Ontological Approach for Effective Generation of Concept Based User Profiles to Personalize Search Results

[1]Prabaharan, S. and [2]R.S.D. Wahidabanu
[1]Research Scholar, Anna University, Chennai, India
[2]Department of ECE, Govt. College of Engineering, Salem, India

Abstract: Problem statement: Ontological user profile generation was a semantic approach to derive richer concept based user profiles. It depends on the semantic relationship of concepts. This study focuses on ontology to derive concept oriented user profile based on user search queries and clicked documents. This study proposes concept based on topic ontology which derives the concept based user profiles more independently. It was possible to improve the search engine processes more efficiently. **Approach:** This process consists of individual user's interests, topical categories of user interests and identifies the relationship among the concepts. The proposed approach was based on topic ontology for concept based user profile generation from search engine logs. Spreading activation algorithm was used to optimize the relevance of search engine results. Topic ontology was constructed to identify the user interest by assigning activation values and explore the topics similarity of user preferences. **Results:** To update and maintain the interest scores, spreading activation algorithm was proposed. User interest may change over the period of time which was reflected to user profiles. According to profile changes, search engine was personalized by assigning interest scores and weight to the topics. **Conclusion:** Experiments illustrate the efficacy of proposed approach and with the help of topic ontology user preferences can be identified correctly. It improves the quality of the search engine personalization by identifying the user's precise needs.

Key words: web personalization, web usage mining, ontological approach, personalized search, user profiling, search engine

INTRODUCTION

Web Search engines do an excellent job when the queries are understandable and exact. Generally, user queries are short, ambiguous and not well-formed in nature (Silverstein *et al*., 1999; Cronen-Townsend and Croft, 2002; Jansen *et al*., 2000). Ambiguous queries confuse the search engine and not satisfy the specific needs of the user. Search engines should provide precise search results to the end user. When queries are issued to the search engines they return the same results to user queries irrespective of topical interest or context. Different users may send query to the search engines that are short and ambiguous. Sometimes the same query may search for different information needs and purposes. But the system will never be able to provide users' precise needs but it provides in general. Personalization of search engine is not effective on some queries. Search engine respond to the list of ranked pages based on the relevance of the query. So that search engines generate user profiles to identify and get the users' actual needs. An effective user profile generation is an important task to customize the search

engine to return outputs related to the personal interest of a user (Shen *et al*., 2005; Dou *et al*., 2007). Search engine personalization is an active research area which deals with automatic generation of user profiles from the query history and browsed documents. User profiles assist search engine to eliminate ambiguous queries and retrieve relevant documents based on users' interest. Effective user profiling strategy has to play key role in search engine personalization.

To overcome this problem, an ontological approach is proposed to optimize the relevance of search engine results. This study presents user profiles by giving interest scores to the topics of topical ontology. User interest may change over time so that profiles are maintained and updated. An improved recommendation system using Profile Aggregation based on Clustering of Transactions (iPACT) shows better prediction accuracy than the previous methods PACT and Hypergraph (Almurtadha *et al*., 2011). To maintain the interest scores of the user profile, spreading activation algorithm is proposed to analyse ongoing browsing activities (Sieg *et al*., 2007a). Ontology affords well meaningful structure to relate user interests and a

Corresponding Author: Prabaharan, S., Research Scholar, Anna University, Chennai, India

wealthy conceptualization among interested topics and permits latest interested topics into the structure (Gauch *et al*., 2003). The access time, browsed pages and mouse activities may determine user interests and content of the document may contain the topics of the user's interest (Bhowmick *et al*., 2010). Topic ontology is utilized for calculating users' topic preferences (positive or negative preferences) based on their queries and visited pages. To classify the user's recent topic preferences exploit semantic similarity amid users' present query and similar query pages. Calculate similarity and interest of query to the topics based on the viewed pages. Ranking function utilizes the preference topic in order to rank the search result. All users might not have equal interests but they can have a little amount of topic preferences in which they have shown their interests. User preferences give major improvements of the search results quality. User interests may be identified by watching user's surfing activities over period (Stamou and Ntoulas, 2009).

The user activities are matched with presented topics in topic ontology and relations among the topics. At first, topics scores in profile would keep on varying. Though, the change in interest scores must be reduced once sufficient information collected for profiling (Pretschner and Gauch, 1999). Depth of ontology is common to signify the user interests for related search activities. Calculated interest score can be assigned by accumulating weights of its topics. Increasing use of web search engines requires mechanisms to select finest matches based on the users' need. Search session allows search queries over a period of time. The user profile is embodied using user search record in a search session. User profile is constructed over a search session to personalize search (Daoud *et al*., 2008).

Previous work: The existing profile-based personalized search approaches are not consistent when compared to click-based method. To allow large-scale assessment of personalized search, an evaluation framework was developed on query logs and this approach has improved search precision on few selected queries but damage other queries and far from finest search (Dou *et al*., 2007). Existing personalized search includes constructing replicas of user framework as ontological profiles through giving interest scores to present ideas in domain ontology. To sustain the interest scores based on the user's enduring activities, spreading activation algorithm is used. Since, it is focused on inherent models for user profiles, profiles have to be adjusted over period (Sieg *et al*., 2007b).

Existing strategy uses user framework to tailor search results by re ranking the outputs returned from search machine. Framework model of user is characterized as examples of indication domain ontology in which ideas are interpreted by interest scores copied and simplified completely based on user's requirements (Gauch *et al*., 2003). In existing ontology oriented user replica approach is proposed in the framework of personalized information access. Static user profile specifies user's interests in a focused way. Dynamic user profiling includes the characteristics of flexibility into it using hybrid method. During the browsing sessions dynamic user profile employs data sources such as usage log and mouse operations. These usage logs are considered to score the concepts. The concept age monitor monitors concepts usage in the user profile. The concepts score that were not considered by the user are reduced (Bhowmick *et al*., 2010).

A user profile is generated over time by exploring browsed sheets to identify content and time. In this, the size of a browsed page may be ignored when the interest in a page is inferred (Pretschner and Gauch, 1999). In existing system, search machine returns document contents based on grouping of identical keyword and concept. Documents are categorized to establish the concepts to which they coordinate. Document contains both identical keyword and concept considered as irrelevant by semantic or content. It returns minimum number of relevant documents for each query and many documents belong to topic is irrelevant (Gauch *et al*., 2004). Many user profiling approaches are evaluated and these make use of click through data to extract from Web-snippets to construct concept-based user profiles. Users' positive and negative preferences were captured. But relationship among users and concepts are not performed. Concept-based user profiles are not integrated with search engine ranking (Leung and Lee, 2010). Existing 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. The most recent developments of mechanize ontology construction and manuscript explanation are potential. For example, proposed 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 (Castells *et al*., 2007). Existing search process integrates users' interests to get better search results. User profiles are ordered as a concept hierarchy and it allows automatic creation of

huge structured user profiles. The length of a browsed page is ignored when the interest in a page is inferred with the proposed strategy (Vallet *et al*., 2006).

An approach was proposed to tailored search that engages structuring replicas of user's framework as ontological profiles via transferring completely resultant interest scores to presented topics in domain ontology. Stability of the user profiles are not evaluated Sieg *et al*. (2007a). Most existing system proposed spreading activation algorithm with domain and reference ontology. This research addresses these problems by proposing topic ontology with spreading activation algorithm. By assessing user browsed pages, user profile is generated over time to identify users' content and time spent on it. When pages are repeatedly visited by the user, it embodies user's interest in subject. The objective is adapting search outcomes to a particular user based on user's interest and topic favourites. Session is introduced to capture users' browsing activities and profile of user refers to interest of user in a specific search time. Spreading of interest scores activate related topics and continued in same search period. This study explores how the user profiles attain the improvements of search engine performance.

MATERIALS AND METHODS

This study builds user profile from user interested topics. A topic consists of various concepts, which form the next stage of the ontology and a concept may belong to one or more topics. The concepts are interrelated within ontology by topic semantic relationships. More accurate information about the users' interests could be done based on users' surfing behaviours. Positive or negative weights of the concepts specify the interestingness or uninteresting ness of the user. Spreading activation algorithm is applied to preserve and modify users' clicking and browsing based on users' ongoing surfing activities and it updates the interest score of the topics. The main function of the spreading activation is scoring user interests, finding negative and positive preferences. It also evaluates session based user interests and topics similarity. Spreading activation is a procedure for retrieving and ranking related information by activating query items and applying their activation along interrelated topics. Profiling can be done based on the search session and change in user interest. User search history denotes profile of user in a period of search. User profile is started through topic ontology for primary query of the session. New topics can also be identified and added to

the profile according to their browsing session. The interest score weighting was chosen to provide weighted topics in the topic ontology.

Problem statements: It is difficult for clients to discover more appropriate information for their search query. This is because of increase of maximum users of internet and amount of web pages. It takes time to search results for users' particular needs. To get relevant information users should go for a public search engine and need to submit their query. But this is also rendering most irrelevant results. So users may be confused with the results and the problem arises here because of not providing users' actual needs in a well formed structure. Previous methods are not accurate in capturing user interest and profiling is not reliable. User interest may (or may not) vary from session to session. Creating of only one user profile and apply the same to all users is not reliable and there will be a problem of providing users actual need. Many user profiling approaches consider only on users positive preferences but they neglect the unclicked documents whenever a web page returns with ranked results. To overcome these problems and meet the users' actual needs and get accuracy of user profiles over time, a topic ontology method is proposed with advanced spreading activation algorithm. It constructs user profiles over time and based on search session. Interest score is propagated to the topics for identifying the weighted topics that can be updated whenever a new (or existing) keyword or query is issued. Negative preferences (i.e.,) users uninterested topics can also be captured.

Proposed work: This research aims at evolving search engine personalization by proposing topic ontology to classify the web pages based on users' content. This study constructs topic ontology with hierarchical relationship among concepts and propose advanced spreading activation algorithm to calculate the interestingness and uninteresting ness of a particular user. In topic ontology, topics are structured hierarchically. More exact topics are presented in the hierarchy and categorize numerous numbers of pages into topic ontology. This directory computes semantic associations among huge records of web pages and topics (Dou *et al*., 2007). It uses spreading activation algorithm with construction of topic ontology to increase in updating topics interest score in user profiles. It is acting as a semantic network. Based on activation values, interest scores are restructured. Spreading activation is used to get relevant topics in topic ontology by giving primary concepts and equivalent primary activation values (Bhowmick *et al*., 2010).

This proposal uses a very specific advanced spreading activation, for the intention of preserving interest scores inside user profile. Activation value is assigned to the specific topics and other adjacent topics as well. It is activated based on collection of weighted relationships throughout transmission. Adjacent activated topics are not presented in precedence queue will be inserted to queue and then restructured (Bhowmick *et al.*, 2010). In this study, concepts of a topic and the relations are monitored. The search engine ranks pages and concepts in the links based on the user profile. Time between clicking activities is important since the interest score of a page mainly depends upon clicking. Users may spend more time for their interesting topics than uninteresting topics. When a user gets an interesting topic and the page has links for corresponding topics user may click on the links. If a topic is interested to the users, link may go in depth. Session is introduced to capture the users browsing behaviours during search. Users' negative preferences are also being captured to generate more precise user profile.

Proposed algorithms: Nowadays, thousands of users search for number of topics and they can have different expectations. User interests can also be changed over time. But web search machines need to identify and suit user expectations effectively. Therefore, search engines personalize the outcome to give interested topics based on their search session. Search engine generates profile for each and every individual, if profile is available for users' query; it ranks topics and returns relevant pages back (Dou *et al.*, 2007). Construction of topic ontology with advanced spreading activation algorithm is to (i) compute topics similarity for evolving user profiles (ii) identify the users' accurate topic preference by assigning interest scores (a) capture the users negative preferences and (b) construct profiles over the session for search engine personalization.

Topic ontology construction: Topic ontology is a graph in which each node represents a topic. It contains the concepts that are organized by semantic relationships. A hierarchical relationship is maintained among the concepts in the topics. In this study, the topics of a user's interest are used to construct user profiles. The user profile consists of the topic's semantic relationship with the use of ontological approach. Topic ontology is built from some terms or keywords. Terms consist of smallest concepts. Topic relevance is calculated through semantic similarity amid ontological concepts. User profile is generated from users' interested topics (i.e.,) search intent. Similarity amid concepts is represented through extent to which they distribute information (Zhou *et al.*, 2006a; 2006b).

The topic ontology is constructed in the following way: Define the keywords and their frequencies. Set of keywords that can be represented as K = (k1, k2,…, k3) and keyword frequencies can be of kf (d, k). Here k in d that is document d and a keyword k. Keyword frequency sets S = {(k, f) │ kεK, f = kf (k, d) > 0}. Here, S is a pattern. Let keywordset (S) = {k │ (k, f) ε S} be the ε set of S. Given a pattern S = {(k1, f1), (k2, f2),…, (kn, fn)}, its usual form {(k1, w1), (k2, w2),…,(kn, wn)} is represented by Eq.(1).

$$w_i = \frac{f_i}{\sum\limits_{j=1}^{n} f_i} \text{ for all } i \le n \text{ and } \ge 1 \tag{1}$$

Topic ontology is represented by T. Relational weight resolves the measure of association between two concepts. Hierarchy is formed by recognizing is-a relationships between the concepts. Hierarchical structure presents an understanding of the relations. If S1 = S2 then 'is-a' association exist between S1 and S2. The hierarchy of all keywords in K can be obtained. T is called a group of primitive objects. The concepts are constructed from the primitive objects and ontology contains the primitive and compound classes and these are inherited by resultant classes (Zhou *et al.*, 2006a; Li and Zhong, 2004). Keywords k1 and k2 score function f through a relative r, depends on an Association Score (AS) between keywords and relation weight. Association score of keyword match up (k1 and k2) is represented through frequent occurrence of keyword. This is given in Eq.(2):

$$AS(k1, k2) = \frac{\log(p(k1, k2) + 1)}{Nf(k1).Nf(k2)} \tag{2}$$

Here, p(k1, k2) represents the probability of keyword pair (k1, k2) and Nf (k) is a normalization feature specifies the amount of keyword minds that exists in keyword k (Stamou and Ntoulas, 2009). Associations initiate relations between topics by generating associations. The most commonly used associations are binary. Associations can include any number of topics and are then said to be "n-ary". Association in the top of ontology is characterized by an association group <support, β> from T such that β(S) = {(k1, w1), (k2, w2),…, (kn, wn)} and β(S) is S's usual type. Association group charts a prototype to a keyword set and presents keyword weight for keywords in a keyword set (Zhou *et al.*, 2006a). Some patterns are discovered from relevant documents that are corresponding to a group of keyword occurrence couples:

d1 = {(java language, 4), (programming, 6)}
d2 = {(c++ language, 5), (programming, 15)}
d3 = {(OOPS, 3), (programming, 7), (others, 10)}

Compound objects are obtained using is-a relationship e1 and 2 from d1-3, where d1→e1, d2→e1 and d3→e2. e1 and e2 are the expanded patterns. Arrow represents the "is-a" relationship in the following fig. The user profile includes a hierarchical structure made of "is-a" links. Here C++ and Java languages are belonging to OOPS also program relates to computer (Li and Zhong, 2004).

Compute topics similarity for evolving user profiles: Figure 1 depicts the user profiling methods based on topic ontology. In general, for each topic, topic ontology provides a way to classify relevant documents. Relevance documents are accessible once, proper performance replicates various characteristics of efficiency of topical search can be determined. Use the Topics Similarity (TS) standards specified by topic ontology to calculate similarity between queries and estimated topics and obtain the average similarity for those topics. The Semantic Similarity (SS) of topics can be calculated between the Sets of Topics (SOT) and the Expected Set of Topics ESOT (Stamou and Ntoulas, 2009) using Eq. (3):

$$SS(SOT, ESOT) = \frac{1}{n1n1} \sum_{\forall Te \in SOT \forall Ta \in ESOT} TS(Te, Ta) \qquad (3)$$

where, nl indicates total number of topics measured. Using cosine similarity measure calculates a keyword vector for document di for each topic Tj in user profile. According to this measure, the similarity among pages in topics that belong to various top-level groups is zero though the topics are obviously relevant. Thus, this measure is used to derive semantic relationships among thousands of web pages stored in this topic ontology (Sieg *et al.*, 2007b). Given topic Tj its similarity with SOT is calculated by Eq.(4):

$$Topics\ score\ (Tj) = \cos\ (di, SOT) \qquad (4)$$

Topic list computes semantic associations amongst huge numbers of topics pair. Categorize the collection of relevant documents for a specified topic in turn to resolve if a topical search is effective. Categorization of clicked pages into topics and their semantic relationship derived from topical ontology.

Identify the users' accurate topic preference by assigning interest score: User interest is constructed for a particular query. Spreading activation updates topics interest score in user profiles.
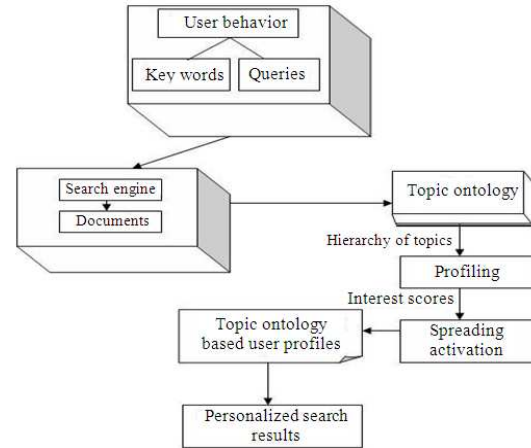

Fig. 1: Topic ontology representation

Initially, spreading activation gets related topics, primary group of topics and its related primary activation values. Hence, topic ontology for user profile acts as semantic web and interest scores are rationalized as per the activation values. Spreading activation methods specify the particular relations between keywords or topics. Obtain interest score in which user has exposed interest via observing user browsing behaviours. Collect the weights of possible topics that can be brought to the peak of user profile demonstration (Sieg *et al.*, 2007a).

Algorithm has primary group of topics from topic ontology based user profiles. Initial activation value is assigned to those topics in user profile. Key target is to trigger other topics following a collection of weighted relations throughout circulation. Finally get a collection of topics and relevant actions. Given topic circulates its activation to its adjacent and to find the activation throughout the network weight of relation between source and terminal is calculated. For each topic the initial activation value is reset to zero in the user profile. Topics similarity score sim (di, Tj) is greater than zero are inserted in a precedence queue, but it is in non rising sequence accordance by topic activation values. Activation rate of topic Tj is consigned to IScore(Tj) _sim(di, Tj), where IScore (Tj) is obtainable interest score to a particular topic.

Uppermost activation rate for topic is deleted from priority queue. Activation quantity that is spread to each adjacent is relative to the weight of the relation. Activated adjacent topics are not presented in precedence queue are then appended to queue and recorded. The process continues until no more topics to be dealt with further. Adjacent spreading topics are measured to be the related topics. Related topics are triggered are appended to a precedence queue, then arranged with activation rates (Sieg *et al.*, 2007b; Haveliwala, 2003).

Spreading activation algorithm:

Input: Topic Ontology for user profile with interest scores and a collection of topics
Output: Topic Ontology for user profile topics with modified activation values
SOT = {T1, ..., Tn}, topics with interest scores
IScore (Tj), interest score
IScore (Tj) = 1, interest information is not accessible
For each di ε I do
Initialize priorityQueue;
   foreach Tj ε SOT do
     Tj.Activation = 0; // activation value reset to zero.
end
   foreach Tj ε SOT do
     Calculate sim(di,Tj);
if sim(di, Tj) > 0 then
     Tj .Activation = IScore (Tj) _ sim(di, Tj);
     priorityQueue.Add(Tj);
else
     Tj .Activation = 0;
end
end
while priorityQueue.Count > 0 do
Sort priorityQueue; // activation values (descending)
Ts = priorityQueue[0]; // first item(spreading concept)
priorityQueue.Desqueue(Ts); // eliminate term
if passRestrictions(Ts) then
     relatedTopics = GetrelatedTopics(Ts);
foreach Tl in relatedTopics do
 Tl.Activation+ = Ts.Activation _ Tl.Weight;
 priorityQueue.Add(Tl);
         end
       end
     end
end

Spreading activation is assigned to input keywords and activation is then sent from node to node throughout the network, over a number of cycles. Web users likely to be submitted the same queries several times. In this study, users' current query is matched with existing queries. Based on users' classified and hierarchy of pages respect to their topics underlie a new user query in estimating the topic preference (Stamou and Ntoulas, 2009). Interest scores for topics are modified using spreading activation. Activation spreads to a wider of topics. Alternatively the huge related set of a topic gets several related pages. Scores of the weight are calculated respect to various topics (Haveliwala, 2003).

Activation weight of word is grouping of topic weights in user queries along with documents correspondingly. Primary activation load propagates via a group with relations initiating at preliminary node. At last, every document node is triggered using topic weights of all topics exist in the document (Salton and Buckley, 1988). For instance, when user searches for information can verify topics of user interests and frequencies. Search engine retrieves a document list that can be attained using keyword in search process. Similarity of relations of user interests can be evaluated and obtain the documents with set of related topics. In particular, each topic and relation in topic ontology would give particular values for representing user interests (Jiang and Tan, 2006).

Capturing negative preferences: Negative preferences may include unclear or inconsistent topics in topic ontology. The users' negative preferences can be captured by considering unclicked pages. If interest score is not assigned for a topic or negative score that can be represented as a negative preference. In this preference, pages can be searched but not clicked and visited by the user that pages may be uninteresting or unrelated topics to the users. Given a set of results for users query, if topic ti, tk are clicked and topic tj is not clicked rank between ti and tk (ti < tj < tk) then topics T (tj) in topic tj is considered less relevant to topics T (tk) and T (ti). Negative preferences are considered as an irrelevant to the user. If interest score is negative (i.e. IScore (Tj) = -1) then no interest information available and if sim (di, Tj) < 0 then no similarity between topics. Finally, negative weight can be added in a queue (Queue.Add (Tj)) (Leung and Lee, 2010).

Constructing the user profile over a search session: This study generates topic ontology based user profiles for their surfing behaviour and giving interest score for the topics in a particular session. Using topic similarity measure, session activity states the following. If user issued query to a search machine at time t, search engine returns a top ten ranked list. Construct user profile in search period based on users interests by clicking and viewing them and it can also be maintained in same period. Once a new query is submitted it identifies a possible session for generating the topic ontology based user profiles. Ranking can be done in same search period based on highest interest scores with topic similarity. Accumulating nodes and edges to user profile allocates topics where user is interested in search session (Daoud *et al.*, 2008).

When a server gets user query and finishing when user quit the website or session timeout is called as session. In session, information is collected from users by the web server. Calculate the query frequency occurrence in a particular session by giving interest scores to them. Session is calculated as the total time for the user to complete a set of transactions. Group of user requests are the structure of URL called as user session. Browsing time is updated based on each user's information access. The request session of each user is transformed into an HTTP request and it is connected within the web server (Speretta and Gauch 2005).

Generally, personalization can be done by creating and maintaining sets of user's interests, stored in profiles to give better results. To get effective personalization, these profiles differentiate between lengthy term and small term user interests. Purpose of user profiles indicating user's preferences rather than user's interests. User search histories are often used to generate interest profiles. User profiles are characterized as topic weighted hierarchies where topics are defined by topic ontology. Search outcomes are also categorized into same topic structure based on session. To compute topical similarity between document and user's interests, document profile is evaluated to user profile (Sampath *et al*., 2004). According to queries frequency occurrence, a Search Session (SS) is defined as in Eq.(5).

$$SS = \{qf1, qf2, qf3\ldots qfn\} \qquad (5)$$

RESULTS

In this study, midstream was developed to group all user information in order to perform the evaluation of user profiling methods. The midstream is intended at facilitating experimentation. The test queries are erratically chosen from 5 different topical classifications to avoid preconception. Table 1 illustrates the topical classification in which the test topics are selected from. The methods developed in this study can be incorporated into any search engine to present personalized user profiles.

There are 100 test queries used to have indistinct representations in Table 2. Human judges determine a typical group for each query. To ensure for their

accuracy, the topics obtained from the above algorithms are evaluated. To explore for the answers of the 100 test topics the 5 users are requested to use midstream. Top 50 results are returned to the users when query is submitted to midstream and users click on the outputs and then find relevant to their queries. Statistics of the clicked data are collected and according to users' needs the user profiles are exploited to collect similar topics together.

Each topic's profile was described by the first 5 related documents carried out and scheduled in Table 3. It is observed that upgrading is much greater when user profile is generated using top ordered documents returned by the search machine regarding topic. This proves that proposed topic ontological user profile reaches an efficient search engines personalization. A session oriented assessment set-up incorporates search period as a series of subtopics produced for a particular topic is also defined. User profiling was evaluated successfully and proposed approach is effective. The performance is calculated by giving average search precision values.

Experimental evaluation: The topic preference pairs obtained from Spreading Activation, Topics similarity and Weight of the topics are evaluated. Then user profiling strategies using those methods over a session are compared. Finally, the performance of the topic ontology with spreading activation algorithm that relies on following things. The effectiveness of search engine personalization within civilizing the excellence of search engine outcomes are evaluated.

Table 1: Topical classification

| Tour package |
| Recipe |
| Computer games |
| Online book stall |
| Geo Informatics Research |

Table 2: Statistics of the clicked data

Sample users	5
Sample test queries	100
Each user query	20
Retrieved URLs	1000
Retrieved topics	2164
Top results retrieved in URLs	50

Table 3: Average search precision (in percentage)

Topics	kf *df (%)	TS = cos (di, SOT)(%)	log(p(k1,k2)+1)(%)	Topic ontology (%)	Evolving (%)
1	82.0	90.0	94.0	100.0	100.0
2	90.0	94.0	94.0	94.0	82.0
3	42.0	38.0	46.0	40.0	38.0
4	98.0	94.0	98.0	94.0	94.0
5	50.0	66.0	90.0	70.0	78.0
Average	72.4	76.4	86.0	79.6	78.4

Table 4: Average precision for spreading activation, topics similarity and weight

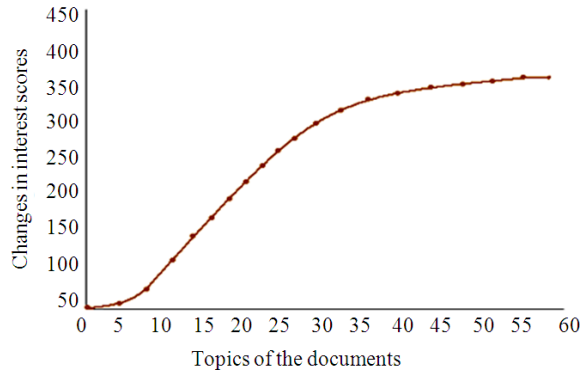|	Average precision
Spreading activation	0.6731
Topic similarity	0.7124
Weighted topics	0.7916

Fig. 2: Changes in Interest scores

To evaluate the efficiency of search engine personalization, query and click sessions are recorded. Then the user profile is evaluated over search session. The user profiles are employed by the similarity method to group similar topics together according to users' precise needs. User profiling methods that integrate negative topic weights return execution positions are extremely close to the optimal points obtained. The finest Interest scores are compared to the standard topic scores using Eq. (6) and (7):

$$\text{Precision (t)} = \frac{\left|\text{T related} \cap \text{T retrieved}\right|}{\text{T retrieved}} \qquad (6)$$

$$\text{Recall (t)} = \frac{\left|\text{T related} \cap \text{T retrieved}\right|}{\text{T retrieved}} \qquad (7)$$

Where:
t = Input
T related = Collection of topics that exists in topic ontology for t
T retrieved = Collection of Interest scores generated by the spreading algorithm

The precision and recall are averaged to design and comparing the effectiveness of the user profiles. Table 4 illustrates the Average Precisions of Topic Preference pairs obtained using Spreading Activation, Topics similarity and weight of the topics.

Evaluating users' accurate topic preferences using spreading activation: Figure 2 illustrates the

evaluation of effectiveness of topic preferences using spreading activation by assigning its interest scores. Interested topics scores in topic ontology oriented user profiles are modified whenever user shown importance in a new web page. Interest topic scores in a profile will keep on changing. Interest score in a user profile was assigned to zero then measure the interest score changes for topic and other topics as well. Finally interest scores are recorded.

Clicked documents are used in profile set for the experimentation. While verifying interest score to users' topic use the user interests from topical group and semantic relevance of topic preferences. While ranking the pages it exploits user topical preferences from users' click record with keywords to recognize the possible topic of a query. Initially, use topic ontology to identify the visited pages' topics and later individual topic preferences are measured. User profiling is highly independent on the categorization in allocating a proper topical group to viewed pages when compare to user profiles in reference ontology.

Topical hierarchy has improved the effectiveness of system that interprets the users' topical preferences and captured the users' topical interests. Topics and interests achieved by the topical ontology have improved relatively to search averaged over topics. Upgrading of interest scores 350 is achieved correspondingly at spreading activation. Evaluation results prove that search engine performs better personalization when the search results are ranked.

Evaluating the topics similarity for evolving user profiles: Figure 3 shows the similarity of topics measure. If topic preferences are valued higher interest scores then it is removed from the priority queue. According to the updating of interest scores, it can be appended to the queue. By the occurrence frequency of keyword the Association Score (AS) of keyword brace (1 and k2) is described. Particularly, associate topic preferences of web pages with mouse clicks on search outcomes. Finally, search outcomes are ranked based on users' topical interests. Topics weight was executed after viewing every page and not later the execution of user session. Average of relevant topics in user profiles ordered by weight of the topics and number of web pages related to topics. Similarity score of similar topics is calculated in the chart based on topic preferences using cosine similarity calculation.

Figure 4 depicts the average weighted topics. Some topics may or may not have same weight. Weighted scores are calculated with respect to various topics. High weighted and less weighted topics are represented as user interestingness and uninterestingness.
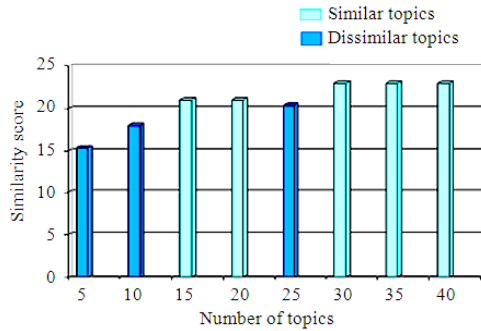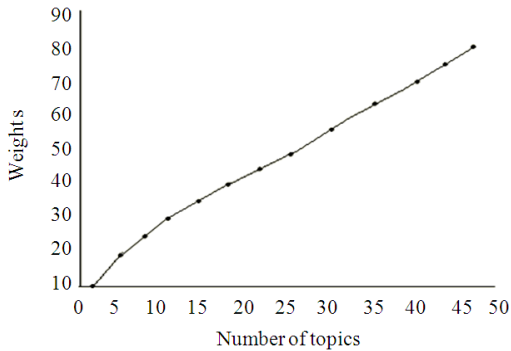
Fig. 3: Similarity measure

Fig. 4: Topics weight

Table 5: Topics with activation score based on session

Topics	Session based activation score
Drawings	30 21 19 44
Recipe	45 53 67 14
School details	65 74 581 93 99
Politics	43 76 57 43

Table 6: Representation of topic preferences

(Sony)	Computer	Television	Mobile	Food
Spreading activation				
Browser 1	1	0	1	-1
Browser 2	1	1	1	-1
Browser 3	1	0	1	-1
Topics similarity				
Browser 1	1	0	1	-1
Browser 2	1	1	1	0
Browser 3	1	1	1	-1
Weighted topics				
Browser 1	1	1	0	0
Browser 2	1	1	1	0
Browser 3	1	1	1	-1

Evaluating the user profiles over the search session:
It is constructed based on score that spreading on set of related topics and uphold in unchanged search session. It recognizes session limitations by semantic relationship measure between topics with keyword relations and permits entering new interested topics. The session based activation score of topics is tabulated in table 5.
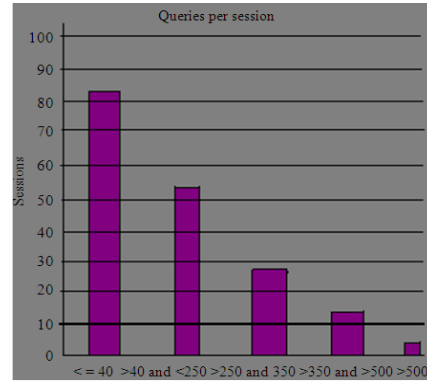

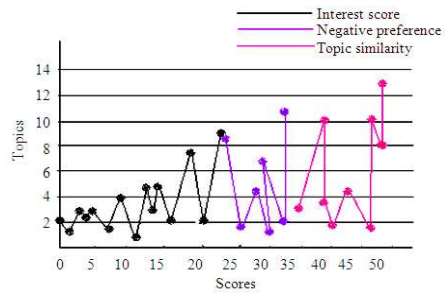Fig 5: Queries per session

Fig. 6: Interest score, negative preferences and topic similarity illustration

For each user session lowest and greatest number of requests has been found. The graph (Fig. 5) explains that majority of the sessions contain small number of (less than or equal to 85) user requests. In this study sessions may have more than 100 user requests. This plays an important role to learn regarding the users browsing behaviours.

Comparing obtained preference pairs for positive, negative preferences and topics similarity: In this experimental setup, evaluate the comparison between the obtained preferences for positive, negative and topics similarity. These are exploited to measure the topic preference pairs from the keyword occurrence frequency. The measured topic preference pairs from different methods are evaluated to get the portion of correct user preference. Then the topic with high interest score in the resulted topic preference is removed from the queue to avoid uncertainty.

The Fig. 6 illustrates the precisions of the topic preferences, negative preferences and topics similarity. From 14 different users the average precisions are obtained. Interest scores and negative preferences are 9-11%. Topics similarity is 13%. Thus, it is capable of finding out more accurate negative preferences. Any changes in similarity values will be updated.

With more accurate negative preferences, a more dependable set of negative topics can be determined. Table 6 demonstrates the topic preferences (i.e., similar and dissimilar topics). The browsing similarity collected from 3 users is shown. The -1 represents similar topics and 0 represents the dissimilar topics.

CONCLUSION

Search engine personalization based on topic ontology was presented for concept based user profiles which identified the user's search interests. This approach works on user's visited pages, also catches users' negative preferences and it can provide an efficient search engine personalization. A major benefit of proposed approach is to calculate more appropriate relevant topics to relate the contents with search profiles of users. Profile shows the user attention in a particular search session. Profiles are modified over time and guaranteed the updates to interest scores exactly. Evaluation results reveals that topic ontology for concept based user profiles attain accuracy of search results. Similarity of the topics is calculated using cosine similarity function. The user profiles were evaluated using different methods and compared with each other. Experimental result shows that profiles capture positive and negative preferences of users. Despite getting better quality of search engine personalization, negative preferences in proposed system also help to detach related and unrelated topics.

REFERENCES

Almurtadha, Y., M.N.B. Sulaiman, N. Mustapha and N.I. Udzir, 2011. IPACT: Improved web page recommendation system using profile aggregation based on clustering of transactions. Am. J. Applied Sci., 8: 277-283. DOI: 10.3844/ajassp.2011.277.283

Bhowmick, P.K., S. Sarkar and A. Basu, 2010. Ontology based user modeling for personalized information access. Int. J. Comput. Sci. Appl., 7: 1-22.

Castells, P., M. Fernandez and D. Vallet, 2007. An adaptation of the vector-space model for ontology-based information retrieval. IEEE Trans. Knowledge Data Eng., 19: 261-272. Doi: 10.1109/TKDE.2007.22

Cronen-Townsend, S. and W.B. Croft, 2002. Quantifying query ambiguity. Proceedings of the 2nd International Conference on Human Language Technology, (HLT' 02), Morgan Kaufmann Publishers Inc. San Francisco, CA, USA., pp: 104-109.

Daoud, M., L. Tamine-Lechani and M. Boughanem, 2008. Learning user interests for a session-based personalized search. Proceedings of the 2nd International Symposium on Information Interaction in Context, (IIC' 08), ACM New York, NY, USA., pp: 57-64. DOI: 10.1145/1414694.1414708

Dou, Z., R. Song and J. Wen, 2007. A large-scale evaluation and analysis of personalized search strategies. Proceedings of 16th International Conference on World Wide Web, (WWW'07), ACM New York, NY, USA., pp: 581-590. DOI: 10.1145/1242572.1242651

Gauch, S., J. Chaffee and A. Pretschner, 2003. Ontology-based personalized search and browsing. Web Intell. Agent Syst., 1: 219-234.

Gauch, S., J.M. Madrid, S. Induri, D. Ravindran and S. Chadalavada, 2004. KeyConcept: A Conceptual search engine. University of Kansas.

Haveliwala, T.H., 2003. Topic-sensitive pagerank: A context-sensitive ranking algorithm for Web search IEEE Trans. Knowledge Eng., 15: 784-796. DOI: 10.1109/TKDE.2003.1208999

Jansen, B.J., A. Spink and T. Saracevic, 2000. Real life, real users and real needs: A study and analysis of user queries on the web. Inf. Proc. Manage., 36: 207-227. DOI: 10.1016/S0306-4573(99)00056-4

Jiang, X. and A.H. Tan, 2006. Learning and inferencing in user ontology for personalized semantic web services. Proceedings of 15th International Conference on World Wide Web, (WWW'06), ACM New York, NY, USA., pp: 1067-1068. DOI: 10.1145/1135777.1136018

Leung, K.W.T. and D.L. Lee, 2010. Deriving concept-based user profiles from search engine logs. IEEE Trans. Knowledge Data Eng., 99: 969-982. DOI: 10.1109/TKDE.2009.144

Li, Y. and N. Zhong, 2004. Capturing evolving patterns for ontology-based web mining. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, Sep. 20-24, IEEE Xplore Press, pp: Australia. Pp: 256-263. DOI: 10.1109/WI.2004.10040

Pretschner, A. and S. Gauch, 1999. Ontology based personalized search. Proceedings of 11th IEEE International Conference on Tools with Artificial Intelligence, Nov. 9-11, IEEE Xplore Press, Chicago, IL , USA., pp: 391-398. DOI: 10.1109/TAI.1999.809829

Salton, G. and C. Buckley, 1988. On the use of spreading activation methods in automatic information. Proceedings of 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (RDIR' 88), ACM New York, NY, USA., pp: 147-160. DOI: 10.1145/62437.62447

Sampath, S., V. Mihaylov, A. Souter and L. Pollock, 2004. A scalable approach to user-session based testing of web applications through concept analysis. Proceedings of the IEEE International Conference on Automated Software Engineering, Sep. 20-24, IEEE Xplore Press, Denmark, pp: 132-141. 10.1109/ASE.2004.1342731

Shen, X., B. Tan and C. Zhai, 2005. Implicit user modeling for personalized search. Proceedings of 14th ACM Conference on Information and Knowledge Management, (IKM'05), ACM New York, NY, USA., pp: 824-831. DOI: 10.1145/1099554.1099747

Sieg, A., B. Mobasher and R. Burke, 2007a. Ontological user profiles for personalized web search. DePaul University.

Sieg, A., B. Mobasher and R. Burke, 2007b. Web search personalization with ontological user profiles. Proceedings of the 16th ACM Conference on Information and Knowledge Management (IKM' 07), ACM New York, NY, USA., pp: 525-534. DOI: 10.1145/1321440.1321515

Silverstein, C., H. Marais, M. Henzinger and M. Moricz, 1999. Analysis of a very large web search engine query log. ACM Special Interest Group Inf. Retrieval, Forum, 33: 6-12. DOI: 10.1145/331403.331405

Speretta, M. and S. Gauch, 2005. Personalized search based on user search histories. Proceedings of the International Conference on Web Intelligent, Sep. 19-22, IEEE Xplore Press, Lawrence, KS, USA., pp: 622-628. DOI: 10.1109/WI.2005.114

Stamou, S. and A. Ntoulas, 2009. Search personalization through query and page topical analysis. User Model. User-Adapted Interact., 19: 5-33. DOI: 10.1007/s11257-008-9056-y

Vallet, D., I. Cantador, M. Fernandez and P. Castells, 2006. A multi-purpose ontology-based approach for personalized content filtering and retrieval. Proceedings of the 1st International Workshop on Semantic Media Adaptation and Personalization, Dec. 4-5, IEEE Xplore Press, Athens, Greece, pp: 19-24. DOi: 10.1109/SMAP.2006.4

Zhou, X., S.T. Wu, Y. Li, Y. Xu and Y.K. Raymond *et al.*, 2006a. Utilizing search intent in topic ontology-based user profile for web mining. Proceedings of the International Conferences on IEEE/WIC/ACM Web Intelligence, Dec. 18-22, IEEE Xplore Press, Hong Kong, pp: 58-564. DOI: 10.1109/WI.2006.186

Zhou, X., Y. Li, Y. Xu and R. Lau, 2006b. Relevence assessment of topic ontology. Proceedings of the Conference on Advances in Intelligent IT: Active Media Technology 2006, (AIAMT' 06), IOS Press Amsterdam, The Netherlands, The Netherlands, pp: 44-51.