Master thesis (30 ECTS)	**MT_00**

Working title

Transformer-based models for automated tabular information extraction from pre-processed civil engineering reports

Project objectives

The objective is to develop a machine learning model, based on transformer architecture, capable of automated extraction and structuring of tabular information from pre-processed technical reports in civil engineering. The thesis starts with a literature review on available transformer architectures for table description and where they are used already. After pre-processing the reports, the next step will be object detection via transformer algorithms like "YOLO", "TrOCR", "Faster R-CNN" to identify key elements. Followed by dividing images into meaningful segments and extracting relevant features. Use transformers for image description (unstructured information extraction). Validate the accuracy of the extracted information and refine the model.

Student has enthusiasm for

Digitalisation, programming, geotechnical data

Requirements on student

Attention to detail, reliability, programming skills (can be acquired during the thesis)

Start (earliest / latest)

February 2024

Project term (min. / max.)

Approx. 6 months full-time

Coop. with external institution

no

Contact person

Paul Unterlaß
unterlass@tugraz.at