Search Using Personal Profiles

Peter CÍCH*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
peter.cich@gmail.com

For Internet users, as well as users of other extensive networks which accommodate vast amounts of different kinds of information, it is at present becoming increasingly more difficult to find the exact piece of information they need. There arises an information overflow which now presents a new, quickly spreading problem. Personalisation of the web provides one way of facing this new informational phenomenon.

This project analyses several possible approaches towards the creation of a personal user profile on the basis of which search results may then be reordered – personalised. It describes the basic principles of their functioning and also the use of a thusly created personal user profile. The personalisation process of the search results and also the user profile creation are set up by using bookmarks of web pages marked by user.

In my proposed solution the user profile is created and constructed on the basis of the user's favourite web pages he has marked as something interesting. These, for user special web pages, can be simply distinguished from the other pages because the very user has voluntarily marked and saved them – they are usually known as the user's bookmarks. Because of that we can say that user's bookmarks represent the user's long-term interests and according to this information we can with some higher probability predict what the user could be interested in, in the near future.

Long-term interests are assumed as a set of bookmarks, representing the list of the web pages that are presumable closer to the user than the others. We can also assume that if the user will not be feeling the necessity of adding the page to his bookmarks, than there is also the expectation that that page is representing only his temporary interest instead of the long-term.

This personalisation of the searching consists of the modification of returned searching results. This kind of personalisation is based on the reordering of standard returned list of query conforming pages – they are ordered by the similarity to the pages stored as user's bookmarks – his profile. When reordering the search results it is

* Supervisor: Pavol Návrat, Institute of Informatics and Software Engineering

Spring 2010 PeWe Workshop, April 18, 2010, pp. 77–78.

also important to realize that according to [2] the standard user at average looks only through the first ten returned results.

Our bookmark user profile consists of words – terms, which are combined as term vectors representing each bookmarked document (web page). These document term vectors and their terms are in user profile divided (marked as) into three groups:

- new bookmarks – newly added bookmarks to the profile
- long-term bookmarks – the oldest bookmarks in profile
- deleted bookmarks – bookmarks removed by user

When the user has added his new bookmark the weight of its key words (terms from appropriate document term vector) has to be higher than the weight of terms that are in profile for longer time. These higher ratings of fresh added pages characterize and capture the most up-to-date interests of user. Therefore the relevant domain that user has assigned to the group of interests has to have the highest impact on influencing the returned results by the next couple of queries.

The weight of terms assigned to the particular bookmark has to decrease by the time because in the most cases also the interests of users changes over time. In the bookmark profile the weight is sequentially lowered to the specified minimal weight – threshold. This threshold weight of terms then represents long-term user interests that are practically creating almost the non-altering core of the user profile.

Under the threshold weight there are terms of deleted bookmarks. These terms are still useful (for specific time) despite the user has removed them from his profile. It is necessary to store these terms for a while because although the page containing them was removed, the user has been interested in that domain in the past (the page was bookmarked by the user). We cannot exactly declare that right at the moment of deletion, the page is totally out of user's interests and everything associated with it should be deleted, too.

The basic part of weight of each term of term vector is computed from normalised weight of term in the given document (known as Euclid normalisation). Then the similarity between the standard results and the user profile, consisting of favourite pages is based on the algorithm of Cosine Similarity Measure.

To the profile similar pages are then pushed to the higher positions in the result list and thus we can say that with this approach there is some kind of personalisation.

References

[1] Abiteboul, S., Preda, M., Cobena, G.: Adaptive On-Line Page Importance Computation. In: *Proc. of the 12th Int. Conf. on World Wide Web* (WWW'03), New York, 2003, s. 280-290.

[2] Guo, Y. Z., Ramamohanarao., Park, L. A. F.: Personalized PageRank for Web Page Prediction Based on Access Time-Length and Frequency. In: *IEEE/WIC/ACM Int. Conf. on Web Intelligence* (WI'07), 2007, s. 687-690.

[3] Sieg, A., Monbaster, B., Burke, R.: Ontological User Profiles for Representing Context in Web Search. In: *Web Intelligence/IAT Workshops 2007*, s. 91-94.