Data Engineering Applied to the Swiss Gig Economy

Collecting, processing and analysing data. This work is focused on the gig economy in Switzerland and aims to make a statement with freely available data on the internet. The technical implementation is in the foreground and is intended to show a solution of how these three steps are possible with modern tools. It also shows the practical use of Natural Language Processing and its difficulties.

Context and goals

This is a thesis focused on applying the methods of data engineering to the information sources powering the Swiss Gig Economy that focus on short-term commitments. The content of this thesis is the utilization of publicly available information as a basis for analyzing the state of the Swiss Gig Economy. The work has been done in cooperation with a research group from BFH Wirtschaft and Bachelor thesis students from University of Bern. The practical aspects cover the application of scraping and crawling and the use of machine learning (specifically natural language processing) to transform data into information.

Content of the thesis

This work is intended to show the way from data acquisition, cleansing and finally deeper analyses, especially in the area of Natural Language Processing. The thesis is structured in 3 parts, scraping, extract and clean and analyses. In an initial step, in order to analyse the Gig Economy in Switzerland, as much data as possible is taken from the freely available area of websites that provide a platform for mediation between employers and employees.
The second part is about getting as much information as possible from this raw data and saving it so that it can be processed further.
The last part is dedicated to the visualisation of data so that information can be extracted from it. The technical implementation is in the foreground here, but some interpretations are dared and the analyses are discussed with the BFH Wirtschaft, where it also emerges that conclusions drawn too quickly must be treated with caution.

Conclusion

Scraping is always individual per target page, which makes the effort high. The approach of collecting data and then examining what can be gained from it is exciting but also holds a level of uncertainty.
In Natural Language Processing, there are many different approaches to choose from and time can be lost in trial and error as the models have several hyperparameters. Sometimes the results are very good, sometimes bad and it is hard to know why, because it is difficult to see behind the model.
With the results, a rough analysis can be made of what the gig economy looks like in late 2020, but in order to make a more general analysis of the gig economy, the data would have to be scraped monthly or even more often from the platforms and the data constantly expanded.
Nevertheless the results of this work have shown to be valuable input for research groups from the university and have allowed researchers to refine their research approach.

Thomas Baumann
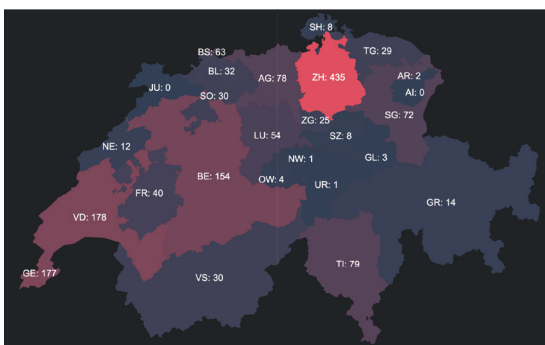
A map of the swiss gig workers from the platforms Upwork and Freelancer

The word cloud generated from the skills of Fiverr