International Journal of Information and Communication Technology Education
Volume 16 • Issue 4 • October-December 2020

Recommendation-Based Meta-Search Engine for Suggesting Relevant Documents Links

A. Salman Ayaz, BSAR Crescent Institute of Science and Technology, Chennai, India

Jaya A Venkat, BSAR Crescent Institute of Science and Technology, Chennai, India

Zameer Gulzar, BSAR Crescent Institute of Science and Technology, Chennai, India

https://orcid.org/0000-0002-7767-2855

ABSTRACT

The information available online is mostly present in an unstructured form and search engines are indispensable tools especially in higher education organizations for obtaining information from the Internet. Various search engines were developed to help learners to retrieve the information but unfortunately, most of the information retrieved is not relevant. The main objective of this research is to provide relevant document links to the learners using a three-layered meta-search architecture. The first layer retrieves information links from the web based on the learner query, which is then fed to the second layer where filtering and clustering of document links are done based on semantics. The third layer, with the help of a reasoner, categorizes information into relevant and irrelevant information links in the repository. The experimental study was conducted on a training data set using web queries related to the domain of sports, entertainment, and academics. The results indicate that the proposed meta-search engine performs well as compared to another stand-alone search engine with better recall.

KEYWORDS

Clustering, Domain, Meta Search Engine, Ranking, Recommendation

1. INTRODUCTION

Content on the web is billowing and growing at a faster rate. Search engines are productive tools to search the relevant information from the web. Due to the rapid growth of information over the internet, the task for finding the relevant information has become very difficult for every individual search engine. The success or failure of a search engine is unswervingly reliant upon the user's satisfaction. The search engine users expect the information to be rendered to them in a small period of time. Users also expect that the results must be relevant and appropriate (Satya Sai & Raghavan, 2001). Most of the time the results returned by the search engines cannot entirely satisfy the requirement of the user and the search results are not very accurate and appropriate (Li et al., 2001).

DOI: 10.4018/IJICTE.2020100106

Recommending relevant information is measured as an imminent factor in searching process nowadays, because current recommendation process is not based on user interest and is still following the predefined and static patterns of retrieved information in-spite of the fact that the resultant information needs to be filtered to meet user's interest and objectives (Gulzar et al., 2019).The lack of any explicit arrangement and a wide variety of data available on the World Wide Web creates a challenge for its users to find the concerned data without extra efforts or without any outer help. It was believed that an individual general-purpose search engine lacks processing capability to cope-up with the amount of information being loaded on web nowadays (Sugiura & Etzioni, 2000; Manning et al., 2008).

There is a Meta-search engine concept which is gaining popularity among users and is built on top of other search engines. The user query in Meta-search engine is run across different components simultaneously, the result generated is ranked and best one is provided to the user (Meng, Yu & Liu, 2002). The primary aim of Meta-search engine is to overcome the inherited differences of individual search engines, and thus provide the finest result from the best search engines. Meta-search engine filters the top N results from individual search engine result and that's why it is able to provide the most inclusive result set which is available on WWW. The traditional search engines crawl the web to retrieve the information, but on the other hand Meta-search do not crawl to provide the search result to the user. The Meta-search send the user query to dissimilar individual search engines at a time and only the top N filtered resultant documents are then visualized by the user in a window.

Meta-search engine too poses a few distinctive challenges in terms of the information which is not similar it gathers from individual engines. The outcome of the search consists of document ranking by individual engines which are also accompanied by a document title, a Snippet and a URL (Fabrizio, 2009; Rashid, 2008). But there are prominent advantages of Meta-search engines against individual engines by increasing the searching coverage on the web to provide high recall. It also increases the information retrieval effectiveness by increasing the precision and solves the problem of scalability of web-search. Search engines help to retrieve the relevant document links, but it is necessary to analyze the validity of the document, website and of the links. The existing search engines fail to apply the cognitive reasoning on the links and the semantics of the text. As a part of Meta-search engine, intelligent algorithms are required to select the relevant links. This research work applies the hierarchical clustering and fuzzy reasoner to club the links and rank the accurate results to cater the need of user requirement. This research work adapts the layered framework for retrieving the links from the web.

The other part of the paper is structured as follows: The following section gives a brief about the related work regarding the recommendation, search engine and meta-search engine. Section 3 explains the framework of the Meta Search engine while as Section 4 shows the parameter calculation followed by section 5 for reasoning and experimental results and conclusion is provided in Section 6.

2. RELATED WORK

This study aims to develop an approach of ranking and personalizing the user information as per interest and the preference of the user. Therefore, a new document recommendation approach based on the intelligent searching concept is presented in the form of a Meta-search engine. This segment will provide an interesting overview of the related work regarding meta-search and recommendation concept. A trained Meta-search was proposed by (Guang-ming & Wen-juan, 2010) based on a neural network (NN) CC4 algorithm for calculating web-page relation obtaining the highest degree of proficient web-pages. The authors provide a basic solution to a big problem related to the users trying to access information in an information pool. The NN algorithm reviews the information related to Web-pages which are then ranked and collected in a trained dictionary in a sorted manner.

Similarly, Raval proposed an engine that was basically powered by the management and mechanization of Google functions to achieve more combination and accurate search results (Vishwas,

& Kumar, 2011). Another Meta-search engine by that works on the principle (Brijeshkumar et al., 2011) of priority and user profiles. The system has a parallel crawling approach which helps to improve the performance of the searching by feeding multiple engines instantly and also reduce the redundancy of the overlapped downloads. The Leonidas et al. have explained the proposed system using a quad-rank mechanism where additional information related to query tokens, data connection and overall result of a meta-search engine is considered. The authors have proved that the proposed technique outshine individual search engine by testing the system for its efficacy and productivity in a real-world scenario at the TREC-2009 Congress (Leonidas et al., 2011).

Moreover, Adeniyiet developed a recommender system that uses web data mining by checking the user's behaviour in the form of data stream clicked by them on RSS web-reader. While surfing the internet the user will get the specific information without even enquiring for it. In this system, they used K- nearest neighbourhood classification algorithm to identify the clicked data stream concerning particular users (Adeniyi, Wei, & Yongquan, 2015). Another hybrid framework for recommender system was developed where user-centric information is being recommended. The basic problem which users are facing which searching information is to choose among the variety of choices which contain both relevant and irrelevant information a right piece of information in their interest. In such situations, if an intelligent searching process is applied that would help in recommending relevant information is of great use as it will help users to make certain decisions to meet their objectives. They integrate different searching techniques to develop a hybrid framework for suggesting relevant information (Zameer & Leema, 2018).

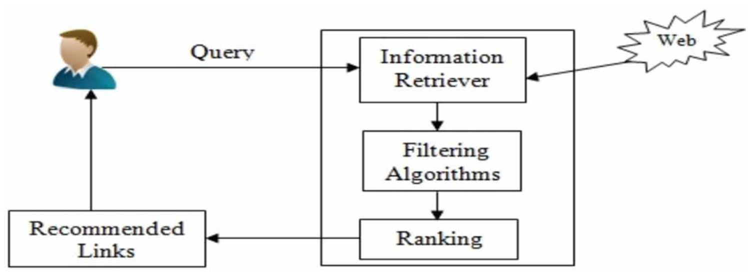
Furthermore, Mishra et al. explained the system which considers sequential information that is present in the web navigating patterns along with the information of the content present in it. The user approval is generated by the SVD algorithm and upper approximation (Rajhans et al., 2015). Zameer developed a query classification technique to recommend relevant information to the users because any recommendation or search engine at core employs information retrieval methods to retrieve and deliver the most useful information. They are classifying the user query in order to obtain the best possible relevant information. The system is able to recognize the user intention in the query itself and based on the applied techniques it generates most relevant recommendations (Zameer & Leema, 2018). Recommendation system based on graphs was suggested which utilize users profile and the positive items ranked in it to generate all undirected, connected and highly connected graphs, with nodes and edges represented by items and correlations respectively. The system finds the recommendation using the entropy and graph connectivity which are new as well as suitable for the users (Kibeom & Kyogu, 2015).

Researchers are advocating the Meta-Search engine capability for providing the qualitative results for different form of domains and suggested a Meta-search engine which determines query relevancy concerned with a web-page and then cluster the retrieved information accordingly. The proposed system is believed to reduce the efforts of users by improving the result quality and performance efficacy to overcome the problem of information overload concerning URL links (Naresh, 2017). Another work where neural networks are used to merge the top computational score effectively and from each individual search engine the top N listed information links for the next stage of processing. They did the ranking based on snippets, titles, position and co-occurrence-based calculation and prove that the ranking methods developed till date were inefficient since only few parameter calculations were done for ranking for each individual engine which later effects precision and recall measures. The response timing was also minimized per 100 links (Vijaya et al., 2016). Likewise, another approach was introduced for reducing the communication and calculation based load from page-ranking algorithm (PRA) and for that purpose they try to gather web-pages regularly and group them inherent scarcity in web. A collective page-rank value was calculated for every group which needs to be shared among the members of the group to provide a scheme of shared update page-ranking along with their merge properties. They kept ranking of errors in a small range to exhibit the intensity of computational reduction (Hideaki et al., 2012).

3. FRAMEWORK FOR META SEARCH ENGINE

Figure 1 describes the general architecture of the recommendation process for Meta-search engine which gives the suitable and authenticated links to the user. The diagram depicts the three-layer architecture developed for the proposed Meta-search engine. The three layers of the generic framework of Meta-search engine consists of Information retriever which is responsible for retrieving links from the various search engines for a given query. The second layer filtered and clustering the retrieved links and also remove redundancy. The third layer reasoning applied fuzzy logic to provide the available link which helps to figure out the relevant links to the user.

Figure 1. General architecture of meta search engine



The architecture of the proposed Meta-search engine as shown in Figure 2. Consists of several components which are explained as follows.

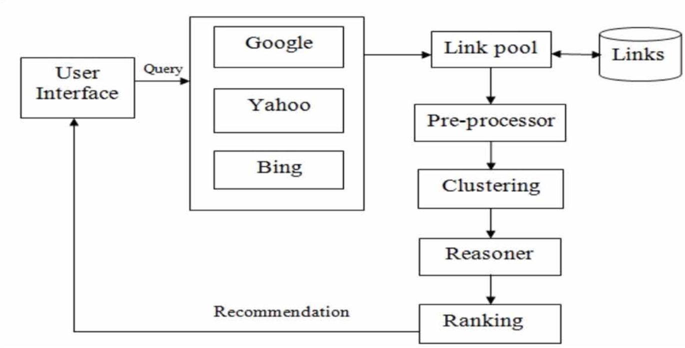
3.1. User Interface

The basic component of the proposed system is the user interface where a user provides an input query which is then fed to each individual engine shown in Figure 2. The link pool component in the proposed system will receive the top N recommended documents from all individual engines which are then rearranged. The document redundancy or duplicate documents are removed on the basis of URLs. The retrieved documents URLs are collected and stored in the database which is later collected and is given for further processing.

3.2. Link Pool

The information dataset which is randomly generated based on the domain of queries given by the users are retrieved in link pool and stored in a database for future use. The individual engines which are used to search the

Figure 2. Proposed meta-search engine framework



information related to a user query are Yahoo, Google, and Bing. In this study, three different domain types are used where user queries are revolved around such as Academic, Sports and Entertainment. As soon as the data set is ready in the link pool, it is then sent to the next phase, pre-processing to filter the retrieved information.

3.3. Pre-Processor

The pre-processing is applied to the whole document to identify the words and phrases of every document which is used for further processing. In pre-processing, the links are extracted from the web pages first and store it as an individual text document. This individual text document is given to the input for the next step. The steps involved in pre-processing are explained below:

Stop word removal → stemming → Top word selection → common words → unique words

User query does not contain all meaningful words, therefore, most of the words don't have semantics and contain irrelevant information which is removed to speed up the search result and also to save memory space. Stemming is a normalization process, which removes the inflexional endings from words or reduces alternative form of words to a general form by restricting the vocabulary space. Stemming can relate words which have different forms based on the same stem or base such as connections, connecting, and connective to common word connect, known as stem or root word. The stemming process improves the effectiveness of the system and reduces the indexing size. Top words (technical word) are selected from the document. From the top n words, common words are separated that is mostly used in the document and then words are compared with common words. From these, only the unique words are selected. The repeated words are removed from the list and unique words are separated.

3.4. Filtering and Clustering

In this layer, the results are filtered and the selected unique words are clustered from the document. In this study, the hierarchical clustering is to combine or cluster the relevant links. The data during the process of clustering is segregated sequentially using a particular distance measure. During the chronological procedure of separation, the algorithm creates a nested partition using a keyword group into an entire cluster tree according to the distance measure without being familiar or recognizing the number of clusters well-ahead. One of the types of clustering known as agglomerative clustering, which uses a bottom-up approach is being applied to combine the links and to form a cluster. The algorithm for the clustering procedure is as follows:

Algorithm
Input: No of nodes with keywords of each node
Output: Cluster hierarchy
Initialization: Number of node (U_i, where i=1,2,..k)
keywords
Disjoint clustering level L (0) = 0
Sequence number n = 0.
Start
Step 1: Assess all pair-wise distances among nodes
Add each node as its own cluster
Step 2: Erect a distance matrix by means of the distance values
Dc [(U_1), (U_2)]
Step 3: Pair the node with the shortest distance
U = ($U_1 \cup U_2$) // Merge the node
Step 4: increase n=n+1,
Set the level to L (n) = dc [(U_1), (U_2)]
Step 5: Revise the distance matrix
Distance between the new clusters
Dc [(U_3), (U)] = min (Dc [(U_3), (U_1)], Dc [(U_3), (U_2)])
Step 6: Replicate till the distance matrix is decreased to a single element
Stop

4. PARAMETER CALCULATION

There are three different parameters used in this study on whose basis the data set is being generated and later clustering and ranking of the unique URL links are prepared. The parameters calculated are title-based calculation, bold based calculation and domain-based calculation.

4.1. Title Based Calculation

In this technique, the calculation is done on the basis of the unique title enclosed in the document link. The title of the document is compared to check its frequency with the user query word tokens and the synonyms generated for those query tokens through WordNet which is shown in Table 1.

In Table 1 W1, W2, W3…., Wa is representing input query token words and 'a' is the entire count of the query token words. The S1 is the first word synonym of the word W1 (W1S1), and WaSb represent the b[th] synonym S of the a[th] token W for a given query, while as C1W1 represents the frequency of the first synonym word present in the unique title link. For each unique link, the title-based calculation is given in Equation 1:

$$T_s\left(p\right)=\sum_{i=1}^{a}\left(\frac{T_sW_i-\max\left(TW_i\right)+1}{\max\left(TW_i\right)}\times w_Q+\sum_{j=1}^{b}\frac{T_sW_iS_j-\max\left(TW_iS_j\right)+1}{\max\left(TW_iS_j\right)}\right) \tag{1}$$

Table 1. Frequency generation for query words and synonyms

	W1	W1S1	W2	W2S1	.	.	.	Wa	WaSb
T1	C1W1	C1W1S1	C1W2	C1W2S1	.	.	.	C1Wa	C1WaSb
T2	C2W1	C2W1S1	C2W2	C2W2S1	.	.	.	C2Wa	C2WaSb
T3	C3W1	C3W1S1	C3W2	C3W2S1	.	.	.	C3Wa	C3WaSb
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
Ts	CsW1	CsW1S1	CsW2	CsW2S1	.	.	.	CsWa	CsWaSb

In Equation 1, *Ts (p)* is the S^{th} calculated value of title based unique links and T_sW_i is the frequency of the i^{th} token word in the document title T *of the* S^{th} unique link. The *max* (TW_i) *is* the frequency *of the* i^{th} token word in whole unique document links titles and $T_sW_iM_j$ is the frequency of j^{th} synonym of i^{th} query token word in whole unique document links title *T* of S^{th} link. The *max* (TW_iM_j) is the maximum frequency of the j^{th} synonym of the i^{th} *token word* in whole unique document links titles whereas '*a*' is the total count of the query token words and '*b'* is total count of the synonym words of the i^{th} query token and W_Q is the weight-age of the query token and W_S is the weight-age of the synonym word.

4.2. Bold Based Calculation

In this technique the words which are bold in the retrieved URL link are being compared with the user query tokens and their synonyms generated from WordNet, and also to check their frequency concerning the words which are bold in each link. Table 2 displayed the frequency of the occurrence of query tokens and synonym tokens with respect to each bold link word.

The B1, B2, …, Bs represents the unique contents in the links and the frequency of a^{th} query token word in content of s^{th} unique link is represented by BsWa. The BsWaSb is the frequency of the b^{th} synonym of a^{th} query token word in the content of s^{th} unique link. The computation of bold based approach is given in Equation 2:

$$B_s\left(p\right)=\sum_{i=1}^{a}\left(\frac{B_sW_i}{\max\left(BW_i\right)}\times w_Q+\sum_{j=1}^{b}\frac{B_sW_iS_j}{\max\left(BW_iS_j\right)}\times w_s\right) \tag{2}$$

Table 2. Frequency of query tokens and synonyms

	W1	W1S1	W2	W2S1	.	.	.	Wa	WaSb
B1	B1W1	B1W1S1	B1W2	B1W2S1	.	.	.	B1Wa	B1WaSb
B2	B2W1	B2W1S1	B2W2	B2W2S1	.	.	.	B2Wa	B2WaSb
B3	B3W1	B3W1S1	B3W2	B3W2S1	.	.	.	B3Wa	B3WaSb
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
Bs	BsW1	BsW1S1	BsW2	BsW2S1	.	.	.	BsWa	BsWaSb

In Equation 2, *Bs* (p) represents the bold word based calculated value of s^{th} unique link and B_sW_i is the frequency of the i^{th} query token word *W* in bold word of s^{th} unique link, while as max (BW_i) is the highest frequency of i^{th} query token word *W* in content of s^{th} unique link. The $B_s\ W_i\ S_j$ is the frequency of j^{th} synonym of i^{th} query token word *W* in the bold word of s^{th} unique link and max (BW_iS_j) is the highest frequency of j^{th} synonym of i^{th} query token word *W* in the bold word of s^{th} unique link.

4.3. Domain Based Calculation

The links retrieved from all individual search engines would fall under a particular domain whose value is calculated for every distinctive URL using that particular domain name for each individual search engine. The formula for calculating the domain value for each unique link is given in Equation 3:

$$U_s\left(p\right)=\log_{10}\left(\frac{2m-1+acc_s}{2m}\right) \tag{3}$$

U_s (*p*) is the value calculated for S^{th} link of a domain whereas acc_s represents the number of links which are unique in a particular domain and *m* refer to the number of individual engines used e.g. if the number of links which are unique is 15 and the links which belong to a particular domain is 8 then the value of acc_s for that link is five.

5. REASONER

Reasoner is responsible for ranking and merging the top URL links generated from different individual search engines on the basis of their sensitivity. The result generated along with their values is used to train the fuzzy logic system by feeding it with the ranked URL links. In recent years fuzzy logic-based interface system emerged as primary classifiers that were used among different fields for a rule-based system because the rule-based systems are used in different applications areas such as medical, entertainment and even error detection systems. In proposing fuzzy interface system, the significant factor is to select the correct rule system but most of the researches have taken this problem in general but in most of the cases there are certain computational methods which are being employed for generating the rules from the regular data.

5.1. Experimental Results

The experiment was conducted using a dataset of user query sessions gathered over the web which was generated in proposed system by capturing the links visited by the users during result retrieved by Meta-search engine. For the generation of dataset the query is input through an interface where the query is passed on to the stack of individual search engine and back again the results generated are displayed on the same interface. In Figure 3 the interface of the proposed system and result for a user query is displayed.

Figure 3. Meta-search engine result for the academic domain

5.2. Performance Analysis

The primary aim of proposed Meta-search engine is to filter the links and rank them using the hierarchical cluster method and fuzzy logic system. The domains were selectively chosen in order to cover a wide variety of queries on internet and the number of randomly chosen test queries for each of the three domains is twenty-five. The queries were fed through the Meta-search engine to all individual search engines and then a group or set of the documents are identifies which are relevant, from all the retrieved documents of each individual search engine. The number of the relevant document retrieved was counter for relevancy by the domain expert and the concerned performance measure values are given in Table 3.

In Table 3 domain-based comparison is provided in the form of average precision and recall for existing and the proposed techniques for 25 test queries are shown in Figure 4. The existing techniques were K-mean clustering and Fuzzy logic based and the proposed technique is based on hierarchical clustering and fuzzy logic. The precision value and the number of queries is represented by y-axis and x-axis respectively. The average precision of all domains in the proposed system is maximum for academic and sports 0.85%, while entertainment 0.75% as compared to the existing systems average precision for academic 0.80%, sports 0.82% and least for entertainment 0.73%. It is also evident from

Table 3. Calculation of precision and recall values

Query Domain	Existing System Techniques		Proposed System Techniques	
	K-Means +FL Based URL Ranking		HC+FL Base URLs Ranking	
	Precision	Recall	Precision	Recall
Academic	0.80	0.60	0.85	0.62
Entertainment	0.73	0.75	0.75	0.80
Sports	0.82	0.80	0.85	0.90

the Figure 4 and Figure 5 graph that the precision and recall results achieved by the proposed system is much better due to more class labels in hierarchical systems comparing to k-mean cluster technique.

Figure 6 to Figure 8 displayed the performance analysis of developed meta-search engine for different domains on the basis of user recommended query. From the graph, it is clearly understood that as the user recommendation result is increased there is a decrease in precision and recall value. The x-axis represents user recommended result and y-axis represents the value of precision and recall.

Figure 4. Precision measure of the proposed methodology

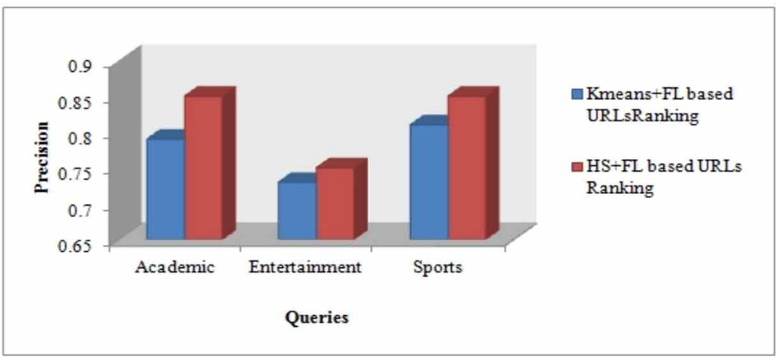


Figure 5. Recall measure of the proposed methodology

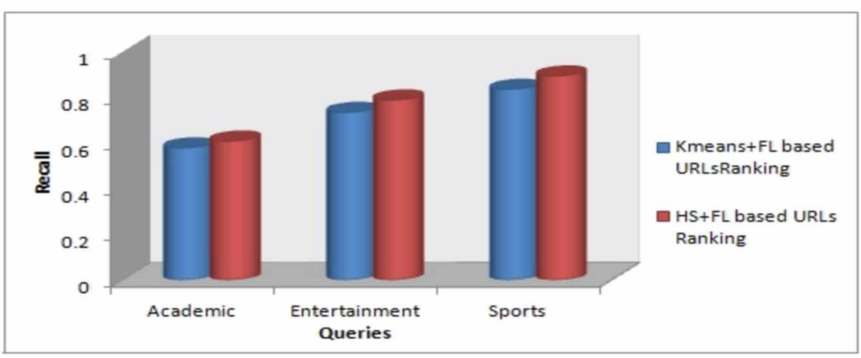


Figure 6. Performance analysis of proposed methodology for academic domain

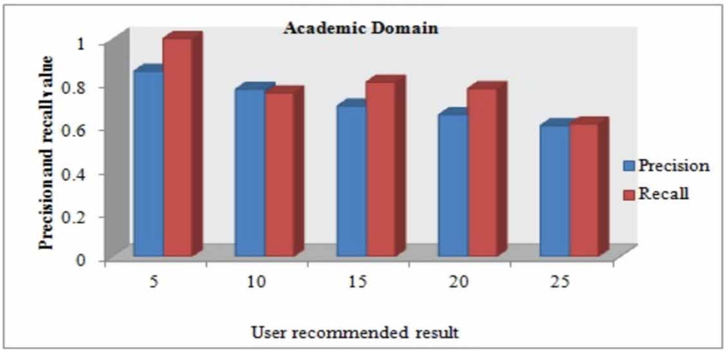


Figure 7. Performance analysis of proposed methodology for entertainment

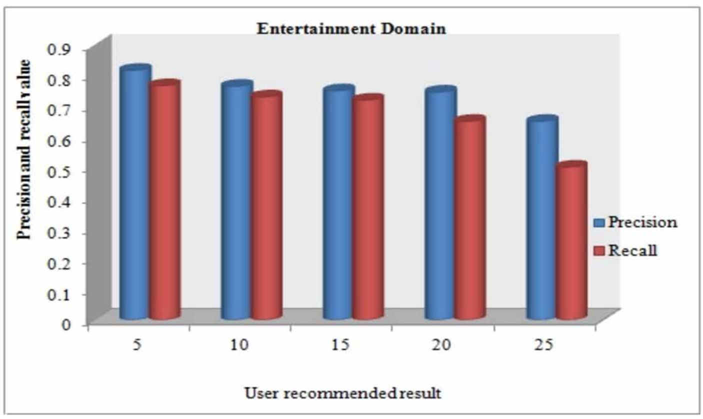
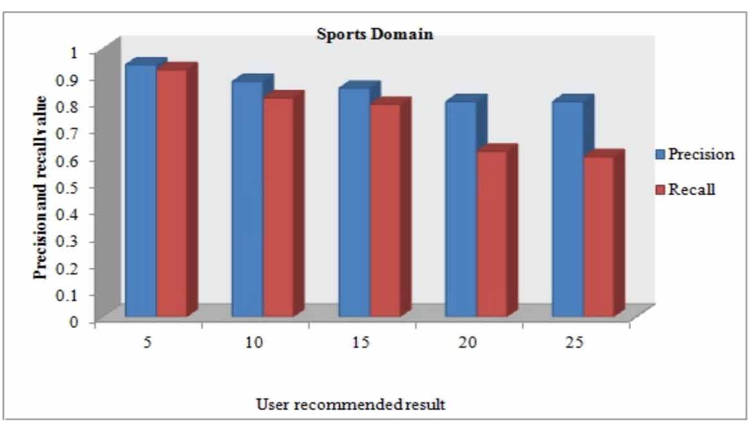


Figure 8. Performance analysis of proposed methodology for sports domain



The proposed research work will also rank the recommended URLs obtained from various search engines. The analysis of three famous search engines Bing, Yahoo, and Google which is based on precision analysis and used in proposed meta-search approach is shown in Table 4. The resulting analysis suggests that the proposed search engine outperforms the individual search engines in terms of better precision score.

Table 4. Average precision for 25 queries in different domain

Domain	Meta-Search	Google	Yahoo	Bing
Sports	0.84	0.75	0.68	0.72
Academic	0.83	0.78	0.70	0.73
Entertainment	0.85	0.77	0.69	0.72

The analysis of the result suggests that the proposed search engine outperforms the individual search engines in terms of better precision score. To check the relevant document links retrieved (First ten document links) with the individual and meta-search engines, a query "data mining" was input to all the engines. The results are displayed in Table 5, which shows that among the three engines the poor performance is shown by Yahoo search engine whose relevant document retrieval score are only 5 among the top 10 documents. On the other hand, Bing and Google were a bit more accurate than Yahoo with 6 relevant links each while as the proposed Meta-search engine tops the ranking list with a qualitative score of 8 relevant document links. The alphabet "R" denotes relevant document link for a particular engine and the symbol "-" denotes an irrelevant link. The relevant document retrieval rate was initially set to 10 and the proposed methodology retrieved 8 relevant documents, while as other individual search engines obtain only maximum of 6 documents links.

Table 5. Top ten relevant document list for a user query

Engine	1	2	3	4	5	6	7	8	9	10	R
Meta-search	R	R	R	R	R	R	-	R	R	-	8
Google	R	R	R	R	R	R	-	-	-	-	6
Bing	R	-	R	-	R	R	-	R	R	-	6
Yahoo	R	-	-	R	-	R	-	R	R	-	5

Table 6 shows the retrieval and ranking performance of different domain using different search engine. The table contains five queries related to sports domain and corresponding retrieval result with ranking of search engine. Based on the retrieval count we rank the search engines. In our proposed work, we utilize the Meta-search engine which is the hybridization of individual engines like Yahoo Google, and Bing. The search engine retrieves the result based on user query and recommended the URLs count.

For example, we have used the user query "Pro kabaddi" and the recommended users URL is "10", so, while using the proposed Meta-search engine the relevant URLs percentage obtained is 50% for Google, 30% for Yahoo and 20% for Bing. Finally, based on the retrieved URLs count we rank the search engine. From the table we understand, among the three search engines "Google" is

Table 6. Retrieval and ranking performance of selected domains

Domain	Keywords	User Recommended Result	Meta-Search			Rank (%)		
			Google	Yahoo	Bing	Google	Yahoo	Bing
sports	Pro kabaddi	10	5	3	2	1 (50)	2 (20)	3(30)
	IPL match	15	7	3	5	1 (67)	3 (20)	2(13)
	Top Athletes	20	9	5	6	1 (45)	3 (25)	2(30)
	Rio Olympics	10	4	4	2	2 (40)	1(40)	3(20)
	Awards of Sachin	20	8	5	7	1 (40)	3(25)	2(35)
Academic	Examination	15	6	5	4	1 (40)	2 (34)	3(26)
	Results	5	3	1	1	1 (60)	3(20)	2(20)
	Syllabus	20	8	7	5	1(40)	2(35)	3(25)
	Universities details	10	4	2	4	1(40)	3(20)	2(40)
	Out Campus interview	15	6	4	5	1(40)	3(26)	2(34)
Entertainment	Latest movies	20	9	5	6	1(45)	3(25)	2(30)
	Korean drama	10	4	4	2	1(40)	2(40)	3(20)
	Pop songs	10	5	3	2	1 (50)	3(30)	2(20)
	Games	15	7	3	5	1(67)	3(20)	2(13)
	Dup mash	20	8	5	7	1(40)	3(35)	2(25)

better other two. Each query we obtain the relevant URLs for different such engines such as Bing, Yahoo and Google. The search engines were ranked based on the retrieved relevant URLs count.

6. CONCLUSION

The meta-search engine was proposed to recommend the relevant information and personalize search engine result related to user's interest. The hierarchical clustering and fuzzy logic based ranking process were developed, unlike individual search engines that use page ranking system. The proposed search engine utilizes a user dependent methodology to rank the resources of search engines and it was also tested against Yahoo, Bing, and other individual engines for availability, consistency, and relevancy. Result comparison indicates that the search results ranking by using the enhanced fuzzy concept networks in the proposed system are showing more relevancy as per the user's interests rather than ranking obtaining through the general concept of networks in individual engines. The main aim of this paper is to develop a novel ranking system such as URL ranking and search engine ranking based on the hierarchical cluster to optimize the performance of the system and for relevant web page recommendation.

REFERENCES

Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Applied Computing and Informatics*, *12*(1), 90–108.

Brijeshkumar, C., Shashi Kant, G., & Soni, R. (2011). Meta search engine based on prioritizor. In Proceedings of the International conference on computational intelligence and communication systems. Academic Press.

Fabrizio, L., Andrea, S., & Claudio, D. (2009). A Relation-Based Page Rank Algorithm for Semantic Web Search Engines. *IEEE Transactions on Knowledge and Data Engineering*, *21*(1).

Guang-ming, L., & Wen-juan, G. (2010). Research and Design of Meta-search Engine Oriented Specialty. In *Proceedings of the International Symposium on Intelligence Information Processing and Trusted Computing*. Academic Press.

Gulzar, Z., Arun, R. L., & Leema, A. A. (2019). Ontology supported hybrid recommender system with threshold based nearest neighborhood approach. *International Journal of Information and Communication Technology Education*, *15*(2), 85–17. doi:10.4018/IJICTE.2019040106

Hideaki, I., Roberto, T., & Er-Wei, B. (2012). A Web Aggregation Approach for Distributed Randomized Page Rank Algorithms. *IEEE Transactions on Automatic Control*, *57*(11), 2703–2717. doi:10.1109/TAC.2012.2190161

Kibeom, L., & Kyogu, L. (2015). Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items. *Expert Systems with Applications*, *42*(10), 4851–4858. doi:10.1016/j.eswa.2014.07.024

Leonidas, A., Dimitrios, K., & Panayiotis, B. (2011). Effective rank aggregation for meta-searching. *Journal of Systems and Software*, *84*(1), 130–143. doi:10.1016/j.jss.2010.09.001

Li, Z., Wang, Y., & Oria, V. (2001). A New Architecture for Web Meta-Search Engines. In *Proceedings of the Seventh Americas Conference on Information Systems*. Academic Press.

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. doi:10.1017/CBO9780511809071

Meng, W., Yu, C., & Liu, K.-L. (2002). Building efficient and effective meta search engines. *ACM Computing Surveys*, *34*(1), 48–89. doi:10.1145/505282.505284

Naresh, K. (2017). Document Clustering Approach for Meta Search Engine. In *Proceedings Series: Materials Science and Engineering*. *IOP Publication*. doi:10.1088/1757-899X/225/1/012291

Rajhans, M., Pradeep, K., & Bharat, B. (2015). A web recommendation system considering sequential information. *Decision Support Systems*, *75*, 1–10. doi:10.1016/j.dss.2015.04.004

Rashid, A., & Sufyan, M. M. B. (2008). *Aggregating Subjective and Objective Measures of We Search Quality using Modified Shimura Technique*. Springer-Verlag Berlin Heidelberg.

SatyaSai, K.P. & Raghavan, S.V. (2001). DLAPANGSE: Distributed Intelligent Agent based Parallel Architecture for Next Generation Search Engines. Indian Institute of Technology Madras, India.

Sugiura, A., & Etzioni, O. (2000). Query routing for Web search engines: Architecture and experiments. *Computer Networks*, *33*(1-6), 417–429. doi:10.1016/S1389-1286(00)00059-1

Vijaya, P., Raju, G., Santosh, K.R. (2016). Artificial neural network-based merging score for Meta search engine. *Journal of Central South University*, *23*(10), 2604–2615.

Vishwas, R., & Kumar, P. (2011). EGG (Enhanced Guided Google) - A Meta Search Engine for Combinatorial Keyword Search. Institute of Technology, Nirma University, Ahmedabad.

Zameer, G., & Leema, A. A. (2018). A Framework for Recommender System to Support Personalization in an E-Learning System. *International Journal of Web-Based Learning and Teaching Technologies*, *13*(3), 51–68. doi:10.4018/IJWLTT.2018070104

Zameer, G., & Leema, A. A. (2018). Course Recommendation Based on Query Classification Approach. *International Journal of Web-Based Learning and Teaching Technologies*, *13*(3), 69–83. doi:10.4018/IJWLTT.2018070105

Salman Ayaz is an Assistant Professor and is currently pursuing his research in the field of semantic web and data mining.

Jaya Venkat is the Professor in the Department of Computer Applications.

Zameer Gulzar is the Assistant Professor in the Department of Computer Applications.