A novel approach for modeling user's short-term interests, based on user queries

Albena Turnina

Information Technology Department, Sofia University
Sofia, Bulgaria

Abstract

In this paper is presented a novel approach for modeling user short-term interests, based on user queries. The proposed user model is represented by a weighted semantic network, composed of nodes and arcs. This semantic network can be used to express relations between query terms submitted by a user in his searches. Our approach is rooted on the idea that there are relations between topics of user's interests, which can be measured and used to provide a search context. We propose an approach for modeling user's interests, based on data taken from his previous queries. We aim to identify relations between topics of user's interests in a set of successive queries. The search personalization features of ShareTec digital library are also presented and the implementation of the proposed model is outlined.

Keywords: *User profile, short-term interests, adaptive query, search personalization.*

1. Introduction

In this work we propose a novel approach for modeling user short-term interests, based on data, taken from user's queries, intended to facilitate a personalized search. The proposed user model is represented by a weighted semantic network, composed of nodes and arcs. This semantic network can be used to express relations between query terms, inputted by a user in his searches. To the best of our knowledge, the same model does not exist, although some of the ideas behind the model have already been realized in existing systems for a personalized search. Our approach is based on the idea that there are relations between topics of user's interests, which can be measured and used to provide a search context. We propose an approach for user modeling, based on data taken from user's previous queries and an identification of relations between topics of interests in successive queries. The proposed model could be able to track dynamic changes in user's interests and to stay updated, representing actual topics of user's interests. We aim to investigate the possibilities for an achieving search personalization using solely the proposed model, which takes into account only usage data, taken implicitly from the queries. The intuition behind our model is that there are relations between different topics of user's interests, which are specific to a particular user and these relations are possible to be detected and used for delivery of personalized search results.

2. Theoretical background

There are three paradigms for information access to a web content in a hypertext environment - searching by browsing, searching by means of queries submitted to search engines and through recommendation systems [18]. The recommendation systems offer topics, analyzing what users with similar interests were chosen in the past. In searching by browsing, the users explore Web pages, one by one by following hyperlinks. The browsing is not a convenient way to find a certain and specific information. In this study we focus on the queries and investigate the approaches to provide a search personalization by adapting the queries.

The two main kind of personalization systems are Content-based systems and Collaborative-filtering systems. The personalization in the Content-based systems is based on the similarity between a user and the documents in a collection. The Collaborative-filtering systems exploit the idea that the users with similar interests are likely to prefer the same resources. There are other personalization techniques and systems, which are based on rules or explore demographic data as well as hybrid systems which combine several techniques together. A short overview of the existing personalization strategies and example systems is presented below.

- Content-based systems – encompasses systems which track user`s browsing behavior and recommend topics similar to those previously chosen by the user. This technique has some shortcomings, such as "over-specialization" and the lack of relations between topics [25], [20]. Example systems are WebWatcher and Letizia.

- Collaborative-filtering systems – includes systems that use cumulative experience of a group of users in order to facilitate searching experience of a certain user. The idea behind this strategy is that users with similar behavior/preferences are likely to have similar interests. Employed algorithms investigate similarities between the users and gives recommendations according to preferences of the most similar neighbor. Collaborative filtering approach is used in popular systems such as Yahoo, Excite, Microsoft Network and Amazon.com.

- Demographic-based systems – encompasses systems that group users on the base of their demographic similarities. In this kind of systems recommendations are given according to demographic characteristics of users, such as age and gender [23].

- Rule-based systems - in these systems the recommendation are given to the users according to their answers to a set of predefined questions [35]. Example system is Broadvision.

- Hybrid systems - combines two or more personalization strategies and approaches [2].

The search personalization techniques are used in systems that provide individualized collections of pages to users. This personalization is based on a user model which presents the user’s interests and search activities [18]. According to (Micarelli et al) the search personalization approaches are subdivided into several major types, which are as follows: Current Context, Search History, Rich User Models, Collaborative approaches, Result Clustering and Hypertextual Data [18].

The personalized information retrieval (PIR) systems can be classified according to different criteria, the main of which are: the type of the system and the personalization approach. The type of the system is based on the application domain or on the type of the service, provided by the system, like Web search systems, Multilanguage search systems, personalized news feeds, e-learning, etc. The personalization approaches are related to the way in which personalization is realized. The personalization in PIR systems can be realized by a query adaptation, by personalization of search results, or both. In multilingual systems the personalization process can include additional steps, needed for translation of the query and results in different languages [21], [22]. According to the chosen approach, the various personalization techniques can be implemented [14].

The query adaptation can be realized by modification of the query, by relaxation of the query or by substitution of the original query with one or more adapted queries. The modification of the query can be performed by wide variety of techniques which encompass: the substitution of the query terms with terms or concepts, taken from a reference vocabulary; the expansion of the query by terms, taken from a user profile; the change of weights of terms and/or relations between them and using the methods of pseudo-relevance feedback and relevance feedback. The modification of the query can be performed explicitly by the user as well.

The expansion terms can be derived from the user profile. These expansion terms are representative for the user’s preferences and serve to give a search context. The exact number of the expansion terms can be predefined or can be selected dynamically. (Chirita et al., 2007) argue that the number of the expansion terms can be tuned according to features of the query, such as the query length, scope of the query, etc. [10]. The process of a dynamic selection of the expansion terms is known as a “selective query expansion”.

The main advantage of the query adaptation approach is the lack of additional processing required in the other two approaches. However, the query adaptation is unlikely to influence significantly the returned results [14].

The personalization of the search results is the process of a selection of the most relevant results and their ranking according to a certain user, a community of users or all users of a system. This personalization can be realized by a number of techniques, which are as follows: pre-ordering of results, filtering of results and result scoring [18].

The techniques of pre-ordering of results and the filtering of results are performed after the initial results list has been extracted from the system. They are realized as an additional processing step over the retrieved set of documents. In contrast, the technique of a result scoring takes place in the process of initial selection of documents where the adaptivity is implemented as an integral part of the result scoring function.

The essential part of each personalization systems is the user model [7], [13], [19]. The user model represents the user’s characteristics and the user’s behavior in a system. There exist different approaches for modeling of a user according to purpose and scope of the system. For example, Web personalization systems use tracking algorithms to follow the user in his surfing, whereas Adaptive hypermedia systems build a user profile, based on the user interaction with the system. The user can be modeled according to his demographic characteristics and specific features such as: level of expertise in a particular domain, cognitive abilities, visual characteristics and etc. (Brusilovsky, 2001) classification of user's characteristics encompass the knowledge, purpose, background, interests, environment and experience with hypertext of a user [8]. (Gasparini et al., 2011) present a new approach for modeling of a user in a domain of e-learning, which takes into account contextual user aspects, such as technological, educational, personal, and cultural characteristics [12]. Related research on the implications of the cultural user's characteristics in the design of user interfaces are conducted by [9], [24].

The distinction between a user profile and a usage profile can be found in different sources. The former is related to user's characteristics and the usage profile is related to the behavior and activities of a user in a system. However, user profile can be found as a representative for both – a user and a usage profile. The user profile is considered to be essential and the most common part of the personalization systems. There are numbers of different techniques for the symbolic representation and the construction of the profile. The degree of a complexity of a user profile depends on purpose of the system. It varies from a simple explicit questionnaire to a complex dynamic structure, containing both explicit and implicit information. The important characteristic of a sophisticated user profile is its ability to be changed dynamically, reflecting the user changes and his shifts of interests. The user profile can take part in a personalization process of the Information Retrieval systems in three different ways - when takes place in the retrieval process, when is used for pre-ordering of the search results and when is used for a query adaptation [14].

The representation of users and documents can be based on various methods and techniques taken from different fields like Information Retrieval or Artificial intelligence. The vector-space model is popular and largely used model for documents and user profiles representation. In this model any document or profile is represented as a set of keywords or as an n-dimensional vector. This representation serves as a basis for comparison between a document and a profile and allows measurement of degree of similarity between them. The document is considered to be relevant to a user if the degree of similarity is over a predefined threshold value. The popularity of the vector-space model is mainly due to its simplicity and proven effectiveness. However, the vector-space model has some shortcomings such as potential loss of information when documents pass linguistic processing in which the keywords are separated. Because of this processing the embedded meaning in sentences and phrases can be lost. The disadvantage of the profile, constructed by means of keywords is that it requires a large amount of user's feedback. This feedback is needed for the selection of the exact words which present a given user. A largely known method for calculation of probability that the document is relevant to a user is by means of Bayesian probabilistic classifier. Other user profile representation techniques include semantic networks, associative concepts or rules although the last serve mainly in the field of a Web log mining [14].

The user profile, based on semantic network/s is composed of nodes (nod) and links (arcs) between them. The semantic network is a graphical notation, used for knowledge representation, based on interconnected nodes and arcs. The semantic networks can be used to represent the knowledge over which the inference can be done. According to (Sowa, 1992) the most popular semantic network types are as follows: Definitional networks, Assertional networks, Implicational networks, Executable networks, Learning networks and Hybrid networks [28]. The Learning network type is of a particular interest for this study because of its ability to extend itself when new knowledge is acquired. This new knowledge is able to modify the existing semantic network, to add new or to remove the existing nodes and links and to change their assigned weights. The semantic network can serves to represent symbolically the relations between terms and/or concepts and their mutual occurrences in texts or documents. Varies weighting schemas can be implemented to measure weights of nodes and links between them.

The semantic network as well as the ontology can be used to model the relation between a word and a concept. This is of a particular importance when the documents are expressed in natural language. The mapping of a word to a concept can be accomplished by means of reference vocabulary such as WordNet, by learning mechanism or manually.

The user profiles, based on semantic networks can have different level of sophistication and can encompass one or multiple networks, each one having its owl level of complexity. Amongst the systems, which implement a user profile, represented as semantic network/s are: WIFS, InfoWeb, SiteIF, ifWeb and PIN. The user profiles in these systems have different level of complexity and sophistication. In more complicated implementations, concepts are represented by nodes called Planets and the words are represented by nodes called Satellites, connected to Planets. In WIFS system the individual semantic network for each user's interest is deployed.

The user model can be represented by specially designed user ontology or by weighted domain ontology. A widely deployed ontology user model is an overlay model in which the user is modeled by domain ontology and his knowledge is represented as a part of an expert knowledge. The ontology allows inference over collected facts and drawn of new facts [11]. An interesting approach is the construction of a user profile, based on domain ontology with weights, assigned to concepts according to user's interests. These weights are updated through spreading activation algorithm [27]. In [15] the authors present a set of statistical methods for learning user ontology from domain ontology by spreading activation algorithm.

The user profile can be constructed by means of a diverse set of techniques emerged from various fields such as machine learning, Information retrieval and Artificial intelligence. Nevertheless not very popular, there are approaches based on the use of genetic algorithms and neural networks.

In this work we present a user model, intended to provide search personalization, based only on usage data taken from user's previous queries. We do not argue that this user model can serve as only source of personalization in a particular system, rather it is intended to complement long-term profiles, modeling user's characteristics. The details of our approach and characteristics of the proposed model are outlined in Section 4 of this work. Section 3 provides brief overview of existing systems, related to our work. The details on the planned implementation and brief presentation of the digital library ShareTec are shown in Section 5. Finally in Section 6 the conclusion and future directions are presented.

3. Overview of the existing systems

Our main focus of interest is directed towards personalization approaches which tackle the problem of a query adaptation. The central part of our work is the development of a novel approach for modeling user's short-term interests. Therefore we have been investigating a number of systems, which explore the problem of a user modeling. In this brief overview the systems and approaches, related to our work are presented. Firstly, we explore the systems that model the user's interests by means of semantic networks. Secondly, we focus on the systems that adapt the original user query in order to provide personalization.

IfWeb is a personalized information retrieval system in which a user profile is represented by a semantic network. The profile is constructed by extracted words which have the highest weight of a set of documents. Each of the extracted word in their approach is used to create a single node in the semantic network. Nodes are connected by links when the words they represent appear together in documents [1]. This approach is extended by SiteIF system where the extracted words are linked to concepts, taken from a dictionary WordNet. The similar approach can be observed in PIN system where the words, extracted from documents are nouns and concepts are learned by means of neural networks. In SiteIF system the user model is represented by weighted semantic network in which nodes are connected by links that have different weights. This system uses WordNet for finding semantic similarity (synonymy) between words [31]. InfoWeb is a filtering system for retrieval of documents in digital libraries. In this system the profile is represented by a semantic network which models the long-term interests of a user. Initially, the semantic network is represented by a collection of unconnected nodes as each node is a separate concept. These nodes called Planets contain single, weighted term that is representative for a given concept. In the process of gathering information about a particular user, his profile is enriched with weighted words, mapped to different concepts. Words are contained in nodes, called Satellites which are connected to conceptual nodes - Planets. The conceptual nodes – Planets, can be linked to each other's. The user model in system WIFS is partially based on semantic networks. The specific of WIFS system is explicitly provided a set of topics of interests by a user during his initial registration. The interests can be implicitly extended further by the system as a new data about a user is gathered in automatic manner. The set of topics, that a given user is interested in, are used to associate the user with a set of stereotypes. Stereotypes themselves are defined by people and experts to describe the knowledge domain of computer science. In this system the user profile models different aspects of the knowledge of the user, which includes personal information and a list of active stereotypes which are associated with the user. Each user's interest is modeled by a separate semantic network. Each semantic network contains a main node called Planet and many Satellites nodes connected to the main node [19].

The conceptual nodes in InfoWeb system are created by the technique of explicit feedback, whereas in WIFS system they are created by human experts. In MiSearch the user profile is represented by means of two alternative models, both based on a vector-space model. The aim of both models is to describe the long-term interests of the user. The first model consists of concepts derived from user's queries, and the second one from concepts derived from snippets of the selected (clicked) documents by the user. The models contain numbers of vectors. Each of them is representative for a particular user's interest. Both models use the ODP hierarchy to categorize the concepts. Mapping of retrieved documents to ODP categories is done by means of a text classifier [29]. (Zhou et al.2012) present a system where the user model is constructed by terms, extracted from the user's tags and bookmarks. They create a statistical model based on these bookmarks to represent the different topics. Their model can be used for an identification of topics in documents, based on user's tags and bookmarks. By means of this model, the most relevant terms for a particular user, can be identified and used to enrich and expand the user query to the system [36]. The query adaptation can be performed on the base of rules. (Koutrika and Ioannidis, 2004) proposed a method, based on rules for rewriting the query by means of which the movies database can by queried. In their approach, the original query is replaced by a number of queries as the process is managed by a set of rules, based on the individual preferences of the user. They propose to connect the different queries through a disjunctive logical operator "OR". For example, if a user likes a particular film genre and make a search for movies issued in a particular year, the system will issue the query that searches the specified film genre in addition to the original search [16]. (Stamou and Ntoulas, 2009) proposed a system in which personalization is performed by re-ordering of search results returned by Google. In their approach they model not only the long-term interests of a user but also short-term interests, taken from a current user query [30]. The long-time interests are modeled on the base of the information gathered from the past user queries as well as from the selected by a user results in the list of returned results. Document and query terms are linked to the concepts, derived from reference ontology. The current user's interests are identified by a current query submitted to the system before the results are presented. After sending the query, the system tries to identify whether that same query was issued in the system before. In such case, the results that were drawn before are presented to a user as a result of a current query. Otherwise, the system determines the degree of similarity between the concepts of the current query and the documents previously classified under different concepts of reference ontology. Thus, the authors model both long-term and short-term interests of a user and by combining both, calculates the user's interests. The relevance of the retrieved documents is a function of user's interests and is calculated as the sum of long-term interests, short-term interests and the weight of the document [30].

According to (Shen et al., 2005) the two main aspects of a personalized search are user's interests and a context of the search (understood as a disambiguation of query). In their system UCAIR they focus on modeling short-term interests of a user, through an approach called eager implicit feedback. In this approach, the current context of a query is output from the previous query within the same session, as well as from results from a previous query, identified as relevant by the user (clicked). In order to determine whether two consecutive queries are related, the system performs two searches - one for the previous query and one for the current query. The lists of results derived from both searches are then compared to determine the similarity between them. If the queries are semantically related, the current query is expanded with terms, taken from the previous query. The results retrieved due to the adapted query, are pre-ordered according to a user model. The user model updates itself dynamically when the user clicks on documents, presented in the result list [26].

A different approach for adapting query and results is presented by (Liu et al., 2004). They propose a system in which the user model is built as a vector consisted of conceptual terms. The conceptual interests are recorded on the base of Google Directory, which in own turn is based on ODP. In their approach, the query adaptation is performed by specifying the category of the query. Thus, the system is trying to identify the concepts, related to the query in order to provide the necessary context of the search [17].

4. Our approach – a short-term user model

4.1 Introduction

In this work is proposed a novel approach for modeling short-term user's interests, needed to provide a search personalization. Through the proposed model we can express in a formal way the relations between topics of user's interests, taken from a data derived from a user searches. Our aim is to provide a search personalization, based on a query adaptation, performed by enrichment of the current query with expansion terms taken from a user profile. The expansion terms are selected according to the weights of links between them and the current query terms as well as to their own weights. These weights serve to represent the importance of the given topics and significance of relations between the different topics to a particular user. The weights are assigned to nodes and links in the user profile according to proposed weighting schema. We argue that keywords (query terms) used by a user in his searches, reflect his information needs and interests. The user in his search activities in the system submits multitude of queries, some of which are related to each other. Each of these quires can be a specification of the previous one. It could be a series of queries that taken together express a user's search intention. Therefore, the discovery of such related terms in a series of queries, submitted by a user is essential. In our model we try to identify the query terms, which often appear together in the user searches, or often appear in the consecutive queries and to recognize the connected topics for a particular user. We believe that such relations exist and can reflect trends in user's interests, which can be useful for providing personalized search results. In our work we have been inspired by so-called method of "previous query" used in Google search engine, the meaning of which is that the previous queries issued by the user influence the search results of the current query. For example, if the user searches for information about the trip, but before that he had shown an interest in a particular country, the system will offer results in which the terms trip and that country appear together [34].

In our model we want to reflect the importance of some terms for the user which appear often in his queries as well as significance of connections between some terms. In order to measure relatedness of terms we use the term "distance" which reflects the distance of terms in a series of a predefined number of consecutive user queries. This "distance" serves as a way to assign weights to links between concept nodes in our model. These weights of the links will be used for determination of the terms, used for the expansion of the current query. The model is built by a predefined number of successive user queries as their exact number is still a subject of investigation. The proposed model is intended to be changed dynamically with acquisition of new knowledge in order to be able to represent the changeable user's information needs. The model is intended to operate in systems in which the searching is performed by means of free words. It has been shown that this is the easiest and most preferred way for users to perform a search, although it can present ambiguity. Through the presented model we proposed one solution to the problem of an identification of a search context which can help in avoiding ambiguity and can serve as an enabler for the search personalization.

4.2 The characteristics of the model

The proposed model is represented by a weighted semantic network composed of two kinds of nodes – one for presenting the concepts (concept nodes) and the other for presenting the query terms (term node). The term nodes can be associated to concept nodes according to semantic similarity between them. The concept nodes in the model can be related by weighted links. The concept nodes and the term nodes have their own weight, assign according to proposed methodology. The presented model has some common features with the user models of InfoWeb and WIFS systems but has its own specific characteristics as well. These characteristics are related to the way in which relatedness between the query terms are modeled and the weights to nodes and links are assigned. The weights of the nodes and the links in the proposed model can be changed dynamically, according to a user searches.

The query terms pass a linguistic processing - stemming before been added to a model as the term nodes. In order to associate the term node to the concept node, some vocabulary or reference ontology has to be used. One of the most viable solutions is WordNet vocabulary. WordNet is a collection of 100,000 words, nearly 80.000 organized in semantic clusters (synsets). But according to a certain implementation other vocabulary can be used as well. In the use case, presented below we propose to integrate the model in ShareTec digital library. In this particular case the education ontology TEO will be used as reference ontology in order to associate the term nodes to concepts. The weight of a term node reflects the frequency of occurrence of a term in a series of user queries. The weight of a concept node is obtained as a sum of weights of term nodes associated to it. Thus, the weight of one concept node is proportional to the number of the term nodes associated to it and their frequency in user queries. This weight reflects the relative importance of that concept for a particular user. At the same time, the weight of a concept node and the weight of a term node are inversely proportional to the time past since the last usage of a given term in the user queries. Thus, the weights of the nodes reflect the actuality of the user's interests. The weights of the links reflect the degree of relatedness of the connected concept nodes and are representative for the particular user's connected interests. The weight of a link is cumulative value which sums multiple occurrences together of terms in the same query, in the previous query or in the predefined number of queries. The connection/link between the concept nodes in the model is made when the term nodes associated to them are related in the user queries. This relation is measured as it is set out below.

After the user inputs a new query, the system first checks whether the terms, contained in the query, already exist in the user profile. If a particular term is new to a profile it is added as a term node and is associated to the most suitable concept node. At the same time the default weight is set to the term node and the weight of the respective concept node is increased. If the term node has already existed in the profile its weight as well as the weight of the concept node, to which it is associated, are increased with a default value. When two or more terms occur together in the same query, the system checks whether there is a link between the concept nodes, to which the terms are associated. If the concept nodes are not linked together in a profile, the new link is created and the default weight is set to it. If the link has already existed, its weight is increased with the predefined value. At the same time, the links are created (if not existed) between the concept nodes of the current query terms and the concept nodes of the previous query terms. But the weights of these links will be lower than the weights of links between concept nodes of a current query terms. This process can recursively be done to the pre-defined number of previous queries. In this process, the more distant backward the query is the lower weights of links between its concept nodes and current query concept nodes will be. The intuition behind the model is that the terms occurring together in the same query are the most semantically related and when taken together, they express the information needs of the user of the system. That is why we propose to assign the higher weights to the links between their concept nodes.

The current query often happens to be a specialization of the previous query and this is the reason why we also assign the weights to a links between the concept nodes of the terms in these queries but they will be lower. And so on as we proceed with a predefined number of previous queries. It appears that when a user inputs a query, the terms contained in it, will be related to each other and to the terms in a given number of previous queries. We propose the following: to assign the value of one (1.00) to a weights of the links between concept nodes of terms that appear in the same query; the value of 0.90 to the links between terms in consecutive queries, the value of 0.80 to the link between concept nodes of terms in current query and query before previous query and so on. At the moment we are not quite sure about certain weighting values because they have to be tuned by means of a set of experiments. Through the proposed model, we are able to express formally the idea that the relations between query terms exist and they can be measured in the range of queries. These relations will get weaker with the increase of the “distance" between the current query and the previous query. At the same time the model will be able to express the relative importance of some query terms for the user, measuring the frequency of their appearance in his searches. We argue that the proposed model could be able to gather information for creating patterns of a user search behavior. It still remains an open research question how many queries to be traced backward in order to build a functional and computationally reasonable profile. The proposed user model will have the ability to updates itself dynamically when new nodes and links are added and removed, and the weights are changed with accordance to the actual information, collected by the system. With each new query, issued by the user, the weights of existing nodes and links will be reduced proportionately. The exact amount of reduction of weights has been not established yet. But it will be in dependence of the number of the queries, tracked backward by the system. If the value of the weight of some node or link is reduced under predefined threshold, the respective node or link will be removed from the model. We believe that could keep the model current and actual by means of the proposed method.

In the process of personalization there is a risk that the personalization could be misdirected or unwanted. Adapting the query can result in increase of relevance of returned results if it is successful but if it is not, the relevance will decrease. In the process of enrichment of a query with terms, taken from a user profile the expectations are towards improving the relevance by providing missing search context. Thus taking into account, the expected benefit of a query adaptation against the risk of an improper personalization we propose the current query to be replaced by a series of queries. This series could include queries, adapted by terms derived from the user profile as well as the original query. The expansion terms are taken from a user profile. The selection of these terms is based on weights of links, connecting terms of a current query with terms from the previous queries and on weights of terms themselves. The last weight serves to express the importance of the terms for a particular user. In this work when we talk about expansion terms we mean the concept nodes which can be used for the expansion. In order to be considered as a valid expansion term, the concept node has to have its own weight greater than predefined threshold value. Moreover, the weight of a link between the concept node used for the expansion and the concept node in the current query also has to be over threshold value. If there are such multiple links, the few rounds of the query adaptations have to be performed. In the first round, the concept nodes with higher cumulative value of the links weights and their own weights are extracted from the model and used for the expansion. In the second round, the concept nodes with lower weights are used for the expansion and so on until the weight value falls under the predefined threshold. We expect that some heuristics will be helpful in the case of having too many expansion concepts.

The search results derived from the adapted queries as well as from the original query will be mixed in a result list. This approach is related to the so-called result diversification, known from the field of IR, which consciously diversified a set of results, returned by the system in response to a user query. The advantage of this method is that the user is offered results retrieved from the adapted query and from the original, formulated by the user query. Thus, the risk that personalization may be misdirected or unwanted is avoided. The user action - clicking on the document, extracted as a result of an adapted query is evidence that personalization was successful. We propose the user feedback to be used not only to validate the effectiveness of a personalization but also to update the user profile. The user action - clicking on documents, returned from an adapted query, will add extra weight to the link between the concept nodes – one to which the term from current query is associated and the other which serves for the expansion.

5. Implementation

The proposed in this work user model is not domain dependent and as such could be implemented in systems operating in different domains. We aim to integrate the model in e-learning environment in order to represent learner short-term interests and facilitate his searches.

The project Share.TEC: Sharing Digital Resources in the Teaching Education Community aims to support teachers and educators by providing access to digital resources, related to educational field. Share.TEC system includes a portal through which an individual user can access available resources [32] [33]. The portal provides a wide range of personalization features which are based on individual users' characteristics and on their national and cultural background. Three elements are crucial in the process of providing personalization: the adaptive approaches and techniques, a system interface and a user model. The main users activities, related to the adaptivity in digital libraries are: setting preferences and searching for digital content [3] [4] [5]. The user model in ShareTec includes three sets of data, by means of which the characteristics of individual user and his behavior in the system can be described and analyzed. They are as follows:

- Explicit preferences defined by the user himself
- Implicit user preferences collected through an analysis of a user behavior in an automated manner
- Summary of the behaviors of all users of the system

In the process of searching and filtering of search results, two clusters of data are created. The first cluster contains a list of search results, whereas the other cluster contains raw data including used in the searching keywords, comments, annotations and ratings. This raw data must be processed and analyzed before being able to describe a user behavior and preferences. In order to use that data, it must be aggregated and processed into a form that can model a user behavior. As stated above, this data has to be analyzed in order to be able to generate the implicit user's preferences from it. In this work we propose a user model, which can be used for modeling a part of a usage data, namely - the query terms. The proposed model aims to examine the effectiveness of personalization, obtained by means of solely information extracted from the user queries. But this model can as well be used together with another user models. In this case we expect an increase of search relevance of results. Realized adaptive features in ShareTec portal include automatic ranking of search results according to a user profile; recommendations to the user and individualized forms for assessments, bookmarks and annotations. The user profile in ShareTec is tightly integrated with the teacher ontology (TEO), developed in the frame of the same project. The search process in the system uses a user model and TEO ontology and is based on search engine Solr. The search component performs semantic query expansion by means of following techniques:

- Expansion based on explicit preference
- Expansion based on ontology - uses a parent-child relation
- Multilingual expansion
- Expansion for recommendation - uses the most concerned resource in accordance to user profile

The analysis of the adaptive, expansive query techniques show that in Share.TEC system does not exist a query adaptation, based on previous user's queries. Share.TEC recommendation system and a search engine use various sources of data like OMM metadata, teacher ontology TEO and implicit information taken from a user model. Solr search engine is configured to use in its indices the individual fields of OMM and TEO. In this work we propose to use TEO ontology concepts for an initialization of the conceptual nodes in presented user model. The intuition behind this is that the query terms used for searching in ShareTec are more likely to be related to concepts of teacher ontology, which models the educational domain. The practical implementation in ShareTec is done with the help of special tasks in Hadoop, working in asynchronous mode with a database in the portal. Our proposal follows the existing implementation of the system and the data processing will be implemented as an additional subtask in Hadoop.

6. Conclusions and future work

Future work is aimed to the practical aspects of the implementation of the proposed model and to answer the research questions, related to it. Amongst the questions that emerge, is the determination of the proper weigh values to links between concept nodes, described in the proposed model. Another research question that needs to be addressed is to be determined the exact numbers of the expansion terms, which can be used for a query adaptation. The proposed model is based on terms, taken from the previous user queries. The numbers of successive queries, which will be used to build the model, is not determined yet. But it appears that the proper number is essential for our model. On one hand, taking too many queries may result in a lack of accuracy and actuality of the model. On the other hand, taking not enough number of queries will lead to inefficiency and inability of the model to represent user's interests. And last but not least, the validation of the proposed model has to be conducted and the determination whether it presents a viable solution for a search personalization has to be provided.

Acknowledgments

"Acknowledge to the European funded Project UNITE (FP7 248583), namely its secondment programme coordinated by UNINOVA-GRIS, that supported the development of various ideas and concepts presented in this paper"
This work was partially supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3.3.06-0052 (2012/2014)

References

[1] Asnicar, F.A., Tasso, C.: ifWeb—a prototype of user model-based intelligent agent for document filtering and navigation in the World Wide Web. In: Adaptive Systems and User Modeling on the World Wide Web, Chia Laguna, Sardinia (1997)

[2] Balabanovic, M. (1998).Learning to Surf: Multiagent Systems for Adaptive Web Page Recommendation. PhD thesis, Department of Computer Science, Stanford University

[3] Boytchev, P., Grigorov, A., Earp, J., Stefanov, K., Georgiev, A. (2010) Adaptability Approaches in Digital Libraries, Second International Conference S3T, September 11-12, 2010, Varna, Bulgaria , pp. 6-13, ISBN 978-954-9526-71-4.

[4] Boytchev, P., Grigorov, A., Sarti, L., Georgiev, A., Stefanov K. and Chechev M. (2010), Recommender systems and repository search: the Share.TEC proposal, Sofia University e-Learning Journal, 2010/3, ISSN 1314-0086.

[5] Bozhilov, D., Stefanov, K., Stoyanov, S. (2009) The Effect of Adaptive Learning Style Scenarios on Learning Achievements, in IJCEELL V19 N4/5/6 2009, Special issue"Stimulating Personal Development and Knowledge Sharing", eds. R. Koper, K. Stefanov and D. Dicheva, pp.381-395.

[6] Brajnik, G., Guida, G., Tasso, C.: User modeling in intelligent information retrieval. Inf. Process.Manag. 23, 305–320 (1987)

[7] Brusilovsky, P., Tasso, C.: Preface to special issue on user modeling for Web information retrieval. User Model. User-Adapt. Interact. 14, 147–157 (2004)

[8] Brusilovsky P. (2001). User Modeling and User-Adapted Interaction, 11: 87-110

[9] Callahan, E. (2005). Cultural similarities and differences in the design of university websites. Journal of Computer-Mediated Communication, 11(1)

[10]Chirita, P.-A., Firan, C.S., Nejdl, W.: Personalized query expansion for the Web. In: 30th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007), pp. 7–14. ACM, Amsterdam (2007)

[11]Dolog, P., Nejdl. W,. Semantic Web Technologies for the Adaptive Web. The Adaptive web. Lecture Notes in Computer Science, 2007, Volume 4321/2007, 697-719, DOI: 10.1007/978-3-540-72079-9_23. p.697-719

[12]Gasparini, I., Weitzel, L., Pimenta, M.S. & Oliveira, J.P.M.d. (2011). Adaptive e-learning for all: integrating cultural-awareness as context in user modeling. In T. Bastiaens & M. Ebner (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2011, pp. 1321-1326.

[13]Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A. : User profiles for personalized information access. In: Brusilovsky, P., Kobsa, A., Nejdl,W. (eds.) The AdaptiveWeb, 1 edn, pp. 54–89. Springer, Berlin (2007)

[14]Ghorab, M., Zhou, D., O'Connor , A., Wade, V., Received:Personalised Information Retrieval: survey and classification < http://link.springer.com/article/10.1007%2Fs11257-012-9124-1>

[15]Jiang, X., Tan, A., (2009). Learning and inferencing in user ontology for personalizedSemantic Web search. Information Sciences 179 (2009) 2794–2808. p. 2794 – 2808

[16]Koutrika, G., Ioannidis, Y.: Rule-based query personalization in digital libraries. Int. J. Digit. Libr. 4, 60–63 (2004)

[17]Liu, F., Yu, C., Meng, W.: Personalized Web search for improving retrieval effectiveness. IEEE Trans. Knowl. Data Eng. 16, 28–40 (2004)

[18]Micarelli, A., Gasparetti, F., Sciarrone, F., Gauch, S., : Personalized Search on theWorld Wide Web < http://citeseer.uark.edu/projects/citeseerX/papers/personalized%20search.pdf>

[19]Micarelli, A., Sciarrone, F.: Anatomy and empirical evaluation of an adaptive Web-based information filtering system. User Model. User-Adapt. Interact. 14, 159–200 (2004)

[20]Mobasher B., DaiH., Luo T.,Nakagawa M., and Wiltshire J. (2002). Discovery of aggregate usage profiles for Web personalization. Data Mining and Knowledge Discovery, Vol. 6 (1), pp. 61–82.

[21]Oard, D.W.: Multilingual information access. In: Encyclopedia of Library and Information Sciences, 3rd edn, Taylor & Francis, Oxford, UK, pp. 3682–3687 (2010)

[22]Oard, D.W., Diekema, A.R.: Cross-language information retrieval. In: Williams M. (ed.) Annual Review of Information Science (ARIST), pp. 472–483. Information Today Inc., Medford (1998)

[23]Pazzani J. M. (2005). A framework for collaborative, content-based and demographic filtering. Artificial Intelligence Review, December 1999, vol. 13, no. 5-6, pp. 393-408(16).

[24]Reinecke, K.; Schenkel, S.; Bernstein, A. (2010) Modeling a User's Culture. In: The Handbook of Research in Culturally-Aware Information Technology: Perspectives and Models, IGI Global.

[25]Shahabi C. and Chen Y. (2003). Web Information Personalization: Challenges and Approaches, In the 3nd International Workshop on Databases in Networked Information Systems (DNIS 2003), Aizu-Wakamatsu, Japan.

[26]Shen, X., Tan, B., Zhai, C.: Implicit user modeling for personalized search. In: 14th ACM InternationalConference on Information and Knowledge Management (CIKM 2005), pp. 824–831. ACM, Bremen (2005)

[27]Sieg, A., Mobasher, B., Burke, R., Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search < http://www.comp.hkbu.edu.hk/~iib/2007/Nov/iib_vol8no1_article1.pdf >

[28]Sowa, J., (1992) : Buiding a semantic network, < http://www.jfsowa.com/pubs/semnet.htm>

[29]Speretta, M., Gauch, S.: Personalized search based on user search histories. In: IEEE/WIC/ACM International Conference onWeb Intelligence (WI 2005), pp. 622–628. Compiegne University of Technology, Compiegne (2005)

[30]Stamou, S., Ntoulas, A.: Search personalization through query and page topical analysis. User Model. User-Adapt. Interact. 19, 5–33 (2009)

[31]Stefani, A., Strapparava, C.: Personalizing access to Web sites: the SiteIF project. In: 2nd Workshop on Adaptive Hypertext and Hypermedia, Pittsburgh, Pennsylvania, USA (1998)

[32]Stefanov, K., Nikolov, R., Boytchev, P., Stefanova, E., Georgiev, A., Koychev, I., Nikolova, N., Grigorov, A. (2011) Emerging Models and e-Infrastructures for Teacher Education, 2011 International Conference on Information Technology Based Higher Education and Training ITHET 2011, paper 33, IEEE Catalog Number: CFP11578-CDR, ISBN: 978-1-4577-1671-3.

[33]Stefanov, K., Boytchev, P., Grigorov, A., Georgiev, A., Petrov, M., Gachev, G., Peltekov, M. (2009), Share.TEC System Architecture, In Proceedings of First International Conference on Software, Services & Semantic Technologies (S3T, 29-29 October 2009), Eds.: D. Dicheva, R. Nikolov and E. Stefanova, Sofia, Bulgaria, 2009, pp. 92-99, ISBN 978-954-9526-62-2

[34] Sullivan D., “Previous Query” Refinement Coming To Hit Google Results <http://searchengineland.com/previous-query-refinement-coming-to-hit-google-results-13743>

[35]Tsianos N., Germanakos P., Lekkas Z., Mourlas C and Samaras G (2009).An Assessment of Human Factors in Adaptive Hypermedia Environments Intelligent User Interfaces, 2009,pp 1-35

[36]Zhou, D., Lawless, S., Wade, V.: Improving search via personalized query expansion using social media.Inf. Retr. 1–25 (2012). doi:10.1007/s10791-012-9191-2

Albena Turnina is a PhD student at Department of Information technology at Sofia University, Bulgaria. She has received many grants amongst which are a research grant from Uninova, a grant for participation in Research Programme at London University as well as grants for participation in PhD courses, seminars and conferences. She holds computer certificates CCNA, Network Security, IBM Certified Database Associate. Her professional experience includes lecturing in the field of computer networks and development of web applications. Her research interests encompass semantic web, search personalization and adaptive search.