Title:
Towards Trustworthy Linked Data Integration and Consumption

Author:
RNDr. Tomáš Knap

Department:
Department of Software Engineering

Supervisor:
RNDr. Irena Holubová, PhD., Department of Software Engineering

Abstract:
We are now finally at a point when datasets based upon open standards are being published on an increasing basis by a variety of Web communities, governmental initiatives, and various companies. Linked Data offers information consumers a level of information integration and aggregation agility that has up to now not been possible. Consumers can now "mashup" and readily integrate information for use in a myriad of alternative end uses. Indiscriminate addition of information can, however, come with inherent problems, such as the provision of poor quality, inaccurate, irrelevant or fraudulent information. All will come with associated costs of the consumed data which will negatively affect data consumer's benefit and Linked Data applications usage and uptake.

In this thesis, we address these issues by proposing ODCleanStore, a Linked Data management and querying tool able to provide data consumers with Linked Data, which is cleansed, properly linked, integrated, and trustworthy according to consumer's subjective requirements. Trustworthiness of data means that the data has associated data provenance, which satisfies the consumer's requirements, has certain data quality required by the data consumer, and is provided by trustworthy agents. We propose in the thesis a novel data fusion component for ODCleanStore which solves conflicts among the consumed heterogenous data and supplements the integrated data with justified quality scores and provenance metadata. Furthermore, to enable expressing and tracking of data provenance on the Web, we propose a novel provenance model for the Web – W3P. We also discuss trust models and their usability to compute trustworthiness of agents in social networks; a factor which contributes to the trustworthiness of consumed Linked Data. The ODCleanStore tool is available under an open license and is planned to be used by the Agile Knowledge Engineering and Semantic Web (AK-SW) research group at the University of Lepzig, by the Department of Computer Science, Systems and Communication at the University of Milan-Bicocca, by the Semantic Web Company, Austria, and at the `http://opendata.cz` portal.