Reading comprehension and question answering are computer science disciplines in the field of natural language processing and information retrieval. Reading comprehension is the ability of the model to read text, process it and understand its meaning. One of its applications is in question answering tasks, which is concerned with building a system that can automatically find an answer in the text to a certain question relied on the content of the text. It is a well-studied task, with huge training datasets in English. However, there are no Czech datasets and models for this task.

This work focuses on building reading comprehension and question answering systems for Czech, without requiring any manually annotated Czech training data. Our main focus is to create Czech training and development datasets, create the models for the Czech question answering system using Czech data, and create the models for the Czech question answering system using English data and cross-lingual transfer and compare the results and select the best model. First of all, we translated freely available English question answering datasets SQuAD 1.1 and SQuAD 2.0 to Czech to create training and development datasets. We then trained and evaluated several BERT and XLM-RoBERTa baseline models used for the question answering task in English. The best results were obtained XLM-RoBERTa model trained on English and evaluated directly on Czech. This model achieved very good results, similar to the model trained on the translated Czech data. However, we consider the result obtained from XLM-Roberta to be overperforming the models trained on Czech because the model does not require any Czech data during training. This also proves that the cross-lingual transfer approach is very flexible and provides reading comprehension in any language, for which we have enough monolingual raw texts to pretrain the language model.