

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**.

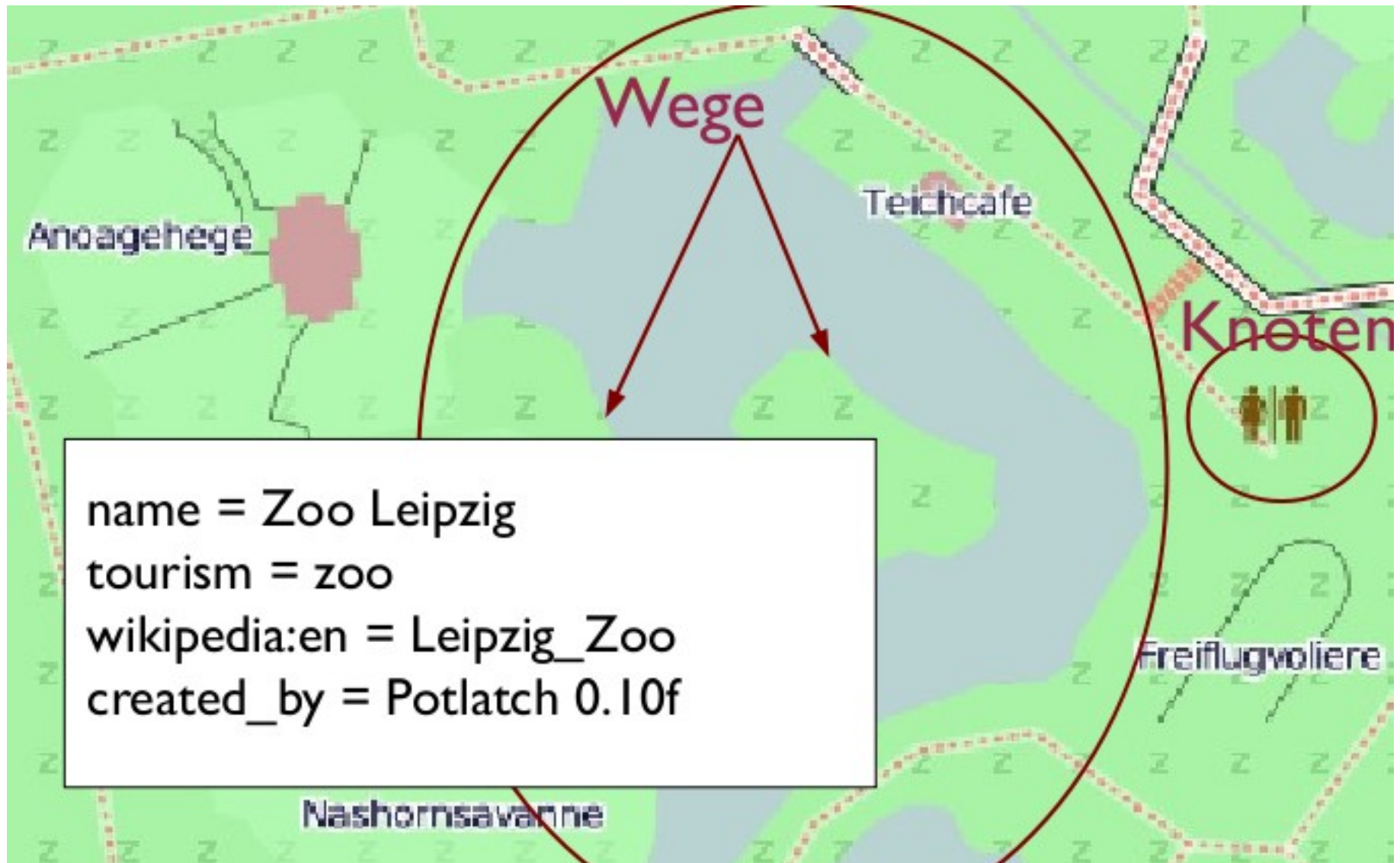
* "Wikipedia for geographic data"

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

- 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.

- 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

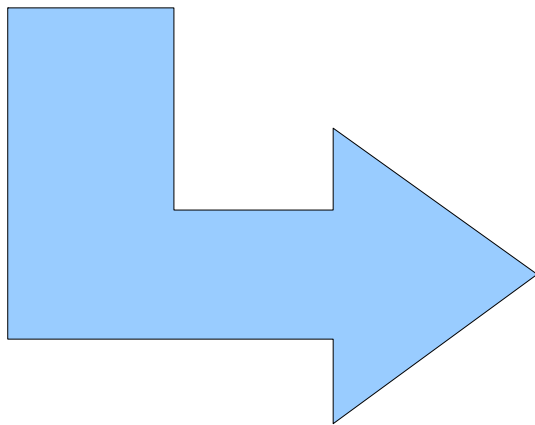


Data/Mapping Example

node_id	k	v
259212302	name	Universität Leipzig, Mathematik und Informatik
259212302	amenity	university
259212302	addr:street	Johannisgasse
259212302	addr:postcode	04103
259212302	addr:housenumber	26
259212302	addr:city	Leipzig

Data/Mapping Example

node_id	k	v
259212302	name	Universität Leipzig, Mathematik und Informatik
259212302	amenity	university
259212302	addr:street	Johannissgasse
259212302	addr:postcode	04103
259212302	addr:houseNumber	26
259212302	addr:city	Leipzig



```

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

- 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`

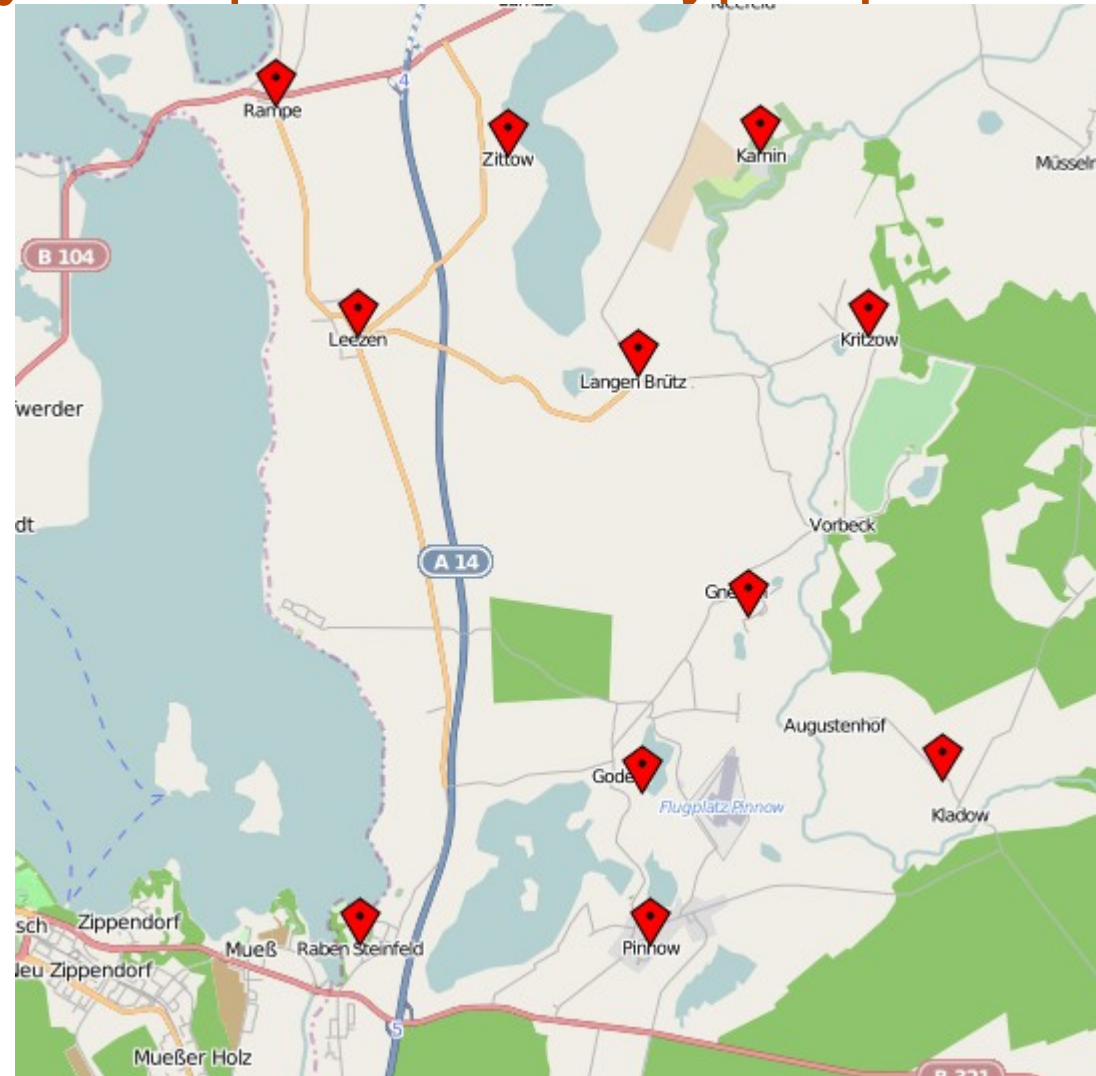
- **Rest Interface** (based on Postgis DB, full osm dataset loaded, > 1billion triples)
 - Supports limited queries (e.g. circular/rectangular area, filtering by labels)
- **Sparql Endpoints** (based on Virtuoso DB, subset of osm dataset loaded, ~222mio triples)
 - **Static** (<http://linkedgeo.org/sparql>)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

- 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):

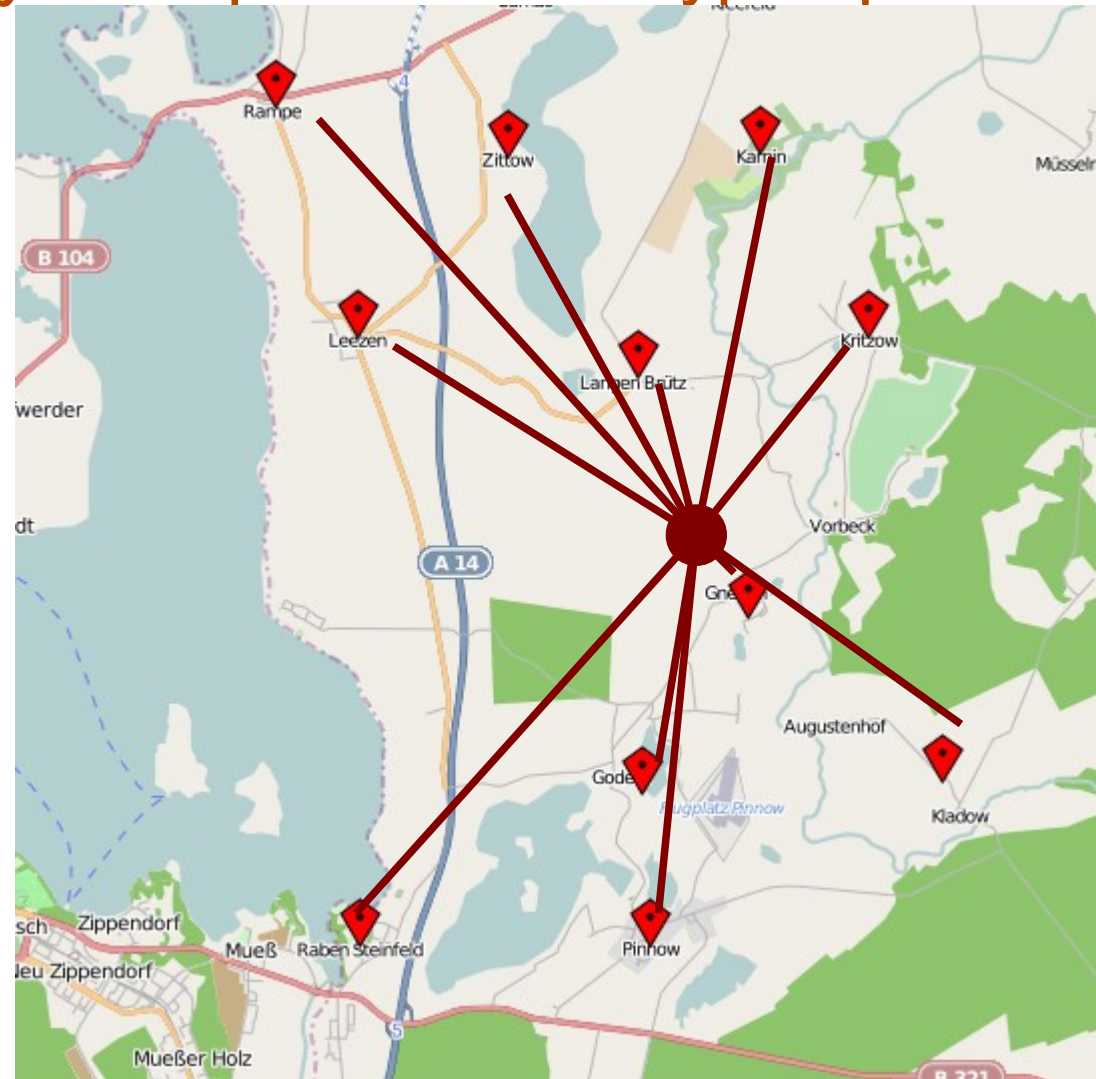
- Compute **spatial score**
- Compute **name similarity** (rdfs:label)



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**.



- 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

- **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



The End

Thanks for your Attention!

