International Journal of Computer Applications (0975 - 8887)
Volume 158 - No.4, January 2017

An Approach of Re-Ranking Search Results based on a Dynamic and Hybrid Modeling of User Profile

Yannick U. Tchantchou Samen
Institute of Mathematics and Physics
The University of Abomey-calavi
and
Faculty of Sciences
The University of yaounde 1

Eugène C. Ezin
Institute of Mathematics and Physics
The University of Abomey-calavi

Charles Awono Onana
National Advanced School of Engineering
The University of yaounde 1

ABSTRACT

The volume of data on the web grew in recent years. Then it becomes increasingly difficult for a user to access the right information in a short time. However, several works have been carried out with the aim of proposing algorithms to re-rank the user's search results on the web by taking into account their profile. In this paper, we propose specific approach of re-ranking user search results based on a dynamic and hybrid modeling of user profile. Our approach takes into account the user interets identified during his browsing session and the history of his search on the web. We use a multi agent system to collect both explicitly and implicitly user data and to process this data to detect the user interets represented as ontological concepts. The experimentation of our model shows that it is able to re-rank user search results with a high accuracy than that given by the google search engine.

General Terms

Information retrieval, text mining

Keywords

User profile, Re-ranking, search result, ontology, multi-agent system

1. INTRODUCTION

The problem of accessing to information has always been at the center of the tasks of researchers. In the past, the difficulty lays in the information availability. However the advent of big data coupled with the recent advances of information and communication technology have brought new challenges in the information retrieval fields. Nowadays, the problem is less about the availability of information but more in the ability of the information retrieval system to select and offer the right information.

The implementation of search engines does not solve the aforementioned problem. On the contrary, they merely offer a multitude of answers to users who feel compelled to search the right information among them. Futhermore, most of these search engines do not take into account the user who is supposed to be the main element of the information search process. Therefore the need to develop information system that can personalize the information search according to specific user needs is useful. To get there, it is important to design and setup user profile in order to collect and detect its preferences and interests. Furthermore, re-rank user's search results according to this interets detected and stored in his profile.

Several approaches are used to collect user interests to personalize user search results on the web. Some authors used ontology to model and build ontological user profiles to retrieve personalised search results. Sieg et al. [1] propose an ontological user profile built by using a spreading activation mechanism and use this profile to re-rank user search results based on his current's interests. Similarly, Hawalah and Falsi [2] proposed a hybrid re-ranking algorithm based on user profiles. This algorithm is based on the combination of different information resources collected from the reference ontology, user profile and original search engines ranking.

A popular approach used the semantic similarity to re-rank search results [3, 4] and another exploited contextual information to identify user interests. In [5] for example, user context is identified based on current user query while in [6], user context is identified based on his browsing behaviour alongside with current user interests at the time of conducting a search.

In this paper, we propose a particular re-ranking algorithm based on a dynamic and hybrid user profiling. This model allows us to gather information about user implicitly and explicitly, to detect changes in these user interests but also to use this interests to personalize the search results ranking. To achieve this, we use the ontology concepts to build the ontological user profile, a multi-agent system to learn user short-term and long-term interests and the history of user search results. We note that this user search history allows us to build a document containing all the web pages visited by the user and having positive feedback, as well as the ontological concept mapped to this web page.

The remainder of this paper is organised as follows: in the Section 2 we propose the main architecture of our dynamic and hybrid model, Section 3 is dedicated to proposed re-ranking approach and the Section give the experimental set-up and evaluation of our proposed model. This paper ends with a conclusion and the outlook in Section 5.

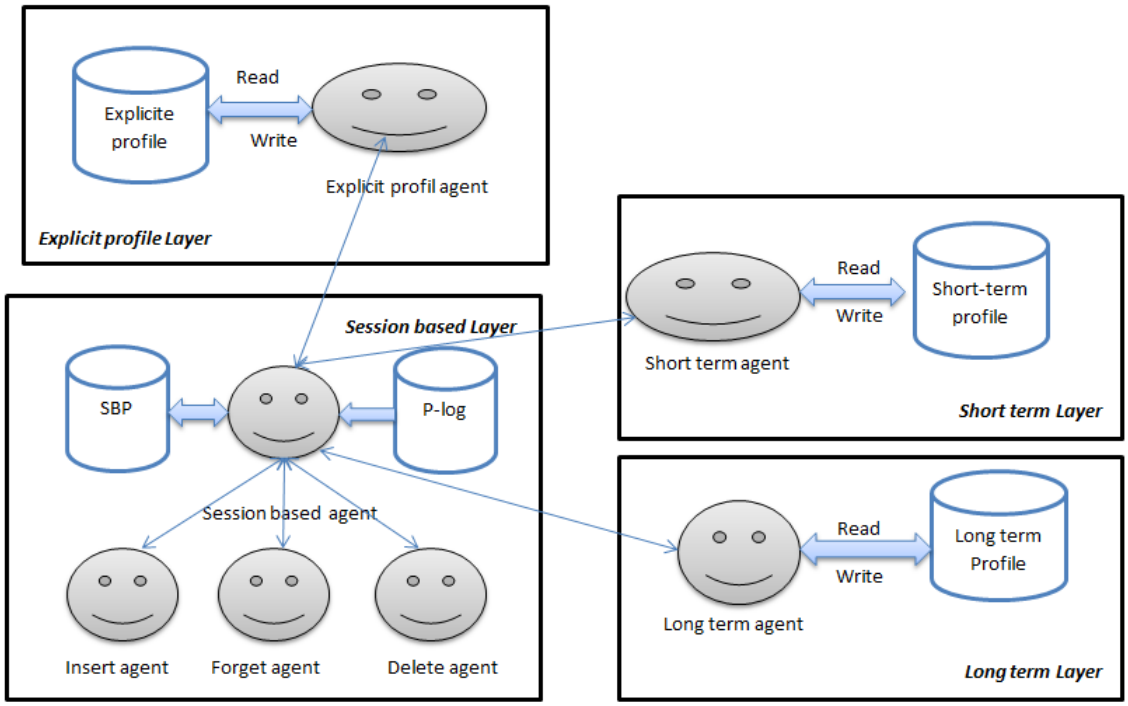

Fig. 1. Architecture of our multi-agent system

2. THE PROPOSED ARCHITECTURE OF USER PROFILE

In this section, we present a dynamic and hybrid user profile that is able to learn and adapt user's interests based on data obtained firstly explicitly from the user and by observing user behaviour and mapped to a reference ontology. The user profile consists of four layers: session based layer, explicit profile layer, short-term layer and long-term layer. Each layer consists of one or more agents that are responsible for a set of tasks. The use of multi-agent system resides in their ability to address the complex problem by dividing it into sub problems which can be handled by agents. The proposed model needs to track user behaviour, add, update, delete user interest and dynamically process explicit user interests. Generally, user interests always change. Indeed, a user may lose an interest in an item or a concept that he was interested in the past. Hence it is important to detect this change in behaviour and to adapt the user profile to improve the user information search. The proposed model is a generalization of the one proposed by Hawalah and Falsi (2015) [8, 7] with the capability to learn and adapt to such user behaviour by collecting data explicitly from user and implicitly by observing user browsing behaviour.

2.1 Explicit profile Layer

This layer deals with the explicit collection, storage and the updating process of the interests inserted explicitly by the user. Each user input instructions concerns his explicit profile are processed first in this layer by the explicit profile agent and stored in the explicit profile before being sent to the session-based agent for the next step.
A **user explicit interest** is an ontological concept of interest insert explicitly by user in his explicit profile.
The explicit profile agent is in charge of certain tasks as:

(1) collection and storage of information explicitly insert by the user;
(2) communication with the session-based agent to advise on all the operations performed by the user in its explicit profile (new information insertion, deletion, etc.);
(3) communication with the user before completing the deletion process of the concept that are still in his explicit profile.

2.2 Session-based layer

This layer has an essential role in the modeling process of our dynamic and hybrid profile. In fact, it contains all the mechanisms associated with learning and adaptation of user interests. For this reason it is related to all other layers in our model.
It receives both explicit concepts inserted by the user and concepts visited by the user during his browsing session and contained in the P-log file (*Processed log file*). Thereafter, it computes and updates the weight of this concept in the session-based profile. Finally it sends the list of this updated concept with their weight to the short-term and long-term layer to determine the short-term interests and the long-term interests. This layer is active during each browsing session to deal with new concepts inserted or visited in order to detect any shift or drift in the user interests. This layer also includes a profile called SBP (session based-profile) and several agents. Each of these agents is responsible for one or more tasks.

Session-based agent. This agent is the core of our multi-agent system and it is in charge of several tasks:

(1) data collection from the P-log file;

(2) data collection from the explicit profile;

(3) communication with other agents to calculate the latest interest weight in a session-based profile;

(4) communication with the short-term and long-term agents to enable them to discover short-term and long-term interests respectively.

Once a session ended, data in the P-log file and the other one inserted explicitly by user in the explicit profile are processed and stored in the session-based profile.

Insert agent. This agent is responsible for processing all concepts with positive status received from the Session-based agent. Unlike Hawalah model [8], we distinguish four different events with positive status: *browsing concept*, *confirmed concept*, *explicit concept* and *explicit confirmed concept*.
The *browsing concept* event is the default one that is assigned to any concept browsed by the user. If such concept appears in two subsequent sessions, it is assigned to *confirmed concept* event. The later event has a higher weight than the *browsing concept*, as concepts appearing in more than one session would likely be of more interest to users than those that appear one.
The *explicit concept* event is any concept inserted explicitly, or manually by the user in its explicit profile. As information given by the user seems much more credible than those implicitly obtained by the system, this event has a higher weight than the *confirmed concept*. If such concept appears implicitly in one or more sessions, it is assigned *explicit confirmed concept* event. The later event has a higher weight than the *explicit concept*, as it is firstly an explicit concept but also confirmed implicitly during the browsing session as an interesting concept.

Forget Agent. This agent handles the behaviour that occurs when user loses interest in a concept. Our system is able to take into account this changes in user behaviour. However, this concept cannot be deleted immediately. they must follow a gradual forgetting process to be confirmed as an uninteresting concept for user.
In our case, the forgetting process depends on several factors:

(1) the *relevance-size* that is associated with each concept. The relevance-size is an indicator of the user's strength of interest in a concept. The value of the relevance-size is essential to determine the pace of the forgetting process as the larger the relevance-size is, the slower will be the forgetting process and vice versa;

(2) the *recency* of a concept, as old concepts are forgotten faster than new ones;

(3) the introduction of new interests to a user profile. If a user has started to lose his interest in a concept, and at the same time started to show interest in new concepts, then this behaviour might indicate that a user has started to drift his interest to new concepts;

(4) the user decides to explicitly delete this concept in his explicit profile.

Delete Agent. As in [8], this agent manages the gradual deletion of a concept from a user profile. When a concept is passed to the Delete Agent, this is removed much faster based on the time of the last appearance of the concept, and until the weight reaches a predefined threshold and then it is removed altogether.

2.3 Short-term Layer

The short-term layer is responsible for the development of learning mechanism of user's short-term interests. This layer includes the short-term profile (STP) use to store all the concepts identified as short-term interests, and the short-term agent (STA). The later is responsible for tasks such as discovering, maintaining and storing short-term interests in the short-term profile.

2.4 Long-term Layer

The goal of this layer is to learn, recognize and store the user long-term interests. It includes also two components: the long-term profile (LTP), which stores all interests recognized as long-term ones; and the Long-term agent (LTA), whose task is to recognize , to maintain and to store long-term interests in the long-term profile.
In our model, any explicit concept cannot be consider as a long-term interest. The reason is that user has just inserted this concept as an explicit concept but not ever browsed it during one of his browsing session.

3. THE PROPOSED APPROACH OF RE-RANKING USER SEARCH RESULTS

In this section, we present how our model takes into account the collected and processed user data (short term and long term interests, and his browsing history) to re-rank the search results of its queries obtained from a search engine. After the learning and adaptation phase, an ontological user profile is created and this profile contains all ontological concepts recognized as interesting to the user.
Each ontological concept is represented by two documents: the **ontology document** containing information derived from a reference ontology and representing this concept, and the **positive document** containing the set of all web pages visited by the user, mapped to this ontological concept and considered as very interesting for the user. We define for this purpose a threshold beyond which any web page visited by the user can be considered as very interesting for him and therefore must be stored in the positive document.
Unlike Hawalah and Fasli [2] for whom this threshold depends on the time spent on all web pages visited during the session, our threshold is fixed but also depends only to the time spent on this page. This allows to make a better selection of web pages deemed positive for the user.
When a query is sent, two processes are triggered simultaneously: on the one hand, the query is mapped to the ontological profile and possibly to the concepts contained in the SBP. On the other hand the query is introduced into a search engine to retrieve potential results.

3.1 Process for mapping query

When a user enters a query, this query is retrieved and the term vector corresponding to this query is determined by using text mining algorithm (remove stop word, stemming, etc). We use the traditional cosine similarity [9] to map this query to each user ontological concept. Firstly, the query is mapped to each concept in the user ontological profile, and the concept with the high cosine similarity is selected with his cosine similarity value. If this value is less than a fixed threshold (in our case 0.5), the mapping process continues, but this time with the ontological concepts in the SBP that have not been identified as either short-term or long-term interest. At the end of this process, the concept with the high cosine similarity value is retained and will be used in the re-ranking search results process.

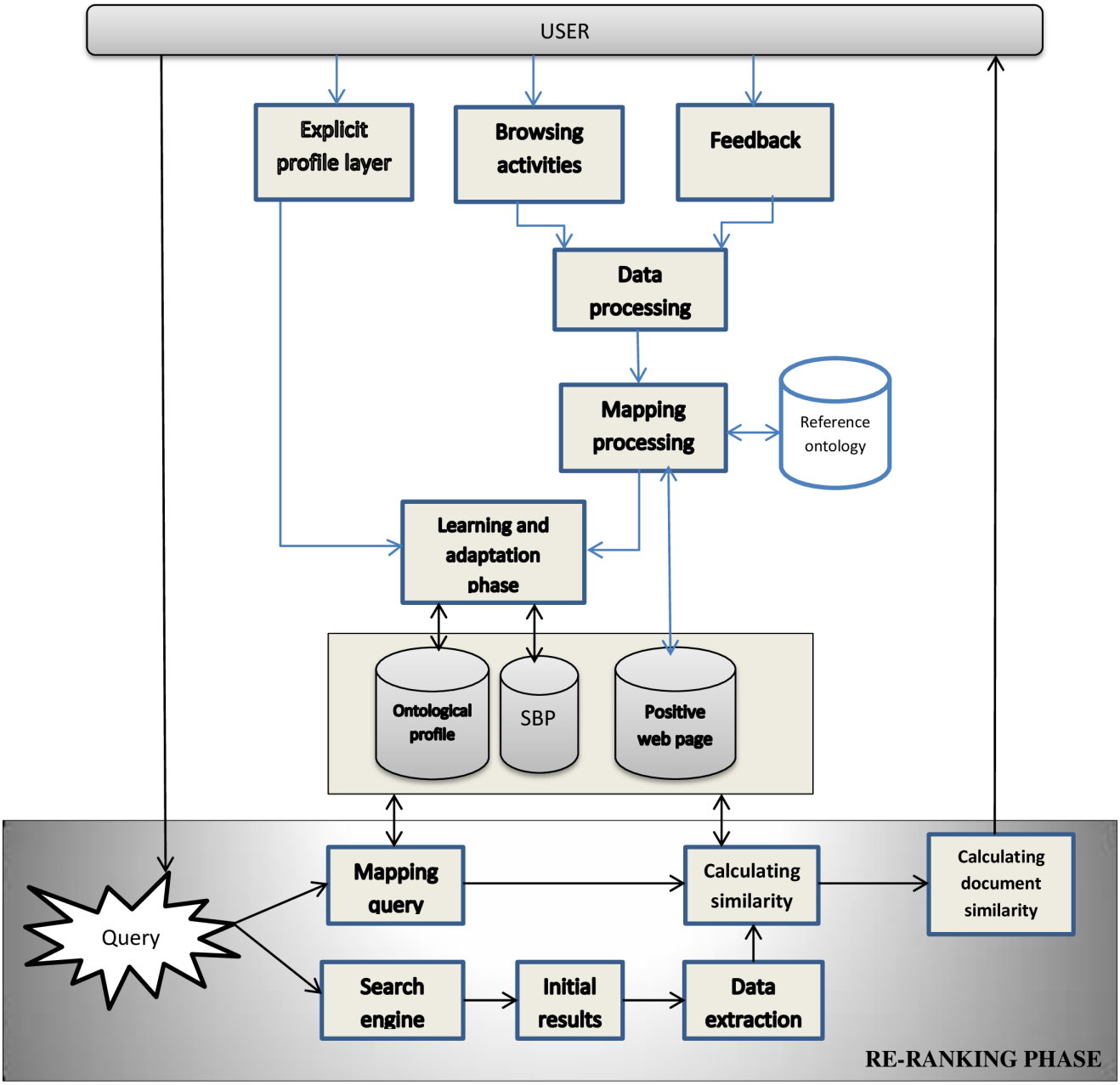

Fig. 2. main architecture

The importance of mapping query to concept in the SBP resides in the fact that the user may be interested in a concept visited in the past and thus no longer being part of his ontological profile. Otherwise, user may be intrested by a concept previously visited but not yet fulfilling the conditions required to be regarded as short-term or long-term interest. At this level, such a concept will be mapped

to this query while for another approach especially that proposed in [2], an erroneous concept will be mapped to this query and the re-ranking process will not be efficient.
Finally, when the query is mapped to the appropriate concept, another mapping is carried out within the positive document associated with this ontological concept, to determine the best term vector (closest to the query) to be used to optimize the re-ranking of the search results.

3.2 Re-ranking search process

When the query is sent to the search engine, the results are extracted and the term vector of each retrieved web pages is determined by using the same text mining algorithm as in the mapping process. After this, each web page is mapped by using cosine similarity to the term vector obtained in the mapping process. Then these are ranked in descending order of their cosine similarity value. The final ranking is obtained by using the personalized ranking and the original ranking proposed by the search engine. For this purpose, we use the following formula.

$$FinalRanking = \alpha \cdot S_i(R_i) + (1 - \alpha) \cdot Rank_{Original} \quad (1)$$

where $S_i(R_i)$ is the rank of the page i obtained by mapping this page to the concept obtained in the mapping process, $\alpha \in [0, 1]$. In our approach, α is the cosine similarity value of the term vector selected in the mapping process. In the case where the query is not mapped to a concept during the mapping process, $\alpha = 0$ and the final ranking depends only of the original ranking proposed by the search engine.

4. EXPERIMENTAL EVALUATION

4.1 Information retrieval phase

To use our evaluation framework, each user must possess an account if it is the first time that he uses it and then he must identify himself using his login and password. During the registration in our framework, the user may insert his interests and other information. This inserted interests are directly collected by the explicit agent and treated by the multi-agent systems.
When the user browses the web, our framework collects implicitly all the visited web pages, duration of visit and stores in the log file database. When the user leaves this web page, our framework extracts this content, processes and maps it to the corresponding ontological concept. Let us note that in our case, we use a reference ontology "Computer" of *open directory project*(ODP) [1].
The content of each visited web page is processed as following: we firstly remove all the stop words by using the Porter algorithm (1997) [10], next by stemming process we reduce each word to his stem. In our experimental framework, the weight of each word in the web page depends on two components: the word position in the web page and the number of occurrences of this word at this position. Indeed, a word located in the title of a web page represents more the content of this page than another word located in the body of this page. For this reason, we define firstly a weight corresponding to the position of the word in the web page. So, for a word located in the title, we attribute a weight of 0.5 while for a word in the metadata we attribute a weight of 0.3 and finally a word in the body a weight of 0.2. The final weight of the word for this page is given by the following formula.

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

$$\omega_{t_i} = \sum_{j=1}^{3} \alpha_{t_i,j} p_{t_i,j} \quad (2)$$

where $p_{t_i,j}$ is the weight of the word t_i in the position j, $\alpha_{t_i,j}$ is the number of occurrences of the words t_i in the position j and finally j represents title, metadata or the body of the web page.
This formula allows us to determine the term vector representing this page with the weight associated to each of this word.
After this, we use the traditional cosine similarity [9] to map the visited web pages to the ontological concept from the reference ontology.

$$Sim_{cosine}(d_1, d_2) = \frac{\sum_{i=1}^{n} \omega_{i_1} \cdot \omega_{i_2}}{\sqrt{\sum_{i=1}^{n} \omega_{i_1}^2} \cdot \sqrt{\sum_{i=1}^{n} \omega_{i_2}^2}} \quad (3)$$

At the end of this process, the ontological concept with the greatest value of cosine similarity is mapped to this page and is stored into the database of processed log file (Plog).
When the session ends, concepts stored in the database during the session is extracted by the session based agent and treated by our proposed multi-agent system.

4.2 Description of the experimental phase

To carry out this experiment, we defined a set of scenarios and used three real users to simulate these different scenarios for a period of six days. These users were asked to research some of the concepts related to the computer domain during the first three days of the experiment. The first user explicitly inserted his concepts of interest into his explicit profile and carried out his research around these concepts. The second user inserted a concept but did not perform any term search. However, his search was accentuated on other concepts not found in his profile. The last user did not insert any concept at the outset; he merely did some search on certain concepts during the first three days.
During the last three days, the first user did his search around the concepts of his choice without inserting them in his explicit profile. The second user searched the concept inserted at the beginning of the experiment. The last user inserted concepts into his explicit profile and continued his search around these concepts.
The framework we implemented for this experimentation allowed us to simultaneously obtain the results personalized and those from the google search engine. This enabled us to easily evaluate the accuracy of our approach of re-ranking search results compared to that proposed by a classic search engine.

4.3 Evaluation of the accuracy of our re-ranking approach

The purpose of this experimentation phase was to compare the accuracy of re-ranking results proposed by our approach with that proposed by google search engine. To this effect, for each query introduced by the user, we evaluated the first ten results proposed by our approach and that proposed by google. For each of the 6 six days, we determined the average of the accuracy for each user. All results are summarized in Figure 3.
Through this figure, we can notice that the results obtained with our approach are clearly more reliable than those obtained by the google search engine. Moreover, through the first user, it can be emphasized that when a user inserts an interest concept into his explicit profile, immediately the system uses this concept in the pro-

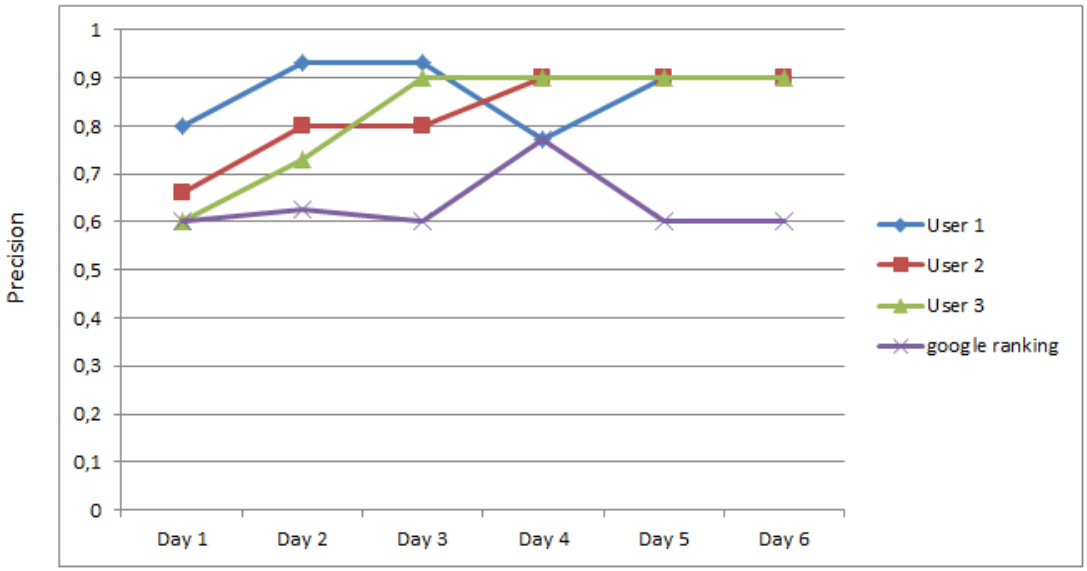

Fig. 3. Accuracy of our re-ranking search results

cess of mapping the queries emitted by the user. This explains the high precision observed during the first day.
In addition, during the second and the third days, we remarked that the precision of our re-ranking is even greater than during the previous days. This is due to the fact that during these days, the system had already taken the time to collect, learn the user interests and also stored the web pages visited and deemed positive for the user. By taking in consideration all collected data, the system can re-rank search results according to this user's interests.
Finally, the weak precision recorded on the fourth day is justified by the fact that at this moment there is a change in the user interests. The SBP does not have information related to these new interests, merely reproduces the same results as that proposed by google. However, this precision increases again during days 5 and 6; reflecting the fact that the system has collected sufficient information on the new preferences of the user.

5. CONCLUSION

In this paper, we have presented an approach allowing to customize the re-ranking search results. This approach is based on a dynamic and hybrid user profile and takes into account the browsing history of this user. Our approach has the advantage of proposing an algorithm exploiting firstly all information collected explicitly (by the user himself) and implicitly during his browsing activities on the web, and on the other hand the web pages visited and judged positive for this user.
The experimentation of our approach shows that the accuracy of our re-ranking search results is better than that obtain directly through the google search engine. However, let us note that the precision of our approach depends strongly on the structure of the ontology used during the mapping process. In the future, it will be interesting to study how to take into account the semantic structure of each ontological concepts for improving the process of building the user ontological profile.

6. REFERENCES

[1] Sieg, A., Mobasher, B., Burke, R., Ontological user profiles for representing context in web search, in: 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops, IEEE. pp. 91 94, 2007.

[2] Hawalah, A., Fasli, M.: A Hybrid Re-ranking Algorithm Based on Ontological User Profiles; Proc. 3rd Computer Science and Electronic Engineering Conference, 50-55, 2011.

[3] Rajpal R. K., and Rathore Y., A Novel Techinque For Ranking of Documents Using Semantic Similarity, International Journal of Computer Science and Information Technologies (IJCSIT), vol. 5, 2014.

[4] Ruofan Wang, Shan Jiang, Yan Zhang, and Min Wang, Reranking search results using semantic similarity. In Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on, volume 2, pages 10471051. IEEE, 2011.

[5] Mohammed, N.U., Doung, T.H. and Jo, G.K. Contextual Information Search Based on Ontological User Profile. ICCCI 2010, pp.490-500, 2010.

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

[7] Hawalah, A., and Fasli, M. A multi-agent system using ontological user profiles for dynamic user modelling. InProceedings of the 2011 IEEE/WIC/ACM international conferences on web intelligence and intelligent agent technology (pp.430-437), 2011.

[8] Hawalah, A., and Fasli, M. Dynamic user profiles for web personalization. Journal of Expert systems with Applications, 42, pp. 2547-2569. Doi: 10.1016/j.eswa.2014.10.032, 2015.

[9] Baeza-Yates, R. and Berthier Ribeiro-Neto. Modern Information Retrieval. AddisonWesley, 1999.

[10] Porter, M.F. An algorithm for suffix stripping, in: Sparck Jones, K., Willett, P. (Eds.), Readings in Information Retrieval. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 313 316, 1997.

[11] Daoud, M., Tamine-Lechani, L., Boughanem, M., Learning implicit user interest using ontology and search history for personnalization. Web information systems engineering-WISE 2007 workshops, vol. 4832, pp. 325-336, 2008.

[12] Li, L., Yang, Z., Wang, B., and Kitsuregawa., M. Dynamic adaptation strategies for long-term and short-term user profiles to personalize search. springer-verlag Berlin Heidelberg. pp. 228-240, 2007.

[13] Rongen P.H.H., Schroder J., Dignum F.P.M., et al. Multi agent approach to interest profiling of users. Proc. of Multi-Agen Systems and Applications IV, pp. 326 355, 2005.

[14] Pannu, M., Anane, R., James, A. Explicit user profiles in web search personalization; computer supported cooperative work in design (CSCWD), 15th international conference on, 8-10 june 2011. pp. 416-421, 2011.

[15] Pannu, M., Anane, R., James, A. Hybrid profiling in information retrieval; computer supported cooperative work in Design(CSCWD), IEEE 17^{th} international conference on computer supported cooperative work in design, 2013.

[16] Razmerita, L., Angehrn, A., Maedche, A. Ontology-based user modeling for knowledge management systems. In: International Conference on User Modeling, pp. 213 217, 2003.

[17] Shen, X., Tom, B., Zhou, C. Implicit user modelling for personalized search. Proceeding of the 14th ACM international conference on information and knowledge management (New York) pp: 824-831, 2005.

[18] Srinvas, S. Explicit user profiles for semantic web search using XML. International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, vol. 2, Issue 6, November-December 2012, pp. 234-241, 2012.