# NLP Assignment #3

Name and surname: Anatoliy Pushkarev
University email: a.pushkarev@innopolis.university
Nickname in CodaLab: anatoliypus
Link to Github repository: https://github.com/anatoliypus/NLP-Assignment-3

---

## Solution 1

My first solution is dictionary approach, where I save entities in the train set with their corresponding labels. After that on the test set saved entities are marked with the most common labels.

Pipeline:

1. Loading and preprocessing the data
2. Saving the entities from the test set
3. Getting tokens and spans with razdel
4. Preprocessing
5. Testing on the test set

Results:

Mention F1: 36.90%
Mention recall: 29.47%
Mention precision: 49.34%
Macro F1: 31.86%

---

## Solution 2

My second solution suggests training pretrained ru_core_news model from Spacy, which is a Russian pipeline, optimized for CPU.

Pipeline:

1. Preprocessing data for fitting to the model
2. Training the model (12 epochs)
3. Testing the model

Results:

Mention F1: 37.20%
Mention recall: 29.76%
Mention precision: 49.59%
Macro F1: 26.46%

## Final words

As a result, Spacy model worked a bit better, probably because it not just saved ners as dict approach, but found some minor dependencies in tokens.