Neural Networks applied in NLP

Lab 4

We will create today a Question Answering system for Romanian using a pre-trained BERT model and the SQuAD dataset.

Stanford Question Answering Dataset (SQuAD) is an English reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable.

Bidirectional Encoder Representations from Transformers (BERT) is a transformer-based machine learning technique for natural language processing developed by Google. It has several pre-trained models you can use, including for Romanian.

Your task for this lab involves several steps:

1. (0.9) Use a fined tuned pre-trained BERT model to predict answers to questions given by the user.
2. (0.1) Translate the answer to Romanian using a freely available translation API.

Reference: https://towardsdatascience.com/question-answering-with-a-fine-tuned-bert-bc4dafd45626