Learning user interests for a session-based personalized search

Mariam Daoud
Institut de Recherche en Informatique de Toulouse
118, Route Narbonne
Toulouse, France
daoud@irit.fr

Lynda Tamine-Lechani
Institut de Recherche en Informatique de Toulouse
118, Route Narbonne
Toulouse, France
lechani@irit.fr

Mohand Boughanem
Institut de Recherche en Informatique de Toulouse
118, Route Narbonne
Toulouse, France
bougha@irit.fr

ABSTRACT

It is now widely assumed in personalized information retrieval (IR) area that user interests can provide substantial clues for document relevance estimation. User interests reflect generally the user background and topics of interests. However most of the proposed personalized retrieval models and strategies do not distinguish between short term and long term user interests and make use of the whole search history to improve the search accuracy. In this paper, we study how to learn long term user interests by aggregating concept-based short term ones identified within related search activities. For this purpose, we tackle the problem of session boundary recognition using context-sensitive similarity measures that are able to gauge the changes in the user interest topics with regard to reference ontology. Finally, the search personalization is achieved by re-ranking the search results for a given query using the short term user interest. Our experimental evaluation is carried out using TREC collection and shows that personalization brings significant improvements in retrieval effectiveness. Moreover, we observe that our context-sensitive session boundary recognition method can, to some extent, find a semantic correlation between the query and the user context across the search sessions.

Categories and Subject Descriptors

[**Personalized and collaborative information access in context**]

General Terms

Personalization, web search

Keywords

Personalization, user interests, user context, session boundaries, ontology

1. INTRODUCTION

Most existing retrieval systems could be characterized as ”one size fits all”. This means that the IR decision is based solely on the query document matching. The query is the only evidence that specifies the user need and information about the user and search context is largely ignored. When the query is ambiguous such as *jaguar*, major search engines return results that deal with multiple issues according to the multiple senses behind the query such as documents on the *jaguar* car and on the *jaguar* cat. Thus, the burden is placed on the user to find the relevant answers among the retrieved results. An effective retrieval system should exploit as much contextual factors as possible in order to tailor search results to a particular user.
Contextual IR has become a promising area for disambiguating web search and improving retrieval effectiveness. Personalized IR is a sub-field of contextual IR that focuses on user background, interests and preferences as the main factors of context. According to [8], the user interests are the most important contextual factor, which can alleviate an ambiguous web search in an ad hoc retrieval task. Some works model the user interests as being a rich repository of personal information extracted from the user search history like past clickthrough data [9], browsing features or desktop information [3], etc. Other works manage this repository by using machine learning strategies in order to model the user profile representing his main topics of interests [4][10]. Generally, short term interests reflect a user information need during a short period of time. They could be inferred from the immediate surrounding information, called the short term user context, which highlights on a user's current information need in a single search session. A search session is defined as sequence of search activities related to the same information need. Long term user interests recflect persistent user information needs generally stable for a long time within the search sessions. They could be inferred from the whole user search history, called the long term user context, by detecting the recurrent topics covered by his past queries.

Our research intention in this paper relies on distinguishing between long term and short term user interests. Indeed, long term user interests cannot be as effective as short term ones in improving search accuracy for a particular search session. So, we aim to identify short term user interests that are relevant to a particular search session in order to personalize the immediate search as well as to learn long term ones. Indeed, the user might look for an information need

is related to only one user interest during a short period of time, in this case identifying a specific user interest that best matches the current user's information need, brings better improvement of search accuracy than the use of a general interest. Each of the short term user interests, called the short term user context, is represented as a concept vector using a reference ontology. It is built and updated across related search activities. Identifying short term user interests requires a session boundary recognition that allows grouping related user search activities in the same search session. We define novel context-sensitive session boundary measures that determine whether a query shares the same information need as the current search session.
The remaining sections are organized as follows: in section 2, we review some related works and highlight our motivation. In section 3, we present a session-based personalized search that consists first of building the short term user context, defining a session boundary recognition mechanism, updating the short term user context and finally personalizing search. In section 4, we evaluate our method of session boundary identification and search personalization. In section 5, we present our conclusion and future works.

2. BACKGROUND AND MOTIVATION

Despite the booming popularity of the search engines, the retrieval process is based on simple query-document matching and is made out of the user interests and preferences context. Thus, the retrieved results using the initial user's query may not be satisfactory especially when the query is ambiguous. Search personalization tackles this problem by means of tailoring search results to a particular user according to his profile. The major challenge in this area is how to model accurately the user profile, considered as the user search context, and how to maintain it based on the user's ongoing behaviors.
In [9], the user context is represented by the query history and summaries of viewed documents used to re-rank the search results. SIS [3] is a system for personal information retrieval, which provides a unified index of information that a person has seen. Here, the user context is defined as global personal information such as email, webpage, document, appointment, etc. Other works focus on the user interests as contextual factors inferred from the overall user search sessions using user profiling methods. User interests are represented in [13][11] by clusters of term-vector using solely user feedback techniques as main sources of evidence. More recent studies use external domain ontology as additional evidence than the user feedback to infer the user profile. Such techniques model the user profile as a set of concepts of ontology [10][7][4]. Indeed, ontologies provide a highly expressive ground for describing user interests and a rich variety of interrelations among them. Our approach lies in the field of ontology-based personalized approaches by using the open directory project (ODP) taxonomy as reference ontology [1].
Liu and al. [7] use the first three levels of the ODP to learn a user profile as a set of categories based on the user's search history. The user profile is then used to disambiguate the user's search query. As opposed to using a set of concepts, an ontological user profile in [4] is learned automatically by assigning weights to existing concepts in the Magellan taxonomy. The user profile learning process consists of mapping each visited Web page into five taxonomy concepts with the highest similarities; the concept weights are accumulated based on user's browsing behaviors. Similar to this last work, Sieg and al.[10] present an approach to personalize search, which involves learning an ontological user profile, where maintaining the interest scores is based on the user's browsing behaviors by using a spreading activation algorithm.
Comparatively to previously cited work, our approach differs in several features. The first one is that we make distinction between long term and short term user interests using a session boundary recognition mechanism and make use of these latter in order to enhance the immediate search accuracy. The second point concerns the representation of the user interests. Indeed, while an ontological user profile holds all possible user interests represented over the overall ontology, we propose to represent the user interests, each one as a set of semantically related concepts of the reference ontology. Since long term user interests are inferred by aggregating short term ones, this implies defining a session boundary recognition mechanism that allows grouping user search activities into the same search session. Few studies have addressed the issue of session boundary detection in a personalized retrieval task. We cite the UCAIR system [12] that defines a session boundary detection based on a semantic similarity measure between successive queries using mutual information. Some approaches dedicated for log file analysis purpose detect session boundaries by grouping the data provided by one user based on IP address or else through close proximity in time [6]. The third feature of our approach concerns a novel session boundary recognition mechanism based on a context-sensitive similarity measure applied between the submitted query and the short term user interest.

3. SESSION BASED PERSONALIZED SEARCH

In this section, we present our approach for a session-based personalized search. Our approach considers the following retrieval setting: a user interacts with a document space D through retrieval sessions bounded using a topical closeness similarity measure. Each search activity expresses the following events: the user U submits at time s a query q^s to a search engine; the latter returns a ranked list of documents, then the user expresses his preferences on the document of interests. We assume that a document retrieved by the search engine with respect to q^s is relevant if it generates some observable user behaviors (page dwell time, click through, saving, printing etc). In this setting, a user U has a *priori* long term interests $I^s = \{C^{s_0}, C^{s_1}, ..., C^{s_n}\}$ that we attempt first to learn using short term user contexts mined in the personalized retrieval task. In this paper, we do not address in this paper a possible change to a prior detected user interest in the user profile. Following this general view, the overall process of our session-based personalized search is set according to the algorithm 1. We detail in the remainder of this section the overall process of our approach which relies on three main components:

1. building the short term user context,
2. defining a session boundary recognition mechanism,
3. updating the user context across a search session,

[1] http://www.dmoz.org/

4. personalizing search by re-ranking the search results.

Algorithm 1 Session-based personalized search algorithm

for each new submitted query q^{s+1} **do**
  Compute topical correlation value: $\Delta I = (q^{s+1} \circ C^s)$
  if $\Delta I \geq \sigma$ **then**
    No session boundary is detected:
    * Re-rank the search results returned with respect to the query q^{s+1} using the short term user context C^s.
    * Build the short term user context C^{s+1} at the end of the search activity as follows:

- build a keyword user context K^{s+1}
- map K^{s+1} on reference ontology
- disambiguate the mapped concepts set using a sub-concepts aggregation scheme.

    * Apply a session-based context updating between C^s and C^{s+1} as detailed in section 3.3.
  else
    There is a session boundary recognition:
    case 1: learn a new user interest.
    case 2: refine a prior detected user interest.
  end if
end for

3.1 Building the short term user context

The short-term user context reflects generally the user's topics of interests during a short period of time. We build a concept-based user context at the end of each search activity by using, as main evidences, the relevant Web pages visited by the user and existing domain ontology. Our method for representing the user context relies on extending a previous approach [13] that represents the user context in a flat representation. A keyword user context contains the most representative terms collected from the relevant documents viewed by the user at a specific search activity. Let q^s be the query submitted by a specific user at time s. Let D^s be the set of relevant documents returned with respect to the query q^s, each represented as a single term vector using the tf*idf weightinfg scheme. The keyword user context K^s is a single term vector that represents the centroid of the documents in D^s, where the weight of a term t is computed as follows:

$$K^s(t) = \frac{1}{|D^s|} \sum_{d \in D^s} w_{td} \quad (1)$$

where w_{td} is the weight of term t in document d. In order to enhance the flat representation of the keyword user context, we build the concept-based user context by first mapping it on reference ontology, then disambiguating the mapped concepts set using a sub-concepts aggregation scheme.

3.1.1 Mapping the keyword user context on the ontology

The keyword user context K^s is mapped on the ODP ontology in order to extract a set of concepts that reflect semantically the short-term user interest. Each concept of the ODP is related to sub-concepts with "*is-a*" relations as illustrated in figure 1 and is associated to a set of web pages classified under that concept. We represent each concept by a single term vector $\vec{c_j}$ extracted from all individual web pages classified under that concept as well as all of its sub-concepts. Strategy involved is detailed in a previous work [2] and briefly consists of creating a super-document Sd_j for each concept c_j by concatenating the first 60 titles and descriptions associated to the web pages classified under that concept. Then We remove stop words and use porter stemming on the collection of super-documents. Finally, each concept c_j is represented as a single term vector $\vec{c_j}$ where each term's weight w_i is computed using ttf*idf weighting scheme. Specifically, ttf is the total term frequency in the superdocument Sd_j as well as in the superdocuments associated to its sub-concepts.
Given a concept c_j of the ODP, represented by the term vector $\vec{c_j}$, its similarity weight $sw(c_j)$ with $\vec{K^s}$ is computed as follows:

$$sw(c_j) = cos(\vec{c_j}, \vec{K^s}) \quad (2)$$

We notice that a specific keyword user context could be matched with different portions of ontology. In the next section we detail our method for disambiguating the mapped concepts set in order to select the most relevant ones describing the user context.

3.1.2 Disambiguating the mapped concepts set

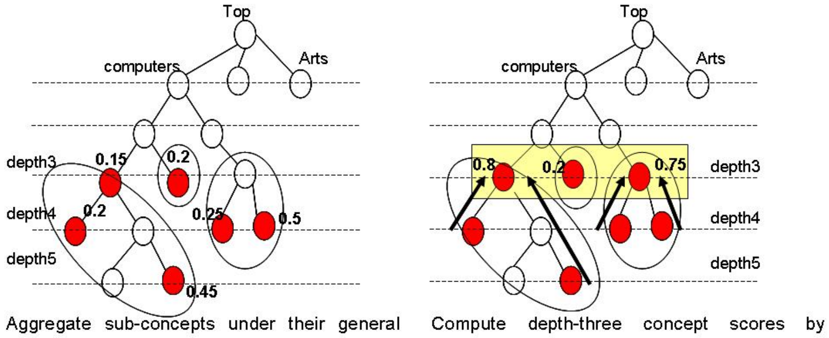
Disambiguating the mapped concepts is carried out using a sub-concepts aggregation scheme, which relies on the assumption that the most relevant concepts are those having a greater number of descendant concepts mapped according to the ontology. We outline that the depth two of the ontology is too general to represent the user interests, and leaf nodes are too specific to improve retrieval precision for related search activities. So we are interested to represent the user context using depth three of the ODP ontology.
As shown in Figure 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 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:

$$sw(c_j) = \frac{1}{n} \cdot \sum_{1 \leq k \leq n \wedge c_k \in S(c_j)} sw(c_k) \quad (3)$$

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

3.2 Session boundary recognition

Our goal in this section is to define a topical-dependant session boundary recognition method that allows determining whether a new query submitted by the user shares the same topic held by the user context built in the current search session. Here, we define dual representations of the query q submitted at time s: the first one is a single term vector, namely $\vec{q_t^{s+1}}$, where terms are weighted according to their frequency in the query. The second one is a concept vector, namely $\vec{q_c^{s+1}} =< w_1, w_2, .., w_i, ... >$ where the concept weight w_i represents the query projection weight on the i^{th} concept c_i of the ODP ontology. Our intuition behind using the query concept-based representation relies on comparing the topics held by the query with the topics held by the user context for a session boundary recognition task. For



Aggregate sub-concepts under their general depth-three concept

(a)

Compute depth-three concept scores by summing its descendant concepts weights

(b)

Figure 1: Disambiguating the mapped concepts of the user context

this purpose, we adopt a context-sensitive weighting scheme by introducing the query frequency (QF) in current search session, which allows ranking concepts which are the closest to the current user interest, in the top of query concept-based vector $\vec{q_c^{s+1}}$.
We propose to compute the the concept weight w_i in the query vector $\vec{q_c^{s+1}}$ by using the concept weight (CW) and the query frequency (QF) as follows:

$$w_i = CW(q^{s+1}, c_i) * QF(c_i) \tag{4}$$

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

$$CW(q^{s+1}, c_i) = cos(\vec{q_t^{s+1}}, \vec{c_i}), QF(c_i) = \frac{|\vec{q}|}{|< q, c_i >|} \tag{5}$$

CW is computed using the cosine similarity measure between the query term-representation $\vec{q_t}$ and the concepts of the ontology. Then, we apply the same disambiguation strategy, used in section 3.1.2, in order to resolve the ambiguity of the mapped concepts set. $|\vec{q}|$ refers to the total number of related queries submitted in the current search session, $|< q, c_i >|$ refers to the number of queries that lead to include the concept c_i in the user contexts built in the current search session.
Using the query concept representation, we define two novel measures that compute a topical correlation degree ΔI between the query submitted at time $s+1$ and the user context performed at time s. The first one is the *Kendall* correlation measure that takes into account the concept ranks between two concept sets and the second one is our modified *WebJaccard* similarity measure that uses the concept counts between two sets.

- *Using the Kendall rank correlation measure*
  We use the Kendall rank order correlation measure to evaluate the topical similarity degree between two sets of concept rank associated to the user query and the current user context. Our choice of this measure relies on the idea that a change of interests across search sessions leads to differences in the rank order of concepts between the query concept representation and the user context.
  So we define the topical correlation degree ΔI between a given query $\vec{q_c^{s+1}}$ submitted at time s and the user context $\vec{C^s}$ performed at time s as follows:

$$\Delta I = Kendall(\vec{q_c^{s+1}} o \vec{C^s}) = \frac{\sum_{c_i}\sum_{c_j} S_{c_ic_j}(\vec{q_c^{s+1}}) \cdot S_{c_ic_j}(\vec{C^s})}{\sqrt{\sum_{c_i}\sum_{c_j} S^2_{c_ic_j}(\vec{q_c^{s+1}}) \cdot \sum_{c_i}\sum_{c_j} S^2_{c_ic_j}(\vec{C^s})}} \tag{6}$$

$$S_{c_ic_j}(\vec{q_c^{s+1}}) = sign(\vec{q_c^{s+1}}(c_i) - \vec{q_c^{s+1}}(c_j)) = \frac{\vec{q_c^{s+1}}(c_i) - \vec{q_c^{s+1}}(c_j)}{\left|\vec{q_c^{s+1}}(c_i) - \vec{q_c^{s+1}}(c_j)\right|}$$

$$S_{c_ic_j}(\vec{C^s}) = sign(\vec{C^s}(c_i) - \vec{C^s}(c_j)) = \frac{\vec{C^s}(c_i) - \vec{C^s}(c_j)}{\left|\vec{C^s}(c_i) - \vec{C^s}(c_j)\right|}$$

  Where, c_i and c_j are two concepts issued from both the query concept representation and the user context. The correlation values ΔI is in the range [-1 1], where a value closer to -1 means that the query and the user context are not similar, and a value closer to 1 means that the query and the current user context are very related to each other.

- *Using our modified WebJaccard similarity measure*
  We modify the traditional *Jaccard* measure [5] for the purpose of measuring topical similarity degree using concept counts between the query concept vector and the concept-based user context. In order to do so, we compute the topical correlation degree ΔI between a given query q^{s+1} submitted at time $s+1$ and the user context C^s performed at time s as follows:

$$\Delta I = WebJaccard(q_c^{s+1}, C^s) = \frac{H(q_c^{s+1} \cap C^s)}{H(q_c^{s+1}) + H(C^s) - H(q_c^{s+1} \cap C^s)} \tag{7}$$

  where $H(q^{s+1} \cap C^s)$ is the number of common concepts in both representations, $H(q^{s+1})$ is the number of the concepts in the query concept vector $\vec{q_c^{s+1}}$ and $H(C^s)$ is the number of concepts in the concept-based user context C^s. The correlation value ΔI is in the range [0 1], where a value equal to 0 means that the query and the user context are not similar, and a value of 1 means that the query and the current user context have all concepts in common and are very related to each other.

3.3 Session-based context updating

Based on our session boundary recognition mechanism, we update the user context using a linear updating scheme. Let C^{s-1} and C^s be respectively the user contexts for successive and related search activities. 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. The new weight of a concept c_j in the user context C^s is computed as follows:

$$sw_{C^s}(c_j) = \begin{cases} \beta * sw_{C^{s-1}}(c_j) + (1-\beta) * sw_{C^s}(c_j) \ if \ c_j \in C^{s-1} \\ \beta * sw_{C^s}(c_j) \ otherwise \end{cases} \tag{8}$$

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. Such user contexts updated across related search activities in the same search session are used further to learn long-term user interests.

3.4 Session-based personalized search

Our goal in personalization is to re-rank the search results returned by a search engine with respect to a given query using the short-term user context. Let's consider the user context C^s performed at time s, containing an ordered set of weighted concepts $< c_j, sw(c_j) >$. Given a related query q^{s+1} submitted at time $s+1$, we re-rank its associated search results by combining for each retrieved result d_k, the initial score S_i returned by the system and a contextual score S_c as follows:

$$S_f(d_k) = \gamma * S_i(q, d_k) + (1-\gamma) * S_c(d_k, C^s) \tag{9}$$

$$0 < \gamma < 1$$

When γ has a value of 0, contextual score is not given any weight, and results are ranked according to the original score. If γ has a value of 1, the original score is ignored and pure contextual score is considered. Both the contextual and the original score could be blended by varying the values of γ.
The contextual score S_c is computed using the cosine similarity measure between the result d_k and the top ranked concepts of the user context C^s as follows:

$$S_c(d_k, C^s) = \sum_{c_j \in C^s \wedge j=1,2,3} sw(c_j) * cos(\vec{d_k}, \vec{c_j}) \tag{10}$$

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

4. EXPERIMENTAL EVALUATION

Our experimental evaluation is designed to evaluate empirically the performance of our proposed approach. Particularly, it addresses the accuracy of the session boundary recognition measures and the effectiveness of our proposed search personalization approach.

4.1 Experimental data sets

The experiments were based on two data sets: the first one is the data provided by TREC collection and the second one is the ODP data set that we created to represent each of the ODP concepts.

Domains	Related 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

Table 1: Trec domains used for simulating the user interests

- TREC data set: as a main test data, we used a TREC data from disks 1&2 of the ad hoc task that contains 741670 documents. We particularly tested queries from $q_{51} - q_{100}$. 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 queries annotated in a particular domain of TREC, and then assume that these queries share the same user information need. We simulated six domains of TREC composed of 25 queries shown in table 1.
- ODP data set: the second data set is the document collection that we created from the ODP ontology. It contains 235331 concepts used to represent the user context. We used "*Mercure*" search engine [1] to index the collection of super-documents.

4.2 Experimental design and results

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

4.2.1 Evaluating session boundary recognition methods

The goals of the session boundary recognition experiments are: (A) analyzing query correlation degrees according to both session identification measures (the *Kendall* coefficient measure and our modified *WebJaccard* measure), (B) comparing the accuracy of both session identification measures in order to identify the best one.
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. So we build the user context for each query in a particular session using 30 of its relevant documents provided by TREC data and update it using formula (8) across related queries of the same domain. The number of relevant documents per query is fixed randomly at 30 for testing purpose. We apply our session boundary identification measures between a particular query and the current user context built or updated after processing previous queries with respect to the query sequence.

(A) Analyzing query-context correlations
In this experiment, we evaluate the correlation degree between a query and the current user context across a given

query sequence for both measures. Figures 2 and 3 show respectively the correlation values computed across a query sequence for the Kendall correlation measure and our modified WebJaccard measure. A query sequence holding four search sessions (represented by TREC domains) is presented on the X-axis, and the corresponding correlation values according to both measures are presented on the Y-axis. A fall of the

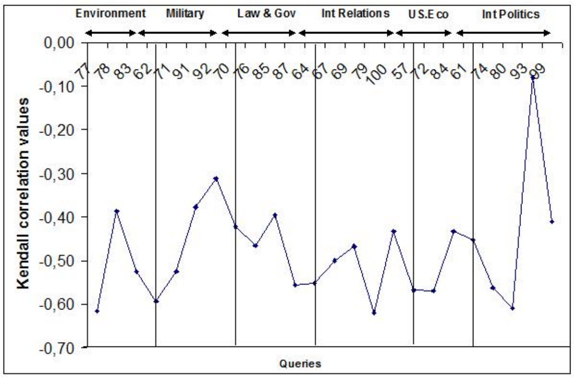


Figure 2: Kendall correlation values computed across a query sequence

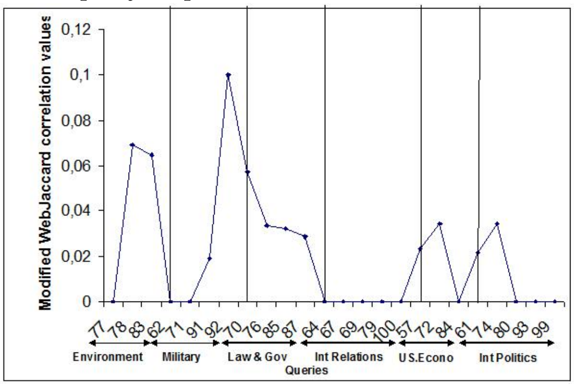


Figure 3: Modified WebJaccard correlation values computed across a query sequence

correlation curve at a particular query 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 at a particular query according to the annotated queries in each TREC domain. We can notice that correlation values vary however between queries in the same domain for both measures. The correlation degree of WebJaccard measure is equal to zero for the sequence of queries from 67 to 57. This is due to the fact that queries allocated in the same TREC domain may not have a topical similarity in the ODP ontology. For example, 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. Comparatively to the Kendall measure, this latter takes into account the concepts ranks in both the query and the user context representation and do not produce -1 correlation values. This makes it tolerant for errors of allocating related queries in different search sessions.
Based on the range of correlation values related to each measure ($[-0.61\ -0.08]$ for the Kendall measure, [0.01 0.07] for our modified WebJaccard measure), we identify a threshold cut-off for each session boundary measure, which allows determining whether a query is correlated to the current search session and then decide whether a session boundary is identified.

(B) Measuring the session boundary measure accuracy
The goal of this experiment is to evaluate the accuracy of the session boundary measures and then identify the best one. Session identification accuracy is computed as follows:

$$P = \frac{|CQ|}{|Q|} \tag{11}$$

Where $|CQ|$ is the number of queries identified as correctly correlated to the current user context according to manually annotated queries of TREC domain, and $|Q|$ is the total number of the queries in the test query sequence.
We show respectively in figures 4 and 5 the accuracy of Kendall correlation measure and our modified WebJaccard measure with varying the threshold in the range of the associated correlation values as mentioned above. An optimal

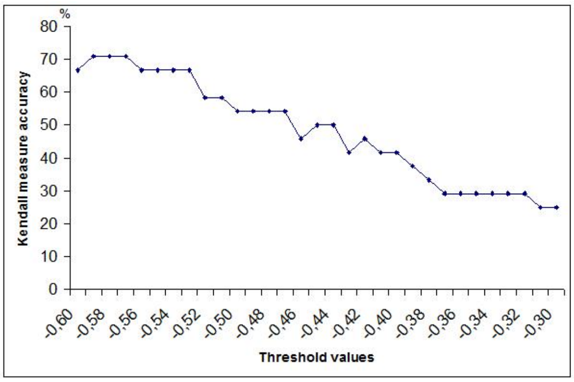


Figure 4: Kendall correlation accuracy as a result of varying the threshold value

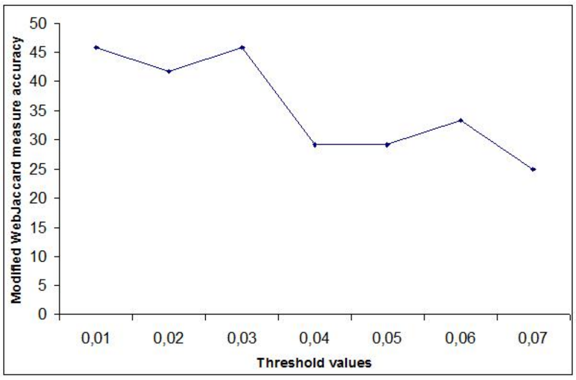


Figure 5: Modified WebJaccard correlation accuracy as a result of varying the threshold value

threshold value for each measure is identified when higher accuracy in identifying correlated search activities is obtained. The evaluation performed on the results reveals that the values -0.58 achieves the optimal accuracy of the Kendall measure (70%) and 0.01 achieves the optimal accuracy of our modified WebJaccard (45.8%).

The Kendall measure gives the better performance comparatively to our modified WebJaccard measure. We explain this by the fact that the WebJaccard measure is considered as a strict measure based on presence/absence of the concepts in both the query and the user context representation. Thus, it is hard to fix a minimal threshold that can alleviate, to some extent, errors of allocating two related search activities in different search sessions. Since the Kendall measure does not produce -1 correlation values, it is easier to define a non-sensitive threshold value reducing such occurring errors.

4.2.2 *Retrieval effectiveness*

Our experimental design for evaluating the retrieval effectiveness consists of evaluating the effectiveness of our personalized approach when using the user context in the IR model for related queries. The evaluation scenario is based on the k-fold cross validation explained as follows:

- for each simulated TREC domain, divide the query set into k equally-sized subsets, and using $k-1$ training subsets for learning the user interests 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 relevance judgements file provided by TREC using formula (1), and then maps it on the ODP ontology using formula (2) in order to extract the user context. Then we disambiguate the mapped concepts set using formula (3),
- update the user context concept weights across an arbitrary order of the queries in the training set using formula (8) and use it for re-ranking the search results of the queries in the test set using formula (9).

We conduct two sets of controlled experiments to examine the effectiveness of our personalization approach: (A) study the effect of the re-ranking parameter γ in the re-ranking formula (9) on the personalized precision improvement, (B) evaluate the percentage of improvement achieved by personalized search compared to the standard search.
We used "*Mercure*" [1] as a typical search engine where the standard search is based on the OKAPI retrieval model. We measure the effectiveness of re-ranking search results in terms of Top-n precision (P@5, P@10) and Mean average precision (MAP) metrics.

(A) Effects of re-ranking parameter γ on the retrieval effectiveness
We present in figure 6 the precision improvement graph obtained for the personalized search compared to the standard search at each cutoff of P@5, P@10 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 and the decay factor used in formula (8) at 0,2. We see that the setting (γ=0,3) produces the best improvement in personalized search since it produces higher precision improvement at P@5 (11,63%). This can be explained by the fact that small values of γ ($\gamma < 0,5$) allows decreasing the original score returned by the system and favouring contextual score in the result ranking.

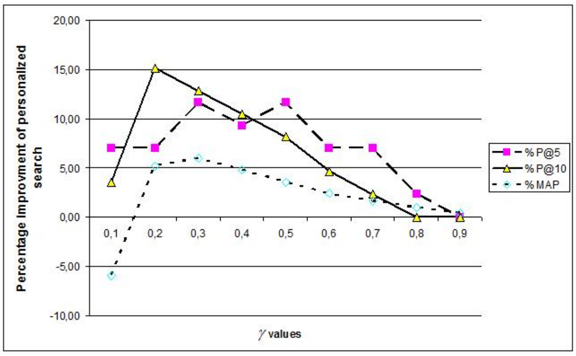


Figure 6: Percentage of improvement achieved by personalized search as a result of varying γ in the re-ranking formula (9)

(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 P@5, P@10 and MAP and averaged over the queries belonging to the same domain. Precision improvement % is computed as follows:

$$\% = (P_{personalized} - P_{baseline})/P_{baseline} * 100$$

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 that maximizes the personalized precision improvement. Results are presented in in table 2, significance of improvement at the 5% level is indicated either by $^{\triangle}$ or $^{\triangledown}$ depending on the direction change compared to the best baseline result; no significance is denoted by $^{\diamond}$.
We see that personalized search improves the retrieval precision of almost 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 paper, we described our approach for a session-based personalized search. It consists of learning long term user interests based on modeling short-term concept-based user contexts identified within related search activities. Unlike most previous related works, we focus on learning the user interests, each one represented as a set of semantically related concepts of reference ontology. Maintaining the concept weights of the user context is achieved across related search activities based on a linear combination scheme. Our approach includes session boundary recognition methods that allow grouping related search activities in the same search session. Our proposed session identification measures

	The OKAPI baseline model			Our personalized retrieval model		
Domain	P@5	P@10	MAP	P@5	P@10	MAP
Environement	0,25	0,32	0,18	0,35	0,37	0,19
Improvement				$^{\triangle}$**20%**	$^{\triangle}$**10,53**%	$^{\diamond}$0,10%
Military	0,25	0,27	0,05	0,35	0,32	0,07
Improvement				$^{\triangle}$**16,67%**	$^{\diamond}$0%	$^{\diamond}$2.27%
Law & Gov	0,40	0,42	0,12	0,50	0,45	0,14
Improvement				$^{\diamond}$0%	$^{\triangledown}$4,55%	$^{\triangle}$**5,34%**
Inter. Rel.	0,16	0,12	0,01	0,16	0,16	0,02
Improvement				$^{\triangle}$**50**%	$^{\triangle}$**57,14%**	$^{\triangle}$**10,09%**
US Eco	0,26	0,30	0,09	0,33	0,36	0,10
Improvement				$^{\triangle}$**50%**	$^{\triangle}$**40%**	$^{\triangle}$**14,48%**
Int. Pol	0,16	0,10	0,05	0,20	0,16	0,07
Improvement				$^{\diamond}$0%	$^{\triangle}$**20%**	$^{\triangle}$**10,74%**

Table 2: Results effectiveness of our personalized search

rely on the topical closeness between the query concept representation and the user context. Personalization is then achieved by re-ranking the search results of related queries using the user context.

Our experimental results show that the session boundary identification based on the Kendall measure achieves better accuracy than our modified WebJaccard measure. Moreover, our experimental evaluation shows a significant improvement of personalized retrieval effectiveness compared to the typical search.

In future work, we plan to gauge changes in the user interests to prior detected ones in the user profile across search sessions. Moreover, we plan to evaluate our session boundary recognition measures using real user data provided by web search engine log file, that can reveal comparison with time-based session boundary recognition approaches.

6. REFERENCES

[1] M. Boughanem, K. Sauvagnat, and C. Laffaire. Mercure at trec 2003 web track - topic distillation task. In *TREC 2003: The Twelfth Text Retrieval Conference*, pages 343–348, 2003.

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

[3] S. Dumais, E. C. J. J. Cadiz, G. Jancke, R. Sarin, and C. R. Daniel. Stuff i've seen: a system for personal information retrieval and re-use. In *SIGIR'03: Proceedings of the International ACM SIGIR Conference*, pages 72–79. ACM Press, 2003.

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

[5] T. H. Haveliwala, A. Gionis, D. Klein, and P. Indyk. Evaluating strategies for similarity search on the web. In *WWW'02:Proc. of the Eleventh International World Wide Web Conference*, pages 432–442, 2002.

[6] B. J. Jansen, A. Spink, and V. Kathuria. How to define searching sessions on web search engines. In O. Nasraoui, M. Spiliopoulou, J. Srivastava, B. Mobasher, and B. M. Masand, editors, *In WEBKDD'06*, volume 4811 of *Lecture Notes in Computer Science*, pages 92–109. Springer, 2006.

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

[8] T. Park. Toward a theory of user-based relevance: A call for a new paradigm of inquiry. *J. Am. Soc. Inf. Sci.*, 45(3):135–141, 1994.

[9] X. Shen, B. Tan, and C. Zhai. Context-sensitive information retrieval using implicit feedback. In *SIGIR '05: Proceedings of the International ACM SIGIR Conference*, pages 43–50, New York, NY, USA, 2005. ACM.

[10] A. Sieg, B. Mobasher, and R. Burke. Web search personalization with ontological user profiles. In *CIKM'07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 525–534, New York, NY, USA, 2007. ACM.

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

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

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