



## **Intelligent Information Retrieval Using Hybrid of Fuzzy Set and Trust**

**SURUCHI CHAWLA**

Shaheed Rajguru College of Applied Science for Women,  
University of Delhi, Vasundhra Enclave, Delhi, India.

\* Corresponding author E-mail: [sur\\_chawla@rediffmail.com](mailto:sur_chawla@rediffmail.com)

<http://dx.doi.org/10.13005/ojcs/10.02.09>

(Received: March 03, 2017; Accepted: April 14, 2017)

### **ABSTRACT**

The main challenge for effective web Information Retrieval (IR) is to infer the information need from user's query and retrieve relevant documents. The precision of search results is low due to vague and imprecise user queries and hence could not retrieve sufficient relevant documents. Fuzzy set based query expansion deals with imprecise and vague queries for inferring user's information need. Trust based web page recommendations retrieve search results according to the user's information need. In this paper an algorithm is designed for Intelligent Information Retrieval using hybrid of Fuzzy set and Trust in web query session mining to perform Fuzzy query expansion for inferring user's information need and trust is used for recommendation of web pages according to the user's information need. Experiment was performed on the data set collected in domains Academics, Entertainment and Sports and search results confirm the improvement of precision.

**Keywords:** Search engines, Intelligent Information Retrieval, Fuzzy Set, Trust, Query Expansion, Web page recommendation.

### **INTRODUCTION**

Search engines are used widely for web information retrieval. The search results precision is low due to imprecise and vague search queries. Research has been done in web information retrieval for improving the precision of search results.<sup>1,2,3,4,5,6,7,8,9</sup> Fuzzy set has been used in the

framework of web information retrieval to deal with vague and imprecise web data.<sup>10</sup> Trust has shown to identify the relevant documents for recommendation and performs effective personalization<sup>11</sup>. In this research a novel approach is proposed for Intelligent Information Retrieval using hybrid of fuzzy set and trust in web query session mining for inferring the user's query information need

and recommend relevant documents for effective information retrieval. The benefit of using Fuzzy set and trust together is because of the following reasons. Fuzzy set deals with imprecise and vague data and has been used for expansion of queries which are imprecise and vague due to limited vocabulary of users and small size. The trust value of clicked URLs selects the relevant document for recommendation. The trusted web pages selected for recommendations based on collaborative filtering will retrieve relevant search results on the assumption that new users preferences will be similar to nearest neighbors preferences on trusted clicked URLs in clusters. Therefore the use of hybrid of Fuzzy set and trust in query session mining manages the retrieval of trusted documents in the presence of vague and imprecise user queries for effective Information Retrieval.

In this research an algorithm is designed for Intelligent Information Retrieval using hybrid of fuzzy set and trust for web page recommendations. There are two phases of processing of the proposed algorithm: Phase I (Offline Processing) and Phase

II (Online Processing). In offline processing, data set is pre-processed to obtain web query sessions. The query session keyword vector is generated using Information Scent and clicked URLs (TF, IDF). Information Scent measure the relevance of document (clicked URLs) to user's information need based on its web usage and is calculated using PF, IPF and Time. The clustering method is used for grouping keyword vectors and therefore identifies the clicked URLs relevant in a specific domain. The trust of clicked URLs in a given cluster is initialized using Information Scent.

The global term document matrix  $W$  and Fuzzy thesaurus  $R$  is formed based on  $tf, idf$  vector of clicked URLs in clustered query session.  $W$  represent the term-document relationship for all distinct clicked URLs set  $D$  present in the entire data set and the terms of matrix is the set of distinct terms set  $T$  found in set  $D$ .  $R$  is built from  $W$  and  $W^T$  in order to create term-term correlation matrix.  $W_j$  is the Fuzzy term trusted document matrix formed for each cluster  $j$  based on trusted clicked URLs.

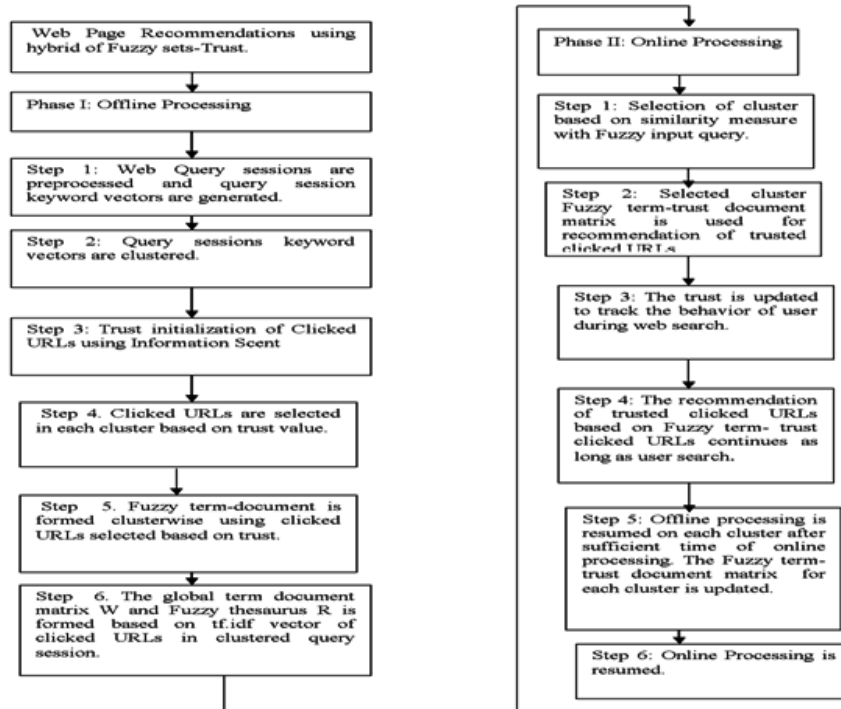


Fig.1: Flowchart of steps of Offline and Online processing of Intelligent IR using Fuzzy sets and trust.

During online processing, the user input query based on Fuzzy Set A using term set T is expanded with related terms using 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 and words synonym. The cluster is selected using Fuzzy expanded input query based on similarity measure and is used to recommend the fuzzy set of trusted documents on set D using term- trusted document matrix. The recommended documents are ranked according to their trust value. The user response to recommended documents is captured to collect the user profile and the trust of clicked URLs is updated. The user profile vector is generated from content and trust of its clicked documents and expanded with related words based on Fuzzy thesaurus. The cluster is selected based on expanded user profile for the recommendation of Fuzzy set of trusted documents. Thus the user profile expansion with related terms for the recommendations of Fuzzy set of trusted documents continues for user's information need satisfaction. Experiment was performed on the data set in Academics, Entertainment and Sports domains and results were statistically analyzed to confirm the effectiveness of proposed approach. The flowchart of steps of offline and online processing is given in Fig 1.

### Related Work

Trust based recommender system for personalized web search had been more effectual than traditional collaborative filtering<sup>12,13,14,15</sup>. In<sup>16</sup> both trust metric and similarity metrics 'are used for trust based recommender system. In<sup>17</sup> collaborative filtering is improved with trust based metrics where trust metric is used for computation of similarity value. In<sup>18</sup> Trust Walker: A Random Walk Model is proposed using both trust-based and item-based recommendation. In<sup>19</sup> trust based web page recommendations is used for personalized web search.

In<sup>20</sup> fuzzy logic is used for adding deductive capability to a search engine. In<sup>21</sup> Fuzzy Reinforcement Learning (FRL) is used for text data mining and Internet search engine. In<sup>22</sup> a new technique is presented using both document

index and perception index for fuzzy web queries refinement. In<sup>23</sup> 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.

In<sup>24,25</sup> modelling of user preferences have been described with fuzzy profiles. In<sup>26</sup> fuzzy IR system uses fuzzy logic to retrieve documents similar to the query document. The system was tested on Arabic documents. In<sup>27</sup> Fuzzy logic is used to deal with imprecise and vague queries for effective information retrieval.

In<sup>28</sup> flexibility of IRS has been increased using fuzzy set theory. In<sup>29,30</sup> Fuzzy generalizations of the Boolean model ( representation of documents and the query language) have been used with the objective of generating discriminated answer for user queries in IRS. In<sup>31</sup> Fuzzy associative mechanisms are used for the representation of documents and users' queries.

In<sup>32</sup> Query Logs for Query Recommendation (QLQR) is proposed that use web page browsing and clicking history for developing user model enables to predict user actions and intends for Intelligent Information Retrieval. Sequential access pattern mining is used in an Intelligent web personalization called WAPPS<sup>33</sup> and SEWEP<sup>34</sup>. Pattern tree based on frequent sequential web access pattern using CS-mining generates online recommendations. In<sup>35</sup> Knowledge based Personalized Recommendation Service System (KBPRSS) is proposed based on filtering or clustering the mass data. The internet users are clustered based on their similarity in usage pattern and a cache of web document is maintained for each of the cluster for identifying the relevance in their perspective.

Hybrid of Fuzzy logic and trust has been applied in<sup>36</sup> where trust between users in web based social network information retrieval system is inferred based on Fuzzy logic. In<sup>37</sup> Fuzzy logic is used to combat web spam with trust rank. Trust rank overcomes the Page Rank problem and uses

the Fuzzy logic in seed selection process for finding the page is spam or not. In<sup>38</sup> Fuzzy based trust measures are developed and validated for online medical diagnosis and symptoms analysis. In<sup>39</sup> combination of Trust based access control strategy and Fuzzy expert systems was used to increase the quality of human computation in crowd sourcing environment.

Hybrid of Fuzzy sets and trust has been used together in various domains and results shows promising. It is found no work has been done using hybrid of trust and fuzzy logic in web query session mining for intelligent web search. The benefit of using trust with fuzzy logic is twofold first is Fuzzy sets enables to infer the user information need from imprecise and vague search queries and secondly clicked URLs selected based on trust value in query sessions mining identify the relevant web pages for recommendation using nearest neighbor in collaborative filtering. Therefore the hybrid of trust and fuzzy sets is proposed in this paper for intelligent Information retrieval using trusted web page recommendations based on fuzzy queries.

## Background

### Trust

The Trust has been used in online recommender systems and gained increase amount of attention in research communities. A trust is a social phenomena and its web model is based on how trust works between people in society.<sup>40,41</sup>

In<sup>42</sup> the general properties of trust in e-services were identified and mentioned as given below:

- Trust is relevant to specific transactions only.
- Trust is a measurable belief.
- Trust is directed.
- Trust exists in time.
- Trust evolves in time, even within the same transaction.
- Trust between collectives does not necessarily distribute to trust between their members.
- Trust is reflexive,
- Trust is a subjective belief.

In<sup>43</sup> the two models for trust called profile and item level were proposed and incorporated into collaborative recommendation process. It was found that trust reduces the prediction error rate and the prediction accuracy is improved.

### Information Scent

Information scent is the measure of sense of value and cost of accessing a web page using perceptual cues with respect to the information need of user. The users click those web pages in search results which seem relevant to information need of search session. Information Scent will be high for those clicked web pages which satisfy the information need of user. The information need can be inferred from surfing pattern based on interactions between user need, user action and content of web.<sup>44,45</sup>

### Information Scent metric

The Inferring User Need by Information Scent (IUNIS) algorithm quantifies the Information Scent  $S_{id}$  of the clicked pages  $P_{id}$  clicked in  $i^{\text{th}}$  query session as follows.<sup>46,47</sup>

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

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

PF.IPf ( $P_{id}$ ): PF is the page 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 is the ratio of total number of query sessions  $M$  in the whole data set to the number of query sessions  $M_{P_{id}}$  that contain the given page  $P_{id}$ .

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  $i$ .<sup>48,49,50,51,52</sup>

The information scent <sub>$id$</sub>  is calculated for all clicked pages in web query sessions mining as given above in Eq(1) and Eq(2).

### Generation of Query sessions keyword vector

The query session keyword vector is generated from query session. The web query

session contains the input query along with the clicked URLs and is represented as follows

$$\text{query session} = (\text{input query, (clicked URLs/ Page)}^+)$$

'+' indicates only those sessions are considered which have at least one clicked Page associated with the input query.

The *i*th query session keyword vector  $Q_i$  is generated using summation of product of content vector of each clicked page  $P_{id}$  scaled by the weight  $s_{id}$  as mentioned below in Eq(3).

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

In the above formula *n1* is the number of distinct clicked pages in session *i*. The  $s_{id}$  is the information scent of clicked page and  $P_{id}$  is the TF.IDF content vector.  $Q_i$  keyword vector models the information need associated with the *i*th query session.

**Clustering of Query session keyword vector**

The *k*- means algorithm has shown good performance for document clustering therefore used for clustering query session keyword vectors.<sup>53,54</sup> In vector space implementation of *k*- means, criterion function is used for measuring the quality of clusters. The criterion function is measured using average similarity between vectors and centroid of the assigned clusters.

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

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

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 *i*th query session belonging to the cluster and *M* is the total number of query sessions in all clusters as defined in Eq(5).<sup>55</sup>

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

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

$$c_p = \frac{(\sum_{v_i \in C_p} v_i)}{|C_p|} \quad \dots(6)$$

where  $|C_p|$  denotes the number of query sessions in cluster  $C_p$  and  $sim(v_i, C_p)$  is calculated using cosine measure.

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

Fuzzy information retrieval methods are based on fuzzy set in order to handle uncertain information. Tools defined in fuzzy logic and fuzzy relations are used to infer the best results to a user query. Fuzzy systems are most effective for dealing with data that require a degree of membership.

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

A fuzzy document—term relation *W* is a fuzzy relation from *D* to *T*. *W* represents the relevance of index terms to individual documents  $W: D \times T \rightarrow [0,1]$  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  $\langle x_i, x_k \rangle \in T$  the degree of association of  $x_i$  with  $x_k$  is expressed using  $R(x_i, x_k)$  in Eq (7) that measure the degree to which the meaning of the index term  $x_k$  is compatible with index term  $x_i$ . The relation  $R(x_i, x_k)$  solves the synonyms problem and used to retrieve the document relevant to user queries containing different terms but synonyms to each other.  $R: T \times T \rightarrow [0,1]$

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

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

$$\text{That is, } A \circ R = B \quad \dots(8)$$

where o is the max-min composition.

The fuzzy set RD defined on D in Eq (9), represents the retrieved document set. The set RD is obtained using composition of augmented inquiry fuzzy set B and term document matrix W as mentioned below.

$$\text{BoW} = \text{RD} \quad \dots(9)$$

The documents represented by Fuzzy set RD is presented to user in whole or selected by some  $\alpha$ - cuts of RD. <sup>56,57</sup>

In [Chawla(2015b)]<sup>27</sup> initial input query expressed by Fuzzy set A on 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 Eq (10).

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

The augmented query B is then used for the selection of cluster relevant 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 Eq (11).

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

The documents in Fuzzy set RD are recommended and clicks to recommended documents are captured in user profile. For next result page, the user profile keyword vector (FUP) is generated and further expanded with related terms FB using Fuzzy thesaurus R in Eq (12) for handling the vagueness due to synonyms and poly- semis of vocabulary used in profile.

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

This expanded user profile fuzzy set FB on set T select the cluster for the recommendations 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 Eq (13).

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

### Use of Trust & Fuzzy set in web mining for Intelligent Information Retrieval using web page recommendations.

In this paper a novel method is proposed for intelligent information retrieval based on web page recommendations using trust & Fuzzy set in web query session mining. The user queries are imprecise due to small size query and limited vocabulary of users. Fuzzy set based query expansion is used for adding related words to user queries and disambiguates query context for inferring the user information need. The trust of clicked URLs is computed based on user's clicks and identifies the relevant documents for web page recommendation. An algorithm is proposed using hybrid of Fuzzy sets and trust in query session mining where query expansion based on fuzzy set disambiguate the user's queries and the cluster is selected similar to user's information need for the recommendation of relevant web pages based on trust. The entire processing of the algorithm proposed for Web Page recommendations using hybrid of Fuzzy set and Trust is divided into two phases Phase I (Offline Processing) and Phase II (Online Processing).

**Offline Preprocessing**

- Preprocessing of web data set generates the Query Sessions.
- Generate the tf.idf vector of each distinct clicked URLs present in the data set.
- The set of distinct terms present in data set is represented by set  $T=\{x1,x2,x3,\dots,xm1\}$  and set of distinct documents is represents by set  $D=\{d1,d2,d3,\dots,dn\}$
- 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.
- A fuzzy thesaurus or fuzzy term—term relation  $R$  is a fuzzy relation from  $T$  to  $T$  identifying the synonym relations using equation (7).
  - Information Scent Metric is calculated for each clicked URLs in the query sessions using Equation. (1).
  - The trust of a given clicked URL  $d$  in session  $i$  is initialized with Information Scent.  
 $trust_{id} = sid, d \in 1..n, i \in 1..m$
  - Query sessions keyword vector is obtained using trust and TF.IDF of Clicked URLs as given below.

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

Where  $n1$  is the number of distinct clicked URLs in a given session  $i$

- k-means algorithm is used for clustering query sessions keyword vectors.
- For all clusters, the mean keyword vector of a given cluster  $i$  is represented by cluster $_i$ \_mean.
- For trust computation ,clicked count and recommended count defined for each distinct clicked URLs are initialized to zero within each cluster .
- The initial trust value of a given cluster  $i$  is calculated as follows  
 $Trust(i)=\{ClickedURLs: Trust(ClickedURLi)>j \in \epsilon_1..n_1$  where  $n1$  is number of distinct clicked URLs in cluster  $i\}$
- Generate Fuzzy term- trusted document matrix  $W_j$  using tf.idf vector of trusted Clicked URLs  $CL_j$  having  $Trust(ClickedURLi) \geq \epsilon$  for clicked URLs present in a given cluster  $i$ .  
 $W : T \times CL_i \rightarrow [0,1]$

During Offline processing, the web query sessions are collected and are preprocessed to contain the query and its associated clicked URL. The information scent of clicked URL in a given query session is calculated based on clicked URLs usage statistics in data set and measure its relevancy for the session information need. The trust value of clicked URLs is initialized using their Information Scent Value. The trust of a cluster is defined to be the number of clicked URLs having trust value above threshold value. The information need of query session is modeled using trust and content of clicked URLs and generates the keyword vector. The query session keyword vectors are clustered to identify the URLs in groups. The trusted

clicked URLs in each cluster are selected based on threshold value and form the Fuzzy term trusted document matrix  $W_j$  based on tf. idf local to a given cluster  $j$ .

The global tf. idf term –document matrix  $W$  is created based on distinct document set  $D$  found in entire data set and distinct terms  $T$  (after stop word removal and stemming) used in documents. Fuzzy thesaurus  $R$  is formed based on  $W$  term-document matrix and identifies the term –term correlation in entire data set.

In Online processing, the input query  $Q$  uses Fuzzy thesaurus  $R$  for expansion with

## Phase II

### Online Processing

1. Final query Q after removing all stop words is represented by Fuzzy set A on T.
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  $B \rightarrow A \circ R$  using equation (10)
4. For each cluster i, the similarity of Fuzzy expanded input query B with cluster<sub>i</sub>\_mean is measured using the formulae  

$$\text{Match Score}_i(B, \text{cluster}_i) = \begin{cases} 2(\text{sim}(B, \text{cluster}_i\text{-mean}) * \text{Trust}(i)) / \text{Trust}(i) + \text{sim}(\text{cluster}_i\text{-mean}, B) & \text{when Trust}(i) \neq 0 \\ \text{sim}(B, \text{cluster}_i\text{-mean}) & \text{when Trust}(i) = 0 \end{cases}$$
5. Select the cluster j with the maximum Match Score.
6. if the  $\text{Trust}(j) = 0$  and  $\text{Match Score}_i(\text{input query}, \text{cluster}_j) < 0.5$   
 Msg "No Trusted Recommendations. Terminate the current session, Initiate new search session" .  
 Goto step 1.  
 Else if  $(\text{Trust}(j) \neq 0$  or  $\text{Match Score}_i(B, \text{cluster}_j) \geq 0.5)$   
 Use the term-trusted document tf.idf matrix W<sub>j</sub> of the cluster j and select the relevant part of the W<sub>j</sub> matrix restricted to support of B and non zero columns for computing the Fuzzy document set RD on D.
7. Compute Fuzzy Document set  $RD \rightarrow B \circ W_j$  using equation (11)
8. Inspect only those document\_ids in RD captured by some  $\alpha$ -cut of RD in which only those documents are filtered for retrieval whose degree of relevance is greater than or equal to  $\alpha$ .
9. For each selected document d in  $\alpha$ -cut of RD is retrieved and stored in L.
10. The list L is recommended to the user in decreasing order of their trust value
11. The Recommended count of the URLs in list L are incremented by 1.
12. The user clicks to the recommended URLs are captured and stored in current user profile.
13. For each clicked recommended URL, the Clicked Count is incremented by 1.
14. The trust value of the recommended URLs is updated as given below.  

$$\text{Trust}(\text{URL}_i) = \{1 - \text{Distrust}(\text{URL}_i) \mid \text{where Recommended count}(\text{URL}_i) \neq 0\}$$
  

$$\text{Distrust}(\text{URL}_i) = \left\{ \frac{(\text{Recommended count}(\text{URL}_i) - \text{Clicked Count}(\text{URL}_i))}{\text{Recommended count}(\text{URL}_i)} \right\}$$
15. Trust value of the selected cluster j is defined as follows  

$$\text{Trust}(j) = \frac{|\text{Correct Set}(j)|}{|\text{RecSet}(j)|}$$
  

$$\text{Correct Set}(j) = \{|\text{Clicked URLs}_j \mid \text{Trust}(\text{ClickedURL}_j) > \epsilon \text{ where Recommended count}(\text{ClickedURL}_j) \neq 0\}$$
  

$$\text{RecSet}(j) \text{ is the total number of recommendations made using cluster } i$$
  

$$\text{RecSet}(j) = \{|\text{Clicked URLs}_j \mid \text{Recommended count}(\text{ClickedURL}_j) \neq 0\}$$
16. For each request of next result page
  - a. The users clicks to the search results captured in user profile generates the keyword vector.
  - b. The user profile keyword vector FUP is generated using the trust and content of the clicked URLs .
  - c. Compute expanded Fuzzy User Profile  $FB \rightarrow FUP \circ R$  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 equation (12).
  - d. OF all clusters i select the j th cluster similar to the information need associated with the FB.  

$$\text{Match Score}_i(\text{cluster}_i\text{-mean}, FB) = \begin{cases} 2 * (\text{sim}(FB, \text{cluster}_i\text{-mean}) * \text{Trust}(i)) / \text{Trust}(i) + \text{sim}(\text{cluster}_i\text{-mean}, FB) & \text{Trust}(i) \neq 0 \\ \text{sim}(FB, \text{cluster}_i\text{-mean}) & \text{Trust}(i) = 0 \end{cases}$$
  - e. The most matching cluster j is selected.
  - f. if the  $\text{Trust}(j) = 0$  and  $\text{Match Score}_i(FB, \text{cluster}_j) < 0.5$



Msg "No Trusted Recommendations. Terminate the current session, Initiate new search session". Goto step 1.

Else if (Trust(j) != 0 or Match Score<sub>i</sub> (FB, cluster<sub>j</sub>) >= 0.5)

Use the term-document tf.idf matrix W<sub>j</sub> of the selected cluster j and select the relevant part of the W<sub>j</sub> matrix restricted to support of FB and non zero columns for computing the Fuzzy document set RD on D.

g. Compute Fuzzy set RD → FB o W<sub>j</sub> using equation (13).

h. Jump to step 8.

Else

Current search session is terminated.

End

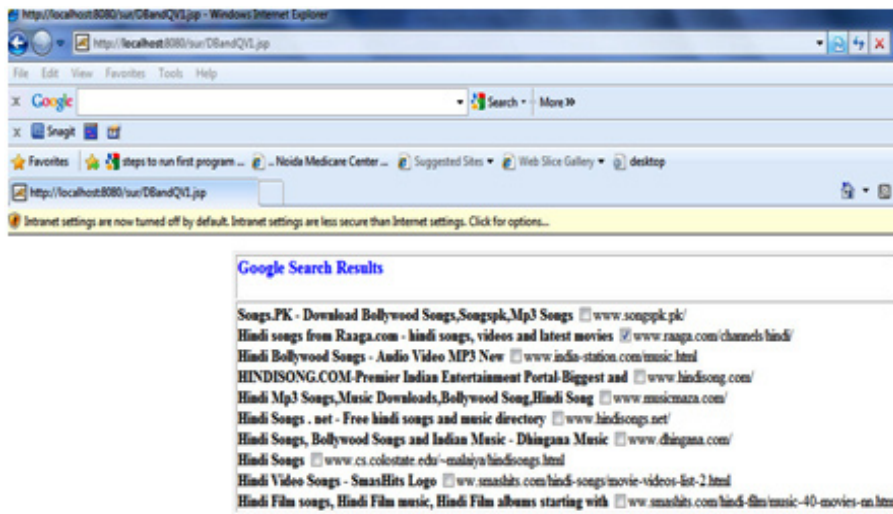


Fig.2: Shows the GUI of architecture displaying Google Search results along with the checkboxes.

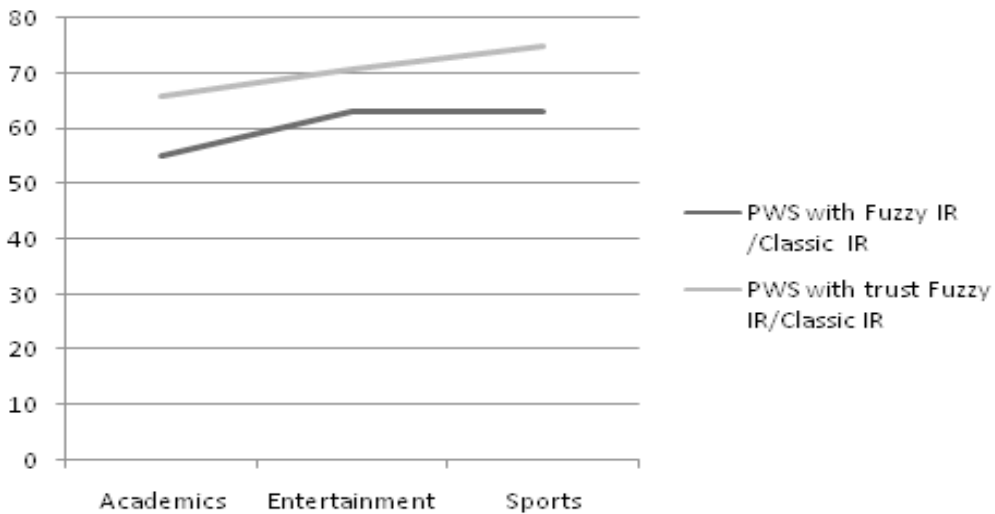


Fig. 3: Plots the percent improvement of PWS with Fuzzy/Trust Fuzzy IR over Classic IR

related keywords to determine the user information need. The Fuzzy expanded query B on term set T is used to select the cluster for trusted web page recommendations. The Fuzzy term-trusted document matrix of the selected cluster is used with Fuzzy expanded query B to retrieve the fuzzy set RD trusted documents on document set D for recommendation. The Fuzzy set RD on D using threshold  $\alpha$  is presented to the user. The user clicks to recommended trusted documents is captured in profile FUP and the trust is updated for each cluster

as well as of recommended clicked web pages. On each request of next web page, the user profile is vectored using trust and content of clicked URLs and expanded based on Fuzzy thesaurus. The fuzzy expanded user profile FB is used to identify the cluster for the recommendations of documents based on Fuzzy trusted document matrix. The recommendation of trusted clicked URLs and user profiling is used for personalizing web search and cater to the user information need.

**Table 1: Shows the test queries used for experimental evaluation of Fuzzy IR(with/without trust)**

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

**Table 2: Shows the average precision of test queries using Fuzzy IR(with/without trust) and percentage improvement over Classic IR(Google Search Engine)**

25 test Queries	Average Precision using Information Retrieval Techniques			% improvement over Classic IR	
	Classic IR	PWS with Fuzzy IR	PWS with Trust FuzzyIR	PWS with Fuzzy IR	PWS with Trusted FuzzyIR
Academics	0.45	0.7	0.75	55	66
Entertainment	0.46	0.75	0.79	63	71
Sports	0.44	0.72	0.77	63	75

### Experimental Study

Experiment was performed on the data set of web query sessions collected through the GUI based architecture in domains Academics, Entertainment and Sports. The GUI based architecture is developed using JADE, JSP and database Oracle and allows the users to enter the input query to retrieve the search results. The search results are displayed with check boxes and user's clicks are captured using check boxes in order to store them as web query session in database for mining. The search results for a given user query are displayed with check boxes in GUI interface as given in Fig2. below.

The experiment was conducted on i3 processor with 2 GB RAM on Windows 8. The clustering agent was developed in JADE in order to execute the clustering algorithm on query session keyword vectors. The Web Sphinx Crawler generated the tf. idf vector of clicked URL and vector was stored in database using or a loader. During the execution of clustering agent, clusters of query session keyword vectors were generated and Fuzzy term-trusted 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 data set. The parameters used for experimental evaluation were Information Scent lies in [0,1] and trust threshold value is set to 0.5.

The performance of the proposed approach was evaluated from the average precision of recommended search results and compared with Fuzzy IR (without Trust). The performance was evaluated using test queries selected in three domains, there were 25 in Academics, 25 in Entertainment and 25 in Sports. The sample of queries used for evaluation is given in Table1. below.

The queries were chosen in a selected domain in order to have wider coverage of web queries. During online web search, the test queries are issued to GUI based interface for Intelligent Information Retrieval using both Fuzzy IR (without trust) in<sup>27</sup> and Fuzzy IR( with trust) the proposed approach. The recommended clicked URLs (with/ without trust) are displayed in decreasing order

of their relevance. The user's response to the recommended search results were captured in user's profile and the trust of clicked URLs was updated.

The average of test queries precision is computed in the selected domains where precision is fraction of retrieved documents that are relevant. The average precision of 25 test queries were calculated during experiment evaluation in academics, entertainment and sports domains using PWS with Fuzzy IR (with/ without trust) and Classic IR is shown in Table 2. and Fig 3.

The Fig. 3 shows the percent improvement in average precision of PWS with Fuzzy IR (with/ without trust) over Classic IR. The results show that percentage improvement of Fuzzy IR (with Trust) over classic IR is more than Fuzzy IR (without trust) over classic IR. The obtained results were statistically analyzed using the paired t-test for average pre-cision of PWS with trusted Fuzzy IR versus both (Classic IR/Fuzzy IR) with 74 degrees of freedom (d.f.) for the combined sample as well as with 24 d.f. in selected categories (Academics, Entertainment and Sports). The t value for average precision of Fuzzy IR (withTrust) versus (ClassicIR) for a combined sample was 84.6, 82.15 in Academics, 35.79 in Entertainment and 110 in Sports. The t value for average precision of Fuzzy IR (withTrust) versus Fuzzy IR (without trust) for a combined sample was 6.76, 4.6 in Academics, 56.12 in Entertainment and 11.09 in Sports.

The computed t value for paired difference of average precision was outside the 95% confidence inter-val in each case. The value of t shows that Web Information retrieval based on Fuzzy trust shows the high t value both over classic IR and Fuzzy IR. Therefore Null hypothesis was rejected and alternate hypothesis was accepted in each case and it was concluded that trusted web page recommendations with fuzzy IR shows the significant improvement in the average precision. Hence proposed approach of Intelligent Information Retrieval using trust based Fuzzy IR proves to be effective based on experimental results. Thus the use of Fuzzy expansion of user queries to infer the information need and recommendations of trusted

web pages to retrieve relevant documents for effective Information retrieval improve the precision of search results.

### CONCLUSION

An approach is proposed using trust in Fuzzy Information Retrieval. The use of fuzzy set for queries expansion and selection of trusted web pages for recommendations retrieves relevant

web pages for information retrieval. An algorithm is proposed for web page recommendations using Trust and Fuzzy IR in query session mining. Experiment was performed on the web query sessions data set in selected domains mainly entertainment, academics and sports. The experimental results confirm the significant increase in the precision of search results therefore s trust based Fuzzy IR personalizes the web search effectively.

### REFERENCES

1. Yu, J.; Liu, F. Mining user context based on interactive computing for personalized Web search. In 2nd International Conference on Computer Engineering and Technology (ICCET),2, 2010; pp. 209-214, IEEE.
2. Pan, X.; Wang, Z.; Gu, X. Context-based adaptive personalized web search for improving information retrieval effectiveness. In 2007 International Conference on Wireless Communications, Networking and Mobile Computing, 2007; pp. 5427-5430, IEEE.
3. Kutub, M.; Prachetaa, R.; Bedekar, M. User Web Search Behaviour. In 3rd International Conference on Emerging Trends in Engineering and Technology (ICETET),2010; pp. 549-554, IEEE.
4. Arzanian B.; Akhlaghian, F.; Moradi, P. A Multi-Agent Based Personalized Meta-Search Engine Using Automatic Fuzzy Concept Networks. In Third International Conference on Knowledge Discovery and Data Mining, 2010; pp. 208-211.
5. Matthijs ,N.; Radlinski, F. Personalizing web search using long term browsing history. In Proceedings of the fourth ACM international conference on Web search and data mining, 2011; pp. 25-34.
6. Xu, S.; Jiang, H.; Lau, F. C. M. Mining user dwell time for personalized web search re-ranking. In In Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, 11,2011; pp. 2367-2372.
7. Chawla, S. Semantic Query Expansion using Cluster Based Domain Ontologies. International Journal of Information Retrieval Research (IJIRR), 2012b; 2(2), pp. 13-28 .
8. Chawla, S. Personalized web search using ACO with information scent. International Journal of Knowledge and Web Intelligence, 2013; 4(2), pp. 238-259.
9. Chawla, S. A novel approach of cluster based optimal ranking of clicked URLs using genetic algorithm for effective personalized web search. Applied Soft Computing, 2016; 46, 90-103.
10. Lee, J.; Kim, E. Fuzzy Web Information Retrieval System,2011.
11. Massa, P.; Avesani, P. Trust-aware Recommender Systems. Proceedings of the ACM Conference on Recommender Systems, 2007; pp. 17-24.
12. Levien, R. Attack-resistant Trust Metrics. Ph.D. thesis, 2004, University of California at Berkeley, USA.
13. Lathia N.; Hailles, S ;Capra, L. Trust-based collaborative filtering. Proceedings of the joint iTrust and PST Conference on Privacy, Trust Management and Security. Springer, 2008; pp. 119-134.
14. Hwang C.,S and Chen, Y. P.Using trust in collaborative filtering recommendation. Lecture Notes in Computer Science, 4570, 2007; 1052-1060. Innovation Network (2009).
15. Peng T.; Seng-cho, T. iTrustU: A blog recommender system based on multifaceted trust and collaborative filtering. Proceedings

- of the ACM Symposium on Applied Computing. New York, NY. 2009; pp. 1278-1285.
16. Massa, P.; Bhattacharjee, B. Using trust in recommender systems: An experimental analysis. *Proceedings of the Second International Conference on Trust Management*, Oxford, UK., 2004; 221-235, (2004).
  17. Jianshu Weng, Chunyan Miao, Angela Goh. Improving Collaborative Filtering with Trustbased Metrics”, SAC’06, April, 2327, Dijon, France, ACM 1595931082/ 06/0004, (2006).
  18. Jamali, M. and Ester M. TrustWalker: A Random Walk Model for Combining Trust-based and Item-based Recommendation”. *Proceedings of the 15th ACM Conference on Knowledge Discovery and Data Mining*. KDD.09, Paris, France, 2009.
  19. Chawla, S. Trust in Personalized Web Search based on Clustered Query Sessions. *International Journal of Computer Applications*, 59(7), 2012a; pp. 37-44.
  20. Zadeh, LA. 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.
  21. Beremji, H. 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.
  22. Choi, D. 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.
  23. Rubens, N. O. The application of fuzzy logic to the construction of the ranking function of information retrieval systems., *Computer Modelling and New Technologies*, 2006; 10(1), pp. 20–27.
  24. Mencar, C.; Torsello, M. A.; Dell’Agnello, D.; Castellano, G.; Castiello, C. Modeling user preferences through adaptive fuzzy profiles. In *2009 Ninth International Conference on Intelligent Systems Design and Applications*, 2009; pp. 1031-1036, IEEE.
  25. Castellano, G.; Dell’Agnello, D.; Fanelli, A. M.; Mencar, C.; Torsello, M. A. A competitive learning strategy for adapting fuzzy user profiles. In *2010 10th International Conference on Intelligent Systems Design and Applications*, 2010; pp. 959-964, IEEE.
  26. Alzahrani S. M.; Salim, N. On the use of fuzzy information retrieval for gauging similarity of arabic documents. In *Second International Conference on the Applications of Digital Information and Web Technologies*, 2009; pp. 539-544, IEEE, 2009.
  27. Chawla, S. Effective Personalization of web search based on Fuzzy Information Retrieval. *International Journal of Computer Science and Information Technologies*, 2015b; 6 (3) , pp. 2831-2837
  28. Van Rijsbergen, C. J. A non-classical logic for information retrieval. *The computer journal*, 29(6), 1986; pp. 481-485.
  29. Bordogna, G.; Pasi, G. Modeling vagueness in information retrieval. In *Lectures on information retrieval* , 2000; 207-241, Springer Berlin Heidelberg.
  30. Kraft, D. H.; Bordogna, G.; Pasi, G. Fuzzy set techniques in information retrieval. In *Fuzzy sets in approximate reasoning and information systems*, 1999; pp. 469-510, Springer US.
  31. Miyamoto S. Information retrieval based on fuzzy associations. *Fuzzy Sets and Systems*, 1990; 38(2), pp.191-205.
  32. Zhang, Z.; Nasraoui, O. Mining search engine query logs for query recommendation. In *Proceedings of the 15th international conference on World Wide Web*, 2006; (pp. 1039-1040). ACM.
  33. Wang, X.; Bai, Y.; Li, Y. An information retrieval method based on sequential access patterns. In *Asia-Pacific Conference on Wearable Computing Systems (APWCS)*, 2010; pp. 247-250., IEEE.

34. Eirinaki ,M.; Vazirgiannis, M.;Varlamis, I.SEWeP:using sitesemantics and a taxonomy to enhance the Web personalization process. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003; pp. 99-108.
35. Liu, X.; Zhou, J. Research on Knowledge-based Personalized Recommendation Service System Retrieval Service. *Energy Procedia*, 13, 2011; pp. 10103-10108.
36. Kim, S.; Han, S.The method of inferring trust in web-based social network using fuzzy logic". In international workshop on machine intelligence research, 2009; pp.140-144.
37. Prakash, A.; Mustafi, D. Fuzzy Logic Approach to Combat Web Spam with TrustRank,International Journal of Innovative Research in Computer and Communication Engineering,2013; 1(4), pp. 1100-1106.
38. Kohli, S. Developing and Validating Fuzzy-Based Trust Measures for Online Medical Diagnosis and Symptoms Analysis. *Fuzzy Expert Systems for Disease Diagnosis*, 302,2014.
39. Folorunso O.and Mustapha, O. A. A fuzzy expert system to Trust-Based Access Control in crowdsourcing environments". *Applied Computing and Informatics*, 2015; 11(2), 116-129.
40. Abdul-Rahman A.; Hailes, S. Supporting trust in virtual communities. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000; (pp. 9-pp). IEEE,2000.
41. DHarrison McKnight N.L.C.What trust means in e-commerce customer relationships: an interdisciplinary conceptual typology. *International journal of electronic commerce*, 6(2), 2002; pp. 35-59.
42. Dimitrakos ,T. A service-oriented trust management framework. In Workshop on Deception, Fraud and Trust in Agent Societies , 2003; pp 53-72, Springer Berlin Heidelberg.
43. O'Donovan, J.; Smyth, B. Trust in recommender systems. In Proceedings of the 10th international conference on Intelligent user interfaces ,2005; (pp. 167-174). ACM.
44. Pirolli, P.Computational models of information scent-following in a very large browsable text collection , Conference on Human Factors in Computing Systems, 1997; pp. 3-10.
45. Pirolli, P. 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.
46. Chi, E H.; Pirolli, P.; Chen, K.; Pitkow, J. 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, 2001; pp. 490-497.
47. Heer J., and Chi, E.H.Separating the Swarm: Categorization method for user sessions on the web", International Conference on Human Factor in Computing Systems, 2002; pp. 243-250.
48. Chawla, S.; Bedi, P.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), 2007; pp. 483-488.
49. Chawla, S. ; Bedi ,P. 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'08,2008; pp. 486-491, IEEE.
50. Chawla, S. Personalised Web Search using Trust based Hubs and Authorities. *International Journal of Engineering Research and Applications*, 7, 2014a; pp. 157-170.
51. Chawla, S. Novel Approach to Query Expansion using Genetic Algorithm on Clustered Query Sessions for Effective Personalized Web Search . *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(11),2014b; pp 73-81.
52. Chawla, S. 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*, 2015a; 4(11), 196-204.
53. Wen J. R.; Nie, J. Y; Zhang, H. J. Query clustering using user logs. *ACM Transactions on Information Systems*, 20(1), 2002; pp. 59-81.
54. Zhao, Y.; Karypis, G. Comparison of agglomerative and partitional document clustering algorithms (No. TR-02-014). MINNESOTA UNIV MINNEAPOLIS DEPT OF COMPUTER SCIENCE,2002.
55. Zhao, Y.; Karypis, G. Criterion functions for document clustering: Experiments and analysis,2001.
56. KARN, B. INFORMATION RETRIEVAL SYSTEM USING FUZZY SET THEORY-THE BASIC CONCEPT.
57. Klir, G.;Yuan, B. Fuzzy sets and fuzzy logic , 4, New Jersey: Prentice hall,1995.