

Using Semantics for Interactive Visual Analysis of Linked Open Data

Gerwald Tschinkel¹, Eduardo Veas¹, Belgin Mutlu¹ and Vedran Sabol^{1,2}

¹ Know-Center gtschinkel@know-center.at | veas@know-center.at | bmutlu@know-center.at | vsabol@know-center.at

² Graz University of Technology

Abstract. Providing easy to use methods for visual analysis of Linked Data is often hindered by the complexity of semantic technologies. On the other hand, semantic information inherent to Linked Data provides opportunities to support the user in interactively analysing the data. This paper provides a demonstration of an interactive, Web-based visualisation tool, the “Vis Wizard”, which makes use of semantics to simplify the process of setting up visualisations, transforming the data and, most importantly, interactively analysing multiple datasets using brushing and linking methods.

1 Introduction

An objective of the CODE³ project is to make Linked Data accessible to novice users by providing easy to use methods for visual data analysis. This is hard to achieve with current Linked Data tools, which require user’s knowledge of semantic technologies (such as SPARQL). This paper demonstrates how semantic information can be used to support the interactive analytical process, without the need for users to understand the complexities of the underlying technology.

Within CODE we use the RDF Data Cube Vocabulary⁴ for describing statistical datasets. Our “Vis Wizard”⁵ tool provides an intuitive, easy to use interface supporting visualisation and interactive analysis of RDF Cubes. In the Vis Wizard we utilise semantic information from Linked Data to support the user in:

1. Selecting and configuring the visualisations
2. Aggregating datasets
3. Brushing and linking over multiple datasets

This paper illustrates the use of semantic technologies in a visual analytics tool that enables novice users to perform complex operations and analyses on Linked Data. The demonstration focuses mainly on step 3, with a screencast of the demonstration also being available⁶.

³ <http://code-research.eu>

⁴ <http://www.w3.org/TR/vocab-data-cube>

⁵ <http://code.know-center.tugraz.at/vis>

⁶ <http://youtu.be/aBfuGhgVaxA>

Related work: A wide range of tools offers functionalities for visualising and interacting with data, but only a few rely on semantic information to support the analytical process. Tableau [6] provides a mighty visualisation toolset, however it does not make use of semantic information for assisting the user. The CubeViz Framework [5] facilitates visual analytics on RDF Data Cubes, but does not use semantics for the user interface. CubeViz supports no brushing, no possibility to compare datasets directly and no automatic selection of visualisations. Cammarano et al. [1] introduces a method to automatically analyse data attributes and map them to visual properties of the visualisation. Even so, this does not include an automatic selection of visualisation types.

2 The Linked Data Vis Wizard

The underlying thought is to make the user capable of visually analysing data without knowing about the concept of Linked Data or RDF Data Cubes. However, the Vis Wizard utilises the available semantic information to support users in interacting with the data and performing analytical tasks.

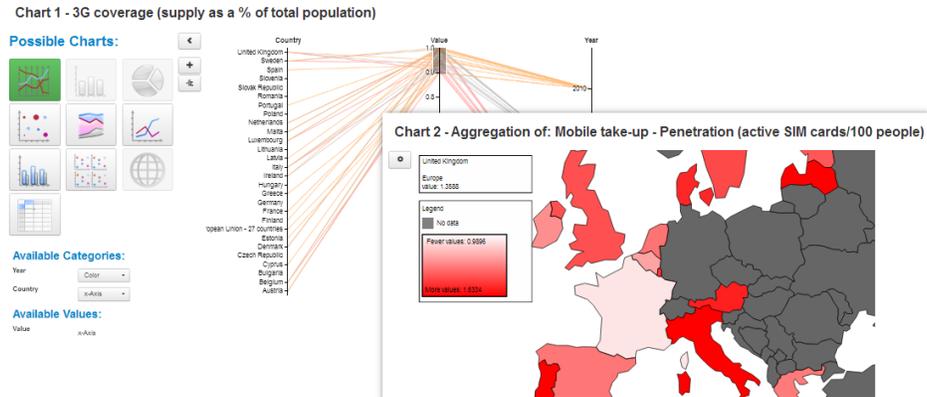


Fig. 1. Two RDF Data Cubes are shown in the Vis Wizard. Brushing the 3G coverage value in the parallel coordinates highlights corresponding countries in the geo-chart.

Scenario: Figure 1 compares two datasets taken from the EU Open Data Endpoint⁷ in the Vis Wizard. The first one, shown in parallel coordinates, represents the 3G coverage in Europe, as percentage value, per country for each year. The second dataset, shown in the geo-chart, contains active SIM cards per 100 people (encoded by colour-grading) for countries in Europe. In the following we use the Vis Wizard to gain insights into the data and ascertain the datasets correlate.

2.1 Interactive Visual Analysis

Selecting and configuring the visualisation: The first step is to find an appropriate visual representation for the given dataset. Within the 10 supported

⁷ <http://open-data.europa.eu>

charts only those are made available which can actively and meaningfully be used with the provided data. For example, the geo-chart is only available if the data contains a geographic dimension. After the chart was selected, the user can adjust the mapping of data onto the visual properties (e.g. axes, colours, item sizes etc.) of the chart, whereby only suitable mappings are offered. Chart selection and the data mapping is computed by an algorithm [3] comparing the semantic information in the RDF Data Cube with the visual capabilities of the chart, which are described using the Visual Analytics Vocabulary⁸.

Aggregation: We provide a dialogue for aggregating the data and creating a new Data Cube. In the scenario shown in Fig 1 the second dataset was averaged over the years and visualised over the countries. Using semantics we differentiate between dimensions and measures and enable validation of the user choices.

For suggesting charts and supporting aggregation we are utilising RDF datatype, occurrence and persistence.

Brushing and Linking: The idea behind brushing and linking is to combine different visualisations to overcome the shortcomings of single techniques [2]. Interactive changes made in one visualisation are automatically reflected in the other ones. Our scenario contains two separate datasets: the first dataset has the dimensions “country” and “year”, the second dataset has only “country”. For conventional tools it is hard to provide interaction over different datasets, because relationships between them are usually not explicitly available. In cases when columns are labelled using equal strings guessing the relationships may be possible, but when labels differ, e.g. a dimension in dataset A is called “Country” while in the dataset B it is called “State”, the relation cannot be established. In such cases the burden of understanding the structure of the datasets and linking them together falls on the user. Within RDF Data Cubes, each dimension has an URI which is (by definition) unique and can be used to establish the connection between datasets, making linking and brushing over different datasets possible.

Applied to our scenario the following interactive analysis is performed (see Fig. 1): The user applies a brush on the first dataset by selecting a specific value range in the “3G coverage” dimension using the parallel coordinates chart. Countries outside of the selected range are greyed out in the geo-chart, which shows the second dataset (SIM card penetration). Obviously, a high 3G coverage correlates with high SIM card penetration (red), with one exception - France.

It should be noted that the functionality of linking data over different datasets, or even different endpoints, depends on the quality of the semantic information: the URIs of the cube-dimensions in different datasets need to be consistent. If datasets use different, domain-specific URI namespaces, linking the data will not be possible.

3 Evaluation

We conducted a formative evaluation to explore if our goals regarding the usability of the Vis Wizard could be achieved and to ascertain that users were able

⁸ <http://code.know-center.tugraz.at/static/ontology/visual-analytics.owl>

to analyse complex datasets. Eight test users participated which executed six tasks, where one task was exclusively about linking and brushing. Test users had a good knowledge of computers, but were not familiar with semantic data. We conducted a quantitative subjective workload test, using the simplified NASA R-TLX, and a qualitative thinking aloud test. More details on the evaluation, including methodology, test users and results are available under [4].

The functionality supporting the choice and configuration of the visualisation was much appreciated, but users pointed out that immediately suggesting the most suitable visualisation would have been even more helpful. The task regarding brushing in the scatterplot had a very high subjective performance of accomplishing (the median was 91.25 on a scale from 0 to 100, 100 being the highest value achievable). The conclusion of our evaluation is that, while several usability issues still need to be fixed, the overall advantage is clearly observable.

4 Conclusion and Future Work

Within this research we have observed a high potential in using semantic information for improving interaction in visual analytics. It has been shown that the user supporting techniques were helpful in gaining insights from the data, without spending much time in selecting and configuring visualisations or analysing how to link the datasets manually.

As for our purpose the correctness of the semantic annotations of the data is essential, the stability of our approach could be improved by implementing the use of URI aliases. We will also explore the possibilities to rank the visualisations in order to, given a particular dataset, automatically show the most suitable one.

Acknowledgements. This work is funded by the EC FP7 projects CODE (grant 296150) and EEX-CESS (grant 600601). The Know-Center GmbH is funded by Austrian Federal Government within the Austrian COMET Program, managed by the Austrian Research Promotion Agency (FFG).

References

1. Cammarano, M., Dong, X.L., Chan, B., Klingner, J., Talbot, J., Halevey, A., Hanrahan, P.: Visualization of heterogeneous data. In: IEEE Information Visualization
2. Keim, D.A.: Information visualization and visual data mining. In: IEEE Transactions on Visualization and computer graphics (2002)
3. Mutlu, B., Höfler, P., Tschinkel, G., Veas, E.E., Sabol, V., Stegmaier, F., Granitzer, M.: Suggesting visualisations for published data. In: Proceedings of IVAPP 2014. pp. 267–275 (2014)
4. Sabol, V., Tschinkel, G., Veas, E., Hoefler, P., Mutlu, B., Granitzer, M.: Discovery and visual analysis of linked data for humans. In: Accepted for publication at the 13th International Semantic Web Conference (2014)
5. Salas, P.E., Martin, M., Mota, F.M.D., Breitman, K., Auer, S., Casanova, M.A.: Publishing statistical data on the web. In: Proceedings of 6th International IEEE Conference on Semantic Computing. IEEE 2012, IEEE (2012)
6. Stolte, C., Hanrahan, P.: Polaris: A system for query, analysis and visualization of multi-dimensional relational databases. IEEE Transactions on Visualization and Computer Graphics 8, 52–65 (2002)