Web Scraping and Natural Language Processing for real estate valuation

Background
Real estate valuation has so far been carried out using cumbersome and lengthy processes. The appraisers spend more than 50% of their time gathering information about the market and the location of the property, formatting tables and diagrams and writing a 100-page report with the same content. The digitization of these processes leads to enormous time savings. Therefore, the aim of this IDP is to develop an application that automates market information gathering and the creation of valuation reports from existing data with the help of Natural Language Generation.

Task
In this project, you will work on the following tasks:

- Design and development of data pipelines from web crawling / scraping to storage.
  - Obtain data from structured and unstructured data sources.
  - Extraction of data (text, numbers, images) from several webpages and documents in different formats.
  - Processing and saving the data in a suitable database.
- Design and development of output templates for automated data export.
  - Conversion of existing data into natural language with the help of natural language generation algorithms.
  - Create graphs and diagrams from existing data.
  - Merging the NLG texts and graphics / diagrams into a pre-formatted output template.
Requirements

- Understanding of back-end developments.
- Experience with Web Scraping, Natural Language Processing and databases.
- Systematic and independent way of working and the ability to work in a team.

Contact

Please do not hesitate to contact me if you have any questions regarding this project:

Dr.-Ing. Christina Mauer                  M.Sc. Maximilian Reiser
Tel.: +49 89 289-22483                  Tel.: +49 89 289-22483
E-Mail: c.mauer@tum.de                  E-Mail: m.reiser@tum.de