

COMPARATIVE ANALYSIS OF SEARCH ENGINES

C. Velmurugan*, K. Ramakrishna Reddy**

Abstract Almost all the people use the Web Search Engines for their everyday activities; especially the academicians and researchers use the Search Engines for their teaching, day to day activities, and research needs. Here, an attempt has been made to rank 5 Search Engines, i.e., Google, AltaVista, Yahoo, MSN and Rediff, based on quantum of records retrieved from the Web for the period from January to December 2011. For this purpose, one book title belonging to engineering field has been used with various combinations of fields, such as, author/s, title, publisher, etc.

Keyword: Search Engines, Google, Yahoo, AltaVista, MSN and Rediff, Components of Search Engines, Comparison of Search Engines, Ranking of Search Engines

INTRODUCTION

Continuous and fast development of the Internet and Library has made it an integral part of everyday life for academicians and organizations. Hence, web search has become very important for everybody to locate required information for teaching, day to day activities, and research needs. According to Internet World Statistics as on 31st March 2011, Internet users reached the number of approximately 2,095 million or 30.2% of the world population. About 85% of Internet users utilize Web Search Engines for their informational needs, while usage of Search Engines is the second most popular web service, after E-mail.

In this paper, an attempt has been made to compare and rank five major Search Engines, i.e., Google, Yahoo, AltaVista, MSN and Rediff, based on retrieved data of short and unfocused queries posed against a massive collection of heterogeneous and hyperlinked documents that change dynamically. Data was collected for the period from January to December 2011 on monthly basis. It is expected that the results will be useful to supplement other methods applied to compare Search Engines that are being tried by search engine researchers.

SEARCH ENGINES

A search engine is a web site that collects and organizes content from all over the Internet as well as the world. Those wishing to locate something would enter a query about what they would like to find and the engine provides links to content that matches what they want.

Components of Search Engines

Search Engines consists of three components, namely:

1. Spider - Programs that traverses the web from link to link, identify and reading pages.
2. Index - Web databases containing a copy of each web page gathered by the spider.
3. Search Engine Mechanism - Software that enables users to query the index and that usually returns results in relevancy ranked order.

OBJECTIVES

The Research is based on the following objectives:

1. To study the features of the identified Search Engines.
2. To formulate various combinations of search queries and searching the web using identified Search Engines on monthly basis for the year 2011.
3. To identify and suggest suitable Search Engines for engineering faculty.

RESEARCH METHODOLOGY

In order to achieve desirable results, following criteria have been followed:

1. Collected the data through the web.
2. Duration of data collection is for one year on monthly basis from January to December 2011.
3. Five Search Engines are considered for data collection, i.e., Google, Yahoo, AltaVista, MSN and Rediff.
4. 45 Queries are framed in advance with various combinations for searches.
5. Queries invoked without any discrepancies.
6. The book mentioned below has been considered for data collection.

* Research scholar, SCSVMV University, Kanchipuram, Tamilnadu, India. Email-id: cvelmuruga72@gmail.com

** Chief Librarian, Acharya Institutes, Bangalore, Karnataka, India. krred@yahoo.com

Author : Palanichamy, M S And Nagan, S
 Title : Engineering Mechanics: Statics and Dynamics
 Edition : 2nd edition
 Publisher : Tata McGraw-Hill Publishing Company Limited
 Place of Publication : New Delhi
 Year of Publication : 2002
 ISBN : 0-07-058830-9

7. With help of bibliographical details mentioned above, following 45 Queries have been framed for searching using the specified 5 Search Engines, i.e., Google, Yahoo, AltaVista, Rediff and MSN.

S.no	Framed 45 Queries
1	engineering mechanics
2	engineering mechanics statics and dynamics
3	engineering mechanics/palanichamy
4	engineering mechanics by palanichamy
5	engineering mechanics by m.s.palanichamy and s.nagan
6	engineering mechanics by palanichamy and nagan
7	engineering mechanics by nagan
8	engineering mechanics by nagan s
9	engineering mechanics statics and dynamics 3rd edition
10	palanichamy ms and nagan s tata mcgraw hill
11	engineering mechanics/nagan
12	engineering mechanics 3rd edition
13	engineering mechanics+3rd edition+tata mcgraw hill
14	engineering mechanics+3rd edition+palanichamy+nagan+tatamcgraw hill
15	engineering+mechanics+palanichamy
16	engineering+mechanics+palanichamy+nagan
17	engineering mechanics with palanichamy and nagan
18	Engineering and Mechanics and palanichamy and nagan
19	palanichamy
20	Dr.m.s.palanichamy
21	m.s.palanichamy
22	m s palanichamy
23	ms palanichamy
24	palanichamy-m.s
25	palanichamy m s
26	palanichamy, m s
27	nagan
28	s nagan
29	s.nagan
30	nagan,s
31	nagan.s

32	nagan-s
33	palanichamy and nagan
34	ms palanichamy and s nagan
35	m.s.palanichamy and s.nagan
36	palanichamy m.s and nagan.s
37	palanichamy ms nagan s tata mcgraw hill
38	tata mcgraw hill
39	tata-mcgraw hill
40	tata-mcGraw hill
41	Tata McGraw-Hill Publishing Company Limited
42	tata mcgraw hill new delhi
43	new delhi
44	0-07-058830-9
45	0070588309

DATA COLLECTION AND CONSOLIDATION OF RETRIEVED RECORDS

The above mentioned queries have been invoked using the 5 Search Engines on monthly basis from January to December 2011. Numbers of records retrieved have been recorded carefully and consolidation of the retrieved data has been carried-out. Details are given below.

Observations

It has been found that there is lot of gap between quantum of records retrieved by the 5 Search Engines. Google retrieved almost double the records retrieved by Yahoo and AltaVista. Both Rediff and MSN retrieved very less records when compare to Google, Yahoo and AltaVista. The results are presented below based on their hierarchy of records retrieved.

Information Explosion Activity

Additionally, due to information explosion, there is tremendous growth in publications from January to December 2011. To substantiate this phenomenon, the results are shown in the Bar diagram.

Bar diagram of Consolidation of Records Retrieved

Ranking

Based on the total records retrieved by the 5 Search Engines, by combining all the 45 queries, ranking has been assigned from 1 to 5. The Search Engine which retrieved maximum number of records has been assigned 1st rank and the Search Engine which retrieved least number of records has been

S.no	Month & Year	No. of Records Retrieved by Search Engines				
		Google	AltaVista	Yahoo	MSN	Rediff
1	January 2011	63,277,241	32,481,328	46,272,656	41,643,789	836,184
2	February 2011	107,775,631	88,015,390	85,118,411	51,652,765	1,006,666
3	March 2011	112,881,140	55,005,361	84,045,916	28,675,351	20,845,346
4	April 2011	110,821,840	85,656,852	72,257,014	23,431,059	1,213,675
5	May 2011	116,331,426	65,298,899	92,028,841	57,582,615	25,781,690
6	June 2011	201,152,397	67,979,637	82,797,587	35,720,162	32,427,134
7	July 2011	218,017,970	67,148,539	94,391,938	48,448,521	10,883,949
8	August 2011	234,571,218	110,335,019	96,350,879	60,677,414	12,323,652
9	September 2011	365,656,831	232,804,231	90,848,865	81,526,554	62,255,657
10	October 2011	433,142,454	255,742,766	84,103,636	70,750,383	31,463,504
11	November 2011	286,035,071	267,971,619	128,254,248	130,389,711	94,465,426
12	December 2011	504,364,419	320,118,391	125,974,935	127,515,543	42,160,912
Total:		2,754,027,638	1,648,558,032	1,082,444,926	758,013,867	335,663,795

S.no	Search Engines	No. of Records Retrieved
1	Google	2,754,027,638
2	AltaVista	1,648,558,032
3	Yahoo	1,082,444,926
4	MSN	758,013,867
5	Rediff	335,663,795
Total:		6,353,866,303

assigned 5th rank. Rests of them have fallen between the ranks 2 to 4. Details are given below.

S.no	Search Engines	No. of Records Retrieved	Rank
1	Google	2,754,027,638	1
2	AltaVista	1,648,558,032	2
3	Yahoo	1,082,444,926	3
4	MSN	758,013,867	4
5	Rediff	335,663,795	5

CONCLUSION

Concluded that there is lot of gap between quantum of records retrieved by the 5 Search Engines. Google retrieved almost double the records retrieved by Yahoo and AltaVista. Both Rediff and MSN retrieved very less records when compared to Google, Yahoo and AltaVista. We conclude that Google gets 1st rank among the 5 Search Engines considered for the purpose.

BIBLIOGRAPHY

Atsaros, G., Spinellis, D., & Louridas, P. (2008). *Site-Specific Versus General Purpose Web Search Engines: A Comparative Evaluation*.

Bitirim, Y., Tonta, Y., & Sever, H. (2002). Information retrieval effectiveness of Turkish search engines. *Advances in Information Systems: Second International Conference*, 93-103.

Baeza-Yates, R; Hurtado, C., & Mendoza, M. (2004). Query recommendation using query logs in search engines. *In Advances in Web Intelligence: Second International Atlantic Web Intelligence Conference, AWIC*, 164-175.

Bradley, P. (2006). Search engines: Where we were, are now, and will ever be. *Ariadne*, 47. Retrieved from <http://www.ariadne.ac.uk/issue47/search-engines/>

Bedi, P., & Chawla, S. (2007). Improving information retrieval precision using query log mining and information scent. *Information Technology Journal*, 6(4), 584-588.

Bradley, P. (2008). Human powered search engines: An overview and round up. *Ariadne*, 54. Retrieved from <http://www.ariadne.ac.uk/issue54/searchengines/>

Bradley, P. (2008). Search Engines: Google Still Growing. *Ariadne*, 56.

Demirci, R.G., Kismir, V., & Bitirim, Y. (2007). An evaluation of popular search engines on finding Turkish documents. *Computer Engineering Department, Eastern Mediterranean University, Second International Conference on Internet and Web Applications and Services*, P 61.

- Dragutsky, P. (2001). *Metasearch Engine Review: Vivisimo*. Retrieved from archive.suite101.com
- Gudivada, V. N., Raghavan V. V., Grosky, W. I., & Kasangottu, R. (1997). Information retrieval on the World Wide Web. *IEEE Internet Computing*, 1(5), 58-68.
- Jansen, B. J., Spink, A., Bateman, J., & Saracevic, T. (1998). Real life information retrieval: A study of user queries on the web. *ACM SIGIR Forum*, 32(1), 5-17.
- Retrieved from <http://realestate.about.com/od/sv/g/def-seacheng.html>
- Retrieved from <http://www.arnoldit.com/lists/intlsearch.asp>
- Retrieved from www.webopedia.com/TERM/s/search_engine.html
- Retrieved from http://whatis.techtarget.com/definition/0,,sid9_gci1146695,00.html