

**STSM Topic:** Keyword search over People with Expertise

**STSM Applicant:** Prof. Yannis Velegrakis, University of Trento, Italy (velgias@unitn.it)

**STSM Host:** Prof. Bogdan Cautis, University of Paris-Sud, France

**STSM Reference:**

**STSM Dates:** 2017-03-27 to 2017-04-28

### **Purpose of the STSM**

The main objective in this STSM was to promote the research on efficient and user-friendly techniques (keyword queries) for answering queries in selecting the best set of people that can form an answer based on their profile. The idea is that users simply specify a set of keywords that describe the desired topic and the system selects the set of people that approximate that topic. The motivation behind this is that often we need to select data from a dataset but think the answer collectively as a whole and not as a set of individual independent items.

Apart from the research objective, there was also the administrative objective that was to strengthen the collaboration between the two institutions and pave the way for future articles and proposals.

### **Description of the work carried out during the STSM**

During the period of the STSM we have performed an extensive literature review that has allowed us to place the problem nicely and clearly within the whole ecosystem of people profiling. We have strengthened the novelty of our work and have formally defined the problem. We have also investigated a number of possible solutions that may be used and have also identified the technical challenges that each of these solutions require.

In parallel to the solutions that provide qualitative results we have also considered efficiency. We have studied how these solutions can be implemented into a spark framework in order to exploit parallelism and have considered the characteristics that we are expecting to have in order to implement such a solution.

### **Description of the main results obtained**

- We have created a first version of the formal problem definition that illustrates the novelty of the approach.
- We have sketched a first modeling of the data in order to have the users represented as points in an N-dimensional space.
- We identified ways through which the N-dimension vectors are derived from things that people say in social media and in particular on twitter.
- We have created a first list of possible algorithmic solutions to the problem of identifying the best set of items (persons in our practical scenario) that collectively constitute the best solution to the problem according to an objective function.

### **Future collaboration with the host institution**

We are planning to pursue the research work developed during this STSM and finalize the paper that we have started and sending it in a related upcoming conference by Aug (in particular we are considering WSDM). Furthermore, the idea wont stop to a single paper but we are currently identifying future extensions that may allow future collaborations on the issue, through the line of research we have initiated.

### **Foreseen publications/articles resulting from the STSM**

As previously mentioned we intend to finalize the draft and submit it to the WSDM this summer.