

Fusing the Web of Data

Chris Bizer

Freie Universität Berlin

Abstract. The public Web is increasingly understood as a global information space, consisting not just of linked documents, but also of linked data. In addition to the maturing of the Semantic Web technology stack, a major catalyst in this transition has been the application of the Linked Data principles, hand-in-hand with the publication and dense interlinking of large-scale data sets distributed across the Web. It is estimated that this Web of Data currently consists of over 10 billion RDF statements covering domains such as geographic information, people, companies, online communities, films, music, books, scientific publications, and an increasing amount of life science and health care related data. Applications that build on top of the Web of Data are confronted with the challenge of fusing information from an unbound set of autonomous data sources. In addition to classic data integration tasks such as identity resolution, inconsistency resolution, and schema mapping, the Web of Data raises new challenges around accountability, provenance tracking, information quality, and data licensing. Within the talk, Chris Bizer will first outline the development and the current topology of the Web of Data. Afterwards, he will discuss the data integration challenges that arise from the Web of Data and will give an overview about current research within the Semantic Web community that aims at handling these challenges.

The Speaker:

Prof. Dr. Chris Bizer explores technical and economic questions concerning the development of global, decentralized information environments. His current research focus lies on the publication and interlinking of structured data on the Web using Semantic Web technologies. The results of his work include the Named Graphs data model, which was adopted into the W3C SPARQL standard; the Fresnel display vocabulary implemented by several Semantic Web browsers, and the D2RQ mapping language which is widely used for mapping relational databases to the Semantic Web. He takes a leading role in the W3C Linking Open Data community effort and the DBpedia project, which both aim at interlinking large numbers of data sources on the Web.