

Protokoll der Realisierung

Daten transformieren

Daten von Geonames transformieren und nutzen

Geonames Conversion / Importer to Sesame Store

Die Daten bei Geonames können nur über Webservices bei Geonames abgefragt werden. Wir wollen die Daten jedoch direkt in Sparql-Queries ansteuern. Der Dump der Daten kann von der Geonames-Webseite heruntergeladen werden, jedoch ist das Datenformat in der Datei kein gültiges RDF-Format. Deshalb wurde dieses Projekt entwickelt, welches diesen Dump einliest und anschliessend in ein Sesame-Repository importiert. Ausserdem werden nur die für die Schweiz relevanten Daten importiert, da es nicht möglich war, den gesamten Dump in den Sesame-Store zu importieren.

↳ Sesame Repository für Geonames anlegen (nur Daten für die Schweiz)

Als erstes muss im Sesame ein Repository für die Daten angelegt werden. Als Repository-Typ wird der **Java Native Store** verwendet. Zusammenfassung des Repository nach dem Anlegen:

```
Repository Location
ID geonames
Title geonames
Location http://linkeddata.fh-htwchur.ch/openrdf-sesame/repositories/geonames
Sesame server http://linkeddata.fh-htwchur.ch/openrdf-sesame
Repository Size
Number of Statements 0
Number of Labeled Contexts 0
```

Vorgehen für das importieren des Geonames-Dump

Daten-Dump von Geonames holen und entpacken (auf dem Server linkeddata.fh-htwchur.ch)

```
cd /tmp/geonames_dump_all_countries
wget http://download.geonames.org/all-geonames-rdf.zip
unzip allCountries.zip
```

Der heruntergeladene Dump hat 20'226'712 Zeilen

Der Dump ist wie folgt aufgebaut (Auszug):

```
http://sws.geonames.org/1/
<?xml version="1.0" encoding="UTF-8" standalone="no"?><rdf:RDF xmlns:cc="http://creativecommons.org/ns#" xmlns:dcterms
http://sws.geonames.org/3/
<?xml version="1.0" encoding="UTF-8" standalone="no"?><rdf:RDF xmlns:cc="http://creativecommons.org/ns#" xmlns:dcterms
...
```

Java-Projekt Importer ausführen:

In der Datei `config.prop` wird die URL und der Login für den Sesame-Store sowie der Pfad zur Geonames Dump Datei konfiguriert:

```
# Filepath for Geonames Dump File
geonamesDump=/tmp/geonames_dump_all_countries/all-geonames-rdf.txt

# URL for the sesame repository
sesameRepoUrl=http://linkeddata.fh-htwchur.ch/openrdf-sesame

# username / password for the sesame server
sesameUsername=[sesameuser]
sesamePassword=[sesamepassword]
```

In das Verzeichniss des GeonamesDumpImporter Projektes wechseln und den Importer (Java) wie folgt ausführen:

```
export JAVA_HOME=/usr/lib/jvm/java-8-oracle
export JRE_HOME=/usr/lib/jvm/java-8-oracle/jre
mvn exec:exec -Dexec.executable="java"
```

Nach dem Durchlauf (Stand Geonames Dump 15. 4. 2016) wurden 47'409 Graphen importiert.

Anzahl Statements im Repository anzeigen: `repositories/geonames/size: 772'712 Triples / Statements`

Weitere Links / Infos

- Dokumentation Geonames Ontologie: <http://www.geonames.org/ontology/documentation.html>

