International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 06, June 2016
www.ijates.com
ijates
ISSN 2348 - 7550

A SURVEY OF GOOGLE, YAHOO AND DOGPILE SEARCH ENGINE IN CONTEXT OF RELEVANT WEB PAGE AND PRECISION

Kamlesh Kumar Pandey[1], Pradeep Kumar Shukla[2], Rajat Kumar Yadu[3]

[1, 2] MCRPVV UTD-Amarkantak, (India)

[3]Department of Computer Science, Mahant Laxmi Narayan Das College Raipur (India)

ABSTRACT

In now day internet are most power full network and searching is a most powerful tool of internet. Internet or World Wide Web is given a space for storing all type of importance information like Text, Audio, Video, Image etc. If anybody, person, any internet user or organization can be wanted to any type of information, data, news or other things he searches to search engine . Role of search engine is only is getting on some query and search to web and given to list of result link.. Search engine given the importance link but some result link is related to search item and some web link is not related to search item or keyword. In present day basically many search engine are used in internet. In this paper is cover searching performance of three search engine this search engine is a Google, Yahoo and Doggpil. Main aim of this research study is judgment on which search engine are given to best result and how many link are related to searched item. In this research paper the searching item or queries are basically divided into three Group . These Group are a simple one word, simple multi word and complex multi word and we taken on each Group in three searching keyword or item or Query. This paper help to they peoples or person or any organization which are searching on any data or any information to all time because we which search engine is given to best relevant result.

Keyword:- *Google, Yahoo,Dogpile, Internet,Precision, Performance, Relevant web page, Search Engine.*

I. INTRODUCTION

web search engine is a Special web application program designed to help for finding on information like as like a files, songs, videos, images, weather information through own interfaces and these information are stored on the World Wide Web. In now day any Internet User want to any type of information then time he open to search engine and given to query to search engine and as soon as search engine given to result list. When search engine given to result list then time he search to query on own database and other search engine database. If search engine use to individual database then time he called to web search engine and when he used to other search engine then time he called to meta search engine. All search engine uses to special software web

crawlers. Web crawlers help to find out the publically available pages on the World Wide Web. crawlers Program seem to web pages follow the links from page to page and bring back the data back to own servers. This process is known as crawling. All result list of web addresses collected from the past crawls and sitemaps provided by the webmasters. In general search engine allow one to ask for comfortable meeting specific criteria and retrieve a catalog of references that match those criteria. Search engines are essentially huge full-text indexes of web pages. They use frequently updated indexes to operate rapidly and professionally. The quality of the indexes and how the engines utilize the data or information they contain is what make or break the excellence of search results. Search engine indexes are comparable but very much more multipart that back of book indexes. meaningful even a little bit concerning how they're build and used can vastly recover the searching skills. additional kinds of search engine are venture search engines, which search on intranets, personal search engines, which search individual personal computers and mobile search engines. a number of search engines also mine data available in newsgroups, huge databases or unlock directories like DMOZ.org. different Web directories which are maintain by human editors, search engines function to algorithmically. the bulk web sites which identify themselves search engines are actually front ends to search engines owned by other companies.

The National Information Standards society defines to boundary and access management is search engine. In this article paper cover to Performance factor analysis of Google, Yahoo and dogpile search engine. these factor are define which search engine have a best capability to given on relevant document then region we calculation on precision and relevant web page.

Google:- Google is product of American multinational technology corporation and it come to internet word in September 4, 1998. It interface searching provide for searching in web, video, image , News, Mail etc. It given to Internet related services and include online advertise technologies, searching, android operating system, Google Mail, Goole cloud computing, and other software. Google Search is written in Java and C++ Programming language and it web address is a www.google.com.

Yahoo:- this Yahoo is Web search Engine and product of American international technology business company and it launch on January 1994 in internet word. It interface searching provide for searching in in web, video, image , News and Answer. Yahoo Search product and service is related to internet and this service are a Yahoo News, yahoo Mail, yahoo Finance, yahoo Sports news , yahoo Search, Messenger, Answer, Flickr, online mapping, video sharing etc. Yahoo Search is written in PHP and Java Programming and it web address is a in.search.yahoo.com.

Dogpile:- Dogpile is Meta search engine and product of Blucora Corporation and launched on November 1995 it fetch on results from Google, Yahoo Search , Yandex and other popular search engine. It provides a various variety of searching in web, video, image, News , shopping and White pages. It product and service is a Category Links, Yellow Pages, White Pages , Web Search Box,Spelling Correction, Search Filter etc. Dogpile web address is a www.dogpile.com.

II. METHODOLOGY

In this paper we selected to most popular search engine Google, Yahoo and dogpile. We taken to Six queries or searching keyword and these keyword is to submitted to yahoo and dogpile search engine. When Google, yahoo and dogpile search engine are given to dynamic result link list then we check to all link and find out or counting total number of relevant link, irrelevant link, not accessing link and after that we stored this information on precision table and relevant web page table. using on these tables we find out precision and relevant web page for these search engine. We taken on help to only a few Research paper, book or other publication for concept and formula for calculation on Precision and relevant web page. We collected on data using on Google, Yahoo and dogpile search engine web sites. Performance evaluation of Google, Yahoo and dogpile are examined to during March 2016 to April 2016 . In this study Google, Yahoo and dogpile are given to own search results and this search result are categorized as five points. These points are first is more relevant, second is less relevant, third is irrelevant, four is links and five point is sites can't be accessed on the basis of the follow criterion and this criterion and point are given in the Chu & Rosenthal 1996[1], Leighton 1996[2], Ding & Marchionini 1996[3], Clarke & Willett 1997[4], B.T. Sampath kumar[6]. We define new criteria on the bases on old criteria and calculation on precision and relevant web page of present time search. These new criteria are identify the above five point and this criteria are.

- If the result of search engine, web page content is closely matched to subject matter of the search item or keyword then this kind of web document is classify to more relevant and we given to 2 number/score.
- If the result of search engine, web page content is not closely matched to subject topic but consists of some related aspects to the subject topic of the search item or keyword then this web page or document is categorize to less relevant and we given to 1 number/score.
- If the result of search engine , web page content is not related to the subject topic of the search item or keyword then this web document or page is categorize as irrelevant and we given to 0 number/score.
- If the result of search engine , web page content is consisted of a complete series of links but some information is required then this web page or document is categorize as link and we set to 0.5 number/score.
- If the result of search engine , web site or web document is can't be accessed or can't be open for a particular URL then its web page is classify to site can't be accessed and we given 0 number/score.

III. RELEVANT SITES ANALYSIS FOR PAGE BY PAGE OF GOOGLE, YAHOO AND DOGPILE SEARCH ENGINE:

In this section we examine to total relevant site of Google, Yahoo and Dogpile search engine. Relevant site or relevant result is totally closely or parsaliy related to searching keyword or query. This is a base for calculation of precision and Relevant Web Page.

A. Relevant sites Analysis Google search engine :- When we search to different 9 query using Google Search engine then it given a total numbers of **13,99,36,00,000 result** sites. we selected to top 900 result

link and selected to top 100 result to each query. We find to total 610 site for relevant sites out of 900 result list. Following Table 1 are shows the calculation of relevant web sites for Google search engine.

B. Relevant sites Analysis Yahoo search engine :- When we search to different 9 query using Yahoo Search engine then it given a total numbers of **2,296,400,000 result** sites. we selected to top 900 result link and selected to top 100 result to each query. We find to total 655 site for relevant sites out of 900 result list. Following Table 2 are shows the calculation of relevant web sites for Yahoo search engine.

***C.* Relevant sites Analysis Dogpile search engine** :- When we search to different 9 query using Bing Search engine then it given a total numbers of **4660 result** sites. We selected to top 900 result link and selected to top 100 result to each query. We find to total 767 sites for relevant sites out of 900 result list. Following Table 3 are shows the calculation of relevant web sites.

Table 1 Relevant sites Analysis Google search engine

Search keyword	Total number of sites	Selected sites	WP 1	WP 2	WP 3	WP 4	WP 5	WP 6	WP 7	WP 8	WP 9	WP 10	Total RS
Simple one word queries													
Computer	**2,41,00,00,000**	**100**	10	10	10	9	9	**7**	7	5	4	4	**75**
Information	**8,37,00,00,000**	**100**	10	10	10	9	9	**8**	8	5	4	4	**77**
Process	**1,93,00,00,000**	**100**	10	10	9	8	7	**6**	5	5	4	2	**66**
Simple multi word queries													
Algorithm Analysis	**11,50,00,000**	**100**	10	10	9	8	9	**8**	7	7	4	5	**77**
Computer science	**50,70,00,000**	**100**	10	10	9	8	8	**7**	7	6	4	3	**72**
Software Engineer	**7,92,00,000**	**100**	10	10	9	8	8	**7**	7	5	5	2	**71**
Complex multi word queries													
Computer science & Application	**24,70,00,000**	**100**	9	9	8	5	7	**6**	6	6	4	1	**61**
Evaluation of Computer Generation	**28,70,00,000**	**100**	9	9	6	7	7	**6**	5	3	2	2	**56**
Introduction to Web and Software Development	**4,84,00,000**	**100**	9	8	8	7	6	**6**	4	3	4	1	**56**

Table 2 Relevant sites Analysis Yahoo search engine

Search keyword	Total number of sites	Selected sites	WP 1	WP 2	WP 3	WP 4	WP 5	WP 6	WP 7	WP 8	WP 9	WP 10	Total RS
Simple one word queries													
Computer	**427,000,000**	**100**	10	9	9	8	7	**7**	6	2	2	2	**62**
Information	**1,060,000,000**	**100**	10	10	9	9	8	**7**	7	5	4	4	**73**
Process	**489,00,000**	**100**	9	9	8	8	7	**6**	5	5	2	2	**61**
Simple multi word queries													
Algorithm Analysis	**31,200,000**	**100**	10	9	9	8	9	**8**	7	7	3	5	**75**
Computer science	**296,000,000**	**100**	10	10	9	8	8	**7**	7	6	4	2	**71**
Software Engineer	**91,400,000**	**100**	10	9	9	8	8	**7**	7	5	5	2	**70**
Complex multi word queries													
Computer science & Application	**228,000,000**	**100**	9	9	8	8	7	**7**	6	6	4	2	**66**
Evaluation of Computer Generation	**26,700,000**	**100**	9	9	7	7	8	**6**	5	4	2	2	**59**
Introduction to Web and Software Development	**87,200,000**	**100**	9	8	8	7	7	**6**	4	4	4	1	**58**

Table 3- Relevant sites Analysis of Dogpile search engine

Search keyword	Total number of sites	Selected sites	WP 1	WP 2	WP 3	WP 4	WP 5	WP 6	WP 7	WP 8	WP 9	WP 10	Total RS
Simple one word queries													
Computer	**610**	**100**	10	10	10	10	9	9	8	8	7	7	**88**
Information	**600**	**100**	10	10	9	9	10	8	8	8	7	7	**86**
Process	**610**	**100**	10	9	10	9	10	8	8	7	7	7	**85**
Simple multi word queries													
Algorithm Analysis	**500**	**100**	10	10	10	8	9	9	8	7	6	6	**83**
Computer science	**510**	**100**	10	10	10	9	9	8	8	7	7	6	**84**
Software Engineer	**490**	**100**	10	10	9	8	9	9	8	7	6	6	**82**

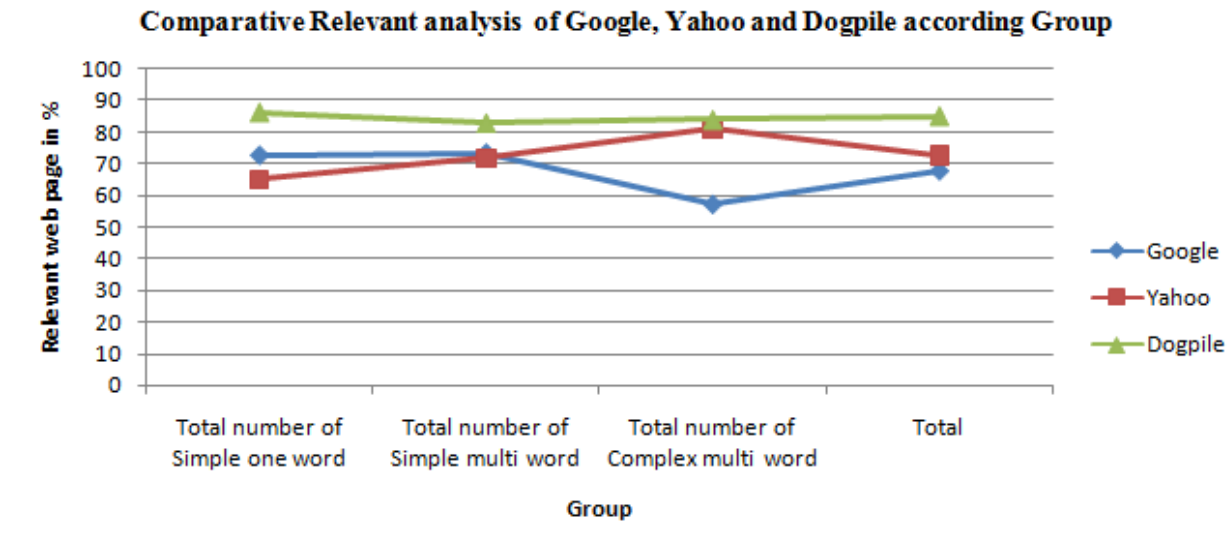
Complex multi word queries													
Computer science & Application	**450**	**100**	10	10	10	8	8	9	9	7	7	7	**85**
Evaluation of Computer Generation	**480**	**100**	10	9	9	9	8	9	7	8	7	7	**83**
Introduction to Web and Software Development	**410**	**100**	10	10	10	9	9	7	8	7	7	7	**84**

After that we calculate to relevant web page analysis according to group. In this research we taken to three groups and each group have a taken to three query. This group is a simple, multi and complex group. We listed to relevant record analysis on according to group in Table 4 and show to comparative relevant analysis of according to group in graph figure 1.

Table 4- Comparative Relevant Table Google, Yahoo and Dogpile according Group

Search Engine	Total number of Simple one word	Total number of Simple multi word	Total number of Complex multi word	Total
Google	218(72.66)	220(73.33)	172(57.33)	610(67.77%)
Yahoo	196(65.33 %)	216(72%)	243(81 %)	655(72.77%)
Dogpile	259(86.33 %)	249(83%)	252(84 %)	767(85.22 %)

Figure 1- Comparative Relevant analysis of Google, Yahoo and Dogpile according Group



IV. PRECISION EVALUATION OF GOOGLE, YAHOO AND DOGPILE SEARCH ENGINE:-

Precision is a ratio of total relevant site retrieve by search engine and total number of site retrieve by search engine including all type of result. In this section we calculation on precision of search engines for each of the search keyword using this formula and five criteria (Eq. 1).

Precision=Sum of the scores of web result s retrieved by a Search Engine / Total number of web result retrieved (Eq. 1)

A. **Precision of Google:- :-** Following Table 5 are shows the total number of more relevant web result sites, less relevant web result sites, irrelevant web result sites, links and sites cannot be accessed of Google search engine in selection of 900 web result or sites. Clear for this table is 38.77% of sites are more relevant, 29% of sites are less relevant, 28.22% of sites are irrelevant, 2.44% of sites are given to link and 1.55% of sites are site cannot be access. Total precision of Bing is a 1.09.

B. **Precision of Yahoo:- :-** Following Table 6 are shows the total number of more relevant web result sites, less relevant web result sites, irrelevant web result sites, links and sites cannot be accessed of Yahoo search engine in selection of 900 sites. Clear for this table is 39.66% of sites are more relevant, 26.44% of sites are less relevant, 29.44% of sites are irrelevant, 2.66% of sites are given to link and 1.77% of sites are site cannot be access. Total precision of Bing is a 1.07.

C. **Precision of Dogpile:-** Following Table 7 are shows the total number of more relevant web result sites, less relevant web result sites, irrelevant web result sites, links and sites cannot be accessed of Dogpile search engine in selection of 900 sites. Clear for this table is 57.77% of sites are more relevant, 26.66% of sites are less relevant, 18.44% of sites are irrelevant, 1.88% of sites are given to link and 0.77% of sites are site cannot be access. Total precision of dogpile is a 1.43

Table 5-Precision Table of Google

Search keyword	Total number of sites	Selected sites	More relevant sites	Less relevant sites	Irrelevant sites	links	Sites cannot be accessed	Repeated link	precision
Simple one word queries									
Computer	**2,41,00,000,000**	**100**	**51**	24	22	2	1	**3**	**1.27**
Information	**8,37,00,000,000**	**100**	**45**	32	20	1	2	**2**	**1.22**
Process	**1,93,00,000,000**	**100**	**30**	36	29	3	2	**2**	**0.97**
Simple multi word queries									
Algorithm Analysis	**11,50,000,000**	**100**	**54**	23	20	1	2	**4**	**1.31**
Computer	**50,70,0**	**100**	**43**	29	25	2	1	**4**	**1.16**

science	**0,000**								
Software Engineer	**7,92,00,000**	**100**	**42**	29	25	2	2	**3**	**1.14**
Complex multi word queries									
Computer science & Application	**24,70,00,000**	**100**	**38**	23	34	4	2	**4**	1.01
Evaluation of Computer Generation	**28,70,00,000**	**100**	**22**	34	40	3	1	**4**	0.79
Introduction to Web and Software Development	**4,84,00,000**	**100**	**24**	31	39	4	1	**4**	0.81
Total	**13,99,36,00,000**	**900**	**349 (38.77)**	**261 (29%)**	**254 (28.22%)**	**22 (2.44%)**	**14 (1.55%)**	**30 (3.33%)**	**1.09**

Table 6-Precision Table of Yahoo

Search keyword	**Total number of sites**	**Selected sites**	**More relevant sites**	**Less relevant sites**	**Irrelevant sites**	**links**	**Sites cannot be accessed**	**Repeated link**	**precision**
Simple one word queries									
Data	**427,000,000**	**100**	**39**	23	33	3	2	**3**	**1.02**
information	**1,060,000,000**	**100**	**41**	32	22	2	3	**2**	**1.15**
process	**489,00,000**	**100**	**27**	34	34	3	2	**2**	**0.89**
Simple multi word queries									
Computer fundamental	**31,200,000**	**100**	**52**	23	21	2	2	**4**	**1.28**
Computer science	**296,000,000**	**100**	**50**	21	25	3	1	**4**	**1.22**
Web Development	**91,400,000**	**100**	**41**	29	26	2	2	**3**	**1.12**
Complex multi word queries									
Computer science & engineering	**228,000,000**	**100**	**45**	21	29	3	2	**3**	1.12
Evaluation of digi	**26,700,0**	**100**	**28**	31	37	3	1	**2**	0.88

tal library	**00**								
Introduction to Design and Analysis	**87,200,000**	**100**	**34**	24	38	3	1	**2**	0.93
Total	**2,296,400,000**	**900**	**357 (39.66%)**	**238 (26.44%)**	**265 (29.44%)**	**24 (2.66%)**	**16 (1.77 %)**	**24 (2.66%)**	**1.07**

Table 7-Precision Table of Dogpile

Search keyword	**Total number of sites**	**Selected sites**	**More relevant sites**	**Less relevant sites**	**Irrelevant sites**	**links**	**Sites cannot be accessed**	**Repeated link**	**precision**
Simple one word queries									
Data	**610**	**100**	**65**	23	8	2	2	**2**	1.54
information	**600**	**100**	**61**	25	9	3	2	**3**	1.48
process	**610**	**100**	**56**	29	12	2	1	**2**	1.42
Simple multi word queries									
Computer fundamental	**500**	**100**	**45**	38	14	2	1	**1**	1.29
Computer science	**510**	**100**	**60**	24	14	2	0	**1**	1.45
Web Development	**490**	**100**	**59**	23	17	1	0	**0**	1.41
Complex multi word queries									
Computer science & engineering	**450**	**100**	**59**	26	14	1	0	**0**	1.44
Evaluation of digital library	**480**	**100**	**58**	25	15	2	0	**0**	1.42
Introduction to Design and Analysis	**410**	**100**	**57**	27	13	2	1	**1**	1.42
Total	**4660**	**900**	**520 (57.77%)**	**240 (26.66%)**	**116 (18.44%)**	**17 (1.88%)**	**7 (0.77%)**	**10 (1.11%)**	**1.43**

We try to show comparative Performance analysis of Google, Yahoo and Dogpile Search engine in Graph Figure 2 Based on Performance criteria. In graph figure 3 are show to Comparative precision analysis of Google, Yahoo and Dogpile based on Searching Keyword

Figure 2- Comparative Performance Analysis of Google,Yahoo and Dogpile

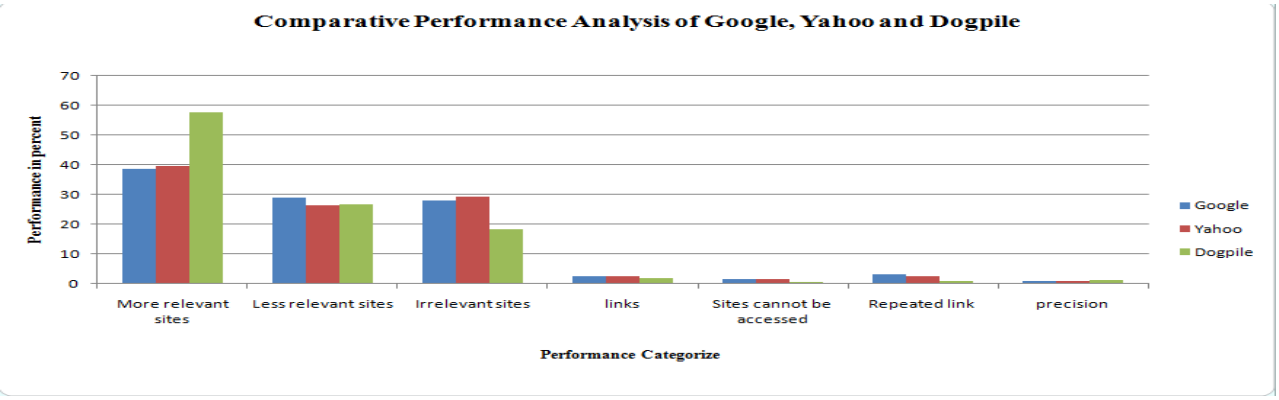
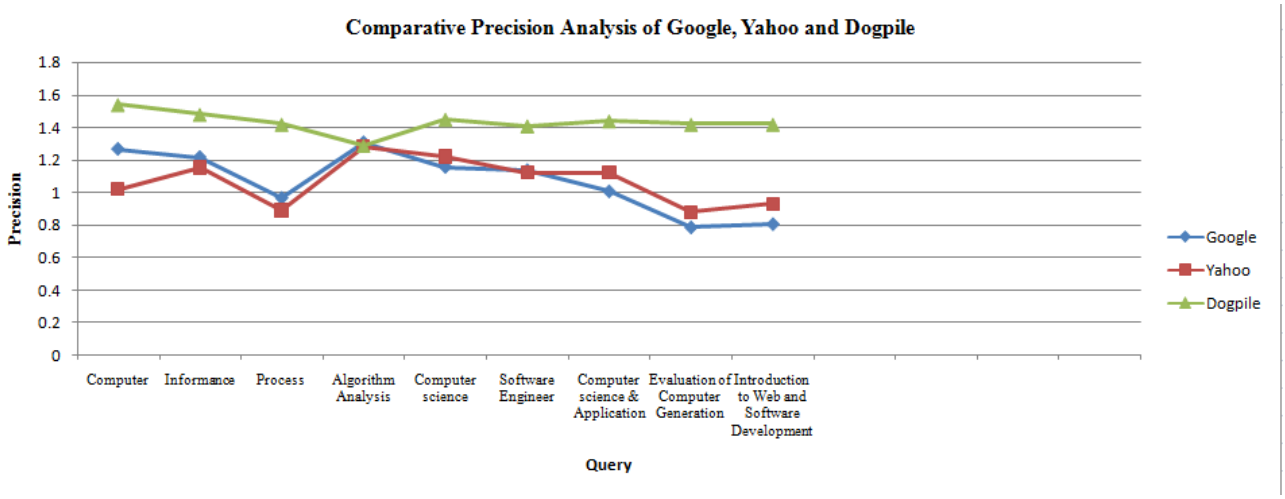


Figure 3- Comparative Precision Analysis of Google, Yahoo and Dogpile

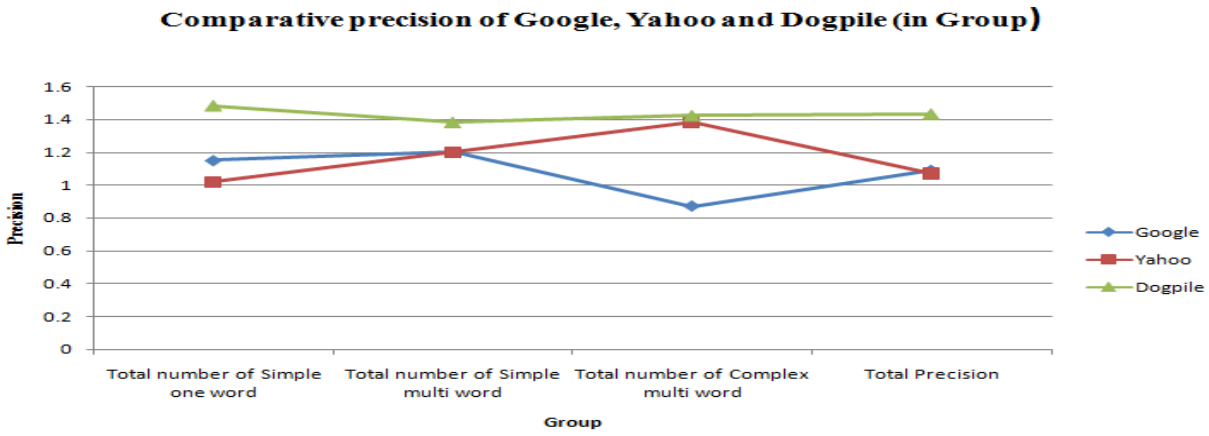


After that we calculate to precision analysis of according to group. In this research we taken to three groups and each group have a taken to three query. This group is a simple, multi and complex group. We listed to precision analysis on according to group in Table 8 and show to comparative precision analysis of according to group in graph figure 4.

Table 8-Comparative precision Table Google, Yahoo and Dogpile

Search Engine	Total number of Simple one word	Total number of Simple multi word	Total number of Complex multi word	Total Precision
Google	1.15	1.20	0.87	1.09
Yahoo	1.02	1.20	1.38	1.07
Dogpile	1.48	1.38	1.42	1.43

Figure 4- Comparative Precision of Google, Yahoo and Dogpile according Group



V. CONCLUSION

In this paper we presented the overview and performance analysis of Google, Yahoo and Dogpile search engine. Yahoo and Google is a web search engine and Dogpile is a meta search engine. The result of study showed that the precision and relevant page of Dogpile is higher the Google and Yahoo because Dogpile is given to less number of web search result and filtration technique is used. When we see to precision on according to group then Dogpile is given to hight precision in simple word group, multi word group and complex group and google is given the less precision and relevant page in a complex Group. Finally conclusion is this research is Google and Dogpile is given a best result in simple and multi group word but Dogpile and Yahoo is given to best result in a complex group word. Overall result is dogpile is best for all type of searching but you want to depth search then yahoo and Google is best because it given to a lot of result list.

REFERENCES

[1] Chu, H., & Rosenthal, M. (1996), "Search engines for the World Wide Web: A Comparative study and evaluation methodology" Proceedings of the ASIS 1996 Annual Conference, vol- 33, pp- 127-35.

[2] Leighton, H. (1996)," Performance of four WWW index services, Lycos, Infoseek, Webcrawler and WWW Worm" Retrieved from http://www.winona.edu/library/webind.htm

[3] Ding, W., & Marchionini, G. (1996)."A Comparative study of the Web search service performance" Proceedings of the ASIS 1996 Annual Conference, vol 33, pp- 136-142

[4] Clarke, S., & Willett, P. (1997) " Estimating the recall performance of search engines", ASLIB Proceedings, vol 49 (7), pp- 184-189.

[5] Shafi, S. M., & Rather, R. A. (2005)," Precision and recall of five search engines for retrieval of scholarly information in the field of biotechnology" Webology, 2 (2), Retrieved from http://www.webology.ir/2005/v2n2/a12.html

[6] B.T. Sampath kumar and S.M.Pavithra "Evaluating the searching capabilities of search engine and metasearch engine:a comparative study" annals of library and information studies vol.57, june 2010, pp-87-97

[7] Kamlesh Kumar Pandey, "Internet Search Engine: A Comparative and Performance Evaluation of Web Search Engine and Meta Search Engine", published in the Proc. ICERCSE 2015 21 November 2015, Published by : UICSA, Rani Durgavati University, Jabalpur, pp 138 143,ISSN No: 2393-9931.

[8] Kamlesh Kumar Pandey "A ANALYSIS OF GOOGLE, YAHOO AND BING WEB SEARCH ENGINE IN PERFORMANCE PARAMETER OF PRECISION AND RELEVANT WEB PAGE", published in the Proc. ICRISEM 2016 27 November 2016, Published by : A R Research publication, Rani Durgavati University, Jabalpur, pp 880-891 ,ISBN No: 978-81-932074-1-3.

[9] Biradar B S, and Sampath Kumar B T, "Internet search engines:a comparative study and evaluation methodology", SRELS Journal of Information Management, volume 43(3) ,2006 pp-231-241

[10] Hossein Jadidoleslamy," INTRODUCTION TO METASEARCH ENGINES AND RESULT MERGING STRATEGIES: A SURVEY" International Journal of Advances in Engineering & Technology, Nov 2011, ISSN: 2231-1963, pp-30-40

[11] K.Srinivas, P.V.S.Srinivas and A.Govardhan. "A Survey on the "Performance Evaluation of Various Meta Search Engines" IJCSI Volume 8, Issue 3, No. 2, May 2011 Pages 359-364.

[12] Lu Y., Meng W., Shu L., Yu C. and Liu K."Evaluation of result merging strategies for metasearch engines" 6th International Conference on Web Information Systems Engineering (WISE Conference); New York;2005.

[13] Yiyao Lu, Weiyi Meng, Liangcai Shu, Clement T. Yu, and King-Lup Liu" Evaluation of result merging strategies for

[14] http://en.wikipedia.org/wiki/search_engine

[15] http://www.searchengineguide.com

[16] https://www.yahoo.com

[17] https://www.dogpile.com

[18] https://www.Google.com