

# ERCIM “Alain Bensoussan” Fellowship Scientific Report

Fellow: Duchateau Fabien  
Visited Location: CWI, Amsterdam  
Duration of Visit: 9 months (01/02/10 to 31/10/10)

## I - Scientific activity

At CWI, I joined the INS2 team, which mainly works on Semantic Web and Information Retrieval research fields. My research activities at CWI involved the continuation of my earlier research topics and new cooperation with the INS2 team, namely in the Petamedia project. In this research project (<http://eventmedia.cwi.nl>), the overall goal is to provide a web-based environment that allows users to discover relationships among events and media. In this context, I decided to work on user's recommendations.

Recommendations aim at providing users with meaningful information such as possible events to participate in. This involves the analysis of user's profiles to determine which events could interest them. However, a user may have a profile on different social networks (on Facebook, Flickr, LastFM) since these networks fulfill different uses (e.g., sharing pictures, keeping in touch with friends). Thus, there is a need for integrating all profiles from the same user to discover his or her interests.

We have proposed an approach based on clusters. The main idea is to gather similar interests in a cluster according to terminology and semantics. We use an external resource (Wordnet dictionary) to match interests with a Wordnet concept. Then, we are able to find the closest common ancestor between two interests based on the Wordnet hierarchy. This common ancestor becomes a cluster label and related interests are linked to this cluster. Interests that have not been matched to a Wordnet concept are terminologically compared with all clusters' labels and added to a cluster if they match with its label. Finally, we compute a weight for each generated cluster to indicate its importance with respect to other clusters. This enables us to rank the clusters and detect which interest(s) are important for a given user.

Finally, we have extended this last work to answer this question: who in your social network can best answer your query ? Let us imagine that you need information about “interesting places in Amsterdam”, then maybe someone in your social network is able to provide you with such information (e.g., a Dutch acquaintance, or someone who visited the city). Due to the multiple user's profiles, integration of these profiles was required too. To do so, we are studying three different strategies that allow the integration of all profiles of the same acquaintance and match it against the query to detect if the acquaintance could answer this query. We expect to submit this work on October, 30<sup>th</sup> at DESWeb workshop.

Thanks to the financial support from ERCIM fellowship and CWI, I was able to attend several conferences where I could both present my results and learn the last trends in my current research topics.

## **II- Publication(s) during your fellowship**

### **Integrating and Ranking Interests From User Profiles**

Fabien Duchateau and Lynda Hardman

LUPAS 2010, workshop at ESWC (<http://www.personal-reader.de/lupas/>)

*Many websites allow their users to personalize their profiles. As users subscribe to many personalization websites, such as social networks or search systems, each user owns different profiles, which are seldom compatible. Yet, there is a strong need for comparing the profiles of different users to discover shared interests, e.g., by integrating all user profiles into a global one. In this paper, we propose a novel method for integrating and ranking user interests from various profiles. Our approach relies on the identification of high-level concepts around which similar user interests are clustered. We compute the weight of each cluster with respect to the other ones, thus enabling the ranking of the most shared user interests between user profiles.*

### **Measuring the Quality of an Integrated Schema**

Fabien Duchateau and Zohra Bellahsène

ER 2010 (<http://www.er2010.sauder.ubc.ca/index.php?p=program&ev=ts#ert08>)

*Schema integration is a central task for data integration. Over the years, many tools have been developed to discover correspondences between schemas elements. Some of them produce an integrated schema. However, the schema matching community lacks some metrics which evaluate the quality of an integrated schema. Two measures have been proposed, completeness and minimality. In this paper, we extend these metrics for an expert integrated schema. Then, we complete them by another metric that evaluates the structurality of an integrated schema. These three metrics are finally aggregated to evaluate the proximity between two schemas. These metrics have been implemented as part of a benchmark for evaluating schema matching tools. We finally report experiments results using these metrics over 8 datasets with the most popular schema matching tools which build integrated schemas, namely COMA++ and Similarity Flooding.*

## **III -Attended Seminars, Workshops, and Conferences**

### **Meeting for the Petamedia project**

11 April 2010, Sophia-Antipolis, France

### **Linking of User Profiles and Applications in the Social Semantic Web (LUPAS) and Extended Semantic Web Conference (ESWC)**

30 May – 3 June 2010, Heraklion, Greece

### **International Conference on Conceptual Modeling (ER)**

1-4 November 2010, Vancouver, Canada

I will attend this conference to present our paper.

## **IV – Research Exchange Programme (12 month scheme)**

Not applicable.