

QyX: QuerY-driven eXample-based discovery of dataspace mappings

STSM SCIENTIFIC REPORT

APPLICANT: Federica Mandreoli

INSTITUTION: University of Modena and Reggio Emilia – Department of Physics, Computer Science and Mathematics

REFERENCE CODE: COST-STSM-ECOST-STSM-IC1302-040715-058928

PERIOD: 05/07/2015 – 21/07/2015

HOST: George H. L. Fletcher

HOST INSTITUTION:

Eindhoven University of Technology - Department of Mathematics and Computer Science

Purpose of the STSM

The purpose of the mission was to accelerate the progress and outcomes of our collaboration to the ongoing project “QyX: QuerY-driven eXample-based discovery of dataspace mappings”. The main aim of the project is to study the problem of mapping discovery in graph dataspaces and to introduce innovative solutions in this field driven by input queries.

Description of the work carried out during the STSM

During the visit, we mainly focused on the following activities:

- Sharing of the activities carried out in the past months on query answering semantics and structural characterization and study of the relationship between the corresponding outcomes;
- Study of the impact of structural characterization of conjunctive dataspace queries on mapping discovery and of an algorithm for deriving mappings from input queries;
- Investigation on funding possibilities and publication plan.

A detailed description of the above activities is provided in the following

Introduction of the relationship between query answering semantics and structural characterization

In the past months, a formal framework for dataspaces has been defined including dataspace definitions and the query answering semantics. To this end, we followed the well-known data

exchange framework. From the other hand, we worked on the structural characterization of object indistinguishability in this setting, where queries and mappings interact.

This activity aimed at a better understanding of the relationship between these two query answering semantics.

Study of the impact of structural characterization of conjunctive dataspaces queries on mapping discovery of algorithms for deriving mappings from input queries

In this activity, we studied how the results obtained in the structural characterization field can be exploited to extract mapping examples from the dataspaces. Examples are a key element of the project to help derive mappings. Then we addressed the mapping discovery problem in the defined framework from an algorithmic point of view. To this end, we devised an algorithm that takes in input a query and exhaustively scans the dataspaces triples consistent with the structural properties of the computed partial answers to build the missing mapping for query answering

Investigation on funding possibilities and publication plan

During the last days of the mission, we investigated funding possibilities including EU calls and Google grants. Moreover, we planned future research activities by considering both theoretical and engineering aspects. As to the latter, we also discussed about the possibilities of involving other researchers in the development phase.

Finally, we defined a publication plan.

Description of the main results obtained

We obtained interesting results for all the activities carried out during the visit.

From the research point of view, the main result we obtained is the algorithm for mapping discovery that relates the required mapping with the need of extending partial answers to total ones. The algorithm is described in a technical report for internal use that will be used as a basis for a first research paper we planned to submit in the next months. The report also outlines the open challenges related to the algorithm and describe a pathway towards the solution of some of them. The next step will be to introduce a notion of candidate mapping and to study the properties of the devised algorithm in terms of completeness and correctness with respect to the set of candidate mappings that can be derived from the input query.

Another concrete result is the draft version of a vision paper we will submit in September. Such a vision paper will describe the QyX vision, outline the process for mapping discovery and briefly describe the key ideas for each step.

Finally, we planned a future visit for preparing a proposal to submit to the next call for Google Faculty research awards.

Future collaboration with the host institution

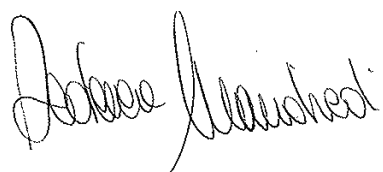
In the future, we foresee to continue to work together to face the theoretical and engineering challenges of the QyX project. Our final aim is to build a prototype that demonstrates the aided-

value of the mapping discovery solutions we are currently devising and we will devise in the future.

Foreseen publications/articles resulting from the STSM

As already mentioned before, we foresee to submit a vision paper in September and to submit further papers at conferences within the next year, reporting on the mapping discovery algorithm and other outcomes of the QyX project. Potential target venues are conferences such as EDBT, ICDE, SIGMOD, and VLDB.

In faith,

A handwritten signature in black ink, appearing to read "Federica Mandreoli". The signature is written in a cursive, flowing style.

Federica Mandreoli