Linked Data Visualization

Laura Po, PhD
Department of Engineering "Enzo Ferrari"
University of Modena and Reggio Emilia
Italy

3th Keystone Training School - Keyword Search in Big Linked Data
Institute for Software Technology and Interactive Systems, TU Wien, Austria
Goal of the Talk

• To provide practical skills required for exploring LOD sources

Outline

• The importance of visualization
• How a Linked Data Visualization Process can be defined
• Practical use of LOD/ RDF browsers and visualization toolkits
Why is visualization of Linked Data important?

- Large and Dynamic Data
  - Efficiently and effectively handle billions of objects within dynamic datasets

- Visual presentation and interaction issues
  - Offer ways to easily explore datasets
  - Proposing summaries and overviews
  - Incremental and progressive techniques

- Variety of Users and Tasks
BOLD – Big Open Linked Data
"The bigger the number, the harder it can be to visualise"

Bratsas et al (2016). Preface on special session “data impact: Big, open, linked data innovations” at 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP) https://doi.org/10.1109/SMAP.2016.7753368


Why visualize data instead of provide statistic analysis?

- Anscombe's quartet of datasets having similar statistical properties but appearing very different when plotted

http://en.wikipedia.org/wiki/Anscombe's_quartet
Users

PRODUCERS

Consumers

domain expert

Lay-users

Technical expert
LOD Visualization

• LOD simplifies accessing and integrating data from different sources
• SPARQL makes it easy to select from, and analyse the data

• It's natural to visualise the data as graphs (networks)
  ... but other forms of visualisation also possible
Example of LOD visualization process

**SPARQL query:** Retrieve number of releases per country of The Beatles

```sparql
SELECT ?country (COUNT(?release) AS ?releases)
WHERE {
  ?release a mo:Release ;
  mo:label ?label .
} GROUP BY ?country
ORDER BY DESC(?releases)
```

**Formatting the names of the countries**

```sparql
?country_code2 := REPLACE(str(?country), "http://ontology.es/place/", "", "")
?country_code := REPLACE(?country_code2, "", "", "")
```

**Selecting the visualization technique (input, output)**

```java
#widget: HeatMap |
input = 'country_code' |
output = {'releases'}
```
Heatmap visualization of The Beatles releases
LOD visualization systems

They can be classified in 6 categories
1. Browsers and Exploratory systems
2. Generic visualization systems
3. Domain vocabulary & device specific systems
4. Graph-based visualization systems
5. Ontology visualization systems
6. Visualization libraries

Evolution over time

<table>
<thead>
<tr>
<th>System category / year</th>
<th>2001...</th>
<th>...2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browsers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Textual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faceted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other paradigms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommenders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type/domain specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-type/domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undisclosed/commercial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>View-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algorithm-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research focus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Exploratory search

- Exploratory search systems (ESS) forms a special category of seeking information on the Web with the purpose of revealing related information to the searcher along with retrievals of what have been searched for.


## Classification and Comparisons

<table>
<thead>
<tr>
<th>Tool Name</th>
<th>Versions</th>
<th>Source Visualization</th>
<th>Querying</th>
<th>Result Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type</td>
<td>Last Update</td>
<td>Classes</td>
<td>Instances</td>
</tr>
<tr>
<td>LODive</td>
<td>live demo [6], desktop</td>
<td>2012</td>
<td>only for the selected classes</td>
<td>only for the selected classes</td>
</tr>
<tr>
<td>LODinilla</td>
<td>online [7]</td>
<td>2014</td>
<td>only for the selected classes</td>
<td>only for the selected classes</td>
</tr>
<tr>
<td>gFacet</td>
<td>live demo [1], desktop</td>
<td>2011</td>
<td>only for the selected classes</td>
<td>only for the selected classes</td>
</tr>
<tr>
<td>LOD Visualizer</td>
<td>online [4]</td>
<td>2012</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>isSPARQL</td>
<td>online [2], desktop</td>
<td>2014</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>SPARKIS</td>
<td>online [9]</td>
<td>2014</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>PAYOLA</td>
<td>live demo [8], desktop</td>
<td>2013</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>LODatio</td>
<td>online [5]</td>
<td>2013</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Linked Data Query Wizard</td>
<td>online [3]</td>
<td>2014</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>VisualBox</td>
<td>demo [10], desktop</td>
<td>2013</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Affective Graph</td>
<td>—</td>
<td>2014</td>
<td>only for the selected classes</td>
<td>no</td>
</tr>
<tr>
<td>LODiX</td>
<td>online</td>
<td>2014</td>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

https://code.research.eu/
EU project 2012-2014

http://code.know-center.tugraz.at/search
TEST

• Using CODE Linked Data Query Wizard search for "Johann Strauss II" within the Dbpedia source

• Explore the result

• Add columns that show some property like "birth place", "given name", "music composer of ", ...
<table>
<thead>
<tr>
<th>Label</th>
<th>Type</th>
<th>Birth date</th>
<th>Birth place</th>
<th>IsPrimaryTopicOf</th>
<th>Music by of</th>
<th>Music composer of</th>
<th>Musician of (old)</th>
<th>GivenName</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johann Strauss II</td>
<td></td>
<td>1825-10-25</td>
<td></td>
<td>Johann Strauss II</td>
<td></td>
<td>Paradise Found (musical)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The Great Waltz</td>
<td>A Night in Venice (1934 film)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A Night in Venice (1953 film)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Boudu Saved from Drowning</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Oh... Rosalinda!</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The Reluctant Sadist</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Vienna Blood (film)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Voices of Spring (1933 film)</td>
<td></td>
</tr>
</tbody>
</table>

- Johann Strauss II
TEST

• Navigate the Linked Jazz cloud
• Change the visualization option (fized, similar, gender, dynamic)
LOD live

LodLive project provides a demonstration of the use of Linked Data standards (RDF, SPARQL) to browse RDF resources. The application aims to spread linked data principles using a simple and friendly interface with reusable techniques.

http://en.lodlive.it/
By using LodLive online to explore dbpedia resources, search for Johann Strauss II [http://en.lodlive.it/](http://en.lodlive.it/)
- who is he?
- where was he born? where did he died?
- Is he the son of Johann Strauss?
- find which type are associated to him
LODmilla


http://lodmilla.sztaki.hu/lodmilla
TEST

Using LODMilla search and add the following node from Dbpedia:

- Johann Strauss II
- Vienna
- The Blue Danube
- Austria
- Johann Strauss I
- Wolfgang Amadeus Mozart
- Composer
- Musician

- Look at the connections between nodes
LODEX

It is a tool for producing a representative summary of a Linked open Data (LOD) source starting from scratch, thus supporting users in exploring and understanding the contents of a dataset.

LODeX extracts statistical indexes that uses to build the representative summary, by querying the SPARQL endpoint of a LOD source.

• LODEX 2.0 (http://www.dbgroup.unimo.it/lodex2) includes the possibility to compose visual queries by selecting objects from the representative summary of a LOD source
• LODEX Cluster (http://www.dbgroup.unimo.it/lodex2/testCluster) provides a more concise schema for huge datasets
LODeX Architecture


Benedetti, et al. (2015), Visual Querying LOD sources with LODeX. Proceedings of the 8th International Conference on Knowledge Capture (ACM)


Benedetti, et al. (2014), A Visual Summary for Linked Open Data sources. ISWC 2014 Posters & Demo Track, Riva del Garda, Italy, ISSN: 1613-0073

Benedetti, et al. (2014), Online Index Extraction from Linked Open Data Sources. Second International Workshop on Linked Data for Information Extraction (LD4IE) @ (ISWC 2014), Riva del Garda, Italy, ISSN: 1613-0073
The information contained in the intensional knowledge can be incomplete or absent.
Schema Summary – Building a Visual Query

LODeX Schema Summary

Organization 1869 instances

Properties

- activeIn ➔ Feature 6.33
- sector ➔ Sector 2.95
- subject ➔ Concept 1.26
- depiction ➔ Image 0.10

Attributes

- name 1.66
- abbreviation 1
- street 0.96
- phone 0.94
- city 0.94
- zipCode 0.88

Linked Clean Energy Data (reegle.info)
Refinement Panel

LODeX Schema Summary

Attribute: name - Mandatory
Class: Select a class - Mandatory
Order: Select a parameter - order condition

SPARQL Query

```
WHERE {
  OPTIONAL { ?Feature a <http://www.geonames.org/ontology#Feature> . }
  OPTIONAL { ?Feature <http://www.geonames.org/ontology#name> ?name1 . }
  OPTIONAL { ?Sector a <http://reagle.info/schema#Sector> . }
  OPTIONAL { ?Sector <http://www.w3.org/2004/02/skos/core#definition> ?definition . }
}
LIMIT 50
```
By using Lodex  [http://www.dbgroup.unimore.it/lodex2/](http://www.dbgroup.unimore.it/lodex2/) find, navigate and explore the following datasets:

• **European Television Heritage**
  • How many classes it has? How many properties it has?
  • How many vocabulary are used?

• **Nobel Prizes**
  • How many vocabulary are used?
  • Define a query that select person (label, gender,name) that won a Nobel Prize , i.e.have an Award (year,label), add also the Category of the Award if it exists
Conclusions

• Analysis of the needs for visualization in the LOD context

• Practical use of some LOD browsers and visualization toolkits

• Navigation and exploration of some datasets and the construction of different visualizations
Actual limitations and challenges

• Most of the LOD visualization tools are still in-lab prototypes

• Lots tools allow the exploration of a limited list of datasets or have limitations in terms of size, format (SPARQL endpoint/RDF dumps) of the datasets they can explore

• SPARQL endpoints might be offline or have bad performance such as taking long time to respond to some queries.

• For dealing with BOLD, graph simplification is needed:
  • reducing size could be possible through filtering or aggregation
My vision

Aggregated View

Schema Summary

RDF Graph
Feel free to contact me at laura.po@unimore.it

You can find more information on my research and my group at www.dbgroup.unimore.it

Slide are available on http://www.slideshare.net/polaura

THANK YOU