

SISTEMAS DE GESTIÓN DOCUMENTAL

BÚSQUEDAS EN EL WEB

LOCALIZACIÓN DE INFORMACIÓN EN EL WEB

El World Wide Web está considerado como la mayor base de datos documental existente en la actualidad. El WWW contienen miles de millones de documentos en varios formatos (html, xml, Word, pdf, ppt, ...) con información de cualquier materia en la que se pueda pensar.

La enorme cantidad de información disponible en el WWW contrasta con la forma en la que está disponible: la información no está estructurada ni organizada, lo que supone un grave inconveniente para su utilización por varios motivos. En primer lugar, si no existe información sobre la localización de un documento, no se puede acceder a él; en segundo lugar, si no se sabe que una información existe, no se puede pretender buscarla, y por tanto no será jamás localizada; en tercer lugar, aún sabiendo que la información está, y está localizada en un lugar, no existen mecanismos de búsqueda eficientes para satisfacer las necesidades de información de los usuarios.

Ante esta situación surgen los motores de búsqueda, que son herramientas que permiten localizar un documento en el WWW que contenga información sobre temas particulares, respondiendo por tanto a consultas realizadas por un usuario. Existen 4 grandes categorías de buscadores: los buscadores, los metabuscadores, los directorios y los buscadores del web invisible. Los primeros son aplicaciones que permiten realizar búsquedas a texto completo sobre los documentos que han sido indexados por el buscador. Los segundos son buscadores de buscadores, es decir, realizan búsquedas en paralelo utilizando varios buscadores, y devuelven los resultados agrupados y con coherencia. Los terceros son aplicaciones que indexan documentos según una categorización previa, y normalmente no permiten búsquedas a texto completo. Los últimos son herramientas que tratan de facilitar la tarea de buscar en bases de datos y aplicaciones que generan contenidos dinámicos que no pueden indexar los tipos de buscadores anteriores.

El objetivo de esta práctica es aprender a utilizar los distintos tipos de buscadores para resolver consultas y necesidades de información de los usuarios según la naturaleza de la necesidad de información. Para ello se proponen unas cuestiones (necesidades de información) que el alumno deberá responder planteando estrategias y eligiendo los buscadores que considere oportuno para responder de la mejor forma posible a la cuestión planteada. Para cada consulta se deberá rellenar una ficha (que se adjunta) que indica los pasos que se han seguido hasta obtener una conclusión razonable para la consulta y cual es la respuesta que el alumno propone para la cuestión planteada.

Parte I. Resolución de necesidades de información

A continuación se plantean varias cuestiones que el usuario deberá tratar de contestar, siguiendo el formato de la Ficha 1 de respuesta que se ofrece en la siguiente sección

1. Obtener la máxima información sobre herramientas de gestión y búsqueda documental (aplicaciones de sistemas de gestión documental).
2. Quiero aprender XML. ¿Dónde puedo encontrar tutoriales, cursos, etc.?
3. Deseo incorporar gráficas basadas en estadísticas en una página HTML que estoy creando, y quiero obtener un servidor web que me genere dichas gráficas sin necesidad de utilizar ninguna aplicación. ¿Dónde puedo conseguir dichas gráficas?
4. Se precisa obtener la mayor información posible sobre temarios de asignaturas que se impartan en alguna universidad del mundo para definir la asignatura Sistemas de Gestión Documental.
5. Obtener un índice alfabético de revistas electrónicas dedicadas a la biblioteconomía y documentación.
6. Obtener la máxima información posible sobre los premios Darwin. Se prefiere obtener un único documento que contenga la máxima información posible.
7. He acuñado una cita: “La verdad es única, pero tiene varias caras”. Quiero saber si esa cita la ha dicho o escrito alguien con anterioridad a mí. ¿Es así o no?
8. Encontrar un servidor web que contenga información sobre tratamientos naturales para el asma, terapias alternativas, consejos, experiencias y cosas que puedan ayudar a un enfermo asmático a mejorar su calidad de vida.
9. Obtener una fotografía de las Islas Baleares desde el espacio.
10. Obtener un listado de referencias sobre las bibliotecas del futuro. Se pretende obtener una lista de referencias que sirvan de ayuda para conocer las tendencias y la planificación de las bibliotecas de un futuro próximo.

Para las cuestiones que figuran a continuación se debe rellenar una ficha del modelo Ficha 2 que contenga exclusivamente la consulta realizada y el buscador, la respuesta a la pregunta, y el tiempo tardado en encontrar la respuesta.

11. La ciudad donde nació el escritor estadounidense galardonado con el premio Nobel de Literatura en 1954 limita al Este con otra ciudad, y la frontera es una calle. ¿Cómo se llama esa calle y cual es la otra ciudad?
12. Obtener información sobre el partido de la final del Mundial de fútbol de España (quien jugaba, quien ganó, que resultado obtuvo, donde se jugó, etc.).
13. Se precisa conocer la edad del jugador más joven y del más mayor que han jugado alguna vez en un partido de un campeonato mundial de fútbol (evidentemente, se trata de la edad que tenían cuando jugaron el partido).

14. En la primera misión espacial tripulada que aterrizó en la Luna, sólo 2 de los 3 tripulantes pusieron pie en el satélite. ¿Cuál fue el tripulante que no pisó la Luna?
15. Deseo comprarme el libro “Las Dos Torres” de la saga de “El Señor de los Anillos”, pero quiero comprar la edición que menos páginas tiene, de las que estén disponibles en España en ediciones en Castellano. ¿Cuál es la edición que debo comprar?
16. ¿Cuál es la galaxia más cercana a la Vía Láctea?
17. Quiero registrar un dominio en Internet del tipo www.nombre.com, pero quiero saber previamente si existe o no el dominio que quiero registrar. Encontrar una página web donde pueda consultar si existe o no el dominio que quiero registrar.
18. Mi Windows XP muestra el menú Inicio->Todos los programas en una columna y yo quiero que lo muestre en varias columnas. ¿Por qué sucede esto? ¿Qué puedo hacer para cambiarlo?
19. Encontrar un remedio para el hipo.
20. ¿Cuál es la longitud máxima que puede tener una Tenia (parásito humano)?
21. ¿Cuál es el animal más longevo?
22. ¿Cuántos pelos tiene una persona, aproximadamente, en la cabeza?

Parte II. Estudio comparativo de buscadores

Se desea realizar un estudio comparativo de los siguientes buscadores:

1. Google (www.google.com)
2. Yahoo (www.yahoo.com)
3. Live Search (MSN Search) (www.live.com)
4. Ask (www.ask.com)

El estudio pretende determinar, para una consulta concreta, los siguientes parámetros:

1. La cantidad de documentos indexados por cada buscador con respecto a la consulta
2. El grado de solapamiento de los buscadores para cada una de las respuestas
3. La distribución de documentos encontrados en un único buscador sobre los buscadores considerados.

Para ello, se pide:

1. Diseñar 10 consultas que proporcionen un conjunto muy limitado de resultados en los buscadores.
2. Ejecutar las consultas sobre los 4 buscadores seleccionados.
3. Calcular los niveles de solapamiento para cada uno de los documentos encontrados, es decir, determinar a través de una matriz documento/buscador la relación existente entre el documento encontrado y el buscador que lo encontró
4. Realizar una tabla en la que, para cada buscador, se especifique el número de documentos únicos encontrados, y representarlos en una gráfica de tipo tarta.

El resultado de esta tarea se debe presentar correctamente formateado en un documento Word, con anexos en Excel que contengan las tablas y gráficas usadas en los cálculos.

Ficha 1 de respuesta

NI:		
Primera Consulta	Consulta	
	Buscador	
Ultima Consulta	Consulta	
	Buscador	
Pasos intermedios más relevantes		
1	Consulta	
	Buscador	
2	Consulta	
	Buscador	
3	Consulta	
	Buscador	
Solución	Respuesta	
	URL	

Ficha 2 de respuesta

NI:		
Consulta	Consulta	
	Buscador	
Tiempo empleado		
Solución	Respuesta	
	URL	

Listado de buscadores que se pueden utilizar

Los buscadores más importantes que se pueden utilizar para realizar las consultas, ordenados por categorías, son los siguientes:

BUSCADORES Y DIRECTORIOS

Google: <http://www.google.com>
AllTheWeb.com: <http://www.alltheweb.com>
Yahoo: <http://www.yahoo.com>
MSN Search: <http://www.msn.com>
Microsoft: <http://www.live.com>
Ask Jeeves: <http://www.askjeeves.com>
HotBot: <http://www.hotbot.com>
Lycos: <http://www.lycos.com>
LookSmart: <http://www.looksmart.com>
Open directory: <http://dmoz.org>
Altavista: <http://www.altavista.com>
Netscape search: <http://search.netscape.com>
Wisnut: <http://www.wisnut.com>
Accoona: <http://www.accoona.com>
Excite: <http://www.excite.com>

METABUSCADORES

Clusty: <http://clusty.com>
Kartoo: <http://www.kartoo.com>
Surfwax: <http://www.surfwax.com>
Fazzle: <http://www.fazzle.com>
Dogpile: <http://www.dogpile.com>
Excite: <http://www.excite.com>
Metacrawler: <http://www.metacrawler.com>
Webcrawler: <http://www.webcrawler.com>
Mamma: <http://www.mamma.com>
Search.com: <http://www.search.com>

WEB INVISIBLE

Librarians Index: <http://lii.org>
Academic info: <http://www.academicinfo.net>
Infomine: <http://infomine.ucr.edu/search.phtml>
Direct search: <http://www.freepint.com/gary/direct.htm>
Invisible-web: <http://www.invisible-web.net>
About.com: <http://www.about.com>
FindLaw: <http://www.findlaw.com>
CompletePlanet: <http://www.completeplanet.com>
Allacademic: <http://www.allacademic.com>
Findarticles: <http://www.findarticles.com>
Magportal: <http://www.magportal.com>