

Guided Link Traversal Query Processing within Contextualized Subwebs

Younes Zeboudj¹ Bas Ketsman¹ Bart Bogaerts¹
Ruben Taelman² Ruben Verborgh²

¹Vrije Universiteit Brussel, Dept. of Computer Science, Brussel, Belgium
firstname.lastname@vub.be

²IDLab, Dept. of Electronics and Information Systems, UGent – imec, Belgium
firstname.lastname@ugent.be

Link-Traversal-based Query Processing (LTQP), in which a SPARQL query is evaluated over a web of interlinked documents rather than a database, is often seen as a theoretically interesting yet impractical technique. However, in a time where the hypercentralization of data has increasingly come under scrutiny, a decentralized Web of Data with a simple document-based interface is appealing because it enables data publishers to easily control their data and access rights. While LTQP allows evaluating complex queries over such webs, it suffers from performance issues (due to the high number of documents containing data) as well as information quality concerns (due to the many sources providing such documents). To mitigate these issues, we created a query semantics for contextualized subwebs, which allows pruning the number of consulted documents based on whether their contents would produce results that are desirable within a given context. This article introduces the theoretical framework and proves several of its properties. We show that a guided LTQP approach generalizes existing reachability semantics, and can deliver results that satisfy distributed content policies while eliminating network requests that would not contribute to the result set. Through this research, we open up opportunities for advanced query processing in decentralized networks, attaining improved performance levels without prior centralization or indexing.

Keywords Web of Data , link-traversal-based query processing , RDF

References

- [1] Tim Berners-Lee. Linked data. <https://www.w3.org/DesignIssues/LinkedData.html>.
- [2] Olaf Hartig and Jorge Pérez. Ldql: A query language for the web of linked data (extended version), 2015.