Personalized Web Search Ranking Based on Different Information Resources

Naglaa Fathy[1], Nagwa Badr[1], and Tarek Gharib[2]

[1] Ain shams University, Faculty of Computer and Information Sciences, Cairo, Egypt
[2] Faculty of Computing and Information Technology King Abdulaziz University Jeddah, Saudi Arabia

Abstract. The goal of personalized search is to provide user with results that accurately satisfy their specific goal of the search. In this paper, a hybrid personalized search re-ranking approach is proposed to provide users with results reordered according to their interests. User preferences are automatically learned into a concept-based user profile. This profile is then employed in the re-ranking process with other information resources to personalize results. Our experiments have shown interesting results in enhancing the quality of web search.

Keywords: Search Personalization, User Profile, ODP, Re-rank, Concept Hierarchy.

1 Introduction

In web search, users usually submit a short query consisting of a few keywords. However, given different interests of users and ambiguities in natural language, it is likely that query words of two users may appear exactly same even though information needs are different. In this paper, a hybrid personalized search re-ranking model is proposed based on learning conceptual user profiles implicitly. The user profile represents user's potential interests from his search history. More specifically, the user profile consists of user's interesting concepts obtained from a concept hierarchy of topics. This user profile is then exploited in the re-ranking process along with other information collected from the reference hierarchy and original search engine's ranking in order to personalize search results for each user.

2 Related Work

Personalized re-ranking considers user's interests to bring more relevant search results on the top few [1], [2], [3]. In [4], a vector of weighted terms from visited URLs is defined to represent each user. However, this approach cannot distinguish words having multiple meanings. The user models in [5], [6] are presented by semantic networks of user's past queries. However, these approaches may become deficient when query words do not exist in the dictionary used by the system.

V.V. Das and P.M. El-Kafrawy (Eds.): SPIT 2012, LNICST 117, pp. 211–214, 2014.
© Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2014

3 Proposed Personalized Re-ranking Architecture

This paper proposes a hybrid personalized search model that involves learning concept-based user profiles implicitly from user's search history as shown in Figure 1.

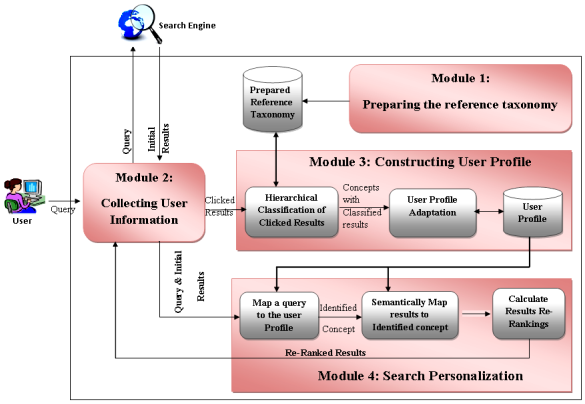

Fig. 1. Proposed Personalized Web Search Architecture

We used the Open Directory Project (ODP) [7], [8] as reference concept hierarchy of topics. Each concept is represented by vector of terms from the first 30 URLs using vector space (TF-IDF) mechanism [9]:

$$\text{Normalized term weight, } \text{ntc}_{ij} = (tf_{ij} * idf_i / \Sigma\, tc_{ij}.) \,. \tag{1}$$

Where tf_{ij} is the frequency of term i in document j and idf_i is the inverse document frequency of documents containing term i. Additionally, we implemented Google Wrapper [10] in order to implicitly collect information from users. The wrapper logs the queries, search results, and clicks per user. As subsequent step, data obtained by observing user search history are used to learn and construct concept-based user profile. The vector space classifier [9] is used in order to hierarchically classify clicked search results into ODP concepts [11]:

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

Where c_i, d_i denote the weight of term *i* associated to concept c and weight of term *i* in the document d respectively. Eventually, each concept in the user profile contains two documents. The first document is called the ***taxonomy document*** which contains the same information as in the reference taxonomy. The second document is the ***Personal document*** and it holds all the information that was extracted from a user's classified clicked search results.

When user submits a query, it is mapped to the taxonomy documents in the user profile to identify the concept that most represents the query by equation 2. Next, the semantic similarity method in [12] is employed to map search result R_i to each

document associated with the query's concept. The importance of concept's taxonomy document C_T and personal document C_P that represent query q is measured as follows:

$$\text{Score}(R_i) = \textit{CosSim}(q, C_T) * \text{SemanticSim}(R_i, C_T) + \textit{CosSim}(q, C_P) * \text{SemanticSim}(R_i, C_P) \quad (3)$$

Finally, Google original ranking is combined with the previous information resources; taxonomy and personal rankings as follows (with $0 \leq \alpha \leq 1$):

$$\text{Final Rank} = (1 - \alpha) * \text{OriginalRank} + \alpha * \text{Score}(R_i) . \quad (4)$$

4 Experimental Evaluation

In order to evaluate the effectiveness of the proposed method, 6 users with different interests were invited to search through our personalized search interface for 10 days. Average Rank [13] and Precision [14] were used as the evaluation metrics. We set α to 0.2 since it produces lower average error and provides the best improvement in personalized search.

4.1 Experimental Results

Figure 2 reports the average improvement of the proposed re-ranking model over original ranking day by day for all users with two different methods for mapping documents to the user profile; semantic similarity and cosine similarity. It is observed that the average improvement of using the semantic similarity for re-ranking over search engine's original ranking is 35.23 %. Figure 3 shows the average precision for the proposed hybrid re-ranking model, concept-only re-ranking model and non personalized Google search results. Results show that the proposed personalized re-ranking method provided better precision at all top-n documents.

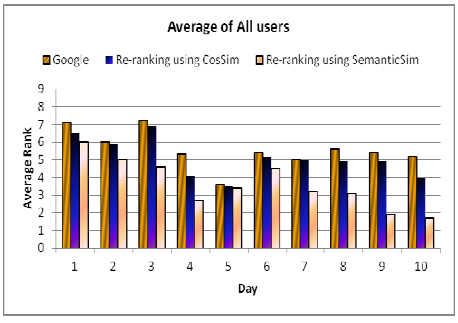

Fig. 2. Quality of Personalized Search day by day (lower is better)

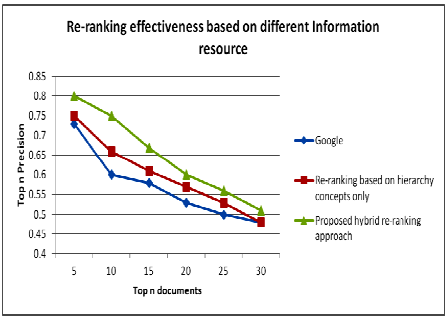

Fig. 3. Average Precision for the top-n documents

5 Conclusion and Future Work

In this paper, a hybrid re-ranking method is employed with concept-based user profile enriched with user's clicked search results to improve web search. We plan in the future to conduct a large scale experiment for longer period with more participants.

References

1. Li, L., Yang, Z., Wang, B., Kitsuregawa, M.: Dynamic Adaptation Strategies for Long-Term and Short-Term User Profile to Personalize Search. In: Proceedings of Joint Asia-Pacific Web Conference / Web-Age Information Management Conference, pp. 228–240 (2007)
2. Mohammed, N.U., Duong, T.H., Jo, G.S.: Contextual information search based on ontological user profile. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCCI 2010, Part II. LNCS, vol. 6422, pp. 490–500. Springer, Heidelberg (2010)
3. Chirita, P., Nejdl, W., Paiu, R., Kohlschutter, C.: Using ODP Metadata to Personalize Search. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 178–185 (2005)
4. Matthijs, N., Radlinski, F.: Personalizing Web Search using Long Term Browsing History. In: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, pp. 25–34 (2011)
5. Mianowska, B., Nguyen, N.T.: A Method for User Profile Adaptation in Document Retrieval. In: Proceedings of 3rd Asian Conference on Intelligent Information and Database Systems, pp. 181–192 (2011)
6. Sheng, Yan, et al.: A Personalized Search Results Ranking Method Based on WordNet. In: Proceedings of 6th International Conference on Fuzzy Systems and Knowledge Discovery, pp. 500–504 (2009)
7. Open Directory Project, `http://www.dmoz.org/` (last visit on April 2012)
8. 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, 2297–2345 (2007)
9. Vector Space Model, `http://en.wikipedia.org/wiki/Vector_space_model` (Last visit on April 2012)
10. Google, `http://www.google.com.eg/` (last visit on April 2012)
11. Pulijala, A., Gauch, S.: Hierarchical Text Classification. In: Proceedings of International Conference on Cybernetics and Information Technologies (2004)
12. Madylova, A., Öğüdücü, Ş.G.: A taxonomy based semantic similarity of documents using the cosine measure. In: Proceedings of the 24th International Symposium on Computer and Information Sciences, pp. 129–134 (2009)
13. Dou, Z., Song, R., Wen, J.-R.: A large-scale evaluation and analysis of personalized search strategies. In: Proceedings of the 16th International Conference on World Wide Web, pp. 581–590 (2007)
14. Precision, `http://en.wikipedia.org/wiki/Precision_information_retrieval` (last visit on March 2012)