

ERCIM “Alain Bensoussan” Fellowship Scientific Report

Fellow: Eugen Popovici

Visited Location : FhG IDMT, Ilmenau, Germany

Duration of Visit: 01/12/2009 – 31/08/2010

I - Scientific activity

In my second visiting period as an ERCIM fellow, between December 1st 2009 and 31st August 2010, I have visited the Semantic Music Technologies Group of the Fraunhofer Institute for Digital Media Technology (FhG IDMT). I have integrated the team of the ongoing GlobalMusic2one¹ project funded by the German Federal Ministry of Education and Research under the "KMU-Innovativ" program.

During this period I have studied Semantic Web technologies and more specifically Linked-Data based approaches for global music recommendation tasks. Additionally I continued to work on recommending locations for placing multimedia content within semi-structured documents within the context of the enrichme² (*ENRICHhing docuMENTS with annotated images*) project initiated in my first visiting period as an ERCIM fellow at the Center for Mathematics and Computer Science (CWI) in Amsterdam.

GlobalMusic2one

The GlobalMusic2one project is developing a new generation of hybrid search engines including new methods of music information retrieval and Web 2.0 technologies. It aims at reaching a better quality in the automated recommendation and online marketing of global music collections. Music recordings are automatically analysed by a self-learning software, which incorporates rhythm, melody and other characteristics. This allows an efficient and exact filing of new content into existing collections. The user may create new categories to allow the system to flexibly adapt to new musical forms of expression and regional contexts. A prototypical work flow consisting of semantic indexing, user interaction, model adaptation as well as search and recommendation functions are currently being implemented.

My contribution within the project concerned the study and development of automatic metadata enrichment and retrieval of music related semantic data in order to develop semantic-based and hybrid content-and-semantic-based music recommendation strategies. For this purpose, the Piranha Music Scout³ dataset used within the GlobalMusic2one project was linked with the Linked Open Data⁴ (LOD) cloud. About 40% of the dataset tracks were successfully linked to the LOD cloud using the tracks ID3 tags and the gnat-0.1-linux library⁵. The results of the matching process were expressed in RDF according to the Music Ontology Specification⁶.

¹ GlobalMusic2one http://www.globalmusic2one.net/en_summary.html

² enrichme <http://media.cwi.nl/enrichme/about.html>

³ Piranha Music Scout dataset <http://musicscout.piranha.de/>

⁴ Linked Open Data (LOD) <http://linkeddata.org/>

⁵ gnat-0.1-linux library <http://sourceforge.net/projects/motools/>

⁶ Music Ontology Specification <http://musicontology.com/>

A configurable semantic crawler prototype was developed in Python based on the RDFLib⁷ library. Starting from the MusicBrainz⁸ identifiers of the tracks and from the identified artists DBpedia⁹ pages a total of 3.7 millions triples were fetched.

A SPARQL endpoint that allows querying the crawled data for recommendation purposes was set up based on the 4store¹⁰ RDF database.

A graphical user interface combining advanced text-search capabilities and faceted navigation was provided and configured for the exploration of relevant music semantic data (artists, tracks, albums, musical genres, ...). The interface is based on the Longwell RDF browser¹¹.

We are currently investigating the content of the crawled data (in combination with GlobalMusic2one content-based recommendations) in order to devise a strategy for automatically identifying semantic patterns relevant for recommending musical items and for generating explanations about the recommended items. Future research involves the use of the identified patterns to build semantic web topical crawlers for the global music domain.

Other professional activities

- Reviewer for the Information Processing & Management - IP&M International Journal, Elsevier.
- Additional Reviewer for the Audio Mostly 2010 - the 5th Conference on Interaction with Sound, September 15-17, in Piteå, Sweden.

Presentations

- Fraunhofer IDMT Metadata Department Get-Together, 16th December 2009, Ilmenau "Presentation of my Research Topics and Ongoing Projects"
- GlobalMusic2one Consortium Meeting, 8th June 2010, Ilmenau "GlobalMusic2one and the Semantic Web"

II- Publication(s) during your fellowship

- E. Popovici, H. Großmann, C. Dittmar, "Exploring Linked Data Patterns for Global Music Recommendation" (paper under preparation).
- E. Popovici, L. Hardman, "Towards Automatic Enrichment of Structured Documents with Annotated Images: An Initial User Study", CWI Technical Report¹².

Abstract. Enriching text documents with images is a common, time consuming, and cognitive intensive task in the day to day activities such as writing a school report or publishing a new entry on a Blog. Automatic methods that assist the user to find relevant images for the document content and recommend where to place them within the document have the potential to reduce the user's time spent on searching, selecting and organizing the images. In this paper we present the results of an initial user study on how users enrich documents with annotated images. We have used Dutch articles covering contemporary history (1946-2007) of the Keesings Historisch Archief and manually annotated images of the 20th century of the Spaarnestad historical photo archive. Text and structure related features that could be used to support the automatic enrichment of structured documents with annotated images are identified and analyzed.

III -Attended Seminars, Workshops, and Conferences

⁷ RDFLib <http://www.rdflib.net/>

⁸ MusicBrainz <http://musicbrainz.org/>

⁹ DBpedia <http://dbpedia.org>

¹⁰ 4store RDF database <http://4store.org/>

¹¹ Longwell RDF browser <http://simile.mit.edu/wiki/Longwell>

¹² Current version available on-line at http://www.cwi.nl/~popovici/pdf/enrichme/enrichme_CWI_TR01.pdf