Discovery and Visual Analysis of Linked Data for Humans

ISWC 2014
22.10.2014, Riva del Garda, Italy

Vedran Sabol
Know-Center GmbH
Austrian Research Center for Data-driven Business and Big Data Analytics
Overview

• Motivation
  – CODE Project
• Two scenarios
  – Requirements
• Tool design
• CODE tools in use
  – Two use cases
  – Integration in other platforms
• Usability evaluation
Motivation

- Huge amount of information in the LOD cloud
- Difficult to access and analyse
  - Knowledge on semantic technologies is required
  - Graph structure: complex to evaluate and interpret
- Goals
  - Empower IT-laymen to search, explore, publish and analyse LD
  - Make LD accessible to the general public
- Design tools to be easy to use
  - Hide the complexities of semantic technologies
  - Exploit the advantages of semantically rich data
• Vision: establish foundations for web-based, commercially oriented ecosystem for LOD
• Focus: research data and publications, open data
Scenarios

1. Discovery and analysis in LOD
   - Searching, transforming/refining data, interactive analysis
   - Target audience: general public, professionals

2. Analysing research data in scientific publications (PDFs)
   - Extracting data, correction/annotating data, Interactive analysis
   - Target audience: researchers

Issues with current methods

- Tedious manual operations are necessary
  - Extracting/correcting data sets
  - Selection/configuration of visualisations

- Searching using SPARQL is too complex
High-Level Requirements

- **NFR1 - ease of use**: employ UI concepts users are acquainted with
- **NFR2 - automation**: eliminate manual steps
- **NFR3 - exploit semantics** to the advantage of the user
- **FR1 - LOD search** in SPARQL endpoints
- **FR2 - Extraction** from publications in PDF format
- **FR3 - Data transformation/refinement** to set up the data set
- **FR4 – Triplification and export** of data sets as RDF (Data Cubes)
- **FR5 – Interactive visualisation** for data analysis
Tool Design

Query Wizard

- Keyword search in LOD endpoints
- Tabular representation of results
  - Rows – subjects; columns – predicates
  - Cell: 0..n objects (literals/entities) depending on row, column

- Data operations
  - Add/remove columns
  - Load more results
  - Filtering
  - Aggregation
  - Open in browser

Viewing and manipulating RDF Data Cubes
Tool Design
PDF Extractor and Data Extractor

- **PDF Extractor**
  - Information extraction from publications (PDFs)
  - Title, table of contents, abstract, authors, named entities, references, figures, **tables**

- **Data Extractor**
  - UI for correcting and annotating tables
  - Data set export as RDF Data Cube

---

Link: [http://kd.know-center.tugraz.at/code-demo/](http://kd.know-center.tugraz.at/code-demo/)
Tool Design

Visualisation Wizard

- Automated Visualisation: selection and configuration of charts
  - Visual Analytics Ontology: describes chart visual properties
    - [http://code.know-center.tugraz.at/static/ontology/visual-analytics.owl](http://code.know-center.tugraz.at/static/ontology/visual-analytics.owl)
  - Map RDF Cubes onto semantically described charts

- Multiple charts
  - Provide different views on data
  - Interactive analysis: coordinated linking & brushing

- Aggregation (grouping)
  - Produces new Data Cube(s)
• Workflow 1 implements discovery and analysis in LOD
• Workflow 2 implements Analysis of Scientific Publication Data
• All tools were public in late 2012
  – Focus on Workflow 1 (currently active)
Use Case 1
Discovery and analysis in LOD

Query Wizard:
1. Searching in LOD
2. Setting up the data set
   - Filtering
   - Choosing columns

Vis Wizard:
3. Selecting a visualisation
4. Aggregating data
5. Interactive analysis

Insight: low funding in Eastern Europe

Link: [http://code.know-center.tugraz.at/search](http://code.know-center.tugraz.at/search)
Video: [http://www.youtube.com/watch?v=mA_v1F7TSE](http://www.youtube.com/watch?v=mA_v1F7TSE)
Use Case 2
Analysing Scientific Publication Data

1. PDF extractor: extract table from a publication
Data Extractor
2. Correcting the table
3. Annotation the table (dimension and measures)
4. Export as RDF Data Cube
5. Vis Wizard: interactive analysis
Integration With Other Platforms
Semantic MindMaps

- MindMeister mind mapping web platform
- Generate mind maps from CODE Tools
  - Insights gained from LOD: data set export, visualisation export
Integration With Other Platforms
42-Data Platform

- **42-data**: data centric Q&A platform
  - Developed by University of Passau
- **Discussions backed by embedded**
  - Data sets - empirical facts in numerical LOD (SPARQL Queries, RDF Cubes)
  - Insights obtained through visual analysis
  - 3rd party resources
- **Data marketplace**: donation-based model

Link: [http://42-data.org](http://42-data.org)
YouTube Channel: [http://www.youtube.com/playlist?list=PLMwJwZJdjUFY9ea6VTC7VUwwMW5v8gJDT](http://www.youtube.com/playlist?list=PLMwJwZJdjUFY9ea6VTC7VUwwMW5v8gJDT)
Evaluation
Formative Study

• Goal: ascertain users can perform interactive analysis
  – Visualisation Wizard (primarily), Query Wizard
• 8 participants (age 24 – 38) no background in semantic technologies
  – Received short training
• 6 tasks testing filtering, aggregation, coordinated brushing
  – Loosely time constrained
  – Included a question – user had to generate an insight!
• Thinking aloud protocol to uncover usability issues
• NASA Task Load Index to measure workload
• Concluding questionnaire (pros, cons, suggestions)
Evaluation

Lessons Learnt

• Generally simple and easy to use, good design
  – Workload: on average below 1/3 of the scale
  – Users quickly learned new concepts
• Data quality and service quality issues
• Purpose: statistical and log data analysis
• Alternative tools: Google, manual copy pasting, spreadsheet
• Usability issues in coordinated brushing
• Appreciated automatic visualisation
  – A non-optimal visualisation shown initially - confusing
• Observation: users ignored data volumes
• Backtracking support missing – need to restart from beginning
Thank you for your attention!

Questions?

Vedran Sabol
vsabol@know-center.at
Area Manager Knowledge Visualisation
Know-Center GmbH

http://code-research.eu/