# Report on completion of the short-stay study abroad program

| | |
|---|---|
| Name: | Nguyen Hoang Khai |
| Receiving univ./institute: | University of Leipzig |
| Period of study: | 2016/08/28 ~ 2016/09/22 |
| Date of report: | 2016/09/23 |

## 1. Research Theme

Time-efficient data access for link discovery

## 2. Summary of the activity

(Note) Describe the following points.
1) Summary of the research activities
2) Role of the program in your research and how to make use of the experience
3) Concrete outcomes including what you learned (discussed) through the program

I worked on LIMES, which is a famous and one of the state-of-the-art frameworks for link discovery on linked data. The main focus of LIMES is to detect the different instances of the same entity. Currently, LIMES has a difficulty dealing with large data and this issue need to be solved to create an efficient framework. My previous work is large scale link discovery, therefore, it is possible to combine the two studies to improve LIMES.

I first investigated the LIMES framework to get to know how LIMES is working. In parallel, we had to think of how to improve LIMES. We finally decided to improve firstly the geographic-data mapping module of LIMES with Spark. Spark is an advanced distributional big data processing framework. In order to implement Spark in LIMES, the existing algorithms in LIMES should be changed as well in order to leverage the advantages of Spark. Therefore, we re-designed some parts of LIMES and implement them in Spark.

The result is that we successfully installed Spark for the geographic-data mapping module. This result is preliminarily important for the further improvements of LIMES. I also learnt new techniques, algorithms, and new libraries while working on LIMES. More important, I had the chance to discuss and make the connection with many researchers in the same field. I also met other interesting persons working on other fields.