Using a concept-based user context for search personalization

Mariam Daoud, Lynda Tamine-Lechani, Mohand Boughanem
{daoud, lechani, bougha}@irit.fr
Paul Sabatier university
IRIT-SIG, 118 Narbonne street, Toulouse, France

Abstract—**Because of the diversity of the user interests and the ambiguity of the user query, current search engines are not very effective. Indeed, they are based on simple query-document matches without considering the user background and interests. Personalized search aims at integrating the user context, defined as a set of user's topics of interests, in the information retrieval (IR) process in order to tailor search results to a particular user. An effective personalization is achieved when an accurate representation of the user context is provided. We present in this paper our approach for learning long term user interests by collecting information from the user's feedback and using existing domain ontology. The learning process is based on the aggregation of the short term user contexts represented as a set of general concepts, where the user context reflect the user's topics of interest in a specific search session. Personalization is achieved by using the user contexts across related search sessions. Our experimental results carried out in TREC collection show that re-ranking the search results based on the concepts weights of the short term user context brings significant improvements in the retrieval precision.**

Keywords: user context, user interest, ontology, personalization

I. Introduction

Current web search features, characterized by information overload and short or ambiguous queries, make the traditional IR systems unable to satisfy the user information needs. Indeed, users have generally different information needs when searching information on the web; by submiting the same query, the search engine returns the same set of results based solely on the user query. Studies in [2] show that the main reason is that traditional search technologies do not take into account the user context in the retrieval process. For example, if a user working in computer science formulates the query "*Java language*", the documents on "*Java island*" will be incorrectly favored. Contextual IR becomes a promising area for disambiguating such web search and improving retrieval effectiveness. In [11] contextual IR is defined 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*. While there are many contextual factors in IR (the user's interests, preferences such as document freshness or language, physical context factors in mobile environment, etc.), the user domains of interest is the most important contextual factor identified that alleviates an ambiguous web search in an ad hoc retrieval task [1]. Distinction has been made between long term and short term user interests [12]. While the long term user interests reflect general domains of interest for the user, the short term ones reflect a specific domain of interest for a user when searching for a specific information need. Since the user is reluctant to provide explicitly information about his personal interests, as in personalized google, MyYahoo, InfoQuest, implicit feedback has attracted much attention recently [14][3] in user profile modeling. By collecting information about the user's needs during search interaction and browsing, users are modeled by their personal profile that reflects a set of user interests. A challenge in personalized search is how to infer the user profile implicitly based on user's ongoing behavior, and how to represent it accurately?

Earlier works in adaptive search systems as Grouplens [4] model a group of users using a collaborative profile, and return the search results for a user according to the profile of the group for which he belongs to. Since the collaborative approaches induce problems in large scale applications, works in personalization converge to the user-oriented based search. Recommendation systems like "*LETIZIA*" [6], "*JITIR*" [22] are types of adaptive systems that exploit information collected from emails or pages viewed by the user to represent the short term user context as being the current user intention, and propose proactively to the user relevant information according to his current task. More recent approaches aim to model more precisely the user profile; while some works use only the user feedback to build the user profile as a set of class vectors [23] or term relations [10], others [7] [9] use a domain ontology as an additional source of evidence to build a semantic representation of the user profile.

In this paper, we address the problem of learning the user profile within the user's ongoing behaviors by using the user feedback and the ODP domain ontology. We learn the long term user interests based on the aggregation of the short term user contexts extracted from the search sessions. Our short term user context is represented by a set of weighted concepts that represent the user's topics of interest at a specific search session. We do not address in this paper the problem of session boundaries by assuming that there is a session boundaries delimitation function that measures the relatedness between search sessions. Search sessions are related in the sense that their user queries are related to the same topic of interest. More precisely, our method runs in two main steps; the first one consists of representing the concept-based user context by mapping a keyword weighted vector, inferred from the user feedback at each retrieval session, on the reference ontology. The second step consists of maintaining the concepts weights across related search sessions. This paper is organized as follows: in the next section, we review some related works and outline our motivation. Section 3 presents our approach for representing the concept-based user context. Section 4 presents our maintaining process of the user context across related search sessions. In section 5, we present our method for exploiting the user context in personalized search. The experimental evaluation and results are presented in section 5. The last section presents our conclusion and points out possible direction for future work.

II. Related works and motivation

The goal of personalized IR is to return search results that better match the user intent. Therefore, personalization requires a user profile modeling component and a personalized IR model that exploit the user profile by means of query reformulation [7], query-document matching [15] or result processing [9],[18]. We review in this section some related works in user profile modeling and show how search personalization is achieved.

Multiple model representation of the user interests are addressed in numerous user profiling approaches. User interests are often represented as a set of keyword vectors [6] [12] or class vectors [23]. In one hand, keyword based representation is generally seen as a flat of words and does not allow to capture the semantic behind the concepts. In another hand, learning the keyword-based user interests is a time consuming process which consists of collecting information across multiple search sessions to finally create the user profile. Indeed, the evidence collected solely from the user does not allow the system to determine the user intention when a new search topic is encountered. When using existing domain ontology, the system can match evidence gathered from the user feedback with concepts of the reference ontology and therefore represent new user's topics of interest. Hence, modeling the user profile using existing domain ontology has become an interesting direction in personalized search. The ARCH system [7] is a personalized IR system that enhance the user query using both of user profile which contains long term user interests, and the yahoo concept hierarchy. The system learns the long term user's contexts that represent the user information needs across the search sessions; each context is represented as a set of pairs by encapsulating the selected concepts and the deselected concepts that are respectively relevant and irrelevant to a specific user information need at a specific search session. Liu and al. [18] build a user profile that consists of a set of categories from a concept taxonomy based on the user's search history. The user profile is then used to map the user's search query onto three depth-two concepts of the ontology. Personalization consists then of categorizing the search results according to the query related categories and then re-ranking the search results using a voting based merging scheme. The OBIWAN system [9] also learns automatically an ontological user profile by assigning weights to existing concepts in the Magellan taxonomy. The learning process consists of mapping each visited Web page into five taxonomy concepts with the highest similarities; the user profile consists then of a list of concepts for which the weights are accumulated based on user's browsing behaviors. This user profile is used to re-rank the search results by combining the original rank of the document and the conceptual rank computed using a similarity between the document and the user profile. An interest-based personalized search in [17] consists of mapping a set of known user interests onto a group of categories in the concepts taxonomy and therefore categorize and personalize search results according to the mapped categories associated to these user interests.

Comparatively to these previous works, our approach consists of representing the user interests, each one as a set of semantically related concepts of reference ontology, while all possible user interests are represented in [9] over all the concepts in the ontology. Another distinctive aspect for our approach is that instead of mapping the web pages browsed by the user as in [9], we map a keyword user context derived automatically from the user feedback onto the ontology. While user interests are mapped in [17] on the ontology as keyword vectors, we note that their representation cannot be derived automatically in such a way that they are far from real world applications.

III. Representing a concept-based user context using the depth three of reference ontology

Our goal is first, to provide the semantic representation of the user context that represents a short term user interest related to a specific search session. Such contexts are then aggregated in order to learn the long term interests. Our method for representing the user context runs in three main steps: (1) representing a keyword user context derived from the user feedback, (2) mapping the keyword user context on the ODP ontology, (3) disambiguating the mapped concepts set using a sub-concepts aggregation scheme, and finally representing the user context by the depth three concepts of the resulting set. The main reason for representing the user context using the depth three of the ontology is that we are interested to represent a user context gathering information fairly general, and that can be able to improve retrieval precision for related search sessions.

A. Representing the keyword user context

A keyword user context reflects a short term user interest in a specific search session. It is represented using the most representative terms derived from the assumed relevant documents in a particular search session. Especially, let q^s be the query submitted by a specific user at the retrieval session S^s performed at time s. We assume that a document retrieved by the search engine with respect to q^s is relevant if it generates some observable user behaviors (page dwell time, click through, saving, printing etc). Let D^s be the related set of assumed relevant documents during the session S^s. Each document d of D^s is represented by a term vector where the relevance value of the term t in document d at time s is computed using the $tf * idf$ weighting scheme as follows:

$$w_{td} = tf(t,d) * log(\frac{n}{n_t}) \quad (1)$$

Where
$tf(t,d)$: the frequency of the term t in d,
n : the number of documents in the collection,
n_t: the number of documents containing the term t.
The keyword user context K^s represent then the centroid of the documents in D^s. The term's weight in the user context is then computed as follows:

$$K^s(t) = \frac{1}{|D^s|} \sum_{d \in D^s} w_{td} \quad (2)$$

B. Mapping the keyword user context on the reference ontology

Once we had the keyword user context, we map it on the ontology in order to extract the most relevant concepts. These mapped concepts are used later to represent concepts of depth three of the user context. Mapping the keyword user context on the ontology requires an aggregate representation of the reference ontology by computing a term vector for each of its concepts.
There are many domain ontology created manually and designed to organize web content for easy browsing by end users. We cite the online portals such as yahoo[1], Magellan[2] and the open directory project[3]. At 31 august 2007, the ODP is a manually edited directory of 4,83 millions URLs that have been categorized into 787774 categories[4]. Since that the Open Directory Project (ODP) is the most widely distributed data base of Web content classified by humans, we use it to get a concept-based representation of the user context. We took advantage of the ODP metadata associated for each concept to represent each one in the vector space model. Indeed, each ODP concept contains some related URL links classified by human editors under this concept, and is related to sub-concepts with "*is-a*" relations. Each of the URL links is annotated by a title and a description that represent the content of the related web page. For each concept, we only use the titles and descriptions of the first 60 URL considered as sufficient data to get an accurate classification of the keyword user context. Works in [19] and [20] use similar concatenation to build topic profiles. In support of our approach, study in [13] proves that using the manually annotated titles and descriptions of the URL in the concept description vector achieves higher classification accuracy than the use of the web pages contents. The procedure for getting the representation of the ODP concepts is explained in details in a previous work [16]. Briefly, each concept in the ontology is represented by a term vector which represents URL links indexed under that concept and also of its sub-concepts as explained below:

1) concatenating the first 60 URL 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,
3) representing each of the ODP concepts C_j, having an associated super-document Sd_j, by a term-based vector V_j computed using the following weighting scheme:

$$w_{ij} = ttf_{ij} * \log(\frac{N}{N_i}) \quad (3)$$

where
N is the number of super-documents,
N_i is the number of super-documents containing the term t_i,
ttf_{ij} is the total frequency of the term t_i in the super-document Sd_j and also in each of the super-documents Sd_k, where C_j has *n* related sub-concepts C_k, each one is represented by Sd_k:

$$ttf_{ij} = \left[(tf_{ij} + \sum_{k=1..n} tf_{ik}) \right] / (n+1) \quad (4)$$

For experimental purposes, we map the keyword user context K^s up to depth five of the ontology using the cosines similarity. Given a concept C_j in the ontology, represented by the term vector V_j, its weight is computed as follows:

$$p(C_j) = cos(V_j, K^s) \quad (5)$$

We note that the mapped concepts set may contain some irrelevant elements that do not reflect the user's search intention. This can be explained by the fact that keyword context terms can be matched to multiple concepts belonging to different portions of the ontology. Indeed, a specific user topic of interest is not exactly matched with a unique portion of the ontology, but it can be represented by concepts extracted from different ones. In the next section we detail our method for disambiguating the mapped concepts set in order to select the most important ones as a set of depth three concepts describing the user context.

C. A sub-concepts aggregation scheme for disambiguating mapped concepts

We aim in this section to represent the user context with general depth three related concepts issued from the ontology. We outline that the depth two of the ontology is too general to represent the user's topic of interest, and leaf nodes are too specific to improve retrieval precision for related search sessions. Our method of disambiguation is based on the assumption that relevant concepts of depth three are those having greater number of related concepts according to the ontology. Thus, aggregating the related concepts weights belonging to each general concept allows to assign to the

[1] http://www.yahoo.com
[2] http://www.mckinley.com
[3] http://www.dmoz.org
[4] http://www.aef-dmoz.org/blog/l-odp-francophone-en-aout-2007/

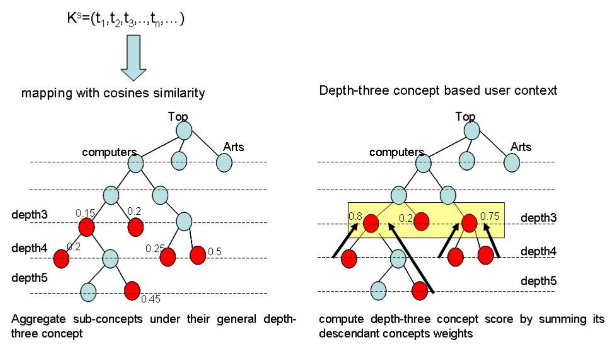

Fig. 1. Disambiguating the mapped concepts of the user context

relevant concepts higher weights. More precisely, as shown in fig.1, we identify a cluster of weighted concepts having a common general depth three concept, the relevancy score of the general concept for each cluster is computed by adding the weights of its related concepts. Thus, greater clusters will be assigned by high weights and will be ordered in the top of the concepts describing the user context. The weight of a general concept C_i having n descendant concepts C_k is computed as follows:

$$p(C_i) = \frac{\sum_k p(C_k)}{n} / \ C_k \in \{Descendant(C_i)\} \quad (6)$$

We finally create the concept-based representation of the user context C^s, related to the search session S^s, based on depth three weighted concepts of the clusters identified as explained above.

IV. Maintaining the user context across related search sessions

Based on user's ongoing behaviors, we maintain the user context by updating its concepts weights across related search sessions using a linear combination formula. Dynamic changes of the user interests lead to differences in the relevant concept rankings of the user context. Let C^{s-1} and C^s be respectively the user contexts for successive and related search sessions S^{s-1} and S^s. Maintaining method is based on the following principles: (1) enhance the weight of possible common concepts that can appear in two successive user contexts. (2) alter the weight of non common concepts using a decay factor β. The new weight of a concept C in the user context C^s is computed as follows:

$$p_{new}(C) = \begin{cases} \beta * p_{C^{s-1}}(C) + (1-\beta) * p_{C^s}(C) \\ if C \in C^{s-1} \\ \beta * p_{C^{s-1}}(C) \ otherwise \end{cases} \quad (7)$$

Where
$p_{C^s}(C)$ is the weight of the concept C in C^s,
$p_{C^{s-1}}(C)$ is the weight of the concept C in C^{s-1}
Such user contexts updated across related search sessions are used further to learn the long term user interests. The learning process is based on session boundaries delimitation mechanism that measures the semantic correlation between search sessions and decides their relatedness.

V. A context sensitive personalized search

At this level, we detail our technical method to personalize the search results for related search sessions using the user context. Let's consider the user context C^s performed at time s, containing an ordered set of weighted concepts $\langle C_j, p(C_j)\rangle$, given a related query q^{s+1}, we re-rank its associated search results by combining for each retrieved result d_k, the initial score returned by the system $Score_i$ and a contextual score computed based on similarity measure between the result and the concepts of the user context. Contextual score $Score_c$ of the result d_k is computed as follows:

$$Score_c(d_k, C^s) = \sum_{C_j \in C^s} p(C_j) * cos(d_k, C_j) \quad (8)$$

Where
C_j : a concept in the user context,
$p(C_j)$: the weight of the concept C_j in the user context C^s.
The final result score is then computed by combining its original score and its contextual score, then the results are re-ranked based on their final score computed as follows:

$$Score_f(d_k) = \gamma * Score_i(q, d_k) + (1-\gamma) * Score_c(d_k, C^s) \quad (9)$$

$$0 < \gamma < 1$$

VI. Experimental evaluation

The goal of our experimental evaluation is to show that re-ranking with the concept-based user context leads to significantly higher retrieval performances comparing with a basic search.

A. Experimental data sets

In order to evaluate the effectiveness of our model, we used two data sets in our experiments: the first one is provided by TREC collection and the second one is the ODP data set that we created to represent each of the ODP concepts.

1) TREC data set: as a main test data, we used a TREC data from disks 1&2 of the ad hoc task containing 741670 documents. We particularly tested the queries among $q_{51} - q_{100}$.

The choice of this test collection is due to the availability of a manually annotated domain for each query. This allows us to enhance the data set with simulated user interests associated for each of annotated TREC domains. We simulate six domains of TREC including 25 queries shown in table *I*. In order to map the query domains to realistic user interests, we applied our method for representing the user context for each query using its 30 relevant documents provided by TREC collection.

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

TABLE I
TREC DOMAINS USED FOR SIMULATING THE USER INTERESTS

2) ODP data set: the second data set is the document collection that we created from the ODP ontology. It contains 235331 concepts used to represent the user context in a particular search session. We used "*Mercure*" search engine to index the collection of super-documents.

B. Evaluation methodology

Our evaluation methodology consists of evaluating the effectiveness of our model when using the user context in the IR model for related search sessions. We take advantage from using the TREC test collection and assume that user contexts generated for queries of the same TREC domain, imply that associated search sessions are related to the same topic of interest. The evaluation scenario is based on the k-fold cross validation explained as follows:

1) for each simulated TREC domain, divide the query set into k equally-sized subsets, and using k -1 subsets for learning the user interests and the remaining subset as a test set,
2) for each query in the training set, an automatic process generates the associated keyword user context based on its 30 relevant documents, and then map it on the ODP ontology and extract the concept-based one,
3) update the user context concept weights associated to the queries in the training set and use it for re-ranking the search results of the queries in the test set.

C. Evaluation results

We conducted two sets of controlled experiments to examine the effectiveness of our personalization approach. We measure the effectiveness of re-ranking search results in terms of Top-n precision (P@5, P@10) and Mean average precision (MAP) metrics. We conduct our experiments using "*Mercure*" as a typical search engine based on the OKAPI retrieval model, and our personalized search model is based on a re-ranking module.

1) Effect of γ on precision improvement: in this experiment, we study the effect of the parameter γ in the re-ranking formula (9) on the precision improvement of the personalized search over all the simulated domains. We fix the decay factor used for updating the concept weights across related search sessions at 0,2 in the formula (7). We present in the figure (2) the precision improvement graph obtained for the personalized system versus the typical system at each of the precisions P@5, P@10 and MAP averaged over the queries belonging to the same domain. We conclude that the 0,3 value of γ produces the better improvement in personalized search. Small values of γ (0,1 or 0,2) do not give optimal performances. Indeed, small values of γ decrease strongly the original score returned by the system and give high confidence for the contextual score. While the ODP concepts used for re-ranking are fairly general, it is difficult to reach an optimal precision for related queries. In another hand, higher values of γ (over than 0,5) do not give better precision improvement versus the 0,3 value of γ, this proves that favoring contextual score when combining it with the original one allows to reach an optimal precision improvement.

2) Retrieval effectiveness: in this experiment, we evaluate the effectiveness of the personalized search over the various simulated user interests by comparing the baseline model with the personalized one. We show in the table *II* the retrieval performance measured in terms of P@5, P@10 and MAP averaged over the queries belonging to the same domain. We fix γ at the best value occurring at 0,3 in the re-ranking formula (9). We see that the personalized search improves the retrieval precision of almost the queries in the six domains simulated. The precision improvement varies however from a domain to another. This is probably due to the accuracy level of the user context representation in one hand, and the correlation degree between the queries of the same domain in another hand. For example, in the environment domain of TREC, some queries are related to the environmental concepts of the ODP, while a specific query (q_{59}) related to weather has no match with the set of the environmental concepts.

VII. Conclusions and outlook

In this paper, we described our approach for learning long term user interests based on modeling concept-based user contexts identified within related search sessions. We used the depth three of the ODP ontology to represent the user contexts. Unlike most previously related works, we focus on learning the user interests, each one represented as a set of semantically related concepts. Maintaining the concept weights of the user context is achieved across related search sessions based on a linear combination formula. Our experimental results show that re-ranking the search results using the user context achieves improvement of the retrieval precision.

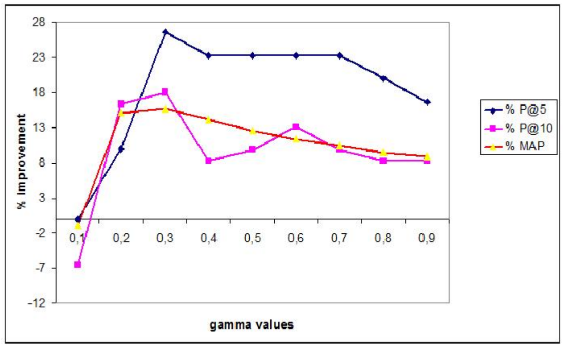

Fig. 2. Precision improvement graph for personalized search

	Baseline			Our model		
Domain	p@5	p@10	map	p@5	p@10	map
Environement %improvement	0,25	0,32	0,18	0,35 **+40%**	0,37 **+15,38%**	0,19 **+1,73%**
Military %improvement	0,25	0,27	0,05	0,35 **+40%**	0,32 **+18,18%**	0,07 **+46,46%**
Law & Gov %improvement	0,40	0,42	0,12	0,50 **+25%**	0,45 **+5,88%**	0,14 **+12,33%**
Inter. Rel. %improvement	0,16	0,12	0,01	0,16 0%	0,16 **+33,33%**	0,02 **+36,59%**
US Eco %improvement	0,26	0,30	0,09	0,33 **+25%**	0,36 **+22,22%**	0,10 **+8,35%**
Int. Pol %improvement	0,16	0,10	0,05	0,20 **+25%**	0,16 **+60%**	0,07 **+42,26%**

TABLE II
RESULTS EFFECTIVENESS OF OUR APPROACH

In our future work, we plan to learn the user profiles reflecting the diversity of the user interests. For this purpose, we plan to simulate changes of the user interests across search sessions and integrate a session boundaries delimitation mechanism that measures the semantic correlation degree between the search sessions. Furthermore, we plan to study the accuracy of the user context representation and also its effect on the retrieval precision.

References

[1] Park , T.K., “Toward a theory of user-based relevance: A call for a new paradigm of inquiry,” *JASIS*, Leuven, Belgium, pp. 135-141 1994

[2] Budzik, J., Hammond, K.J., “Users interactions with everyday applications as context for just-in-time information access,” *Proceedings of the 5th international conference on intelligent user interfaces*, pp. 41-51, 2000

[3] Shen, X., Tan, B., Zhai, C., “Ucair: Capturing and exploiting context for personalized search,” *Proceedings of the Information Retrieval in Context Workshop, SIGIR IRiX*, Salvador, Brazil, 2005

[4] Miller, B., Konstan, J., Matlz, D., Herlocker, J.L., Gordan, L., Riedl, A., “GroupLens: applying collaborative filtering Usenet news,” *Communications of ACM*, 1997

[5] Saracevic, T., “The stratefied model of information retrieval interaction: extension and applications,” *In proceedings of the 60th annual meeting of the American Society for Information Science*, Medford, NJ, pp. 313-327, 1997

[6] Lieberman, H., “Autonomous interface agents,” *In ACM Conference on Human-Computer Interface*, ACM Press, pp. 67-74, 1997

[7] Sieg, A., Mobasher, B., Burke, R., Prabu, G., Lytinen, S., “Representing user information context with ontologies,” *In uahci05*, 2005

[8] Sieg, A., Mobasher, B., Burke, R., Prabu, G., Lytinen, S., “Using Concept Hierarchies to Enhance User Queries in Web-based Information Retrieval,” *Artificial Intelligence and Applications(AIA)*,2004

[9] Challam, V., Gauch, S., Chandramouli, A., “Contextual Search Using Ontology-Based User Profiles,” *Proceedings of RIAO 2007*,Pittsburgh, USA, 2007

[10] Koutrika, G., Ioannidis, Y., “A Unified User Profile Framework for Query Disambiguation and Personalization,” *Proceedings of Workshop on New Technologies for Personalized Information Access PIA 2005*, Edinburgh, Scotland, UK, 2005

[11] Allan, J., al., “Challenges in information retrieval and langage modelling,” *Workshop held at the center for intelligent information retrieval*, 2002

[12] Tamine, L., Boughanem, M., Zemirli, W. N., “Inferring the user’s interests using the search history,” *Workshop on information retrieval, Learning, Knowledge and Adaptatbility (LWA 2006)*, Hildesheim, Germany, pp. 108-110, 2006

[13] Shen, D., Chen, Z., Yang, Q., Zeng, H., Zhang, B., Lu , Y., Ma, W., “Web-page classification through summarization,” *In Proceedings of the 27th Annual International ACMSIGIR Conference on Research and Development in Information Retrieval*, Sheffield, South Yorkshire, UK pp. 242-249, 2004

[14] Joachims, T., “Optimizing search engines using clickthrough data,” *In Proceedings of SIGKDD 2002*, 2002

[15] Tamine, L., Boughanem, M., Zemirli, W.N., “Exploiting Multi-Evidence from Multiple User’s Interests to Personalizing Information Retrieval,” *In IEEE International Conference on Digital Information Management(ICDIM 2007)*, 2007

[16] Daoud, M., Tamine, L., Boughanem, M., Chebaro, B., “Learning Implicit User Interests Using Ontology and Search History for Personalization,” *In International Web Information Systems Engineering - International Workshop on Personalized Access to Web Information (WISE-PAWI 2007)*, Nancy, France, 2007

[17] Ma, Z., Pant, G., Sheng, O., “Interest-based personalized search,” *ACM Transactions on Information Systems*, V25, N1, 07

[18] Liu, F., Yu, C., Meng, W., “Personalized Web Search For Improving Retrieval Effectiveness,” *IEEE Transactions on Knowledge and Data Engineering*, V16, N1, pp. 28-40, 04

[19] Menczer, F., Pant, G., Srinivasan, P., “Topical Web crawlers: Evaluating adaptive algorithms,” *ACM Trans. Internet Techn.*, V4, N4, pp. 378419, 04

[20] Srinivasan, P., Menczer, F., Pant, G., “A general evaluation framework for topical crawlers,” *Inform. Retriev.*, V8, N3, pp. 417447, 05

[21] Fuhr, N., *information retrieval: introduction and survey*, post-graduate course on information retrieval, university of Duisburg-Essen, Germany, 2000

[22] Rhodes, B., *Just-In-Time Information Retrieval*, PhD thesis, MIT Media Laboratory, Cambridge, MA, June 2000

[23] Mc Gowan, J.P., *A multiple model approach to personalized information access* Master Thesis in computer science, Faculty of science, University College Dublin, February 2003