National Library of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/pubs/factsheets/medline.html>.

²⁰ NLM.Unified Medical Language System. National Library of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/research/umls/umlsmain.html>.

ses de datos bibliográficas, bases de datos factuales, y sistemas expertos. Está compuesto por cuatro fuentes de conocimiento: UMLS Metathesaurus; SPECIALIST (TM) Lexicon; UMLS SemanticNetwork y UMLS Information Sources Map.

Sólo mencionaremos aquí el UMLS Metathesaurus, que contiene información sobre conceptos y términos biomédicos de un número determinado de vocabularios controlados (tesauros) y establece los nombres, significados, contextos jerárquicos, atributos y relaciones entre los términos; añade cierta información básica a cada concepto y establece nuevas relaciones entre los términos de las diferentes fuentes de vocabularios. Le da información al sistema para que interprete las preguntas del usuario, identificando qué bases de datos contienen la información que se solicita, transformando el lenguaje libre en lenguaje controlado mediante calificadores, para evitar la ambigüedad de los términos de aquél. El UMLS está organizado por conceptos o descriptores a los que se asocian términos equivalentes (sinónimos, reenvíos, variantes léxicos, y traducciones). Cada concepto contiene una ayuda para definir su significado, la cadena semántica o de categorías a la que pertenece, su posición en el contexto jerárquico de varios lenguajes documentales, y para muchos conceptos, una definición. Aplica además, relaciones de asociación de términos (subencabezamientos y expresiones asociadas), y añade otras relaciones de términos asociados del mismo nivel jerárquico.

En 1998 el Metatesauro contiene 331.756 conceptos biomédicos que corresponden a 739.439 términos diferentes de más de 40 lenguajes documentales. El más importante de todos ellos es el tesauro MeSH²¹ que utiliza la NLM para indizar los artículos de unas 4.000 revistas en sus bases de datos, y que incluyen la catalogación de libros, documentos y audiovisuales. Se aplica también a la lista de encabezamientos de materias del Index Medicus. Cada referencia bibliográfica se asocia a un conjunto de términos MeSH que describe el contenido. Los términos MeSH se asignan a todo el documento, no sólo al título y al abstract. Se actualiza periódicamente por especialistas y documentalistas expertos.

Pues bien, si las bases de datos de la NLM son el paradigma de la eficacia en cuanto a la sistematización de la información para su posterior búsqueda y recuperación, no ocurre lo mismo con la mayoría de los motores de búsqueda de los servidores de INTERNET, cuyos creadores debieron pensar que el hipertexto lo solucionaba todo. La realidad es muy dis-

²¹ NLM. Medical Subject Headings (MeSH). National Library of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/pubs/factsheets/mesh.html>.

tinta, el hipertexto es una herramienta de acceso rápido en modo hipertexto de la aplicación Web y no un índice en sí mismo. Por este motivo, la mayoría de ellos incluyen una clasificación por categorías, generalmente arbitraria, de los documentos que contienen. Por otra parte, muchos de estos servidores carecen de una base de datos y ofrece la información en formato de texto, de forma que el usuario sólo puede intuir el contenido de los iconos, imágenes y otros tipos de hipertextos y la única manera de conocer el contenido es «navegar» sin rumbo por la página Web en espera de atinar con no se sabe qué, ni cuándo, ni cómo. Generalmente, el resultado es poco efectivo y muy costoso en tiempo y dinero.

El primer motor de búsqueda, Yahoo, creado a mediados de 1994 por David Filo y Jerry Yang, nació de la necesidad de estos dos estudiantes de Standford, de organizar sus datos y direcciones con una base de datos y un pequeño *software* al que se aplicaron unos determinados criterios de búsqueda. En la actualidad, existen más de 2.000 buscadores y metabuscadores²² que se clasifican en índices y motores de búsqueda, según que dispongan de una clasificación por categorías o de una casilla de búsqueda. En ambos casos, el resultado de la búsqueda es bastante ineficaz. En los del primer tipo, la búsqueda es más sencilla porque el árbol jerárquico permite al usuario «navegar» por el área de su interés, especialmente si el buscador es de contenido temático. Sin embargo, estas divisiones y subdivisiones, en la mayoría de los casos, son arbitrarias y no se corresponden con ningún principio de clasificación del conocimiento, por lo que el resultado de la búsqueda ofrece un bajo índice de exhaustividad. En los del segundo tipo, el problema es aún mayor por cuanto la búsqueda se realiza en texto libre en millones y millones de documentos, de forma que ni la ayuda de los operadores booleanos y otros tipos de nomenclatura para acotar la búsqueda, evitan el bajo índice de pertinencia y el muy elevado índice de exhaustividad. En lenguaje coloquial, podríamos decir que «ni están todos los que son, ni son todos los que están».

Los buscadores temáticos son más útiles en cuanto que la información está más sistematizada, pero en la mayoría de ellos, la clasificación del contenido tampoco se ajusta a lenguaje de indización alguno. Este problema atañe igualmente a la información contenida en todo tipo de recursos y fuentes de información de los servidores de INTERNET y es la principal dificultad para lograr una adecuada precisión, rapidez y relevancia en los resultados de la búsqueda.

²² GRAU, A.: "Buscar por la Web", *IWorld*, 1998, 10, 28-42.

El segundo problema, y no menos importante concierne al lenguaje. La mayoría de los servidores del área de Ciencias de la salud son de habla inglesa y la correcta traducción de los términos al español exige un esfuerzo añadido que no siempre se logra con éxito. Las malas traducciones dan lugar a conceptos imprecisos, equívocos, e incluso falsos. Por otra parte, la acuñación de nuevos términos por la tecnología informática y de las telecomunicaciones es constante y no siempre se encuentra el término adecuado en castellano para expresar el concepto idóneo.

La necesidad de resolver este problema ha dado lugar a la creación de proyectos financiados por instituciones bibliotecarias y dirigidos por profesionales del área de Documentación Científica de Ciencias de la Salud. Las primeras listas de recursos de fuentes de información de Medicina se deben a Lee Hancock, que en 1991 la volcó en el FTP del servidor Suranet. Tres años después, Gary Malet la amplió con el nombre de Medical List y le dio una orientación clínica. En la actualidad, Malet²³ mantiene el servidor Medical Matrix que forma parte de la unidad de proyectos de investigación de la American Medical Informatics Association Working Group.

El servidor escocés BUBL fundado por el Scottish Higher Education Funding Council utiliza la DDC (Decimal Dewey Classification) para clasificar los recursos y fuentes de información del Reino Unido en INTERNET.

Hay que señalar que, en la actualidad, la preocupación por aplicar un lenguaje normalizado a los documentos es común a la mayoría de las instituciones académicas con servidores Web y a las entidades productoras de bases de datos. De las primeras se pueden citar las páginas Web de la National Network Libraries of Medicine²⁵, que indiza los recursos de la red de bibliotecas de los EE.UU. con el vocabulario controlado MeSY y el UK.ofrecen su información en la Web, la pag uk ttata. De las segundas, son un claro exponente las bases de datos de la National Library of Medicine y las bases de datos EMBASE producidas por Excerpta Medica, Elsevier Science Publishers con sede en Amsterdam y que disponen de tres vocabularios controlados: EMCLAS (clasificación decimal de termas médicos

²³ MALET, G.: *Medical Matrix*. Healthtel Corporation. Available from: URL:

<http://www.medmatrix.org/index.asp>.

²⁴ NN/LM.US National Network of Libraries of Medicine Home page. National Network Libraries of Medicine 1998. Available from: URL: <http://www.nlm.nih.gov/>.

²⁵ BUBL. BUBL LINK: 610 Medical sciences, medicine. BUBL Information Service, Andersonian Library, Strathclyde University, Glasgow (Scotland). Available from: URL: <http://link.bubl.ac.uk/medicine>.

a partir de las secciones de los repertorios), MALIMET (lista normalizada que comprende más de 200.000 términos preferenciales y un cuarto de millón de sinónimos) y EMTAG (tabla de dos centenares de términos normalizados de carácter general).

CONCLUSIONES

De la elaboración del presente trabajo se quiere dejar constancia de la necesidad de que los servidores de INTERNET y las bases de datos de Ciencias de la Salud dispongan de algún tipo de clasificación normalizada que cumpla las dos elementales características de los vocabularios controlados, el principio de entropía y la eliminación de los accidentes lingüísticos. En los dos ámbitos existen ya excelentes modelos a seguir fruto de magníficos proyectos de investigación como la ya mencionada aplicación del UMLS y en especial del vocabulario controlado MeSH a las bases de datos de la NLM, y a las páginas Web de la National Network Library of Medicine.

Es de señalar también para los hispano-parlantes la importancia de una correcta traducción de los términos anglosajones, a fin de evitar barbarismos, anglicismos y otros tipos de vicios del lenguaje, innecesarios en la mayoría de las ocasiones, y que dan lugar a malas interpretaciones y confusión de los conceptos.