Joint Workshop on Personalised Information Access - PIA 2014

Ben Steichen[1], Tuukka Ruotsalo[2], Maristella Agosti[3], Giulio Jacucci[4], Séamus Lawless[5], Peter Brusilovsky[6], Vincent Wade[5], Samuel Kaski[2], Oswald Barral[4]

[1] University of British Columbia, Canada
steichen@cs.ubc.ca

[2] Aalto University, Finland
tuukka.ruotsalo@tkk.fi, samuel.kaski@aalto.fi

[3] University of Padua, Italy
agosti@dei.unipd.it

[4] University of Helsinki, Finland
{giulio.jacucci, barralme}@cs.helsinki.fi

[5] Trinity College Dublin, Ireland
{seamus.lawless, vincent.wade}@scss.tcd.ie

[6] University of Pittsburgh, USA
peterb@pitt.edu

1 Preface

We are pleased to introduce the proceedings of the Joint Workshop on Personalized Information Access, held in conjunction with the 22nd Conference on User Modeling, Adaptation and Personalization, UMAP 2014, on the 7th of July 2014. This unique workshop was a result of merging two workshops with overlapping topics - the First Workshop on Personalized Multilingual Information Access (PMIA 2014), and the First Workshop on Personalizing Search - From Search Engines to Exploratory Search Systems (PESE 2014):

- The PMIA 2014 workshop was designed to share, discuss, and combine ideas for novel solutions that support users according to their particular language abilities, as well as other characteristics (e.g. culture, domain expertise) and contexts (e.g. intent, topic) that influence what and how information should be retrieved, composed, and presented.
- The PESE 2014 workshop was designed to explore another subtopic of personalized information access: addressing the challenges in user modeling when aiming to bring personalization to complex exploratory search tasks.

During the reviewing process, the organizers discussed the overlapping nature of the workshops and a broader scope of interesting submissions, and decided that it would be most appropriate to merge these workshops under a broader topic of personalized information access.

This volume contains the revised accepted papers from among those submitted to PMIA 2014 and PESE 2014. The organizers hope that the workshop results will directly influence the design of personalized applications that support more effective access to knowledge and deliver users search experiences which are tailored to their information needs and contexts.

The organizing committee would like to thank those institutions and individuals who have made this workshop possible: the UMAP 2014 Conference, and in particular Rosta Farzan (University of Pittsburg, USA) and Robert Jäschke (University of Hannover, Germany) who in their role as UMAP 2014 Workshop Chairs have supported us with their critical comments and suggestions; the Program Committee members for their valuable work of evaluation of the submissions that was timely, despite the associated time pressures. We would also like to thank our institutions for having supported us in this endeavor: the University of British Columbia (Canada), Aalto University (Finland), the University of Padua (Italy), the University of Helsinki (Finland), Trinity College Dublin (Ireland) and the University of Pittsburgh (USA).

2 Program Committee

- Nick Belkin (Rutgers University, USA)
- Peter Brusilovsky (University of Pittsburgh, USA)
- Luanne Freund (University of British Columbia, Canada)
- Giulio Jacucci (University of Helsinki, Finland)
- Samuel Kaski (Aalto University, Finland)
- Alessandro Micarelli (Roma Tre University, Italy)
- Tuukka Ruotsalo (Aalto University, Finland)
- Ben Steichen (University of British Columbia, Canada)
- Maristella Agosti (University of Padua, Italy)
- Séamus Lawless (Trinity College Dublin, Ireland)
- Vincent Wade (Trinity College Dublin, Ireland)
- Oswald Barral (University of Helsinki, Finland)
- Omar Alonso (Microsoft Research, USA)
- Leif Azzopardi (University of Glasgow, UK)
- Paul Buitelaar (Insight, National University of Ireland, Galway, Ireland)
- David Carmel (Yahoo Labs Haifa, Israel)
- Ed Chi (Google, USA)
- Philipp Cimiano (University of Bielefeld, Germany)
- Nicola Ferro (University of Padua, Italy)
- Bo Fu (University of Victoria, Canada)
- Norbert Fuhr (Universität Duisburg-Essen, Germany)
- Julio Gonzalo (UNED, Spain)
- Donna Harman (NIST, USA)
- Gareth Jones (Dublin City University, Ireland)
- Jaap Kamps (University of Amsterdam, The Netherlands)
- Noriko Kando (NII, Japan)

- Udo Kruschwitz (University of Essex, UK)
- Killian Levacher (Trinity College Dublin, Ireland)
- Walid Magdy (Qatar Computing Research Institute, Qatar)
- Mandar Mitra (Indian Statistical Institute, India)
- Jian-Yun Nie (University of Montréal, Canada)
- Andreas Nuernberger (Universitaet Magdeburg, Germany)
- Alex O'Connor (Trinity College Dublin, Ireland)
- Doug Oard (University of Maryland, USA)
- Katharina Reinecke (University of Michigan, USA)
- Fabrizio Silvestri (Yahoo Labs Barcelona, Spain)
- Ryen W. White (Microsoft Research, USA)
- Dong Zhou (Hunan University of Science and Technology, China)

References

1. Lamche, B., Pollok, E., Woerndl, W., and Groh, G. Evaluating the Effectiveness of Stereotype User Models for Recommendations on Mobile Devices. Proceedings of the Joint Workshop on Personalized Information Access (PIA 2014), in conjunction with the 22nd conference on User Modeling, Adaptation and Personalization (UMAP 2014).
2. Kunaver, M., Dobravec, S., Pozrl, T., and Kosir, A. Increasing Top-20 search results diversity through recommendation post-processing. Proceedings of the Joint Workshop on Personalized Information Access (PIA 2014), in conjunction with the 22nd conference on User Modeling, Adaptation and Personalization (UMAP 2014).
3. Barral, O., Kosunen, I., and Jacucci, G. Influence of Reading Speed on Pupil Size as a Measure of Perceived Relevance. Proceedings of the Joint Workshop on Personalized Information Access (PIA 2014), in conjunction with the 22nd conference on User Modeling, Adaptation and Personalization (UMAP 2014).
4. Athukorala, K., Oulasvirta, A., Glowacka, D., Vreeken, J., and Jacucci, G. Supporting Exploratory Search Through User Modeling. Proceedings of the Joint Workshop on Personalized Information Access (PIA 2014), in conjunction with the 22nd conference on User Modeling, Adaptation and Personalization (UMAP 2014).
5. De Nart, D., Tasso, C., Degl'Innocenti, D. Users as crawlers: exploiting metadata embedded in web pages for user profiling. Proceedings of the Joint Workshop on Personalized Information Access (PIA 2014), in conjunction with the 22nd conference on User Modeling, Adaptation and Personalization (UMAP 2014).
6. Calvo, I., Arruarte, A., Elorriaga, J.A., and Larrañaga, M. Work in Progress: Multicultural Concept Map Editor. Proceedings of the Joint Workshop on Personalized Information Access (PIA 2014), in conjunction with the 22nd conference on User Modeling, Adaptation and Personalization (UMAP 2014).
7. Ghorab, M.R., Lawless, S., O'Connor, A., and Wade, V. Does Personalization Benefit Everyone in the Same Way? Multilingual Search Personalization for English vs. Non-English Users. Proceedings of the Joint Workshop on Personalized Information Access (PIA 2014), in conjunction with the 22nd conference on User Modeling, Adaptation and Personalization (UMAP 2014).

Evaluating the Effectiveness of Stereotype User Models for Recommendations on Mobile Devices

Béatrice Lamche, Enrico Pollok, Wolfgang Wörndl and Georg Groh

Technische Universität München,
Boltzmannstr. 3, 85748 Garching, Germany
{lamche,pollok,woerndl,grohg}@in.tum.de
http://www.11.in.tum.de

Abstract. Mobile recommender systems have been proven as a promising approach in mobile scenarios to support the decision making process of users by suggesting beneficial items in a certain mobile context. The main goal of this paper is to examine whether a stereotype user model leads to better recommendations as part of such a system. For this purpose, we developed and tested a prototype for a shopping scenario. Research on fashion stereotypes led to a user model containing ten different stereotypes. The stereotype classification is performed by computing the proximity of each stereotype to the user's properties. Results of a user study show that a user model based on stereotypes generates better results than a recommender system without a stereotype-based user model. Moreover, stereotype-based user models allow personalized recommendations right away thus contributing to alleviating the cold start problem.

Keywords: mobile recommender systems, stereotypes, user modeling

1 Introduction

Mobile recommender systems support the decision making process of users by providing suggestions for items that are of potential use for them in a certain mobile context [1]. Stereotype user modeling was one of the earliest approaches to user modeling and personalization in general [2]. A stereotype-based system maps the individual features for the recommendation process to one of several equivalence classes, whose profiles are then used for computing the recommendations. Stereotypes are usually organized in a directed acyclic graph to allow for generalizations. Each stereotype corresponds to a certain set of features characteristics. If the characteristics of users change they may be reassigned to a different stereotype. In order to match a stereotype to a person, the system needs to have specific *triggers* - events that signal the appropriateness of a particular stereotype and in turn activate it. For one person, several stereotypes can be active. Once activated, the characteristics of the stereotype are incorporated into the user model [2]. Several approaches for constructing user models exist. One approach are *keyword user profiles* which usually extract several keyword vectors from a specific source (e.g. the browser history of the user) using different weighting schemes or algorithms. The user's explicit and implicit feedback is used in order to build the user profile [3].

Examining related research, most mobile recommender systems do not explicitly state the user model used behind their recommendation algorithm (e.g. [4]). It may be simple or implicitly part of the recommendation algorithm. In order to provide personalized mobile recommendations even in the cold start phase, a user modeling approach based on stereotypes is suitable. most people can be associated with a specific style that barely changes (e.g. *casual* vs. *elegant*), so that stereotypes can be easily predefined and an already existing user data base is not required. Moreover, the use of a stereotypical user model allows for a quick characterization of users, particularly important for a mobile scenario. So far, no research was found which tried to combine a stereotypical user model with a recommender system on a mobile device. This work will therefore examine the effectiveness of a mobile recommender system with a user model based on stereotypes. The main goal of this paper is to examine whether a stereotype user model leads to better recommendations as part of a mobile recommender system. The rest of the paper is organized as follows: We first introduce important foundations of user modeling and summarize related work. Next, we explain our prototype. We then present the results of our user study that showed that the recommender system using a stereotype user model performed overall better. We close by suggesting opportunities for future research.

2 Designing the Prototype

The scenario in mind is that of a fashion recommender system on a mobile device. Going shopping is an exploratory scenario and the users most often do not have a specific item in mind. We therefore develop a system that delivers recommendations right from the beginning without having to specify a search query. Since the system is used in a mobile environment, the mobile context such as the user's current location and time should be considered for the calculation of personalized recommendations. To be more precise: When users are going to town to look for clothing items and start the application, the system should recommend items of open stores nearby suitable to the user's taste right away and provide information about these items and corresponding stores.

There is little academic research on stereotypes for fashion styles. Therefore our knowledge on publicly perceived stereotypes was limited to information found on the world-wide web, e.g. [5]. We compared the most frequently classified stereotypes based on their given definition and finally identified the following ten fashion stereotypes: *Indie/Hipster*, *Emo*, *Preppy*, *Gothic*, *Urban*, *Athlete/Jock*, *Skater*, *Girly*, *Classy* and *Mainstream*. For the allocation of items to stereotypes we use a *weighted keywords* approach (see *section 1*). Out of the features identified for the various stereotypes, a limited set of attributes consisting of colors, brands and general descriptions for the clothing was identified. Each stereotype has manually been given a rating on a scale of 0 to 10 for each attribute, representing the *weight* to which the feature is related to the stereotype.

The application was written for the Android API version 19 and supports all devices running Android API version 8 or higher. The first screen of the application is a form in which users are asked to provide the data necessary for determining their stereotype (such as age, gender, profession and music taste). The music taste is taken into account because studies found out that it is highly related to the individual fashion style [6]. The user profile thus contains a user ID and a stereotype that is based on the users gender, age, job and music taste. After filling out the form, the application computes the three most relevant stereotypes based on the information that has previously been provided by the user. Each stereotype has been given a weight for all available age groups, jobs and music styles. The stereotype algorithm iterates through all stereotypes available and adds up the likelihood that this stereotype has for each of the properties age, job and music. The resulting three stereotypes are presented to the user in a picture. An extract of the two corresponding user interfaces can be seen in *figure 1*. As soon as the user selects the preferred stereotype, stereotype-based recommendations are calculated and shown in a grid view.

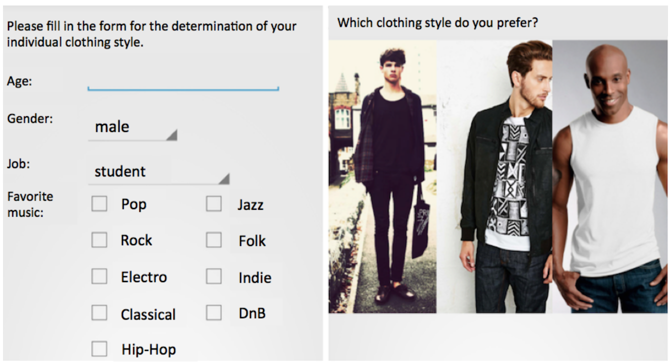


Fig. 1. The stereotype determination interfaces.

The *recommendation* algorithm sorts the items by their expected interest for the user. It first gets all attributes and their values for the active stereotype and then scans each item for the attributes color, brand and description. If the checked item contains one of the stereotype attributes, the specific *attribute weight* is added to the proximity measure. All weight values for the found attributes are thus added up and then divided by the number of found attributes. The result is a value for each clothing item which indicates the expected interest for a user with the selected stereotype. These values will be added to a map, sort in descending order and then presented as clothing recommendations to the user. It is worth noting that we give found brand names double the weight compared to other attributes, as we found out during testing that they provide the most reliable indicator for the attractiveness of a clothing item to a person. The user can scroll down in the recommendations view as long as necessary.

An implemented drop-down menu allows for filtering the results. Users can select or exclude specific values of features such as type of clothing, color, brand or price. A text field above the results is always visible, listing all the filters that have already been set. Clicking on a recommendation opens a new screen and details about price, colors, brand, images and the stores that sell them are listed. Once users have found an item they would like to purchase they can select it and the recommendation process terminates.

3 Evaluation

The main goal of the evaluation is to find out whether personalized recommendations can be improved through the use of stereotypes. To keep the number of testers at a reasonable size the study was designed as *within-subject*, one group of people tested both variants. The first system used in the user study is our developed mobile recommender system that uses a stereotypical user model. The user's stereotype is determined as described above. To successfully test the developed system, we need to establish a *baseline* to compare against. The second system that is tested is therefore a mobile recommender system without this stereotypical user model, or more clearly without any other algorithm supporting it but the users can still filter the results based on preferred features. The complexity of the experiment is kept low by asking users to choose one item they like for each approach. A potential user bias by using one approach before the other and thereby being aware of the choices available, is reduced by not making the user aware which approach is currently used and randomly switching the order of execution. All variables other than the algorithm used are kept fix. After having performed the task for each approach, candidates are asked to fill out a demographic questionnaire and to rate statements about the system design, the perceived ease of finding information and effort required to use the system, the usefulness of the system, the perceived accuracy of the suggestions, the satisfaction with the user's choice and intention to actually buy the product and reuse the system.

For the study, *participants* using mobile applications or showing an interest in using the described application were recruited. The user study finished with 32 participants, 27 male and 5 female, with an average age of 28 years, ranging from 22 to 54.

The *data set* used for this study was extracted from the now deprecated Google API for Shopping to retrieve the clothing item data and the Google Places API to retrieve information about shopping stores. The raw information from the API was rather limited with most information having to be extracted from the item and store description. To generate the data set of clothing items, the Shopping Search API was queried for keywords associated with types of clothing (e.g. simply 'dress') without any adjectives, to avoid leaning into a particular style as much as possible. The dataset built contains 668 different clothing items of 263 different brands. Items were associated with the following features: an id, one of 13 types of clothing, one of 15 colors, the price, the sex, a description and the link to an image of the item.

The *analysis* of the results is based on the user evaluation framework as described by Chen and Pu [7]. The data was analyzed using averages, standard deviations and student's t-test for determining distribution differences. A one-tail paired t-test was performed to calculate the p-value. *Table 1* shows the means for the most important metrics of the two systems, the standard deviation, as well as the p-value.

Table 1. A comparison of the user study's results.

	stereotype mean	stdev	baseline mean	stdev	p value
objective accuracy	0.47	0.34	0.32	0.34	0.036
perceived accuracy	3.5	0.53	2.6	0.52	0.00036
time consumption	47.82 s	35.83	64.26 s	33.68	0.001
perceived effort	57.9 %	-	42.1 %	-	-

Objective accuracy refers to the estimate of how likely it is that the user will select an item from a ranked list. The system sorts items according to their expected use, so each successive item in a list should be less likely to be selected by the user with an exponential decay. Objective accuracy can be measured using the *R-Score* which is based on the assumption that the value of a recommendation declines exponentially with the position of an item [8]. A higher *R-score* refers to a better ranking of the item. Calculating the *R-score* for the selected items leads to a mean of 0.47 (σ = 0.34) in stereotype mode and 0.32 (σ = 0.34) in the baseline. So we conclude that the stereotype-based approach is significantly more accurate at a 0.05 level (p-value = 0.036).

To determine the *perceived accuracy*, users were asked whether they would purchase the item they last selected. The answers to the question were put on a five-point Likert scale (from 1, strongly disagree to 5, strongly agree with 3 being neutral). The stereotype iteration was rated better in a median of 3.5 (σ = 0.53), compared to the baseline which was rated in a median of 2.6 (σ = 0.52), being statistically significant at a 0.05 level (p-value = 0.00036).

Objective effort is measured in terms of the time a user needs to find a satisfying item and go through the cycles. On average users took less time to complete the task when supported by a stereotype-based user model, in particular 47.82 seconds (σ = 35.83) versus 64.26 seconds for the baseline (σ = 33.68). The t-test confirms the difference of the samples at a significance level of 0.05 with a p-value of 0.001.

Perceived effort refers to the difficulty a subject has during the performance of the task in terms of information processing. 57.9 % of the participants preferred the stereotype round and 42.1% preferred the baseline.

The analysis of the open questions showed that the participants were overall very satisfied with the design of the application (62% with 28% feeling neutral about it) and 91% understood the usage of the application quickly.

4 Conclusion and Future Work

This work investigated the development and effectiveness of recommendations provided by a mobile prototype in the domain of fashion. Recommendations are generated by exploiting a stereotype user model and combining it with a mobile recommender system. The goal of the prototype was to provide the means to measure the effectiveness of its recommendations. 10 fashion stereotypes were identified and included in the user model. The app offers the user the possibility to criticize clothing items by clothing type, color, brand and price. Finally, a user study was conducted among 32 participants. The recommendation system using a stereotype user model performed overall better. Future research could use a more advanced approach to user modeling than the static stereotype user model, e.g. a form of an overlay model or a semantic network. Modeling aspects such as the mobile context may also lead to improved results, as well as a more sophisticated recommendation algorithm based on a dynamic user model which is able to learn how users behave.

References

1. Lamche, B., Trottmann, U., Woerndl, W.: Active Learning Strategies for Exploratory Mobile Recommender Systems. In: Proc. Decisions@CaRR workshop, ECIR, Amsterdam, Netherlands (2014)
2. Rich, E.: User Modeling via Stereotypes. In: Cognitive science, vol. 3(4), pp. 329–354 (1979)
3. 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. Lecture Notes in Computer Science, vol. 4321, pp. 5489. Springer, Berlin Heidelberg (2007)
4. Gavalas, D., Kenteris, M.: A web-based pervasive recommendation system for mobile tourist guides. In: Personal and Ubiquitous Computing, vol. 15(7), pp. 759–770. Springer, London (2011)
5. Angerosa, O.: Fashion stereotypes (2011), `http://people.rit.edu/~ona8039/wf/stage3/index.html`
6. Na, Y., Agnhage, T.: Relationship between the preference styles of music and fashion and the similarity of their sensibility. In: International Journal of Clothing Science and Technology vol. 25(2), pp. 109–118 (2013)
7. Chen, L., Pu, P.: Interaction design guidelines on critiquing-based recommender systems. In: User Modeling and User-Adapted Interaction, vol. 19(3) , pp. 167–206. Springer, Netherlands (2009)
8. Breese, J. S., Heckerman, D., Kadie, K.: Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th conference on Uncertainty in artificial intelligence, UAI98, pp. 43–52, San Francisco, CA, USA (1998)

Increasing Top-20 search results diversity through recommendation post-processing

Matevž Kunaver[1], Štefan Dobravec[1], Tomaž Požrl[1], and Andrej Košir[1]

[1]University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, Ljubljana, Slovenia
matevz.kunaver@fe.uni-lj.si,stefan.dobravec@fe.uni-lj.si,,tomaz.pozrl@fe.uni-lj.siandrej.kosir@fe.uni-lj.si

Abstract. This paper presents three different methods for diversifying search results, that were developed as part of our user modelling research. All three methods focus on post-processing search results provided by the baseline recommender systems and increase the diversity (measured with ILD@20) at the cost of final precision (measured with F@20). The authors feel that these methods have potential yet require further development and testing.

Keywords: Recommender systems, Diversity, ILD, F-measure, User modelling

1 Introduction

The focus of recommender systems (RSs) is moving from generating recommendations (providing personalized data retrieval and search results) without any additional situational data about the user to generating recommendations that also consider the user's context [1][3] and personality in order to improve the recommendation results[7]. All these improvements serve to present the user with a selection of results that will be the most appropriate for the situation in which the user desires to review the selected result. Recommendation results can be further improved by paying attention to the diversity [4] [8] [5] [11] of results presented to the user.

1.1 Motivation and Goal

The purpose of our study is to determine whether we can increase the diversity of results generated and presented to the user by a baseline RS by introducing three methods that post-process these results. Each of these methods uses a different diversification approach yet all three aim to maintain a high level of user satisfaction (measured by evaluating the accuracy of the modified RS). While 'search results' cover a wide array of possible items, we focused our research on movie search results as we had two different working RSs developed as part of our previous research in movies domain [9][10] and could therefore immediately focus on diversification method development.

2 Materials and Methods

In this section we describe the dataset, the baseline RSs used to generate recommendations, the developed diversification methods and the evaluation methods.

2.1 Dataset

For the purposes of our research we used the *Context Movie Dataset* (LDOS-CoMoDa), that we have acquired in our previous work. The dataset was collected using an on-line application for rating movies (www.ldos.si/recommender.html) that enabled the users to track the movies they have watched and to obtain recommendations from several RS algorithms. In addition, the application features a questionnaire whose purpose is the collection of the contextual data describing the situation during the item consumption.

The dataset currently consists of 4237 ratings given by 184 users to 1782 items. Each rating is also annotated with associated contextual variables. Each user is described with basic demographic data (age, sex, location) provided on a voluntary basis. Each item is described with several attributes: genre, director, actor, language, country, budget and release year.

The on-line application is still available and in use. Additional information about LDOS-CoMoDa can be found in [2] and [3].

2.2 Recommender System

For this paper we implemented our diversification methods on two different RSs: a hybrid RS and a content-based RS.

Hybrid RS[9]: The hybrid RS used for this experiment was developed as part of our previous research [9]. It is a collaborative RS that selects nearest neighbours based on genre preferences instead of their ratings. Each preference indicates the user's interest for one specific genre (25 in total). By using these preferences we are able to select nearest neighbours who perfectly match the active user in preferences without having a single overlapping item (i.e. item rated by both users). This increases the recommendation pool and the overall quality of the RS.

The hybrid RS generates recommendations for each user by performing the following steps: (i) Calculate genre preferences for the user based on his/her previous ratings, (ii) Find 20 users whose preferences are the most similar to the active user, (iii) Create a pool of potential recommendations from all of the items rated by these users, (iv) Calculate the predicted rating for each item using the Bayesian estimator, (v) Present the user with the top 20 items.

Content-based RS[10]: The content-based RS used in this paper developed as part of our previous research [10] as well and is based on a rule-based approach that considers all attributes available in the dataset. We defined a special similarity function that enables us to detect attribute values in the description of the item that user has a very high preference towards. If we detect such attribute values, we assign a high similarity value between the attribute value and the model of the user.

The content-based RS performs the following steps: (i) Generate content-based user model from items the user has already watched and rated, (ii) For each item not yet rated calculate attribute similarity values for attributes in item metadata using content-based user model, (iii) First, calculate similarity for each attribute value and then combine these similarities to a similarity of the attribute, (iv) Classify a vector of similarities of attributes into one of the rating values using 'M5Rules' decision rule classification method.

2.3 Diversification method

We aimed to develop methods that could be implemented in existing RSs without requiring a direct change in the way those RSs work. We therefore focused on diversifying the top 20 lists generated by those RSs. In our case the diversification process is following next steps (as shown in figure 1): for every user's list of recommended items (i) prepare ordered (descending) list of recommendations and split it into top 20 recommendations list and the remainder of the set, (ii) find exchange candidates in such manner that the diversity of top 20 items increases without significant harm to the accuracy of the system, and (iii) exchange the items to yield diversified list of recommended items. As indicated in figure 1 the second and the third step can be performed iteratively.

In our experiment we developed and tested three variations of the diversification process that differ mainly in the way how the exchange candidates are picked.

The first method swaps up to three items in a single step (no iteration). It starts by assessing the worst items in the top 20 list. It calculates the ILD value of the list while excluding one item (exchange candidate) of the list at a time. Effectively this means calculating ILD@19. Higher values of ILD@19 indicate better exchange candidates. Next, it searches for the best replacement candidates from the first 20 items of the remainder of the set. The method calculates the ILD@20 after exchanging every combination of up to three items. Final result is the top 20 list with best ILD@20 score after the exchange.

The second method uses the same approach as the first one to determine exchange candidates in the top 20 list. The best item, which yields highest ILD@19 is then replaced with an item from the first 20 of the remainder of the set. The final exchange is done using the replacement candidates that gives best ILD@20 score. In this case we shuffle only a single item at a time, but repeat the process K-times. It can be expected that increasing value of K would favour list diversity in trade-of to lowering list accuracy.

The third method, just like the second one, replaces single item at a time. The difference is, it considers a joint score in form of $a * avgPR + b * nILD$ instead of a pure ILD value. In this formulation $avgPR$ stands for the average prediction rating of the list and $nILD$ for the normalized ILD value of the same list. Parameters a and b allow balancing the top 20 list from more accurate / less diverse towards less accurate / more diverse. The shuffling procedure is repeated until best top 20 list (in term of joint score) is achieved.

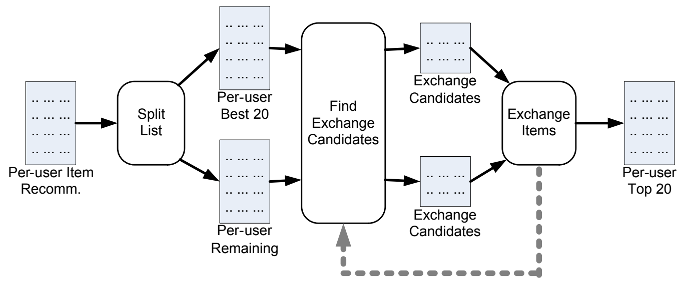


Fig. 1: Diversification process

2.4 Evaluation methods

In order to evaluate our methods and compare them to the control (non-diversified) RS we had to consider accuracy as well as diversity of each generated top list.

We evaluated **accuracy** using the F-measure at top ranking position 20 (F-measure@20) [6] as it is one of the most often used measures of accuracy in recommender systems. In order to evaluate the **diversity** of our recommendations we used the intra-list diversity [11] (ILD@20) calculating the diversity value of each top list based on the following metadata descriptions of each item: genre, director, actor, language and country.

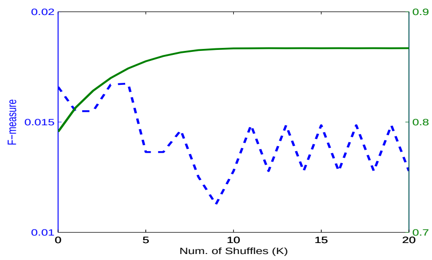
3 Results

Table 1 shows the results of our evaluation (F@20 and ILD@20) of both baseline recommender systems and for all three developed diversification methods.

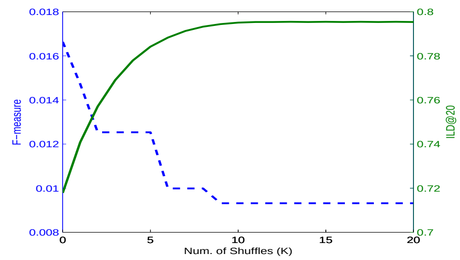
Table 1: Evaluation results

	Hybrid RS		Content-based RS	
method	F@20	ILD@20	F@20	ILD@20
non-diversified	0.011	0.772	0.020	0.717
diversified - method 1	0.007	0.818	0.0122	0.764
diversified - method 2	0.015	0.867	0.0125	0.784
diversified - method 3	0.018	0.915	0.0151	0.878

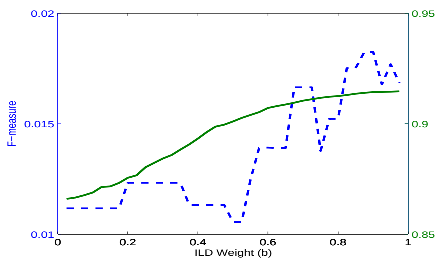
As methods 2 and 3 featured additional parameters (number of iterations / joint score settings) we also present their results in figure 2, where we show how different parameter settings impact the systems accuracy / diversity.



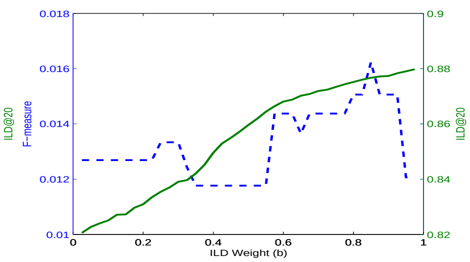
(a) Hybrid RS - method 2



(b) Content-based RS - method 2



(c) Hybrid RS - method 3



(d) Content-based RS - method 3

Fig. 2: Evaluation results - method 2 and 3

4 Conclusion and Further Work

The results presented in this paper show promise as all three diversification methods increased the overall top 20 list diversity by at least 6% with the best increase being by method 3 which increased the diversity of content-based recommendations by 22%. The main difference between all three methods is that the first method is a non-iterative one and therefore requires a single run to diversify all top 20 lists while methods 2 and 3 require several iterations to provide the best results in addition to requiring and extra training run to determine the best parameter values.

The real surprise however came when we measured the impact on accuracy for each method. While we saw a decrease in accuracy in content-based RS (from 25% to 40% as expected) we actually found that diversifying our hybrid RS with method 2 or 3 increases the overall accuracy by as much as 60%. We think that this might be due to the small number of ratings per user in our dataset (meaning that shuffling the top items managed to hit a few additional items in the test set, thus increasing the R@20 and P@20 values) and that using the same method on a different dataset might yield different results. However, we also believe that we should use additional accuracy evaluation methods in our future experiments and see if they support the findings from this paper.

We have nevertheless started a study in post-processing diversification that shows promise and we plan to further expand our understanding by addressing these key issues:

- Determine whether the number of replaced items from the top list can be fixed or must be calculated iteratively for each user each time the RS generates recommendations.
- The number of replacement candidates to be considered.
- Perform a series of statistical tests in order to determine whether our results are really significantly different from those of a non-diversified RS.
- Determine the optimal values of parameters a and b for the third method.
- Perform an A/B test to determine how the lower accuracy impacts the actual user satisfaction.
- Perform a study of method efficiency to determine which of the three methods performs best in which circumstances - when can we afford the extra iterations required by methods 2 and 3 and when we cannot.

References

1. Dey A. and Abowd G. Towards a better understanding of context and context-awareness. pages 304–307, 199.
2. Košir A., Odic A., Kunaver M., Tkalcic M., and Tasic J. F. Database for contextual personalization. *Elektrotehniški vestnik*, 78(5):270–274, 2011.
3. Odic A., Tkalcic M., and Košir A. Managing irrelevant contextual categories in a movie recommender system. *Human Decision Making in Recommender Systems (Decisions@ RecSys 13)*, page 29, 2013.
4. Smyth B. and McClave P. Similarity vs. diversity. In *Case-Based Reasoning Research and Development*, pages 347–361. Springer, 2001.
5. Jannach D., Lerche L., Gedikli F., and Bonnin G. What recommenders recommend–an analysis of accuracy, popularity, and sales diversity effects. In *User Modeling, Adaptation, and Personalization*, pages 25–37. Springer, 2013.
6. Ricci F., Rokach L., Shapira B., and Kantor P.B. *Recommender Systems Handbook*. Springer, 2011.
7. Adomavicius G. and Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.
8. Adomavicius G. and Kwon Y. Improving aggregate recommendation diversity using ranking-based techniques. *Knowledge and Data Engineering, IEEE Transactions on*, 24(5):896–911, 2012.
9. Kunaver M., Košir A., and Tasič J. F. *Hybrid recommender for multimedia item recommendation*. Lambert Academic Publishing, 2011.
10. T Požrl, M Kunaver, M Pogačnik, A Košir, and JF Tasič. Improving human-computer interaction in personalized tv recommender. *International Journal of Science and Technology, Transactions of Electrical Engineering*, 36(E1):19–36, 2012.
11. Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pages 22–32. ACM, 2005.

Influence of Reading Speed on Pupil Size as a Measure of Perceived Relevance

Oswald Barral, Ilkka Kosunen, and Giulio Jacucci

Helsinki Institute for Information Technology HIIT,
Department of Computer Science,
University of Helsinki, Finland
{oswald.barral,ilkka.kosunen,giulio.jacucci}@helsinki.fi

Abstract. Depending on the task or the environment, we read texts at different speeds. Recently, a substantial amount of literature has risen in the field of predicting relevance of text documents through eye-derived metrics to improve personalization of information retrieval systems. Nevertheless, no academic work has yet addressed the possibility of such measures behaving differently when reading at different speeds. This study focuses on pupil size as a measure of perceived relevance, and analyses its dependence on reading speed. Our results are followed by a discussion around the need of taking into account reading speed when using eye-derived measures for implicit relevance feedback.

Keywords: pupillometry, perceived relevance, reading behavior

1 Background

1.1 Reading behavior

When using information retrieval systems to seek for information, the user adopts different reading behaviors, depending on several factors. The task to achieve, the environment or time pressure are some of them. The main component of reading behavior addressed in this study is reading speed.

Different reading speeds are usually associated to different reading tasks. Skimming can be helpful when there is a need to address a large amount of information and retain the most relevant parts of it. However, reading at fast rates involves less comprehension [1, 2]. If the goal of the reading process is to comprehensively understand the text, a normal reading speed will be adopted. On the other side, if there is a reduced available time and the amount of information is large, a faster reading speed will be more adequate, in order to focus just on the relevant parts of the text. The information seeker will therefore always adopt an optimal reading speed for every situation.

Having said that, as the amount of information available increases, the users tend to adopt faster reading rates, especially when seeking for information. Liu made an extensive survey addressing the changes of reading behavior in people ranging between 30 and 45 years old [3]. The participants in the study were asked to answer a set of questions regarding how their reading characteristics had changed over the past ten years. One of the outcomes of the survey was that 80% of the participants reported to have increased the time spent scanning and browsing, which are reading behaviors that imply high reading rates.

1.2 Pupil size as a measure of perceived relevance

Eye tracking technologies have been used in the field of information retrieval and personalized access over the past years as eye-derived metrics have proven to be useful to indicate users subjective perception of relevance [4–6]. In the goal of personalizing results, these implicit metrics are highly valuable as they provide an intrinsically individualized feedback.

Studies have shown a relationship between pupil size and user attention [7, 8]. It is well known that pupil size and cognitive load are highly correlated, different researches having approached the matter. Experiments have ranged from mathematical operations to search tasks [9]. Interestingly, Oliveira et al. [10] showed how pupil size could be of special interest when analyzing relevance in web search results. They studied both relevance of images and documents. Focusing on changes in pupil diameter, they were able to claim pupil size to be a carrier of interest-related information. Their experiments were on a very controlled level, letting the demonstration of similar conclusions in less controlled experiments as future research.

2 The present study

Given the above-mentioned reading behaviors, especially the increasing trend to read at fast reading rates, we consider highly relevant to study eye-derived implicit measures of relevance under the influence of different factors. In the present study, we focus on pupil size under the influence of reading speed. We designed an experiment in order to study whether reading speed has a direct impact on the ability of pupil size to indicate perceived relevance in documents.

2.1 Apparatus

The machine used to run the experiment was a 64bit processor Intel Core i73930k 3.20GHz 3.20GHz 16GB RAM, OS Windows 7 Enterprise SP1 with NVIDIA GEForce GTX580 GPU. The display device was a Dell 1703FPt 17" LCD Monitor at a 1280x1024 resolution. The experiment was developed using ePrime Software. The texts were displayed in an 85% window (I.e. 1088x870.4 pixels) with a 22-point font size. The subject was asked to sit 40-50 cm away from the screen approximately and to take a comfortable position. A Mirametrix S2 eye tracker operating at 60 Hz was situated under the screen and slightly moved to best fit to the subject eyes according to his natural and more comfortable position. The number of clock ticks since the booting of the operative system was used as reference for the synchronization between the Mirametrix S2 eye tracker and the ePrime software.

A first eye tracking calibration procedure was carried out at the beginning of the experiment and another one at the middle of the experiment. Each calibration procedure lasted for about 5 minutes, depending on the subject. The process was repeated up to five times to ensure optimal calibration (average error < 40 pixels). If the threshold was not reached within the first attempts, the average error margin was augmented in 10 pixels. The subject was rejected if after 5 additional attempts the average error was not fewer than 50 pixels. Two subjects out of ten were rejected due to calibration impossibility.

2.2 Participants and Procedure

Ten students (four undergraduate and six master's) participated in the experiment. Two of them were women. Eight participants reported to have advanced English reading level, and two reported a medium English reading level. None of them was a native English speaker. All of them had normal or corrected to normal vision. As already pointed out, two of the participants did not overcome the calibration procedure due to technical difficulties and their data was rejected. At the beginning of the experiment the participants were asked to sign a consent form and to indicate basic information about themselves. The data was saved anonymously in order to preserve participants privacy.

The participants were first conducted through a training session. The training consisted of two parts. The first one intended to get the users familiar with the three different speeds. As the reading speed is relative to the user's expertise or abilities, among other factors, instead of using an absolute word per minute rate for each of the speeds, an approach similar to the one by Dayson and Haselgrove was implemented [2]. The participants were first asked to read a document at a comfortable reading speed in order to be able to understand everything. They were instructed to reproduce that speed when they would be asked to read at a normal speed. They were then presented another text and asked to read it as twice as fast as the first text. If the time spent reading was higher than 70% of the previous one, they were presented a new text and asked to read faster, until they managed to spend less than 70% of the original time reading the text. They were then instructed to reproduce that speed every time they would be asked to read at a fast speed. An homologous procedure was used to train the skimming speed. Different texts were used in each of the phases in such a way that the familiarity with the text could not influence the reading speed. The participants were told explicitly to try to do their best to reproduce each of those speeds during the experiment. The second part of the training consisted of using the actual system until the participants explicitly recalled to have fully understood how they were supposed to interact with the system.

We decided to split the recording session into two parts as the participants of a pilot study reported to feel tired after having gone through the whole sequence of abstracts. Also, this allowed the recalibration of the eye-tracking device, avoiding the accumulation of systematic error [11]. Each of the two parts consisted of three topics. For each of the topics, the participants were asked to read in a given speed a sequence of abstracts. For each abstract, they were asked to assess as soon as possible using the left and right arrows whether the text was relevant to the topic (*binary-rating*). The participants were asked to keep reading until the end of the text at that given speed and to press space when done. Then, they were asked to grade, in a scale from 0 to 9, how relevant was the abstract to the topic (*scale-rating*) and how confident they felt about their answer (*confidence-rating*) .

For each of the six topics six abstracts were shown, half of them being relevant and the other half being non-relevant. The participants had to read two of the abstracts at a normal speed, two at a fast speed and two at a skimming speed. The order of the topics and the abstracts, as well as the reading speeds, was randomized. The topics were selected to be of common understanding and the participants were allowed to ask to the experimenter any question regarding the understanding of those. The topics were also selected in a way that their semantic meaning would not overlap. The relevant abstracts were selected not to be too obvious in the first lines. The non-relevant abstracts were selected to be completely non relevant to any of the topics.

3 Analysis and Results

For each abstract we took a time window of 10 seconds (i.e. five seconds before and five seconds after *binary-rating*) and averaged the values of the pupil each 500 milliseconds. We normalized the pupil data in each text by subtracting the mean of the pupil size over the entire text. Only the data of texts where the *binary-rating* and the *scale-rating* were congruent, and where *confidence-rating* was higher than 6 were taken into account (i.e. *valid-trials*). In these cases we observed a clear spike in the pupil size about 1 to 1.5 seconds after assessing the *binary-rating*. This was not surprising as the maximal pupil dilation has been reported between the event attracting attention and 1.3 seconds after [8].

In order to test for statistical significance between the spikes when assessing texts as relevant and when assessing texts as non-relevant we first took, for every abstract, the average value of the normalized pupil size in the time window of 0 to 1.3 seconds after the response time. Then, for the overall texts, as well as for each speed and each condition (the user answered relevant or answered non-relevant) we averaged the values within subjects. Finally, we performed Wilcoxon signed-rank text on the resulting paired samples.

In overall, pupil size was significantly higher when assessing texts as relevant (Mdn = 0.8) than when assessing texts as non-relevant ($Mdn = 0.66$), $z = -2.366$, $p < 0.05$, $r = -0.63$. When analyzing the texts read at normal speed, pupil size was also found to be significantly higher when assessing relevant ($Mdn = 0.93$) than when assessing non-relevant ($Mdn = 0.8$), $z = -2.197$, $p < 0.05$, $r = -0.59$. However, when analyzing the texts read at fast speed – relevant ($Mdn = 0.91$), non-relevant ($Mdn = 0.7$), $z = -1.690$, $r = -0.45$– and

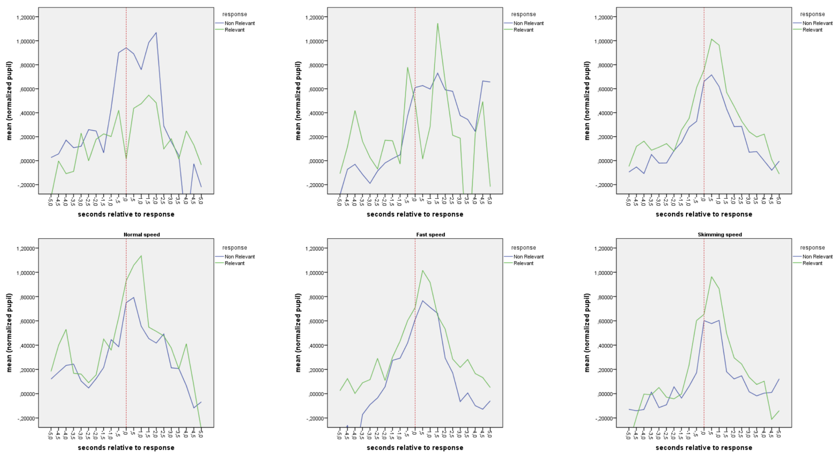


Fig. 1. Beginning from top-left: Pupillary response when *confidence-rating* is below 6; Pupillary response when *binary-rating* and *scale-rating* are not congruent; Pupillary response in the *valid-trials*. Beginning from bottom-left: Pupillary response for *valid-trials* read at normal speed, fast speed and skimming speed. The red line indicates the moment of *binary-rating*. The blue line represents the non-relevant and the green line represents the relevant texts. The plotted values are normalized within trials and averaged across participants.

skimming speed –relevant ($Mdn = 66$), non-relevant ($Mdn = 0.59$), $z = -0.676$, $r = -0.18$– no statistical significance was found.

4 Discussion

The results showed a clear relationship between the pupil dilation and the participants' subjective judgments. On top of that, the analysis of pupil size confirmed our hypothesis that its behavior would differ when reading documents at different speeds. When looking at the data without taking into account the speed in which the document was read, statistical analysis showed a significantly bigger response-related spike when the user perceived the document as relevant than when perceiving it as irrelevant. Nevertheless, when having a look at the same data but splitting the analysis by reading speed, the data showed statistical significance only when the user was reading at normal speed. That is, when the subject was given the instruction to read at faster rates than the comfortable normal reading speed, the response-related spike in the pupil size did not carry statistically relevant information regarding the judgement of the participant.

With this study we aim to raise a discussion around the fact that, when dealing with documents, different reading behaviors might have a direct impact on the reliability of our eye-derived measures. Thus, reading behaviors should be controlled and studied in order to have more accurate implicit feedback and, consequently, better personalization. As with pupil size, we believe that fixation-derived features used to infer relevance in documents will also behave differently when reading at different speeds and, therefore, need a closer look when the aim is to build realistic personalized search engines based on implicit feedback [12]. We encourage researchers to study the behavior of information seekers, and to apply such knowledge in the design of personalized information retrieval systems. We believe that a main element of the information seeking behavior that need to be understood is how the texts are addressed, studying which components have an influence on the application of implicit relevance measures. In the presented work we identified reading speed as one of these components affecting pupil size but, surely, in order to apply implicit metrics to enhance personalization in realistic systems, other measures and components of reading behavior need to be carefully studied.

References

1. Masson, M.E.: Cognitive processes in skimming stories. Journal of Experimental Psychology: Learning, Memory, and Cognition 8(5) 400 (1982)
2. Dyson, M., Haselgrove, M.: The effects of reading speed and reading patterns on the understanding of text read from screen. Journal of Research in Reading 23(2) 210–223 (2000)
3. Liu, Z.: Reading behavior in the digital environment: Changes in reading behavior over the past ten years. Journal of Documentation 61(6) 700–712 (2005)
4. Puolamäki, K., Salojärvi, J., Savia, E., Simola, J., Kaski, S.: Combining eye movements and collaborative filtering for proactive information retrieval. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ACM 146–153 (2005)
5. Ajanki, A., Hardoon, D.R., Kaski, S., Puolamäki, K., Shawe-Taylor, J.: Can eyes reveal interest? implicit queries from gaze patterns. User Modeling and User-Adapted Interaction 19(4) 307–339 (2009)
6. Loboda, T.D., Brusilovsky, P., Brunstein, J.: Inferring word relevance from eye-movements of readers. In: Proceedings of the 16th international conference on Intelligent user interfaces, ACM 175–184 (2011)
7. Beatty, J., Lucero-Wagoner, B.: The pupillary system. In: Handbook of psychophysiology, New York, NY, US, Cambridge University Press 142–162 (2000)
8. Just, M.A., Carpenter, P.A.: The intensity dimension of thought: Pupillometric indices of sentence processing. Canadian Journal of Expérimental Psychology/Revue canadienne de psychologie experimentale 47(2) 310 (1993)
9. Hess, E.H., Polt, J.M.: Pupil size in relation to mental activity during simple problem-solving. Science 143(3611) 1190–1192 (1964)
10. Oliveira, F.T., Aula, A., Russell, D.M.: Discriminating the relevance of web search results with measures of pupil size. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM 2209–2212 (2009)
11. Hornof, A.J., Halverson, T.: Cleaning up systematic error in eye-tracking data by using required fixation locations. Behavior Research Methods, Instruments, & Computers 34(4) 592–604 (2002)
12. Barral, O.: Exploring the applicability of implicit relevance measures in varying reading speed for adaptive I.R. systems. Master's thesis, Universitat Politècnica de Catalunya/Helsingin Yliopisto, Spain/Finland (2013)

Supporting Exploratory Search Through User Modeling

Kumaripaba Athukorala ‡, Antti Oulasvirta †, Dorota Glowacka ‡, Jilles Vreeken †, Giulio Jacucci ‡

‡ Helsinki Institute for Information Technology HIIT
‡ Department of Computer Science, University of Helsinki
† Max Planck Institute for Informatics and Saarland University

Abstract. Exploratory search is becoming more common as the web is used more increasingly as a medium for learning and discovery. Compared to traditional known-item search, exploratory search is more challenging and difficult to support because it initiates with poorly defined search goals, while the user knowledge and information-needs constantly change throughout the search process. Although information-retrieval algorithms have improved greatly in providing users with relevant information to specific queries, there is still room to improve our understanding of users' exploratory information-seeking strategies as well as design of systems supporting exploratory information-seeking. Modeling the user behavior and predicting dynamically changing information-needs in exploratory search is hard. Over the past decade there has been increasing attention on rich user interfaces, retrieval techniques, and studies of exploratory search. However, existing work does not yet support the dynamic aspects of exploratory search. The aim of this research is to explore different aspects of how to understand and support exploratory search, including user studies, intent visualization and user modeling.

Keywords: User Modeling, Exploratory Search, Scientific Information-Seeking

1 Introduction

Search can be broadly divided into two categories: known-item search and exploratory search. In known-item search the user has a specific search result in mind. On the other hand, in exploratory search the goal is ill-defined and changes as the search progresses [1]. Traditional information retrieval techniques concentrate mostly on known-item search. However, exploratory search is becoming more important as the web is becoming a major source for learning and discovery [2].

Exploratory information-seeking is known to be complex and hard to support due to its inherently open-ended and dynamic nature [3]. It arises in situations where there is a need to find information from a domain in which the user has a general interest but not specific knowledge [1]. Exploratory search has also been defined based on the distinct characteristics of the search process such as submitting tentative queries, selectively seeking and passively obtaining cues about the next steps, and iteratively searching with evolving information-needs. In this paper, we concentrate on exploratory search in the scientific information-seeking context. To be more precise, we use a scientific essay writing scenario, where a student has to write an essay on a research topic in which she has a general interest but lacks knowledge to formulate queries to gather the necessary literature. This type of search involving exploration of unfamiliar research areas for the purpose of learning is found to be one of the most challenging literature search purposes [4].

Over the last decade many techniques have been proposed to provide better support for exploratory information-seeking, such as results clustering [5], relevance feedback [21], faceted search [7], and novel visualizations to support the creation of unfamiliar information spaces [8]. Even though these solutions help in improving exploration, exploratory search involves many different phases. For example, it begins with an imprecise query and then through several successive iterations of exploring the retrieved information and reformulating queries, the scope of the information need narrows down. This iterative and evolving nature of exploratory search makes it very difficult to identify the constantly changing information needs of the user and different phases of exploration.

Systems that suggest queries, provide interactive keyword visualizations, cluster results, and provide similar help to better support exploration need to "know" whether the suggested queries/keywords and selected clusters are too narrow or too broad for the current information need of the user. Hence, in exploratory search it is important to predict which stage of exploration the user is in with respect to the evolving state of his or her knowledge. One way to address this problem is by understanding user behaviors with queries with varying specificity in exploratory searching, which, in turn, will allow us to build a user model to predict whether a given query is too broad or too specific for the current information need of the user. Another method is by providing visualizations of systems interpretation of user needs and allow the user to provide feedback. The main goals of this research is to investigate exploratory search behaviors of academics, and build interaction models and visualizations that allow information retrieval systems to infer the state of exploration from the observable aspects of user interactions.

2 Related Work

Over the past decade researchers from, among others, information retrieval (IR), human-computer interaction (HCI), and cognitive science communities have made many attempts to better support the user in tasks involving exploratory search by developing retrieval techniques, user interfaces and conducting studies aim at understanding user behaviors in exploratory search.

In the context of information retrieval, existing contributions include relevance feedback based retrieval [21], faceted search [7], and result clustering [5].

However, evidence from user studies shows that results clustering and relevance feedback based methods are rarely used due to high cognitive overload of selecting relevant results and providing feedback, and the problem of the context trap [21]. Faceted search is found to be overly demanding as users have to go through a large number of options [7]. Furthermore, studies have shown that exploratory search requires more active user engagement with the search results [9]. The lack of success of systems such as relevance feedback is often attributed to user interface designs failing to conveniently provide feedback at suitable levels of granularity [7].

In response, a number of new techniques were designed to visualize search results and capture user feedback. Some of them include rich user interfaces combined with learning algorithms to support users to comprehend the search results [8], and visualization and summarization of results [10]. All these solutions are giving users more control, however, they fail to take the moment-by-moment information-needs of the user into consideration [11]. This is where user modeling can greatly improve existing approaches to exploratory search.

User behavior in exploratory information-seeking is studied with intents: predicting cognitive styles [12], identifying search and query formulation strategies [13], and constructing user models to predict the domain knowledge [14]. Early studies showed emergence of different search strategies depending on the users familiarity with the topic. Crucially, user studies show that users spend more time evaluating unfamiliar topics than familiar ones [14], domain knowledge and experience with a search tool impact search behavior [15], and that search strategies change over time when domain knowledge increases [16]. Existing models are useful in customizing results according to user preference [17] and knowledge, however, they do not capture situations where domain experts search information in narrower sub-fields of a familiar domain. Information Foraging Theory (IFT) provides several quantitative models of user search [18], yet existing work on IFT does not consider the effect of evolving user knowledge and queries. Overall, behavioral studies clearly point to the dynamic nature of the exploratory information-seeking process and the effect of prior knowledge on users' search strategies, which lends support to the assumptions behind the models we develop in this research. Our aim is to design user models that predict moment-by-moment information-needs of the user through observable user behaviors to improve the performance of retrieval algorithms.

3 Information-Seeking Behaviors and Intent Visualization

Exploratory search is very common among academics. Therefore, we conducted a study to investigate how academics search for scientific information and what challenges they face. This was a mixed method study involving interviews, diary logs, user observations, and a survey. The findings suggested that exploring unfamiliar research areas was one of the most common purposes of scientific information-seeking and it is the most difficult task to perform [4]. Results of this study provided useful insights into the problem of exploratory search.

In the initial stages of exploratory search, users have poor knowledge about the information space. Therefore, visualizations of the underlying information space can help the users to make better sense of the search topic. As a part of this research a prototype search tool called SciNet was developed [19–21]. With SciNet, the user can perceive the state of user model through the interactive visualization and provide feedback by moving keywords. User studies that compared SciNet indicated that it helps users to more effectively find relevant, novel and diverse results.

4 Distinguish Exploration from Navigational Search

We also conducted a study to compare exploratory search and navigational/known-item search. The results indicated that unlike in known item search in exploratory search there is a higher percentage of fixations even on results at the bottom of the ranked list of Search Engine Results Pages (SERPs) (See Figure 1). These results are useful in building a model to distinguish exploratory search from navigational search.

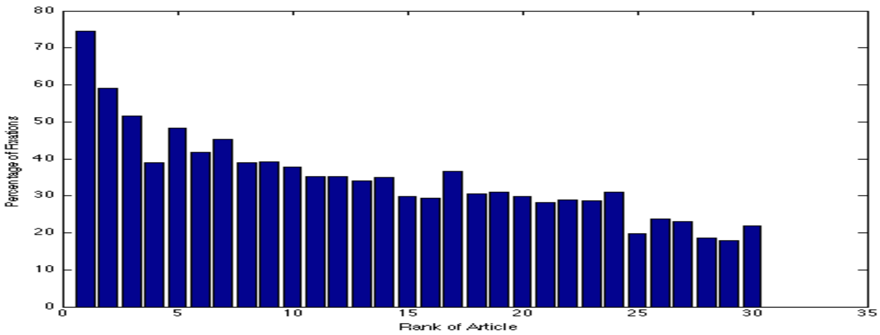


Fig. 1. Percentage of fixations at each article in the SERPs displaying 30 results in an exploratory search task. This figure indicates that in exploratory search users fixate more on results than in known item search.

5 Interaction Model to Predict Stages of Exploration

We also designed a model by combining insights from research into exploratory search and Information Foraging Theory (IFT) [18]. According to IFT, information gain can be modeled as a linear function of time when the results are ordered by relevance to the query. Further, IFT states that this information gain function will qualitatively shift towards a diminishing returns curve if new interface elements, such as result clustering, are introduced. Hence, IFT shows how information gain is affected by the user interface changes.

Our research is motivated by this model. If we keep the user interface constant, the information gain function should change according to the stage of exploration. We define the stages of exploratory search as broad, intermediate, and specific. In the broad stage the user has very little knowledge about the search topic and will issue a vague search query addressing a very broad information space. In the intermediate stage, the user would have some idea about the topic and would reformulate queries referring to sub-areas in the search topic. In the specific stage, the user would have gained a good enough knowledge and would use queries referring to very specific search topics in the area. We refer to this as subjective specificity. Sometimes a user might start the search with a very specific search query without having any knowledge about the area. Such a situation might arise when a search engine suggests keywords to the user, or when the user picks up some new terms randomly from the results without actually learning about them. If a search engine can predict the subjective specificity of the search results then it would be very useful to personalize the search results.

Our model captures how information gain in exploratory search is affected by this subjective specificity. The key idea is that the same search result can have very different information content for a user depending on how well it matches their current information needs. Consider two users who differ in the specificity of their goals and the extent of previous knowledge about a given topic, for example an undergraduate student writing a short overview essay on a well-known topic versus an experienced researcher gathering information about the latest developments in a specialized field. Their responses would differ, the former user probably spending more time on every item and the latter quickly scanning for the most informative items.

Empirical evaluation shows that our model captures the effects of query-specificity as well as the known effect of both prior knowledge and experience. Through a preliminary study we show the feasibility of using our model in a running IR system for predicting query-specificity.

6 Future Contributions

An important future challenge is to investigate in a real exploratory information-seeking scenario the performance of the formal model that we developed to predict the specificity of search results. We have already conducted a preliminary classification study which found that a system using only a simple classifier can obtain informed estimates on the specificity of a query while the user is interacting with its results. In the future, we will incorporate our model in a running IR system and further validate its usefulness in enhancing performance of exploratory search tasks.

References

1. White, R. W., Kules, B., Drucker, S. M.: Supporting exploratory search. J. Comm. ACM, 49(4), 36–39 (2006)

2. Teevan, J., Alvarado, C., Ackerman, M. S., Karger, D. R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. In: CHI, 415–422 (2004)
3. Marchionini, G.: Exploratory search: from finding to understanding. J. Comm. ACM, 49(4), 41–46 (2006)
4. Athukorala, K., Hoggan, E., Lehtio, A., Ruotsalo, T., Jacucci, G.: Informationseeking behaviors of computer scientists: Challenges for electronic literature search tools. In: AGM of Assoc. Inf. Sci. Tech. (2013)
5. Ferragina, P., Gulli, A.:A personalized search engine based on Web-snippet hierarchical clustering. J. Software: Practice and Experience, 38(2), 189–225 (2008)
6. Kelly, D., Fu, X.: Elicitation of term relevance feedback: an investigation of term source and context. In SIGIR, 453–460 (2006).
7. Yee, K. P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: CHI, (2003)
8. Chau, D. H., Kittur, A., Hong, J. I., Faloutsos, C.: Apolo: making sense of large network data by combining rich user interaction and machine learning. In CHI, pp 167–176 (2011)
9. Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., Yee, K.: Finding the flow in web site search. J. Commun. ACM, pp. 42-49 (2002)
10. Kki, M.:Findex: search result categories help users when document ranking fails. In CHI, pp 131–140 (2005)
11. Teevan, J., Alvarado, C., Ackerman, M. S., Karger. D. R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. In CHI, pp 415–422 (2004)
12. Saito, H., Miwa, K.: A cognitive study of information seeking processes in the www: the effects of searchers knowledge and experience. In: WISE, pp 321327 (2001)
13. Payne, S. J., Howes, A.: Adaptive interaction: A utility maximization approach to understanding human interaction with technology. In Synt. Lec. Hum.-Cent. Inf., pp 1–111 (2013)
14. Holscher, C., Strube, G.: Web search behavior of internet experts and newbies. J. Computer Networks, 33(1), pp 337-346 (2000)
15. Vakkari, P., Pennanen, M., Serola, S.: Changes of search terms and tactics while writing a research proposal: A longitudinal case study. In: Inf. Proc. Manag., 39(3), pp 445–463 (2003)
16. Wildemuth, B. M.: The effects of domain knowledge on search tactic formulation. J. Assoc. Inf. Sci. Tech., 55(3), pp 246–258 (2004)
17. Kelly, D., Teevan, J.: Implicit feedback for inferring user preference: a bibliography. In: SIGIR, pp. 18–28. (2003)
18. Pirolli, P., Card, S.: Information foraging in information access environments. In: CHI, pp 51–58, (1995)
19. Głowacka, D., Ruotsalo, T., Konyushkova, K., Athukorala, K., Kaski, S., Jacucci, G.: Directing exploratory search: Reinforcement learning from user interactions with keywords. In: IUI (2013)
20. Ruotsalo, T., Athukorala, K., Glowacka, D., Konyushkova, K., Jacucci, G., Kaski, S.: Supporting Exploratory Search Tasks with Interactive User Modeling. In:ASIST (2013)
21. Ruotsalo, T., Peltonen, J., Eugster, M., Głowacka, D., Konyushkova, K., Athukorala, K., ...and Kaski, S. Directing exploratory search with interactive intent modeling. In CIKM, pp. 1759–1764, ACM. (2013)

Users as crawlers: exploiting metadata embedded in Web pages for user profiling

Dario De Nart, Carlo Tasso, Dante Degl'Innocenti

Artificial Intelligence Lab
Department of Mathematics and Computer Science
University of Udine, Italy
{dario.denart,carlo.tasso}@uniud.it, dante.deglinnocenti@spes.uniud.it

Abstract. In the last years we have witnessed the rapid growth of a broad range of Semantic Web technologies that have been successfully employed to enhance information retrieval, data mining and user experience in real-world applications. Several authors have proposed approaches towards ontological user modelling in order to address different issues of personalized systems, such as the cold start problem. In all of these works, non-structured data such as tags are matched, by means of various techniques, against an ontology in order to identify concepts and connections between them. However, due to recent popularity of semantic metadata formats such as microformats and RDFa, structured data are often embedded in many Web contents, with no need to "guess" them using a support ontology which may not be coherent with the actual content and the original goals of the author. In this paper we propose a novel approach towards ephemeral Web personalization based on extraction and enrichment of semantic metadata embedded in Web pages. The proposed system builds, at client-side, a rdf network that can be queried by a content provider in order to address personalized content.

Key Words: User Modeling, Semantic Web, Ephemeral Personalization, RDFa

1 Introduction

Personalization is one of the leading trends in Web technology today and we have all stumbled upon it in a way or in another while surfing the Web. Most of the times the process is evident, for instance when web sites require us to sign in and ask for our preferences in order to maintain an accessible user profile. But in other cases personalization is more subtle and is hidden to the user
Ephemeral personalization [9], for instance, aims at providing personalized content fitting only short-term interests that can expire after the navigation session. Most of the times there is no need for the user to sign in order to exploit ephemeral personalization, since all the information needed to determine which content should be presented may be

found in his/her browsing cache and/or content providers are not interested in modelling and archive such short-term interests. An example of ephemeral personalization is targeted adversing, that is providing personalized ads to users as they browse the Web. This task is currently accomplished by checking which cookies are present in the client's browser cache and selecting candidate ads accordingly. This process, however, in most cases results in a particular ad from a previously visited site, "stalking" in such way the user throughout all his/her browsing activities. As the authors of [7] suggest, this may generate a revenue for the advertiser by encouraging customers to return, but can also be extremely annoying and the users may perceive their privacy attacked. Other forms of ephemeral personalization are guided by contextual information derived from the IP address of the client or by analysing the content of the pages that the client requests, like in Amazon's product pages, however these are very shallow forms of personalization and do not involve an explicit and persistent user model.

In this work, we claim that there is another way to address ephemeral personalization that, to the best of our knowledge, has never been explored yet. Our approach consists in collecting semantic metadata contained in visited web pages in order to build a client-side user model to be queried by content providers. By doing this the user has total control over his/her user model and the content provider does not need to save and maintain user profiles, therefore privacy risks are significantly reduced.

Before proceeding forth into the technical matter we would like to point out that our approach heavily relies on the availability of semantic metadata embedded in Web pages: the more metadata available, the more detailed the user profile will be; vice versa, if visited pages do not contain metadata, no user profile can be built. Luckily, according to a recent study [2], a huge number of Web sites actually provides semantic annotations, consisting of Microformats, Microdata, or RDFa data, mostly conformed to Facebook's Open Graph metadata protocol[1], hCard, or the Schema.org[2] vocabulary. Since its announcement in 2010, Open Graph caused many concerns about its privacy and service-dependency issues [14], but however, it is greatly contributing to link the Web together. After its adoption by several major players of the WWW, including Google and all its related services, it has affirmed as a de facto standard for RDFa metadata, it has been integrated in CMSs such as Drupal, and can be found in almost any noteworthy site.

The rest of the paper is organized as follows: in Section 2 we briefly introduce some related works; in Section 3 we present our system; in Section 4 we illustrate our data model; in Section 5 we discuss some experimental results and, finally, in Section 6 we conclude the paper.

[1] http://ogp.me/

[2] http://schema.org/

2 Related Work

Several authors have already addressed the problem of generating, parsing, and interpreting structured metadata embedded in Web sites. Automatic metadata generation has been widely explored and can be achieved in many ways: extracting entities from text, inferring hierarchies from folksonomies [13], or exploiting external structured data [8]. Interoperability issues among various metadata formats have been discussed as well: for instance, the authors of [1] propose a metadata conversion tool from microformats to RDF.
Other authors have discussed how Semantic Web tools, such as ontologies and RDF, can be used to model users' behaviours and preferences in Recommender Systems [5]. However, the field on which most research efforts are focused is Personalized Information Retrieval. For instance in [12] is presented an approach towards Ontological Profile building exploiting a domain ontology: as the user interacts with the search engine, interest scores are assigned to concepts included in the ontology with a spreading activation algorithm. The authors of [4] discuss a system that builds a user model aggregating user queries raised within a session and matching them with a domain ontology. Finally, the authors of [3] and [10] suggest that ontological user models can be built as "personal ontology views", that are projections of a reference domain ontology deduced by observing user interest propagation along an ontology network. However, in all these works, user profiles are specializations or projections of a domain ontology and therefore their effectiveness relies on the availability, scope, and quality of such asset.
A recent patent application [15] also claims that the so-called targeting advertising can greatly benefit from the use of semantic user models extracted from Web usage data. The authors, however, do not provide any hint on their extraction technique, focusing, instead, on the architecture and deployment issues of their system. Though many authors have discussed the issues above mentioned, no one, to the best of our knowledge, has ever discussed how to exploit semantic metadata for building personalized interest profiles.

3 System Architecture

In order to support our claims, we developed an experimental system consisting in a client and a server module built using well-known open source tools such as Apache Jena and Semargl. Figure 1 shows the workflow of the system. The basic idea behind our work is that user interests can be identified by observing browsing activity and by analysing the content of visited Web sites, thus our goal is to exploit the user himself as an intelligent Web crawler to provide meaningful data for building his/her personal profile, therefore the project was named *Users As*

Crawlers (herein *UAC*). A compact OWL2 ontology, herein referred as *UAC ontology*, was developed as well in order to introduce new modelling primitives and to allow classification of instances. Among others, the primitives defined in the UAC ontology are: relatedTo, which associates Web pages with DBpedia entities named in the metadata, nextInStream, which associates a page with the next one visited by the user, and previousInStream, which is the inverse of nextInStream.

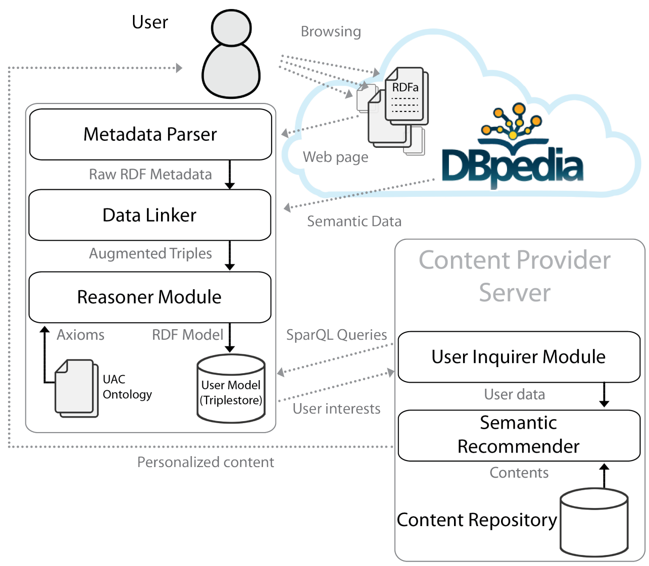


Figure 1. Work flow of the System.

The client module is made of a *Metadata Parser*, a *Data Linker* module, a *Reasoner Module*, and a compact triplestore. The Matadata Parser reads the header sections of the visited web pages and extracts RDF triples from available metadata. Due to its large availability, the preferred metadata format is OpenGraph RDFa, however other formats are allowed as well, as long as they can be converted into RDF. The Data Linker receives the collected triples as input and adds new triples linking visited pages with DBpedia entities. This task is accomplished by both expanding URIs pointed by object properties and by analysing the content of datatype properties such as *tag*, *title*, and *description* with basic NLP techniques in order to find possible matches with DBpedia entries. Finally, the augmented set of triples is processed by a Reasoner module, performing logic entailments in order to classify visited pages. In our prototype the reasoning task is performed by the OWL Lite Reasoner that comes bundled with Apache Jena, but any other OWL reasoner (e.g: Pellet) could fit as well. The result of this process is a semantic user model, built incrementally as the user visits Web pages, in which visited pages are classified and have a hopefully high number of semantic properties linking them each other and to DBpedia. In our prototype system the client part is a standalone application, however, in a production scenario it could be a Web browser plug-in, in order to incrementally build the user profiles as pages are downloaded by the Web browser.

The server part of the system is designed to simulate a content provider scenario and consists in two modules,a *Semantic Recommender*, and a *User Inquirer*, and in a content repository. We assume each content to be addressed towards a specific user stereotype, which is a realistic assumption since many e-commerce companies already do market segmentation analysis. We exploit such knowledge in order to map user characteristics into a specific stereotype and therefore contents to be recommended. More specifically, in our current experimental system we use a decision tree for classifying the user, as shown in Figure 2. Each node is associated with a specific SPARQL query and each arc corresponds to a possible answer to the parent node's query. Stereotypes are identified on the leaves of the tree. When a client connects, it receives the SPARQL

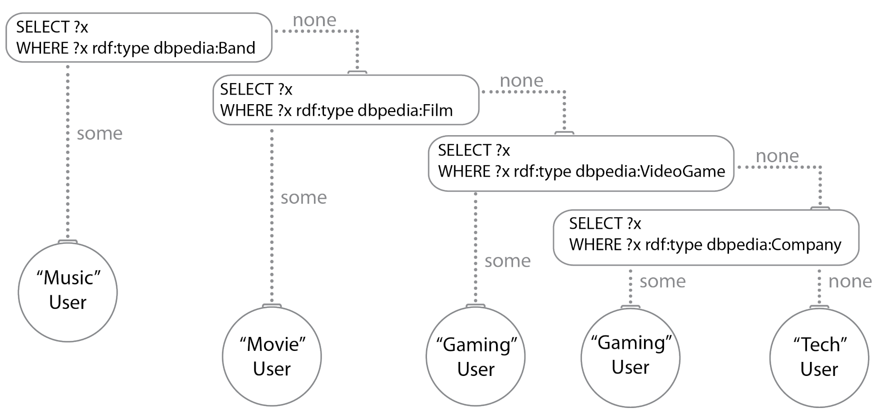


Figure 2. A decision tree with SPARQL queries on the nodes and user stereotypes on the leaves.

query associated with the root node in order to check whether a specific characteristic is present in the user model. The Semantic Recommender module handles the client's answer to the query and fetches content or further queries from the content repository. Due to the hierarchical nature of the decision tree, we expect the number of queries to be asked to the client before being able to identify relevant content to be very small: indeed, in our experimental setting in the worst case six queries were needed.

4 Structured data augmentation and classification

Metadata commonly embedded in Web pages actually provide a very shallow description of the page's content: the Open Graph protocol itself specifies only four attributes as mandatory (title, image, type, and url) and six object classes (video, music, article, book, profile, and website). However, these informations are a good starting point, especially when a few optional properties too are specified, providing "hooks" to more descriptive ontologies.

Instead of focusing on a particular domain ontology, in this work we have chosen to adopt a general purpose and freely available reference ontology: DBpedia. This choice is motivated by three factors: (i) in a realistic scenario it is impossible to restrict users' Web usage to a particular domain, (ii) authors may describe their contents in ways non compliant to a single taxonomy crafted by a domain expert, therefore, the ontology needs to be the result of a collaborative effort, and (iii) since the modelling task is to be accomplished at client-side, we need an ontology freely accessible by anyone.

The Data Linker module of the system analyses the RDF data extracted from the pages in order to find "hooks" to DBpedia, that are named entities present in DBpedia either linked by an extracted Object Property or present as strings in the body of some Datatype Property. To this aim, properties such as *title* and *tags* are particularly useful since they clearly identify relevant entities. Another interesting property is *description* which contains a very short text summarizing the content of the page: this can easily be processed by means of stopword removal and POS tagging in order to extract all its meaningful substrings that match DBpedia entries. Once these entities have been identified, they are linked to the Web page RDF representation with a *relatedTo* property, defined in the UAC Ontology. All the *rdf:type*, *dc:subjec*, and *db:type* attributes of the linked entity are then imported into the RDF data, in order to provide further information about the contents of the page and to support the classification task.

The classification task is entirely performed by the *Reasoner Module*, which entails the evidence provided by both extracted and augmented statements with class and property axioms provided by the Open Graph specification and by the UAC ontology. The Open Graph specification, as mentioned above, provides six classes, each of them has a unique set of properties: for instance, the OGP "article" class has the "author", "section", and "published_time" properties, so a page including one or more of those properties in its metadata can be easily labelled as a "article" page. On the other hand, the UAC Ontology provides a "relatedTo" property connecting Web pages (classified as "webSite" objects) to DBpedia entries and a "nextInStram" property linking each page to the next one requested by the user. Related DBpedia entries and adjacent (successively and previously visited) pages are then exploited by the reasoner to infer new *rdf:type* attributes for page elements. The result, as shown in Figure 3, is a twofold classification of visited pages, which are labelled according to the entities they are "relatedTo" and to the form of the content, specified by Open Graph metadata.

```
1  <rdf:Description rdf:about="http://www.sonatype.com/request/2014-
     developer-survey?s2=arba2">
2    <uac:relatedTo rdf:resource="http://dbpedia.org/resource/
       Open_Source"/>
3    <uac:relatedTo rdf:resource="http://dbpedia.org/resource/
       Sonatype"/>
4    <uac:relatedTo rdf:resource="http://dbpedia.org/resource/
       Development"/>
5    <og:title xml:lang="en">4th Annual Open Source Development
       Survey - Sonatype.com</j.1:title>
6    <og:locale xml:lang="en">en_US</j.1:locale>
7    <uac:nextInStream rdf:resource="https://www.surveymonkey.com/s
       /2014_OpenSource?s2=arba2"/>
8    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
9    <og:type xml:lang="en">website</j.1:type>
10 </rdf:Description>
```

Figure 3. A snippet of generated RDF data.

In the example illustrated in 3 it can be noticed how Open Graph properties (lines 5, 9 and 9) are maintained and the original RDF Description element of the Web page is enriched with UAC properties (lines 2, 3, 4, and 7). In this case, the Open Graph property "og:type" (line 9) has the value "website", classifying the item as a Web portal, and the UAC "relatedTo" property has, among others, the "Open_Source" value, which, in DBpedia, is connected to entities such as "Standards" and "Free_Software". By doing so, our model provides a semantic representation of the content visited so far by the user and its form (e.g: website, article, video, ...).

5 Evaluation

Formative tests were performed in order to evaluate the accuracy of the proposed method. In our experiment, we asked a number of volunteers (mostly university students) to let us use their browsing history data, in order to have real-world data. In order to avoid biases, browsing data was asked to be relative to sessions occurred in the five days before the test subjects were asked to supply data, moreover all test subjects were completely unaware of the real purpose of the experiment. After supplying the data, volunteers were asked to review their own browsing history in order to identify different sessions and to point out what they were actually looking for. At the end of a process we were able to identify six user stereotypes, much like market analysts do when performing segmentation analysis. Since we had no real content to provide in this experiment, we only classified users. The six identified stereotypes are: (i) people mostly interested in economics (nicknamed *business*), (ii) mostly interested in courses, seminars, summer schools and other educational events (*student*), (iii) mostly interested in films and tv series (*moviegoer*), (iv) mostly interested in music (*musician*), (v) mostly interested in videogames (*gamer*), and, finally, (vi) people whose main interests are hardware, programming, and technology in general (*techie*). Three iterations of the data gathering and testing process were performed, each time with different volunteers, in order to test our approach with different users with different browsing habits, and different size of the training set. In the first iteration 36 browsing sessions were collected and labelled, in the second 49 and in the third 69.

Over the three iterations, the average number of Web sites visited in a single browsing session was 31.5 and the average number of triples extracted from a browsing session was 472.8.

During each iteration of the evaluation, the *rdf:type* properties of the visited Web pages were considered as features and used to train a Decision Tree algorithm. In this experiment the J48 algorithm [11] was used; in Figure 4 we show an example of a generated tree, built during the third iteration. The nodes of the tree were then replaced with SPARQL queries and then this structure was used to classify a validation set of user models. A ten-fold cross validation approach was used to estimate the accuracy of the system. Table 1 shows the results of the classification

```
Band <= 0
|   OfficeHolder <= 0
|   |   VideoGame <= 0
|   |   |   Person <= 0
|   |   |   |   Magazine <= 0
|   |   |   |   |   EducationalInstitution <= 0: techie (29.0/15.0)
|   |   |   |   |   EducationalInstitution > 0: student (2.0)
|   |   |   |   Magazine > 0: gamer (3.0/1.0)
|   |   |   Person > 0
|   |   |   |   Organisation <= 0: moviegoer (12.0/2.0)
|   |   |   |   Organisation > 0
|   |   |   |   |   Magazine <= 0: moviegoer (2.0/1.0)
|   |   |   |   |   Magazine > 0: business (2.0)
|   |   VideoGame > 0: gamer (5.0)
|   OfficeHolder > 0: business (5.0)
Band > 0: musician (9.0)
```

Figure 4. A decision tree built during the third iteration of the experiment

over the three iterations of the data set. Our system was compared with the ZeroR predictor, which always returns the mode value of the training set in order to have a baseline. For this formative experiment, only the precision metric (defined as the number of correctly classified instances over the total number of instances) was considered. Though precision

Table 1. Average precision of the UAC system and of a ZeroR classifier on the considered data sets.

Data Set size	ZeroR precision	Tree precision
36	0,306	0.639
49	0,195	0.601
69	0,217	0.623

values are not very high, it is important to point out two limitations of the performed tests: the number of considered browsing sessions is extremely low, due to the fact that only a handful of volunteers let us analyse and use freely their browsing history data; in fact many volunteers dropped out as soon as they realized that their actual browsing history and not some laboratory activity was needed. Secondly, these results were obtained by considering only the rdf:type attribute as feature when building the decision tree. Evaluation and development are ongoing and further experiments, with more test users, more stereotypes, and a richer RDF vocabulary are planned.

6 Conclusion

In this paper we presented a new approach towards ephemeral personalization on the Web, relying on semantic metadata available on the web and, even though the presented results are still preliminary, the overall outcome is very promising. With the growth of the Web of Data, we expect in the next few years to be able to raise the average number of extracted triples from a browsing session and therefore build more detailed user profiles.

In our opinion this approach could fit particularly well to the application domain of targeted advertising because of three major advantages over the actual cookie-based techniques: (i) our approach can predict whether a user may like a content he/she has never seen before, rather than associate a user with a set of already visited (and potentially disliked) contents (ii) the explicit decision model of the decision tree can easily be reviewed by domain experts, supporting market analysis and knowledge engineering, and (iii) by deploying the user model at client side, the user has total control over his/her own data, addressing many privacy concerns. However, the proposed approach has one major drawback: in order to receive personalized contents, users have to install a client, which may be either a browser plug in or a standalone application. Anyway, this seems to be necessary for providing real privacy and also other works aimed at addressing the privacy issues of online advertising have stated the need of a software agent [6].

References

1. Adida, B.: hgrddl: Bridging microformats and rdfa. Web Semantics: Science, Services and Agents on the World Wide Web 6(1), 54–60 (2008)
2. Bizer, C., Eckert, K., Meusel, R., Mühleisen, H., Schuhmacher, M., Völker, J.: Deployment of rdfa, microdata, and microformats on the web–a quantitative analysis. In: The Semantic Web–ISWC 2013, pp. 17–32. Springer (2013)
3. Cena, F., Likavec, S., Osborne, F.: Propagating user interests in ontology-based user model. AI* IA 2011: Artificial Intelligence Around Man and Beyond pp. 299–311 (2011)
4. Daoud, M., Tamine-Lechani, L., Boughanem, M., Chebaro, B.: A session based personalized search using an ontological user profile. In: Proceedings of the 2009 ACM symposium on Applied Computing. pp. 1732–1736. ACM (2009)
5. Gao, Q., Yan, J., Liu, M.: A semantic approach to recommendation system based on user ontology and spreading activation model. In: Network and Parallel Computing, 2008. NPC 2008. IFIP International Conference on. pp. 488–492 (Oct 2008)
6. Guha, S., Cheng, B., Francis, P.: Privad: practical privacy in online advertising. In: Proceedings of the 8th USENIX conference on Networked systems design and implementation. pp. 13–13. USENIX Association (2011)
7. Lambrecht, A., Tucker, C.: When does retargeting work? information specificity in online advertising. Journal of Marketing Research 50(5), 561–576 (2013)
8. Liu, X.: Generating metadata for cyberlearning resources through information retrieval and meta-search. Journal of the American Society for Information Science and Technology 64(4), 771–786 (2013)
9. Mizzaro, S., Tasso, C.: Ephemeral and persistent personalization in adaptive information access to scholarly publications on the web. In: Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems. pp. 306–316. AH '02, Springer-Verlag, London, UK, UK (2002), http://dl.acm.org/citation.cfm?id=647458.728228
10. Osborne, F.: A pov-based user model: From learning preferences to learning personal ontologies. In: User Modeling, Adaptation, and Personalization, pp. 376–379. Springer (2013)
11. Quinlan, J.R.: C4. 5: programs for machine learning, vol. 1. Morgan kaufmann (1993)
12. Sieg, A., Mobasher, B., Burke, R.D.: Learning ontology-based user profiles: A semantic approach to personalized web search. IEEE Intelligent Informatics Bulletin 8(1), 7–18 (2007)
13. Tang, J., Leung, H.f., Luo, Q., Chen, D., Gong, J.: Towards ontology learning from folksonomies. In: IJCAI. vol. 9, pp. 2089–2094 (2009)
14. Wood, M.: How facebook is ruining sharing. Weblog post 18 (2011)
15. Yan, J., Liu, N., Ji, L., Hanks, S.J., Xu, Q., Chen, Z.: Indexing semantic user profiles for targeted advertising (Sep 10 2013), uS Patent 8,533,188

Work in Progress: Multicultural Concept Map Editor

Iñaki Calvo, Ana Arruarte, Jon A. Elorriaga, Mikel Larrañaga

University of the Basque Country (UPV/EHU)
{inaki.calvo, a.arruarte, jon.elorriaga, mikel.larranaga}@ehu.es

Abstract. This paper presents a project that aims to develop a collaborative Concept Map Editor that will provide the necessary functionality to be used by multicultural teams.

1 Introduction

Internationalization and globalization have become familiar terms in current developed societies. The political, economic, cultural, and social changes caused by globalization have made an impact on society especially with the incorporation of Information and Communication Technologies (ICTs). Government strategies focus their attention on ensuring citizens equal opportunities to use ICTs. But ICTs should be prepared also to support cultural diversity. Possibly the single biggest factor that global teams need to address is cultural difference [5]. It is imperative to ICTs that encourage mutual understanding and bridge the difference in cultures [3]. But when developing such tools, it is also important to maintain some of the differentiation allowed by modern information technology to preserve such differences [6].

Since Novak [4] placed concept mapping on the educational agenda, it has become an increasingly popular advanced teaching and learning tool. The fundamentals of concept mapping are in Ausubel's learning theory [1]. A Concept Map (CM) is a graphical way of representing and organising knowledge. It is comprised of nodes and links, arranged in some order to reflect the domain information being represented. Nodes symbolize concepts, and links represent relationship between concepts.

This paper presents the functionality that has been included in a collaborative Concept Map editor to allow multicultural concept mapping.

2 Identifying functionality for multicultural concept mapping

In [2], a survey-based cross-cultural study was presented. The objectives of the study were the identification of the requirements for concept mapping editors when considering multicultural issues. Eleven university students from seven countries participated in the experiment. Participants of the experiment were asked to use a Concept Map editor to adapt a base CM to their culture and afterwards to complete a survey. From the analysis of the resulting CMs and the responses to the questionnaire, some conclusions were drawn.

Language was found an essential factor when working with multicultural issues. Participants thought that language tools such as dictionaries, translators, spellers, thesauri, should be integrated in a multicultural Concept Map editor. Images and colours were identified as important factors. Finally, spatial distribution of the CM elements was not considered relevant when adapting the CM.

3 From a multilingual CM Editor to a multicultural one

Elkar-CM (Arellano et al., 2006) is a multilingual collaborative CM editor that allows synchronous collaboration based on token-passing. Elkar-CM has been designed following the internationalization-localization guidelines. The tool can be localized not only at interface level but also regarding the final CMs it generates. With this feature multicultural CMs can be drawn using the view mechanism. A CM can have different views, one for each culture. All the views share the same structure, i.e. the same nodes and relationships. However, each view can have its own way of representing nodes and relationships, labels, images, etc. In addition, Elkar-CM provides a chat that is synchronised with the actions performed by the users.

Considering the results of the study mentioned above, some new functionality is being added to Elkar-CM to improve the tools offered to support multilingualism and other mechanisms to allow multicultural concept mapping:

Dictionary: a set of on-line dictionaries has been already included in Elkar-CM. It includes defining dictionaries, bilingual dictionaries for different languages and thesauri. Elkar-CM also provides local dictionaries to improve efficiency: a general dictionary, dictionaries attached to a CM and multilingual dictionaries. It is planned to implement a dictionary adapted to each user.

Translator: the translation functionality has been added. Manual translation is based on the used of multilingual dictionaries and automatic translation uses on-line translators. Thus, CM labels and chat interventions can be translated.

Speller: it is programmed to include spellers for different languages.

Multimedia management: multimedia files attached to the nodes and relations can be localized to different cultures.

Transformation rules: when clear correspondences are found between visual characteristics in different cultures (e.g. colour) transformation rules could be defined to help in the adaption of the CM to other cultures. A simple transformation mechanism has been implemented but it has to be improved.

Elkar-CM will be tested with multicultural teams composed of people from close cultures and more diverse teams.

Acknowledgements: This work is supported by the Basque Government (GIC12/79) and the University of Basque Country (UPV/EHU) (UFI11/45).

References

1. Ausubel, D.P.: Educational Psychology; A cognitive View. N.Y.: Holt, Rinehart and Winston (1968).
2. Calvo, I., Arruarte, A., Elorriaga, J.A., Larrañaga, M., McCalla, G.: Identifying Meaningful Concept Map Elements from a Cultural Perspective. IEEE 13th Int. Conf. on Advanced Learning Technologies, IEEE Computer Society, pp. 250-252 (2013).
3. Hachimura, K., Ishida, T., Tosa, N.: Proceedings of the Second International Conference on Culture and Computing, IEEE Computer Society (2011).
4. Novak, J.D.: A theory of education. Cornell University, Ithaca, NY. (1977)
5. Olson, G.M., Olson, J.S.: Distance Matters. Human Computer Interaction,15(2),139-178 (2000).
6. Van Alstyne, M. & Brynjolfsson, D.: Widening Access and Narrowing Focus: Could the Internet Balkanize Science? Science, 274, 1479-1480 (1996).

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].

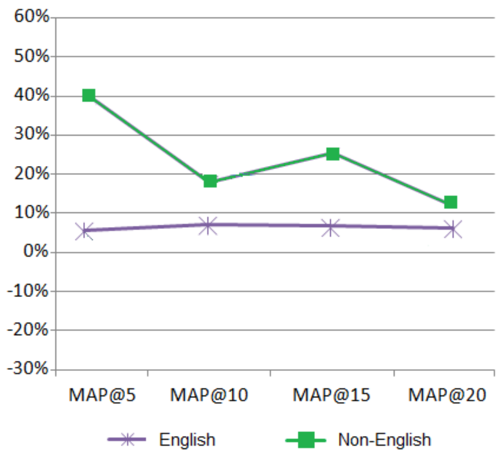


Fig. 1. MAP improvements for English vs. Non-English

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)