

B4M33OSW: Checkpoint 0

October 18, 2017

1 Selected subtopic

For the semestral work, I have decided to integrate timetables of urban public transport in two cities, specifically in Prague and Pilsen.

1.1 Goals

The goal is to compare rush hours in the two cities, and possibly to find out parts of the cities with the highest density of the urban public transport network.

2 Data sources

During research, I haven't found any data source containing data for both Prague and Pilsen simultaneously, and the data sources I have discovered differ in their schema or format.

2.1 Prague

As mentioned in related data sources on CourseWare, the city of Prague provides open data at [1].

On the website, there is an urban public transport data source created by the Prague Public Transit Company. The data is distributed in the GTFS format [2]. Information such as names of stops or arrival and departure times is present in the GTFS feed.

Since the data is well structured and distributed in an open, well-documented format, I expect it to be machine-processed easily.

2.2 Pilsen

Pilsen provides open data, too. The portal is available at [3] and is very similar to the Prague one. There is a list of all public transport stops, but unfortunately, in this case, there isn't data on either arrival or departure times.

I have found the stop times of public transport in Pilsen at [4]. On the contrary to the Prague open data, here we have only departure time instead

of a pair of arrival and departure time, and furthermore, the data is published directly on the web page in several HTML tables. It will probably be necessary to parse those HTML tables and transform them into some more machine-friendly format.

References

- [1] <http://opendata.praha.eu/>
- [2] <https://developers.google.com/transit/gtfs/reference/>
- [3] <https://opendata.plzen.eu/>
- [4] <http://jizdnirady.pmdp.cz/LinesList.aspx>