American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)
ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

http://asrjetsjournal.org/

Quantitative & Qualitative Evaluation of Three Search Engines (Google, Yahoo, and Bing)

Mohammad Asmaran*

Al-Balqa'Applied University , Al-Salt 19117, Jordan
Email: m_asmaran@yahoo.com

Abstract

As Internet evolution becomes bigger every day, people become more interested in finding services and information spread all over the web. This is a difficult process because of the neuromas number of sites released every day. This brings the importance of search engines to be very high to web users so they can surf the web easily. Search engines are required to provide appropriate and accurate results to users looking for any information. This research evaluates three of the top search engines over the internet (i.e. Google, Yahoo, and Bing) using well know qualitative and quantitative approaches. According to the research results, Google is the best of the three engines in term of quality of search results while Yahoo is the best in terms of search speed.

Keywords: Information Retrieval; Data Ranking; Search Engine Evaluation; Search Engines Comparison.

1. Introduction

Web search engines become very important as internet reaches to almost every detail of life. Users are always looking for services and information over the web. These needs lead to an extensive use of search engines looking for appropriate information. As the number of search engines increases day by day, an evaluation criteria should be exist to evaluate the quality of search results so that we can decide which engine is the best choice to use. This research has applied an evaluation methodology to evaluate three search engines Google, Yahoo, and Bing [1, 2, 3]. Evaluation involves representing results quantitatively. Quantitative representation of search results helps evaluating the quality of search results. While quality of search engines can be measured by two factors: Effectiveness and Efficiency. Effectiveness describes the quality of the results while efficiency represents how long it takes the user to obtain intended results from the search engine [4, 5, 6, 7, 8, 9, 10, 11].

---

* Corresponding author.

1.1. Theoretical Background

As described earlier in the introduction, there are two main factors to evaluate web search engine these factors are Effectiveness and Efficiency [4, 5, 6, 7, 8, 9, 10, 11].

a. Effectiveness: Indicates to the quality of search engine results. This can be achieved by applying one of the standard evaluation equations such as Harmonic Mean, Average Precision, or Discounted Cumulative Gain (DCG), etc. [4, 5, 6, 7, 8, 9, 10, 11].

Discounted Cumulative Gain (DCG) has been selected as an evaluation equation in this research. This metric is the best suited in this type of research since other factors measure according to two main factors called Recall and Precision. Recall and Precision are calculated based on the total number of relevant documents over the web, number retrieved documents, in addition to the number of relevant documents in the search results. These values can't be determined in case of web search to go further in the evaluation process. Though, Discounted Cumulative Gain (DCG) is best suitable for the web search engines since it evaluates only top twenty results based on the relevancy level that varies between 0 (Worst) and 5 (Best). Note that we can calculate DCG_p using Equation 1, note that p refers to the result position index (first, second, third, up to twentieth result) [4, 5, 6, 7, 8, 9, 10, 11].

$$DCG_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{log_2(i)} \qquad (1)$$

Where rel_i refers to the relevancy level ($0 =< rel_i <= 5$) of the i^{th} result position. So, if we apply this equation, there will be twenty DCG values for every search query. This indicates that as much better search result is obtained as much higher DCG value is determined. So, better search results would obtain higher DCG value [4, 5, 6, 7, 8, 9, 10 , 11].

Another factor is used to evaluate results called *normalized discounted cumulative gain (NDCG)* which is the division of DCG value over the ideal DCG value of the result position so that it indicates how much search engine results are optimal at a specific result position [4, 5, 6, 7, 8, 9, 10 , 11].

b. Efficiency: Indicates to the speed of search engine processing and response to the user [4, 5, 6, 7, 8, 9, 10, 11].

2. Materials and methods

In this research, statistical approach is used to evaluate selected engines (i. e. Google, Yahoo, Bing). This is done by trying 40 random search queries manually over the selected three search engines. Each query top 20 results are evaluated by the user to determine their relevancy level by assigning each of them a value that varies between 0 and 5 as described earlier in the theoretical background section. After the completion of the search results evaluation by the user, DCG value has been calculated for all the queries results across all search engines. Finally, average value for each search engine result position is calculated to do comparison. Search engines spiders are evaluated by setting fault and error pages to have 0 relevancy level. So, if search engine

crawl the web regularly in relatively short periods, number of fault pages would be less.

To determine Efficiency, search engine speed has to be determined. In Google search engine search time is shown next to the search results count. But in Yahoo and Bing search engines, this time is not shown at all. So, another method is used to determine search engine speed. This method depends on the actual experimental speed by determining load time of the first results web page. But the problem is that this time is affected by the internet speed of the client computer which is usually not stable in our experimental environment.

To overcome the problem of speed measurement and client instability of line speed, a web tool called Pingdom Website Speed Test [12] is used to do so. This tool determines several performance factors of any website. Two of them are used in this research to determine speed; these factors are Load Time in Seconds and Page Size in KB. So, Speed can be calculated by the following equation:

$$Speed = \frac{Page\ Size * 1024 * 8}{Load\ Time} \qquad (2)$$

The above equation calculates speed in bit per second so that it can be divided by 1024*1024 to be calculated in term of Mbps. So, the resultant equation is:

$$Speed = \frac{Page\ Size * 1024 * 8}{Load\ Time * 1024 * 1024} = \frac{Page\ Size}{Load\ Time * 128} \qquad (3)$$

Moreover, Pingdom performs site testing using several servers in different locations such as Amsterdam, New York City, or Dallas.

In order to get more accurate results, site is configured to do the test from fixed server that is located in Amsterdam, Netherlands.

3. Experimental Results

3.1 Effectiveness

As described earlier in previous section, 40 search queries are tried and their results relevancy are evaluated manually while DCG values are calculated as illustrated in Quality sheet in the associated excel file. Finally, average DCG & NDCG values are calculated as illustrated in Table 1.

Figure 1 shows the comparison chart of the DCG values of the three search engines while Figure 2 shows the comparison chart of the NDCG values of the three search engines. As illustrated in the figures, Google appears to have better quality of search results followed by Yahoo and Bing respectively.

3.2 Efficiency

As described earlier in the theoretical background section, 40 search queries are tested using Pingdom web tool

and the average speed results are illustrated in Table 2.

As seen in Table 2 the best efficiency is achieved by Yahoo followed by Bing and Google respectively. Although Google achieves best qualitative results, it obtains the worst efficiency result of the three search engines.

3.3 Quantitative Results

In addition to the effectiveness and efficiency, other factors are calculated to give some additional clear view of search results. In Table 3 probability of user first click is calculated according to the test results and these results are illustrated in Figure 3.

As we can see in the Figure 3, the best probability of the first position to be clicked by the user is achieved by Yahoo followed by Google and Bing respectively. This indicates that user is most likely to be satisfied by the first result position in Yahoo more than Google and Bing search engines.

Table 1: Average DCG & NDCG values of the search results.

Position	Summary $NDCG_p$			Summary $NDCG_p$		
	Google	Yahoo	Bing	Google	Yahoo	Bing
1	4.58	4.53	4.5	0.916	0.906	0.9
2	9.28	8.78	8.48	0.928	0.878	0.848
3	12.15	11.44	11.06	0.923628	0.869655	0.840767
4	14.43	13.63	13.05	0.921771	0.870668	0.833618
5	16.36	15.2	14.83	0.918687	0.853547	0.83277
6	18.12	16.74	16.19	0.917826	0.847926	0.820067
7	19.7	18.33	17.69	0.915286	0.851634	0.821899
8	21.15	19.68	18.99	0.912031	0.848642	0.818888
9	22.49	21.05	20.22	0.908051	0.84991	0.816398
10	23.83	22.22	21.39	0.907033	0.845752	0.81416
11	25.13	23.13	22.03	0.906638	0.834482	0.794796
12	26.25	24.18	22.8	0.901674	0.830571	0.783168
13	27.38	25.26	23.9	0.898775	0.829184	0.78454
14	28.44	26.2	24.83	0.894988	0.824497	0.781384
15	29.51	27.17	25.75	0.892708	0.82192	0.778964
16	30.56	28.12	26.66	0.890787	0.819664	0.777107
17	31.54	29.01	27.57	0.887701	0.816493	0.775964
18	32.53	29.93	28.49	0.885675	0.814886	0.77568
19	33.58	30.8	29.37	0.885873	0.812534	0.774809
20	34.47	31.63	30.11	0.882421	0.809718	0.770806

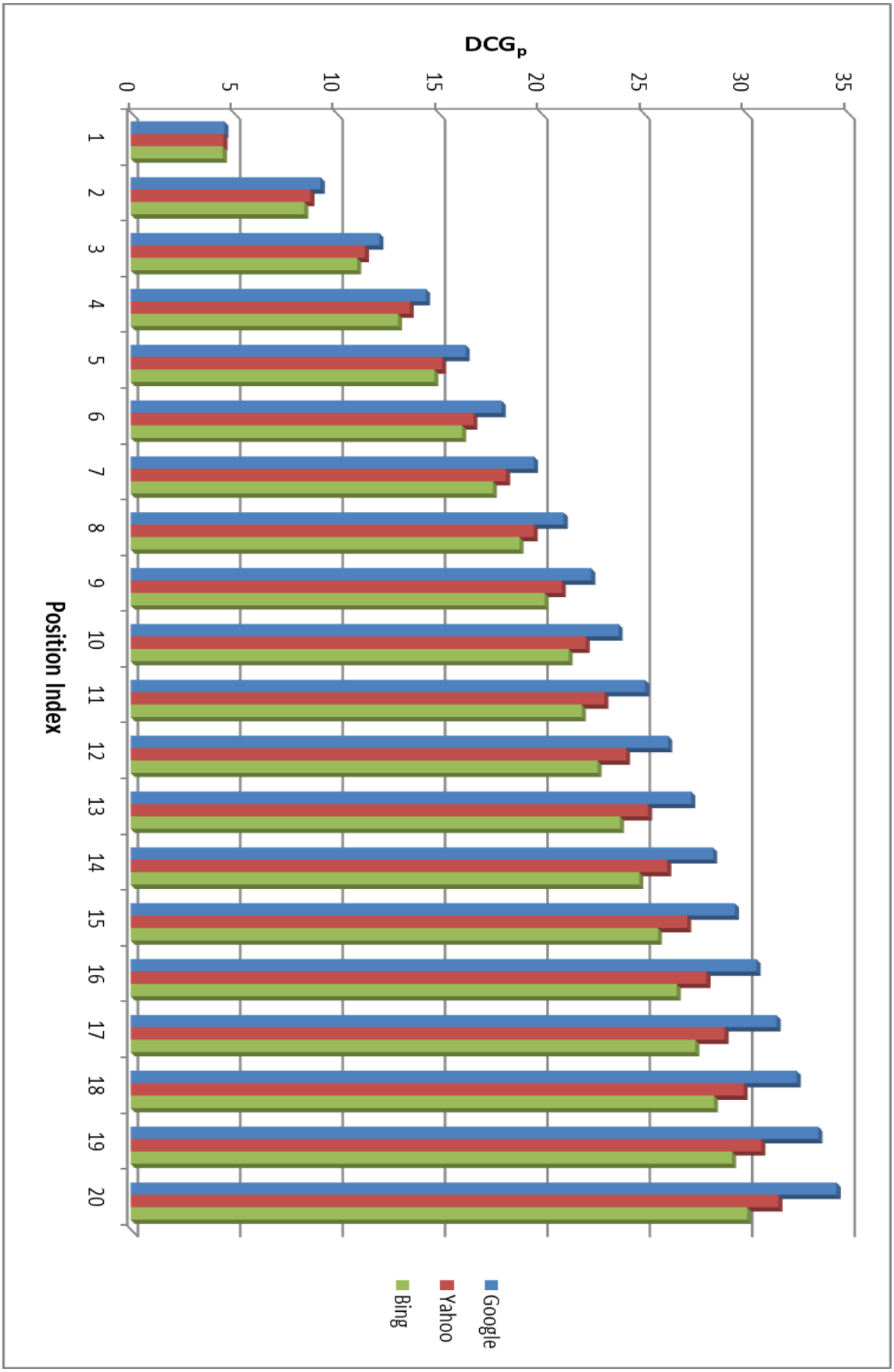

Figure 1: Average DCG_p versus Position index for all search engines.

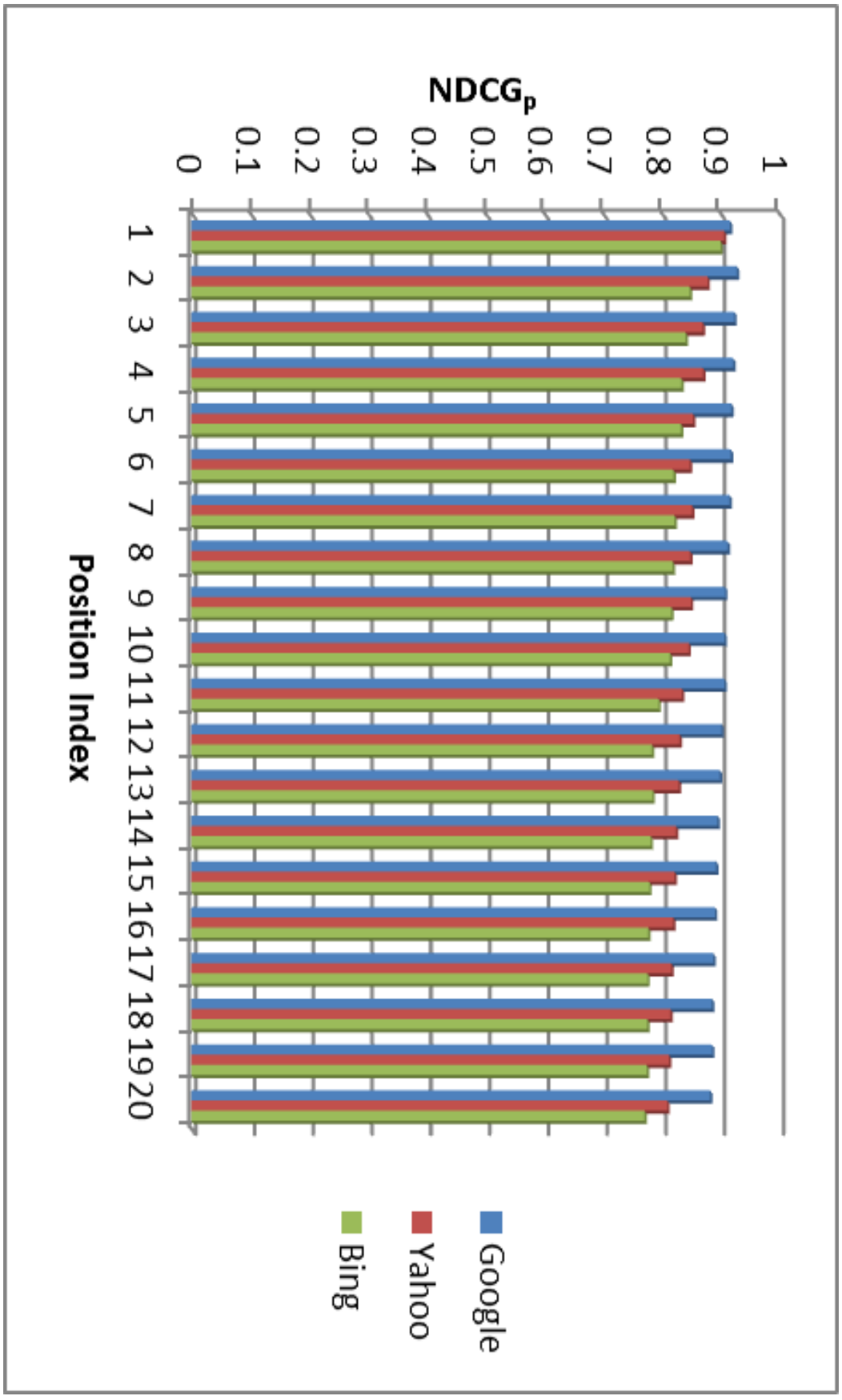

Figure 2: Average $NDCG_p$ versus Position index for all search engines.

Table 2: Speed Test of Queries Using Pingdom Web Testing Tool.

No	Google Load Time (Seconds)	Page Size (kB)	Speed (Mbps)	Yahoo Load Time (Seconds)	Page Size (kB)	Speed (Mbps)	Bing Load Time (Seconds)	Page Size (kB)	Speed (Mbps)
1	1.06	139.90	1.03	0.85	271.20	2.49	0.38	50.10	1.04
2	1.01	139.90	1.08	0.99	271.00	2.13	0.36	51.30	1.10
3	1.06	139.90	1.03	0.90	270.20	2.35	0.38	51.00	1.04
4	1.04	141.20	1.06	0.82	271.30	2.57	0.44	50.40	0.89
5	1.05	139.90	1.04	0.84	273.50	2.54	0.39	98.40	1.95
6	1.01	140.10	1.08	1.08	274.20	1.98	0.50	97.50	1.52
7	1.02	139.90	1.07	0.72	271.70	2.94	0.62	51.20	0.65
8	1.06	139.90	1.03	0.74	270.80	2.87	0.61	50.80	0.66
9	1.00	139.90	1.10	1.12	270.50	1.89	0.48	50.40	0.82
10	1.04	139.90	1.05	0.91	274.10	2.36	0.42	52.20	0.97
11	1.04	139.90	1.05	1.15	279.00	1.90	0.44	50.10	0.88
12	1.04	139.90	1.05	1.13	276.70	1.91	0.34	52.10	1.20
13	1.07	139.90	1.02	0.86	283.90	2.59	2.01	109.40	0.43
14	1.04	139.90	1.05	0.93	270.40	2.27	0.40	51.80	1.00
15	1.05	140.10	1.04	0.85	281.60	2.60	0.39	53.00	1.06
16	1.50	139.90	0.73	0.83	274.30	2.57	0.36	51.60	1.11
17	1.04	139.90	1.05	1.34	379.50	2.21	0.46	122.70	2.07
18	1.03	139.90	1.06	0.87	274.30	2.46	0.41	52.10	0.99
19	1.03	139.90	1.06	1.20	395.00	2.57	0.41	95.90	1.85
20	1.02	139.90	1.07	0.96	278.50	2.26	0.64	94.90	1.16
21	1.05	139.90	1.04	1.30	342.20	2.06	0.52	124.60	1.89
22	1.00	139.90	1.10	0.97	314.30	2.53	0.42	86.60	1.62
23	1.03	140.10	1.06	0.93	276.60	2.33	0.43	98.60	1.80
24	1.00	139.90	1.10	1.02	401.10	3.07	0.45	126.70	2.22
25	1.07	140.10	1.02	1.01	388.10	3.00	0.33	51.70	1.24
26	1.06	139.90	1.03	0.77	275.90	2.81	0.37	50.30	1.05
27	1.03	139.90	1.06	1.24	458.60	2.89	0.48	123.80	2.03
28	1.03	139.90	1.06	1.00	280.10	2.20	0.29	47.90	1.29
29	1.20	139.90	0.91	1.30	402.80	2.42	0.67	122.70	1.43
30	1.06	139.90	1.03	1.22	405.40	2.60	0.46	123.20	2.09
31	1.03	139.90	1.06	1.30	353.00	2.12	0.44	122.30	2.19
32	1.08	139.90	1.01	1.32	382.30	2.26	0.44	96.10	1.73
33	1.01	139.90	1.08	1.19	462.30	3.04	0.49	127.90	2.05
34	1.02	139.90	1.07	1.19	394.30	2.59	0.61	50.10	0.65

35	1.03	139.90	1.06	0.98	273.10	2.18	0.38	50.90	1.06
36	1.02	139.90	1.07	0.92	273.80	2.32	0.43	122.40	2.21
37	0.99	139.90	1.10	1.13	372.10	2.57	0.42	99.60	1.85
38	1.11	139.90	0.98	1.61	398.40	1.93	0.59	127.20	1.68
39	1.00	139.90	1.09	1.08	276.70	2.00	0.45	95.30	1.67
40	1.04	139.90	1.05	1.05	270.50	2.01	0.46	50.10	0.85
Avg:	**1.05**	**139.95**	**1.04**	**1.04**	**317.83**	**2.41**	**0.49**	**80.87**	**1.37**

Table 3: Probability of User First Click.

Probability of User First Click			
Position	**Google**	**Yahoo**	**Bing**
1	0.7	0.775	0.6
2	0.25	0.15	0.225
3	0.05	0.025	0.125
4	0	0.025	0.025
5	0	0.025	0

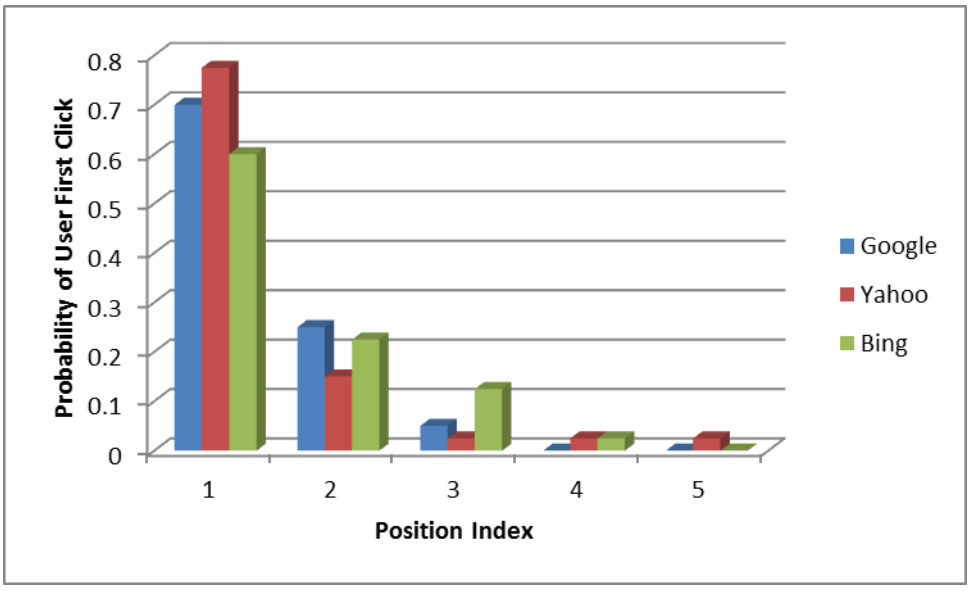

Figure 3: Probability of User First Click.

Another factor is calculated to show how much search results are retrieved for each search query by the search engines. This is illustrated in Figure 4 that shows logarithmic values of search results counts. As we can see from the figure, no clear view can be shown from these results, so probability of maximum search results count

is calculated for each search engines as illustrated in Figure 5. As we can see in the figure Yahoo search engine is the most likely search engine that would retrieve maximum search results.

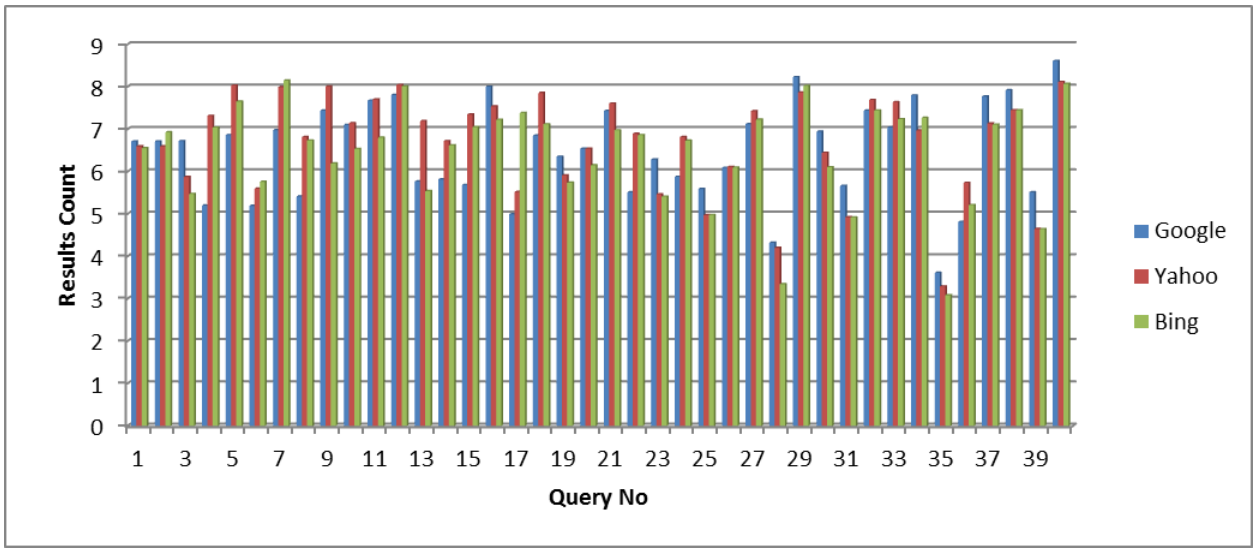

Figure 4: Search Results Counts.

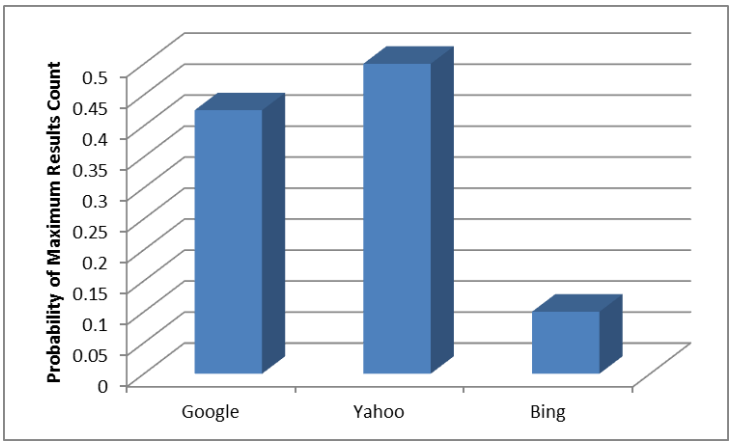

Figure 5: Probability of Maximum Search Results Count of Search Engines.

4. Results & Conclusions

In this research, statistical approach is used to evaluate three of the top search engines over the web (Google, Yahoo, and Bing). Evaluation process is carried using quantitative & qualitative approach to show that high quality search engines could perform in lower speed such as in Google. It's clear that quality of Google results are more qualified than other two search engines. Although it is slower than the others, it is the most used search engine over the web [13]. Moreover, this research shows that Yahoo search engine shows better performance and maximum search results count while it is not the best quality search engine to use. Although search engine spiders are evaluated by evaluating down and fault pages to be 0, in future work, search engines spiders should be evaluated by registering public website and specifying how much it takes the search engine to find it. Moreover, logical connectors of search statements are not evaluated in addition to semantic search. So, it is

recommended in future work to emphasis semantic search to evaluate the intelligence of search engines in more exact way.

References

[1] Google. "Google" Internet: http://www.google.com, [01-01-2016]

[2] Yahoo. "Yahoo Search" Internet: https://search.yahoo.com/, [01-01-2016]

[3] Bing. "Bing" Internet: http://www.bing.com/, [01-01-2016]

[4] Bruce Croft,Donald Metzler,Trevor Strohman,Search Engines: Information Retrieval in Practice, 2010

[5] Dawei Yin, Yuening Hu, Jiliang Tang, Tim Daly Jr., Mianwei Zhou, Hua Ouyang, Jianhui Chen, Changsung Kang, Hongbo Deng, Chikashi Nobata, Jean-marc Langlois, Yi Chang. "Ranking Relevance in Yahoo Search" ACM SIGKDD Conference On Knowledge Discovery And Data Mining (KDD 2016), 2016

[6] C. D. Manning, P. Raghavan, and H. Schütze. Introduction to Information Retrieval. New York, NY, USA, Cambridge University Press, 2008.

[7] K. Järvelin, J. Kekäläinen. "Cumulated gain-based evaluation of ir techniques". ACM Transactions on Information Systems, 20, pp. 422–446, Oct. 2002.

[8] S. Robertson and H. Zaragoza. "The probabilistic relevance framework: Bm25 and beyond". Foundations and TrendsR in Information Retrieval, 3, pp. 333–389, Apr. 2009.

[9] R. Srikant, S. Basu, N. Wang, and D. Pregibon. "User browsing models: relevance versus examination". ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 2010), 2010, pp. 223-232.

[10] T. Joachims. "Optimizing search engines using clickthrough data". ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 2002), 2002.

[11] S. Jiang, Y. Hu, C. Kang, T. Daly Jr., D. Yin, Y. Chang and C. Zhai. "Learning Query and Document Relevance from a Web-scale Click Graph". the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR 2016), 2016, pp. 185-194.

[12] Pingdom. "Website speed test" Internet: http://tools.pingdom.com/fpt/, [01-01-2016]

[13] comScore. “comScore Releases February 2016 U.S. Desktop Search Engine Rankings” Internet: https://www.comscore.com/Insights/Rankings/comScore-Releases-February-2016-US-Desktop-Search-Engine-Rankings, [16-03-2016]