

Sematic Web - Projektabschluss

Alexander Müller

7. Juli 2015

Fakultät
Informatik, Mathematik und Naturwissenschaften

Fragestellung

„Welcher ist der familienfreundlichste Stadtteil Leipzigs“

Kriterien:

- Schulen (Grundschulen, Gymnasien, Realschulen)
- Kindergärten
- Spielplätze
- Schwimmbäder
- Sportplätze
- Sporthallen



Familienfreundliche Stadt



„Wir wollen, dass junge Menschen gern in Leipzig leben und unsere Stadt wählen, um eine Familie zu gründen. Helfen Sie mit, dass uns dies weiterhin gelingt! Gestalten Sie Leipzig kinder- und familienfreundlich mit!“

Oberbürgermeister Burkhard Jung

➤ <http://www.leipzig.de/jugend-familie-und-soziales/familienfreundliche-stadt/>

Relevante Datenquellen

➤ Adressen, Ortsteile und Schulen der Stadt Leipzig

Link: <http://leipzig-data.de/Data/> (OntoWiki)

Datenformat: **RDF**

Schnittstelle: **HTTP**

Lizenz: **CC0 1.0 Universal**

Open Data: ★★★★★



Relevante Datenquellen

- Kindergärten, Spielplätze, Schwimmbäder, Sportplätzen, Sporthallen der Stadt Leipzig

Link: <http://www.leipzig.de/suchergebnisse-adressdatenbank/>

Datenformat: HTML

Schnittstelle: HTTP, (interne API)

Lizenz: urheber- bzw. leistungrechtlich geschützt

Open Data: ★★★★★



Daten Extraktion

Beispiel-Query

[http://www.leipzig.de/suchergebnisse-dressdatenbank/
?tx_ewerkaddressdatabase_pi\[showAll\]=1
&tx_ewerkaddressdatabase_pi\[topics\]=105
&tx_ewerkaddressdatabase_pi\[query\]=
&tx_ewerkaddressdatabase_pi\[action\]=list
&tx_ewerkaddressdatabase_pi\[controller\]=Address](http://www.leipzig.de/suchergebnisse-dressdatenbank/?tx_ewerkaddressdatabase_pi[showAll]=1&tx_ewerkaddressdatabase_pi[topics]=105&tx_ewerkaddressdatabase_pi[query]=&tx_ewerkaddressdatabase_pi[action]=list&tx_ewerkaddressdatabase_pi[controller]=Address)

tx_ewerkaddressdatabase_pi[topics]=

- 105 – Kindergärten
- 111 – Spielplätze
- 375 – Sporthallen
- 377 – Sportplätze
- 433 & 435 – Schwimmbäder

Daten Extraktion

Java Applikation für Daten von www-leipzig.de

1. Extraktion der Daten mit JSOUP

```
Document doc = Jsoup.connect("http://www.leipzig.de/suchergebnisse-adressdatenbank/")
    .data("tx_ewerkaddressdatabase_pi[showAll]", "1").data("tx_ewerkaddressdatabase_pi[action]", "list")
    .data("tx_ewerkaddressdatabase_pi[controller]", "Address")
    .data("tx_ewerkaddressdatabase_pi[topics]", itemId).timeout(0).get();

Elements results = doc.getElementsByClass("search-result-list");

for (Element result : results) {
    Elements entries = result.getElementsByTag("h3");
    // ...
}
```

2. Speicherung als JSON-Dateien

```
[
  {
    "name": "Abenteuerland Kombinierte Tageseinrichtung",
    "url": "abenteuerland-kombinierte-tageseinrichtung",
    "strasse": "Heilemannstraße 19",
    "plz": "04277",
    "ort": "Leipzig",
    "ortsteil": "Connewitz"
  },
  {
    "name": "Am Gutenbergplatz Kombinierte Kindertageseinrichtung",
    // ...
  }
]
```

Daten Extraktion

3. Normalisieren der Daten

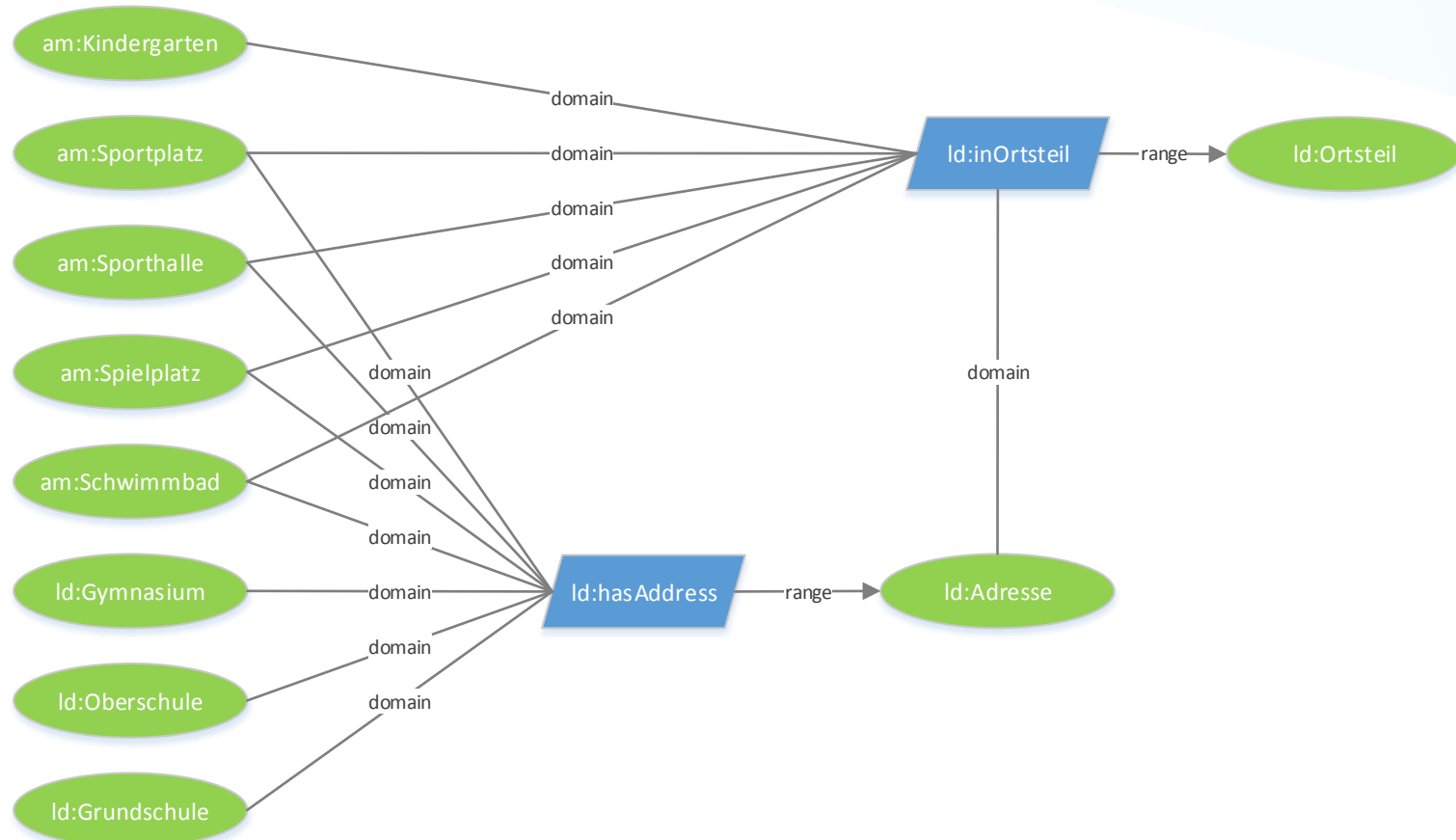
4. Erstellen der RDF-Dateien mit *Jena Framework*

```
am:Kindergarten a owl:Class ;  
    rdfs:label "Kindergarten in Leipzig" .
```

```
<http://www.imn.htwk-leipzig.de/~amuelle3/Data/Kindergarten/abenteuerland-kombinierte-tageseinrichtung>  
    a am:Kindergarten ;  
    rdfs:label "Abenteuerland Kombinierte Tageseinrichtung" ;  
    ld:hasAddress <http://leipzig-data.de/Data/04277.Leipzig.Heilemannstrasse> ;  
    ld:inOrtsteil <http://leipzig-data.de/Data/Ortsteil/Connewitz> .
```

```
<http://www.imn.htwk-leipzig.de/~amuelle3/Data/Kindergarten/kindertageseinrichtung-karl-heine-strasse-95>  
    a am:Kindergarten ;  
    rdfs:label "Kindertageseinrichtung Karl-Heine-Straße 95" ;  
    ld:hasAddress <http://leipzig-data.de/Data/04229.Leipzig.Karl-Heine-Strasse> ;  
    ld:inOrtsteil <http://leipzig-data.de/Data/Ortsteil/Plagwitz> .
```


RDF-Schema



PREFIX am: <<http://www.imn.htwk-leipzig.de/~amuelle3/Data/Model/>>

PREFIX Id: <<http://leipzig-data.de/Data/Model/>>

Auswertung

➤ SPARQL-Anfrage an den Triple Store *Apache Fuseki*

```
1 PREFIX cc: <http://creativecommons.org/ns#>
2 PREFIX am: <http://www.imn.htwk-leipzig.de/~amuelle3/Data/Model/>
3 PREFIX ld: <http://leipzig-data.de/Data/Model/>
4 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
5 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
6 PREFIX owl: <http://www.w3.org/2002/07/owl#>
7 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
8
9 SELECT ?Ort ?grund ?gym ?ober ?kiga ?spielplatz ?schwimmbad ?sportplatz ?sporthalle
10 where
11 {
12     ?ort a ld:Ortsteil .
13     ?ort rdfs:label ?Ort .
14     OPTIONAL {
15         select ?ort (count(?g) as ?grund )
16         where {
17             ?g rdf:type ld:Grundschule.
18             ?g ld:hasAddress ?a.
19             ?a ld:inOrtsteil ?ort.
20         }
21         GROUP BY ?ort
22     }
```

Probleme

- nicht normalisierte Daten
 - Zahlendreher in der PLZ (04**2**55 → 04155)
 - Straßennamen (Voll**l**bedingstraße → Vol**l**bedingstraße)
- Komplexität der Sparql-Query
 - Time Out bei Fuseki und Virtuoso

The query timed out (restricted to 2400000 ms) (#503)

```
Error 503: The query timed out (restricted to 2400000 ms)
```

```
Fuseki - version 2.0.0 (Build date: 2015-03-08T09:49:20+0000)
```

Ergebnisse

Die familienfreundlichsten Stadtteile

PLATZ	Ortsteil	Grund- schulen	Gymnasie n	Ober- schulen	Kinder- gärten	Spielplätze	Schwimm- bäder	Sport- plätze	Sport- hallen	GESAMT
1	Reudnitz- Thonberg	2	1	1	8	14	0	1	1	28
2	Stötteritz	1	1	0	8	13	2	1	2	28
3	Engelsdorf	1	1	0	6	15	0	1	0	24
4	Zentrum- Südost	2	1	1	12	7	1	0	0	24
5	Neustadt- Neuschönefeld	3	0	0	8	11	0	0	1	23
6	Paunsdorf	3	0	1	10	5	1	3	0	23

Ergebnisse

Die „familienUNfreundlichsten“ Stadtteile

PLATZ	Ortsteil	Grund- schulen	Gymnasie n	Ober- schulen	Kinder- gärten	Spielplätze	Schwimm- bäder	Sport- plätze	Sport- hallen	GESAMT
63	Grünau- Siedlung	0	0	0	0	1	0	0	0	1
62	Baalsdorf	0	0	0	1	2	0	0	0	3
61	Heiterblick	0	1	0	1	2	0	0	0	4
60	Miltitz	1	0	0	1	1	0	1	0	4
59	Meusdorf	1	0	0	1	2	0	0	1	5
58	Zentrum	0	0	0	0	2	0	0	3	5

Sematic Web - Projektabschluss

Alexander Müller

7. Juli 2015

Fakultät
Informatik, Mathematik und Naturwissenschaften