Multilingual Ontology-based User Profile Enrichment

Ernesto William De Luca, Till Plumbaum, Jérôme Kunegis, Sahin Albayrak
DAI Lab, Technische Universität Berlin
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
{ernesto.deluca, till.plumbaum, jerome.kunegis, sahin.albayrak}@dai-labor.de

ABSTRACT

In this paper, we discuss the possibility of enriching user profiles with multilingual information. Nowadays, the English language is the de facto standard language of commerce and science, however users can speak and interact also in other languages. This brings up the need of enriching the user profiles with multilingual information. Therefore, we propose to combine ontology-based user modeling with the information included in the RDF/OWL EuroWordNet hierarchy. In this way, we can personalize retrieval results according to user preferences, filtering relevant information taking into account the multilingual background of the user.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms

RDF/OWL, Web 2.0, Multilingualism, EuroWordNet

Keywords

Multilingual Semantic Web, User Modeling

1. INTRODUCTION

At present most of the demand for text retrieval is well satisfied by monolingual systems, because the English language is the de facto standard language of commerce and science. However, there is a wide variety of circumstances in which a reader might find multilingual retrieval techniques useful. Being able to read a document in a foreign language does not always imply that a person can formulate appropriate queries in that language as well. Furthermore, dealing with polysemic words seems to be more difficult in multilingual than in monolingual retrieval tasks.

Every text retrieval approach has two basic components: the first for representing texts (queries and documents) and the other for their comparison. This automated process is successful when its results are similar to those produced by human comparison between queries and documents. Queries and documents often differ from its length however. While the query is often quite short, documents might be up to hundreds of pages long. Moreover, users frequently adopt a

vocabulary that is not contained in the documents, known as the paraphrase problem.

Multilingual Retrieval. When working in a multilingual environment, words have to be disambiguated both in the native and in the other languages. In this case the combination of multilingual text retrieval and word sense disambiguation (WSD) approaches is crucial [2]. In order to retrieve the same concept in different languages, some relations between the searched concept and its translations have to be built. WSD is used to convert relations between words into relations between concepts; sense disambiguation can be acquired for words, but it is more difficult for documents. To have accurate WSD, we need a larger coverage of semantic and linguistic knowledge than is available in current lexical resources.

Because we focus on multilingual concepts, we decided to use EuroWordNet [6], a variant of the most well-known available lexical database WordNet. In previous work, we extended the RDF/OWL WordNet representation [5] for multilingualism, leading to our own RDF/OWL EuroWordNet representation [3].

Ontology-based User Modeling. With the advent of the Web 2.0 and the growing impact of the Internet on our every day life, people start to use more and more different web applications. They manage their bookmarks in social bookmarking systems, communicate with friends on Facebook[1] and use services like Twitter[2] to express personal opinions and interests. Thereby, they generate and distribute personal and social information like interests, preferences and goals [4]. This distributed and heterogeneous corpus of user information, stored in the user model (UM) of each application, is a valuable source of knowledge for adaptive systems like information filtering services. These systems can utilize such knowledge for personalizing search results, recommend products or adapting the user interface to user preferences. Adaptive systems are highly needed, because the amount of information available on the Web is increasing constantly, requiring more and more effort to be adequately managed by the users. Therefore, these systems need more and more information about users interests, preferences, needs and goals and as precise as possible. However, this personal and social information stored in the distributed UMs usually exists in different languages due to the fact that we communicate with friends all over the world. Also, today's adaptive systems are usually part of web applications and typically only have access to the information stored in that specific ap-

---

[1]http://www.facebook.com/
[2]http://twitter.com/

plication. Therefore, we enhance the user model aggregation process by adding valuable and important meta-data which leads to better user models and thus to better adaptive systems. For this reason, we propose a combination of RDF/OWL EuroWordNet within ontology-based aggregation techniques.

2. PROPOSED SEMANTIC USER MODELING AGGREGATION

RDF/OWL EuroWordNet opens new possibilities for overcoming the problem of language heterogeneity in different user models and thus allows a better user modeling aggregation. Therefore, we propose an ontology-based user modelling approach that combines mediator techniques to aggregate user models from different applications and utilize the EuroWordNet information to handle the multilingual information in the models. Based on this idea, we define some requirements that we have to fulfill.

Requirement 1: Ontology-based profile aggregation. We need an approach to aggregate information that is both application independent and application overarching. This requires a solution that allows us to semantically define relations and coherences between different attributes of different UMs. The linked attributes must be easily accessible by applications such as recommender and information filtering systems. In addition, similarity must be expressed in these defined relations.

Requirement 2: Integrating semantic knowledge. A solution to handle the multilingual information for enriching user profiles is needed. Hence, we introduce a method to incorporate information from semantic data sources such as EuroWordNet and to aggregate complete profile information. We decided to use an ontology as the conceptual basis of our approach to meet the first requirement explained above. Therefore a meta-ontology is used to link attributes of different UMs that contain equal or similar content.

The definition of a meta-model based on the meta-ontology can be divided into two steps. First, we define a concrete meta-model for a specific domain we want to work with, such as music, movies or personal information. The meta-model can be an already existing model, like FOAF[3] or a proprietary model that only certain applications understand. Next, we decribe how to connect multilingual attribute information stored in different user models.

3. MULTILINGUAL ONTOLOGY-BASED AGGREGATION

To enrich the user model with multilingual information, as described above, we decided to utilize the knowledge available in RDF/OWL EuroWordNet [3]. We want to leverage this information and use it for a more precise and qualitatively better user modeling. We treat the semantic external resources as a huge semantic profile that can be used to enrich the user model and add valuable extra information (see Figure 1). The aggregation of information into semantic profiles and user models is performed similarly to the approach described in [1], by using components that mediate between the different models. We extend this approach by using a combined user model, aggregated with the proposed ontology.
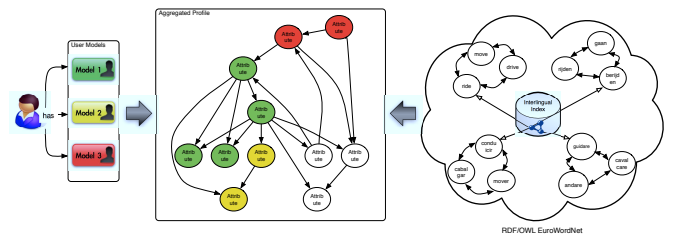
---

[3]http://www.foaf-project.org/

Figure 1: Integrating semantic knowledge about multilingual dependencies with the information stored in the user models.

To use the information contained in RDF/OWL EuroWordNet, we developed a framework that allows us to define several mediators that take the information from user models and trigger different sources in the Semantic Web for more information. These mediators are specialized components that read a user model and collect additional data from an external source.

4. CONCLUSION

In this paper, we presented the possibility of enriching user profiles with information included in the RDF/OWL EuroWordNet hierarchy to better filter results during the search process. This aggregated information can be used in our multilingual semantic information retrieval system that has been described in more details in [2]. In this work, we have shown that we can handle the high heterogeneity of distributed data, especially concerning multilingual heterogeneity, using aggregated user profiles that have been enriched with information contained in the RDF/OWL EuroWordNet representation. This gives us the possibility to personalize retrieval results according to user preferences, filtering relevant information taking into account the multilingual background of the user.

5. REFERENCES

[1] Shlomo Berkovsky, Tsvi Kuflik, and Francesco Ricci. Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction*, 18(3):245–286, 2008.

[2] Ernesto William De Luca. *Semantic Support in Multilingual Text Retrieval*. Shaker Verlag, Aachen, Germany, 2008.

[3] Ernesto William De Luca, Martin Eul, and Andreas Nürnberger. Converting EuroWordNet in OWL and extending it with domain ontologies. In *Proc. Workshop on Lexical-semantic and Ontological Resources*, 2007.

[4] Till Plumbaum, Tino Stelter, and Alexander Korth. Semantic web usage mining: Using semantics to understand user intentions. In *Proc. Conf. on User Modeling, Adaptation and Personalization*, pages 391–396, 2009.

[5] Mark van Assem, Aldo Gangemi, and Guus Schreiber. WordNet in RDFS and OWL. Technical report, W3C, 2004.

[6] Piek Vossen. Eurowordnet general document, version 3, final, 1999.