IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

Future of Results in Select Search Engine

Peerzada Mohammad Iqbal[1], Dr. Abdul Majid Baba[2], Aasim Bashir[3]

Professional Asst, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir (SKUAST-K), India[1]

Head, Department of Library and Information Science, The University of Kashmir, India[2]

Assistant Professor, Department of Computer Science, The University of Kashmir, India[3]

Abstract: The paper is an attempt to evaluate projected data of select search engine on the basis of seasonal data collection. The attempt is to generate the future of results in select search engines using simple keyword "Indexing" from the field of Library and Information Science. The Data collection was carried out for 100 Days to generate 50 days of predicted data using project trend method. The search engines viz., Google, Bing, Yahoo! & Baidu selected on certain parameters show tremendous variations in results. The evaluation reveal that Bing shows a significant positive secular trend while Google, Yahoo! and Baidu show a downward or negative secular trend.

Keywords: Future, Search engine, Fluctuation, Indexing, Counter.

INTRODUCTION

The internet is considered as the first hand source of information for daily usage, covering every aspect of our life. From research to navigation every bit of information is covered, searched and accessed [1,2] by various search engines [3]. However the results yielded for a number of queries rank in several thousand or even in millions due to the availability of infinite amount of information. However many studies show that only first few results are browsed by the users[4,5,6,7], which determines the success of a search engine therefore result ranking holds utmost importance in this regard. Result ranking was merely based on term frequency and the inverse document frequency in case of classical IR system[8].Various parameters are taken into account in Web search results ranking as number of links pointing to a given web page[9,10], the anchor text of the links pointing to the web page, the placement of the search terms in the document (terms occurring in title or header may get a higher weight), the distance between the search terms, popularity of the page (in terms of the number of times it is visited), the text appearing in metatags[11], subject specific authority of the web page[12.13], recently in search index and exactness of the hits[14]. There is always an ongoing competition between search engines and Web page authors for users and high ranking respectively, which is why the algorithm ranking are kept a secret by the search engine companies as Google states[14], "Due to the nature of our business and our interest in protecting the integrity of our search results, this is the only information we make available to the public about our ranking system". Apart from this search engines keep on updating and upgrading their algorithm so to improve their ranking of results.

Nowadays search engine optimization industries are present which design and redesign Web pages in order to enhance their rankings within a specific search engine (e.g., search engine optimization Inc., www.seoine.com/). Therefore in the crux it can be concluded that the First ten results retrieved for a query have major chances of being visited by the users. In addition to the examination of changes overtime for the top ten results related to a query of the largest search engine, which at the times of first data collection were Google, yahoo and Tacoma (MSN search came out if beta on Feb 1st 2005 in the midst of data collection for the second round[15]. However various transformations between the user's "visceral need" (a fuzzy view of the information problem in user's mind) and the "compromised need" (the way the query is phrased taking into account the limitations of the search tool at hand)[16]. Above all the fluctuation of a result related to a query can only be judged by the user while some researchers claim that it is impractical due to the presence of a large number of documents related to a query and all of them can't be viewed by the user, hence for checking fluctuation a panel of judges is required [17, 18].

PROBLEM

In the beginning of the web, finding information on the web was direct. With some informative programming framework that some of the time charge driven as opposed to utilizing a graphical interface. With the multiplication of information, frameworks similar to Archie, Gopher and speedwell turned out to be continuously not able to jog out huge information. the landing of the numerous styles of web crawlers gave determination to writing seek exploitation mathematician administrators, Closeness watching out, Trump cards, Truncation and so forth a few web indexes grew new forms and methods to understand some sensibly modernity anyway all haven't sent the instance of access and watching out from researcher's point of view.

Other than keeping noticeable option methods for arrangement the web, web indexes work in a few routes in which and recover archives in a few requests. Further, it

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

doesn't filter information from researcher's motivation of read i.e., it recovers information on a particular point from totally distinctive perspectives like advancing, advertizement, news and diversion blended with some investigation papers. the instructional exercise group tries to appear to be entirely for significant information on his theme important to have yield/recovery best as far as thoroughness and dejected of changes and so forth.
The present investigation attempts to evaluate the performance of the select search engines in terms of result fluctuation captured in two phases to check the consistency of search engines.

Objectives:
- To select search engines.
- To select search term for the study.
- To collect data for 100 days.
- To compare trending by forecasting of time series analysis.

Methodology
The ISO (International Standard Organization) certifies 230 search engines working on the web covering different subjects and working with different platforms [19]. These search engines are of various types like general search engine, robotic search engine, Meta search engine, directories and specialized search engines. Most users prefer robotic search engines as they allow the users to compose their own quires rather than simply follow pre specified search paths or hierarchy as in case of directories. Moreover, robotic search engines locate data in a similar way i.e., by the use of crawlers or worms. This distinguishing feature differentiates them form web directories like Yahoo! Where collections of links to retrieve URL's are created and maintained by subject experts or by means of some automated indexing process. However some of these services are also include a robot driven search engine facility. But this is not their primary purposes. This due to this feature Yahoo! Was included for the study.

Meta search engine e.g., Dogpile etc don't have their own database. These access the database of many robotic search engines simultaneously. Thus these were excluded for the study. Still hundreds of robotic general search engines navigate the web, in order to limit the scope of study after preliminary study, following criteria was laid down for selection of general search engines:-

a) Availability of automated indexing
b) Global coverage to data.
c) Quick response time.
d) Availability of result counter.

Following two general search engines were selected for the study for meeting all the criteria and being comprehensive in nature.
a) Google.
b) Baidu.

Since the study relates to the field of Library and Information Science but there is no specialized search engine in the subject so another specialized search engine which relates to the subject area i.e., Bing was taken for stydy. Thus the search engines undertaken for evaluation of study are:- 1) Google.2) Bing, 3) Yahoo!, and 4) Baidu.

SELECTION OF TERMS
Selection of terms is not directly possible in development and multidimensional field like Library and Information Science. Therefore, classification schemes like DDC (18^{th}) and DDC (22^{nd}) were consulted to understand Broad/Narrow structure of Library and Information Science. It helped to get five terms/Fields i.e.,

a) Information System.
b) Digital Library.
c) Library Automation.
d) Library Services.
e) Librarianship.

These terms were then browsed in "LC list of subject Headings" which provided many other related terms (RT) and Narrow terms (NT). Further NT and RT attached to each other preferred or standard terms were also browsed which retrieve a large number of Library and Information Science terms. At first instance 140 Library and Information Science related terms were identified.
Some terms occurred more than once and duplication removed. It reduced the number to 100. Later terms were divided into three broad groups under:
a) Application. b) Transformation. c) Inter-relation.

"Application" denotes utility of Library and Information science in various fields and about 50 terms came under this group. "Transformation" refers to a method of developing or manufacturing library services into practical market and 30 terms fall under this group. "Inter-relation" means transformation/dependence of one subject onto another and 20 terms came under this group.

Further each category is sub-divided into groups.
"Application" into four i.e., "Reference service", "Informatics", "Information Retrieval" & "Information Sources". "Transformation" into two i.e., "Digitization" & "Consortia". "Inter-relation" into two i.e., "Library Network" & "Information System".

The terms in each group were arranged alphabetically and each term was given a tag. Later 19% of the terms were selected from each group using "Systematic Sampling" (i.e., first item selected randomly and next item after specific intervals). It further reduced the number to 19. Finally the selected terms were classified into three groups under "Simple", "Compound" & "Complex Terms" **(Table:-1.0)**. This was done in order to investigate how search engines control and handle simple and phrased terms.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

"Simple Terms" containing a single word were submitted to the search engine in the natural form i.e., without punctuating marks. "Compound Terms" consisting of two words were submitted to the search engines in the form of phrases as suggested by respective search engines and "Complex Terms" composed of more than two words or phrases, were sent to the search engine with suitable Boolean operator "AND" & "OR" between the terms to perform special searches. From the Simple terms the 4th Keyword **"Indexing"** was taken for the study as the other keywords are already taken for other studies.

Table 1.0: Keywords

S. No	Simple terms	Compound Terms	Complex Terms
1	Catchwork	Bibliometric Classification	Digital Library Open Source Software
2	Citation	Citation Analysis	Health Information System
3	Dublincore	Comparative Librarianship	Library Information System
4	Indexing	Digital Preservation	Library Information Network
5	Manuscript	Electronic Repositories	Multimedia Information Retrieval
6	Plagiarism	Library Automation	
7	Reprints	Semantic web	

Change in Indexing

Search engine constantly update their database as documents are added, removed and modified continuously, effecting the database index. These quantitative and qualitative changes are expressed as fluctuations. The quantitative changes are expressed as "Result Fluctuations" and the qualitative changes are expressed as "Document" and "Indexing Fluctuations". A fluctuation may show decrease or increase in number of documents. However, growth in size of the database is a continuous and usual routine of the search engines. Thus increase and decrease is taken into account here.

A "Result Fluctuation" appears when a search engine show increase/decrease in total number of results for a query that is searched at two different intervals of time. In other words the total number of results retrieved for a query in second observation may be less as retrieved in the first observation. Thus result fluctuation appears when there is increase/decrease in the number of results for a query tested over time i.e., the number of results in succeeding observation may be more or less than the results of the preceding observation.

The Future of Search Engine

Event achieved by systematically combining and casting forward in a predetermined way from the data about the past is "Forecasting". It is simply a statement about the future prediction. Forecasts are possible only when a history of data exists. The study collected 100 days of data samples from four search engine out of seven as result-counter was available with Google, Bing, Yahoo and Baidu. The data collection was carried on 15th May, 2016 and ended on 18th of August, 2016 collecting 100 samples for keyword "Indexing" in four search engines **Table:-1.1.** For forecasting process few points were taken into consideration as:

1) Fluctuation of search results and sustainability
2) 100 days of data sampling were taken into consideration **(Table:- 1.1).**
3) As the data is seasonal, Trend Projection Method was taken into consideration.
4) Total results were taken from result search counter of search engine.
5) A forecast of 50 days was generated (**Table:-1.2**).
6) The results were evaluated on a scattered graph with regression line.

Table 1.1:- Time series data for forecasting of Select Search engines for the keyword "Indexing"

	Google			Bing			Yahoo!			Baidu		
Days (t)	Result (Y_t)	Multiplication of Days and Results (tY_t)	Square of Days $(t)^2$	Result (Y_t)	Multiplication of Days and Results (tY_t)	Square of Days $(t)^2$	Result (Y_t)	Multiplication of Days and Results (tY_t)	Square of Days $(t)^2$	Result (Y_t)	Multiplication of Days and Results (tY_t)	Square of Days $(t)^2$
1	47700000	47700000	1	11100000	11100000	1	12900000	12900000	1	6630000	6630000	1
2	47700000	95400000	4	11000000	22000000	4	12800000	25600000	4	6680000	13360000	4
3	47600000	142800000	9	10900000	32700000	9	12800000	38400000	9	6680000	20040000	9
4	74200	29680000	16	10900	43600000	16	39100	1564000	16	5110	20440000	16

DOI 10.17148/IJARCCE.2016.5868

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

	000	0		000			000	00		000		
5	74000000	370000000	25	10900000	545000000	25	39000000	195000000	25	6680000	33400000	25
6	74300000	445800000	36	10700000	642000000	36	39100000	234600000	36	5030000	30180000	36
7	73300000	513100000	49	10700000	749000000	49	39200000	274400000	49	6610000	46270000	49
8	47900000	383200000	64	10700000	856000000	64	12900000	103200000	64	6610000	52880000	64
9	47900000	431100000	81	10800000	972000000	81	12900000	116100000	81	6610000	59490000	81
10	48400000	484000000	100	10900000	1090000000	100	12800000	128000000	100	4920000	49200000	100
11	73700000	810700000	121	10900000	1199000000	121	39200000	431200000	121	6370000	70070000	121
12	73700000	884400000	144	10800000	1296000000	144	39000000	468000000	144	4910000	58920000	144
13	48600000	631800000	169	11000000	1430000000	169	12800000	166400000	169	4870000	63310000	169
14	74000000	1036000000	196	10900000	1526000000	196	39000000	546000000	196	4870000	68180000	196
15	73900000	1108500000	225	10900000	1635000000	225	38800000	582000000	225	6230000	93450000	225
16	49500000	792000000	256	10900000	1744000000	256	12800000	204800000	256	5040000	80640000	256
17	49500000	841500000	289	10900000	1853000000	289	12800000	217600000	289	5040000	85680000	289
18	74200000	1335600000	324	10700000	1926000000	324	38400000	691200000	324	4820000	86760000	324
19	49400000	938600000	361	10700000	2033000000	361	12800000	243200000	361	6210000	117990000	361
20	49500000	990000000	400	10900000	2180000000	400	12800000	256000000	400	5040000	100800000	400
21	73900000	1551900000	441	11100000	2331000000	441	38400000	806400000	441	4890000	102690000	441
22	73900000	1625800000	484	10900000	2398000000	484	38300000	842600000	484	4890000	107580000	484
23	48600000	1117800000	529	11000000	2530000000	529	12800000	294400000	529	4870000	112010000	529
24	74000000	1776000000	576	10900000	2616000000	576	39000000	936000000	576	4870000	116880000	576
25	49400000	1235000000	625	11000000	2750000000	625	12800000	320000000	625	6300000	157500000	625
26	74300000	1931800000	676	11100000	2886000000	676	38500000	1001000000	676	6300000	163800000	676
27	74400000	2008800000	729	11100000	2997000000	729	38400000	1036800000	729	6200000	167400000	729
28	74700000	2091600000	784	11200000	3136000000	784	38700000	1083600000	784	4820000	134960000	784
29	74800000	2169200000	841	11300000	3277000000	841	38700000	1122300000	841	4820000	139780000	841
30	50300000	1509000000	900	11300000	3390000000	900	13000000	390000000	900	6190000	185700000	900
31	75200000	2331200000	961	11200000	3472000000	961	38600000	1196600000	961	6170000	191270000	961
32	74300000	2377600000	1024	11100000	3552000000	1024	38500000	1232000000	1024	6300000	201600000	1024
33	74400000	2455200000	1089	11100000	3663000000	1089	38400000	1267200000	1089	6200000	204600000	1089
34	74700000	2539800000	1156	11200000	3808000000	1156	38700000	1315800000	1156	4820000	163880000	1156
35	51100000	1788500000	1225	11000000	3850000000	1225	12800000	448000000	1225	6150000	215250000	1225
36	51000000	1836000000	1296	10900000	3924000000	1296	12800000	460800000	1296	6270000	225720000	1296
37	51000	18870000	1369	10900	40330000	1369	12800	4736000	136	6270	23199000	1369

DOI 10.17148/IJARCCE.2016.5868

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

	000	00		000	0		000	00	9	000	0	
38	51100 000	19418000 00	1444	11000 000	41800000 0	1444	12800 000	4864000 00	144 4	6270 000	23826000 0	1444
39	75700 000	29523000 00	1521	11000 000	42900000 0	1521	38500 000	1501500 000	152 1	6270 000	24453000 0	1521
40	76100 000	30440000 00	1600	11000 000	44000000 0	1600	38600 000	1544000 000	160 0	6270 000	25080000 0	1600
41	51100 000	20951000 00	1681	11000 000	45100000 0	1681	12800 000	5248000 00	168 1	6150 000	25215000 0	1681
42	51000 000	21420000 00	1764	10900 000	45780000 0	1764	12800 000	5376000 00	176 4	6270 000	26334000 0	1764
43	51000 000	21930000 00	1849	10900 000	46870000 0	1849	12800 000	5504000 00	184 9	6270 000	26961000 0	1849
44	49700 000	21868000 00	1936	11000 000	48400000 0	1936	12800 000	5632000 00	193 6	6310 000	27764000 0	1936
45	49100 000	22095000 00	2025	11000 000	49500000 0	2025	12800 000	5760000 00	202 5	6310 000	28395000 0	2025
46	74100 000	34086000 00	2116	11100 000	51060000 0	2116	38700 000	1780200 000	211 6	6310 000	29026000 0	2116
47	49100 000	23077000 00	2209	11000 000	51700000 0	2209	12800 000	6016000 00	220 9	6310 000	29657000 0	2209
48	73900 000	35472000 00	2304	11300 000	54240000 0	2304	38800 000	1862400 000	230 4	6310 000	30288000 0	2304
49	48800 000	23912000 00	2401	11300 000	55370000 0	2401	12900 000	6321000 00	240 1	3400 00	16660000	2401
50	48700 000	24350000 00	2500	11200 000	56000000 0	2500	12800 000	6400000 00	250 0	6340 000	31700000 0	2500
51	73300 000	37383000 00	2601	11200 000	57120000 0	2601	38700 000	1973700 000	260 1	6260 000	31926000 0	2601
52	48400 000	25168000 00	2704	11300 000	58760000 0	2704	13000 000	6760000 00	270 4	6260 000	32552000 0	2704
53	73300 000	38849000 00	2809	11200 000	59360000 0	2809	38700 000	2051100 000	280 9	6260 000	33178000 0	2809
54	72700 000	39258000 00	2916	11200 000	60480000 0	2916	38500 000	2079000 000	291 6	6260 000	33804000 0	2916
55	72900 000	40095000 00	3025	11200 000	61600000 0	3025	38600 000	2123000 000	302 5	6260 000	34430000 0	3025
56	72300 000	40488000 00	3136	11200 000	62720000 0	3136	38600 000	2161600 000	313 6	6020 000	33712000 0	3136
57	72500 000	41325000 00	3249	11200 000	63840000 0	3249	12900 000	7353000 00	324 9	6020 000	34314000 0	3249
58	71100 000	41238000 00	3364	11300 000	65540000 0	3364	38600 000	2238800 000	336 4	6040 000	35032000 0	3364
59	47300 000	27907000 00	3481	11000 000	64900000 0	3481	13000 000	7670000 00	348 1	6170 000	36403000 0	3481
60	63500 000	38100000 00	3600	11000 000	66000000 0	3600	38900 000	2334000 000	360 0	6170 000	37020000 0	3600
61	64200 000	39162000 00	3721	11100 000	67710000 0	3721	39000 000	2379000 000	372 1	6170 000	37637000 0	3721
62	63500 000	39370000 00	3844	11100 000	68820000 0	3844	39100 000	2424200 000	384 4	6190 000	38378000 0	3844
63	62600 000	39438000 00	3969	11100 000	69930000 0	3969	39600 000	2494800 000	396 9	6190 000	38997000 0	3969
64	46700 000	29888000 00	4096	11100 000	71040000 0	4096	13100 000	8384000 00	409 6	6190 000	39616000 0	4096
65	46700 000	3035500 000	4225	11100 000	7215000 00	422 5	11500 000	7475000 00	422 5	6180 000	40170000 0	4225
66	46600 000	3075600 000	4356	10900 000	7194000 00	435 6	11100 000	7326000 00	435 6	6180 000	40788000 0	4356
67	58700 000	3932900 000	4489	10900 000	7303000 00	448 9	11100 000	7437000 00	448 9	6180 000	41406000 0	4489
68	58800 000	3998400 000	4624	10600 000	7208000 00	462 4	11600 000	7888000 00	462 4	4810 000	32708000 0	4624
69	58800 000	4057200 000	4761	10600 000	7314000 00	476 1	11600 000	8004000 00	476 1	6310 000	43539000 0	4761
70	58900 000	4123000 000	4900	11400 000	7980000 00	490 0	28000 000	1960000 000	490 0	6270 000	43890000 0	4900
71	58000	4118000	5041	11400	8094000	504	27900	1980900	504	4800	34080000	5041

DOI 10.17148/IJARCCE.2016.5868

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

	000	000		000	00	1	000	000	1	000	0	
72	46200000	332640000	5184	114000000	820800000	5184	128000000	921600000	5184	48000000	3456000000	5184
73	46200000	337260000	5329	114000000	832200000	5329	128000000	934400000	5329	48600000	3547800000	5329
74	51600000	381840000	5476	115000000	851000000	5476	27800000	20572000	5476	48600000	3596400000	5476
75	59100000	44325000	5625	116000000	870000000	5625	28000000	21000000	5625	62200000	4665000000	5625
76	46200000	35112000	5776	119000000	904400000	5776	128000000	972800000	5776	62200000	4727200000	5776
77	52600000	40502000	5929	119000000	916300000	5929	27800000	21406000	5929	61900000	4766300000	5929
78	46200000	36036000	6084	119000000	928200000	6084	118000000	920400000	6084	57200000	4461600000	6084
79	46200000	36498000	6241	118000000	932200000	6241	128000000	101120000	6241	57500000	4542500000	6241
80	46200000	36960000	6400	119000000	952000000	6400	118000000	944000000	6400	57200000	4576000000	6400
81	46200000	37422000	6561	118000000	955800000	6561	28000000	22680000	6561	49800000	4033800000	6561
82	46300000	37966000	6724	118000000	967600000	6724	128000000	104960000	6724	57500000	4715000000	6724
83	46200000	38346000	6889	119000000	987700000	6889	118000000	979400000	6889	57200000	4747600000	6889
84	46200000	38808000	7056	118000000	991200000	7056	28000000	23520000	7056	49800000	4183200000	7056
85	68000000	57800000	7225	115000000	977500000	7225	27700000	23545000	7225	56700000	4819500000	7225
86	68300000	58738000	7396	113000000	971800000	7396	27800000	23908000	7396	48500000	4171000000	7396
87	68600000	59682000	7569	113000000	983100000	7569	11300000	98310000	7569	56400000	4906800000	7569
88	68500000	60280000	7744	113000000	994400000	7744	27500000	24200000	7744	56400000	4963200000	7744
89	68600000	61054000	7921	113000000	100570000	7921	11300000	100570000	7921	48600000	4325400000	7921
90	69200000	62280000	8100	113000000	101700000	8100	27300000	24570000	8100	48700000	4383000000	8100
91	69200000	62972000	8281	113000000	102830000	8281	27300000	24843000	8281	48700000	4431700000	8281
92	69500000	63940000	8464	114000000	104880000	8464	27200000	25024000	8464	56000000	5152000000	8464
93	69600000	64728000	8649	113000000	105090000	8649	27400000	25482000	8649	56000000	5208000000	8649
94	69500000	65330000	8836	114000000	107160000	8836	27200000	25568000	8836	56000000	5264000000	8836
95	69500000	66025000	9025	113000000	107350000	9025	27500000	26125000	9025	47300000	4493500000	9025
96	69500000	66720000	9216	113000000	108480000	9216	27600000	26496000	9216	54200000	5203200000	9216
97	69600000	67512000	9409	112000000	108640000	9409	27800000	26966000	9409	47300000	4588100000	9409
98	70000000	68600000	9604	112000000	109760000	9604	28000000	27440000	9604	47300000	4635400000	9604
99	70600000	69894000	9801	111000000	109890000	9801	28400000	28116000	9801	54200000	5365800000	9801
100	44700000	44700000	10000	110000000	110000000	10000	12700000	12700000	10000	54100000	5410000000	10000
$\sum t$	**$\sum(Y_t)$**	**$\sum tY_t$**	**$\sum(t)^2$**	**$\sum(Y_t)$**	**$\sum tY_t$**	**$\sum(t)^2$**	**$\sum(Y_t)$**	**$\sum tY_t$**	**$\sum(t)^2$**	**$\sum(Y_t)$**	**$\sum tY_t$**	**$\sum(t)^2$**
5050	**6067500000**	**304513300000**	**338350**	**1115400000**	**568742000000**	**338350**	**2449100000**	**119286400000**	**338350**	**569620000**	**2835365000000**	**338350**

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

Prediction

There are various methods for time-series forecasting. Here Projection Trend Methods fits for the study as a trend line expresses the prediction to a series of historical data points and then projects the line into the future for medium- to long range forecasts. The research has described the trend component with a line visually to a set of points on a graph. The graph, however, is subject to slightly different interpretations. There are three types of trend projection viz.,

1) Positive Secular Trend or Upward Secular Trend:- it describes the data into a upward or raising trend line.
2) Negative Secular Trend or Downward Secular Trend:- it describes the data into lowering trend line
3) Neutral Secular Trend or Straight Secular Trend:- no changes the data is consistent.

For the study 400 samples were taken into account to generate 200 results of projected data which are described in graphs.

The formula derived for the study is:-

$t_t=b_0 + b_1t$

b_0 and b_1 can be derived as:

$$b_0 = \bar{y} - b_1\bar{t}$$

$$b_1 = \frac{n\Sigma ty_t - \Sigma t\Sigma y_t}{n\Sigma t^2 - (\Sigma t)^2}$$

Where

t = days

y_t = Result of the search query

The projected result **Table 1.2**, shows a vast fluctuation both in terms of positive Secular trend and negative secular trend. The estimate is given by a trending line.

Table 1.2:- Projected data using trend projection method for 50 days for the keyword "Indexing"

Days	Google	Bing	Yahoo!	Baidu
1	59526242	11485212	21828485	5446406
2	59181635	11504071	21421810	5451127
3	58798290	11521609	20973364	5458715
4	58371528	11537535	20482614	5468338
5	58513326	11553836	20552647	5443979
6	58679861	11570525	20648054	5454549
7	58884560	11582764	20775288	5427389
8	59098297	11594821	20936772	5437011
9	58698648	11606668	20463093	5448781
10	58256167	11620884	19945080	5462847
11	57781393	11637812	19377241	5434305
12	57949614	11655092	19478915	5442544
13	58149934	11669944	19607038	5411845
14	57669658	11690532	19022567	5377073
15	57880697	11708907	19148352	5339042
16	58124916	11727668	19302536	5337925
17	57668130	11746822	18705025	5301781
18	57166084	11766379	18054321	5262538
19	57394160	11780045	18153706	5213062
20	56864543	11793399	17464112	5203912
21	56287586	11812945	16715046	5157040
22	56470296	11839517	16755889	5101363
23	56689669	11860434	16824288	5041052
24	56074469	11885216	16044814	4975108
25	56293800	11907298	16120471	4903813
26	55673925	11933402	15295015	4877944
27	55901835	11964007	15334958	4851865
28	56175847	11995933	15404786	4821897
29	56507225	12033010	15521814	4739476
30	56893341	12075803	15679009	4650489
31	56384394	12121044	14882780	4607627
32	56806618	12164946	15027437	4562591
33	57252247	12207156	15211666	4521218
34	57761752	12251353	15438511	4474773

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

35	58348475	12301773	15727725	4369948
36	58019512	12346525	14989324	4312621
37	57649260	12388933	14186088	4258108
38	57238674	12432853	13313354	4201786
39	56789075	12482723	12366114	4143573
40	57381697	12534760	12476026	4083382
41	58068572	12589086	12636337	4021121
42	57703475	12645831	11674326	3951238
43	57292915	12700527	10630958	3883976
44	56837794	12757391	9500273	3814293
45	56273203	12821269	8275860	3743958
46	55620981	12888106	6950823	3671172
47	56114047	12962929	6773499	3595817
48	55449694	13036735	5355683	3517767
49	55949483	13129013	5114982	3436885
50	55257808	13226827	3599723	3052720

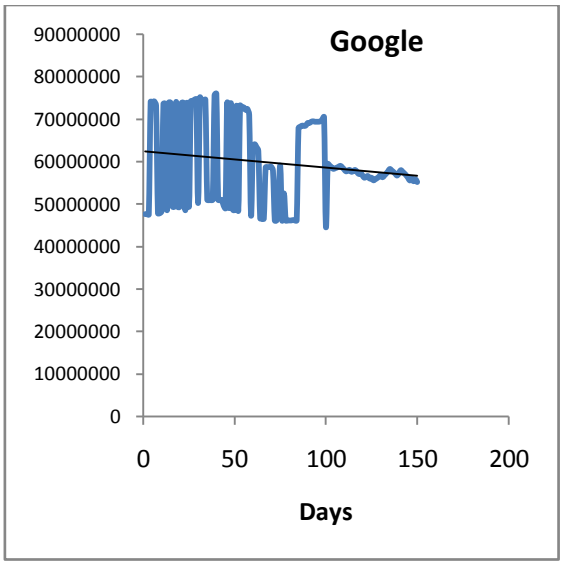

Fig 1.3:- Negative Secular Trend of Google for the keyword "Indexing"

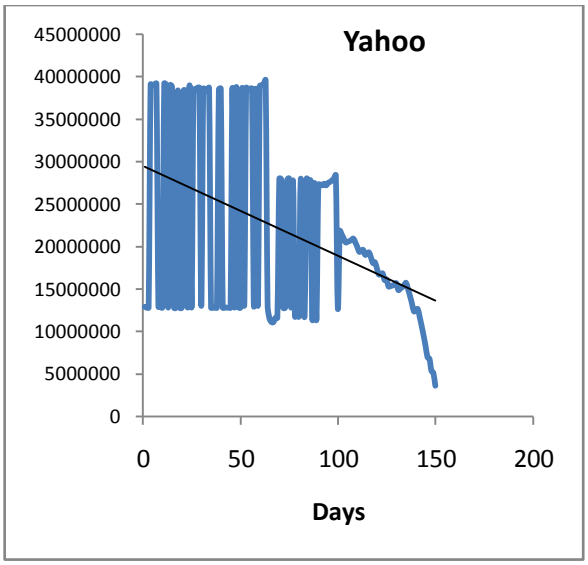

Fig 1.5:- Straight Secular Trend of Yahoo! for the keyword "Indexing"

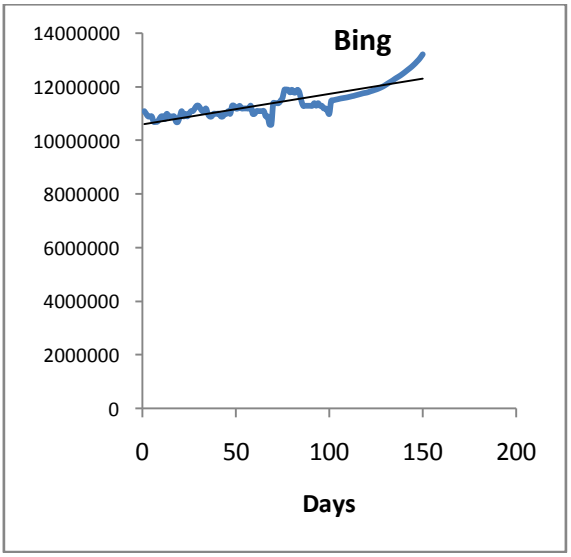

Fig 1.4:- Negative Secular Trend of Bing for the keyword "Indexing"

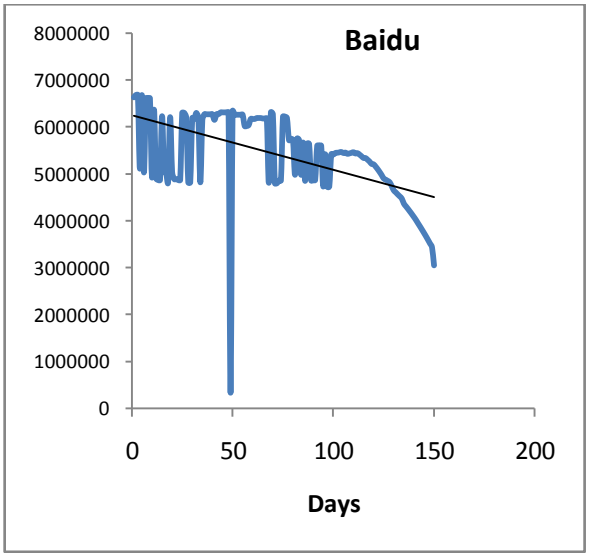

Fig 1.6:- Positive Secular Trend of Baidu for the keyword "Indexing"

DOI 10.17148/IJARCCE.2016.5868

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

REFERENCES

1. Sullivan D, Nielsen Netratings search engine ratings, In Search Engine Watch Reports, 2005. Available at: http://searchenginewatch.com/reports/article.php/2156451
2. Madden M, America's online pursuits: The changing picture of who's online and what they do, PEW Internet & American Life Project, 2003 Available at: http://www.pewinternet.org/pdfs/PIP_Online_Pursuits_Final.PDF
3. Fallows D, The Internet and daily life, PEW Internet & American Life Project, 2004 Available at: http://www.pewinternet.org/pdfs/PIP_ Internet_and_Daily_Life.pdf
4. Silverstein C, Henzinger M, Marais H, and Moricz M, Analysis of a very large Web search engine query log, ACM SIGIR Forum, 33(1) (1999), Available at: http://www.acm.org/sigir/forum/F99/Silverstein.pdf
5. Spink S, Ozmutlu, Ozmutlu H C, and Jansen B J, U.S. versus European Web searching trends, SIGIR Forum, Fall 2002 (2002). Available at: http://www.acm.org/sigir/forum/F2002/spink.pdf
6. Jansen B J, and Spink A, An analysis of Web searching by European Alltheweb.Com Users. Information Processing and Management , 41(6) (2004), 361- 381.
7. Jansen B J, Spink A, and Pedersen J, A temporal comparison of AltaVista Web searching. Journal of the American Society for Information Science and Technology, 56(6) (2005), 559-570.
8. Baeza-Yates R A, and Ribeiro-Neto B A, Modern information retrieval. ACM Press, Addison Wesley: Harlow, England, 1999.
9. Brin S, and Page L, The anatomy of a large-scale hypertextual Web search engine. In Proceedings of the 7th International World Wide Web Conference, April 1998, Computer Networks and ISDN Systems, 30 (1998), 107 - 117. Available at: http://www-db.stanford.edu/pub/papers/google.pdf
10. Google. Google information for Webmasters, 2016. Available at: http://www.google.com/webmasters/4.html
11. Yahoo, Yahoo! Help: How do I improve the ranking of my website in the search results, 2016. Available at: http://help.yahoo.com/help/us/ysearch/ranking/ranking- 02.html
12. Kleinberg J. M, Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5), (1999) 604-632.
13. Teoma (2005). Adding a new dimension to search: The Teoma difference is authority. Accessed on June 26, 2016, from http://sp.teoma.com/docs/teoma/about/searchwithauthority.html
14. MSN Search, Web search help: Change your search results by using Results Ranking, 2005. Available at: http://search.msn.com/docs/help.aspx?t=SEARCH_PROC_BuildCustomizedSearch. htm
15. Payne C, MSN Search launches, 2015. Available at http://blogs.msdn.com/msnsearch/archive/2005/01/31/364278.aspx
16. Taylor R S, Question-negotiation and information seeking in libraries. College and Research Libraries, 29 (May) (1968),178-194.
17. Gordon M, and Pathak P, Finding information of the World Wide Web: The retrieval effectiveness of search engines. Information Processing and Management, 35(1999),141–180.
18. TREC, Data – English relevance judgements, 2014. Available at http://trec.nist.gov/data/reljudge_eng.html.
19. Promote3.com, Top Search Engine Ranking Search Engine Optimization. IDV International: California (2016), Retrieved from http://www.promote3.com/search-engine-230.htm

BIOGRAPHIES

Peerzada Mohd Iqbal Professional Assistant in SKUAST-K, Srinagar, Pursuing Ph.D in Library and Information Science, Masters In Library Science, Masters in Information Technology, Double M.Phil in Library Science and B.ED from Kashmir University. Areas of research: Content Management, Digital Libraries, and System Analysis and Designing, Knowledge management process, Data Mining and Internet Research.

Dr. Abdul Majid Baba Head, DLIS and University Librarian, University of Kashmir. M.A, M.Lib. Sc., NET, M.Phil. (Library Science) Delhi University, Ph.D (Library Science) Kashmir University. Area of interests Digital Libraries, Consortium, Data warehouse, Digitization of rare Manuscripts, Reviewer of research papers etc.

Aasim Bashir Masters In Technology (Computer Science) from Lovely Professional University, Phagwara, India. Masters in Computer Application from Lovely Professional University. He has received his bachelor's degree from Kashmir University. Currently working as Assistant Professor in Department of Computer Sciences, University of Kashmir. His area of research is data mining and knowledge management process, internet technologies, Search Engine Optimization.