

## **Home Project**

### **“Linked Data Programming”**

This home project exercise can be completed at home, using the expertise and skill sets acquired in previous lectures. All students are invited to present solutions in the exercises for grade improvement. In order to participate, please proceed as follows:

- Prepare your solution (task description below) at home
- Provide the source code and generated output (e.g. HTML files) via email to [dietze@l3s.de](mailto:dietze@l3s.de) by 07.06.2016 (nearly 2 weeks time)
- Present your solution (approach, results) to the class during the KESW16 lecture on 09.06.2016 (max. 10 mins presentation, presentation details below)

Presenting and submitting a correct solution (on time) will result in a 0.3 grade improvement in the final exam.

### **Involved Endpoints/Datasets**

- LAK Dataset:
  - Endpoint: <http://lak.linkededucation.org/request/lak-conference/sparql>  
(Note: there is no HTML query conductor on this URL. You should use the HTTP GET request option as shown below).
  - Example request:  
[http://lak.linkededucation.org/request/lak-conference/sparql?query=select%20distinct%20?Concept%20where%20{\[\]%20a%20?Concept}%20LIMIT%2010](http://lak.linkededucation.org/request/lak-conference/sparql?query=select%20distinct%20?Concept%20where%20{[]%20a%20?Concept}%20LIMIT%2010)  
(further information at <http://lak.linkededucation.org>)
- DBLP:
  - Endpoint: <http://dblp.l3s.de/d2r/sparql>
  - Example request: `http://dblp.l3s.de/d2r/sparql?query=select%20distinct%20?Concept%20where%20{[]%20a%20?Concept}%20LIMIT%2010`
- DBpedia (optional):
  - Endpoint: <http://dbpedia.org/sparql>
  - Example request: `http://dbpedia.org/sparql?query=select%20distinct%20?Concept%20where%20{[]%20a%20?Concept}%20LIMIT%2010`

## Task Description

Build an application which:

1. Finds all authors which are represented in both the LAK Dataset and DBLP  
(note: the important problem to tackle is the one of coreference resolution, i.e. to find the correct representation of a particular LAK author in DBLP)
2. Retrieves for those authors their publications from both the LAK and DBLP datasets
3. Optional: retrieves – if available – the URL and key facts from DBpedia about that author (if available)
4. Orders authors ranked according to their LAK publication count
5. Presents all selected data in an HTML page (simple tables or more sophisticated visualization, feel free to be creative):
  - Author: name, organisation/affiliation
  - Their publications in LAK Dataset (and count)
  - Their publications in DBLP (and count)
  - Additional information (optional)

Note 1: for several of these steps, different approaches are feasible. It is your choice on how to design and implement the solution. While some are more efficient than others, in this case the results themselves matter (evaluation as below), rather than the approach.

Note 2: the LAK Dataset already contains links to author instances in DBLP. Please note that your approach should NOT make use of these references but use its own approach for identifying matching authors. This could in the simplest case use simple regex matching, while more sophisticated approaches are very welcome too.

## Presentation

For your project presentation, you have approximately 10 mins. A good presentation should cover the following aspects:

- overall approach
- used schema terms from both datasets
- used SPARQL queries
- coreference resolution approach
- results (descriptive statistics, eg total amount of matched authors, average/min/max publications from LAK respectively DBLP etc)
- evaluation: did you assess the quality of your results (precision and/or recall), if so, how? What are the results?

- demo (if available)

Note: this is not a table of contents but a list of items to cover.

## **Evaluation**

The presented solution should be able to retrieve at least 50 correct matches of LAK authors with DBLP authors with an accuracy of at least 70%. All presented solutions which meet these criteria will receive the grade improvement as described above.