

DATA 622 | HOMEWORK 4

Questions

Use the Gettysburg Address by Abraham Lincoln.

1. Tokenization

Tokenize both sentences into words using spaCy. Print the list of tokens for each sentence. Also use the *benepar* library.

2. Part-of-Speech Tagging

Print the part-of-speech (POS) tag for each token in the first sentence.

3. Dependency Parsing

Print the dependency relation and head word for each token in the second sentence.

4. Constituent Parsing

Using the *NLTK* and *benepar* libraries, print the constituency (phrase structure) parse tree of the first sentence.

5. Extract Noun Phrases

Using spaCy, extract all noun phrases (noun chunks) from both sentences.

5. CRF and HMM

Why do you use CRF and HMM? How do they differ? Please summarize in less than 50 words.