

Neural Networks applied in NLP

Lab 3

N-gram Language Model:

1. (0.25p) Use an API to collect a corpus of Romanian texts (blogs, Wikipedia articles, news, reviews etc.) of at least 1000 unique words.
2. (0.50p) Implement an N-gram language model for the Romanian language, using a smoothing technique at your choice.
3. (0.25p) Compute the probability of a new text, given at input, using the n-gram model you developed.