

# Bachelor-Forschungsprojekt Informatik: Relevante OSM-Tags vorschlagen

Marco Hildebrand, XXXX, stXXXX@stud.uni-stuttgart.de  
Lukas Baur, 3131138, st141998@stud.uni-stuttgart.de  
Felix Bühler, 2973410, st117123@stud.uni-stuttgart.de

October 8, 2018

## Abstract

Die vom *Institut für Formale Methoden der Informatik Stuttgart* entwickelte textbasierte Suchmaschine *OSCAR*, die OpenStreetMap-Daten auf Eingabe von OSM-Tags durchsucht, liefert unbefriedigende Ergebnisse auf anderweitige textuelle Eingaben. Im Rahmen unseres Bachelor-Forschungsprojekt Informatik sollte diese Lücke geschlossen werden, indem eine Anfrage an das von uns entwickelte System eine Menge an damit verwandten, relevanten Tags zurückgibt.

# 1 Einleitendes

## 1.1 Projektrahmen

Die Arbeit wurde im Rahmen des *Bachelor-Forschungsprojekts Informatik* in der Zeit vom April bis Oktober 2018 angefertigt. Diese Ausarbeitung stellt die inhaltliche Dokumentation des entwickelten Moduls dar.

## 1.2 initiale Problemstellung

Grundlage für unsere Arbeit war die Suchmaschine *OSCAR*, die vom *Institut für Formale Methoden der Universität Stuttgart* entwickelt wurde.

OSCAR durchsucht auf Eingabe eines *OpenStreetMap-Tags* die zugehörige Datenbank nach passenden Einträgen und bereitet das Suchresultat grafisch auf. Ein *Tag* ist in OpenStreetMap wie folgt definiert:

$$\textit{key}=\textit{value}$$

Ein *key* wird benutzt, um ein Themenbereich zu charakterisieren, es repräsentiert einen Typ oder beschreibt ein Feature. Außerdem werden Tags vereinzelt als Namespaces verwendet [1].

Der *value*-Teil stellt ein Wert des Features da. Typische Werte sind Eigenschaften oder Zahlen [1]. Beispiele für Tags sind *building=yes*, *building=house* oder *highway=service* [2][3].

Da die Eingabe auf Tags beschränkt ist, benötigt ein User zur Suche einen passenden Tag. Diese Lücke soll mithilfe dieses Projekts geschlossen werden. Das zu entwickelnde System soll auf Eingabe eines natürlichen Wortes der englischen Sprache möglichst eng verwandte, relevante OpenStreetMap-Tags vorschlagen.

## 1.3 Abgrenzungen

Unsere Arbeit konzentriert sich auf die Suche der relevanten Tags zu einem eingegebenen Wort. Formaler ausgedrückt besteht unsere Eingabe aus genau einem Wort der englischen Sprache, das nicht in der zugrundeliegenden Stop-Word-Liste enthalten ist.

## 1.4 Planungsaspekte

Zu Beginn unserer Arbeit gliederten wir unser Projekt in

# 2 notizen

Phase 1: Planung - Tags und dazugehörige semantische Beschreibung holen - in Struktur bringen - Suchanfrage an Daten - vorhanden/nicht vorhanden -> bewerbung fehlt - tf-idf -> gut, aber Problem: Mehrere Links auf dieselbe Seite -> Duplikate entfernen -> hohe Gewichtung für kleine Seiten -> Multiplizieren mit log/oder Wurzel 2

# 3 Einleitung

## 3.1 Projektbeschreibung

## 4 Vorgehensweise

Anschauen von wiki xml dump

unbrauchbare daten, da viel untereinander verlinkt ist.

herunterladen der tags: <https://taginfo.openstreetmap.org/>

## 5 Gettings started

### 5.1 languages

einfach eine liste aller sprachen bekommen mithilfe `taginfo-wiki.db`.

Die kann man von <https://taginfo.openstreetmap.org/download> herunterladen.

### 5.2 export-links

herunterladen der osm-wiki sitemap <https://wiki.openstreetmap.org/sitemap-index-wiki.xml>

davon interessiert uns nur `sitemap-wiki-NS_0-0.xml` der rest enthält daten zu den nutzern, diskussionen und historie

## 6 crawl

alle gesammelten link in die `links.txt` legen

```
scrapy crawl osmWiki -t json -o keys.json
```

### 6.1 pretty json

```
python -m json.tool keys.json > keys-pretty.json
```

## 7 Anhang

### References

- [1] wikibooks. Tag. <https://wiki.openstreetmap.org/wiki/Tags>. Accessed: 2018-10-08.
- [2] taginfo. building. <https://taginfo.openstreetmap.org/keys/buildingvalues>. Accessed: 2018-10-08.
- [3] taginfo. building. <https://taginfo.openstreetmap.org/tags/highway=service>. Accessed: 2018-10-08.