Journal of Universal Computer Science, vol. 20, no. 9 (2014), 1232-1258
submitted: 17/2/14, accepted: 29/8/14, appeared: 1/9/14 © J.UCS

A Personalized Approach for Re-ranking Search Results Using User Preferences

Naglaa Fathy
(Ain Shams University, Faculty of Computer and Information Sciences, Cairo, Egypt
naglaa_fathy@cis.asu.edu.eg)

Tarek F. Gharib[1, 2]
([1]King Abdulaziz University, Faculty of Computing and Information Technology
Jeddah, Saudi Arabia
tfgharib@kau.edu.sa)

Nagwa Badr
([2]Ain Shams University, Faculty of Computer and Information Sciences, Cairo, Egypt
dr.nagwabadr@gmail.com)

Abdulfattah S. Mashat
(King Abdulaziz University, Faculty of Computing and Information Technology
Jeddah, Saudi Arabia
asmashat@kau.edu.sa)

Ajith Abraham
(Machine Intelligence Research Labs, Scientific Network for Innovation and Research Excellence, P.O. Box 2259, Auburn, Washington 98071, USA
ajith.abraham@ieee.org)

Abstract: Web search engines provide users with a huge number of results for a submitted query. However, not all returned results are relevant to the user's needs. Personalized search aims at solving this problem by modeling search interests of the user in a profile and exploiting it to improve the search process. One of the challenges in search personalization is how to properly model user's search interests. Another challenge is how to effectively exploit these models to enhance the search quality. In this paper, an effective hybrid personalized re-ranking search approach is proposed by modeling user's search interests in a conceptual user profile, and then exploiting this profile in the re-ranking process. The user profile consists of concepts obtained by hierarchically classifying user's clicked search results into categories. These categories are extracted from the taxonomy of concepts called The Open Directory Project (ODP) where each concept represents a category. Additionally, each concept in the user profile consists of two types of documents; taxonomy document and viewed document. Taxonomy document is used to represent the user general interests as it contains information from web pages originally associated with such ODP category. Viewed document is used to represent the user specific interests as it contains information from web pages clicked by the user. Finally, the re-ranking process of search results is performed by semantically integrating user's general and specific interests from the user profile together with rankings of the traditional search engine. Experimental results show that semantic identification of user's search interests improves re-ranking quality by providing users with the most relevant results at the top of the search results list.

Keywords: Search engine, User profile, Personalization, Taxonomy, Open Directory Project, Re-rank.
Categories: H.3.3

1 Introduction

Different users with different needs submit queries with one or more keywords to web search engines through simple user interfaces. Search engines depend on keyword matching for searching against a collection of web pages to find the pages that would be returned. Therefore, current retrieval systems are not adaptive enough to satisfy user's search needs.

Furthermore, some keywords could be ambiguous and have different meanings as in the search query "Ajax". For such query, users might have various goals and prefer different answers, i.e. "Ajax web based development", "Dutch football team Ajax Amsterdam", or "cleaning product Ajax". However, users often submit short queries in searching the web that does not provide adequate information to identify user needs.

Moreover, users might not choose the right words that best identify their needs. A recent study demonstrated that users with more than 7 years of online searching experience obtained much more relevant documents than users with less experience, [Al-Maskari and Sanderson 2011]. Accordingly, search engines to cover users' different interests provide a variety of search results.

The goal of web search personalization is to consider the user's search preferences and interests in the search process to provide each user with the results that are most relevant to his interests. One of the challenges to personalization is how to identify such users' interests. Another Challenge is to effectively exploit these interests in the retrieval system to improve search results.

In particular, personalized search can be achieved by re-ranking search results returned by a traditional search engine according to the user profile which might be constructed from the user's search or browsing behavior. The user profile information is used to identify user's interests and can be collected explicitly by directly asking the user about his search preferences.

However, not all users are willing to provide their search intentions for each query. A more complex method is by implicitly collecting user information from his visited web pages as well as monitoring his browsing activities (i.e. bookmarks).

Several personalization techniques have been proposed to model the users' preferences from their click-through data and browsing behaviors. These personalization techniques represent each user with a set of terms extracted and weighted from a user's visited pages. Then, re-ranking search results is achieved by computing the scores of the results' snippets, [Matthijs and Radlinski 2011].

Another approach is proposed by [Hoeber and Massie 2009] in which the results are categorized by different topics and user's clicked results for user current query are observed to re-order results according to the user's current needs.

The user profiles constructed with reference to a topical ontology to categorize user's visited pages were presented in [Mianowska and Nguyen 2011], [Chirita et al. 2005], [Sieg et al. 2007a], [Sieg et al. 2007b]. Re-ranking is performed by computing the numerical scores to check the relevance of search results for a given query against the user profile.

The user profile in [Li et al. 2009] is composed of queries submitted by the user associated with URLs and topics of the clicked results for each query. Re-ranking is done by identifying queries from the user profile that are similar to user's current query, then comparing topics of these relevant queries with the topics of search results.

In this paper, an effective hybrid personalized re-ranking search approach is proposed by modeling user's search interests in a conceptual user profile, and then exploiting this profile in the re-ranking process. The user profile consists of concepts obtained by hierarchically classifying user's clicked search results into categories. These categories are extracted from a concept hierarchy called The Open Directory Project (ODP) where each concept represents a category. Any structural noise is removed from the ODP to obtain a more accurate concept hierarchy. Furthermore, each concept in the user profile consists of two types of documents; taxonomy document and viewed document. Taxonomy document is used to represent the user general interests as it contains information from web pages originally associated with such ODP category. Viewed document is used to represent the user specific interests as it contains information from web pages clicked by the user. Finally, the hybrid re-ranking process of search results is performed by semantically integrating user's general and specific interests from the user profile together with rankings of the traditional search engine.

Subsequent sections in this paper are arranged as follows. Related work is presented in Section 2. The proposed Architecture is described in Section 3. Experimental evaluation is explained in Section 4 and Finally, the conclusions and future works are discussed in section 5.

2 Related Work

Most personalization approaches are based on constructing a user profile that aims to collect information about the user's topics of interest to improve the quality of information retrieval. In order to build a user profile, information may be collected either explicitly or implicitly.

Explicit information is collected directly by asking the user about his/her interests. However, implicit information is collected by monitoring the user activity, [Baeza-Yates et al. 2011]. Profiles that are adapted to the user's changing interests are called dynamic, whereas profiles that maintain same information are considered static, [Teevan et al. 2005].

Additionally, short-term and long-term interests might be distinguished in user profiles when taking time into consideration as in [Kim et al. 2003], [Mobasher 2007], [Perkowitz and Etzioni 1998]. User interests that are not changing frequently over time might be represented by *long-term* profiles. User's current interests that are changing quickly are represented by *short-term* profiles and are more difficult to identify.

In personalized search systems, [Micarelli et al. 2007], user profiles can enhance web search quality in one of 3 phases, namely: "*Part of the retrieval process, Query modification, or Re-ranking". In the first phase, user profiles are built into the search process, and are utilized to score web documents. However, this method of that search systems is forced by time constraints, and is considering the personalization process as a time-consuming process.

In the query modification phase, user profiles are extended only to the submitted keywords in the query without changing the ranking procedure. Therefore, lists of result are not highly affected by query modification phase. In the re-ranking phase, user profiles are used to re-order search results retrieved from a non-personalized search engine while considering the user's interests.

One of the main essential forms of representing user profiles is by setting weighted keywords. In keyword profiles, the user can directly provide the system with his interesting keywords or the system can extract the keywords from the user's visited pages. The score, number of the user's interests represents weighted keywords. However the main problem with the keyword profiles is the ambiguity exists in words having more than one meaning which might affect the accuracy of keyword profiles, [Gauch et al. 2007].

Another form of user profile representation is the semantic network-based profiles, where each node represents a concept which represents the user's specific interest in a collection of words and its synonyms. However, constructing such semantic network profiles is not easy because terms that represent each concept are not predefined.

For example, a user profile represented by semantic network was proposed by [Mianowska and Nguyen 2011], where concepts are extracted from the WordNet ontology [Miller 1995]. They represent each node in the semantic network-based profile as a set of (synonyms) terms extracted and weighted from a user's history. Additionally, each concept is associated with a time stamp to identify the last time this concept appeared in a user's query, in order to update the user profile. The main drawback in that work is that some query words are missing from WordNet.

A more efficient model for representing user's interests is the concept profile. Such profiles are constructed with a predefined matching between concepts and vocabulary, [Gauch et al. 2007], [Andhare and Mahajan 2014]. In the concept-based profiles, Nodes do not represent specific words or synonym words as occurs in semantic network profiles. Instead, concept profiles represent abstract concepts (topics) that are interesting to the user. Relationships between these concepts help to disambiguate the vocabulary of terms.

An example for concept user profile is presented in [Kumar and Sharan 2014]. The profiles are created from user's browsing history where web pages are classified into certain categories obtained from the Open Directory Project. These categories have a fixed number of weights calculated as the number of pages visited by the user for each category. They additionally created an enhanced profile by adding the most relevant URLs in ODP for each category in the user profile. Relevant URLs from ODP are obtained by measuring the cosine similarity between each visited web page in the user profile and the URL from ODP in the corresponding category.

One of the disadvantages of the above approach is that the weights of the categories in the user profile are fixed. However, user's interest in certain category might increase/decay over time. Another drawback in the enhanced profile is that they ignore the semantic relationship between the URLs in ODP and the visited web pages in the Profile. As a result, some important URLs from ODP might not be added in the enhanced profile, simply because they have fewer words in common to the profile pages.

A personalized approach for concept-based user profile based on user's browsing behavior is proposed by [Sieg et al. 2007a], [Sieg et al. 2007b]. Each concept represents a topic from ODP with numerical scores to identify user interests mapping to the vector space classifier used to categorize browsed pages. These heaps are then used for the re-ranking process. This approach ignores the hierarchical relationship between ODP concepts when classifying the web pages. Consequently, it is still challenging to identify commonalities between subtopics of a specific class, and distinguish between them.

Persona, [Tanudjaja and Mui 2002], uses ODP to build taxonomy of user interest and disinterest to represent the user profile. Users were asked to explicitly provide feedback on the results they clicked for queries submitted to the ODP. The user profile is then updated for each concept with the number of negative and positive interests, according to the results that are already pre-classified into concepts from the directory. However, the user profile can grow heavily and include a large number of concepts that are so close to each other because the hierarchy of the ODP has so many levels with several concepts.

Another personalization method in [Li et al. 2007] presented two types of user profiles adapted to the user's changing interests. The first type is long-term profiles that store visited pages' topics as part of the Google Directory together with the number of visits for each topic. The second type is a short-term model that stored user's history of recently visited pages. Considering the entire search history, Re-ranking is achieved by computing the similarity between the user profile hierarchical topics and current search results' topics. However, not all information in the user profile reflects the user's current search interest for a given query.

In [Oishi et al. 2008], a personalized search system is proposed by creating user profiles from user's bookmark folders. For a submitted query, the user has to choose the most related user profile among a set of constructing user profiles. The cosine similarity is measured between the selected user profile and each answer to re-rank search results. However, if different users use the computer, then the same profile will be shared among all users even though they have different interests.

Additionally, in [Hawalah and Fasli 2011], web pages of the user browsing behavior (i.e. browsed pages, favourites.etc) are captured and defined in an ontological user profile based on the ODP. Then, the user profile is exploited with other information sources to provide a hybrid re-ranking method to personalize the search results. On the other hand, the hierarchical semantic structure of ODP concepts is not considered by this approach for classifying web pages visited by the user. Another drawback is that user's browsing history exploited for creating the user profile could be available from a single computer only. Even though, the user profile might be imprecisely if different users utilize the same computer.

Other personalization approach in [Mohammed et al. 2010] constructs 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 is created by expanding keywords into a semantic hierarchy from WordNet associated with results 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.

Furthermore, in [Chirita et al. 2005], the user profile is constructed by storing the hierarchy of topics of interests chosen explicitly by the user from the ODP. For each new query, search results are mapped to topics from the ODP, and then the hierarchical distance is computed between topics of search results and topics in the user profile. However, the user profile is still static and does not adapt to the user's changing interests.

[Antoniou et al. 2012] proposed a concurrent re-ranking of search results with no need to store user's search history. As the user selects a result, the information included on that page is used to identify user's search needs. However, it has been proven that such strategies used for immediate updates are not matched by users' interest, even when they give more accurate results [He et al. 2007].

A recent study by [Bibi et al. 2014] proposed the user profile based on concepts which are groups of words that co-occur frequently in web snippets of visited web pages. Concepts are organized in the profile as a tree with the relationship between these concepts. These relationships include similar or parent/child relationship. Weights are assigned/ incremented to concepts found in the clicked web snippets and to concepts having a relationship with this concept. Re-ranking is done by assigning scores to current web snippets for a given query based on the aggregation of its concepts' weights.

However, concepts in web snippets of a new submitted query might not exist in the user profile though they might be semantically similar to other concepts in the profile. Besides, polysemy[1] and synonymy are not considered in weight ancestor /descendant concepts for a given concept. More specifically, a concept *c* might have different ancestors/descendants according to the context of the snippet that contains this concept i.e. "apple" concept which might refer to "apple fruit" once and "apple computer" other time.

3 Proposed Software Service Architecture

In this paper, we propose a personalized search system that involves creating concept-based user profiles from user search history with reference to ODP concept hierarchy. In the proposed approach, the user profile is enriched with two different types of information for each concept: taxonomy document, and viewed document. The taxonomy document includes keywords from documents originally associated with topics from the ODP directory. The viewed document includes terms from user's clicked search results. Furthermore, re-ranking is based on user's general interests and matches in certain query' topic as well as considering the ranks of the non-personalized search engine.

The proposed system consists of four main modules as shown in Figure 1:

- Module 1: Preparing the reference taxonomy (or concept hierarchy)
- Module 2: Collecting user information
- Module 3: Learning and constructing the user profile

[1] Polysemy refers to a word that might have different meanings. While in synonymy, different words can be used to represent similar information.

• Module 4: Search personalization by exploiting the user profile to re-rank search results.

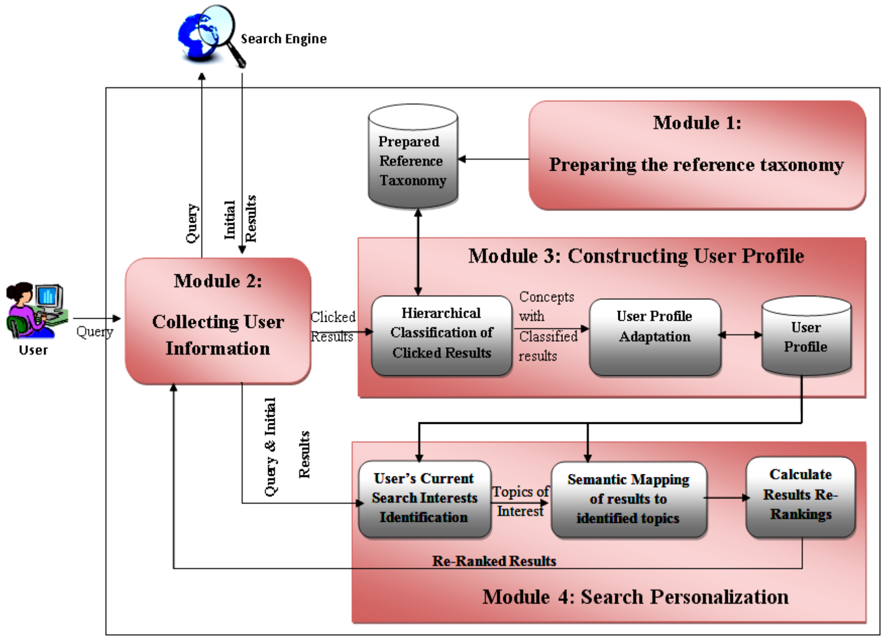

Figure 1: Personalized Search Proposed Architecture

3.1 Module 1: Preparing Reference Taxonomy

In this paper, the user profile is constructed with reference to a concept hierarchy or taxonomy of topics. For this purpose, Open Directory Project (ODP) is utilized as our reference taxonomy. The Open Directory Project is an open content directory of the web that is produced and preserved by a group of volunteer editors, [Dmoz 2012]. Topics in the ODP and web pages that belong to these topics are organized using hierarchical ontology schema as shown in Figure 2.

In order to get a precise concept hierarchy, some changes should take place because some parent-child links are not conceptual. For example, some topics are divided geographically, while others are divided alphabetically to separate content. Furthermore, some topics may have fewer children while others may have hundreds. Additionally, some topics may be associated with many web pages, while others may have fewer pages. Therefore, in order to improve the profiling accuracy, parent-child topics that are not conceptually related are eliminated together with those topics that have too few Web pages linked to them, [Gabrilovich and Markovitch 2007].

In order to represent the reference taxonomy, we choose the first 30 URLs for each concept based on the order in which they are represented by ODP. Terms from the 30 pages are collected in one document for each concept. The (Term Frequency – Inverse Document Frequency, TF-IDF) mechanism [VSM 2012] is then used to weigh each term from 0 to 1 in each document Eq.(1) which is then normalized by the vector magnitude because documents are not the same length Eq.(2)

$$\text{Term weight, } tc_{ij} = (tf_{ij} * idf_i) \tag{1}$$

Where

tf_{ij} is the frequency of term i in document j,

idf_i =*Log (Number of documents in D / Number of documents in D that contain* t_i*)*

D = the collection of documents that represent the ODP concepts i.e. one document for each concept.

$$\text{Normalized term weight, } ntc_{ij} = (tc_{ij} / vector\_length_j) \tag{2}$$

$$\textit{Where} \qquad vector\_length_j = \Sigma\, tc_{ij} \tag{3}$$

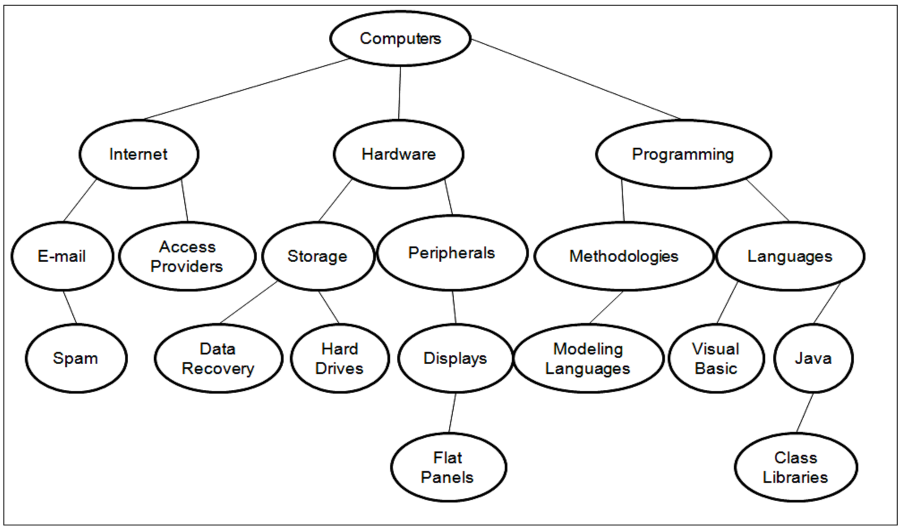

Figure 2: Portion of the open directory Project

3.2 Module 2: Collecting User Information

In order to implicitly collect information about users, we implemented Google Wrapper; a wrapper around the Google search engine [Google 2012]. In particular, the wrapper stores the following information: the user's submitted queries, returned search results, and user clicks. Google Ajax API [Google API 2012] and the .Net Library for Google search [Dotnet API 2011] were used for the implementation.

Users are identified through cookies created on their local machines when they register with their email addresses. Cookies are used to store and retrieve their user Id. Users are notified in case the cookie was lost so that they could login to reset the cookie.

Google wrapper is the linking point among the user, the traditional search engine, and the main modules of the proposed personalized search system as shown in Figure 3. More specifically, Queries submitted by users are redirected by the Google Wrapper to the Google search engine. ***The wrapper then performs*** the following:

- Capture the results returned from the search engine,
- Record them together with the query and the user ID,
- Pass the query with the returned results to Search Personalization module to apply the proposed re-ordering method,
- Then show the re-ordered results to the user.
- If a user clicks on a result, the wrapper records the clicked page in conjunction with the user ID in the log, prior to redirecting the browser to the proper web page. This log is then exploited in the User Profile Construction module to update the user profile.

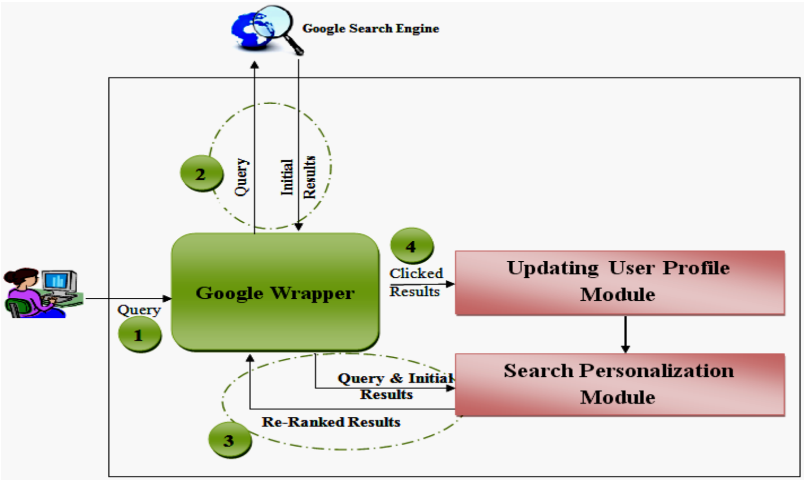

Figure 3: Collecting User Information with Google Wrapper

3.3 Module 3: Constructing the user profile

In this module, data obtained by observing user search history from module 2 in Section 3.2 is used to learn and construct concept-based user profile. This profile is mainly an instance of the ODP reference taxonomy. More specifically, search results clicked by the user are classified into concepts from ODP which are then used together to build the profile.

It is worth mentioning that ODP classifies only 0.03% of the pages that are known to the search engines, [Gulli and Signorini 2005]. So, we used the hierarchical classification method in [Pulijala and Gauch 2004] in order to classify clicked search results into ODP concepts. Hierarchical classification starts by matching document to the best category (concept) at the top level and then “stepping down” the concept hierarchy by matching the document into subcategories of that category only. This method, [Pulijala and Gauch 2004], provides better accuracy of the highest matching (concept) category, (70% using hierarchical classifiers versus 46% using flat classifier). For each clicked search result, a set of processes is applied so as to construct the profile as shown in Figure 4.

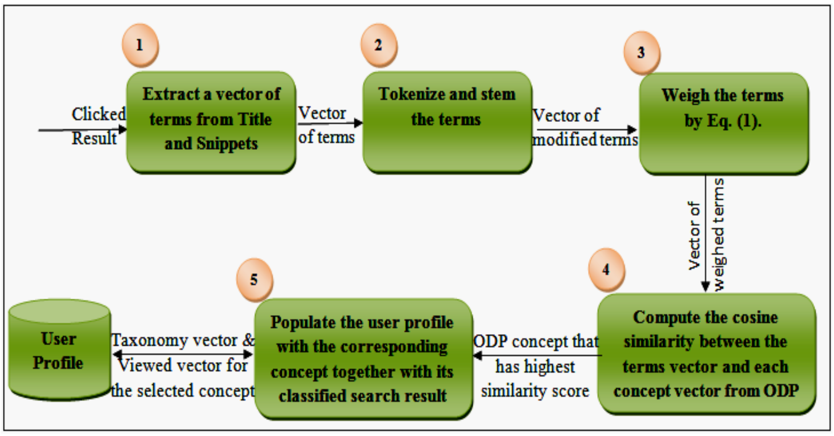

Figure 4: Steps for constructing/updating the user profile for a clicked result

In the second process, Porter Stemmer [Porter 1980] is used to stem the terms of each clicked result. In the fourth process, the hierarchical classification method by [Pulijala and Gauch 2004] is used in order to classify search results into the appropriate concepts from the ODP .For this purpose, the vector space model [VSM 2012] is used for computing the cosine similarity between the concept vector and the result vector as follows:

$$\cos(\vec{c},\vec{d}) = \frac{\vec{c}\bullet\vec{d}}{|\vec{c}||\vec{d}|} = \frac{\vec{c}}{|\vec{c}|}\bullet\frac{\vec{d}}{|\vec{d}|} = \frac{\sum_{i=1}^{|V|} c_i d_i}{\sqrt{\sum_{i=1}^{|V|} c_i^2}\sqrt{\sum_{i=1}^{|V|} d_i^2}} \tag{4}$$

Where c_i represents the weight of term *I* associated with concept *c,* and

d_i represents the weight of term *i* in document *d*.

In the last process, the user profile is populated with the clicked results and their corresponding concepts. If the concept already exists in the profile, the new classified result is concatenated with the past clicked results under this concept and terms weights are normalized by Eq. (2) to create one document called **viewed document**.

Finally, the concept-based user profile contains a taxonomy document and a viewed document for each concept:

- The ***taxonomy document*** includes a vector of weighted terms of information originally collected from the reference taxonomy discussed in part 3.1. This kind of document shows an overview of various topics categorized into an ODP concept.
- The ***viewed document*** includes a vector of weighted terms taken from a user's clicked search results which were classified into this concept as shown in Figure 5. This kind of document represents a user's specific interest at a particular concept.

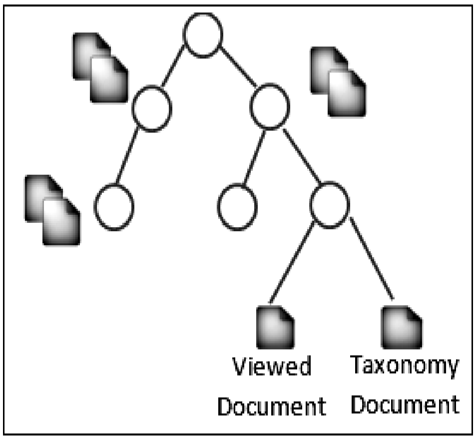

Figure 5: Enhanced Concept-based User Profile

3.4 Module 4: Search Personalization

In this module, a hybrid personalized re-ranking methodology is applied to provide users with more relevant search results for the top. For a given query, search Personalization is achieved in three steps as follows:

i. Identifying user's topics of interest for current search

ii. Semantic Mapping of search results to the identified topics

iii. Calculating search results' re-ranking scores

3.4.1 Identifying user's topics of interest for current search

As a first step, the query submitted by the user is matched to the user profile to choose concepts that are highly similar to a user's for the current query. For this purpose, the cosine similarity is computed between the query and user profile's taxonomy documents as shown in Eq. 4. Taxonomy documents are basically obtained from the reference taxonomy; therefore, they could clearly describe the correlated concept. Additionally, in order to take advantage of the user profile more effectively, queries are matched to the conceptual user profile which includes user's interesting concepts only rather than the entire reference taxonomy.

3.4.2 Semantic Mapping of Search Results to the identified topics

After selecting the concepts that represent the user's query, search results are semantically mapped to these concepts. This step is necessary to measure the relevance of each result with the concepts selected from the user profile. Many approaches use the cosine similarity measure to map query's results to the user profile concepts. However, this measure does not take into account that different documents which have less common terms might have semantically related words. Therefore, in this paper, the semantic similarity method proposed in [Madylova and Oguducu 2009] is employed to map search results to a query's concept.

[Madylova and Oguducu 2009] created a semantic vector for each document by extending the document vector with parent terms extracted from an IS-A taxonomy of words (e.g. WordNet [Miller 1995]). And then the cosine similarity is calculated between the newly formed semantic vectors. This method brings down the computational time needed to compute the semantic similarity between two documents. This is because they compute the cosine similarity once between the vectors of the two documents rather than having to compare each individual word from one document against each word from the other document.

In this work, semantic vectors are constructed by the adopted method for each search result of a given query and for the taxonomy document of the concepts that represent that query. Then the cosine similarity is computed between the semantic vector of the search result and the semantic vectors of the query concepts. Steps for constructing a semantic vector for a search result using the method in [Madylova and Oguducu 2009] are presented in Figure 6**.** Parent words extracted from the Is-A taxonomy for each term in a search result is weighed by Eq. (5).

$$p_{wjm} = w_{ij} \times (10 - m) \times 0.1 \qquad (5)$$

Where

p_{wjm} is the weight of the mth parent word of term t_j for a given search result, r_i

w_{ij} is the weight of term t_j in search result, r_i

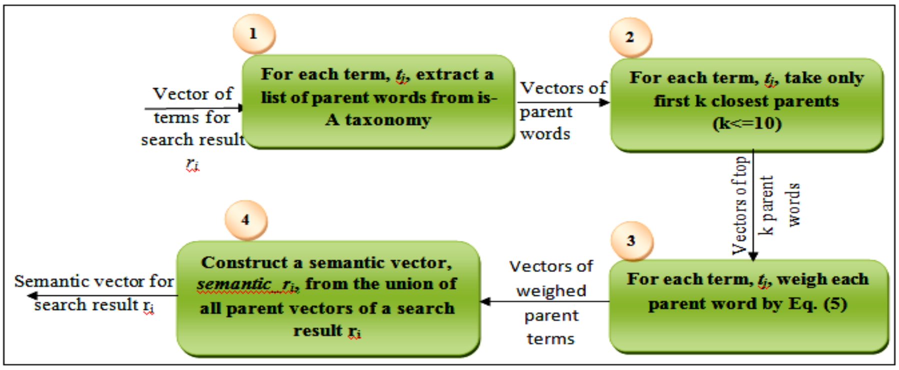

Figure 6: Steps for constructing a semantic vector for a given search result

To explain the semantic classification process, consider a simple example described in Table 1, where a user submits the query "program interface". In this example, two concepts are selected from the user profile to represent that query; "computer/programming/languages" and "computer/software/operating systems". It is noticeable, that the cosine similarity between terms *(browser, interface)* in a search result $\boldsymbol{r_i}$ and terms *(editor, GUI)* in the taxonomy document for concept1 is 0. Also, the cosine similarity between terms *(browser, interface)* in a search result $\boldsymbol{r_i}$ and terms *(Linux, Windows)* in the taxonomy document for concept2 is 0. This is because search result $\boldsymbol{r_i}$ *does* not have words in common with either taxonomy document, although these documents are extremely semantically related to r_i.

Query, q	program interface
Terms vector for a search result, r_i	(*browser,* 0.30), (*interface,* 0.20)
Query's Selected Concepts	Concept1: computer/programming/languages Concept2: computer/software/operating systems
Taxonomy Document for Concept1	(*editor,* 0.30), (*GUI,* 0.20)
Taxonomy Document for Concept2	(*Linux,* 0.30), (*Windows,* 0.20)
Parent Vectors for search result, r_i (from Figure 7)	$P_{browser, 0.30}$= { (*browser*, 0.30),(application, 0.27),(program, 0.24),(software, 0.21), (code, 0.18), (coding system, 0.15) } $P_{interface, 0.20}$= { (*interface*, 0.20), (program, 0.18), (software, 0.16), (code, 0.14),(coding system, 0.12), (writing, 0.10)}
Semantic Vector for search result, r_i	$semantic_r$= { (*browser*, 0.30), (*interface*, 0.20), (application, 0.27), (program, 0.42), (software, 0.37), (code, 0.32), (coding system, 0.27), (writing, 0.10) }
After forming semantic vectors for concept1 and concept2 in same behavior: **the semantic similarity between result, r_i, and concept1 :** 0.62 **the semantic similarity between result, r_i, and concept2:** 0.46	

Table 1: Simple example for semantic mapping of search results to query's concepts

Therefore, by applying the adopted method and by setting k=5, parent vectors form for each term in a search result, r_i from Figure 7. Then the semantic vector for r_i is constructed from these parent vectors as shown in the table. After calculating the semantic similarities, it is obvious that result r_i is more related to concept1.

3.4.3 Calculating Search Results Re-ranking Scores

Re-ranking search results is the last step in the proposed personalized web search approach. It is worth mentioning that each document for each concept in the profile represents the user interests from different perspectives. The viewed documents identify the user's specific preference of a concept. For example, a user may be interested in certain parts of a concept. In this case, the viewed documents should be considered greatly when re-ranking search results. Nevertheless, personalized search results could be provided only if such viewed documents hold adequate information about a user's interests. In such case, taxonomy documents should be weighted heavily when re-ranking results. For this reason, we need to measure the significance of viewed and taxonomy documents of each concept. This could be done by computing *cosine similarity* between the query and each document separately for a given concept as shown in Figure 8. In case the query is further related to the taxonomy document of the result's concept, then more weight will be given on the taxonomy document's ranking. Otherwise, the viewed document will be assigned more weight, as in the following equation:

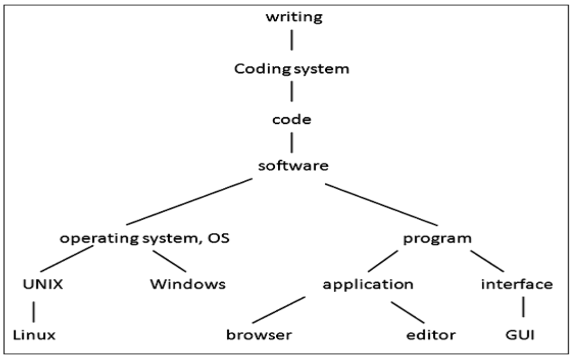

Figure 7: Fragment of WordNet

$$\text{Score}(R_i) = CosSim(q, C_{Ri\_T}) * \text{SemanticSim}(R_i, C_{Ri\_T}) + CosSim(q, C_{Ri\_V}) * \text{SemanticSim}(R_i, C_{Ri\_V}) \quad (6)$$

Where

R_i is the search result,

q is the query,

C_{Ri_T} and C_{Ri_V} are the taxonomy and viewed documents respectively, of concept c that represent result *re*.

Finally, taxonomy and viewed rankings together with the ranking of the non-personalized search engine (i.e. Google) are combined as in Eq.(7):

$$\text{Final Rank} = (1 - \alpha) * \text{OriginalRank} + \alpha * \text{Score}(R_i) \quad (7)$$

Where $0 \leq \alpha \leq 1$, in this way, if $\alpha = 1$ then personalized ranking is only considered while neglecting the traditional search engine ranking.

Input: Concepts C from the user profile that represent the user's current search interests for current query q, Search Results $R_i \in R$ represented by document vectors

Output: Re-ranked Search Results

Begin

Initialize TaxonomyQueue

Initialize ViewedQueue

// 1- Calculate the semantic similarity between the results and the user's general interests represented by the taxonomy documents to find the concept C_{Ri} that highly represent each result

foreach $R_i \in R$ **do**

Calculate Maximum_SemSim (R_i,$C_{j\_taxonomy\_document}$) ; $\forall C_j \in C$

TaxonomyQueue.Add (SemSim (R_i, C_{Ri_T}))

//C_{Ri_T} is the taxonomy document of the concept C with the highest Sem_Sim score with result R_i

foreach $R_i \in R$; $C_{Ri} \in C$ **do**

//2- Calculate the semantic similarity between the results and the user's specific interests represented by the viewed document of concept C_{Ri}

Calculate SemSim (R_i, C_{Ri_V}) // C_{Ri_V} is the viewed document of the concept C that represents result R_i

ViewedQueue.Add (SemSim (R_i, C_{Ri_V}))

end-for

//3- Calculate Results Re-ranking Score

Initialize Final_Queue

foreach $R_i \in R$ **do**

Calculate CosSim (q, C_{Ri_T}).

Calculate CosSim (q, C_{Ri_V}).

R_i.TaxonomyRanking = TaxonomyQueue.FindRank (R_i)

R_i.ViewedRanking = ViewedQueue.FindRank (R_i)

Score(R_i) = CosSim (q, C_{Ri_T}) * R_i.TaxonomyRanking + CosSim (q, C_{Ri_V}) * R_i.PersonalRanking.

Final_Rank = (1- α) * original_rank + α * Score (R_i).

Final_Queue.Add (R_i, Final_Rank)

end-for

End

Figure 8: The proposed Re-ranking Algorithm

It is also noticed that we used in the semantic similarity between the search results and concept documents to improve accuracy of the personalization. However, for measuring the similarity between the query and concept documents, we used only the cosine similarity. This is because dictionaries (i.e. WordNet) have poor coverage of common queries [Song et al. 2010].

4 Experimental Evaluation

In general, personalized search systems are evaluated by conducting user studies with a specific number of people taking part in the evaluation process over a period of time. The user profiles might be automatically learned from search histories or manually identified by the participants themselves [Dou et al. 2007]. In this section, we evaluate the effectiveness of the proposed personalized re-ranking approach as follows:

- *Evaluating the user profile accuracy* in terms of effectively ordering the concepts according to their degree of relevance to user's needs. This step is essential because personalized search efficiency is greatly affected by the user profile accuracy.

- *Evaluating the personalized search effectiveness* of the proposed re-ranking approach against a typical search performed by a traditional search engine, i.e. Google. Traditional search engines do not consider the user's search context in the search process. As the details of Google's personalized search algorithms are not publicly available [Matthijs and Radlinski 2011], our work will only compare to the default search engine ranking and not the personalized version.

- *Comparing the proposed personalized re-ranking approach to other ranking models.*

4.1 Experimental Setup

In order to evaluate the effectiveness of the proposed approach, experiments were designed with part of the AOL 2006 dataset as well as a set of data from 6 users invited to search through the proposed personalized search interface as follows:

4.1.1 AOL real dataset

To evaluate the overall effectiveness of the proposed approach from a wide set of users, we used the click-through data taken from an AOL log of real search data released to the public in August 2006 [Pass et al. 2006]. In this collection, we stored only queries with at least 5 unique clicks. Also, we only kept users who submitted more than 50 unique queries to construct more accurate user profiles. Overall, we extracted an AOL search log of 30 users with a total number of 2035 distinct queries and 15960 clicked results.

4.1.2 The users' dataset

In this section, we constructed our own user dataset because of it was impossible to ask the original users of the AOL query log to evaluate the degree of relevance of the profile to their interests. For this purpose, 6 users were invited to search through our personalized search interface. For each search, Google API returned the order of the top 50 results. To avoid result's position bias, the results were positioned at random orders. Participants are divided into 3 types:

- *Users with Clear* Queries: searching for one-meaning queries.
- *Users with Semi-ambiguous* Queries who search for queries with 2 or 3 meanings.
- *Users with Ambiguous* Queries who search for queries with more than 3 meanings.

Examples of ambiguous queries selected from the Wikipedia disambiguation page are shown in Table 2. Ambiguous queries were exploited to evaluate the personalized re-ranking quality of search results.

Query	Meaning
Eagle	- American musical group - Kind of Birds - The British comic book
Opera	- A web browser that is very commonly used - A form of musical and dramatic work
Race Track	- A purpose-built facility for the conducting of races. - A paper and pencil game - Memory, a device for storing bits in a magnetic racetrack

Table 2 : Some ambiguous queries selected from the Wikipedia disambiguation page

Participants were asked to submit queries related closely to their preferences and subjects of study. More specifically, users submit queries about their subjects of study in the first four days. They input queries on their avocations in the next 3 days. In the last 3 days, the users were requested to submit some repeated queries. Repeated queries were employed to evaluate our personalization approach efficiency in case of re-finding known information. After the ten-day period, we collected all logs for the 6 participants shown in Table 3.

User	U1	U2	U3	U4	U5	U6
# queries	33	29	27	35	30	25
# clicked pages	50	46	36	56	47	39

Table 3 : Total number of queries and clicked pages over 10 days

4.2 Evaluation Metrics

We evaluate the proposed personalized approach using the following information retrieval metrics:

- ***Precision at K*** (P@ (K)): to compute the fraction of retrieved documents that are relevant in the top K results. The position of relevant documents within the top K results is not considered; therefore this metric measures the overall user satisfaction with the top K results defined as the number of relevant documents retrieved divided by the total number of documents retrieved.

• ***Recall at K***(R@ K): to compute the fraction of relevant documents that are successfully retrieved in the top k results. This is defined as the number of relevant documents retrieved divided by the total number of relevant documents.

• ***Average Rank:*** to assess the effectiveness of the proposed re-ranking approach in terms of placing results that are most relevant to the user on top of the returned results. For a query *q* sent by user *u* with *r* defined as a collection of results returned for *q*, the Average rank is calculated as follows:

$$\text{Average Rank } (u, q) = \frac{\sum_{p \in r} p.position}{total\ number\ of\ p} \tag{8}$$

Where *p.postion* is the position of a page *p* in the ranking list and (*total number of p)* is the total number of results that are clicked by the user.

4.3 User Profile Accuracy Evaluation Results

For each user, we analyze the user profile accuracy in this experiment in terms of the average rank of non-relevant concepts for using our own dataset. The profiles were presented to each user, and they were asked to identify concepts that exist in their profiles but are not relevant to their interests. Then, each concept is ordered according to the vector length of its viewed document. As shown in table 4, each user is presented with the following data:

- The profile size (#Concepts), that is the number of concepts of the profile associated with each user.

- The average rank of non-relevant concepts in the profile associated to each user (AvgRank) using Eq. (8)

- The normalized average rank Norm_AvgRank computed by dividing the (AvgRank) of the non relevant concepts over the profile size [Pass et al. 2006].

From Table 4 , the user profile of (User 3) contains the minimum number of concepts (#Concepts), 19, with an AvgRank of non-relevant concepts, 14.73. While the user profile of (User 4) contains the maximum number of concepts, 42, with AvgRank of 26.48. Non-relevant concepts are advanced down the list when an ordering of a concept generates a large value of (AvgRank).

Users	#Concepts	AvgRank	Norm_AvgRank
User 1	38	18.99	0.51
User 2	30	17.25	0.58
User 3	19	14.73	0.78
User 4	42	26.48	0.63
User 5	36	21.07	0.59
User 6	25	16.91	0.68

Table 4 : Average Rank of non relevant concepts per user

The Norm_AvgRank evaluates the concept ordering quality with an average rate of inserting concepts that are not relevant to the user at the bottom of the conceptual user profile. For all the users, Norm_AvgRank is above 50%, and reaches the maximum value from (User 3) at 78%. Approximately, the results Fed show that the user profiles were relevant, at least at the highly weighted concepts.

We address that accuracy of mapping a page to a certain concept in the taxonomy extremely affects the user profile quality measured in terms of the AvgRank of concepts that are not relevant to the user. Obviously, a better user profile quality is achieved by accurately classifying the pages in the search history into the taxonomy. Accurate page classification results in ranking down non-relevant concepts or excluding them from the user profile representation.

4.4 Evaluation results for personalized search effectiveness

In this experiment, our own dataset is also used to discuss the effect of re-ranking search results based on their semantic similarity with concepts from the user profile discussed in section 3, against using the cosine similarity exploited by other concept-based approaches. Experiments are conducted on Windows Server 2008 R2 operating system with Intel core i5 CPU 2.4 GHz and 3 GB RAM. Given this limited working environment, the average time required for the re-ranking process is 18 seconds. In the future, we are planning to perform our experiments in an environment with more available resources to enhance the efficiency of our approach in terms of the average re-ranking response time.

The total average rank for all users is calculated to define the best value of α. It has been observed that when α is set to 0.34, it produces the best enhancement. In order to retrieve the senses of words in WordNet, we used WordNet 2.1 [Wordnet 2012] and WordNet.Net library [Simpson and Crowe 2012].Figure 9 summarizes the performance improvement of the proposed re-ranking model day by day for all users using two different methods for mapping documents to the user profile; semantic similarity and cosine similarity. Google original ranking is too used as baseline.

It is noticed that the performance improvement of the proposed re-ranking model based on the semantic similarity in Figure 9 (b) is 50.27%, which is better than those in both Figure 9(a) (35.21%) and Figure 9(c) (17.69%). The slight enhancement in Figure 9(c) for the “Clear Users” shows that traditional search engine, Google, has performed well with clear queries. However, in Figure 9(a) and Figure 9(b), the significant improvements for the “Semi-ambiguous User" and the "Ambiguous User" demonstrates that traditional search engine performs poorer than both of the similarity methods used in re-ranking.

Figure 9(d) reports the average improvement for all users. As a result of asking the participants to change the queries from their subjects of study to avocations from day 5, it is observed that the Average Rank values show a sudden increase from day 4 to day 5. However, after 3 days of learning the changes, the proposed re-ranking model based on semantic similarity produces much better results than both Google and re-ranking based on cosine similarity. More accurately, the proposed semantic based re-ranking model outperforms the cosine-based re-ranking with a 67.31% improvement from day 10.

For day 5, the improvement is only around 5.56%. This divergence demonstrates that changing user preferences will reduce the improvement that the proposed method could achieve. However, semantic-based re-ranking mechanism still greatly improves over Google and cosine-based re-ranking method overall. The average improvement of the proposed method over Google is 35.23 %.

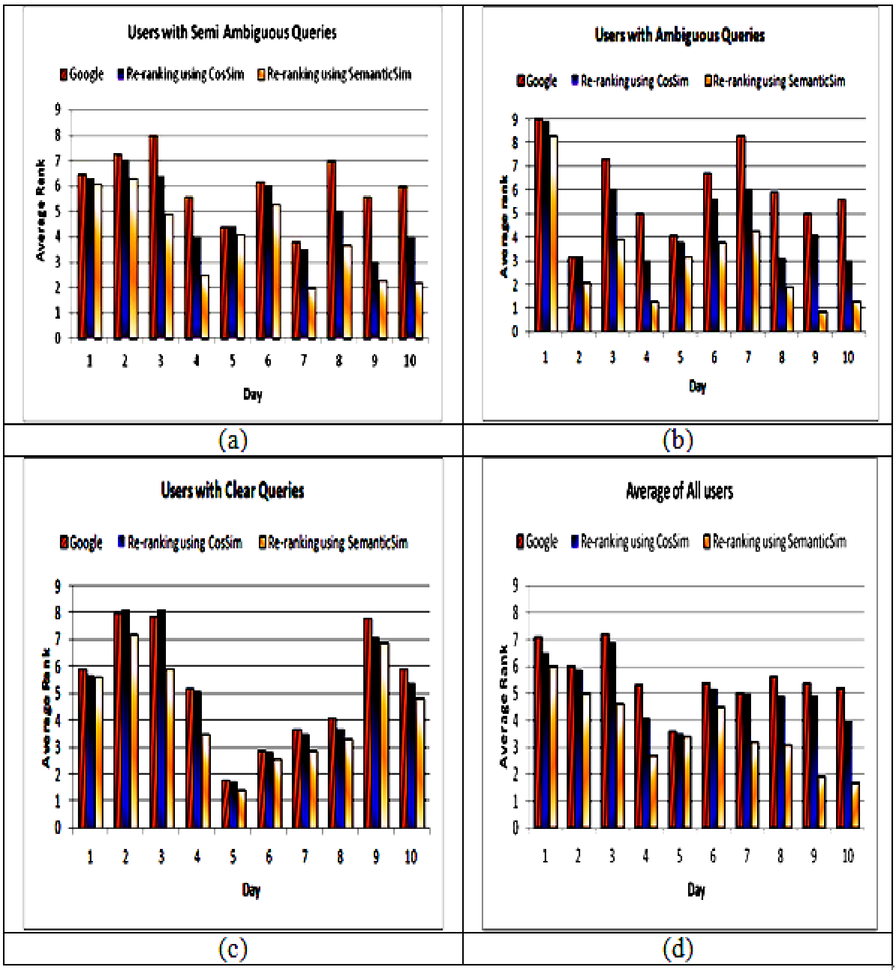

Figure 9: personalized search quality day by day (using our dataset)

4.5 Comparing the proposed personalized re-ranking approach to other ranking methods using AOL real dataset

In this experiment, we study the effect of using various information resources in order to re-rank results. The proposed re-ranking model depends on three different information resources:

- User's general interests represented by taxonomy documents from the concept hierarchy (ODP),

- User's specific interests represented by viewing documents from the user profile,

- The original ranking scores of the retrieved results.

To evaluate the efficiency of our proposed re-ranking method, we compare our approach with two different approaches in the literature according to the information resource they exploited to collect user information to re-rank search results as shown in Table 5. We also compared the proposed approach with Google original rankings. The AOL dataset is used for the purpose of this experiment assuming that the clicked results are relevant to the user.

	Utilizing clicked results only, [Antoniou et al. 2012]	**Utilizing Taxonomy documents only,** [Mohammed et al. 2010]
Description of the method used for collecting user information	Executes a concurrent re - ranking of search results of a given query when the user clicks the results with no need to store user's search history. As the user selects a result, the information included on that page is used to identify user's search needs.	Collects the web search history of a particular user implicitly which is utilized with the reference ontology to construct an initial user profile for a certain period of time. WordNet hypernyms[2] are used to extend the query content into concept hierarchy. Search documents are matched to the concepts of the query according to similarity.
Re-ranking process	Re-ranking depends on measuring the similarity of the clicked results against other results in addition to the similarity between the (ODP) categories of results.	Matching query ontology and personal profile ontology is applied to filter the search results and re-rank them by similarity scores

Table 5 : Research approaches used for comparison with our proposed approach

Figure 10 and Figure 11 show, respectively the average precision and recall for the proposed re-ranking method. The re-ranking method in [Mohammed et al. 2010] referred in the figures as "Utilizing Taxonomy docs only". Also, the re-ranking method of [Antoniou et al. 2012] referred in the figures as "Utilizing clicked results only "after the user selects the second result (RR2) and the non-personalized Google search results at top n documents. The results show that in all top-n documents, the proposed personalized re-ranking approach provided better precision and achieve the best at the top5 documents 0.8.

[2] *Y* is a hypernym of *X* if every *X* is a (kind of) *Y*, (e.g. *color* is a hypernym of *red*).

Additionally, seven levels of recall are used to calculate the graph of Precision Recall as shown in Figure 12. From this Figure, it can be observed that the proposed approach for search results re-ranking works efficiently better than other approaches, and achieves more precise results.

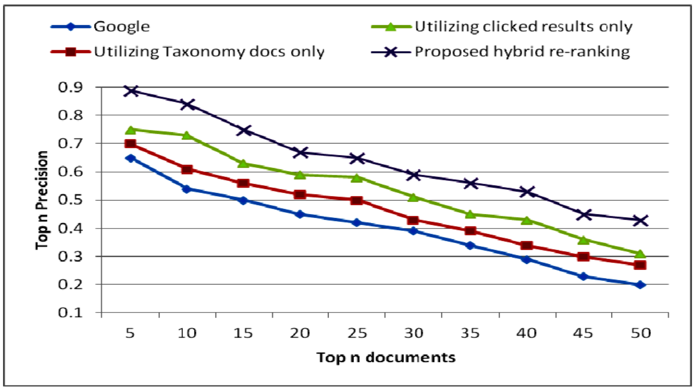

Figure 10: Average Precision for the top-n documents (using AOL dataset)

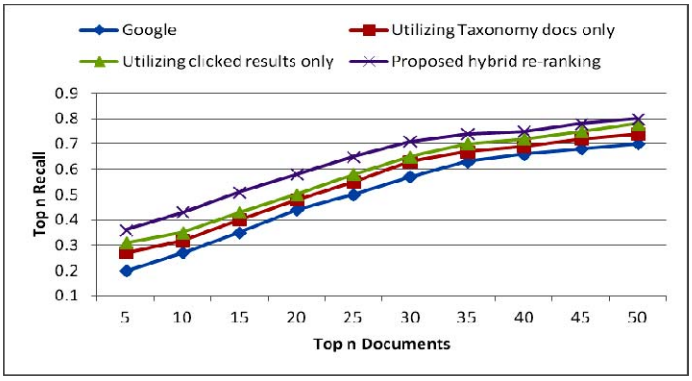

Figure 11: Average Recall for the top-n documents (using AOL dataset)

4.5.1 Discussing the benefits derived from the real world with the proposed work

From the experiment conducted above on AOL real dataset, it can be noticed that the proposed approach works effectively better than other approaches, and produces more precise results. This is because the proposed approach builds a dynamic user profile from different information sources which efficiently adapts to the changing user's preferences. Additionally, semantic mapping of search results to the user profile successfully managed to give an effective, personalized search results that meet the user's search interests.

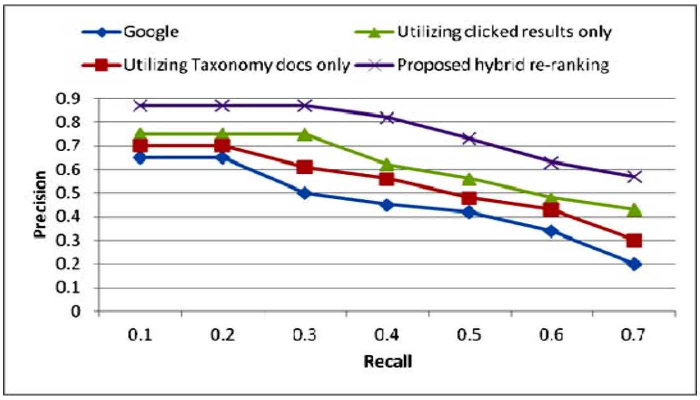

Figure 12: Precision-Recall Graph (using AOL dataset)

On contrary, the profile in [Mohammed et al. 2010] doesn't represent the updated interests and preferences of the user as it is constructed after a certain period of time. Also the personalization effectiveness in [Mohammed et al. 2010] could be negatively affected if the user submits a query word that not exists in WordNet. In addition, re-ranking is done by mapping the query context to the entire user profile without considering user's different levels of interests in different concepts of the profile.

As for the approach in [Antoniou et al. 2012], since the re-ranking is done as the user makes a choice (click), it can be expected that false choices might be clicked by users, therefore the re-ranking efficiency could be negatively affected. Even though there is another chance for the user to click again on a relevant result and correct his/her choice to re-order the results accordingly, lower positions will be assigned to the relevant results in the new ranking, and could be much lower, in case the user keeps clicking on false results. Additionally, this approach relies on the similarity between categories of two results as one of the parameters used for re-ranking search results. It returns the ODP category in which a search result belongs to. However, not all web pages are listed in the ODP categories. In this case, the personalization efficiency is affected when no categories are found for certain search results. Furthermore, this technique reorders search results as each page is viewed, but it has been proven that such strategies used for immediate updates are not well-accepted by users, even when they give more accurate results [He et al. 2007].

5 Conclusions And Future Work

Personalized web search provides users with results that accurately satisfy their specific goal and intent of the search. In this paper, a hybrid personalized search, re-ranking approach is proposed based on constructing a conceptual user profile and exploiting it in re-ranking search results. The user profile consists of concepts obtained by hierarchically classifying user's clicked search results into categories from the concept hierarchy, Open Directory Project. Each concept in the user profile consists of two types of documents; taxonomy document and viewed document. Taxonomy document is used to represent the user general interests as it contains information from web pages originally associated with such ODP category. Viewed document is used to represent the user specific interests as it contains information from web pages clicked by the user. Finally, for a given query, search results are re-ranked by semantically mapping them to the general user and specific interests from the profile together with rankings of the basic search engine.

From the experimental results, there is a significant precision improvement of the proposed personalized search system compared to the basic search performed by standard search engine. This shows the effectiveness of the user profile modeling and the effectiveness of the personalized search re-raking using different information resources.

Additionally, representing the user profile with concepts from reference taxonomy together with user's clicked pages is more accurate and reduces the ambiguity than using concepts only**.** It provides sufficient information for representing user interests for either wide topics (i.e. computer science) or particular elements (i.e. a programming language).

It is also noticed that mapping search results to a user's profile using the semantic similarity improves the personalization effectiveness over using the cosine similarity with 29%. Also, the overall improvement of the proposed model over the non-personalized search engine (Google) is 35.2%.

Finally, re-ranking search results using a hybrid of the viewed documents, taxonomy documents and the original ranking advances more relevant results on the top. This proofs that the proposed dynamic user profile efficiently adapts to the user's preferences, and successfully managed to give an effective personalized search results.

In the future, we plan to perform a large-scale experiment for longer period with more users. We can also learn other implicit information such as mouse movement, the time interval between two clicks, etc. to effectively update the user profiles. Furthermore, we plan to examine the effect of other semantic relations in the concept hierarchy on the re-ranking quality.

Acknowledgements

We would like to thank Prof. Mohammed J. Zaki for his valuable suggestions on the manuscript".

References

[Al-Maskari and Sanderson 2011] Al-Maskari, A., Sanderson, M.: "The effect of user characteristics on search effectiveness in information retrieval"; Journal of Information Processing & Management, 47, 5, (2011), 719-729.

[Andhare and Mahajan 2014] Andhare, A.A., Mahajan, N.V.: "Personalization of Web Knowledge Using ontology model"; International Journal of Advance Research in Computer Science and Management Studies, 2, 2, (2014), 348-353.

[Antoniou et al. 2012] Antoniou, D., Plegas, Y., Tsakalidis, A., et al.: "Dynamic refinement of search engines results utilizing the user intervention"; Journal of Systems and Software, 85, 7, (2012), 1577-1587.

[Baeza-Yates et al. 2011] Baeza-Yates, R., Boldi, P., Bozzon, A., et al.: "Trends in search interaction"; In: Ceri, S., Brambilla, M. (ed.), Search Computing, Springer- Verlag, Berlin, Heidelberg, (2011), 26–32.

[Bibi et al. 2014] Bibi, T., Dixit, P., Ghule, R., Jadhav, R.: "Web search personalization using machine learning techniques"; Proc. Advance Computing Conference, IEEE, (2014), 1296-1299.

[Chirita et al. 2005] Chirita, P.A., Nejdl, W., Paiu, R., Kohlschutter, C.: "Using ODP Metadata to Personalize Search"; Proc. 28th annual international ACM SIGIR conference on Research and Development in Information Retrieval, (2005), 178-185.

[Dmoz 2012] Open Directory Project, 2012, Http://www.dmoz.org

[Dotnet API 2011] Google API for dotnet, 2011, http://code.google.com/p/google-api-for-dotnet/#Google_Search_API_for_.NET_0.4_alpha

[Dou et al. 2007] Dou, Z., Song, R., Wen, J.R.: "A large-scale evaluation and analysis of personalized search strategies"; Proc.16th international conference on World Wide Web, (2007), 581-590.

[Gabrilovich and Markovitch 2007] Gabrilovich, E., Markovitch, S.: "Harnessing the expertise of 70,000 human editors: knowledge-based feature generation for text categorization"; Journal of Machine Learning Research, 8, (2007), 2297-2345.

[Gauch et al. 2007] Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: "User profiles for personalized information access"; In: Brusilovsky, P., Kobsa, A., Nejdl, W. (ed.), The Adaptive Web, Springer-Verlag, Berlin, Heidelberg, (2007), 54-89.

[Google 2012] Google Search Engine, 2012, http://www.google.com.eg/

[Google API 2012] Google Ajax API, 2012, http://googleajaxsearchapi.blogspot.com/

[Gulli and Signorini 2005] Gulli, A., Signorini, A.: "The indexableweb is more than 11.5 billion pages"; Special interest tracks and posters of the 14th international conference on World Wide Web, (2005), 902-903.

[Hawalah and Fasli 2011] Hawalah, A., Fasli, M.: "A Hybrid Re-ranking Algorithm Based on Ontological User Profiles"; Proc. 3rd Computer Science and Electronic Engineering Conference, (2011), 50-55.

[He et al. 2007]He, D., Brusiloviksy, P., Grady, J., et al.: "How up-to-date should it be? The value of instant profiling and adaptation in information filtering"; Proc. IEEE/WIC/ACM International Conference on Web Intelligence, IEEE Computer Society, (2007), 699–705.

[Hoeber and Massie 2009] Hoeber, O., Massie, C.: "Automatic topic learning for personalized re-ordering of web search results"; Proc. 6th Atlantic Web Intelligence Conference, (2009), 105-116.

[Kim et al. 2003] Kim, H. R., Chan, P. K.: "Learning implicit user interest hierarchy for context in personalization"; Proc. 8th International Conference on Intelligent User Interfaces, (2003), 101-108.

[Kumar and Sharan 2014] Kumar, R., Sharan, A.: "Personalized web search using browsing history and domain knowledge"; Proc. International Conference on Issues and Challenges in Intelligent Computing Techniques, IEEE, (2014), 493-497.

[Leacock and Chodorow 1998] Leacock, C., Chodorow, M.: "Combining local context and WordNet similarity for word sense identification"; In: Fellbaum, C., (ed.), WordNet: An Electronic Lexical Database, MIT Press, Cambridge, MA, (1998).

[Li et al. 2009] Li, L., Yang, Z., Kitsuregawa, M.: "Rank optimization of personalized search"; Detroit Baptist Seminary Journal, 4, 3, (2009), 666–671.

[Li et al. 2007] Li , L., Yang , Z., Wang , B., Kitsuregawa, M.: "Dynamic adaptation strategies for long- term and short-term user profile to personalize search"; Proc. Joint Asia-Pacific Web Conference and Web-Age Information Management Conference, (2007), 228-240.

[Madylova and Oguducu 2009] Madylova, A., Oguducu, S. G.: "A taxonomy based semantic similarity of documents using the cosine measure"; Proc. 24th International Symposium on Computer and Information Sciences, IEEE, (2009), 129-134.

[Matthijs and Radlinski 2011] Matthijs, N., Radlinski, F.: "Personalized web search using long term browsing history"; Proc. 4th ACM international conference on Web Search and Data Mining, (2011), 25-34.

[Micarelli et al. 2007] Micarelli, A., Gasparetti, F., Sciarrone, F., Gauch, S.: "Personalized search on the World Wide Web"; In: Brusilovsky, P., Kobsa, A., Nejdl, W., (ed.), The Adaptive Web, Springer-Verlag, Berlin, Heidelberg, (2007), 195–230.

[Miller 1995] Miller, G.: "WordNet: a lexical database for English"; Communications of the ACM, 38, 11, (1995), 39-41.

[Mobasher 2007] Mobasher, B.: "Data mining for web personalization"; In: Brusilovsky, P., Kobsa, A., Nejdl, W., (ed.), The Adaptive Web, Springer-Verlag Berlin, Heidelberg, (2007), 90-135.

[Mohammed et al. 2010] Mohammed, N. U., Duong, T. H., Jo, G. S.: "Contextual information search based on ontological user profile"; Proc. 2nd international Conference on Computational Collective Intelligence, (2010), 490-500.

[Oishi et al. 2008] Oishi, T., Kambara, Y., Mine, T., et al.: "Personalized search using odp-based user profiles created from user bookmark"; Proc.10th Pacific Rim International Conference on Artificial Intelligence, (2008), 839-848.

[Pass et al. 2006] Pass, G., Chowdhury, A., Torgeson, C.: "A picture of search"; Proc. 1st international Conference on Scalable Information Systems, (2006).

[Perkowitz and Etzioni 1998] Perkowitz, M., Etzioni, O.: "Adaptive web sites: automatically synthesizing web pages"; Proc. 15th national/10th conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence, (1998), 727-732.

[Porter 1980] Porter, M.: "An algorithm for suffix stripping"; Program: Electronic Library & Information Systems, 40, 3, (1980), 211-218.

[Pulijala and Gauch 2004] Pulijala, A., Gauch, S.: "Hierarchical text classification"; Proc. International Conference on Cybernetics and Information Technologies, Systems and Applications, (2004).

[Sieg et al. 2007a] Sieg, A., Mobasher , B., Burke, R.: "Learning ontology-based user profiles: a semantic approach to personalized web search"; IEEE Intelligent Informatics Bulletin,8, 1,(2007), 7- 18.

[Sieg et al. 2007b] Sieg, A., Mobasher, B., Burke, R.: "Web search personalization with ontological user profiles"; Proc. 16th ACM conference on Information and Knowledge Management, (2007), 525-534.

[Simpson and Crowe 2012] Simpson, T., Crowe, M.: "WordNet.Net"; 2012, http://opensource.ebs-wift.com/WordNet.Net

[Song et al. 2010] Song, R., Dou, Z., Hon, H.W., Yu, Y.: "Learning query ambiguity models by using search logs"; Journal of Computer Science and Technology, 25, 4, (2010), 728–738.

[Tanudjaja and Mui 2002] Tanudjaja, F., Mui, L.: "Persona: A Contextualized and Personalized Web Search"; Proc. 35th Hawaii International Conference on System Sciences, IEEE Computer Society, (2002), 1232-1240.

[Teevan et al. 2005] Teevan, J., Dumais, S. T., Horvitz, E.: "Personalizing search via automated analysis of interests and activities"; Proc. 28th annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (2005), 449–456.

[VSM 2012] (Vector Space Model), 2012, http://en.wikipedia.org/wiki/Vector_space_model

[Wordnet 2012] WordNet lexicon, 2012, http://wordnet.princeton.edu/