

Projekte Softwareprojekt WS 23/24

Raphael Schumann
Doktorand Computerlinguistik
Universität Heidelberg

October 17, 2023

- **Hintergrundinformationen für beide Projekte**
- **Vorstellung Projekt 1**
- **Vorstellung Projekt 2**

Text-to-OverpassQL

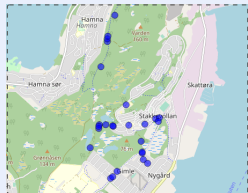
Natural Language Input:

Bike lanes 500 meters around the top of a hill or mountain in Troms.

Overpass Query Language:

```
1 | {{geocodeArea:"Troms"}}->.searchArea;  
2 | (  
3 |   node["natural"="peak"](area.searchArea);  
4 | )->.peaks;  
5 | way["highway"="cycleway"](area.searchArea)(around.peaks:500);  
6 | out center;
```

Query Execution Results:



```
{ "elements": [  
  { "type": "way",  
    "id": 4013565,  
    "center": { "lat": 69.6900047,  
               "lon": 18.9971581 },  
    "tags": { "highway": "cycleway",  
              "surface": "asphalt",  
              "segregated": "no" },  
  },  
  ...  
]
```

Daten:

- ▶ 6352 Instanzen im Trainingsset
- ▶ 1000 Instanzen im Developmentset
- ▶ 1000 Instanzen im Testset

Evaluationsmetriken:

- ▶ String Ähnlichkeit von generierten Queries und Referenz Queries
- ▶ "Execution Accuracy" Übereinstimmung der extrahierten Elemente
 - Unterschiedliche Queries können zum gleichen, richtigen Ergebniss führen

Model	OQS	EM	#Errors	EX	EX _{SOFT}
GPT-4	73.2	23.4	24	40.4	53.0

Large Language Model Prompting

The OverpassQL language allows one to formulate questions to the OpenStreetMap database.

Input:

castle in Deutschland

Overpass Query:

◊

Naive Anfrage an Large Language Model einen Overpass Query für "castles in Deutschland" zu generieren.

In-Context Learning (ICL)

The OverpassQL language allows one to formulate questions to the OpenStreetMap database.

Here are a few examples:

Input:

Find every castle in Luxemburg, Neatherlands and Belgium.

In-context Example 1

Overpass Query:

```
[out: json][timeout: 120]; {{{{geocodeArea: "Belgium"}}->.be; {{{{geocodeArea: "Luxembourg"}}->.lu; {{{{geocodeArea: "Nederland"}}->.nl; }}->.benelux; (node["historic"="castle"]["name"](area.benelux);););out center;
```

Input:

Castles in current view.

In-context Example 2

Overpass Query:

```
[out: json][timeout: 25]; (node["historic"="castle"]({{bbox}}); way["historic"="castle"]({{bbox}}); relation["historic"="castle"]({{bbox}}););out;>;out skel qt;
```

Input:

castles in Tuscany.

In-context Example 3

Overpass Query:

```
[out: json][timeout: 250]; {{{{geocodeArea: "Tuscany"}}->.searchArea; (node["historic"="castle"](area.searchArea); way["historic"="castle"](area.searchArea); relation["historic"="castle"](area.searchArea);););out;>;out skel qt;
```

Input:

castle in Deutschland

New Input

Overpass Query:



In-Context Learning (ICL)

Präsentiere Text und Query Paare aus dem Trainingsset als In-context Beispiele in der LLM Anfrage (k-shot in-context learning). **Systematisches Auswählen der Trainingsbeispiele ist meist besser als zufälliges Auswählen.**

- ▶ Vektorrepräsentation aller Anfragetexte im Trainingsset
- ▶ Vektorrepräsentation des Anfragetextes
- ▶ Suche k ähnlichste Anfragetexte aus dem Trainingsset
- ▶ Nutze diese Text und Query Paare als in-context Beispiele

Model	Setting	Overpass Query Similarity				EM	Execution Accuracy	
		chrF	KVS	TreeS	OQS		EX	EX _{SOFT}
5-Shot In-Context Learning								
GPT-4	random	63.8	57.5	61.1	60.8	5.1	25.4	39.5
GPT-4	retrieval-BLEU	74.3	66.1	72.4	71.0	22.9	38.5	50.7
GPT-4	retrieval-sBERT	75.7	69.9	74.0	73.2	23.4	40.4	53.0

Projekt 1: Similarity-Diversity In-context Example Selection

Beschreibung: Die In-context Beispiele sollen nicht nur nach Ähnlichkeit (Similarity) zum Anfragetext ausgesucht werden, sondern auch nach Vielfältigkeit (Diversity) untereinander.

- ▶ Vektorrepräsentation aller Anfragetexte im Trainingsset
- ▶ Vektorrepräsentation des Anfragetextes
- ▶ Suche 100 ähnlichsten Anfragetexte aus dem Trainingsset
- ▶ Erstelle k Clusters innerhalb der 100 Vektoren
- ▶ Wähle die k Zentroid-Instanzen als In-context Beispiele aus

Projekt 1: Similarity-Diversity In-context Example Selection

Ziel: Die Generierung von Overpass Queries mit Hilfe von LLMs soll verbessert werden.

Variationen:

- ▶ Auswirkung verschiedener Clustering Methoden
- ▶ Unterschiedliche Anzahl an Vektoren für das Clustering
- ▶ Ausprobieren verschiedener Vektorrepräsentation/Ähnlichkeitsmaße
 - tf-idf
 - word2vec
 - sentence BERT
 - ...

Projekt 1: Similarity-Diversity In-context Example Selection

Literatur für dieses Projekt:

Text-to-OverpassQL: A Natural Language Interface for Complex Geodata Querying of OpenStreetMap

Michael Staniek, Raphael Schumann, Maike Züffle, Stefan Riezler
2023, <https://arxiv.org/pdf/2308.16060.pdf>





Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

Nils Reimers, Iryna Gurevych
2019, <https://arxiv.org/pdf/1908.10084.pdf>

What Makes Good In-Context Examples for GPT-3?

Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, Weizhu Chen
2022, <https://aclanthology.org/2022.deeLIO-1.10.pdf>

Projekt 2: Documentation Augmented In-context Learning

Key	Value	Element	Comment
Sustenance			
amenity	bar		Bar is a purpose-built commercial establishment that sells alcoholic drinks to be consumed on the premises. They are characterised by a noisy and vibrant atmosphere, similar to a party and usually don't sell food. See also the description of the tags amenity=pub;bar;restaurant for a distinction between these.
amenity	biergarten		Biergarten or beer garden is an open-air area where alcoholic beverages along with food is prepared and served. See also the description of the tags amenity=pub;bar;restaurant . A biergarten can commonly be found attached to a beer hall, pub, bar, or restaurant. In this case, you can use biergarten=yes additional to amenity=pub;bar;restaurant .
amenity	cafe		Cafe is generally an informal place that offers casual meals and beverages; typically, the focus is on coffee or tea. Also known as a coffeehouse/shop , bistro or sidewalk cafe . The kind of food served may be mapped with the tags cuisine=* and diet:*=* . See also the tags amenity=restaurant;bar;fast_food .
amenity	fast_food		Fast food restaurant (see also amenity=restaurant). The kind of food served can be tagged with cuisine=* and diet:*=* .

https://wiki.openstreetmap.org/wiki/Map_features

Projekt 2: Documentation Augmented In-context Learning

Beschreibung: Nutzung der OpenStreetMap Dokumentation um dem LLM explizit nützliche Informationen über die aktuelle Anfrage bereitzustellen.

Die relevantesten Informationen aus der Wiki sollen automatisch gefunden, und dem LLM Prompt angehängt werden.

- ▶ Vektorrepräsentation aller Beschreibungen von Key-Value Paaren in OpenStreetMap Wiki
- ▶ Vektorrepräsentation des Anfrage Textes
- ▶ Suche 5 ähnlichsten Key-Value Beschreibungen zu Anfrage Text
- ▶ Hänge die Key-Value Paare mit Beschreibungen an den LLM Prompt

Projekt 2: Documentation Augmented In-context Learning

Ziel: Die Generierung von Overpass Queries mit Hilfe von LLMs soll verbessert werden.

Variationen:

- ▶ Ausprobieren verschiedener Vektorrepräsentation/Ähnlichkeitsmaße
 - tf-idf
 - word2vec
 - sentence BERT
 - ...
- ▶ Erweiterung auf die komplette Wiki von OpenStreetMap (weniger strukturiert als die Map Features Tabelle)

Projekt 2: Documentation Augmented In-context Learning

Literatur für dieses Projekt:

Text-to-OverpassQL: A Natural Language Interface for Complex Geodata Querying of OpenStreetMap

Michael Staniek, Raphael Schumann, Maike Züffle, Stefan Riezler
2023, <https://arxiv.org/pdf/2308.16060.pdf>

Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

Nils Reimers, Iryna Gurevych
2019, <https://arxiv.org/pdf/1908.10084.pdf>

What Makes Good In-Context Examples for GPT-3?

Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, Weizhu Chen
2022, <https://aclanthology.org/2022.deeLIO-1.10.pdf>

Ende

Ende