

MI-SWE semestrální práce

Kyryl Bulat

Data

- Parkomaty – 984 instancí
- Stanoviště tříděného odpadu – 20816 instancí
- Veřejné toalety – 247 instancí
- Každý dataset je zaměřen na instance v rámci města Prahy. Každá instance obsahuje informaci o zeměpisné šířce a zeměpisné délce objektu.

Ukázka dat

```
{
  "type" : "Feature",
  "geometry" : {
    "type" : "Point",
    "coordinates" : [ 14.415883054000062, 50.077568300000005 ]
  },
  "properties" : {
    "OBJECTID" : 1,
    "POSKYT" : "HMP-TSK",
    "PA" : "1000065",
    "PX" : "P1",
    "STREET" : "Myslíkova",
    "CODE" : 1000065
  }
},
```

Parkomat

```
{
  "type" : "Feature",
  "geometry" : {
    "type" : "Point",
    "coordinates" : [ 14.530531866000047, 50.02791253100003 ]
  },
  "properties" : {
    "OBJECTID" : 1,
    "STATIONID" : 1308,
    "TRASHTYPENAME" : "Čiré sklo",
    "CLEANINGFREQUENCYCODE" : 41,
    "CONTAINERTYPE" : "3350 Atomium Reflex SV"
  }
},
```

Stanoviště tříděného odpadu

```
{
  "type": "Feature",
  "geometry": {
    "type": "Point",
    "coordinates": [
      14.456224317,
      50.0823295810001
    ]
  },
  "properties": {
    "OBJECTID": 1,
    "ADRESA": " ",
    "OTEVRENO": "nonstop",
    "CENA": "jednotná cena 5 Kč"
  }
},
```

Veřejná toaleta

Linkování

- Linkování probíhalo na základě údajů o zeměpisné šířce a zeměpisné délce objektů
- Parkomaty – tříděný odpad: **247 linků**
- Parkomaty – veřejné toalety: **6 linků**
- Tříděný odpad - veřejné toalety: **15 linků**
- Tříděný odpad – externí seznam objektů v Praze (oříznutý dataset z DBpedia): **55 linků**

localhost:9000/linking/cp_2/parkomaty_wc/editor

Silk Workbench

Start Workspace Status Editor Generate Links Reference Links About

Undo Redo Get LinkSpec Help

Property Paths ↻

Source: parkomaty

- (custom path)
- /schema:address
- /schema:branchCode
- /schema:containedInPlace

Target: wc

- (custom path)
- /schema:address
- /schema:geo
- /kbr:cena

Transformations Recommended ▾

- Constant
- Lower case
- Tokenize

Comparators Recommended ▾

- Equality
- Jaccard
- Levenshtein distance

Aggregators Recommended ▾

- Average
- Maximum
- Minimum

```
graph LR; subgraph Source [parkomaty]; direction TB; longitude1["longitude1  
/schema:geo/schema:longitude  
Path (Source)"]; longitude2["longitude2  
/schema:geo/schema:longitude  
Path (Target)"]; latitude1["latitude1  
/schema:geo/schema:latitude  
Path (Source)"]; latitude2["latitude2  
/schema:geo/schema:latitude  
Path (Target)"]; end; subgraph Target [wc]; direction TB; longitude_num["longitude_num  
Required  
Threshold 1.0E-4  
Weight 1  
minValue -Infinity  
maxValue Infinity  
Numeric similarity (Compare)"]; latitude_num["latitude_num  
Required  
Threshold 1.0E-4  
Weight 1  
minValue -Infinity  
maxValue Infinity  
Numeric similarity (Compare)"]; end; subgraph Aggregators; direction TB; average1["average1  
Required  
Weight 1  
Average (Aggregate)"]; end; longitude1 --> longitude_num; longitude2 --> longitude_num; latitude1 --> latitude_num; latitude2 --> latitude_num; longitude_num --> average1; latitude_num --> average1;
```

Použití nástroje SILK pro linkování dat

Nové souvislosti

- Po provedení linkování jsem schopen zjistit odpovědi na následující dotazy:
- „Existuje-li veřejná toaleta vedle parkomatů v různých částech Prahy?“
- „Kolik stanovišť tříděného odpadu existuje v Praze a jejich grupování podle typu kontejneru?“
- „Doplňte adresy pro veřejné toalety, které se nacházejí vedle parkomatů podle adresy parkomatu“.

Apache Jena Fuseki

datasetmanage datasetshelp

Server status:

Dataset: /all2

queryupload fileseditinfo

SPARQL query

To try out some SPARQL queries against the selected dataset, enter your query here.

EXAMPLE QUERIES

Selection of triplesSelection of classes

PREFIXES

rdfrdfswlxsdd

SPARQL ENDPOINT

CONTENT TYPE (SELECT)

CONTENT TYPE (GRAPH)

/all2/sparqlJSONTurtle

1

2PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

3PREFIX schema: <http://schema.org/>

4PREFIX kbr: <http://www.kyrylo.bulat.com/resource/>

5PREFIX owl: <http://www.w3.org/2002/07/owl#>

6PREFIX kbrwcp: <http://www.kyrylo.bulat.com/resource/wcprice/>

7PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

8

9SELECT

10?trashtypename (if(count(?tridodpad) > 4000, "More than 4000", count(?tridodpad)) as ?count)

11WHERE {

12?tridodpad a schema:Place;

13kbr:objectId ?objectId;

14kbr:trashtypename ?trashtypename;

15kbr:centisestupa ?centisestupa;

QUERY RESULTS

TableRaw Response

Showing 1 to 5 of 5 entries

Search: Show 100 entries

	trashtypename	count
1	kbr:trashtypename/5	"More than 4000"
2	kbr:trashtypename/4	"More than 4000"
3	kbr:trashtypename/3	"More than 4000"
4	kbr:trashtypename/2	"2951"^^xsd:integer
5	kbr:trashtypename/0	"1806"^^xsd:integer

Showing 1 to 5 of 5 entries

Použití nástroje Apache Fuseki jako triple store a SPARQL endpoint

Webová aplikace

- Nad SPARQL triple storem byla postavená webová aplikace, která dotazuje se na data pomocí SPARQL endpointu
- Pro vytvoření webové aplikace byla použita knihovna **apache-jena**.
- Webová aplikace realizuje následující funkčnosti:
- Zobrazení seznamu parkomatu, stanovišť tříděného odpadu
- Zobrazení detailu parkomatu a stanovišť tříděného odpadu, které se nacházejí vedle a vice versa
- Zobrazení odpovědi na tři SPARQL dotazy v rozumné čitelné podobě

Detail zaznamu	
Uri	http://www.kyrylo.bulat.com/resource/parkomaty/0
Object id	1
Poskytovatel	HMP-TSK
Část Prahy	P1
Adresa	Myslíkova
Branch code	1000065
Tříděný odpad vedle nachazehíci	<ul style="list-style-type: none">http://www.kyrylo.bulat.com/resource/tridodpad/743http://www.kyrylo.bulat.com/resource/tridodpad/5964http://www.kyrylo.bulat.com/resource/tridodpad/15458

Detail parkomatu

Detail zaznamu	
Uri	http://www.kyrylo.bulat.com/resource/tridodpad/743
Object id	744
Station id	4264
Trash type	Papír
Cleaning frequency code	11
Container type	120 normální HV
Parkomaty vedle nachazehíci	<ul style="list-style-type: none">http://www.kyrylo.bulat.com/resource/parkomaty/0

Detail stanoviště tříděného odpadu

Subject	
Entity: http://www.kyrylo.bulat.com/resource/wc/163 is a PublicToilet. It is situated on this address: "Vodičkova", "v průchodu, pod MÚ Praha 1, Vodičkova 18", it's price is "zdarma". Internal object id of this instance is: "164". Days and hours of work of this instance are "po a st 8-18 h; út a čt 8-16 h, pá 8-14 h".	
Entity: http://www.kyrylo.bulat.com/resource/wc/105 is a PublicToilet. It is situated on this address: "Národní", "Národní 63/26", it's price is "zdarma". Internal object id of this instance is: "106". Days and hours of work of this instance are "po-so 8:00-21:00, ne 9:00-20:00".	
Entity: http://www.kyrylo.bulat.com/resource/wc/215 is a PublicToilet. It is situated on this address: "Korunní", " ", it's price is "Jednotná cena 5 Kč". Internal object id of this instance is: "216". Days and hours of work of this instance are "po-pá 6-21 h; so-ne 8-21 h".	
Entity: http://www.kyrylo.bulat.com/resource/wc/10 is a PublicToilet. It is situated on this address: "Vinohradská", " ", it's price is "jednotná cena 5 Kč". Internal object id of this instance is: "11". Days and hours of work of this instance are "nonstop".	
Entity: http://www.kyrylo.bulat.com/resource/wc/104 is a PublicToilet. It is situated on this address: "Národní", "Národní 63/26", it's price is "zdarma". Internal object id of this instance is: "105". Days and hours of work of this instance are "po-so 8:00-21:00, ne 9:00-20:00".	
Executed query:	

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX schema: <http://schema.org/>
PREFIX kbr: <http://www.kyrylo.bulat.com/resource/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX kbrwcp: <http://www.kyrylo.bulat.com/resource/wcprice/>
```

```
CONSTRUCT {
?wc a schema:PublicToilet;
kbr:objectid ?objectid;
schema:address ?address, ?parkomatAddress;
kbr:openingHours ?openingHours ;
kbr:cena ?price.
}
WHERE {
?wc a schema:PublicToilet;
kbr:objectid ?objectid;
kbr:openingHours ?openingHours ;
kbr:cena ?price.
```

```
?parkomaty a schema:ParkingFacility;
skos:related ?wc;
schema:address ?parkomatAddress
```

```
OPTIONAL {
?wc schema:address ?address.
}
```

CONSTRUCT dotaz