CHAPTER 2

LITERATURE REVIEW

2.1 PERSONALIZATION

Personalization is the provision to the individual of customized products, service, information or information relating to a product or service. It covers recommender systems, customization, and adaptive Web sites. It is a process used to customize the services based on the individual’s needs, characteristics or preferences. It is commonly based on the attributes of users. Personalization can be applied to Web content or Web applications to reach customers quickly and easily. For example, Web search engine is personalized based on the past queries, clicked documents, and browsing activities of an individual to target marketing activities. The key requirement is the understanding of users’ need to tailor the Web services to maximize the customer satisfaction. Personal user information must be collected to personalize any Web application.

2.1.1 Types of Personalization

There are three categories in Personalization (Eirinaki and Vazirgiannis 2003). They are

- Content based
- Rule based
- Collaboration based
Content-based methods are based on information about and characteristics of the items that are going to be recommended. Basically, these methods maintain user profiles to fulfill the needs of the end users. It provides the highest value to an individual for individualizing the profile. Personal information can be collected from the users based on two approaches. Explicit profile generation is static in nature and it collects information from the users by filling up the forms, feedbacks, preferences and ratings. Users’ interest may change over time which is not reflected or updated to the profile. Therefore, the major drawback is static and inaccurate user profiles. Implicit approach collects information from the search and browsing activity of the user. Any change in the user interest is reflected into the profile and it is dynamic in nature (Balabanovic and Shohan 1997).

In rule based system, site owner derives the set of predefined rules to perform recommendations according to the requirement of a user. A set of business regulations is derived from the attributes related to the item, user demographics, and their experiences. It resolves to decide what category of content is shown when a certain profile type visits the site (Eirinaki and Vazirgiannis 2003). In this approach, site owner derives the set of questions from a decision tree and ask the user to answer the question according to the requirement. Items are recommended based on the criteria mentioned by the user. It provides good recommendation when the user requirement is relatively stable and makes clearly visible to the site owner. Rules-based personalization involves a broad, research-intensive setup. Particularly in the case of large record, it is hard to control the difficult rule maintenance. Basically, site owner must devise new rules for each new product evolved by them in the market.

The foremost goal of Collaborative filtering approach is to recommend the specific item based on prediction from previous experiences
of similar minded users. In aforementioned user profile based approach, items are recommended based only on user preferences predicted from content of the items (Konstan et al 1994). The major drawback of this approach is sparsity problem. It performs well only when several users provide the experiences about the item. It is hard to recommend the new item through this approach because sufficient feedback is not available to predict the similarities between different users.

2.2 USER PROFILES

User profile contains the list of concepts for which weights are accumulated based on user browsing behaviours. Concept, which is often browsed by users, is extracted from their search record, and hierarchical structure is constructed. A user profile must be created and maintained using collected information to improve conceptual results. Conceptual rank is obtained from the similarity between document and user profile. Set of keyword vectors is represented by user interests (Sumathi et al 2010). User profiling is a basic and necessary process for effective personalization. User profiles are constructed using the user interest predicted through implicit or explicit feedback. Depending upon the user feedback, user profiling strategies are divided into two major approaches such as explicit and implicit user profiling.

In explicit profiling, user preferences are predicted through raising the direct questions explicitly in the form of an item rating. However, if the user is not interested in responding to the explicit request for feedback, then it leads to imprecise user profile that cannot reflect the users changing preferences over time.

In Implicit profiling, user profiles are constructed through predicting user preferences implicitly based on the users interactions within the system.
It is possible by user browsing history and click based approach. There are two approaches in implicit user profiling. They are (i) concept based approach and (ii) document based approach.

In Concept based approach, user profiles are constructed using extracted user preferences from topical categories. The topical categories automatically perform the mapping between user browsing history and the browsed document. On the other hand, in document based approach, user profiles are constructed using user preferences extracted from the click through data. The preferences of users are represented in the form of set of weighted factors predicted from the user behaviour model (Leung and Lee 2010).

### 2.2.1 Concept based User Profile

In concept based method, topical interests of the user are automatically derived from the browsed documents in user’s search history. It is represented in the form of set of categories. Each category has a set of keywords with weighted score. The user profiles are constructed using the top four levels of the concept hierarchy. Hierarchical User profiles are constructed using reference ontology and concepts related to topical categories are extracted from the user browsed documents in their browsing history. Concept hierarchy is dynamically constructed that mainly allows extending the user profiles with newly extracted concepts.

In most of the user profiling strategies, profiles are constructed using concept based method because it can construct user profile dynamically with high accuracy. In the creation of user profiles, text classification techniques are deployed to classify the user browsed documents into concepts in the reference ontology. However, it includes some restrictions and weaknesses. It points out that the user profiling performance considerably
relies on the precision of predefined categories. If the negotiation of a specified category is wrong, the performance is ruined (Tao and Li 2009). Another challenging problem, specified to as “cold start”, arises when there are insufficient number of exercise samples accessible to learning classifiers. The concept-based user profiling systems rely on the assumption that the content of user browsed Web documents is sufficient to construct descriptions for classification. (Leung and Lee 2010)

2.2.2 Document based User Profile

In Document based user profile, user clicking and browsing behavior are recorded from users’ click through data. In Web search engines, click through data is the important feedback mechanism provided by the user through clicking. This type of user profile is modeled using preference mining and machine learning tools. User Document preferences are computed using spying technique with novel voting procedure which is used to determine the user document preferences based on the user click through data. After determining the user behaviour from the document preferences, Ranking Support Vector Machine (RSVM) model is deployed to determine the user behaviour model from the set of weighted feature. Then user profile is associated with the set of weighted features and their concepts.

2.3 MINING ALGORITHMS FOR USER PROFILE CREATION

Data mining process determines the inherent relationships among Web data which are expressed in the form of textual, linkage, or usage information, via analysing the features of the Web and Web-based data using data mining techniques. Due to influential ability of non linear range of data sources, it is extensively used. It consists of different cluster based approaches performed on various data inspection model. The data inspection represents the feasibility of pattern mining. It is mainly concentrated on discovering
Web usage pattern by Web usage mining, and then make use of the
discovered usage knowledge for representing Web users with more
personalized Web contents. The post processing is proceeded by clustering
results to attain user profiles, and at last ends with tracking profile progression
(Kuflik and Shoval 2004; Sugiyama et al 2004). The repeated recognition of
user profiles is a knowledge discovery assignment consisting of intermittently
mining of the user contact log files and is concise in the following steps:

- Pre-process Web log file to mine user sessions.
- Cluster user sessions via Hierarchical Unsupervised Niche
  Clustering (H-UNC).
- Reiterate session clusters /categories into user profiles.
- Improve the user profiles with added feature by using extra Web
  log data and peripheral domain knowledge.
- Follow current profiles aligned with existing profiles.

User profile creation can be represented through various data mining
algorithms:

2.3.1 K-Means Algorithm

K-means algorithm is an effective data mining algorithm used to
cluster the datasets iteratively into a user specified number of clusters. It is
performed on the set of $d$-dimensional vectors $V= \{X_i \mid i=1, 2, 3\ldots N\}$,
where $X_i$ denotes the $i$-th data point. This algorithm is initialized through
randomly selecting the $k$ number of points from the $d$ number of dimensional
vectors and $k$ representative points are grouped into clusters and also called as
centroids. Here, $k$ points are randomly selected from the data sets using
specific sampling techniques and also clusters the small subset of data points.
(Wu et al 2008). Clustering is performed through iteratively comparing the $N$ number of data points and $k$ number of randomly selected $k$ data points ($N \times k$). The number of iterations required to maintain the convergence is predicted from the number of data points in the dimensionality vector. It is performed in two steps.

The first step is data assignment in which each data point in data sets is assigned to the closest centroid. Data partitioning is performed through breaking the datasets randomly. Then data points are grouped into user specific clusters. The second step is relocation in which representative of the cluster is relocated to the center of all data points randomly assigned to it. Depending upon the weight (probability measure of data points) relocation of representative in cluster is performed with some expectations. This algorithm allows convergence when the data points in clusters are static without any dynamic change (Dhillon et al 2004).

### 2.3.2 Naïve Bayes Classifier

This algorithm is a supervised learning algorithm that classifies the Web documents with a given set of user known objects belongs to the specific class. User acquired the knowledge about the class where object is situated and then number of vector variables possessed by the object. The main aim of this classifier is to construct user profile based on the set of rules that allow the user to assign the object for the class. Certain rules are developed to perform classification using supervised learning. This type of Naïve Bayes classification is also called as idiot’s Bayes, simple Bayes, and independence Bayes. It is ubiquitous approach that can construct user profile easily without any complexity. It can be readily available for a huge number of data sets and also performs quite well. The main disadvantage is that it does not perform classification in some particular application (Hand and Yu 2001).
The Naïve Bayes model is tremendously attractive due to its simplicity, sophistication, and robustness. It is one of the oldest formal classification algorithms, and even now available in its simplest form. It is mostly used in the field of text classification and spam filtering. In order to make this classifier as flexible, large number of modifications has been introduced, by the statistical, data mining, machine learning, and pattern recognition communities.

2.3.3 K-Nearest Neighbor (KNN) Classification

Rote classifier is one of the simplest and trivial classifiers used to remember the entire training set of data and also classification is performed according to the matching between the attributes of the test object and any one of the training samples exactly. The main problem in Rote classifier is that test records will not be classified in the situation that no objects are matched with the training set. In order to alleviate this problem, a more sophisticated approach called as K-Nearest Neighbor (KNN) is evolved (Han et al 1999). It can determine the group of \( k \) number of objects in the training set that is approximately matches with the test object. Depending upon the similarity, predominance score is assigned as a label to the object. There are three key elements that are mainly useful in this approach called as set of labeled objects, similarity score and the value of \( k \). Set of labeled objects are nothing but the set of stored records. Similarity score is determined from the distance between the set of objects. It is used to classify the set of unlabeled objects. The class label of the object is determined from the class labels of the nearest neighbor. KNN classifier is a slow learning supervised classifying system that cannot be explicitly earners, that is, models are not built explicitly unlike eager learners (e.g., decision trees, SVM, etc.).
Building the model is cheap, but classifying unknown objects is relatively expensive since it requires the computation of the k-nearest neighbors of the object to be labeled. In general, distance of the unlabeled object is computed for the complete set of objects in the labeled set. But it is more expensive in the case of large training sets. Large number of techniques is evolved to compute the distance between the K-Nearest Neighbors. But it is particularly applicable for low dimensional data and also provides less computational cost without affecting classification accuracy.

2.4 ONTOLOGY

Ontology is exactly stated as “formal, explicit specification of a shared conceptualization”. It is a logical theory explained through a set of logical axioms designed with an explicit design of conceptualization (Guarino 1998). Conceptualization represents the suitable formalization that provides the conceptual relation that exists between the set of entities in a domain. It is mainly intended to provide formal vocabulary with knowledge, proper representation of the world in program code. It indirectly reflects the underlying conceptualism evaluated from the intended semantic models (Guarino 1995). It is also used in the field of information technology such as information retrieval and integration (Mayfield and Finin 2003).

Semantic Web is stated as an Ontology which is formally constructed through semantic relationship between concepts in domain entities and represent static domain knowledge. Ontology must need to specify the description for the following kinds of concepts such as

- Classes: These are concepts in domains of interest.
- Relationships: Semantic relation between the set of concepts in the domain.
Properties: These are the attributes possessed by the concepts in certain domain.

Figure 2.1 Ontology Modeling

Figure 2.1 shows the model to design the ontology with their necessary concepts and its corresponding relationship. Ontologies are classified into many types such as Reference Ontology (Gauch et al 2003), Domain Ontology (Bhowmick et al 2010), Personalized Ontology (Tao et al 2011), Topic Ontology (Zhou et al 2006), User Ontology, and so on.

2.4.1 Need for Topic Ontology

It is very difficult to understand and access all the information enclosed in a large corpus of document. Normally, the techniques like information retrieval and text mining relies on keyword matching. It would not consider the semantic similarity within the corpus. Topic ontology can overcome the problem through extracting the semantically related concepts from the large document corpus. The topic ontology holds the topic in a particular document signifies the subject of the terms (Zhou et al 2006).
2.4.2 Topic Ontology

Topic ontology is an unambiguous design of set of topics and the conceptual semantic relationship that exists between the set of topics. Topic ontology is constructed in the form of tree or graph like structure that categorizes Web information into predefined topics. Each node in the graph represents concepts in the topic. It can describe the basic structure and their functions created by conceptual topics of relationship (Wu et al. 2008). Therefore, conceptual hierarchical relationship is maintained between the set of concepts in topic ontology (Zhou et al. 2006).

Topic ontology is formed using the set of terms and its primitive classes. The primitive classes are the least concepts that cannot be accumulated from other classes. It may be inherited by derived concepts or their subclasses. From a set of primitive classes, compound classes are generated. The procedure of mining ontology from Web documents has been developed efficiently. In this mining procedure, the base spine and the top spine are engaged to attach patterns with each other. The base spine is used for the connection between primitive classes as the top spine for the linkages among compound classes. The method of constructing the topic ontology needs both the base spine and the top spine generation.

Figure 2.2 illustrates the topic ontology model with four layers such as Super topic layer, Latent topic layer, Index Topic layer and Page Document layer. The conceptually related set of topics are clustered together in the form of tree structure, and then whole ontology needs to be structured in concept hierarchy since the semantic relationships such as “is -a” and “part-of” exist among page documents, index terms, and latent topics.
This topic ontology is built from the user browsed documents from their browsing history (Grcar et al 2005). If the user browses particular topics, it can be extracted from user browsed histories and ontology is used for hierarchical structure of profiles (Sieg et al 2007).

2.4.3 Topic Scoring

Topic scoring is the essential component to detect the topics that are interested to the users. If topic is detected for a given query exactly, then construct many accurate profiles and then personalize the outcome more efficiently. Conversely in click-based topic learning method, personalized ranking is not effective. This is mainly due to the personalization that is performed through predicting the fascinated topics from user browsing history. Therefore, it leads to return irrelevant results, which are entirely dissimilar from the actual requirement of the user. Every user represents their topic preferences during query submission. This type of submitted topic preferences is used to evaluate the performance of personalization. Sequentially, score is assigned to the topics according to the number of times the pages related to that topic is viewed by the user. Finally, topics are listed out in the form of descending order from higher score topics to lower score topics.
Interest score is applied on the topics that provide the ranked collection of documents as a result. Scoring is computed according to the relevancy between topics of documents and requirement of user information. Ranking could be done by having higher scores of user independent topics. Results set are personalized for each topic and provided by the personalized search engine. The interest score weight is calculated for each topic, and based on the weight those topics are provided as a result set of documents. Spreading Activation algorithm is suitable on the huge amount of topics to be ranked. It is more useful in finding a precise user interests rather than any scoring algorithm. It renders the result quality and also helps in personalization and recommendation of user profiles more accurately. Scoring is good at capturing user interests for ranking the most appropriate topics for user profiling (Thiel and Berthold 2010).

2.5 SPREADING ACTIVATION TECHNIQUE

Spreading Activation technique was initiated by Collins and Loftus (1975) and it was applied to semantic priming and psycho linguistics. Later on, this technique was implemented in computer science, particularly in information retrieval process. The ideas of Spreading Activation were also employed in theory of information for aging. Spreading Activation comprises the type called Constrained Spreading Activation (CSA). The main application of CSA is to generate recommendation. For this purpose, activation paths are stored to provide recommendations. The essential design of Spreading Activation is all related information is plotted on a graph as nodes with a particular activation level (Hussein and Ziegler 2008). For mapping, all the related concepts on a graph, spreading algorithm is exploited. A link between nodes is to specify the related concepts in a concept hierarchy. Relations between two concepts are characterized by a link between the equivalent nodes. Associative networks and neural networks are searched in
Spreading Activation method. It identifies and propagates the nodes linked together iteratively. It terminates when two paths reach the same node (Vilches-Blazquez et al 2009).

Activation level is raised while activating the nodes, and initial activation value is 0 for all the nodes. Mouse click of a certain topic is utilized as an initial node and it initiates the activation. Thus, concepts are associated with the current one encompass a high activation value. Activation can be initiated for adjacent nodes too. All nodes have some activation value which specifies the degree of relevance (Hussein and Neuhaus 2010). Average rate of increase and average variation in interest scores together express that interest scores in user profiles are getting updated accurately. Existing concepts are updated by Spreading Activation. Interest score rates are updated using this activation (Shirazi et al 2009). Spreading Activation is utilized to get the user present interests in concepts. To provide better concept based user profiles, the outcomes have to be combined with long term interests of user. Spreading Activation value of every node spreads to its adjusting nodes of each concept in a network and this value is related with their relations (Jiang and Tan 2006).

2.5.1 Spreading Activation in Information Retrieval (IR)

Information Retrieval System (IRS) is a knowledge based approach that endeavors to store and permit rapid access to a large amount of information. It may be a textual, visual or auditory. IRS is a tool to store and retrieve information in the future. A user sends query and retrieves information which suits the query through IRS. It enables only the textual information or documents retrieval.

This system is said to be good only when it provides relevant information because some of them will provide irrelevant information. The
use of network depiction structure makes this method particularly pleasing for associative IR. The network knowledge depiction structure is a Semantic Network that uses Spreading Activation as its dispensation hypothesis (Crestani 1997).

Spreading Activation (SA) in information retrieval is based on survival of maps representing the relations among the terms or documents. There is no uniformity in the network, it can represent anything. The node relations are represented by the links. The availability of delegate network signifies the efficiency of the IR process. It also represents the effectiveness of the useful relations. It is hard to construct and maintain the network up-to-date. So it requires broad domain knowledge which provides authorized application. It is very expensive and time consuming. This problem, frequently present in applications of SA to IR, is the basis for the growing interest in techniques for automatically building network illustration and, in meticulous, for the application of machine learning techniques in IR. Entire SA system that crafts use of varied link types and of spreading rules with expanse and fan-out constraints has never been applied with ordinary document collections. Different prototype systems are presented in the IR fiction but no business system based on a SA model actually exists. The principle of this review is not to state in particulars about such prototypes, but only to explain advantages and problems associated to the practical application of SA to IR.

2.5.2 Working Principle of Spreading Activation

According to the theories of Spreading Activation, the quantity of activation that spreads between a prime and a target should be a function of the amount of arbitrating links between the prime and target in a semantic network and the power of those links. The sum of activation should decide the quantity of facilitation given by a prime to a target in lexical decision. It is
essential to compute the associative links among prime and target in cache for forecasting the facilitation (McKoon and Ratcliff 1992). Spreading Activation algorithm constructed in the following way. Nodes $[1...M]$ is populated in a directed graph with a real number range of $[0.0 \ldots 1.0]$ is assigned for each value $U[a]$. Source node $[a]$ with target node $[b]$ is connected to the link $[a, b]$. Each link has its weight $V[a, b]$, a real number in the range of $[0.0 \ldots 1.0]$.

**Parameters:**

- Let $T$ be a Firing threshold, a real number in the range $[0.0 \ldots 1.0]$.
- Let $F$ be a Decay factor, a real number in the range $[0.0 \ldots 1.0]$.

**Steps:**

- Set all activation values $U[a]$ to zero. Define one or more nodes to activation value greater than $T$. normally the starting value is 1.0.
- If $U[a]>T$ for each unfired node$a$ connecting $[a]$ with $[b]$, Adjust $U[b] = U[b] + (U[a]*V[a, b]*F)$.
- If $[b]$ receives the adjustment value which exceeds $U[a]$ to 1.0, reset $U[a]$=1.0 and set $[b]<=0.0$ to receive $U[b]$=0.0.
- A node should not re-fire again. Mark $T$ as firing when nodes receive a new activation value for next Spreading Activation cycle.
- If $U[a]$ points from more than one node, changes in algorithm allow to distinguish the path over graph.
- Terminates when no nodes to fire and reach the path. Changes in algorithm allow repeating nodes to fire and activation loop in the graph. After strong activation state, terminates the process when it reaches maximum iterations.

### 2.5.3 Advantages

The node activation method used in the spreading model starts by inserting a specified weight at some preliminary term or document node. Normally the starting node might signify a term integrated in a primary query formulation or a document retrieved in a search operation. The starting activation weight then spreads through the network with the links originating the starting node. The spreading action first affects those nodes located closest to the starting node and spreads through the network one link at-a-time. The efficiency of the method is, however, crucially dependent on the accessibility of a delegate node relationship map, and on the use of activation rules that can discriminate the functional from the superfluous nodes. It remains to be seen whether representative term and document maps can be premeditated for the subject areas enclosed by regular document collections, and whether the polished spreading rules prepared potential by the system can essentially be translated into practicable node activation methods (Crestani 1997).

### 2.6 ONTOLOGICAL USER PROFILE

Ontology provides a highly meaningful structure for relating user interests, a rich conceptual among them, and allows new topics of interests into the structure (Pretschner and Gauch 1999; Sieg et al 2007). The main intension is to provide enriched structure of ontology for concept based user profiles.
Figure 2.3 Ontological user profile for the user interested in music

Figure 2.3 shows the ontological user profile for the user interested in music. The profile pages are listed in a sequential order and it is possible to be viewed by the users against other profiles inside the session group. User visiting trend is increased due to the existence of user interests and access. According to the score of the topics, user profiles are distinguished from others. However, profiles which are generated are flexible to profit more Web applications such as prediction and recommendation.

Ontological model observes the user behaviors and acknowledges the user needs and preferences. It discovers user interests from the query they have submitted, their frequency, and clicked documents. It also constructs the user profiles. Profiles are built with semantic relationship since topic ontology is used to provide these relationships between ontological topics. Chan (1999) proposed an approach which is mainly directed on assisting users by tailoring and presenting the profiles to each user. The major advantage is that it also obtains the negative preferences based on the interest scores.
2.6.1 Updating User Interest by Spreading Activation

Ontological concept based user profile is created over time by analyzing browsed pages to recognize their content and by associating that content with the length of the document and time that was spent on it. When the pages are viewed again and again, this could be considered as a hint of user interest. User needs to interact with the system for surfing behaviour other than surfing no user interaction with the system is necessary (Pretschner and Gauch 1999). A Spreading Activation algorithm is used to update the concept of interest score in user profile incrementally.

Hence, semantic network is indulged on ontological user profile with interest score updating on activation values. Usually, Spreading Activation method is used in information retrieval which depends upon the subsistence of maps representing the subsistence of particular terms or concepts. It is used to identify the related concepts in an ontology specified set of concepts and equivalent initial activation values. In this approach, very precise pattern of Spreading Activation is used, for the individual reason to preserve interest scores within a user profile (Sumathi et al 2010).

For each process, the algorithm has an opening set of concepts from an ontological user profile. These concepts are allocated with an initial activation value. The major thought is to trigger other concepts subsequent a set of weighted relations through dissemination and at the last part get a set of concepts and their relevant activations. Since some given concept disseminate its activation to its relations, the weight of the relation between the source concept and the end concept which acts a main role in the sum of activation that is approved through the network.

Therefore, once calculation of the weights for the relations in the network is required. The nodes are constructed into a concept hierarchy
resulting from the domain ontology. It computes the weights for the relations among each concept and all of its subordinate concepts using assess of restraint. The restraint weight makes a range of values among zero and one such that a value of zero specifies no overlie among the nodes while a value of one specifies complete overlie (Sumathi et al. 2010).

Most existing systems employed Spreading Activation algorithm with domain and reference ontology, not with topic ontology. Topic ontology is not used for profiling process (Sieg et al. 2007). Ontologies are mainly used in the task of information retrieval. Nevertheless, there is no efficient analysis on enhancing ontology facets or search performance (Strasunskas and Tomassen 2008). Internet users are not satisfied with the results of existing search engines since they provide irrelevant results for user query (Wen et al. 2001). Personalization is not efficient on queries of distinct users (Dou et al. 2007). Existing profiling approaches use only strong relationships and others are discarded. Classifiers are also used to categorize the documents based on user search preferences.

User profiles are updated once user interest is changed from time to time. User interested documents are monitored carefully and Spreading Activation algorithm is applied to update the user profile for each user search session. Interest scores are updated whenever user shows their new interests. Track of changes in user interest are kept and identified that which user has shown more interest or less interest. If interests are closer to user meaning that the user would search for more content related to their interests in between updates. Therefore, interest score weight is assigned to the vectors on the beginning of downloaded and browsed documents in order to obtain the user current interest. To the extent, interest score weight can be assigned comparative to the amount of documents assigned to each topic on the basis of topic similarity. Most frequent topics are recommended to the user. It also
explains the information of how user preferences or interests are applied to personalize the search results (Sieg et al 2007a).

\subsection*{2.6.2 Interested Profiles}

Comprehensive personalization is useful for user profile recommendation relevant to interests of the user. When user is functioning on a specific subject, if the user is interested to surf desktop directories associated to individual’s interests, then it can be classified to the short term interested profiles. Profile is created from repeatedly user visited pages throughout browsing time and the user local directory documents. The reality is that a user clicked link is also taking into account of an interest for the recent document. These information sources are exploited to generate the user interested profiles.

Download history of the users is also observed to make use for creation of user profiles, since it contains specific user interested downloads. User downloads from a specific domain or Uniform Resource Locator (URL) is ordered in a hierarchy which represents user directory preferences for their resources. Requested URL and embedded page are compared to identify the user clicking habits on search results. Similar topics are recommended to enhance the user interested profiles. Web search results are indicated using vectors. Vector is then accepted to give scores based on the similarity of interests. The results with their highly activated scores are then passed to rank the topics and change the order that is ultimately offered to the user (Bloedorn and MacMillan 1996).

\subsection*{2.6.3 Browsing History of User}

User click information matches user surfing history with past surfing activities that generated the collection of document result. This is also
one of the most widespread bases of implicit feedback. However, topic ontology is utilized to learn the user interests and past user history also utilized to recommend the profile model. User click through information is used to surf identical data from preceding users by merging likelihood of preceding user clicks with other users and by calculating a topic similarity among documents. Implicit interests of user are taken into the account of visiting time and user prefers to access more information regarding the document.

It comprises user negative preferences in a way that the user is not following results after submitted a query. From the user local directory information, the profile is modeled. Top-n results are returned whenever a user requests to the search engine and this query is added to the history. Based on the browsing data in a session, long term and short term topic preferences of users are recommended. A collection of related documents is employed to modify the usual ranking and provide a top rank score to documents associated to the preliminary collection. This collection of related documents is extracted and considered as a good quality basis of user interests (Matthijs and Radlinski 2011).

### 2.6.4 Tracking Evolving User Profile

The technique used to generate better user access patterns of different period of time and different profile events is called tracking (Nasraoui et al 2008). Generally profiles and click streams are evolved in summarizing user session. Profiles are compatible which is essential for tracking purpose. This analysis can develop the better understanding of the user activity preference and find seasonality in the access pattern over long time. It also caches the frequently accessed profiles. The following steps describe the main outline of tracking profile algorithm.
Steps:

- For each profile, the beginning and end time period which are stored in a database are extracted.
- Set of URL and its scale is assigned as a profile.
- Start tracking of time periods for each profile as tracking trail.
- Increment the loop till last profile and time period. Compute the distance of the profile.
- Store the record if distance is lesser than scale of the particular profile.

This determines the basic analysis of profile tracking. Hence user profile creation can be performed in an effective manner.

Merits:

It automatically identifies the user profile for knowledge discovery process and for mining the new content of user log file. The mining analysis focuses truly on new session. On using data mining algorithm user session can be clustered automatically and quickly and also for grouping the user interest.

Demerits:

It did not focus on scalability for better updating of profiles. Sessions that pursue a Power Law distribution, the mainstream of concepts or URLs are at the long tail of the allocation and thus have very low probabilities
that can bound the applicability of the IC assess as a untainted similarity measure.

**2.7 SEARCH SESSION PERSONALIZATION BASED ON USER ACTIVITIES**

User action is done when they perform searching and browsing. User actions include submitting a query, mouse clicking on a ranked result, surfing a Web page and etc. In the creation of user profiles, users browsing model in a way they have visited, sessions and clicking activities are identified. Clicks are specific interests since they permit rebuilding of user direction-finding patterns. User interests are captured from the browsing sessions to improve the keyword description of user interests to represent the semantic in a topical hierarchy.

The history of the pages browsed by user and topical hierarchy are utilized to capture and maintain the user’s interests. Usually, a personalized search engine will automatically learn the user’s interests or preferences without any human interference. User preferences are captured using the search or click history relevant to the user interests. This approach improves the search quality and personalization quality. Search provides ranked pages which the user clicks (Daoud et al 2007). Search engine is personalized according to the user current session. Personalization improves the quality of the retrieval process and session data is thin to personalize perfectly. Session is a time closest order of actions done by the same user.

The session does not permit personalization process before submitting the next query in each session. List of terms or search queries, visited pages, past search history and clicked pages represent the user profile. Topic weight is computed by assigning interest scores for each topic and personalization is attracted by the group of people. User model is built using
specific activities of user on browsing habits. Profile is generated and top results are ranked using personalized search engine. It yields good search output without changing the human setting and without having the human perception. Interest score is assigned to the topics of viewed pages, search related information and the other local directory information. These data are activated with score and relevant topics are ranked (Matthijs and Radlinski 2011).

2.8 CONCEPT BASED PERSONALIZATION

Concept based Personalization method is an efficient approach to personalize the Web search using concepts extracted from the user browsing history. Four effective steps are performed in this approach such as i) user submitted the query and then concept relation graph are constructed through extracting the relation between Concepts from the Web snippets. ii) Conceptual user preferences are predicted from the user click through data. iii) Personalized Query Clusters are constructed using two phase personalized agglomerative clustering and then conceptually relative queries are identified from the conceptual graph. (iv) The most related queries are recommended to the user to perform effective Web searching. This approach (Leung et al 2008) provides better precision and recall than other already existing approaches.

Personalized Ontologies are conceptualization model that properly express and state user profiling knowledge. From annotations in daily life, it is established that Web users force to have dissimilar outlook for the same search query. For example, for the topic “New York,” business travelers can insist different information from free time travelers. On occasion, even the same user may comprise dissimilar expectations for the same search query if applied in a different situation. A user may turn into a business traveler when setting up for a business trip, or a leisure traveler when scheduling for a
family holiday. Based on this scrutiny, an idea is formed that Web users have a personal concept model for their information needs.

2.9 TECHNIQUES AND METHODS IN CONCEPT BASED PROFILING

Leung and Lee (2010) proposed six concept based user profiling methods that considered both users’ positive and negative preferences.

- $P_{Joachims-C}$
- $P_{mJoachims-C}$
- $P_{SpyNB-C}$
- $P_{Click+Joachims-C}$
- $P_{Click+mJoachims-C}$
- $P_{Click+SpyNB-C}$

2.9.1 Joachims-C Method ($P_{Joachims-C}$)

This method assumes that the user performs the searching to examine the search results from top to bottom. In order to skip from one document to other document, user must examine the document whether to click on it. If one document has high user preferences than other document, then user can skip to document in top to bottom approach. Document preferences predicted from Joachim’s original method are used to decide whether to click or skip from the later document (Joachims et al 2007). If a user bounces a document at certain rank before clicking the other document at the rank, it must be traced before skipping. This method takes input as a conceptual relation between topics and generates an output called as more accurate user profile consist of list of clicked documents with positive and negative preferences.
2.9.2 mJoachims-C Method (P_mJoachims-C)

mJoachims-C method is extended from the aforementioned Joachim’s method. It considers the documents which are not clicked by the user. That is, it takes documents which relies on above and below of preferences predicted by the user. Joachims-C considers only page reference pairs whereas this approach is extended from that through considering the concept-preference pairs.

2.9.3 SpyNB-C Method (P_SpyNB-C)

Both Joachims and mJoachims make a big assumption that documents, which are not clicked by the user, are considered as irrelevant document to the user query. SpyNB (Ng et al 2007) assumes that the unclicked pages may contain either relevant or irrelevant document related to the user query. Therefore, this method considers clicked pages as positive samples and then unclicked pages as unlabeled samples in the working process. This approach deploys novel voting procedure into a Naïve Bayes classifier to identify and separate out the irrelevant document in unlabeled samples with negative preferences.

2.9.4 Click+Joachims-C Method (P_Click+Joachims-C)

P_Click method is an effective method to capture user’s positive preferences. Like that Joachims-C method is an appropriate method to predict the user negative preferences. In order to predict both positive and negative preferences, this method combines both click and Joachims-C Method. In this P_Click and Joachims-C Method, User profiles are represented in terms of weighted concept vectors and combined into a single unit (Leung and Lee 2010).
2.9.5 **Click+mJoachims-C Method** ($P_{Click+mJoachims-C}$)

This method is proposed through the hybrid combination of both $P_{Click}$ and $P_{mJoachims-C}$ method. It is expressed as a single unit. The weighted concept based user profile is constructed through combining the user preferences in both methods. If the negative preferences are acquired in any one of the methods, then it must be reflected in the weight of the user profile. For example, the concept in $P_{Joachims-C}$ method acquires negative preferences, then the negative weight is appended to the concept of the weight profile even $P_{Click}$ has positive weight on the specified concept.

2.9.6 **Click+SpyNB-C Method** ($P_{Click+SpyNB-C}$)

Similar to Click+Joachims-C and Click+mJoachims-C methods, this method is used to create a hybrid profile $P_{Click+SpyNB-C}$ that combines $P_{Click}$ and $P_{SpyNB-C}$. In aforementioned methods, same user profile is used for all of the queries requested by the user. The main drawback of these methods is that many users consider only the user interested document which represents the positive preferences and skip the uninterested documents which represent the negative documents, in which the positive preference alone cannot provide fine grain interests of the user. The majority of the existing personalization methods focused on the creation of a single profile for a user and utilized the same profile for all of the queries requested by the user (Sieg et al 2007a).

2.9.7 **Limitation in Concept Based Approach**

Concept based user profiling approaches make use of click through data to extract from Web snippets. User positive and negative preferences were captured. The relationship among users and concepts are not performed. Concept based user profiles are not integrated with search engine ranking (Leung and Lee 2010). Ontology based retrieval model exploited complete
domain ontology and information base to sustain semantic search in storage places of document. This method was indirect relation with quantity and excellence of information inside knowledge base.

In most recent developments, mechanized ontology construction and manuscript explanation are potential. The marginal note weight method is not enchanting benefit of document relevance fields. Furthermore, difficulties occur once interoperation associations amongst various arrangements from dissimilar sources are concerned. Multi purpose ontology based search process integrates user interests to get better search results. User profiles are ordered as a concept hierarchy and it allows automatic creation of huge structured user profiles (McKoon and Ratcliff 1992).

2.10 TOWARDS SEMANTIC BASED APPROACH

In concept based methods, insufficient information and also lack of semantic relationship between concepts lead to imprecise construction of user profile. Therefore, ontological approach is considered as the best approach and also can address the various problems in existing concept based profiling. It also facilitates the efficiency and accuracy of information retrieval process with effective personalization.

User profiles are constructed in the form of ontology where high, rich semantic relationship that exists between the set of user interested concepts predicted from the user. The content of the document is semantically annotated and constructed in to ontology. Conceptual relation between domain entities is constructed and then the concepts are identified and then instance are subsequently integrated. In order to attain semantic interpretation, document pre-processing or indexing is performed. The weight of the semantically annotated concepts is computed through combining the structural weights and frequency of occurrence of terms in the document.
In current developed Web technology, Web search engines can afford various advanced features to the user for facilitating an efficient Web search. Therefore, user can effectually represent the query in the form of title or text of the page, URL and links to the page (Sridevi and Nagaveni 2010).

### 2.10.1 Ontology based Personalization

Ontology is a disclaimer of concepts and associations among them. It describes content specific accord on glossary usage, distribution, and reprocess of knowledge. It is used to improve the communication tribulations between systems due to indefinite usage of dissimilar terms. When originally learning user interests, systems usually execute poorly until they accumulate sufficient and appropriate information. Using ontology based profile, the problem like existing concept matching of user’s initial behaviour information with their relations can be solved easily. Open Directory Project concept hierarchy is used for reference ontology (Gauch et al 2003).

The user profile for each concept is explained with initial value of interest score. By viewing or selecting new documents user can interact with the system. Existing concepts annotations can be updated or modified using Spreading Activation method. Therefore, it provides incremental retain of user context on the basis of user behaviour. Correct information about the user’s interests must be composed and defined with nominal user interference. This can be done by submissively monitoring the user’s browsing behaviour in excess of time and collecting Web pages in which the user has shown interest (Daoud et al 2009).
An ontological approach for user profiling is considered as the best approach and also can address the various problems obtained in personalization. User profiles are constructed using domain ontology that coordinates the active concepts in the corresponding domain. The interest score of concepts in the profile is dynamically changing according to the preferences of the user. Therefore, the interest score may be increased or decreased depending upon the user context. After the completion of sufficient information to construct efficient profile, then the sum of interest score change must be reduced and also that is considered as a long term interest (Sumathi et al 2010).

The Figure 2.4 shows the ontological user profile and its corresponding interest score is assigned to each concept in its class hierarchy. The initial query is customized based on the user’s interaction by means of a concept hierarchy which detains the domain knowledge. This domain knowledge is used to disambiguate the user context. In the current framework, the user context is specified by means of an ontological user profile that annotates occurrence of reference ontology. When disambiguating the context, the domain knowledge inherent in existing reference ontology is called upon as a source of key domain concepts.

Every ontological user profile is originally an occurrence of the reference ontology. Each concept in the user profile is explained through interest score with initial value of one. Spreading Activation algorithm is used to update the user profile with proper semantic annotation. (Shen and Zhai 2005).
2.10.2 Representing Ontological User Profile

This approach describes the information that has been developed for profiling and the number of change in interest scores, which are utilized, as long term interest. User profile is characterized as an occurrence of reference
domain ontology in which concepts are explained by interest scores and modified entirely based on the user’s browsing behavior (Shirazi 2009). This specification can be denoted as ontological user profile. Normally it is said to be an occurrence of the reference ontology. The user profile annotated with interest score for each concept has the initial value one. On viewing or selecting the new documents, the user interacts with system, updates ontological user profile, through Spreading Activation, existing concepts are modified. Therefore, based on user’s ongoing behavior, the user context has been maintained and updated incrementally. With nominal user interference, accurate information of the user is gathered and specified. This can be made by inactively monitoring the user’s browsing behavior in excess of time and gathering Web pages in which the user has exposed interest.

The factors such as the total amount of time spent on the page, frequency of visited page and other behavior like book marked pages can be used automatically for collecting such documents. However, based on the user’s behavior of various associations, the user interest score is increased or decreased through contextual data. Once an ontological user profile is built, the core user context can subsequently be utilized for an assortment of information entree activities such as browsing, searching and filtering (Sumathi et al 2010).

2.11 USER PROFILE CONSTRUCTION TECHNIQUES

User profile typically comprises the set of user interested topics that can be customized or modified according to the interests and preferences of the user. The main objective of user profiling is to construct the user profile dynamically according to the requirement of user. The user profile has information which can differentiate one user from a multiple of other users. This may consist of personal information concerning age, name, location etc. It also consists information concerning interests and preferences of a user.
Profiles usually comprise topics of interests but may also contain topics of unconcern by taking into account relevant and nonrelevant documents. Profiles can be static. This means information that specifies a user does not change over time. Similarly, a profile can be active implication and the information specifying a user can be customized or improved to present interests and preferences of the user. To construct a user profile information is able to acquire explicitly (the user gives feedback) or absolutely, where user's interests are not predicted through observing the activity of the user (Singh 2007). User profiles are constructed based on two types such as keyword based and concept based approaches.

2.11.1 Keyword based Approach

In keyword based approach (Crabtree and Soltysiak 1998), user profile construction process is so modest. User profile is constructed with the set of keywords that represent user interest. The user interest is mostly predicted from their feedback provided to the system. The keywords in the user profile are supposed to be more general words. Some feedback mechanisms are deployed to refine the general keywords. Normally, only single keyword is assigned to per interest. It leads to the situation of less sufficient and rich enough information to capture and express the particular interest of the user.

In order to improve, set of top n keywords are selected as a descriptor for every interest of the user. It provides more information than a single keyword based approach but it is too hard to derive a suitable descriptor for the set of keywords by keyword extraction process. A keyword user context reproduces a short term user interest in a definite search session (Daoud et al 2008). It is denoted using the mainly used terms derived from the assumed relevant documents in a specific search session. Particularly, let the query be submitted by a particular user at the retrieval session achieved at a
definite time. Assume that a document reclaimed by the search engine with esteem to query is significant if it produces some noticeable user behaviours. Each document is represented by a term vector where the consequence value of the term in document at a time is calculated using the weighting.

2.11.2 Concept based Approach

Concept based profile (Aroyo and Dicheva 2004) is entirely different from the keyword based profile, the concepts are represented in the form of feature vector that are used to construct user profile instead of keywords (Leung and Lee 2010). For finding common interests of user, a hierarchical profile is favored to a horizontal one. The weight also assigned to the feature vector according to the user preferences on specified topic. The constructed topic ontology also may be static or dynamic. It mainly depends upon the user interest. Type of weighted concept profile encloses big vocabulary to specify a user’s present and future interests. The concepts are constructed in the form of hierarchy based on user's interest which may be specified using either explicit or implicit behaviour. The interest score of a document is determined through the retrieving the keywords from the user browsing history, profiles and bookmarks. The weight is assigned to the keywords or concepts according to the level of interest in that topic.

2.11.3 Multi Strategy Machine Learning Approach

This machine learning approach is a hybrid approach that constructs user profile with system’s user modeling component. It is the first initiative approach that constructs user profile with separate models for both long-term and short-term interests. It also investigates the utility of explicitly modeling information that the system has already presented to the user. News Dude (Billsus and Pazzani 2007) is an intelligent agent that is used to construct the
user profile according to the changing users’ interests through two separate user models such as short-term and long-term interests.

The short-term model is learned from the most recent observations only, whereas the long-term (default) model represents the user’s general preferences. In an active profile, the interests are normally recognized as long term interests (does not change frequently) and short term interests (changes recurrently). In order to group the similar short term interest, Nearest Neighbor Algorithm is evolved to compute similarity measure between two short term interests of the user. Similarity measure is determined by the $TF \times IDF$ i.e. (Term Frequency x Inverse Document Frequency) weighting. It uses cosine similarity measure to compute the similarity of the two vectors.

In the long term interest model, the user interest is modeled from the user general preferences and also it does not change recurrently. In the aspect of modeling the long term interest, probabilistic learning algorithm called as the Naïve Bayesian classifier can model the user general preferences with the set of background knowledge from the set of domain specific features.

### 2.11.4 Interest Based Clustering Approach

In order to perform effective Web personalization, user profile must be constructed with the specific interest of the user. Normally, interest of the user is specified using one or more keywords. It cannot express the interest of the users effectively. In this interest based clustering approach, User Profiles is constructed through actively monitoring the user’s Web browsing and e-mail habits (Crabtree and Soltysiak 1998). It deploys efficient clustering algorithm to group the user interest according to the similarity in between them. Therefore, the user interest is categorized into several interest themes and also that user profiles can adaptively change according to the interest of the users over time.
User profiles are constructed through semantic relationship exists between keywords and then suitable weights are assigned to the keywords according to the user interest. User interest is predicted from the user browsing history through active monitoring of their email transactions behaviour. Information retrieval techniques are used to generate document vectors from the predicted user interest. Generated document vectors are grouped into clusters vectors. Cluster vectors have the same representation as that document vectors. Therefore, same clustering technique is used to group the cluster vectors into theme vectors. This theme vectors are used to represent probably the overall interest of the user and also reasonably assigns the interest of the user with specific interest of the user. Similarity between the interests of the user is tracked for every specific time intervals.

2.11.5 **Enriching Web Directory**

Ontological user profile is constructed in the form of concept hierarchies through enriching the set of categories in directories like ODP (ODP 2009), Yahoo etc. In this type of hierarchical construction of user profiles using concept based approach, Web pages in Web directories are categorized into specific categories and it also used to assist the appropriate page of interest in the appropriate subset (Ravindran and Gauch 2004). The irrelevant information in Web directories is leftover while constructing concept hierarchies.

The more enhanced user profile is constructed through high wealthy semantic relationship between concepts in the ontology. Ontology with high affluent semantic information can lead to serious bottleneck. In order to assign weight to the concept according to the user interest, different methods are applicable. If the user visits the page within the concept hierarchy, the weight for the concept is incremented according to the user preferences. The weight for the concept in the user profile is decremented when negative
feedback is provided by the user. Therefore, it is clearly visible that the weight of the concept mainly depends on the type of feedback either positive or negative feedback provided by the user.

2.12 ONTOLOGY BASED USER PROFILING

Various types of Ontologies are used to construct user profiles. In this section, three types of Ontologies such as Domain ontology, Reference ontology, and Topic ontology are discussed.

2.12.1 Domain Ontology

Concept based user profiles using ontology is efficient in providing more relevant results. Domain ontology plays a key role in conceptualization (Sieg et al 2007). Concept relationship is also compared. Ontology is applied to share information in a particular domain and enables the reuse of domain. Domain or application specific concepts and their relationships could be captured using domain ontology (Bhowmick et al 2010). Ontological user profile approach was proposed to tailor search that engages structuring replicas of user framework as ontological profiles via transferring completely resultant interest scores to presented topics in domain ontology (Sieg et al 2007).

Domain ontology is used as a basic component of semantic knowledge. To get a user context, this semantic knowledge is used. In domain ontology, user behaviours and their relationships are matched with existing concepts using user ontology (Shirazi 2009). This improves developing concept based user profiles. Domain ontology acts as a backbone of Semantic Web. It makes smooth of information distribution and exchange by rendering a conceptualization. In fact, domain ontology confines the valuable knowledge of domain and information resources. In information retrieval
process, there might be many systems utilize domain ontology (Jiang and Tan 2006).

![Figure 2.5 Domain Ontology](image)

**Figure 2.5 Domain Ontology**

Figure 2.5 shows the domain ontology that provides the effective conceptual relation between the set of classes and their associated events in a certain domain of user profile.

Domain ontology has proven to be very helpful in different levels of ontology structure process. This includes information sources with quality, characteristics and ontological richness. Domain ontology specifies the invariant conditions of domain of interests. This domain ontology to concepts ensures the domain specific quality and specifies domain axiom in terms of concepts, relations between concepts and the rules that direct these relations. It provides quality assurance rather than the reuse of domain knowledge and also provides conceptualization of domains (Aroyo et al 2006).

Semantic user preferences may be represented as a vector of weights, representing strength of user interest for each concept in domain ontology (Vallet et al 2006). Categorization can be expressed in a hierarchical structure of class in ontology definition. Domain ontology describes the
concepts in user profiles. Ontology is a type of knowledge base that describes concepts through definitions that are adequately in depth to confine the semantics of a domain. Domain ontology represents domain terms, their relationships and provides a personalized environment to a user. This enriches central domain ontology and it is a specific for domain and performs independent tasks in domain (Gailly and Poels 2010).

2.12.2 Reference Ontology

The Figure 2.6 represents the structure of reference ontology. Concept based user profiles represented in the form of ontology is to represent the relation between the concepts and provide the structured format based on user ongoing browsing behaviours (Gauch et al 2003; Sumathi et al 2010). Rule out the concepts from reference ontology if there are a small number of Web pages sufficiently train the classifier. A classifier is used to classify user browsed documents into concepts. Hence, this type of user profile is known as reference ontology (Sieg et al 2007; Sumathi et al 2010). Concept based user profiles are constructed by classifying the collected documents with respect to reference ontology. User browsed topics that can be extracted from Web pages, explain the semantics of concepts. To identify the user interests in ontological concept based method, reference ontology has been created using concepts from the top level of concept hierarchy. Group of related Web pages are used for classifying the user Web pages into concepts in reference ontology. These pages are treated as instruction data in classifier.

Reference ontology is used in Open Directory Project (ODP 2009) to represent the hierarchy of topics. A term vector is applied to each concept in a hierarchy with the representation of reference ontology. Each concept vector specifies in a hierarchy, all documents indexed under the concept and sub concepts well.
2.12.3 Topic Ontology based User Profile

Topic ontology based User profile is constructed based on the semantic relationship between set of concepts in the domain. Mainly, topic ontology is constructed based on the set of keywords or terms extracted from the user browsing history. Relevance between concepts is determined through determining the semantic similarity between ontological concepts. The user profile is generated from the topics of a user’s interest i.e., search objective. The topic in a specific document encompasses the terms which denotes the subjects (Aroyo and Dicheva 2004).

Based on user needs and search goal, topic ontology based user profile captures the user interested topics. Therefore, topic ontology is constructed to build user profiles by applying interest scores to existing topics in topic ontology. Interest scores based search results in personalized topic ontological user profile is efficient in providing most appropriate results to the user.

Topics and their relations are extracted from Web pages and these extracted topics are given as inputs. A list of Web pages is returned after
requests query to a search engine. If a particular query or topic occurs frequently in Web pages, it can be determined as related topic to the user query. User topics are extracted from clicked documents. Search topics of user and surfing histories are automatically drawn into a set of topical categories.

Topic ontology based user profiling methods intends at getting conceptual needs of a user. A necessity for constructing ontological concept based user profile is to represent the user interests in a hierarchical structure. It represents frequently occurring terms and existing terms. A given concept was considered to be overlapping with another concept if a specific term appeared in term vectors of both concepts (Vilches-Blazquez et al 2009).

2.13 REPRESENTING CONCEPT BASED USER CONTEXT

The goal is to present a semantic description of the user context which denotes the short term user interest associate to a particular search session. These contexts are then collected for learning the long term interests. Three main methods are used to represent the user contexts runs. (i) Keyword user context representation, (ii) Mapping the keyword user context on the reference ontology, (iii) Disambiguating the mapped concepts set using a sub-concepts aggregation scheme, and finally defining the user context by the depth three concepts of resulting set. The major cause for denoting the user context by means of the depth three of the ontology is that are interested to denote a user context collecting information moderately common, and that can be proficient to develop retrieval exactitude for associated search sessions (Speretta 2005).
2.13.1 Representing the Keyword User Context

A keyword user context reproduces a short term user interest in a definite search session (Daoud et al 2008). It is denoted using the mainly used terms derived from the assumed relevant documents in a specific search session. Particularly, let the query be submitted by a particular user at the retrieval session achieved at a definite time. It is assumed that a document reclaimed by the search engine with esteem to query is significant if it produces some noticeable user behaviours. Each document is represented by a term vector where the consequence value of the term in the document at the time is calculated using the weighting.

2.13.2 Mapping the Keyword User Context on the Reference Ontology

In order to extract the main relevant concepts from reference ontology, the keyword user context is mapped on the ontology (Davies et al 2012). It is later used to denote the concepts of depth three user contexts. It needs a collective signification of the reference ontology by calculating term vector of each concept. Each of the URL links is explained by a title and a description that signifies the content of the related Web page.

For each concept, titles and descriptions of the URL is used as enough data for obtaining an accurate categorization of the keyword user context. Note that the mapped concepts collection may have some unrelated elements that do not reproduce the user’s search target. This can be explicated by the information that keyword context terms can be coordinated to numerous concepts belonging to dissimilar portions of the ontology. Certainly, a particular user topic of interest is not precisely matched with a distinctive segment of the ontology, but it can be denoted by concepts extorted from different ones.
2.13.3 Aggregation of Sub-Concepts Scheme

The main aim of this method is to represent the user context with general depth three related concepts subjected from the ontology. This method is too common for signifying the topic of user interests and specifying the leaf nodes to develop the retrieval correctness of associated search session. This disambiguation method is found on the supposition that the significant concepts of depth three which has the maximum number of related concepts in order to the ontology.

Hence, combining the related concepts weights depend on each common concept permits to allocate to the relevant concepts maximum weights. It identifies a cluster of weighted concepts containing a frequent general depth three concept. The related score of the common concept for each cluster is calculated by adding the weights of its associated concepts. Hence, higher clusters will be allocated by greater weights and will be regimented in the top of the concepts relating the user context (Speretta 2005).

2.13.4 Maintaining the User Context in Search Session

The context of the user is automatically updated according to the concepts based on user’s ongoing behaviors. For updating, linear combination formula is used to assign weight to the concept in ontology. The relevant concept ranking of the user context is varies accordingly to the dynamic change of the user context. The user context with according to varying interest is represented for every search session. In order to maintain the user context for every related search session, the steps followed are (i) the weight of possible common concepts are enhanced between two successive user context (ii) the weight of the uncommon concepts are decayed using specific decay factor called as $\beta$. The weight of the new concept in the specific user
context also computed. In order to learn the long term interest of the user, user context in related search session is properly updated for every concept in ontology. The semantic correlation between search sessions is learned to use session delimitation mechanism.

2.13.5 A Context Sensitive Personalized Search

In this level, Web personalization is performed to return more effective search results for every related search session in the user context. The context of the user at specified time is represented using an ordered set of weighted concepts related to the specific query. The associated search results are re-ranked through combining the initial score of the system for each retrieved results. The contextual score is computed based on the similarity measure between the results and related concepts of the user context. Thus, the result score of the concept is computed by combining both the original score and contextual score of the system. Then, results are re-ranked according to the result score

2.14 SPREADING ACTIVATION OVER PERSONAL ONTOLOGIES

Spreading Activation algorithm is mainly used as a cognition model and also a most important concept in semantic network related computation research. It plays a more important role in various applications and also especially in the area of information retrieval. In ontological user profile, Spreading Activation algorithm is used to update and maintain the interest scores of the concepts in user profile. Since, the interest of the user may be varying periodically depending upon the user behaviour and also reflected in the concept of the user profile with weighted interest score. In order to support search engine personalization, interest score is assigned to the weight of the topics in ontology.
This scheme uses a very definite advanced Spreading Activation, for the purpose of conserving interest scores within user profile. Activation value is allocated to the definite topics and other neighboring topics in addition. It is activated founded on an assortment of weighted associations throughout communication. Neighboring activated topics are not obtainable in preference queue will be there added to queue and then reorganized. In this scheme, topic concept and associations are observed.

The search engine ranks pages and concepts in the links are derived from the user profile. Instance between clicking activities is essential as the interest score of a page primarily depends upon clicking. Users may expend much time for their fascinating topics than unexciting topics. After a user acquire an interesting topic and the page has links for matching topics, user might click on the links. If a topic is interested to the users, link may go deeply. Session is initiated to detain the users browsing behaviours through search. Users’ negative preferences are being captured to create more accurate user profile (Dix 2010).

The use of Spreading Activation on automatically built hypertext networks in order to maintain browsing within these networks. Constrained Spreading Activation is used in order to avoid spreading through the entire network. Spreading Activation is achieved on the resulting network in turn to re-rank the search grades of a definite query, also permitting faster integration of newly constructed pages into search results by constructing similarity sets. Whereas this method might improve the effectiveness of searches, it recommends only results at document granularity, where in a number of applications, with task-based interaction, entity-level granularity is far more useful. Moreover, only terms emerging in user queries are measured, which do not inevitably wrap the full span of ontological resources, particularly when the extent of the application is an individual user’s interaction. It is also
importance noting that though the works deem Spreading Activation, they do not agree with the diverse timescales of recall.

Ambiguous queries confuse the search engine and not influence the particular needs of the user. Search engines must present accurate search results to the end user. While queries are concerned to the search engines, they produce the similar results to user queries irrelevant of topical interest. Unlike users may throw query to the search engines that are small and ambiguous. Occasionally the same query might search for various information needs and intention. However, the system will not at all be able to give users’ accurate needs, but it gives in general. Search engine personalization is not effectual on some queries.

The search engine reacts to the set of ranked pages based on the significance of the query. Thus, the search engine produces user profiles to recognize and obtain the users’ actual needs. An efficient user profile construction is an essential task to modify the search engine to revisit outputs associated to the personal interest of a user. Search engine personalization is an energetic research area which agreed with automatic construction of user profiles from the query history and browsed documents. User profiles support search engine to remove ambiguous queries and extract related documents based on users’ interest.

Encompassing a personal ontology that retains the features of interest to the specific user and the correlation between them, it can be able to replicate the Spreading Activation method to recognize the features which will be the interest of the user in a specific context (Katifori et al 2010). The fundamental design can be described as, when the user performs an activity, some entities in the personal ontology can be denoted in the context of this activity.
Suppose the user reads an e-mail, the sender, other recipients of the same e-mail or a project whose name is cited in the e-mail body are such candidate entities. These entities are said to get instant activation. Later, through the relationships established in the ontology, part of this activation can be spread to other connected entities within the ontology. When the algorithm finishes entities that have an adequately high activation, which is over a definite threshold or the top-\(k\) ones will be considered as the main candidates for the user to present succeeding activities.

2.15 REVIEW ON ONTOLOGY BASED SEMANTIC SIMILARITY

In this section, review is carried out on approaches that determine the semantic similarity between the set of topics in ontology. Use of ontology in semantic searching is a best approach that facilitates to return more relevant results.

2.15.1 Similarity Assessment of Topic Ontology

The hierarchical ontology structure is constructed based on two methods such as node based and edge based approach. In edge based approach, ontology is represented in the form of a graph where concepts are denoted as nodes and then relation between the nodes is marked as the link in the graph. Then, the relevance between the two concepts in edge based approach is computed through distance or length between the two concepts in graph. If the distance between two nodes in the graph is small, then high similarity exists between the concepts in ontology. When multiple paths are linking two nodes, in that situation average or shortest distance are only suppose to be consider to compute the similarity between the two concepts.
The major limitation in this approach is that inaccurate assumptions are carried out based on the taxonomy hierarchical structure with variable link densities. Another approach called as node based approach in which similarity is computed conceptually. In order to determine the conceptual similarity, object and part relationship between the concepts are measured. Similarity is mainly computed through identifying to which extent the concepts are sharing their information. If two concepts can share common information, then it is identified that more similarity exists between the two concepts. Therefore, the way to compute similarity is different in both edge and node approaches.

Depending upon the user search intention, Similarity assessment mainly relies on two factors such as Specificity and Exhaustivity. Specificity is defined as the extent where searching is accurately focused on the user required topic. Here, more accurate results are returned to the user. Then, Exhaustivity is defined as the extent where searching pattern paid their attention on the general information or discussion about the required topic. Therefore, common information about the user required topic is returned to the user. Therefore, depending upon the user required searching information whether it may be general or accurate, the searching system chooses to access the relevance of the pattern.

2.15.2 Topic Preference Learning

In order to construct more accurate user profile, must identify the topics which are more suitable for user query and also personalize the results effectively. Already existing approach called as click based topic learning method can predict the interest of the user from the browsing history of the user. In this type of learning method, topics which are supposed to be learn is completely different from the present actual requirement of the user. In that situation, personalized ranking is not that much effectual (Stamou and Ntoulas 2009).
This is mainly due to the personalization that is performed through predicting the fascinated topics from user browsing history. Therefore, it leads to return irrelevant results which are entirely dissimilar from the actual requirement of the user. In order to alleviate the problems in this approach, every user represents their topic preferences during query submission. This type of submitted topic preferences is used to evaluate the performance of personalization method where topic preferences are considered as position truth. Sequentially, score is assigned to the topics according to the number of times the pages related to that topic is viewed by the user. Finally, topics are listed out in the form of descending order from higher score topics to lower score topics.

2.15.3 Ontology based Semantic Similarity

Large number of essential terms and words are spread out through the document. The relevance score of the document is assigned according to the term frequency and equivalency between the requested query and corresponding document. It is very important to evaluate the precision of return results which mainly rely on the similarities exists between queries and retrieve document. It is achieved through assigning tag weights to the concepts in the domain. The order of the document is mostly defined as the total number of times the make of the instance emerge in the document if the document is interpreted with instance or zero. The similarity between the entities is measured through association of semantic relationship between concepts in domain ontology. In order to obtain the automatic retrieval of information through proper filtering mechanism, ontology based concept similarity mechanism is deployed.

The main aspect in ontology based personalized search is that involves modeling the user context as ontological profiles by assigning implicitly derived interest scores to existing concepts in the domain ontology.
A Spreading Activation algorithm is used to maintain and incrementally update the interest scores based on the user’s ongoing behaviour (Sieg et al 2007b). Precision of the returned results to the user mainly relies on the similarities exists between query and document. In order to determine similarity in ontological profile, tag weights are assigned to the concepts in domain. Tag weights are normally determined from the total number of times that feature vector of corresponding words is located as an instance in the document. Initially tag weight of the concepts is initialized as zero. Then, depending upon the locality of concepts, weight is assigned to the concepts in ontology.

The similarity between the concept entities is measured through association of semantic relationship (Resnik 1999) between concepts in domain ontology. In order to obtain the automatic retrieval of information through proper filtering mechanism, ontology based concept similarity mechanism is deployed. This approach merges the ontology similarity distance (Smeaton and Quigley 1996) with explanation scheme to rank the interpreted documents. The ontology based distance method and annotation scheme significantly improves the retrieval performance especially for the top ranked document (Shamsfard et al 2006).

2.15.4 Weighted Profiling Approach

This scheme of a weighted profiling semantic search is depending upon multi valued verdict for extracting related information in order to user needs. The weight of the ontology base domains is computed using the confidence weights represented with the user profile keywords. The assurance weight of every profile keyword is calculated using the frequency number allocated to this keyword through the search process which denotes how it is nearer to user interests. Besides, the proposed scheme utilizes obvious profile, query terms and uses ontology base for achieving its task.
Semantic based search assumes profiling method and using ontology knowledge needs a logical search system. Initializing such an inclusive system is away of scale of this work. This method mainly initiates the important structural elements of such a system that could assist achieving the proposed method and enhances the searching process. The search system involves grouping of three essential, related elements. These elements are query, profiles, and an ontology base. Accumulating such a proposed system in each entity search engine can help search engine for presenting accurate results in order to user requirements (Hikmat et al 2008).

In this approach, a user profile that keeps the keywords and frequencies are used. The keywords denote the user preferences and the frequencies specifying the weights of these keywords. That is each keyword has a frequency number which specifies the number of occurrence of that keyword. These frequencies can be used to give information on how many during his search history.

### 2.16 SESSION BASED ONTOLOGICAL USER PROFILE

This approach presents a search personalization method that is used to construct and maintain the ontological user profile in the form of combined graph for every identical search session. The search session can be defined as a session at a specific time as a series of associated search activities executed by queries submitted correspondingly with time. The user profile has a tree like structure through the links and a non tree like structure through the cross links of various types available in the ontology which is considered as directed graph (Daoud et al 2009a). The event expressed by each search activity when a user submits a query at a time to search engine, it returns a list of ranked documents later, and after this, the user states his interest on document preferences. It produces some observable user behaviours by
assuming the document retrieved by a search engine with esteem to a time of a query is relevant.

The three main stages used for building the user profile are constructing the query over search activity, assigning and maintaining user profile over search session and finding a feasible session boundary on submitting query (Daoud et al 2008).

In the first stage, the initial event is query profiling by gathering user document interest resulted with respect to the query submitted at the time. The query context is created of documents centroid with interest, and the weight of the term is calculated. Then the query context is mapped on ontology using cosine similarity measure. The single term vector is denoted for each concept of the ontology issued from the Web pages. The similarity score is computed for each concept with query context. Based on this, the query profile is concluded using one-hop score propagation. This process can be presented using graph based query profile algorithm. The perception behind this algorithm is to denote the semantically representation of related concepts of the user interest. The interconnected concepts are clustered together in order to provoke a single weighted graph. The query profile is denoted at last by the most relevant graph between the created ones. Certainly, it is presumed that the major relevant graph has a larger number of related concepts, primarily weighted or activated in order to the ontology. For this effort, a graph relevance evaluation is based on integrating the weights of concept nodes of the graph (Sumathi et al 2010).

In the second stage, assigning and maintaining user profile over search session, the user profile is assigned by the graph-based ontological profile when the initial query is submitted in the search session. It sustains the user profile in the similar search session when newly submitted query is
associated to the current user profile. User profile maintaining method is based on merging query submitted and current user profile as follows:

- Collecting the weights of feasible general concepts which permit bringing them to the top of the user profile specification.

- Accumulating nodes and edges to the user profile, which permits captivating into account all feasible concepts in which the user has shown interest in the search session.

In the third stage, session boundary identification method, the conceptual association among user profile and the query is computed using correlation measure. Then the threshold values and queries are assigned from the same session when the threshold is greater. Now, the term based query vector is recorded onto the ontology in order to denote the concept vector. The query vector and its similarity have been computed. The association values are in the range of [-1…1], where an association value nearer to -1 that the query and the user profile are not comparable, and a value nearer to 1 then they are highly related (Daoud et al 2008).