

Exploiting and Enriching Highly Dynamic Linked Data

Luggen Michael from Switzerland

The ever-faster process of data creation, integration and consumption prompts ever more frequent updates to data schemas. Traditional data models were not meant for such rapid changes and are impractical to continually amend or update. *Linked Data*, on the other hand, is often used as an umbrella term to refer to data models that are interlinked and support decentralized, agile data management. With *Linked Data*, both the data and schemas remain agile, and hence can adapt to new requirements on-the-fly. In this thesis, we investigate the challenges and opportunities arising in this context, when both the application development process and the data itself are dynamic and continuously evolving. First, we propose a new data flow process for end-user applications that takes advantage of highly dynamic data represented as *Linked Data*. By leveraging the original data structures up to the final view layer instead of projecting subsets of the data, user interfaces can automatically adapt to the various changes in the schema. We provide a prototype end-user application engine to demonstrate the viability of our approach. Furthermore, we introduce techniques to stabilize and enrich both the datasets and schemas automatically in the context of fast-changing application data. Collaborative encyclopedic *Knowledge Graphs*, such as *Wikidata*, represent a good example of such application data. By assuring a high coverage of potential instances per class recorded, it is possible to improve the stability of a Knowledge Graph. To that end, we introduce a completeness estimator based on the author's edits from *Wikidata*. Our estimator identifies areas in the *Knowledge Graph* that are complete and can point out areas that need further attention. *Entity Linking* is a class of methods that identify concepts in unstructured text and links them to a defined vocabulary, nowadays mostly in the form of a *Knowledge Graph*. To further enrich the data, *Entity Linking* can be used to identify new facts from unstructured text. We provide a new entity linking approach by analyzing the links and their respective labels in the target *Knowledge Graph*. We focus on high-precision linking with the goal to prepare the output for fully automated downstream tasks. Finally, we investigate how to stabilize the properties in the data schema. We introduce a novel method that predicts missing properties for concepts in a *Knowledge Graph*. Our method learns from a high-dimensional representation of unstructured text, namely embedded *Wikipedia* articles. Through this selection of new methods, we provide algorithms and tools to handle highly dynamic data and accelerate the adoption of new properties in the schema, with the ultimate goal of stabilizing highly dynamic data in the long run.

Jury:

Prof. Dr. Philippe Cudré-Mauroux, University of Fribourg (Switzerland), thesis supervisor.
Prof. Dr. Bernhard Anrig, Bern University of Applied Sciences (Switzerland), co-supervisor.
Prof. Dr. Ruben Veborgh, Ghent University (Belgium), external examiner.
Prof. Dr. Djellel Eddine Difallah, New York University (Abu Dhabi), external examiner.
Prof. Dr. Ulrich Ultes-Nietsche, University of Fribourg (Switzerland), president of the jury.