Exercise Evaluation

Natural Language Processing

Ulrich Heid, Christian Wartena, Johannes Schäfer

Summer Term 2022

The goal of this exercise is to build a small dataset for evaluation. First you should have two models. E.g. you could load two different pre-trained vector sets into Gensim (see e.g. `https://huggingface.co/models?other=gensim`) or you could train two Gensim models with different parameter settings yourself using an English corpus e.g. from `https://wortschatz.uni-leipzig.de/en/download/English` or from `https://www.corpusdata.org/`.

Now construct a data set to evaluate your models. The data set should have between 50 and 200 examples and it should be able to differentiate between the two models. you can select data by hand, but it would be good to base yor decisions on external data sources, like Wikipedia or Wordnet. Make sure that in your data sets only words are used for which both models have vectors.

You have two options, either

1. make a list of related (not necessarily synonyms) and non-related words, or

2. make a list of words that belong to a few classes, like emotions, processes, actions, fruits, materials, buildings, etc.

Hand in a description of your results and the list of words in tsv-format (utf8-encoded plain text; columns separeted by a tab).

We will evaluate two models with you datasets. Next week we will see and discuss whether each data set evaluates the models in the same way.