27.12.2022

AYDIN ADNAN MENDERES UNIVERSITY

CSE 431 NATURAL LANGUAGE PROCESSING WITH MACHINE LEARNING

THIRD TERM PROJECT

Due Date: 17.01.2023 23:59

Your third term project is related to named entity recognition (NER) tagging using Hidden Markov Model (HMM) in Turkish.

Initially, you need a labeled dataset (MilliyetNER dataset) [1] involving the words with their NER tags in Turkish. Then calculate the emission and transition probabilies properly for NER. After obtaining the related probability values, apply Viterbi algorithm to get the best NER tag for each word in a given dev/test sentence. [2] involves the steps and the sample code in detail for an English dataset. Show the accuracy, precision, recall and F1 scores for sentence and word levels in dev/test sets.

[1] https://data.tdd.ai/#/effafb5f-ebfc-4e5c-9a63-4f709ec1a135

[2] https://medium.com/mlearning-ai/introduction-to-hidden-markov-model-hmm-with-simple-ner-d1353ff35842