CODE
Commercial Empowered Linked Open Data Ecosystem in Research

presented by Florian Stegmaier
University of Passau
2012-10-04
Some basic facts...

- Budget: 2,4M €, funded by the European Commission
- Started in May 2012 with a runtime of 2 years
Current situation

• Data is being produced in an immense rate:
  • Evaluation campaigns (e.g., CLEF campaign)
  • Benchmarking communities (e.g., TPC)
  • Researchers (e.g., proceedings, journals or slides)

Most data remains unstructured and sophisticated access methods are missing!
Why is this a problem?!

- From a data perspective...
  - ...what is the quality of data?
  - ...how to deal with missing values?
- From a user perspective...
  - ...how can I compare this data? baseline?
  - ...are there contradicting facts?

The semantics of documents must be unleashed to make them accessible and processible!
The long way to knowledge...
Step 1: Analyze data

• Analysis of documents has to find:
  • Structural elements (TOC, images, etc.)
  • Extract facts and numerical measures
  • Disambiguate facts (from „string“ to „object“)
• Automatic annotation is defective or not complete
• Crowdsourced annotation of documents
• Marketplace offers revenue for expert knowledge
Step 2: Lift and extend data

- Extracted and disambiguated data will be lifted into the Linked Data cloud
- Interlink with already existent data of the cloud
- Enrich data with provenance information (increase quality estimations)
- Perform OLAP queries on data cubes (e.g., time series)

Enriched and aggregated data is exposed as Linked Data endpoint.
Step 3: Interact with data

- Query wizard will focus on:
  - Excel based interaction possibilities
  - On the fly creation of statistical analyses on a federated dataset
  - Marketplace encourages users to interact with data

Non-IT (but maybe domain) experts are able to create visual analytics as well as create new data cubes.
...does all this actually work?

Current analysis of PDFs is able to discover basic table of contents, reading direction, as well as specific objects.
...how are the users involved?

One possible way to engage users in annotating data is the Mendeley Desktop. (early stage)
Basic triplification chain established to lift table based data into a Semantic Web compatible data cube.

...what about lifting data?
...how can i find data?

The first prototype of the query wizard is able to show and interact with retrieved data in a Excel-like manner.
...and the marketplace?

The data can be exposed in several ways, just like in mind maps to help things getting structured. (example shows biggerplate)
Thank you for your attention!

http://www.code-research.eu/
https://www.facebook.com/CODEresearchEU
#CODEresearchEU (Twitter)

Thanks to Michael Granitzer, Christin Seifert, Kai Schlegel and Sebastian Bayerl for supporting me with input, figures and slide templates and last but not least our consortium for the prototype screenshots ;)}