

Title:

Uncovering the Hidden in Large Size Knowledge Graphs

Abstract (max. 200 words):

Discovering hidden facts in steadily growing and complex knowledge graphs is challenging. It's not that a proper SPARQL query can't retrieve valuable information. A mission critical task often is to get an idea of those queries and the graph. This becomes even more challenging when the underlying data or schema is not well known. For instance, in case of an empty SPARQL result, either the query is misspelt, whose detection needs full schema awareness or some schema research, or there is simply no such data.

We will introduce SemSpect, which enables even non-SPARQL users to build sophisticated queries by interacting with a visual representation of the data. Since querying and exploring with SemSpect is data and schema driven users are guided throughout their analysis not suffering foregoing problems. The talk will motivate the approach with the help of use cases and provides practical insight to the task of exploring and visually analyzing complex and large knowledge graphs such as the Panama Papers. This is supported by industry use cases and live demoing during the presentation.

Description (max. 400 words):

Large knowledge graphs are hard to comprehend but occur within almost all business relevant missions. A SPARQL endpoint or classical BI tools are valuable where the basic structure of the data and the query idea is known beforehand. But how to proceed when facing complex, cross-linked data pools or when the insightful queries are not transparent because of a lack of the big picture? For instance, how to identify the real beneficiaries of an offshore entity within the Panama Papers whose shares are intentionally blurred without knowing the obfuscation patterns?

The presentation starts with investigating this challenge with the help of examples from the Panama Papers and lessons from industry projects. It turns out, that for quite a number of application domains typical visualization approaches and query interfaces fall short to provide meaningful insight into large knowledge graphs.

The talk will then introduce SemSpect, a visual analytics solution for RDFS/OWL knowledge graphs. SemSpect allows to visually explore knowledge graphs interactively following the „overview first, details on demand“ strategy. It's a client-server solution with a Web UI that aggregates objects and relationships in contrast to traditional network layouts for reasons of clarity and comprehensibility. Instead of successively writing and evaluating SPARQL queries users are guided by the data itself and can on-demand:

- render or list the objects of a class
- show details of one or more objects
- visualize a relationship between single or groups of objects
- filter any group to restrict the current exploration
- define custom classes from any groups within the exploration graph
- save and re-use custom classes or explorations
- export exploration renderings and selected objects resp. groups

Because of its “context zoom” that groups objects and relationships graphically and detail information on user demand SemSpect is eminently suitable for large data sets containing several millions of objects and relationships.

The presentation reports on lessons learned and compares the commonalities as well as different application purpose of SPARQL and SemSpect. We will showcase demos from the LOD space such as SpringerNature’s SciGraph publication data at

<http://scigraph.semspect.de> and the Panama Papers at <http://panama.semspect.de>

The SemSpect back-end can load RDF or OWL data in any of its formats and will soon be able to directly import data from Neo4j databases. The system is already in use at various domains and a community version is planned to be released at SEMANTiCS 2017.

Speakers bio:

Thorsten Liebig is CEO of derivo GmbH, a company that faces the challenges of more and more complex, extensive and distributed data pools with a semantic management of knowledge. In his role he is consulting customers from the machine or publishing industry or from the intelligence domain with respect to sustainable semantic solutions. derivo is in charge also of award winning reasoning systems such as Konclude or GraphScale. Prior to derivo he has lead a working group at Ulm University. He has lead the OWLLink international working group to define a communication protocol for OWL, that became a W3C member submission in 2010.