International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 04, April 2016
www.ijates.com
ijates
ISSN 2348 - 7550

A COMPARATIVE PERFORMANCE ANALYSIS OF GOOGLE INDEX WEB SEARCH ENGINE AND YAHOO DIRECTORIES WEB SEARCH ENGINE

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

[1, 3] *MCRPVV UTD-Amarkantak, (India)*

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

ABSTRACT

A search engine are planned to search information and data on the World Wide Web or FTP server. The Search Engine is given the helps in the process of retrieving the information, data or other importance thing as required to user or person. The main goal of this paper is to analysis the efficiency of index and directory search engines and determine the best one search engine . Now day search engine is a most popular application of internet and it used by everyone because all importance subject related information, data, news is given by WWW. Search engine are given the result as a format of URL and URLs are consist of web pages, data, images, video, audio, information and other type of file. Result of Search engine is given some result web link is related to search item and some result link is not related for search item. In this research papers are cover to evaluating the performance of index web search engine Google and directory search engine Yahoo on their searching result. This paper compares the retrieval effectiveness of the Google and Yahoo. Precision and relative recall of Web search engine was considered for evaluating the Performance. We get the performance in based on General Queries were tested. We divided on searching keyword in a simple one word, simple multi word and complex multi word group and we taken on each group two searching keyword.

Keyword: Search Engine, Directories search, Index Search, Crawler, World Wide Web, Precision, Relative Recall.

I. INTRODUCTION

The help of Web we can be quick and direct get to any type of information all over the world. The Information of the found in the Web then time information needs to be filters and may include huge non-relevant information. Performance of presently existing search engines is improving continuously on day by day with powerful search capabilities. In present time Search engines play a critical role in the process of retrieve information from the Web. When the user or person gives a Query or searching Keyword then Search engine returns a list of relevant results in ranked order. The results of search query is a same type of search Keyword ,different type of subtopics, meanings of a query mixed jointly in the result list. An Effective approach of the retrieval of the information on the Web in latest years is by using the Meta Search Engine as a substitute of simply a Search Engine.

World Wide Web present day is a store address of huge quantity of information but not any of the searches engine can be explore added than 16 % of available web information or data. Though million of web pages are presented but the result of the each search is only some web pages. For a search engine it's very hard to decide which web pages are mainly important for user . Here we discuses to search engines in the bases of it special type and then we analysis on some parameters like more relevant search, less relevant search, irrelevant search, link, Repeated link, site can't be access , total resultant document, time taken for search, current status of search engine, efficiency. Total resulting URLs refers to the total number of web pages being display by the search engine as a result of the search keyword or query entered by the user.

Web Search engines search to all information and all things in internet like a user files, multimedia element like songs, videos, images, web page, web sites, weather information through various interfaces means through search engine. In present day a lot of search engine are available and each search engine are used a different language interface, techniques for searching and indexing, searching algorithm, web services and interface for searching keyword. The National Information Standards society defines to boundary and access management is web search engine. Search engine is a web application they handle on the particular IT Company and this origination given to a unique website name. User given a some searching query or keyword on the specific search engine and search engine are create a dynamically searching keyword related web page listing. in general three types of web search engines are accessible they are a Index base web search engines, Directories base web search engines and Meta search engine.

A. Index Web Search Engines

Indexes are used to Spiders or robots search program and used to huge amount of data in database. This database create a dynamically listing for searching time. Index search engines are use autonomous software means it type of web search engines search to all subject related information in Web. AlltheWeb, Google, AltaVista, Teoma, Wisenut are popular Index web search engines and this research Google search engine are taken for performance analysis.

Google:- Google is product of American multinational technology company. Google is given to Internet related services and include online advertise technologies, searching for web, android operating system for mobial, E-Mail, cloud computing for accessing for data and more software. Google is come to internet word in September 4, 1998.

B. Directories Web Search engines

This type of search engine are classify web documents or sites into a subject classification, yellow pages scheme for a all type of Entertainment, Arts , Business, Computers and Internet. This type of search engine are regularly compiled by some type of rational order and using on small database uses as compared to Computer generated indexes. directories search engine manually place Web page into specific category means directories search engines are search in web only specific subject related information. Yahoo is a directories web search engines. We have taken to Yahoo search engine for performance analysis.

Yahoo:- Yahoo is product of American international technology industry and it come to January 1994 in internet word. Yahoo product and service is related to internet. yahoo service are a Yahoo News, Yahoo Mail,

Finance news , Sports news , Search for web, image, video, Messenger, yahoo Answer, online mapping, video sharing etc

II. PRECISION EVALUATION OF SEARCH ENGINE

First factor of performance is precision. in this section we computation on precision of search engines for each of the search keyword using this procedure and used to five criterion (Eq. 1).

Precision=Sum of the scores in web sites or page retrieve by a Search Engine / Total number of sites or page retrieved (Eq. 1)

Performance evaluation of Google and Yahoo are examined to during November 2015 to 10 January 2016. In this study Google and Yahoo are given to search results and this search result we are categorize as five points. These points are first is more relevant web page, second is less relevant web page, third is irrelevant web page , four is links and five point is web page or web sites can't be accessed in the basis of following criterion and this criteria and points are given in the B.T. Sampath kumar 2010[1], Ding & Marchionini 1996[2], Chu & Rosenthal 1996[3], Clarke & Willett 1997[4], Leighton 1996[5]. These criteria using on the calculation of precision and relative recall of present time search. this criteria are.

- .If the web page contented is closely matched the searching query or searching subject topic then we categorize this web page is as more relevant and given to 2 point or score.
- If the web page contented is not matched the closely for subject topic but some aspects are related to the searching query or searching subject topic then it web page categorize as less relevant and given to 1 point or score .
- If the web page contented are not related to searching query or searching subject topic then it web page categorize as irrelevant and given to 0 point or score.
- If the web page contented is are given to series of web links of another web page but something is required for subject topic then it web page categorize as links and it score or point are a 0.5.
- If the web site is can't be accessed or not open for a particular web address or URL then its web page categorize as site can't be accessed and given to 0 point.

Using of these point we compute precision and relative recall. First we calculate a precision and after we calculate relative recall.

a. **Precision of Google (Index web search engine*):-*** Total numbers of 7,67,57,00,000 sites are founded for different 12 keyword and we selected to 1200 sites for precision calculated. Following Table 1 are shows the total number of more relevant web page, less relevant web page, irrelevant web page, links and sites cannot be accessed of Google in selection of 1200 sites. Clear for this table is 40.33% of sites are less relevant and 33.58% of sites are more relevant. Precision mean of Google is 1.11 found.

b. **Precision of yahoo (Directory web search engine):-** Total numbers of 5,473,600,000 sites are founded for 12 keyword and we selected to 900 sites for precision calculated. Following Table 2 are listed the total number of more relevant web page, less relevant web page, irrelevant web page, links and sites cannot be accessed of yahoo in selection of 1200 sites. Clear for this table is 37.41% less relevant sites, 13.16% irrelevant sites and only 38.50% of sites are more relevant. Yahoo precision mean is 1.17.

International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 04, April 2016
www.ijates.com
ijates
ISSN 2348 - 7550

TABLE 1- Precision calculation 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	Searching Time	Precision
Simple one word queries/Keyword										
Computer	2,30,00,00,000	100	43	23	16	12	6	**7**	0.40 se	1.15
Database	99,80,00,000	100	39	36	18	6	1	**4**	0.58 se	1.17
Multimedia	64,70,00,000	100	36	32	14	9	9	**4**	0.52 se	1.09
Program	78,26,00,000	100	40	43	9	4	4	**2**	0.55 se	1.25
Simple multi word queries										
What is search engine	35,90,00,000	100	32	41	18	5	4	**2**	0.60 se	1.075
Computer science	28,90,00,000	100	45	39	12	3	1	**1**	0.39 se	1.44
Digital India	43,30,00,000	100	31	46	13	6	4	**2**	0.53 se	1.38
Operating System	36,40,00,000	100	39	43	13	4	1	**2**	0.38 se	1.23
Complex multi word queries/ Keyword										
Internet and web designing	3,23,00,000	100	26	42	19	11	2	**9**	0.52 se	0.99
Evaluation of digital library	1,45,00,000	100	23	46	14	13	4	**8**	0.52 se	0.98
Computer science & engineering	22,40,00,000	100	22	44	15	12	7	**8**	0.51 se	0.94
Window and Linux Operating System	23,23,00,000	100	27	49	9	12	3	**7**	0.56 se	1.09
Total	**7,67,57,00,000**	**1200**	**403 (33.58%)**	**484 (40.33%)**	**170 (14.16%)**	**97 (8.08%)**	**46 (3.83%)**	**44 (3.66%)**		**1.11**

TABLE 2- Precision calculation of Yahoo

Search keyword	Total numberof sites	Selected sites	More relevant sites	Less relevant sites	Irrelevant sites	links	Sites cannot be accessed	Repeated link	Searching Time	precision
Simple one word queries										
Computer	539,000,000	100	42	44	9	3	2	1	0.38 se	1.29
Database	436,000,000	100	36	36	13	8	7	4	0.55 se	1.12
Multimedia	89,800,000	100	35	29	17	9	10	4	0.51 se	1.03
Program	458,000,000	100	38	34	12	6	10	4	0.52 se	1.13
Simple multi word queries										
What is search engine	3,740,000,000	100	29	36	14	11	10	7	0.54 se	0.99
Computer science	16,300,000	100	36	48	8	7	1	5	0.38 se	1.23
Digital India	12,700,000	100	32	38	17	9	4	3	0.58 se	1.06
Operating System	28,800,000	100	34	36	16	12	2	4	0.36 se	1.10
Complex multi word queries										
Internet and web designing	109,000,000	100	42	36	17	3	2	0	0.54 se	1.21
Evaluation of digital library	10,300,000	100	46	36	12	4	2	0	0.55 se	1.3
Computer science & engineering	16,300,000	100	49	37	11	2	1	0	0.52 se	1.36
Window and Linux Operating System	18,400,000	100	43	39	12	4	2	1	0.57se	1.27
Total	**5,473,600,000**	**1200**	**462 (38.5%)**	**449 (37.41%)**	**158 (13.16%)**	**78 (6.50%)**	**53 (4.41%)**	**29 (2.41)**		**1.17**

We selected the first top 100-result link given by Google and Yahoo search engine. We try to show comparative precision analysis of Google and Yahoo show in graph figure 1 in the bases of searching Keyword. Comparative Performance analysis of Google and Yahoo show in graph figure 2 in the base of searching Keyword and precision. Finally Table 3 are summaries the total precision of simple one word, simple multi word and complex multi word group of Google and Yahoo and graph figure 3 are show to comparative precision on the basis this three group. We try in graph figure 4 are show to repeated link in the basis of searching link and graph figure 5 are show to searching time and according to searching Keyword.

International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 04, April 2016
www.ijates.com
ijates
ISSN 2348 - 7550

Figure 1- Comparative Precision Analysis of Google and Yahoo

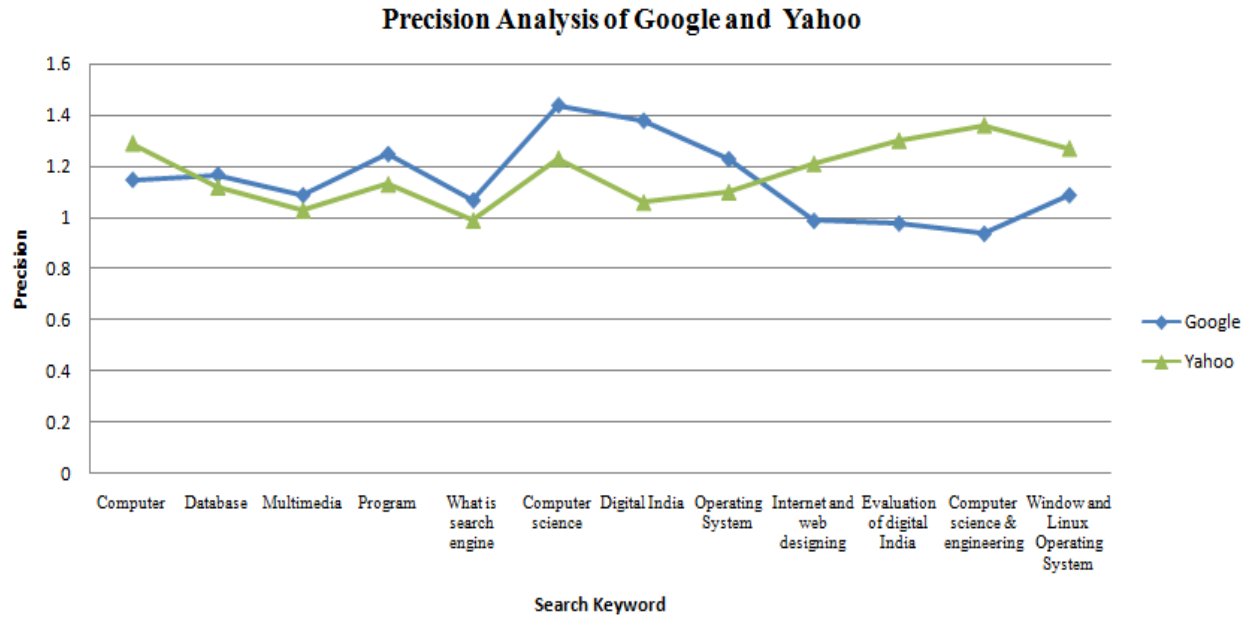


Figure 2- Comparative Performance Analysis of Google and Yahoo

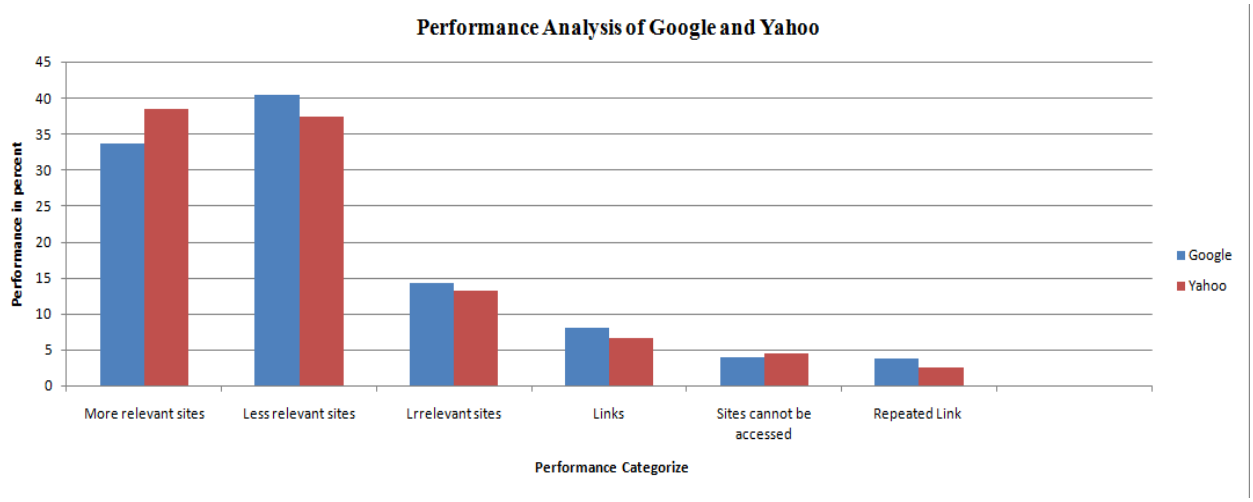


Table 3 –Comparative precision of Google and Yahoo

Search Engine	Total number of Simple one word	Total number of Simple multi word	Total number of Complex multi word	Total Precision
Google	1.16	1.18	1.00	1.11
Yahoo	1.14	1.07	1.28	1.17

Figure:-3 Comparative precision analysis according to word group

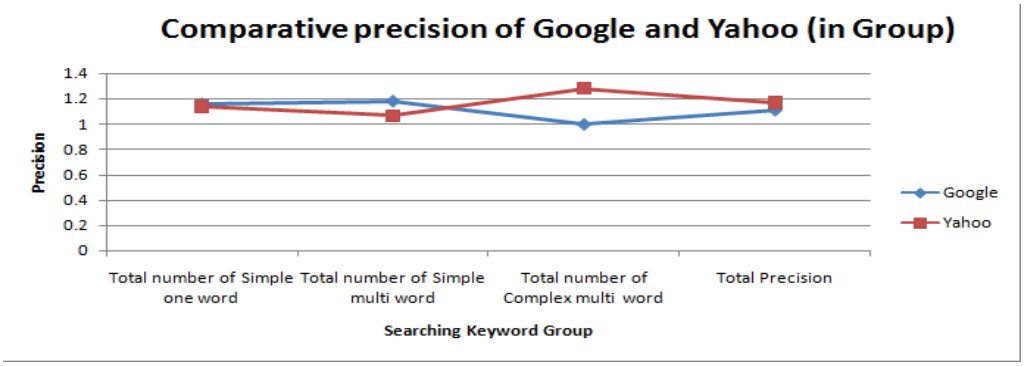


Figure 4- Repeated link Analysis of Google and Yahoo

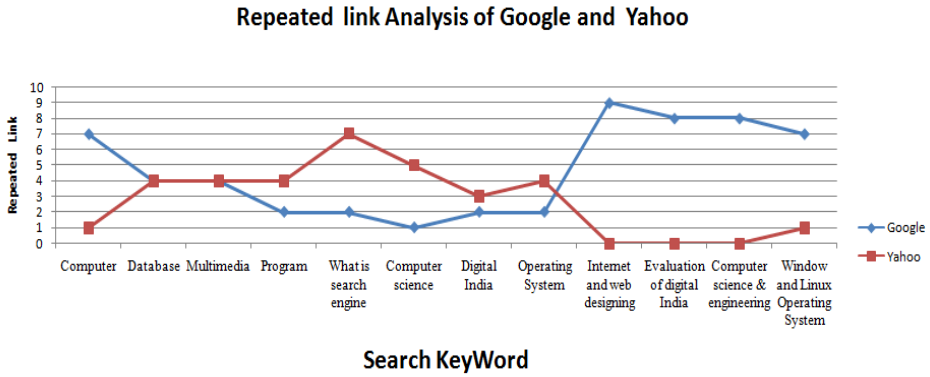
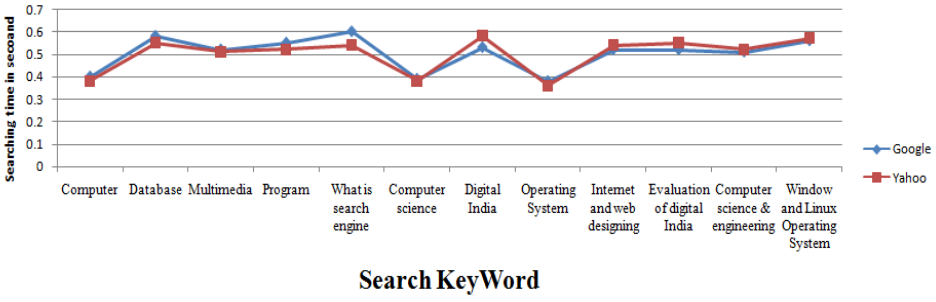


Figure 5- Searching Time Analysis of Google and Yahoo



III. RELATIVE RECALL EVALUATION OF SEARCH ENGINE

Second factor of performance is relative recall. Recall is a retrieval system and it achieve all or most relevant documents in the collection means recall is the ratio of the amount of relevant records retrieve to the search engine and total number of relevant records in the database. Calculating on relative recall using this formula and this formula (Eq. 2).

Relative Recall =Total number of web sites or web page are retrieve by a search engine/ Sum of sites retrieved by the all search engine (Eq. 2)

a. **Relative Recall of Google (Index web search engine):** - Total numbers of **7,67,57,00,000** sites are founded for different 12 keyword. Google is given to relative recall is 0.99 in all group but it given in

Simple one word group have it recall is 0.99, Simple multi word group have it recall is 0.99 and Complex multi word have it recall is 0.99.

b. **Relative Recall of Yahoo (Directory web search engine):-**Total numbers of 4,968,400,000 sites are founded for different nine keyword. Yahoo is given to relative recall is 0.48 in all group but it given in Simple one word group have it recall is 0.22, Simple multi word group have it recall is 0.78 and Complex multi word have it recall is 1.64.

The relative recall of the Google and Yahoo is calculated and show the Table 4 in the base of searching keyword and graph figure 6 shows to comparative analysis for relative recall. We try to summaries the total relative recall of simple one word, simple multi word and complex multi word group of Google and WebCrawler in Table 5 and graph figure 7 are show to comparative relative recall on the basis this three group.

Table -4 Relative recall of the Google and Yahoo

Searching Keyword	Google		Yahoo	
	Total No. of Sites	Relative Recall	Total No. of Sites	Relative Recall
Simple one word				
Computer	2,30,00,00,000	0.81	539,000,000	0.18
Database	99,80,00,000	0.69	436,000,000	0.30
Multimedia	64,70,00,000	0.87	89,800,000	0.12
Program	78,26,00,000	0.63	458,000,000	0.36
Simple multi word				
What is search engine	35,90,00,000	0.08	3,740,000,000	0.91
Computer science	28,90,00,000	6.39	16,300,000	0.36
Digital India	43,30,00,000	0.97	12,700,000	0.02
Operating System	36,40,00,000	0.92	28,800,000	0.07
Complex multi word				
Internet and web designing	3,23,00,000	0.23	109,000,000	0.77
Evaluation of digital India	1,45,00,000	0.58	10,300,000	0.41
Computer science & engineering	22,40,00,000	0.93	16,300,000	0.67
Window and Linux Operating System	23,23,00,000	0.92	18,400,000	0.73
Total	**7,67,57,00,000**	**0.58**	**5,473,600,000**	**0.41**

Figure 6- Comparative Relative Recall analysis of Google and Yahoo

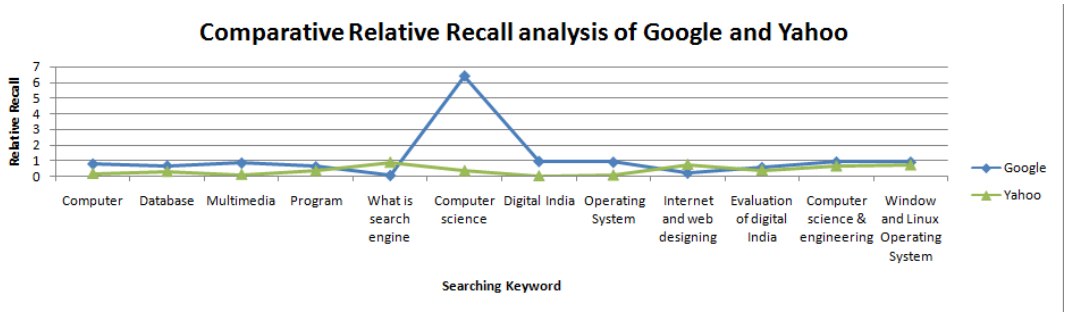
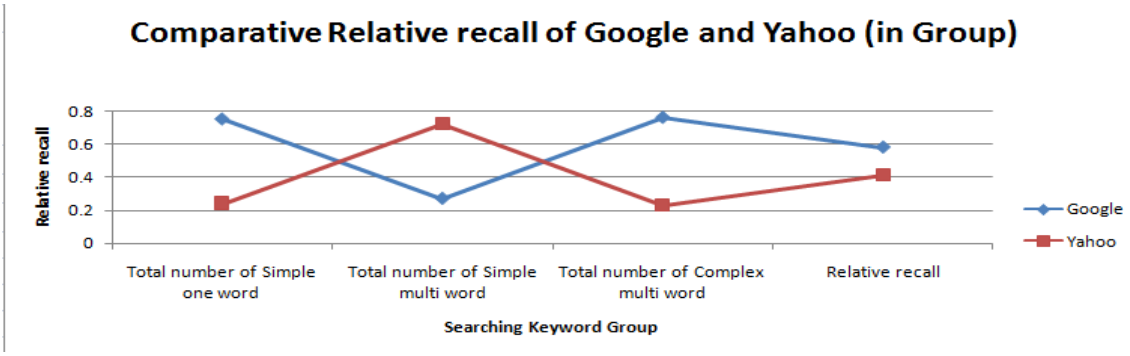


Table -5 Comparative Relative recall of Google and Yahoo

Search Engine	Total number of Simple one word	Total number of Simple multi word	Total number of Complex multi word	Relative recall
Google	0.75	0.27	0.76	0.58
Yahoo	0.24	0.72	0.23	0.41

Figure 7- Comparative Relative Recall analysis according to word group



IV. CONCLUSION

In this paper we presented the overview and performance (precision and relative recall) of Index web search engine Google and Directory web search engine Yahoo. The present study estimated the precision and relative recall of Google and Yahoo. The result of this study is show the precision of Google is high for simple one word, Multi Word but Google Precision for complex multi word queries are less then Yahoo Directory Web search engine. Overall precision of Yahoo is always greater then Google precision. Relative Recall of Google was high for simple one word and complex multi word queries but Relative Recall of Google was less for Multi Word queries as compare to Yahoo Relative Recall. Overall Relative Recall of Google is always greater then Yahoo Relative Recall. Finally we find Google is given to large amount of result and it best for deep Search and complex Multi Word Group and Yahoo best for simple queries.

REFERENCES

[1] 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

[2] 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

[3] 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.

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

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

[6] 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.

[7] Kamlesh Kumar Pandey "A ANALYSIS OF GOOGLE, YAHOO AND BING WEB SEARCH ENGINE IN PERFORMANCE PARAMETER OF PRECISION AND RELATIVE RECALL", 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.

[8] 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

[9] 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.

[10] Felipe Bravo-Marquez, Gaston L'Huillier, Sebasti´an A. R´ıos, Juan D. Vel´asquez, "A Text Similarity Meta-Search Engine Based on Document Fingerprints and Search Results Records" 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pp-146-153

[11] K.Srinivas, P.V.S. Srinivas, A.Govardhan," Web Service Architecture for a Meta Search Engine", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 10, 2011,pp-31-36

[12] Dorn J. and Naz T."Structuring Meta-search Research by Design Patterns", Institute of Information Systems,Technical University Vienna, Austria; International Computer Science and Technology Conference; San Diego; April, 2008

[13] Subarna Kumar Das," ROLE OF META SEARCH ENGINES IN WEB- BASED INFORMATION SYSTEM: FUNDAMENTALS AND CHALLENGES" 4th Convention PLANNER -2006, INFLIBNET Centre, Ahmedabad, Mizoram Univ.,Aizawl, 09-10 November, 2006, pp-445-454

[14] 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.

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

[16] metasearch engines" In WISE, volume 3806 of Lecture Notes in Computer Science, pages 53–66. Springer,2005

[17] Meng W., Yu C. and Liu K. "Building efficient and effective metasearch engines" In ACM Computing Surveys; 2002

[18] Javed A. Aslam and Mark Montague. Models for metasearch. In SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, pages 276–284, New York, NY, USA, 2001. ACM.

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

[20] https://www.google.com