



INTERNET SEARCH ENGINE: PRECISION AND RELATIVE RECALL ANALYSIS OF GOOGLE, YAHOO AND WEBCRAWLER SEARCH ENGINE BASED ON THEIR SEARCHING CAPABILITY

Kamlesh Kumar Pandey¹, Rajat Kumar Yadu², Pradeep Kumar Shukla³

^{1,3}MCRPVV UTD, Amarkantak, (India)

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

ABSTRACT

In Present day internet are most power full network and it given a space for storing all type of importance information. If anybody, person or organization can be wanted to any type of information, data, news or other things he searches to WWW through search engine. World Wide Web is a storing to all type of data and Search engine given the searched item related list of link and person select the one link at a time. Search engine given the some importance link they related to search item and some link is not related to search item. In present day basically three type of search engine are used first is Index base search engines search second is Directories search engines and third is Meta search engine. Main aim of this research paper is finding on which search engine are given to best result and how many link are related to searched item. This paper help to they person or organization which are searching on data or information to all time because we try to calculation on performance of Index base search engines, Directories search engines and Meta search engine.

Index Terms: Search Engine, Meta Search Engine, Directories search, Index Search, Crawler, World Wide Web, Precision.

I. INTRODUCTION

If anybody or organization can be want to any type of information in present time in commonly he go to internet and search this information through search engine in WWW. In present time many searching technique are used in internet or WWW and one specific own search algorithm are used to search engine. Web Search engines search to all things in World Wide Web like a files, songs, videos, images, web sites, weather information through various interfaces. In present day a lot of search engine used different type of searching language, techniques, searching algorithm, web services and interface. The National Information Standards Organization (NISO) uses the terms Federated Search and Meta search interchangeable to explain this web search model means boundary and access management are define to NISO is any type of search engine. Search engine is a web application program they handle on the particular organization and this origination given to a unique website name. User given a searching keyword on the specific search engine web site and it's are creating on dynamically keyword



related web page listing. Normally three types of search engines are available in present time they are a Index search engines, Directories search engines and Meta search engine.

1.1 Index Web Search engines

Indexes are used to Spiders or robots search programs and used large database they create a dynamically listing. Index search engines are use autonomous software means it search engines search to all subject. AlltheWeb, Google, AltaVista, Teoma, Wisenut are popular Index web search engines and we taken to Google search engine for performance analysis.

Google:- Google is an American multinational technology corporation it given to Internet related services and include online advertise technologies, searching, android operating system, Mail, cloud computing, and software. Google is come to internet word in September 4, 1998.

1.2 Directories Web Search engines

Directories classify web documents or sites into a subject classification, yellow pages scheme like a Arts, Business, Computers and Internet, Entertainment, Government they are usually compiled by some type of logical order and small database uses compared to Computer-generated indexes. and directories search engine manually place Web sites and pages into specific categories means directories search engines are search to only one subject. Yahoo,Open Directory are popular directories web search engines. and we taken to Yahoo search engine for performance analysis.

Yahoo:- Yahoo is an American international technology business company and it come to January 1994 in internet word. Yahoo product and service is related to internet and this service are a Yahoo News, Mail, Finance, Sports, Search, Messenger, Answer, Flickr, online mapping, video sharing etc

1.3 Meta search engine

Meta search engine is a special type of web search engine and this run on the particular own website. User given a searching keyword on the specific Meta search engine web site and it's are creating on dynamically keyword related web page listing using on difference search engine database not a own database. Multi search engines don't carry out the crawling or maintain own database like web search engines but filtering the results found as a replacement for Based on a specific algorithm and remove duplicate and given the rank and listing on results from their sources. Meta search engine is a use an index and directory web search engine then religion it a best way of advertising on using yellow page and white page. Yellow page is a telephone directory of production organized by group rather than alphabetically through business name and in which advertising is sold and White page environmental area to service give by the society that publish the directory. Its reason is to allow the phone number of a subscriber known by name and address to be found. WebCrawler ,MetaCrawler, Dogpile, Brainboost are example of popular meta search engine.. WebCrawler blends the search results from Google, Yahoo Search WebCrawler was the first Web search engine and it given full text search. MetaCrawler blends the web search results from Google, Yahoo, Bing, Ask and other popular search engines. and we taken to WebCrawler search engine for performance analysis.

WebCrawler:- WebCrawler is a register trade mark of InfoSpace and it come to April 20, 1994. it produced by Brian Pinkerton at the University of Washington. WebCrawler provides to users the option for search is images, audio, video, news, yellow pages and white pages.

II. EFFICIENCY OR PERFORMANCE EVALUATION OF SEARCH ENGINE

Performance evaluation of Google, Yahoo and WebCrawler are examined to during November 2015 to 2 January 2016. In this study Google, Yahoo and WebCrawler are given to search results and this search result we 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 following criteria and this criteria and points are given in the Chu & Rosenthal 1996[1], Leighton 1996[2], Ding & Marchionini 1996[3], Clarke & Willett 1997[4], B.T. Sampath kumar 2010[5]. We define new criteria on the bases on old criteria and calculation on precision and relative recall of present time search. These new criteria are identify the above five point and this criteria are.

- If the web page content is closely matched the subject topic of the search keyword then we categorize this web page is as more relevant and given to 2 point.
- If the web page content is not closely matched the subject topic but some aspects related to the subject topic of the search keyword then it web page categorize as less relevant and given to 1 score.
- If the web page content are not related to the subject topic of the search keyword then it web page categorize as irrelevant and given to 0 point.
- If the web page content is are given to complete series of links of another web page but some information is required then it web page categorize as links and given to 0.5 point.
- If the web site is can't be accessed or open for a particular URL then its web page categorize as site can't be accessed and given to 0 point.
- We calculate precision and relative recall using these five points first we calculate a precision and after we calculate relative recall.

A. **Precision calculation**:- First factor of performance is precision so this section we calculation on precision of search engines for each of the search keyword using this formula and used to five criteria (Eq. 1).

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

a. ***Precision of Google (index web search engine)***:- Total numbers of 5,29,68,00,000 sites are founded for different nine keyword and we selected to 900 sites for precision calculated. Following Table 1 are shows the total number of more relevant web sites, less relevant web sites, irrelevant web sites, links and sites cannot be accessed of Google in selection of 900 sites. Clear for this table is 38.78% of sites are less relevant and 33% of sites are more relevant. Precision mean of Google is 1.09 found.

b. ***Precision of yahoo(directory web search engine)***:- Total numbers of 4,968,400,000 sites are founded for nine keyword and we selected to 900 sites for precision calculated. Following Table 2 are shows the total number of more relevant sites, less relevant sites, irrelevant sites, links and sites cannot be accessed of



yahoo in selection of 3000 sites. Clear for this table is 37.78% less relevant sites, 13.11% irrelevant sites and only 38.56% of sites are more relevant. Yahoo precision mean is 1.18.

- c. **Precision of WebCrawler (Meta search engine):-**Total numbers of 2870 sites are founded for nine keyword and we selected to all 2870 sites for precision calculated. Following Table 3 are shows the total number of more relevant sites, less relevant sites, irrelevant sites, links and sites cannot be accessed of selection of 2870 sites. Clear for this table is 40.13% less relevant sites and 53.73% of sites are more relevant. WebCrawler precision mean is 1.48.

TABLE 1- Precision calculation of Google

Search keyword	Total number of sites	Selecte d sites	More relevant sites	Less relevant sites	Irrelevant sites	links	Sites cannot be accessed	Repeated link	Precision
Simple one word queries									
Computer	2,30,00,00,000	100	43	23	16	12	6	7	1.15
Database	99,80,00,000	100	39	36	18	6	1	4	1.17
Multimedia	64,70,00,000	100	36	32	14	9	9	4	1.09
Simple multi word queries									
What is search engine	35,90,00,000	100	32	41	18	5	4	2	1.075
Computer science	28,90,00,000	100	45	39	12	3	1	1	1.44
Digital India	43,30,00,000	100	31	46	13	6	4	2	1.38
Complex multi word queries									
Internet and web designing	3,23,00,000	100	26	42	19	11	2	9	0.99
Evaluation of digital library	1,45,00,000	100	23	46	14	13	4	8	0.98
Computer science & engineering	22,40,00,000	100	22	44	15	12	7	8	0.94
Total	5,29,68,00,000	900	297	349	139	77	38	33	1.09



			(33 %)	(38.78%)	(15.44%)	(8.56 %)	(4.22%)	(3.67%)	
--	--	--	--------	----------	----------	----------	---------	---------	--

TABLE 2- Precision calculation of Yahoo

Search keyword	Total number of sites	Selecte d sites	More relevan t sites	Less relevan t sites	Irreleva nt sites	links	Sites cannot be accesse d	Repeat ed link	precisi on
Simple one word queries									
Compute r	539,000,000	100	42	44	9	3	2	1	1.29
Database	436,000,000	100	36	36	13	8	7	4	1.12
Multimed ia	89,800,000	100	35	29	17	9	10	4	1.03
Simple multi word queries									
What is search engine	3,740,000,000	100	29	36	14	11	10	7	0.99
Compute r science	16,300,000	100	36	48	8	7	1	5	1.23
Digital India	12,700,000	100	32	38	17	9	4	3	1.06
Complex multi word queries									
Internet and web designing	109,000,000	100	42	36	17	3	2	0	1.21
Evaluatio n of digital library	10,300,000	100	46	36	12	4	2	0	1.3
Compute r science & engineeri ng	16,300,000	100	49	37	11	2	1	0	1.36



Total	4,968,400,000	900	347 (38.56%)	340 (37.78%)	118 (13.11%)	56 (6.22%)	39 (4.33%)	24 (2.67%)	1.18
--------------	----------------------	------------	------------------------	------------------------	------------------------	----------------------	----------------------	----------------------	-------------

TABLE 3- Precision calculation of WebCrawler

Search keyword	Total number of sites	More relevant sites	Less relevant sites	Irrelevant sites	links	Sites cannot be accessed	Repeated link	precision
Simple one word queries								
Computer	210	126	76	3	4	1	0	1.57
Database	250	145	102	1	1	1	0	1.57
Multimedia	240	123	91	12	12	2	0	2.79
Simple multi word queries								
What is search engine	100	53	42	3	2	1	0	1.49
Computer science	580	294	234	26	13	13	0	1.43
Digital India	120	66	49	3	2	1	0	1.52
Complex multi word queries								
Internet and web designing	570	286	256	25	2	1	0	1.45
Evaluation of digital India	450	226	197	19	6	2	0	1.45
Computer science & engineering	350	223	105	12	6	4	0	1.29
Total	2870	1542 (53.73%)	1152 (40.13%)	104 (3.62%)	48 (1.67%)	26 (0.91%)	0 (0.0%)	1.48

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



Yahoo and WebCrawler 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

Figure 1- Comparative Precision Analysis of Google, Yahoo and WebCrawler

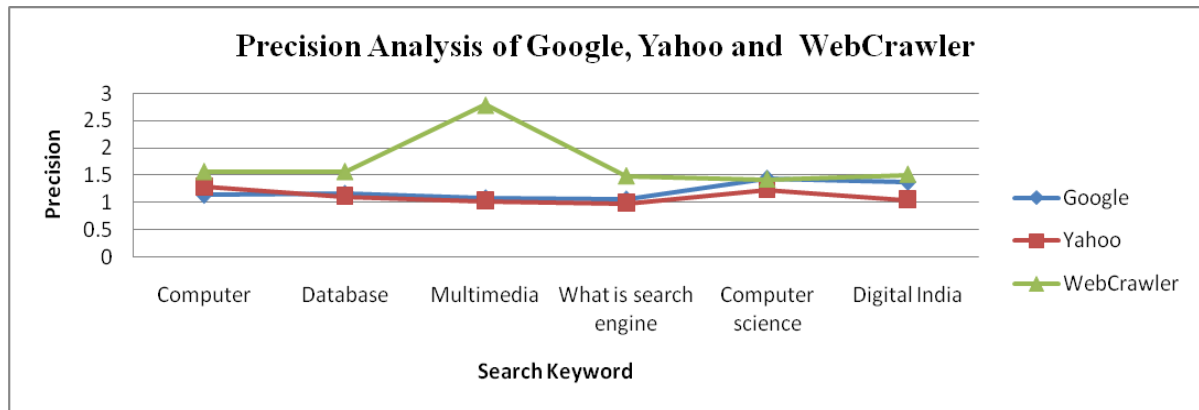


Figure 2- Comparative Performance Analysis of Google, Yahoo and WebCrawler

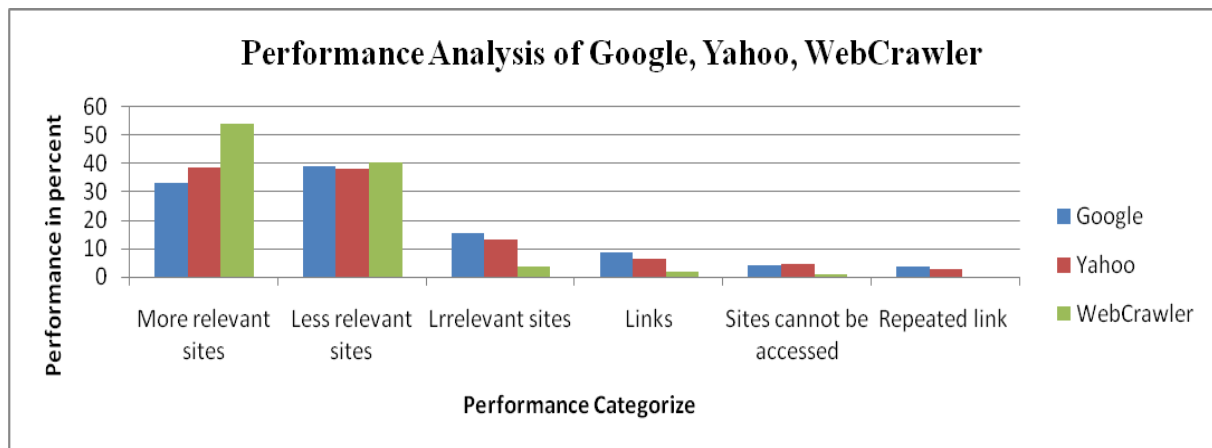


Table 4 –Comparative precision of Google, Yahoo and Bing

Search Engine	Total number of Simple one word	Total number of Simple multi word	Total number of Complex multi word	Total Precision
Google	1.36	1.16	0.97	1.09
Yahoo	0.79	1.06	1.29	1.18
WebCrawler	1.52	1.45	1.49	1.48

Figure:-3 Comparative precision analysis according to word group

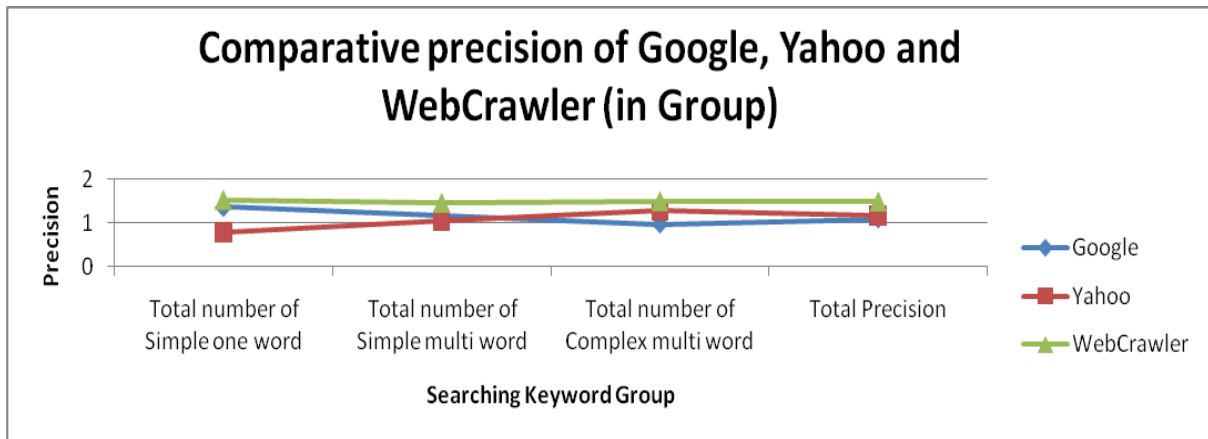
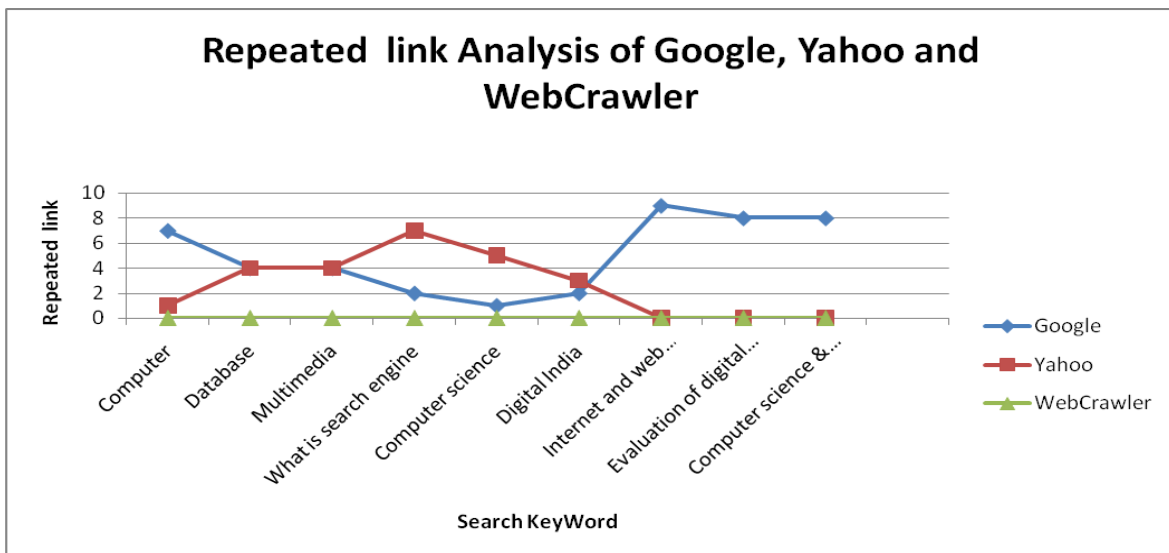


Figure 4- Repeated link Analysis of Google, Yahoo and WebCrawler



B. **Relative Recall** :- 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).

$$\text{Relative Recall} = \frac{\text{Total number of web sites retrieve by a search engine}}{\text{Sum of sites retrieved by the all search engine}} \text{ (Eq. 2)}$$

a. **Relative Recall of Google (index web search engine)**:- Total numbers of 5,29,68,00,000 sites are founded for different nine keyword. Google is given to relative recall is 0.51 in all group but it given in Simple one word group have it recall is 0.78, Simple multi word group have it recall is 0.21 and Complex multi word have it recall is 1.39.

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.

c. **Relative Recall of WebCrawler (Meta search engine):-** Total numbers of 2870 sites are founded for different nine keyword. Yahoo is given to relative recall is 2.79 in all group but it given in Simple one word group have it recall is 0.66, Simple multi word group have it recall is 0.33 and Complex multi word have it recall is 0.61.

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

Table -5 Relative recall of the Google, Yahoo and WebCrawler

Searching Keyword	Google		Yahoo		WebCrawler	
	Total No. of Sites	Relative Recall	Total No. of Sites	Relative Recall	Total No. of Sites	Relative Recall
Simple one word queries						
Computer	2,30,00,00,000	0.81	539,000,000	0.19	210	7.39
Database	99,80,00,000	0.69	436,000,000	0.30	250	1.74
Multimedia	64,70,00,000	0.87	89,800,000	0.12	240	3.26
Simple multi word queries						
What is search engine	35,90,00,000	0.08	3,740,000,000	0.91	100	2.44
Computer science	28,90,00,000	6.39	16,300,000	0.36	580	1.28
Digital India	43,30,00,000	0.97	12,700,000	0.02	120	2.69
Complex multi word queries						
Internet and web designing	3,23,00,000	0.23	109,000,000	0.77	286	2.02
Evaluation of digital India	1,45,00,000	0.58	10,300,000	0.41	226	9.11
Computer science & engineering	22,40,00,000	0.93	16,300,000	0.67	223	9.28
Total	5,29,68,00,000	0.51	4,968,400,000	0.48	2870	2.79



Figure 5- Comparative Relative Recall analysis of Google, Yahoo and Bing

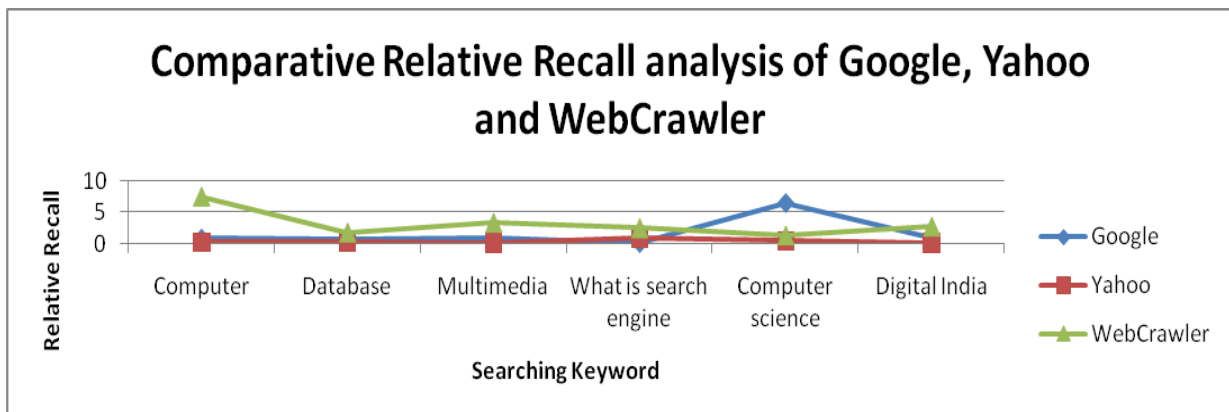
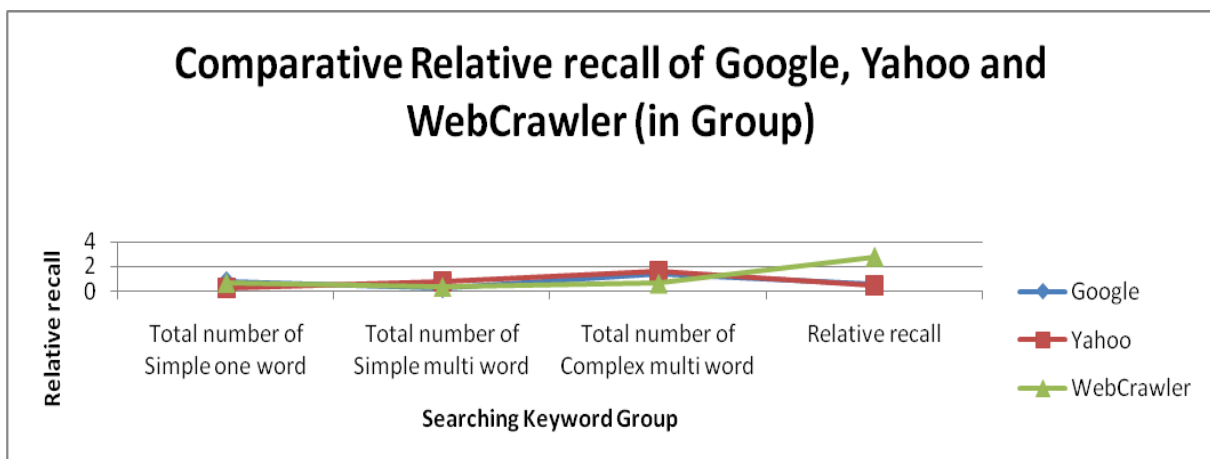


Table -6 Comparative Relative recall of Google, Yahoo and WebCrawler

Search Engine	Total number of Simple one word	Total number of Simple multi word	Total number of Complex multi word	Relative recall
Google	0.78	0.21	1.39	0.51
Yahoo	0.22	0.78	1.64	0.48
WebCrawler	0.66	0.33	0.61	2.79

Figure 6- Comparative Relative Recall analysis according to word group



III. CONCLUSION

In this paper we presented the overview and performance (precision and relative recall) of web search engine and Meta search engine. Main aim of this study is evaluating the performance of Google, Yahoo, WebCrawler search engine and comparing the theirs precision and relative recall. Meta search engine are not using one search engine but uses many search engine at one time to utilize of web searching effectively. Finally result of this study shows on precision of Google is 1.09, Yahoo is 1.18 and WebCrawler is 1.48 of all group. Then over



all WebCrawler precision is higher than Google and Yahoo, Yahoo precision higher than Google but we taken on group then WebCrawler and Google given to high precision then Yahoo in Simple one word and multi word but Complex multi word group Google have a low Precision. Second factor of performance is Relative Recall (most relevant document) and this result is of Google is 0.51, Yahoo is 0.48 and Bing is 2.79 of all group so WebCrawler have a higher Relative Recall and Yahoo is given to lower Relative Recall in first group, second group Google have a low Relative Recall . we also check to repeated link in all search keyword we find Google database is large and it sometime some link is repeated but WebCrawler is not repeated to any link. So Meta search engine (WebCrawler) retrieve less number of sites for all search keyword. Mean precision of Meta search engine is comparatively high as compared to index and directory search engine. It clearly shows Meta search engine are achieve higher precision as a compare search engines

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] 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
- [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] 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
- [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'ios, 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] MARK LEVENE, “AN INTRODUCTION TOSEARCH ENGINES AND WEB NAVIGATION”, ISBN 978-0-470-52684-2 (pbk.) year-2010, Published by John Wiley & Sons, Inc., Hoboken, New Jersey
- [20] Kim-An Phan, “Meta-Search Engine Analysis” Seminar Thesis, University of Fribourg Faculty of Economics & Social Sciences Information Systems Research Group
- [21] <https://www.webcrawler.com>
- [22] <https://www.yahoo.com>
- [23] <https://www.google.com>