



ERCIM "ALAIN BENSOUSSAN"
FELLOWSHIP PROGRAMME



Scientific Report

First name / Family name	Simon Scerri
Nationality	Malta
Name of the <i>Host Organisation</i>	Fraunhofer Institute for Applied Information Technology
First Name / family name of the <i>Scientific Coordinator</i>	Sören Auer
Period of the fellowship	15/02/2014 to 14/02/2015

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

A. *Social Media Filtering based on Linked Open Data:*

Thematic Filtering: Investigation of Ontology-based (Semantic) Thematic Extension techniques, to extend the keywords/topics given an initial seed term. Public data from repositories on the Linked Open Data (LOD) cloud, specifically DBpedia, was used to identify semantically-related entities and return social media posts with links to text-based documents referring to one or more of the resultant entities. These efforts were supported by a Master student under the fellow's supervision.

Spatial Filtering: Investigation of the value of LOD content (DBpedia) to filter social media streams with respect to a seed term based on the origin of the geo-tagged posts, or posts from location-aware networks (e.g. FourSquare). These efforts were also supported by a Master student under the fellow's supervision.

The above techniques will be integrated as extensions to the RESA online tool¹. Future work will combine both approaches to enable LOD-driven spatio-thematic filtering of social media.

B. LOD-enhanced Sentiment Analysis of Social Media

Stock Market Prediction: Investigation of the impact of considering additional related entities when monitoring social media sentiment for stock market prediction. Using DBPedia as the knowledge base, the conventional company tracking monitoring was extended with DBPedia-derived known products and personnel for the company in question. An experiment that assigned different weights to the additional two entity types determined the optimal proportion for the three variables, and preliminary investigations indicate a higher correlation with stock market variations over the company-only based tracking. These efforts were also supported by a Master student under the fellow's supervision.

C. Ontology-based Longitudinal Analysis of Social Media (ongoing)

Semantic Clustering of Relevant Social Media Streams: In this effort, we have begun investigating whether semantically-annotated social media posts resulting from the described spatio-thematic filtering can be clustered in a non-conventional manner, combining the RDF-based data structure with Formal Concept Analysis. The hypothesis is that the resulting concept lattices can then be reduced to produce a series of time-stamped snapshots of relevant social media. These efforts are being supported by a PhD student who acquired funding to carry out this research under the fellow's supervision. Future work will include the investigation of longitudinal analysis over the resulting social media snapshots to identify and visualize trends for human consumption.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

- Extracting Information for Context-aware Meeting Preparation. Simon Scerri, Behrang Q Zadeh, Maciej Dabrowski, Ismael Rivera. Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014), Reykjavik, Iceland.

ABSTRACT: People working in an office environment suffer from large volumes of information that they need to manage and access. Frequently, the problem is due to machines not being able to recognise the many implicit relationships between office artefacts, and also due to them not being aware of the context surrounding them. In order to expose these relationships and enrich artefact context, text analytics can be employed over semi-structured and unstructured content, including free text. In this paper, we explain how this strategy is applied and partly evaluated for a specific use-case: supporting the attendees of a calendar event to prepare for the meeting.

¹ <http://ok-spotlight.iais.fraunhofer.de:5555/>

- Ontology-Enabled Access Control and Privacy Recommendations. Marcel Heupel, Lars Fischer, Mohamed Bourimi, Simon Scerri. Chapter 3 in Mining Big Data in Social Media and the Web, Volume 8940 in Lecture Notes in Computer Science series (LNAI 8940) 12/2014; Springer-Verlag.

ABSTRACT: Recent trends in ubiquitous computing target to provide user-controlled servers, providing a single point of access for managing different personal data in different Online Social Networks (OSNs), i.e. profile data and resources from various social interaction services (e.g., LinkedIn, Facebook, etc.). Ideally, personal data should remain independent of the environment, e.g., in order to support flexible migration to new landscapes. Such information interoperability can be achieved by ontology-based information representation and management.

- LOD-enhanced Social Media Analysis for Stock Market Prediction. Priyanka Dank, Simon Scerri, Stefan Nann, Sören Auer. Submission in progress for ISWC 2015 (pending).

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

- LREC Conference, May 2014, Reykjavik, Iceland (full paper presentation).
- EUROMED Conference, October 2014, Limassol, Cyprus (invited Keynote speaker for a Big Data session).
- In addition during the course of the fellowship the fellow was involved in as many as 7 project proposals targeting BMBF (German) and H2020 (EU) calls, one of which was accepted (2 are still pending). As part of the experience in the EIS group at Fraunhofer IAIS, the fellow managed teaching for the Lab (practical group work) and Seminar (presentations prepared by students) modules offered by the group at the affiliated Computer Science department of the University of Bonn.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

During the course of the fellowship, the fellow visited the Wimmics² team, headed by Fabien Gandon, which is a joint initiative between Inria Sophia Antipolis - Méditerranée and I3S (CNRS and Université Nice Sophia Antipolis), in Nice, France.

Over the course of two weeks, the fellow exchanged research ideas and objectives with a large number of researchers in this team, learning about many diverse topics and new methods that were previously not considered for research contributions. The very fruitful experience was made even more positive by the discovery of a secondary team (EURECOM³), based in the same campus, with ambitions in common research areas.

² <http://wimmics.inria.fr/>

³ <http://www.eurecom.fr/en/directory/service/mm>