
Mining and Managing Large-Scale Linked Open Data

[Abstract]

Ansgar Scherp
University of Kiel
Christian-Albrechts-Platz 4
24118 Kiel
asc@informatik.uni-kiel.de

ABSTRACT

Linked Open Data (LOD) is about publishing and interlinking data of different origin and purpose on the web. The Resource Description Framework (RDF) is used to describe data on the LOD cloud. In contrast to relational databases, RDF does not provide a fixed, pre-defined schema. Rather, RDF allows for flexibly modeling the data schema by attaching RDF types and properties to the entities. Our schema-level index called SchemEX allows for searching in large-scale RDF graph data. The index can be efficiently computed with reasonable accuracy over large-scale data sets with billions of RDF triples, the smallest information unit on the LOD cloud. SchemEX is highly needed as the size of the LOD cloud quickly increases. Due to the evolution of the LOD cloud, one observes frequent changes of the data. We show that also the data schema changes in terms of combinations of RDF types and properties. As changes cannot capture the dynamics of the LOD cloud, current work includes temporal clustering and finding periodicities in entity dynamics over large-scale snapshots of the LOD cloud with about 100 million triples per week for more than three years.

About the Author

Ansgar Scherp is Professor for Knowledge Discovery with ZBW - Leibniz Information Centre for Economics and Kiel University since January 2014. He is expert on analyzing Linked Open Data, the approach to publish and interlink data on the web. Ansgar has been EU Marie Curie Fellow at UC Irvine, CA, USA from 2006 to 2007. Subsequently, he has led work packages in EU projects such as WeKnowIt and Social Sensor at the University of Koblenz-Landau. Currently, Ansgar is scientific coordinator of the EU H2020 project MOVING (<http://moving-project.eu/>) on training users from all societal sectors to improve their information literacy by training how to choose, use, and evaluate data mining methods in connection with their daily research tasks and to become data-savvy information professionals.