

# Time Series Analysis of Google, Bing, Yahoo! & Baidu Using Simple Keyword “Plagiarism”

Peerzada Mohammad Iqbal<sup>1</sup>, Dr. Abdul Majid Baba<sup>2</sup>, Aasim Bashir<sup>3</sup>

<sup>1</sup>Professional Assistant, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir (SKUAST-K), India

<sup>2</sup>Head, Department of Library and Information Science, The University of Kashmir, India

<sup>3</sup>Assistant Professor, Department of Computer Science, The University of Kashmir, India

**Abstract-** This paper provides a comprehensive research on time series analysis of four search engines viz., Google, Bing, Yahoo! and Baidu using simple keyword “*Plagiarism*” from the field of Library and Information Science. The time series analysis is used to forecast result fluctuation using series of data which was collected on daily basis for about 100 Days, to generate 50 days of projected data, and latter a trend line was used to compare the forecast of select search engines. The evaluation reveal that Bing shows a positive secular trend while Baidu, Yahoo! and Google show a downward or negative secular trend.

**Index Terms-** Time series, Fluctuation, Plagiarism, Search engine, Result, Counter, Index.

## I. INTRODUCTION

Information on the web can be searched and access via search engines<sup>1</sup>, Depending on the usability and type of information needed Web has procured as an important source of information in research oriented society. The major activity performed on web is searching information for research purposes mainly via these engines<sup>2,3</sup>. 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<sup>4,5,6,7</sup>, 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<sup>8</sup>. Various parameters are taken into account in Web search results ranking as number of links pointing to a given web page<sup>9,10</sup>, 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<sup>11</sup>, subject specific authority of the web page<sup>12,13</sup>, recently in search index and exactness of the hits<sup>14</sup>. 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<sup>10</sup>, "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/](http://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 1<sup>st</sup> 2005 in the midst of data collection for the second round<sup>15</sup>. 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)<sup>16</sup>. 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<sup>17,18</sup>.

## II. PROBLEM

In early days the internet was simple, restricted and direct. With help of some command driven software finding information was limited and not user friendly. The advent of many types of search engines provided solution for literature search using Boolean operators, Proximity searching, Wild cards, Truncation etc. Many search engines developed new versions and techniques to achieve some kind of sophistication but all have not helped to forward the case of access and searching from scholar's perspective. Besides keeping in view different ways of indexing the internet, search engines operate in different ways and retrieve documents in different orders. Further, it does not sift information from scholar's point of view i.e., it retrieves information on a particular topic from different aspects like marketing, advertisement, news and entertainment mixed with some research papers. The academic community attempts to look purely for scholarly information on his topic of interest to have output/ retrieval best in terms of comprehensiveness and devoid of fluctuations etc.

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

The following objectives are laid down for the study:

- 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.

## Method

A search engine is a website dedicated to find information on other websites. Before Google, there were hundreds of search engines on the internet but with the development of Google the number started decreasing to few hundreds. As certified by International Standard Organization there are only 230 search engines (**Promote3.com, 2016**). 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 queries 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 from 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 study. Thus the search engines undertaken for evaluation of study are:-

- a) Google (General)
- b) Bing (Specific)
- c) Yahoo! (Directory)
- d) Baidu (Country Specific General Search engine)

## III. 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<sup>th</sup>) and DDC (22<sup>nd</sup>) 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.

"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 6<sup>th</sup> term “*Plagiarism*” was taken for the study as the other keywords are already taken for other studies.

<b>S. No</b>	<b>Simple terms</b>	<b>Compound Terms</b>	<b>Complex Terms</b>
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	

*Table 1.0:* Keywords

### Changes in Search Engine Results

The Result meter of a search engine displays total number of documents it has retrieved for a given query but when the same query is entered again the search engines displays a different result because information on the web keeps on changing as documents are added, removed or modified. 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.

### Time Series Analysis

In forecasting there are numerous ways to project data based on prior collection. Time series analysis is a method of collecting data where data is collected in a series of time, a gap which may be based on daily, weekly, quarterly, monthly, yearly etc. The forecast is an estimate of a future event achieved by systematically combining and casting forward in predetermined way from the data about the past. Forecasts are possible only when a history of data exists. The study collected 100 days of time series 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 15<sup>th</sup> May, 2016 and ended on 18<sup>th</sup> of August, 2016 collecting 100 samples for keyword “*Plagiarism*” 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

Many methods are used in Time-series forecasting but for the study trend project method was chosen as it fits a trend line 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:-  $t_t = b_0 + b_1 t$

$b_0$  and  $b_1$  can be derived as:

$$b_0 = \bar{y} - b_1 \bar{t}$$
$$b_1 = \frac{n\sum ty_t - \sum t \sum y_t}{n\sum t^2 - (\sum t)^2}$$

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 1.1:- Time series data for forecasting of Select Search engines for the keyword “Plagiarism”**

Days (t)	Google			Bing			Yahoo!			Baidu		
	Result (Y <sub>t</sub> )	Multiplication of Days and Results (tY <sub>t</sub> )	Square of Days (t) <sup>2</sup>	Result (Y <sub>t</sub> )	Multiplication of Days and Results (tY <sub>t</sub> )	Square of Days (t) <sup>2</sup>	Result (Y <sub>t</sub> )	Multiplication of Days and Results (tY <sub>t</sub> )	Square of Days (t) <sup>2</sup>	Result (Y <sub>t</sub> )	Multiplication of Days and Results (tY <sub>t</sub> )	Square of Days (t) <sup>2</sup>
1	32900000	32900000	1	10900000	10900000	1	6140000	6140000	1	2220000	2220000	1
2	33100000	66200000	4	10600000	21200000	4	6100000	12200000	4	2160000	4320000	4
3	33100000	99300000	9	10500000	31500000	9	6040000	18120000	9	2160000	6480000	9
4	33000000	132000000	16	10500000	42000000	16	6030000	24120000	16	2230000	8920000	16
5	33100000	165500000	25	10300000	51500000	25	6080000	30400000	25	2230000	11150000	25
6	33400000	200400000	36	10200000	61200000	36	6170000	37020000	36	2180000	13080000	36
7	33300000	233100000	49	10200000	71400000	49	6200000	43400000	49	2180000	15260000	49
8	33400000	267200000	64	10100000	80800000	64	6200000	49600000	64	2180000	17440000	64
9	33500000	301500000	81	10000000	90000000	81	6160000	55440000	81	2180000	19620000	81
10	33900000	339000000	100	10300000	103000000	100	6180000	61800000	100	2170000	21700000	100
11	34000000	374000000	121	10400000	114400000	121	6220000	68420000	121	2170000	23870000	121
12	34700000	416400000	144	10400000	124800000	144	6560000	78720000	144	2160000	25920000	144
13	35000000	455000000	169	10400000	135200000	169	6720000	87360000	169	2160000	28080000	169
14	35600000	498400000	196	10400000	145600000	196	6960000	97440000	196	2160000	30240000	196
15	35400000	531000000	225	10500000	157500000	225	6850000	102750000	225	2160000	32400000	225
16	36400000	582400000	256	10500000	168000000	256	7410000	118560000	256	2180000	34880000	256
17	36400000	618800000	289	10500000	178500000	289	7410000	125970000	289	2180000	37060000	289
18	36300000	653400000	324	10300000	185400000	324	7220000	129960000	324	2230000	40140000	324
19	36400000	691600000	361	10300000	195700000	361	7230000	137370000	361	2180000	41420000	361
20	36400000	728000000	400	10500000	210000000	400	7410000	148200000	400	2180000	43600000	400
21	36300000	762300000	441	10600000	222600000	441	7560000	158760000	441	2230000	46830000	441
22	37000000	814000000	484	10600000	233200000	484	7640000	168080000	484	2270000	49940000	484
23	35000000	805000000	529	10400000	239200000	529	6720000	154560000	529	2160000	49680000	529
24	35600000	854400000	576	10400000	249600000	576	6960000	167040000	576	2160000	51840000	576
25	36800000	920000000	625	10600000	265000000	625	7290000	182250000	625	2280000	57000000	625
26	36000000	936000000	676	10500000	273000000	676	6700000	174200000	676	2260000	58760000	676
27	35900000	969300000	729	10600000	286200000	729	6540000	176580000	729	2260000	61020000	729
28	35300000	988400000	784	10700000	299600000	784	6390000	178920000	784	2280000	63840000	784
29	35200000	1020800000	841	10700000	310300000	841	6310000	182990000	841	2280000	66120000	841
30	35200000	1056000000	900	10700000	321000000	900	6190000	185700000	900	2200000	66000000	900
31	35000000	1085000000	961	10700000	331700000	961	5990000	185690000	961	2270000	70370000	961
32	36000000	1152000000	1024	10500000	336000000	1024	6700000	214400000	1024	2260000	72320000	1024
33	35900000	1184700000	1089	10600000	349800000	1089	6540000	215820000	1089	2260000	74580000	1089
34	35300000	1200200000	1156	10700000	363800000	1156	6390000	217260000	1156	2280000	77520000	1156
35	33000000	1155000000	1225	10500000	367500000	1225	5930000	207550000	1225	2190000	76650000	1225

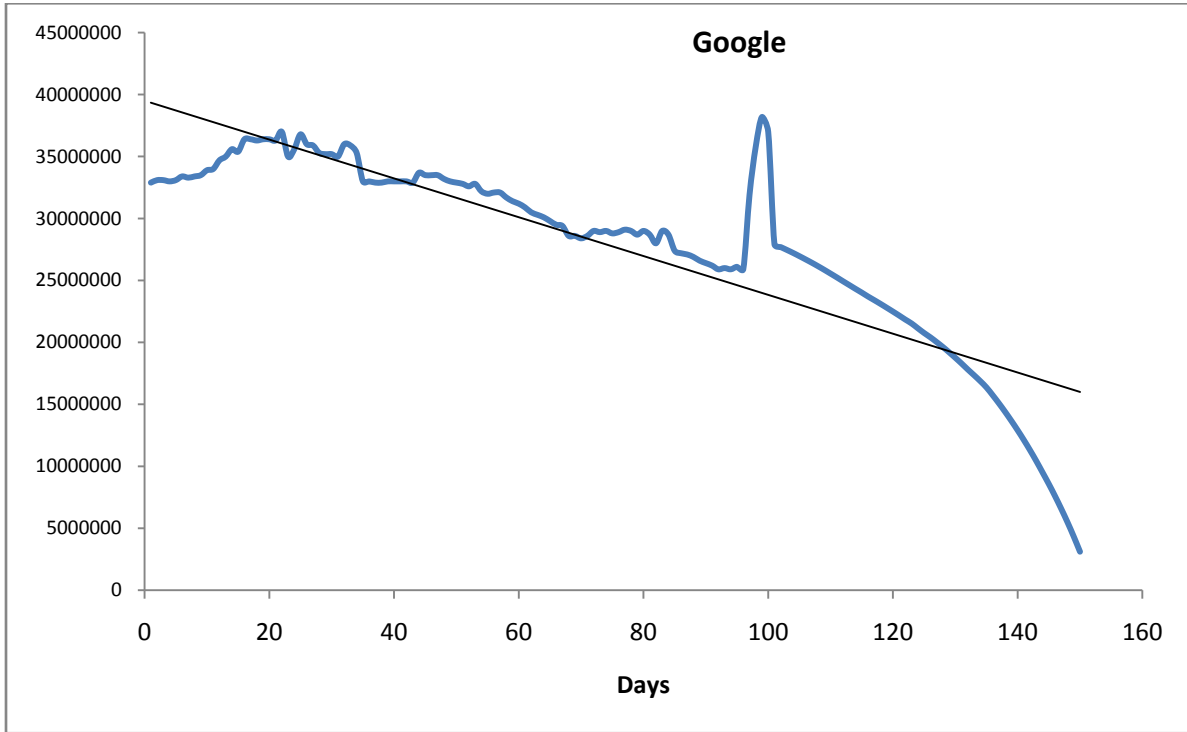
36	33000000	1188000000	1296	10500000	378000000	1296	5830000	209880000	1296	2190000	78840000	1296
37	32900000	1217300000	1369	10500000	388500000	1369	5830000	215710000	1369	2190000	81030000	1369
38	32900000	1250200000	1444	10500000	399000000	1444	5850000	222300000	1444	2130000	80940000	1444
39	33000000	1287000000	1521	10500000	409500000	1521	5860000	228540000	1521	2130000	83070000	1521
40	33000000	1320000000	1600	10600000	424000000	1600	5890000	235600000	1600	2130000	85200000	1600
41	33000000	1353000000	1681	10500000	430500000	1681	5930000	243130000	1681	2190000	89790000	1681
42	33000000	1386000000	1764	10500000	441000000	1764	5830000	244860000	1764	2190000	91980000	1764
43	32900000	1414700000	1849	10500000	451500000	1849	5830000	250690000	1849	2190000	94170000	1849
44	33700000	1482800000	1936	10500000	462000000	1936	5970000	262680000	1936	2220000	97680000	1936
45	33500000	1507500000	2025	10400000	468000000	2025	5960000	268200000	2025	2140000	96300000	2025
46	33500000	1541000000	2116	10400000	478400000	2116	5940000	273240000	2116	2140000	98440000	2116
47	33500000	1574500000	2209	10400000	488800000	2209	5960000	280120000	2209	2140000	100580000	2209
48	33200000	1593600000	2304	10500000	504000000	2304	5890000	282720000	2304	2150000	103200000	2304
49	33000000	1617000000	2401	10600000	519400000	2401	5870000	287630000	2401	2150000	105350000	2401
50	32900000	1645000000	2500	10600000	530000000	2500	5850000	292500000	2500	2090000	104500000	2500
51	32800000	1672800000	2601	10600000	540600000	2601	5880000	299880000	2601	2090000	106590000	2601
52	32600000	1695200000	2704	10800000	561600000	2704	6000000	312000000	2704	2100000	109200000	2704
53	32800000	1738400000	2809	10600000	561800000	2809	5880000	311640000	2809	2090000	110770000	2809
54	32200000	1738800000	2916	10900000	588600000	2916	5910000	319140000	2916	2050000	110700000	2916
55	32000000	1760000000	3025	10900000	599500000	3025	5920000	325600000	3025	2050000	112750000	3025
56	32100000	1797600000	3136	10900000	610400000	3136	6010000	336560000	3136	2030000	113680000	3136
57	32100000	1829700000	3249	10900000	621300000	3249	6090000	347130000	3249	2030000	115710000	3249
58	31700000	1838600000	3364	11100000	643800000	3364	6120000	354960000	3364	2030000	117740000	3364
59	31400000	1852600000	3481	10800000	637200000	3481	6060000	357540000	3481	2070000	122130000	3481
60	31200000	1872000000	3600	10800000	648000000	3600	6050000	363000000	3600	2070000	124200000	3600
61	30900000	1884900000	3721	10800000	658800000	3721	5980000	364780000	3721	2070000	126270000	3721
62	30500000	1891000000	3844	10800000	669600000	3844	5940000	368280000	3844	2200000	136400000	3844
63	30300000	1908900000	3969	10800000	680400000	3969	5990000	377370000	3969	2200000	138600000	3969
64	30100000	1926400000	4096	10900000	697600000	4096	4790000	306560000	4096	2120000	135680000	4096
65	29800000	1937000000	4225	10900000	708500000	4225	4790000	311350000	4225	2120000	137800000	4225
66	29500000	1947000000	4356	10900000	719400000	4356	4790000	316140000	4356	2190000	144540000	4356
67	29400000	1969800000	4489	10700000	716900000	4489	4790000	320930000	4489	2190000	146730000	4489
68	28600000	1944800000	4624	10700000	727600000	4624	4710000	320280000	4624	2130000	144840000	4624
69	28600000	1973400000	4761	10700000	738300000	4761	4710000	324990000	4761	2130000	146970000	4761
70	28400000	1988000000	4900	11100000	777000000	4900	4710000	329700000	4900	1940000	135800000	4900
71	28600000	2030600000	5041	11100000	788100000	5041	4810000	341510000	5041	1940000	137740000	5041
72	29000000	2088000000	5184	11100000	799200000	5184	5180000	372960000	5184	1980000	142560000	5184
73	28900000	2109700000	5329	11100000	810300000	5329	5110000	373030000	5329	1980000	144540000	5329
74	29000000	2146000000	5476	11100000	821400000	5476	5070000	375180000	5476	1980000	146520000	5476
75	28800000	2160000000	5625	11400000	855000000	5625	5050000	378750000	5625	1980000	148500000	5625
76	28900000	2196400000	5776	11600000	881600000	5776	5040000	383040000	5776	2080000	158080000	5776
77	29100000	2240700000	5929	11600000	893200000	5929	5180000	398860000	5929	2080000	160160000	5929
78	29000000	2262000000	6084	11500000	897000000	6084	5180000	404040000	6084	1990000	155220000	6084

79	28700000	2267300000	6241	11500000	908500000	6241	6650000	525350000	6241	1990000	157210000	6241
80	29000000	2320000000	6400	11500000	920000000	6400	6650000	532000000	6400	1990000	159200000	6400
81	28700000	2324700000	6561	11500000	931500000	6561	6410000	519210000	6561	1980000	160380000	6561
82	28000000	2296000000	6724	11400000	934800000	6724	4880000	400160000	6724	1950000	159900000	6724
83	29000000	2407000000	6889	11500000	954500000	6889	4880000	405040000	6889	1990000	165170000	6889
84	28700000	2410800000	7056	11500000	966000000	7056	6410000	538440000	7056	1980000	166320000	7056
85	27400000	2329000000	7225	11000000	935000000	7225	5140000	436900000	7225	1960000	166600000	7225
86	27200000	2339200000	7396	10700000	920200000	7396	5070000	436020000	7396	1930000	165980000	7396
87	27100000	2357700000	7569	10700000	930900000	7569	5070000	441090000	7569	1930000	167910000	7569
88	26900000	2367200000	7744	10700000	941600000	7744	4980000	438240000	7744	1920000	168960000	7744
89	26600000	2367400000	7921	10700000	952300000	7921	4980000	443220000	7921	1920000	170880000	7921
90	26400000	2376000000	8100	10800000	972000000	8100	4930000	443700000	8100	1770000	159300000	8100
91	26200000	2384200000	8281	10800000	982800000	8281	4910000	446810000	8281	1770000	161070000	8281
92	25900000	2382800000	8464	10900000	1002800000	8464	4920000	452640000	8464	1780000	163760000	8464
93	26000000	2418000000	8649	10900000	1013700000	8649	4930000	458490000	8649	1780000	165540000	8649
94	25900000	2434600000	8836	10900000	1024600000	8836	4920000	462480000	8836	1780000	167320000	8836
95	26100000	2479500000	9025	10900000	1035500000	9025	5060000	480700000	9025	1740000	165300000	9025
96	25900000	2486400000	9216	10900000	1046400000	9216	5290000	507840000	9216	2040000	195840000	9216
97	31900000	3094300000	9409	11100000	1076700000	9409	5290000	513130000	9409	2010000	194970000	9409
98	35800000	3508400000	9604	11200000	1097600000	9604	5290000	518420000	9604	2010000	196980000	9604
99	38200000	3781800000	9801	11300000	1118700000	9801	5290000	523710000	9801	2010000	198990000	9801
100	37000000	3700000000	10000	11200000	1120000000	10000	5290000	529000000	10000	2040000	204000000	10000
<b>Σt</b>	<b>Σ(Y<sub>t</sub>)</b>	<b>Σ tY<sub>t</sub></b>	<b>Σ(t)<sup>2</sup></b>	<b>Σ(Y<sub>t</sub>)</b>	<b>Σ tY<sub>t</sub></b>	<b>Σ(t)<sup>2</sup></b>	<b>Σ(Y<sub>t</sub>)</b>	<b>Σ tY<sub>t</sub></b>	<b>Σ(t)<sup>2</sup></b>	<b>Σ(Y<sub>t</sub>)</b>	<b>Σ tY<sub>t</sub></b>	<b>Σ(t)<sup>2</sup></b>
5050	3200700000	154899400000	338350	1074800000	55038500000	338350	588410000	28148380000	338350	209840000	10303340000	338350

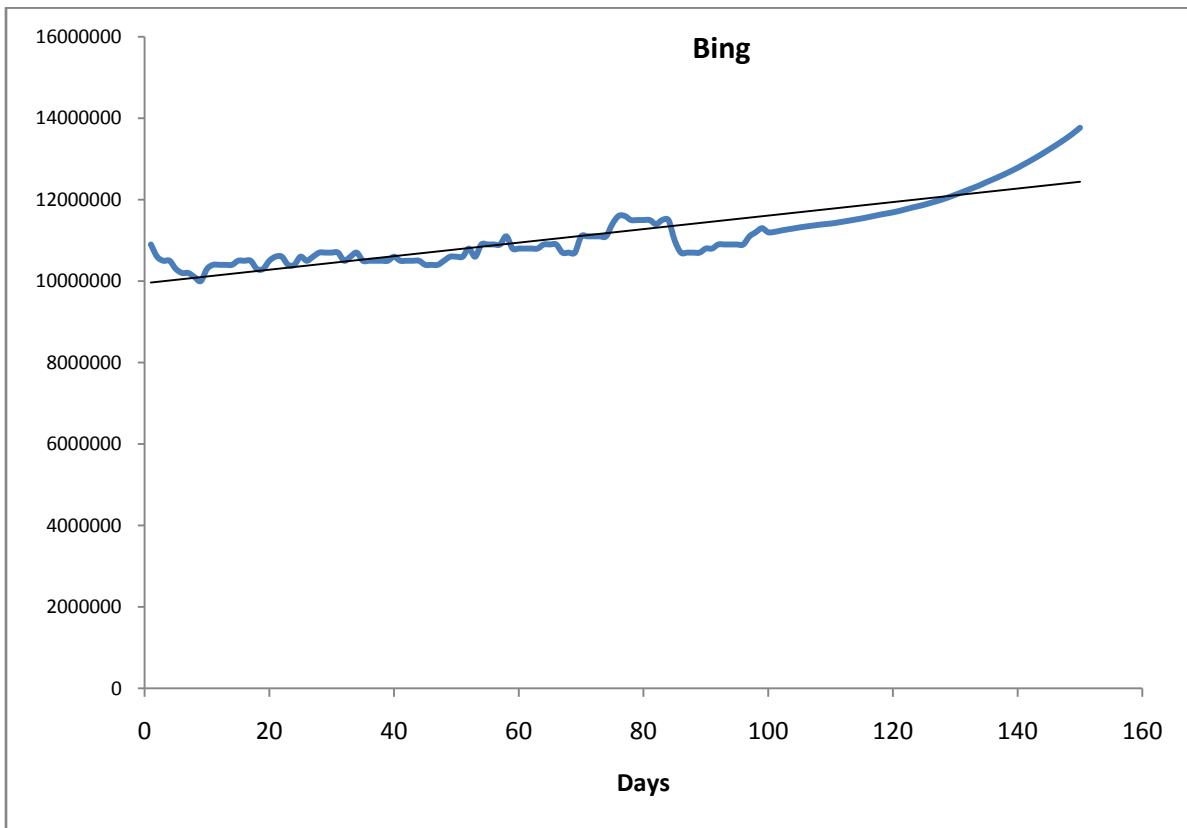
**Table 1.2:- Projected data using trend projection method for 50 days for the keyword “Plagiarism”**

<b>Days</b>	<b>Google</b>	<b>Bing</b>	<b>Yahoo!</b>	<b>Baidu</b>
<b>1</b>	27924606	11209273	4934812	1920473
<b>2</b>	27696988	11240356	4882907	1912306
<b>3</b>	27463725	11266712	4827933	1902619
<b>4</b>	27220509	11291913	4769199	1892546
<b>5</b>	26964546	11317973	4707531	1883683
<b>6</b>	26699698	11340204	4644174	1874597
<b>7</b>	26430602	11360362	4580127	1864070
<b>8</b>	26147815	11380578	4514073	1853158
<b>9</b>	25855514	11398290	4445222	1841843
<b>10</b>	25553466	11413111	4372390	1830106
<b>11</b>	25249429	11435293	4296883	1817662
<b>12</b>	24936113	11460456	4219157	1804724
<b>13</b>	24630051	11486192	4147561	1790990
<b>14</b>	24321996	11512517	4078129	1776680
<b>15</b>	24021400	11539448	4013718	1761765
<b>16</b>	23706499	11569970	3944828	1746216
<b>17</b>	23412535	11601479	3891295	1730604
<b>18</b>	23112382	11634018	3838084	1714362
<b>19</b>	22802747	11661326	3779298	1699034
<b>20</b>	22489487	11689020	3720648	1681587
<b>21</b>	22169592	11723644	3667796	1663419
<b>22</b>	21839581	11762740	3620601	1646161
<b>23</b>	21526009	11803408	3577472	1629656
<b>24</b>	21138586	11838819	3504399	1608877
<b>25</b>	20759544	11875170	3438141	1587167
<b>26</b>	20412576	11919644	3382995	1568767
<b>27</b>	20030210	11962309	3307091	1549100
<b>28</b>	19634424	12010175	3223635	1528787
<b>29</b>	19205635	12063797	3132188	1508553
<b>30</b>	18759625	12119954	3034596	1487703
<b>31</b>	18299212	12178791	2928770	1463107
<b>32</b>	17815895	12240458	2810778	1440245
<b>33</b>	17355986	12297118	2715769	1416220
<b>34</b>	16879259	12360087	2611807	1391353
<b>35</b>	16364113	12430007	2498380	1366429
<b>36</b>	15735724	12495027	2361579	1336911
<b>37</b>	15078424	12562752	2213982	1306094
<b>38</b>	14386259	12633327	2058958	1273903
<b>39</b>	13661271	12706908	1896934	1237645
<b>40</b>	12905962	12783660	1727103	1199551
<b>41</b>	12114366	12868245	1549919	1159511
<b>42</b>	11284328	12952341	1365475	1120133
<b>43</b>	10413541	13040207	1167000	1078877
<b>44</b>	9494874	13132061	957984	1035633
<b>45</b>	8567625	13228136	744378	991698
<b>46</b>	7585942	13323891	519144	941861
<b>47</b>	6555554	13423821	281034	889382
<b>48</b>	5473448	13528168	31124	834090
<b>49</b>	4321488	13642157	-235752	776303
<b>50</b>	3099306	13766709	-518406	715387

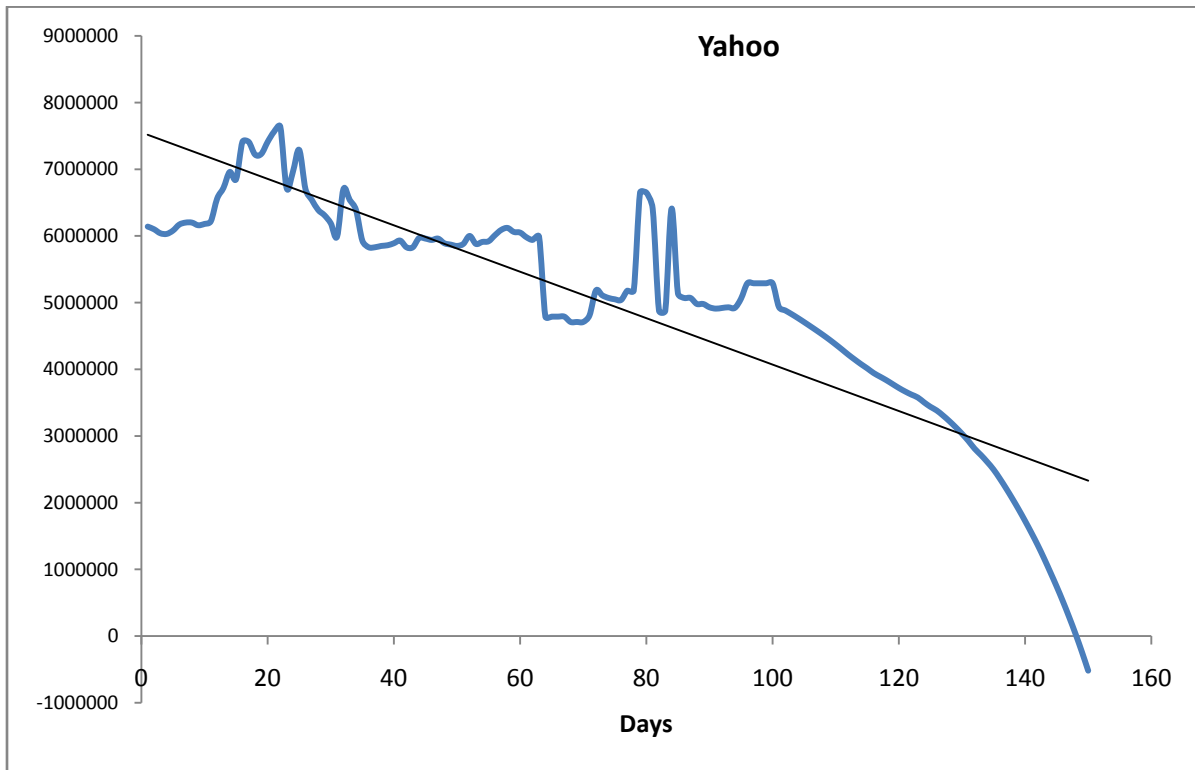




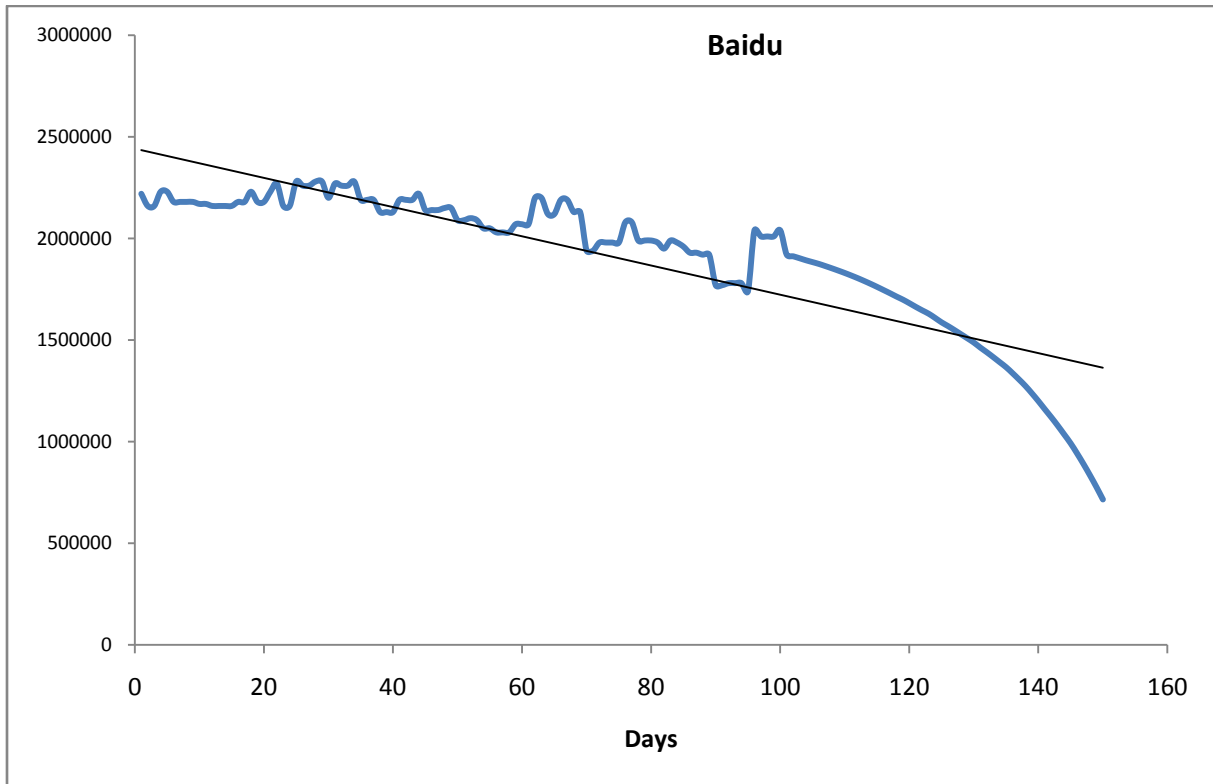
**Fig 1.3:- Negative Secular Trend of Google for the keyword "Plagiarism"**



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



**Fig 1.5:- Straight Secular Trend of Yahoo! for the keyword “Plagiarism”**



**Fig 1.6:- Positive Secular Trend of Baidu for the keyword “Plagiarism”**

#### IV. CONCLUSION

The forecasting of the select search engines reveal that Bing shows a positive forecasting towards its database growth.

The projected data reveal an upward trending securing a better growth in database. While Google, Baidu, and Yahoo! shows a negative secular trend in forecasting and the trending line show a downward projection resulting in loss of database, the results

show a downward trend as the results are consistently falling for the latter three search engines.

#### 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](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](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\\_BuildCustomize dSearch.htm](http://search.msn.com/docs/help.aspx?t=SEARCH_PROC_BuildCustomize dSearch.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](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>

#### AUTHORS

**First Author** – Peerzada Mohammad Iqbal, Professional Assistant, Faculty of Fisheries, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir (SKUAST-K)

Mail:- [iqlis2007@gmail.com](mailto:iqlis2007@gmail.com), +919906496608

**Second Author** – Dr. Abdul Majid Baba, Head, Department of Library and Information Science, & University Librarian The University of Kashmir, Mail:- [abmbaba@kashmiruniversity.ac.in](mailto:abmbaba@kashmiruniversity.ac.in), +919419014203

**Third Author** – Aasim Bashir, Assistant Professor, Department of Computer Science, The University of Kashmir, [aazim207@gmail.com](mailto:aazim207@gmail.com), +919906866673

