

Handbuch
LIMES via GUI
Installation und Bedienung

Projektverantwortlicher: Sascha Hahne

18. Mai 2015

Inhaltsverzeichnis

1	Einleitung	3
2	Installationsanweisungen	3
3	Bedienung	3
3.1	Erstellung einer neuen Konfiguration	3
3.1.1	Anlegen von Endpoint-Properties	3
3.2	Erstellung einer Metrik	4
3.3	Start des Mapping Prozesses	10
3.4	Import und Export	10
3.5	Automatische Erstellung von Metriken	12

1 Einleitung

Links zwischen Dateninstanzen sind sowohl im Semantic Web von zentraler Bedeutung als auch für verschiedenste Anwendungen wie Datenintegration, gebündelte Abfragen und Knowledge Retrieval. Mit der Entdeckung dieser Links beschäftigt sich Link Discovery(LD).

Es gibt zwei fundamentale Probleme bei der Suche nach Links zwischen Datensätzen.

Zum einen sorgt die naive Lösung, der Vergleich jeder Instanz einer Wissensbasis mit jeder Instanz der anderen Wissensbasis, für eine quadratische Zeitkomplexität. Zum anderen muss die Qualität der Links garantiert werden.

Um das erste Problem zu lösen, wurden zeiteffiziente Algorithmen, wie SILK und LIMES entwickelt um die Anzahl der Vergleiche zu reduzieren. Für die zweite Problemstellung sind außerdem 'supervised' und 'unsupervised' machine learning Ansätze entwickelt wurden.

Diese GUI bietet dem Anwender eine simple Verwendung des LIMES-Frameworks.

Im folgendem wird erläutert, wie man mit diesem Programm eine Link-Discovery Anfrage gestalten und starten kann.

2 Installationsanweisungen

Die GUI wird durch Ausführung der Link-Discovery.jar über die Kommandozeile folgendermaßen gestartet:

```
java -jar path-to/link-discovery.jar
```

Es ist zu beachten, dass Link-Discovery JAVA 8 benötigt, alle anderen Abhängigkeiten sind direkt in der Jar-Datei bereitgestellt.

3 Bedienung

3.1 Erstellung einer neuen Konfiguration

3.1.1 Anlegen von Endpoint-Properties

Die Erstellung einer neuen Konfiguration erfolgt über die Menu-Leiste durch das Klicken auf:

→

Danach öffnet sich ein Dialog(sh. 1). In diesem kann man entweder die URLs zu den SPARQL-Endpoints im Internet angeben oder den Pfad zu lokalen Endpoints. Optional sind die Felder *ID/namespace*, *Graph* und *Page size* zu befüllen, wenn sie leer bleiben, werden Default Werte verwendet. Nach dem Klick auf *Next* werden die Endpoints geladen und ausgewertet. Dieser

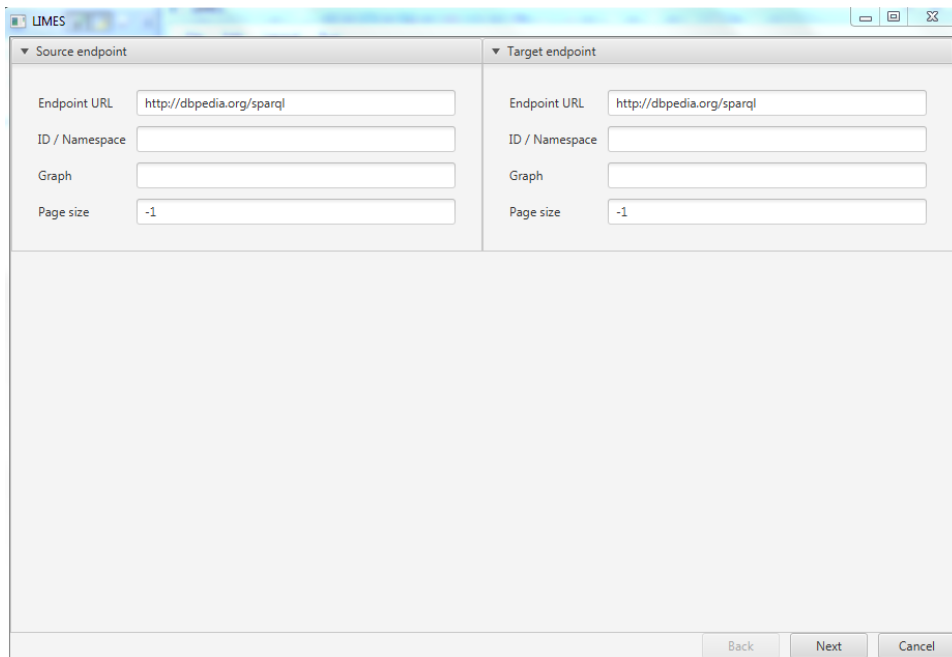


Abbildung 1: Dialog Eingabe Endpoint URL

Prozess kann eine gewisse Zeit in Anspruch nehmen. Nachdem der Prozess abgeschlossen ist, öffnet sich der Classmatching Dialog 2. Hier kann man nun die zu vergleichenden Klassen durch einfaches Klicken auswählen. Nachdem die richtigen Klassen gefunden und markiert wurden, klickt man wiederum auf den Button *Next*.

Nun beginnt der Property Matching Dialog 3. Zu Beginn ist dieser leer. Durch Klicken auf *Add* 4 kann man paarweise neue Properties hinzufügen. Es ist natürlich auch möglich durch markieren einer Property-Zeile und dem Klicken auf *Remove* vorher ausgewählte Property-Zeilen direkt wieder zu löschen. Wenn der User mit den gewählten Properties zufrieden ist, kann er mit dem Button *Finish* den Prozess abschließen.

3.2 Erstellung einer Metrik

Nachdem man Source- und Target-Endpoints erstellt 3.1.1 oder eine Linkspezifikation importiert 3.4 hat, kann man mit dem Erstellen der Abfragemetrik beginnen. Auf der linken Seite der Oberfläche 5 stehen dafür die Tools zur Verfügung, welche durch einfaches Klicken der Modellierungs-Oberfläche hinzugefügt werden können. Bei Source-Properties findet man die eingeladenen Endpoint Properties der Source und bei Target die entsprechend festgelegten Properties des Targets. Unter Metrics werden die verfügbaren

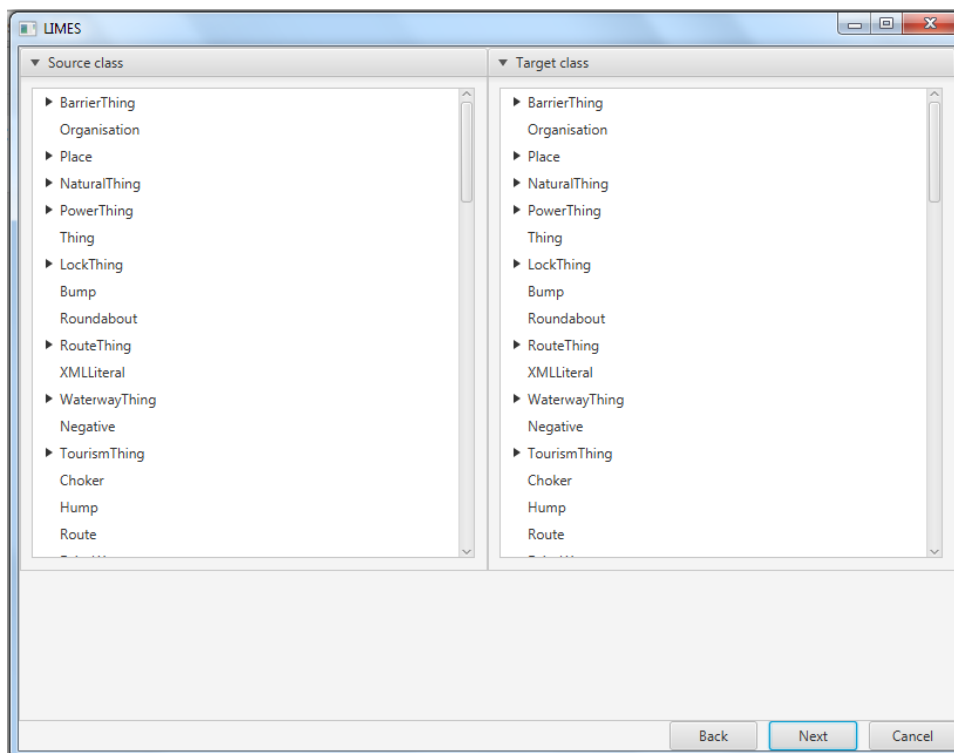


Abbildung 2: Dialog Class Matching

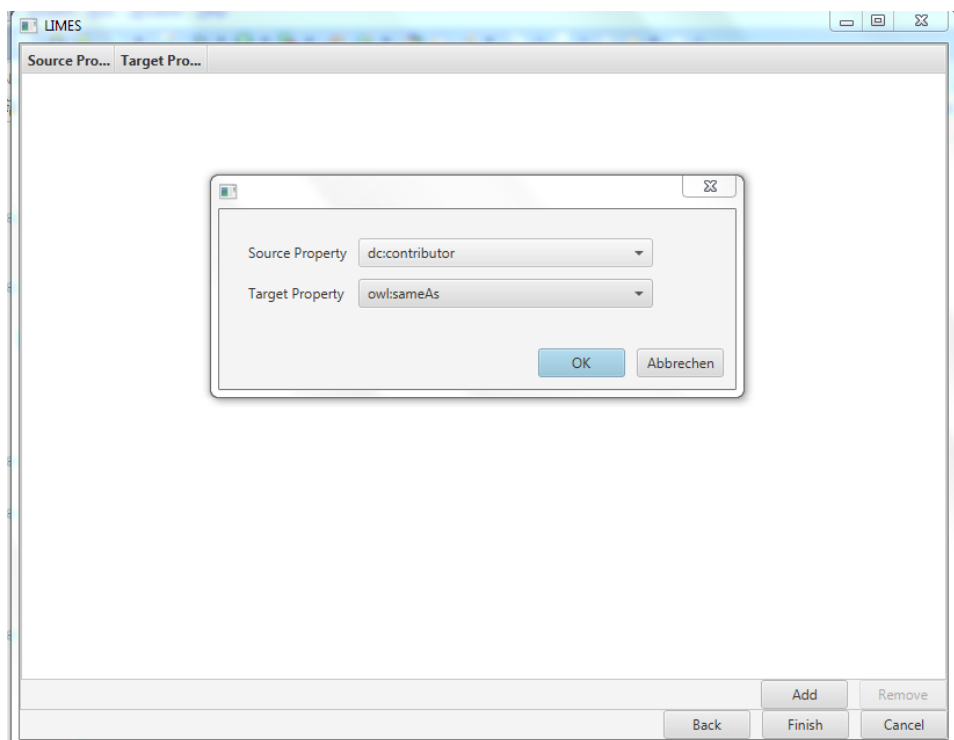


Abbildung 4: Dialog Hinzufügen von Property-Match

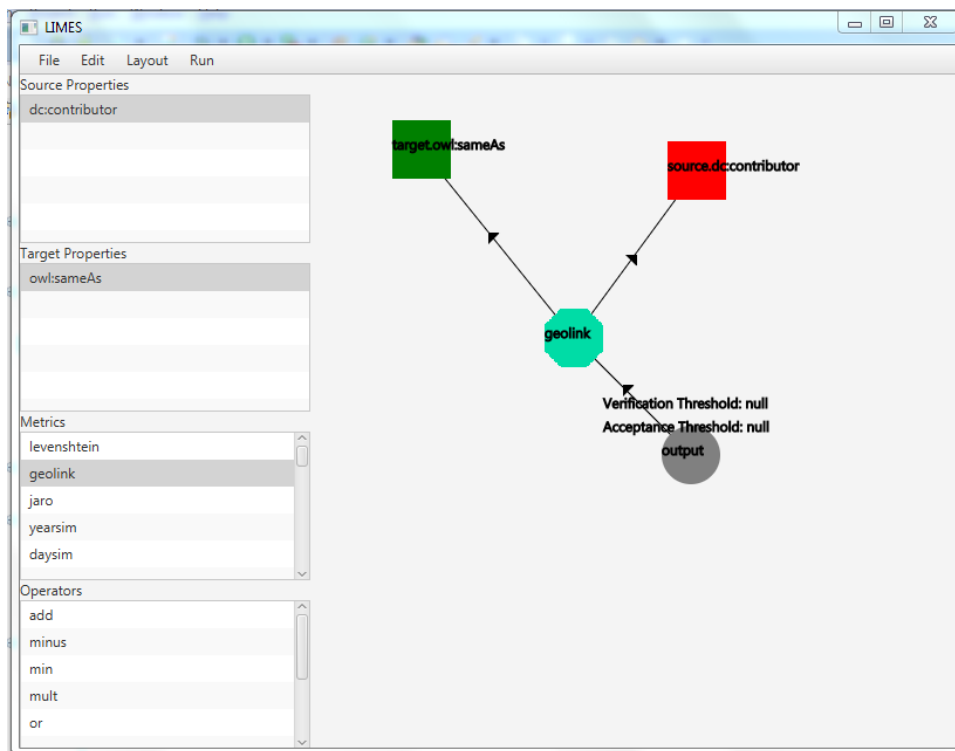


Abbildung 5: Modellierung der Metrik

Metriken zum Vergleich von zwei Properties gelistet. Aktuell sind dies:

- cosine
- euclidean
- jaccard
- levenshtein
- overlap
- trigrams
- jaro
- hausdorff
- orthodromic
- frchet
- qgramsimilarity
- exactmatch
- symmetrichausdorff
- frechet
- geosummn
- surjection
- fairsurjection
- geolink
- geomean
- datesim
- daysim
- yearsim
- geomn
- geomx
- geoavg

In der Box Operators befinden sich logische Operatoren, die die Ergebnisse zweier Metriken, Operatoren oder eines Operators und einer Metrik kombinieren können. Hierbei stehen dem User aktuell folgende zur Verfügung:

- add
- and
- diff
- max
- min
- minus
- mult
- or
- xor

Durch klicken und ziehen auf die einzelnen Elemente auf der Modellierungs-Oberfläche kann die Position dieser verändert werden. Beim Rechtsklick auf ein Element öffnet sich ein Kontextmenu, hier kann eine Node komplett gelöscht werden, wenn diese nicht mehr benötigt wird oder mit einem validem Kind verbunden werden.

3.3 Start des Mapping Prozesses

Wenn der User die Modellierung der Metrik 3.2 abgeschlossen hat kann er nun einen Mapping-Prozess starten. Dies geschieht über klicken in der Menu-Leiste auf die Punkte:

→

Ein Dialog öffnet sich und durch klicken auf *Start* wird die LIMEs Anfrage ausgeführt. Wenn diese fertig ist, erscheint ein Fenster mit den entdeckten Links der Abfrage 6. Durch Auswählen der einzelnen Zeilen werden auch die betreffenden Properties angezeigt. Um die Ergebnisse der Abfrage zu speichern, klickt der User in der Menu-Leiste auf:

→

3.4 Import und Export

Es ist möglich schon vorhandene LIMEs Abfragen in der Form von Link-specs in die GUI zu importieren, um sie im nachhinein zu verändern oder noch einmal auszuführen. Dafür klickt der User in der Menu-Leiste auf:

→

File			
Property	Value	Property	Value
rdfs:label	inge walthemate	name	inge walthemate
Source URI	Target URI	value	
http://dbpedia.org/resource/Jasmin_Schornberg	http://dbpedia.org/resource/Jasmin_Schornberg	1.0	
http://dbpedia.org/resource/Inge_Walthemate	http://dbpedia.org/resource/Inge_Walthemate	1.0	
http://dbpedia.org/resource/Ralf_Wienand	http://dbpedia.org/resource/Ralf_Wienand	1.0	
http://dbpedia.org/resource/Waltraud_Schale	http://dbpedia.org/resource/Waltraud_Schale	1.0	
http://dbpedia.org/resource/Walter_Horn_(canoeist)	http://dbpedia.org/resource/Walter_Horn_(canoeist)	1.0	
http://dbpedia.org/resource/Susanne_Erbers	http://dbpedia.org/resource/Susanne_Erbers	1.0	
http://dbpedia.org/resource/Petra_Setzkorn	http://dbpedia.org/resource/Petra_Setzkorn	1.0	
http://dbpedia.org/resource/Hans_Rein	http://dbpedia.org/resource/Hans_Rein	1.0	
http://dbpedia.org/resource/Wolfgang_B%C3%BCchner	http://dbpedia.org/resource/Wolfgang_B%C3%BCchner	1.0	
http://dbpedia.org/resource/Michael_Scheuer_(kayaker)	http://dbpedia.org/resource/Michael_Scheuer_(kayaker)	1.0	
http://dbpedia.org/resource/Robert_Nuck	http://dbpedia.org/resource/Robert_Nuck	1.0	

Abbildung 6: Anzeigen der Ergebnisse

Danach öffnet sich ein Dialog in dem der User die betreffende Konfiguration auswählen und speichern kann.

Zum Export einer selbst erstellten oder veränderten Konfiguration klickt der User auf:

→

3.5 Automatische Erstellung von Metriken

Die GUI und LIMES implementieren verschiedenste Möglichkeiten eine Metrik automatisch zu erstellen. Nach erfolgreichem Import 3.4 oder Anlegung von Endpoints 3.1.1 ist es dem User möglich über:

Run → Self Configuration

Verschiedene Machine Learning Algorithmen zu starten oder über:

Run → Active Learning

die Implementierung von Active Learning zu nutzen um LIMES selbständig die Metrik erzeugen zu lassen.