LinkedGeoData

Adding a Spatial Dimension to the Web of Data

http://linkedgeodata.org

Jens Lehmann, Claus Stadler
AKSW Research Group
University of Leipzig
Conversion, interlinking and publishing of OpenStreetMap.org* data sets as RDF.

* "Wikipedia for geographic data"
Structure

- Motivation
- OpenStreetMap
- Mapping to RDF/OWL
- **Access** (How LinkedGeoData is published)
- DBpedia Mapping
- Conclusion & Future Work
Motivation

• Ease information integration tasks that require spatial knowledge, such as
  • Offerings of bakeries next door
  • Map of distributed branches of a company
  • Historical sights along a bicycle track
  • Therefore use RDF/OWL in order overcome structural and semantic heterogeneity.
    • Requires a vocabulary – which we try to establish.

• LOD cloud contains data sets with spatial features
  • e.g. Geonames, DBpedia, US census, EuroStat
  • But: they are restricted to popular or large entities like countries, famous places etc.
    • Therefore they lack buildings, roads, mailboxes, etc.
OpenStreetMap - Datamodel

• Basic entities are:
  • **Nodes** Latitude, Longitude
  • **Ways** Sequence of nodes
  • **Relations** Associations between any number of nodes, ways and relations.

• Each entity may be described with **tags** (= key-value pairs)
Example: Leipzig's zoo

name = Zoo Leipzig
tourism = zoo
wikipedia:en = Leipzig_Zoo
created_by = Potlatch 0.10f
<table>
<thead>
<tr>
<th>node_id</th>
<th>k</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>259212302</td>
<td>name</td>
<td>Universität Leipzig, Mathematik und Informatik</td>
</tr>
<tr>
<td>259212302</td>
<td>amenity</td>
<td>university</td>
</tr>
<tr>
<td>259212302</td>
<td>addr:street</td>
<td>Johannisgasse</td>
</tr>
<tr>
<td>259212302</td>
<td>addr:postcode</td>
<td>04103</td>
</tr>
<tr>
<td>259212302</td>
<td>addr:housenumber</td>
<td>26</td>
</tr>
<tr>
<td>259212302</td>
<td>addr:city</td>
<td>Leipzig</td>
</tr>
</tbody>
</table>
Data/Mapping Example

<table>
<thead>
<tr>
<th>node_id</th>
<th>k</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>259212302</td>
<td>name</td>
<td>Universität Leipzig, Mathematik und Informatik</td>
</tr>
<tr>
<td>259212302</td>
<td>amenity</td>
<td>university</td>
</tr>
<tr>
<td>259212302</td>
<td>addr:street</td>
<td>Johannisgasse</td>
</tr>
<tr>
<td>259212302</td>
<td>addr:postcode</td>
<td>04103</td>
</tr>
<tr>
<td>259212302</td>
<td>addr:housenumber</td>
<td>26</td>
</tr>
<tr>
<td>259212302</td>
<td>addr:city</td>
<td>Leipzig</td>
</tr>
</tbody>
</table>

```
lgd:node259212302
  a                         lgdo:University ;
  rdfs:label                "Universität Leipzig, Mathematik und Informatik" ;
  lgdo:hasCity              "Leipzig" ;
  lgdo:hasHouseNumber       "26" ;
  lgdo:hasPostalCode        "04103" ;
  lgdo:hasStreet            "Johannisgasse" ;
  georss:point              "51.3369334 12.385401" ;
  geo:lat                   51.3369334 ;
  geo:long                  12.385401 .
```
Mapping Types

- Three Mapping Types
  - Text
    - (5, name, Leipzig) → lgd:node5 rdfs:label "Leipzig"
    - (5, name:de, Leipzig) → lgd:node5 rdfs:label "Leipzig"@de
  - Datatypes
    - (6, seats, 4) → lgd:node6 lgdo:seats "4"^^xsd:integer
  - Classes/Object Properties
    - (7, place, city) → lgdn:7 a lgdo:City
    - (7, religion, pastafarian) → lgdn:7 lgdo:religion lgdo:Pastafarian
Access

- **Rest Interface** (based on Postgis DB, full osm dataset loaded, > 1 billion triples)
  - Supports limited queries (e.g. circular/rectangular area, filtering by labels)
- **Sparql Endpoints** (based on Virtuoso DB, subset of osm dataset loaded, ~222 mio triples)
  - Static (http://linkedgeodata.org/sparql)
  - Live (http://live.linkedgeodata.org/sparql)
- **Downloads** (http://downloads.linkedgeodata.org)
  - Monthly updates on the above datasets envisioned
LinkedGeoData Live

- OpenStreetMap provides full dumps and minutely **changesets** for download
  - Changesets are numbered, e.g. “001/234/567.osc.gz”

- We also convert the changesets to sets of **added** and **removed** triples (relative to our store) and publish them
  - 001/234/567.added.nt.gz
  - 001/234/567.removed.nt.gz

- Advantage: Other users could easily **sync** their RDF store with LinkedGeoData
Given a DBpedia point, query LGD points within type specific maximum distance

Basic idea (performed with Silk):

- Compute spatial score
- Compute name similarity (rdfs:label)
Given a DBpedia point, **query LGD points within type specific maximum distance**

Basic idea (performed with Silk):

- Compute **spatial score**
- Compute **name similarity** (rdfs:label)
- Combine both scores
- Depending on final score, either **automatically accept/reject links or mark for manual verification**.
Statistics (2011-Feb-23)

- 222,539,712 Triples
  - 6,666,865 Ways
  - 5,882,306 Nodes
- Among them
  - 352,673 PlaceOfWorship
  - 60,573 RailwayStation
  - 59,468 Recycling
  - 50,955 Town
  - 30,099 Toilet
  - 7,222 City
Conclusion

- **OpenStreetMap**
  - immensely successful project for collaboratively creating free spatial data
  - Community uses **key value structures**, which provide a **rich source of information**
  - Key strength: **broad coverage**

- **LGD Contributions**
  - Established **mapping to Dbpedia**
  - **Geonames mapping** partially done (37 different entity types cities, churches, ...)
  - **Facet-based LGD Browser** provides an interface for OSM/LGD, which highlights its **structural aspects**
  - **Live sync**

- **Goal**: Make LGD as useful (successful) as DBpedia for the geospatial domain
Thanks for your Attention!