CHAPTER – 7

FINDINGS AND CONCLUSION

7.1 Findings

The evaluation of results found a statistical difference in the performance of the five search engines in retrieving web resources. The performance of the search engines has been measured by evaluating the coverage, results relevancy, false links, duplicate links, uniqueness of the results, and the stability of the retrieved results within a time frame. The findings of the research are summarized as per the objectives of the study as mentioned in the chapter one.

In chapter 2, 3, 4 and 5, an exhaustive description has been provided to have a broad overview of different types of search engines. In these chapters it has been discussed about the components of search engines, history and trends of development of search engines, utility and importance of different types of search engines, an overview of most used search engines year over year (YOY), and the search strategies of the five most used search engines on the web. From the discussion it is clear that with the tremendous growth of web information the importance of search engines is increasing day by day. As a result the search engines are also indexing billions of information to incorporate into their databases. With the increasing popularity of search engines and in order to return the relevant results against an input query, the search engines are incorporating different strategies to their search environment. It is clear from the analysis, as mentioned in chapter 4, that the popularity status of
search engines are almost remaining same among the five search engines year over year. In terms of coverage and user statistics Google is still remaining the most popular search engine on the web followed by Yahoo as reported by most of the search engines analysis company.

From the broad analysis of search strategies of the five search engines as mentioned in the chapter 5, it is obvious that the basic concepts of search strategies and searching processes of all the five search engines is remain same, that relates to the coordination of items in order to formulate the actual search statement. The first web information services were based on traditional information retrieval algorithms, which were originally developed for smaller more coherent collections than the web. But due to the web's continued growth, today's web search services are incorporating new techniques to their search strategies to exploiting and extending linkages among web pages. All the five search engines followed different indexing method and the relevance ranking system is based on their owned unique algorithms. Almost all five search engines, as described above, provides basic text search facilities like Boolean search, Proximity search, Phrase search, Truncation, Field-specific search, Numeric data search, Word Stemming, Title search, URL search, Customizable Result Formatting, and Limiting search. There are some specific features which are unique to a particular search engines. Google is unique for its 'Search Within' feature. The 'Page translation' is possible only with Google and Yahoo. The 'Wildcard' feature is available with Yahoo, Ask and AOL. Again, 'Related Search' is possible only with Yahoo and Live.
The search operators vary from one search engine to another. It may be noted that major improvements in the web search environment has been in the design interfaces that allow users to conduct complex search queries without having to remember or use complex search syntax or operators. The five described search engines provides enough room for users to formulate comparatively complex search expressions simply by choosing the appropriate options and entering the query terms in the advanced search option in their search page. All the five search engines provide enough help information for user to complete their search on the screen itself. In addition to those retrieval features that are common to the traditional text retrieval systems, all the five web search services provide some unique search facilities that have been designed specifically to retrieve various available web information resources. These services make the search engines more interesting and luring.

In order to evaluate the performance of search engines in retrieving internet resources the five most used search engines are evaluated in terms of their coverage, relevancy, uniqueness, and stability, which is described in the chapter 6. From the present analysis it is clear that there is significant difference of performance among the evaluated search engines. The analysis reveals that there is significant difference among the search engines in terms of coverage, relevancy, uniqueness, and stability. Although there is difference among the search engines in its retrieval effectiveness, Google occupied the top rank in terms of coverage, relevancy, and uniqueness of it retrieved results during both the study, which were conducted in two different time frames. Google seems to satisfy information needs best, because it delivers the highest numbers of relevant results for the single queries when compared with the
other four engines. So the present study coincides with the survey results reported by different search engines research agency, i.e. Neilson Netranking, ComScore, etc. In terms of the analysis of results stability, it is found that Yahoo has the highest mean stability as compared with the other four search engines. This point also as contradict as it proven the inability of search engines to update their databases frequently. The present study also proves that no search engine is free from its faulty result hits. All the five major search engines retrieved duplicate and false links along with it relevant hits. Although the major search engines are claiming of its uniqueness of its retrieved results, at present no search engine is able to filters the duplicate and false links from its retrieved results.

From the overall analysis of the findings of the matrices we can concludes that the Google has statistically higher rate of performance in retrieving web resources as compared with the other four search engines, as Google has highest mean relevance, highest unique results, maximum coverage, less number of irrelevant hits, and less number of false links. Google is followed by Yahoo in terms of measuring the retrieval performance. The statistical analysis of the derived data reveals that Yahoo occupied the second place in terms of relevant hits, result stability, uniqueness of results, and coverage. The other three search engines, i.e. AOL, Live, and Ask, have not performed satisfactorily as compared with Google and Yahoo. There was also no significant statistical difference among these three search engines. Another conclusion of this evaluation is that, users could, and should, contemplate Yahoo as a real alternative to Google, especially because both engines were able to answer all of the fifty queries with highest mean of relevant hits. Beyond that, Yahoo was found to retrieve the greatest number of hits, less number of bad links, great stability
in retrieving relevant hits, and maximum number of unique hits, after the Google. This indicates at least Yahoo's performance is not worse than Google's performance.

**Hypothesis 1: There is a significant difference in the number of relevant and irrelevant links retrieved by different search engines.**

The study revealed that all the five search engines retrieved different number of relevant hits as well as different number of irrelevant hits in both the study. After analysis the first ten results retrieved by the five search engines against each query, it has been noticed that Google retrieved maximum number of relevant hits in both the study followed by Yahoo. And again, Google retrieved less number of irrelevant hits in both the study followed by AOL. Ask and Live occupied the bottom place in both the study in terms of retrieving relevant hits. The statistical analysis of the findings also revealed that there is a significance difference among the all search engines in retrieving relevant and irrelevant results. The study thus proves the first hypothesis.

**Hypothesis 2: No two-search engines are identical in terms of their coverage and their search results.**

As stated in section 6.6.1 of the chapter 6, it is obvious that Google retrieved the maximum number of results for the fifty tested queries in both the two studies. Google was followed by Yahoo, Ask, and Live. Google's coverage is much more than the other three search engines. Again in terms of uniqueness of the retrieved results, it is found that there is a significant mean difference of unique hits returned by the five search engines in both the study. Therefore the study proves the second
hypothesis, i.e. no two-search engines are identical in terms of their coverage and their search results.

**Hypothesis 3: There is significance fluctuation in retrieved results among top rated search engines.**

After testing the consistency of five search engines it is found that search engine fluctuations in the results sets can no longer be neglected and that search results stability should be considered as a performance measure of web search engines. The present study revealed that there was sudden increase in the number of pages retrieved by three search engines, i.e Google, Yahoo, and Ask, within the interval of one month period. Again the study also revealed that all the five search engines have retrieved different relevant hits in two different studies. The Ask and Live have shown more instability in retrieving relevant results during the study period. The statistical analysis of the finding revealed that there is significant difference among the search engines in terms of the stability in retrieving results, as stated in section 6.5.5 of chapter 6. Therefore, the study proves the third hypothesis.

**Hypothesis 4: There are significant differences among the top rated search engines in their retrieving performance.**

The overall analysis of the findings of different matrices revealed that all the five search engines have different mean result for the matrices. Google has statistically higher rate of performance in retrieving web resources as compared with the other four search engines. It has the highest mean relevance, highest unique results, maximum coverage, less number of irrelevant hits, and less number of false links. The statistical analysis of the derived data reveals that Yahoo occupied the second
place in terms of relevant hits, result stability, uniqueness of results, and coverage. The other three search engines, i.e. AOL, Live, and Ask, have not performed satisfactorily as compared with Google and Yahoo. Although there were differences among these search engines in their search performances, the study found no significant statistical difference among AOL, Ask, and Live. Thus, the present study partially proves the hypothesis.

Given these findings, we can consider that this study will provide important insight into the effectiveness of five major types of search engines and their support in retrieving relevant internet resources. The strengths of the research are that I used five very popular search engines, ensuring that the results would have a broad impact, the use of a large number of simple to complex queries, and finally, I employed a rigorous methodology to limit the possibility of bias in the evaluation.

The present research tried to achieve results as valid and reliable as possible within the given resources. The available knowledge and evaluation experience was used to consider web-specific retrieval peculiarities and evaluation standards in a sufficient manner. The goal was to develop the experimental design in the best possible way, within the given resources. It is assumed that the results are valid and reliable enough to get some helpful insights into the effectiveness of the investigated search engines. The mentioned problem fields show, on the one hand, the constraints of this evaluation and, on the other hand, give some hints on how to improve further investigations.

This study has produced key findings that are important for all Web search engine users and researchers, and the Web industry. The key finding of this large-scale
study is that first results returned by the five major Web search engines included in
this study differ from one another. The study results highlight the fact that different
web search engines, which use different technology to find and present web
information, yield different first page search results. Web search engine's first page
results are primarily unique, meaning the other engines did not return the same result
on the first result page for a given query. The fact that no one web search engine
covers every page on the Internet. The findings suggest that many web technical
and user related characteristics, such as overlap, number of queries entered, etc. are
not dramatically changing over time and further highlights the value of studying web
search trends to gauge the true impact of technological changes.

7.2 Conclusion

The performance and capabilities of web search engines is an important and
significant area of research. Millions of people world wide use Web search engines
every day. In this study it has been tried to evaluate the performance of web search
engines. The performance of search engines has been measured after examining the
first ten results returned by the search engines and checks the consistency between
them. The experiments show a high level of statistically significant result between
the search engines in terms of effectiveness and also in terms of selecting the best
and worst performing search engines. Knowing the most effective web search
engines satisfying the current information need is important both at a personal and
business enterprises level. However, the definition of “the best” changes due to the
changing information need of users and the changing quality and nature of search
engines. For example individuals and organizations may work on the solution of
different problems at different time instances, and search engines may change their
indexing or ranking strategies and their web coverage. Accordingly, for different information needs or with different search technologies, the most effective search engine may be different. Hence, search engines performance needs to be tested and this should be done quite often.

Even after continuous work for several years, web search is still in its infancy and technology around web search will continue to evolve. The work done to date has uncovered strong editorial voices for web search based on unique ways of capturing and ranking search results. Google is different than Yahoo, Yahoo is different than Ask, Ask is different than MSN. And MSN is different than AOL. These differences contradict the notion that all search engines are the same, and searching through one search engine will yield the absolute best results of the web. This study quantifies the similarities and most importantly the differences among the leading single search engines.

Each of the five single source search engines measured has a unique voice and does a good job returning results they deem relevant based on that voice. The differences in indexing the web and ranking results across these engines prevents users from feeling confident that they have found the best results for their search through the use of just one single search engine. There are good search engines, but there is no perfect search engine on the web. This is because intent is subjective to the end user, making it possible for any each search engine to understand every person’s intent correctly each and every time.
The results of this study highlight the fact that the top search engines, Google, Yahoo, MSN, Ask, and AOL, have built and developed proprietary methods for indexing the web and their ranking of keywords driven search results differs greatly. However we can suggest that, by using meta search engines users can reduce the number of search engines they need to consult to one, making it easy to find the best results of the web, and installing confidence that they have performed the most comprehensive search of the web with one click. Through the relationship meta search engines have built with these engines, it is able blend these differences and provides the best results from the top web search engines to the end users. Meta search technology harness the collective content, resources and ranking capabilities of all top search engines and delivers web searchers a more comprehensive results set that brings the best results from top engines to the first results page. Since web content is not static there are barriers for any one engine’s ability to cover the entire web all of the time. As indices change and web content emerges, the metasearch solution will be able to keep better pace with the web as a whole than any single source search engine.

Search-engine problems are connected with each component of the engine’s architecture and each process it performs—search engines can’t update indexes at the same speed at which the Web evolves. Another problem is the quality of the search results. We’ve already looked at their lack of stability, heterogeneity, high linking, and duplication. Today’s search engines are amazingly apt at uncovering the depth of information buried within the billions of pages that encompass the web. Yet, despite their prowess, today’s search engines still produce irrelevant search results for many queries. In recognition of these limitations, search engines are
constantly innovating to make search more relevant. Some are providing a means to personalize search results with shared knowledge, some are experimenting with a new and different results page, and others want to improve relevance with the human touch. These results imply that Web search engine designers are doing a proper job of designing Web interfaces and ranking algorithms that accommodate the searching patterns of their customers.

Currently, the predominant business model for commercial search engines is advertising. The goals of the advertising business model do not always correspond to providing quality search to users. Since it is very difficult even for experts to evaluate search engines, search engine bias is particularly insidious. This type of bias is much more insidious than advertising, because it is not clear who deserves to be there, and who is willing to pay money to be listed. Furthermore, advertising income often provides an incentive to provide poor quality search results. In general, it could be argued from the consumer point of view that the better the search engine is, the fewer advertisements will be needed for the consumer to find what they want. However, the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm.

7.3 Issues for Further Research

Web searching involves not only selection of terms, but also the construction of queries and selection of search engines. So, future work in this area involves research from the perspectives of both the search engine designer and the search engine user. The most important avenues of further research to pursue are those that would provide further insight into the issues of relevance and ranking, since these
are of more concern to most users than coverage. There are several avenues for future investigation in the field of search engines industry. Considering the area of coverage and study made in this work, additional research can be conducted to some new innovative areas. The following are few vital areas which deserve the attention for further research.

- In order to strengthen the results of study a larger set of complex queries and a longer period of time interval be undertaken to reveal more interesting result pattern.

- It is necessary to examine the effect of longer queries with different supportive operators including the negative operators, provided by the search engines to know the impact of operators on result relevancy.

- An obvious direction for future work is in the area of search interface usability. Variations in coverage resulting from use of different interface and search options could also be considered, along with whether or not increases in coverage result in greater relative precision.

- Further studies are needed to gather data on different evaluative measures and to asses their compatibility so that fair comparisons of search engine performance can be made in future.

- And finally, an exciting research area to explore would be to study the effectiveness of the meta search engines in retrieving relevant web resources as compare with the other major single source search engines.