Modeling Ontology-based User Profiles from Company Knowledge

Silvia Calegari, Matteo Dominoni
Department of Informatics, Systems and Communication (DISCo)
University of Milano-Bicocca
Milan, Italy
e-mail: {calegari,dominoni}@disco.unimib.it

***Abstract*—This paper presents a model based on the ontology paradigm to represent the knowledge of users who work in companies. The model guarantees to follow specific guidelines that can be used in defining user knowledge. The goal is to improve the development of processes and applications aimed at leveraging the information stored in companies to support users during their interactions with the system by considering all the advantages of the ontology-knowledge.**
This research was developed in the KBMS 2.0 project where an advanced portal has been defined with the aim to manage and search information effectively. The model to represent user knowledge has been defined by analysing the information of the KBMS 2.0 project stored in the database.

User Profile, Ontology, Company Knowledge.

I. INTRODUCTION

In the last few decades, many research activities have focused on the management of the growing amount of information on the Web or in any big repository. There are several problems that can afflict this issue, ranging, for example, from architectural aspects, i.e., how to define a robust architecture for addressing issue of scalability, to personalization aspects, i.e., how to model user profiles to represent the user's interests and preferences. The former allows to help systems retrieve information faster, whereas the latter is used to filter knowledge which is not related to the user's interest needs.

In this paper, we focus on the definition of user profiles in the context of big companies, where thousands of pieces of information are stored in the repository. In this scenario, the problem is to support the user during his/her research activities, not only with the local search engine for retrieving relevant documents, but also with all the knowledge accessible in the company portal, such as news, emails, external links, etc., where the information is geographically distributed in the several locations of the company. This means that in a big company, besides the standard problems with a search engine on the Web, there are complex problems due to the several information access points of a user. In addition, in a company a user belongs to a hierarchical structure, where for each level users have a different role, meaning different access to the information. Thus, the definition of processes that exploit the knowledge represented in the user profile assumes a key role for managing knowledge in companies.

The accurate definition of a user profile plays a central role in effective approaches to personalization: only if a user profile faithfully represents the information related to a user, can a system rely on it. Three main activities have to be considered during the building of a user profile, such as: (1) the identification of the knowledge which represents the user's interests, (2) the choice of a formal language used to represent this knowledge, and (3) a strategy for updating it. Regarding the first point, in the literature explicit and implicit approaches [1] are adopted in order to capture the user's interests and preferences. With the explicit approach the user must explicitly specify his/her preferences to the system by filling in questionnaires and/or by providing short textual descriptions. With the implicit approach, the user's preferences are automatically gathered by monitoring the user's actions, thanks to the use of click-through data analysis, log analysis, etc. The second point is on the choice of suited data structures for modeling a user profile; in the literature bags of words, vectors, graph-based representations, and some external knowledge sources (i.e., WordNet, or Open Directory Project) have been mainly used to define users' profiles. Ontologies are a recent powerful tool for knowledge definition, modeling and representation; and they allow to give a more structured and expressive knowledge representation with respect to the above mentioned approaches. In fact, they allow to enrich the expressiveness of the information represented in a profile by using formal languages like Resource Description Framework Schema (RDFS) or Ontology Web Language (OWL). The third point is the most complex as it has to guarantee a dynamic user profile, where the knowledge is always updated and related to the user's recent interests and preferences. The objective is to support the processes that exploit this fresh information to help users during their interactions with the system.

In the literature, the existing models that build user profiles based on ontologies are mainly focused on approaches either relying on data mining techniques [2] or adopting external reference knowledge [3] to capture the meaning of the user's preferences that are represented in RDFS or OWL. With the use of formal ontological languages it is possible to define user profiles, thanks to the adoption of suitable logical

 ISBN: 978-1-61208-334-6

language constructs, but they do not give explicit guidelines on how to use them in order to solve specific tasks. The Ontology Design Patterns (ODP) [4] are a recent solution to make ontology as reusable solutions, and they are defined as a "*reusable successful solution to a recurrent modeling problem*". ODP are aimed at reducing mistakes in ontologies, detecting uncovered requirements, improving qualities of produced ontologies, etc. [5].

In this paper, for the first time in the literature we make use of the ODP solution to define the skeleton of user profiles that can be reusable in any domain where users are involved. Our objective is to represent a richer knowledge of user profiles than the existing models previously defined that mainly acquire initial information from textual information or from a set of keywords related to the user's interests. To help us in the definition of the ODP for user profiles, we considered the knowledge gained from a research project with the Enel SpA energy company. In this case study, user knowledge is obtained by analysing the database of the company where only the portion of information useful for the definition of some crucial parts of the ODP-user profile is extracted. In addition, we have completed same portions of the defined ODP-ontology by considering the basic notion of the ODP solution, where it is possible to compose pieces of knowledge by considering other ODPs previously defined. An ODP is used across ontologies to define a richer knowledge based on ontologies, whereas an ontology is used across applications. Our intent is to obtain an accurate ODP-user profile that can be used as a model for representing the user knowledge in any enterprise. The next phase of this work will consider the validation of the obtained user profile; at the moment this aspect is out of topics with respect to the goal of this paper.

The paper is organised as follows. Section II describes how the main pieces of knowledge for defining the ODP-user profile have been acquired. Section III presents the generic ODP-user profile for companies. Finally, in Section IV conclusion and future work are given.

II. TOWARDS THE DEFINITION OF A ODP-USER PROFILE

This section gives the steps performed to elicit knowledge used to define the ODP-user profile. We started from the knowledge stored in the database of the KBMS 2.0 project (KBMS 2.0: Knowledge-Base Management System 2.0 is a research project that has involved the University of Milano-Bicocca, Italy, together with the Enel SpA energy company) [6]. The KBMS 2.0 project is aimed at providing an advanced knowledge-based portal to manage knowledge innovatively, with the objective to support people in all the phases of their searches and also in the emergence of new knowledge in the system. This project is based on the open source edition of the Liferay portal and is based on several modules (by developing new modules or by extending some existing ones) in order to satisfy the requirements of the Enel SpA stakeholders. The most important modules developed for the KBMS 2.0 portal are: a workflow to guarantee a certified quality of the information, a personalised search engine to retrieve relevant documents for users, a grid of navigation to allow users immediate access to documents classified, thanks to the support of specific categories, a newsletter to send and filter news according to specific rules, spot-news to highlight emergent news, and my-links to personalise access to external web-links. All the information used by these modules is stored in the database. In detail, the Relational Database Management System (RDBMS) Oracle 10g version has been used for the project by defining a total of 253 tables. Most of the tables have been automatically defined by Liferay during its installation, while some of them are not used for the project such as tables on market-place, wikipedia, kaleo, etc. Other tables, instead, have been introduced to support newly defined modules for activities not defined in the standard edition of Liferay.

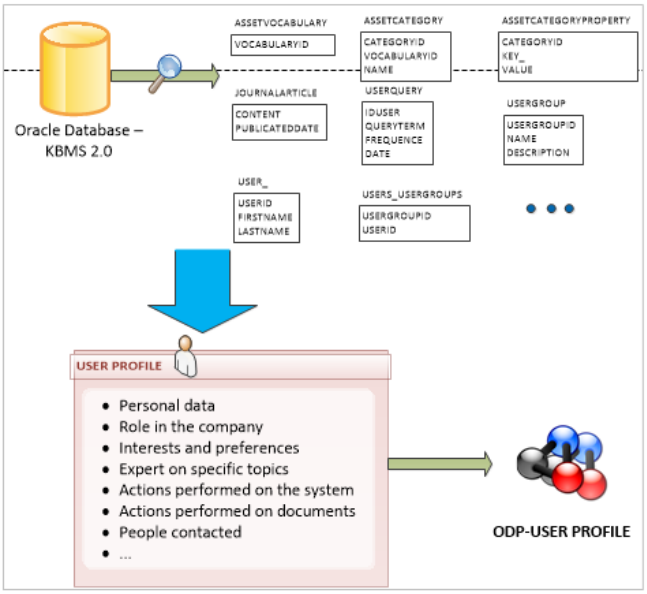

Figure 1. Logical path for the definition of a ODP-User Profile.

The database has been analysed to design for the extraction of some portions of knowledge useful in defining the skeleton of the user profile presented in Section III. Figure 1 shows the phases considered for our analysis that are: (1) view the tables stored in the database in order to consider only the subset related to the knowledge of users, (2) list the user's knowledge, and (3) define the ODP-user profile based on point (2).

The list of user knowledge assumes a key role for the definition of the user profile model, which has been grouped in the following macro-areas:

- **Personal data:** describes personal user characteristics, such as name, address and age.
- **User's company data:** describes user characteristics of the company, such as login, password, nickname and job/role.

 ISBN: 978-1-61208-334-6

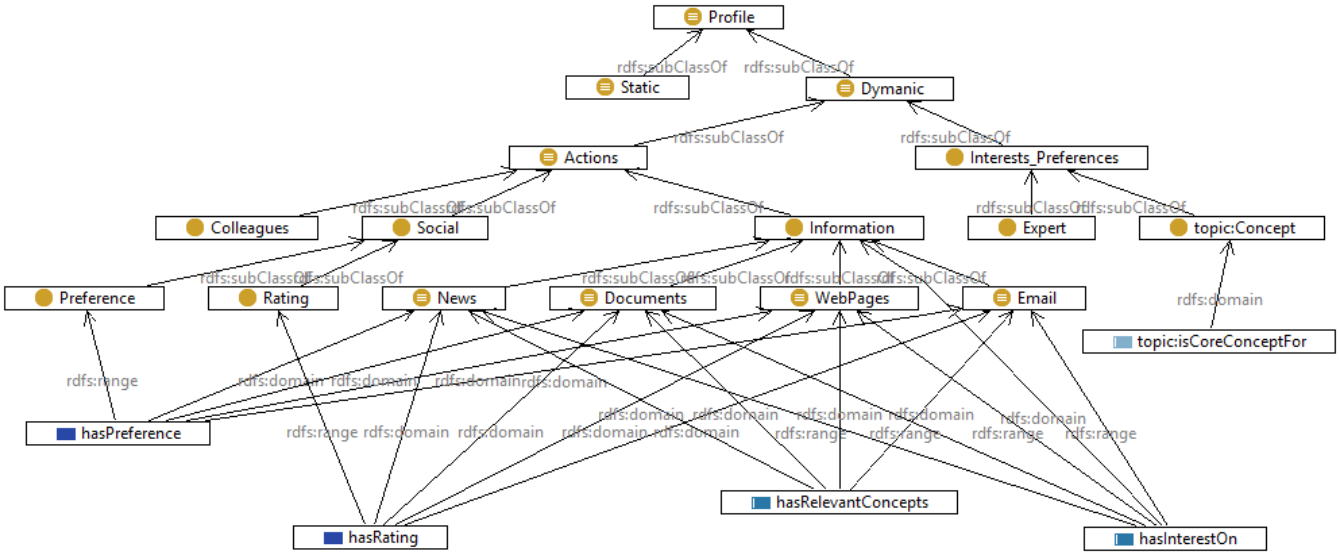

Figure 2. Example of a TWO-COLUMN figure caption: (a) this is the format for referencing parts of a figure.

- **Company data:** describes company information, such as name and address.
- **Interests and preferences:** describes user interests and preferences on topical arguments. They can be obtained by analysing, for example, the documents saved or viewed by the user.
- **Expert on specific topics:** indicates which topics a user is considered as an expert. Several processes can be adopted in order to establish the user's expertise, such as methodologies of voting or use of metrics defined in the literature.
- **Actions performed on the information:** monitors the actions on the information stored in the system, such as the pages/documents viewed, the time spent on a page/document, emails sent, etc.
- **Social Actions:** monitors the user's action on the system or documents, such as a methodology of rating on documents retrieved by a search engine, or on people contacted for suggestions, directed preferences on the information such as the use of like/not-like.
- **People contacted:** monitors the people who are working in the company that have been contacted by a user, for example, by emails or by using social applications.

The knowledge considered from these macro-areas can be gathered by using both explicit and implicit approaches, as explained in Section I. In this phase, it is not important to understand the complex processes that exploit the data stored in the log files, generally used to monitor the user's actions (i.e., actions on information and social actions), in order to understand what information can be of interest to a user. Our objective is to analyse such preferred information to design the ODP model used for establishing the knowledge that has to be represented in a generic user profile (see Section III).

III. ODP-USER PROFILE FOR COMPANIES

The ODPs are used in the community to define reusable logical models with the aim to design conceptualizations of knowledge. In this section, we present the first ODP model defined to represent the knowledge of users who work in companies. Figure 2 shows the class diagram of the user profile-ODP where the main concepts are reported. The main concepts are organized by the taxonomic relation IS-A, whereas the other knowledge is represented by different relations (i.e., *ObjectProperties*) as described in this section. For the sake of readability, we have not introduced the pre-defined *Instances* (e.g., *male* and *female* instances for the concept *Gender*), and the relations (domain and range) established with the *DatatypeProperties* and the *ObjectProperties*.

The user profile is logically divided into two logical areas that are: *Static* and *Dynamic* with the intent of representing the two key aspects of the user's profile knowledge. The *Static* class represents long-term interests of user knowledge. The *Dynamic* class represents short-term interests of user knowledge, where the changes can happen every time that a user interacts with the system. This class identifies the most interesting part of the ODP-user profile, since it needs to represent the interactions of a user with the information stored in the system and the people who work in the company. Thus, we have classified the information into four categories that are: *Documents*, *Webpages*, *Email*, and *News*. For these categories the following common *ObjectProperties* have been defined:

hasInterestedOn, *hasPreference*, *hasRating*, and *hasRelevantConcepts:topicConcept*.

The *hasInterestedOn* defines user interests on specific information stored in the system, the *hasPreference* defines

the social actions on the information related to the *like/not-like* preferences, and the *hasRating* defines the social actions on the information related to the rating preferences given, for example, by the *star rating* method.

The *hasRelevantConcepts:topicConcept* deserves a closer look, as it links the documents or email or news or webpages with the relevant concepts extracted from them. The relevant concepts are not defined, as bags of words or vectors, etc., as they are in many works presented in the literature (see Section I). In our work, the relevant concepts are represented as ontology-based knowledge, not extracted by external reference knowledge (i.e., Yet Another General Ontology (YAGO) or Open Directory Project or domain ontologies). To define the knowledge for representing the relevant concepts of user preference information, we have used the ODP schema on *Topic* [7] download from the official ODP's repository [4] with the intent of following the logic of the ODP models, i.e., to reuse pre-defined logical models in new ODP models.

In the end, the class *Colleagues* defines the users who work in the same company that have interacted with the user. The idea is to create a network of users to establish, for example, people who share common interests.

A. How to use the ODP-User Profile

When the ODP-User Profile is defined, it is possible to use it in order to represent the knowledge of a user profile represented as an ontology. The next phase consists in modeling an ontology in the OWL language that is an instance of the ODP-User Profile. This step is easily performed by importing the ODP schema in the configuration part of the OWL file as follows:

```
<?xml version="1.0"?>
<rdf:RDF
  ...
  xmlns="http://www.ontologydesignpatterns.org/cp/
  examples/userprofile/enel.owl#"
  xmlns:up="http://www.ontologydesignpatterns.org/cp/
  owl/userprofile.owl#"
  xml:base="http://www.ontologydesignpatterns.org/cp/
  examples/agentrole/enel.owl">
    <owl:Ontology rdf:about="">
      ...
      <rdfs:comment>It encodes the following:User Profile
       in the Enel SpA company</rdfs:comment>
     <owl:imports rdf:resource=
      "http://www.ontologydesignpatterns.org/cp/owl/
      userprofile.owl"/>
    </owl:Ontology>
    <up:Profile>
       ...
    </up:Profile>
</rdf:RDF>
```

IV. Conclusion and Future Work

In this paper, we have presented the ODP model for representing the user profile logical schema. An ODP model is a new frontier of research, as it allows to provide specific guidelines for defining knowledge based on specific logical schemas. In fact, the defined ODP-user profile can be used to represent the knowledge of users in companies to facilitate the definition of applications and of processes by exploiting all the well known advantages of the ontology paradigm. The ODP-user profile has been designed by taking support from the knowledge stored in the database of the KBMS 2.0 project. In this project, an advanced portal has been defined with the aim of satisfying the user's information needs by providing a certified quality of information and by supporting a user during his/her searches, with the goal of retrieving relevant documents faster.

In the future, we plan to improve the ODP-user profile by including more details on the defined concepts and by defining axioms to classify user interests on information and on people to support the processes that are defined for this goal. Furthermore, we will consider its validation in order to establish the good quality of the obtained ODP-user profile.

Acknowledgment

The authors would like to thank all the people who allowed us to make an in depth study for the development of the KBMS 2.0 system.

References

[1] J. Teevan, S. T. Dumais, and E. Horvitz, “Personalizing search via automated analysis of interests and activities”. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, ser. SIGIR ’05. New York, NY, USA: ACM, 2005, pp. 449–456, doi:10.1145/1076034.1076111.

[2] X. Tao, Y. Li, and N. Zhong, “A personalized ontology model for web information gathering”. IEEE Trans. on Knowl. and Data Eng., vol. 23, April 2011, pp. 496–511, doi: 10.1109/TKDE.2010.145.

[3] S. Calegari and G. Pasi, “Personal ontologies: Generation of user profiles based on the yago ontology”. Inf. Process. Manage., vol. 49, no. 3, 2013, pp. 640–658, doi: http://dx.doi.org/10.1016/j.ipm.2012.07.010.

[4] Ontology Design Pattern, website. [Online]. Available: http://ontologydesignpatterns.org/wiki/Main Page (October, 2013)

[5] V. Presutti, E. Blomqvist, E. Daga, and A. Gangemi, “Pattern-based ontology design”. In Ontology Engineering in a Networked World, M. del Carmen Suarez-Figueroa, A. Gomez-Perez, E. Motta, and A. Gangemi, Eds. Springer, 2012, pp. 35–64, doi:10.1007/978-3-642-24794-1_3.

[6] S. Calegari, M. Dominoni, and E. Panzeri, “Towards the design of an advanced knowledge-based portal for enterprises: the KBMS 2.0 project”. The 27th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems, 2013, Unpublished.

[7] Ontology Design Pattern, Topic schema, website. [Online]. Available:http://www.ontologydesignpatterns.org/cp/owl/topic.owl (October, 2013)

 ISBN: 978-1-61208-334-6