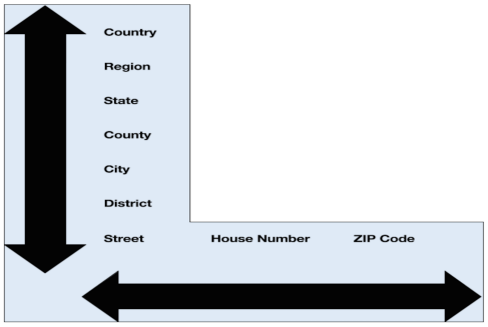
Andreas Egloff

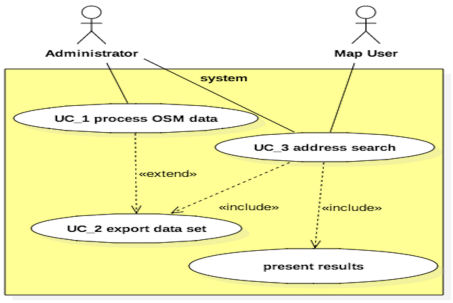
Graduate Candidate	Andreas Egloff
Examiner	Prof. Stefan F. Keller
Co-Examiner	Prof. Stefan F. Keller
Subject Area	Software and Systems
Project Partner	Klokan Technologies GmbH

OSMNames

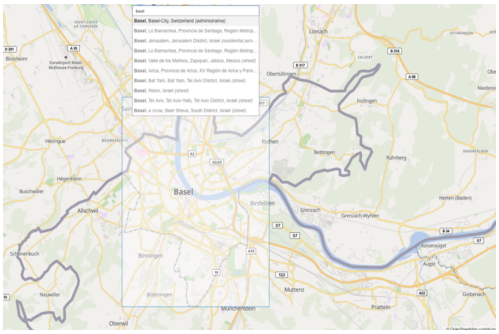
Local Open Street Name Database



Street names and postal address overview (Source: Stefan Keller).



Use cases of OSMNames.



OSMNames used for geo names search (geocoder).

Problem: There is a need for a data set consisting of street names (geo names) of the world. Such gazetteer data however is either not available for every country or is not in a suitable format. Furthermore, if such data is found, it often has no open license. Another problem we face is the formatting of postal building addresses. Each country has its own set of formatting rules and in order to have a database of addresses one needs to bring these different formats to a common format and vice versa. Finally, such a dataset can be used in conjunction with a geocoder in various applications.

Approach/Technologies: The main objective is creating a global and structured data set with street names with the help of OpenStreetMap (OSM) data. There are two ways to look at this assignment as depicted in the L-shaped figure. One way is to look at the vertical axis representing the hierarchy and the street names, the other being the horizontal, contextual axis representing a postal address. The focus of this work will be on creating a global data set representing the vertical axis. The tasks can be summarized as follows:

- Analyze similar products, mostly Nominatim in order to get familiar with the topic.
- Create a Docker workflow that imports OSM data and does the ranking of the features.
- Refine data quality especially when it comes to street segments that belong together.
- With the help of the ranking from above the features should get hierarchized.
- An export should be provided so it can be used in conjunction with a geocoder.
- The data should be made available to download in a usable format.

The whole Docker workflow can be run regularly at a later date.

Result: First of all, a data set containing planet data could be provided. A secondary target, importing additional house numbers for each street, could not be met in scope of this project. The workflow could be setup in such a way, that it can easily be setup anywhere (via Docker). The simplicity of the installation and, most of all, the clear arrangement of code is a big advantage over other products like Nominatim. After all, the processing run times are way faster than the latter, which takes up to several days for a global data set. The TSV file from the planet export includes 21'055'840 entries. The current data export can be downloaded at https://osmnames.org. The resulting data have been successfully integrated by Klokan Technologies with a SphinxSearch powered geocoder. The result includes osm2vectortiles-generated vector tiles and can be seen here: https://osmnames.klokantech.com/.