

Improving Data Quality using Domain-Specific Data Types

Summary

The JValue open data service collects various open data from governmental data sources and imports them into one large database. The raw source data comes in various forms and qualities. This thesis designs and implements an approach for improving data quality by utilizing domain-specific data types. An example of a domain-specific data type is Coordinate, consisting at a minimum of two attributes, longitude and latitude, with associated domain semantics. The thesis implements various data types following the Value Object pattern and utilizes them to improve data quality in the JValue open data service.

Work Results

- Research question
 - How and why to improve data quality using domain-specific data types?
- Literature review
 - <http://dirkriehle.com/computer-science/research/2006/plop-2006-value-object.html>
 - <http://www.riehle.org/computer-science/research/1998/ubilab-tr-1998-10-1.html>
 - Relevant literature on improving data quality in database systems
- Research approach
 - Design and implementation of data types within JValue open data service
 - Evaluation of design by comparison with situation not using data types
- Expected results
 - Design and implementation of data types with JValue ODS using JValue data types
 - Conclusion on suitability of approach to improving data quality within JValue ODS

Supervisor

Prof. Dr. Dirk Riehle, dirk.riehle@fau.de

Open Source Research Group
Computer Science Department
Friedrich-Alexander University

More information: <http://osr.cs.fau.de/theses/resources/>