Does Personalization Benefit Everyone in the Same Way? Multilingual Search Personalization for English vs. Non-English Users

M. Rami Ghorab, Séamus Lawless, Alexander O'Connor, Vincent Wade

Centre for Next Generation Localisation, Knowledge & Data Engineering Group,
School of Computer Science & Statistics, Trinity College Dublin, Ireland
[rami.ghorab, seamus.lawless, alex.oconnor,
vincent.wade]@scss.tcd.ie

Abstract. The web community is witnessing an increase in the amount of available multilingual content and the number of multilingual web users. With this variety, personalized search systems are needed to connect people with relevant content, regardless of the language in which the content is provided, and taking into consideration the user's language capabilities and preferences. Therefore, search personalization algorithms should be developed with the aspect of multilinguality in mind, part of which involves understanding the effect of personalization algorithms on the user's search experience. This leads to an important question: given that users come from different linguistic backgrounds and have different language preferences, would personalization benefit all search users in the same way? This paper addresses this question by conducting an experiment to: (1) evaluate the effectiveness of the multilingual personalization algorithms (multilingual user modeling and multilingual result adaptation); and (2) determine whether multilingual personalization algorithms achieve the same degree of effectiveness for users who have different language preferences.

Keywords: Personalization, Multilingual Search, Information Access.

1 Introduction

The web is becoming increasingly multilingual, with respect to both content[1] and users[2]. Nearly half of the content available on the web is provided in languages other than English, such as Russian (6%), Spanish (5%), Chinese (4%), Japanese (4%), and Arabic (3%). The best answer to a user's query may not necessarily be available in his/her own language, but may reside in the diverse, multilingual corpora of the web. This calls for solutions that not only assist the users in finding relevant information,

1 http://en.wikipedia.org/wiki/Languages_used_on_the_Internet
2 http://www.internetworldstats.com/

but also enables them to easily and readily access this information if it was provided in a language that they don't comprehend.

Web search engines (and in general, Information Retrieval systems) employ various personalization techniques in order to satisfy the user's query (information need) [1, 2]. With multilinguality becoming an important dimension of the information finding/access process, search personalization techniques, in turn, have to be extended into the multilingual dimension. In specific, they have to be extended with respect to users and search results. In terms of users, the system has to cater for the user's language preferences/capabilities and adapt to the user's search interests across these languages (multilingual search interests). In terms of search results, the system has to take into consideration that, in multilingual search [3], the search results come from multiple languages; thus, the system has to adapt the way these multilingual search results are presented to the user –for example by blending the results into a single list and/or translating the results to the user's preferred language where necessary [4].

A key question facing the extension of personalized search into the multilingual dimension is: given that users come from different linguistic backgrounds and have different language preferences, *would personalization benefit all users in the same way?* In other words, *would the search personalization algorithms achieve the same degree of improvements for all queries, regardless of query language?*

This paper addresses this research question by carrying out an experiment to evaluate the retrieval effectiveness of multilingual search personalization algorithms with respect to English vs. Non-English user queries. This entails developing algorithms for multilingual user modeling and multilingual result-list adaptation, and evaluating these algorithms via a user study. The study involves users coming from different linguistic backgrounds, engaging with a web search system that facilitates access to search results from multiple languages. The evaluation results show that Non-English users benefit more from the search personalization process than English users.

The rest of this paper is organized as follows. Section 2 provides background and related work. Section 3 presents the algorithms for constructing user models that cater for the user's search interests across languages and for re-ranking multilingual search results. Section 4 discusses the experimental setup and the experimental results. Finally, conclusion and future work are presented in Section 5.

2 Background and Related Work

Textual search is prominent on the web, being used in search engines [5], digital libraries [6], or local search facilities provided on numerous websites. A natural characteristic of traditional search systems is that if different users submit the same query, the system would yield the same list of results, regardless of the user. Personalized search systems, on the other hand, include the user in the equation [2, 7]; they retrieve results that are not only relevant to the query alone, but that are also relevant to the user. This can be achieved by keeping track of the user's interests and preferences, and then using this information to adapt the search results. This personalization approach has shown success in several studies in the literature [8, 9].

A key component of personalized search systems is the user model, which keeps track of information about the user such as demographic data, prior knowledge, and search interests [10-12]. Some systems represent this information in an individualized manner [13, 14], while other systems maintain an aggregate view of usage information across the cohort of system users [9, 15]. For personalized search systems, the user's search interests are inferred by analyzing the user's search history: extracting keywords from queries that the user submitted and results that the user clicked [8].

The user models in the aforementioned studies represented the users' search interests in a monolingual fashion. It is not an uncommon case in today's world to have users who are familiar with multiple languages. For example, many internet users from various countries are familiar with English in addition to their native language. Moreover, some countries, such as Switzerland, South Africa, and Canada are naturally multilingual. This paper argues that taking the aspect of multilinguality into consideration significantly affects the way user information is gathered, modelled, and employed for the delivery of a personalized search service. Furthermore, the paper argues that the differences in the users' linguistic backgrounds –and accordingly, their language preferences– affects the degree to which they benefit from personalization.

3 Personalization Algorithms for Multilingual Search

3.1 Modeling the User's Search Interests Across Languages

For Multilingual Information Access systems in general, and Multilingual Search systems in specific, multilinguality is present in two aspects: *(1) users:* in terms of the languages they understand and in terms of their choice of query language when using the search system; and *(2) content:* in terms of the documents that are retrieved from multiple languages. In order for user models to cater for the interests and attributes of multilingual search users, they have to be re-designed with multilinguality in mind.

The user model proposed in this research, which was briefly presented in [16], captures two types of information about the user: demographic information and interest information (i.e. terms that represent the user's search interests across languages – inferred from the user's search history). The nature of this information affects the kind of **attributes** represented in the model as well as the **structure** of the model. **In terms of attributes**, the proposed design includes the following set of attributes:

1. *Native Language:* the user's native language.
2. *Familiar Languages:* a list of languages that the user understands[3].
3. *Preferred Language:* this language is used for the following in the experiment:
   (a) Search results that come from languages that the user is not familiar with are translated to this language.
   (b) The search interface is displayed in this language (menu items, labels, etc.).

[3] In the experiment reported in Section 4, the users were asked to enter a list of languages in which they had moderate proficiency or higher.

In terms of structure (i.e. how the user's multilingual search interests are stored in the user model), the interest terms are grouped by language and are maintained in their original form (without translation). That is, the user model stores the terms in multiple languages, where a term is maintained in the same language of the document or query from which it was extracted. Thus the model is made up of language fragments (language groups); each fragment holding interest terms that correspond to its language. The terms within a fragment are divided along one or more clusters of related terms. The underlying assumption of this user model representation is that the users' search interests are language-biased (distributed across languages), and therefore more effective personalization may be achieved if the user model reflects this phenomenon. Accordingly, the design of the result adaptation algorithms involves making dynamic decisions regarding which fragment(s) to use in the personalization process (when attempting the re-ranking of the multilingual search results).

3.2 Adapting Multilingual Search Results

Result adaptation involves merging and/or re-ranking the search results coming from multiple languages (e.g. operating on three lists of results: English, French, and German) based on the user's interests. It also involves translating the results before displaying them to the user –where necessary.

The result adaptation algorithm performs merging (interleaving) and re-ranking of the results based on the similarity between the search results and the user model interests. To do this, each result is assigned a score based on its textual similarity with the interest terms present in the corresponding language fragment of the user model (e.g. scoring the results of the French list against the group of French terms in the model, and therefore, no translation is required). All the results are then gathered together in a single list and then sorted in descending order of the assigned scores.

In multilingual search, translation plays a crucial role in the adaptation and presentation of results to the user, where the snippets (titles and summaries of the results) and the whole documents may have to be instantly translated to a target language.

4 Evaluation

4.1 Objectives

The objectives of this experiment are:

- *To quantitatively evaluate the retrieval effectiveness of the multilingual personalization algorithms discussed in the previous section.* This is measured using the Mean Average Precision (MAP) metric, which is a well-known IR metric that rewards lists where relevant documents appear at higher positions.
- *To determine whether the multilingual personalization algorithms achieve the same effectiveness for users who have different language preferences* (with respect to the language attributes mentioned in subsection 3.1). This is made evident by comparing the MAP scores at cut-off points for English vs. Non-English queries.

4.2 Experiment

Experimental System. The experiment was conducted online using the framework described in [17], which is a system for the delivery and evaluation of Personalized Multilingual Information Retrieval services. The framework was set to provide a multilingual Web-search service, where it was configured to interface with the Search API of one of the major search engines. Furthermore, the framework was also configured to carry out machine translation using the Translation API of that company.

Experimental Setup. The experiment took place over three phases. In the first phase, 76 users from different linguistic backgrounds (participants from different countries) were asked to use the multilingual web-search system to complete a number of search tasks. This was a baseline system that provided textual, non-personalized search results from three languages: English, French, and German, where the results were merged (interleaved) on a round-robin basis. The system logged the submitted queries and the clicked results.

The second phase took place without user participation. In this phase, the last query submitted by the user in each task was reserved for testing. The remaining queries, along with their associated clicked results, were used to train the user models. A pool of results was then automatically generated for each test query by submitting it to the search system multiple times using the baseline algorithm and various personalization algorithms that adapt the query and the results. This paper is specifically concerned with the result adaptation algorithm mentioned in Section 3.2, and so the other personalization algorithms tried in the experiment are not discussed in this paper.

The third phase involved the participation of the same users of the first phase. Each user was shown his/her test queries along with the associated pool of results. The users were asked to judge the degree of relevance of each result in the pool. The judgments were carried out on a 4-point scale (*not relevant*, *somewhat relevant*, *relevant*, or *very relevant*)[4]. Finally, the retrieval effectiveness of each personalization algorithm was evaluated according to the relevance judgments provided by the users.

Table 1. Final dataset description

Item	Number
Total Users	76
English	*56*
French	*10*
German	*10*
Total Test Queries	98
English	*75*
French	*12*
German	*11*
Total results judged	6,775

[4] As MAP operates on binary relevance judgments, the 4-point-scale judgments were converted to 2-point by taking the higher 2 judgments as Relevant and the lower 2 as Irrelevant.

Dataset Description. Some data cleaning operations were carried out. For example, the following queries were deleted from the dataset: malformed queries (e.g. character encoding issues) and queries for which no results were clicked (assumption of incomplete search sessions). Table 1 reports a breakdown of the number of users and queries in the final dataset; the language of the user refers to the language that he/she specified as the *preferred language* when he/she signed up with the system[5]).

4.3 Results

To analyze and compare the MAP scores, the users were grouped into the following two subsets, based on the preferred language:

1. *English users:* these are the users who selected English as their preferred language.
2. *Non-English users:* users who selected French or German as preferred language.

Fig. 1 shows the comparison of the MAP improvement percentages over the baseline for various result-list positions (cut-off points: @5 to @20)[6].

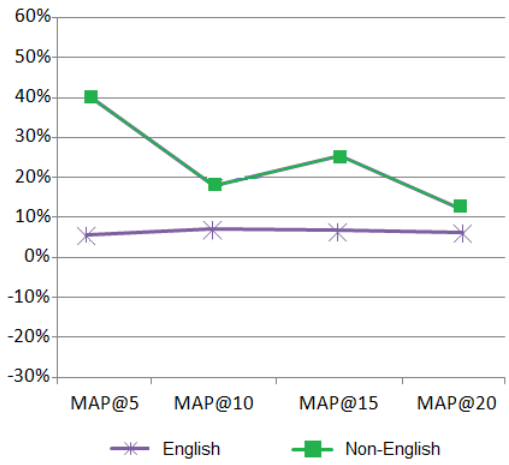

Fig. 1. MAP improvements for English vs. Non-English

The evaluation shows higher improvements for the adaptation algorithms with Non-English users (more than double the improvement at some points). This indicates that personalization benefits Non-English users much more than it benefits English users.

In order to gain more insight into this observation, the retrieval effectiveness of the baseline (non-personalized) algorithm was examined using the Precision metric. Table 2 reports the Precision scores of the baseline lists for English and Non-English users at various list positions.

---

[5] Caveat: the selection of a preferred language does not necessarily imply the native language (or the linguistic background) of the user. It is also important to mention here that the system only allowed users to choose 1 of 3 preferred languages: English, French, or German.

[6] The improvements reported for MAP@20 for English and MAP@15 for Non-English are statistically significant as per the 2-tailed T-test, with p=0.05.

Table 2. Baseline Precision scores

List Position	English	Non-English	Percentage of English over Non-English
P@5	0.58	0.45	29.15%
P@10	0.55	0.49	11.54%
P@15	0.51	0.45	14.46%
P@20	0.50	0.48	3.71%

The baseline Precision scores show that Non-English users received results with lower relevance than for English users. This suggests that there was more room for the adaptation algorithms to improve over the baseline for Non-English. In other words, one of the reasons why a lower improvement percentage was exhibited for English is that the effectiveness of the baseline algorithm was relatively higher; this provided less opportunity for the adaptation algorithm to improve over the baseline.

5 Conclusion and Future Work

This paper evaluated the effectiveness of multilingual search personalization algorithms with respect to English vs. Non-English users. The study involved 76 users from different linguistic backgrounds. The paper posed the question of whether personalization benefits all users in the same way; to which the evaluation showed that this is not the case for users of multilingual search. Based on this, we recommend that personalized search system adopt different personalization strategies for certain languages or groups of languages (as in, what works for one language, may not necessarily work for another). Future work involves developing the personalization algorithms further. In terms of user modeling, a viable extension to this study would be to use concept-based user models, where the user's interests are not stored as just words, but rather mapped on a specific vocabulary of an ontology; one that encompasses the aspect of multilinguality. In terms of result adaptation, future work involves investigating alternative ways of merging the multilingual results based on the user model.

Acknowledgements

This research is supported by the Science Foundation Ireland (Grant 12/CE/I2267) as part of the Centre for Next Generation Localisation (www.cngl.ie) at Trinity College, Dublin.

References

1. Stamou, S., Ntoulas, A.: Search Personalization Through Query and Page Topical Analysis. User Modeling and User-Adapted Interaction 19, 5-33 (2009)

2. Micarelli, A., Gasparetti, F., Sciarrone, F., Gauch, S.: Personalized Search on the World Wide Web. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web, vol. 4321, pp. 195-230. Springer (2007)
3. Peters, C., Braschler, M., Clough, P.: Multilingual Information Retrieval - from research to practice. Springer Berlin Heidelberg (2012)
4. Petrelli, D., Beaulieu, M., Sanderson, M., Demetriou, G., Herring, P., Hansen, P.: Observing users, designing clarity: A case study on the user-centered design of a cross-language information retrieval system. Journal of the American Society for Information Science and Technology 55, 923-934 (2004)
5. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval: The Concepts and Technology behind Search (2nd Edition). Addison-Wesley (2011)
6. Agosti, M.: Digital Libraries. In: Melucci, M., Baeza-Yates, R. (eds.) Advanced Topics in Information Retrieval, vol. 33, pp. 1-26. Springer Berlin Heidelberg (2011)
7. Silvestri, F.: Mining Query Logs: Turning Search Usage Data into Knowledge. Foundations and Trends in Information Retrieval 4, 1-174 (2010)
8. Ghorab, M.R., Zhou, D., O'Connor, A., Wade, V.: Personalised Information Retrieval: Survey and Classification. User Modeling and User-Adapted Interaction 23, 381-443 (2013)
9. Agichtein, E., Brill, E., Dumais, S.: Improving Web Search Ranking by Incorporating User Behavior Information. 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 19-26 ACM, Seattle, Washington, USA (2006)
10. Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A.: User Profiles for Personalized Information Access. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web, vol. 4321, pp. 54-89. Springer (2007)
11. Brusilovsky, P.: Adaptive Hypermedia. User Modeling and User-Adapted Interaction 11, 87-110 (2001)
12. Hu, J., Chan, P.K.: Personalized Web Search by Using Learned User Profiles in Re-ranking. Workshop on Web Mining and Web Usage Analysis (WebKDD 2008), pp. 84-97. ACM, Las Vegas, Nevada, USA (2008)
13. Speretta, M., Gauch, S.: Personalized Search based on User Search Histories. IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005), pp. 622-628, Compiegne University of Technology, France (2005)
14. Zhang, H., Song, Y., Song, H.-t.: Construction of Ontology-Based User Model for Web Personalization. Lecture Notes in Computer Science. 11th International Conference on User Modeling (UM 2007), pp. 67-76, Corfu, Greece (2007)
15. Smyth, B., Balfe, E.: Anonymous Personalization in Collaborative Web Search. Information Retrieval 9, 165-190 (2006)
16. Ghorab, M.R., Lawless, S., O'Connor, A., Zhou, D., Wade, V.: Multilingual vs. Monolingual User Models for Personalized Multilingual Information Retrieval. In: Carberry, S., Weibelzahl, S., Micarelli, A., Semeraro, G. (eds.) 21st Conference on User Modeling, Adaptation and Personalization (UMAP 2013), vol. 7899, pp. 356-358. Springer Berlin Heidelberg, Rome, Italy (2013)
17. Ghorab, M.R., Leveling, J., Lawless, S., O'Connor, A., Zhou, D., Jones, G.J.F., Wade, V.: Multilingual Adaptive Search for Digital Libraries. International Conference on Theory and Practice of Digital Libraries, vol. 6966, pp. 244-251. Springer Berlin Heidelberg, (2011)