International Journal of Soft Computing 11 (6): 382-390, 2016
ISSN: 1816-9503

An Automatic Arabic Web Personalization Search Engines and Information Retrieval Systems

Safaa I. Hajeer, Rasha M. Ismail, Nagwa L. Badr and M.F. Tolba
Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

Abstract: Over the years, several achievements on the improvement of web personalized searching based on user's interests, preferences and contextual information have been made, unfortunately, most of them are concerned with the static profile approach, preferences or weight values and not changed once the user preference profile is created and this might be unacceptable by users. Also, they didn't consider usage features like the changing in user's attributes, ontology, description and features over time. In addition, its rare that any of them are considered in Arabic Language Personalization Re-ranking through the search engines. Therefore, in this study we proposed a New Automatic Semantic Personalization Re-ranking (NASPR) approach. The objective of this NASPR algorithm is to overcome the drawbacks of ranking algorithms and improve the efficiency of web searching. The NASPR approach was applied to 242 Arabic Corpus to measure its performance and the results show improvements in the recall and precision by using the new personalized approach.

Key words: Information Retrieval (IR), personalized search engines, automatic user profile, machine learning, semantic web ranking, ontology, Web Ontology Language (OWL), personalized information retrieval, collaborative filtering, contextual-based personalization, arabic language

INTRODUCTION

The amount of information in the world has and will continue to increase exponentially over the years, new books, journal articles and conference proceedings have been coming out each year. Searching within this flood of information to provide personalized information to users is an important issue in the information retrieval field.

Really, each user on the Internet has a distinct background and a specific goal when searching for information on the web. Currently, most search engines still serve a generic user in a "one size fits all" fashion returning the same set of results whilst being unaware of individual preferences which in turn can vary with individual working environment times (Rohini and Varma, 2007). Only a few of the available search engines like Google, Bing, etc. tried as much as possible to comfort various users and meet their search needs in the real world. Unfortunately, some of them are unsuitable to provide a perfect personalized search results to users, because of limited user's personal information such as hobbies, preferences and interests. The perfect idea to solve this problem is to provide the preference profile for each user to help in the personalization of search results. The user preference profiles consist of the results of semantic analysis, the input query that may hold ambiguity, the documents that were clicked by the user, the queries that were used by the user in the past and some weight values. However, a user preference profile that includes incorrect user preferences only gives a headache to users. Incorrect user preferences are generally obtained by the static profile approach. In this static profile approach, preferences or weight values are static and not changed once the user preference profile is created. Most existing personalized IR systems usually model items as static user profiles, i.e., unchanging in attributes, user's ontology, description and features. Because user's preferences vary over time, place, context or domain, the static profile approach has a high chance of having incorrect user's preferences. To solve this problem, the automatic user profile is the solution and is based on machine learning techniques.

Another problem facing the search engines is the user's queries, these queries which are often short, consist one or two terms, that make it to be ambiguous; For example, the word "Java" has different meanings in different domains. In computers, it means one of an object oriented programming language in coffee, it means a beverage consisting of an infusion of ground coffee beans and in geography, it's an island in Indonesia to the

Corresponding Author: Safaa I. Hajeer, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

South of Borneo. An additional example is the term "Apple" it may belong to a technology company Apple Inc., while a farmer would possibly think of it as a fruit. Also, the term "Virus" has different meanings in different domains. In Biology, it means a simple sub-microscopic parasites of plants, animals and bacteria that often cause disease whereas in computers, it means a program or a piece of code thats loaded onto your computer without your knowledge and runs against your wishes. Examples in the English language are numerous, this rule may not apply for another language like Arabic; Arabic is a highly inflected language and has a complex morphological structure. For Example, "wehdat" is a term in Arabic: This term may mean a unit of measurement, we say in Arabic "wehdatkeyas", also "Wehdat" may belong to a blood in Arabic says "wehdat dam", furthermore, it may be used to sign a place in Arabic like "wehdatwamarkezgamaa" or housing units in Arabic "wehdatsaknya". Really, the users while entering a query would expect results belong to their domains. Unfortunately, this ambiguity will lead the search result to contain different documents far away from user needs. Therefore, often users are unable to understand the obtained result. For this, the user must refine the query in order to get the required results but this is a time consuming process. Personalized search is an approach that re-ranks search results based on user ontological profiles and combines the information in all documents to provide personalized search results. Each concept in an ontological user profile is built using ontology documents.

Therefore, to overcome the pervious problems, this study proposes a hybrid automatic semantic personalization re-ranking algorithm to utilize the usage features. This algorithm is called NASPR (New Automatic Semantic Personalization Re-ranking). The objective of this algorithm is to improve the user's ranked list that's provided from Arabic search engines. This improvement is important as it will affect the effectiveness and the performance of personalized Information retrieval systems and web search engines.

Literature review: The target of most search engines is to serve their users as much as possible, only a few of them like Google, Bing … etc. tried as much as possible to comfort various users and meet their search needs in the real world. This pushed the researchers to be in a race to discover the most effective and accurately personalization models that serve the user's needs as much as it can.

As an early study in 2005 (Sun *et al*., 2005) proposed a novel approach CubeSVD to improve the performance of Web Search. Their approach was utilized by focusing on click-through data and contains different types of objects (user, query and web pages). According of analysis for these data they attempted to discover web users' interests and the patterns that users locate information. They did their experimental evaluations using a real-world data set collected from an MSN search engine show that CubeSVD achieved to improve the performance of search results in comparison with other methods (Latin Semantic Indexing (LSI) and Collaborative Filtering (CF). Jeon *et al*. (2010) states that the analysis by Sun *et al*. (2005) approach is very complex and it is difficult to apply commonly used search engines.

A contextual model was presented by Jrad *et al*. (2007). This model used for tourism usage. The idea of it was that the users and their context are in an extensible way, this information can be interpreted and used for personalization. Each user profile is created based on the ontology. The ontology includes the combination of the context and user behaviors in the past with influences user current context and behaviors. The harder and effient thing of this model is the combination of various filtering and reasoning algorithms in the process of personalization, so this model may be unsuitable for some search engines.

A study was conducted in 2009 by Carmel *et al*. (2009), they investigated personalized social searches based on the user's social relations. The re-ranking search results are based on the relationship strength with the user's related people and topics. The results show the high effectiveness of this approach for social search and implies that the social relations used for personalization, as derived from the user's social network are highly reliable in predicting user interests and preferences.

Mohammed *et al*. (2010), a personalization approach to construct the user profile as part of the ODP by storing the concepts related to user's clicked pages only. For a given query, a query profile was created by expanding keywords into a semantic hierarchy from WordNet and associated results are matched to each node. Re-ranking is done by mapping results in the query profile with topics in the user profile, however, the user profile is considered static and might be inaccurate as it is not updated according to the user's changing interests over time.

An adaptive user profiling approach by Jeon *et al*. (2010). The idea of this approach based on Collaborative Filtering (CF) technique, this technique is using dynamic updating policy considering the change of the user preferences over time and domain. The personalized search results for this approach that for each user through automatic creation, maintenance and personalization of user preference profiles that include search patterns for each user.

A study was presented by Ghorab (2011), this study concerned improving the personalization techniques in Multi-lingual Information Retrieval (MIR) Systems. The study investigates how to model different aspects of a multi-lingual search user. Information about users can be demographic information such as language and country or information about the user's search interests. This information is gathered explicitly by asking the user to supply the required information or implicitly by inferring the information from the user's search history. Unfortunately, this study was in two pages at the ACM's conference and did not show the results.

Moawad *et al.* (2012), the research presented a model for web search personalization for multi-agent systems, the model was built and maintained the user profile dynamically. The semantic process is based on WordNet ontology. This model was tested on two users to search the web using the same query but with different profiles. The results show that their model had increased the precision of both Google and Bing search engines. This research has a weakness in using WordNet ontology which is poor in semantic of some words.

An effective hybrid personalized re-ranking search approach is proposed by Fathy *et al.* (2014), this approach models user's search interests in a conceptual user profile and then exploiting this profile in the re-ranking process. Each concept in the user profile consists of two types of documents; taxonomy document and viewed document. The results show that semantic identification of user's search interests improves the re-ranking quality by providing users with the most relevant results at the top of the search results list. That means, the re-ranking search results using a hybrid approach of the viewed documents, taxonomy documents and the original ranking give more relevant results on the top. This hybrid approach based on click through which may in accurate indicator and done on English Language only.

Vicente-Lopez a comparative study between six different user profile representation approaches, based on the content of the documents of the Andalusian Parliament. Then in the same research the authors provided a hybrid approach having a two level user profile representation, the first level is represented by subjects which are from a thesaurus. Note that subjects here are manually assigned to the initiatives in the documents by documentalists. The second level is by terms which are from representing these subjects. The results show that the best case reached up to 80.67% of improvement with respect to the original non-personalized model.

Bibi *et al.* (2014) proposed the concept based personalization approach, the user profile is based on concepts which are groups of words that occur frequently in web snippets of visited web pages. Concepts are organized in the profile as a tree with the relationship between these concepts. Weights are assigned/ incremented to concepts found in the clicked web snippets and to concepts having a relationship with this concept. Re-ranking is completed by assigning scores to current web snippets for a given query based on the aggregation of its concept's weight. As a result, the re-ranked results adapt to various users.

Liu *et al.* (2004) proposed a technique in personalized web search for improving retrieval effectiveness. Web search personalization is to carry out retrieval for each user incorporating his/her interests. This research proposed a novel technique to learn user profiles from users search histories. The user profiles and general profiles were then used to improve retrieval effectiveness in Web search.

A technique has been suggested by Shou *et al.* (2014) for supporting privacy protection in personalized web search. This existing system focused on profile based personalization in passive profiling technique. It has collected the implicit information from the user side and maps this into the Open Directory Project (ODP). The implicit information depicts the user's interests and preferences. It may be the previous query issued by the user, desktop files, amount of time spent on the SERPs, clicked Uniform Resource Locators (URLs)and so on.

In a recent study done by Fancy and Rajamurugan, 2015), it focused on two techniques, namely active and passive personalization to personalize the web search for individual users. The main objective of this study is to provide the web content personalization thataffects the relevant Search Engine Result Pages (SERPs) for the users according to their interests and preferences. Then the ranking of SERPs is performed by using the Personalized Page Rank (PPR) algorithm. The results show that the passive profiling technique gives accurate results in a short time compared to active profiling, because in passive profiling there is no need for user interaction to refine the SERPs. This personalized study based on PageRank algorithm which concern in the structure of links in the pages more than its content.

From previous studies most researches were concerned with the static profile approach, preferences or weight values that are not changed once the user preference profile is created, but this might be unacceptable by users. Furthermore, most of them usually

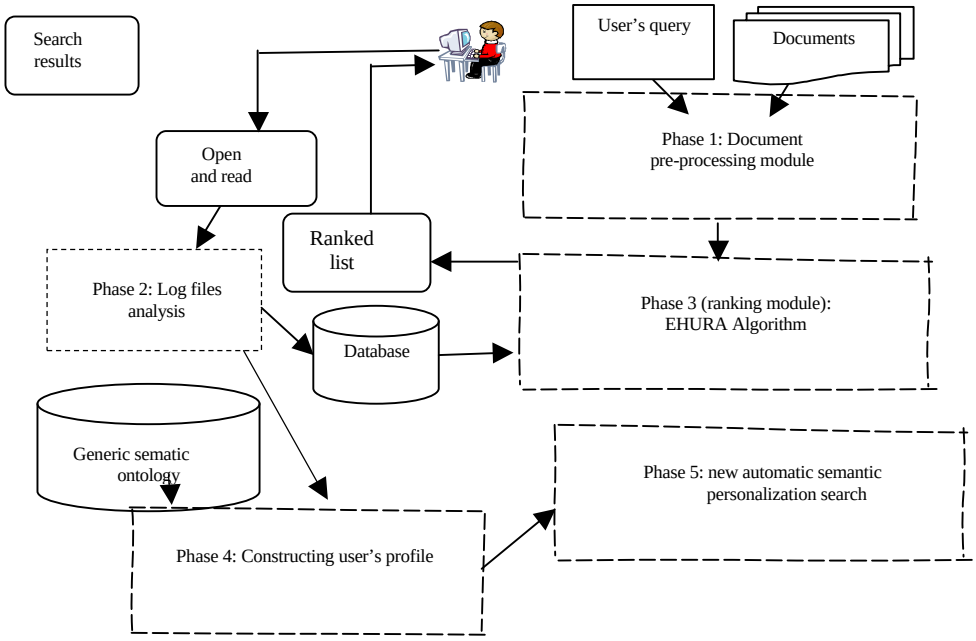

Fig. 1: The system architecture

did not consider the changing in user's attributes, ontology, description and features over time because user's preferences vary over time, the reasons might be the changing place, context or domain. In addition, it is rare for them to be considered in Arabic language personalization re-ranking through the search engines. Therefore, we decided to develop a new hybrid automatic Arabic semantic personalized re-ranking algorithm that utilizes the usage features like; user's time spent and user's frequency of visits. This hybrid personalized re-ranking algorithm is based on our work EHURA (Hajeer *et al*., 2015a, b) ranking, semantic analysis, machine learning and usage features which are:

- Frequency of visit that the numbersf visits of a web page by the user
- Time spent that shows the time spent here is the real time spent on a page

MATERIALS AND METHODS

The system architecture: This study discusses the New Automatic Semantic Personalization Re-ranking (NASPR) approach for Arabic Language searching. This approach was applied to the following system architecture and is shown in Fig. 1.

According to Fig. 1, our system consists of five main phases. Each phase has several working modules in it. These phases and it modules are explained as follows in the next subsections.

Phase 1 (Document pre-processing module): This phase consists of the following modules.

Module 1 (Tokenization): this stage is for breaking a stream of text into words and keeping the words in a list called a word's list.

Module 2 (data cleaning): Removes useless words from the Word's List, these useless words are stored in a stop words database as appears in the figure. The database of Arabic has 1459 stop words with a size 10 kb.

Module 3 (stemming): In this stage, we applied a hybrid affix removal algorithm (Arafat and Saad, 2008) for Arabic language.

Module 4 (indexing): Indexing is a process for describing or classifying a document by index terms; index terms are the keywords that have a meaning of their own (i.e., which usually has the semantics of the noun). These index terms are grouped in an indexer and the stemmer is service at this stage by improving the group of these keywords in the indexer.

Phase 2: Log files analysis: This phase is for removing irrelevant records from log files and may contain lots of them, so, in order to enhance the efficiency of the usage based retrieval algorithm by using relevant records only. This phase consists of a series of processes like

data cleaning, user identification, session identification (Lingras and Akerkar, 2010). Each process is explained as follows.

Data cleaning is the process of removing unnecessary records like graphics, video and formatted information, i.e., css. In addition, this process also removes the records of failed HTTP status codes. User identification is the process of identifying users and user agent fields of log entries and is consider on:

- Different IP addresses refer to different users
- The same IP with a different operating system or different browser should be considered as a different user
- While the IP operating system and browsers are all the same, a new user can be determined as to whether the requesting page can be reached by accessing the pages before according to the topology of the site
- A user session is considered to be all of the pages accessed that occur during a single visit to a web site. In session identification the process for defining users that may access the site more than once

Phase 3 (ranking module: EHURA Algorithm): EHURA Algorithm consists of the following modules:

Module 1 (content-based ranking): This part focuses on the ranking algorithm and is based on the content of documents and queries. It simply tries to find the similarity between the contents of documents and queries. We applied here the cosine similarity measure, this selection is based on studies that are represented in (Hajeer, 2012a, b) which prove that the cosine measure is the most efficient one in compared with other statistical measures.

The Cosine measure calculates the angel between two documents (between document and user's query which is treated as a document) representation vectors. Thus, a cosine value of zero means that the query and document vector were orthogonal to each other and means that there was no match or the term simply did not exist in the document being considered. To know the cosine relations between two documents (document D and query Q) see Eq. 1:

$$\text{Cosine (D, Q)} = \frac{|D \cap Q|}{\sqrt{|D| \times |Q|}} \quad (1)$$

Where:

Cosine (D, Q) = The Cosine Similarity relationship between document D and user's query Q

D = Refers to the document in the collection

Q = Refers to the user's query

After calculate the similarity measure, the ranked list appears to the user as an answer of his/her query. This list is arranged from the highest value of cosine measure to the lowest one as a weight inthe ranked list. ranking algorithms according to the most relevant to the user's query.

Module 2 (usage based parameters): In this stage, the system calculates two usage based parameters as the following.

Frequency of visits that determine the relevance of a web page by its selection frequency in order to find the frequency weight which is the admittance frequency of page u is the number of times the page is visited and the page rank which appears in the ranked list from the previous stage. The frequency weight formula is:

$$FW = \frac{\text{Number of visit on a page}(u)}{\text{TotalNumber of visit on all page}} \times PR(u) \quad (2)$$

Where:

FW = Frequency Weight

PR(u) = The page rank of a page u

Time spent that shows how long users spend on a page after removing the download time of the page, because a user generally spends more time on a more useful page and does not waste more time on screening the page and rapidly skipping to another page. So, it's an important parameter to indicate usefulness of the pages, this parameter is considered to calculate the real time spent on a page by taking the value of time taken (time spent on the page) from the log file, subtracting from Download time in order to find a time spent weight as the following:

$$TW = \frac{\text{Timespent on a page}(u) - \text{DownloadTime}(u)}{\max(\text{Timespent on a page}(u) - \text{DownloadTime}(u))} \quad (3)$$

where, TW is time spent weight:

$$\text{Download time}(u) = \frac{\text{Size of a page}(u)}{\text{Transfer rate for page}(u)} \quad (4)$$

Module 3 (usage based re-ranking): This is the final stage in our EHURA algorithm and is basically uses the two parameters that are calculated in the previous stage to find the usage based weight which is equal to the new weight for each page. This weight is then used to re-rank

the pages and the effective reflects on the previousrank list to obtain a new ranked list. So as a result, a new search engine results appear to the user.

Phase 4: Constructing a user's profile

Module 1 (Domain Hierarchal): This stage creates a sequence of levels, going from many specific terms at the lower levels to a few generic terms at the top by combining the potential hierarchies in different domains. This stage is based on creating levels of generic semantic ontology General Ontology Web Language (OWL) or the domain hierarchal extracts from user's profile if the same query has been entered before.

Module 2 (usage features): Extracts some usage features like; user's time spent on web pages and the user's frequency of visits from log files.

Module 3: User profile adaption: this stage is to model and track users' interests and their changes, to address that, Content Based Filtering (CBF) has been explored in this system. User interests involve the interests on fixed categories and dynamic events.

Phase 5: new automatic semantic personalization search:

Module 1: User's interest identification: this module holds two implicit stages.

Identifying user's topics of interest for search by using semantic mapping from a Generic Semantic ontology database or from the user's profile if the user searched before for the query. The Generic Sematic ontology (General Ontology Web Language (OWL) holds a sequence of levels, going from many specific terms at the lower levels to a few generic terms at the top which may hold an ambiguity concept. For Example, "Ain" in Arabic as a query may mean "Ain Shams" or "Ain" as part of a body (organs of sight). "Ain" is in the top level in the Generic semantic ontology and the lowest level is "Ain Shams" and "part of a body (organs of sight)", the decision of which branch to take for the low level is based on user behaviors and their interest before, the Generic Sematic ontology combines the potential hierarchies in different domains and not only for "Ain" as in the example.

Usage features: Extract the frequency of visits, at this time the frequency of visits is the numbers of visits per page by the user for each visit for the page, the system adds one to the frequency of visits automatically. In addition to the usage features extracts here, is the time spent, the time spent here is the real time spent on a page by taking the value of time taken (time spent on the page) from the log file.

Module 2 (calculate the new weight): In order to calculate the new weight, the new weight should pass several steps to complete. The first step calculates the score of the term, this can be done by computing the cosine similarity (Hajeer, 2012a) between the query and each document separately for a given concept. As following:

$$\text{Score} = \text{Sim}(Q, \text{Doc}) \tag{5}$$

Where:

Sim = The cosine similarity
Q = User's query
Doc_i = The semantic documents saved in the user's profile

The second step utilizes the usage features, these usage features are the frequency of visits and time spent which are extracted from the previous module, then sum of values for each document in terms USF as in the following equation:

$$\text{USF} = \text{Freq} + \text{Time Spent} \tag{6}$$

Where:

USF = Usage Features of document i
Freq_i = Frequency of visits for document i
Time spent_i = Time taken (time spent on the document i) from the log file

The final step based on EHURA ranked list weights, i.e., bring the pervious weights coming from ranking algorithm EHURA and compare each document in the user profile with the EHURA's ranked list weights. Finally, the equation of the new weight for each document is calculated by:

$$\text{New Weight} = (1-\alpha)\times \text{EHURAWeight} + \alpha\times(\text{Score} + \text{USF}) \tag{7}$$

Where:

New Weight = New Weight for document i
EHURAWeight = EHURA Weight for document i
α = Threshold ($0 \le \alpha \le 1$)
Score = The result of Eq. 5
SF = the result of Eq. 6

Module 3 (personalization re-ranking): This is the final stage in New Automatic Semantic Personalization Re-ranking (NASPR) approach, it basically uses the new weight coming for each page (Pages presented as documents), this weight is used to re-rank the pages from the highest to lowest values of the new weight. This re-ranking provides a new ranked list and the effectiveness reflects on the previousrank list to get a new ranked list. So as a result, new search engine results appear to the user based on his/her interests and behavior.

In order to see if these results make search engines more efficient, the system was tested using IR performance measures, this will be explained in the next section.

Performance analysis: In order to study the performance of our system, we used different evaluation measures. These measures are discussed. Then, the data sets used and the experimental results.

Evaluation of the proposed system: This study is to evaluate the performance of our IR system. The performance measured by the recall and precision measurements and other measures which are represented in the following equation:

$$\text{Precision} = \frac{\left|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}\right|}{\{\text{retrieved documents}\}} \quad (5)$$

$$\text{Recall} = \frac{\left|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}\right|}{\{\text{retrieved documents}\}} \quad (6)$$

Fall-out is the proportion of non-relevant documents that are retrieved out of all non-relevant documents available:

$$\text{Fall} - \text{out} = \frac{\left|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}\right|}{\{\text{non} - \text{retrieved documents}\}} \quad (7)$$

$$\text{F} - \text{Measure} = \frac{\left|2 \times \text{Precision} \times \text{Recall}\right|}{\text{Precision} + \text{Recall}} \quad (8)$$

where, F-measure is the weighted harmonic means of precision and recall.

$$\text{Avep} = \frac{\sum_{i=1}^{N_0} P_i(r)}{N_q} \quad (9)$$

Where:

AveP = Average precision at recall level r

$P_i(r)$ = The precision at the recall level r for the ith query

N_q = The number of queries used

RESULTS AND DISCUSSION

For testing our system, it was applied on Ain Shams Arabic corpus. The Arabic corpus belongs to the Modern Standard Arabic type; it contains 242 documents with different sizes and we tested the system with 20 queries in order to evaluate the IR system performance.

The log files stored 622 MB of data and we have 323 MB data after pre-processing by analyzing those log files using one of the analyzer tools called Deep Log Analyzer. Deep Log Analyzer did the series of processes that are explained in section 3, i.e. data cleaning, user identification and session identification with several statistics about usage data like; numbers of Hits, numbers of successful Hits, numbers of repeated visitors, most visitors come from which country by percentage and value, etc.

The user's profiling approach was tested by extracting 7 different IP addresses belonging to different users from the log files and studying their behaviors over 4 week. Figure 2 shows the number of retrieved relevant documents over the retrieved top 10 documents for the different users. It's clear from the figure that the number of relevant documents in the top 10 for a given query is increased for most users that participated in this testing over time, unfortunately, in some cases the decrease of the number of relevant documents occasionally happened, including the case for "user7 from the second to the third week. This decreasing means that a user uses a query using new terms and finds new documents whenever doing search. Also, in zooming more for the behaviors of "user3" and "user4", the figure shows that they kept the same behavior during the time between the second week and the third one that means that they were using the same query within this period and keep the same visiting numbers and time spent, so the number of relevant documents are kept the same. This situation rarely happens.

Our system was tested using IR evaluation measures which was mentioned in the evaluation section. Figure 2 shows the precision and recall results for each query for the ranking algorithm: EHURA in comparison with the new automatic semantic personalization re-ranking

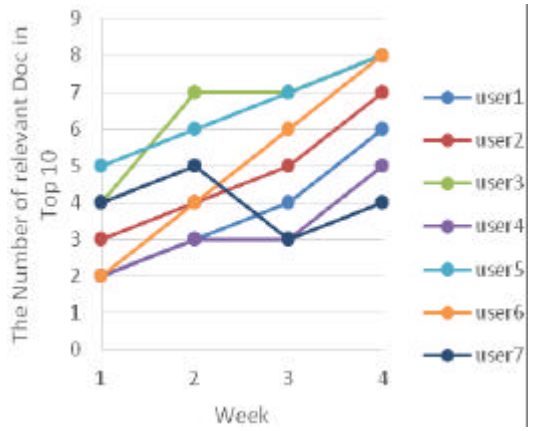

Fig. 2: The relevant documents for top 10 search results

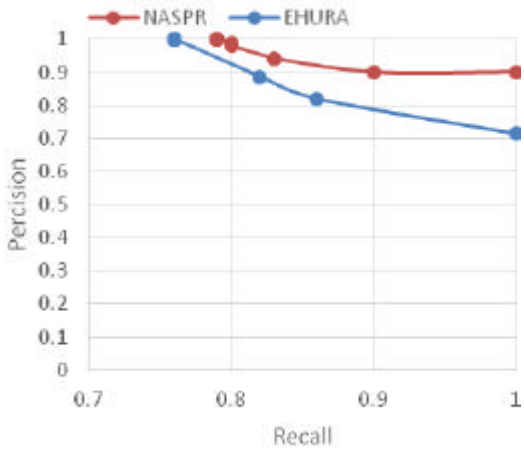

Fig. 3: Precision and recall for ranking against the Arabic Ain Sham's corpus 20 queries

(NASPR) Algorithm. It's clear that NASPR reached a better result overall than the EHURA one. The average precision of our new approach (NASPR) reached 98% while the precision of the EHURA ranking algorithm is 97%, the results are shown in Table 2. Therefore, our proposed NASPR algorithm improves the precision over the EHURA ranking algorithm by about 1.14% while it also improves the recall percentage by nearly 3%.

The proportion of non-relevant documents retrieved (fall-out) from the system using the EHURA algorithm reached 22% while our proposed NASPR algorithm reached 18.9%

Figure 3 shows the F-Measure for using the EHURA algorithm and the NASPR and it is clear from the figure that the NASPR algorithm improved the F-Measure over the EHURA algorithm by 2.3% shown in Fig. 4.

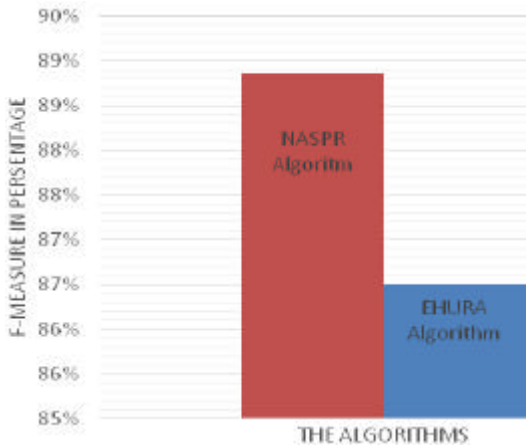

Fig. 4: F-measure for ranking against the Arabic Ain Sham's corpus

Table 1: Evaluation measures for our system for arabic language

Parameters	Precision	Recall	Fall-out	F-Measure
EHURA ranking	0.9711	0.7800	0.220	0.8651
NASPR re-ranking	0.9825	0.8110	0.189	0.8885

CONCLUSION

Searching becomes a normal behavior of our life. Millions of users interact with search engines daily. Many of the existing information retrieval systems still serve a generic user in a "one size fits all" fashion returning the same set of results whilst unaware of individual preferences which in turn can vary with individual working environment times. Only a few of the available search engines like Google, bing, etc., tried as hard as possible to comfort various users and meet their search needs in the real world. Unfortunately, some of them are unsuitable to provide a perfect personalized search result to users. Unfortunately, those search engines are suffering from static user profiles with limited user's personal information such as hobbies, preferences, behavior and interests. In addition, theyare not concerned insolving the ambiguity of user's queries. Furthermore, its rare for researches to consider Arabic language personalization re-ranking through the search engines.

Thus, in this study we propose an efficient new automatic semantic personalized search engine using a new automatic semantic personalization re-ranking algorithm called NASPR. The objective of this NASPR algorithm is to overcome the drawbacks of ranking algorithms and improve the efficiency of web searching.

The system was applied to Ain Shams Arabic corpus for testing and the results show that the NASPR algorithm improves the performance of the information retrieval system in respect to the recall and precision measures. So, NASPR improves the efficiency of search engines and information retrieval system by a good percentage.

REFERENCES

Arafat, S. and S. Saad, 2008. An affix removal stemming algorithm for Arabic language. Int. J. Intell. Comput. Inf. Syst., 8: 141-153.

Bibi, T., P. Dixit, R. Ghule and R. Jadhav, 2014. Web search personalization using machine learning techniques. Proceeding of the 2014 IEEE International Conference on Advance Computing (IACC), February 21-22, 2014, IEEE, Pune, India, ISBN: 978-1-4799-2572-8, pp: 1296-1299.

Carmel, D., N. Zwerdling, I. Guy, O.S. Koifman and N. Harel et al., 2009. Personalized social search based on the user's social network. Proceedings of the 18th ACM Conference on Information and knowledge Management, November 02-06, 2009, ACM, New York, USA., ISBN:978-1-60558-512-3, pp: 1227-1236.

Fancy, L. and A. Rajamurugan, 2015. Improving personalized web search quality in information retrieval. Int. J. Innovative Res. Comput. Commun. Eng., 3: 3057-3064.

Fathy, N., T.F. Gharib, N.L. Badr, A.S. Mashat and A. Abraham, 2014. A personalized approach for re-ranking search results using user preferences. J. UCS., 20: 1232-1258.

Ghorab, M.R., 2011. Improving query and result list adaptation in personalized multilingual information retrieval. Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, July 24-28, 2011, ACM, New York, USA., ISBN: 978-1-4503-0757-4, pp: 1323-1324.

Hajeer, S.I., 2012a. Comparison on the effectiveness of different statistical similarity measures. Int. J. Comput. Appl., 53: 14-19.

Hajeer, S., 2012b. Vector space model: Comparison between euclidean distance and cosine measure on Arabic documents. Int. J. Eng. Res. Appl., 2: 2085-2090.

Hajeer, S.I., R.M. Ismail, N.L. Badr and M.F. Tolba, 2015a. AEHURA: A new ranking algorithm for Arabic web search engines. Asian J. Inf. Technol., 14: 105-110.

Hajeer, S.I., R.M. Ismail, N.L. Badr and M.F. Tolba, 2015b. An Efficient Hybrid Usage-Based Ranking Algorithm for Arabic Search Engines. In: Computational Science and its Applications, Gervasi, O., B. Murgante, S. Misra, M.L. Gavrilova and A.M.A.C. Rocha *et al.* (Eds.). Springer, Berlin, Germany, ISBN:978-3-319-21403-0, pp: 382-391.

Jeon, H., T. Kim and J. Choi, 2010. Personalized information retrieval by using adaptive user profiling and collaborative filtering. AISS., 2: 134-142.

Jrad, Z., M.A. Aufaure and M. Hadjouni, 2007. A Contextual user Model for Web Personalization. In: Web Information Systems Engineering, Weske, M., M.S. Hacid and C. Godart (Eds.). Springer, Berlin, Germany, ISBN:978-3-540-77009-1, pp: 350-361.

Lingras, P. and R. Akerkar, 2010. Building an Intelligent Web: Theory and Practice. Jones & Bartlett Publishers, New York, USA., ISBN:13:978-0-7637-4137-2, Pages: 325.

Liu, F., C. Yu and W. Meng, 2004. Personalized web search for improving retrieval effectiveness. IEEE Trans. Knowl. Data Eng., 16: 28-40.

Moawad, I.F., H. Talha, E. Hosny and M. Hashim, 2012. Agent-based web search personalization approach using dynamic user profile. Egypt. Inf. J., 13: 191-198.

Mohammed, N.U., T.H. Duong and G.S. Jo, 2010. Contextual Information Search Based on Ontological user Profile. In: Computational Collective Intelligence, Pan, J.S., S.M. Chen and N.T. Nguyen (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-16731-7, pp: 490-500.

Rohini, U. and V. Varma, 2007. A novel approach for re-ranking of search results using collaborative filtering. Proceeding of the ICCTA'07 International Conference on Computing: Theory and Applications, March 5-7, 2007, IEEE, Hyderabad, India, ISBN: 0-7695-2770-1, pp: 491-496.

Shou, L., H. Bai, K. Chen and G. Chen, 2014. Supporting privacy protection in personalized web search. IEEE Trans. Knowl. Data Eng., 26: 453-467.

Sun, J.T., H.J. Zeng, H. Liu, Y. Lu and Z. Chen, 2005. CubeSVD: A novel approach to personalized web search. Proceedings of the 14th International Conference on World Wide Web, May 10-14, ACM Press, New York, USA., pp: 382-390.