

Semi-automatisches Lernen von Inter-Links mit Geolink

Kevin Shrestha, Martin Stoffers

07. Juli 2015

Gliederung

Einführung in GeoLink

- Aufgabenstellung

- Aufgaben

Technische Umsetzung

- Softwarebestandteile

- Demo

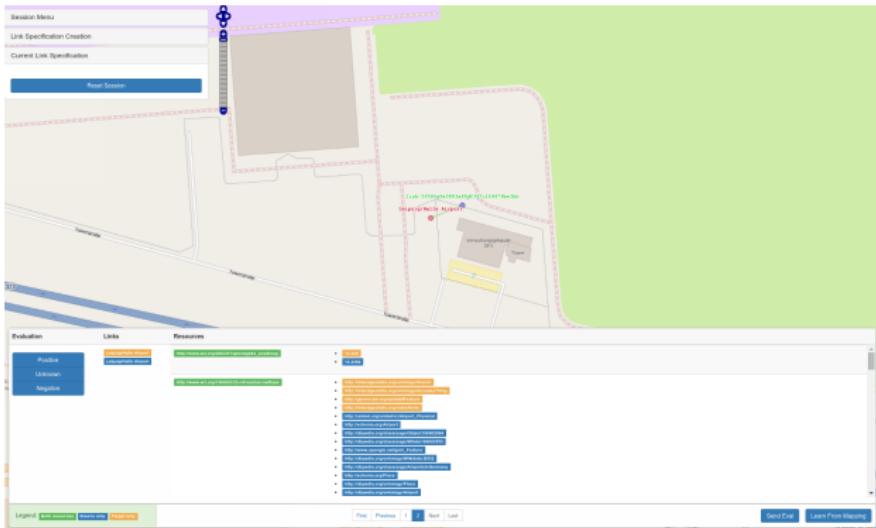
- Auswertung

Ausblick

Einführung in GeoLink

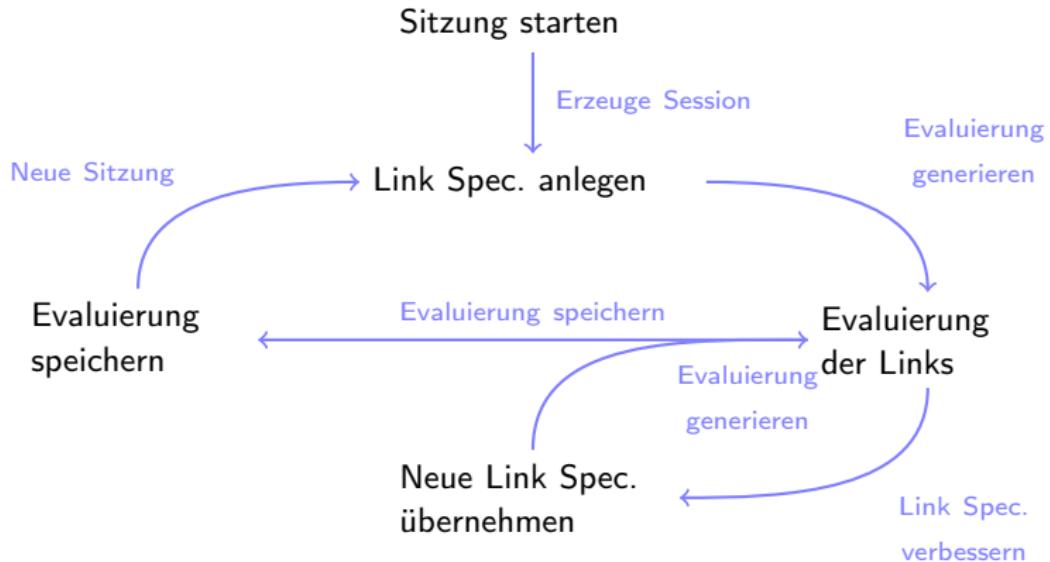
Aufgabenstellung

- Erstellen einer Webanwendung zur Evaluierung von Interlinks
- Verbesserung der Link Spezifikation mittels eines Supervised Learners
- Speichern der erzeugten Evaluierung



Einführung in GeoLink

Aufgabenstellung



Einführung in GeoLink

Aufgaben

1. Generierung einer Link Spezifikation im Browser
2. Evaluierung der erzeugten Interlinks durch den Benutzer
3. Verbesserung der Link Spezifikation durch einen Supervised Learner
4. Speichern der endgültigen Evaluierung in einem Triplestore
5. Möglichkeit des User-Managements soll beachtet werden
6. XML-Import/Export der Link Spezifikation

Technische Umsetzung

Softwarebestandteile

Webserver

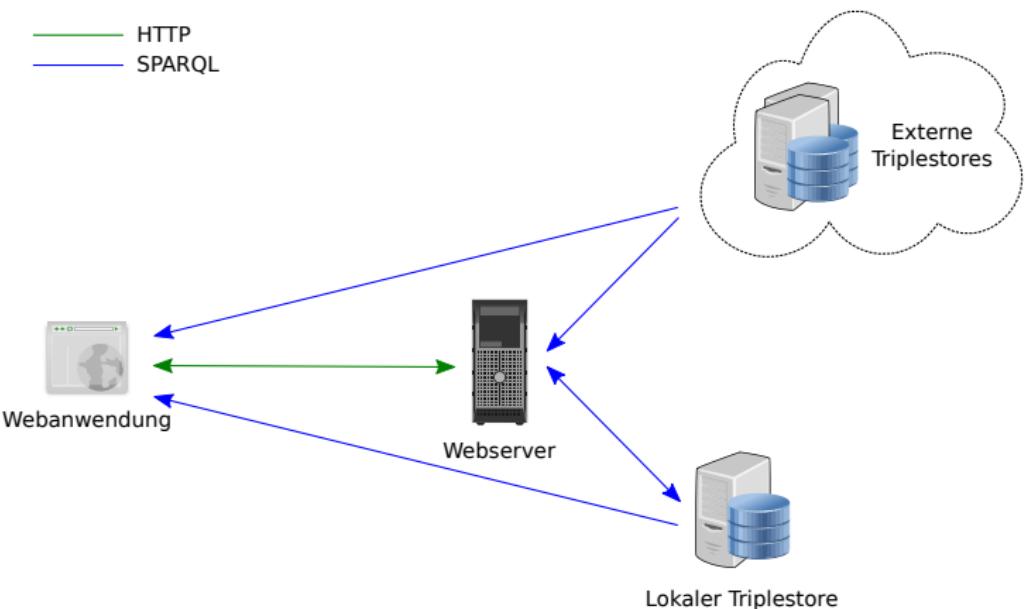
- Jetty Server mit Spring Framework
- Jena-Framework
- LIMES
- Virtuoso Triplestore

Webanwendung

- HTML5
- Angular.js
- Angular Bootstrap (GUI)
- Jassa-Framework

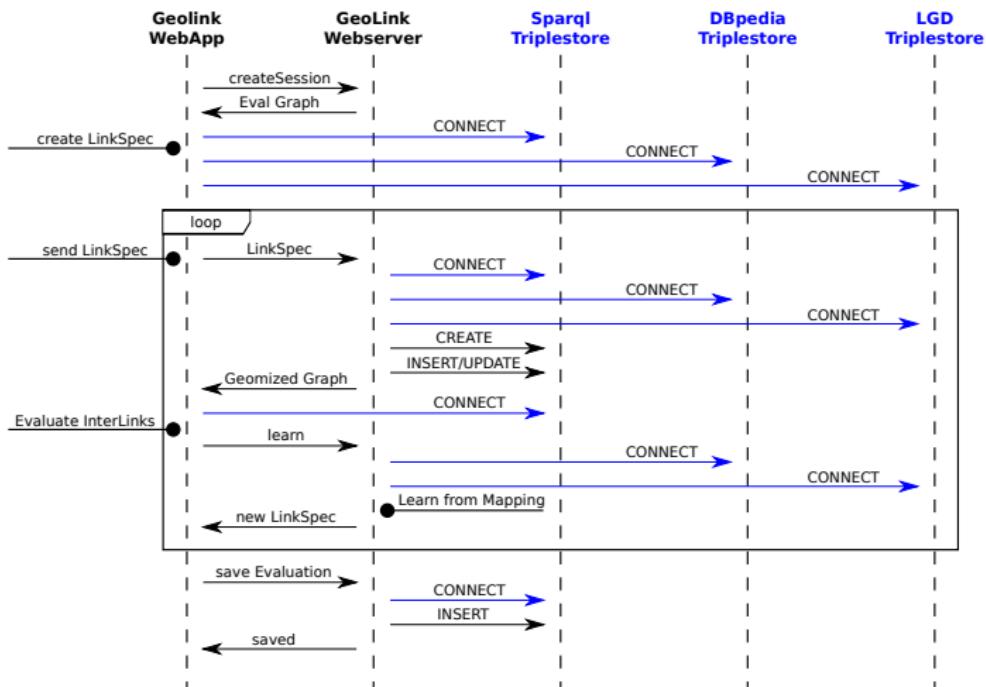
Technische Umsetzung

Systemübersicht



Technische Umsetzung

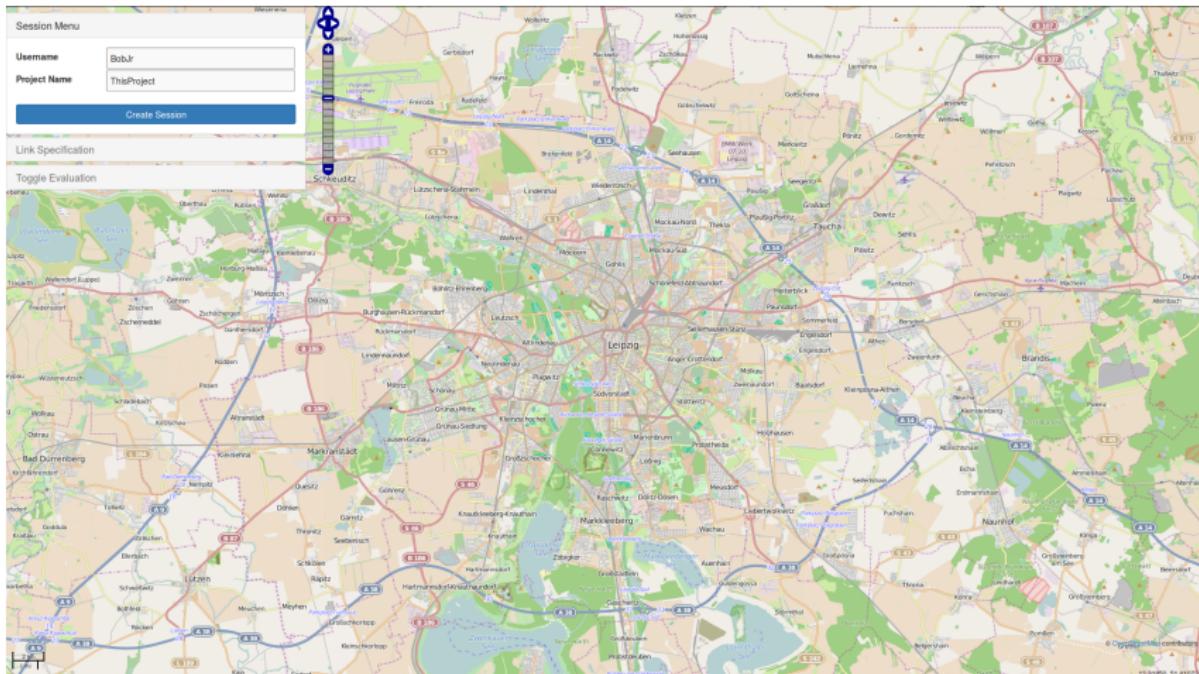
Kommunikationsablauf



Technische Umsetzung



Technische Umsetzung



Technische Umsetzung

Session Menu

Link Specification

Fill with Source Template -> FR DBP Min

Endpoint: http://fastreboot.de:8890/sparql
Graph: http://fastreboot.de/tbpitest
Restrictions: ?x a [http://dbpedia.org/ontology/Airport](#)
Var: ?x
Properties: rdfs:label

Fill with Target Template -> FR LGD Min

Endpoint: http://fastreboot.de:8890/sparql
Graph: http://fastreboot.de/tgttest
Restrictions: ?y a [http://linkedgodsdata.org/ontology/Airport](#)
Var: ?y
Properties: rdfs:label

Metric: trigrams(x.rdfs:label, y/rdfs:label)
Threshold: 0.95

Show Data
Send LinkSpec

Evaluation graph

Toggle Evaluation

Technische Umsetzung



Technische Umsetzung

The screenshot shows a geospatial application interface with the following components:

- Top Left:** Session Menu, Link Specification, Toggle Evaluation.
- Top Right:** Evaluation graph, Geomized graph.
- Map Area:** Aerial view of a highway interchange with roads labeled A 14, A 50, and A 51. A purple polygon labeled "Aeron 3" covers a specific area. A legend indicates "Leistungsfähige Autobahn" (Highway) and "Ruhigfahrende Autobahn" (Quiet Highway). A red dashed line highlights a specific route segment.
- Evaluation Panel:** Shows Positive, Unknown, and Negative buttons. Below them is a list of URLs corresponding to the highlighted area:
 - <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
 - <http://linkeddata.org/sparql#Product>
 - <http://linkeddata.org/sparql#Node>
 - <http://linkeddata.org/sparql#Thing>
 - <http://linkeddata.org/sparql#UnknownThing>
 - <http://www.w3.org/2000/01/rdf-turtle>
 - <http://schema.org/Place>
 - <http://linkeddata.org/sparql#WebLink>
 - <http://www.w3.org/2003/01/geo/gml/Location>
 - <http://www.opengis.net/gml/Feature>
 - <http://schema.org/Feature>
 - http://linkeddata.org/sparql#Airport_Passenger
 - <http://www.w3.org/2003/01/geo/gml/pointProperty/Thing>
 - <http://dbpedia.org/class/yago/Airport10267903>
 - <http://dbpedia.org/class/yago/Airport2242227>
 - <http://dbpedia.org/class/yago/Airport10261819>
- Bottom Left:** Legend: Both resource, Source only, Target only.
- Bottom Center:** Navigation buttons: Find, Previous, Next, Last. Page number 81.
- Bottom Right:** Save, Learn From Mapping, ID 24115, 51.40417.

Technische Umsetzung

Session Menu

Link Specification

Fill with Source Template → FR DBP Min

Endpoint: <http://fastreboot.de:8890/sparql>

Graph: <http://fastreboot.de/dbpediaTest>

Restrictions: ?x a <http://dbpedia.org/ontology/Airports>

Var: ?x

Properties: rdfs:label

Fill with Target Template → FR LGO Min

Endpoint: <http://fastreboot.de:8890/sparql>

Graph: <http://fastreboot.de/lgdTest>

Restrictions: ?y a <http://linkedgodata.org/ontology/Airport>

Var: ?y

Properties: rdfs:label

Metric: $\text{trigrams}(x.\text{rdfs:label}, y.\text{rdfs:label})$

Threshold: 0.95

Show Data

Send LinkSpec

Reject New LinkSpec

Accept New LinkSpec

Toggle Evaluation

Evaluation graph
Geomized graph

Technische Umsetzung

Session Menu

Link Specification

Fill with Source Template -> FR DBP Min

Endpoint: http://fastreboot.de:8890/sparql
Graph: http://fastreboot.de/tbpitest
Restrictions: ?x a [http://dbpedia.org/ontology/Airport](#)
Var: ?x
Properties: rdfs:label

Fill with Target Template -> FR LGD Min

Endpoint: http://fastreboot.de:8890/sparql
Graph: http://fastreboot.de/tgttest
Restrictions: ?y a [http://linkedgodata.org/ontology/Airport](#)
Var: ?y
Properties: rdfs:label

Metric: trigrams(x.rdfs:label, y/rdfs:label)
Threshold: 0.95

Show Data
Send LinkSpec

Evaluation graph

Evaluation graph

124101, 51.3789

Toggle Evaluation

15 of 19

Auswertung

Verwendete Datensätze

- Auswahl der Datensätze über Polygon
- Teilgraphen aus DBpedia und LinkedGeoData (LGD)
- Messung der Zeit für das Generieren der InterLinks
- Betrachtung des RAM-Verbrauchs

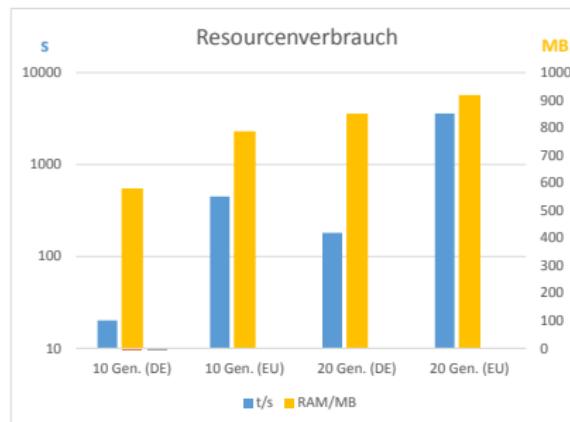
Datensätze	DBpedia	LGD
Airports DE	2251	1314
Airports EU	19498	16196

Tabelle : Extrahierte Ressourcen

Auswertung

Analyse des Interlinking Prozesses

- 22-fache Rechenzeit bei Verzehnfachung der Datenvolumen
- 9-fache Rechenzeit bei Verdopplung der Generationen
- RAM-Verbrauch stark von Anzahl der ausgewählten Ressourcen abhängig



Ausblick

Bisher nicht umgesetzt

- XML-Export/Import
- User-Management

Erweiterungsmöglichkeiten

- Optimierung auf größere Datenmengen
- Performance der Webanwendung verbessern
- Versionierung der LinkSpezifikationen über den Triplestore
- Integration von Facete 2

