

Assignment 3: Social Media NER with Transfer Learning

For this assignment, you will create a tool for tagging named entities in social media. It's likely that you won't have much data - so use transfer learning of some kind to get a good tagger.

The tagger

Implement the tagger in any language and framework of your choice. Perhaps you'd like to start from one of the examples mentioned in the transfer learning tutorial, or perhaps you have another idea. You're also welcome to use CRFSuite and build social media features manually.

The data

Download some named entity recognition data for a language of your choice. Then, find some social media data tagged for named entities as well. If you're using English, you could try Ontonotes for the main data and the Broad Twitter corpus for the social media corpus. You might also like to use a social media platform API to extract some data for training or fine-tuning a language model.

The analysis

What did the tagger do well? What kind of mistakes did it make? Find what the most common errors were, and give some examples of them - as well as some examples of the tagger working correctly. Report F1 overall and for each entity type.

Assignment hand-in

Essay

Description of what you did and why, describing your general code, and the answers to all of the above questions in the work description. About 1000 words.

Code

Include your code. A link to a Colab notebook is best. Test the whole notebook first. I will run it myself.

How?

By Innopolis Moodle.

When?

The end of December 6. I must deliver final marks to the administration by December 9, so no extension is possible.