ISSN:0975-9646

Effective Personalization of web search based on Fuzzy Information Retrieval

Dr. Suruchi Chawla

Assistant Professor, Department of Computer Science ,
Shaheed Rajguru College of Applied science for Women
University of Delhi ,INDIA

Abstract— **Information on the web is growing huge in size and retrieval of relevant information from the web is a big challenge for search engines. Search engines retrieve huge set of documents out of which very few are relevant due to small size input query having few keywords. User information needs are vague or imprecise and changes his query as users are not conscious of their exact needs of information and thus making it difficult to understand the information need of the user and hence the precision of search results is low. Fuzzy Logic is an ideal tool to manage imprecise and vague information. In this paper an approach is proposed using Fuzzy logic to Personalize the web Search based on clustered query sessions. The initial user input query is represented by fuzzy set and expanded with related words based on Fuzzy thesaurus. This Fuzzy expanded input query is used to select the cluster which is most similar to the information need of the user input query. The selected cluster is used to recommend the Fuzzy ranked set of documents. The user's response to recommended search results is tracked to capture the user profile. The process of expansion of user profile based on fuzzy thesaurus and recommendations of Fuzzy ranked set of documents continue till the search is personalized to the information need of the user. Experiment was conducted on the data set of user query sessions captured in Academics, Entertainment and Sports domain and results confirm the improvement of precision of search results.**

Keywords— **Fuzzy Information Retrieval, Search Engines, Fuzzy Sets, Personalized Web Search, Information Scent.**

I. INTRODUCTION

Web is a huge repository of data which can be accessed using search engines. The user input query considered as the bag of words is matched against the textual representation of web data and retrieves millions of documents out of which very few are relevant. The ratio of relevant documents to the total retrieved documents is less and is responsible for the low precision of search results. The reason for low precision of search results is imprecise and vague input query as users are not conscious of their exact needs of information.

Hence to provide better service, a search engine should go for the understanding of the information need rather than depending on key words. This understanding will enable a search engine to correlate well with the user query.

Fuzzy logic has been applied in Information Retrieval to improve the precision of search results. Fuzzy logic is an ideal tool to manage imprecise and vague information and it has been widely used in the semantic web field.[7][9] In this paper Fuzzy logic has been applied for effective Personalization of web search based on clustered query sessions. The entire processing of the proposed approach is divided into two phase Offline and online. During offline processing, the data set is preprocessed to query sessions where each query session contains the user query and the associated clicked URLs. The term document matrix W global to entire data set is created using tf.idf vector of the distinct clicked documents present in the user query sessions captured on the web. 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. Query sessions are transformed into keyword vectors using tf.idf content and Information Scent of clicked URLs. These query sessions keyword vector are clustered in order to group user query sessions with similar need at one place. Each cluster groups query sessions with similar information need in a specific domain. The term-document matrix local to each cluster is built from global matrix W.

In 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 is used with the term document matrix of the selected cluster in order to identify the fuzzy set of ranked documents on set D where documents are ranked according to their membership in Fuzzy set. The user response to recommended Fuzzy set of ranked documents is tracked to collect the user profile. The user profile is vectored using content of its clicked documents and is expanded with related words based on Fuzzy thesaurus. This expanded 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.

Experiment was conducted on the data set of user query sessions collected in three domains Academics, Entertainment and Sports. The results verified statistically shows the improvement in precision of Personalized web search based on Fuzzy IR in comparison to classical IR and PWS(without Fuzzy IR).

II. RELATED WORK

In [2][8] modelling of user preferences have been described with fuzzy profiles. Their work was focused on a system where multiple (often diverging) sets of interests of an individual are modeled in a user profile. In order to personalize web search results, an adaptive algorithm is designed to learn these multiple profiles of users. But it was realized it makes little sense to keep unrelated profiles from past searches. In [19] a fuzzy ontology is combined with the TF-IDF measure for document ranking.

A novel classification approach was developed in [29] [26] in order to conceptualize documents into concepts using FFCM (Fuzzy Formal Conceptualization Model). In [6] Fuzzy logics were used to summarize text for extracting the most relevant sentences.

In [17] fuzzy logic is considered as a necessity to add deductive capability to a search engine. In [18] qualitative approach is expressed towards adding deduction capability to the search engine based on the concept and framework of protoforms. In [12] fuzzy conceptual graph is implemented in the machine-learning framework. In [13] the basic concept of fuzzy Bayesian Nets is presented for user modeling, message filtering and data mining. In [30] the fuzzy conceptual graph is presented for the semantic web. It is concluded that the usage of conceptual graph and fuzzy logic is complementary for the semantic web. In [31] conceptual matching of text notes to be used by search engines is presented. In [11] Fuzzy Reinforcement Learning (FRL) for text data mining and Internet search engine is used. In [3] a new technique is presented which integrates document index with perception index and can be used for refinement of fuzzy queries on the Internet. In [20] the extended profiles containing additional information related to the user are used to personalize and customize the retrieval process as well as the web site. Fuzzy clustering of these extended profiles is carried out and fuzzy rules are constructed. Fuzzy inference was used to modify queries and extract knowledge from profiles with marketing purposes within a web framework. In [24] the expressiveness of the queries can be increased with the use of fuzzy aggregation methods for intelligent search. In [4] the use of fuzzy ontology is proposed in search engines and it is built automatically from a collection of documents. Fuzzy ontology of term relations can be used for query refinement and to suggest narrower and broader terms suggestions during user search activity. In [10] fuzzy logic for rule-based personalization is introduced and can be implemented for personalization of newsletters.

In [7] fuzzy set model has been used to define fuzzy queries. In [28] [32] fuzzy relationship between query terms and documents is introduced.

In [16] three templates for the representation of keyword importance are proposed. In [25] fuzzy IR system uses fuzzy logic to retrieve documents similar to the query document. The system was tested on Arabic documents. In [21], 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. A third variable that considers the document structure was added which is the title variable that reflects the term frequency in the document title. idf was calculated by log(N/n), where N is the number of documents in the corpus and n is the number of documents that had the considered term.

III. BACKGROUND

A. *Fuzzy Set Theory in Information Retrieval(IR) System*

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 = $\{x_1,x_2,.....x_{m1}\}$ and other is a set of n relevant documents, D = $\{d_1, d_2, ..., d_n\}$

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 $< x_i, x_k > \in T$, $R(x_i, x_k)$ 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.

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

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

$$\forall x_1, x_2 \in T \quad (1)$$

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 is generated from initial input query using max min composition operator with set A and Fuzzy Thesaurus R.

That is, $AoR = B$ (2)

where o is the max-min composition.

The retrieved documents expressed by a fuzzy set RD defined on D, is then obtained by composing the augmented

inquiry, expressed by fuzzy set B, with the term document matrix W. That is

$BoW = RD$ (3)

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.[1][9]

B. 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. [22][23]

1) ***Information Scent metric***: A 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 i[th] query session. [5] [14]
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 follow

$s_{id} = PF.IPF(P_{id}) \times Time(P_{id}) \forall i \in 1..m \; \forall d \in 1..n$ (4)

$$PF.IPF(P_{id}) = \frac{f_{P_{id}}}{\max_{d \in 1..n} f_{P_{id}}} \times \log\left({}^{M}/_{m_{P_d}}\right) \quad (5)$$

$PF.IPF(P_{id})$: 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. [27]

2) ***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. That is

$$Q_i = \sum_{d=1}^{n1} s_{id} * P_{id} \qquad \forall\, i \in 1..m \quad (6)$$

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 eq 4. 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 (6). This vector is modeling the information need associated with the i[th] query session.

3) **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. [15] [33] 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 follows:

$$I = {}^{1}/_{M} \quad \sum_{p=1}^{k} \sum_{v_i \in C_p} sim(v_i, c_p) \quad (7)$$

where C_p be a cluster found in a k-way clustering process (p∈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 below . [34]

$$M = \quad \sum_{p=1}^{k} |C_p| \quad (8)$$

The centroid c_p of the cluster C_p is defined as below:

$$c_p = \left(\sum_{v_i \in C_p} v_i\right) \Big/ |C_p| \quad (9)$$

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.

IV. Use of Fuzzy Information Retrieval for effective Personalized Web Search based on clustered query sessions

In this paper an approach is proposed which is based on using fuzzy sets for effective Personalized Web Search. The use of Fuzzy sets is mainly due to imprecise and vague information need of the web users which is responsible for the low precision of search results. In this approach initial input query is represented by Fuzzy set A on the term set T is augmented with related terms using max-min composition with Fuzzy thesaurus R for better representing the information need of the user.

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

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.

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

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 using Fuzzy thesaurus R 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 where ranking is generated using composition of Fuzzy set FB and Fuzzy term document matrix W_j associated to a given cluster.

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

Thus an algorithm is proposed for personalized web search based on Fuzzy Information Retrieval. The entire processing of the algorithm is divided into two phases: Phase I and Phase II.

In Phase I offline processing is performed, the user query sessions collected on the web contains its associated clicked URLs. The content(tf.idf) of each distinct clicked URL present in the data set is fetched using crawler and loaded into database in the form of document-term matrix W for further processing. The Fuzzy thesaurus which is term-term correlation matrix is generated using W and W^T. Thus each query sessions is transformed into keyword vector using Information Scent and Content of clicked URLs. The resulting keyword vectors are clustered in order to group similar information need query sessions in one region. Each cluster is associated with term-document matrix of the clicked documents present in it. The stepwise execution of offline processing is given below.

Phase I
Offline Processing

1. Data Set of web queries and the associated clicked URLs is collected on the web and further preprocessed to get the Query Sessions.
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 term document relation W a fuzzy relation from D to T where |D|=n and |T|=m1 where D refers to the distinct clicked URLs collected from the 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 eq (1).
6. For each clicked URLs in a given query session, the Information Scent Metric is calculated which is the measure of the relevancy of the clicked URLs with respect to the information need of the user query session using eq(4).
7. Query sessions keyword vector is generated from query sessions using Information Scent and content of Clicked URLs(TF.IDF) using eq (6).
8. k-means algorithm is used for clustering query sessions keyword vector.
9. Each cluster j is associated with the mean keyword vector $clust\_mean_j$.
10. Generate Fuzzy term- document matrix W_j using tf.idf vector of distinct Clicked URLs CL_j present in a given cluster.

$$W_j : T \times CL_j \rightarrow [0,1]$$

In Phase II online Processing is performed. Due to imprecise and vague initial input query, the input query is augmented with related keywords based on Fuzzy thesaurus which also reduces the impact of uncertainty and vagueness of user's information need expressed with the input query. 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 associated with the selected cluster is used to determine another fuzzy set RD which is ranked set of documents which are more expressive than crisp relevance. The Fuzzy set RD on D using threshold α is presented to the user. The user clicks to recommended documents is tracked to capture the user profile. The user profile is then converted to keyword vector using information scent and content of clicked URLs. The user profile keyword vector is also augmented with related words based on Fuzzy thesaurus. This fuzzy set representing the augmented user profile is used to select the most similar cluster for the generation of next set of ranked documents based on Fuzzy set using max-min composition operation. This process of user profile expansion and recommendation of ranked set of documents based on fuzzy sets continues till the user information need is satisfied.

The stepwise execution of online processing is given below.

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 tif.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 eq (10)
4. Find the cluster j which is most similar to expanded input Fuzzy expanded input query B.

5. Use the term-document tf.idf matrix W_j 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 eq(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 ordered according to their degree of membership in set RD.
9. 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 information scent 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 eq (12).
    d. Select the j th cluster which is most similar to the information need associated with the FB.
    e. Identify the term-document tf.idf matrix W_j associated with the selected cluster j. The relevant part of the Wj matrix restricted to support of FB and non zero columns is used for computing the Fuzzy document set RD on D .
    f. Compute Fuzzy set RD ← FB o Wj using eq (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 ordered according to their degree of membership in set RD.
    i. Goto step 9.

else

Current search session is terminated.

V. Experimental Study

The experiment was conducted on a data set of user query sessions collected on the web. System architecture is developed in this work to collect the data set of user query sessions by capturing the user's clicks on Google search results. The user is required to enter the input query through a GUI based interface of the architecture which is then passed on to the Google search engine API, to retrieve the search results displayed with the check boxes on the user interface. A SnapShot of GUI interface of the architecture showing the Google search results for the input query "hindi song" is given below in Fig 1.

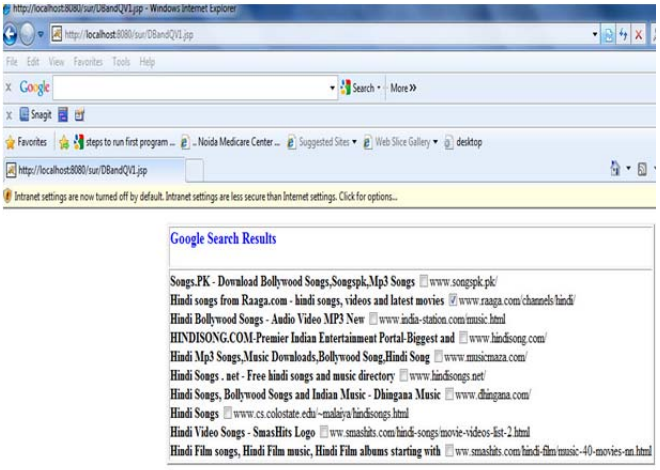

Fig. 1 Screen SnapShot of architecture displaying Google Search results along with the checkboxes.

The user clicks on the retrieved search results are captured through the check boxes displayed on the GUI and stored in the database in the form of query sessions. The captured user query sessions on the web are processed further to find the query session keyword vector using Information Scent and content of clicked URLs. The k-means algorithm is then applied to group the similar information need query session keyword vector in clusters.

In order to evaluate the performance of personalized web search based on fuzzy information retrieval, the experiment was performed on Pentium IV PC with 120 GB RAM under Windows XP using JSP, JADE , Oracle and MATLAB.

During offline preprocessing, the tf.idf vector of the clicked URLs of the query sessions are fetched using the web sphinx crawler and loaded into database using Oraloader. The clustering agent developed in JADE is executed to generate the clusters of query session keyword vectors.

The approach proposed for PWS using Fuzzy Information retrieval based on the clustered query sessions was compared with Classical IR/ PWS (without Fuzzy IR) in [27] based on clustered query sessions.

During online processing, the input query is issued to GUI based interface designed for both PWS with/ without Fuzzy Information Retrieval. In PWS with Fuzzy based

Information Retrieval, the augmented input query based on fuzzy set is used to select the cluster most similar to the information need of the user and the selected cluster is used to identify the fuzzy ranked set of documents. The ranked documents based on fuzzy sets are displayed with checkboxes to capture the user's clicks where the value of α=0.5.

The user's clicks to the fuzzy ranked documents are tracked to capture the user's profile and dynamically update the user's clicked profile during the search session of the user. When the user requests for the next result page, this captured user's profile is transformed into keyword vector and is augmented with related keywords based on fuzzy thesaurus. This augmented user profile is used to select the cluster similar to the information need of the current user profile. The selected cluster is used to generate the next set of ranked documents based on fuzzy sets. This process of generation of user profile and recommendations of ranked set of documents based on fuzzy set continues till the user search is personalized to the need of the user.

The performance of PWS based on Fuzzy IR is evaluated using the average precision of Personalized Search Results and compared with average precision of ranked Documents using Personalized Search Results (without Fuzzy IR) and Classical IR(Google Search Results) in each of the selected domains (Academics, Entertainment and Sports). In order to evaluate the performance, the 25 test queries were selected randomly in each of the domains Academics, Entertainment and Sports. The purpose of selecting the queries in these three domains is to cover wide range of queries on the web. The relevancy of the documents was decided by the experts in the domain to which the queries belong.

During online searching, the test queries were issued in each of the selected domain to the GUI based interface to retrieve the personalized search results(with/without Fuzzy IR) . The precision of a query is computed using the fraction of retrieved documents which are relevant in the personalized search results. The average of the precision of the collection of queries in a given domain is computed for comparing the performance results showing the average precision of test queries computed in the domains of academics, entertainment and sports using both PWS with / without Fuzzy IR/classical IR are shown in Fig 2.

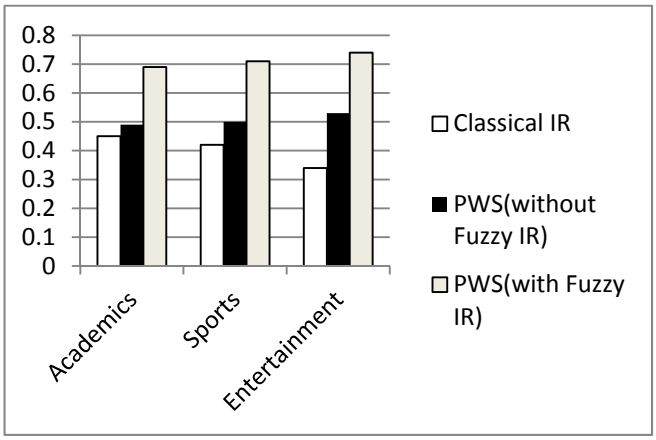

Fig.2. Shows the avgprecision of Classical IR/PWS (without/with Fuzzy IR) in Academics, Sports and Entertainment.

The average precision is improved in each of the selected domains using personalized web search with ranked documents(with fuzzy IR). The obtained results were analyzed using the statistical paired t-test for average precision of PWS (with/without Fuzzy IR). The comparison was done on the basis of data set of 25 queries in each of selected domain 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 was 58.09 for the combined sample. Value of t for paired difference of average precision was 39.90 for academics, 25.34 for entertainment and 46.38 for the sports categories. It was observed that the computed value for paired difference of average precision lies outside the 95% confidence interval in each case. Hence Null hypothesis was rejected and alternate hypothesis was accepted in each case and it was concluded that average precision improved significantly when personalized web search using Fuzzy IR in comparison to improvement in average precision of search results (without Fuzzy IR).

This proves that use of Fuzzy set for queries augmentation with related keywords and document ranking for PWS retrieves higher number of relevant clicked URLs up in top ranked clicked documents and increases their probability of being clicked by the users. The increase in the ratio of relevant documents to the total documents retrieved is responsible for the improvement in the precision in each of the selected domains. Thus the use of Fuzzy set for personalization of web search satisfies the user information need effectively. The experimental results which were also verified statistically confirm the significant improvement in precision in comparison to PWS (without Fuzzy IR) and classical IR.

VI. CONCLUSION

In this paper an approach is proposed for personalized web search based on fuzzy IR. The significance of using Fuzzy IR is to deal with vagueness and impreciseness of input queries which arises as users are not aware of their actual information need and as a result of which precision of search results are low. An algorithm is proposed for Personalized Web Search based on Fuzzy IR. The Fuzzy set is used for both query augmentation with related keywords and document ranking in order to disambiguate the context of input queries for better understanding the information need of the user and increases the probability of bringing relevant documents up in ranking. The process of recommendation of ranked documents and augmentation of queries with related keywords based on fuzzy set continue till the search is personalized to the information need of the user. Experiment was conducted on the data set of query sessions captured on the web in the domains: Academics, Entertainment and Sports. The results confirm the improvement in the average precision of search results using Personalized Web Search with Fuzzy IR.

REFERENCES

[1] B. Karn, "Information retrieval system using Fuzzy set theory-The Basic concept", Birla Institute of Technology.

[2] 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.

[3] D Choi , "Integration of document index with perception index and its application to fuzzy query on the Internet". 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.

[4] D. Widyantoro, and J.Yen. "Incorporating fuzzy ontology of term relations in a search engine". In *Proceedings of the BISC Int. Workshop on Fuzzy Logic and the Internet* , pp. 155-160,2001.

[5] E. H.Chi, P. Pirolli, K.Chen, and J. Pitkow. "Using information scent to model user information needs and actions and the Web." Proceedings of the *SIGCHI conference on Human factors in computing systems*, 490-497, ACM, 2001.

[6] 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.

[7] 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.

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

[9] G.J. Klir, and Yuan, 60, "Fuzzy Sets and Fuzzy Logic Theory and Application" , PHI 1994, Pp. 379-387.

[10] 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.

[11] H Beremji ,"Fuzzy reinforcement learning and the internet with applications in power management or wireless networks". 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.

[12] Ho KHL , " Learning fuzzy concepts by example with fuzzy conceptual graphs". In 1st Australian Conceptual Structures Workshop, Armidale, Australia, 1994.

[13] JF Baldwin, and SK Morton . "Conceptual Graphs and Fuzzy Qualifiers in Natural Languages Interfaces", University of Bristol, 1985.

[14] J. Heer, and E. H.Chi. "Separating the swarm: categorization methods for user sessions on the web." Proceedings of the *SIGCHI Conference on Human factors in Computing Systems*, 243-250, ACM,2002.

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

[16] K.Nowacka, S.Zadrozny, and J.Kacprzyk. "A new fuzzy logic based information retrieval model." Proceedings of Information Processing and Management of Uncertainty in Knowledge-Based System(IPMU'08), pp.1749-1756.June 22-27, 2008

[17] 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.

[18] LA Zadeh , "A Prototype-Centered Approach to Adding Deduction Capability to Search engines – the concept of protoform", BISC Seminar, UC Berkeley,2002.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] S. Miyamoto. Fuzzy sets in Information Retrieval and Cluster Analysis. Kluwer Academic Publishers, 1990.

[29] 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.

[30] TH Cao , "Fuzzy conceptual graphs for the semantic web". 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.

[31] T Takagi, and M Tajima ,"Proposal of a search engine based on conceptual matching of text notes". 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.

[32] 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.

[33] Y. Zhao, and G. Karypis." Comparison of agglomerative and partitional document clustering algorithms", *SIAM Workshop on Clustering High-dimensional Data and its Applications,* 2002a.

[34] Y. Zhao, and G. Karypis. "Criterion functions for document clustering: Experiments and Analysis". *Technical report*,University of Minnesota, Minneapolis, MN,pp. 01-40, 2002b.