Bachelor/Master Thesis

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Wikipedia List Extraction

Work with us on an innovative approach for extending the DBpedia Knowledge Graph.

Wikipedia is the most popular online encyclopaedia and has become an essential asset on the Web today. Besides its primary content, i.e. article pages on any kind of subject or entities, Wikipedia also contains aggregated "List Pages", i.e. pages that contain lists of entities that share something in common, as e.g. [1]. **DBpedia** is an online knowledge base that is extracted from Wikipedia content on a regular basis.[2] However, DBpedia information on List Pages is rather shallow and has to be further enriched.

Subject of this thesis is the analysis of Wikipedia List Pages and the enrichment of DBpedia with knowledge extracted from Wikipedia List pages.

The work comprises:

1. Extraction of Wikipedia List pages from Wikipedia Dump files
2. Connecting Wikipedia lists with underlying DBpedia entities
3. Extract (raw) information from single Wikipedia list pages
4. Encode Wikipedia lists via given ontology schema as RDF for use in DBpedia

For a Master thesis, information extraction from Wikipedia lists should apply and extend already existing approaches [3] with a special focus on completeness and data quality.

This thesis will be supervised by **Prof. Dr. Harald Sack, Information Service Engineering at Institute AIFB, KIT, in collaboration with FIZ Karlsruhe**.

[1] https://en.wikipedia.org/wiki/Lepsius_list_of_pyramids
[2] Lehmann, J. et al. "DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia." Semantic Web Journal 6 , no. 2 (2015): 167--195.
[3] https://github.com/dbpedia/list-extractor

Which prerequisites should you have?

- Good programming skills preferably in Python
- Interest in Knowledge Representation
- Interest in Data Mining
- Interest in Natural Language Processing

Contact person:
Prof. Dr. Harald Sack
harald.sack@kit.edu
harald.sack@fiz-karlsruhe.de

DBpedia

Wikipedia
The Free Encyclopedia

aifb
Institute of Applied Informatics and Formal Description Methods
http://www.aifb.kit.edu/

Karlsruhe Institute of Technology