Journal of Medical Systems (2019) 43: 243
https://doi.org/10.1007/s10916-019-1350-1

TRANSACTIONAL PROCESSING SYSTEMS

Collaborative Search Engine for Enhancing Personalized User Search Based on Domain Knowledge

Senthilkumar N C[1,2] · Pradeep Reddy Ch[1,2]

Received: 11 March 2019 / Accepted: 20 May 2019 / Published online: 23 June 2019

Abstract
In the fast moving world, users cross over large amount of data for their daily life. Due to the misinterpretation of the context, user cannot retrieve the proper context or failure to retrieve the information. The main aim of this paper is to design and implement a personalized search engine which works based on the domain of the user with the specific constraints suggested by the user. In this paper, the proposed system, build a search engine with web content which get information from the document corpus for the domain through the cloud databases. Web search engine re-ranks the generic results based on a ranking of a context linked with the domain. In this system, collaborative search service helps to improve the relevancy of the search results and to reduce the overtime on bad links and hence caters to customized needs with collaborative feedback using fuzzy decision tree based on fuzzy rules.

Keywords Web search engine · Fuzzy decision tree · Fuzzy rules

Introduction

In this information age, online information resources are abundantly available and widely used. Particularly, in the field of academia, several Terabytes of academic content is uploaded on the Internet every week, and the demand for such resources is always on the rise. Tragically, access to this information using a generic search engine is not satisfactory in terms of the relevance of links and the overtime on bad links. This can be attributed to several factors, the most important being the absence of identification of context of the search and lack of customization based on users' profiles.

This article is part of the Topical Collection on *Transactional Processing Systems*

✉ Senthilkumar N C
ncsenthilkumar@vit.ac.in

Pradeep Reddy Ch
pradeep.ch@vitap.ac.in

[1] School of Information Technology and Engineering, VIT, Vellore, India

[2] Department of Computer Science and Engineering, VIT-AP, Vellore, India

In web search applications, queries are submitted to search engines to represent the information needs of users. Sometimes queries may not exactly represent users' specific information needs since many ambiguous queries may cover a broad topic and different users may want to get information on different aspects when they submit the same query. User search goals are to be analyzed to know their interest and intent based on the information on different aspects of a query that user wants to obtain. Information need is a user's particular desire to obtain information to satisfy his/her need. User search goals can be considered as the clusters of information needs for a query. Advantage of such approach is the provision of relevant information to the user. Moreover, users with different search goals can easily find what they want. Second, user search goals represented by some keywords can be utilized in query recommendation. Thus, the suggested queries can help users to form their queries more precisely. Third, the distributions of user search goals can also be useful in applications such as web security and re-ranking web search results that contain different user search goals.

Collaborative information retrieval environment support teams seeking and retrieving information they need to accomplish work together [1].

The sharing of knowledge within the group is through Relevance Feedback. It acts more like collaborative filtering than like synchronous collaborative searching which is not

dynamic and not time dependant. The aim of this paper is to provide a collaborative web search engine which re-ranks the generic search engine based on the learned preferences of expertise users identified by the engine based on the user's implicit feedback collected from his search behavior being used effectively. The expertise level is upgraded automatically and dynamically based on how many users use and how effectively it influences their search. The expert level is domain specific enabling the same user to be at different levels of expertise at different periods of time enabling the dynamic swapping of roles. Section 2 gives the details of previous related work of the search engine system. The section 3 describes the proposed system and its advantages over the previous systems. Section 4 talks about detailed description of all the modules. Section 5 talks about the algorithms used in each module along with the implementation details of these algorithms and also their end results. Finally, Section 6 gives an overview about the future work to be done in the same area.

Literature survey

Classification of documents make the search engine easier to provide better search results. Some of the existing works concentrates on providing better search results based on feedback mechanism. Feedback mechanisms classified into implicit and explicit based on the features used for get the feedback. And also, some other works used the context representation of the user query and personalized their search based on log maintenance. This section gives the abstract of the existing related works with this works.

Based on [2] searching techniques that most people used are: i) Search Engine based Web search ii) Web Directory based Web search iii) Hyperlink based Web search. But this approaches does not read the actual content of the web page. These techniques are all traditional approaches simply retrieve the web pages if the query word is mentioned in the content without analyzing the semantic similarity or relevant of the context [3].

Information on the Web is dynamic, heterogeneous and rapidly increasing, modified, and moved out. It is also un-structured and not self organized, so surfing needed information is dangling in nature [4, 5]. This work also deals about the modified the unstructured content into structured content which is needed for the retrieved user. User's information needs are not satisfied by a single search process, there is a mismatch between the retrieved documents and the required information. Information encountered at one point in a search may lead to a new unanticipated direction, so the user's required information and consequently their queries continually shift. The user model will translate information required by the user into a query form in the language supported by the system, which conveys some meaning of the system model. Keeping track of user's choice made during their search process allows them to return to temporarily unwanted contents, jump from one category of the content to another and, to retain information and the context across search sessions [6, 7].

Jesus Serrano-Guerrero et.al [8] developed a recommender system for students' study environment, which is able to suggest personalized activities for each student in order to reinforce their competences in a subject. The entire recommendations are computed through the characteristics of competences from each subject and the designed activities modeled by fuzzy linguistic labels. This approach is the core of a recommender system based on fuzzy linguistic modeling which allows students to receive personalized activities to practice competences that have to be reinforced in order to pass a subject. Since this system is not connected with social networking medium it is hard to determine their affinity with other students.

Ming Li et.al [9] proposed an expert recommendation to assist the user to find the required experts. This method adopts the fuzzy linguistic method to construct the expert profile, that is, to model expert's expertise. In addition, the fuzzy text classifier is used to get the relevant degree of the document to each knowledge area when the document is registered, which is the base for user profile construction. Then, the user profile consisting of time and the relevance factor of the rated documents is constructed to derive the overall knowledge level of the user.

Decision tree construction and its representation of data set with decision is shown in [10–12]. These papers will show the importance of decision tree to provide the optimized search results. And also, decision tree incorporated with fuzzification to provide the better classification for uncertain queries. The user model will translate information required by the user into a query form in the language supported by the system, which conveys some meaning of the system model. Keeping track of user's choice made during their search process allows them to return to temporarily unwanted contents, jump from one category of the content to another and, to retain information and the context across search sessions [13]. Selvakumar et al. [19] proposed a clustering algorithm for grouping the users based on their interest. K-Means approach is used to cluster the users by mapping the similarity between the keywords given by them for search.

Based on the literature, the search needs both the implicit and explicit parameters for assess the interest of the users for providing the efficient search engine. At initial stage, the weblogs are used for retrieving the data of browsing history. And then, applying the C4.5 with fuzzy rules for classifying the users into the groups based on their interest.

Proposed system

The system contains two major components as document corpus phase and collaborative search phase. For the Document corpus formation phase, the documents of different domains are processed to extract keywords and their relevance with their

respective documents. A decision tree is constructed using the above mentioned documents preprocessing which has nodes of documents linked to their keywords with links storing the respective relevance values and nodes of terms forming a term dependence tree where the links from one term to another stores a value reflecting the respective terms' dependence on each other.

In the collaborative search phase, when a user enters a search query, the subject of the query is identified after refining the query using stop lists, stemming followed by query expansion by permutations and combinations of the query's keywords and different tenses of the keywords. The query's keywords and subject are used to search the query tree structure to extract similar past query sub-trees which provides collaborative feedback to find the best links by prioritizing preferences of collaborators. The search behavior of the current user is tracked and updated in the query tree.

Document nodes, which represents a particular document and the representation nodes, which represent a particular concept. Typically, this is a term contained within one or more documents. A document tree is constructed using maximum spanning tree. It is a weighted tree with related terms and documents linked by arcs with the strength of the relationship stored in the arcs. The Expected Mutual Information Measure (EMIM) $I(T_j,T_i)$ gives the term dependence between T_i and T_j.

Collaborative search service

This is the search component which uses the learned preferences of collaborators to re-rank the results of a generic document tree ranking of the documents to facilitate collaborative information seeking. A User Queries Tree is used to log user behavior in terms of the documents he has downloaded, scrolled or just clicked for a given query by users of specific levels of expertise in the subject of the query. When a user enters a search string, the User Query Tree is queried with selection criteria based on a maximization function to provide the most appropriate results for the particular user by means of collaborative feedback. The search service is enhanced by including two different ranking services mentioned Collaborative Feedback Ranking and Decision tree re-ranking.

Collaborative feedback ranking

Ranking the search results is provided efficiently by grouping the user's interest. The user logs in and gives a query to search. The query keywords are extracted by applying stop lists, stemming, different forms of the keywords, permutations and combinations of the keywords is all extracted and document tree ranking is done. Now if there are no similar past query subtrees then the term tree is considered to find the most relevant terms to the current query terms and then the User Query Tree is searched for these terms and then the feedback ranking is done. The conjunctor combines both the rankings and gives the final ranking. The user's search behavior is tracked and updated in the User Query Tree.

Using evidence from past queries:

Here the left half represents the current query returning the documents d_{kj} as ranked as mentioned above.

The right half has the set v which has the past similar queries' roots where C_l, l lies between 1 to p, is the past query which is related to the current query C_0 if the set of keywords in the query C_l covers, at least partially, the set of keywords in the query q.

C_0 is the current query.

Q_{cov} is a vector to quantify the relationship of each query with the present query the influence of each query cl on the node qcov is considered separately, one at a time.

$$Q_{cov} = \frac{\text{No of terms common for past and current query}}{(\text{total no of terms in current query})^{0.5} * (\text{total no of terms in current query})^{0.5}}$$

$$Fb_{ranking(dj/ql)} = Q_{cov} * \text{expertlevel} * \text{weightage of the document}$$

Fuzzy decision tree based fuzzy rule generation

Re-ranking is done on the search results taken form collaborative feedback ranking. The user logs in and gives a query to search. The ranked results of collaborative ranking is seeded as input and tree construction is carried out based on the individual user's interest on each group. Classification rules are an important tool for discovering knowledge from personalized dataset. Integrating fuzzy logic algorithms into dataset allows to reduce uncertainty which is connected with data stored in database and to increase discovered knowledge's accuracy. Proposed user model analyze some possible variants of making classification rules from a fuzzy decision tree based on cumulative information. Decision trees, which make use of fuzzy sets and fuzzy logic for solving the introduced uncertainties, are called Fuzzy decision trees (FDTs) [14,15–17].

Based on the parameters mentioned in [18], fuzzy rules are generated to construct the decision tree using C 4.5 algorithm for classification. C4.5 is an extension of ID3 improves computing efficiency, deals with continuous values, handles attributes with missing values, avoids over fitting, and performs other functions.

Table 1 Experimental dataset information

Sl. No	Search Information	Total	Avg Access
1	Number of Users	50	–
2	Number of search sessions	126	2.52
3	Number of search queries	486	9.72
4	Number of Web pages visited	864	17.28

Experiments and result analysis

This model collets 50 distinct user's search data consists 1730 pages which they are accessed during their search process. The number of queries, number of sessions and number of pages visited information's are shown in Table 1 [18]. Also it provides the ratio of the each user is calculated and represented.

The collected user data contains the randomness as well as uncertainty due to drift in their search process. It is tested and measured through frequency test and various statistical measures. A hypothesis test use sample data to test a hypothesis about the population from which the sample was taken. It makes inference about one or more population when sample data are available.

Each page visited by the user consists of set of scrolling speeds recorded by the browser. The average, maximum and minimum scrolling speed on a page visited by the user is computed from the set of scrolling speeds. The click through on a page is calculated whenever there was the change in address box in the browser. Change in the content of the address box occurs whenever the user clicks a link/URL's available in the page that is currently being visited. The Time/Page-Size ratio is derived from the list of pages visited and time spent by the user in the page. A sample pre-processed data is shown in Fig. 1.

The decision tree is constructed based on the classification carried out by C 4.5 by analyzing user's interest is shown in Fig. 2. The main goal of a generating group of classification rules is to determine the value of the class attribute from the attribute's values of new instances. This section explains the mechanism of making fuzzy rules from the Fuzzy Decision Tree and the mechanism of their use for classification. The single user accessed 20 pages related fuzzified values and its equivalent linguistic labels are represented in the Tables 2. Here LMD, MED, HGH labels represents Low-medium, Medium and High respectively. It is converted into 25 rules and more than 100 rule conditions. User's given feedback label based sample training dataset is shown in Table 3.

Using these linguistic variables as mentioned in Table 2, various fuzzy if-then rules are generated. A sample rule is given below: The following sample rule gives cleared representation of the various antecedents and consequent part existing in it, also shows the multiple antecedents connection using the operator ***AND.***

IF Time−Spent = low **AND** Scrolling−Speed

= low **AND** Click−Through = low **AND** Page−Size

= low **AND** Time/Page−Size

= high **THEN** Interest−Label = Not−Interest.

From the above rule in between ***IF*** and ***THEN*** variables such as: Time-Spent, Scrolling-Speed, Click-Through, Page-Size,

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	PAGE_NO	USER_NAME	IP_ADDRESS	CLICK TROUGH	MAX SCROLL	MIN SCROLL	NO.OF PAGES	AVG_SCROLLING	SIZE OF THE PAGE(KB	IME_SPENT(SEC	IME/PAGE SIZ	SAMPLE KEYWORDS	URL_REFERRED
2	1	USER_1	10.6.156.247	1	2384.6154	30.209518	1	508.0923354	143	99.828	0.698097902	Web, Mining,ask, service,browse,minin	http://www.ask.com/web?q=about+web+m
3	2	USER_1	10.6.156.247	1	6000	6.7215958	2	1061.177159	59	85.735	1.453135593	web, mining,relationship,web,mining,w	http://searchcrm.techtarget.com/definitio
4	3	USER_1	10.6.156.247	1	12400	6.7215958	1	946.5634091	150	36.547	0.243646667	Web, Mining,ask, service,browse,minin	http://www.ask.com/web?q=about+web+m
5	4	USER_1	10.6.156.247	1	12400	6.7215958	1	946.5634091	129	11.171	0.086596899	Web, Mining,ask, service,browse,minin	http://www.ask.com/web?q=about+web+m
6	5	USER_1	10.6.156.247	1	11625	17.153924	1	882.6134444	132	10.906	0.082621212	Web, Mining,ask, service,browse,minin	http://www.ask.com/web?q=about+web+m
7	6	USER_1	10.6.156.247	2	12400	6.7215958	2	853.3021823	7	6.657	0.951	Web, Mining,ask, service,browse,minin	http://informatics.indiana.edu/fil/Class/b6
8	7	USER_1	10.6.156.247	2	12400	6.7215958	2	810.5354534	4	69.093	17.27325	web,mining,usage mining, patterns, log	http://informatics.indiana.edu/fil/Class/b6
9	8	USER_1	10.6.156.247	1	12400	6.7215958	2	729.5507003	1675	262	0.15641791	semantic, web mining,www,ontogogies	http://www.imap.websemanticsjournal.or
10	9	USER_1	10.6.156.247	1	12400	6.7215958	1	738.2065203	128	10.031	0.078367188	Web, Mining,ask, service,browse,minin	http://www.ask.com/web?q=about+web+m
11	10	USER_1	10.6.156.247	1	12400	6.7215958	2	713.7313274	9	48.266	5.362888889	contents,books_catalog,Unsupervised,I	http://www.cse.iitb.ac.in/soumen/mining_
12	11	USER_1	10.6.156.247	2	12400	6.7215958	2	645.1727913	1191	106.39	0.089328296	WebContentMining,Web mining,structu	http://www.cs.uic.edu/~liub/WebMiningBo
13	12	USER_1	10.6.156.247	1	2384.6154	16.464548	2	339.7438681	125	66.016	0.528128	web,mining,class,search,Effects,Negati	http://www.ask.com/web?q=web+AND+mi
14	13	USER_1	10.6.156.247	2	2384.6154	16.464548	2	324.2375758	114	49.422	0.433526316	web,mining,class,search,Effects,Negati	http://www.ask.com/web?q=web+AND+mi
15	14	USER_1	10.6.156.247	2	2384.6154	16.464548	1	316.870609	114	29.25	0.256578947	web,mining,class,search,Effects,Negati	http://www.ask.com/web?q=web+AND+mi
16	15	USER_1	10.6.156.247	1	2384.6154	16.464548	1	301.2777468	114	24.094	0.211350877	web,mining,class,search,Effects,Negati	http://www.ask.com/web?q=web+AND+mi
17	16	USER_1	10.6.156.247	2	2384.6154	5.7199148	1	364.7874279	36	111.984	3.110666667	Web,social,network,Lab,text,Mining,Un	http://weblab.com.cityu.edu.hk/blog/
18	17	USER_1	10.6.156.247	1	11250	5.7199148	1	584.0221515	61	76.953	1.26152459	webcontentmining,web,mining,data,kn	http://webcontentmining.com/
19	18	USER_1	10.6.156.247	1	6960	5.7199148	1	576.6906373	47	9.235	0.196489362	browswer,automatically,web,mining,da	http://boston.lti.cs.cmu.edu/classes/11-6
20	19	USER_1	10.6.156.247	1	6960	7.973251	1	557.7648687	5	4.875	0.975	business,intelligence,data,Mining,custo	http://www.cs.cmu.edu/~callan/Teaching/
21	20	USER_1	10.6.156.247	1	1488	24.048096	1	427.607176	121	37.625	0.310950413	web,mining,data,WebMiningBook,know	http://www.ask.com/web?q=web+mining+
22	21	USER_1	10.6.156.247	1	1488	24.048096	1	427.607176	126	1.047	0.008309524	web,mining,data,WebMiningBook,know	http://www.ask.com/web?q=web+mining+
23	22	USER_1	10.6.156.247	1	1488	24.048096	1	507.8493725	111	10.906	0.098252252	web,mining,data,WebMiningBook,know	http://www.ask.com/web?q=web+mining+
24	23	USER_1	10.6.156.247	1	2307.6923	24.048096	1	506.5196757	114	50.828	0.445859649	web,mining,data,WebMiningBook,know	http://www.ask.com/web?q=web+mining+
25	24	USER_1	10.6.156.247	1	2307.6923	24.048096	1	531.6981686	115	205.735	1.789	web,mining,data,WebMiningBook,know	http://www.ask.com/web?q=web+mining+
26	25	USER_1	10.6.156.247	1	2307.6923	8.2381079	1	474.3123509	14	81.625	5.830357143	Mining,Web,stanford,introduction,asso	http://infolab.stanford.edu/~ullman/cs345
27	26	USER_1	10.6.156.247	1	2307.6923	8.2381079	1	459.4567562	115	10.735	0.093347826	web,mining,data,WebMiningBook,know	http://www.ask.com/web?q=web+mining+
28	27	USER_1	10.6.156.247	1	2307.6923	8.2381079	1	456.9732062	150	10.938	0.07292	Database,Systems,sigir2008,medical,St	http://www.informatik.uni-trier.de/~ley/db
29	28	USER_1	10.6.156.247	2	12400	8.2381079	3	1736.645911	5298	180.219	0.034016421	data, mining, classification, supervised,	http://www.dais.unive.it/~dm/New_Slides_
30	29	USER_1	10.6.156.247	2	12400	8.2381079	2	1736.645911	6	90.203	15.03383333	data, mining, extraction, knowledge, pa	http://www.dais.unive.it/~dm/
31	30	USER_1	10.6.156.247	1	2384.6154	12.504202	1	830.8607586	128	96.422	0.753296875	mining,web,applications,Social,Networ	http://www.ask.com/web?q=%22web+min
32	31	USER_1	10.6.156.247	1	11625	12.504202	1	1073.072032	122	30.141	0.247057377	mining,web,applications,Social,Networ	http://www.ask.com/web?q=%22web+min
33	32	USER_1	10.6.156.247	1	11625	12.504202	1	1487.048007	124	19.797	0.159653226	mining,web,applications,Social,Networ	http://www.ask.com/web?q=%22web+min
34	33	USER_1	10.6.156.247	1	11625	2.0189331	3	1207.484593	75	157.641	2.10188	book,Web,Mining,Usage,Analysis,Work	http://www.springer.com/computer/ai/bo
35	34	USER_1	10.6.156.247	1	6000	2.0189331	1	1117.2042	103	18.531	0.179912621	mining,web,applications,Social,Networ	http://www.ask.com/web?q=%22web+min
36	35	USER_1	10.6.156.247	1	11625	2.0189331	1	1119.008258	5	61.75	12.35	pattern, knowledge,numbers,discover,	http://www.articlesbase.com/internet-art
37	36	USER_1	10.6.156.247	1	11625	2.0189331	1	1079.210087	118	30.391	0.257550847	mining,web,applications,Social,Networ	http://www.ask.com/advancedsearch?o=0
38	37	USER_1	10.6.156.247	1	11625	2.0189331	1	1263.953587	118	92.484	0.783762712	mining,web,applications,Social,Networ	http://www.ask.com/web?q=%22web+min
39	38	USER_1	10.6.156.247	1	11625	2.0189331	2	1814.771198	16	204.14	12.75875	ebooks,computer,science,data,mining,	http://onlinelibrary.wiley.com/doi/10.100
40	39	USER_1	10.6.156.247	1	11625	2.0189331	2	1676.54854	16	120.406	7.525375	web,mining,content,structure, usage, p	http://www.dcc.uchile.cl/docs/CC72T-2007

Fig. 1 Snapshot representation of sample user data after the pre-processing

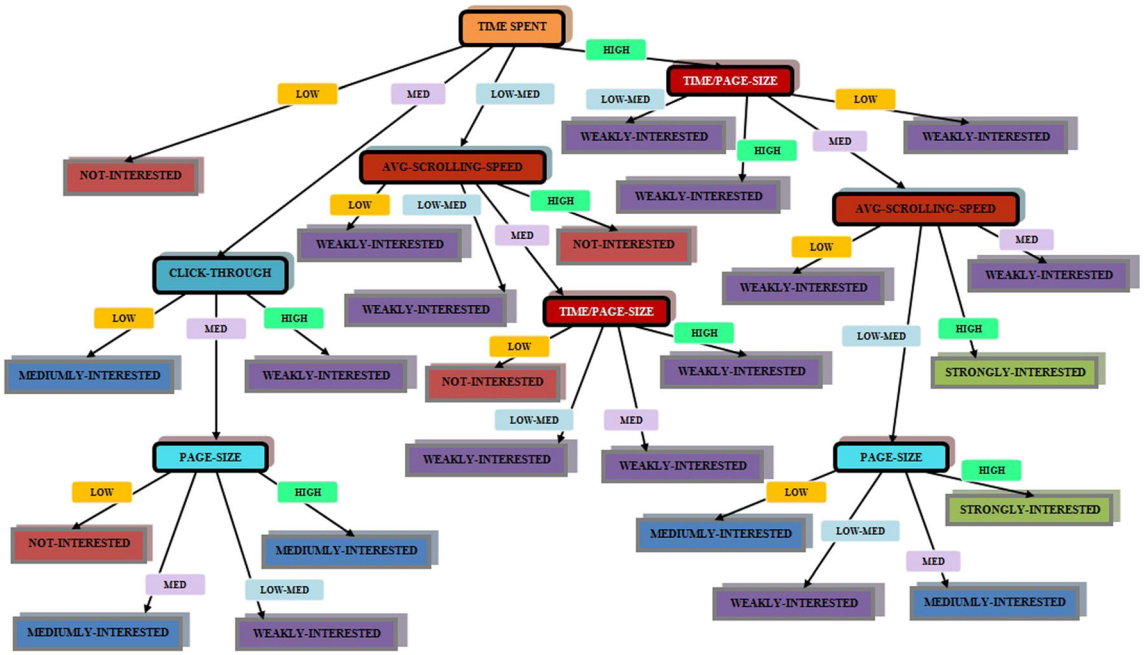

Fig. 2 Pictorial representation of decision tree for user interest classification

Max-Scrolling-Speed, Min-Scrolling-Speed, and Time/Page-Size are called different antecedents are approaching this user model and their corresponding consequent part as "Not-Interest".

In this empirical research work apply the above procedure on the various user attributes and generate 25 rules and set of sample rules are shown in the Table 4. The same rules are applied for 10 different users in order checking its completeness and consistency.

Each user's interest may vary according to their different and depends on several factors. This user model considers the user interest as the decision variable. The attributes that are playing major role in the user interest classification and the notations that are used in this work is normalized and presented in Table 4. Also represent a selected number of fuzzy rules among the huge set of rules which are generated using linguistic variables which are related with various users' parameters. Some of the rules (Rules 1, 2 and 3) can be interpreted in the following form:

Rule 1: ***IF*** *Time-Spent = medium* ***AND*** *Scrolling-Speed = low* ***AND*** *Click-Through = low* ***AND*** *Page-Size = low* ***AND*** *Time/Page-Size = medium* ***THEN*** *Interest = Medium-Interest.*

Rule 2: ***IF*** *Time-Spent = low-medium* ***AND*** *Scrolling-Speed = medium* ***AND*** *Click-Through = medium* ***AND*** *Page-size = medium* ***AND*** *Time/Page-Size = low-medium* ***THEN*** *Interest = Weak-Interest.*

Rule 3: ***IF*** *Time-Spent = high* ***AND*** *Scrolling-Speed = high* ***AND*** *Click-Through = medium* ***AND*** *Page-Size = high* ***AND*** *Time/Page-Size = medium* ***THEN*** *Interest = Weak-Interest.*

From the above set of fuzzy rules any one is said to be fired if any of the preceding parameters (low, low-medium, medium, high) estimate to true (1); otherwise, if all the parameters evaluate to false (0) then it does not fire.

After the estimation of categories, this model approaches fuzzy based classification on the specific query based pages which is referred by the user. Also generate the user grouping based on the class labels which are assigned to each page which are referred (like Not-Interest, Weak-Interest, Medium-Interest, and Strong-Interest). This model process users accessed web pages and perform the both interests based classification as well as the clustering based on their class labels. Also it performed other types of grouping using users' queries and its related pages as well as categorization based grouping.

In order to simplify the grouping process single user accessed 58 pages are processed and grouped according to its interest class labels as well as other groupings. The entire pages have belonged to any one of the class label and it is showing. Each page is represented as separate edge and there is a link between the corresponding classes it belongs to. Similarly all other groups are formed and represented as a graph and shown in Fig. 3.

This model also performs the grouping of user accessed pages and its related queries based on their categories. This approach provides the inference of the users' major interest area in order to provide their needed content in future based on their interest. Here the USER_1 accessed 58 Web pages are grouped and the result is shown in Fig. 4. This user accessed most of the Web pages are belongs to C_1. So it infers that the USER_1 is interested in "data_mining/databases/software/computers" category based contents shown in Fig. 5.

Table 2 User personalized parameters with linguistic form

Page no	Time spent	Avg scroll	Click through	Page size	Max scroll	Min scroll	Time/Page size ratio
001	LMD	LOW	LOW	LOW	LOW	LOW	LMD
002	LMD	LMD	MED	LOW	MED	MED	MED
003	LMD	MED	MED	MED	MED	LOW	LMD
004	LMD	HGH	MED	LMD	MED	LOW	HGH
005	LOW	LOW	LOW	LOW	LOW	LOW	HGH
006	LOW	LOW	LOW	LOW	LOW	LOW	HGH
007	HGH	MED	LOW	HGH	HGH	LOW	LMD
008	MED	LOW	LOW	LOW	LOW	LOW	MED
009	LOW	LOW	LOW	LOW	LOW	LOW	HGH
010	LMD	LOW	LOW	LOW	LOW	LOW	LMD
011	MED	LMD	MED	HGH	MED	MED	LMD
012	HGH	MED	LOW	HGH	HGH	LOW	LMD
013	LMD	LOW	LOW	LOW	LOW	LOW	LMD
014	LMD	LOW	LOW	LOW	LOW	LOW	LMD
015	HGH	LOW	LOW	LOW	LOW	LOW	HGH
016	MED	LMD	MED	MED	MED	MED	LOW
017	LMD	MED	LOW	LOW	MED	MED	MED
018	LMD	MED	LOW	LOW	MED	LOW	LOW
019	LOW	LOW	LOW	LOW	LOW	LOW	HGH
020	LMD	LOW	LOW	LOW	LOW	LOW	LMD

Table 3 Sample user training dataset approached in decision tree generation process

Page no	Time spent	Avg scroll	Click through	Page size	Max scroll	Min scroll	Time/Page size ratio	Feedback class label
1	LMD	LOW	LOW	LOW	LOW	LOW	LMD	WI
2	LMD	LMD	MED	LOW	MED	MED	MED	WI
3	LMD	MED	MED	MED	MED	LOW	LMD	WI
4	LMD	HGH	MED	LMD	MED	LOW	HGH	NI
5	LOW	LOW	LOW	LOW	LOW	LOW	HGH	NI
6	LOW	LOW	LOW	LOW	LOW	LOW	HGH	NI
7	HGH	MED	LOW	HGH	HGH	LOW	LMD	WI
8	MED	LOW	LOW	LOW	LOW	LOW	MED	MI
9	LOW	LOW	LOW	LOW	LOW	LOW	HGH	NI
10	LMD	LOW	LOW	LOW	LOW	LOW	LMD	WI
11	MED	LMD	MED	HGH	MED	MED	LMD	MI
12	HGH	MED	LOW	HGH	HGH	LOW	LMD	MI
13	LMD	LOW	LOW	LOW	LOW	LOW	LMD	WI
14	LMD	LOW	LOW	LOW	LOW	LOW	LMD	WI
15	LMD	LOW	LOW	LOW	LOW	LOW	LMD	WI
16	MED	LMD	MED	MED	MED	MED	LOW	MI
17	LMD	MED	LOW	LOW	MED	MED	MED	WI
18	LMD	MED	LOW	LOW	MED	LOW	LOW	NI
19	LOW	LOW	LOW	LOW	LOW	LOW	HGH	NI
20	LMD	LOW	LOW	LOW	LOW	LOW	LMD	WI

Table 4 Fuzzy based rules for classification of users'

Rule no	Time spent	Avg scroll	Click through	Page size	Time/Page size ratio	Interest
1	Low	Low	Low	Low	High	NOT INTEREST
2	Low Medium	Low	Low	Low	Low Medium	WEAKLY INTEREST
3	Medium	Low	Low	Low	Medium	MEDIUM INTEREST
4	High	Low	Low	Low	High	STRONGLY INTEREST
5	Low Medium	Medium	Medium	Medium	Low Medium	WEAKLY INTEREST
6	Medium	Low Medium	Medium	High	Low Medium	MEDIUM INTEREST
7	High	High	Medium	High	Medium	WEAKLY INTEREST
8	Medium	High	Medium	Low	Medium	NOT INTEREST
9	Medium	Medium	High	Medium	High	WEAKLY INTEREST
10	Low Medium	High	Medium	Low Medium	High	NOT INTEREST
11	High	Low Medium	High	High	Medium	STRONG INTEREST
12	Low Medium	Low Medium	Medium	Low	Medium	WEAKLY INTEREST
13	Medium	Low Medium	Medium	Medium	Low	WEAKLY INTEREST
14	High	Low Medium	High	Medium	Medium	MEDIUM INTEREST
15	Medium	Low	High	High	Low Medium	WEAKLY INTEREST
16	Low Medium	Medium	Low	Low	Low	NOT INTEREST
17	High	Low Medium	Low	Low	Medium	MEDIUM INTEREST
18	Low Medium	Medium	Low	Low	Medium	NOT INTEREST
19	High	Medium	Medium	Medium	Medium	MEDIUM INTEREST
20	High	Medium	Low	High	Low Medium	WEAKLY INTEREST

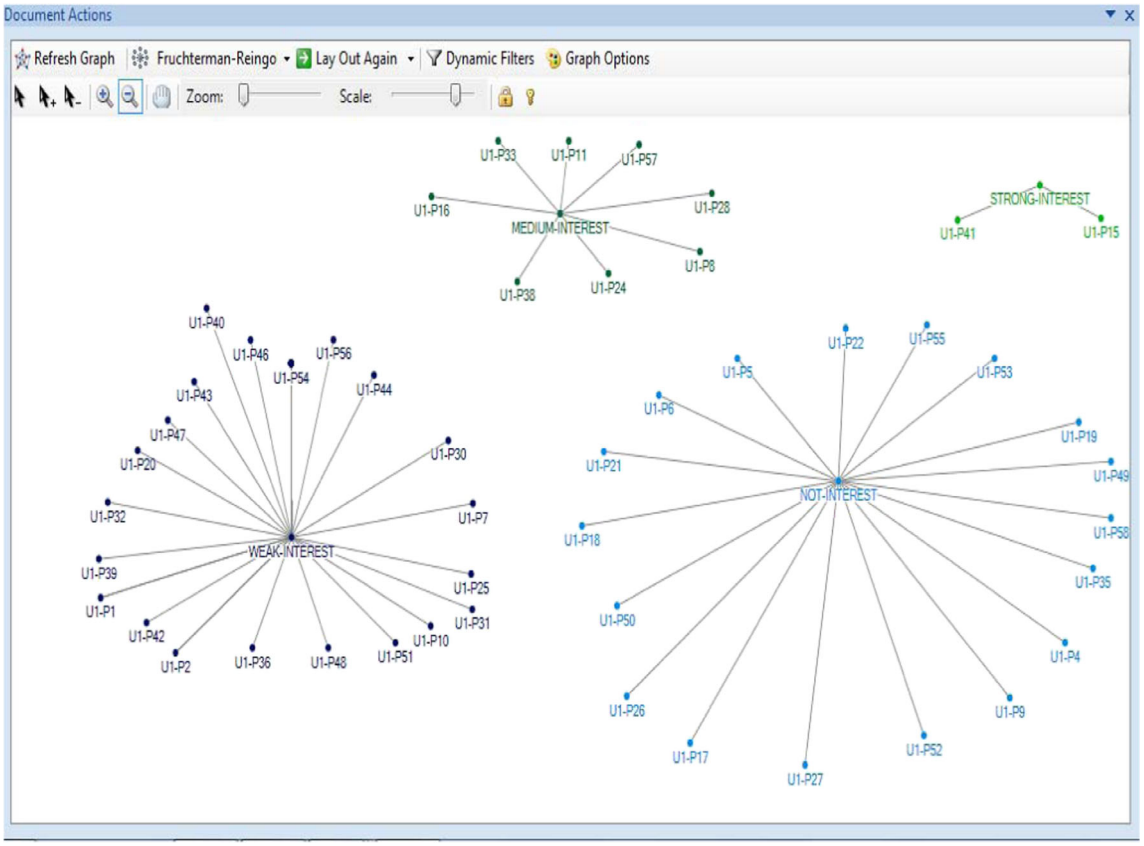

Fig. 3 Pictorial representation of User grouping based on interest class labels

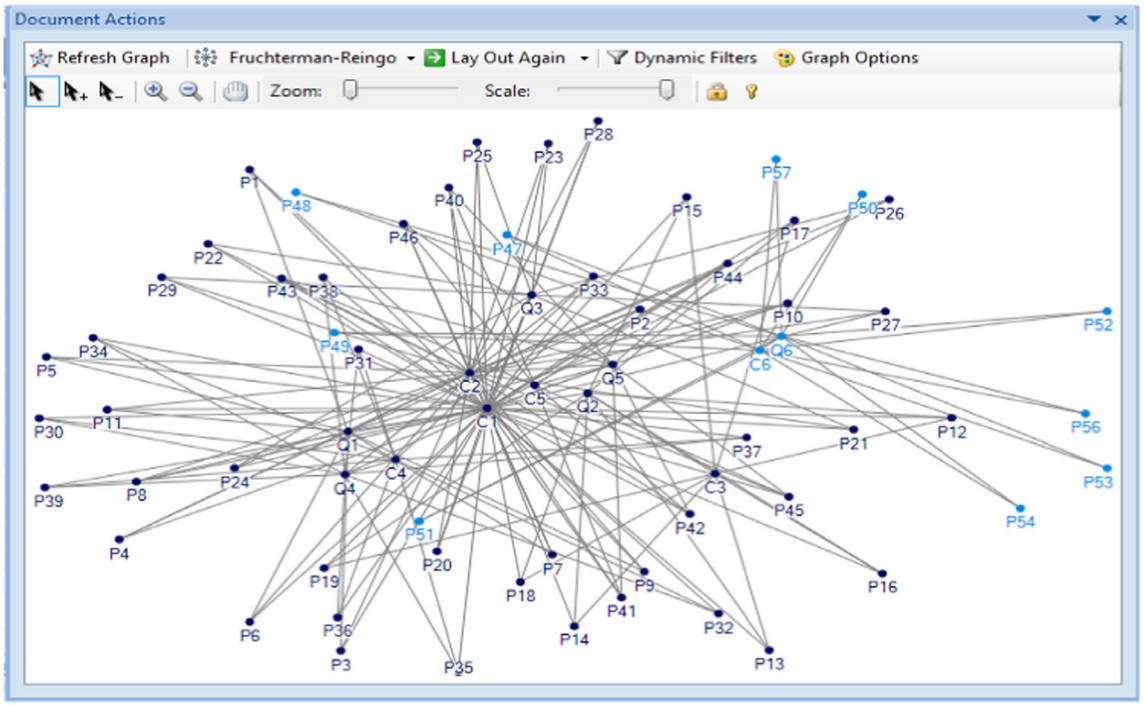

Fig. 4 Category based USER_1 accessed Web pages and its query grouping

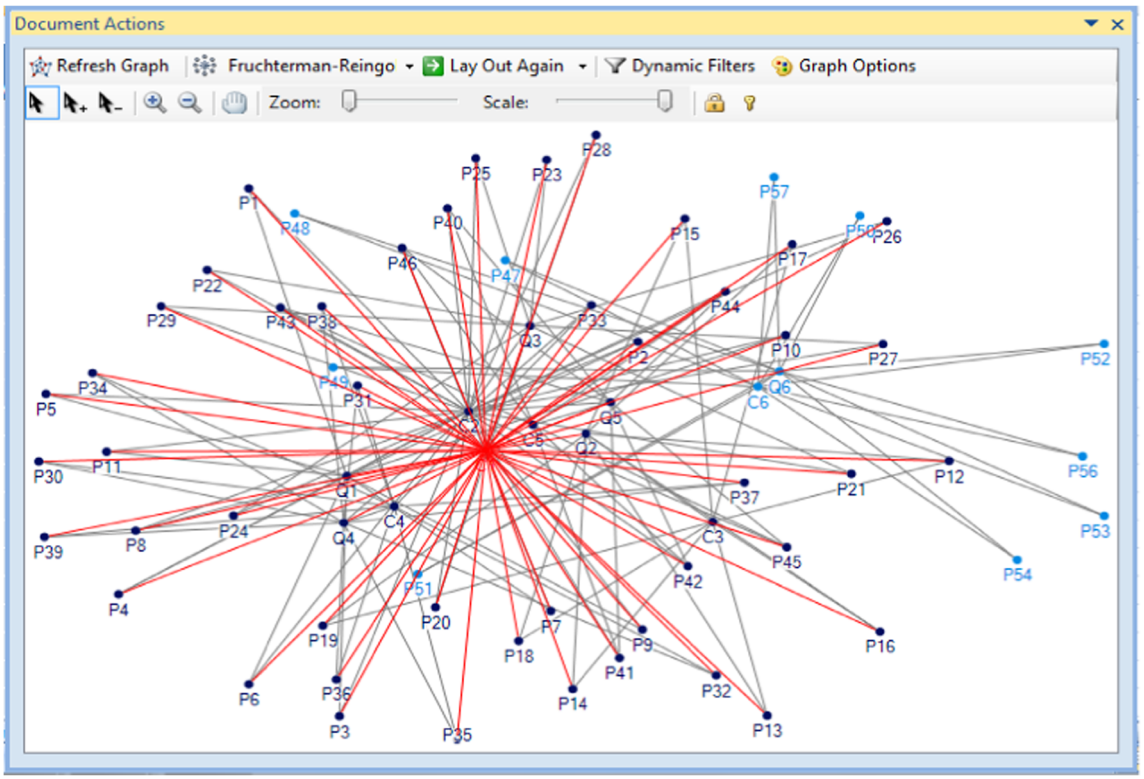

Fig. 5 Category C_1 of USER_1 accessed Web pages and its query grouping

Conclusion

In spite of huge data volumes in web, the user needs an expertise web search engine to satisfy the user requirements in quick response time. In this paper, the proposed system attempts to achieve collaboration by implicitly identifying and reflecting search behavior of collaborators The fuzzification functions are playing a major role for handling uncertainty data in such wage environment. Here the fuzzification is performed based on specific membership function and the selection of a specific membership function is based on the nature of the search data. Based on the class labels, users' irrelevant pages are filtered and recommended Medium and Strong-Interest pages. Finally, users are grouped according to their class labels. It also makes the user to provide a feedback and make the user to promote their levels based on their expertise.

Compliance with Ethical Standards

Conflict of Interest This paper has not communicated anywhere till this moment, now only it is communicated to your esteemed journal for the publication with the knowledge of all co-authors.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Collaborative Information Retrieval Environment., Integration of Information Retrieval with Group Support Systems by Nicholas C. Romano, Dmitri Roussinov,Jay F. Nunamaker, Jr,Hsinshun Chen, IEEE 1999.
2. Herrera, M. R., de Moura, E. S., Cristo, M., Silva, T. P., and da Silva, A. S., Exploring features for the automatic identification of user goals in web search. *J. Inform. Process. Manag.* 46:131–142, 2010.
3. Yu, J., Gong, J., and Liu, F., Generation of semantic interactive environment for personalized search. IEEE International Conference on Computer and Information Technology September. Computer Society Washington, DC, USA. IEEE. 443–448, 2011.
4. Dinh, D., and Tamine, L., Towards a context sensitive approach to searching information based on domain specific knowledge sources. *J. Web Seman.: Sci. Services Agents World Wide Web* 12(13):41–52, 2012.
5. Siva kumar, P., Premchand, D. P., and Govardhan, D. A., Query-based summarizer based on similarity of sentences and word frequency. *Int. J. Data Mining Knowl. Manag. Process* 1:1–12, 2011.
6. Kim, J. Y., and Collins-Thompson, K., Characterizing Web Content, User interests, and search behavior by Reading level and topic. ACM 2012 conference on web search and data mining Feb 8, 2012; ACM new York, NY, USA: ACM. 213–222, 2012.
7. Verma, R., and Pathak, K., Deriving personalized concept and fuzzy based user profile from search engine queries. *Int. J. Sci. Eng. Res.* 4:1–11, 2013.
8. Serrano-Guerrero, J., Romero, F. P., and Olivas, J. A., Hiperion: A fuzzy approach for recommending educational activities based on the acquisition of competences. *J. Inform. Sci.* 248:114–129, 2013.
9. Li, M., Liub, L., and Li, C.-B., An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems. *J. Expert Syst. Applic.* 38: 8586–8596, 2011.
10. Pulkkinen, P., and Koivisto, H., Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms. *J. Approx. Reason.* 48:526–543, 2008.
11. Kazemia, A., and Mehrzadeganb, E., A new algorithm for optimization of fuzzy decision tree in data mining. *J. Optim. Indust. Eng.* 7:29–35, 2011.
12. Adidela, D. R., Jaya, S. G., and Lavanya, D. G., Construction of fuzzy decision tree using expectation maximization algorithm. *J. Comput. Sci. Manag. Res.* 1:416–424, 2012.
13. Ramesh, L. S., Ganapathy, S., Bhuvaneshwari, R., Kulothungan, K., Pandiyaraju, V., and Kannan, A., Prediction of user interests for providing relevant information using relevance feedback and re-ranking. *Int. J. Intell. Inform. Technol. (IJIIT)* 11(4):55–71, 2015.
14. Sulthana, R., and Ramasamy, S., Context based classification of Reviews using association rule mining, fuzzy logics and ontology. *Bull. Electric. Eng. Inform.* 6, no. 3, 2017
15. Dong, M., and Kothari, R., Look-ahead based fuzzy decision tree induction. *IEEE Trans. Fuzzy Syst.* 9:461–468, 2001.
16. Mitra, S., Konwar, K. M., and Pal, S. K., Fuzzy decision tree, linguistic rules and fuzzy knowledge-based network: Generation and evaluation. *IEEE Trans. Syst. Man Cybernet.—Part C: Applic. Rev.* 32:328–339, 2002.
17. Vitaly, L., and Penka, M., Fuzzy decision tree for parallel processing support. *J. Inform. Contrl Manag. Syst* 3:45–52, 2005.
18. Senthilkumar, N.C. and Pradeep Reddy, Ch., (Classification) grouping the users based on their web search using fuzzy logic. International Journal of Pure and Applied Mathematics, vol. 116, no. 21, 2017.
19. Selvakumar, K., Sai Ramesh, L., and Kannan, A., Enhanced K-means clustering algorithm for evolving user groups. Indian Journal of Science and Technology 8, no. 24, 2015.