

# Impact of Meta Search Engines on Information Retrieval Effectiveness: A Case Study



Nagaraju Mamillapally<sup>1</sup>, Trivikram Mulukutla<sup>2</sup>

<sup>1</sup>Lecturer, Department of Computer Science, Adarsh Degree & PG College, Mahabubnagar, Andhra Pradesh, India, nagaraj.mavilla29@gmail.com

<sup>2</sup>Lecturer, Department of Computer Science, Adarsh Degree & PG College, Mahabubnagar, Andhra Pradesh, India, mulukutlatrivikram@yahoo.in

**Abstract:** WWW is a warehouse with millions of websites, billions of web pages and tons of data, and is growing exponentially by the order of the day. It is almost impossible to access right content for which user is looking for in this ocean of information, unless he/she knows the exact location or URL. Here web search becomes a key technology and one of the important purposes of the web. This is where the role of a search engine becomes meaningful. Meta search engines extract some thousands of documents from different search engines, filter them as per the user requirement and display them. Selecting a right Meta Search Engine among plenty of available search engines is a very big challenge. This is because, not all the search engines produce documents which are most relevant to the user query.

Keeping this point in view, in this paper we have consider Zapmeta, Google, Yahoo and Search from the top most Meta search engines list made a case study based investigation to identify how Meta search engines selection shows a greater impact on the information retrieval effectiveness. The results showed that the precision of Yahoo was high for simple one-word queries (9.75) and Search had comparatively high precision for simple multi-word queries (9.46).

**Keywords:** WWW, Warehouse, URL, Meta Search Engines, Zapmeta, Google, Yahoo, Search, Information Retrieval.

## INTRODUCTION

Internet can be said as virtual world on which millions of people fulfilling their information needs across the globe. Such popularity of internet is because of WWW or Web Service, where millions if websites available which consists tons of information. Today no information is there that is not in the web. The platform for finding a required website is the search engine. Among most of them, user is facing difficulty in finding a website that provides the desired information against the search request. Solution for this problem is the selection of a right Meta search engine, which meets the user needs.

A search engine is a program on internet to extract information stored on the web. In simple words, search engine is an interface between user and the web. For a same query, different search engines display different results.

Search engines are categorized into various types based on criteria.

- Search engines based on the method used for building the database are

- Pure Search Engines: Robot-based search engines use automated software agents that visit all websites and index the information on each page.
  - Directories: These search engines build their indexes by facilitating manual submission of web pages.
  - Hybrid Search Engines: These are also called mixed-result search engines. These search engines populates its catalog or index by using both spiders as well as manual indexing.
- Search engines based on the scope of search they provide like
    - Primary Search Engines: These search engines has a wider scope of conducting search for a given query. They scan the entire web and build a database of all active web pages for later use to conduct search.
    - Meta Search Engines: These do not own any database or index. They conduct search on multiple search engines for a given query. It accumulates and filters the results provided by various search engines before presenting them to users.
    - Subject Guides: These search engines have a broader focus on a specific subject. These are small indexes essentially built with more human intern mention.
    - Specialized Search Engines: These are search engines for specific purpose like to search jobs, people, places, products, news, reports, medicines and images etc.

Search engines consists of three components like Web Crawler who visits the links on the web and updates search engine index periodically, Index which also known as catalog is a collection web pages from which a search engine fetches the result for a user query, Search interface and relevancy software which provides user-friendly environment for searching and also organize the results based on relevancy.

Search Engine works with three components which accept user query, validate it prompts the user with correct or popular spelling. If misspelled, then checks if the query is relevant to other vertical search databases and provide relevant links to a few items of search query along with regular search results, fetches a list of relevant pages for search results and rank such results based on the page content, usage data etc., requests a list of relevant ads to place near the search results.

In this section, we proposed an algorithm to calculate the precision and relative recall and used the same for the case study.

---

#### ALGORITHM:

---

Step 1: Set site\_no=0, userquery\_no=0, searchengine\_no=0, more\_relevant=0, relevant=0, links=0, irrelevant=0,  
 Step 2: Define number of search engines as **s**, number of queries as **n**, number of sites from beginning as **nosb**, total number of links retrieved by all selected search engines as **sum\_links\_searchengines**.  
 Step 3: Read **n**, **s**, and **nosb**.  
 Step 4: Select and open a web browser.  
 Step 5: Compute searchengine\_no = searchengine\_no + 1.  
 Step 6: Compute userquery\_no = userquery\_no + 1.  
 Step 7: Enter user query to search information from related websites.  
 Step 8: Record number of sites retrieved by search engine and assign it to **srse** to calculate Relative Recall.  
 Step 9: Consider a web page link and Add 1 to site\_no.  
 Step 10: If user query appears in link then  
     Compute more\_relevant = more\_relevant + 1  
     else  
     If user query appears in link description then  
         Compute relevant = relevant + 1.  
     else  
     If links appear other than user query then  
         Compute links = links + 1.  
     else  
         Compute irrelevant = irrelevant + 1.  
 Step 11: Repeat Step 9 through Step 11 until site\_no = **nosb**.  
 Step 12: Record the count of more\_relevant, relevant, links, and irrelevant.  
 Step 13: Calculate precision as (more\_relevant\*2 + relevant\*1 + links\*0.5 + irrelevant\*0) / **nosb**.  
 Step 14: Record precision.  
 Step 15: Repeat Step 6 through Step 15 until userquery\_no=n.  
 Step 16: Repeat Step 5 through Step 16 until searchengine\_no = **s**.  
 Step 17: Reset userquery\_no=0, searchengine\_no=0.  
 Step 18: Compute userquery\_no = userquery\_no + 1.  
 Step 19: Set **sum\_links\_searchengines**=0.  
 Step 20: Compute searchengine\_no = searchengine\_no + 1.  
 Step 21: Read **srse**  
 Step 22: Calculate **sum\_links\_searchengines = sum\_links\_searchengines + srse**.  
 Step 23: Calculate Relative Recall by dividing **srse** with **sum\_links\_searchengines**.  
 Step 24: Record Relative Recall.  
 Step 25: Repeat Step 20 through Step 25 until searchengine\_no = **s**.  
 Step 26: Repeat Step 18 through Step 26 until userquery\_no= **n**.

---

Zapmeta, Google, Yahoo and Search were considered to examine the results of precision and relative recall for some selected queries. A total of 10 queries in the technology and research discipline were selected for the study. All the search queries were classified into two categories by the level of search capabilities like simple one-word queries and simple multi-word queries. (See Appendix 1)

In the present study, the search results which are retrieved by 04 search engines are categorized as “more relevant”, “relevant”, “irrelevant” and “links” on the basis of following criteria.

- If the user query appears in web page link then it is categorized as “more relevant (MR)” and given a score of 2.
- If the web page is not closely matched to the user query but consists of some relevant content, then it is categorized as “relevant (R)” and given a score of 1.
- If links appears other than user query for a particular web page link then it was categorized as “links (L)” and given a score of 0.5.
- If the web page is not at all related to the user query then it was categorized as “irrelevant (IR)” and given a score of 0.

This study would measure the relevance of the websites retrieved for each search query. Advanced search options were used for retrieving sites. Only English language pages were searched for each query since the web pages in other languages would be difficult to assess the relevancy. It was specified that the search query must appear in the “title of the web page” or in its short description displayed below the web page link. Since the number of search results retrieved was large, only the first 100 sites were selected for analysis.

#### Precision of Meta Search Engines

Precision is the ratio of the number of relevant documents retrieved and the total number of irrelevant & relevant documents retrieved.

$$\text{Precision} = \frac{\text{Sum of scores of sites retrieved by search engines}}{\text{Total number of sites selected for evaluation}}$$

**Table 1: Precision of Zapmeta for Simple One-Word Queries**

Search Queries	Total Sites Evaluated	MR	R	IR	L	Precision
Q 1.1	100	90	4	6	0	1.84
Q 1.2	100	92	0	8	0	1.84
Q 1.3	100	98	2	0	0	1.98
Q 1.4	100	98	2	0	0	1.98
Q 1.5	100	94	6	0	0	1.94
Total	500	472	14	14	0	1.92
%		94.4	2.8	2.8	0	

Table 1 showed that 94.4% of sites retrieved by Zapmeta were more relevant followed by relevant (2.8%) and irrelevant (2.8%). It was observed that none of the sites are links. The Precision of Zapmeta was calculated using the

Table 5: Precision of Yahoo for Simple One-Word Queries

Search Queries	Total Sites Evaluated	MR	R	IR	L	Precision
Q 1.1	100	96	2	0	2	1.95
Q 1.2	100	96	2	0	2	1.95
Q 1.3	100	96	4	0	0	1.96
Q 1.4	100	96	4	0	0	1.96
Q 1.5	100	94	4	0	2	1.93
Total	500	478	16	0	6	1.95
%		95.6	3.2	0	1.2	

Table 5 showed that 95.6% of sites retrieved by Yahoo were more relevant followed by relevant (3.2%) and irrelevant (2.8%). It was observed that none of the sites are irrelevant and 1.2% of sites are links. The Precision of Yahoo was calculated using the above formulae. The overall Precision of Yahoo was 1.95. The highest Precision was 1.96 for search queries Q 1.3 and Q 1.4 and lowest Precision of 1.93 for search query Q 1.5.

Table 6: Precision of Yahoo for Simple Multi-Word Queries

Search Queries	Total Sites Evaluated	MR	R	IR	L	Precision
Q 2.1	100	86	12	0	2	1.85
Q 2.2	100	90	6	2	2	1.87
Q 2.3	100	74	26	0	0	1.74
Q 2.4	100	100	0	0	0	2
Q 2.5	100	80	20	0	0	1.8
Total	500	430	64	0.4	0.8	1.85
%		86	12.8	0.4	0.8	

Table 6 shows that the search results of Yahoo for simple multi-word queries. From the table it is clear that 86% of sites retrieved are more relevant followed by 12.8% sites are relevant while 0.4% of sites retrieved are irrelevant and 0.8% of sites are links. The overall Precision of Yahoo is 1.85. The highest Precision query is 2 for search queries Q 2.4 and the lowest Precision is 1.74 for search query Q 2.3.

Table 7: Precision of Search for Simple One-Word Queries

Search Queries	Total Sites Evaluated	MR	R	IR	L	Precision
Q 1.1	100	96	2	2	0	1.94
Q 1.2	100	88	2	10	0	1.74
Q 1.3	100	98	2	0	0	1.98
Q 1.4	100	82	10	8	0	1.74
Q 1.5	100	88	2	10	0	1.78
Total	500	452	18	30	0	1.84
%		90.4	3.6	6	0	

Table 7 showed that 90.4% of sites retrieved by Search were more relevant followed by irrelevant (6%) and relevant (3.6%). It was observed that none of the sites are links. The Precision of Search was calculated using the above formulae. The overall Precision of Search was 1.84. The highest Precision was 1.98 for search queries Q 1.3 and lowest Precision is 1.74 for search queries Q 1.2 and Q 1.4.

Table 8 shows that the search results of Search for simple multi-word queries. From the table it is clear that 89.2% of sites retrieved are more relevant followed by 10.8% sites are relevant while none of sites retrieved are irrelevant and links. The overall Precision of Search is 1.89. The highest Precision query is 1.96 for search queries Q 2.1 and the lowest Precision query is 1.78 for search query Q 2.3.

above formulae. The overall Precision of Zapmeta was 1.92. The highest Precision was 1.98 for search queries Q 1.3 and Q 1.4. The lowest Precision was 1.84 for search queries Q 1.1 and Q 1.2.

Table 2: Precision of Zapmeta for Simple Multi-Word Queries

Search Queries	Total Sites Evaluated	MR	R	IR	L	Precision
Q 2.1	100	90	4	6	0	1.76
Q 2.2	100	92	0	8	0	1.94
Q 2.3	100	98	2	0	0	1.68
Q 2.4	100	98	2	0	0	1.96
Q 2.5	100	94	6	0	0	1.8
Total	500	472	14	14	0	
%		94.4	2.8	2.8	0	

Table 2 shows that the search results of Zapmeta for simple multi-word queries. From the table it is clear that 84.8% of sites retrieved are more relevant followed by 13.2% sites are relevant while 2% of sites retrieved are irrelevant. The overall Precision of Zapmeta is 1.83. The highest Precision query is 1.96 for search queries Q 2.4 and the lowest Precision query is 1.76 for search query Q 2.1.

Table 3: Precision of Google for Simple One-Word Queries

Search Queries	Total Sites Evaluated	MR	R	IR	L	Precision
Q 1.1	100	90	4	4	2	1.85
Q 1.2	100	94	4	2	0	1.92
Q 1.3	100	98	2	0	0	1.98
Q 1.4	100	90	8	2	0	1.88
Q 1.5	100	92	4	2	2	1.89
Total	500	464	22	10	4	1.9
%		92.8	4.4	2	0.8	

Table 3 showed that 92.8% of sites retrieved by Google were more relevant followed by relevant (4.4%) and irrelevant (2%). It was also observed that 0.8% of the sites are links. The Precision of Google was calculated using the above formulae. The overall Precision of Google was 1.9. The highest Precision was 1.98 for search queries Q 1.3 and lowest Precision was 1.85 for search query Q 1.1.

Table 4: Precision of Google for Simple Multi-Word Queries

Search Queries	Total Sites Evaluated	MR	R	IR	L	Precision
Q 2.1	100	84	12	2	2	1.81
Q 2.2	100	94	4	0	2	1.93
Q 2.3	100	82	16	0	2	1.81
Q 2.4	100	88	10	0	2	1.87
Q 2.5	100	94	4	0	2	1.93
Total	500	442	46	2	10	1.87
%		88.4	9.2	0.4	2	

Table 4 shows that the search results of Google for simple multi-word queries. From the table it is clear that 88.4% of sites retrieved are more relevant followed by 9.2% sites are relevant while 0.4% of sites retrieved are irrelevant and 2% sites are links. The overall Precision of Google is 1.87. The highest Precision query is 1.93 for search queries Q 2.2 and Q 2.5. The lowest Precision query is 1.81 for search queries Q 2.1 and Q 2.3.

Table 8: Precision of Search for Simple Multi-Word Queries

Search Queries	Total Sites Evaluated	MR	R	IR	L	Precision
Q 2.1	100	96	4	0	0	1.96
Q 2.2	100	88	12	0	0	1.88
Q 2.3	100	78	22	0	0	1.78
Q 2.4	100	96	4	0	0	1.96
Q 2.5	100	88	12	0	0	1.88
Total	500	446	54	0	0	1.89
%		89.2	10.8	0	0	

Table 10: Relative Recall for Simple One-Word Queries

Search Query	Zapmeta		Google		Yahoo		Search	
	Total No. Of Sites	RR	Total No. Of Sites	RR	Total No. Of Sites	RR	Total No. Of Sites	RR
Q 1.1	347000000	0.05	3690000000	0.85	3490000000	0.77	519000000	0.07
Q 1.2	476000000	0.09	662000000	0.13	4470000000	3.65	87500000	0.02
Q 1.3	1619000000	0.6	2000000000	0.86	486000000	0.13	232000000	0.06
Q 1.4	504000000	0.24	1390000000	1.17	467000000	0.22	212000000	0.09
Q 1.5	504000000	0.05	8460000000	2.74	1640000000	0.17	942000000	0.09
Total	6573000000	1.03	49412000000	5.75	41963000000	4.94	6663500000	0.33

Mean Precision of Meta Search Engines

The mean precision of Dogpile, Clusty, Zapmeta, Google and Search are 9.37, 7.13, 9.16, 9.44 and 9.39. Yahoo had the highest mean precision of 9.50 which is just 0.6 higher than Google as shown in the table 9.

Table 9: Mean Precision of Zapmeta, Google, Yahoo and Search

Search Engine	Simple one-word Queries	Simple multi-word Queries	Mean Precision
Zapmeta	9.58	8.74	9.16
Google	9.52	9.35	9.44
Yahoo	9.74	9.26	9.50
Search	9.22	9.56	9.39

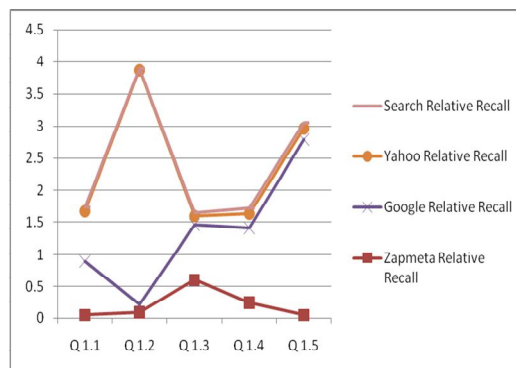


Fig 2: Relative Recall for Simple One-Word Queries

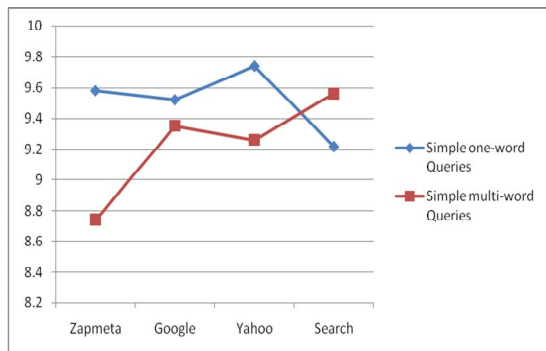


Fig 1: Mean Precision of Zapmeta, Google, Yahoo and Search for the two search queries

Table 11: Relative Recall for Simple Multiple-Word Queries

Search Query	Zapmeta		Google		Yahoo		Search	
	Total No. Of Sites	RR	Total No. Of Sites	RR	Total No. Of Sites	RR	Total No. Of Sites	RR
Q 1.1	122000000	0.05	1860000000	3.62	143000000	0.06	247000000	0.12
Q 1.2	415500000	0.49	550000000	0.78	154000000	0.14	137000000	0.12
Q 1.3	127500000	0.35	205000000	0.7	108000000	0.28	56200000	0.13
Q 1.4	273000000	0.68	249000000	0.59	125000000	0.23	274000000	0.04
Q 1.5	134000000	0.32	224000000	0.68	144000000	0.35	50500000	0.1
Total	4627000000	1.89	22069000000	6.37	3086000000	1.06	2987700000	0.51

Relative Recall of Meta Search Engines

Recall is the ratio of the number of relevant documents retrieved by a search engine, to the total number of relevant documents retrieved by all selected search engines.

$$\text{Relative Recall} = \frac{\text{Total no. of sites retrieved by search engine}}{\text{Sum of sites retrieved by all search engines}}$$

Mean Relative Recall of Meta Search Engines

The mean relative recall of Zapmeta, Yahoo and Search are 1.46, 3 and 0.42. Google had the highest relative recall of 6.06 as shown in the table 6.

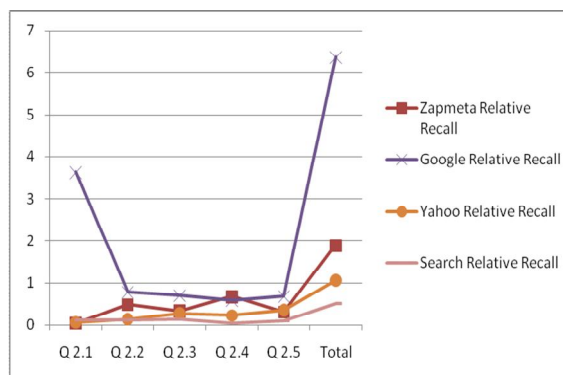


Fig 3: Relative Recall for Simple Multi-Word Queries

Table 12: Mean Relative Recall of Zapmeta, Google, Yahoo and Search

Search Engine	Simple One-Word Queries	Simple Multi-Word Queries	Mean Relative Recall
Zapmeta	1.03	1.89	1.46
Google	5.75	6.37	6.06
Yahoo	4.94	1.06	3
Search	0.33	0.51	0.42

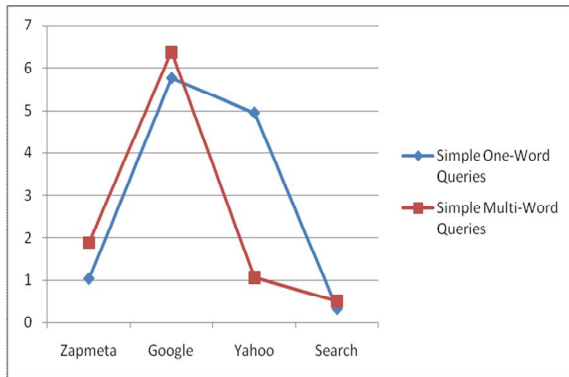


Fig 4: Mean Relative Recall of Zapmeta, Google, Yahoo and search for the two search queries

Table 13: Correlation of Zapmeta

Queries	Simple One-Word (A)	Simple Multi-Word (B)	AA	AB	BB
Q 1.1	1.84	1.76	3.39	3.24	3.10
Q 1.2	1.84	1.94	3.39	3.57	3.76
Q 1.3	1.98	1.68	3.92	3.33	2.82
Q 1.4	1.98	1.96	3.92	3.88	3.84
Q 1.5	1.94	1.8	3.76	3.49	3.24
Total	9.58	9.14	18.38	17.51	16.77

$$\text{Correlative Coefficient } r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

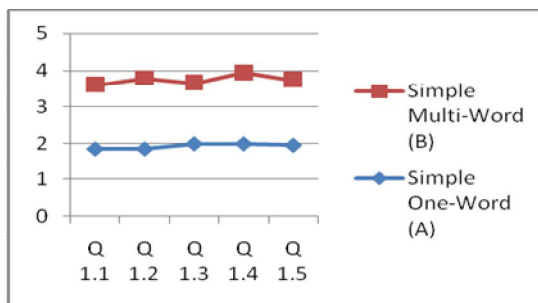


Fig 5: Correlation between Simple One-Word and Simple Multi-Word Queires of Zapmeta

Table 14: Correlation of Google

Queries	Simple One-Word (A)	Simple Multi-Word (B)	AA	AB	BB
Q 1.1	1.85	1.81	3.42	3.35	3.28

Q 1.2	1.92	1.93	3.69	3.71	3.72
Q 1.3	1.98	1.81	3.92	3.58	3.28
Q 1.4	1.88	1.87	3.53	3.52	3.50
Q 1.5	1.89	1.93	3.57	3.65	3.72
Total	9.52	9.35	18.14	17.80	17.50

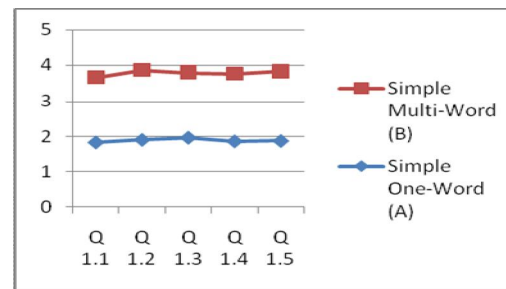


Fig 6: Correlation between Simple One-Word and Simple Multi-Word Queires of Google

Table 15: Correlation of Yahoo

Queries	Simple One-Word (A)	Simple Multi-Word (B)	AA	AB	BB
Q 1.1	1.95	1.85	3.80	3.61	3.42
Q 1.2	1.95	1.87	3.80	3.65	3.50
Q 1.3	1.96	1.74	3.84	3.41	3.03
Q 1.4	1.96	2	3.84	3.92	4.00
Q 1.5	1.93	1.8	3.72	3.47	3.24
Total	9.75	9.26	19.01	18.06	17.2

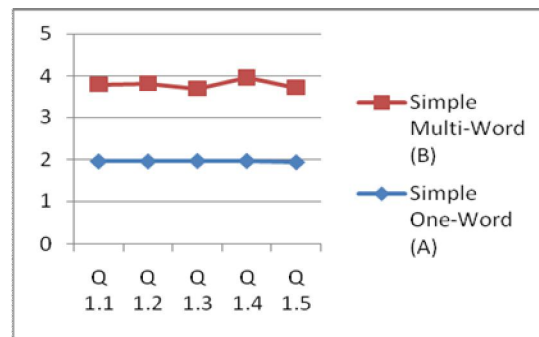
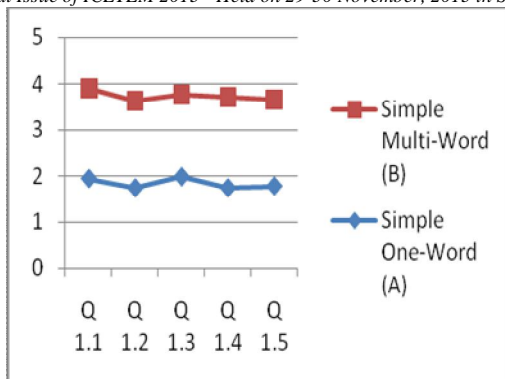


Fig 7: Correlation between Simple One-Word and Simple Multi-Word Queires of Yahoo

Table 16: Correlation of Search

Queries	Simple One-Word (A)	Simple Multi-Word (B)	AA	AB	BB
Q 1.1	1.94	1.96	3.76	3.80	3.84
Q 1.2	1.74	1.88	3.03	3.27	3.53
Q 1.3	1.98	1.78	3.92	3.52	3.17
Q 1.4	1.74	1.96	3.03	3.41	3.84
Q 1.5	1.78	1.88	3.17	3.35	3.53
Total	9.18	9.46	16.91	17.35	17.92



**Fig 8: Correlation between Simple One-Word and Simple Multi-Word Queires of Search**

Correlation Coefficients were calculates as per the above formulae. Correlation between Zapmeta, Google, Yahoo and Search are 0.5, 0, 1.05 and 0.2. Correlation between A and B for Zapmeta are positive and near to 1, so both are strongly correlated. Correlation between A and B of Google is 0, so this is not strongly correlated. Correlation of Yahoo between A and B is also positive and equal to 1, so both are strongly correlated. Correlation of Search between A and B are almost equal to 0, so this is not strongly correlated.

## I. APPENDIX 1: SEARCH QUERIES

### 1. Simple One-Word Queries

- Q 1.1 Database
- Q 1.2 Internet
- Q 1.3 Precision
- Q 1.4 Recall
- Q 1.5 Intranet

### 2. Simple Multi-Word Queries

- Q 2.1 Information Technology
- Q 2.2 Research Methodology
- Q 2.3 Information Retrieval System
- Q 2.4 Search Engines
- Q 2.5 Web Usability

## ACKNOWLEDGEMENT

I would like to thank my partner, Trivikram Mulukutla, for his love, kindness and support he has shown during the research and preparation of this paper. Furthermore I would also like to thank my parents for their endless love and support. I would like to thank all the people who supported me as well for their assistance for this paper.

## CONCLUSION

The present study estimated the precision and relative recall of Zapmeta, Google, Yahoo and Search Meta Search

Engines. The results of study showed that the precision of yahoo was high for simple one-word queries, and Search for simple multi-word queries. Relative recall of Google was high for both simple one-word and simple multi-word queries. It was observed that none of the Meta search engines correlation coefficient is negative and are not correlated.

## REFERENCES

- [1]. Nagaraju Mamillapally, Trivikram Mulukutla – Performance Evaluation of Meta Search Engines from User's Prespective, Proceedings of National Conference on Advances in Signal Processing, Communications and Networking, Vol 1, Issue 1, p.p.1-5, 29<sup>th</sup> & 30<sup>th</sup> August 2013.
- [2]. B.T.Sampath Kumar, J.N. Prakash – —Precision and Relative Recall of Search Engines: A Comparative Study of Google and Yahoo!, Singapore Journal of Library & Information Management, Volume 38, 2009.
- [3]. Tauqeer Ahmed Usmani – —A Comparative Study of Google and Bing Search Engines in Context of Precision and Relative Recall Parameter!, International Journal on Computer Science and Engineering, Vol 4 No.1 January 2012.
- [4]. Clarke, S., & Willett, P. (1997). Estimating the recall performance of search engines. *ASLIB Proceedings*, 49 (7), 184-189.
- [5]. 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*, 33, 127-35.
- [6]. Ding, W., & Marchionini, G. (1996). A Comparative study of the Web search service performance. *Proceedings of the ASIS 1996 Annual Conference*, 33, 136-142.
- [7]. Leighton, H. (1996). Performance of four WWW index services, Lycos, Infoseek, Webcrawler and WWW Worm. Retrieved from <http://www.winona.edu/library/webind.htm>.
- [8]. 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>
- [9]. Wu, G., & Li, J. (1999). Comparing Web search engine performance in searching consumer health information: Evaluation and recommendations. *Bulletin of the Medical Library Association*, 87 (4), 456-461.