Improving Query and Result List Adaptation in Personalized Multilingual Information Retrieval

M. Rami Ghorab

Centre for Next Generation Localisation,
Knowledge & Data Engineering Group, Trinity College Dublin, Dublin 2, Ireland.
ghorabm@scss.tcd.ie

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval.

General Terms

Algorithms, Experimentation, Languages.

Keywords

Personalized Multilingual Information Retrieval, User Modeling.

1. Extended Abstract

A general characteristic of Information Retrieval (IR) and Multilingual IR (MIR) [5] systems is that if the same query was submitted by different users, the system would yield the same results, regardless of the user. On the other hand, Adaptive Hypermedia (AH) systems operate in a personalized manner where the services are adapted to the user [1]. Personalized IR (PIR) is motivated by the success in both areas, IR and AH [4]. IR systems have the advantage of scalability and AH systems have the advantage of satisfying individual user needs. The majority of studies in PIR literature have focused on monolingual IR, and relatively little work has been done concerning multilingual IR.

This PhD research study aims to improve personalization in MIR systems, by improving the relevance of multilingual search results with respect to the user and not just the query. The study investigates how to model different aspects of a multilingual search user. Information about users can be demographic information, such as language and country, or information about the user's search interests. This information can be gathered explicitly by asking the user to supply the required information or implicitly by inferring the information from the user's search history. The study will then investigate how to exploit the modeled user information to personalize the user's multilingual search by performing query and result list adaptation. The main research questions that are addressed in this study are: how to improve the relevance of search results with respect to individual users in PMIR and how to construct profiles that represent aspects and interests of a multilingual search user.

So far, the work carried out for this study included: (1) a proposed framework for the delivery and evaluation of PMIR [3]; and (2) exploratory experiments with search history and collection (result) re-ranking on a dataset of multilingual search logs [2]. The next stage of experimentation will involve the investigation and development of algorithms for: (1) constructing multilingual user profiles; (2) pre-translation and post-translation query expansion based on terms from the user profile; and (3) result list re-ranking based on the user's interests, and preferred language.

Two types of experiments will be conducted in an *in-lab* setting, with a group of users from different linguistic backgrounds. In the first set of experiments, users will be asked to use a baseline web search system for their daily search activities over a period of time. The baseline system will be wrapped around one of the major search engines. Interactions with the system will be logged, and part of this information will be used for training the system (constructing user profiles from text of queries and clicked documents); the other part (remaining queries) will be used for testing the effectiveness of the query adaptation and result list adaptation algorithms, where the users will be asked to provide some personal relevance judgements. In the second set of experiments, the users will be asked to use the PMIR system to fulfill a number of defined search tasks.

Quantitative and qualitative techniques will be used to evaluate different aspects of the experiments, including: (1) retrieval effectiveness, which can be measured using standard IR metrics; (2) user's performance on search tasks, which can be measured in terms of time and number of actions needed to fulfill the tasks; (3) user profile accuracy, which can be assessed by questionnaires that indicate how well the user profile depicted the users' search interests; and (4) usability and user satisfaction, which can be assessed using standard system usability questionnaires.

2. ACKNOWLEDGMENTS

3. REFERENCES

[1] Brusilovsky, P., Kobsa, A. and Nejdl, W. (eds.). *The Adaptive Web - Methods and Strategies of Web Personalization*. Springer, 2007.

[2] Ghorab, M.R., Leveling, J., Zhou, D., Jones, G.J.F. and Wade, V. Identifying Common User Behaviour in Multilingual Search Logs. in Peters, C., Di Nunzio, G., Kurimo, M., Mandl, T., Mostefa, D., Peñas, A. and Roda, G. eds. *Lecture Notes in Computer Science (6241/2010), Multilingual Information Access Evaluation I*, Springer, 2010.

[3] Ghorab, M.R., Zhou, D., O'Connor, A. and Wade, V. A Framework for Cross-language Search Personalization *4th International Workshop on Semantic Media Adaptation and Personalization (SMAP)*, San Sebastian, Spain, 2009, 15-20.

[4] Micarelli, A., Gasparetti, F., Sciarrone, F. and Gauch, S. Personalized Search on the World Wide Web. in Brusilovsky, P., Kobsa, A. and Nejdl, W. eds. *The Adaptive Web*, Springer, 2007, 195-230.

[5] Oard, D.W. Multilingual Information Access. *Encyclopedia of Library and Information Sciences, 3rd Edition*. 3682 – 3687.