

Web Search Personalization Using Ontological User Profiles

Kretika Gupta and Anuja Arora

Abstract In web, users with different interest and goal enter queries to the search engine. Search engines provide all these users with the same search results irrespective of their context and interest. Therefore, the user has to browse through many results most of which are irrelevant to his goal. Personalization of search results involves understanding the user's preferences based on his interaction and then re-ranking the search results to provide more relevant searches. We present a method for search engine to personalize search results leading to better search experience. In this method, a user profile is generated using reference ontology. The user profile is updated dynamically with interest scores whenever, he clicks on a webpage. With the help of these interest scores in the user profile, the search results are re-ranked to give personalized results. Our experimental results show that personalized search results are effective and efficient.

Keywords Personalization · Ontological user profile · Re-ranking

1 Introduction

The amount of information in world wide web has seen a phenomenal increase in the past years. In 1994, one of the first web search engines had to index 110,000 web pages approximately. Today, search engines need to deal with more than 25 billion documents. Search results retrieved by internet search engines display the same result irrespective of who has queried. A user looking for “apple” maybe interested in apple

K. Gupta (✉) · A. Arora
Department of Computer Science, Jaypee Institute of Information Technology,
Noida, India
e-mail: kretika@hotmail.com

A. Arora
e-mail: anuja.arora@jiit.ac.in

as a fruit instead of apple the company. A user has to go through irrelevant search results before he finds his required results. This irrelevant information is due to the one size fits all policy of the search engines [1]. Identical queries from different users with different interest generate same search results. Another main reason of irrelevant search results is ambiguity in query. Ambiguity can be attributed to polysemy, existence of many meanings for a single word, and synonymy, existence of many words with the same meaning. Ontology is defined as an explicit specification of conceptual categories and relationships between them [2]. Therefore, to personalize the search results, a user profile is required to map the user interest. Re-ranking of webpages is done using user profile. Many approaches have been developed to personalize web search. User preference based on the analysis of past click history was discussed in detail by Pretschner and Gauch [3] and Sugiyama et al. [4]. Short-term personalization based on a current user session was discussed by Sriram et al. [5].

2 Methodology

Reference ontology is built by using Open Directory project. A user profile is generated by annotating interest scores in the concepts provided by the reference ontology. The interest scores in the user profile created is updated dynamically whenever he clicks on a webpage. With the help of the user interest the search results are re-ranked.

2.1 User Profile Generation

The User profile is an instance of reference domain ontology. The reference domain ontology is created with the help of a web directory, Open Directory Project (ODP) [6]. A portion of ODP has been shown in Fig. 1. In this, the concepts are

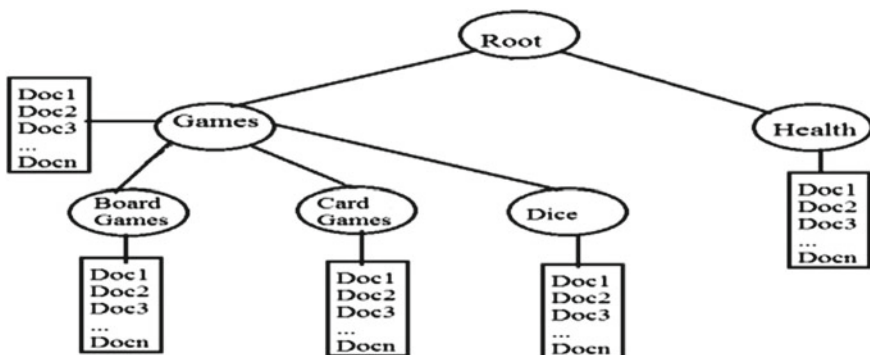


Fig. 1 Portion of an ontological profile. Each node has documents associated with it

annotated with an interest score which is updated dynamically each time the user clicks on a webpage. Open directory project is considered as the “largest human-edited directory of the web”. The data structure is organized in Directed Acyclic Graph. Each category has a set of documents associated with it which were used as a training set for classification. Text classification is required to find out under which category the content of the webpage lies in. For text classification, all the documents classified under one category in the ODP structure is merged under one super document. Whenever a user clicks on a webpage, a page vector is computed and then compared with each category’s vector in the DAG to calculate the similarities. Trajkova and Gauch [7] have calculated the similarity between Web pages visited by the user and the concepts in an ontology. The page vector is computed with the help of the title of the web page, Metadata Description Unigrams, and Metadata Keywords Unigrams associated with the webpage [8].

2.2 Updating User Profile

The User Profile for a given user saves his interests in the particular categories determined by the ODP structure. The user does not have to choose his interest areas explicitly [9]. This is automatically generated using various features which will be further discussed. The user profile is dynamic and keeps updating over time. As, whenever a user clicks on given link, the interest score is determined and updated. Since the profile is dynamically updated it takes into consideration the changing interests of a user.

Interest score is calculated with the help of the time spent, length, and subject similarity of the webpage. Time denotes the user’s duration of viewing the webpage, length denotes the number of characters in a webpage. Subject similarity denotes the similarity between the webpage’s content and the category defined by the ODP structure. As shown in Fig. 2.

$Sim(d, c_i)$ refers to the similarity of match between the content of document (d) and category (c_i) defined by ODP. Adjustment of the interest of a user in category (c_i) is $\delta(i, c_i)$. The interest score is updated with the help of the following equation, according to [3].

$$\delta(i, c_i) = \log(\text{time}/(\log \text{length})) * Sim(d, c_i) \tag{1}$$

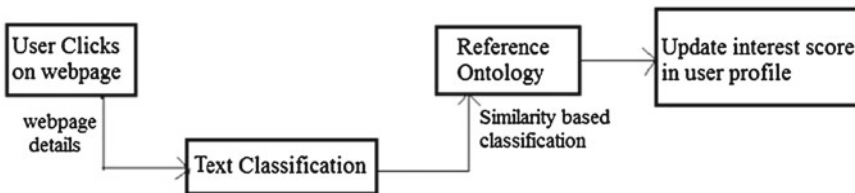


Fig. 2 Updating user profile

It can be noted that the above equation takes length into less consideration as the users can tell from a glance that the webpage is not relevant and move on to the next webpage swiftly irrespective of the length.

2.3 Re-ranking Search Results

Web search API: many commercial search engines have provided their API's so third party tools can access their search results (index). Google custom search API is used to retrieve search results for a query given by the user. These search results are retrieved with their index and are then used to re-rank web pages according to the interest scores in the generated user profile of that user.

The pages are re-ranked by a similarity matching function that computes the similarity of the retrieved result's document with each concept in the user profile's ontology to find the best matching concept.

$$\text{CSim}(\text{UserProfile}_i, \text{Result}_j) = \sum_{k=1}^N \text{wp}_{i,k} * \text{wd}_{j,k} \quad (2)$$

where,

$\text{Wp}_{i,k}$ represents the weight of concept k in the user profile,

$\text{Wd}_{j,k}$ represents the weight of concept k in the result j .

As Google applies its own PageRank algorithm, to rank websites based on their importance, we have incorporated Google's original ranking score as well. This will keep a check that we do not miss important webpages.

$$\begin{aligned} \text{FinalRank}(\text{UserProfile}_i, \text{Result}_j) \\ = \gamma * \text{CSim}(\text{UserProfile}_i, \text{Result}_j) + (1 - \gamma) \text{GRank}(\text{Result}_j) \end{aligned} \quad (3)$$

where GRank is the original rank. γ is used to combine the two ranking measures. We consider γ as 0.5 to give equal weightage to both the ranking mechanisms. If γ is 0, ranking will be done based on Google search results and if γ is 1 the ranking is done purely according to context. Each time, a user clicks on the links of the search results; the interest score is updated dynamically to determine the user's preferences. This has been represented in Fig. 3.

3 Experiments

To evaluate the effectiveness of personalized search results we need to find:

Research Question 1: (RQ1): Do the interest scores for individual concepts in ontological profile converge?

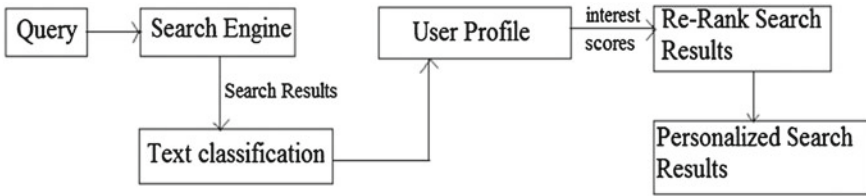


Fig. 3 Re-ranking results

Research Question 2: (RQ2): Can the interest scores maintained by the ontological profile be used to re-rank Web search to give personalized search results?

3.1 Experiment 1

With this experiment we want to evaluate RQ1, if the rate of increase in the user’s interest scores for all categories stabilizes over incremental updates [10]. The categories are defined by the user’s ontology. Each time the user clicks on a webpage the user interest are updated in the ontological user profile. Initially, the interest scores for the categories in the user profile will continue to change rapidly. However, once enough information has been collected and processed, the rate of change interest scores should decrease. Hence, we wanted to find out if over time the concepts with the highest interest scores would become relatively stable or not. For conducting the experiment, 15 users were asked to use the personalized search engines over a period of 20 days. Their user profile was monitored during these days. The number of categories the profiles converged to, changed according to the user, mainly it was in the range of 48 and 180. The Fig. 4 shows the convergence for a sample of 4 users. We can see that over time the user profile converges and becomes stable.

3.2 Experiment 2

In this experiment, we determined if the users found the personalized search results more relevant than standard web search results for RQ2. Experiment has been performed manually. To conduct this comparative experiment, whenever the user clicked on a given webpage for a query, we asked the user to mark the page as relevant or irrelevant. 15 users entered several queries over a period of 20 days. On a single search query, 12 webpages from each of the standard search engine and personalized search engine was randomly presented to the user. Few pages were marked as “both”, if they were common to both the search engines. By looking at the log of the user, it was determined how many relevant webpages the users clicked on from each.

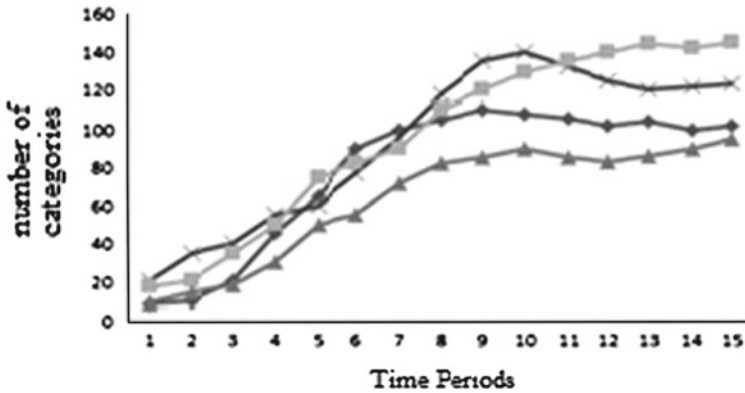


Fig. 4 Convergence of profiles

The proposed personalized search results were 55% more relevant than the normal search results for the user searches.

4 Conclusion

This paper proposed a method for a search engine to personalize search results based on a user's preferences. The user preferences were mapped to a user profile. It was shown with the help of experiments that over time, the interest got converged. With the help of the user profile, web search results can be re-ranked leading to more relevant results for the users. In future, we plan to optimize our search engine for more relevant results. We would also look into the location based information of user to provide better search results.

References

1. Allan, J., et al.: Challenges in information retrieval and language modeling. *ACM SIGIR Forum* 37(1), 31–47 (2003)
2. Sieg, A., Mobasher, B., R. Burke.: Web search personalization with ontological user profiles. In: *Proceedings of CIKM* (2007)
3. Pretschner, A., Gauch, S.: Ontology based personalized search. In: *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence*. Chicago, IL, pp. 391–298. IEEE Computer Society (1999)
4. Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive web search based on user profile constructed without any effort from user. In: *Proceedings of the 13th International Conference on World Wide Web*. New York, pp. 675–684. (2004)
5. Sriram, S., Shen, X., Zhai, C.: A session-based search engine. In: *Proceedings of SIGIR(2004)*
6. Open Directory Project - <http://dmoz.org>

7. Trajkova, J., Gauch, S.: Improving ontology-based user profiles. In: Proceedings of the Recherched'Information Assiste par Ordinateur, RIAO 2004, pp. 380–389. University of Avignon (Vaucluse), France, April (2004)
8. Chirita, P., Firan, C., Nejdl, W.: Summarizing local context to personalize global web search. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM 2006, pp. 287–296. Arlington, VA, November (2006)
9. Teevan, J., Dumais, S., Horvitz, E.: Personalizing search via automated analysis of interests and activities. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2005, pp. 449–456. Salvador, Brazil, August (2005)
10. Liu, F., Yu, C., Meng, W.: Personalized web search for improving retrieval effectiveness. *IEEE Trans. Knowl. Data Eng.* **16**(1), 28–40 (2004)