Contribution ID: **300**

Type: **not specified**

Evaluation of named entity recognition program packages for data mining

Thursday, 6 July 2023 17:45 (15 minutes)

Natural language processing technologies are one of the key areas in the field of data analysis. Natural language processing performs a plenty of tasks, which include the task of named-entity recognition. It provides an opportunity to get value information from a large amount of data. The study is devoted to select better program packages for named-entity recognition from Russian news text.
To choose the program packages, a corpus of 70 news articles has been collected from different Internet resources. “Natasha”, “SpaCy”, “Stanza”, “DeepPavlov”’s models («ner rus bert probas», «ner rus bert», «ner ontonotes bert mult») were selected for conducting an experiment. Named entities were extracted manually and using the program packages. After receiving the result of the packages performing, the data was processed and metrics precision, recall, f-measure were calculated.
According to the results of the experiment, the packages “Natasha” and “SpaCy” were selected for their accurate recognition of entities. It was concluded that “Natasha” better recognizes entities such as “PER” (person) and “LOC” (location), “SpaCy” is able to recognize data such as “ORG” (organization) without breaking the semantic part. The obtained result can be used to create an algorithm to recognize entities from Russian-language articles for further data analysis.

Summary

Primary author: SOKOLOV, Ivan (NRNU MEPHI)

Co-authors: ARTAMONOV, Alexey (National Research Nuclear University MEPHI); ANTONOV, Evgeniy (Plekhanov Russian University of Economics, NRNU MEPHI)

Presenter: SOKOLOV, Ivan (NRNU MEPHI)

Session Classification: Big Data, Machine Learning and Artificial Intelligence

Track Classification: Big Data, Machine Learning and Artificial Intelligence