Chapter 1

DETECTING SESSION BOUNDARIES TO PERSONALIZE SEARCH USING A CONCEPTUAL USER CONTEXT

Mariam Daoud, Mohand Boughanem, Lynda Tamine-Lechani
IRIT, University of Paul Sabatier
118, Route de Narbonne
Toulouse, France
Email: $\{daoud, bougha, lechani\}$ *@irit.fr*

Abstract Most popular Web search engines are carachterized by "one size fits all" approaches. Involved retrieval models are based on the query-document matching without considering the user context, interests ang goals during the search. Personalized Web search tackles this problem by considering the user interests in the search process. In this chapter, we present a personalized search approach which adresses two key challenges. The first one is to model a conceptual user context across related queries using a session boundary detection. The second one is to personalize the search results using the user context. Our experimental evaluation was carried out using the TREC collection and shows that our approach is effective.

1. INTRODUCTION

Most popular web search engines accept keyword queries and return results that are relevant to these queries. Involved retrieval models use the content of the document and their link structure to assess the relevance of the document to the user query. The major limitation of such systems is that information about the user is completely ignored. They return the same set of results for the same keyword query even though these latter are submitted by different users with different intentions. For example, the query *python* may refer to *python* as a snake as well as the programming language.
Personalized Web search aims at tailoring the search results to a partic-

ular user by considering his profile, which refers to his interests, preferences and goals during the search. One of the key challenges in personalized web search are how to model accurately the user profile and how to use it for an effective personalized search.
User profile could be inferred from the whole search history to model long term user interests [Tan et al., 2006] or from the recent search history [Sriram et al., 2004] to model short term ones. According to several studies [Sriram et al., 2004], mining short term user interests is more effective for disambiguating the Web search than long term ones. User profile representation model has also an impact on the personalized retrieval effectiveness. Involved models could be arranged from a very simple representation based on bags of words to complex representation based on concept hierarchy, namely the ODP [Liu et al., 2004], [Sieg et al., 2007].
This chapter focuses on learning short term user interests to personalize search. A short term user interest is represented by the user context in a particular search session as a set of weighted concepts. It is built and updated across related queries using a session boundary identification method. Search personalization is achieved by re-ranking the search results of a given query using the short term user context.
The chapter is organized as follows. Section 2 presents some related works of search personalization approaches and session boundary identification propositions. Section 3 presents our contribution of search personalization and session boundary detection. In section 4, we present an experimental evaluation, discussion and results obtained. In the last section, we present our conclusion and plan for future work.

2. RELATED WORKS

Personalized Web search consists of two main components: the user profile modeling and the search personalization processes.
User profile is usually represented as a set of keyword or class of vectors [Tamine-Lechani et al., 2008],[Gowan, 2003] or by concept hierarchy issued from the user search history [Micarelli and Sciarrone, 2004],[Kim and Chan, 2003]. As we build the user context using the ODP ontology, we review some related works based on the same essence. The ODP ontology is used in [Liu et al., 2004] to learn a general user profile applicable to all users represented by a set of concepts of the first three levels of the ontology. Instead of using a set of concepts, an ontological user profile in [Gauch et al., 2003] is represented over the entire ontology by classifying the web pages browsed by the user into its concepts. Similar to this last work, an ontological user profile is described in [Sieg

et al., 2007] where the concept weights are accumulated using a spreading activation algorithm that activates concepts through the hierarchical component of the ontology.
The user profile is then exploited in a personalized document ranking by means of query reformulation [Sieg et al., 2004], query-document matching [Tamine-Lechani et al., 2008] or result re-ranking [Gauch et al., 2003]. Mining short term user interests in a session based search requires a session boundary mechanism that identifies the most suitable user interests to the user query. Few studies have addressed this issue in a personalized retrieval task. We cite the UCAIR system [Sriram et al., 2004] that defines a session boundary detection based on a semantic similarity measure between successive queries using mutual information.
Unlike previously related works, our approach has several new features. First, we build a semantic user context as a set of concepts of the ODP ontology and not as an instance of the entire ontology [Gauch et al., 2003]. The main assumption behind this representation is to prune non relevant concepts in the search session. Second, we build the user profile across related queries, which allow using the most suitable user interest to alleviate an ambiguous web search.

3. DETECTING SESSION BOUNDARIES TO PERSONALIZE SEARCH

Our general approach for search personalization consists of modeling a conceptual user context built across a search session to improve the web search accuracy of related queries. We define a session boundary recognition method that allows grouping a sequence of related queries in the same search session. Involved approach is detailed in next sections by: (a) modeling the user context and use it to personalize search, (b) describing the session boundary recognition mechanism.

3.1 A CONCEPTUAL USER CONTEXT MODELLING FOR PERSONALIZING SEARCH

We present in this section how to build and update the user context in a search session and how to exploit it in a personalized document re-ranking.

3.1.1 Building the user context using the ODP metadata.

We build a short-term user context that refers generally to the user's topics of interests during a search session. It is built for each submitted query using the relevant documents viewed by the user and the ODP ontology. The first step consists of extracting the keyword user context K^s for a submitted query q^s. Let D_r^s be the set of documents returned with respect to query q^s and judged as relevant by the user, each represented as a single term vector using the tf*idf weighting scheme. The keyword user context K^s is a single term vector where the weight of term t is computed as follows:

$$K^s(t) = \frac{1}{|D_r^s|} \sum_{d \in D_r^s} w_{td} \tag{1.1}$$

w_{td} is the weight of term t in document d. The concept-based user context is built by first mapping it on the ODP ontology, then disambiguating the obtained concept set using a sub-concept aggregation scheme.

(A) Mapping the keyword user context on the ontology
The user context K^s is matched with concepts of the ODP ontology using the cosine similarity measure. Each concept of the ODP is related to sub-concepts with "*is-a*" relations and is associated to a set of web pages classified under that concept. We represent each concept by a single term vector $\vec{c_j}$ where terms are extracted from all individual web pages classified under that concept as detailed in a previous work [Daoud et al., 2008]. Given a concept c_j, its similarity weight $sw(c_j)$ with $\vec{K^s}$ is computed as follows:

$$sw(c_j) = cos(\vec{c_j}, \vec{K^s}) \tag{1.2}$$

We obtain a set of concepts that contain relevant concepts at different levels and having different weights. In the next section, we proceed by disambiguating the concept set in order to rank the most relevant concepts of general level in the top of the user context representation.

(B) Disambiguating the concept set
We aim at disambiguating the obtained concept set using a sub-concept aggregation scheme. We retain the level three of the ontology to represent the user context. Indeed, the level two of the ontology is too general to represent the user interests, and leaf nodes are too specific to improve the web search of related queries.
Then, we recomputed the weight of each weighted concept by summing the weights of its descendants. We are based on the assumption that the most relevant concepts are those having a greater number of descendant concepts weighted according to the ontology. As shown in figure 1.1(a), we identify a cluster of weighted concepts having a common general depth-three concept; we assign to this latter a relevance score computed by adding the weights of its descendant concepts as shown in figure

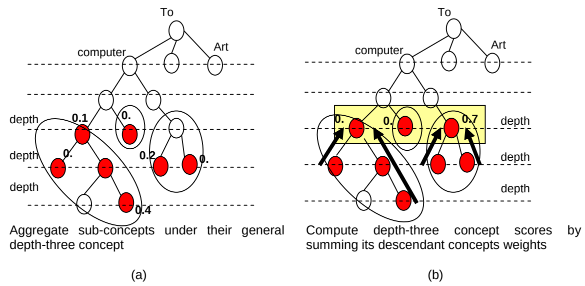

Figure 1.1 Disambiguating the user context

1.1(b). The weight of a general concept c_j, having a set of n related descendant concepts $S(c_j)$, is computed as follows:

$$sw(c_j) = \frac{1}{n} \cdot \sum_{1 \leq k \leq n \wedge c_k \in S(c_j)} sw(c_k) \tag{1.3}$$

We finally represent the user context C^s performed at time s as a set of depth-three weighted concepts, noted $< c_j, p(c_j) >$.

3.1.2 A session-based context updating. We update the user context across queries identified as related using a session boundary recognition mechanism described in the next section. Let C^{s-1} and C^s be respectively the user contexts for successive and related queries. Updating method is based on the following principles: (1) enhance the weight of possible common concepts that can appear in two successive user contexts, (2) alter the weight of non-common concepts using a decay factor β. This allows taking into account the most recent concepts of interests to the user in the search session. The new weight of a concept c_j in the user context C^s is computed as follows:

$$sw_{C^s}(c_j) = \begin{cases} \beta * sw_{C^{s-1}}(c_j) + (1-\beta) * sw_{C^s}(c_j) \ if \ c_j \in C^{s-1} \\ \beta * sw_{C^s}(c_j) \ otherwise \end{cases} \tag{1.4}$$

where $sw_{C^{s-1}}(c_j)$ is the weight of concept c_j in context C^{s-1}, $sw_{C^s}(c_j)$ is the weight of concept c_j in context C^s.

3.1.3 Personalizing search using the user context. We personalize search results of query q^{s+1} related to the user context C^s represented as an ordered set of weighted concepts $< c_j, sw(c_j) >$ by combining for each retrieved document d_k, its initial score S_i and its contextual score S_c as follows:

$$S_f(d_k) = \gamma * S_i(q, d_k) + (1-\gamma) * S_c(d_k, C^s) \tag{1.5}$$

$$0 < \gamma < 1$$

Contextual score S_c of document d_k is computed using the cosine similarity measure between d_k and the concepts of the user context C^s as follows:

$$S_c(d_k, C^s) = \frac{1}{h} * \sum_{j=1..h} sw(c_j) * cos(\vec{d_k}, \vec{c_j}) \tag{1.6}$$

Where c_j is a concept in the user context, $score(c_j)$ is the weight of concept c_j in the user context C^s.

3.2 HOW TO DETECT SESSION BOUNDARIES

We propose a session boundary recognition method using the *Kendall* rank correlation measure that quantifies the conceptual correlation ΔI between the user context C^s and the query q^{s+1}. We choose a threshold σ and believe the queries are from the same session if the correlation is above the threshold.

Here, the term-based query vector $\vec{q_t^{s+1}}$ (where terms are weighted according to their frequency in the query) is matched with concepts of the ontology in order to represent the concept vector $\vec{q_c^{s+1}}$. We adopt a context-sensitive query weighting scheme by introducing the query frequency (QF) in the current search session, in order to rank concepts in the top of $\vec{q_c^{s+1}}$ when they are close to the user context. Indeed, query vector $\vec{q_c^{s+1}} =< w_1, w_2, .., w_i, ... >$ is computed as follows:

$$w_i = CW(q^{s+1}, c_i) * QF(c_i) \tag{1.7}$$

where the query frequency (QF) and the concept weight (CW) are formally defined as:

$$QF(c_i) = \frac{|\vec{q}|^S}{< |\vec{q}|^S, c_i >}, CW(q^{s+1}, c_i) = cos(\vec{q_t^{s+1}}, \vec{c_i}) \tag{1.8}$$

$|\vec{q}|^S$ is the total number of related queries submitted in search session S, $< |\vec{q}|^S, c_i >$ is the number of user contexts built in the current search session S and containing concept c_i.

Thus, the similarity ΔI gauge the changes in the concept ranks between the query and the user context as follows.

$$\Delta I = (\vec{q} o \vec{C^s}) = \frac{\sum_c \sum_{c'} S_{cc'}(\vec{q}) \times S_{cc'}(\vec{C^s})}{\sqrt{\sum_c \sum_{c'} S^2_{cc'}(\vec{q}) \times \sum_c \sum_{c'} S^2_{cc'}(\vec{C^s})}} \tag{1.9}$$

$$S_{c_i c_j}(\vec{v}) = sign(\vec{v}(c_i) - \vec{v}(c_j)) = \frac{\vec{v}(c_i) - \vec{v}(c_j)}{|\vec{v}(c_i) - \vec{v}(c_j)|}$$

Where, c_i and c_j are two concepts issued from both the query and the user context, $\vec{q_c^{s+1}}(c_i)$ (resp. $\vec{C^s}(c_i)$) is the weight of the concept c_i in $\vec{q_c^{s+1}}$ (resp. $\vec{C^s}$).

4. EXPERIMENTAL EVALUATION

Our experimental evaluation is designed to evaluate empirically the accuracy of the session boundary recognition measure and the effectiveness of the search personalization approach.

4.1 EXPERIMENTAL DATA SETS

The experiments were based on the test data provided by TREC collection especially from disks 1&2 of the ad hoc task that contains 741670 documents. We particularly tested topics from $q_{51} - q_{100}$ presented in table 1.1. The choice of this test collection is due to the availability of a manually annotated domain for each query. This allows us, on one hand, to enhance the data set with simulated user interests associated for each TREC domain. On the other hand, we can define a search session as a set of related queries annotated in the same domain of TREC.

Table 1.1 TREC domains with annotated queries.

Domains	*Queries*
Environment	59 77 78 83
Military	62 71 91 92
Law and Government	70 76 85 87
International Relations	64 67 69 79 100
US Economics	57 72 84
International Politics	61 74 80 93 99

4.2 EXPERIMENTAL DESIGN AND RESULTS

Our experiments consist of two particular stages: (1) evaluate the accuracy of the session identification measure; (2) evaluate the personalized retrieval effectiveness of the session-based personalized search.

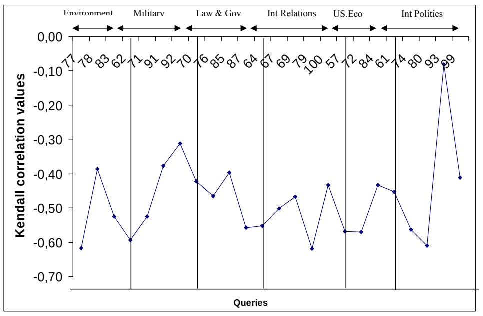

Figure 1.2 Kendall correlation values computed across a query sequence

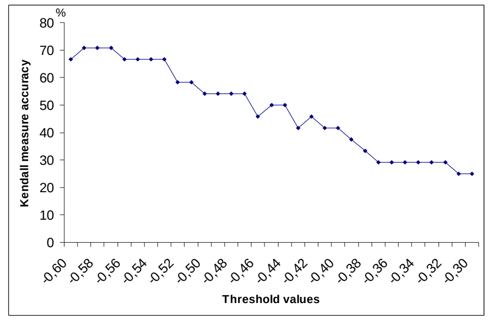

Figure 1.3 Kendall correlation accuracy with varying the threshold value

4.2.1 Evaluating the session boundary recognition method.

The goals of the session boundary recognition experiments are: (A) analyzing query-context correlation values (ΔI) according to the *Kendall* coefficient measure, (B) computing the accuracy of the session identification measure and identifying the best threshold value (σ).
For this purpose, we apply a real evaluation scenario that consists of choosing a query sequence holding six successive sessions related to six domains of TREC listed in table 1.1. For testing purpose, we build the user context for each query using 30 of its relevant documents listed in the TREC assessment file and update it using formula (8) across related queries of the same domain using $\beta = 0.2$.

(A) Analyzing query-context correlations

In this experiment, we computed the query-context correlations values between a particular query and the user context built across previous and queries related to the same TREC domain with respect to the query sequence. Figure 1.2 shows a query sequence holding six search sessions presented on the X-axis and the query-context correlation values on the Y-axis. A fall of the correlation curve means a decrease of the correlation degree with the previous query and possible session boundary identification. A correct session boundary is marked by a vertical line according to the annotated queries in each TREC domain. We can notice that correlation values vary between queries in the same domain. Indeed, in the domain "Environment" of TREC, some queries are related to the environmental concepts of the ODP, while a specific query (q_{59}) is related to the "Weather" topic that has no match with the set of the environmental concepts.
Based on the range of correlation values $[-0.61\ \ -0.08]$, we identify in

the next paragraph the best threshold cut-off value (σ).

(B) Measuring the session boundary measure accuracy
The goal of this experiment is to evaluate the accuracy of the session boundary detection measure. It is computed for each threshold value σ as follows:

$$P(\sigma) = \frac{|CQ|}{|Q|} \tag{1.10}$$

$|CQ|$ is the number of queries identified as correctly correlated to the current user context along the query sequence, and $|Q|$ is the total number of correlated queries in the query sequence.
We show in figure 1.3 the accuracy of the Kendall correlation measure with varying the threshold in the range of ($[-0.61\ \ -0.08]$. The optimal threshold value is identified at -0.58 achieving the optimal accuracy of 70%. We can conclude that the Kendall measure achieves significant session identification accuracy. Indeed, it takes into account the concept ranks in both the query and the user context representations, which makes it tolerant for errors of allocating related queries in different search sessions.

4.2.2 Retrieval effectiveness. Our experimental design for evaluating the retrieval effectiveness consists of comparing the personalized search performed using the query and the suitable user context to the standard search performed using only the query ignoring any user context.
We conducted two sets of controlled experiments: (A) study the effect of the re-ranking parameter γ in the re-ranking formula (9) on the personalized precision improvement, (B) evaluate the effectiveness of the personalized search comparatively to the standard search.
We used "*Mercure*" as a typical search engine where the standard search is based on the BM25 scoring formula retrieval model. We measure the effectiveness of re-ranking search results in terms of Top-n precision (P5, P10) and Mean average precision (MAP) metrics.
The evaluation scenario is based on the k-fold cross validation explained as follows:

- for each simulated TREC domain in table 1.1, divide the query set into k equally-sized subsets, and using $k-1$ training subsets for learning the user context and the remaining subset as a test set,
- for each query in the training set, an automatic process generates the associated keyword user context based on its top n relevant documents listed in the TREC assessment file provided by TREC

using formula (1), and then maps it on the ODP ontology to extract the semantic user context,

- update the user context concept weights across an arbitrary order of the queries in the training set using $\beta = 0.2$ in formula (8) and use it for re-ranking the search results of the queries in the test set using $h = 3$ in formula (9).

(A) Effects of re-ranking parameter γ on the retrieval effectiveness
We present in figure 1.4 the precision improvement graph obtained for the personalized search compared to the standard search at each cutoff of P5, P10 and MAP averaged over the queries belonging to the same domain. In this experiment, we fix the number of relevant documents per query used to represent the user context randomly at 30 for testing purpose. We see that the setting (γ=0.3) produces the best improvement in personalized search since it produces higher precision improvement at P5 (11.63%). This proves that favoring contextual score returned by the system in the result ranking using small values of $\gamma(\gamma < 0.5)$ against the original score gives better performance for improving the web search ranking.

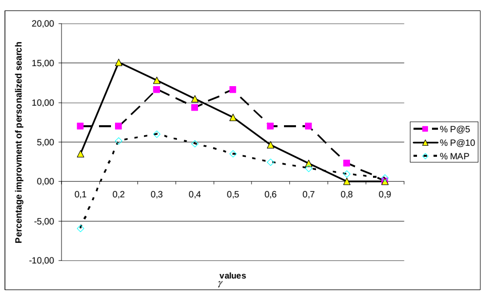

Figure 1.4 Percentage of improvement achieved by personalized search as a result of varying γ in the re-ranking formula (9)

(B) Personalized retrieval effectiveness
In this experiment, we evaluated the effectiveness of the personalized search over various simulated domains. We have computed the percentage of improvement of personalized search comparatively to the standard search computed at P5, P10 and MAP and averaged over the queries belonging to the same domain. Precision improvement ($Impro$) is computed as follows: $Impro = (P_{personalized} - P_{baseline})/P_{baseline} * 100$. We fixed γ at the best value occurring at 0.3 in the re-ranking formula (9). Then, we ran experiment to identify the best number of the relevant document used to represent the user context per query and get 20 at the

best value. Results are presented in Table 1.2.
We see that personalized search improves the retrieval precision of al-

	The baseline model			**Our personalized retrieval model**					
Domain	P5	P10	MAP	P5	Impro	P10	Impro	MAP	Impro
Environment	0.25	0.32	0.18	0.35	40%	0.37	15.3%	0.19	1.7%
Military	0.25	0.27	0.05	0.35	40%	0.32	18.1%	0.07	46.4%
Law Gov	0.40	0.42	0.12	0.50	25%	0.45	5.8%	0.14	12.3%
Inter. Rel.	0.16	0.12	0.01	0.16	0%	0.16	33.3%	0.02	36.5%
US Eco	0.26	0.30	0.09	0.33	25%	0.36	22.2%	0.10	8.3%
Int. Pol	0.16	0.10	0.05	0.20	25%	0.16	60%	0.07	42.2%

Table 1.2 Result effectiveness of the personalized search

most the queries in the six simulated domains. However, the precision improvement varies between domains. This is probably due on one hand, to the accuracy level of the user context representation, and on the other hand to the correlation degrees between queries of the same domain. Indeed some queries annotated in the same domain of TREC may not share concepts of the ontology, then re-ranking search results with not related concepts influences the precision improvement and probably reduce the retrieval performance especially for *Law&Gov* TREC domain.

5. CONCLUSION AND OUTLOOK

In this chapter, we described our approach for a session-based personalized search. It consists of learning short term user interests by aggregating concept-based user contexts identified within related queries. Our experimental results show that the session boundary identification based on the Kendall measure achieves significant accuracy. Moreover, our experimental evaluation shows a significant improvement of personalized retrieval effectiveness compared to the typical search.

In future work, we plan to enhance the user context representation with semantically related concepts according to the ontology. We plan also to evaluate the accuracy of the user context representation as well as its impact on the retrieval effectiveness. Moreover, we intend to evaluate the session boundary recognition measure using real user data provided by web search engine log file, which can reveal comparison with time-based session boundary recognition approaches.

References

[Daoud et al., 2008] Daoud, M., Tamine-Lechani, L., and Boughanem, M. (2008). Using a concept-based user context for search personalization. In *ICDMKE'08: Proceedings of the International Conference of Data Mining and Knowledge Engineering*, pages 293–298. IAENG.

[Gauch et al., 2003] Gauch, S., Chaffee, J., and Pretschner, A. (2003). Ontology-based personalized search and browsing. *Web Intelli. and Agent Sys.*, 1(3-4):219–234.

[Gowan, 2003] Gowan, J. (2003). *A multiple model approach to personalised information access*. Master thesis in computer science, Faculty of science, Universit de College Dublin.

[Kim and Chan, 2003] Kim, H. R. and Chan, P. K. (2003). Learning implicit user interest hierarchy for context in personalization. In *Proceedings of IUI '03*, pages 101–108, New York, NY, USA. ACM.

[Liu et al., 2004] Liu, F., Yu, C., and Meng, W. (2004). Personalized web search for improving retrieval effectiveness. *IEEE Transactions on Knowledge and Data Engineering*, 16(1):28–40.

[Micarelli and Sciarrone, 2004] Micarelli, A. and Sciarrone, F. (2004). Anatomy and empirical evaluation of an adaptive web-based information filtering system. *User Modeling and User-Adapted Interaction*, 14(2-3):159–200.

[Sieg et al., 2007] Sieg, A., Mobasher, B., and Burke, R. (2007). Web search personalization with ontological user profiles. In *Proceedings of the CIKM'07 conference*, pages 525–534, New York, NY, USA. ACM.

[Sieg et al., 2004] Sieg, A., Mobasher, B., Burke, R., Prabu, G., and Lytinen, S. (2004). Using concept hierarchies to enhance user queries in web-based information retrieval. In *The IASTED International Conference on Artificial Intelligence and Applications. Austria*.

[Sriram et al., 2004] Sriram, S., Shen, X., and Zhai, C. (2004). A session-based search engine. In *SIGIR'04: Proceedings of the International ACM SIGIR Conference*.

[Tamine-Lechani et al., 2008] Tamine-Lechani, L., Boughanem, M., and Zemirli, N. (2008). Personalized document ranking: exploiting evidence from multiple user interests for profiling and retrieval. to appear. *In Journal of Digital Information Management*.

[Tan et al., 2006] Tan, B., Shen, X., and Zhai, C. (2006). Mining long-term search history to improve search accuracy. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 718–723, New York, NY, USA. ACM.