Article

Journal of Information Science
XX (X) pp. 1-23

DOI: 10.1177/016555150000000
jis.sagepub.com
SAGE

A personalized search using a semantic distance measure in a graph-based ranking model

Mariam Daoud
IRIT, Paul Sabatier University, 118 Route Narbonne F-31062, Toulouse cedex 9, France

Lynda Tamine; Mohand Boughanem
IRIT, Paul Sabatier University, 118 Route Narbonne F-31062, Toulouse cedex 9, France

Abstract
The goal of search personalization is to tailor search results to individual users by taking into account their profiles, which include their particular interests and preferences. As these latter are multiple and changing over time, personalization becomes effective when the search process takes into account the current user interest. This paper presents a search personalization approach that models a semantic user profile and focuses on a personalized document ranking model based on an extended graph-based distance measure. Documents and user profiles are both represented by graphs of concepts issued from predefined web ontology, namely the ODP[1]. Personalization is then based on reordering the search results of related queries according to a graph-based document ranking model. This former is based on using a graph-based distance measure combining Minimum Common Supergraph (MCS) and maximum common subgraph (mcs) between the document and the user profile graphs. We extend this measure in order to take into account a semantic recovery at exact and approximate concept level matching. Experimental results show the effectiveness of our personalized graph-based ranking model compared to Yahoo[2] and to different personalized ranking models performed using classical graph-based measures or vector space similarity measures.

Keywords
Personalization; user profile; graph formalism; ontology

1. Introduction

Traditional information retrieval (IR) approaches consider that user needs are described fully by the user query. These approaches are characterized as 'one size fits all' providing 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. In more general context, the relevance decision is made independently of what the user has searched before, thus by considering the query as the main clue that specifies the user information need.

Personalized search aims at enhancing the user query with the user profile to better meet the individual user needs. The user profile describes the user interests and preferences that could be explicitly set by the user or gathered implicitly from the user search history. The most challenging tasks for an effective search personalization are: (1) how to model an accurate user profile and, (2) how to exploit it to improve the search. User profile representation models vary from very simple representation models based on bag-of-words [1] to more elaborated ones based on external semantic knowledge such as ontologies or concept hierarchies [2, 3, 4]. A survey on the research approaches that exploit the information extracted from Web logs (or query logs) in personalized user ontologies for personalizing search is presented in [43]. Tailoring the search results to a particular user consists of exploiting his/her profile in a

[1] http://www.dmoz.org
[2] http://www.yahoo.com

Corresponding author:
Mariam Daoud, IRIT, Paul Sabatier University, 118 Route Narbonne F-31062, Toulouse cedex 9, France, mariam_daoud@live.fr

personalized document ranking by means of query reformulation techniques [2, 5], query-document matching [6, 7, 8] or result reordering [9, 4, 3].

In this paper, we exploit a semantic graph-based user profile in a personalized document ranking model. In this model, the documents and the user profile are both represented as graphs of concepts and then personalization consists of reordering the documents using a graph-based distance measure. This latter is based on a semantic extension of the graph-based distance measure combining maximum common subgraph (mcs) and Minimum Common Supergraph (MCS). We extended this measure by introducing an approximate concept matching considered through cross links connecting the two graphs. In this paper, we extend our previous work [42] by (1) studying the accuracy of the user profile modeling (2) tuning different parameters of the systems and studying their impact on the personalized search effectiveness, (3) comparing our approach to other graph-based measures used for personalized reranking as well as to a vector space matching model where documents and the user profile are represented either by term lists or concept lists.

The rest of this paper is organized as follows. In section 2, we present some related works on personalized document ranking models and give an overview of exploiting a graph formalism for improving the search process. In section 3, we present a general overview of our contribution; the graph representation model for documents and the user profile and highlight our research objectives. In section 4, our personalized graph-based document ranking model is detailed. In section 5, an experimental evaluation setup and appropriate evaluation protocols for the user profile accuracy and the personalized search effectiveness are described. In section 6, the evaluation results of the user profile accuracy are presented. In section 7, the results of the personalized search effectiveness evaluation compared to Yahoo search including a comparative study with other ranking models are presented. Finally, section 8 concludes the paper.

2. Related works

In this section, we overview some significant related works on representing user profile models, personalizing search using a user profile and exploiting a graph formalism for improving the retrieval effectiveness.

2.1. On modelling the user profile

A common aspect in most of user modeling techniques is to collect relevant sources of evidence from the user search history in order to model the user profile. Generally, the user search history is composed by the user's documents of interest such as recently browsed or judged/assessed web pages [9, 4], a repository of interesting information such as emails, browsing features or desktop information [10]. These data are then exploited in algorithms of data mining [4, 9, 11] or machine learning [12, 3, 13] strategies in order to build the user profile. The user profile could represent either short term user interests [11, 4] or long term ones [14] and could be represented by simple representation models based on simple terms [15, 6] or using ontologies and concepts hierarchies [4, 9, 3]. Several works make distinction between short term user interests [11, 4, 16] used to better meet the user needs in a current task, and long term user interests [14] used to improve the search for any query. Generally, a short-term user profile refers to the user interests during a short period of time and inferred from the recent search history. Some works such as [11, 4] consider only the pre-query session activity to define the search context. White et al. [16] consider the user interest model for the current query in addition to the session context defined by pre-query session activity, and their combination to refer to the user context, called intent model. A long-term user profile holds persistent user interests generally stable for a long time and inferred from the whole user search history.

Concerning the user profile representation models, they are often based on a set of keyword vectors [15, 6] or classes of keyword vectors [1]. To overcome the limitations of the keyword-based representation models (absence of interrelations between keywords), some approaches model the user interests by term and term relations [5] or hierarchical clusters of terms [17, 18, 19] issued from the user's documents of interest. The benefit behind such representation models is that the links between terms allow improving the retrieval effectiveness for similar queries. Even if such representations are well elaborated, they are leveraged from the user history that is often limited and not sufficient to encounter new user need. In order to integrate new knowledge in the user profile model, some approaches make use of predefined semantic resources to represent the user profile as a set of concepts issued from predefined ontology [3, 20] or an instance of reference ontology [4, 9]. It is well known that ontology-based user profile representations present several advantages compared to other representation models as they provide explicit semantic knowledge that facilitates the inference of new knowledge [21].

2.2. On personalizing search using the user profile

Search personalization is achieved by integrating the user profile in the query reformulation process [2, 5, 11], query-document matching [6, 8, 7], personalized document categorization [22] or document reordering [4, 3, 9, 23].

Most of query reformulation techniques used in personalized IR are based on adding or reweighting terms issued from the user profile [2, 5, 11]. In [2], the query is matched with the most appropriate pair of concepts of the user profile; the first concept is the most similar one to the user query, and it is used to add or enhance the weight of the query terms while the second concept is the less similar one, it is used to eliminate non relevant terms in the search. A more elaborated user profile based on terms connected by different edge relations [5] (negation, substitution, etc.), allows adding, eliminating or substituting the query terms with relevant ones. Query refinement in [11] consists of adding weighted terms to the query from the user context defined by the user search history using Rocchio [24] algorithm. Added terms are selected according to a language model that calculates the weight of a term in the previous queries and the snippets of the clicked documents.

Few works incorporate the user profile in a query-document matching model. This latter consists of calculating the relevance score of a document relative to both the user query and the user profile. A bayesian model integrating a document relevance score function is proposed in [6] in order to increase the score of the document when its vocabulary matches the user profile one. In the same context, the proposed approach in [7] adjusts the scoring function parameters to the user context using genetic programming.

Personalization based on result re-ranking consists of combining the content-based document score with the personalized document score [9, 4, 3] or with the personalized PageRank of the document [23]. The personalized document score is computed in [9, 4] using the cosine similarity measure with the most similar concepts of the user profile. A variant technique of personalized document ranking is the personalized categorization of the search results. It consists of grouping the search results into categories that describe the user interests [22]. Personalization in [3] consists of combining personalized categorization and result re-ranking using a voting-based merging scheme. Mapping the query into a set of related concepts 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 concept associated to the list and the similarity of the concept with respect to the query. Personalized result-reranking based on PageRank is described in [23]. The user selects his preferred pages from a set of hub pages and a personalized PageRank vector is computed for each user interest used to redirect the returned web pages to the preferred ones. Chirita et al. [20] use a system in which a user manually selects ODP categories as entries in his/her profile. Reranking search results is based on measuring the similarity between a search result and the user profile using the node distance in a taxonomy concept tree, which means the search result must associate with an ODP category.

2.3. On using graphs for document ranking

The use of a graph formalism for document ranking is investigated in several works for improving the retrieval effectiveness. We categorize these works into two broad categories based on the following principles:

- using the graph as a semantic representation model for queries and documents and exploiting it in a semantic document retrieval model using graph-based measures [25, 26, 27, 28,44],
- using the graph to formalize hyperlinks between documents, clickthrough relationships between queries and documents [29], or semantic relationships between terms and documents [30, 31, 32] and then exploit the graph as an additional resource for improving the retrieval effectiveness.

Concerning the first category of works, graphs of documents or queries are represented by concepts of WordNet linked by association rules, where each concept is defined by a set of descendant concepts [25] or by fuzzy conceptual graphs used for medical text indexing [44]. Ranking the documents with respect to a query is based on calculating a conceptual similarity using a graph-based measure that combines structural and relational similarity degrees on WordNet.

In the same context, queries and documents are represented by concept subtrees of WordNet [26]. Ranking the documents with respect to the user query is based on calculating the minimal subtree that contains the document subtree H_d and the query subtree H_q at the same time and then calculating the relevance document score using a graph-based distance measure. This latter makes use of fuzzy logic association rules such as Diene, Gödel, or Lukasiewicz.

In [27], a document is represented as a subgraph of a predefined ontology graph and an approximate graph matching approach is used for content (address) extraction. The ontology construction for the address domain utilized the DMTI GIS database, which provided all instance concepts, such as “Street Name”, “Street Type”, “City” and “Province”, used Quebec and Ontario, Canada and the www.yellowpages.ca is used as a test Web document set. The involved graph matching model considers a graph template for the address domain (city, country, ..) and calculates a similarity between a graph representing a segment text according to the number of shared nodes as well as the number of shared edges between graphs.

Instead of using only the shared nodes between graphs, a conceptual graph matching for semantic search in [28] exploits a similarity between concepts representing the entries of graphs and a similarity between relations based on their respective positions in the concept hierarchy from which they are issued (namely WordNet). The similarity between concepts is calculated in terms of the length path through the closest common parent concept. While the similarity score between two relations is 1 depending on whether a relation is a supertype of another relation and 0 otherwise.

Concerning the second category of works, several graph-based ranking models similar to HITS algorithm [33] or PageRank [34] have been proposed in recent years, by focusing on using the graph formalism as an additional resource for improving the retrieval effectiveness. While the graph in HITS and PageRank formalizes relationships between the documents through the hyperlinks, it formalizes in some other approaches the relationships between queries and clicked documents in the same searching session [29], or relationships between terms and documents [30, 31, 32].

3. Research overview and objectives

As mentioned above, the main challenging questions for an effective search personalization are (1) how to model a user profile and (2) how to exploit it in an effective way. Concerning the user profile modeling, we consider a semantic graph-based user profile [35] that represents a short-term user interest in a specific search session. The user profile is inferred from the ODP ontology[3], which is considered as a highly expressive ground to describe a semantic representation model of the user profile. In order to exploit the user profile in an effective way, we propose a personalized graph-based ranking model, which is the focus of the paper. Our research objectives and a general overview of our search personalization approach are given below.

3.1. Research objectives

Our objective in this paper focuses on a personalized document ranking model where documents are reordered according to their similarity with the user profile using a graph-based distance measure. Compared to other related works that model ontology-based user profiles [4, 9, 3, 22], our work is the first attempt to involve a graph-based representation for both the documents and the user profile and a personalized graph-based ranking model for reordering the documents. Our intuition behind this model is to capture term dependencies between terms in the document content as well as between the concepts of interests in the user profile model and to rank semantically the documents according to an exact and approximate concept level matching. Indeed, the main limitation of document-profile keyword-based matching depends relatively of the document length by treating multiple topics in the documents in similar way. Concept-based matching allows overcoming the limitations of the term-based matching by exploiting the semantics in the document and by discovering concept dependencies between multiple topics that could exist between the document and the user profile instead of estimating the shared terms between the document and the user profile.

Compared to other related works, our approach has the new following features:

- Most of personalized document reranking approaches represent the documents and the user profile by term vectors. Whereas our approach represents both the documents and the user profile by graphs of concepts,
- Most of other personalized reranking methods ignore the semantic relationships between concepts of the user profile by ranking the documents according to their similarity to the user profile using the cosine measure. Whereas our approach ranks semantically the documents compared to the user profile according to a graph-based distance measure.

The main research questions that highlight our contribution in the present paper are the following:

- How to exploit the graph formalism of the reference ontology to represent semantically the documents and the user profile?
    - o Concerning the user profile modeling, we have previously proposed [35] a graph-based representation model, inferred from the ODP ontology,
    - o Concerning the document graph, it is inferred by mapping the document content on the ODP ontology and extract its associated graph.
- How to design a graph-based ranking model that ranks semantically the documents with respect to the user profile?

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

- We use a semantic graph-based measure based on combining MCS and mcs measures [36] between the document and the user profile graphs.
- We extend this measure to take into account a semantic recovery between the document and the user profile graphs at exact and approximate concept level matching.

3.2. General approach

Our search personalization approach aims at exploiting the user profile in a result reranking model. It consists of (1) representing both the documents and the user profile, as graphs of concepts and then (2) reordering the search results of a given query according to a graph-based document ranking model using the user profile. This latter is built across previous queries related to the current user information need, which are grouped in the same search session. The graph-based representation model of the user profile and the documents and a general overview of our session-based personalized search using a graph-based measure are detailed below.

3.2.1 A graph-based user profile representation model

The user profile represents the user interest in a current search session and is represented as a graph of weighted and interrelated concepts inferred from the ODP ontology [35]. It is built across related queries by mapping the user's documents of interest on reference ontology and then applying spreading concept activation for extracting the graph of the most relevant concepts. The overall process of generating the user profile is detailed in two main steps: (1) building the ontological query profile, (2) maintaining the user profile in the same search session.

- Building the ontological query profile: the query profile represents the concepts extracted from the reference ontology with respect to a user query and the associated user relevance feedback at the end of a search activity. Building the query profile is based on two main steps.
  - *Extracting a keyword vector*: a keyword vector is extracted from the user's documents of interest D_R and then is mapped on the reference ontology to extract an initial weighted concept set. Each document in D_R is represented as a single term vector where the weight w_{td} of term t in document d is computed as follows: $w_{td} = tf_d * \log(N/n_t)$. 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 represent then the keyword vector by averaging the documents in D_R. In order to represent the user interests semantically as a set of concepts derived from the ontology, we map the keyword vector on the ODP ontology using the cosine similarity measure. The result of mapping the query context on the ontology is an initial set of concepts, called θs = {(c1, score(c1), ..(ci, score(ci),)}.
  - *Inferring the graph-based query profile*: based on the set of concepts θs, we attempt to build a graph of semantically related concepts of ontology using a spreading activation strategy. Spreading activation consists of propagating the weights of concepts having non-zero values through different edge types of the ontology. We distinguish the role of different edges in activating linked concepts in the score propagation. A set of graphs is then identified by grouping interrelated concepts together and the query profile is finally represented by the most relevant graph consisting of the most highly interrelated concepts. More details regarding the spreading activation process could be found in [35].
    We mention that the 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 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. 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.

- Maintaining the user profile: it consists of combining query profiles issued from the same search session. More precisely, it is based on accumulating the weights of possible common concepts and adding edges linking related concepts across related queries.

Formally, the graph structure G=(V,E) of the user profile has a hierarchical (tree) component 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 and R corresponds to the non-hierarchical component made of "related" cross links.

Figure 1 illustrates an example of a user profile/document that corresponds to the computer language programming interest. In this example, the user/document profile G is defined by the following sets:

V = {(c1, score(c1)), (c2, score(c2)), .., (c8, score(c8))},
S = {(c5, c4), (c5, c8), (c5, c6)},
T = {(c1, c2), (c1, c3), (c2, c4), (c2, c5), (c3, c6), (c3, c7), (c4, c8)},
R = {(c5, c3)}.

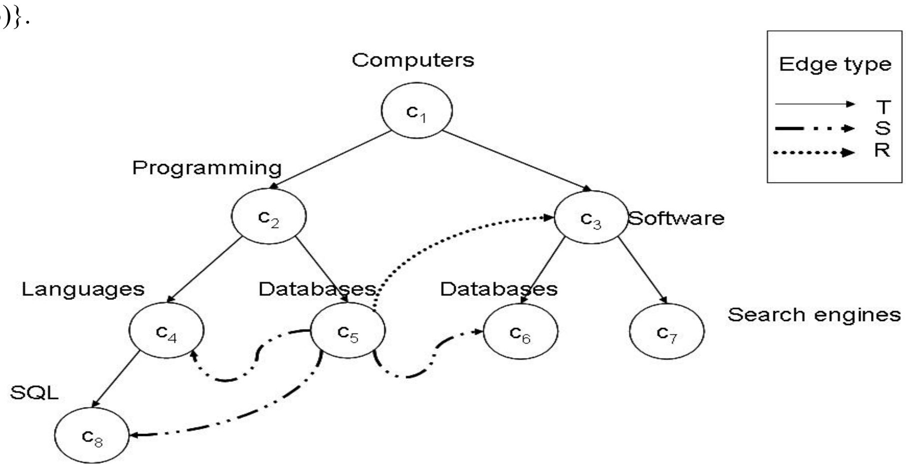

Figure 1: A portion of an ontological user profile in a graph-based representation

3.2.2 A graph-based document representation model

Documents returned with respect to a user query are represented by graphs of concepts and reordered according to a graph-based distance measure with respect the user profile. Inferring the document graph is described by the following steps:

- representing the document content by a term vector $\overrightarrow{d_k}$,
- mapping the document content on the ODP ontology by calculating the weight of document d_k in concept c_j as follows:

$$weight(c_j) = \cos(\overrightarrow{d_k}, \overrightarrow{c_j})$$

  where c_j is a concept of the ontology represented as a vector of terms issued from the web pages associated to the concept and *weight*(c_j) is the weight associated to concept c_j. More details on ontology concept representation are presented in [35].
- connecting the top *n* weighted concepts using different edge types of the reference ontology, and activating the associated general concepts in order to represent the document content at both general and specific concept levels.

Figure 1 could also represent an example of a document representation model using the reference ontology.

3.2.3 Our session-based personalized search using a graph-based distance measure

Our search personalization approach integrates a short term user profile modeling for representing the user interest in a current search session. Search personalization is then based on exploiting the user profile in a graph-based document ranking model for reordering the search results of a related query.

The overall process of our session-based personalized search could be described in algorithm 1 as follows. The algorithm includes a session boundary recognition mechanism for grouping related queries in the same search session, a user profile modeling across related queries and a personalized ranking model that represents documents by graphs and makes use of a graph-based distance measure for reordering the documents.

- Session boundary recognition: each new submitted query is considered in a session identification method proposed in our previous work [35] using the Kendall rank correlation measure [35]. This latter quantifies a conceptual correlation ΔI between the user profile G_u^s built in the current session and the new submitted query q^{s+1} submitted by the user at time $s + 1$. We choose a threshold σ and consider that the queries are from the same session if the correlation is above the threshold.

Algorithm 1: A general approach for a session-based personalized search using a graph-based user profile

q^{s+1} : a query submitted by the user at time s + 1

R^{s+1} : the search results returned by the system with respect to user query q^{s+1}

G_u^s : a user profile built at time s in a search session

θ : a reference ontology

FOR EACH new submitted query q^{s+1} by user *u* do

Session boundary test: Compute query-profile conceptual correlation

$\Delta I = (q^{s+1} \circ G_u^s)$

If ΔI ≥ σ then // No session boundary is detected:

1) Re-rank the search results R^{s+1} of the query q^{s+1} using the user profile G_u^s

FOR EACH document d in R^{s+1} do

Build the document graph representation G_d using the reference ontology θ

$G_d = project(d, \theta)$

Calculate the personalized document score $S_p(G_d, G_u^s)$

$S_p(G_d, G_u^s) = GraphDis(G_d, G_u^s)$

Calculate the final document score $S_f(d)$

$S_f(d) = \gamma * S_i(q^{s+1}, d) + (1-\gamma) * S_p(G_d, G_u^s)$

END FOR

2) Update the user profile using the query profile $G_u^{s+1} = G_u^{s} \cup G_q^{s+1}$

ELSE

A new session is detected:

Reinitialize the user profile by the query profile $G_u^{s+1} = G_q^{s}$

END IF

END FOR

- User profile modeling: the user profile is built across related queries using the relevant documents assessed by the user as described in section 3.2.1. When a new query is grouped in the current session, the user profile is maintained by exploiting the recent user relevance feedback. Otherwise, the user profile is reinitialized and rebuilt in the new session.
- Search personalization: when the query is related to the search session, the user profile built in the current session is used to rerank the returned documents according to a graph-based distance measure. The document reordering process is achieved as follows:
  - build the document graph G_d using the ODP ontology (cf. section 3.2.2),

- calculate the personalized document score $S_p(G_d, G_u^s)$ of each retrieved document d using a graph-based distance measure,
- combine the original document score S_i returned by the system according to a content-based ranking model and its personalized score S_p using a reranking parameter $0 \leq \gamma \leq 1$. When γ has a value of 1, personalized score is not given any weight. If γ has a value of 0, the original score is ignored and pure personalized score is considered.

4. A personalized graph-based ranking model using a semantic user profile

Our search personalization approach exploits the user profile in a personalized graph-based document ranking model using a semantic extension of the distance measure combining minimum common supergraph (MCS) and maximum common subgraph (mcs). We review in this section the most common known graph-based distance measures, our motivations for using the combined distance measure and our personalized graph-based document ranking model.

4.1. Background: Graph-Based Distance Measures

The most common known graph-based distance measures are the maximum common subgraph (mcs) [37], the minimum common supergraph (MCS) [37], or the combined measure using MCS and mcs [36].
The maximum common subgraph (mcs) of two graphs g_1 and g_2, is a subgraph g of g_1 and g_2 based on common concepts and has among all the subgraphs, the maximum number of nodes [38]. The graph-based distance measure based on the subgraph mcs [37] is given by the following formula:

$$d(g_1, g_2) = 1 - \frac{mcs(g_1, g_2)}{\max(|g_1|, |g_2,|)} \qquad (2)$$

Where $|g_1|$ (resp. $|g_2|$) is the the number of nodes in g_1 (resp. in g_2). This formula gives lower distance for graphs having large mcs.

The minimum common supergraph (MCS) of two graphs, g_1 and g_2, is a graph g that contains both g_1 and g_2 as subgraphs and that has the minimum number of nodes and edges [38]. The distance measure based on only the supergraph (MCS) [38] is given in the following formula:

$$d(g1, g2) = 1 - \frac{|g_1| + |g_2| - |MCS(g_1, g_2)|}{\max(|g_1|, |g_2|)} \qquad (3)$$

The distance measure based on combining MCS and mcs is given in [36] as follows:

$$d_{MMCS}(g_1, g_2) = |MCS(g_1, g_2)| - |mcs(g_1, g_2)| \qquad (4)$$

According to this measure, the distance between graphs is lower when the size of the supergraph is smaller and the size of the subgraph is larger.

4.2. Motivations

Our choice of using the graph-based distance measure combining MCS and mcs in our personalized search model is based on the following assumption: "a document is ranked higher if it recovers the maximum of concepts of the user profile at both specific and general levels". According to this assumption, documents presenting common general concepts are ranked lower than documents presenting general and specific concepts with the user profile. Indeed, our choice could be justified by the following reasons:

- The Minimum Common Supergraph (MCS) allows measuring the similarity between the document and the user profile at general levels even if the graph do not have common concepts. Indeed, the supergraph of two graphs is smaller when the root nodes of both graphs are close in the ontology.
- The maximum common subgraph (mcs) allows measuring the similarity between the document and the user profile at specific levels. Indeed, the subgraph is larger when the two graphs have common concepts at specific levels, and consequently this enlarge the subgraph with common concepts at general levels.

- Thanks to the ODP ontology, the following assumption for calculating the combined distance measure based on MCS and mcs remains correct: "The bigger is the subgraph mcs between two graphs, the smaller is the supergraph MCS".

Compared to previous cited work, our method overcomes the limitations of document-profile term-based matching [9, 4]. Indeed, while term-based matching introduces the impact of noisy terms from both the document and the user profile, our approach alleviates this limitation by performing a graph-based matching at concept-level. Moreover, our approach integrates both common and similar concepts into the personalized score. The personalized approach in [20] use a distance measure between multi-concept profile derived from the ODP and a document/url represented by one concept of the ontology, in which it is already classified. In this approach, the document collection is restricted to the Web pages in the ontology and the document-profile distance makes use of average distance in terms of nodes using is-a relationships. While the main advantage of our ranking is to consider a multi-concept representation of the document and a semantic similarity between the concepts via different kind of links instead of taking into account only is-a links. This ranking gives higher score for a document that has a graph similar or close to the user profile graph than a document that has a distant graph from the user profile graph.

4.3. A graph-based ranking model based on a document-profile semantic recovery

We propose a graph-based ranking model that ranks semantically the documents with respect to a semantic user profile. The involved ranking model calculates a personalized score of the document using an extension of the combined distance measure. We mention that the basic distance measure combining MCS and mcs in formula 4 can't deal with an approximate matching between the document and the user profile. Indeed, this measure gives similar distance for documents that have cross links with the user profile and others that do not have cross links nor common concepts. This is due to the exact recovery assumption using the common concepts to build the subgraph. That is why we argue to extend semantically the subgraph of two graphs by taking into account a relative document-profile semantic recovery. Our intuition at this level is to consider two types of recovery in the semantic distance measure:

- *Exact recovery*: refers to an exact similarity between the document and the user profile. It is calculated by the number of common concepts.
- *Approximate recovery*: refers to a relative similarity between the document and the user profile. It is calculated by the number of common related concepts connecting the two graphs, in other terms, those linked with cross links.

We take into account these two types of recovery when calculating the personalized document rank. We extend first each of the document's graph and the user-profile's graph and then compute the distance measure based on the combination of MCS and mcs of the two extended graphs by giving more weight to the direct common concepts.

4.3.1 Extending the document-profile common subgraph using cross links

Formally, let G_u and G_d the graphs representing respectively the user profile and a document. As shown in figure 2, G_d is a graph of concepts representing document d and G_u is a graph of concepts representing the user profile. The set of concepts of graph G_d connected to graph G_u with cross links represent the extension of graph G_u. Formally, we define the extended graph G_u^{d*} of graph G_u with respect to G_d, as follows:

$$G_u^{d*} = G_u \cup \{c_i \in G_d / \exists c_j \in G_u \wedge e_{ij} \in S \cup R\} \quad (5)$$

Where e_{ij} is the edge linking concept c_i to concept c_j, $S \cup R$ is the set of symbolic and related concepts of the ODP ontology (cf. Sect.3). We create the extended graph G_d^{u*} by the same manner as the graph G_u^{d*}. We obtain two extended graphs G_u^{d*} and G_d^{u*} that will be used in the personalized document ranking model. Figure 2 presents an illustrative example of an extended graph of g_1 from graph g_2 where concepts c_{11} and c_{14} from g_2 and their links to graph g_1 compose the extension.

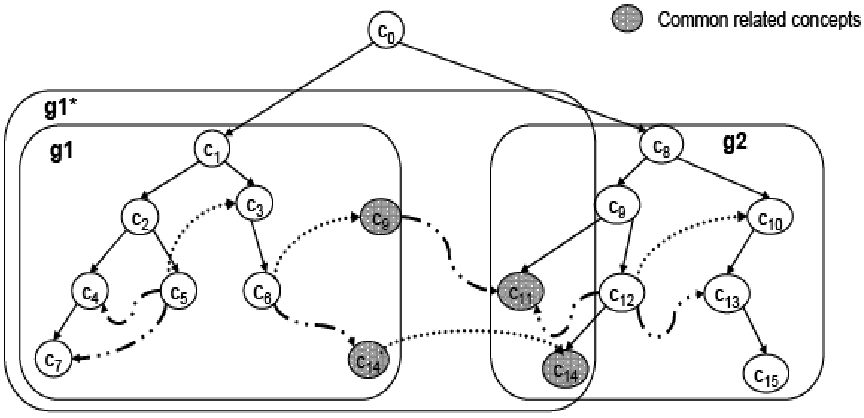

Figure 2: A semantic extended graph through cross links

4.3.2 Calculating the personalized document score

We use the extended graphs G_u^{d*} and G_d^{u*} to calculate the personalized relevance score of a document based on a semantic distance combining MCS and mcs between the document and the user profile. The mcs of the two extended graphs contains initially common concepts called mcs_{cc} and the activated concepts issued from the graph extension called mcs_{ca} that are either the concepts linking the two graphs through cross links or the inner concepts linking the concepts of the subgraph together.

In order to distinguish the role of the related concepts connected through the cross links relative to the direct common concepts between graphs, we used a decay factor f_{ca}. This factor is calculated automatically based on the following assumption: "*the weight of the activated concepts compared to the common ones must be reduced as much as we have symbolic or related edges connecting the graphs*". Indeed, the number of activated concepts is high when the number of links between graphs or the number of common concepts is high. f_{ca} is given as follows: $f_{ca} = \frac{L_R}{1+L_R}$, L_R is the set of cross edges linking concepts of the two graphs. Finally, in order to take into account the graph size difference between the documents compared to the user profile, we normalized the semantic distance measure between the graphs by dividing over MCS as follows:

$$S_p(G_u, G_d) = \frac{\left|MCS(G_u^{d*}, G_d^{u*})\right| - \left(\left|mcs(G_u^{d*}, G_d^{u*})\right| + fca * \left|mcs_{ca}(G_u^{d*}, G_d^{u*})\right|\right)}{\left|MCS(G_u^{d*}, G_d^{u*})\right|} \tag{6}$$

5. Experimental evaluation setup

In this section, we present the experimental evaluation setup concerning the evaluation objectives, the dataset and the evaluation protocol.

5.1. Evaluation objectives

The experimental evaluation of our approach is undertaken through a hybrid evaluation combining context simulation and user study as adopted in [39] where we used real user data issued from the search log of Exalead [4] web search engine. The evaluation methodology and measures are adopted for the evaluating the effectiveness of contextual search technologies in the context of the QUAERO project[5]. Experiments are conducted to achieve the following objectives:

- *Evaluate the user profile accuracy*. This step is necessary as the user profile accuracy has an important impact on the personalized search effectiveness. The goal is to assess the relevance of the concepts that represent the user profile with respect to the associated queries. Analyzing the user profile accuracy is achieved in terms of

[4] http://www.exalead.com

[5] http://www.quaero.org/modules/movie/scenes/home/

the quality of the concept rank ordering, the precision and the recall considering different percentages of the top ranked concepts of the user profile.

- *Evaluate the personalized search effectiveness*. The goal is to evaluate the effectiveness of our personalized search in front of a typical search performed by Yahoo, and that do not integrate any contextual element in the search process.
- *Compare our personalized search approach to other ranking models*. A comparative study is conducted to evaluate the effect of our extended graph-based ranking model to other ranking models based on using basic graph-based measures or vector-space similarity measures.

5.2. Dataset

As there is no publically available dataset for personalized search evaluation purpose, we exploited a search log of a commercial web search engine, namely Exalead, and we extracted the search history of 10 users collected along three months. Descriptions of search sessions, queries and document collection are given below.

5.2.1 Search sessions

As our approach is based on personalizing search across sessions, each session is defined by a sequence of related queries, where we have selected 25 user search sessions for all the 10 users as follows:

- Each session contains three queries related to the same user information need and are submitted by the same user in a chronological order.
- Each query has at least one clicked document as it is the only source of evidence to build the user profile in a search session. We assume that a document is relevant if it is clicked by the user.

A preliminary cleaning of the web search log leads to maintain only relevant clicked documents in the search sessions so as the user profiles which are created from non biased resource.

5.2.2 Queries

Queries are extracted from the user search sessions. In order to test the personalized search effectiveness along a user search session, we have divided the query set per session into a training query set and a testing query set.

- *Training query set*: it is used to learn the user profile. The first two queries of each search session are part of the training query set. Therefore this set contains a total of 50 queries.
- *Test query set*: it is used to evaluate the retrieval performance. The last formulated user query of each session is part of the testing query set. Therefore this set contains a total of 25 queries, which is the minimum number required for validating experimental findings [45].

Queries have an average length of 3 terms and the user intent behind these queries is mostly informational (“Risques auditifs”) or transactional (“Le bourg d’oisans hotel”). We notice that related testing and training queries in the same session are different where in some few cases a single common keyword exists and in other cases, the queries have completely different keywords. In addition, as these queries are extracted from a Web search log, this make the query set consistent for evaluating real personalized Web search.

Table 1 presents some example of training and testing queries derived from the same search sessions.

Training queries	**Tesing query**
Q1: Ouie, Q2: Oreille interne	Q3: Risques auditifs
Q1: Pollinisation, Q2: Vanille culture	Q3: Abeille et gaucho
Q1:Cocktail menthe - rhum, Q2: cocktail mojito	Q3: Cocktail margarita
Q1: Arbre orme Q2Graphiose	Q3: Lichens bioindicateurs

Table 1: Examples of training/ testing queries derived from search sessions.

JIS Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

5.2.3 Document collection

The document collection consists of collecting the top 50 results retrieved from the publicly available Yahoo API[6] for each testing query. Results are crawled and each one is represented by its complete content. In our evaluation setting, these documents are used only for reranking the search results using the user profile. In order to evaluate the retrieval effectiveness, the relevance assessments for the testing queries were given through a user study. To do this, 5 computer science students of our lab[7] were presented with the set of top 50 results retrieved from Yahoo. Each participant was considered as the user who has formulated the query and asked to judge whether each document was relevant or not according to the subject of a subset of testing queries.

5.3. Evaluation protocol

As stated above, we evaluate both the user profile accuracy and the personalized search effectiveness. In order to achieve this objective, we propose appropriate evaluation protocols described as follows.

5.3.1 How to evaluate the user profile accuracy?

The goal of evaluating the user profile is to validate their accuracy and their exploitation for personalizing search as this evaluation has a strong impact on the effectiveness of the personalized approach. Hence, we only look at evaluating the user profile accuracy for validation purpose without comparing its derivation technique to other methods. The protocol for evaluating the user profile accuracy is inspired from [40] where the evaluation is made at mono-profile level assigned for each user. We adapt this protocol to take into account multi-profiles assigned for each user where the evaluation measures are averaged over the profile built for each user. The methodology for evaluating the user profile accuracy and the evaluation measures are given below.

Methodology. We recall that each user u_i is associated to multiple profiles built across the user search sessions. Each profile is built in a search session across the training queries and is represented as a graph of weighted concepts issued from the reference ontology. We are interested in evaluating the graph content (the concepts representing the nodes of the graph) by ignoring its structure as it is issued from reference ontology.

Formally, we consider for each user u_i, the set of n profiles $Pfl_{u_i} = \{Pfl_1^{u_i}, Pfl_2^{u_i}, Pfl_3^{u_i}, ..., Pfl_n^{u_i}\}$. Each profile is composed of a list of ordered weighted concepts $Pfl_j^{u_i} = \{C_{j1}, C_{j2}, .., C_{jm}\}$. For each user u_i, the user profile evaluation protocol consists of the following steps:

- *Assessment step:* for each profile $Pfl_j^{u_i}$, one participant assessed each concept C_{jk} as relevant or non-relevant with respect to the associated queries in the search session.
- *Local accuracy evaluation step*: this step consists of computing the accuracy of each profile $Pfl_j^{u_i}$.
- *Global accuracy evaluation step*: this step consists of computing the average accuracy of the profiles in $Pfl_j^{u_i}$ built for a user u*i* across the whole search history.

Evaluation measures. In contrast to the direct use of some common known IR measures for evaluating the user profile accuracy [40], where each user has a mono-profile, we have computed for each user ui an average accuracy value of the set of profiles built across the user search sessions. The evaluation measures used to calculate the user profile accuracy are the following: the average rank of non relevant concepts (*ARnkNR(u_i)*), the precision P@x(u_i), the recall R@x(u_i) at the top *x* concepts of the user profile for each user u_i and the F-measure averaged over all the users. Given a user profile, precision at top *x* concepts is computed as the fraction of the number of relevant concepts among the top *x* over *x*, and recall at top *x* concepts is computed as the fraction of the number of relevant concepts among the top *x* over the total number of relevant concepts in the user profile.

5.3.2 How to evaluate the personalized search effectiveness?

The protocol for evaluating the personalized search effectiveness consists of a training step and a testing step.

- *Training step*: This step consists of learning the user profiles for each testing query using the clicked documents of the corresponding training queries belonging to the same user.

[6] search.cpan.org/perldoc?Yahoo::Search

[7] www.irit.fr

- *Testing step*: This step consists of evaluating the personalized retrieval effectiveness for each testing query using the user profile compared to the baseline search performed by Yahoo search using only the testing query. Personalized search is based on reranking the top 50 results of Yahoo for each testing query using the appropriate user profile and by combining for each document its original rank (sorted by Yahoo) and its personalized rank calculated using our graph-based distance measure. We use the precision at top 10 and top 20 documents (P10, P20) as an evaluation metric. We pool together the queries and judgments of all the ten users, so that the evaluation result will be an average over the whole testing queries.

6. User profile accuracy evaluation results

We have conducted two series of experiments in order to evaluate (1) the concept ranking quality and (2) the effect of the concept depth of the reference ontology on the user profile accuracy.

6.1. Concept ranking quality results

In this experiment, we evaluate the concept ranking quality of the user profile according to the evaluation protocol presented in section 5.3.1 where concepts are ordered according to their weights in the user profiles. We then analyze the evaluation results in terms of the rank of non relevant concepts, the precision, the recall and the F-measure.

6.1.1 Analyzing the concept rank ordering

Similar to the study conducted in [40], we analyze the accuracy of the user profile in terms of the average rank of non relevant concepts for each user u_i. Table 2 presents for each user u_i, the average profile size ($Avgc(u_i)$) which is the average number of concepts of the profiles in Pfl_{ui} associated to user u_i, the average rank of non relevant concepts averaged over the set of profiles associated to user u_i ($ARnk_{NR}(u_i)$), and the normalized average rank $NoARnk_{NR}(u_i)$ calculated by dividing the average rank of the non relevant concepts over the average profile size.

A concept rank ordering that produces a large value of $ARnk_{NR}(u_i)$ ranks non relevant concepts further down the list. As shown in table 2, the average number of concepts of the user profile ($avgc(u_i)$) varies between 19 (User 4) and 83 (User 10) and the average rank of non relevant concepts varies between 14 (User 4) and 43 (User 10) with respect to the user profile size ($Avgc(u_i)$). As a measure of the quality of the concept ordering, the normalized $ARnk_{NR}(u_i)$ gives an average rate of placing non relevant concepts further down the user profile representation. For all the users, $NoARnk_{NR}(u_i)$ is above 50%, and reaches the highest value of 78% (User 4). In average, the obtained values highlight that the user profiles were relevant at least at the highly weighted concepts.

We mention that the quality of the user profile measured by the average rank of non-relevant concepts is highly affected by the document-concept matching accuracy. Indeed, a more precise classification of the documents in the session history into the ontology contributes to a better user profile quality where non-relevant concepts would be ranked down or excluded from the user profile representation.

Users	Avgc(ui)	ARnkNR(ui)	NoARnkNR(ui)
User 1	36.6	21.04	0.63
User 2	55	28.1	0.51
User 3	40.5	23.08	0.63
User 4	19.5	14.68	0.78
User 5	33.5	20.81	0.63
User 6	33.67	18.7	0.53
User 7	32.67	17.24	0.53
User 8	38	18.99	0.5
User 9	28	15.91	0.57
User 10	83	43.1	0.51
	Avg	**22.165**	**0.582**

Table 2: Average rank of non relevant concepts per user

JIS Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

6.1.2 Analyzing the user profile precision and recall

In this section, we analyze the user profile accuracy for each user u_i in terms of the average precision and the average recall calculated at different percentages of the top ranked concepts of the user profile. As it is difficult to judge all the concepts of the ODP ontology whether each one is relevant and not included in each user profile, we computed the recall as proposed in [40] where recall at top x concepts is computed as the fraction of the number of relevant concepts among the top x over the total number of relevant concepts in the user profile. This could be an approximation of the global recall calculated over the entire ontology.

Figure 3 and Figure 4 show respectively the precision and the recall calculated considering 10%, 20%, 30%, ..., 100% of concepts of the user profiles. Results show that the precision of the user profiles is quite different between the users and it is variable when considering differrent cutoff concept points.

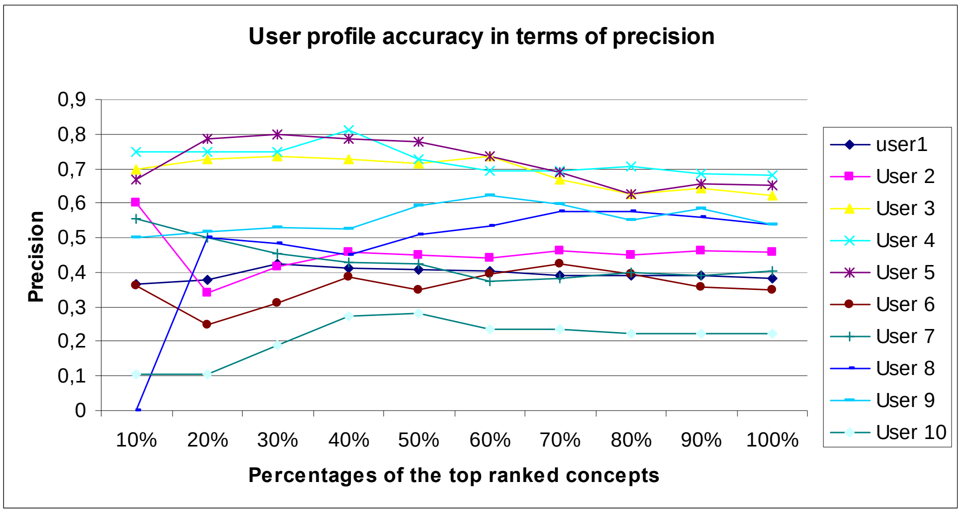

Figure 3: Precision calculated for each user at different concept cutoff points

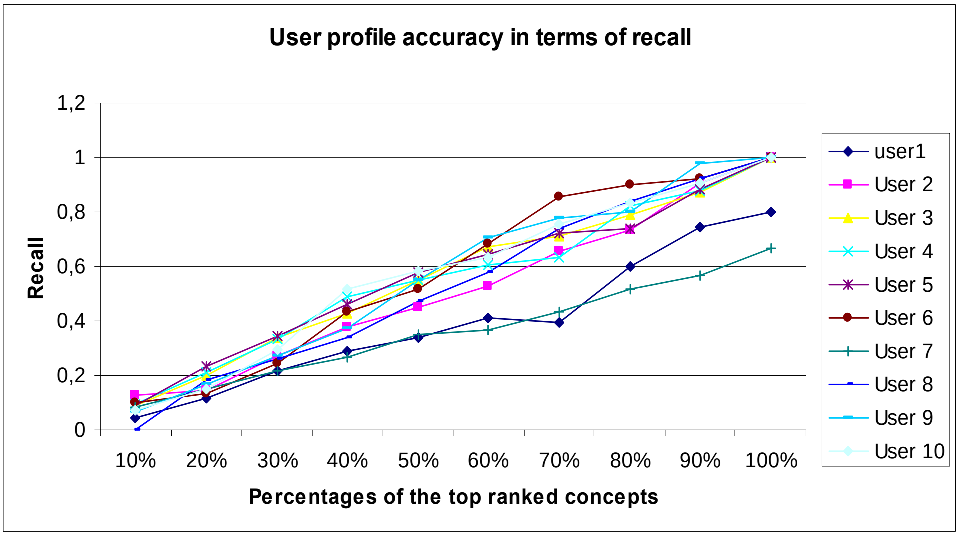

Figure 4: Recall calculated for each user at different concept cutoff points

Figure 5 shows the average F-measure calculated over all the users and considering different percentages of the top ranked concepts of the user profiles (10%, 20%, 30%, ..., 100%). Results confirm the increasing curve of the user profile accuracy where accuracy reaches the highest value when we consider all the concepts of the user profiles. Generally, the results obtained in this experiment deal with those obtained in [40] and justify the use of all the concepts of the user profile to calculate the personalized document score.

JIS Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

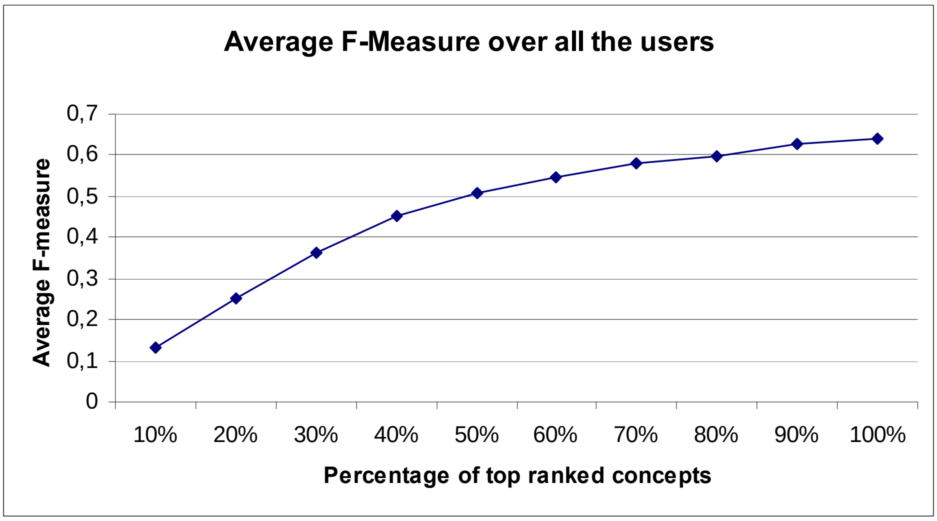

Figure 5: User profile accuracy in terms of average F-measure over all the users

6.2. Effect of the concept depth on the user profile accuracy

In this experiment, we study the effect of the concept depth of the reference ontology on the accuracy of the user profile, by considering depth 1, 2 and 3. For each user, we have filtered the concepts per depth and have applied the same evaluation protocol presented in section 5.3.1 for evaluating the user profile accuracy per depth.
The first level of the ODP ontology contains 15 concepts, which is the maximum number of concepts of this depth. As shown in figure 6, the average number of concepts in the profiles increases with the depth, allowing for an average of 9 concepts per user if one level is used, 90 concepts per user if two levels are used and 190 concepts per user if three levels are used.

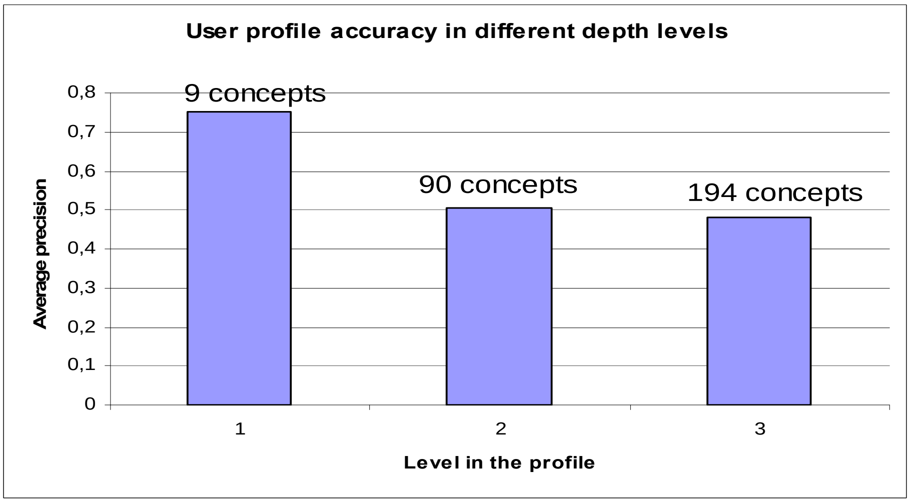

Figure 6: User profile accuracy in terms of F-measure averages over all the users

We can see that the precision at the first level is the highest one and it drops from 75% to 50.4% when the profile expands from one level to two. When the profile is expanded from two levels to three, the precision reveals a small drop (48.13%). As the drop in precision is small when the profile expands to deeper levels and the specificity of the user profile increases with the depth, we have used in our work concepts of general and specific levels to represent the user profile.

7. Personalized search effectiveness results

In this section, we present the results of evaluating the personalized search effectiveness. We consider the following objectives:

- Evaluating the effect of the document graph length, in terms of the number of concepts graph, on the personalized retrieval performances.
- Evaluating the effect of combining the original ranking and the personalized ranking on the retrieval effectiveness.
- Evaluating the effect of extending the document-profile subgraph for producing the personalized ranking.
- Evaluating the effectiveness of our personalized graph-based ranking model compared to other ranking models performed using different graph-based measures and vector space similarity measures.

7.1. Effect of document graph length on the retrieval effectiveness

In our personalized result reordering process, each document to be ranked is mapped on the ODP ontology to infer its concept graph. We evaluate the effect of the document graph length defined by the number of concepts used to represent the document graph on the personalized retrieval effectiveness, where the documents are ranked according to our graph-based ranking model. For this aim, we varied the number of concepts used to represent the document graph by considering 5, 10, 20, 30, 40, 50 and all of the obtained concepts. Figure 7 shows the average precision at the top 10 and the top 20 documents with varying the number of concepts used to represent the document graph.

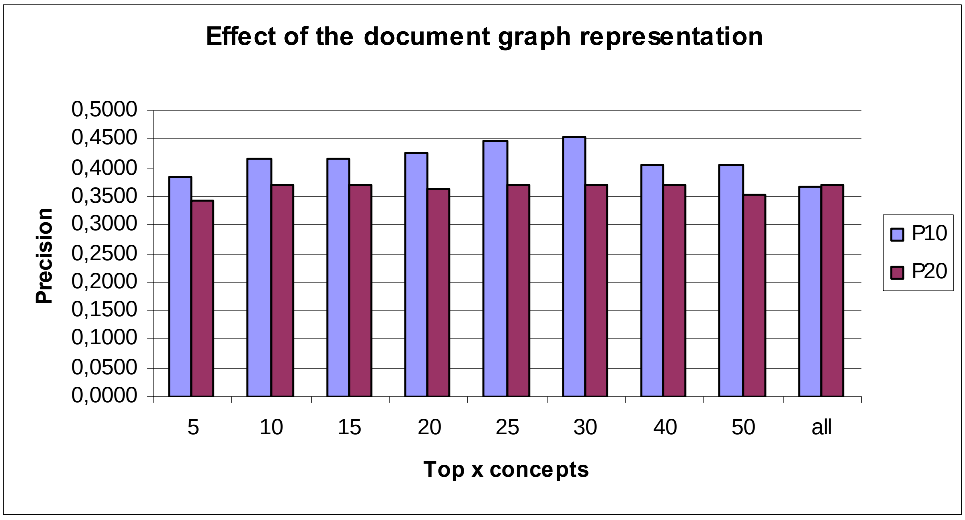

Figure 7: Effect of the document graph length on the personalized search effectiveness

Using few concepts or all the concepts gives the lowest precisions, while using 30 concepts is the best to perform the highest precision. This could be due to a poor representation of the document content in the first case or the spread of non relevant concepts in lower ranks in the second case. Hence, we retain 30 concepts to represent the document graph in the rest of our experiments.

7.2. Tuning the combination parameter between original and personalized document ranks

In this experiment, we study the effect of combining the original document's rank of Yahoo (corresponding to the initial document score (cf. algorithm 1) and the personalized document rank obtained according to our graph-based ranking model, on the retrieval effectiveness. The personalized document score is calculated using the distance between the extended graphs of the document and the user profile. Figure 8 shows the improvement of our personalized search in terms of P10 and P20 by varying the combination parameter γ in the interval [0 1].
Results show that the best performance is obtained when γ is 0, i.e., when the original search engine rankings are ignored altogether. This result conforms the experimental findings in several previous personalized approaches [40]. This is likely due to the fact that all the results on the top 50 match the query well and thus the distinguishing feature is how well they match the user profile.

JIS Accepted for Publication
By the Journal of Information Science: http://jis.sagepub.co.uk

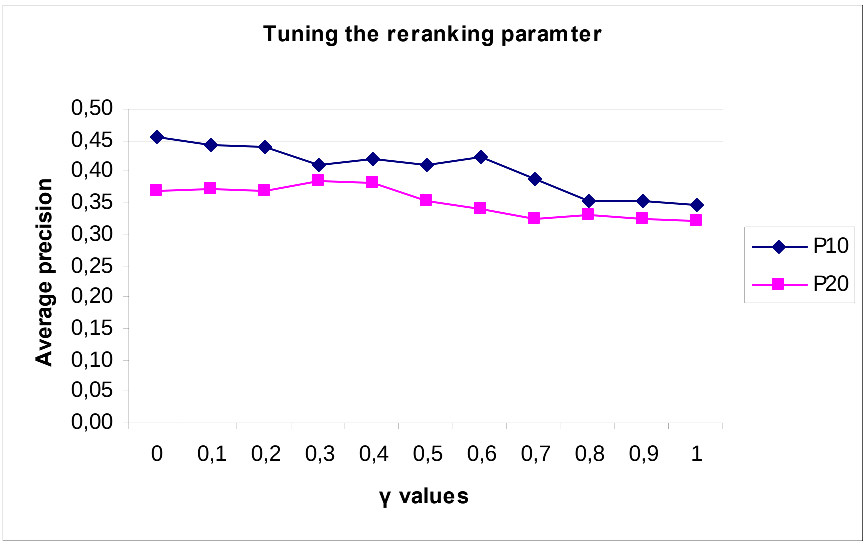

Figure 8: Effect of the reranking parameter on the final rank

7.3. Effect of the extended distance measure on the personalized search

In this experiment, we study the effect of extending the document-profile common subgraph, using the cross links on the personalized search effectiveness. Figure 9 shows the average precision of our personalized search in terms of P10 and P20 by giving static values for the decay factor f_{ca} between 0 and 1.

When $f_{ca} = 0$, the common subgraph contains only direct common concepts. A value of 1 gives the same weight for the activated concepts compared to the direct common ones. Figure 9 shows that there is not a cutoff value of the decay factor that produces high performances of the personalized search effectiveness. Hence we compare in Figure 10 our personalized search effectiveness using an adaptive decay factor in the distance measure (f_{ca} = LR/LR + 1) to the ranking model performed using the extended measure where we tune f_{ca} in [0 1]. Results show that extending the subgraph with activated concepts using an adaptive decay factor increases the precision P10, which proves that the extension of the distance measure enables to bring in the top of the result list presented to the user, the most related documents to the user profile. We have also calculated the Wilcoxon signed-rank test between each personalized ranking and the baseline search performed by Yahoo. We choose the Wilcoxon test as the main criteria for choosing the test are whether the sample set is sufficiently large as stated in [21].

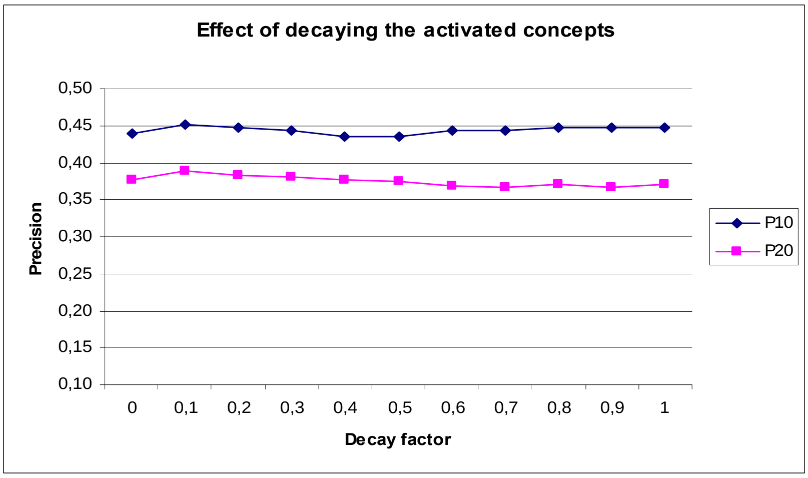

Figure 9: Effect of decaying the activated concepts of the document-profile common subgraph

As shown in Figure 10, our proposed document ranking model (Extension parameter corresponds to “Authomatic”) achieves the best performance at P10 and has shown a significant p-value according to the Wilcoxon test at P10, while all other personalized ranking models performed without extension or using static values of decay factor in [0 1] did not achieve a significant improvement according to the statistical Wilcoxon test. We also notice that using a small value for the extension parameter by giving more weights to the direct common concepts achieves higher precision than large values.

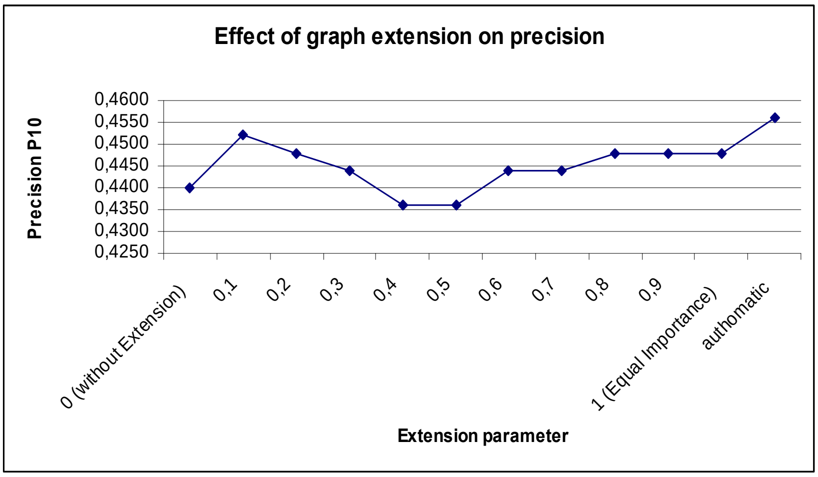

Figure 10: Effect of extending the common subgraph between the document and the user profile graphs

7.4. Comparative study with other personalized ranking models

We compare our personalized ranking model to other personalized ones defined as follows:

- Personalized ranking models based on using basic graph-based measures, namely, the distance measure based only on mcs, MCS or the distance measure combining mcs and MCS.
- Personalized ranking models based on using a vector-space similarity measure, namely the cosine measure, used in a keyword-based matching model [35] on one hand and a concept-based matching model on the other hand.

7.4.1 Comparing with graph-based ranking models

We compare our personalized ranking model to different ranking models based on using classical graph-based measures. Classical measures are: (a) the distance measure based on mcs (formula 2), (b) the distance measure using only MCS (formula 3), (c) the basic distance measure combining MCS and mcs (formula 4)

For all the graph-based measures, we calculate a normalized distance between the extended graphs of both the document and the user profile. We recall that for each testing query, the personalized search is based on considering the top 30 concepts to represent the document graph and reranking the top 50 results in descending order. Table 3 presents the improvement of each measure compared to Yahoo search. Since we have a limited distribution of precision values related to 25 queries. We assume that the difference between two rankings is significant if $p < 0.1$ (noted *) and very significant if $p < 0.05$ (noted **).

	P10	**P20**
Yahoo search (baseline)	0.3480	0.3220
Our semantic extended measure	0,4560	0,3700
Improvement	**31.03% ***	14.91%
Mcs	0,4480	0,3720
Improvement	28.74%	15.53%
MCS	0,4320	0,3720
Improvement	24.14%	15.53%
The basic combined measure MCS+mcs	0,4480	0,3700
Improvement	29%	15%

Table 3: Comparison of different personalized search ranking based on different document-profile similarity measures

Results show a significant improvement of our personalized search at P10. This indicates the effectiveness of ranking semantically the documents with respect to the user profile using our semantic graph-based distance measure. Indeed, our measure gives higher ranks for documents that are semantically related to the user profile by bringing them to the top 10 results presented to the user.

Compared to other distance measures used for producing the personalized ranking, we notice the following:

- The distance measure based only on the subgraph (mcs) or the supergraph (MCS) gives lower improvement compared to the combined measure especially at P10. This shows that the combined semantic distance at general and specific levels allows performing a better personalized search improvement.
- Compared to the basic combined distance measure (formula 3), our model performs better precision at P10. This proves the positive effect of using a decay factor calculated automatically depending on the cross links between graphs compared to the use of a decay factor that considers the activated concepts as the same manner as the direct common ones ($f_{ca} = 1$). This proves that the direct common concepts should be given more weights than the related ones.

In order to get a statistical view of the obtained results, we have also calculated the Wilcoxon test for each personalized ranking compared to the baseline search performed by Yahoo. Results show a significant test for only our model at P10, which proves the consistency of our measure compared to the other ones.

7.4.2 Comparing with vector-space ranking models

The goal of comparing our approach to vector space ranking models is to study (1) the effect of using concept-based matching rather than keyword-based matching for calculating the personalized document score (2) the effect of using a graph ranking model rather than a vector-space ranking model in a concept based matching model. The vector-space ranking models used in the comparative study and the obtained results are presented below.

Vector-space ranking models. We compare our personalized ranking model to two personalized ranking models performed using the cosine similarity measure. These models are (1) a keyword-based matching proposed in our previous works [35, 41], (2) a concept-based matching on the other hand.

- Keyword-based matching model: this model [35] is based on calculating the personalized score of the document using the cosine measure between the document and the most highly weighted concepts of the user profile. Documents and concepts of the user profile are represented by term vectors where terms are extracted respectively from the document content and the web pages classified under the concept of the reference ontology. Hence, the personalized document score is computed as follows:

$$S_p(d_k, G_u) = \frac{1}{h} . \sum_{j=1..h} score(c_j) * \cos(\overrightarrow{d_k}, \overrightarrow{c_j}) \quad (7)$$

  Where d_k and c_j are term-based vectors representing respectively document d_k and concept c_j, score(c_j) is the weight of concept c_j in the user profile G_u and h is the number of concepts considered in the personalized search.
  In order to set a reliable comparison between the different personalized ranking models, we have conducted a preliminary experiment in order to identify the best number of concepts used to calculate the personalized score of the document in formula 7. In this experiment, we calculate the average precision of the personalized search with varying the number of concepts used to calculate the personalized score of the document. We calculated the precision in terms of P10 and P20 using 3, 5, 7, 10 and 20 concepts. We observed that the best improvement (17, 24% at P10 and 8, 07% at P20) is obtained when using 7 concepts of the user profile. We retain this value for comparing the retrieval effectiveness of the keyword-based matching model using the cosine measure with other personalized ranking models.
- *Concept-based matching model*: This model is based on representing both the document and the user profile by concept-based vectors. To do so, each document is matched with each concept of the ontology using to the cosine similarity measure applied on their term-based representations. A set of weighted concepts of the ODP ontology is obtained to represent a concept-based vector representing the document. The weight of each concept is the accumulation of its similarity score with the mapped documents. Also, the user profile is represented as a vector of weighted concepts of the ontology, where concepts are derived from the user profile graph by keeping only the nodes. Calculating the personalized score of the document is based on using the cosine measure between the document and the profile concept-based vectors as follows:

$$S_p(d_k, G_u) = \cos(\overrightarrow{d_k}, \overrightarrow{G_u}) \quad (8)$$

where d_k and G_u are the concept-based vectors representing respectively the document and the user profile. The choice of this model is motivated by our interest to compare graph-based ranking to vector space ranking at concept level.

Results. Table 4 presents the precision at 10 and 20 documents for our personalized search model compared to the keyword-based matching model and the concept-based matching model mentioned above using the cosine measure. According to the results, we notice the following:

- The keyword-based matching model performs a reasonable improvement at P10. However, it performs the lowest improvement at P20 where the improvements at both P10 and P20 are not significant according to the Wilcoxon test. This shows the effectiveness of using semantic resources for representing both the documents and the user profile and then perform a matching based on these representations.
- Compared to the concept-based matching model, the precision improvement of our model is significant according to the Wilcoxon test at P10, while it is not for the concept-based matching at the same level. The inverse is noticed at P20. According to these findings, we can assume that the performance of our model seems comparable to the concept-based matching model. The high performance of the concept-based matching model is not surprising due to the importance of the concept weights in both the document and the user profile for calculating the personalized score of the document.

	P10	**P20**
Yahoo search	0.3480	0.3220
Our personalized search	0.4560	0.3700
Improvement	31.03% *	14.91%
Cosine (term-based matching)	0.4080	0.3480
Improvement	17.24%	8.07%
Cosine (concept-based matching)	0.4640	0.4160
Improvement	33.33%	29. 19% **

Table 4: Comparison of different personalized search ranking models based on different document-profile similarity measures

In order to get a more detailed understanding of the effects of our personalized search compared to the concept-based matching using vector space model, we examined the results on a query by query basis at P10. Figure 11 presents the performance variability per query for our personalized search compared to the concept-based matching model.

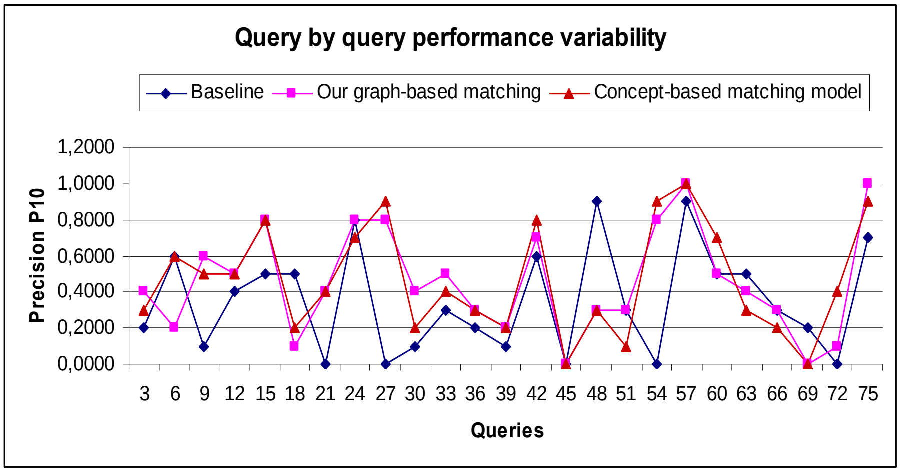

Figure 11: Query by query performance variability.

For the 25 testing queries reranked using our graph-based measure, 15 (60%) showed an improvement, 5 (20%) were negatively impacted and 5 (20%) were unchanged. According to concept-based matching, among the 25 testing queries, 16 (64%) showed an improvement, 7 (28%) were negatively impacted and 2 (8%) were unchanged. Thus, our personalized reranking helped 3 times as many queries as it hurts while concept-based matching helpe only 2 times as many queries as it hurts. This shows that our approach outperforms the concept-based matching model using vector-space model in terms of robustness.

In general, the variability of the performance between queries is probably due to the user profile accuracy and the difference of the clarity degree between the queries. Indeed, we outline that testing queries are more precise than training queries as each testing query is the last one formulated in a search session. This would reduce the impact of personalization on the retrieval effectiveness compared to the use of ambiguous or difficult queries. In conclusion, compared to other ranking models, our model presents the benefit of: (1) conceptualizing the similarity between the document and the user profile according to the distance of their graph structures in the ontology, (2) excluding non relevant documents represented by irrelevant topics/concepts, which are generally distant from the concepts of the user profile in the ontology. Indeed, creating the supergraph is done by connecting the irrelevant concepts of the documents to those in the user profile according to the ontology, which causes a big MCS and a small mcs and lead to have a low personalized document score.

8. Conclusion and outlook

We have presented in this paper an approach for personalizing search integrating a graph-based document ranking model. Personalization is achieved by reranking the search results using an extended graph-based distance measure combining MCS and mcs.

Compared to other related works that model ontology-based user profiles [9, 4, 3], the key differentiating aspect in our approach is the use of a personalized graph-based ranking model that takes into account the semantic relationships between concepts of the document and the user profile for reordering the search results. The benefit of such model is twofold: (1) representing semantically the document content by a graph of concepts rather than terms (i.e. keywords), which excludes noise terms that usually affect the term-based document-profile matching, (2) calculating a semantic graph-based distance between the document and the user profile graphs by integrating an approximate matching level considered through the cross links between graphs.

The experimental evaluation of our approach was done on real user data provided from Exalead web search log. Results show a significant precision improvement of our personalized search compared to the basic search performed by Yahoo. This shows the effectiveness of the user profile modeling and the effectiveness of the personalized search using a semantic graph-based distance measure where exploiting the semantic relationships between concepts enhances the quality of the search. We can confirm also that using few training queries and exploiting few resources for creating the user profile in a search session (the clicked web pages) allowed us to perform an effective personalized search.

The comparative study with other ranking models has shown that our model outperforms basic graph-based ranking models and the vector-space ranking models where documents are represented by term vectors. The performance of our model seems comparable to the vector-space matching model where the documents and the user profile are represented by vectors of concepts as the use of the concept weights has shown its importance. However, we believe that graph-based distance measure could perform better when weighted nodes are considered in the graph matching.

In future work, we plan to (1) study the effect of the concept depth in the document and the user profile graphs on the effectiveness of the personalized graph-based ranking model, (2) study the effect of extending the subgraph mcs between the document and the user profile graphs, using a spreading activation at multiple hops through the cross links, (3) exploit a weighted graph-based distance measure that takes into account the weight of the concepts in the document and the user profile for producing the personalized document ranking.

9. Acknowledgments

This work was in part supported by OSEO under the Quaero program.

10. References

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

[2] A. Sieg, B. Mobasher, R. Burke, G. Prabu, S. Lytinen, Using concept hierarchies to enhance user queries in web-based information retrieval, in: The IASTED International Conference on Artificial Intelligence and Applications. Innsbruck, Austria, 2004.
[3] F. Liu, C. Yu, W. Meng, Personalized web search for improving retrieval effectiveness, IEEE Transactions on Knowledge and Data Engineering 16 (1) (2004) 28–40.
[4] S. Gauch, J. Chaffee, A. Pretschner, Ontology-based personalized search and browsing, Web Intelli. and Agent Sys. 1 (3-4) (2003) 219–234.
[5] G. Koutrika, Y. Ioannidis, A unified user profile framework for query disambiguation and personalization, in: Proceedings of Workshop on New Technologies for Personalized Information Access, 2005.
[6] L. Tamine-Lechani, M. Boughanem, N. Zemirli, Personalized document ranking: exploiting evidence from multiple user interests for profiling and retrieval, In Journal of Digital Information Management.
[7] W. Fan, M. D. Gordon, P. Pathak, Discovery of context-specific ranking functions for effective information retrieval using genetic programming, IEEE Trans. on Knowl. and Data Eng. 16 (4) (2004) 523–527.
[8] C. Lin, G. Xue, H. Zeng, Y. YU, Using probabilistic latent semantic analysis for personalised web search, in: Proceedings of the APWeb conference, 2005, pp. 707–711.
[9] A. Sieg, B. Mobasher, R. Burke, Web search personalization with ontological user profiles, in: CIKM'07: Proceedings of the sixteenth ACM conference on information and knowledge management, ACM, New York, NY, USA, 2007, pp. 525–534.
[10] S. Dumais, E. C. J. J. Cadiz, G. Jancke, R. Sarin, C. R. Daniel, 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 (SIGIR'03), ACM Press, 2003, pp. 72–79.
[11] X. Shen, B. Tan, C. Zhai, Context-sensitive information retrieval using implicit feedback, in: SIGIR '05: Proceedings of ACM SIGIR conference, ACM, New York, NY, USA, 2005, pp. 43–50.
[12] G. Webb, M. Pazzani, D. Billsus, Machine learning for user modeling, User Modeling and User-Adapted Interaction 11 (1-2) (2001) 19–29.
[13] C. Ding, J. C. Patra, F. C. Peng, Personalized web search with selforganizing 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, 2005, pp. 144–147.
[14] B. Tan, X. Shen, C. Zhai, Mining long-term search history to improve search accuracy, in: KDD '06: Proceedings of the 12th ACM SIGKDD international conference, ACM, New York, NY, USA, 2006, pp. 718–723.
[15] H. Lieberman, Autonomous interface agents, in: CHI, 1997, pp. 67–74.
[16] R. W. White, P. N. Bennett, S. T. Dumais, Predicting short-term interests using activity-based search context, in: J. Huang, N. Koudas, G. Jones, X. Wu, K. Collins-Thompson, A. An (Eds.), CIKM, ACM, 2010, pp. 1009–1018.
[17] I. M. Begg, J. Gnocato, W. E. Moore, 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, New York, NY, USA, 1993, pp. 211–214.
[18] A. Micarelli, F. Sciarrone, Anatomy and empirical evaluation of an adaptive web-based information filtering system, User Modeling and User-Adapted Interaction 14 (2-3) (2004) 159–200.
[19] H. R. Kim, P. K. Chan, Learning implicit user interest hierarchy for context in personalization, in: IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces, ACM, New York, NY, USA, 2003, pp. 101–108.
[20] C. P. Alexandru, N.Wolfgang, P. Raluca, K. Christian, 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, 2005, pp. 178–185.
[21] S. Staab, R. Studer, Handbook on Ontologies (International Handbooks on Information Systems), SpringerVerlag, 2004.
[22] Z. Ma, G. Pant, O. R. L. Sheng, Interest-based personalized search, ACM Trans. Inf. Syst. 25 (1) (2007) 5.
[23] G. Jeh, J. Widom, Scaling personalized web search, in: WWW '03: Proceedings of the 12th international conference on World Wide Web, ACM, New York, NY, USA, 2003, pp. 271–279.
[24] J. Rocchio, Relevance feedback in information retrieval, in: The SMART retrieval system-experiments in automatic document processing, ACM, New York, NY, USA, 1971, pp. 313–323.

[25] F. Boubekeur, M. Boughanem, L. Tamine, Semantic Information Retrieval Based on CP-Nets,, in: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), London, 23/07/07-26/07/07, IEEE.

[26] M. Baziz, M. Boughanem, Y. Loiseau, H. Prade, Fuzzy Logic and Ontology-based Information Retrieval, in: P. Wang, D. Ruan, E. Kerre (Eds.), Studies in Fuzziness and Soft Computing, Vol. 215/2007, Springer, 2007, pp. 193–218.

[27] W.-T. Cai, S.-R.Wang, Q.-S. Jiang, Address extraction: a graph matching and ontology-based approach to conceptual information retrieval, in: Proceedings of the Third International Conference on Machine Learning and Cybernetics, 2004, pp. 1571–1576.

[28] J. Zhong, H. Zhu, J. Li, Y. Yu, Conceptual graph matching for semantic search, in: ICCS '02: Proceedings of the 10th International Conference on Conceptual Structures, Springer-Verlag, London, UK, 2002, pp. 92–196.

[29] J. Yi, F. Maghoul, Query clustering using click-through graph, in: WWW '09: Proceedings of the 18th international conference on World wide web, ACM, New York, NY, USA, 2009, pp. 1055–1056.

[30] H. Cui, J.-R. Wen, J.-Y. Nie, W.-Y. Ma, Query expansion by mining user logs, IEEE Trans. on Knowl. and Data Eng. 15 (4) (2003) 829–839.

[31] Y. Huang, L. Sun, Z. Wang, A unified graph-based iterative reinforcement approach to personalized search, in: AIRS '09: Proceedings of the 5th Asia Information Retrieval Symposium on Information Retrieval Technology, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 193–204.

[32] X. Wan, J. Yang, J. Xiao, Towards an iterative reinforcement approach for simultaneous document summarization and keyword extraction, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 552–559.

[33] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (5) (1999) 604–632.

[34] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. 30 (1-7) (1998) 107–117.

[35] M. Daoud, L. Tamine, M. Boughanem, Towards a graph based user profile modeling for a session-based personalized search, Knowledge and Information Systems 21:3 (2009) 365–398.

[36] M.-L. Fern´andez, G. Valiente, A graph distance metric combining maximum common subgraph and minimum common supergraph, Pattern Recogn. Lett. 22 (6-7) (2001) 753–758.

[37] G. Levi, A note on the derivation of maximal common subgraphs of two directed or undirected graphs, Calcolo 9 (1972) 341–354.

[38] H. Bunke, X. Jiang, A. Kandel, On the minimum common supergraph of two graphs, Computing 65 (1) (2000) 13–25.

[39] O. Bouidghaghen, L. Tamine, M. Daoud, C. Laffaire, Contextual evaluation of mobile search (regular paper), in: B.-L. Doan, J. Jose, M. Melucci, L. Tamine (Eds.), Workshop on Contextual Information Access, Seeking and Retrieval Evaluation, Milton Keynes, 28/03/2010-28/03/2010, Vol. 569

[40] J. Trajkova, S. Gauch, Improving ontology-based user profiles, in: C. Fluhr, G. Grefenstette, W. B. Croft (Eds.), RIAO, CID, 2004, pp. 380–390.

[41] M. Daoud, L. Tamine, M. Boughanem, B. Chebaro, A Session Based Personalized Search Using An Ontological User Profile, in: ACM Symposium on Applied Computing (SAC 2009), Haiwai (USA), ACM, 2009,pp. 1031–1035.

[42] Mariam Daoud, Lynda Tamine, Mohand Boughanem. A personalized graph-based document ranking model using a semantic user profile (regular paper). In Conference on User Modeling, Adaptation and Personalization (UMAP 2010), Hawai, 20/06/2010-24/06/2010, Springer, p. 171-182, 2010.

[43] Calegari, S., & Pasi, G. (2010). Ontology-Based Information Behaviour to Improve Web Search. Future Internet, 2(4), 533-558.

[44] Loïc Maisonnasse, Jean-Pierre Chevallet, Catherine Berrut. Incomplete and Fuzzy Conceptual Graphs to Automatically Index Medical Reports. NLDB'2007. pp.240-251.

[45] Chris Buckley and Ellen M. Voorhees. 2000. Evaluating evaluation measure stability. In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '00). ACM, New York, NY, USA, 33-40.

