

Jiří Helmich

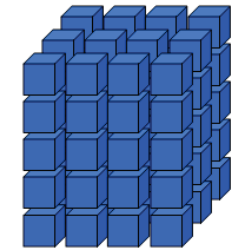
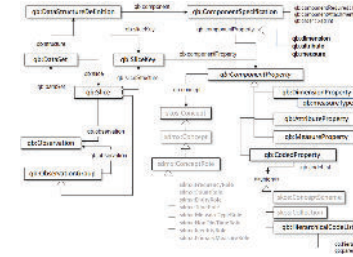
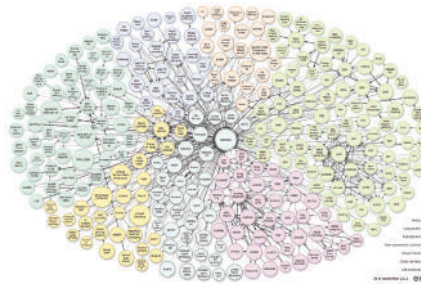
supervisor: Martin Nečaský, PhD.

Analysing and Visualizing Statistical Linked Data

<http://datacube.payola.cz>

<http://helmich.cz/master-thesis/>

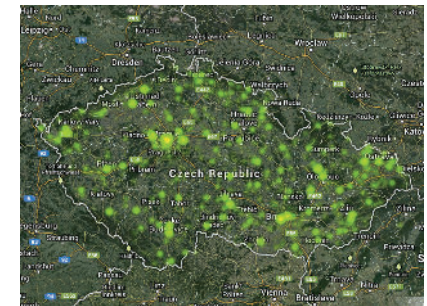
The thesis describes several means of processing statistical data in the ambience of Linked Data and is in particular focused on the utilization of Data Cube Vocabulary metaformat. Its content offers a description of tools related to analysis and visualization of RDF data not only from the statistical view. An indivisible part of this work is the depiction of the Payola tool on whose development is the author still working on. The outcome of this thesis is mainly proposal and consequential implementation of the system that enables a conversion of RDF data in compliance with the DCV vocabularies. The designed system was implemented and integrated to the Payola application. Several other extensions of the system were also implemented by the author.



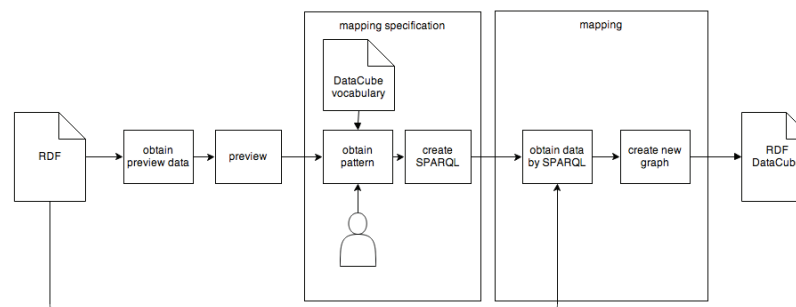
While utilizing the RDF Data Cube Vocabulary W3C recommendation, the system enables the user to map arbitrary RDF data into a form of a data cube.

To make it so, the system implements the well-known query-by-example principle. Therefore, it enables the user to provide an exemplary mapping graph pattern, which is then applied to the whole input dataset.

[red nodes correspond to a data cube dimension, green ones make the pattern more accurate, blue ones are not considered]



While taking advantage of the analytical pipeline Payola feature, it is possible to create a visualization based on a more sophisticated statistical dataset. E.g. a visualization of where the money is spent on Czech public contracts.



A schema of the system proposed in the thesis. The user provides an input RDF graph and an arbitrary Data Cube vocabulary. The system makes a preview of the input data to allow the user to select an exemplary pattern.