Towards a graph-based user profile modeling for a session-based personalized search

**Mariam Daoud · Lynda-Tamine Lechani ·
Mohand Boughanem**

Abstract   Most Web search engines use the content of the Web documents and their link structures to assess the relevance of the document to the user's query. With the growth of the information available on the web, it becomes difficult for such Web search engines to satisfy the user information need expressed by few keywords. First, personalized information retrieval is a promising way to resolve this problem by modeling the user profile by his general interests and then integrating it in a personalized document ranking model. In this paper, we present a personalized search approach that involves a graph-based representation of the user profile. The user profile refers to the user interest in a specific search session defined as a sequence of related queries. It is built by means of score propagation that allows activating a set of semantically related concepts of reference ontology, namely the *ODP*. The user profile is maintained across related search activities using a graph-based merging strategy. For the purpose of detecting related search activities, we define a session boundary recognition mechanism based on the *Kendall* rank correlation measure that tracks changes in the dominant concepts held by the user profile relatively to a new submitted query. Personalization is performed by re-ranking the search results of related queries using the user profile. Our experimental evaluation is carried out using the HARD 2003 TREC collection and showed that our session boundary recognition mechanism based on the *Kendall* measure provides a significant precision comparatively to other non-ranking based measures like the *cosine* and the *WebJaccard* similarity measures. Moreover, results proved that the graph-based search personalization is effective for improving the search accuracy.

Keywords   Personalization · Graph-based user profile · Ontology · Search session · Session boundaries

M. Daoud (✉) · L.-T. Lechani · M. Boughanem
IRIT, Paul Sabatier University, Toulouse, France
e-mail: daoud@irit.fr

1 Introduction

Most popular Web search engines use the content of the Web documents and their link structures to assess the relevance of the document to the user's query. With the growth of the information available on the web as well as the ambiguity of typical user queries, it becomes difficult for such Web search engines to satisfy the user information need. The involved retrieval approaches are characterized as '*one size fits all*' that provides the same results for the same keyword queries 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 *python* programming language. Thus, the relevance decision is made out of the user background and interests' context by considering the keyword query as the only clue that specifies the user information need.

The first techniques addressed in IR aim at improving the search accuracy using relevance feedback techniques [5,6,38] and word sense disambiguation techniques [44]. Unfortunately, these techniques may be limited in real world applications [25] as they require an explicit user information feedback. Explicit user feedback is either provided by specifying the intention behind the query in the case of disambiguation techniques or by specifying the documents of interest in the case of relevance feedback. Another range of techniques are based on predicting the query difficulty or clarity [56] by measuring the ranked list robustness for improving the query performance.

User-focused IR approaches begins with collaborative filtering used in recommender systems [28,29] which aims at providing personalized recommendations to users based on their previously expressed preferences and on those of other similar users. Collaborative filtering has some limitations potentially related to the volume and the diversity of the information in large scale environment [39]. Indeed, greater is the group of similar users, higher is the time to associate a user to a particular group.

More sophisticated approaches have been proposed in the field of contextual IR, which is defined in [3] as follows:*"combine search technologies and knowledge about query and user context into a single framework in order to provide the most appropriate answer for a user's information need"*.

A contextual retrieval system should involve a mechanism that exploits as much contextual factors as possible in order to tailor search results to a particular user. Personalized IR is a field of contextual IR that models the user profile by his general interests and preferences. The amount of information extracted from the user's retrieval environment constitutes a rich repository managed in [37] by algorithms of data mining [54] or machine learning [53] strategies in order to model the user profile. The most challenging tasks in the field are how to model the user profile and how to personalize the document ranking in order to filter out irrelevant information and rank results in the top when they are the most suitable to the user interests.

User interests are usually represented as a set of keyword vectors [30,49], a topic preference vector [13], class of vectors [16] or by concept hierarchy issued from the user's documents of interests [4,26,36]. More recent studies use external domain ontology as an additional evidence to model the user profile as a set of concepts issued from predefined ontology [1,32] or an instance of predefined ontology [15,43].

Tailoring the search results to a particular user is then carried out through a personalized document ranking using the user profile. Some personalized techniques make distinction between long term and short term user profile and investigates either short term ones [15,41] or long term ones [42] in a personalized document ranking. Generally, a short term user profile refers to the user interests during a short period of time and inferred from the recent

search history. Long term user profile hold persistent user interests generally stable for a long time and inferred from the whole user search history.

Our research intuition in this paper relies on building a short term user profile in a particular search session defined as a sequence of queries that are related to the same user information need. The user profile is represented as a weighted graph of semantically related concepts of predefined ontology, namely the ODP.[1] For each submitted query by the user, we exploit the documents judged relevant by the user to build a query profile represented also as a weighted graph. Query profile is used to initialize and update the user profile across related queries in the same search session. In order to detect related queries, we define a session boundary recognition mechanism that tracks changes in the dominant concepts between the user profile and the query using the *Kendall* rank correlation measure. Finally, the user profile built across related search activities is used to re-rank the search results returned by the system with respect to new queries allocated in the same search session.

The rest of this paper is organized as follows. In Sect. 2, we review related work and highlight our contribution. In Sect. 3, we present terminology and notations, an overview of the graph-based representation of the user profile and the general approach in our contribution. In Sect. 4, we describe our approach for building the user profile and setting a session boundary recognition mechanism. In Sect. 5, we present our search personalization method. The experimental evaluation and results are presented in Sect. 6. In the last section, we present our conclusion and plan for future work.

2 Related work

Personalized search is a promising way in improving the web search rankings by making the user out of hiding during the search. An effective personalized search relies on two main challenging tasks, which are the user profile modeling and the search personalization. Generally, the user profile refers to the user knowledge, interests and preferences that could be inferred during the search. Mining short term user profile in the search personalization task involves a session boundary identification that allows using the most suitable information sources from the search history to infer the user profile. Then, the goal of the search personalization is to tailor search results to a particular user according to his profile.

We review in next sections some user profile modeling techniques, session boundary recognition algorithms and search personalization techniques.

2.1 User profile modeling

User profile modeling approaches have matured in the last decade and can be distinguished by several criteria concerning the temporal dimension of the user profile as being short term or long term and the user profile representation models. A key aspect in most of the user modeling techniques is to collect relevant sources of evidence from the user search history in order to model the user profile. Often, the user search history is composed by the user's documents of interest such as recently browsed or viewed web pages [15,43], a repository of interesting information such as emails, browsing features or desktop information [12]. An integrated platform for gathering, organizing and personalizing the information is proposed in [50] and aims at building personalized information portfolios. The information is gathered

---

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

and clustered in terms of the content as well as the organizational structure according to the user preferences.

Concerning the temporal dimension of the user interests, distinction has been made between long term ones and short term ones. A wide range of personalized IR approaches exploited long term user interests [51] in the search personalization task. Another range exploited short term ones [12,15,21,41]. We outline here that these latter could be inferred either from the recent search activities that could hold multiple user information needs [12,15] or from a single search session defined as sequence of related search activities belonging to the same user information need [41].

According to the representation model, the user interests are often represented as a set of keyword vectors [30,49], class of vectors [16]. To overcome the limitations of the keyword-based representation of the user interests (absence of interrelations between user interests), some approaches model the user interests by term relations [27] or a concept hierarchy [4,26,36] issued from the user's documents of interest. A fuzzy user profile modeling was proposed in [24] where the user profile is represented by a set of categories for which the weights provide the knowledge about the user indicating the membership of a category to some degree.

Even if such representations are complex, they are leveraged from the user knowledge that is often limited and not sufficient to encounter new user intention. To alleviate the weakness of such representations, some approaches make use of a predefined semantic resources to represent the user profile as a set of concepts issued from reference ontology [32] or an instance of reference ontology [15,43]. Ontology-based representation of the user profile has shown several advantages in improving personalized search effectiveness. On one hand, ontology provides a highly expressive ground for describing user interests and a rich variety of interrelations among them. On the other hand, using predefined domain knowledge allows encountering new possible topics of interest without requiring time consuming process for collecting information across multiple search sessions.

Our research work follows ontology-based personalized search approaches that use the *ODP* ontology for representing the user profile. We focus on works that were based on the same essence [15,32,43].

[32] use the ODP ontology to learn a general user profile as a set of concepts of the first three levels of ontology and the user's search history to learn a personal user profile as a set of categories represented by keywords. The user profile is learned using algorithms of text categorization of the user's documents of interest, which allow representing the user profile as a category-term matrix. General and personal user profiles are then used to map the user query to a set of categories exploited in a search personalization task. Using unsupervised neural network algorithm, Ding et al. [11] model the user profile as the user search history is represented in a document-term matrix where documents are labeled by categories. It is used as a training set for classifying new queries into a set categories used further to personalize search.

Instead of using a set of concepts, the user profile in [15] is represented as an instance of the ODP ontology using a supervised classification of the Web pages browsed by the user. Each concept is thus assigned a weight that represents the amount of the web pages classified into that concept. Similar to this last work, Sieg et al. [43] present an approach to personalize search, which involves learning an ontological user profile. While the concept weights in [15] are accumulated along the user's browsing search history, they are accumulated in [43] using a spreading activation algorithm that activates concepts through the hierarchical component of ontology.

2.2 Session boundary recognition algorithms

In order to mine useful information from user navigation patterns, it is more appropriate to cluster user sessions first. A session is defined as a group of queries made by a single user for a single navigation purpose [19]. A user may have a single session or multiple sessions during a period of time. In our case, the purpose of clustering is to identify the most suitable sources of evidence for improving the search accuracy of a given query. Various approaches have been introduced in the literature to cluster user sessions. The most commonly used session identification method is called timeout [18], in which a user session is usually defined as a sequence of queries from the same user such that two consecutive queries are separated by an interval less than a predefined threshold. This session identification method suffers from the problem of setting the best time threshold. Indeed, different users may have different navigation behaviors, and their time intervals between sessions may be significantly different. Even for the same user, intervals between sessions may vary. Results reported on two sets of Web logs show that a time range of 10–15 min was an optimal session interval threshold.

Some other approaches dedicated for log file analysis purpose detect session boundaries by grouping the data provided by the user based on IP address, cookies and also session time interval [22]. A transaction identification method called reference length for clustering search sessions was proposed in [8]. This method assumes that the time spent by the user on a page is correlated with whether the page is an "auxiliary" or "content" page for that user. Once pages are classified as either auxiliary or content pages, a session boundary is detected whenever a content page is met. The lack in this model is that users may obviously look at more than one content page for a single retrieval purpose.

Another session identification method, referred to as maximal forward reference, was proposed in [7]. In this approach, each session is defined as the set of pages from the first page in a request sequence to the final page before a backward reference is made. Here, a backward reference is defined to be a page that has already occurred in the current session. This method treats sessions as sets of visited pages within a time period and don't consider the sequence of the click-stream visitation. The clustering approach discussed in [52] was based on the sequence alignment method. They took the order of page accesses within the session into consideration when computing the similarities between sessions. Another range of session identification approaches are based on statistical language modeling that presents the advantage of learning a language model without well-labeled training data [20].

Few studies have addressed the issue of session boundary recognition in a personalized retrieval task. The UCAIR system [47] defines a session boundary detection based on a semantic similarity measure between successive queries using mutual information. Mutual information is defined by the number of documents indexed by terms issued from both queries. In our approach, we define a session identification method based on conceptual correlation measure that seems to be much more appropriate by introducing the context of queries submitted in the search session.

2.3 Search personalization

Search personalization aims at tailoring the search results to a particular user using the user profile. It may be performed by means of query refinement [42,44], query-document matching [31,49], personalized result categorization [33] or result re-ranking [15,23,32,43].

Query refinement consists in [42] of adding terms to the query from the user context defined by the user search history using Rocchio [38] algorithm. In [44], terms added to the

query are derived from concept hierarchies where the selected and deselected concepts are disambiguated relatively to the user profile.

Few works incorporate the user profile in the query-document matching model. Personalized Bayesien retrieval model [31,49] consists of query evaluation based on computing the document score by considering its relevance to the user query and also the correspondence relevance with regard to the user's topic of interest represented by a keyword vector.

Personalization based on result re-ranking usually consists on combining the content-based document score with the contextual document score. This latter is computed in [15,43] using the *cosine* similarity measure between each returned document and the most similar concept of the user profile. Personalized result-reranking of the PageRank is already described by [23] focusing on user profiles. The user selects his preferred pages from a set of hub pages and one personalized PageRank vector is computed for each user used to redirect the returned web pages to the preferred ones.

A variant technique of personalized document ranking is the personalized categorization of search results. It consists of grouping the search results into categories that describes the user interests [33]. A hybrid technique in [32] consists of combining personalized categorization and result re-ranking using a voting-based merging scheme. Mapping the query into a set of related categories allows categorizing the search results into multiple lists of retrieved documents. The new rank of a result is computed based on its rank in the list, the rank of the category associated to the list and the similarity of the category with respect to the query.

3 A session-based personalized search using a graph-based user profile: terminology and general approach

Our general approach for search personalization relies on building a short term user profile in a particular search session. The user profile is represented as a graph of semantically related concepts issued from the ODP ontology. We summarize below the terminology and notations used in our contribution, then we detail our approach.

3.1 Terminology and notations

– *Search activity*
  A search activity is defined by a set of events such as user querying, information retrieval (IR) and user feedback made by the user and the system to complete a searching task. Each search activity expresses the following events: the user u submits a query to a search engine; the latter returns a ranked list of documents, and then the user expresses his preferences on the document of interests. Thus, the main elements defining a search activity are a single query q^s submitted by the user at time s, the list of documents D^s returned by the system with respect to the query and the subset D_r^s of documents in D^s judged implicitly by the user as relevant. We assume that a document retrieved by the search engine is judged as relevant by the user if it generates some observable user behaviors (page dwell time, click through, saving, printing, etc).
– *Search session*
  A search session is defined as a sequence of search activities related to the same user information need. Usually, users submit queries that could be grouped into search sessions using a session boundary recognition mechanism. Formally, we define a search session S at time s as a sequence of related search activities performed by queries $\{q^0, \ldots, q^{s-1}, q^s\}$ submitted, respectively, at time $\{0, \ldots, s-1, s\}$.

–   *Query profile*
    The query profile refers the user's concepts of interest for a particular submitted query.
    It is represented at time s as a graph G_q^s of interrelated concepts issued from the ODP
    ontology and built using the user's documents of interest D_r^s returned with respect to
    query q^s.
–   *User profile*
    The user profile refers to the user's concepts of interest over an overall search session. It
    is also represented as a set of semantically related concepts issued from the ontology and
    initialized by the query profile G_q^0 of the first query q^0 submitted in the search session.
    Then it is updated by merging the query profiles built across the queries of the same search
    session. At time s, the user profile is represented as a graph G_u^s obtained by merging the
    previous one G_u^{s-1} and the query profile of the last submitted query G_q^s.

3.2 Graph representation

Query and user profiles are represented as graphs of weighted and interrelated concepts issued
from the ODP ontology. The graph structure $G = (V, E)$ has a hierarchical (tree) compo-
nent composed of "is-a" links, and a non hierarchical component composed of cross links of
different types predefined in the ODP ontology, where:

– V is a set of weighted nodes, representing the user's concepts of interest,
– E is a set of edges between nodes in V, partitioned into three subsets T, S and R, such
    that:

    · T corresponds to the hierarchical component of the user profile made of "is-a" links,
    · S corresponds to the non-hierarchical component made of "symbolic" cross links,
    · R corresponds to the non-hierarchical component made of "related" cross links.

Figure 1 illustrates an example of a user/query profile inferred from the ODP ontology and cor-
responding to the *computer language programming* interest. In this example, the user/query
profile G is defined by the following sets:
$V = \{(c_1, score(c_1)), (c_2, score(c_2)), \ldots, (c_8, score(c_8))\}$,
$S = \{(c_5, c_4), (c_5, c_8), (c_5, c_6)\}$,
$T = \{(c_1, c_2), (c_1, c_3), (c_2, c_4), (c_2, c_5), (c_3, c_6), (c_3, c_7), (c_4, c_8)\}$,
$R = \{(c_5, c_3)\}$.

Fig. 1 A portion of an ontological user/query profile in a graph-based representation
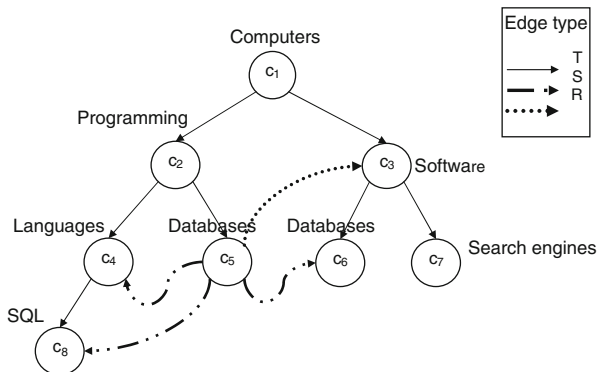
3.3 General approach

Our approach attempts to build a short term user profile by aggregating graph-based query profiles of the same search session. In particular, we assume that the user interest built in a search session contains the most suitable information that could achieve an effective personalized search. Search personalization is achieved by re-ranking the search results returned with respect to a query using the short term user profile. The overall process supporting our approach for search personalization is detailed in Algorithm 1. We build a graph-based ontological query profile G_q^s for query q^s submitted at time s. The user profile G_u^0 is initialized by the graph-based ontological profile G_q^0 of the first query submitted in the search session.

Algorithm 1 General approach for personalizing search across sessions using a graph-based user profile

---

for each new submitted query q^{s+1} **do**
  Compute conceptual correlation value: $\Delta I = (q^{s+1} \circ G_u^s)$
  if $\Delta I \geq \sigma$ **then**
    No session boundary is detected :
    * Re-rank the search results of the query q^{s+1} using the user profile G_u^s
    * build the graph-based ontological query profile G_q^{s+1}
    * update the user profile: $G_u^{s+1} = G_u^s \cup G_q^{s+1}$
  else
    There is a session boundary recognition and new session is detected:
    case 1: build a new user interest
    * build the graph-based ontological query profile G_q^{s+1}
    * reinitialize the user profile : $G_u^{s+1} = G_q^{s+1}$
    case 2: refine a prior discovered user interest
  end if
end for

The algorithm considers in turn each new submitted query q^{s+1} in a session boundary recognition mechanism. We propose a session identification method using the *Kendall* rank correlation measure that quantifies the conceptual correlation ΔI between the user profile G_u^s and a new submitted query q^{s+1}. We choose a threshold σ and believe that queries are from the same session if the correlation is above the threshold.

When the computed correlation is above the threshold, we consider that there is no session boundary and the user profile G_u^s is used to re-rank the search results of the new query q^{s+1}. After the user clicks or view interesting documents, the query profile G_q^{s+1} is built and possibly contains common or semantically related concepts with the user profile G_u^s. In this case, the user profile is updated by merging it with the graph-based query profile. The updated user profile contains new concepts issued from the query profile and possible new edges linking concepts derived from the user profile and concepts derived from the query profile. As an example, let's consider a user searching about computer programming languages. He submits a query about the "SQL" language and clicks on documents in which he is interested. The user profile is built and contains general concepts of computer language programming and specific ones such as the "SQL" concept of the ontology (see Fig. 1) ranked in the top of its representation. The user submits a new query concerning "DATABASES" as topic of interest to the system; the query is matched with common and general concepts of the ontology existing in the user profile. In addition, new specific concepts concerning "DATABASES" are ranked in the top of the query representation. The topical correlation

between the user profile and the new query is based on associating the top ranked concepts of their concept-based representation using the *Kendall* rank correlation measure. Hence, as common general concepts are ranked in the top of both the query and the user profile representation, the correlation value must be high and allows assuming that there is no topic change.

If there is a session boundary, we start a new search session in which the user profile is re-initialized by the graph-based profile of the new submitted query.

With respect to this general view, we address in the remainder of this paper the following research questions:

– How to build and maintain the graph-based user profile in a specific search session?
– How to detect related search activities and possible session boundaries?
– How to personalize the search accuracy using the short term user profile?

4 A graph-based ontological user profile

The overall process of generating the user profile is detailed in three main steps: (1) building the ontological query profile over a search activity, (2) detecting a possible session boundary when a new query is submitted, (3) maintaining the user profile when there is no session boundary encountered. We notice that even if there is no topic change held by the new query, the user profile is updated by means of adding concepts/edges from the query profile or updating the weights of common concepts.

4.1 Building the ontological query profile over a search activity

Building the ontological query profile is based on two main steps:

– initializing the query context as a set of keywords extracted from the user's documents of interests in a search activity, and then mapping it on the reference ontology to build an initial weighted concept set,
– inferring the graph-based query profile using a score propagation strategy applied on the initial weighted concept set.

First, we present our current implementation of the reference ontology. Then we present how to create the graph-based query profile.

4.1.1 Representation of the reference ontology

We use the ODP ontology as fundamental predefined domain knowledge in our user profiling component. The ODP is the most widely distributed Web directory that classifies millions of web pages into 787,774 categories.[2] Each concept is related to a set of sub-concepts through "is-a" links except the leaf nodes. Concepts of the ODP are interrelated with different relationship types. An important distinction between taxonomies and ontology such as the ODP ontology is that edges in a taxonomy are all of the same type ("is-a" links). While in the ODP ontology, edges can have additional semantic types (e.g., "symbolic", "related"), called cross links.

All of the directory's data are available free to the public[3] in RDF, a common format for describing web data. These are organized in two RDF files: the first one is "structure.rdf" that

---

contains all type relations (symbolic and related) linking the concepts in the ontology. The second one is "content.rdf" that lists the web pages classified under each concept of the ontology. Symbolic links are used for multiclassification in ODP. This multiclassification link, like all multiclassification links, brings the user across the directory into a different category. It prevents the user being at a page in ODP from having to traverse back (or up) some steps to the root of the directory and drills down from there to the topic the user search for. Another link types are cross-references such as those annotated with "see also:" and frequently used in ODP. Such links are not considered as symbolic links in the ODP RDF markup where they are represented with the "related" tag. The label of a "see also:" cross-reference link informs the user of the link target a priori, i.e., before the user commits to following it.

In order to use the ODP in the user profile representation, we represent the concepts by keyword vectors. For this, we used only the hierarchical relationship of the ODP and kept the transversal links between concepts in an aggregate form in order to use them further to activate semantically related concepts for representing the user profile. Geographic concepts classified under "Top/local" and all non English concepts and web pages are excluded from the ODP representation. Each concept c_j is associated to a set of web pages classified under that concept and each of the web pages is annotated by a title and a description that explain its content. For each concept, we only use the titles and descriptions of the first 60 web pages listed in the online available RDF metadata file "content.rdf" of the ODP in order to represent it by vector of single terms $\vec{c}_j$. Using this amount of data is sufficient to get an accurate text classification into the ontology. Several studies in [35] and [46] use similar concatenation to build the concept representation. In support of our approach, study in [40] proves that using the manually annotated titles and descriptions of the web page in the concept description vector achieves higher text classification accuracy than the use of the web pages contents. The procedure for getting the representation of the ODP concepts is described as follows:

1. concatenating the first 60 web pages classified under each concept in a super-document Sd_j in order to obtain a collection of super-documents, one per concept,
2. removing stop words and applying porter stemming on the super-documents,
3. computing the frequency tf_{ij} of each term t_i in super-document Sd_j,
4. applying score and term propagation from super-documents of sub-concepts to the super-documents of concepts. Given a general concept c_j having n related sub-concepts c_k, we recomputed the new term frequency ttf_{ij} of term t_i in super-document Sd_j using the following weighting scheme:

$$ttf_{ij} = \left[\left(tf_{ij} + \sum_{k=1,\ldots,n} tf_{ik} \right) \right] \Big/ (n+1) \tag{1}$$

5. representing each super-document Sd_j, by a single term-based vector $\vec{c}_j$ where the weight w_{ij} of term t_i in $\vec{c}_j$ is computed using the following weighting scheme:

$$w_{ij} = ttf_{ij} \times \log \left(\frac{N}{N_i} \right) \tag{2}$$

where N is the total number of super-documents, and N_i is the number of super-documents containing term t_i.

4.1.2 Query context initialization

We assume that the user profile could be inferred across related search activities using the user's documents of interest. Our goal here is to create the query context that holds the

user interest as the most relevant terms occurring in the relevant documents judged by the user. Creating the query context starts by collecting a set of relevant documents D_r^s returned with respect to query q^s submitted at time s. Each document in D_r^s is represented as a single term vector where the weight w_{td} of term t in document d is computed as follows: $w_{td} = tf_d \times log(N/n_t)$. tf_d is the frequency of term t in document d, N is the total number of documents in the test document collection and n_t is the number of documents containing term t. We create the query context K^s as being the centroid of the documents in D_r^s. The weight of term t in K^s is computed as follows:

$$K^s(t) = \frac{1}{|D_r^s|} \sum_{d \in D_r^s} w_{td} \tag{3}$$

In order to represent the user interests semantically as a set of concepts derived from the ontology, we map K^s on the ODP ontology using the *cosine* similarity measure. Given a concept c_j of the ODP, represented by term vector $\vec{c}_j$, its similarity score with $\vec{K}^s$ is computed as follows:

$$score(c_j) = cos(\vec{c}_j, \vec{K}^s) \tag{4}$$

The result of mapping the query context on the ontology is an initial set containing the top-50 weighted concepts, called $\theta^s = \{(c_1, score(c_1), \ldots (c_i, score(c_i),)\}$. For experimental purpose, we used the top-50 concepts matched with the query context and assume that this number is sufficient to include concepts of interests to the user. Based on this set, we attempt to build a graph of semantically related concepts of ontology using a score propagation detailed in the next section.

4.1.3 Inferring the graph-based query profile using score propagation

We infer the graph-based representation of the query profile using one-hop score propagation applied on the concept set θ^s. Our intuition behind the graph representation of the query profile relies on representing the user interests with semantically related concepts located in different portions of the ontology. We describe in Algorithm 2 the process of inferring the graph-based query profile. We distinguish the role of different edges in activating linked concepts in the score propagation. Indeed, we re-use the edge weight setting adopted in [34] in our score propagation as follows: $w_{ij} = \alpha_S$ for $e_{ij} \in S \cup T$, $w_{ij} = \alpha_R$ for $e_{ij} \in R$, where e_{ij} is the edge linking concept i to concept j. We set $\alpha_S = 1$ because *symbolic* links seem to be treated as first-class link "is-a" in the ODP web interface, and we set $\alpha_R = 0.5$ because *related* links are treated differently on the ODP web interface, labeled as "see also" topics. We didn't consider "is-a" links in score propagation because we assume that concepts linked by "is-a" relations are activated in the initial concept set, as specific concepts have common terms with their general concepts and they are both matched with the query terms.

At a starting point of the algorithm, θ^s is the initial set of weighted concepts of the ODP ontology. The algorithm considers in turn each of the concepts c_i in θ^s as as a starting node likely to induce a graph G_i. In order to group initially weighted and non symmetric interrelated concepts in θ^s in the same graph G_i, we define $queue_i$ that allows processing such concepts in the same graph. Each concept c_i propagates its weight to all its linked concepts c_k (made of "is-a", "related" and "symbolic" edges) extracted in a list in ℓ_j according to the ontology. When the concept is activated by multiple concepts, its score is recomputed by accumulating its weight with the propagated one. Interrelated concepts are grouped together

Algorithm 2 Building a graph-based ontological query profile using a score propagation strategy

---

input: θ^s an initial set of activated concepts
output: $G_q^s = (V s_q, E s_q)$ an ontological query profile as a concept-based graph
$\theta^s = \{c_1, c_2, .., c_n\}, ListGraphs = \emptyset$
for each concept $c_i \in \theta^s$ **do**
  $Queue_i = \{c_i\}$
  //initialization of the graph induced by concept c_i
  $G_i = (V_i, E_i), V_i = \{c_i\}, E_i = \emptyset, w(G_i) = score(c_i)$
  while $Queue_i.HasElement()$ **do**
    $c_j = Queue_i.PopElement()$
    //extract all linked concepts (*is-a, symbolic, related*)
    $\ell_j = GetLinkedConcepts(c_j)$
    for each concept $c_k \in \ell_j$ **do**
      if $e_{jk} \in S$ the edge type is *symbolic* **then**
        $\alpha = \alpha_S$
      else if $e_{jk} \in R$ the edge type is *related* **then**
        $\alpha = \alpha_R$
      end if
      //score propagation pattern for all linked concepts
      $score(c_k) = (\alpha * score(c_j) + score(c_k))/(\alpha + 1)$
      $V_i = V_i \cup c_k, E_i = E_i \cup e_{jk}, w(G_i) = w(G_i) + score(c_k)$
      if $c_k \in \theta^s$ **then**
        $\theta^s = \theta^s - \{c_k\}$
        $Queue.PushElement(c_k)$
      end if
    end for
  end while
  $ListGraphs = ListGraphs \cup \{G_i\}$
end for
//if there are induced graphs G_m, G_n having some common concepts
for each $G_m, G_n \in ListGraphs$ **do**
  if $V_m \cap V_n \neq \emptyset$ **then**
    $E_m = E_m \cup E_n, V_m = V_m \cup V_n, w(G_m) = w(G_m) + w(G_n)$ // merge the graphs together
  end if
end for
$G_q^s = argmax_{G_i \in ListGraphs}(w(G_i));$

---

in order to create a single or disconnected weighted graphs G_i. The weight $w(G_i)$ of the graph G_i is computed by summing the scores of its concept nodes.

As the score is propagated at one hop from an initially weighted concept set θ^s, created graphs may have common concepts. So we proceed by combining these graphs together in a single one by merging the node and edge sets as well as their weights. Finally, the ontological query profile G_q^s is represented by the highly weighted graph among the created ones. The user profile is initialized by the query profile G_q^s if the query is the first one submitted in the search session S.

4.2 Detecting session boundary

Our goal in this section is to define a topical-dependant session boundary recognition method that allows determining whether a new query q^{s+1} submitted at time $s + 1$, shares the same user interest held by the user profile G_u^s of the current search session.

Most of existing session boundaries algorithms are based either on grouping session based on user IP address, cookies, time, term similarity, semantic-based similarity based on term

co-occurrence in a document collection, user feedback and clickthrough data. We assume that detecting session boundaries is highly dependant on the change of the concept ranks across the search sessions, for this, we gauge the correlation between the user profile and each new submitted query using a conceptual correlation measure that considers the concept ranks in the two rankings.

Here, we define dual representations of query q^{s+1}. 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 $\vec{q}_c^{s+1}$ representing the query projection onto the ontology. Our intuition behind using the concept-based representation of the query relies on comparing the concepts of interest held by the query with those held by the user profile in a session boundary recognition task.

To obtain the concept vector $\vec{q}_c^{s+1}$ of the query, we map $\vec{q}_t^{s+1}$ on the ontology using the *cosine* similarity measure. We obtain a concept set where each concept c_i is weighted as follows:

$$CW(q_t^{s+1}, c_i) = cos(\vec{q}_t^{s+1}, \vec{c}_i) \tag{5}$$

For experimental purpose, we propagate the scores of the top-50 weighted concepts of the query in the same manner as explained in Sect. 3.1.3. This allows activating much more concepts that are likely to be matched with those of the user profile. Then, we compute the concept vector $\vec{q}_c^{s+1} = \langle w_1, w_2, \ldots, w_i, \ldots \rangle$ using a context-sensitive weighting scheme by introducing the frequency of queries that are matched with recurrent concepts across the search session.

The main reason of using the query frequency in the search session is to bring to the top of the query representation, the concepts that are recurrent in the search session and disambiguating the set of concepts matched with the query terms. Indeed, an ambiguous query could be matched with relevant and irrelevant concepts of the ontology. The relevant concepts are possibly recurrent in the search session, so we proceed by ranking in the top of the query representation the concepts that appear in multiple user profiles built across the same search session. Particularly, at time s, the query frequency of concept c is computed as the number of the user profiles built at time 0, 1, up to s belonging to the same search session and containing the concept c. Pruning the non relevant concepts is achieved by bringing them to low ranks of the query representation.

Hence, the weight w_i of concept c_i in the concept vector $\vec{q}_c^{s+1}$ is calculated using the concept score $(CW(q^{s+1}, c_i))$ and the query frequency (QF) as follows:

$$w_i = CW(q_t^{s+1}, c_i) \times QF(c_i) \tag{6}$$

where the query frequency (QF) is formally defined as:

$$QF(c_i) = \frac{|\vec{q}|^S}{\langle |\vec{q}|^S, c_i \rangle} \tag{7}$$

$|\vec{q}|^S$ refers to the total number of related queries submitted in the search session S, $\langle |\vec{q}|^S, c_i \rangle$ refers to the number of queries that lead to include the concept c_i in the user profiles $\{G_u^0, \ldots, G_u^s\}$ in the search session S.

Using the concept-based representation of the query, the *Kendall* correlation measure computes a conceptual correlation degree ΔI between the query submitted at time $s + 1$ and the user profile performed at time s. We compared it to the *WebJaccard* and the *cosine* similarity measures in order to show the effect of the concept ranks relatively to the concept

count (in the WebJaccard measure) and to the concept weights (in the *cosine* measure) for detecting correct session boundaries and correct related queries.

4.2.1 Using the Kendall rank correlation measure

Our choice of this measure relies on tracking changes of user interests across search sessions by measuring the differences of the concept ranks between the concept vector of the query and the user profile. We define the conceptual correlation degree ΔI between query $\vec{q}_c^{s+1}$ and user profile $\vec{G}_u^s$ performed at time s as follows:

$$\Delta I = Kendall(\vec{q}_c^{s+1}, \vec{G}_u^s) = \frac{\sum_{c_i} \sum_{c_j} S_{c_i c_j}(\vec{q}_c^{s+1}) \times S_{c_i c_j}(\vec{G}_u^s)}{\sqrt{\sum_{c_i} \sum_{c_j} S_{c_i c_j}^2(\vec{q}_c^{s+1}) \times \sum_{c_i} \sum_{c_j} S_{c_i c_j}^2(\vec{G}_u^s)}} \quad (8)$$

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

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

Where, c_i and c_j are two concepts issued from both the query and the user profile, $\vec{q}_c^{s+1}(c_i)$ is the weight of the concept c_i in $\vec{q}_c^{s+1}$, and $\vec{G}_u^s(c_i)$ is the weight of the concept c_i in the user profile $\vec{G}_u^s$.

This formula allows measuring the degree of correspondence between two rankings and assessing the significance of this correspondence. Indeed, for each couple of concepts c_i and c_j, $S_{c_i c_j}(\vec{q}_c^{s+1})$ and $S_{c_i c_j}(\vec{G}_u^s)$ defines the ranking order of the concepts, respectively, in the query and the user profile representation. Having the same order means an agreement between the two ranking concerning this couple. If the agreement between the two rankings is perfect (i.e., the two rankings are the same) the coefficient has value 1. If the disagreement between the two rankings is perfect (i.e., one ranking is the reverse of the other) the coefficient has value -1. For all other arrangements the value lies between -1 and 1, and increasing values imply increasing agreement between the rankings.

4.2.2 Using the WebJaccard similarity measure

We modify the traditional *Jaccard* measure [17] for the purpose of measuring topical similarity degree ΔI based on the concept counts between the query q_c^{s+1} submitted at time $s+1$ and the user profile G_u^s performed at time s.

The original WebJaccard is used to detect session boundaries based on intersections between these sets. It measures the degree of common visited pages in successive search iterations to be compared [14]. As we have to measure the degree of correspondence between two sets of concepts representing the query and the user profile, we re-used the *WebJaccard* measure by replacing the number of common web pages by the number of common concepts. Hence, the topical correlation degree ΔI according to *WebJaccard* is computed as follows:

$$\Delta I = WebJaccard(q^{s+1}, G_u^s) = \frac{H(q^{s+1} \cap G_u^s)}{H(q^{s+1}) + H(G_u^s) - H(q^{s+1} \cap G_u^s)} \quad (9)$$

where $H(q_c^{s+1} \cap G_u^s)$ is the number of common concepts in both representations, $H(q_c^{s+1})$ is the number of the concepts in the concept vector q_c^{s+1} and $H(G_u^s)$ is the number of concepts

in the user profile G_u^s. The correlation value ΔI is in the range $[0, 1]$, where a value equal to 0 means that the query and the user profile are not similar, and a value closer to 1 means that the query and the user profile have all the concepts in common and are very related to each other.

4.2.3 Using the cosine similarity measure

We used the *cosine* similarity [17] for the purpose of measuring the topical similarity degree ΔI between the query q^{s+1} submitted at time $s + 1$ and the user profile G_u^s performed at time s. This measure takes into account the concept weights between two vectors. Here, the query and the user profile are represented, respectively, by the concept vectors $\vec{q_c^{s+1}}$ and $\vec{G_u^s}$. A concept that appears in only one vector has a weight of zero in the other vector. The topical correlation degree ΔI according to *Cosine* measure is computed as follows:

$$\Delta I = Cosine(\vec{q_c^{s+1}}, \vec{G_u^s}) = \frac{\vec{q_c^{s+1}}.\vec{G_u^s}}{\left\| q_c^{s+1} \right\| . \left\| \vec{G_u^s} \right\|} \tag{10}$$

The correlation value ΔI is in the range $[0, 1]$, where a value equal to 0 means that the query and the user profile are not similar, and a value closer to 1 means that the query and the user profile have all the concepts in common and are very related to each other.

We expect that the *Kendall* measure is the most appropriate measure to detect more precisely session boundaries relatively to the *Cosine* or the *WebJaccard* measure. Indeed, we assume that when the user changes his interests across the search sessions, this is formally defined by the change of the rank of concepts in the user profile representation across the search sessions. The *cosine* and the *WebJaccard* measures take into account, respectively, the concept weights and the concept count between two sets of weighted concepts and ignore the change of the ranking between the two sets. This means that whatever is the change in the rank of the common concepts between the query and the user profile, the correlation value according to the *cosine* or the *WebJaccard* is the same. As an example, when the ranks of the recurrent concepts (ranked in the top of the user profile representation) is low in the query representation, the correlation degree with the user profile must be less than when they are ranked high in the query representation.

4.3 Maintaining the ontological user profile over a search session

We maintain the user profile in the same search session when a new submitted query q^{s+1} is correlated to the current user profile G_u^s according to the session boundary recognition mechanism. We use a graph-based merging scheme that produces the user profile $G_u^{s+1} = (V_u^{s+1}, E_u^{s+1})$ at time $s + 1$ by merging the previous one G_u^s with the query profile G_q^{s+1}. We assume that a query related to the search session, performs an ontological query profile that reveals common concepts with the current user profile. We present in Fig. 2 an example that shows two related graph-based profiles for successive queries searching about computer programming languages. The figure reuse the example presented in Sect. 3.3, where updating the user profile aims at adding new concepts that appear in the query profile and update the weights of possible common concepts between the query and the user profile.

Let $G_u^s = (V_u^s, E_u^s)$ and $G_q^{s+1} = (V_q^{s+1}, E_q^{s+1})$ be, respectively, the user profile and the ontological query profile performed, respectively, at time s and $s+1$. User profile maintaining process is based on the following principles:
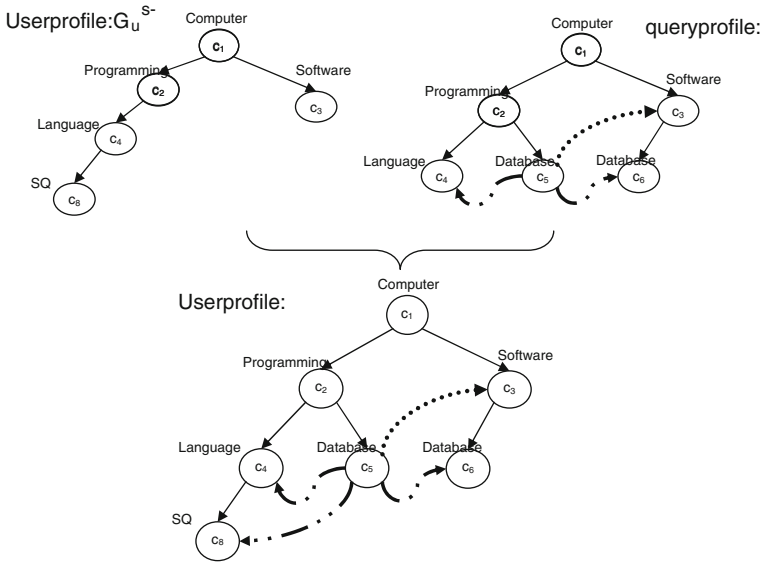
Fig. 2 Maintaining the user profile across related search activities

- accumulating the weights of possible common concepts c_i that can appear in query and user profile as follows:

$$\vec{G_u^{s+1}}(c_i) = \vec{G_u^s}(c_i) + \vec{G_q^{s+1}}(c_i)$$

where $\vec{G_u^s}(c_i)$ is the weight of concept c_i in user profile G_u^s, $\vec{G_q^{s+1}}(c_i)$ is the weight of concept c_i in query profile G_q^{s+1}. Accumulating the weights of common concepts allows bringing them to the top in the weighted concept vector of the user profile.
- merging the graph-based user profile as follows:

$$V_u^{s+1} = V_u^s \cup V_q^{s+1}, \quad E_u^{s+1} = E_u^s \cup E_q^{s+1}$$

This allows taking into account all possible concepts in which the user has shown interest in the search session.

5 Search personalization

Our major hypothesis is that using the user profile built across related search activities in the same search session can help improve search accuracy of a related query. Let us consider the user profile $G_u^s = (V_u^s, E_u^s)$ performed at time s. Let q^{s+1} be a query assumed to be related to the user profile ($\Delta I > \sigma$ according to the session boundary recognition mechanism). The query is submitted to the search engine that returns an initial list of documents. The results are re-ranked by combining for each retrieved result d_k, the initial score S_i and a contextual score S_c as follows:

$$S_f(d_k) = \gamma \times S_i(q, d_k) + (1 - \gamma) \times S_c(d_k, G_u^s) \quad 0 < \gamma < 1 \tag{11}$$

The contextual score S_c is computed using the *cosine* similarity measure between the result d_k and the top h weighted concepts of the user profile G_u^s as follows:

$$S_c(d_k, G_u^s) = \frac{1}{h} . \sum_{j=1,...,h} score(c_j) \times cos(\vec{d_k}, \vec{c_j}) \qquad (12)$$

where c_j is a concept in the user profile, $score(c_j)$ is the weight of concept c_j in the user profile G_u^s.

6 Experimental evaluation

In this section, we empirically evaluate the performance of our proposed search personalization approach. Our experimental evaluation is based on contextual simulation that uses queries and its associated relevance judgments provided by the TREC evaluation track. The evaluation of IR systems in context is already treated in [48,55] and shows the merits of the laboratory-based evaluation though its lack of realism in daily-life IR tasks. Particularly, it is a controlled framework that allows repeatable experiments and comparable evaluation.

Earlier experiments were carried out using the TREC adhoc collection [9] and the HARD TREC 2003 [10]. Here we extend the evaluation on the HARD TREC 2003 and study the effect of different parameters involved in our approach.

Our experimental methodology is designed to evaluate three particular topics:

– The accuracy of our proposed session boundary recognition mechanism,
– The effectiveness of score propagation in the user profile representation,
– The effectiveness of our personalized search approach comparatively to typical search and to a personalized approach described in [15].

6.1 Experimental data set: the TREC HARD TRACK collection

As the queries of average web users tend to be short and ambiguous [45], we used query topics provided by TREC[4] 2003 HARD Track [2]. Indeed, such queries are difficult and we expect to improve them by means of personalized search. The HARD corpus is a combination of NewsWire text from the 1999 portion of the AQUAINT corpus and the US Government documents containing 1,033,442 documents. HARD topics follow the basic TREC style, but are more richly annotated with metadata that describe the searcher and the context of the query. The title of the topic is a short phrase and is used as a query to the retrieval system.

We excluded topics that achieve zero values in the standard mean average precision (MAP) as our personalized approach is based on re-ranking search results. We also excluded topics that have low number (less than 30) of relevant documents in the relevance assessment file (Qrels provided by TREC). Indeed, for each topic, we need to create the user profile using a sufficient subset of its associated relevant documents and test the personalized search on the rest of documents.

As our approach relies on building a user profile across related search activities in a given search session, this requires a session based evaluation scenario. For this aim, we generated our own query set (subtopics) using 30 of the HARD topics of TREC. We assume that the topics are unrelated and consider that each one can represent a single search session where its subtopics perform related search activities.

The adopted strategy for generating the *sub-topics* of a topic consists of:

---

[4] Text REtrieval Conference: http://trec.nist.gov.

Table 1 Example of generated subtopics

Topic HARD-77	Insect-borne illnesses
Subtopic 1	Encephalitis, lyme, state
Subtopic 2	Encephalitis, mosquito, spray
Subtopic 3	State, encephalitis, nile
Search terms given by TREC	Insects, lyme disease, ticks, West Nile virus, mosquitos

- extracting a document *profile set* that consists of the first *r* relevant documents listed in the relevance assessment file given by TREC and associated to the topic ($r = 30$ in our experiments).
- dividing the document *profile set* into equally-sized three *profile subsets*.
- creating the centroid vector of each profile subset using formula (3) by representing each document as an ordered term vector using $tf \times idf$ weighting scheme.
- building each *sub-topic* by selecting the top three terms of the centroid vector.

An example of generated subtopics is given in Table 1 where texts about diseases outside of the US are off-topic. Thus, we generated a total of 90 queries (subtopics), three queries (subtopics) are generated per topic in order to define a session length equal to three. We generate three subtopics per topic in order to use an average session length close to real web environment. According to several studies [18,22], more than 79% of the sessions were three or fewer queries according to the user IP address and cookies to define session boundaries. Moreover, this allows evaluating the personalized search using a user profile built across few queries.

For each *subtopic*, the associated relevant documents in the collection are those of its main topic. The relevant document profile sets of all the topics are excluded when evaluating the performance of our search experiments.

In order to validate the reliability of the automatically generated *subtopics*, we compute first the average percentage of relevant overlapping documents returned by the system at Top-1,000 with respect to the *subtopic* comparatively to their main topic. The results are
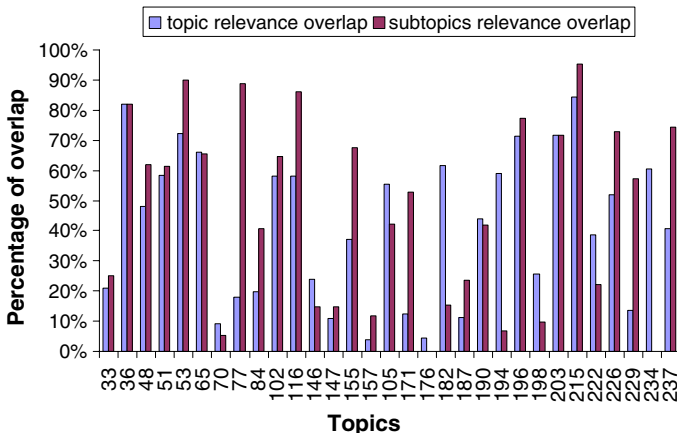

Fig. 3 Percentage of relevant document overlap of the *subtopics* comparatively to their main topic
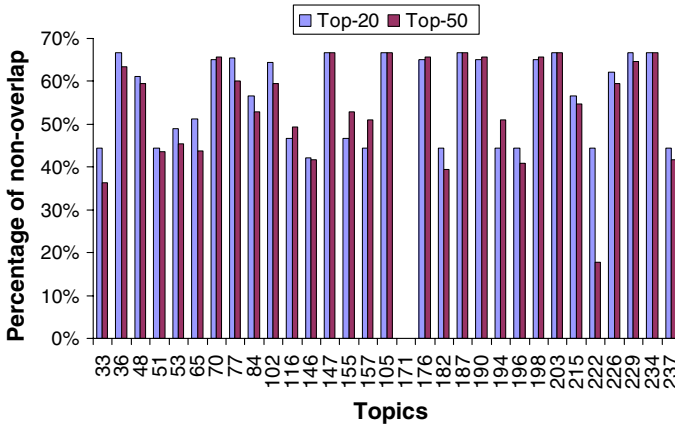
Fig. 4 Percentage of average non-overlapping documents over the Top-*n* between the *subtopics* of a main topic

shown in Fig. 3 where 30 topics of the HARD TRACK are presented on the *X*-axis and the percentages of overlap achieved by the topic compared to its associated *subtopics* are presented on the *Y*-axis. We notice that percentages of average overlap achieved by most of the *subtopics* are higher or equal to those achieved by the topic. This proves that most generated *subtopics* contain significant terms able to return relevant documents as the main topic.

Moreover, we computed the percentage of average non-overlapping documents over the Top-*n* documents (Top-20 and Top-50) between the *subtopics* themselves. Results are shown in Fig. 4 where 30 topics are presented on the *X*-axis and the percentage of average non-overlapping documents at Top-20 and Top-50 are presented on the *Y*-axis. We notice that values of non-overlapping are significant between the *subtopics* (higher than 40% at the Top-20). This proves that even though the *subtopics* were built relatively to the same topic, they do not return the same documents.

6.2 Experimental design and results

Our experimental methodology is designed to evaluate the accuracy of our session boundary recognition mechanism and the effectiveness of our personalized search approach. For this purpose, we divided the HARD topics into two topic sets. The first one is a *training topic set* used to train the system on the session boundary recognition mechanism in order to identify the optimal session boundary threshold value σ. The second one is a *test topic set* used for testing our personalized search approach making use of the optimal threshold value identified in the training step. In our experimental evaluation, we make use of the created *subtopics* of the HARD topics along *subtopic sequences*.

6.2.1 Evaluating the session boundary recognition mechanism

In order to evaluate our session boundary recognition mechanism, we defined a real evaluation scenario that consists of simulating search sessions by aligning successively subtopics of 15 main topics of the HARD track along a *training subtopic sequence*. We notice that we intentionally created the training query sequence in order to tune an average correlation value that allows getting the maximum precision of detecting correct session boundaries and

correct correlated queries. This value is then used on the test query sequence, for which evaluation performance was calculated.

We assume that related search activities are performed by the *subtopics* of the same topic and session boundaries should be identified between the main topics. We conduct experiments to achieve two goals: (A) analyzing the *subtopic correlation* degrees with the user profile using *Kendall*, *WebJaccard* and *cosine* measures along the *training subtopic sequence*, (B) measuring the session boundary recognition accuracy to identify the best threshold value.

(A) Analyzing subtopic-profile correlations

In this experiment, we computed the correlation degrees between a *subtopic* in the *training subtopic sequence* and the user profile built across previous processed *subtopics* of the same search session. They are computed according to the *Kendall* measure and compared to the *WebJaccard* and the *Cosine* measures.

As the subtopic ranks in a sequence might influence the identification of the best threshold value σ, we defined a critical *subtopic sequence* by aligning successively subtopics of the most correlated topics using the *Kendall* rank correlation measure. Such sequence is considered difficult as the optimal session boundary threshold is identified based on correlation values computed at boundaries, which could overlap with correlation values computed between related search activities. The reason of choosing such critical *training subtopic sequence* is to test our system using a threshold value issued from the most difficult *training subtopic sequence*. The *subtopic sequence* is aligned as follows:

– choosing randomly the first topic and align its subtopics.
– running an iterative process that builds the user profile for the last selected topic using its associated *subtopics*. More precisely, the graph-based user profile of each topic is built by merging the query profile of its *subtopics* as explained in Sect. (3.1). The query context is initialized by using its associated relevant document *profile subset* as explained in Sect. 3.1.2.
– For each of the remaining training topics, we compute the correlation degrees between each of its *subtopics* and the user profile of the last selected topic. The topic that has the maximum *subtopic-profile* correlation value is selected to be the next one by aligning its *subtopics* in the *training subtopic sequence*.

Once the critical *training subtopic sequence* is selected, we computed *subtopic-profile* correlations values where the user profile built across previous *subtopics* of the same topic along *training subtopic sequence*. We identify the range of correlations values according to each measure (*Kendall* and *WebJaccard* measures). Figure 5 shows, respectively, *Kendall*, *WebJaccard* and *cosine* correlation values computed across the *training subtopic sequence*. Correlation values are presented on the Y-axis and the *training subtopic sequence* is presented on the X-axis holding 15 topics of the HARD TRACK. *Subtopics* are assigned the number of the topic dotted by the number of the *subtopic* {1, 2, 3}.

A fall of the correlation curve at a particular *subtopic* means a decrease of its correlation degree with the user profile built across *subtopics* of the same main topic and indicates possible session boundary identification. A correct session boundary is marked by a vertical line at a particular *subtopic*.

We notice that when the correlation degree tends to be very low (large negative values in Kendall and null correlation values in *cosine* and *WebJaccard*), the probability of detecting a session boundary becomes high. In contrast, to this assumption, there are null correlation values computed across related *subtopics* of topics 105, 234, 182, 187 according to *WebJaccard* or *cosine* measures. This is due to the fact that subtopics related to the same topic may not have common concepts in the ODP ontology, which results null *cosine/WebJaccard*
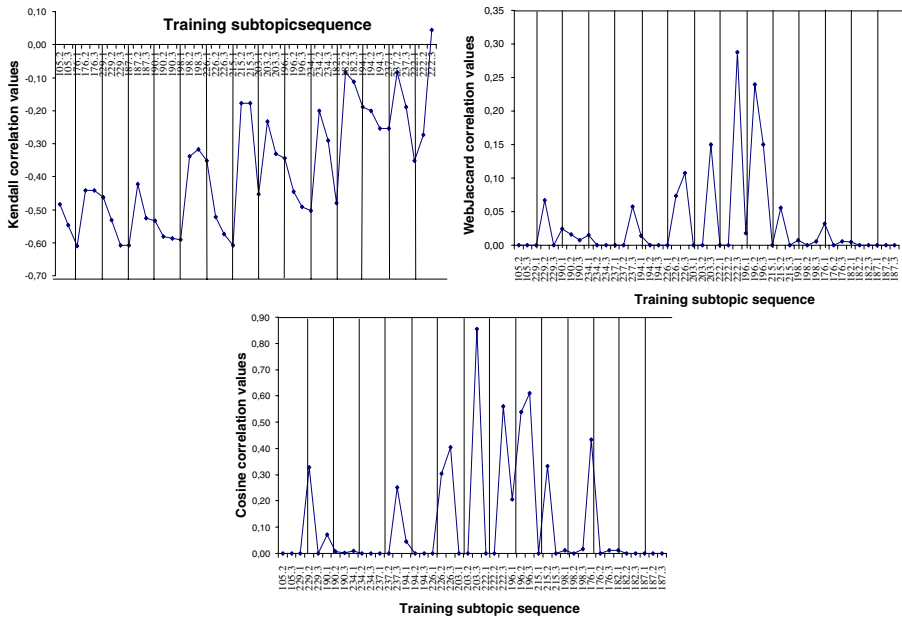
Fig. 5 *Kendall*, *WebJaccard* and *cosine* correlation values computed across the training *subtopic sequence*

correlation values. For the *Kendall* measure, most of correlation values are negative and large negative values are computed at boundaries. Indeed, the concepts matched with a new query are partially present in the user profile concept set. Concepts that exist in only one concept set (either the query or the user profile) are assigned null weights in the other concept set. Consequently, negative values inside a session do not mean that the query and the user profile are negatively correlated (which is considered by default in the *Kendall* measure) but correspond to a minimal correlation value between the query and the user profile due to the rank proximity of the common concepts (ranked usually in the top ranks) over the entire concepts. This allows computing a correlation value based on the changes of the concept ranks instead of the concept weights or the concept count and then considering possible semantic correlations between two sets of concepts.

Large negative values are gained at boundaries, because there are no common concepts between the two ranking and this do not reflect the fact that the concepts in the new query and the concepts of the user profile are in reverse order (considered by default in the *Kendall* measure) as they do not have originally the same concept set.

Based on the range of correlation values $[-0.6 \ +0.01]$, $[0.01 \ 0.3]$ and $[0.1 \ 0.9]$ related, respectively, to *Kendall*, *WebJaccard* and *Cosine* measures, we computed the session boundary accuracy in order to identify the optimal session boundary threshold for each measure as detailed in the next experiments.

(B) Measuring the session boundary recognition accuracy

The goal of this experiment is to evaluate the accuracy of the session boundary recognition mechanism on the *training subtopic sequence* using an optimal session boundary threshold value σ.

In order to measure the session identification accuracy, we define two measures: the first one is the precision $P_{intra}(\sigma)$ of allocating related search activities into the same search
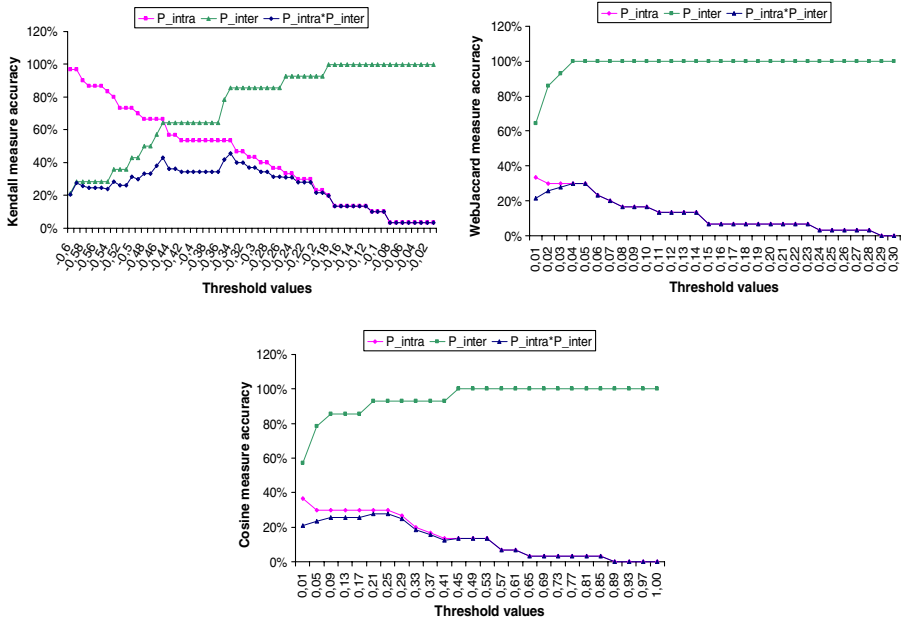
Fig. 6 *Kendall*, *WebJaccard* and *cosine* correlation accuracy as a result of varying the appropriate threshold values

session, the second one is the precision $P_{inter}(\sigma)$ of allocating unrelated search activities into different search sessions and so detecting session boundaries. These accuracy measures are defined as follows:

$$P_{intra}(\sigma) = \frac{|RQ|}{|TRQ|}, \quad P_{inter}(\sigma) = \frac{|BQ|}{|TBQ|} \tag{13}$$

Where $|RQ|$ is the number of *subtopics* identified as correctly correlated, and $|TRQ|$ is the total number of *subtopics* that should be identified as correlated in the *subtopic sequence*, $|BQ|$ is the number of *subtopics* indicating correct session boundaries according to the *subtopic sequence*, and $|TBQ|$ is the total number of session boundaries in the *subtopic sequence*. The optimal session boundary threshold value is identified when both measures ($P_{intra}(\sigma)$ and $P_{inter}(\sigma)$) reach the highest accuracy. Indeed, the threshold value is identified for the maximum value of the product of both measures as follows:

$$\sigma* = argmax_\sigma (P_{intra}(\sigma) \times P_{inter}(\sigma)) \tag{14}$$

We show in Fig. 6 the accuracy of the *Kendall* (resp. *WebJaccard* and *Cosine*) correlation measure in terms of $P_{intra}(\sigma)$ and $P_{inter}(\sigma)$ and the product of these precisions ($P_{intra}(\sigma)) \times P_{inter}(\sigma)$) with varying the threshold value in the range of the associated correlation values as mentioned above $[-0.6 \ + 0.01]$ (resp. $[0.01 \ 0.3]$ and $[0.1 \ 0.9]$). In our *training subtopic sequence*, there are 14 session boundaries (TBQ = 14) and 30 *subtopics* (TRQ = 30) where two *subtopics* must be encountered as correctly correlated per topic.

The accuracy evaluation performed using *Kendall* measure (resp. *WebJaccard* and *Cosine*) reveals that the threshold value -0.34 (resp. 0.04 and 0.21) achieves the optimal accuracy of the precision product 45.71% (resp. 30 and 27.85%) maximizing both the precision measures P_{intra} at 53.33% (resp. 30 and 30%) and P_{inter} at 85.71% (resp. 100 and 92.85%).

We notice here that the optimal threshold value is potentially dependant of the correlation degrees between the main topics in the training subtopic sequence given by TREC. In our experiments, we consider that topics are unrelated, while it possibly exists some topical relatedness between them. This does not affect the test performance because we use the optimal correlation threshold on the test subtopic sequence.

According to the results, the *Kendall* measure gives better performance than the *WebJaccard* and the *Cosine* measure. The *WebJaccard* measure is based on the number of common concepts between two rankings and ignores their ranks. This means that whatever is the rank of the common concepts between the query and the user profile, the correlation based on WebJaccard has the same value, which does not reflects the fact that the user changes his interests. The *cosine* measure takes into account the concept weights between two rankings and gives different *cosine* values for two different weighted vectors having the same ranking. This case could happen when the user has shown lower interest in a concept comparatively to other concepts by keeping the same ranking of the concepts. Hence, the computed score are biased and could reflect a change of the user interest without any consideration of the concept ranking.

Using the *Kendall* measure, it considers the concept order to track a change in the user interest. In effect, it computes higher correlation value when the concepts of the user profile are ranked in the top of the query representation relatively to the case when they are ranked low with the same order. Consequently, we use the identified optimal threshold value ($\sigma = -0.34$) of the *Kendall* measure for testing our search personalization presented in the next section.

6.2.2 Evaluating search personalization

Our experimental design for evaluating search personalization concerns two main objectives. The first one is to evaluate the effect of score propagation on the personalized retrieval performance. The second one is to test our personalized approach comparatively to the standard search.

Experiments were conducted on the *test subtopic sequence* aligned using the *test topic set*. With respect to the numbers of topics done by HARD TREC considered to be randomly, we aligned the associated *subtopics* randomly in the *test subtopic sequence*. Along this *test subtopic sequence*, our experimental methodology could be described as follows:

1. Creating the ontological query profile for each *subtopic* using its appropriate relevant document *profile subset* in the query context initialization step [formula (3)].
2. Initializing the user profile by the ontological profile of the first query of the session.
3. Applying the session boundary recognition mechanism using an appropriate threshold value (σ^*) when a new *subtopic* has to be processed. If the *subtopic* is correlated to the current user profile, this latter is used to re-rank the results returned by the system with respect to the *subtopic*. The user profile is then updated using the profile of the new subtopic.

In our search experiments, we compared the standard retrieval performance using only the query (ignoring any user profile) comparatively to the personalized search performed using the user profile built for queries allocated in the same search session. Our baseline search is based on the BM25 relevance scoring formula given as follows:

$$w_{td} = tf_d \times \frac{log\left(\frac{N-n_t+0.5}{n+0.5}\right)}{K_1 \times \left((1-b) + b \times \frac{dl}{avgdl}\right) + tf} \tag{15}$$

where tf_d is the frequency of term t in document d, N is the total number of documents in the test document collection and n_t is the number of documents containing term t, $K_1 = 2$ and $b = 0.75$. We fix γ at the value 0.3 in the re-ranking formula (10). We have shown in earlier experiments [9] that $\gamma = 0.3$ is the best value among the range of values [0.1 0.9], achieving the highest retrieval performance. For experimental purpose, we used the top three concepts of the user profile [$h = 3$ in formula (11)] to compute the contextual score of the document. According to several studies in word sense disambiguation techniques [32,44], three concepts are sufficient to disambiguate the web search.

Evaluation measures are the Top-n recall and Top-n precision computed as follows:

$$Top-nRecall = \frac{RelDoc_n}{RelDoc_{total}}, \quad Top-nPrecision = \frac{RelDoc_n}{n} \tag{16}$$

where $RelDoc_n$ is the number of relevant documents that appear within the top n search results, $RelDoc_{total}$ is the total number of relevant documents for the given *subtopic*, by excluding all the relevant documents profile set of the topics.

(A) Evaluating the effect of score propagation on the retrieval effectiveness

The goal of this experiment is to study the effect of adding semantically related concepts of ontology on the retrieval effectiveness. For this purpose, we measured the effect of score propagation by comparing the personalized retrieval effectiveness on the *test subtopic set* using the user profile built within applying the score propagation (*propagated profile*) comparatively to the one built using the user profile without applying the score propagation (*non propagated profile*).

In order to set a comparative evaluation, we exclude two features that may influence on the result analysis. The first one is the best threshold value of the session boundary mechanism obtained from the training subtopic sequence. The second one is the selection of the most highly weighted graph among the created ones to describe the user profile (Sect. 4.1.3). Thus, the evaluation protocol consists of the following principles:

1. Representing the non propagated user profile by the set of concepts issued from the query context projection onto the ODP ontology. The propagated profile is represented by the same set of concepts enhanced with interrelated concepts using score propagation.
2. Identifying the best threshold value of the session boundary recognition mechanism on the *test subtopic sequence* according for each method (propagated profile and non propagated profile). This allows reliable comparison between the methods by setting the best conditions to each one when running experiments.

Following this evaluation scenario, we computed first the *subtopic-profile* correlations for each *propagated* and *non propagated profile* on the *test subtopic sequence* in the same manner of evaluating the session boundary mechanism. Based on the range of the correlation values, we identify the values -0.51 (resp. -0.55) as the optimal threshold value for the propagated profile (resp. the non propagated profile). The best value of threshold is then used in a search personalization task on the test subtopic sequence. We recall that using the best threshold value corresponding to each method and not the threshold issued from the training subtopic sequence allows a reliable comparison between methods by setting each one in the best evaluation conditions.

Figure 7 shows a comparison between personalized search using the *propagated profile* and personalized search using the *non propagated profile* on the *test subtopic* sequence. It can be shown that the high average precision computed at most of Top-n documents was achieved by the personalized search based on *propagated profile* except at Top-10 and Top-20 average precision. This can be explained by the fact that expanding the profile representation with
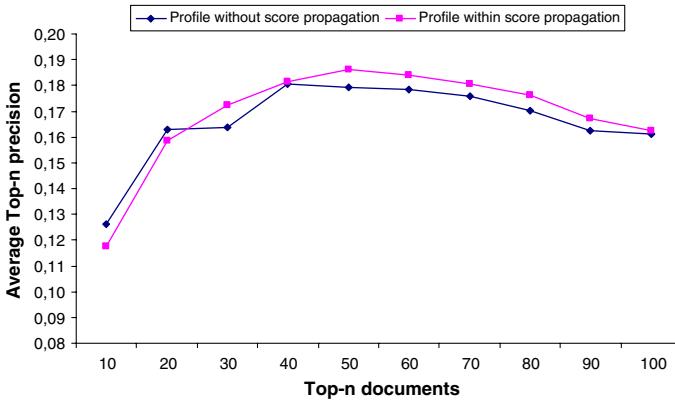
Fig. 7 Comparison of the average Top-*n* precision of the personalized search using the profile built without score propagation and within score propagation

new relevant concepts (propagated profile) could be weakly accurate to improve the retrieval effectiveness at Top-10 and Top-20 documents comparatively to the non propagated one. However, it achieves higher recall at most of the Top-*n* precision by bringing more relevant documents to the top of the returned ranked list.

(B) Evaluating personalized retrieval effectiveness on the test subtopic sequence

We conduct two sets of experiments to evaluate the personalized retrieval effectiveness. The first one consists of using the first 30 relevant documents from the relevance assessment file (Qrels) given by TREC to define the profile set of a topic. In the second experiment, the profile set of a topic is defined by the first 9 relevant documents ranked in the top 40 ranked documents returned by the system with respect to the topic. This allows setting an evaluation scenario close to real world search engines.

Experiment I:

In this experiment, we used the optimal threshold value ($\sigma^* = -0.34$) in the session boundary recognition mechanism based on the *Kendall* measure (identified in Sect. 6.2.1). According to the evaluation scenario based on computing subtopic-profile correlations along the test subtopic sequence, we obtained a precision of allocating related search activities in the same search session equal to 71.42% and a precision of detecting correct session boundaries on the *test subtopic sequence* equal to 40.47%.

Figure 8 shows the average Top-*n* precision and Top-*n* recall achieved by personalized search comparatively to the standard one on the *subtopic* sequence. In the evaluation task, we consider that a subtopic shares the same set of relevant documents as the main topic and we exclude the relevant document profile set of all topics from search experiments and evaluation task. We see that personalized search achieves higher retrieval precision and recall comparatively to the standard search. We have also computed the percentage of improvement in average Top-*n* precision and Top-*n* recall achieved by the personalized search as shown in Fig. 9. The re-ranking of documents using the concepts of high interest to the user in the same search session produced significant performance increase over all cutoff precision and recall. In particular, the biggest improvement occurred at Top-70 average precision (11.95%) and Top-10 average recall (23.62%).

Experiment II:

In this experiment, we exclude topics that do not return a minimum of 9 relevant documents in the top-40 returned documents. The profile set of each topic containing 9 documents
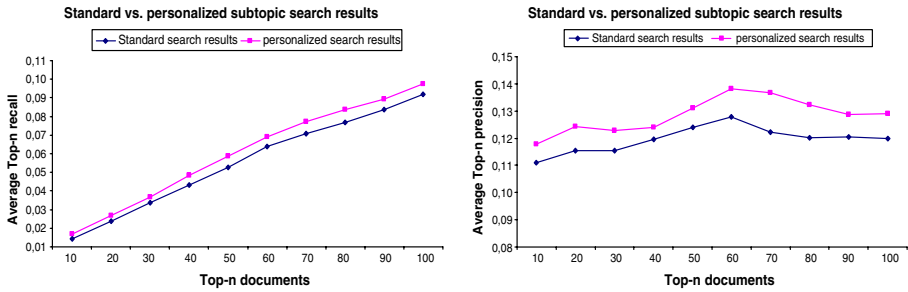
Fig. 8 Average Top-*n* recall and Top-*n* precision comparaison between the personalized search and the standard search on the subtopic sequence
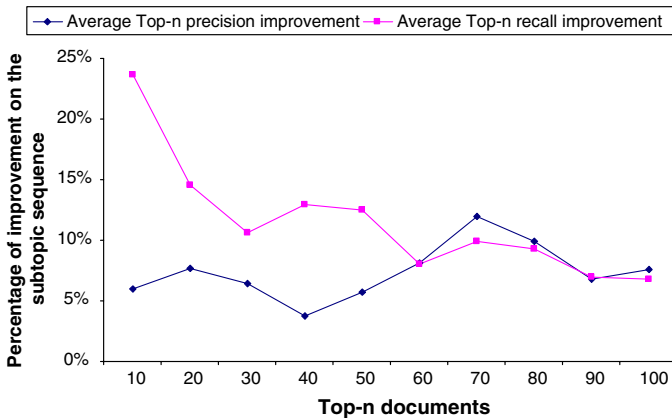

Fig. 9 Percentage of improvement in average Top-*n* recall and Top-*n* precision achieved by personalized search relative to the standard one on the subtopic sequence

is equally divided into three subsets used to create three subtopics. We adopt the same evaluation scenario presented for evaluating the personalized search effectiveness on the subtopic sequence. A total of 8 topics generated 24 subtopics where the training subtopic sequence of 12 subtopics allows identifying the optimal session boundary identification $\sigma^* = -0.41$).

Figure 10 shows the average Top-*n* precision and Top-*n* recall achieved by personalized search comparatively to the standard one on the *subtopic* sequence. Results prove that personalized search achieves higher retrieval precision and recall comparatively to the standard search. Especially, the best performance is achieved by the personalized search in terms of top-10 precision (0.1273) and top-10 recall (0.0057) comparatively to the standard search having lower top-10 precision (0.0182%) and lower top-10 recall (0.0010). We notice that the improvement is much higher where the user profile is built using the top ranked documents returned by the system with respect to the topic. This confirms that our method achieves an effective personalization in real world search engines where we use the top ranked documents returned with respect to the topic to create the user profile and then to personalize search.

6.3 Comparative study for evaluating search personalization on the test topics

In this experiment, we evaluated the retrieval effectiveness of our approach comparatively to the approach described in [15] on the *test topic set*. We compared the retrieval effectiveness
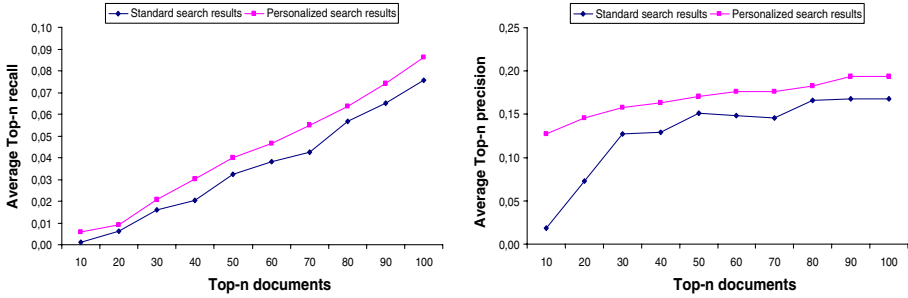
Fig. 10 Average Top-*n* recall and Top-*n* precision comparaison between the personalized search and the standard search on the subtopic sequence: profile built from the top ranked documents of the topic

achieved by the standard search using only the topic to the one achieved by personalized search using the topic and the corresponding user profile in each method. In this case, we consider that the topic is the last query submitted by the user in the search session, which allows evaluating the personalized search on the originally difficult/ambiguous queries given by TREC.

6.3.1 Overview of [15] approach

In [15], the user profile refers to the user interests identified in a specific searching time. It is represented using a reference ontology created on the basis of subject hierarchies and associated web pages from Yahoo, Lycos, Magellan and the ODP directory project.

Each concept of the ODP ontology is represented in the vector space model. A collection of super-documents were created based on concatenating the content documents of each concept. The super-documents were then pre-processed and stemmed. Then each concept c_j is represented by a term vector $\vec{c}_j$, where the weight of a given term is computed using $tf \times idf$ weighting scheme. Creating the user profile consists of the following strategy:

1. Collecting the web pages browsed or visited by the user from the user's web browser cache folder. These web pages were considered the main sources of evidence that could leverage the user interests and preferences.
2. Classifying the web pages into the appropriate concepts in the reference ontology. For each of the visited pages, a document vector was calculated using the $tf \times idf$ weighting scheme. The similarity between the document vector and the concept vector was calculated using the *cosine* similarity measure. As pages are classified with respect to the reference ontology, the similarity values computed between the concepts and the web pages are added to the top five concept's weights. Thus, the weight of a concept in the ontology represents its aggregated similarity scores with the pages classified into this concept.

Personalization consists then of re-ranking the search results of a given query using the user profile. For each document *r* returned in the result list, a new score of the document is computed based on the original match value returned by the search engine, the similarity between the document and its top concepts in the user profile, and the weight of these concepts in the user profile. The new score new_{wt_r} of a document r is computed as follows:

$$new_{wt_r} = wt_r \times \left(0.5 + \frac{1}{4} \sum u_{crl}\right) \tag{17}$$

where wt_r is the score of the result r returned by the search engine, u_{crl} is the weight of the concept c_{rl} in the user profile, c_{rl} is the lth most highly weighted concept for result r.

6.3.2 Experimental design and results

Our purpose is to compare the performance results obtained with two personalization appr-oaches: our search personalization approach and [15] approach. In order to achieve an accurate comparative evaluation, we set an evaluation scenario that consists of using the same infor-mation to build the user profile in both approaches. In our approach, the user profile related to a tested topic is built by merging the query profile of its *subtopics*. For the same tested topic, we consider the same relevant document profile set to create the user profile an instance of the reference ontology according to [15]. The user profile is then used to personalize the search results returned by the search engine with respect to the tested topic.

We present in Fig. 11 the average Top-*n* precision and Top-*n* recall achieved by personal-ized search according to both approaches comparatively to the baseline search. Results show that both personalized search approaches perform higher Top-*n* precision and Top-*n* recall than the standard search.

Moreover, we see that our approach gives higher performance than [15] model. The per-centage of improvement in average Top-*n* precision and Top-*n* recall in both approaches is presented in Fig. 12. Maximum improvements achieved by our approach are 66.7% at aver-age Top-10 precision and 188.24% at average Top-10 recall. While maximum improvements achieved by the second approach are 44% at average Top-30 precision and 144.4% at average Top-10 recall (resp. [15] model).
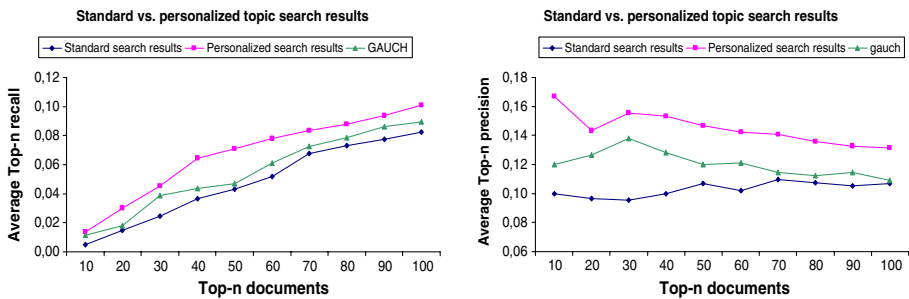

Fig. 11 Average Top-*n* recall and Top-*n* precision achieved by the standard search, personalized search using our approach, personalized search using Gauch et al. [15] approach on the HARD topics

Fig. 12 The percentage of improvement in average Top-*n* recall and Top-*n* precision achieved by the standard search, personalized search using our approach, personalized search using Gauch et al. [15] approach on the HARD topics
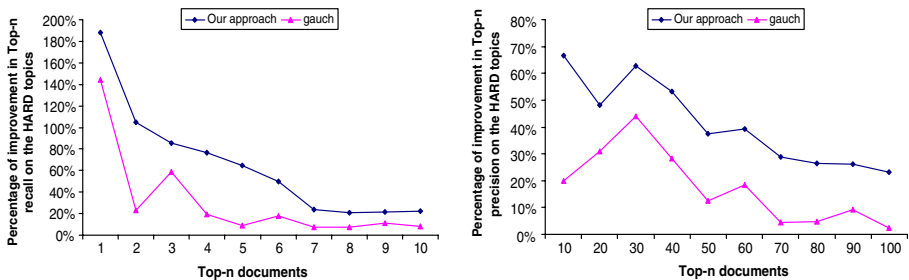
We notice that there are several factors that could influence the result performance between approaches. We cite the impact of the user profile representation and its accuracy on one hand and the re-ranking technique on the other hand. First, our user profile representation enhances the user's concepts of interests with semantically related ones using score propagation, which allows defining more accurately the user profile. Second, re-ranking search results consists in our approach of using both contextual score and original score of the document as well as the weight of the top ranked concepts in the user profile instead of using only the weights of the most similar concepts to the document as explained in [15]. Based on the overall evaluation results, the first conclusion we can made is that our approach has shown effectiveness in improving the search accuracy, especially for difficult queries that were selected from the HARD topics of TREC. Second, the comparative study of our approach with the Gauch et al. [15] model proves the effectiveness of features involved in our approach such as score propagation and the user profile representation as being the most highly weighted graph of the reference ontology.

7 Discussion

Our research work relies on personalizing search for related queries using the short term user profile. Our intuition was based on the assumption that the short term search history collected in a single session highlights on the user's current information need is more effective than the long term one in improving the search accuracy. For this purpose, our user profile modeling relies on building and maintaining an ontological user profile across related search activities in the same search session so as to hold short term user interests.

Following this general view, our approach could be distinguished by several features in the personalized web search community. The first one concerns the user profile representation, the second one concerns the user profile modeling technique and the third one concerns the session boundary recognition mechanism for mining the short term user profile in the search personalization task.

Regarding the user profile representation, it consists of an instance of the ODP ontology in previously listed works [15,43]. Unlike these works, our user profile is represented as a set of interrelated concepts of the ODP ontology organized in a graph-based representation. The main assumption behind this representation is that we aim at representing the user interest as a set of related concepts located in different portions of the reference ontology. Hence, it cannot be represented as an instance of the overall ontology that holds all possible user interests inferred across related and unrelated search activities. An example where our user profile representation can be validated is as follows. Consider a user interested in soccer video games. Involved user interest could be matched with interrelated concepts of different portions of the ontology such as concepts under different broader topics such as *Sports* and *Games*. Irrelevant concepts that could be matched with the involved user interest and have no relationship with the relevant ones must be excluded.

The second feature of our approach concerns the strategy adopted for representing the user profile. The graph-based representation pf the user profile is inferred using a score propagation through both hierarchical (*is-a* links) and cross links (*symbolic* and *related* links) of ontology. While in [43], only the hierarchical links of ontology are used for score propagation. The main intuition behind score propagation is to activate as much interrelated concepts as possible, even though they are not directly linked in the ontology. We outline that we activate semantically related concepts at only one hop from an initial weighted concept set, instead of propagating the concept scores over the entire ontology as adopted in [43]. This

allows describing the user profile by the most relevant interrelated concepts in a particular search session.

Furthermore, we maintain the user profile using a graph-based merging scheme by combining it with the related graph-based query profile in the same search session. This presents the advantages of enhancing the weights of the recurrent concepts with interrelated ones across related search activities and guarantee profile stability by bringing to the top the most relevant concepts during the search.

Because of detecting related search activities, our approach is characterized by an additional feature which is the session boundary recognition mechanism. Unlike previous proposed session boundaries recognition mechanisms, we proposed a mechanism based on tracking changes in the dominant concepts held by the query and the user profile. We used the *Kendall* rank correlation measure and proved that it is more accurate than the *WebJaccard* or the *Cosine* measures. In addition, we aim to set a session boundary recognition based on combining conceptual correlation and time-based session length.

Finally, our search personalization is based on re-ranking the search results using contextual score and original score. Contextual score is computed using a similarity measure between the result and the top ranked concepts of the user profile. Comparatively to other approaches [15,43], contextual score is computed using the most similar concept to the result. This could be considered as a time consuming process especially when the user profile is represented as an instance of reference ontology including huge number of weighted concepts.

8 Conclusion and outlook

We have presented in this paper a personalized search approach using a graph-based user profile issued from the ODP ontology. User searches are bounded using a session boundary recognition mechanism and the user profile refers to the user interest in a particular search session. Our experimental evaluation was carried out using the TREC 2003 HARD TRACK. We defined a session-based evaluation scenario by simulating search sessions along a *subtopic* sequence created using the main topics of the HARD TRACK. Experimental results showed that the session boundary recognition mechanism based on the *Kendall* measure achieves higher significant accuracy than the use of *WebJaccard* or *cosine* as correlation measures.

The evaluation of personalized search effectiveness addressed the impact of several aspects involved in our approach. First, results showed that the enhanced representation of the user profile using the score propagation is helpful to get an accurate user profile and then improve the personalized search effectiveness. Second, the comparative study between our personalized search approach and another one described in [15] proved that our approach achieves significant performance increase on the main HARD topics comparatively to [15] approach and also to the baseline search. This confirms the effectiveness of our approach, especially the accuracy of the graph-based representation of the user profile.

In future work, several aspects in user modeling and session boundary recognition mechanism merit to be evaluated in our approach. Regarding the user modeling, it is important to evaluate the accuracy of the user profile representation as it influences on the retrieval effectiveness. Concerning the session boundary recognition mechanism, we plan to improve the involved mechanism by combining it with time-based session identification using an adaptive threshold value. In this setting, evaluation of session boundary mechanism must be then carried out using real user data provided by web search engine log file.

Moreover, we plan to address an evaluation scenario that reveals the changes in the user interests to prior discovered ones across the search sessions. Indeed, the user profiles built across multiple search sessions are aggregated to define prior discovered user profile/interests.

Finally, we plan to carry out a user study, which allows evaluating our approach in real web environment search.

References

1. Alexandru CP, Wolfgang N, Raluca P, Christian K (2005) Using odp metadata to personalize search. In: SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval, pp 178–185
2. Allan J (2003) Hard track overview in trec 2003: high accuracy retrieval from documents. In: TREC, pp 24–37
3. Allan J et al (2003) Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent information retrieval, University of Massachusetts Amherst, September 2002. SIGIR Forum 37(1):31–47
4. Begg IM, Gnocato J, Moore WE (1993) A prototype intelligent user interface for real-time supervisory control systems. In: IUI '93: Proceedings of the 1st international conference on Intelligent user interfaces. ACM Press, New York, pp 211–214
5. Boughanem M, Chrisment C, Mothe J, Soul-Dupuy C, Tamine L (2000) Connectionist and genetic approaches to achieve IR. In: Crestani F, Gabriella P (eds) Soft computing in information retrieval techniques and applications. Springer, Berlin, pp 173–198
6. Boughanem M, Tamine L (2002) A study on genetic niching for query optimisation in document retrieval. In: Europeen colloquium on information retrieval, Glasgow, 25–27 March 2002, pp 135–149
7. Chen M-S, Park JS, Yu PS (1998) Efficient data mining for path traversal patterns. IEEE Trans Knowl Data Eng 10(2):209–221
8. Cooley R, Mobasher B, Srivastava J (1999) Data preparation for mining world wide web browsing patterns. Knowl Inf Syst 1:5–32
9. Daoud M, Tamine L, Boughanem M (2008) Learning user interests for session-based personalized search. In: ACM Information Interaction in context (IIiX). ACM Press, London, pp 57–64
10. Daoud M, Tamine L, Boughanem M, Chebaro B (2009) A session based personalized search using an ontological user profile. In: ACM symposium on applied computing (SAC), Haiwai (USA). ACM Press, London, pp 1031–1035
11. Ding C, Patra JC, Peng FC (2005) Personalized web search with self-organizing map. In: EEE '05: Proceedings of the 2005 IEEE international conference on e-Technology, e-Commerce and e-Service (EEE'05) on e-Technology, e-Commerce and e-Service. IEEE Computer Society, Washington, DC, pp 144–147
12. Dumais S, Cadiz ECJJ, Jancke G, Sarin R, Daniel CR (2003) Stuff i've seen: a system for personal information retrieval and re-use. In: SIGIR'03: Proceedings of the 26th annual international ACM SIGIR conference on research and development in informaion retrieval (SIGIR '03). ACM Press, London, pp 72–79
13. Feng Q, Junghoo C (2006) Automatic identification of user interest for personalized search. In: WWW '06: Proceedings of the 15th international conference on world wide web, pp 727–736
14. Foss A, Wang W, Zaane OR (2001) A non-parametric approach to web log analysis. In: Workshop on webmining in first international SIAM conference on data mining, pp 41–50
15. Gauch S, Chaffee J, Pretschner A (2003) Ontology-based personalized search and browsing. Web Intell Agent Syst 1(3–4):219–234
16. Gowan J (2003) A multiple model approach to personalised information access. Master thesis in computer science, Faculty of science, Universitt de College Dublin
17. Haveliwala TH, Gionis A, Klein D, Indyk P (2002) Evaluating strategies for similarity search on the web. In: WWW'02: Proceedings of the eleventh international world wide web conference, pp 432–442
18. He D (2000) Detecting session boundaries from web user logs. In: Proceedings of the BCS-IRSG 22nd annual colloquium on information retrieval research, pp 57–66
19. Huang X, Peng F, An A, Schuurmans D (2004) Dynamic web log session identification with statistical language models. J Am Soc Inf Sci Technol 55(14):1290–1303
20. Huang X, Yao Q, An A (2006) Applying language modeling to session identification from database trace logs. Knowl Inf Syst 10(4):473–504

21. Jaime T, T., DS, Eric H (2005) Personalizing search via automated analysis of interests and activities. In: SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on research and development in information retrieval, pp 449–456
22. Jansen BJ, Spink A, Kathuria V (2006) How to define searching sessions on web search engines. In: Advances in web mining and web usage analysis, 8th international workshop on knowledge discovery on the web, WebKDD 2006, Philadelphia, pp 92–109
23. Jeh G, Widom J (2003) Scaling personalized web search. In: WWW '03: Proceedings of the 12th international conference on world wide web. ACM Press, New York, pp 271–279
24. John RI, Mooney GJ (2001) Fuzzy user modeling for information retrieval on the world wide web. Knowl Inf Syst 3(1):81–95
25. Kelly D, Teevan J (2003) Implicit feedback for inferring user preference: a bibliography. SIGIR Forum 37(2):18–28
26. Kim HR, Chan PK (2003) Learning implicit user interest hierarchy for context in personalization. In: IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces. ACM Press, New York, pp 101–108
27. Koutrika G, Ioannidis Y (2005) A unified user profile framework for query disambiguation and personalization. In: Proceedings of workshop on new technologies for personalized information access
28. Leung CW, Chan SC, Chung F (2006) A collaborative filtering framework based on fuzzy association rules and multiple-level similarity. Knowl Inf Syst 10(3):357–381
29. Lieberman H (1995) Letizia: an agent that assists web browsing. In: IJCAI 95: Proceedings of international joint proceedings of the fourteenth international joint conference on artificial intelligence, pp 924–929
30. Lieberman H (1997) Autonomous interface agents. In: CHI, pp 67–74
31. Lin C, Xue G, Zeng H, YU Y (2005) Using probabilistic latent semantic analysis for personalised web search. In: Proceedings of the APWeb conference, pp 707–711
32. Liu F, Yu C, Meng W (2004) Personalized web search for improving retrieval effectiveness. IEEE Trans Knowl Data Eng 16(1):28–40
33. Ma Z, Pant G, Sheng ORL (2007) Interest-based personalized search. ACM Trans Inf Syst 25(1):5
34. Maguitman AG, Menczer F, Roinestad H, Vespignani A (2005) Algorithmic detection of semantic similarity. In: WWW '05: Proceedings of the 14th international conference on World Wide Web. ACM Press, New York, pp 107–116
35. Menczer F, Pant G, Srinivasan P (2004) Topical web crawlers: evaluating adaptive algorithms. ACM Trans Interet Technol 4(4):378–419
36. Micarelli A, Sciarrone F (2004) Anatomy and empirical evaluation of an adaptive web-based information filtering system. User Model User-Adapt Interact 14(2–3):159–200
37. Mobasher B (2007) Data mining for web personalization. In: Brusilovsky P, Kobsa A, Nejdl W (eds) Lecture notes in computer science. Springer, New York
38. Rocchio JJ (1971) Relevance feedback in information retrieval. In: Salton G (ed) The SMART retrieval system—experiments in automatic document processing. Prentice-Hall, Englewood Cliffs
39. Shahabi C, Chen Y-S (2003) Web information personalization: Challenges and approaches. In: Bianchi-Berthouze N (ed) DNIS. Lecture notes in computer science, vol 2822. Springer, Berlin, pp 5–15
40. Shen D, Chen Z, Yang Q, Zeng H-J, Zhang B, Lu Y, Ma W-Y (2004) Web-page classification through summarization. In: SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. ACM Press, New York, pp 242–249
41. Shen X, Tan B, Zhai C (2005a) Context-sensitive information retrieval using implicit feedback. In: SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. ACM Press, New York, pp 43–50
42. Shen X, Tan B, Zhai C (2005b) Implicit user modeling for personalized search. In: CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management. ACM Press, New York, pp 824–831
43. Sieg A, Mobasher B, Burke R (2007) Web search personalization with ontological user profiles. In: CIKM'07: Proceedings of the sixteenth ACM conference on conference on information and knowledge management. ACM Press, New York, pp 525–534
44. Sieg A, Mobasher B, Burke R, Prabu G, 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. Innsbruck, Austria
45. Spink A, Ozmutlu S, Ozmutlu HC, Jansen BJ (2002) U.S. versus European web searching trends. SIGIR Forum 36(2):32–38
46. Srinivasan P, Menczer F, Pant G (2005) A general evaluation framework for topical crawlers. Inf Retr 8(3):417–447

47. Sriram S, Shen X, Zhai C (2004) A session-based search engine. In: SIGIR'04: Proceedings of the international ACM SIGIR conference
48. Tamine L, Boughanem M, Daoud M (2009) Evaluation of contextual information retrieval: overview of issues and research. Knowl Inf Syst (Kais) (to appear)
49. Tamine L, Boughanem M, Zemirli WN (2008) Personalized document ranking: exploiting evidence from multiple user interests for profiling and retrieval. J Digit Inf Manage 6(5):354–365
50. Tan A-H, Ong H-L, Pan H, Ng J, Li Q-X (2004) Towards personalised web intelligence. Knowl Inf Syst 6(5):595–616
51. Tan B, Shen X, 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. ACM Press, New York, pp 718–723
52. Wang W, Zanane OR (2002) Clustering web sessions by sequence alignment. In: Proceedings of the 13th international workshop on database and expert systems applications (DEXA 2002). Springer, Aix-en-Provence, pp 394–398
53. Webb G, Pazzani M, Billsus D (2001) Machine learning for user modeling. User Model User-Adapt Interact 11(1–2):19–29
54. Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou Z-H, Steinbach M, Hand DJ, Steinberg D (2007) Top 10 algorithms in data mining. Knowl Inf Syst 14(1):1–37
55. Zhicheng D, Ruihua S, Ji-Rong W (2007) A large-scale evaluation and analysis of personalized search strategies. In: WWW '07: Proceedings of the 16th international conference on World Wide Web, pp 581–590
56. Zhou Y, Croft WB (2008) Measuring ranked list robustness for query performance prediction. Knowl. Inf. Syst. 16(2):155–171

Author Biographies

Mariam Daoud is currently a Ph.D. student at the Laboratory of Computer Science IRIT (France). She has received a M.Sc. degree in computer science from the university of Toulouse III in 2006 and a B.Sc. degree in computer science from the Lebanese University 2005. Her research interests are in contextual information retrieval and personalization, user profiling and semantic data mining. She has served as a program committee member in RJCRI (CORIA 2008) and an additional reviewer in ECIR 2009.

Lynda Tamine-Lechani is Assistant Professor at SIG/RI team in the Laboratory of Computer Science IRIT (France). She received her Ph.D. in computer science from the University of Toulouse (France) in 2000, and a M.Sc. in computer science from the same University in 2008. Her research interests are in theoretical models, experimental evaluation and applications design in IR and focus around contextual IR and flexible IR. She has been involved in projects on IR such as Quaero (http://www.quaero.org), ANR-AMPD (http://apmd.prism.uvsq.fr) and she has regularly been invited as a PC member of conferences on IR (SAC, ECIR, FQAS, CORIA).

Mohand Boughanem is Professor at the University of Toulouse III and the leader of the Information Retrieval and filtering (IRf) group at IRIT/SIG team. His current research focuses on IR models, XML information retrieval, Contextual/personalized IR, Ontology/concept-based IR. He has been involved in a number of projects and working groups on IR, the recent ones are Quaero (http://www.quaero.org), ANR-AOC (http://www.irit.fr/AOC), ANR-AMPD(http://apmd.prism.uvsq.fr) and he has regularly participated in major IR evaluation forums (TREC, CLEF, INEX). He has served as a programme committee member or chairman of the major IR conferences. In 2007, he was appointed Editor-in-Chief of the I3 journal (http://www.revue-i3.org). He has been invited as reviewers by several related-IR journals (IR journal, IP&M, JASIST). He is one of the founders of ARIA (the French Association of Information Retrieval) and CORIA (the annual French conference in information retrieval), both were founded in 2004. He was the chairman of ARIA during 2004–2007.