International Journal on Emerging Technologies **11**(4): 72-77(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Evidence Phrase based Combined Relevance for Web Information Filtering

***Mu. Annalakshmi*[1] *and A. Padmapriya*[2]**
[1]*Research Scholar, Department of Computer Science, Alagappa University, Karaikudi (Tamil Nadu), India.*
[2]*Associate Professor, Department of Computer Science, Alagappa University, Karaikudi (Tamil Nadu), India.*

(Corresponding author: A. Padmapriya)
(Received 28 May 2020, Revised 19 June 2020, Accepted 02 July 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: In this information age people depend on Internet for everything. Large volume of data is added to it every day and this leads to information overload in web. Filtering out the required information from the overloaded web has been difficult for the users due its vast size and unstructured nature. Though search engines help the users to some extent, they too return millions of web pages for a query. This makes finding of required information difficult and time consuming for the users. Moreover, the results are not customized to meet the information needs of the user. This paper proposes a method to filter relevant web pages from search engine results by identifying the evidence phrases. The occurrence of the query keywords and their corresponding synonyms in the HTML source code of the web pages along with the factors in the user profiles such as visit count, bookmarks and downloads determine the evidence phrases. The user profiles are implicitly collected from the browser history and semantics of the query keywords are obtained from WordNet. The evidence phrases play an important role in ranking the web pages. Weights are assigned to evidence phrases based on their occurrence and score is calculated for each webpage. This score determines the rank of the web page in the result list. Experimental studies show that the proposed method is more precise than Google as it ranks more relevant pages at the top of the result list.

Keywords: Combined Relevance, Evidence Phrase, Information Filtering, Search Engine, Semantic Relevance, User profiles, User Relevance, Web Mining.

Abbreviations: IF, Information Filtering; ICD, International Classification of Diseases; ODP, Open Directory Project; IEPKCB, Information Extraction and Prediction using Partial Keyword Combination and Blends Measure; MAP, Mean Average Precision; SEO, Search Engine Optimization; SR, Semantic Relevance; SVM, Support Vector Machine; UR, User based Relevance.

I. INTRODUCTION

Ever since the inception of web in 1989 by Tim Berner's-Lee, it has been loaded with enormous amount of data. Now with the advent of smart phones, this information overload has increased drastically since data can be loaded anywhere anytime with lesser or no technical skills.

The major reasons for information overload [1] are

– Large volumes of new information are being created

– Exponential increase in information including videos, images and audio from visual and print media, social networks, ecommerce sites, etc.

– Possibility of creating, duplicating and sharing of information online.

– Information is often redundant or contradictory or old.

– Difficult to find relationships between groups of information since it is unstructured.

One of the most important problems of information overload is the difficulty in finding out the desired and relevant information from the huge pile of information in the web. Even though the information relevant to the user is available, the user is not able to retrieve it easily.

Information filtering (IF) has recently emerged as a technique for effective delivery of the required and also relevant information. Information Filtering differs from information retrieval by eliminating redundancy as well as unwanted information through (semi)automated or computerized methods before exhibiting them to the user.

Some of the key filtering techniques [2] include keyword vectors, n-grams, hyperlink structures and data mining techniques. The prime application of information retrieval is the search engine which in deployed for large scale of text collections.

Semantic web mining is about turning web pages into formal knowledge structures and extending the functionality of web browsers and knowledge manipulation and reasoning tools. It eases the information filtering process with the help of semantic structures of the web pages. Our previous paper dealt [4] with how semantic structures obtained from WordNet aided the user in information filtering. In addition to semantic structures, information reflecting the user behaviour can be used in effective information filtering. Such information can be gathered from user profiles.

The user profiles may be obtained from the user either implicitly or explicitly. In case of implicit method, the users' browsing behaviour such as frequency of visit, time spent in a page, whether the user has printed, etc are collected from the browser. In short, the use preferences are learnt without the explicit involvement of the user. The explicit method requires the users to submit information about their preferences, area of interest, etc to create their profiles.

The search engine helps the users to find web pages that satisfy their requirements. The user requirement is submitted to the search engine in the form of a query.
The search engine returns a list of web pages that match the query depending on various search factors [3]. This list usually contains several thousands of web pages from which the user has to find his needed page. This is really a cumbersome process. Semantic web along with user profiles can provide a major support in this information filtering process.
Though search engines help the user in retrieving the information from the web engines, the users are not satisfied with the accuracy of the search results. There is a necessity to filter the search results in a better way suiting the needs of the users. The proposed system has been developed to include semantics and user personalization for providing better search results to the user.

II. BACKGROUND STUDY

Search Engine Optimization SEO plays a remarkable role in improving the visibility of the web pages. Kavitha *et al.*, (2018) highlighted that the browsing habits of the users is the prime goal of SEO. This determines the results to be displayed in the first page of the search results. The commonly accepted practice about search engine results is that, the pages displayed in the top will be the ones accessed most by the users [19].
Preprocessing of text plays a very important part in information filtering since it improves the efficiency of information retrieval and reducing the size of the text to be processed. Preprocessing involves tokenization, stop words removal and stemming. Tokenization is the process of breaking the text into individual tokens [5] thereby removing punctuation symbols. It is used to identify the basic meaningful units to be processed.
Stop words are the most common words in the documents namely articles, prepositions which are not important in document classification or relevance determination. They also increase the overhead of processing and hence they are removed from the documents [6].
Stemming [7] refers to finding the root or base of a given word. For example, the words present, presented, presentation and presenting are converted to root word present. Their suffixes are removed to get the root word present. This improves the information retrieval performance by clustering words according to their root.
Kumar and Padmapriya (2016) applied an evidence based algorithm for extracting information related to subsequent diseases from the dataset. Here they employed key word based search [14]. This work presented the information extracted to the user without being filtered. The dataset used here is ICD-10. The dataset is preprocessed to remove redundancy. But this is not possible for the entire domain.
Ramesh and Chandrasekar (2016) proposes a semantic similarity technique which expands the user query by extracting the synonyms from WordNet. With the help of query terms and the synonyms extracted, the search results are rearranged. Evaluations show that the proposed system has a better performance than traditional non WordNet based search [8].
A hybrid conceptual pattern extraction [9] uses the snippets of the search engine's result documents for the given query concept. The terms in each of the snippet which are similar to the query concept are replaced by the concept. The frequency of the concept in the snippet is calculated and then the results are re-ranked according to the semantic similarity between the query and the document snippets.
Wu & Hu (2005) use three different types of keywords namely obvious, hidden and logical keywords and also Cellular Neural Network (CNN) like word nets to illustrate the semantic features and relations. The semantic features are extracted, trained and classified using Support Vector Machine(SVM). The web sensitive information is filtered based on these semantic features [17].
An artificial immune system [10] personalizes the user search by building the user profiles implicitly from the user browsing pattern. It uses an affinity function to compare the documents with user profile keywords. This affinity function is used in ranking the web pages.
Abdeljaber (2018) the author uses A* heuristic technique to find the closest horizontal root domain by making use of user profile and query keywords. Then the query is expanded by adding the concept extracted which disambiguates the initial query and this expanded query is resubmitted to the search engine to get better results [11].
A personalized search approach on internet [12] maps users known interests onto a group of categories with Open Directory Project (ODP), which aids in classification of user interests. Each webpage in the search result is assigned to a category with the nearest cosine similarity value. The categorized web pages are displayed to the users so that they can choose them based on their interests.
An Information Extraction and Prediction using Partial Keyword Combination and Blends Measure (IEPKCB) is proposed [13]. This is used to extract data related to the user query which is an unstructured input. The extracted information is then used for prediction. Here keyword based search is employed. The performance of this proposed work is assessed through precision, recall and F-measure.
An ontology based personalization framework [16] uses semantic information and also incorporates customization of user interests in re-ordering the search results. It is constructed based on the history of user actions and the explicit information provided by the user which is the (meta) information of the user. These concepts are compared with the ontology based semantic structures and the relevant information is retrieved.
The research work [18] uses hierarchical and semantic relationship between concepts to identify more user interests and also updates user preferences based on the subset of correlated concepts. A fuzzy semantic intersection between the user interests and semantic runtime context called contextualized user preferences is used to refine the process of personalized information filtering.

III. EVIDENCE BASED COMBINED RELEVANCE FOR WEB INFORMATION FILTERING

The proposed Evidence Phrase based Combined Relevance for Web Information Filtering uses two relevance metrics to re-rank the search results. The metrics are Semantic Relevance (SR) and User based Relevance (UR). The workflow of the proposed architecture is depicted in the following Fig. 1.

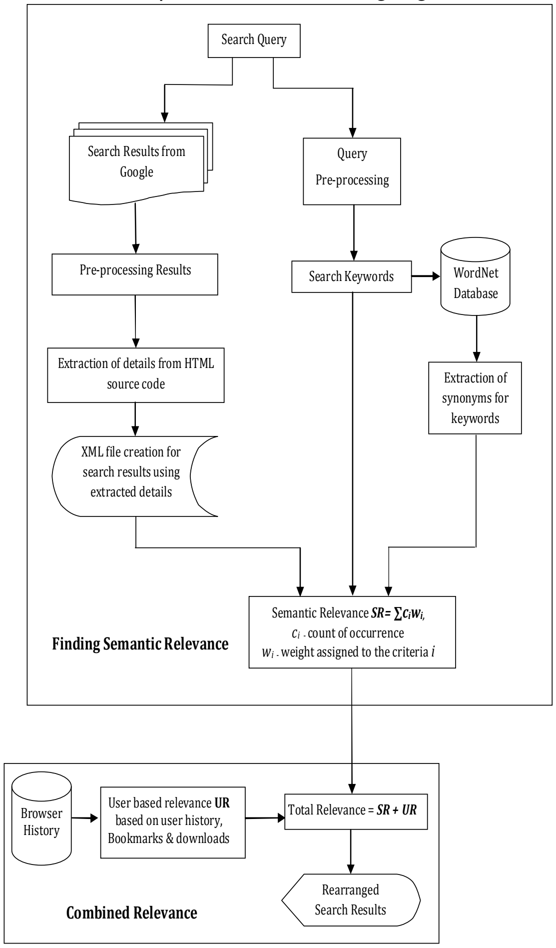

Fig. 1. Work flow of Evidence based Combined Relevance for Web Information Filtering.

The Semantic Relevance (SR) is calculated based on [4]. The research work [4] retrieved the results from the search engine for the given user query and ranked them in the decreasing order of semantic relevance. The semantic relevance is determined by the occurrence of the search query terms or their corresponding synonyms in the HTML source code of the webpage. Different weights were assigned to the matching terms in URL of the web page and different tags such as title, meta tags, headings and image tags as well as to the occurrence of the keywords in the body of the code as free text and co-position of the keywords.

A relevance score (based on semantics) is calculated for each of the web pages based on the weights which determines the order in the results. Information filtering using semantic relevance score alone may not be suitable for all types of users.

The preferences vary with respect to the users. The user preferences can be identified from the browsing history which is recorded in the form of user profiles.

The proposed work collects the user profiles implicitly from the browser since users may find it inconvenient to specify their preferences explicitly. The information about the web pages such as frequency of visiting the corresponding page, the time spent in the pages, whether the page has been downloaded or bookmarked are obtained from the browser the user uses. These factors can also be considered for the host of the web pages also. Weights are assigned proportionately with preferences in the order of factors like

- Successful downloads from the site
- Bookmarks on the site
- Paused downloads from the site
- Successful downloads from the host
- Bookmarks on the host
- Visit to the site
- Paused downloads from the host and
- Visit to the host

Most preference is given to successful downloads from the site and least preference is given to host visit. The calculated User profile based relevance is termed as UR. Then the combined relevance can be calculated as

Total Relevance (TR) = SR + UR

This sum of semantic relevance and user profile relevance gives the total relevance of each of the web pages which is used in the ordering of search results. The search results are re-ranked based on the user preferences and personalized according to the user. Each time when a query is entered, the semantic and user profile relevance are calculated. So the current user preferences are considered without disturbing the users unlike explicit user profile collection which requires updating of user profiles to reflect their current interests.

The algorithmic steps are listed below

— The search query is given to the search engine Google and the results are retrieved from the top 30 search results.

— An XML file is created [15] for each of the retrieved web pages. The details of the web pages such as URL name, title, meta tags, heading tags, image tags and number of word occurrences and their position of occurrence for each of the words in the webpage are retrieved after pre-processing which involves tokenization, removal of stop words and stemming to root words.

— The words in the search query are also pre-processed.

— The synonyms for each of the keywords in the search query are retrieved from WordNet and pre-processed.

— The evidence phrases include the count of occurrence of the keywords in the query and also their synonyms in URL, Title, meta tags, heading tags, image tag and normal text occurring as free text. In case of normal text the nearness of occurrence of the query words is also taken into account.

— Weights are assigned to each of these occurrences.

— The semantic relevance of the web page to the given query is calculated by sum of product of the count and weight of the keywords and their respective synonyms.

— The user profiles are implicitly collected from the browser the user uses. Also the occurrence of evidence phrases such as visit count, bookmarks and downloads are collected from the user's browsing history stored by the respective browser.

— The user based relevance is calculated based on these occurrences and the corresponding weights assigned.

— The relevance score of each of the web page which is the sum of semantic relevance and the user relevance determines the order of the corresponding web page in the search engine results with respect to the query.

IV. RESULTS AND DISCUSSION

As mentioned [19], the most commonly used search engines are Google and Bing. This research work compares the results obtained from the proposed method with the Google results.

The proposed work is implemented using .NET framework. The top 30 search results of the search engine Google for the given 10 different queries related to the field of Computer Science were distributed to users.

Table 1: Sample relevance ratings of two users for the query "Transposition Ciphers".

S. No	Website	User1	User2
1	https://en.wikipedia.org/wiki/Transposition_cipher	5	4
2	crypto.interactive-maths.com/simple-transposition-ciphers.html	4	3
3	https://www.britannica.com/topic/transposition-cipher	3	3
4	https://learncryptography.com/classical-encryption/transposition-ciphers	2	4
5	https://www.geeksforgeeks.org/columnar-transposition-cipher/	3	4
6	https://www.dcode.fr/transposition-cipher	5	3
7	https://www.staff.uni-mainz.de/pommeren/Cryptology/Classic/88_Transpos/Examples.html	2	4
8	https://nrich.maths.org/7940	3	5
9	https://www.youtube.com/watch?v=0um-_4SvPg0	5	4
10	https://www.youtube.com/watch?v=sHsnH1u03e4	5	3
11	https://www.ti89.com/cryptotut/transposition.htm	5	2
12	https://en.wikibooks.org/wiki/Cryptography/Transposition_ciphers	4	2
13	pi.math.cornell.edu/~mec/2003-2004/cryptography/transposition/transposition.html	3	3
14	https://www.slideshare.net/daniyalqureshi712/transposition-cipher-65475317	4	4
15	https://www.slideshare.net/AntonyAlex1/transposition-cipher	4	5
16	https://www.codemiles.com/java/how-row-transposition-encryption-algorithm-work-t613.html	4	5
17	practicalcryptography.com › Ciphers	3	4
18	www.tech-faq.com/substitution-and-transposition-ciphers.html	3	3
19	https://www.hackerrank.com/challenges/keyword-transposition-cipher	4	3
20	https://www.merriam-webster.com/dictionary/transposition%20cipher	4	2
21	https://www.cs.uri.edu/cryptography/classicaltransposition.htm	4	3
22	https://www.coursera.org/lecture/basic-cryptography-and-crypto-api/columnar-transposition-ciphers-cUsuc	4	4
23	rumkin.com/tools/cipher/coltrans.php	4	3
24	tholman.com/other/transposition/	4	3
25	www.ques10.com/p/13446/explain-with-examples-keyed-and-keyless-transpos-1/	4	4
26	https://crypto.stackexchange.com/questions/15748/frequency-analysis-of-transposition-ciphers	3	4
27	scienceblogs.com/goodmath/2008/08/24/transposition-ciphers/	3	3
28	www.counton.org/explorer/codebreaking/transposition-ciphers.php	5	4
29	https://kifanga.com/what-is-transposition-cipher/	5	3
30	https://www.coursehero.com/file/p29tg5v/Example-Row-Transposition-Ciphers-46-Row-Transposition-Ciphers-Plaintext-is/	5	4

The users were asked to rank the relevance of the web pages in a 5 point scale with 5 being more relevant and 1 being least relevant. Also their browsing history for last 10 days was collected from the corresponding browser used by the user. Since the evaluation requires the browsing history of the users, the number of users is limited to 50.

The users were from the field of Computer Science ranging from Post Graduate Students, Research Scholars to Professors. Sample browsing histories were not disclosed in order to preserve privacy of the users. The Table 1 given below depicts the relevance score of two random users from the sample space.

A. Performance Metrics

The performances of the proposed work are analyzed using three metrics namely precision, recall and mean average precision.

Precision [5] (P) is the fraction of retrieved documents that are relevant whereas Recall (R) [5] is the fraction of relevant documents that are retrieved.

$$\text{Precision} = \frac{\text{Relevant items retrieved}}{\text{\#retrieved items}}$$

$$\text{Recall} = \frac{\text{Relevant items retrieved}}{\text{\#retrieved items}}$$

Mean Average Precision (MAP) [5] provides a single-Fig. measure of quality across recall levels. When compared to the other metrics, the MAP is having good discrimination and stability. If the set of relevant documents for information need

$q_j \in Q$ is $\{d_1, d_2 \ldots\ldots d_{mj}\}$

qnd R_{jk} is the set of ranked retrieval results from the top result until you get to document d_k, then

$$MAP\,(Q) = \frac{1}{|Q|}\,\Sigma \sum_{j=1}^{|Q|} \frac{1}{mj} \sum_{k=1}^{mj} \text{Precision}\,(Rjk)$$

B. Performance Evaluation

This proposed work is compared with the result of Google Search Engine. Google is the most commonly used search engine. Hence comparing the performance of the proposed method with that of Google will definitely justify the performance of the proposed method. The calculated precision, recall and mean average precision are listed in the Table 2 and 3.

Table 2: Precision and Recalls value of User 1.

S. No of Webpage	Google		Proposed	
	Precision	Recall	Precision	Recall
1	1	0.05	1	0.05
2	1	0.1	1	0.1
3	0.666667	0.1	1	0.15
4	0.5	0.1	1	0.2
5	0.4	0.1	0.8	0.2
6	0.5	0.15	0.666667	0.2
7	0.428571	0.15	0.714286	0.25
8	0.375	0.15	0.625	0.25
9	0.444444	0.2	0.666667	0.3
10	0.5	0.25	0.7	0.35
11	0.545455	0.3	0.636364	0.35
12	0.583333	0.35	0.666667	0.4
13	0.538462	0.35	0.615385	0.4
14	0.571429	0.4	0.642857	0.45
15	0.6	0.45	0.666667	0.5
16	0.625	0.5	0.625	0.5
17	0.588235	0.5	0.647059	0.55
18	0.555556	0.5	0.666667	0.6
19	0.578947	0.55	0.684211	0.65
20	0.6	0.6	0.7	0.7
21	0.619048	0.65	0.666667	0.7
22	0.636364	0.7	0.636364	0.7
23	0.652174	0.75	0.608696	0.7
24	0.666667	0.8	0.625	0.75
25	0.68	0.85	0.64	0.8
26	0.653846	0.85	0.653846	0.85
27	0.62963	0.85	0.666667	0.9
28	0.642857	0.9	0.642857	0.9
29	0.655172	0.95	0.655172	0.95
30	0.666667	1	0.666667	1

The metrics of User 2 is shown in Table 3.

Table 3: Precision and Recall values of User 2.

S. No of Webpage	Google		Proposed	
	Precision	Recall	Precision	Recall
1	1	0.066667	1	0.066667
2	0.5	0.066667	1	0.133333
3	0.333333	0.066667	1	0.2
4	0.5	0.133333	1	0.266667
5	0.6	0.2	1	0.333333
6	0.5	0.2	0.833333	0.333333
7	0.571429	0.266667	0.714286	0.333333
8	0.625	0.333333	0.625	0.333333
9	0.666667	0.4	0.555556	0.333333
10	0.6	0.4	0.6	0.4
11	0.545455	0.4	0.545455	0.4
12	0.5	0.4	0.583333	0.466667
13	0.461538	0.4	0.538462	0.466667
14	0.5	0.466667	0.5	0.466667
15	0.533333	0.533333	0.466667	0.466667
16	0.5625	0.6	0.5	0.533333
17	0.588235	0.666667	0.470588	0.533333
18	0.555556	0.666667	0.5	0.6
19	0.526316	0.666667	0.526316	0.666667
20	0.5	0.666667	0.5	0.666667
21	0.47619	0.666667	0.47619	0.666667
22	0.5	0.733333	0.454545	0.666667
23	0.478261	0.733333	0.478261	0.733333
24	0.458333	0.733333	0.5	0.8
25	0.48	0.8	0.52	0.866667
26	0.5	0.866667	0.5	0.866667
27	0.481481	0.866667	0.518519	0.933333
28	0.5	0.933333	0.5	0.933333
29	0.482759	0.933333	0.517241	1
30	0.5	1	0.5	1

Table 2 shows the precision and recall calculation (based on the relevance score given by the user1) of Google and the results of the proposed system. The table shows the precision and recall values of the data being filtering from the web with respect to User 1. Here as mentioned earlier two random users are drawn from the sample space. The values in Table 2 are the precision and recall value of the User 1.

The precision and recall for the User 1 obtained using the proposed method as well as the Google search engine are shown in Fig. 2 and 3. Here precision and recall are shown for all the 30 websites.

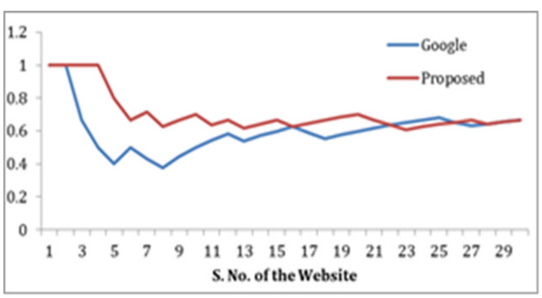

Fig. 2. Precision Comparison for User 1.

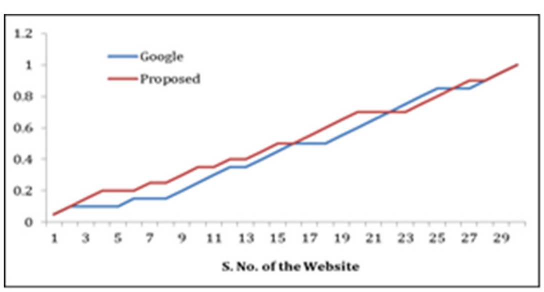

Fig. 3. Recall Comparison for User 1.

The precision and recall for the User 2 obtained using the proposed method as well as the Google search engine are shown in Fig. 4 and 5.

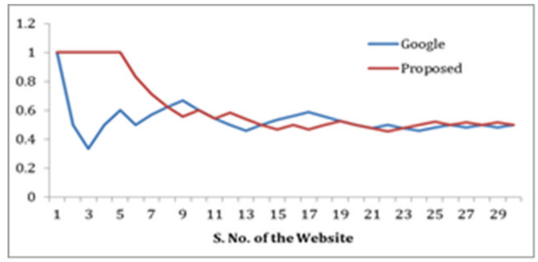

Fig. 4. Precision Comparison for User 2.

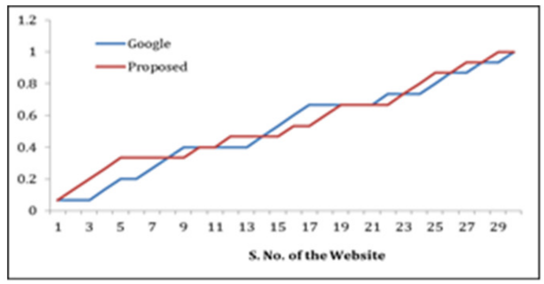

Fig. 5. Recall Comparison for User 2.

The mean average precision calculated for Google search results is 0.4869 and that of the proposed system is 0.6667. This shows that the proposed system has an average increase of 17.98% in precision than Google.

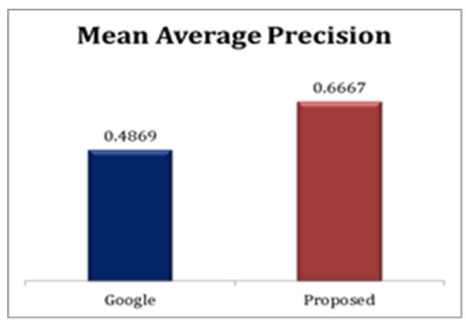

Fig. 6. Mean Average Precision Comparison.

V. CONCLUSION

This proposed Evidence Phrase based Combined Relevance for Web Information Filtering combines term based relevance with user profile based relevance. Both of them are determined by the presence of evidence phrases. This combined relevance is used to re-rank web pages of the search results of Google according to user profiles. Experimental studies show that the proposed method produces more precise ordering of search results. Web pages that are more relevant to the users appear in the top of the list which makes it easy for them to find the required information.

VI. FUTURE SCOPE

The evidence phrases are mainly based on keywords, synonyms and user profiles. Some more factors such as hypernyms, hyponyms can also be included in determining the evidence phrases. This research work evaluates the proposed methodology in terms of precision, recall and mean average precision. The experiment is conducted with fifty users. This can be further extended to more number of users and other metrics like interpolated precision, precision at K, R-Precision, Discounted Cumulative Gain, etc. can also be considered for evaluation.

ACKNOWLEDGEMENTS

This article has been written with the financial support of RUSA – Phase 2.0 grant sanctioned vide Letter No.F.24-51/ 2014-U, Policy (TNMulti-Gen), Dept. of Edn. Govt. of India, Dt. 09.10.2018

Conflict of Interest. There is no Conflict of Interest in this work.

REFERENCES

[1]. Zanarini, P. (2017). Information Overload, Why it Matters and How to Combat It. *Retrieved from internacton-design. org: https://www. interaction-design. org/literature/article/information-overload-whyit-matters-and-how-to-combat-it*.

[2]. Bohté, S. M., Langdon, W. B., & Poutré, H. L. (2000). On Current Technology for Information Filtering and User Profiling in Agent-Based Systems, Part I: A Perspective. *CWI, Centre for Mathematics and Computer Science*, 1-12

[3]. Annalakshmi, M., & Padmapriya, A. (2016). Search Factors used by search engines, *International Journal of Advanced Research Trends in Engineering and Technology*, *3*, 643-646.

[4]. Annalakshmi, M., & Padmapriya, A. (2018). Term Based Weight Measure for Information Filtering in Search Engines, *International Journal of Engineering and Technology*, *10*(1), 23-28.

[5]. Manning, C. D., Raghavan, P., & Schutze, H. (2008). Introduction to infor-mation retrieval? Cambridge University press, *20*, 405-416.

[6]. Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, *5*(1), 7-16.

[7]. Moral, C., de Antonio, A., Imbert, R., & Ramírez, J. (2014). A survey of stemming algorithms in information retrieval. *Information Research: An International Electronic Journal*.

[8]. Ramesh, S., & Chandrasekar, A. (2016). Semantic based Re-ranking model Using WordNet for retrieving Web URL in E-learning.

[9]. Abdullah K-K. A, Robert A.B.C., Adeyemo A.B. (2016). Semantic Indexing Techniques on Information Retrieval of Web Content. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(8), 347- 352.

[10]. Rastegari, H., & Shamsuddin, S. M. (2010). Web search personalization based on browsing history by artificial immune system. *International Journal of Advances in Soft Computing and Its Applications*.

[11]. Abdeljaber, H. A. (2018). Profile-Based Semantic Method using Heuristics for Web Search Personalization. *International Journal of Advanced Computer Science and Applications*, *9*(9), 191-198.

[12]. Ma, Z., Pant, G., & Sheng, O. R. L. (2007). Interest-based personalized search. *ACM Transactions on Information Systems (TOIS)*.

[13]. Sathish Kumar, L., & Padmapriya, A. (2019). Information Extraction and Prediction Using Partial Keyword Combination and Blends Measure. *IETE Journal of Research*, *65*(2), 164-171.

[14]. Kumar, L. S., & Padmapriya, A. (2016). Evidence based subsequent disease extraction from EMR Health Record by Grade Measure. In *2016 Online International Conference on Green Engineering and Technologies (IC-GET),* 1-5. IEEE.

[15]. Annalakshmi, M., & Padmapriya, A. (2019). Structuring of Web Pages using XML Framework for Information Filtering. *Asian Journal of Computer Science and Technology*, *8*(2), 35-38.

[16]. Vallet, D., Castells, P., Fernández, M., Mylonas, P., & Avrithis, Y. (2007). Personalized content retrieval in context using ontological knowledge. *IEEE Transactions on circuits and systems for video technology*, *17*(3), 336-346.

[17]. Wu, O., & Hu, W. (2005). Web sensitive text filtering by combining semantics and statistics. In *2005 International Conference on Natural Language Processing and Knowledge Engineering,* 663-667. IEEE.

[18]. Cantador, I., Fernández, M., Vallet, D., Castells, P., Picault, J., & Ribiere, M. (2008). A multi-purpose ontology-based approach for personalised content filtering and retrieval. In *Advances in Semantic Media Adaptation and Personalization* (pp. 25-51). Springer, Berlin, Heidelberg.

[19]. Kavitha, S. N., Thanuja D. V., & Shravya, M. (2018). Survey on Search Engine Optimization. *International Journal of Electrical, Electronics & Computer Science Engineering, 5*(2), 63-64.

How to cite this article: Annalakshmi, M. and Padmapriya, A. (2020). Evidence Phrase based Combined Relevance for Web Information Filtering. *International Journal on Emerging Technologies*, *11*(4): 72–77.