

SAN JOSÉ MONTANO, BLANCA (BIBLIOTECARIA), SAGÜILLO ANTOLÍN, MILAGROS (ENFERMERA); GÓMEZ MAZO, NATIVIDAD (ENFERMERA)
Hospital de Móstoles

Análisis y evaluación de buscadores en la web sobre información de Enfermería Urológica

Palabras clave: Buscadores web, enfermería urológica, análisis, evaluación e información.

Key words: Searches Web, urológic nursing, analysis, evaluation and information.

INTRODUCCIÓN

Los servidores de búsqueda en Internet son sistemas basados en la habitual arquitectura cliente/servidor de Internet, estos servidores están compuestos por dos grandes subsistemas: un programa (spider) que utiliza para localizar los documentos y una base de datos para gestionar la información. A su vez, la base de datos se divide en otras dos partes: un motor de análisis e indización y un sistema de consulta e interrogación (1).

Los servidores de búsqueda funcionan a través de un programa denominado robot que explora Internet para localizar los documentos y direcciones de documentos, luego un sistema automático que analiza e indiza lo localizado por el robot y por último un sistema de consulta (también a disposición del usuario para expresar sus necesidades de información).

El lenguaje de consulta es la parte más importante del sistema de interrogación, ya que mediante él, el usuario puede expresar mejor o peor sus necesidades de información, establecer filtros y condiciones, limitar, navegar por la lista de resultados...

El rendimiento global de los servidores de búsqueda depende de tres cosas:

- Eficacia del robot en descubrir y mantener los documentos.
- Cantidad de información que guardan.
- Versatilidad y potencia del lenguaje de consulta.

Hay que tener en cuenta que muchas veces lo que se busca puede no estar en Internet y lo que se encuentra a veces no es información de calidad; además, a veces no se encuentra la información mediante una consulta sino que hay que acudir directamente a una fuente de información.

Cuando se quiere realizar una búsqueda en la web, el objetivo es encontrar información sobre un tema determinado, para ello se emplean palabras claves o frases cortas que definan el tema (siempre teniendo en cuenta en qué

idioma se busca la información. Si la buscamos en inglés, la encontraremos en inglés, y si la buscamos en castellano, la encontraremos en castellano).

Cuando se va a utilizar un buscador para realizar una búsqueda en la web hay que tener una serie de variables en cuenta, como son: la composición de la base de datos, su cobertura y actualización, capacidad de búsqueda (operadores booleanos, truncamiento...), ejecución de la recuperación, precisión, exhaustividad, tiempo de respuesta...

La relevancia o pertinencia de los documentos recuperados dependerá del grado de conocimiento de las prestaciones que poseen los distintos buscadores. Una de las diferencias más importantes entre los buscadores es su fórmula para calcular la relevancia de la información recuperada, que ordena la presentación de los resultados. Algunos buscadores se basan en el análisis de frecuencia de palabras en el texto, otros en el análisis de la frecuencia con que las páginas están ligadas a otras...

Es muy importante la forma en que los buscadores presentan sus resultados obtenidos, este es un factor clave para todos los interfaces, cada motor de búsqueda emplea un algoritmo de relevancia para ello que es un secreto para el usuario.

OBJETIVO

El objetivo principal del trabajo es evaluar los servidores de búsqueda más importantes actualmente en Internet, en base a la búsqueda de información sobre enfermería urológica.

Los objetivos secundarios que nos llevarán a esto son:

- Evaluar los buscadores con sus características de base de datos y lenguaje de recuperación.
- Evaluar los resultados obtenidos con las estrategias desarrolladas sobre «enfermería urológica».
- Cotejar los resultados de la evaluación de todos los buscadores.

METODOLOGÍA

Hemos utilizado el modelo «Big Six Skills» (2) para abordar la recuperación de información y su evaluación en seis fases:

1. Definición de la tarea (identificar problema información)

Necesitamos encontrar información sobre enfermería urológica, sus técnicas y cuidados, preferiblemente en español. Sobre todo lo que respecta a pacientes urostomizados y nefrectomizados, sus cuidados y tratamientos, tanto protocolos de actuación como material utilizado y sus consejos de utilización. Incluso sería muy útil encontrar información sobre recomendaciones o formación del paciente en el autocuidado.

Las tres estrategias de búsqueda se han realizado dando prioridad a la búsqueda de información en español (3).

Las preguntas serían:

1. Enfermería urológica, cuidado del paciente urológico, técnicas de enfermería urológica.
2. Cuidados de enfermería en el paciente urostomizado, técnicas de enfermería, educación y autocuidado.
3. Cuidados de enfermería en el paciente nefrectomizado, técnicas de enfermería, educación y autocuidado.

2. Estrategias de búsqueda de información (seleccionar fuente de información)

Vamos a buscar información en los tres servidores de búsqueda en Internet más importantes actualmente, como son: Altavista, Google y Alltheweb.

Las búsquedas se han realizado el mismo día y en el mismo período (unas tras otras siguiendo el orden), las estrategias utilizadas se han ejecutado usando la opción de la búsqueda avanzada y utilizando los campos que dan para limitar la búsqueda el propio interfaz, que en los tres buscadores es muy similar. Seleccionamos la búsqueda avanzada, con la opción de construcción de consulta. La opción de operadores booleanos no la hemos tomado para poder realizar las búsquedas con estrategias similares, ya que no todos los buscadores tenían las mismas opciones de operadores booleanos, truncamiento, o de uso de paréntesis. Además, hemos buscado los términos en minúscula y eliminando los acentos.

Las búsquedas realizadas han sido las que se reflejan en el cuadro nº 1.

3. Localización y acceso (localizar fuente y la información en ellas)

Para su posterior evaluación se ha creado una tabla para baremar la presencia o ausencia de alguna de las características de cada buscador, esta puntuación se refleja en el cuadro nº 2.

CUANDO SE QUIERE REALIZAR UNA BÚSQUEDA EN LA WEB, EL OBJETIVO ES ENCONTRAR INFORMACIÓN SOBRE UN TEMA DETERMINADO, PARA ELLO SE EMPLEAN PALABRAS CLAVES O FRASES CORTAS QUE DEFINAN EL TEMA

Cuadro 1. Estrategias de búsqueda	
	ESTRATEGIA
GOOGLE http://www.google.com	1. Con todas las palabras urologia enfermeria Con la frase exacta enfermeria urologica Con alguna de las palabras técnicas OR cuidados
	2. Con todas las palabras urostomia cuidados pacientes Con alguna de las palabras enfermeria OR enfermero
	3. Con todas las palabras nefrectomia cuidados pacientes Con la frase exacta enfermeria Con alguna de las palabras técnicas OR autocuidado OR educacion
ALTAVISTA http://www.altavista.com	1. Todas estas palabras urologia enfermeria Esta secuencia exacta enfermeria urologica Cualquiera de estas palabras tecnicas OR cuidados
	2. Todas estas palabras urostomia cuidados pacientes Cualquiera de estas palabras enfermeria OR enfermero
	3. Todas estas palabras nefrectomia cuidados pacientes Esta secuencia exacta enfermeria Cualquiera de estas palabras autocuidado OR tecnicas OR educacion
ALL THE WEB http://www.alltheweb.com	1. All the words "enfermeria urologica" AND (tecnicas OR cuidados) Words filters must include in the text enfermeria Words filters must include in the text urologia
	2. All the words urostomia AND cuidados AND pacientes AND enfermeria
	3. All the words nefrectomia AND cuidados AND pacientes AND enfermeria

Cuadro 2. Puntuación para baremar características de los buscadores		
	PRESENCIA	AUSENCIA
BASE DE DATOS		
Tamaño en URLs		
<50 millones	1	
>50 millones	3	
Actualización		
Diaria	3	
Semanal/quincenal	2	
Mensual	1	
Indización		
Texto completo		-3
Título y primeros párrafos	1	
Evaluación recursos	3	
Memoria caché	3	
INTERFAZ		
Consulta		
Ayuda		-3
Documentación general		-3
Contextualizada	3	
En pantalla	2	
Ejemplos	2	
Búsqueda avanzada		-3
Definir y guardar preferencias	1	
Mirrors	1	
Interfaz en diferentes lenguas	2	
Claridad de diseño e interacción	2	
MOTOR BÚSQUEDA		
Truncamiento	1	
No discrimina mayúsculas y minúsculas	1	
OPERADORES		
AND		-3
OR		-3
NOT		-2
PROXIMIDAD	2	
PARÉNTESIS	1	
POR DEFECTO AND	2	
FRASES		-3
DIRECTORIO		-3
AYUDA		
TESAURO-CONCEPTO	2	
FORMULARIO	2	
LIMITAR		
TIEMPO	1	
FORMATO	2	
DOMINIO	1	
ETIQUETAS XML	1	-2
LENGUA	1	

4. Uso de la información (comprender y extraer información relevante)

Para su posterior evaluación se ha creado una tabla (ver cuadro nº 3) para baremar los resultados.

Cuadro 3. Puntuación para baremar los resultados		
	RESULTADOS	
	PRESENCIA	AUSENCIA
Frecuencia de las palabras		-1
Frecuencia y etiqueta	1	
Datos externos a la página	3	
Documentos similares	1	
Corrector ortográfico	2	
Traducción	2	
Conceptos relacionados para ampliar la consulta	2	
Agrupación de los resultados	2	
Búsqueda en los resultados	2	
Agrupación de los resultados por sede web	2	
Total		
Relevancia media		
Cobertura		
Exhaustividad		

Entendiéndose la relevancia como la medida de la relación entre lo que dice el autor del documento y lo que dice el usuario que necesita, siendo el concepto fundamental de la recuperación de información.

5. Síntesis (organizar y presentar información)

6. Evaluación (juzgar el producto resultante -efectividad y eficiencia-)

En la evaluación hemos utilizado medidas de funcionamiento objetivas y subjetivas y otras características relacionadas con el usuario, basando la evaluación en la experiencia real del usuario.

RESULTADOS

La información obtenida sobre las características de cada buscador analizado ha sido obtenida bien mediante la información que proporciona el propio buscador, en directorios que analizan buscadores y en búsquedas de dicha información a través de publicaciones en la red, esta información ha facilitado el baremo de características que refleja el cuadro nº 4.

Cuadro 4. Puntuación de los buscadores analizados			
	ALTAVISTA	GOOGLE	ALLTHEWEB
BASE DE DATOS			
Tamaño en URLs			
<50 millones	-	3	-
>50 millones	-	-	-
Actualización			
Diaria	-	-	-
Semanal/quincenal	-	-	-
Mensual	-	-	-
Indización			
Texto completo	-	-	-
Título y primeros párrafos	1	1	1
Evaluación recursos	3	3	-
Memoria caché	0	3	0
INTERFAZ			
Consulta			
Ayuda	-	-	-
Documentación general	-	-	-
Contextualizada	3	3	-
En pantalla	2	2	2
Ejemplos	2	2	2
Búsqueda avanzada	-	-	-
Definir y guardar preferencias	0	1	1
mirrors	1	1	1
Interfaz en diferentes lenguas	2	2	0
Claridad de diseño e interacción	2	2	2
MOTOR BÚSQUEDA			
Truncamiento	1	-	-
No discrimina mayús. y minús.	-	1	-
OPERADORES			
AND	-	-	-
OR	-	-	-
NOT	-	-2	-
PROXIMIDAD	2	-	-
PARÉNTESIS	1	0	1
POR DEFECTO AND	2	2	2
FRASES	-	-	-
DIRECTORIO	-	-	-
AYUDA			
TESAURO-CONCEPTA	2	2	-
FORMULARIO	2	2	2
LIMITAR			
TIEMPO	1	1	1
FORMATO	2	2	2
DOMINIO	1	1	1
ETIQUETAS XML	-	-	-
LENGUA	1	1	1
TOTALES	31	33	19

Los resultados obtenidos de las búsquedas han sido:

Cuadro 5. Total resultados		
	TOTAL RESULTADOS	
GOOGLE	1	32
	2	2
	3	71
TOTAL		105
ALTAVISTA	1	13
	2	9
	3	22
TOTAL		44
ALL THE WEB	1	17
	2	9
	3	24
TOTAL		50

El análisis de los resultados obtenidos ha sido (cuadros 6 y 7):

Cuadro 6. Puntuación de los buscadores analizados			
	ALTAVISTA	GOOGLE	ALLTHEWEB
Frecuencia de las palabras	-	-	-
Frecuencia y etiqueta	1	1	1
Datos externos a la página	3	3	-
Documentos similares	1	1	-
Corrector ortográfico	-	2	-
Traducción	2	2	-
Conceptos relacionados para ampliar la consulta	2	2	-
Agrupación de los resultados	-	-	-
Búsqueda en los resultados	2	2	2
Agrupación de los resultados por sede web	-	-	-
TOTAL	11	13	3
TOTAL FINAL CON RESULTADOS ANTERIORES de características	34	36	21

CONCLUSIONES

Las conclusiones que se deducen son:

En cuanto a las bases de datos que conforman los buscadores es evidente que realizar las búsquedas en español produce una gran limitación, ya que la mayoría de la información que contienen está en inglés y, además, las URLs localizadas en español, un gran número pertenecen a información comercial o referencial, siendo pocos los directorios sobre el tema o los documentos completos en PDF o Word.

En cuanto a las características de los buscadores, sus lenguajes de recuperación y sus servicios de valor añadido,

Cuadro 7. Análisis resultados de las búsquedas										
RESULTADOS										
		Fecha (4)	Resultados inactivos	Resultado duplicado	Resultado Relevante/resultados totales	Relevancia resultados (5)				
						0	1	2	3	4
GOOGLE	1	12-4-04		3	12/20		2	5	3	2
	2	12-4-04		0	2/2			1		1
	3	12-4-04	1	2	8/20		2	3	3	
TOTAL				5	22		4	9	6	3
ALTAVISTA	1	12-4-04		2	11/13		4	3	4	
	2	12-4-04		0	9/9		4	4	1	
	3	12-4-04		2	5/20		2	3		
TOTAL				4	25		10	10	5	
ALL THE WEB	1	12-4-04		4	11/17		5	4	2	
	2	12-4-04		0	6/9			5	1	
	3	12-4-04		2	7/20		6	1		
TOTAL				6	24					

encontramos sistemas de búsqueda como Google (36) y Altavista (34) que tienen casi el mismo número de puntos con respecto a la características que tienen en presencia, siendo Google el que más servicios de valor añadido tiene, como son el corrector ortográfico o la documentación sobre su sistema; mientras que Altavista desarrolla más los operadores booleanos de búsqueda como es el uso de paréntesis, NOT o proximidad... Tanto uno como otro incorporan dos servicios que pueden ser muy útiles para el usuario final enfermero, como son los enlaces en cada URL a la traducción del documento y a documentos similares.

En cuanto a los resultados obtenidos en las búsquedas, aunque Google es el que mayor número de documentos recupera (105), su recuperación en las tres búsquedas no es equilibrado. Este dato se vuelve a poner en evidencia al medir su relevancia media que es más alta en Altavista con un 0,68 (de 44 documentos), al igual que su cobertura de 0,34 y la exhaustividad de 0,568 (frente a Google con 0,009). Es decir, que de los tres buscadores el de mayor cobertura, exhaustividad y relevancia es Altavista, que aunque no recupera tantos documentos éstos son en mayor cantidad más relevantes.

También con referencia a la relevancia de los documentos recuperados se puede observar que los resultados de Google tienen una relevancia más alta que los recuperados por Altavista o Alltheweb, con mayor número de documentos con relevancia 4 y 3. Mirando los documentos que alcanzan esta relevancia tan alta de 4 en Google, se observa que son documentos que sólo se encuentran en Google, la razón puede ser que los creadores de ellos los registran en los buscadores más importantes, mientras que no lo hacen con el resto de los buscadores.

La conclusión final es que se deben utilizar buscadores que tengan una gran base de datos y un potente buscador que facilite al usuario final no sólo la búsqueda, sino también la recuperación de esa información útil.

BIBLIOGRAFÍA

- Codina, Lluís. Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte I). Information World en español 1997, 6 (5): 22-27.
- Codina, Lluís. Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte II). Information World en español 1997, 6 (6): 18-27.
- González Bonorino, Adina; Rodríguez, Daniel. Guía de buscadores y búsqueda en Internet. http://www.cyta.com.ar/biblioteca/bddoc/bdlibros/guía_buscadores_búsqueda_en_Internet (Consultado 16-02-2004).
- Hernández Pérez, Antonio. La búsqueda y recuperación de información en Internet. Caridad Sebastián, Mercedes, coordinadora. La sociedad de la información: política, tecnología e industria de los contenidos. Madrid: Centro de Estudios Ramón Areces; 1999; pp. 213-242.
- Moreno Jiménez, María Pilar. Estrategias y mecanismos de búsqueda en la web invisible. http://biblio.colmex.mx/recelec/web_invisible.htm (consultado 16-02-2004).
- Search Engine showdown. The user's guide to web searching. <http://www.searchengineshowdown.com> (consultado 8-3-2004)

NOTAS

- (1) Lluís Codina. Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte I). Information World en español 1997, 6 (5): 22-27.
- (2) Hernández Pérez, Antonio. La búsqueda y recuperación de información en Internet.
- (3) Denominamos el idioma español y no castellano, cogiendo el término genérico con el que denominan el lenguaje en los buscadores.
- (4) Las búsquedas han sido realizadas el día 4-04-2004, pero hemos tenido que volver a repetir las con fecha 12-04-2004.
- (5) Los resultados relevantes los hemos baremado del 0 al 4, siendo 0 muy poco relevante y 4 muy relevante. Señalando el total de relevantes por el grado de baremo.