



LIRMM

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier

Konstantin Todorov
LIRMM / University of Montpellier

September 11, 2017

TO:

IC1302 Action Chair: Prof. Francesco Guerra

STSM Coordinator: Dr Hussain Mahdi

CC: Host Dr. Stefan Dietze

Short Term Scientific Mission (STSM) Scientific Report

STSM title: Detecting fake news: web data analysis for fact verification.

Reference : ECOST-STSM-IC1302-250617-088575

STSM dates: from 04-09-2017 to 08-09-2017

My visit to L3S in Hanover in the framework of the COST action IC1302 has been completed a few days ago. I am writing this report as an account of that visit. My special thanks goes to my host Stefan Dietze and to Pavlos Fafalios for their availability and time.

Purpose of the Visit

This STSM request follows the line of a long term collaboration between my working group in LIRMM, Montpellier (recently renamed FADO) and the L3S group lead by Stefan Dietze that has resulted in a series of common publications in top ranked Semantic Web conferences and journal.

The visit had as a goal to provide the possibility to prepare a proposal for a joint call of the French National Research Agency (ANR) and the German Research Foundation (DFG). The proposal focuses on the development and application of semantic web technologies and machine learning techniques to fact verification and fake news checking on web documents, based in part on the ongoing research on website classification and controversiality detection, lead by the L3S and the work on data linking carried out by the FADO team at LIRMM.

Work Carried Out

The visit was organised in the following manner.

A first meeting between the project consortium members (Stefan Dietze and Pavlos Fafalias from L3S and Dino Ienco (via Skype) and myself from LIRMM) took place in the beginning of the stay. This meeting aimed, in the first place, at coming up with an action plan for the rest of the week, defining the preliminary goals/objectives of the project proposal and brainstorming on the different issues regarding the methods and technologies to be applied in order to fulfil these objectives. The meeting also aimed at clarifying the administrative constraints of the ANR/DFG call. It has been decided to dedicate the forthcoming days on research on the state-of-the-art on automatic fact checking with an emphasis on two main points: 1) definitions of main the concepts (bias, controversiality and fact-checking) in existing works, and 2) identification and categorization of the features that have been used in existing approaches for the automatic description and classification of facts.

In the following days I have been working principally on the collection of state of the art papers regarding these two points. Several discussions with Pavlos Fafalias and Stefan Dietze throughout the week helped converge ideas.

A second meeting of the four members of the consortium took place on Thursday afternoon in order to discuss the advances made. We came up with narrowed down objectives list for the proposal, focusing on the assessment of the trustworthiness of emerging facts exploiting structured and unstructured web data. This includes research on the features that characterise these facts, the different types of facts that we will need to look into and the sources of these facts (e.g., unstructured documents combined with markup data). Several applications of the proposal have been discussed, such as knowledge base augmentation, entity linking, or improving web search.

Results

Among the tangible results of the visit is a clear plan for action for the preparation and submission of a joint ANR/DFG proposal. After the end of the stay, we have a better mutual understanding of the main notions that we will be dealing with, the objectives that the project will pursue, as well of the means to achieves them.

Future Actions

The submission deadline for the proposal is December 2017. A next confcall meeting of the consortium is scheduled for September 26. In the meanwhile, the Montpellier team will be working on the related work of the proposal as well as on breaking the project down to work packages, while the Hanover team will be developing further the objectives and scope of the proposal. A first draft of the proposal is due by the end of September.



LIRMM

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier