

Motores de Búsqueda Web - Tarea 2

José Alberto Benítez Andrades

71454586A

Motores de Búsqueda Web

Master en Lenguajes y Sistemas Informáticos - Tecnologías del Lenguaje en la Web

UNED

07/12/2010

7 de diciembre de 2010

Tarea 2

Enunciado del ejercicio

El lugar básico para realizar búsquedas bibliográficas en la red será scholar.google.com. Para familiarizaros con este servicio, os proponemos que busquéis los cinco artículos de más impacto (es decir, que han sido más referenciados) en relación con la búsqueda Web. Además de dar una lista como resultado de vuestro trabajo, debéis documentar cuáles fueron vuestros criterios de búsqueda y cómo fue vuestra interacción con el sistema.

Respuesta:

Después de realizar la búsqueda: *web search* en **scholar.google.com**, los 5 artículos más relevantes que se encontraron fueron los siguientes:

The screenshot shows the Google Scholar search interface. At the top, there are navigation links: La Web, Imágenes, Vídeos, Maps, Noticias, Shopping, Correo, Más. The search bar contains the text 'Web search' and a 'Buscar' button. Below the search bar, there are options to 'Buscar en la Web' (selected) or 'Buscar sólo páginas en español'. A green bar below the search bar contains the text 'Académico' and 'en cualquier momento'. Below this, there are links for 'incluir citas' and 'Crea alerta de correo electrónico'. The search results are listed below, each with a title, author, year, journal, and a PDF link. The results are:

- The anatomy of a large-scale hypertextual Web search engine*** 1 [PDF] de psu.edu
S Brin... - Computer networks and ISDN systems, 1998 - Elsevier
To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of Web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on ...
Citado por 7814 - Artículos relacionados - Las 458 versiones
- A taxonomy of web search** [PDF] de psu.edu
A Broder - ACM Sigir forum, 2002 - portal.acm.org
A central tenet of classical information retrieval is that the user is driven by an information need. Schneiderman, Byrd, and Croft [SBC97] define information need as "the perceived need for information that leads to someone using an information retrieval system in the first ...
Citado por 861 - Artículos relacionados - Las 29 versiones
- Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search** [PDF] de psu.edu
TH Haveliwala - IEEE transactions on knowledge and data ..., 2003 - computer.org
Abstract—The original PageRank algorithm for improving the ranking of search-query results computes a single vector, using the link structure of the Web, to capture the relative "importance" of Web pages, independent of any particular search query. To yield more ...
Citado por 1128 - Artículos relacionados - Las 89 versiones
- Analysis of a very large web search engine query log** [PDF] de psu.edu
C Silverstein, H Marais, M Henzinger... - ACM SIGIR Forum, 1999 - portal.acm.org
Page 1. Analysis of a Very Large Web Search Engine Query Log Craig Silverstein Google Inc. 2400 Bayshore ... query. This suggests that traditional information retrieval techniques may not work well for answering web search requests. The ...
Citado por 724 - Artículos relacionados - Las 23 versiones
- From e-sex to e-commerce: Web search changes** [PDF] de tripod.com
A Spink BJ Jansen, D Wolfram... - Computer, 2002 - ieeexplore.ieee.org
SCOPE OF STUDY Excite (http://www.excite.com) is a major Internet media company offering Web searching and a personalization portal. Its searches are based on the exact terms a user enters in a query. Capitalization is disregarded with the exception of logical commands ...
Citado por 376 - Artículos relacionados - Las 28 versiones

7 de diciembre de 2010

Es decir:

[The anatomy of a large-scale hypertextual Web search engine* 1](http://linkinghub.elsevier.com/retrieve/pii/S016975529800110X)

<http://linkinghub.elsevier.com/retrieve/pii/S016975529800110X>

[A taxonomy of web search](http://portal.acm.org/citation.cfm?id=792552)

<http://portal.acm.org/citation.cfm?id=792552>

[Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search](http://www.computer.org/portal/web/csdl/doi/10.1109/TKDE.2003.1208999)

<http://www.computer.org/portal/web/csdl/doi/10.1109/TKDE.2003.1208999>

[Analysis of a very large web search engine query log](http://portal.acm.org/citation.cfm?id=331405)

<http://portal.acm.org/citation.cfm?id=331405>

[From e-sex to e-commerce: Web search changes](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=989940)

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=989940

Son los 5 artículos más citados relacionados con la búsqueda en la web.

1. Interacción con la herramienta scholar.google.com.

Scholar.google.com, al igual que el buscador principal de Google, sigue un sistema de ranking de artículos basado en distintos algoritmos (semejanza de la búsqueda con el contenido de los artículos, citas de los mismos, algoritmos de pagerank, etc).

Entre otras cosas, cuando uno busca "web search" en este buscador, realmente scholar busca el mayor número de artículos relacionados con estos 2 términos en su conjunto. Pero, a medida que pasamos páginas en la búsqueda, la importancia de los artículos o la concordancia con lo que estamos buscando, no es la correcta, como sucede en el buscador real, ya que, influyen muchos factores en la ordenación de los mismos.

Podríamos decir que si buscamos "Web search engines" o "Web search algorithms", es muy probable encontrar mejores resultados con el contenido que realmente buscamos, aunque sean también un número menor. Es decir, a más palabras clave que busquemos, menores serán los resultados, pero más acertados en lo que necesitamos.

La **búsqueda web** son dos términos clave que abarcan muchísimos temas sobre ellos mismos. Es por ello que, se puede decir, que los artículos con más relevancia en este tema y que más citas tienen, son los 5 anteriormente dados. Pero también cabe destacar, que pueden existir artículos en los que se nombre la búsqueda web, pero que no estén tan intensamente relacionados con estos términos, que posean un mayor número de citas que los anteriormente descritos.

En definitiva, para buscar bien sobre un tema, debemos desglosar el mayor número de palabras clave posible, dentro de un límite.