Traitement Automatique du Langage Naturel

TP 3 Named entity recognition (NER) Token classification

Start notebook TP 3

swedish medical NER Dataset

Swedish BERT Models

Using the start notebook provided in the link, build and train your own NER with swedish_medical_ner dataset using the Swedish BERT Models.

you will be evaluated according to the performance of your model

1 place 5
2 place 4.5
3 place 4
4 places 3.5

5 places 3

Token classification

This generic task encompasses any problem that can be formulated as “attributing a label to each token in a sentence,” such as:

- **Named entity recognition (NER)**: Find the entities (such as persons, locations, or organizations) in a sentence. This can be formulated as attributing a label to each token by having one class per entity and one class for “no entity.”
- **Part-of-speech tagging (POS)**: Mark each word in a sentence as corresponding to a particular part of speech (such as noun, verb, adjective, etc.).
- **Chunking**: Find the tokens that belong to the same entity. This task (which can be combined with POS or NER) can be formulated as attributing one label (usually `B-`) to any tokens that are at the beginning of a chunk, another label (usually `I-`) to tokens that are inside a chunk, and a third label (usually `O`) to tokens that don’t belong to any chunk.