International Journal of Computer Networks and Communications Security

VOL. 3, NO. 1, JANUARY 2015, 11–15
Available online at: www.ijcncs.org
E-ISSN 2308-9830 (Online) / ISSN 2410-0595 (Print)

Distributed Backup of User Profiles for Information Retrieval

ABDELBAKI Issam[1], CHARKAOUI Salma[2], LABRIJI Amine[3] and BEN LAHMAR El habib[4]

[1, 2, 3, 4] Faculty of Sciences Ben M'SIK, Department of mathematics and informatics, Casablanca, Morocco

E-mail: [1]i.abdelbaki@gmail.com, [2]charkaoui.salma@gmail.com, [3]labriji@yahoo.fr, [4]h.benlahmer@gmail.com

ABSTRACT

The information research systems tends mainly to model the user according to profile and then integrate it into the chain of access to information, to better meet their specific needs. Given the large number of user profiles available on the internet, the safeguarding becomes problematic. This paper presents a technic of safeguard and of implicit construction of the user profile that is part of a distributed backup approach and a formal construction method using the user behavior as a source for predicting implicitly its need.

Keywords: *User profile, Formal context, Personalization, Information research systems.*

1 INTRODUCTION

The generalists information research models are based on the assumption that the user need is represented by its request, thus, for a given query, the information research systems (IRS) return the same results list, however users have different information needs. Work is now moving towards a broader definition of the user. It is a stream of research that seeks the implementation of user-centric systems by representing him by a profile.

The Analysis of user behavior reveals particular importance. Indeed, it is with full knowledge of how the user will elaborate his strategies for information research, that it will be possible to propose to him the significant information for his research. The modelisation of profiles and how to adapt them to different users who do not have a clear idea of the information they seek, allows us to provide personalized access to content of scientific papers based on the exploitation of the user profile.

However, with the significant growth of the number of web user, the storing the user profile has become problematic. Generally, the information search systems store the users profiles in a central knowledge base, however the user must identify themselves to determine their profile, other systems store the profile in the user but if he changes his workstation or he deletes the historic of his navigation, the system loses his profile. Other parts, the use of profiles of other users with the same area of interest appears interesting.

So, with the event of peer-to-peer (P2P) systems and their deep exploitation in sharing media files, motivated us to operate such architectures to create a user profile. The aim is that the information research system uses the current user profile and detects its area of interest in order to use the profiles of users with the same area of interest, such moneys are stored in a distributed manner among users.

2 MODELING THE USER

Without user model, an information search system will behave exactly the same way with all users, but these are different: they have different knowledge, different preferences and needs and different interest centers. All of these variations can be grouped under the user profile term.

Different definitions have been proposed of user profile, according to [10] a user profile (or user model) is a set of data concerning the user of a computer service. It is a source of knowledge that contains acquisitions on all aspects of the user that can be useful for system behavior. The goal of the

personalization of the information consists on modeling the user in the form of a profile, and then integration of the latter in the process of access to information.

The user modeling is a process at different stages namely, a naive representation of interests centers is based on keywords, as in the case of web portals MyYahoo, InfoQuest, etc. There are other more elaborate representations to illustrate the Interests centers of the user. [2] and [3] represent the Interests centers as vectors of terms weighted, on the other hand [4] present them semantically according to weighted concepts of a general ontology, or as matrices of concepts by [5].

[2] and [3] proposed a modeling of the user profile in a class of vectors each of which represents a center of interest of the user, thus, the classes centroids represent the user interest centers. The Semantic representation approaches exploit a reference ontology for representing user Interests centers as vectors of weighted concepts of the ontology used. We quote the hierarchy of concepts of "Yahoo" or of ODP as sources of evidence most often used in this type of approach. [4] built the user profile on a technique of supervised classification of documents deemed relevant according to a measure of vectorial similarity with ontology concepts of the ODP. This classification allows on multiple search sessions, to associate with each concept of the ontology, a weight calculated by aggregating the similarity scores of documents classified under this concept. The user profile will consist of all the concepts with the highest weights representing the user interests centers. On the other hand [11] operate simultaneously Interests centers of the user represented according to vectors of weighted terms and the hierarchy of concepts "Yahoo". The user profile will be composed of contexts; each context is formed of adequate concepts to research and concepts to exclude from the search.

A matrix representation of the user profile is adopted in [5], the matrix is constructed from the search history of the user incrementally, in order to establish categories representing the Interests centers of the user and the terms associated weighted reflecting the degree of interest of the user for each categories.

Once the choice of representation is made, the phase of profile's construction is the collection of information that represent it and this in an explicit way, based on information provided by the user [6], for example, when the user views a document, it indicates his opinion on the degree of relevance of the document with respect to to his request, or implicitly, from the consulted documents and the user behavior (time reading a document, saving, printing, etc.) [7].

3 DISTRIBUTED USER PROFILE

We propose architecture of distributed backup of the user profiles represented by Figure 1. The goal is to generate profiles and save them in the corresponding user. Only addresses and categories of the user are stored in the knowledge base of our IRS, thus each profile is referenced by all of these categories and accessible via the address of the user.

Furthermore, when a user submits a query, the IRS extracts the concepts of the query in order to infer its categories (a concept is a category for the ODP ontology). Then, it uses all the profiles of users with one of the categories of the current user. So the IRS can use all the recovered profiles including profile of the current user, in one of the access to information process (reformulate the query, sort results …).

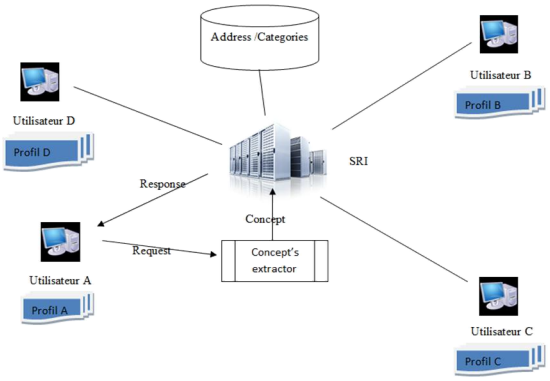

Fig. 1. General Architecture

In that section we detail main axes of our approach, namely our extraction method of categories of the request using the ODP ontology then we present the different phases of construction used of the user profile.

3.1 Extraction of categories

The goal is to extract all the concepts related to the query using domain ontology ODP (Open Directory Project). It is regarded as a source of semantic knowledge in our process of building the user profile.

Each category defines a concept that represents an area of interest of a user. We use a vector representation of all categories, so we extract the concepts of the query by a search in the vector space using a vectorial similarity measure between

vectors representing all categories of the ODP noted V(Ci) and the vector representing the query noted V(R).
The article [1] describes in detail our concept extraction process.

3.2 Construction of the user profile

As part of our work, we need user profiles for the meta-search engine, so we will focus on two information's, namely the relationship between the concepts of the query and documents and the relationship between the concepts of the query and the search engines. We use a formal approach using the user behavior as a source for predicting implicitly its need. We distinguish three main phases, the first phase is the acquisition of information from the browsing history of the user, the second is the construction of the formal context using data retrieved in the previous step. The third is the generation of profile from formal contexts previously generated.

3.2.1 Acquisition of users data

This phase is to collect relevant information to instantiate the user's profile. We focus on user interactions with the system. Indeed, the system saves in the log files the historic of user interactions, namely the query, the weighted concepts related to the query, the consulted documents and search engines associated to this documents. Indeed, when the user enters a query, he consults certain documents, so search engines that gave as results these documents is deduced. These search engines and documents are called assets in relation to this request.

To summarize, each request has a list of weighted concepts and a set of search engines and active documents in relation to the query.

3.2.2 Generation of formal contexts

This is an intermediate step that involves manipulating the history of users in order to generate subsequently the knowledge's. These latter will be stored in our system to provide the necessary elements to define the user's profile. Formal concept analysis (FCA) seeks to study the concepts when they are formally described to make them precisely defined.

The AFC allows to classify within formal concepts subset of concepts and its documents and search engines active. We take O a set of objects, P a set of property and R a binary relation between P and O. A formal context is defined by the triplet (O, P, R). The elements of O are called objects and the elements of P are known as context properties. To express that an object o of O is related to a property p of P, we write oRp. This means that object o has the property p.

In our case, concepts are objects, the properties are either active documents or active search engines, so we define two types of context:

- Context Document Concept "CDC": defines a relationship between a set of weighted query concepts (objects) and a set of documents (property).
- Context Engine Concept "CEC": defines a relationship between a set of concepts (objects) and a set of motors (property).

In our case, we say that an object Oi has the property Pj when this latter is always presents in the presence of the object Oi. It can be represented by a matrix where 1 means that the object Oi has the property Pj and 0 otherwise.

Table 1: Example of a Matrix Showing the Relationship between Object and Property.

	O1	**O2**	**O3**	**O4**	**O5**
P1	1	1	1	0	0
P2	1	0	0	1	1
P3	0	1	1	1	1
P4	1	1	0	1	0

3.2.3 Generation of user profiles

From contexts CEC and CDC we have two types of profile; the first is the link between all concepts weighted of past queries and search engines asset called "Profile Engine Concept" (PEC), the second is the link between weighted concepts of past queries and the active documents called "Concept Document Profile" (CDP), they are defined as follows : ({m1, . . . , mi}; {c1, . . . , cj}), respectively, ({d1, . . . , dt}; {c1, . . . , ck}), such as {m1, . . . , mi} is a set of search engines that have in common the set of concepts {c1, . . . , cj} and {d1, . . . , dt} is a set of documents that have in common all the concepts {c1, . . . , ck}.

All profiles represent a cover, in our case, we have two types of coverage, one for PEC denoted C1 and the other for CDP denoted C2, ces deux this two covers represent our knowledge base generated during the learning phase denoted B(C1,C2).

In Table 1 objects {O1, O2, O4} have the properties {P2, P3, P4}, so we can define a profile P = ({O1, O2, O4}, {P2, P3, P4}).

Example

Suppose for a given query, IRS extract the concepts (C1, C2, C3). The IRS consults its knowledge base to retrieve the list of addresses (A1, A2) of connected users with one of the concepts (C1 or C2 or C3), so he uses their profiles to return to the user the results list. We consider that the user has viewed some documents (D1-D2), since the engines (E1-E3-E4) gave in results these documents, then these search engines and these documents are considered active with the concepts of the application previously extracted.

We schematize this example by the following Figure.

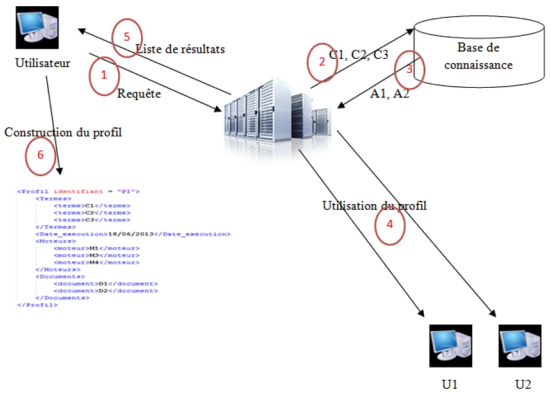

Fig. 2. Distributed backup example

4 CONCLUSION AND PROSPECTS

We presented through this paper a method for distributed backup of user profiles. He is inspired from the peer-to-peer model where a node can be both a client and a server, in our case the user shares his profile and uses the profiles of other users belonging to his field of interest. We use a formal representation method of the user profile.

We plan to use our backup and construction method of the user profile to classify the results in our meta-search engine.

5 REFERENCES

[1] I.Abdelbaki, E.Benlahmar, E.Labriji, Z.Rachik, "Automatic Extraction of Concepts of the Request Submitted to the IRS Based on Ontology", International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 8, August 2013.

[2] J. Gowan. "A multiple model approach to personalised information access". Master thesis in computer science, Faculty of science, Université de College Dublin, February, 2003.

[3] Sieg, B. Mobasher, R. Burke, G. Prabu, and 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.

[4] V. Challam, S. Gauch, A. Chandramouli, "Contextual Search Using Ontology-Based User Profiles", Proceedings of RIAO 2007, Pittsburgh USA, 30 may - 1 june 2007.

[5] F. Liu, C. Yu, and W. Meng. "Personalized web search for improving retrieval effectiveness". IEEE Transactions on Knowledge and Data Engineering, 16(1) :28–40, 2004.

[6] F. Maghoul, C. Chang, "contextual search at the point of inspiration", In CILM '05: Proceedings of the 14th ACM international conference on Information and knowledge management, New York, NY, USA, pp. 816±823, October 2005.

[7] S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-based personalized search and browsing". Web Intelligence and Agent Systems», 1(3-4) , pp. 219± 234, 2003.

[8] H. Fu, E. M. Nguifo, "Etude et conception d'algorithmes de génération de concepts formels", Revue Ingénierie des Systèmes d'Information, vol. 9, no 3-4, p. 109–132, Hermès-Lavoisier, 2004.

[9] A. L. Floc'h, C. Fisett, R. Missaoui, P. Valtchev, R. Godin, " JEN : un algorithme efficace de construction de générateurs pour l'identification des règles d'association", Numéro spécial de la revue des Nouvelles Technologies de l'Information, Vol. 1 No. 1, Editions Cépaduès, p. 135–146, 2003.

[10] W. Wahlster, A. Kobsa, "Dialogue-based user models", In Proceedings of IEEE, Vol. 74(7), pp. 948-960, 1986.

[11] A. Sieg, B. Mobasher, R. Burke, "Web search personalization with ontological user profiles", CIKM'07, Proceedings of the sixteenth ACM conference on information and knowledge management, ACM, New York, NY, USA, p. 525-534, 2007.

[12] R. Mghirbi, K. Arour, Y. Slimani et B. Defude, "Un modèle comportemental d'interclassement de résultats dans un système de recherche d'information P2P", Actes du XXVIII° congrès INFORSID, Marseille, mai 2010.

[13] P. De Bra, A. Kobsa, D. Chin, ”User Modeling, Adaptation, and Personalization”, 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010.