CHAPTER – 2

LITERATURE REVIEW

Over the years most information resources have appeared in electronic form – on CD Rom, remote online databases, and now on the World Wide Web. So there are large varieties of electronic information sources on the web. And to cover all these information sources that are available on the web, various web search tools and web-based information services have developed. In order to help information practitioners and students learn about the variety of electronic information sources vis-à-vis the information retrieval techniques, various literatures have been published by different authors expert on these fields. Various studies have been conducted by researchers, on the performance of search engines since its inception. The existing literature on the Web and Search Engines performance centers on several issues. We have witnessed rapid growth in the search engine market since its inception. Search engines continue to attract large number of web searchers and consistently rank as some of the heavily visited sites in the market in terms of number of visitors. The review of the related literature carried out for the purpose of the study has been divided in to following sections. After classifying documents into particular section, they are cited in chronological order, irrespective of their types.
2.1 World Wide Web

The World Wide Web is a massive collection of pages of information, stored on millions of computers across the worlds that are linked by the internet. The development of web began in 1989 with the works of Tim Berners Lee and his colleagues at CERN. Now web is a huge reservoir of information filled with billions of documents and many new information services are available through the web. The web has brought tremendous changes in the way we create, search for and use information in our day to day lives. A large of numbers of literatures in different forms is available on web and web services. Some of the important documents, which dealt with WWW, are as follows:

"The World Wide Web: a very short personal history" by Tim Berners Lee (2006). In this one page article Tim Berners Lee, the father of WWW, looking back on the development of the web from the point of his personal view. This is one of the most original lectures on www by its creator. Berners Lee briefed the history of World Wide Web and his involvement with the project. He thinks that there is immense potential of the mixture of human and machine working together and communicating through the web in the near future.

"Information sources and searching on the WWW" by G G Chowdhury and S Chowdhury (2001). This book gives an overview of the various information sources and services available on the web, and the corresponding tools and techniques that are required to search and retrieve the information. The book also describes the features of various search and retrieval tools that are available. Also discusses on the trends related to the various information services available on the web.
"The semantic web: opportunities and challenges for next generation web applications" by Shi Yong Lu, Ming Dong and Farchad Fotouhi (2002), described about next generation web, that is Semantic web – where the content of the web is structured in a semantic way so that it is meaningful to computers as well as human. It also reports a survey on recent research on the semantic web. Also discussed the changes that it will bring towards web services, semantic based search engines, semantic based digital libraries.

"Trends in the evolution of the public web", by Edward T O’Neil (2003). In this paper, three key trends in the evolution of the public were examined, based on the annual surveys conducted by OCLC’s Web Characterization Project. It examined the rate of growth of the public web, internationalization of the public web, and search and retrieval effectiveness.

"Beyond Google: the invisible web in the academic library", by Jane Devine and F Egger-Sider (2004). This article analyzes the concept of the invisible web and its implication for academic library. It offers a guide to tools that can be used to mine the invisible web and discusses the benefits of using the invisible web to promote interest in library services.

"The invisible web: an empirical study of cognitive invisibility", by Nigel Ford and Yazdan Mansourian (2006). The purpose of this paper was to report an empirical investigation into conceptions of the “invisible web”. The paper presented a conceptual model that was intended to be a useful reference point for researchers.
wishing to investigate user-based aspects of search failure and the invisible web. A distinction was drawn between technical objective conceptions of the “invisible web” that commonly appear in the literature, and a cognitive subjective conception based on searchers’ perceptions of search failure. Besides these literatures, there are numbers of web sites covering different aspects of World Wide Web.

2.2 Web Search Services

There has been much discussion of Web search tools in both popular and academic literature. Following are some of the most original works on Web search services.

“A history of search engines”, by Wes Sonnenreich and Others (2000). This is a very valuable information available on the web, and published by the John Wiley & Sons Inc. The authors have described the lesson, the evolution of the search engine, from the very beginning of Archie to the meta search, SavvySearch.

“Sink or Swim: Internet search tools & techniques”, by Ross Tyner (2001). In this article, the author has elaborately described about different types of search engine, their searching strategies, and a comparative study of some individual and multi-threaded search engines.

“The extreme searcher’s guide to web search engine”, by Randolph Hock (2001). The book is one of the most comprehensive, authoritative and a just downright useful guide to what goes on under the hood of the major search engines. The book covers all the major chapters in search engine development, like from the history of search engine, how they operates, in-depth examination of major search engine or
directory, and the crucial subject of keeping up with all of the changes in the world of web search.

"A distributed search engine for fresh information retrieval", by N Sato, M Uehara and Y Sakai (2001). In general, a conventional search engine has the problem that the update interval is very long because it is based on centralized architecture, which collects documents by using robots. The authors have developed a distributed Cooperative Search Engine, in order to reduce the update interval. In this paper, they have described the framework of this Cooperative search engines and its efficiency.

"Discovering the representation of a Search Engine", by King-Lup-Liu and others (2002). In this conference paper, the authors have discussed the importance of Metasearch engine to find out internet resources. This paper shows that the needed information can be estimated from an uncooperative search engine with good accuracy. Two pieces of information which permit accurate search engine selection are the number of documents indexed by the search engine and the maximum weight of each terms. In this paper, the authors presented techniques for the estimation of these two pieces of information.

"Web search: how the web has changed information retrieval", by Terrence A Brooks (2003). This paper surveys web technology with an eye on assessing the appropriateness of web pages as hosts for topical metadata. The survey reveals web pages to be both transient and volatile. Also discusses closed web and open web in relation to the use of topical metadata.
"A brief history of search engine promotion on the web", by Gradiva Couzin (2003). Here in this paper the author has elaborated the indexing programs of the earlier search engines and the techniques used by the present day search engines. Also gives a brief description of the trends of the search engine optimization.

"Major search engines and Directories", by Danny Sullivan (2004). In this online published document, the Search Engine Watch has provided a guide to the major search engine and the directory available on the web. This is mainly based on their well-known and well-used by the surfers.

"Delving deep inside the searcher’s mind", by Heather Lloyd-Martin (2004). In this paper, the author presents the discussion of various search engines expert on user-search related topics, like which search engines do searchers prefer and why, how do users use search engines to find information. Also depicts some research results on user’s search behaviour.

"The extreme searcher’s Internet handbook", by Randolph Hock (2004). It is both a compendium of high quality web sites and a treasure trove of tips and tricks for finding information online. The book is an excellent follow-on-to Hock’s previous book. This is an excellent book for people of any skill level, with detailing basics for becoming a serious searcher and how to go about publishing information online.

"A brief History of search engines", by Lee Underwood (2004). In this article, Lee gives a comprehensive exposure of the history of the Internet, WWW and the Search Engine, which wrapped up backgrounds of all the major search engines.
“The web search garage”, by Tara Calishain (2004). Every now and then a book comes along that’s a must-read for every serious searcher and Tara Calishain’s “The web search garage” falls squarely into that category. There are ten principles expounded in the book, illustrated with examples that demonstrate the power of taking a thoughtful approach to search rather than simply punching in keywords and praying for satisfactory results. Most chapters feature both techniques for finding specific type of information using the major search services as well as description of specialized search engines and directories, many of which are invisible web resources.

“Search for better Search Engine” by Amardeep Gupta (2005). It describes the importance of Search engine, chronology of Search Engine, some loopholes of present generation search Engine, the challenges to the Search Engine and the future of Search Engines. Also narrates how present generation search engine overlook information. He remarks that the challenges before us are not information overload but information overlook.

“Domain knowledge, search behaviour, and search effectiveness of engineering and science students: an exploratory study”, by Xiangmin Zhang, Hermina G B and Xiaojun Yuan (2005). This study sought to answer three questions i) would the level of domain knowledge significantly affect the users search behaviour, ii) would the level of domain knowledge affect search effectiveness, iii) what would be the relationship between search behaviour and search effectiveness, This paper reveals that the level knowledge seems to have an effect on search behaviour, but not on
search effectiveness, and search behaviour does not seem to be related to search effectiveness.

"The online ad attack" by The Economist (2005). Describes how search engine's new advertising service made the Internet an even more valuable marketing medium. The combined advertising revenues of Google and Yahoo will rival the combined prime-time ad revenues of America's three big television networks, ABC, CBS & NBC. Also narrates that now big firms are allocating more percent of their marketing budgets to the online advertising.

"Copyrights, Trademarks and Search Engines", by Grant Crowell (2005). Web site owners must be on constant guard with copyrighted and trademarked materials, both to make sure competitors don't "borrow" content and to avoid provoking search engines in legal gray areas. How can web site owners get links to materials infringing their copyrighted materials out of search engines results? What resource do web site owners have if their sites were removed unfairly from search engines results due to an unfair or unfounded accusation of copyright infringement? Do web site owners have trademarks protection rights and responsibilities in search engine advertisement? All these issues and some recent rulings and existing search engines policies were explored in this article.

"Google power: unleash the full potential of Google", by Chris Sherman (2005). This is a book on how to get most out of the Google search engine. It gives a regular introduction to how the search engines works, the Google interfaces and advanced
search syntax, elaborate study on search operators and an interesting chapter on the additional services provided by Google.

"Understanding Search Engine Patents", by Christine Churchill (2005). This is a special report from the Search Engine Strategies conference held on August 8-11, 2005, San Jose, California. Search related patents provide insight into what's going on in search engine algorithms, and a search marketer who understands these "rules of ranking" are better positioned to win top position in search results. In the reports, experts from different disciplines offered a well rounded interpretation of the recent proliferation of search related patents.

"The Search: how Google rewrite the rules of business and transformed our culture", by John Battelle (2005). This is an ambitious work on search engine industry. The book is really a much broader look at both the history of the web search industry and the profound effects and changes it is having on our social lives.

"What cooking in search engine labs", by Chris Sherman (2005). The article discussed the research and development efforts going on at the major search engines. It gives an extensive list of publications posted by different search engine company experts.

"Next-generation networks: the evolution of web search", by David M Ewalt (2005). David has discussed the problems with the present generation search tools. He pointed out the importance of semantic web in this regards. He also discussed that
how the semantic web would allow search engines to act more intelligently, and making it easier to find specific things.

“Yahoo & MSN closing the Google gap”, by Chris Sherman (2005). In this article, Chris reported the results of a study conducted by Vividence for the search engine industry, based on research with 2000 consumers as they interact with AskJeeves, Google, Lycos, MSN Search, and Yahoo Search.

“A short and easy search engine tutorial”, by Per Koch and Susanne Koch (2005). This tutorial covers all the essential aspects of web searching and search engines. It includes different types of web search tools, how to choose these tools, and various tips and recommendation for internet searching.

“Survey: searchers are confident, satisfied & clueless”, by Chris Sherman (2005). This article reports the findings of a survey conducted by the Pew Internet & American Life project about the self perception of search engine users, which was based on observation and questioning of people as they searched.

“The expanding role of search”, by Paul J Bruemmer (2006). As 2005 ends, it was definitely a Google year, although rivals Yahoo and MSN also made moves. Many new technologies are introduced, expanding the role of search. This year also saw the advent of consumer control, largely made possible due to the role of search in RSS and blogging, as well as manual emphasize on behavioral targeting and personalized search. With multimedia search and local mobile now widely available, search is not only ubiquitous, its simplifying life for all of us.

“Multilingual Search – site covering search engines world wide” by Pandia Search Engine News (2006). This paper discussed about the search engine market outside North America. Search engine in other than English language. It also discussed about “Multilingual search” which gather information on search sites that speaks a different language that English. It is a reference point on search engines and internet statistics worldwide for marketer working globally. It also discussed search engine development worldwide, Asia, Australia and outskirts of Europe.

“The Multimedia Search Engine Quaero, Europe’s answer to Google”, by Lars Vage (2006). Discussed on the progress of the European search engine Quaero, a multimedia search engine, developed in French announced by the French president in last summer in 2005. This search engine utilizes techniques for recognizing, transcribing, indexing and automatic translation of audiovisual documents and it will operate in several languages. Several companies and research institutes are involved in this project. It has more advanced technology then Yahoo and Google.
"How Search Engine ranked web pages", by D Sullivan (2006). The crawler based search engine follow a set of rules. Known as an algorithm, to determining the relevancy, when confronted with hundreds of millions of web pages to sort through. However, how a particular search engine's algorithm works is a closely kept trade secret. In this article Sullivan gives some general rules follow by all major search services.

"How Search Engine works", by D Sullivan (2006). In this article the author writes a brief description of different search engines, their main parts and the job behind each part. It has the links to other related articles on this issue.

"The Infinite Library", by Wade Roush (2006). In this article, it has been discussed the search engine giant Google’s plan to digitize the millions of print books into searchable Web Pages. Discussed views of experts in this regards. Also discussed the direct and indirect impact of this ambitious plan on library and library professionals. It has also pointed out the probable function of a future library in the digitized knowledge.

2.3 Search Engine Evaluation
For locating information on the Web, search engines and directories are the most frequently used Web portals. Several sources report that more than 80% of web visitors use a search engines as a starting point. How good are these search engines at locating relevant information? One may need to evaluate search engines often due to either the changing needs of users or the dynamic nature of search engines, like, changing of total coverage and the ranking technology of search engines. There have
been an increasing number of studies evaluating search engines with many of these focusing on establishing a standard set of metrics generally in line with those for general searching system evaluation. These metrics can be grouped into categories of relevance evaluations, ranking, stability of ranks, uniqueness and coverage. These articles are valuable for further study of available web search services.

**Search Engine Coverage Evaluation**

"Search Engine results over time: a case study on search engine stability", by Judit Bar-Ilan (1998). This paper examines the performance of search engines over time. Search engines lose information, relevant URLs that were retrieved at a given time by a certain search engine, were not retrieved by the same search engine at a later time, although they continued to exist and to be relevant. All these aspects of the performance of search engines have been studied in this article.

"Accessibility of information on the web", by S Lawrence and C L Giles (2000). Here authors discussed about the importance of search engines to retrieve web information, the indexable information covered by the major search engines, and the performance of six major search engines based on the responses of the search engines to real-world queries.

"Coverage, Relevance, and Ranking: The Impact of Query Operators on Web Search Engine Results", by C M Eastman and B J Jansen (2003). It is often assumed that use of query operators improves the effectiveness of web searching. It this article the authors have tested this assumption by examining the effect of query operators on the performance of three major search engines. The use of most query operators had no
significant effect on coverage, relative precision, or ranking, although the effect varied depending on the search engine.

"How are we searching the World Wide Web? A comparison of nine search engine transaction logs", by Bernard J. Jansen, Amanda Spink (2006). This paper reports results from research that examines characteristics and changes in Web searching from nine studies of five Web search engines based in the US and Europe. According to this study, (1) users are viewing fewer result pages, (2) searchers on US-based Web search engines use more query operators than searchers on European-based search engines, (3) there are statistically significant differences in the use of Boolean operators and result pages viewed, and (4) one cannot necessarily apply results from studies of one particular Web search engine to another Web search engine. Also discussed the implications of the findings for the development of Web search engines and design of online content.

Search Engine Relevance Evaluation

"Search Engines for the World Wide Web: a comparative study and evaluation methodology", by H Chu and M Rosenthal (1996). The authors used three search engines (AltaVista, Excite, and Lycos) and 10 queries to examine precision with a three level relevance score for the top 10 links. The authors also proposed a methodology for evaluating web search engines.

"Estimating the recall performance of web search engines", by S Clarke and P Willett (1997). In this paper, the authors examined recall value of AltaVista, Excite, and Lycos by using a pooling technique.
“First 20 precision among World Wide Web search services (search engines)”, by H V Leighton and J Srivastava (1999). In this study five search engines, Altavista, Excite, HotBot, Infoseek and Lycos were compared for precision on the first 20 results returned for 15 queries, adding weight for ranking effectiveness. The researchers determined relevance themselves, after cloaking the identity of the search engine that retrieved the links. The analysis shows that AltaVista, Excite and Infoseek are the top three services with their relative rank changing depending on how one operationally defines the concept of relevance.

“Finding information on the World Wide Web: the retrieval effectiveness of search engines”, by M Gordon and P Pathak (1999). In this research paper, the authors evaluated eight search engines (AltaVista, Excite, Infoseek, Open text, HotBot, Lycos, Magellan, and Yahoo) using 33 queries and top 200 links. The researchers used four level relevance judgments and one evaluator. In an interesting twist for search engine evaluation, the authors had expert searchers translate the information need into a series of queries, selecting the optimal query over a set of time period.

“Precision among World Wide Web search services: Altavista, Excite, Hotbot, Infoseek and Lycos”, by H V Leighton and J Srivastava (1999). In this study five search engines were compared for precision on the first twenty results returned for fifteen queries. The study shows that AltaVista, Excite and Infoseek are the top three services with their relative rank changing depending on how one interpreted the concept of “relevant”.

29
The value of delivering relevant records in response to a query has been assumed by information retrieval research paradigms otherwise differentiated (the cognitive and the physical). An enhanced capacity for informed choice is advocated as an alternative principle for system evaluation and design. This broadly corresponds to: the exploratory capability discussed in recent information retrieval research; the value of discriminatory power in classification and indexing; Giambattista Vico’s critique of the unproductivity of Aristotelian methods of categorisation as routes to new knowledge; and, most significantly, to ordinary discourse conceptions of the value of information retrieval systems. Finally, the substitution of the principle of enhanced choice exemplifies the development of a true science, in which previous paradigms are absorbed into new as special cases.

“Selected results from a large study of web searching: the Excite study”, by Amand Spink and Jack L Xu (2000). This paper reports selected findings from an ongoing series of studies analyzing large-scale-data sets containing queries posed by Excite users. The findings presented report on Queries length and frequency, Boolean Queries, Query reformation, Phrase searching, Search term distribution, Relevance feedback, Sexually related searching, Image queries and Multi-lingual aspects.

“Search Engines Evaluation Using Precision and Document-Overlap Measurements at 10-50 Cutoff Points”, by A Ismail, T M T Sembok, and H B Zaman (2000). In this article the precision measurement has been used to evaluate ten major search engines using ten queries at cutoff points of 10, 20, 30, 40, and 50. The article also
introduced an overlap measurement to determine the commonality of documents between the hits lists of various search engines.

"Measuring search engine quality", by D Hawking and Others (2001). Hawking and his team evaluated the effectiveness of 20 search engines using TREC. They pooled the top 20 links from each search engine and presented them to one evaluator. The researchers examined various performance measures, including precision at various document cut-off values, mean reciprocal rank of first relevant document, and TREC-style average precision.

"Form and function: the impact of query term and operator usage on web search results", by Wendy Lucas and Heikki Topi (2001). Conventional wisdom holds that queries to information retrieval systems will yield more relevant results if they contain multiple topic related terms and use Boolean and phrase operators enhance interpretation. In this study, search engines users formulated queries and expert formulated queries were submitted to the user-specified search engines and provide a basis for comparing relevancy rating across search engines. The results shows that the difference in the number of terms between expert and non expert searches, the percentage of matching terms between those searches, and the erroneous use of non supported operators in non expert searches explain most of the variation in the relevance of search engines. These findings highlight the used for designing search engines interfaces that provide greater support in the areas of term selection and operator usage.
"Boolean Searching", by Danny Sullivan (2001). This page covers how Boolean commands are implemented at the major search services. Also describes the basic functionality of Boolean commands and how to use them in different search engines.

"Search Assistance Features", by Danny Sullivan (2001). The major search engines have many features meant to assist beginning searchers, and even professional researchers may find these aids helpful. This article summarizes and illustrates some of the major search assistance features that are available.

"On caching search engine query results", by E P Markatos (2001). In this paper the author explore the problem of caching of search engine query results in order to reduce the computing and input/output requirements needed to support the functionality of a search engine. The study compare the effectiveness of static and dynamic caching and conclude that although dynamic caching can use large caches more effectively, static caching can perform better for small caches.

"In search of the perfect search engine", by Judy Salpeter (2001). In this paper the author has presented some criteria to consider in comparing the search tools, to discover how they measured up to each of them.

"The parallel evolution of search engines and digital libraries: their convergence to the Mega-Portal", by U Hanani and A J Frank (2001). There are many and varied search engines on the internet but it is still hard to locate and concentrate only on materials relevant to a specific task. Digital libraries could better provide such services on the web. By comparing them here and analyzing their characteristics, the
authors discovered that they actually share much in common. Finally, the authors noted the expected incorporation of intelligent techniques and knowledge management in future generation search engines and digital libraries and the expected convergence of their interfaces and structures in the fifth generation the mega portal.

"The role of individual differences in internet searching: an empirical study", by N. Ford, D. Miller and N. Moss (2001). This article reports the result of a study of the role of individual differences in internet searching. The dimensions of individual differences forming the focus of the research consisted of: cognitive styles; levels of prior experience; internet perceptions; study approaches; age; and gender. Factor analysis and multiple regressions revealed interesting differences. retrieval effectiveness.

"Precision evaluation of search engines", by Y. Shang and L. Longzhuang (2002). Here authors evaluated six popular search engines (Altavista, Fast, Google, Go. iWon, and NorthernLight), with 3000 queries from two domains, using a generally automatic test design. The researchers computed relevance scores using three difference relevance algorithms and statistical comparisons of the ranking.

"Methods for measuring search engines performance over time", By Judit Bar-Ilan (2002). This study introduces methods for evaluating search engines performance over a time period. Search measures are defined, which as a whole described search engine’s functionality over time. The set of measures introduced here may serve as a guideline for the search engines for testing and improving their functionality.
“Automatic evaluation of world wide web search services”, by Abdur Chowdhury and Ian Soboroff (2002). In this article author has presented a method for comparing search engines automatically based on how they rank known item search results. The study has been conducted by comparing five well known search engines, i.e. Lycos, Netscape, Fast, Hotbot, and Google.

“In search of the relevancy figure”, by Danny Sullivan (2002). While relevancy is the most important feature a search engine can offer, there sadly remains no widely accepted measure of how relevant the different search engines are. Sullivan has pointed out different ways to test search engine depending on a particular situation.

“Web search relevance test” by Inktomi Corporation (2003). The Veritest.com had evaluated the relevancy of Google, Wisenut, Fast, Teoma and AltaVista, and compared the top 10 links using 100 randomly selected queries from the Inktomi engine transaction log. Three independent evaluators judged relevancy. The researchers conducted no statistical analysis comparing search engines.

“Measuring user perceptions of web site reputation”, by E G Toms and AR Taves (2003). The study has compared a search tool, TOPIC, with three other widely used tools that retrieve information from the web. The results of this research include insight into the factors that web users consider in formulating perceptions of web site reputation, and insight into which search tools are output able sites for web users.

"Effectiveness evaluation and comparison of web search engines and meta-search engines", by S L Wu and J Y Li (2004). In this research article the authors had evaluated the effectiveness of four search engines (Google, AllTheWeb, HotBot, and AltaVista) and four meta search engines (MetaCrawler, ProFusion, MetaFind and MetaEUREKA). Experimental results showed that on average the performances of selected meta search engines and search engines are very close, although there were no statistical tests evaluating search engine differences performed.

"Automatic performance evaluation of Web Search Engines", by Fazil Can, Rabia, and A B Seydik (2004). In this study, the authors have introduced automatic web search engine evaluation method as an efficient and effective assessment tool of such systems. The experiments based on eight web search engines, 25 queries, and binary user relevance judgments show that the method provides results consistent with human-based evaluation.

"Evaluation of three German Search Engine: Altavista.de, Google.de and Lycos.de", by Joachim Griesbaum (2004). The goal of this study was to investigate the retrieval effectiveness of three popular search engines. For this purpose the three search engines were compares with each other in terms of the precision of their top
twenty results. The basic findings were, i) Google reached the best result marks. There were no significant differences between Google & Lycos. In term of top twenty precision the experiment showed similar outcomes to the preceding retrieval test in the year 2002. It shows that the gap between the engines is closing now. ii) There are big deviations between the relevance assignments based on the judgments of the results themselves and those based on the judgments of the result descriptions on the search engine result pages.

"New measurement for search engine evaluation proposed and tested", Liwen Vaughan (2004). A set of measurement is proposed for evaluating web search engine performance and to evaluate search engine’s stability, an issue unique to web information retrieval systems. These new measurement were tested by using three search engines, Google, AltaVista and Teoma and it was found that the proposed measurement are able to distinguish search engines performance very well.

"Towards next generation web information retrieval", by Wei-Ying Ma and others (2004). In this paper, the authors reviewed the current status of search engines, and then presented the recent works on building next generation web search technologies. Also discussed about, how to extract data records from web pages using vision-based approach. Also discussed the goal of continually advancing web search to next level by applying data mining, machine learning, and knowledge discovery techniques into the process of information analysis, organization, retrieval, and visualization.
“A subjective measure of web search quality”, by M M S Beg (2005). The paper has outlined a procedure for assessing the quality of search results obtained through several search engines. It gives emphasis to demonstrate the procedure of quality measurement than to carry out the actual performance measurement of these search engines. It also measures the satisfaction a user gets when presented with the search results.

“Searching the web: operator assistance required”, by Heikki Topi and Wendy Lucas (2005). This study examines the effect of the search interface and satisfaction. The study reveals that the assisted search term had a significant positive effect on performance, satisfaction, and confidence. Promoting the use of advanced search features is therefore in the best interest of both web search providers and users. The study reports that, coupling the assisted interface with Boolean training was no more effective than either treatment alone.

“Expectations versus reality – Search engine features needed for Web research”, by Judit Bar-Ilan (2005). The major search tools are commercial and are oriented towards the "average" user and not towards the Web researcher, and therefore are unable to meet all the requests. Here the author describes some features for the ideal search engine, explain the need for the specific features and examine whether the currently existing major search engines can at least partially fulfill the requirements of the ultimate search tool.

“Search engines evaluation”, by R Kumar, P K Suri, and R K Chauhan (2005). In this study a metric ‘Ranked Precision’ is proposed to evaluate the performance of search engines. Also discussed the existing methods available to measure the search
engines. In order to identify the limitation of precision, two aspects are taken into consideration, one is, behaviour of search engine users and keyword spamming.

"A study of results overlap and uniqueness among major web search engines", by A Spink, B J Jansen, C Blakely, and S Koshman (2006). This paper reports the results of a major study examining the overlap among results retrieved by multiple web search engines for a large set of more than 10,000 queries. The main goal of the study was to conduct a large-scale study to measure the overlap of search results on the first result page across the four most popular web search engines, i.e. MSN Search, Google, Yahoo and Ask Jeeves.

"The effectiveness of web search engines for retrieving relevant ecommerce links", by B J Jansen and P R Molina (2006). The purpose of this research paper was to evaluate the effectiveness of search engines in the retrieval of relevant ecommerce links. The study examines the effectiveness of five different types of search engines in response to ecommerce queries by comparing the engines quality of ecommerce links using topical relevancy ratings.

"The influence of task and gender on search and evaluation behaviour using Google", by Lori Lorigo and others (2006). In this paper the authors compared the search behaviour variability with respect to different classes of users and different classes of search tasks to reveal whether user models or task models may be greater predictors of behaviour. The study discovered that gender and task significantly influence different kinds of search behaviours.
"A decision-theoretic approach to the evaluation of information retrieval systems", by Ye Diana Wang and Guisseppi Forgionne (2006). This research paper attempts to narrow the gap by providing a comprehensive and integrated multiple criteria decision-theoretic approach for the evaluation of IR systems. The novelty of this approach lies in the focus on the user aspect and the application of decision-making theories in the IR field.

Search Engine Ranking Evaluation

Ranking has been shown to improve user satisfaction with information retrieval systems. There has been limited research in this area of search engine. The result ranking has more importance in a web information retrieval system.

"An investigation into the use of simple queries on web information retrieval systems", by B J Jansen (2000). In this article Jansen compared the links among several search engines, noting that regardless of query formulation, there was an approximately 60% overlap on search engines.

"Ranking retrieval systems without relevant judgments", by I Soboroff, C Nicholas, and P Cahan (2001). Authors have suggested an automatic method that maps queries with a random set of documents.

"Automatic evaluation of World Wide Web search services", by A Chowdhury and I Soboroff (2002). Here, the authors used five search engines (Lycos, Netscape, Fast, Google, HotBot), and compared their ranking performances automatically, finding that the performance for most search engines are statistically equivalent. The
researchers used 2000 queries taken from AOL transaction logs, pairing these queries with known web documents.

"Automatic performance evaluation of Web search engines pages", by N Can and Sevdik (2004) proposed an automatic Web search engine evaluation method as an efficient and effective assessment tool of such systems. Using eight Web search engines (AllTheWeb, AltaVista, HotBot, InfoSeek, Lycos, MSN, Netscape, and Yahoo!), 25 queries, and the top 20 links, the researchers used binary user relevance judgments, determined by the information requester. Using Pearson’s r correlation, the researchers report that their method provided results consistent with human-based evaluations.

"Different Engines, different results", by Dogpile.com (2005). The study measure the overlap and ranking differences of the leading web search engines in order to gauge the benefits of using a met search engine to search the web. The study highlights the facts that there are vast differences between the four most popular search engines, i.e. Google, Yahoo, MSN and AskJeeves. It also reveal that there is compelling evidence that web searching are not always finding what they are looking for with their search engine.

"Methods for evaluating dynamic changes in search engine rankings: a case study" by Judit Bar-Ilan, Mark Levene and Mazlita Mat-Hassan (2006). The papers compare rankings of the top-ten results of the search engines Google and AlltheWeb on ten identical queries over a period of three weeks. In order to assess the changes in the rankings, three measures were computed for each data collection point and
each search engine. The findings in the paper show that the rankings of AlltheWeb were highly stable over each period, while the rankings of Google underwent constant yet minor changes, with occasional major ones. The paper shows that because of the abundance of information on the web, ranking search results is of extreme importance. The paper compares several measures for computing the similarity between rankings of search tools, and shows that none of the measures is fully satisfactory as a standalone measure. It also demonstrates the apparent differences in the ranking algorithms of two widely used search engines.

Search Engine Result Stability Evaluation

There is a growing interest in the stability of search engine results, and stability being the re-occurrence of web documents within search engine results to the same query over time. Following are some of the literature available on evaluation of search engine result stability.

"Search Engine results over time: a case study on search engine stability", by Judit Bar-Ilan (1998). This paper examines the performance of search engines overtime. Search engines loose information, relevant URLs that were retrieved at a given time by a certain search engine, were not retrieved by the same search engine at a later time, although they continued to exist and to be relevant. All these aspect of the performance of search engines has been studied in this article.

"The responsiveness of search engine indexes", by Mike Thelwall (2001). This paper reports on an experiment to investigate the effect of link count on the indexing of 1000 sites in three search portals over a period of seven months. It was found that, although all engines added sites during the period of the survey, only Google
showed evidence of being very responsive to the existence of links on the test site, whereas AltaVista's results were very stable over time.

"Internet search engines: fluctuations in document accessibility", by W. Mettrop and P. Nieuwenhuysen (2001) tested the consistency of links from 13 search engines, concluding that content stability should be considered as a performance measure for web search engines.

"Internet search engines: fluctuation in document accessibility", by W Mettrop and P Nieuwenhuysen (2001). An empirical investigation of the consistency of retrieval through Internet search engines is reported. Thirteen engines are evaluated: AltaVista, EuroFerret, Excite, HotBot, InfoSeek, Lycos, MSN, NorthernLight, Snap, WebCrawler and three national Dutch engines: Ilse, Search.nl and Vindex. The focus is on characteristics related to size: the degree of consistency to which an engine retrieves documents. The authors observed and identified three types of fluctuations in the result sets of several kinds of searches, many of them significant. These should be taken into account by users who apply an Internet search engine, for instance to retrieve as many relevant documents as possible, or to retrieve a document that was already found in a previous search, or to perform scientometric/bibliometric measurements.

"Methods for assessing search engine performance over time", by J Bar-Ilan (2002) defined several metrics for evaluating search engines functionality over time. He used six search engines (AltaVista, Exite, HotBot, Infoseek, NortherLight, Lycos) and one query during a 6 month time period. The researchers reported that Exite was the least stable of the search engines.
"Assessing bias in search engines", by A. Mowshowitz and A. Kawaguchi (2002) measured the performance of 12 search engines with 12 queries and the top 20 links, using the overlap of URLs of the matching pages. The researchers stated that Northern Light was consistently the most biased search engine in terms of stability of links.

In addition to the above-mentioned articles a few bibliographies have also been published relating to various aspects of design and evaluation criteria for web search services. Even most of the sources of information on this research area are also available on Internet. Presently, different search engine expert has developed a numbers of dedicated web sites. Amongst them the most popular are,

http://www.searchenginewatch.com
http://www.searchengineland.com
http://www.searchenginejournal.com

There are also some journals which are regularly publishing works on the subjects: Design, Development, and Evaluation of web search services; Comparative study on Search Engines; Relevancy measurement of Search Engine; Trends on web Search Service. The list of the selected journals is attached as Annexure IV.

There are also surveys and studies, conducted to ascertain the evaluation of search engines. These studies also produced considerable amount of literatures. In sum, previous studies on web searching provide a multifaceted picture of web searching, and they identify a number of cognitive, affective and situational factors influencing
the search process. These studies suggest that web searching is a dynamic and context-bound activity based on the combined use of browsing links and using search engines. Web searching seems to be strongly guided by clues provided by the search environment. In most cases, the judgment of the relevance of information being found is based on the assessment of what is "good enough", and various kind of decision rules direct the search process. In these studies, however, insufficient attention is paid to the nature of problems that interrupt or halt search activities during search sessions. Similarly, attention is seldom paid to the ways in which searchers try to overcome discontinuities during search processes. The literature review presented so far shows that there is a sustainable and significant literature on web search services in general. These literatures will be pertinent and relevant for carrying out a study on a particular field. In other words, these literatures can be applied for studying a particular web search services. But it is quite clear that there is a gap in the literature pertaining to specific comparative study on the coverage, relevancy, uniqueness and performance of the major web search engines. The present study has been undertaken to fill up this gap. Overall, previous studies demonstrated various loopholes having with the most rated search engines on the web.