



Describing the Cloud of Linked Datasets

While there is a growing number of published and offered as openly available Linked Open Data datasets¹; many challenges remain to be tackled on making such data usable to large extents for the different application purposes. One of the biggest challenges is knowing what the data is all about? with most datasets having poor and superficial metadata describing content or at best linked to reference datasets like DBpedia. A detailed and structured description describing the main concepts and topics covered by those datasets is highly valuable and necessary information to facilitate and expand usage of the available Linked Open Datasets.

M.Sc./Diploma thesis project

L3S Research Center, under supervision of Prof. Dr. Nejdil offers a M.Sc. or Diploma thesis in the scope of LinkedUp project². Taking into account the research context described above, the aim of the thesis is to provide methods that create useful and representative descriptions of Linked Data datasets, using a variety of methodologies and tools like: Named Entity Recognition and Disambiguation and as well extraction of related information about such concepts from knowledge graphs like DBpedia, and establishing of relationship of datasets under the extracted information.

**Are you interested in working with information extraction, semantic web and search technologies?
Want to be part of an international research team working with a new exciting research topic?**

The tasks:

- Research state of the art named entity extraction from reference datasets
- Measures on extracting related information from entities extracted from DBpedia such as categories describing the topics covered those entities cover
- Apply different scoring mechanisms based from Social Network Theory on extracted categories, conveying their usefulness and representativeness w.r.t to an entity and dataset
- Procedures that scale into the currently large amount of offered Linked Open Data

You should be:

- Familiar with data modelling, Semantic Web and Linked Data
- Knowledgeable about graph-based data representation
- Familiar with programming languages (ideally Java)
- An independent thinker and willing to learn

Are you interested or have questions? Contact us:

Besnik Fetahu, Email: fetahu@l3s.de, Forschungszentrum L3S, Appelstr. 9a, 30167 Hannover

¹ <http://datahub.io/lodcloud>

² <http://linkedup-project.eu>