CHAPTER 2

LITERATURE SURVEY

2.1 GENERAL WEB SEARCH

General web search is the process of searching for information in web pages / documents stored in web servers based on previously compiled and indexed information in large databases. Web search can be conducted by any one of the following three approaches: (1) using search engines, (2) using web directories, (3) browsing the web. However, general web search do not take into account the users interests/preferences and previous search history while presenting the results to the user. The pages returned by any search system are ranked based on various factors like keywords, hyperlink structures, and context.

2.1.1 Keyword based Web Search

The users construct query with keywords and submit to search engines. Broad keywords or shorter keywords will often result in a large number of hits, many of them irrelevant. Conversely, using the appropriate keyword(s) will result in high degree of relevance in the returned set, thereby resulting in a highly effective search. In order to take advantage of currently available search engines to return user-specific relevant results, an intelligently crafted list of keywords and phrases needs to be formed. To develop such a list, it is necessary to have sufficient knowledge of the domain
of context. Unless the user is a subject expert, finding the relevant and related terms of a domain may be difficult or even erroneous.


Search templates (Fang and Salvendy 1999) assist users in formulating better search queries. A template consists of predefined keywords that help users select their search terms. The keywords can be used directly, or can help users formulate their own queries. Each template should be organized as a hierarchy of predefined keywords that could help to restrict the users’ initial search sets, and improve the relevance of the returned hits. Keyword based web search return results that do not take into account context of search, which is more important from user perspective.

2.1.2 Context based Web Search

In order to gain contextual knowledge, ontologies, defined as ‘specification of conceptualization,’ can be used to discover relevant information about the domain of interest (Sowa 1999). By navigating through
ontologies, users are assisted in finding a relevant set of key terms that will aid the search engines in narrowing, widening, or refocusing a web search.

Several kinds of search limitation mechanisms have been supported by most popular search engines in order to aid users to select the target results from several thousands of result pages. One can use Yahoo! (2009) to search only one category at a time. InfoSeek (1994) and Lycos (2009) support ‘Search within results’ and ‘Find similar pages’ functionalities. Few search engines support fielded search mechanisms, such as searching only in links, only in titles etc. WebGlimpse (2009) allows search to be limited to the neighborhood of a current page (Note: Neighbourhood is defined as all pages within a certain distance following outgoing links).

Another approach to enhance web search is to make use of additional structural information. Link analysis is a well known method associated with structural information. Many popular search engines utilize structural information like in/out links to a page for computing the page scores.

2.1.3 Hyperlink based Web Search

In addition to traditional text matching techniques, modern web search engines employ the importance scores of pages for ranking search results. The most famous example is the PageRank algorithm (Brin and Lawrence 1998), which is the basis for all web search tools of Google. By utilizing the linkage structure of the web, PageRank computes the corresponding importance score for each page. These importance scores will affect the final ranking of search results. Structural information has been so far used for enhancing relevance judgments and ranking web pages. Google and Clever (2009) use weighted link popularity as a primary criteria
in their ranking mechanism. The link structure in the web is dynamic, i.e., pages present today may not be available tomorrow and many new pages may be introduced.

The Hyperlink Induced Topic Selection (HITS) algorithm (Kleinberg 1999) is a well known approach in information retrieval which covers the dynamism in link structure. According to HITS algorithm a node that has many outgoing links is a good hub, whereas a node that has many incoming links is a good authority. The number of links pointing to and pointing out of a node determines the node’s authoritativeness and hubness, respectively. However, HITS does not account for the freshness of a web page. Carriere and Kazman (1997) use the number of neighbors (without regard to the directions of links) of a page as a method of ranking pages.

Several link-based similarity functions have been suggested over the web graph. To exploit the information in multi-step neighborhoods, SimRank (Jeh and Widom 2003) and the Companion (Dean and Henzinger 1999) algorithms are introduced by adapting link-based ranking schemes. All the above-mentioned techniques give least preference to user’s search context or their interests. Any form of web search (general information search or product search) will be effective only when it takes into account user’s preferences/interests.

### 2.2 PERSONALIZED WEB SEARCH

Web search that incorporates user’s interests/preferences is known as personalized web search. Personalization becomes a popular remedy to customize the Web environment for users. Personalization can be of two types: context oriented and individual oriented. Context includes factors of
personalization like the nature of information available, the information currently being examined, the applications in use, when, and so on. Individual oriented search encompasses other elements of personalization like user’s goals, prior and tacit knowledge, and prior information seeking behaviors. These factors are more generic and are applicable to any Personalized Information Retrieval system. However, there are many other factors of personalization that can significantly facilitate web search. The factors and their importance are discussed in the following sub section.

2.2.1 Factors affecting Web Search Personalization

The factors that affect web search personalization can be broadly divided into two categories: 1) Spatial factors and 2) Temporal factors.

1. Spatial Factors
   a. Search queries used
   b. Pages Visited
   c. Semantics between search queries and visited pages
   d. Order/sequence of page access
   e. Browsing behaviors/actions
   f. Context of search
   g. User interests
   h. Relation between context of search and the information currently being examined

2. Temporal Factors
   a. Page-view time
   b. Query usage time
   c. User’s shift in interests
Page hit-count, which indicates frequency of the page visits during a session, has been traditionally considered as an informative indicator of the user preferences. In addition, order or sequence of page accesses is recently identified as an important piece of information. In addition to spatial features, temporal features such as page view time are also of significant concern, especially in the context of web personalization applications. The context of search can be derived from the terms used in a search query. The individual search behaviors like the pages visited, order of visit and actions performed on a visited page can be used to confirm the context of search derived from the search query. Utilization of all the above mentioned factors of personalization will produce an effective personalization. It is this focus on the user and their context within the application of search that makes personalized search a compelling area to explore.

With the cost of running a large scale search engine already very high, it is likely that such a full-scale personalization is currently too expensive for the major web search engines. Most major search engines do not even provide an alerting service that notifies users about new pages matching specific queries.

Though personalization seems a simple concept, it is difficult to implement. By its very nature, personalization means different things to different people. Feature accessibility is another important consideration. In many cases, personalization features (as mentioned previously) are available, but they are buried deep within a site and difficult for users to locate. Another potential roadblock is users' willingness to reveal personal information to fine-tune personalization features. Privacy policy statements are often confusing and arouse suspicion in users’ minds, leading to reluctance to share personal information. All these factors contribute to a relative scarcity of personalization features on search engine sites.
A personalized search system that uses the search results returned by any of the existing search engine can be used to achieve small scale personalized web search for individual as well as to small sized group users. Such a personalized search could be either server or client-based. A server-based search engine like Google (2009) could keep track of a user’s previous queries and selected documents, and use this information to infer user interests. A client-based personalized search service can keep track of all of the documents edited or viewed by a user, in order to obtain a better model of the user’s interests. This way, an effective personalization search system could decide autonomously whether or not a user is interested in a specific webpage and, in the negative case, prevent it from being displayed or, the system could navigate through the web on its own and notify the user if it found a page or site of presumed interest. To make such decisions, users search history is very essential. Search history houses many hidden facts like user browsing patterns and user interests which are considered important parameters for a personalized search system.

2.2.2 Search Histories

Classical search histories, does not house any association between the relevant pages retrieved and the given search queries. Such search histories only grow in size and make the recommendations more complex thus providing an ineffective personalization.

A search engine that knows the entire user’s previous requests and interests, and uses that information to tailor results is very essential for the web users to identify their information need from the vast repository on the WWW. Some search engines like Google (2009), ODP (2009) and Yahoo! (2009) perform personalized search based on the explicit user interaction for obtaining user interests. These interests are then recorded in the servers. Both
relevant documents and categories according to the user specified interests are then provided to the users. However getting explicit user feedbacks regarding the user interests will consume the user’s valuable time. Northern Light (2009) and WiseNut (2009) cluster their results into categories and Vivisimo (2009) groups’ results dynamically into clusters. Teoma (2009) clusters its results and provides query refinements.

A lot of research in metasearch and distributed retrieval (Gravano and Garcia-Molina 1995; Gauch et al 1996; Howe and Dreilinger 1997; Dolin et al 1998; Fuhr 1999; Xu and Croft 1999; Powell et al 2000; Yu et al 2001) also investigates mapping user queries to a set of categories or collections. However, all of the above techniques return the same results for a given query, regardless of who submitted the query. This can be interpreted as having a general profile.

A local personalization strategy that allows the building of a search index in order to assist web search for well-known and new web sites is very essential for developing an effective personalized search system. The main idea is to define fundamental associations between queries and relevant results according to the concept and use those associations for future access in a personalized ranking list.

A personalized representation is missing in order to present the results depending on previous relevant search results of all users. Hence, for an accumulation, which leads to a wider base of ranked and validated results, identical queries and relevant results collected by all users have to be identified. Search histories allow users to aid content and collaborative search which is the working principle behind recommender systems.
2.2.3 Recommender Systems

Among all personalization tools, recommendation systems are the most employed tools in web search. Recommender systems form a specific type of information filtering technique that attempts to present information items (movies, music, books, news, images, web pages, etc.) that are likely of interest to the user. Typically, a recommender system compares the user's profile to some reference characteristics, and seeks to predict the rating that a user would give to an item they had not yet considered. These characteristics may be from the information item (the content-based approach) or the user's social environment (the collaborative approach). To date, most recommendation systems are designed as either content-based or collaborative recommendation systems.

Unobtrusive monitoring provides positive examples of what the user is looking for, without interfering with the users’ normal activity. Heuristics can also be applied to infer negative examples, although generally with less confidence (Staab and Studer 2004). This idea has led to content-based recommender systems, which unobtrusively watch users browse the web, and recommend new pages that correlate with a user profile.

2.2.3.1 Content-based Recommendation

In content-based recommendation one tries to recommend items similar to those a given user has liked in the past. The content-based approach to recommendation has its roots in information retrieval, and employs many of the same techniques. Text documents are recommended based on a comparison between their content and user profile. A pure content-based recommendation system is one in which recommendations made for a user is based solely on a profile built up by analyzing the content of items which that
user has rated in the past. Examples of such systems are InfoFinder (Krulwich and Burkey 1996), NewsWeeder (Lang 1995), and systems developed for the routing task at the TREC conferences (Harman 1994).

Personal web-based agents such as Letizia (Lieberman 1995), Syskill and Webert (Pazzani et al 1996) and Personal Webwatcher (Mladenic 1996) track the users browsing and formulate user profiles. Profiles are constructed from positive and negative examples of interest, obtained from explicit feedback or heuristics analyzing browsing behavior. Later they suggest which links are worth following from the current web page by recommending page links most similar to the user’s profile. News filtering agents such as NewsWeeder (Lang 1995) and News Dude (Billsus and Pazzani 1999) recommend news stories based on content similarity to previously rated examples. Systems such as CiteSeer (Bollacker et al 1998) use content-based similarity matching to help search for interesting research papers within a digital library. Ontologies are also used to improve content-based search, as seen in OntoSeek (Guarino et al 1999).

However, content-based methods can only be applied to a few kinds of content, such as text and image, and the extracted features can only capture certain aspects of the content. In addition, these systems provide recommendations merely based on user profiles. Therefore, users have no chance of exploring new items that are not similar to those items included in their profiles. Another way to recommend pages is based on the ratings of other people who have seen the page before. Collaborative recommender systems do this by asking people to rate explicitly pages and then recommend new pages that similar users have rated highly.
2.2.3.2 Collaborative Recommendation

Collaborative recommendation identifies users whose tastes are similar to those of the given user and recommends items they have liked. Collaborative recommendation works based on the assumption that if user x’s interests are similar to user(s) y’s interests, the items preferred by y can be recommended to x. A pure collaborative recommendation system is one which does no analysis of the items at all; in fact, all that is known about an item is a unique identifier. Recommendations for any user are made solely on the basis of similarities to other users. Examples of systems taking this approach include GroupLens (Resnick et al 1994), the Bellcore video recommender (Hill et al 1995), and Ringo (Shardanand and Maes 1995).

In addition, since other user profiles are also considered, user can explore new items. The nearest-neighbor algorithm is the earliest collaboration based technique used in recommendation systems (Resnick et al 1994; Shardanand and Maes 1995). With this algorithm, the similarity between users is evaluated based on their ratings of products, and the recommendation is generated considering the items visited by nearest neighbors of the user. In its original form, the nearest-neighbor algorithm uses a two-dimensional user-item matrix to represent user profiles. However, collaborative based recommendation systems suffers from problems like the following: 1) if the number of users are very less when compared to the amount of information collected, then there is a danger of the coverage of ratings becoming very sparse and 2) for a user whose tastes are unusual compared to the rest of the population there will not be any other users who are particularly similar, leading to poor recommendations.
2.2.3.3 Hybrid Recommendation

A hybrid recommendation system is one which simultaneously utilizes the advantages of content and collaborative recommendation approaches. Yoda (Shahabi et al 2001) is one such system which utilizes client-side web usage data and generates cluster recommendation lists based on clustering and content analysis techniques. This approach addresses the problems associated with collaborative recommendation system. In addition, Yoda collects information about user interests from implicit web navigation behaviors.

Utility-based and knowledge-based recommender systems require that knowledge about how a particular object satisfies the user needs and can use this knowledge to search the information space for objects relevant for the user in a particular situation. These systems do not attempt to build long-term generalizations about their users. Utility-based systems can be considered a special case of knowledge-based systems where a utility function is defined for each user (automatically or manually, by the user himself). These types of recommender systems suffer from different weaknesses like content limitation, over specialization, synonymy and scalability (and have different strengths like capturing real user navigations, accurate and implicit user data collection) so, usually, they are combined in order to increase performance.

All the above-mentioned systems work well for one type of web search called the product search through the web and for page recommendations within a web site. But, our work is about general search through the web which is identified as the second largest online activity in the World Wide Web (Jupiter Communications 2007). Alternatively, recommendation to users can also be made by modeling user search behavior utilizing the underlying conceptual knowledge from the visited web pages.
2.2.4 User Modeling

User modeling is typically either knowledge-based or behavior-based. Knowledge-based approaches engineer static models of users and dynamically match users to the closest model. Behavior-based approaches use the users’ behavior itself as a model, often using machine-learning techniques to discover useful patterns of behavior. Kobsa (1990) provides a good survey of user modeling techniques. In-depth ontological representations are also seen in knowledge-based systems, which use relationships between web entities (bookmarks, web pages, page authors etc.) to infer facts about given situations. Ontology is used to investigate how domain knowledge can help in the acquisition of user preferences. Kelly and Nicholas (2002) describes longitudinal, naturalistic study designed to model characteristics of the user’s problematic situation by considering several characteristics, such as topic familiarity and persistence, task endurance and problem solving stage.

As it becomes possible to gather and store more historical information about a user’s interactions, we need to develop models that span tasks and applications. Such a model would be in contrast to the short-term, single-task models that are more common to date. In the area of personalized Web search, WebMate (Chen and Sycara 1998) utilizes user profiles to refine user queries, but no experimental results are given. Watson (Budzik and Hammond 1999) refines queries using a local context, but does not learn the user profile. Inquirus 2 (Glover 2001) utilizes users’ preferences to choose data sources and refine queries, but it does not have user profiles and requires the users to provide their preferences of categories.

Works by Pretschner and Gauch (1999) and Liu et al (2004b) adopt automatic learning of users’ profiles from their surfing histories and reranks/filters documents returned by a meta-search engine based on the
profiles. User modeling in dynamic environment like WWW helps to predict the interests of individuals during web search. Such predictions aid effective user categorization and hence lead to community based search.

2.2.5 Group Search

In general, searches conducted in an organization or institution will be in groups. Recommendation of pages can also be done with respect to the interests and web usage of other users who participate in the group search within the organization. Grouping user searches can help to identify similar searches which are very useful for recommending pages/ search paths for a new user/novice user. For a first time user of a personalized search system there will not be any recorded browsing history or profile. Hence the new user cannot be recommended with personalized search results. But analysis of search groups helps to solve the above-mentioned problem by recommending pages by identifying similar search queries which were previously used in a group. Any IR task including web search demand or implicitly requires much more sophisticated models of users. There is therefore a significant need for research in how to include characteristics of such elements as users’ goals, tasks, and contexts in user models, as well as how to make use of this knowledge in affecting IR system design and response.

The search community should support one or more groups in creating such a data-gathering laboratory, defining what types of information need to be acquired (interaction histories, annotations, explicit preferences, etc), and sharing all results. This process would be cooperative among a large number of sites. As a result, it will not be necessary for every group to gather these types of data in order to do experiments.
2.2.6  Web Search Evaluation

Even when advances are made in the ranking of search results, proper evaluation of these improvements is a non-trivial task. In contrast to traditional IR evaluation methods using manually classified corpora such as the TREC collections, evaluating the efficacy of web search remains an open problem. Traditional IR evaluation measures usually require relevance feedback from users. However, obtaining relevance feedback explicitly from users for personalized web search systems is extremely challenging and time consuming due to the large size of WWW. Commercial search engines often make use of various manual and statistical evaluation criteria in evaluating their ranking functions (Joachims 2002). There has also been work along the lines of using decision theoretic analysis (i.e., maximizing users’ utility when searching, considering the relevance of the results found as well as the time taken to find those results) as a means for determining the ‘goodness’ of a ranking scheme. One or more standard experimental data sets need to be created that are heavily annotated with information about users, context, etc. Such a research collection should allow evaluations that focus on the user in more detail, but in a way that does not require expensive user studies. Such a dataset would allow for some degree of comparability and comparison between results from different groups.

An evaluation methodology that explicitly leverages information about the user and the context of information retrieval process is essential for effective web search. Such an evaluation would have the goal of dramatically improving effectiveness for a particular user rather than providing merely satisfactory results for an unknown user.
2.3 PERSONALIZATION FOR GENERAL INFORMATION SEARCH

A variety of techniques have been proposed for personalized web search systems. All these approaches incorporate user preferences/interests for recommending pages to the user. In order to incorporate user preferences into web search, three major approaches are proposed: user profiles, query refinement and personalized page ranking/importance.

2.3.1 User Profile Creation

User profiles provide abstract information about user’s interests and preferences. Researchers have used various methodologies like: 1) by getting explicit user inputs like favorite pages, feedbacks about pages, 2) utilize implicit indicators of user’s interests like user actions, page-view time, hits and 3) exploit user navigational behaviors, for creating user profile.

2.3.1.1 User Profile Creation Using Explicit User Feedbacks

The user profiles could be explicitly entered by users or implicitly learned from user behaviors. Most personalization systems are based on user profiles, most commonly a set of weighted keywords. The client-side web search tool proposed by Chau et al (2001) requires direct inputs about interesting phrases from users. Personalised PageRank Vector introduced by Jeh and Widom denotes the personalized view of importance of pages on the web (Jeh and Widom 2003). The users must explicitly specify user preferences so as to set up Personalised PageRank Vector. The Personalised PageRank Vector is then used for final page recommendations. However, getting explicit feedbacks increases the work load of user and affects the normal search process.
2.3.1.2 User Profile Creation Using Implicit Indicators of User Interests

WebMate (Chen and Sycara 1998) automatically learns users' interested domains through a set of interesting examples. Montebello et al (1998) present a system that reuses the information generated from search engines together with previously developed systems, and adapts it, by generating user profiles, to better meet the needs and interests of users by improving recall and precision measures. They assume that normally, when searching or even browsing, a user bookmarks a page of interest and proceeds with the activity he/she is performing. Taking this activity into perspective all that is required is to take into consideration what the user bookmarks, and utilize this information to generate the profile.

Citeseer (Rucker and Polanco 1997) utilizes each user’s bookmarks as an implicit declaration of interest in the underlying content, and the user’s grouping behavior as an indication of semantic coherency or relevant groupings between subjects. Over time, Citeseer learns each user’s preferences and the categories through which they view the world, and at the same time it learns for each Web page how different communities or affinity based clusters of users regard it. Citeseer then delivers personalized recommendations of online content, Web pages, organized according to each user’s folders. Widyantoro et al (2001) utilized temporal information like page view time to distinguish between a user’s short term and long term interests. SmartPush (Kurki et al 1999) uses concept hierarchies for user profiles.

Personalized search systems are required to adapt to dynamic changes in user interests. Amalthea (Moukas and Zacharia 1997) and NewT (Sheth 1993) are systems that consider this issue by employing an
evolutionary algorithm and represent user interests as a population of profiles. These profiles evolve based on the user’s relevance feedback to adapt to the drift of user interests over time. Due to the nature of the algorithm, however, a high computational cost and a great amount of effort from the user to rate documents are required. Thus, a practical and cost-effective approach to address this issue remains to be developed.

Thus, implicit profile creation based on observations of the user’s actions is used in most recent projects. Utilization of conceptual knowledge for generating user profiles is another important dimension in user profile creation. One such example is the use of ontology to represent the salient relationships and entities in a given domain as well as to model the user. One key benefit in using ontological models is that they can support the design of sharable, agreed-upon models of a web search user’s context. Gauch et al (2003) have developed a relationship-based system in which the user profile is created automatically and implicitly while the users browse the web pages.

Profiles are generated by analyzing the surfing behavior of the user, specifically the content, length, and time spent on each web page they visit. The web pages that the user visits are automatically classified into concepts contained in the reference ontology and the results of the classification are accumulated. Another example of a relationship-based approach is the Personal Web Context (PWC). The goal of the PWC approach is to analyze known relationships in order to infer and explicate implicit relationships between resources. The ontology in the PWC is exploited to support operation in a homogeneous environment and support inference processes for extending and exploiting the user’s personal web. The core of this approach is a match-making process that supports discovering sequences of relationships from
which a single relation can be compositionally inferred; for this purpose, rules are used.

Chan (2000) considers the frequency of visits to a page, the amount of time spent on the page, how recently a page is visited and whether or not the page is bookmarked. User profile creation based on implicit indicators of user’s interests prove to be effective and less time consuming than explicit feedbacks that hinder the normal search process.

### 2.3.1.3 User Profile Creation Using Navigational Behavior

Many personalization projects have focused on navigation. Persona (Tanudjaja and Mui 2002) learns the taxonomy of user interests and disinterests from user's navigation history; the system proposed by Liu et al (2002) can learn user's favorite categories from his/her search history.

Syskill and Webert (Pazzani et al 1996) get explicit feedbacks from the user and they can recommend other links on the page in which they might be interested. This recommendation agent is designed to help a user with long-term information seeking goals. Letizia (Lieberman 1995) is a similar individual system that assists a user while browsing by suggesting various links. These links might be of interest to the user and/or related to the page the user is currently visiting. The system relies on implicit feedback including links followed by the user or pages and/or bookmarked pages. WebMate (Chen and Sycara 1998) is an individual system based on a stand-alone proxy that can monitor a user’s actions to automatically create a user profile. Later, the user can enter an URL and WebMate will download the page, check for similarity with the user’s profile, and recommend further similar pages.
2.3.2 Query Refinement by Incorporating User Profiles

Instead of modifying the ranking algorithms one easy technique is query refinement based on user interests/preferences. Such systems first adjust the input query based on the corresponding user profile. Subsequently, the modified query is given to search engines. The system proposed by Liu et al (2002) maps the input query to a set of interesting categories based on the user profile and confines the search domain to these categories. Websifter (Scime and Kershberg 2000) is another such system which formulates the query based on user’s search taxonomy and then submits the query to multiple search engines.

After receiving the query results from the search engine, the systems refine the response. Occasionally, some search systems would further filter the irrelevant pages. For example, in the Persona system (Tanudjaja and Mui 2002), the search results are ranked according to authoritativeness with a graph based algorithm. The returned set in Persona only contains the top n documents. Furthermore, Persona would refine the results if the user provides positive or negative feedback on the response.

In general, maintaining efficiency is the major challenge of query refinement approach. The time complexity of the proposed techniques grows with the size of user profiles, e.g., the number of interested categories, keywords, and domains. Hence an alternative to user profile is tracing user’s navigational behavior.

2.3.3 Personalized Page Importance

The most famous example that employs the importance scores of pages for ranking the search results is the PageRank algorithm, which is the
basis for all web search tools of Google (2009). By utilizing the linkage structure of the web, PageRank computes the corresponding importance score for each page. These importance scores will affect the final ranking of the search results. Therefore, by modifying the importance equations based on user preference, the PageRank algorithm can create a personalized search engine.

Basically, personalized importance scores are usually computed based on a set of favorite pages defined by users. Works like topic-sensitive PageRank (Haveliwala 2002) first pre-computes web pages based on the categories in Open Directory. Next, by using the pre-computation results and the favorite pages, the system can retrieve ‘topic-sensitive’ pages for users. But topic-sensitive PageRank is not scalable, since the number of favorite pages is limited to sixteen.

Personalized PageRank Vector (Jeh and Widom 2003) can be considered as a personalized view of the importance of pages. User explicitly mentions their favorite pages are and priority is given to those pages linked by the favorite pages. Therefore, by incorporating Personalised PageRank Vector during the selection process, the search engine can retrieve pages closer to user preferences. Since these techniques require direct inputs from users, the system increases the usage overhead.

Modifying the ranking algorithm of large scale search engines like Google is not advisable since it is not cost effective. But results returned by existing search engines might be collected, re-ranked according to the user’s preferences and returned to the user which proves to be cost effective. Moreover the techniques discussed above touch upon a few aspects of personalization and do not provide an integrated approach that encompasses the entire spectrum of factors that affect personalization.
All the above-mentioned systems tend to miss another important factor – the pages missed by search engines, but explored by user. Such pages prove to be the direct answers to user’s information need. A user while searching for information might discover a relevant page which is not listed by search engine. Identification and recommendation of such informative pages is an important feature that any personalized search system should possess.

Although the previously discussed systems involve certain factors of personalization they do not cover the entire range of factors that affect personalization. Integration of such factors will be more promising to achieve improved personalization in search systems. Considering this in mind we have proposed an integrated framework of web search personalization which takes into account the conceptual relations between search queries and the visited pages.