

# Experimenting ML techniques in Keyword search for structured databases

---

*STSM Scientific Report - COST-STSM-IC1302-24963*

## ***STSM Applicant***

Dr Francesco Guerra, University of Modena and Reggio Emilia, Modena(IT) ,  
francesco.guerra@unimore.it

## ***Host***

Julian Szymanski, Gdansk University of Technology, Gdansk (PL), julian.szymanski@eti.pg.gda.pl

## ***Period***

31/06/2015 – 11/06/2015

## **1. Purposes of the STSM**

The main aim of the STSM is to develop research projects involving two researchers with complementary expertise: dr. Guerra's background is mainly oriented toward the development of techniques for database management, semantic integration and search; dr. Szymanski's research mainly includes the development and application of machine learning techniques. The applicative scenario where we started to join our work is the one provided by the topic of keyword search for structured database.

The STSM can be divided in two phases: in the first part of the STSM we started to share our background. We had several meetings where the current research activities of both the teams have been described, discussed and analyzed. The goal was not only to find a defined research task for achieving research results in a "short time" but also to find a research theme around which to build a more complex project.

The possible research themes identified in the first phase of the STSM were:

1. Development of techniques for the integration of heterogeneous services and retrieval;
2. Study of techniques for the recommendation of web pages based on URL analysis;
3. Automatic shopping list creation, based on analysis of users' behaviors (adopting techniques for local and global/user profiling) and external knowledge;
4. Development of NN-based techniques for indexing databases
5. Study of techniques for "database documentalization", i.e. for transforming a database into a set of documents.

In the second part of the STSM, we started to work in the development of some ideas to achieve short term results (i.e., a paper to be submitted in a conference) and long term results (i.e., an idea for a research project to be developed in the future). In particular, we decided to extend and improve a previous research activity performed by the University of Gdansk, where they have implemented a threshold-based k-NN classifier. During the STSM we identified some applicative scenarios for the technique that was also improved by introducing a mechanism for learning. The resulting paper has been submitted to the International KEYSTONE Conference. The project around which to build some durable joint research

activities concerns the development of a querying-answering system for querying DBpedia. The idea is to create a middle-ware able to interpret the users' request and to transform them into SPARQL queries.

## 2. Main results

The main results of the STSM are

1. The definition of a proposal for a joint research project on keyword interpretation / disambiguation, aiming at creating a querying-answering system for DBpedia.
2. A paper submitted to the KEYSTONE Conference 2015.

During the stay, Dr. Guerra also delivered some lectures about techniques for keyword search in structured databases to the students of the Artificial Intelligence course and to some research members of the Computer Science Department of the Gdańsk University of Technology.

## 3. Future collaborations

The opportunities provided by European and National frameworks for supporting joint research activities were analyzed. Unfortunately no close opportunity had been found.

- Italy and Poland signed an agreement supporting visiting activities. The deadline of the call issued in 2015 expired during the stay and we were not able to submit any idea. The next call will be on 2016.
- H2020 includes some "Call for Twinning" to address networking gaps and deficiencies between the research institutions of the low performing Member States and regions and internationally-leading counterparts at EU level. There is no open call now, but new calls are foreseen in the future.
- There are not suitable calls in the remaining part of the H2020 work plan 2014-2015.

For this reason we planned to analyze the next calls, once the H2020 work plan 2016 -2017 will be issued, to find some opportunities. For the moment, we planned to work with already achieved research funds.

dott. ing. Francesco Guerra



Modena, 17/06/2015