



Natural Language Processing – Fall 23/24

Final Project

Prof. Mohammad Azzeh

Out date: 26-12-2023
Due date: 11-Jan-2024

Part-of-speech (POS) tagging simply means labeling words with their appropriate Part-Of-Speech so it explains how a word is used in a sentence. The following example explains what does POS mean.

word	POS
جاء	verb
في	Preposition
التعاون	Noun

The following list contains all possible tags in any language:

Tag	Description
ADV	Adverb
ADJ	Adjective
CONJ	Conjunction
DET	Determiner
NOUN	Noun
NSUFF	Noun suffix
NUM	Number
PART	Particle
PREP	Preposition
PRON	Pronoun
PUNC	Punctuation
V	Verb
ABBREV	Abbreviation
CASE	Alef of tanween fatha
JUS	Jussification attached to verbs
VSUFF	Verb Suffix
FOREIGN	Non-Arabic as well as non-MSA words
FUR_PART	Future particle "s" prefix and "swf"
PROG_PART	Progressive particle
NEG_PART	Negation particle
HASH	Hashtag
EMOT	Emoticon/Emoji
MENTION	Mention
URL	URL

POS is considered as sequence-to-sequence problem which can be implemented using RNN or LSTM. You are requested to build Arabic POS tagging model based on the following dataset (https://huggingface.co/datasets/arabic_pos_dialect) from huggingface website. You should learn how to download this dataset before you use it. Your model should use **RNN Encoder-Decoder with attention**.

The inference model should be able to receive Arabic sentence and returns POS tag for each word as follows:

- `Input`: ["بس", "يلي", "بدي", "افهموا", "إنوا", "أنا", "يسأل", "الدكتور", "سؤال", "ليش", "كل", " ", "للكاترة", "رفقائي", "بيتحولوا", "EOS"]

- `Segmentation`: ["بس", "يلي", "بدي", "افهموا", "إنوا", "أنا", "ب+سأل", "ال+دكتور", " ", "للكاترة", "رفقائي", "سؤال", "ليش", "كل", "رفق+ات+ي", "ب+يتحولوا", "ال+دكاترة", "EOS"]

- `POS`: ["PART", "PART", "NOUN+PRON", "V+PRON", "PART+PRON", "PRON", "PROG_PART+V", "DET+NOUN", "NOUN", "PART", "NOUN", "NOUN+NSUFF+PRON", "PROG_PART+V+PRON", "PREP+NOUN+NSUFF", "EOS"]

What to submit:

1. A report that explains all actions and configurations you did to complete the experiments in addition to the results of comparison.
2. a Jupyter Notebook with complete code.