ERCIM "Alain Bensoussan" Fellowship Scientific Report

Fellow: Duchateau Fabien
Visited Location :   NTNU, Trondheim
Duration of Visit: 9 months

I - Scientific activity

I joined Trond Aalberg's team, which is specialized in the Digital Libraries domain. A challenging task deals with the conversion of existing cultural data into a semantic model called FRBR. The use of this model would result in improved user searches and navigation for instance. In addition, the notion of reusability for entities and their relationships is currently achieved through the Linked Open Data cloud (LOD), and librarians are interested in this reusable feature due to the management of large collections of data.

In this context, we have first proposed a method to discover FRBR entities from Web product descriptions. Our goal was to demonstrate that the FRBR model is not only suitable for library catalogs, but also for other collections such as those stored on the Web. Experiments with Amazon dataset clearly highlight this feature [3]. The next step consists in enriching this data by automatically discovering the LOD entity which corresponds to the generated FRBR entities (e.g., artistic work, author). To fulfill this goal, matching techniques have been used and our experiments show that our approach is able to find the correct LOD entity in most cases [1]. An online application summarizes the whole process for FRBRizing a Web product and linking the generated FRBR entities to Linked Open Data [7].

In libraries, the most common format is MARC and its alternative forms. Librarians are interested in providing a well-defined semantic to their records for improving user searches, which implies to convert the data into a semantic format such as FRBR. Due to the large amount of data, the conversion is a difficult and error-prone process. Our idea was to propose an intermediate format with the following properties: a method for identifying FRBR entities from MARC records, a semantic enrichment process using LOD, a correction and disambiguation step, a conversion from and back to MARC, and the design of metrics to check the quality of the conversion. The full method along with experiments using records from the National Library of Norway is described in a journal paper [2].

During my stay, I was also able to finalize two book chapters about schema matching [5, 8] and to collaborate to the writing of a demo paper in another domain [4].

II- Publication(s) during your fellowship

[1] Linking FRBR Entities to LOD through Semantic Matching
TPDL, to appear in 2011
Naimdjon Takhirov and Fabien Duchateau and Trond Aalberg

[2] FRBR-ML: A FRBR-based Framework for Semantic Interoperability
Semantic Web Journal, to appear in 2011
Naimdjon Takhirov and Fabien Duchateau and Trond Aalberg and Maja Zumer

[3] Supporting FRBRization of Web Product Descriptions
TPDL, to appear in 2011
Naimdjon Takhirov and Fabien Duchateau and Trond Aalberg

[4] A Cloud-based and Social Authoring Tool for Video
Document Engineering, to appear in 2011
Naimdjon Takhirov and Fabien Duchateau

[5] On Evaluating Schema Matching and Mapping (book chapter)
Schema Matching and Mapping, 2011
Angela Bonifati and Zohra Bellahsene and Fabien Duchateau and Yannis Velegrakis

[6] Who Can Best Answer a Query in my Social Network ?
DESWeb - Data Engineering meets the Semantic Web, 2011
Fabien Duchateau

[7] FRBRPedia: a Tool for FRBRizing Web Products and Linking FRBR Entities to DBpedia
JCDL, 2011
Fabien Duchateau and Naimdjon Takhirov and Trond Aalberg

[8] Tuning for Schema Matching (book chapter)
Schema Matching and Mapping, 2011
Zohra Bellahsene and Fabien Duchateau

III -Attended Seminars, Workshops, and Conferences

Thanks to the financial support of ERCIM and NTNU, I attended the JCDL conference in Ottawa to demonstrate the capabilities of our tool. In September, I may also attend the TPDL conference in Berlin to present one of our accepted papers.

IV – Research Exchange Programme (12 month scheme)

Not relevant.