



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG

NEUPHILOLOGISCHE FAKULTÄT

- INSTITUT FÜR COMPUTERLINGUISTIK -

Semantic Parsing as Monolingual Machine Translation

– Anleitung zur Korpuserweiterung –

Autor:

Nadja HEINZEN

E-Mail: heinzen@cl.uni-heidelberg.de

Betreuer:

Prof. Dr. Stefan RIEZLER

30. Juli 2016

1 grundsätzliche Vorgehensweise

Zur Korpuserweiterung geht man wie folgt vor:

Zuerst sucht man sich ein Objekt, welches man hinzufügen möchte, und gibt die dazugehörige Frage in Carolins Interface NLMaps¹ ein. Erzielt das Interface einen Treffer, so sucht man ein neues Objekt. Denn unsere Annahme war hier, dass wenn das System die Objekte schon kennt, sei es, weil die Frage so in Carolins Korpus² steht oder weil das System die MRL neu kombiniert hat, dann brauchen wir dieses Objekt nicht zum Korpus hinzufügen. Erzielt das System keinen Treffer oder im Falle des Fragentyps least den default-Wert 0 (dieses Ergebnis wird oft erzielt, die dazugehörige MRL hat aber oftmals gar nichts mit der geforderten Frage zu tun), dann kann man eine neue MRL erstellen. Dazu sucht man sich eine ähnliche, funktionierende MRL und ändert die relevanten **Eigenschaften** ab.

Beispiel:

How many parkings are there in Mannheim?

```
query(area(keyval('name','<nom>3600062691</nom>')),nwr(keyval('name','parking'))  
,qtype(count))
```

Diese Eigenschaften findet man mit Hilfe von OpenStreetMap³. Wenn man sich auf der OpenStreetMap-Seite anmeldet, kann man sich die Eigenschaften von Objekten anzeigen lassen. Diese kann man dann in die MRL eintragen.

Auch muss man die **Orts-ID** ändern. Dazu kann man in dem NLMaps Interface eine Frage stellen, die in den meisten Fällen funktioniert. Im Rahmen des Projektes wurde die Frage: “where is a bank” verwendet, da solche Objekte am häufigsten gefunden werden. Dann gibt man in dem Feld **“Reference point”** den gewünschten Ort ein und lässt das System suchen. Hat das System die Objekte gefunden, erscheint rechts ein Feld: **“Was that helpful?”** mit der Option **“more”**. Unter dieser Option kann man sich dann die MRL anzeigen lassen, die die Orts-ID beinhaltet.

Städte wurden mit Hilfe einer Großstädteliste⁴ ausgewählt, da man davon ausgehen kann, dass größere Orte auch mehr Systemanfragen erhalten. Ebenso wurden die Sehenswürdigkeiten für den Fragetyp dist mit Hilfe einer Liste von beliebtesten Sehenswürdigkeiten⁵ ausgewählt.

¹<http://nlmaps.cl.uni-heidelberg.de/index.html>.

²A Corpus and Semantic Parser for Multilingual Natural Language Querying of OpenStreetMap; Haas; Carolin and Riezler; Stefan; 2016.

³<http://www.openstreetmap.de/karte.html>.

⁴<http://www.localpedia.de/deutsche-grossstaedte>.

⁵<http://www.germany.travel/de/staedte-kultur/top-100/germany-travel-attractions.html?page=6>.