

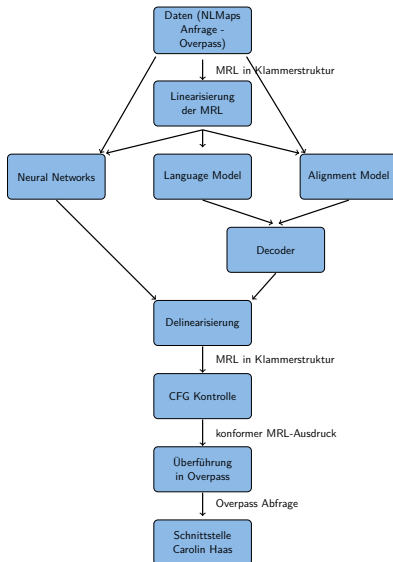
Semantic Parsing as Monolingual Machine Translation

NLMaps Gruppe

31. Juli 2016

- 1 Korpuserweiterung
- 2 (De-) Linearisierung
- 3 Baseline System
- 4 Neural Machine Translation
- 5 Evaluation

Projektverlauf



Korpuserweiterung

Korpuserweiterung - Überblick

- **Gesamterweiterung** ~ 700 Queries
- **5 Fragetypen** dist, latlong, findkey, count & least
- **zu jedem Fragentyp 2 Fragearten** 1x lange Frage, 1x kurze Frage
- **Orte** ~ 50 neue Orte
- **Topics** ~ 80 Themen wie Schwimmbad, Optiker, Carsharing,...
- **Beschränkung** auf Deutschland
- **Auswahl der Orte / Sehenswürdigkeiten** durch Top 100 Listen

Korpuserweiterung - Vorgehensweise

- ① Objekt / Frage aussuchen, welche(s) hinzugefügt werden soll
- ② in NLMaps Frage eingeben
 - ① Frage erzielt eine Antwort -> weiteres Objekt testen
 - ② Frage erzielt keine Antwort -> Juhu :)
- ③ **Eigenschaften** des Objektes in OpenStreetMap ansehen und in die MRL eintragen
 - `query(area(keyval('name','<nom>3600062691</nom>')) ,nwr(keyval('name','parking')) ,qtype(count))`
- ④ Orts-ID suchen, indem man in NLMaps den gewünschten Ort als Referenzpunkt angibt und eine Frage testet, die immer funktioniert -> unter "more" findet man die Orts-ID



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

NLmaps

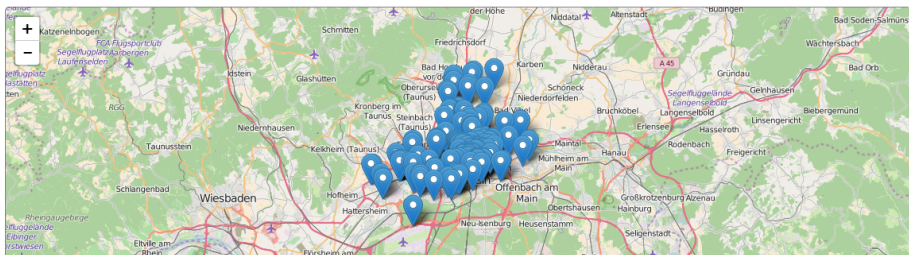
Query NLmaps

Reference point (optional):

OR

Your objects of interest are marked below

Was that helpful?





UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

where is a bank

Reference point (optional):

Frank

Submit

at helpful?

Yes

No

More

Feedback



Could you provide us with some more feedback?

Just fill out as much as you want, every question you answer will help us to improve!

Did we get the location right?

Yes

No

[name=<nom>3600062400</nom>]

Did we select the correct question type?

(Options are: count, findkey, latlong, least, dist)

Yes

No

latlong

Did we select the correct OSM tags?

Yes

No

[amenity=bank]

Can you correct the Overpass query?

Yes

No

area(3600062400)->.a;(node(area.a)[amenity=bank].way(area.a)[amen

Can you correct the MRL?

Yes

No

query(area(keyval('name','<nom>3600062400</nom>')),nwr(keyval('am

Submit



OpenStreetMap

Bearbeiten



Chronik

Export

GPS

Suchen

Wo bin ich?

Los

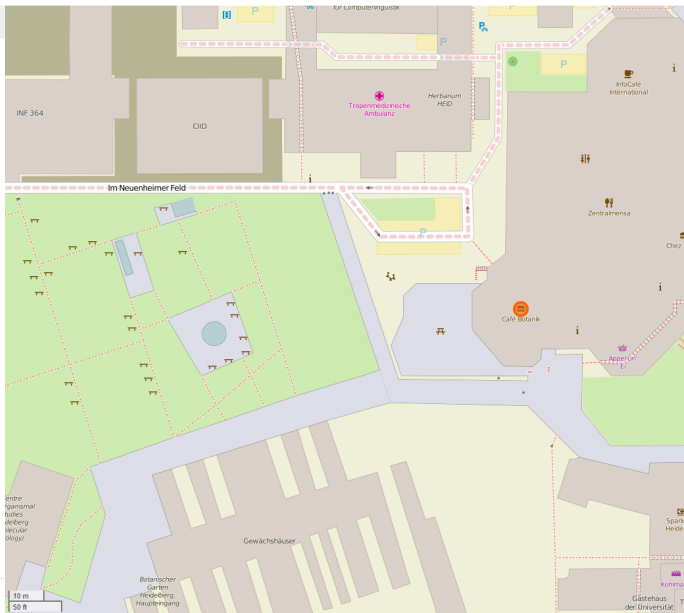
**Knoten: Café Botanik (303195887)**

Cafe Botanic was there two times. Deleted the less detailed one

Bearbeitet vor 2 Monate von [Mrlindum](#)
Version #14 · Änderungssatz #39476136
Standort: [49.4153134, 8.6703781](#)

Attribute

access	no
addr.city	Heidelberg
addr.housenumber	304
addr.postcode	69120
addr.street	Im Neuenheimer Feld
amenity	fast_food
delivery	no
drink.clubmate	retail
drive_through	no
internet_access	wlan
name	Café Botanik
opening_hours	Mo-Th 08:00-24:00; Fr 08:00-23:00
phone	+49 6221 54 26 44
smoking	no
website	http://www.studentenwerk-heidelberg.de
wheelchair	yes

[XML herunterladen](#) · [Chronik anzeigen](#)


Korpuserweiterung - Fragetyp latlong

- sucht nach Objekten und zeigt die **GPS Punkte** auf der Karte an
- **126 Queries**
- **10 Orte** Frankfurt, Aachen, Trier,...
- **18 Topics** Schwimmbad, Frisör, Schule,...
- **Beispielfragen**
 - ① where is a bus stop in Munich?
 - ② bus stop in Munich
- **zugehörige MRL**

```
query(area(keyval('name',' < nom > 3600062428 < /nom >'  
')), nwr(keyval('highway',' bus_stop')), qtype(latlong))
```

Korpuserweiterung - Fragetyp count

- liefert die **Anzahl** der erfragten Objekte
- **126 Queries**
- **9 Orte** Nürnberg, Leipzig, Mannheim,...
- **19 Topics** Optiker, Zoo, Parkplatz,...
- **Beispielfragen**
 - ① How many parkings are there in Mannheim?
 - ② number of parkings in Mannheim
- **zugehörige MRL**

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

Korpuserweiterung - Fragetyp least

- liefert einen **boolschen Wert** und fragt danach, ob ein Objekt existiert
- **192 Queries**
- **12 Orte** Wuppertal, Bochum, Bielefeld,...
- **16 Topics** Reitstall, Tierarzt, Carsharing,...
- einziger Fragetyp, der **Fuzzy Language** beinhaltet (Himmelsrichtungen)

Korpuserweiterung - Fragetyp least

• Beispielfragen

- ① Is there a bottle bank in the east of Wiesbaden?
- ② does a bottle bank in the east of Wiesbaden exist?
- ③ Is there a bottle bank in Wiesbaden?
- ④ does a bottle bank exist in Wiesbaden?

• zugehörige MRLs

```
query(east(area(keyval('name',' < nom > 3600062496 < /nom >'
)), nwr(keyval('amenity',' recycling'), keyval('recycling :
glass',' yes'))), qtype(least(topx(1))))
```

```
query(area(keyval('name',' < nom > 3600062496 < /nom >'
)), nwr(keyval('amenity',' recycling'), keyval('recycling :
glass',' yes'))), qtype(least(topx(1))))
```

Korpuserweiterung - Fragetyp dist

- berechnet die **Distanz zwischen 2 Objekten**
- **122 Queries**
- **schwierigster Fragetyp**, da zwei MRLs miteinander verknüpft werden (lange MRLs)
- **Beispielfragen**
 - ① How far away is Messeturm from Messe Frankfurt in Frankfurt am Main?
 - ② distance between Messeturm and Messe Frankfurt in Frankfurt am Main?
- **zugehörige MRL**

```
dist(query(area(keyval('name',' < nom > 3600062400 < /nom >'))
, nwr(keyval('name',' Messeturm')), qtype(latlong)), query(area(keyval
('name',' < nom > 3600062400 < /nom >')
)), nwr(keyval('name',' MesseFrankfurt')), qtype(latlong)))
```

Korpuserweiterung - Fragetyp findkey

- gibt den **Value** zu einem Key aus
- **141 Queries**
- **Beispielfragen**
 - ① What is the short name of Miniatur Wunderland in Hamburg?
 - ② short name of Miniatur Wunderland in Hamburg
- **zugehörige MRL**

```
query(area(keyval('name',' < nom > 3602618040 < /nom >'
)), nwr(keyval('name',' MiniaturWunderland')), qtype
(findkey('short_name')))
```

(De-) Linearisierung

Linearisierung

- **MRL** *query(area(keyval('name',' < nom > 3600062422 < /nom > ')), nwr(keyval('office',' newspaper')), qtype(count))*
- **Linearisieren** query@3 area@1 keyval@2 name@0 fjbbbjdgdd nwr@1 keyval@2 office@0 newspaper@s qtype@1 count@0
- **Probleme** *query(area(keyval('name',' Heidelberg'), keyval('de : place',' city')), nwr(keyval('tower : type',' communication'), keyval('operator',' Südwestrundfunk(SWR)')), qtype(least(topx(1))))*
- **Weiterleiten SMT/NMT**

Stemming

- **NL** number of zoos in Mannheim
- **Gestemmt** number of zoo in Mannheim

Delinearisierung

- **LMRL** query@3 area@1 keyval@2 name@0 fjbbbjdgdd nwr@1 keyval@2 office@0 newspaper@s qtype@1 count@0
- **Delinearisieren** *query(area(keyval('name',' < nom > 3600062422 < / nom >')), nwr(keyval('office',' newspaper')), qtype(count))*
- **Weiterleiten zur Evaluation**

Baseline System

Baseline: Statistical Machine Translation

Moses Decoder

Implementiertes Verfahren zur maschinellen Übersetzung.

- Trainingsdaten
- Testdaten
- Tuning Daten zur Verfeinerung der Gewichte

Baseline: Moses Decoder

Language Model Training

Das language model (LM) dient zur Sprachoptimierung, es wird mithilfe der Zielsprache erstellt (hier: MRL)

- Language Model: Trigrammmodell
- Alignment (IBM4): GIZA++

Neural Machine Translation

Übersetzung mit Seq2Seq

Das Vorbild

Grammar as a Foreign Language

Paper einer Google Arbeitsgruppe

- 'Syntactic parsing as machine translation' erreicht F1 score $> 90\%$
- Wesentlicher Beitrag dazu leistet der **attention mechanism**

Übersetzung mit Seq2Seq

- **LSTM**: RNNs, die long-term dependencies gut lernen
- Der Seq2Seq Decoder generiert den Output Wort für Wort
- Der **Attention mechanism** erlaubt dabei die Suche nach bestimmtem Context in der Input-Sequenz

Übersetzung mit Seq2Seq

- Wir nutzen **TensorFlow** (C++ low-level, runs on CPU or GPU) für eine effiziente Implementierung des LSTM+A Modells
- Translate.py Implementierung aus Tensorflow
- Anpassungen auf unsere Daten
- Ein training step dauert $\sim 3s$ mit `batch_size=64`

Backtracking

- nicht alle Übersetzungen entsprechen den syntaktischen Anforderungen
- Decoding/Backtracking hinzugefügt
- Im decoding step erzeugen wir mit beam search mehrere MRLs, die mit CFG auf Wohlgeformtheit überprüft werden

nmt Experimente

Erster Versuch

- **Beamsize** 5
- **learning Rate** 0.5/0.99 decay
- **Buckets** (5, 10), (10, 15), (20, 25), (40, 50)

Bester Versuch

- **Beamsize** 10
- **learning Rate** 0.5/0.99 decay
- **Buckets** (5, 15), (10, 20), (20, 30), (40, 50)

Evaluation

Grundlage

- keine MRL-Ähnlichkeit
- Ergebnisse der System- und Gold-Anfragen bei der Datenbank werden verglichen

Grundlage

- kein Ergebnis kann erhalten werden (KE)
- das Ergebnis unterscheidet sich von dem Gold-Ergebnis (FE)
- das Ergebnis ist richtig (RE)

Metriken

Recall: Anteil richtiger Ergebnisse von allen

$$R = \frac{RE}{RE + FE + KE}$$

Precision: Anteil richtiger Ergebnisse von den korrekten Anfragen

$$Pr = \frac{RE}{RE + FE}$$

F1-Score: Das harmonische Mittel

$$F1 = 2 * \frac{Pr * R}{Pr + R}$$

Ergebnisse

System	Recall	Precision	F1
Baseline	0.36	0.88	0.51
NMT	0.55	0.70	0.61

Anmerkung: Bei jedem neuen Trainieren wird ein etwas anderes Modell erzeugt, deswegen können sich die Ergebnisse auch variieren. Hier ist die Evaluation eines Durchlaufs dargestellt.

Ergebnisse: Fragetypen

System	count	dist	findkey	latlong	least	insgesamt
SMT(übersetzt)	6	9	2	8	2	27
SMT(davon richtig)	6	6	2	8	2	24
NMT(übersetzt)	10	3	13	9	17	52
NMT(davon richtig)	9	0	6	6	15	36
Daten	11	13	15	9	18	66

Ergebnisse: Fragetypen

System	count	dist	findkey	latlong	least	insgesamt
SMT(übersetzt)	6	9	2	8	2	27
SMT(davon richtig)	6	6	2	8	2	24
NMT(übersetzt)	10	3	13	9	17	52
NMT(davon richtig)	9	0	6	6	15	36
Daten	11	13	15	9	18	66

Ergebnisse: Fragetypen

System	count	dist	findkey	latlong	least	insgesamt
SMT(übersetzt)	6	9	2	8	2	27
SMT(davon richtig)	6	6	2	8	2	24
NMT(übersetzt)	10	3	13	9	17	52
NMT(davon richtig)	9	0	6	6	15	36
Daten	11	13	15	9	18	66

Ergebnisse: Fragetypen in %

System	count	dist	findkey	latlong	least	insgesamt
SMT(übersetzt)	.55	.69	.13	.30	.07	.41
SMT(davon richtig)	.55	.46	.13	.89	.11	.36
NMT(übersetzt)	.91	.23	.87	1.0	.94	.79
NMT(davon richtig)	.81	0.0	.40	.67	.83	.55
Daten	1.0	1.0	1.0	1.0	1.0	1.0

Ergebnisse: Lange vs. kurze Fragen

System	kurz	lang	insgesamt
SMT(übersetzt)	18	9	27
SMT(davon richtig)	15	9	24
NMT(übersetzt)	16	36	52
NMT(davon richtig)	10	26	36
Daten	25	41	66

Ergebnisse: Lange vs. kurze Fragen

System	kurz	lang	insgesamt
SMT(übersetzt)	18	9	27
SMT(davon richtig)	15	9	24
NMT(übersetzt)	16	36	52
NMT(davon richtig)	10	26	36
Daten	25	41	66

Ergebnisse: Lange vs. kurze Fragen

System	kurz	lang	insgesamt
SMT(übersetzt)	18	9	27
SMT(davon richtig)	15	9	24
NMT(übersetzt)	16	36	52
NMT(davon richtig)	10	26	36
Daten	25	41	66

Ergebnisse: Lange vs. kurze Fragen in %

System	kurz	lang	insgesamt
SMT(übersetzt)	.72	.22	.41
SMT(davon richtig)	.60	.22	.36
NMT(übersetzt)	.64	.88	.79
NMT(davon richtig)	.40	.63	.55
Daten	1.0	1.0	1.0

Zusammenfassung

- NMT übersetzt mehr, dafür weniger präzise
- Das NMT-System schwächelt deutlich bei *dist* Fragen, die die höchste MRL-Durchschnittslänge haben
- Dafür ist es signifikant besser bei *findkey* und *least* ($p < 0.05$)
- Ausformulierte Form der Fragen kann mit NMT viel öfter übersetzt werden als mit SMT