Correlating disconnected Web content
Uncovering relationships between concepts in reference datasets

The most recent evolution of the Semantic Web is the Linked (Open) Data (LOD) approach which has led to an ever growing cloud of LOD datasets\(^1\) on the Web, covering all sorts of domains and areas, where prominent examples include the data.gov.uk initiative\(^2\), DBpedia\(^3\), or Geonames\(^4\). However, while interlinking of data is one of the fundamental principles of LOD, mainly explicit links are considered so far. For instance, this covers explicit references between a Web resource about “Berlin” and the corresponding DBpedia entry (http://dbpedia.org/resource/Berlin). However, while those direct relationships are good starting points to perform initial classification, for instance, to retrieve all resources which refer to http://dbpedia.org/resource/Berlin, less strict and more implicit links can be derived from those explicit relationships based on the inherent structure of the involved graphs (in this case DBpedia). For instance, one might cluster all resources which refer to German cities even though no particular reference to http://dbpedia.org/resource/Germany is given. Reasoning on the DBpedia graph to identify nodes which are classified as German cities could uncover those otherwise hidden relationships and allow a less strict clustering approach.

M.Sc /Diploma thesis project

L3S Research Center, under supervision of Prof. Dr. Nejdl offers a M.Sc. or Diploma thesis in the scope of the ARCOMEM\(^5\) and mEducator\(^6\) projects. Taking into account the research context described above, the aim of the thesis is to provide methods to uncover and rank implicit relationships between distinct concepts (nodes) in the same dataset (graph) which in turn will allow the clustering of arbitrary Web content. As initial dataset, DBpedia is recommended.

Are you interested in working with data mining and search technologies? Want to be part of an international research team working with a new exciting research topic?

The tasks:

I. Research state of the art in relationship discovery and path exploration

II. Develop a relationship discovery and evaluation method (e.g. based on DBpedia) to be integrated into clustering techniques

---

\(^1\) [http://lod-cloud.net/state](http://lod-cloud.net/state)
\(^2\) [http://data.gov.uk](http://data.gov.uk)
\(^3\) [http://dbpedia.org](http://dbpedia.org)
\(^4\) [http://www.geonames.org/](http://www.geonames.org/)
\(^5\) [http://www.arcomem.eu/](http://www.arcomem.eu/)
\(^6\) [http://www.meducator.net](http://www.meducator.net)
You should be:
- Familiar with data modeling, 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:
Stefan Dietze, Email: dietze@l3s.de, Forschungszentrum L3S, Appelstr. 9a, 30167 Hannover