

Personalized Information Retrieval: Application to Virtual Communities

Azza Harbaoui¹, Sahbi Sidhom², Malek Ghenima¹, and Henda Ben Ghezala¹

¹ Laboratory RIADI-GDL, National School of Computer Sciences, University of Manouba
2010 la Manouba, Tunisia

{azza.harbaoui, malekghenima, hbg.hhbg}@gmail.com

² KIWI Research Team, LORIA, Lorraine University (Nancy), France
sahbi.sidhom@gmail.com

Abstract. Internet has become the largest library through the history of humanity. Having such a big library made the search process more complicated. In fact, traditional search engines answer users by sending back the same results to different users having expressed different information needs and different preferences. A significant part of difficulties [1],[4] is due to vocabulary problems (polysemy, synonymy...). Such problems trigger a strong need for personalizing the search results based on user preferences. The goal of personalized information [11] is to generate meaningful results to a collection of information users that may interest them using user's profile. This paper presents a personalized information retrieval approach based on user profile. User profile is built from the acquisition of explicit and implicit user data. The proposed approach also presents a semantic-based optimization method for user query. The system uses user profile to construct virtual communities. Moreover, it uses the user's navigation data to predict user's preferences in order to update virtual communities.

Keywords: personalized information retrieval, user modeling, user profile, virtual communities.

1 Introduction

With the large volume of information available on the web, browsing this content became a difficult task, especially to keyword-based search engines. User is more and more unsatisfied by web search results. [15]

However, these search engines do not address vocabulary problems such as polysemy and synonymy. Polysemy is multiple meanings for a single word. For example, when a user searches for the word "Apple", the retrieved results may be related to "Apple fruit" or "Apple computer". Synonymy refers to the same meaning as another word in the same language. For example, when a user searches for the word "little", all results related to another word like "small" would be processed although the two words nearly have the same meaning.

Personalized information retrieval can bring solution to the above problems by focusing on the most relevant results of a user query[14] that takes into account his/her preferences identified in his/her user profile. Existing personalization approaches try to determine users' preferences in order to assist them while searching for information. It is an emergent research field with the aim of facilitating the use of web content, and assisting the user to obtain the most relevant result.[18]

In this paper, a new approach (based on user modeling (multidimensional representation), for personalizing web search result is proposed.

In the proposed approach (figure 1), a user profile is built based on basic information (explicit, extracted from human user interface) and implicit through user feedback (history of sessions, printed documents ...etc.).Consequently, user profile is used to build virtual communities. In addition, and in order to keep up-to date user profile, we propose a semantic query enrichment technique based on user profile-related query, Wordnet Ontology and ODP ontology domain[9] (**O**pen **D**irectory **P**roject).

2 The Notion of User Profile

The term "user profile", appeared around the 80s, with interface agents, mainly because of the need to create custom applications adapted to user needs [5]. User profile is at the heart of personalized RI. Unlike context which covers contextual elements, profile is defined by contextual elements directly related to the user (his interests, preferences.). Several definitions [16] of profile have been discussed in the literature, of which we retain a few:

"All changes that characterize a user or group of users can be grouped under the term user profile "[2]

"User profile is a structure of heterogeneous information covering broad aspects such as users' cognitive, social and professional environment, this information is usually used in order to clarify their intentions during a search session"[3]

In summary, we can define RI profile as the set of all the dimensions that describe and/or infer their intentions and perception of relevance.

3 Modeling User Profile

Modeling the user is at the centre of the implementation of a personalized information search process. The goal of user modeling is to select the most relevant information that reflects users' interests[13]. This modeling consists of designating a structure in which we store information that describes essentially:

- User interests;
- Preferences;
- Context;

- Expected goal of the search;
- Individual traits;
- Experience

There are several definitions of user modelling in the literature [6][8]. We retain some below:

«A user model is a knowledge source in a natural-language-dialogue system which contains explicit assumptions on all aspects of the user that may be relevant for the dialogue behavior of the system"[17]

"User model is an explicit representation of the system of a particular user's characteristics that may be relevant for personalized interaction."[10]

"The process of gathering information about the users of computer systems [7]and of making this information available to systems which exploit it to adapt their behavior or the information they provide to the specific requirements of individual users has been termed as user modeling."

Several techniques were developed in the literature to model the user. They differ according to the approach of profile representation and construction [12]. We present in this section the data acquisition techniques and profiles construction techniques as proposed by several representation models: global, connectionist, semantic or multidimensional.

The objective of this research is to provide architecture of an information research system which should be:

- User-centred: by taking into account user's profile, preferences and interests in order to provide results, the most suitable to their needs.
- Interactive: through using a dialogue mechanism allowing a "natural" interaction with users during expression and refinement of their requests.

Therefore, we develop an approach that combines two processes:

An information search process and a user profiles building process in order to evaluate users' contribution (ratings, tags) in improving the information search process.

To this end, we set the following objectives:

- The definition of an approach that integrates to user profiles construction;
- The definition of a collaborative process: to determine the manner with which both processes (RI and profiles management) will complete each other;
- The definition of ontologies to build (domain ontology and profile ontology) to use for personalized search;
- The integration of the social aspect in constructing user's virtual communities.
- The study of the impact of the proposed approach on improving information search (in terms of evaluation metrics).

All these proposals are detailed in the next section which describes the holistic approach integrating the different objectives and contributions of this article.

4 Presentation of the Approach

This section describes the overall approach entitled "Personalized information search Approach based on influential networks". The approach we propose is a modular approach is composed of the following main modules:

Query Reformulation Module. This module will reformulate the query based on the initial user profile P_0 . The query will be enriched with new concepts (new content added by user), users' ratings of the results generated from the RI classic process (bad, good, very good ...)

User Modeling Module. The construction of user profile will be based on certain criteria such as: explicit acquisition that will safeguard profile information and search motivation (user is motivated, less motivated), also an updating module (user-initiated update and an automatic update)

Navigation Data Acquisition Module. This module captures all user navigation data during search sessions namely: printed, saved, tagged documents and calculation of participation rate.

This module is powered by the already built user profile. It allows for grouping user profiles into similar profiles communities in order to classify new profiles in relevant communities. This classification could help new users by reasoning through similar search.

Once constructed, influential networks will be deduced from the communities in order to identify the most influential user profiles.

4.1 Formal Framework

User profile is represented under the following dimensions:

1. Interests;
2. Ontology domain ODP (Open Directory Project) is a widely used ontology. It represents the most complete web directory edited by humans and often used as a source of semantic knowledge)
3. Explicit and implicit feedback;
4. Virtual Community;
5. History: Past search and navigation data;
6. Personal data
7. User profile, noted P_u , includes:

$$P_u = \{D_p, D_{ci}, D_{cv}, D_N, F_{ci}\}$$

With:

$$D_p = \{I, D_d\}$$

$$D_{ci} = \{C_i, q, d_{NI}\}$$

$$D_{cv} = \{tp, tc, P_{cv}\}$$

$$D_N = \{_l, _s, _i, _tag\}$$

$$Fei = \{V_u, D_e, D_i\}$$

Q	User query
P0	User profile
PB	Preferences base
Tq	Set of a query terms
Si	Similarity between profile in vc
D	Document
Test	detector (of interest...)
DU	Usability of new documents
U	Usability threshold to launch an update.
Nbq, Nbd	Respectively number of queries and documents
Dp	Personal data
Dci	Interests data
Dcv	Virtual community data
DN	Navigation data
Fei	Explicit and implicit Feedback
I	Identity
Dd	Demographic data
dNI	Detector new interests
Ci	Concept
qi	Query during current session
tp	Participation rate
tc	Trust rate
al	Reading time
Ai	Number of printed documents
As	Number of saved documents
atag	Number of tagged documents
PRq	Set of profiles similar to one query

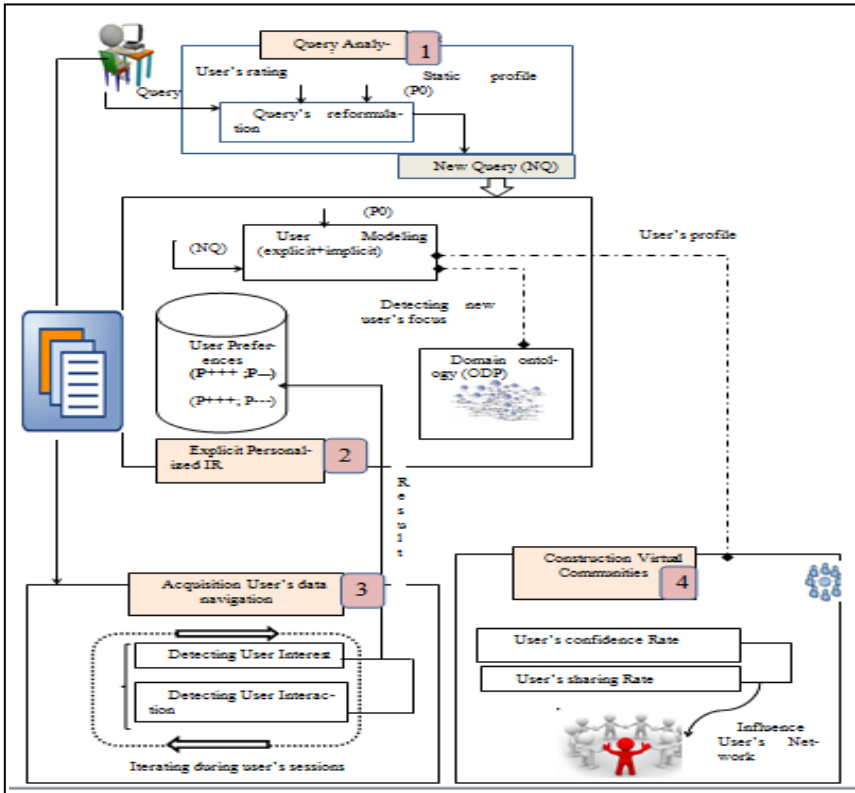


Fig. 1. The Proposed approach

4.2 Proposed Algorithms

We detail In this section the principle algorithms of the proposed approach

Algorithm 1: Algorithm for Building User Profile.

Entries: $Q_t = (w_1, w_2, w_3 \dots w_n)$

Outputs:

$P_u = \{D_p, D_{ci}, DCV, DN, Fei$

g

Start

If a user with an initialization session of a P_0 profile

for each query submitted do

Save navigations or search history;

Calculate interest;

Infer or predict interest dimension;

Calculate I ;

Inf-Interest (D_{ci}, q_i, P_u)

```
Pn
Case of an unregistered user in the System
Assign the user to a profile
For each query submitted do
Return to start
End
Return Pu
End
```

Algorithm 2: GetNewInterest Algorithm.

```
Entries:
Data: S: confidence level of new interest with the profiles database
BP: Base profile
qi: user query
Pi: Initial profile
Output: Result: Flag: flag value specifying the new interest CI
Start
Flag = 0
Score (qi) = GetSimilarity (qi, Pu)
Flag = 1
if Score (qi) < S then
Flag = 1
End
Return Flag
End
```

5 Conclusion

Although a considerable number of works focused on information retrieval, some important challenges for the research community still remain, the ultimate goal of personalization system being the satisfaction of the user. To reach this goal, the user has to be implied in the construction process in order to add the semantic value to the information retrieval. This paper proposed a personalized research information approach based on user modeling and the construction of virtual communities. The aim is to generate social relations from constructed community, which will allow us to infer influence user networks that relate to each other through their relationships and their sharing spirit. We plan to test the approach on real corpus of users of social networks.

References

1. Research, I.A., Zien, J., Meyer, J., Tomlin, J.: Web query characteristics and their implications on search engines. In: Zien, J., Meyer, J.O., Tomlin, J. (eds.) Proceedings of the 10th International WWW Conference, Hong Kong (2001)

2. Robertson, S.: The probability ranking principle in modern information retrieval. *Journal of Documentation* 33(4), 294–304 (1977)
3. Allan, J., et al.: Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent information retrieval. SIGIR. University of Massachusetts Amherst (September 2002)
4. Mianowska, B., Nguyen, N.T.: Tuning user profiles based on analyzing dynamic preference in document retrieval systems. *Multimedia Tools and Applications* (2012), <http://dx.doi.org/10.1007/s11042-012-1145-6>
5. Boudighaghen, O., Tamine, L., Boughanem, M.: Personalizing mobile web search for location sensitive queries. In: *Proceedings of the 2011 IEEE 12th International Conference on Mobile Data Management*, Lulea, Sweden, vol. 01, pp. 110–118 (2011)
6. Ghosh, R., Dekhil, M.: Discovering user profiles. In: *Proceedings of the 18th International Conference on World Wide Web*, pp. 1233–1234. Polytechnic University in Madrid (2009)
7. Treur, J., Umair, M.: An agent model integrating an adaptive model for environmental dynamics. *International Journal of Intelligent Information and Database Systems* 5(1), 201–228 (2012)
8. Tanudjaja, J., Mui, L.: Persona: A contextualized and personalized web search. In: *Proc. 35th Hawaii International Conference on System Sciences*, Big Island, Hawaii, p. 53 (January 2002)
9. Trajkova, J., Gauch, S.: Improving ontology-based user profiles. In: *Proceedings of the 8th Conference of Recherche d'Information Assistée par Ordinateur*, April 26–28, pp. 380–389. University of Avignon, Vaucluse (2004)
10. Wen, J., Lao, N., Ma, W.Y.: Probabilistic model for contextual retrieval. In: *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Sheffield, United Kingdom, pp. 57–63 (August 2004)
11. Tamine, L., Boughanem, M., Zemirli, W.N.: Exploiting multi-evidence from multiple user's interests to personalizing information retrieval. In: Badr, Y., Chbeir, R., Pichappan, P. (eds.) *IEEE International Conference on Digital Information Management (ICDIM 2007)*, Lyon, France, pp. 7–12. IEEE Engineering Management Society (October 2007)
12. Micarelli, A., Gasparetti, F., Sciarone, F., Gauch, S.: Personalized search on the World Wide Web. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) *Adaptive Web 2007*. LNCS, vol. 4321, pp. 195–230. Springer, Heidelberg (2007)
13. Min, J., Jones, G.J.F.: Building user interest profiles from Wikipedia clusters. In: *The Workshop on Enriching Information Retrieval (ENIR 2011) at Special Interest Group on Information Retrieval (SIGIR)*, Beijing, China (July 2011)
14. Lalmas, M., MacFarlane, A., Rüger, S.M., Tombros, A., Tsirikia, T., Yavlinsky, A. (eds.): *ECIR 2006*. LNCS, vol. 3936. Springer, Heidelberg (2006)
15. Maleszka, M., Mianowska, B., Nguyen, N.-T.: A heuristic method for collaborative recommendation using hierarchical user profiles. In: Nguyen, N.-T., Hoang, K., Jedrzejowicz, P. (eds.) *ICCCI 2012, Part I*. LNCS (LNAI), vol. 7653, pp. 11–20. Springer, Heidelberg (2012)
16. Stermsek, G., Strembeck, M., Neumann, G.: User profile refinement using explicit user interest modeling. In: *GI-Jahrestagung Conference*, pp. 289–293. Technical University in Berlin (2007)
17. Esparza, S.G., O'Mahony, M.P., Smyth, B.: Mining the real-time web: a novel approach to product recommendation. *Knowledge-Based Systems* 29, 3–11 (2012)
18. Formoso, V., Fernandez, D., Cacheda, F., Carneiro, V.: Using profile expansion techniques to alleviate the new user problem. *Information Processing and Management* (2012), <http://dx.doi.org/10.1016/j.ipm.2012.07.005>