ISSN No. 0976-5697

Volume 8, No. 5, May – June 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Use of hybrid of Fuzzy set and ACO for effective personalized web search

Dr. Suruchi Chawla
Assistant Professor
Shaheed Rajguru College of Applied Science for Women
Delhi, INDIA

Abstract: Personalized web search techniques have been applied with success for effective information retrieval. The users search queries are vague and imprecise due to limited vocabulary of users and therefore the precision of search results is low. Fuzzy set has been used in research to infer the user's information need from imprecise and vague queries. Ant Colony Optimization techniques(ACO) have been applied to optimize the search results in order to increase the relevant documents and improve the precision of search results. In this paper hybrid of Fuzzy set and ant colony optimization technique have been used together and an algorithm is proposed for recommendation of relevant web pages according to user's information need. Experiment was conducted on the data set captured in three domains Academics, Entertainment, Sports and results confirm the improvement of precision of search results.

Keywords: Fuzzy set; Information Retrieval; Ant Colony Optimization; Personalized Web Search; Search Engines; Pheromone.

I. INTRODUCTION

Personalization of web search aims at customizing the web search according to the information need of the user. Research has been done for effective personalization of web search based on optimization techniques such as ACO, Fuzzy set and their hybrid. [1][2][3][4][5] In this paper hybrid of ACO and Fuzzy sets are used together for effective information retrieval. The benefit of using both Fuzzy set and ACO together because of the following reasons Fuzzy set is used to infer the user's information need from vague and imprecise queries. The ACO technique is quite amenable to capture the user searching behaviour on the web. A user is like an artificial ant searching on the web for a specific information need. The user response to clicked URL is captured using pheromone. The pheromone of the clicked URLs increases as more and more users visit that page. Thus the Ant Colony optimization technique uses pheromone value to optimize the set of documents relevant to user's information need. Thus an approach is proposed using hybrid of Fuzzy set and ACO to recommend web pages relevant to the information need of the user.

In this paper an algorithm is proposed for personalized web search using hybrid of Fuzzy set and ACO. The entire processing of the algorithm is divided into two phase Offline and Online Processing. During Offline processing, query session keyword vector is generated from query session using Information scent and content of clicked URLs. Information Scent measure the relevance of clicked URLs based on its web usage data. The keyword vectors are clustered where each cluster groups query sessions with similar information need in a specific domain. The term-document matrix local to each cluster is built based on high pheromone clicked URLs selected in a given cluster. The term document matrix W global to entire data set is created using tf.idf vector of the clicked documents present in the entire data set. The terms of matrix W is the set of distinct terms found in the vocabulary of the clicked documents. Fuzzy thesaurus R is built from W and W^T in order to create term-term correlation matrix Fuzzy thesaurus R.

During online processing, the user input query is represented by Fuzzy Set A which is expanded with related terms based on Fuzzy thesaurus R. The addition of related terms reduces the impreciseness and vagueness of input query which arises due to limited vocabulary of user query. The Fuzzy expanded input query is used to select the most similar cluster and the term document matrix of the high pheromone clicked URLs is used to identify the fuzzy set of ranked documents on set D where documents are ranked according to their pheromone value. The user response to recommended documents is tracked to collect the user profile and the pheromone is updated using ACO. The user profile is vectored using content of its clicked documents and is expanded with related words based on Fuzzy thesaurus. The user profile is used to select the cluster for the recommendation of Fuzzy ranked set of documents. This process of user profile expansion with related terms and recommendations of Fuzzy ranked set of documents continues till the user information need is satisfied. The stepwise execution of offline and online processing is given below in Fig 1.

Experiment was conducted on the data set captured in three domains Academics, Entertainment and Sports. The results were compared with Fuzzy IR [6]. The proposed approach confirms the improvement of precision of search results.

II. RELATED WORK

In [7][8] modeling of user preferences had been described with fuzzy profiles. In [9] a fuzzy ontology is combined with the TF-IDF measure for document ranking.

A novel classification approach was developed in [10] [11] in order to conceptualize documents into concepts using FFCM (Fuzzy Formal Conceptualization Model). In [12] Fuzzy logics were used to summarize text for extracting the most relevant sentences.

In [13] fuzzy logic is considered as a necessity to add deductive capability to a search engine. In [14] a new technique was presented which integrates document index with perception index for the refinement of fuzzy queries on the Internet. In [15] the extended profiles containing additional information related to the user were used to personalize and customize the retrieval process as well as the web site. Fuzzy clustering of these extended profiles was carried out and fuzzy rules were constructed. Fuzzy inference was used to modify queries and extract knowledge from profiles with marketing purposes within a web framework. In [16] the expressiveness of the queries can be increased with the use of fuzzy aggregation methods for intelligent search. In [17] fuzzy logic for rule-based personalization was introduced and implemented for personalization of newsletters.

In [18] fuzzy set model had been used to define fuzzy queries. In [19] [20] fuzzy relationship between query terms and documents was introduced. In [21] fuzzy IR system uses fuzzy logic to retrieve documents similar to the query document. The system was tested on Arabic documents. In [22] the fuzzy logic model was actually used in information retrieval and came up with a ranking model. The ranking model had rules for fuzzification based on three fuzzy variables; tf, idf and overlap.

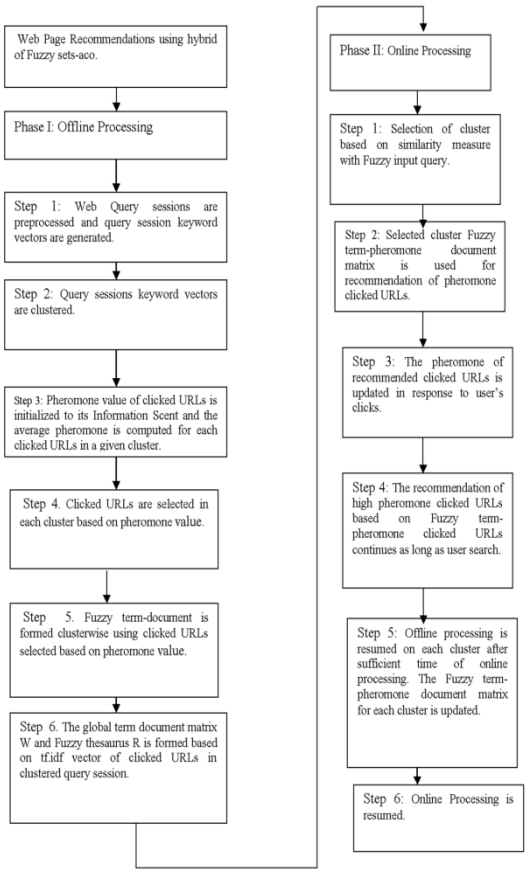

Fig 1. Shows the steps of offline and online processing

In [23] Ant Colony Optimization (ACO) was applied to build query association graphs from the query logs for the purpose of query recommendation. ACO was applied to query logs as an adaptive learning process.

In [24] recommender system based on ant colony optimization (ACO) was proposed which recommends to the users the shortest path to the target pages using ACO. The target pages are the pages which the users search for. In [25] Ant colony optimization was applied to predict the next request of the user and make recommendations on the basis of these predictions. The recommender model was updated according to the changing user needs using the advantages of both the Click Stream Tree Model [26] and the ACO method. In [3] the user profile based on interest scores was built using the pheromone deposited by the ants. The user's behavior was reflected from user profile as the pheromone being accumulated or evaporated. In the mean time the content keywords of user's profile was classified in a reference concept hierarchy. A set of experiments was conducted to determine the effectiveness of personalized search using the proposed approach and obtained the satisfactorily results.

In [27] model based on the learner's cognitive level, learning ability, cognitive ability, can adjust teaching methods, teaching strategies, achieve individualized to help the user personalized distance learning. The clustering based on ants was used to enable access to a variety of collaborative learning agent groups so as to fully mobilize the enthusiasm of collaborative learning team members. In [28] the ant colony optimization was applied on the log data to build an adaptive domain model automatically in order to satisfy user's information request effectively in more structured collections such as digital libraries, local Web sites, and intranets. The entire user's population searching as well as navigation was learned and this learned knowledge was incorporated in a constantly adapting domain model. This domain model assisted the user in search process and reflected the collection characteristics.

In [29] approach was proposed which combines ant based clustering and fuzzy c-means. Their model was employed in [30] where a fuzzy ant-based recommender system was presented that provides online users with a list of recommendations based on the comparison of the user's navigational behavior with other user's data. In this model ACO algorithms and fuzzy logic were combined to generate the recommendations. In [31] an ACO algorithm called Ant-Recommender was developed with the aim of recommending items within clusters of user profiles. In this method the ants sense pheromone found on clusters rather than on individual paths to determine the best cluster to provide a recommendation.

In [32] the ant colony metaphor was also used for selecting optimal solutions in his hybrid recommendation method. In [33] a recommender system was presented based on the collaborative behavior of ants.

In [34] it was found that Information Scent of the clicked page guides user in finding the relevant information on the web in the same way as the pheromone guides the ant in identifying the shortest path to food source. The performance of the Personalized Web Search was improved when converted into optimization problem and solved using Ant Colony optimization techniques by replacing the pheromone in ACO with Information Scent. In this research it was motivated to apply the hybrid of Fuzzy set and ACO for effective personalized web search since the use of ACO for optimal ranking of clicked pages retrieved based on fuzzy set will bring more and more relevant documents up in ranking for the improvement of precision of search results. Hence thus the precision of search results can be improved using Fuzzy set with ACO for web page recommendation.

III. BACKGROUND

A. *Information Scent*

Information scent is the sense of value and cost of accessing a page based on perceptual cues with respect to the information need of user. The users on the web tend to click those pages in the retrieved search results on the web which seem to satisfy the user's information need. More the page is satisfying the information need of user, more will be the information scent perceived by the user associated to it and more is the probability that the page is clicked by the user. The interactions between user need, user action and content of web can be used to infer information need from a pattern of surfing. [35][36]

- Information Scent metric: The Inferring User Need by Information Scent (IUNIS) algorithm is used to quantify the Information Scent s_{id} of the pages P_{id} clicked by the user in ith query session. [37][38]

The page access PF. IPF weight and Time are used to quantify the information scent associated with the clicked page in a query session. The information scent s_{id} is calculated for each clicked page P_{id} in a given query session i for all m query sessions identified in query session mining as follows in (1) and (2).

$$s_{id} = PF.IPF(P_{id}) \times Time(P_{id}) \forall i \in 1..m \; \forall d \in 1..n \quad (1)$$

$$PF.IPF(P_{id}) = \frac{f_{P_{id}}}{\max_{d \in 1..n} f_{P_{id}}} \times \log\left(M/m_{P_d}\right) \quad (2)$$

$PF.IPF(P_{id})$: PF correspond to the page P_{id} normalized frequency $f_{P_{id}}$ in a given query session i where n is the number of distinct clicked page in session i and IPF correspond to the ratio of total number of query sessions M in the whole data set to the number of query sessions m_{P_d} that contain the given page P_d.

$Time(P_{id})$: It is the ratio of time spent on the page P_{id} in a given session i to the total duration of query session i. [39][40][41][42][43]

- Generation of Query sessions keyword vector: Each query session keyword vector is generated from query session which is represented as follows

  query session=(input query,(clicked URLs/Page)+)

where clicked URLs are those URLs which user clicked in the search results of the input query before submitting another query ; '+' indicates only those sessions are considered which have at least one clicked Page associated with the input query.

The query session vector Q_i of the i^{th} session is defined as linear combination of content vector of each clicked page P_{id} scaled by the weight s_{id} which is the information scent associated with the clicked page P_{id} in session i as given in (3).

$$Q_i = \sum_{d=1}^{n1} s_{id} * P_{id} \quad \forall i \in 1..m \quad (3)$$

In the above formula n1 is the number of distinct clicked pages in the session i and s_{id} (information scent) is calculated for each clicked page present in a given session i as defined in (3). The content vector of clicked page P_{id} is weighted using TF.IDF. Each i^{th} query session is obtained as weighted vector Q_i using formula (3). This vector is modeling the information need associated with the i^{th} query session.

- Clustering of Query session keyword vector: The k-means algorithm is used for clustering query sessions keyword vectors since its performance is good for document clustering. [44] [45] The vector space implementation of k-means uses score or criterion function for measuring the quality of resulting clusters. The criterion function is computed on the basis of average similarity between vectors and centroid of the assigned clusters.

The criterion function I is defined as given in (4):

$$I = 1/M \quad \sum_{p=1}^{k} \sum_{v_i \in C_p} sim(v_i, c_p) \quad (4)$$

where C_p be a cluster found in a k-way clustering process ($p \in 1..k$) , c_p is the centroid of p^{th} cluster , v_i is the vector representing some query session belonging to the cluster C_p and M is the total number of query sessions in all clusters as defined in (5). [46]

$$M = \sum_{p=1}^{k} |C_p| \quad (5)$$

The centroid c_p of the cluster C_p is defined as in (6):

$$c_p = \left(\sum_{v_i \in C_p} v_i\right) \Big/ |C_p| \quad (6)$$

where $|C_p|$ denotes the n**u**mber of query sessions in cluster C_p and $sim(v_i, c_p)$ is calculated using cosine measure.

B. Fuzzy Set Theory in Information Retrieval (IR) System

The Fuzzy information retrieval methods are based on fuzzy set in order to handle uncertain information. It utilizes the tools defined in fuzzy logic and fuzzy relations to infer the best results to a user query. Unlike Boolean systems, fuzzy systems are most effective when dealing with data that may display a degree of membership. In fuzzy systems, objects described in terms of their properties which characterize the objects are assigned relational membership values to show relevancy from properties to objects or vice versa.

In order to implement the Information retrieval based on the concept of Fuzzy Sets, two finite crisp sets are defined, one is the set of m1 recognized index terms, T = $\{x_1, x_2, \ldots\ldots x_{m1}\}$ and other is a set of n relevant documents, D = $\{d_1, d_2, \ldots, d_n\}$

A fuzzy document—term relation W is a fuzzy relation from D to T.W represents the relevance of index terms to individual documents $W: DxT \rightarrow [0,1]$ such that membership value $W(d_j, x_i)$ specifies for each $x_i \in T$ and $d_j \in$D the grade of relevance of index term x_i to document d_j . $W(d_j, x_i)$ can be obtained in a probabilistic manner by counting frequencies in the so called TF.IDF approach.

A fuzzy thesaurus or fuzzy term—term relation R is a fuzzy relation from T to T. R is a reflexive relation on T. For each pair of index term $< x_i, x_k > \in T$, $R(x_i, x_k)$ in (7) expresses the association of x_i with x_k that is the degree to which the meaning of the index term x_k is compatible with meaning of the given index term x_i . The role of this relation is to deal with the problem of synonyms among index terms. The relationship helps to identify relevant documents for a given query that otherwise would not be identified. This happens whenever a document is characterized by an index term that is synonymous with an index term contained in the query.

$$R: T \times T \rightarrow [0,1]$$

$$R(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} W(d_i, x_1) * W(d_i, x_2) \quad \forall x_1, x_2 \in T \quad (7)$$

The Fuzzy set A representing the initial input query is defined on the set of index terms T. The augmented query represented by Fuzzy set B in T in (8) is generated from initial input A using max min composition operator with Fuzzy Thesaurus R.

That is, $AoR = B$ (8)

where o is the max-min composition.

The retrieved documents expressed by a fuzzy set RD defined on D in (9), is then obtained by composing the augmented inquiry expressed by fuzzy set B, with the term document matrix W. That is

$$BoW = RD \quad (9)$$

The user can now decide whether to inspect all documents capture by the support of RD or to consider only documents captured by some α- cuts of RD. [47][48]

In [6] initial input query is represented by Fuzzy set A on the term set T is augmented with related terms using Fuzzy thesaurus R based on max-min composition for better representing the information need of the user in (10).

$$B(x_j) = \max_{x_i \in T} \min(A(x_i), R(x_i, x_j)) \forall x_j \in T \quad (10)$$

The augmented query B is then used to select the cluster which is most similar to the information need of query. The term-document matrix W_j constructed based on tf.idf(term frequency inverse document frequency) associated with the selected cluster is used for max-min composition with augmented input query B in order to identify the fuzzy Document set RD on D in (11).

$$RD(d_j) = \max_{x_i \in T} \min(B(x_i), W_j(x_i, d_j)) \forall d_j \in D \quad (11)$$

The documents in Fuzzy set RD are recommended to the user. The user response to recommended documents is tracked to capture the user profile. On the request of next result page, the user profile is transformed into keyword vector FUP and is further expanded with related terms FB using Fuzzy thesaurus R in (12) for handling the vagueness due to synonyms and poli-semis of vocabulary used in profile.

$$FB(x_j) = \max_{x_i \in T} \min(FUP(x_i), R(x_i, x_j)) \forall x_j \in T \quad (12)$$

This expanded user profile fuzzy set FB on set T is used to select the cluster for the recommendations of set of ranked documents RD where ranking is generated using composition of Fuzzy set FB and Fuzzy term document matrix W_j associated to a given cluster as given in (13).

$$RD(d_j) = \max_{x_i \in T} \min(FB(x_i), W_j(x_i, d_j)) \forall d_j \in D \quad (13).$$

C. ANT COLONY OPTIMIZATION

ACO is inspired from the social behavior of ant colonies. Ants communicate with each other indirectly through chemical called pheromone released by the ants on their path to food source. [49][50][51]

The objective of ACO is to construct the path of the ant from source to destination by making local optimal choice at each point of decision using probability rule as given in (14) [52].

$$P_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha * [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}(t)]^\alpha * [\eta_{il}]^\beta}, j \in N_i^k$$

$$P_{ij}^k(t) = 0, j \notin N_i^k \quad (14)$$

Where $P_{ij}^k(t)$: is the probability of the k^{th} ant to move from node i to node j at the t^{th}

iteration/time step.

.

N_i^k is the set of nodes in the neighborhood of the k^{th} ant in the i^{th} node.

$P_{ij}^k(t) = 0, j \notin N_i^k$ means the ants are not allowed to move to any node not in their neighborhood. The neighborhood definition is problem-specific.

$[\tau_{ij}(t)]^\alpha$ is the pheromone amount on the arc connecting node i and node j, weighted by α.

$[\eta_{ij}]^\beta$ is the heuristic value of arc connecting node i to j weighted by β.

α and β are weight parameters that control the relative importance of the pheromone versus heuristic information.

In the beginning of the optimization, the pheromone value of all arcs on the constructed path is initialized to the constant value τ_0. The pheremone trails are updated in two ways as defined in (15) and (16) given below. This pheromone updation process increases the probability of the quality paths to be followed by the more ant in future while finding solution to the problem. The quality paths include solution component that were either used by the many ants in the past or was at least followed by the ant which produces high quality solution. [53]

$$\tau_{ij}(t+1) \leftarrow (1-\rho) * \tau_{ij}(t) + \sum_{k=1}^{m} \Delta\tau_{ij}^k(t) \quad \forall i, j \in A, 0 \leq \rho < 1 \quad (15)$$

$$\Delta\tau_{ij}^k(t) = Q/C^k(t), (i,j) \in T^k(t) \quad (16)$$

$$\Delta\tau_{ij}^k(t) = 0, \text{ otherwise}$$

Where Q is an application-specific constant, m is the number of ants, A represents all arcs of the problem of construction graph, $C^k(t)$ is the overall cost function of tour $T^k(t)$constructed by the kth ant at the t^{th} iteration, and $T^k(t)$ is the set of all arcs visited by ant k at the iteration t. Other variations of ACO, however, restrict pheromone depositing to the arcs of the best tour T^{best} only.

IV. PERSONALIZATION OF WEB SEARCH USING FUZZY IR AND ACO

In this paper an approach is proposed for effective personalized web search using hybrid of Fuzzy set and ACO. The advantage of using the hybrid of Fuzzy set and ACO is to infer the information need from user's imprecise and vague queries and optimize the search results with more and more relevant documents. The entire processing of the proposed approach is divided into two phases: Phase I(Offline) and Phase II(Online). During Offline processing, the user query sessions and its associated Clicked URLs are collected on the web. The tf.idf vector of clicked URLs is fetched using crawler and forms the document-term matrix W. The Fuzzy thesaurus which is term-term correlation matrix is generated using W and W^T. The query session keyword vector is generated using Information Scent and tf.idf content of clicked URLs. The resulting keyword vectors are clustered in order to group similar information need query sessions. The pheromone value of clicked URLs associated with each cluster is initialized using Information Scent. Information Scent measures the relevance of clicked URLs based on its usage statistics. Each cluster j is associated with term-document matrix Wj of clicked URLs selected based on pheromone threshold. The user's clicks to the web pages are captured using pheromone updation and identify the web pages relevant to the users information need.

During online processing, the user input query which is vague and imprecise due to small length is augmented with related keywords based on Fuzzy thesaurus. This augmented initial input query based on Fuzzy set is used to select the cluster closer to the information need of the query.

The Fuzzy term-document matrix of clicked URLs associated with the selected cluster is used to determine another fuzzy set RD which is ranked set of documents. The Fuzzy set

RD on D using threshold α is presented to the user in decreasing order of relevance based on pheromone value. The user clicks to recommended documents is tracked to capture the user profile. The user's response to recommended web pages are used to optimize the recommended set of web pages with the updation of pheromone of clicked URLs. The optimization of clicked web pages using ACO identifies the colony of relevant web pages based on user search behavior. The user profile is then converted to keyword vector using pheromone and content of clicked URLs and augmented with related words based on Fuzzy thesaurus. This process of user profile expansion based on fuzzy set and recommendation of optimal ranked set of documents using ACO continues till the user information need is satisfied. The stepwise execution of offline & online processing is given below.

Phase I :Offline Preprocessing

1. Data Set Collected on the Web is preprocessed to get the Query Sessions where each query session contains the user input query and associated clicked URLs.
2. Generate the tf.idf vector of each distinct clicked URLs present in the data set.
3. The set of distinct terms present in data set is represented by set T={x1,x2,x3,……,x_{m1}} and set of distinct documents is represents by set D={d1,d2,d3`…….,d_n}
4. Generate the global term document relation W a fuzzy relation from D to T where |D|=n and |T|=m1 where n is the distinct document(clicked URLs) collected from the entire data set and m1 is the set of distinct terms present in data set for the construction of fuzzy thesauri.
5. A fuzzy thesaurus or fuzzy term—term relation R is a fuzzy relation from T to T identifying the synonym relations using (7).
6. For each clicked URLs in the query session, the Information Scent Metric is calculated using. (1) which is the measure of the relevancy of the clicked URLs with respect to the information need of the user associated with the query session.
7. For each cluster i the initial value of pheromone is calculated as follows
   a. Each clicked URLs of the user query session is associated with the initial pheromone value $\tau_{pheremoneclickedURLs}\ (0)$= Information Scent of the Clicked URLs. $\Delta\tau_{URLs}$ =0. where $\Delta\tau_{URLs}$ is the quantity trail substance (pheromone in real ants) laid on Clicked URLs by the k^{th} user/ant between time t and t+n;
   b. For each distinct clicked URL in the given cluster identify $\tau_{avgp\ heremoneURLs}\ (0)$ which is calculated over all the query sessions present in the given cluster.
8. Query sessions keyword vector is generated from query sessions using pheromone and content of Clicked URLs where content of clicked URLs is TF.IDF weighted vector as given below.
$$Q_i = \sum_{d=1}^{n1} \tau_{avgp\ heremoneURLid}\ (0) * P_{id} \qquad \forall\, i \in 1..m$$
Where n1 is the number of distinct clicked URLs in a given session i
9. k-means algorithm is used for clustering query sessions keyword vector.
10. Each cluster i is associated with the mean keyword vector clusteri_mean.
11. For each cluster j maintain the list of clicked URLs in list CL_j whose avgpheremone >= threshold(ε).
12. Generate Fuzzy term- document matrix W_j using tf.idf vector of Clicked URLs CL_j having $\tau_{avgp\ heremoneURLid}\ (0) >= \varepsilon$ for clicked URLs present in a given cluster j.
$$W_j : T \times CL_j \rightarrow [0,1]$$

Phase II :Online Processing.

1. Consider a query Q. Ignore all the stop words from the search query. Final query is represented by Fuzzy set A on T based on tf.idf.
2. Collect the fuzzy thesaurus R restricted to the support of A and non zero columns for the expansion of input query A.The support of A is the set of terms belonging to set A and used in expressing query Q.
3. Compute expanded input query B ← A o R using (10)
4. Find the cluster j which is most similar to Fuzzy expanded input query B.

MatchScorej(B, $cluster_j$)= sim(B , $cluster_j$_mean)

5. Use the term-document tf.idf matrix W_j of the clicked URLs whose avgpheremone >= threshold(ε).associated with the selected cluster j and select the relevant part of the W_j matrix restricted to support of B and non zero columns for computing the Fuzzy document set RD on D.
6. Compute Fuzzy Document set RD ← B o W_j using (11)
7. Inspect only those document_ids in RD captured by some α-cut of RD in which only those documents are filtered for retrieval whose degree of relevance is greater than or equal to α.
8. For each selected document d in α-cut of RD is retrieved and stored in L.
9. The list L is presented to the user in decreasing order of their pheremone value
10. The user response to the recommended URLs is tracked and stores it in current user profile.
11. For each i^{th} clicked URL of the current user session if present in the selected cluster j, then pheromone value of the clicked URLs w.r.t current user profile will be added to the average pheromone value of the corresponding clicked URLs in each selected cluster j in order to update its average pheromone value.
$$\tau_{avgp\ heremoneURLs\ \_ij}(t) = (1-\rho) \times \tau_{avgp\ heremone\ URLs\_ij}(t-1) + \tau_i(t)$$
where i is the URLs clicked by the current user at time t and present in the selected cluster j. This is for all i where i is the clicked URLs of the current user and i ∈ selected cluster j and $\tau_i(t)$ is the information scent of clicked URL i in the current user session. $\tau_{avgp\ heremoneURLs\ \_ij}(t)$ is the average pheromone of the i^{th} clicked URL in j^{th} cluster.
12. If the user request for the next result page
    a. The users clicks to the search results on the current page have been tracked and user selected input query are stored in user profile.
    b. Model the partial information need of the current user profile using the pheromone and content of the URLs clicked so far in his partial user profile and obtain the user session keyword vector FUP.
    c. Compute expanded Fuzzy User Profile FB ← FUP o R where the part of fuzzy thesaurus R, restricted to the support of FUP, and non zero columns, is relevant for the expansion of user profile FUP using (12).
    d. Select the j th cluster which is most similar to the information need associated with the FB.
    e. Use the term-document tf.idf matrix Wj of the clicked URLs associated with the selected cluster j and select

the relevant part of the W_j matrix restricted to support of FB and non zero columns for computing the Fuzzy document set RD on D.

f. Compute Fuzzy set RD ← FB o W_j using (13).

g. Inspect only those document_ids in RD captured by some α-cut of RD in which only those documents are filtered for retrieval whose degree of relevance is greater than or equal to α.

h. For each selected document d in α-cut of RD is retrieved and stored in L.

i. Goto step 9.

Else

Current search session is terminated.

End

V. EXPERIMENTAL STUDY

Experiment was conducted on the data set of user query sessions collected on the web. The architecture is developed using JADE, JSP and Oracle database. The dataset of user query sessions is captured on the web using the GUI of the architecture, the user enter the input query through a GUI to retrieve the search results. The search results are displayed along with the check boxes and the user's clicks to search results are captured using checkboxes as given below in Fig 2. The captured clicked URLs of user query sessions are stored in the database for query session mining.

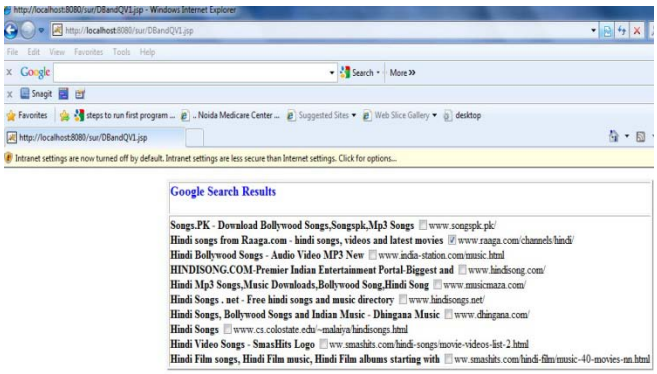

Fig 2: Shows the GUI of the proposed architecture displaying the Google search results with check boxes.

The experiment was performed on i3 processor with 2 GB RAM on Windows 8. The clustering agent developed in JADE was executed to generate the query sessions keyword vector using the tf.idf vector and pheromone of the clicked URLs. The tf.idf vector of clicked URL was fetched using the Web Sphinx Crawler and loaded into database using oraloader. During the execution of clustering agent, clusters of query session keyword vectors were generated and Fuzzy term-high pheromone document matrix was formed local to each cluster. The Fuzzy thesaurus is formed using tf.idf document matrix based on distinct documents found in whole dataset.

The parameters used for experimental evaluation were Information Scent lies in [0,1] and threshold value of pheromone ε is set to 0.5.The following parameters set as follows the pheromone evaporation factor ρ =0.5 as the experiments was conducted with different value of parameters based on the preprocessed collected data set, where the value of $\rho \in [0,1)$ (Dorigo, 1992. It is further found by setting the value of ρ below 0.5, the evaporation rate of the pheromone was not effective enough to the capture the changing user's need in the clustered query sessions, therefore the pheromone evaporation factor was set at ρ =0.5.

During Online processing, the input query was issued to GUI based interface designed each for both Personalized Web Search with fuzzy IR(with/without ACO) based on the same clustered query sessions dataset. In Personalized Web Search with fuzzy IR(with ACO) the input query was used to find the cluster most similar to the information need of the current user. The resultant set of the clicked URLs associated with the selected cluster were recommended and displayed in decreasing order of their pheromone value.

The user's clicks to the personalized search results were tracked to capture the user's profile and dynamically update the pheromone associated with the stored clicked URLs. This dynamic updation of the pheromone of the clicked URLs using ACO optimizes relevant set of clicked URLs for each identified information need associated with the clusters.

The performance of the Personalized Web Search using fuzzy IR (with ACO) was evaluated from the average precision of search results and compared with Personalized Search Results using Fuzzy IR(without ACO). In Personalized Web Search(PWS) with fuzzy IR(without ACO) the recommended search results were displayed in decreasing order of their Fuzzy membership function.

In order to evaluate the performance, the test queries were chosen in three domain Academics, Entertainment, Sports. The purpose of selecting the queries in a given domain is to cover wide range of user queries on the web. Sample of queries is given below in Table1.

Table I. Shows the test queries used for experimental evaluation of Fuzzy IR(with/without ACO).

Entertainment	■ free pics ■ online audio stores ■ free download mp3 ■ skies of arcadia pictures ■ vcd files ■ mpeg movies
Sports	■ grand american road racing series ■ arena football ■ south dakota wrestlings ■ major league baseball tryout ■ kit car arena football
Academics	■ cgi perl tutorial ■ sql tutorial ■ tutorial oracle ■ windows 2000 tutorial ■ macros ■ templates ■ weblogs

The number of test queries in the three domains was 25 in Academics, 25 in Entertainment and 25 queries in Sports. During online searching, these test queries were issued in each of the selected domain to the GUI based interface to retrieve the personalized search results using fuzzy IR(with/without ACO). The average of precision of test queries is computed in each of the selected domain where precision is fraction of retrieved documents that are relevant.

The experimental results showing the average precision of 25 test queries computed in the domains of academics,

entertainment and sports using PWS with Fuzzy IR(with/ without ACO) and their percent improvement over ClassicIR is given in Fig 3 and Table II.

Table II. Shows the average precision of test queries using Fuzzy IR(with/without ACO) and percentage improvement over Classic IR(Google Search Engine).

25 TEST QUERIES	**Average Precision using Information Retrieval Techniques**			**% improvement over Classic IR**	
Domains	**Classic IR**	**PWS with Fuzzy IR**	**PWS with ACO Fuzzy IR**	**PWS with Fuzzy IR**	**PWS with ACO Fuzzy IR**
Academics	0.45	0.7	0.76	55	68
Entertainment	0.46	0.75	0.79	63	71
Sports	0.44	0.72	0.78	63	77

The obtained results were analyzed using the statistical paired t-test for average precision of PWS with ACO Fuzzy IR versus both (ClassicIR/Fuzzy IR) with 74 degrees of freedom (d.f.) for the combined sample as well as in all three categories (Academics, Entertainment and Sports) with 24 d.f each. The observed value of t for average precision of Fuzzy IR(with ACO) versus (ClassicIR) for a combined sample is 89.54, 96.43 in Academics, 35.79 in Entertainment and 117 in Sports. The value of t for average precision of Fuzzy IR(with ACO) versus Fuzzy IR(without ACO) for a combined sample is 7.2, 5.9 in Academics, 11.09 in Entertainment and 11.47 in Sports.

It was observed that the computed t value for paired difference of average precision lie outside the 95% confidence interval in each case. The value of t shows that Web Information retrieval based on Fuzzy ACO shows the high t value both over classic IR and Fuzzy IR. Hence Null hypothesis was rejected and alternate hypothesis was accepted in each case and it was concluded that average precision is improved significantly using web page recommendations with fuzzy IR(with ACO). It is shown that personalized search results using fuzzy IR(with ACO) recommend Fuzzy clicked URLs with high pheremone value since the use of high pheremone clicked urls increases the likelihood of generating more relevant documents for recommendations. The pheremone is updated for each click to recommended URLs and will identifies the colony of webpages relevant to the information need of the user and personalizes the web search more effectively. Thus the PWS with Fuzzy IR(ACO) shows the significant improvement in precision in comparison to PWS with Fuzzy IR (without ACO).

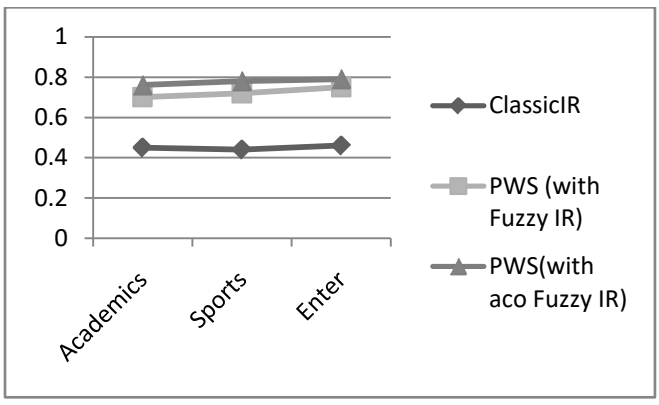

Fig 3. Compares the average precision of search results using Fuzzy IR(with ACO), Fuzzy IR with Classic IR.

VI. CONCLUSION

In this research an approach is proposed for personalized web search using Fuzzy set and ACO. The user query expansion based on fuzzy set infers the information need of user query which is otherwise difficult to infer. The association of pheromone with web pages and its updation using ACO identifies the relevant web pages for recommendations. Thus the use of hybrid of Fuzzy set and ACO recommend relevant web pages according to the information need of the user. Experiment was conducted on the data set of web query sessions collected in the three domains mainly entertainment, academics and sports. The results were compared and confirm the significant improvement in the precision of search results. The proposed algorithm proves to be effective for web search personalization according to the information need of the user.

REFERENCES

[1] K. Selvakumar Sendhilkumar, and G.S. Mahalakshmi,"Applications of fuzzy logic for user classification in personalized web". International Journal on Cybernetics & Informatics (IJCI) Vol. 3, No. 3, June 2014.

[2] Kim ,Kyung-Joong and Cho ,Sung-Bae, "Personalized mining of web documents using link structures and fuzzy concept networks". Elsevier- Fuzzy Sets and Systems ,September 2005.

[3] P.,Phinitkar and P. Sophatsathit, "Personalization of search profile using ant foraging approach". In International Conference on Computational Science and Its Applications (pp. 209-224). Springer Berlin Heidelberg,2010.

[4] M. Göksedef, G. N. Demir, and S. Gündüz-Ögüdücü, "A web recommender system based on ant colony optimization". In IADIS: Proceedings IADIS European Conference on Data Mining,(pp. 535-540, 2007 Lisbon, Portugal: IADIS Press.

[5] S. Nadi, , M. Saraee, , A. Bagheri, , & M. Davarpanh Jazi, "FARS: Fuzzy ant based recommender system for web users". International Journal of Computer Science Issues, 8(1), 203-209,2011.

[6] S Chawla, "Effective Personalization of web search based on Fuzzy Information Retrieval". International Journal of Computer Science and Information Technologies, 6 (3) , pp. 2831-2837,2015b

[7] C. Mencar, M. Torsello, D. Dell'Agnello, G. Castellano, and C. Castiello, "Modeling user preferences through adaptive fuzzy profiles". In 9th International Conference on Intelligent Systems Design and Applications, ISDA 2009, pp 1031 –1036, Nov. 30-Dec. 2 2009.

[8] G. Castellano, D. Dell'Agnello, A. M. Fanelli, C. Mencar, and M. A. Torsello, "A competitive learning strategy for adapting fuzzy user profiles". In10th International Conference on Intelligent Systems Design and Applications, ISDA 2010, pages 959–964, Nov. 29-Dec. 1 2010.

[9] M. Holi, E. Hyvnen, and P. Lindgren, "Integrating tf-idf weighting with fuzzy view-based search." In Proceedings of the ECAI Workshop on Text-Based Information Retrieval (TIR-06), 2006.

[10] Sankar K.Pal, Saroj K. Meher, and Soumitra Dutta, "Class-dependent rough-fuzzy granular space, dispersion index and classification." Pattern Recognition, 45, no. 7, pp 2690-2707,2012.

[11] Sheng-Tun, Li, and Fu-Ching Tsai, "A fuzzy conceptualization model for text mining with application in opinion polarity classification." Knowledge-Based Systems, 39, pp. 23-33, 2013.

[12] F. Kyoomarsi, H. Khosravi, E. Eslami, and M. Davoudi, "Extraction-based text summarization using fuzzy analysis." Iranian Journal of Fuzzy Systems, 7, no. 3, pp 15-32, 2010.

[13] LA Zadeh. "The problem of deduction in an environment of imprecision, uncertainty, and partial truth." In: M Nikravesh, B Azvine (eds), FLINT 2001, New Directions in Enhancing the

Power of the Internet, UC Berkeley Electronics Research Laboratory, Memorandum No. UCB/ERL M01/28,2001.

[14] D Choi , "Integration of document index with perception index and its application to fuzzy query on the Internet". In Proceedings of the BISC International. Workshop on Fuzzy Logic and the Internet, pp. 68-72,2001.

[15] MJM Batista, "User profiles and fuzzy logic in web retrieval". In: Nikravesh M, Azvine B (eds), FLINT 2001, New Directions in Enhancing the Power of the Internet, UC Berkeley Electronics Research Laboratory, Memorandum No. UCB/ERL M01/28,2001.

[16] R Yager, "Aggregation methods for intelligent search and information fusion". In: Nikravesh M, Azvine B (eds), FLINT 2001, New Directions in Enhancing the Power of the Internet, UC Berkeley Electronics Research Laboratory, Memorandum No. UCB/ERL M01/28, 2001.

[17] G Presser, "Fuzzy personalization". In: M Nikravesh, B Azvine (eds), FLINT 2001, New Directions in Enhancing the Power of the Internet, UC Berkeley Electronics Research Laboratory, Memorandum No. UCB/ERL M01/28 ,2001.

[18] G Bordogna.and G.Pasi , "Handling vagueness in information retrieval systems." In: Proceedings of the Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, Nov. 20-23, 1995, pp.110-114.

[19] S. Miyamoto, "Fuzzy sets in Information Retrieval and Cluster Analysis". Springer Science & Business Media, 2012.

[20] Y. Ogawa, T. Morita, and K.Kobayashi, "A fuzzy document retrieval system using the keyword connection matrix and a learning method". Fuzzy Sets and Systems, 39(2), pp.163-179,1991.

[21] Salha Mohammed Alzahrani, and Naomie Salim, "On the Use of Fuzzy Information Retrieval for Gauging similarity of Arabic Documents", In Second International Conference on Applications of Digital Information and Web Technologies(ICADIWT'09),2009,pp. 539-544, IEEE.

[22] N. O. Rubens, "The application of fuzzy logic to the construction of the ranking function of information retrieval systems.", Computer Modelling and New Technologies, 10(1), pp.20–27, 2006.

[23] M.-D., Albakour, U.Kruschwitz, , N. Nanas, D.Song, M. Fasli and A. De Roeck, 'Exploring Ant colony optimsation for adaptive interactive search' in ICTIR'11: Proceedings of the International Conference on the theory of Information Retrieval, September, Bertinoro, Italy,pp 213-224, 2011.

[24] W. M. Teles, Li, Weigang and C. G. Ralha, "AntWeb—The Adaptive Web Server Based on the Ants' Behavior" in IEEE/WIC: Proceedings of the 2003 International Conference on Web intelligence,Washington, DC, USA, pp 558-561,2003s.

[25] M. Göksedef, G. N. Demir, and S. Gündüz-Ögüdücü, A web recommender system based on ant colony optimization. In IADIS: Proceedings IADIS European Conference on Data Mining ,Lisbon, Portugal: IADIS Press, pp. 535-540,2007.

[26] S. Gündüz. , and M. Özsu, "A web page prediction model based on click-stream tree representation of user behavior" in KDD: Proceedings of Ninth ACM International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, pp. 535–540,2003.

[27] R. Zeng and Y. Y. Wang, "Research of personalized Web-based intelligent collaborative learning, Journal of Software",Vol.7 No.4, pp. 904-912,2012.

[28] U. Kruschwitz, M-D. Albakour, J. Niu, J. Leveling, N. Nanas, Y. Kim, D. Song, M. Fasli, and A. De Roeck, "Moving towards Adaptive Search in Digital Libraries". Springer Berlin Heidelberg, pp 41-60,2011.

[29] P.M. KANADE AND O. L. HALL, "Fuzzy Ants as a Clustering Concept". In Proc. of the 22nd Int. Conf. of the North American Fuzzy Information Processing Soc.,pp. 227–232, IEEEs.

[30] S. NADI, , M. H. SARAEE, M. D. JAZI and A. BAGHERI, FARS: Fuzzy Ant based Recommender System for Web Users", International Journal of Computer Science Issues, 8(1), pp. 203-209,2011.

[31] R. SHARMA, M. SINGH, R. MAKKAR, H. KAUR, AND P. BEDI, "Ant Recommender: Recommender system inspired by ant colony", In Proceedings of International Conference on Advances in Computer Vision and Information Technology, pp. 361-369,2007.

[32] J. SOBECKI. "Colony Metaphor Applied in User Interface Recommendation". New Generation Computing. 26(3), pp.277-293,2008.

[33] P. BEDI and R. SHARMA. "Trust based recommender system using ant colony for trust computation". Expert Systems with Applications, 39(1), pp. 1183-1190,2012, Tarrytown, NY, USA.

[34] S. Chawla. "Personalized web search using ACO with information scent". International Journal of Knowledge and Web Intelligence, 4(2), pp. 238-259, 2013.

[35] P Pirolli. "Computational models of information scent-following in a very large browsable text collection" , Conference on Human Factors in Computing Systems, pp. 3-10, 1997.

[36] P. Pirolli." The use of proximal information scent to forage for distal content on the world wide web", Working with Technology, Mind: Brunswikian. Resources for Cognitive Science and Engineering, Oxford University Press, 2004.

[37] E H Chi, P. Pirolli, K. Chen and J. Pitkow, " Using Information Scent to model User Information Needs and Actions on the Web", International Conference on Human Factors in Computing Systems, New York, NY, USA, pp. 490-497,2001.

[38] J. Heer and E.H , Chi ,"Separating the Swarm: Categorization method for user sessions on the web", International Conference on Human Factor in Computing Systems, pp. 243-250,2002.

[39] S. Chawla, and P. Bedi, "Personalized Web Search using Information Scent", International Joint Conferences on Computer, Information and Systems Sciences, and Engineering, Technically Co-Sponsored by: Institute of Electrical & Electronics Engineers (IEEE), University of Bridgeport, published in LNCS (Springer), pp. 483-488,2007.

[40] S. Chawla, and P. Bedi ," Improving information retrieval precision by finding related queries with similar information need using information scent". In First International Conference on Emerging Trends in Engineering and Technology, ICETET'08, pp. 486-491,2008, IEEE.

[41] S. Chawla, " Personalised Web Search using Trust based Hubs and Authorities. International Journal of Engineering Research and Applications, 7, pp. 157-170,2014a.

[42] S. Chawla, "Novel Approach to Query Expansion using Genetic Algoirthm on Clustered Query Sessions for Effective Personalized Web Search" . International Journal of Advanced Research in Computer Science and Software Engineering, 4(11), pp 73-81,2014b.

[43] S. Chawla, Domainwise Web Page Optimization Based On Clustered Query Sessions Using Hybrid Of Trust And ACO For Effective Information Retrieval, International Journal of Scientific and Technology Research, 4(11), pp. 196-204,2015a.

[44] J. R Wen., J. Y Nie,, H. J.Zhang," Query clustering using user logs. ACM Transactions on Information Systems", 20(1), pp. 59-81,2002.

[45] Y.Zhao and G. Karypis, ,"Comparison of agglomerative and partitional document clustering algorithms" (No. TR-02-014). MINNESOTA UNIV MINNEAPOLIS DEPT OF COMPUTER SCIENCE,2002.

[46] Y , Zhao and G ,Karypis. Criterion functions for document clustering: Experiments and analysis,2001.

[47] G Klir, B. Yuan, Fuzzy sets and fuzzy logic , 4, New Jersey: Prentice hall, 1995.

[48] B. KARN, INFORMATION RETRIEVAL SYSTEM USING FUZZY SET THEORY-THE BASIC CONCEPT.

[49] M. Dorigo, V. Maniezzo,. and A. Colorni. "Positive feedback as a search strategy",Technical Report 91-016, 1991.

[50] M. Dorigo . "Optimization, learning and natural algorithms". Ph. D. Thesis, Politecnico di Milano, Italy, 1992.

[51] A. Colorni, M. Dorigo, V. Maniezzo, and M. Trubian, "Ant system for job-shop scheduling". Belgian Journal of Operations Research, Statistics and Computer Science, Vol 34 , No 1, pp.39-53, 1994.

[52] K.O. Jones, and A. Bouffet. "Comparison of ant colony optimisation and differential evolution". In Proceedings of the 2007 international conference on Computer systems and technologies ,p. 25, ACM, 2007.

[53] M., Dorigo, and K. Socha.” An introduction to ant colony optimization”. Handbook of approximation algorithms and metaheuristics, pp.26-1, 2006.