KEYPHRASE BASED USER PROFILES IN SHORT-TERM AND SESSION-TERM QUERY LOGS

Sara Abri and Rayan Abri
Department of Computer Engineering, Hacettepe University. Ankara, Turkey

ABSTRACT

Personalization depends on prior knowledge about information retrieval and web search and aims to build accurate and detailed user models. Thus, in the first step, it has to present a definition of a user model and in the next step, the contexts of what type of data is used in user-profiles and the models to represent them have to be provided. The structure of the user profile always is an important issue because of its impact on ranking performance. It is clear that if the algorithms used in the user model are more accurate and robust, the user model and personalized services will result in better efficiency and quality. Therefore, creating an efficient user profile is a challenge. Our motivation is to develop a keyphrase-based profile that operates on documents to improve personalization. These profiles are created using the keyphrase based models on the query log, as long-term, short-term and session-term to consider user interest in different time intervals to compare efficiency. Besides, we conduct comparative research on topic-based user profiles, intending to compare keyphrase-based and topic-based profiles in the personalization process. The results obtained more accuracy in session-based models by 13% in mean reciprocal rank and 14% in normalized discounted cumulative gain than long-based models.

KEYWORDS

Personalized Web Search, Keyphrase Extraction,Short-Term User Profiles, Session-Term User Profiles

1. INTRODUCTION

Personalization using a user profile is investigated in lots of research in recent years. Although personalization can improve the returned results by search engines, it indicates different performance in different time ranges. The structure of the user profile is an essential issue because of its impact on ranking performance. It is clear that if the user profile structure is more robust, the user profile and personalized services result in better efficiency. In keyphrase-based user profiles, a set of keywords can indicate the content or hidden topics of a document. Extracted keyphrases can be used in the fields such as information retrieval (Medelyan, O. et al,

(2009)), summarization (Dredze, M. et al, (2008)), clustering (Hammouda, K.M. et al, (2005)), categorization (Hulth, A. and Megyesi, B.B. (2006)) and indexing (Gutwin, C. et al. (1999)).

Keyphrase extraction methods are divided into supervised and unsupervised. Various unsupervised methods consist of statistical approaches, topic-based clustering groups, and graphed-based methods. In unsupervised methods, it is used several graph-based methods since those are the most commonly used. From supervised methods, we implemented Kea and WINGNUS to compare with unsupervised methods. Unsupervised methods consist of corpus-dependent and corpus-independent. Most of the corpus-independent methods are graph-based. Graph-based keyphrase extraction introduced with TextRank by Mihalcea and Tarau (2004) is applied to rank words based on their associations in the graph. Adjacent keywords in the document form keyphrases. The edges may be weighted, like in SingleRank (Xiaojun, W. and Jianguo, X. (2008)), using the number of co-occurrences as weights. The nodes are scored using some node ranking metric, such as degree centrality or PageRank.

Keywords and keyphrases can be used to enrich the presentation of search results and help users to understand the main topics of a Web page. Using keywords in creating user profiles is explored in various research; however, there is little research on investigating the state of the art methods in keyphrase extraction to create keyphrase based user profiles in the process of personalization. For example, Hulth (2004) described a system along the keywords extraction process for web pages returned by Google search engine. We conduct comparative research on keyphrase-based user profiles to consider keyphrase-based user profiles in the personalization process. These profiles are created using the keyphrase based models on the query log, divided into long-term, short-term and session-term to consider user interest in different time intervals to compare efficiency. We also conduct comparative research on topic-based user profiles to consider keyphrase-based and topic-based profiles in the personalization process.

The organization of the paper is as follows. Section 2 discusses the related work on keyphrase extraction and personalization. The method, including keyphrase based user profiles and temporal based user profiles are presented in Sections 3 and 4. Evaluation methodology and re-ranking process are presented in Section 5. Section 6 concludes the discussion.

2. RELATED WORK

The purpose of the user profiling process is to collect user information based on user needs and interests. There is an overview of the user profiling process in personalization (*Eline, R. (1999))*. The system developed by (Tang, J. et al. (2010)) has divided the user profiling process into three main works that consist of data extraction, data integration, and discovery. In a similar approach by (Stuart M., et al. (2004)) the authors added one additional phase visualization to the process. Zhiheng et al. (Qi Z. et al. (2009)) proposed a multi-interest modeling method to complete a hierarchy structure to improve efficiency. Based on their research, the scheme consists of three modules: user modeling module for managing data, text classification training module, and semantic similar network training module.

A user profiling process can be conducted into the following subprocess, collecting information such as experience, interests, behaviors, pre-processing of collected data, and creating a user model. Finally, the process is completed using applications to provide personalization services. There are many most popular techniques in collecting, representing and creating user models (Gauch, S. and Speretta, M. (2007)).

Besides the user profile process, there is also a range of content information such as click-through data or query log analysis, mouse movement, scrolling, browsing the history, desktop information analysis, dwelling or display time, etc. For example, there are some studies with a focus on creating user profiles by the content of browsing history (Pretschner, A. and Gauch, S. (1999), Kazunari, S. et al. (2004), Chen, L. and Sycara, K. (1998), Asnicar, F. and Tasso, C. (1997)). As discussed by Chen and Sycara (Chen, L. and Sycara, K. (1998)), they have used multiple TFIDF vectors for creating the WebMate system. Pretschner et al. (Pretschner, A. and Gauch, S. (1999)) also used browsing history and user preferences over time. Kazunari et al. (Kazunari, S. et al. (2004)), Karimi et al. (Karimi, S. and Abri, R. (2016)) and Gauch et al. (Asnicar, F. and Tasso, C. (1997)) have built a user profile based on browsing history to re-rank results from a search engine. Also, Gauch et al. (Trajkova, J. and Gauch, S. (2004)) with OBIWAN, have built the user profile by analyzing the browsing history.

There are several modified versions of keyword-based user models. For example, Wei, et al. (2009) and (Chirita, A. et al. (2005)) tried to solve a lack of semantic information of keywords and designed a user profile based on the category knowledge base. In this paper, the authors attempted to improve the user model by combining the keywords and ontology concepts. To employ a keyword-based model for web personalization, it is needed to acquire keywords from documents visited by the users, the same as conducted work by Barla (Barla, M. and Bieliková, M. (2010)). They have developed the JKeyExtractor library to extract keywords from web pages to make user models.

Collecting data from different sources such as web pages, email, and documents are used by Dumais (Dumais, S. (2003)). Matthijs et al. (Matthijs, N. and Radlinski, F. (2011)) collected the data from several components such as the visited pages in the search process and the time opening on the page and Xueping (Xueping, P. et al. (2012)) used the user search result concerning Google Directory.

Social networks can also be used to collect information about users; for example, Jiwei (Jiwei, L. et al. (2014)) has extracted user information from Twitter, Google Plus, and Facebook. In this research, the author used a supervised approach to extract user profiles from Twitter. Kelly et al. (Kelly, D. and Teevan, J. (2003)) have used implicit feedback to recommend a suitable web page to the user. The paper uses some unique implicit feedback such as text tracing and selection, link pointing, printing a page, window movement, and resizing. Extending on this idea, Nazim et al. (Nazim, M. et al. (2012)) also used the social web to collect information to create profiles. So, the first commitment of a system is how to collect data for the user model, and the second is to identify users uniquely.

According to Dumais et al. (Teevan, J. et al. (2010)) clicked item rank and URL by users are the most critical feedback used by researchers for the personalization field. So, in this research, we used the click data collected by users. We used a query log collected by the AOL search engine as a large amount of click-through. Using implicit data sets, we can study many different users' interactions while it will be infeasible with explicit data.

As time is changing, user profiles are also changing. In some studies (Shen, X. et al. (2005), Vu, T. et al. (2017) and Bennett, P. et al. (2012)) used the different vectors to model the user interests as short-term and long-term. In another study by Bennett et al. (Bennett, P. et al. (2012)), the weights of clicked URLs are used to represent how interests change over time. Table 1. summarise some publications of several approaches in the literature as short-term and long-term profiles. Vu, T. et al. (2017) and Momtazi, S. and Lindenberg, F. (2016) created temporal user profiles using the user's clicked documents, and the profile was used for the ranking process. In another research, Vu, T. et al. (2015) also used three typical time scales for building a long-term profile, a daily profile, and a session profile. The results were improved using click entropy and query position in a search session.

Table1. Some publication of several approaches in the literature as short-term and long-term profiles

Publications	Long-term	Short-term
Shen, X. et al. (2005)		*
Kazunari, S. et al. (2004)		*
Chirita, A. et al. (2005)	*	
Kelly, D. and Teevan, J. (2003)	*	
Bennett, P. et al. (2012)	*	*
Vu, T. et al. (2017)	*	*
Vu, T. et al. (2015)	*	*
Nanas et al. (2009)	*	
White et al. (2004)		*
Harvey et al. (2013)	*	

Since there is not complete research on keyphrase based user profiles in different time ranges, we tried to reflect the personalization results for session-term, long-term and short-term user profiles. The proposed models are discussed in the methodology section.

3. KEYPHRASE BASED USER PROFILES

A user profiling process can be conducted into the following subprocess, collecting information like behaviors, pre-processing collected information, and analyzing information to create a user model. Older systems gathered data explicitly from the users, while recent research is focused on implicit methods or behavioral user profiling for collection information. There are two significant matters of collecting information: The content of the data collected and the techniques to collect information. The data's content includes browsing data, bookmarking data, click-through data, while methods to collect data are divided into explicit and implicit methods. In more research, implicit methods are preferred because of the access to large data sources in contrast to explicit methods with a few data. Using an implicit dataset, we can study many different user interactions while infeasible with explicit data. In this paper, we use clicked pages by a user or collected query log to create a user profile.

After collecting relevant information, it needs to preprocess to remove duplicates and clean up the data. Before cleaning the data set, we should uniform the language in available data. Since the query log is obtained from a search engine with different geographies, we only retained the queries in the English language. Then, we cleaned the dataset by only keeping some queries. It means that without clicked results queries are dropped. Then the data is filtered by removing URLs clicked less than one-hundred times.

Besides, the extracted URL or document also needs to be cleaned. To clean them, we extracted clicked pages by users and re-open them using a search engine. Then we did a pre-processing on extracted documents. To do so, we did a collection of word tokenization, sentence segmentation, steaming, stop-word removing, and parsing process provided by the python Natural Language Toolkit (nltk) for English. The candidate keyphrases are stemmed using porter steaming on words to reduce the number of mismatches.

Then we create user models with the collected and preprocessed information from users. As we mentioned before, there are some approaches and techniques to represent and structuring user profiles. In this part, by investigating user profile structure, we create user profiles as keyword-based, keyphrase-based, and topic-based, and in the following, we explain in detail.

We create user profiles using keyphrases extracted of clicked documents by users by keyphrase extraction methods. We use some models such as TF-IDF, RAKE (Stuart, R., et al. (2010)), TopicRank (Bougouin, A. et al. (2013)), TextRank (Mihalcea and Tarau (2004)), Kea(Keyphrase Extraction Algorithm), Witten, IH. et al. (1999) and WINGNUS (Thuy, D. et al. (2010)) to extract keyphrases. Current methods in unsupervised are divided into statistical models and graph-based models. Among statistical models, TF-IDF is used successfully, and we intended to use it as a basic algorithm for the comparison.

In keyphrases extraction algorithms, in the first step, keyphrase candidates are extracted from the document's content, then they are ranked using a weighting mechanism or a machine learning technique as unsupervised or supervised. Finally, the top-K highest weighted candidates are selected. More on unsupervised keyphrase extraction methods, we apply efficient graph-based models such as RAKE, TopicRank, and TextRank. In graph-based models, a graph is built based on words or phrases where the edges' weights are computed using co-occurrence counts (Xiaojun, W. and Jianguo, X. (2008) and Yutaka, M., and Mitsuru, I. (2004)).

The RAKE algorithm proposed by Stuart, R., et al. (2010) is a language-independent method that extracts keywords by analyzing word frequency and its co-occurrence with other words in the text. RAKE focuses on finding multi-word phrases containing frequent words. First, RAKE splits the text into sentences using punctuation signs and generates the candidates. All terms listed in the stop-word file will be treated as phrase boundaries. RAKE is based on the theory that keyphrases frequently contain multiple words but rarely contain standard punctuation or stop words or other words with minimal lexical meaning; however, it won't work in cases where the stop-word is part of the phrase.

The TextRank algorithm proposed by Mihalcea and Tarau (2004) is an algorithm based on PageRank and the same as RAKE extracts key phrases by a co-occurrence graph. In this algorithm, documents are split into sentences and words with specific tags such as noun, prop, and verb. Then a graph of words is created, and the weight for each node is calculated. It collects the influence of each of its connections and determines the new score for the node. In this way, TextRank considers the similarity between each sentence to all other sentences.

The TopicRank algorithm proposed by Bougouin, A. et al. (2013) is an improvement of the TextRank that extracts key phrases using a topical representation of the document. Candidate keyphrases are clustered into topics and used as vertices in a graph. A graph-based ranking model is applied to assign a score to each topic. In the TopicRank approach, ranking topics instead of words is a more straightforward way to identify the set of keyphrases that covers the main topics of a document. Also, the use of a complete graph captures the semantic relations between topics.

Given a complete graph as E=$\{(v_1,v_2)|\forall v_1,v_2 \in V\}$, $v_1 \neq v_2$ of V $\times$ V, Vertices are topics, and the edge between two topics t_i and t_j is weighted according to the strength of their semantic relation. t_i and t_j have a strong semantic relation if their keyphrase candidates often appear close to each other in the document. Therefore, the weight $w_{i,j}$ of their edge is defined as defined in Equations 1 and 2 by Bougouin, A. et al. (2013):

$$w_{i,j} = \sum_{c_i \in t_i} \sum_{c_j \in t_j} dist(c_i, c_j) \qquad (1)$$

$$dist(c_i, c_j) = \sum_{p_i \in pos(c_i)} \sum_{p_j \in pos(c_j)} \frac{1}{|p_i - p_j|} \qquad (2)$$

Where $dist(c_i, c_j)$ is the reciprocal distance between the offset positions of the candidate keyphrases c_i and c_j and $pos c_i$ represents all the offset positions of the candidate keyphrase c_i. After creating a graph, the ranking model TextRank, is used to rank the topics. This ranking model considers a score to topics based on the concept of voting that high scoring topics contribute more to the score of their connected topic t_i as Equation 3.

$$S(t_i) = (1 - \lambda) + \lambda \times \sum_{t_j \in V_i} \frac{w_{i,j} \times S(t_j)}{\sum_{t_k \in V_j} w_{j,k}} \qquad (3)$$

Where v_i are the topics voting for t_i and λ is a damping factor generally defined to 0.85. To complete the benchmark, we evaluate supervised models like Kea and WINGNUS as baseline methods. Kea builds a classifier based on the Bayes Theorem using training documents, and it uses the classifier to extract keyphrases from new documents. In the training and extraction phases, KEA analyzes the input document depending on orthographic boundaries (such as punctuation marks, newlines, etc.) and exploits two features: TF-IDF and the first occurrence of the term. In WINGNUS keyphrase, candidates are simplex nouns and noun phrases detected using a set of POS filtering rules. Keyphrases are then selected using a Naive Bayes classifier with a large set of features including document logical structure information. Finally, extracted keyphrases are stored in the keyphrase based user profiles in the form of XML and in the next step to be applied in the ranking process.

4. TEMPORAL USER PROFILES

There is much research (Vu, T. et al. (2017), Vicente-López, E. et al. (2014)) in creating user models into ranking algorithms in various areas effectively, but more research has a problem that they have ignored that the user interests change over time. As time goes on, the user becomes reluctant to some topics while starting attention to other topics.

Since during a search session, user interests and search intentions are changing, so the long-term and short-term profiles were also discussed (Vu, T. et al. (2017), Bennett, P. et al. (2012)). For example, Vu et al. (Vu, T. et al. (2017)) created temporal user-profiles and used them for the re-ranking process. In similar research, Bennett (Bennett, P. et al. (2012)) built different temporal user-profiles, and the re-ranking results were improved using metrics like click entropy.

Therefore in this research, we consider time in creating user profiles as session-term, short-term and long-term. In this way, to build long-term, short-term and session-term user profiles, we consider user profiles for three month, one month and one session. In this order, we can consider the keyphrase-based user profile as sessions of time.

5.   RE-RANKING PROCESS FOR PERSONALIZATION

Personalization is the task of re-ranking the retrieved document set concerning the user profile. We discuss two ranking methods, including ranking based on the keyword/keyphrase profiles and re-ranking based on the topical user profiles. The purpose is to make a comparison between keyword/keyphrase based models and topic-based models. In this way, ranking approaches based on the topical model are used in our evaluations, as a generic scoring function based on topic models without any personalization using the collected query log.

In this paper, evaluation is made on two datasets to investigate the precision of methodology. We used the AOL data set in the first experiment, including the query log collected from 650k users. To clean the data, we first selected those queries, which resulted in a click on a URL. In the second experiment, to achieve a more dominant result, we used the data set provided by TREC 2014 Session Track. The Session Track consists of 1021 query sessions for 60 different topics along with the clicked result and user id.

5.1 Evaluation Parameters

To evaluate the personalized model, we divided the dataset into 95% for training and the last 5% of queries for testing. The performances of personalization are evaluated in terms of success at rank k (S@k), the mean reciprocal rank (MRR), and discounted cumulative gain (DCG) up to rank 10 when a maximum of 10 key phrases are extracted (k = 10) to create user profiles. To compare with a baseline method, we use a non-personalized model as a generic ranking without personalization created based on Latent Dirichlet Allocation (LDA) topic model. Building on the same framework introduced by Harvey, M. et al., (2013) and the results presented by Abri S. et al. (2020), documents are ranked with respect to the LDA model *P(d/q)* called NonPTM (Non-Personalized Topic Model) in conducted work by Abri, S. et al. (2020) as Equation 6.

$$NonPTM(d, q) = P(d|q) \propto P(d)P(q|d) \quad (4)$$
$$= P(d) \prod_{w \in q} P(w|d) \quad (5)$$
$$= P(d) \prod_{w \in q} \sum_{z} P(w|z)P(z|d) \quad (6)$$

Where *P(d)* is the prior document probability and *z* is the topic latent variable estimated using LDA. *P(w/z)* and *P(z/d)* are obtained from the LDA topic model. The number of topics used for LDA is considered 40 topics in the AOL dataset and 30 topics for the TREC2014 dataset.

We compare the performance of the proposed method with the state-of-the-art algorithms using an open-source python-based keyphrase extraction toolkit, called pke. The pke toolkit is open source and available at https://github.com/boudinfl/pke. To investigate the effect of various parameters in rake, we explored these parameters depending on the text. It has shown in Figure 1. the parameters that performed best on the current dataset with at most three keywords in each phrase.
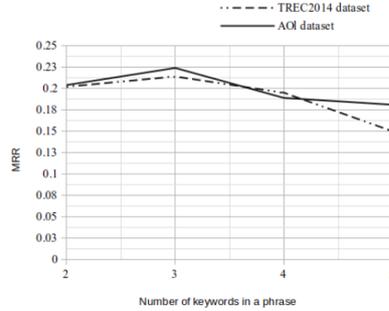
Figure 1. The MRR changes in phrase length in Rake

5.2 Evaluation Result

Re-ranking documents for a new query using a keyphrase based user profile, we measure the similarity between the extracted keyphrases of each document in the query log and the keyphrase based user profile for each user by cosine similarity. The cosine similarity is useful because of measures changing between similar documents regardless of the document's size. It can be measured using Equation 7.

$$cos(\theta) = \frac{\vec{a}.\vec{b}}{\parallel \vec{a} \parallel \parallel \vec{b} \parallel} = \frac{\Sigma a_i b_i}{\sqrt{\Sigma a_i^2}\sqrt{\Sigma b_i^2}} \qquad (7)$$

Where $\vec{a}.\vec{b} = \Sigma a_i.b_i = a_1 b_1 + a_2 b_2 + ... + a_n b_n$ is the dot product of the two vectors. For the evaluation re-ranking process using keyphrase based user profile, we tried a re-ranking algorithm on the data set and the results are presented. Besides, we need to consider time in creating user profiles. To build long-term and short-term user profiles, we considered user profiles for three and one month and for session-term profiles, one session is considered.

To evaluate the proposed model, we provided two experiments with different datasets. In the first experiment, we used the AOL query log data set as a big resource of explicit query log data. In both of the data sets, clicked documents are a reference in building user profiles. We will present the result of the experiments in the following. The models are called personalized keyphrase extraction(PKM), as Abri, et al.(2020) conducted.

Tables 2. and 3. report the MRR, S@1, S@10, and nDCG@10 scores for the Non-personalized topic model(Non-PTM) and personalized keyphrase extraction(PKM) models in AOL and Session Track 2014 datasets. In Table 2., the columns represent the used methods, including Non-PTM, TextRank, Rake, and TF-IDF. To consider the time dimension in the process of user profiling, we separated profiles into long-term, short-term and session-term user profiles. As depicted in Table 2., the best results are obtained with the Non-PTM and TextRank methods for all time dimensions. In Personalization using the TextRank(PKM-TextRank) model, the MRR is about 32% for the short-term, 34% for the session-term and 26% for the long-term while it needs more time to run rather than Rake and Tfidf. It can be seen that although the results between short-term and session-term in all methods are close, the results based on the session-term profiles are much more efficient than the short-term and long-term profiles.

Table 2. Ranking Performance of Non-PTM and personalized keyphrase based models on the AOL data set

	Long-Term				Short-Term				Session-Term			
	S@1	S@10	MRR@10	DCG@10	S@1	S@10	MRR@10	DCG@10	S@1	S@10	MRR@10	DCG@10
Non-PTM	0.205	0.371	0.267	0.289	0.282	0.436	0.330	0.360	0.308	0.459	0.360	0.387
PKM-TextRank	0.201	0.353	0.259	0.290	0.280	0.410	0.325	0.353	0.290	0.419	0.339	0.379
PKM-Rake	0.181	0.310	0.224	0.257	0.250	0.368	0.297	0.325	0.275	0.394	0.318	0.346
PKM-Tfidf	0.172	0.308	0.218	0.249	0.255	0.358	0.296	0.326	0.284	388	0.326	0.357

Table 3. represents the ranking score when using non-personalization and personalization based on keyphrase-based profiles using a set of supervised and unsupervised approaches. Among supervised approaches, Kea and WINGNUS are used as basic approaches while the best results are obtained using created models using these methods. In unsupervised methods, we also tried TextRank, Rake, and TopicRank as graph-based methods and TF-IDF as a feature-based unsupervised method. As shown in Table 3., PKM-Tfidf resulted in the best performance in nDCG@10 among all approaches. As results show, supervised approaches (Kea and WINGNUS) have provided more successful results in all time dimensions; however, those are needed for training data. The bias towards the domain in which they are trained is a critical issue. As obtained results of AOL dataset, in Session Track 2014 also the tendency in performance in session-based profiles is more than short-term and long-term profiles in all personalized models.

Table 3. Ranking Performance of the Non-PTM and personalized keyphrase based models on the Session Track 2014 using nDCG metric

	Long-Term				Short-Term				Session-Term			
	S@1	S@10	MRR@10	DCG@10	S@1	S@10	MRR@10	DCG@10	S@1	S@10	MRR@10	DCG@10
Non-PTM	0.216	0.340	0.262	0.291	0.315	0.422	0.356	0.388	0.344	0.451	0.380	0.426
PKM-TextRank	0.197	0.318	0.238	0.263	0.304	0.322	0.350	0.371	0.326	0.350	0.379	0.395
PKM-Rake	0.175	0.303	0.214	0.249	0.278	0.419	0.321	0.358	0.293	0.429	0.340	0.361
PKM-Tfidf	0.213	0.335	0.257	0.280	0.321	0.440	0.366	0.389	0.340	0.462	0.385	0.412
PKM-TopicRank	0.204	0.326	0.240	0.272	0.315	0.438	0.354	0.376	0.333	0.460	0.379	0.398
PKM-Kea	0.257	0.390	0.304	0.348	0.369	0.497	0.410	0.454	0.390	0.517	0.437	0.480
PKM-WINGNUS	0.260	0.395	0.310	0.352	0.378	0.500	0.428	0.468	0.404	0.529	0.460	0.491

6. CONCLUSION

In this paper, to investigate keyphrase-based user profiles in the personalized web search, it is considered how integration between keyphrase extraction and personalization by state-of-the-art approaches. The personalization methods are created using supervised and unsupervised keyphrase extraction methods. The profiles are created using the keyphrase based models on the query log, as long-term, short-term and session-term to consider user interest in different time intervals to compare efficiency. For evaluation of the model, keyphrase-based user profiles using the re-ranking algorithms are applied using different datasets. The results showed more accuracy in session-based models by 13% in mean reciprocal rank and 14% in normalized discounted cumulative gain than long-based models.

REFERENCES

Abri, S. and Abri, R. 2020. Providing a personalization model based on fuzzy topic modeling. *Arab J Sci Eng (2020)*.

Abri, S. et al. 2020. Group-based Personalization Using Topical User Profile. *UMAP '20 Adjunct Proceedings of the 28th* ACM Conference on User Modeling, Adaptation and Personalization. July 2020, pp 181–186.

Abri S. et al. 2020. Personalized Web Search Using Key Phrase-Based User Profiles. *In the 19th International Conference on WWW/Internet (ICWI 2020).*

Abri S. et al. 2020. A Classification on Different Aspects of User Modelling in Personalized Web Search, *In 4th International Conference on Natural Language Processing and Information Retrieval.*

Abri S. et al. 2020. Estimating Personalization Using Topical User Profile, *In 12th International Conference on Knowledge Discovery and Information Retrieval.*

Asnicar, F. and Tasso, C. 1997. ifweb: a prototype of user model based intelligent agent for documentation filtering and navigation in the world wide web. *In Proceedings of the 6th Intl. Conf. On User Modeling.*

Barla, M. and Bieliková, M. 2010. Ordinary Web Pages as a Source for Metadata Acquisition for Open Corpus User Modeling. *In: Proceedings of the IADIS International Conference WWW/Internet 2010*, pp. 227-233.

Bennett, P. et al. 2012. Modeling the impact of short- and long-term behavior on search personalization. *In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pp. 185-194.

Bougouin, A. et al. 2013. Topicrank: Graph-based topic ranking for keyphrase extraction. *In Proceedings of IJCNLP*, pp. 543–551.

Chen, L. and Sycara, K. 1998. Webmate: A personal agent for browsing and searching. *In Proceedings of the 2nd International Conference on Autonomous Agents*, pages 132–139.

Chirita, A. et al. 2005. Using ODP metadata to personalize search. *In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp 178-185.

Chris, M. et al. 2015. Corpus- independent Generic Keyphrase Extraction Using Word Embedding Vectors.

Dredze, M. et al, 2008. Generating summary keywords for emails using topics. *In Proceedings of the 13th International Conference on Intelligent User Interfaces*, pp. 199–206.

Dumais, S. 2003. Sigir 2003 workshop report: Implicit measure of user interests and preferences. S*IGIR Forum*, 37(2).

Eline, R. 1999. Users are individuals: individualizing user models. *International Journal of Human-Computer Studies, vol. 51, No. 2, pp. 323-338.*

Gauch, S. and Speretta, M. 2007. Aravind Chandramouli, and Alessandro Micarelli. User profiles for personalized information access. *In The Adaptive Web 2007,* pages 54–89. Springer.

Gutwin, C. et al. 1999. Improving browsing in digital libraries with keyphrase indexes. *Decision Support Systems,* Vol. 27, No. (1-2), pp 81–104.

Hammouda, K.M. et al, 2005. CorePhrase: Keyphrase extraction for document clustering. *In Proceedings of the 4th International Conference on Machine Learning and Data Mining in Pattern Recognition*, pp. 265–274.

Harvey, M. et al, 2013. Building user profiles from topic models for personalised search. *In Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pp. 2309-2314.

Hulth, A., 2004. Combining machine learning and natural language processing for automatic keyword extraction.

Hulth, A. and Megyesi, B.B. 2006. A study on automatically extracted keywords in text categorization. *In Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, pp. 537–544.

Jiwei, L. et al. 2014. Weakly supervised user profile extraction from twitter. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*.

Karimi, S. and Abri, R. 2016. Improvement of Semantic Search Results with Providing an Updatable Dynamic User Model, *In International Journal of Computer Applications(IJCA),* Vol. 155 , No. 4, pp. 7-14.

Kazunari, S. et al. 2004. Adaptive web search based on user profile constructed without any effort from users. *In Proceedings 13th International Conference on World Wide Web*, pages 675–684.

Kelly, D. and Teevan, J. 2003. Implicit feedback for inferring user preference: a bibliography. *ACM SIGIR Forum*, 37(2):18–28.

Matthijs, N. and Radlinski, F. 2011. Personalizing web search using long term browsing history. *In Proceedings of the fourth ACM international conference on Web search and data mining*, pp 25-34.

Medelyan, O. et al, 2009. Human-competitive tagging using automatic keyphrase extraction. *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 1318–1327.

Mihalcea, R. and Tarau P., 2004. Textrank: Bringing order into texts. *In Proceedings of EMNLP 2004*, pp. 404–411.

Momtazi, S. and Lindenberg, F. 2016. Generating query suggestions by exploiting latent semantics in query logs. *Journal of Information Science*, pp 437-448.

Nazim, M. et al. 2012. Construction of semantic user profile for personalized web search. *In Proceedings of the 4th international conference on Computational Collective Intelligence: technologies and applications,* pages 99–108.

Pretschner, A. and Gauch, S. 1999. Ontology based personalized search. *In Proceeding of 11th IEEE International Conference on Tools with Artificial Intelligence*, pages 391–398. IEEE.

Qi Z. et al. 2009. User multi-interest modeling based on semantic similar network in personalized information retrieval. *Issues in Information Systems,* 10(2):257–266.

Shen, X. et al. 2005. Implicit user modeling for personalized search. *In Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM 2005, pages 824–831.

Stuart M., et al. 2004. Ontological user profiling in recommender systems. *ACM Transactions on Information Systems,* 22(1):54– 88.

Stuart, R., et al. 2010. Automatic keyword extraction from individual documents. *Text Mining: Applications and Theory*, pp. 1–20.

Tang, J. et al. 2010. A Combination Approach to Web User Profiling. *In: ACM Transactions on Knowledge Discovery from Data*, Vol. 5, No. 1.

Teevan, J. et al. 2010. Potential for personalization. *In ACM Transactions on Computer-Human Interaction (TOCHI) TOCHI*, pages 1–31. ACM.

Trajkova, J. and Gauch, S. 2004. Improving ontology-based user profiles. *In Proceedings of the RIAO '04 Coupling approaches, coupling media and coupling languages for information retrieval*, pages 380–390.

Thuy, D. et al. 2010. Wingnus: Keyphrase extraction utilizing document logical structure. *In Proceedings of SemEval*, pp. 166–169.

Vicente-López, E. et al. 2014. Personalization of Parliamentary Document Retrieval using different User Profiles. *In Proceedings of the 2nd International Workshop on Personalization in eGovernment Services and Applications.*

Vu, T. et al. 2015. Temporal latent topic user profiles for search personalisation. *In: ECIR:37th European Conference on IR Research*, pp. 605-616.

Vu, T. et al. 2017. Personalised Query Suggestion for Intranet Search with Temporal User Profiling. *In Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pp. 265-268.

Witten, IH. et al. 1999. Kea: practical automatic keyphrase extraction. *In: ACM conference on digital libraries*. New York, NY, USA, pp. 254-255.

Xiaojun, W. and Jianguo, X. 2008. Single Document Keyphrase Extraction Using Neighborhood Knowledge. *In AAAI*, Vol. 8. pp 855–860.

Xueping, P. et al. 2012. Personalized web search using clickthrough data and web page rating. *Journal of Computers*, 7(10):2578–2584.

Yutaka, M. and Mitsuru, I. 2004. Keyword extraction from a single document using word co-occurrence statistical information. *International Journal on Artificial Intelligence Tools,* vol. 13, No. 1, pp. 157–169.