Mission Report

Name

GlobeSec/EMM

Object of the mission

The objective of the mission was to deliver an invited talk at the Deggendorf Forum for Digital Information Analysis (DFDDA; http://www.dfdda.de) and to attend the yearly conference.

Place, date, organizations visited and contacted persons

Garching, Germany, 9-10 June 2016
Deggendorf Forum zur Digitalen Datenanalyse
Institute for Advanced Study, Technical University Munich.

Other Participants

Of the establishment

Distribution List

Obligatory(*) +

(*), JRC
globesec-secretariat@jrc.ec.europa.eu
(*), JRC
globesec-ipsc@jrc.it

Summary

The objective of the mission was to deliver an invited talk at the Deggendorf Forum for Digital Information Analysis (DFDDA; http://www.dfdda.de) and to attend the yearly conference, which took place in Garching, near Munich, Germany. DFDDA mostly targets auditors and it includes new technologies that may become important for their work. For the program of the event, see http://www.dfdda.de/dfdda2016/dfdda20160.html.

The one-hour JRC presentation was one of seven presentations at the forum. It had the title 'Supporting the analyst by turning multilingual text into structured data’. In the presentation, I gave an overview of EMM and OSINT Tools, highlighting the benefits of large-scale data analysis, showing various ways to visualise data, and making users aware of the dangers of fully-automatic analysis. I aimed to address the specific needs of auditors and I tried to show what an EMM-like analysis can do for this type of users and what it cannot do. Finally, I presented the multilingual linguistic resources created as by-products of EMM and downloadable from the JRC’s Science Hub pages, as well as from the EU Open Data Portal, with a special focus on JRC-Names and the JRC Eurovoc Indexer JEX.

The talk was received extremely well. There were many questions regarding functionality, availability of the software and the possible usage of the linguistic resources.

The DFDDA event was useful to get a good insight into the needs of auditors. The presentation I found most useful for our own purposes was the talk by [Name of Person] from Passau University (Germany), who talked about combining formal and distributional semantics to represent...
words and texts, showing the benefit for information retrieval and question answering. While in formal
semantics, meaning aspects of concepts are intellectually identified and described (e.g. a ‘dog’ has a
name, a colour, a race, a size, etc.), the individual values for each of these meaning aspects can be
found in large amounts of texts, using distributional methods. The commonalities and the differences
between the concepts ‘apple’ and ‘pear’ can thus be identified, for instance, by searching which
adjectives and which verbs (subject of, object of) are found with the equivalent words. Also
presented word clouds around the country name ‘Iraq’ (associated words) and how they changed over
time. This latter kind of observation could also be useful as an analysis aspect to be displayed in EMM.