

Ranking of Search Engines in View of Information Explosion

Dr. K Ramakrishna Reddy

Chief Librarian, Acharya Institutes, Acharya Dr. Sarvepalli Radhakrishnan Road, Hesaraghatta Main Road, Soldevanahalli,
Bangalore, India.

and

Haritha Raja Harinath

MLIS Student, Indira Gandhi National Open University, New Delhi, India.

Abstract:

Purpose – Deals with collection of number of records published on Web, on 23 Main Classes of Colon Classification from 1st January to 31st March 2014 on daily basis, using 5 Search Engines – Google, Lycos, Microsoft, Rediff and Yahoo. Ranking of 5 Search Engines, 23 Main Classes, and the months of January, February and March 2014 has been carried-out, based on number of records they retrieved during the period of study.

Design/Methodology/Approach – Searching of Web, using 23 Main Classes of Colon Classification, on daily basis, has been adopted, and number of records retrieved have been analysed and presented.

Findings – The findings are presented under following 3 headings in the body of the article based on the number of records retrieved: (1) Ranking of 5 Search Engines, (2) Ranking of January, February and March, and (3) Ranking of 23 Main Classes of Colon Classification.

Originality/Value – The present work is original in nature. The study will be useful to the users for deciding the Search Engine/s of their choice.

Keywords - Search Engines - Google, Lycos, Microsoft, Rediff and Yahoo; Information Explosion; Definitions, Data Collection; Main Classes of CC; Consolidation of Records; Ranking of Search Engines, Main Classes. and months.

Paper Type – Research paper.

1. Introduction

The Libraries around the world are collecting, processing, organizing, and disseminating information recorded in the documents. The sources of information, few centuries ago, were in the form of papyrus, parchment, palm and birch leaves, tablets, stone and rock carvings, etc., until invention of paper by China. For several centuries, paper has been used for documenting information in the form of Books. But, at a later date, in addition to Books, paper has been extensively used to bring-out other forms of documents, such as, Periodicals, Reports, Conference documents, Thesis and Dissertations, Patents, Standards and Specifications, Souvenirs, etc.

It is a well known fact that, modern day Libraries are not mere repositories of Books, Periodicals and other publications, limiting themselves to just circulation among users, but their tentacles reached the users in various means. The type of documents held by Libraries has been changed

to a great extent. For example, the Libraries are emphasising on collection of documents in the form of Electronic resources, such as, E-Books, E-Journals, E-Texts, E-Databases, etc. Libraries in order to reach out to maximum number of customers, to increase usage of documents, to satisfy their customers, and to bring publications closer to the users, offers variety of innovative services, such as, Alerting service, Bibliographic and Literature search service, Full text service, Document delivery service, Reference and Referral services, etc.

Due to emergence of Internet facility, Libraries are satisfying needs of the users at a faster rate by searching the World Wide Web (WWW), utilising readily available Search Engines, such as, AltaVista, Google, Inktomi, Lycos, Microsoft, MSN, Netscape, Rediff, Yahoo etc. Earlier, the information seeker had to contact the Librarian for required information or had to know exact address of the website hosted on the web or location of the file on the web and use

FTP service to search and download the file or would have to ask the Librarian to locate the source in database, etc. But, now-a-days, it is observed that Search Engines have come close to providing precise or appropriate information to the users. Simply, the user has to enter the keywords and the Search Engine will retrieve appropriate information.

1.1 Information Explosion

Online New World Encyclopaedia defines Information Explosion as “rapidly increasing amount of published information and the effects of this abundance of data”. As the amount of available data grows, managing the information becomes more difficult, which can lead to information overload. In-turn, Information overload leads to the following:

- Overloading of the internal processing system.
- Decrease in the quantity and/or quality of the information that one generates or outputs.
- Rarely, a system shut down is also possible.

In this paper, efforts are made to study the most popular form of information retrieval, i.e., through Search Engines. For the study, 5 Search Engines, namely, Google, Lycos, Microsoft, Rediff and Yahoo have been considered.

1.2 Objectives of the Study

- To find-out how much Information Explosion is occurring in the 23 Main Classes of Colon Classification designed by Dr. S. R. Ranganathan.
- To find-out which Main Class tend to generate more information on daily basis.
- To find-out number of records captured by the 5 Search Engines on each of the 23 Main Classes.
- To find-out number of records retrieved during each of the 3 months

1.3 Methodology

- 5 Search Engines, i.e., Google, Lycos, Microsoft, Rediff and Yahoo are considered for study.
- 23 Main Classes included under 3 broad categories

(Natural Sciences, Humanities, and Social Sciences) of Universe of Knowledge have been considered for the study.

- The web was searched on the 23 Main Classes by using designated 5 Search Engines on daily basis and results were recorded.
- Based on the recorded results, comparative analysis of search results has been carried-out and provided.

1.4 Scope of the Study

- 5 Search Engines, namely, Google, Lycos, Microsoft, Rediff, and Yahoo are considered for the study.
- The study is limited to 23 Main Classes enumerated in the Colon Classification.
- The study is limited to 3 months from January to March 2014.

2. Data Collection, Analysis and Findings

2.1 Criteria for Data Collection

Here, an attempt has been made to study Information Explosion pattern based on the quantum of records retrieved against 23 Main Classes. In order to achieve desirable results, following criteria have been followed.

- Collected data through web.
- Duration of data collection is for 3 months, on daily basis, from January to March 2014.
- For data collection, 5 Search Engines, i.e., Google, Lycos, Microsoft, Rediff and Yahoo have been considered.
- 23 Main Classes have been selected for data collection.
- Queries are invoked without any discrepancies.

2.2 Main Classes Selected for the Study

The Colon Classification (6th edition) developed by Padmashree Dr. S R Ranganathan listed number of Main Classes of Subjects. Out of which, only 23 Main Classes, as mentioned below, were selected randomly, for the present study.

Main Class CC No.	Main Class	Main Class CC No.	Main Class	Main Class CC No.	Main Class
B	Mathematics	K	Zoology	T	Education
C	Physics	L	Medicine	U	Geography
E	Chemistry	N	Fine Arts	V	History
F	Technology	O	Literature	W	Political Science
G	Biology	P	Linguistics	X	Economics
H	Geology	Q	Religion	Y	Sociology
I	Botany	R	Philosophy	Z	Law
J	Agriculture	S	Psychology		

2.3 Data Collection

The above mentioned Main Classes have been invoked using the 5 Search Engines on daily basis from 1st January to 31st March 2014. Number of records retrieved was recorded carefully under each of the Search Engines. However, for convenience, all the above mentioned Main

Classes are arranged alphabetically while presenting the collected data. As it is very cumbersome to accommodate number of records retrieved by all 5 Search Engines, every day of 3 months, details of first day of each of the three months are shown below as samples.

January 2014	Day 1 to 31				
	Google	Lycos	Microsoft	Rediff	Yahoo
Agriculture	112,000,000	89,400,000	206,000,000	11,819	219,000,000
Biology	72,600,000	96,500,000	200,000,000	5,029	220,000,000
Botany	9,510,000	25,500,000	26,100,000	989	25,400,000
Chemistry	1,330,000,000	96,700,000	230,000,000	4,607	224,000,000
Economics	90,000,000	91,100,000	190,000,000	38,979	188,000,000
Education	825,000,000	1,040,000,000	1,110,000,000	44,772	1,100,000,000
Fine Arts	405,000,000	97,300,000	206,000,000	3,518	206,000,000
Geography	48,000,000	102,000,000	225,000,000	2,077	227,000,000
Geology	18,600,000	53,600,000	60,600,000	2,555	54,400,000
History	1,060,000,000	2,870,000,000	4,130,000,000	55,917	4,470,000,000
Law	441,000,000	431,000,000	600,000,000	63,230	542,000,000
Literature	181,000,000	122,000,000	276,000,000	4,713	273,000,000
Mathematics	102,000,000	95,900,000	176,000,000	2,653	162,000,000
Medicine	224,000,000	238,000,000	433,000,000	10,274	381,000,000
Mining	71,900,000	78,600,000	101,000,000	21,233	112,000,000
Philosophy	94,800,000	92,900,000	198,000,000	3,501	211,000,000
Physics	91,200,000	96,000,000	203,000,000	13,716	214,000,000
Political Science	946,000,000	83,000,000	142,000,000	1,689	94,100,000
Psychology	76,500,000	120,000,000	119,000,000	4,335	211,000,000
Religion	214,000,000	125,000,000	270,000,000	5,492	260,000,000
Sociology	20,000,000	57,300,000	72,500,000	1,232	72,600,000
Technology	778,000,000	455,000,000	536,000,000	60,989	502,000,000
Zoology	9,440,000	22,400,000	24,600,000	1,317	28,500,000
Total:	7,220,550,000	6,579,200,000	9,734,800,000	364,636	9,997,000,000

<i>February 2014</i>	<i>Day 1 to 28</i>				
	<i>Google</i>	<i>Lycos</i>	<i>Microsoft</i>	<i>Rediff</i>	<i>Yahoo</i>
Agriculture	107,000,000	113,000,000	210,000,000	13,975	199,000,000
Biology	71,000,000	116,000,000	216,000,000	5,681	211,000,000
Botany	11,500,000	16,200,000	30,800,000	1,005	28,000,000
Chemistry	91,200,00	119,000,000	210,000,000	5,388	203,000,000
Economics	92,000,000	106,000,000	196,000,000	47,547	185,000,000
Education	793,000,000	547,000,000	1,080,000,000	54,095	1,020,000,000
Fine Arts	483,000,000	117,000,000	200,000,000	3,958	183,000,000
Geography	46,400,000	122,000,000	210,000,000	2,321	205,000,000
Geology	17,900,000	35,900,000	55,700,000	2,693	54,800,000
History	1,030,000,000	1,970,000,000	4,260,000,000	66,098	3,400,000,000
Law	430,000,000	331,000,000	535,000,000	76,247	561,000,000
Literature	180,000,000	147,000,000	304,000,000	5,362	267,000,000
Mathematics	93,000,000	161,000,000	245,000,000	2,876	229,000,000
Medicine	218,000,000	223,000,000	352,000,000	12,178	391,000,000
Mining	76,400,000	57,200,000	93,700,000	25,312	102,000,000
Philosophy	93,900,000	112,000,000	178,000,000	3,939	197,000,000
Physics	89,800,000	116,000,000	205,000,000	16,121	191,000,000
Political Science	1,180,000,000	68,200,000	88,800,000	1,797	99,300,000
Psychology	69,000,000	118,000,000	190,000,000	4,933	188,000,000
Religion	204,000,000	142,000,000	269,000,000	6,336	293,000,000
Sociology	21,100,000	39,900,000	52,900,000	1,281	60,000,000
Technology	767,000,000	291,000,000	616,000,000	74,085	556,000,000
Zoology	9,560,000	18,900,000	29,900,000	1,381	30,200,000
Total:	6,083,560,000	5,087,300,000	9,827,800,000	434,609	8,853,300,000

<i>March 2014</i>	<i>Day 1 to 31</i>				
	<i>Google</i>	<i>Lycos</i>	<i>Microsoft</i>	<i>Rediff</i>	<i>Yahoo</i>
Agriculture	114,000,000	108,000,000	200,000,000	16,458	196,000,000
Biology	62,000,000	116,000,000	183,000,000	6,369	185,000,000
Botany	9,540,000	17,000,000	29,700,000	1,030	31,300,000
Chemistry	92,000,000	116,000,000	229,000,000	5,791	196,000,000
Economics	97,300,000	102,000,000	156,000,000	54,408	91,500,000
Education	897,000,000	756,000,000	1,390,000,000	63,030	750,000,000
Fine Arts	457,000,000	114,000,000	210,000,000	4,589	210,000,000
Geography	44,900,000	115,000,000	202,000,000	2,409	219,000,000
Geology	17,000,000	32,300,000	45,700,000	2,960	51,600,000
History	1,040,000,000	1,620,000,000	3,050,000,000	79,159	3,350,000,000
Law	435,000,000	325,000,000	636,000,000	89,258	554,000,000
Literature	185,000,000	140,000,000	290,000,000	5,830	279,000,000
Mathematics	91,600,000	155,000,000	247,000,000	3,130	229,000,000
Medicine	221,000,000	214,000,000	386,000,000	13,772	365,000,000
Mining	71,700,000	55,400,000	102,000,000	29,588	94,400,000
Philosophy	96,100,000	113,000,000	157,000,000	4,313	180,000,000
Physics	90,700,000	113,000,000	216,000,000	18,291	115,000,000
Political Science	1,130,000,000	64,300,000	96,300,000	2,094	84,100,000
Psychology	70,100,000	115,000,000	92,500,000	5,455	184,000,000
Religion	275,000,000	141,000,000	250,000,000	7,082	271,000,000
Sociology	22,100,000	38,100,000	61,400,000	1,351	63,100,000
Technology	785,000,000	289,000,000	297,000,000	86,481	423,000,000
Zoology	8,270,000	16,300,000	29,900,000	1,411	22,400,000
Total:	6,312,310,000	4,875,400,000	8,556,500,000	504,259	8,144,400,000

2.4 Consolidation of Records Retrieved

The retrieved records have been consolidated and a

statement on monthly basis, from January to March 2014,

using the 5 Search Engines is given below:

S.N	Search Engines	No. of Records Retrieved by Search Engines from Jan – Mar 2014			
		January	February	March	Total
1	Google	7,220,550,000	6,083,560,000	6,312,310,000	19,616,420,000
2	Lycos	6,579,200,000	5,087,300,000	4,875,400,000	16,541,900,000
3	Microsoft	9,734,800,000	9,827,800,000	8,556,500,000	28,119,100,000
4	Rediff	364,636	434,609	504,259	1,303,504
5	Yahoo	9,997,000,000	8,853,300,000	8,144,400,000	26,994,700,000
	Total:	33,531,914,636	29,852,394,609	27,889,114,259	91,273,423,504

2.5 Ranking of Search Engines

In order to study the Information Explosion activity, as preamble, the 5 Search Engines are ranked from 1 to 5 in

descending order, based on total number of records retrieved.

S.N.	Search Engines	No. of Records Retrieved	%	Rank
1	Microsoft	28,119,100,000	30.80	1
2	Yahoo	26,994,700,000	29.57	2
3	Google	19,616,420,000	21.49	3
4	Lycos	16,541,900,000	18.12	4
5	Rediff	1,303,504	00.02	5
	Total:	91,273,423,504	100.00	

2.5.1 Findings

- It has been found that Microsoft retrieved highest number of records (30.80%) and Rediff retrieved least number of records (00.02%). Hence, Microsoft got 1st Rank.
- Yahoo (29.57%), Google (21.49%) and Lycos (18.12%) retrieved records that are in between Microsoft and Rediff. Hence, Yahoo, Google and Lycos are placed on 2nd Rank, 3rd Rank and 4th

Rank, respectively. As Rediff retrieved very less number of records (00.02%) placed at 5th Rank.

2.6 Information Explosion Activity

Due to Information Explosion, there is tremendous growth in number of publications. Analysis has been made on number of records received on monthly basis from January to March 2014, and the results are given below.

S.N.	Month	No. of Records Retrieved	%	Rank
1	January	33,531,914,636	36.74	1
2	February	29,852,394,609	32.71	2
3	March	27,889,114,259	30.55	3
Total:		91,273,423,504	100.00	

2.6.1 Findings

- Based on the number of records retrieved, it has been observed that the Information Explosion activity during all the 3 months is in the decreasing order of January (36.74%), February (32.71%) and March (30.55%).

- Hence, the 3 months ranked in the same order.

2.7 Ranking of Main Classes

An attempt has also been made to rank the Main Classes based on total number of records retrieved from January to March 2014. Results are shown below.

S.N	Main Class	January	February	March	Total	%	Rank
1	History	12,530,055,917	10,660,066,098	9,060,079,159	32,250,201,174	35.33	1
2	Education	4,075,044,772	3,440,054,095	3,793,063,030	11,308,161,897	12.39	2
3	Technology	2,271,060,989	2,230,074,085	1,794,086,481	6,295,221,555	6.90	3
4	Law	2,014,063,230	1,857,076,247	1,950,089,258	5,821,228,735	6.38	4
5	Political Science	1,265,101,689	1,436,301,797	1,374,702,094	4,076,105,580	4.47	5
6	Medicine	1,276,010,274	1,184,012,178	1,186,013,772	3,646,036,224	3.99	6
7	Chemistry	1,880,704,609	532,005,388	633,005,791	3,045,715,788	3.34	7
8	Fine Arts	914,303,518	983,003,958	991,004,589	2,888,312,065	3.16	8
9	Religion	869,005,492	908,006,336	937,007,082	2,714,018,910	2.97	9
10	Literature	852,004,713	898,005,362	894,005,830	2,644,015,905	2.90	10
11	Mathematics	535,902,653	728,002,876	722,603,130	1,986,508,659	2.18	11
12	Agriculture	626,411,819	629,013,975	618,016,458	1,873,442,252	2.05	12
13	Geography	602,002,077	583,402,321	580,902,409	1,766,306,807	1.94	13
14	Biology	589,105,029	614,005,681	546,006,369	1,749,117,079	1.92	14
15	Physics	604,213,716	601,816,121	534,718,291	1,740,748,128	1.91	15
16	Philosophy	596,703,501	580,903,939	546,104,313	1,723,711,753	1.89	16
17	Economics	559,138,979	579,047,547	446,854,408	1,585,040,933	1.74	17

					4		
18	Psychology	526,504,335	565,004,933	461,605,455	1,553,114,723	1.70	18
19	Mining	363,521,233	329,325,312	323,529,588	1,016,376,133	1.11	19
20	Sociology	222,401,232	173,901,281	184,701,351	581,003,864	0.64	20
21	Geology	187,202,555	164,302,693	146,602,960	498,108,208	0.54	21
22	Botany	86,510,989	86,501,005	87,544,000	260,555,994	0.28	22
23	Zoology	84,941,317	88,561,381	76,871,411	250,374,109	0.27	23
Total:		33,531,914,636	29,852,394,609	27,889,114,259	91,273,423,504	100.00	

2.7.1 Findings

- Maximum Information Explosion has been observed under the Main Class History (35.33%) followed by Education (12.39%).
- The Main Class Zoology recorded least number of records (0.27%) when compared to other Main Classes.
- Rest of the Main Classes falls between History and Zoology. Their rankings are shown in the above table.

3. Conclusion

When compared to number of records retrieved by the 5 Search Engines, Microsoft occupied top position (30.80%), followed by Yahoo (29.87%), Google (21.49%), Lycos (18.12%) and Rediff (00.02%). Number of records retrieved by Microsoft and Yahoo are close to each other, but there is lot of gap when compared to Google, Lycos and Rediff. Rediff constitutes negligible percentage occupying last position in ranking.

When ranking of 23 Main Classes is observed, it is seen that higher numbers for the first seven Main Classes, namely - History, Education, Technology, Law, Political Science, Medicine and Chemistry making them probably the most sought out ones by users on the Web. Secondly, it is observed that Fine Arts, Religion, Literature, Mathematics, Agriculture, Geography, Biology, Physics, Philosophy, Economics, Psychology, Mining, Sociology and Geology are also reasonably popular Main Classes in the same sequence. Botany and Zoology are least popular Main Classes.

Another interesting fact is that it is quite surprising to see the amount of information available on the Web. There is information explosion given the number of records retrieved in so many different Main Classes, but when it comes to information explosion in individual Main Classes, it is not a continuous growth as thought, but growth is seen happening in pockets as shown in the data from January to March 2014, where some Main Classes retrieved more in the month of January, than in March and in many Main Classes the February month retrieved lowest number of records. It is concluded that there is lot of information on the Web and many Databases, Search Engines, Portals etc., have indexed to make it simpler for user to search the vast Web and continue to do so as Search Engines do not capture the entire Web / Internet.

References

- [1] "A brief history of search engines". <<http://www.livemint.com/Consumer/ZPKAhHlIjQa79t4OMQy2I/A-brief-history-of-search.html>>
- [2] **Arora, Jagdish**. "Search engines. In: Information communication technologies: applications. Part II", *New Delhi, Karan Press, 2006*
- [3] **Asadi, Saeid and Jamali, Hamid R.** "Shifts in search engine development: a review of past, present and future trends in research on search engines", *Webology, 1(2), 2004, Article 6*. <<http://www.webology.org/2004/v1n2/a6.html>>
- [4] **Hanani, Uri and Frank, Ariel J.** "The parallel evolution of search engines and digital libraries: Their convergence to the mega-portal". <<http://u.cs.biu.ac.il/~ariel/download/ird665/kyoto.pdf>>
- [5] **New World Encyclopaedia.** "Information Explosion, 2014". <https://www.newworldencyclopedia.org/entry/Information_explosion>

-
- [6] **Rather, Rafiq Ahmed; Lone, Fayaz Ahmed and Shah, Gulam Jeelani.**
- [7] “Overlap in web search results: a study of 5 search engines, 2008”. *Library*
a. *Philosophy and Practice*
- [8] <<http://www.webpages.uidaho.edu/~mbolin/rather-lone-shah.htm>>
- [9] **Seymour, Tom; Frantsvog, Dean and Kumar, Satheesh.** “History of search engines”, *International Journal of Management & Information Systems*, 2011, 15(4), p47-58.
<<http://journals.cluteonline.com/index.php/IJMIS/article/view/5799/5877>>
- [10] **White, Jacquelyn P.** “The effects of the information explosion on information literacy, 2009”.
- [11] <<http://jacquelynpwhite.files.wordpress.com/2009/09/theeffectsoftheinformatio-nexplosiononinformationliteracyedits-doc.pdf>>
- [12] **Wikipedia.** “Lycos, 2014”.
a. <<http://en.wikipedia.org/wiki/Lycos>>
- [13] **Wikipedia.** “Microsoft, 2014”
a. <<http://en.wikipedia.org/wiki/Microsoft>>
- [14] **Wikipedia.** “Rediff, 2014”.
- [15] <<http://en.wikipedia.org/wiki/Rediff>>
- [16] **Wikipedia.** “Yahoo!, 2014”.
a. <<http://en.wikipedia.org/wiki/Yahoo!>>
- [17] **Wall, Aaron.** “AltaVista, © 2004-2013”.
a. <<http://www.search-marketing.info/search-engines/major-search-engines/altavista.htm>>