The Linking Open Data
W3C SWEO Community Project

Chris Bizer
Freie Universität Berlin

Basic Ideas

1. Apply the Linked Data principles to large open datasets
   - Publish open datasets as Linked Data on the Web
   - Set RDF Links between data items from different datasets

2. Provide a forum for working on the tasks arising from having lots of Linked Data on the Web
   - Browsers and other user interfaces
   - Semantic Web crawlers and search engines
   - RDF repositories and reasoning engines
   - Linked Data publishing tools
   - Record linkage algorithms
   - Data cleansing tools
**Spirit of the Project**

- Don’t talk about the Semantic Web, but do it, NOW!
- Get as much Linked Data as possible online, NOW!
- Get as many people engaged as possible.
- Solve the easy problems first, do the hard ones later.

**Infrastructure**

- **Wiki Page**
  - [http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData](http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData)
- **Mailing List**
- **Participating in the project**
  - Put your name on the Wiki page
  - Subscribe to the mailing list
  - Do something useful
What did we achieve so far?

- About 600 million triples served as Linked Data
- About 100,000 RDF links between different data sets

Where will we be in 2 month?
Example Data Source: DBpedia

- DBpedia.org is a community effort to
  - extract structured information from Wikipedia
  - make this information available on the Web under an open license
  - interlink the DBpedia dataset with other datasets on the Web

- Contributors
  - Freie Universität Berlin (Germany)
  - Universität Leipzig (Germany)
  - OpenLink Software (UK)
  - Linking Open Data Community (W3C SWEO)

Extracting Structured Data from Wikipedia

http://en.wikipedia.org/wiki/Calgary

<http://dbpedia.org/resource/Calgary>
  dbpedia:native_name "Calgary" ;
  dbpedia:altitude "1048" ;
  dbpedia:population_city "988193" ;
  dbpedia:population_metro "1079310" ;
  mayor_name
    dbpedia:Dave_Bronconnier ;
  governing_body
    dbpedia:Calgary_City_Council ;
  ...

- Altogether 9,100,000 RDF triples extracted from 754,000 infoboxes
The DBpedia Dataset

- 1,600,000 concepts
- including
  - 58,000 persons
  - 70,000 places
  - 35,000 music albums
  - 12,000 films
- described by 91 million triples
- using 8,141 different properties.
- abstracts in 10 different languages
- 557,000 links to pictures
- 1,300,000 links to relevant external web pages
- 207,000 Wikipedia categories
- 75,000 YAGO categories

Multi-Lingual Abstracts

- The dataset contains a short and a long abstract for each concept.
- Short abstracts
  - English: 1,637,622
  - German: 246,791
  - French: 206,085
  - Dutch: 133,746
  - Polish: 118,874
  - Italian: 113,950
  - Spanish: 112,417
  - Japanese: 106,610
  - Portuguese: 104,842
  - Swedish: 100,267
  - Chinese: 54,991
Example RDF Links

**Out-Bound RDF Link**

<http://dbpedia.org/resource/Berlin> owl:sameAs
<http://sws.geonames.org/2950159> .

<http://dbpedia.org/resource/Tim_Berners-Lee> owl:sameAs

**In-Bound RDF Links**

<http://www.w3.org/People/Berners-Lee/card#i> owl:sameAs

<http://richard.cyganiak.de/foaf.rdf#cgyri> foaf:topic_interest

Browsing DBpedia together with Linked Data
Browsing In and Out of DBpedia with DISCO

Label of the displayed resource

Navigation box

Go to URI button

Sources of each piece of information

Resource description

List of all source graphs

Link for displaying the session cache