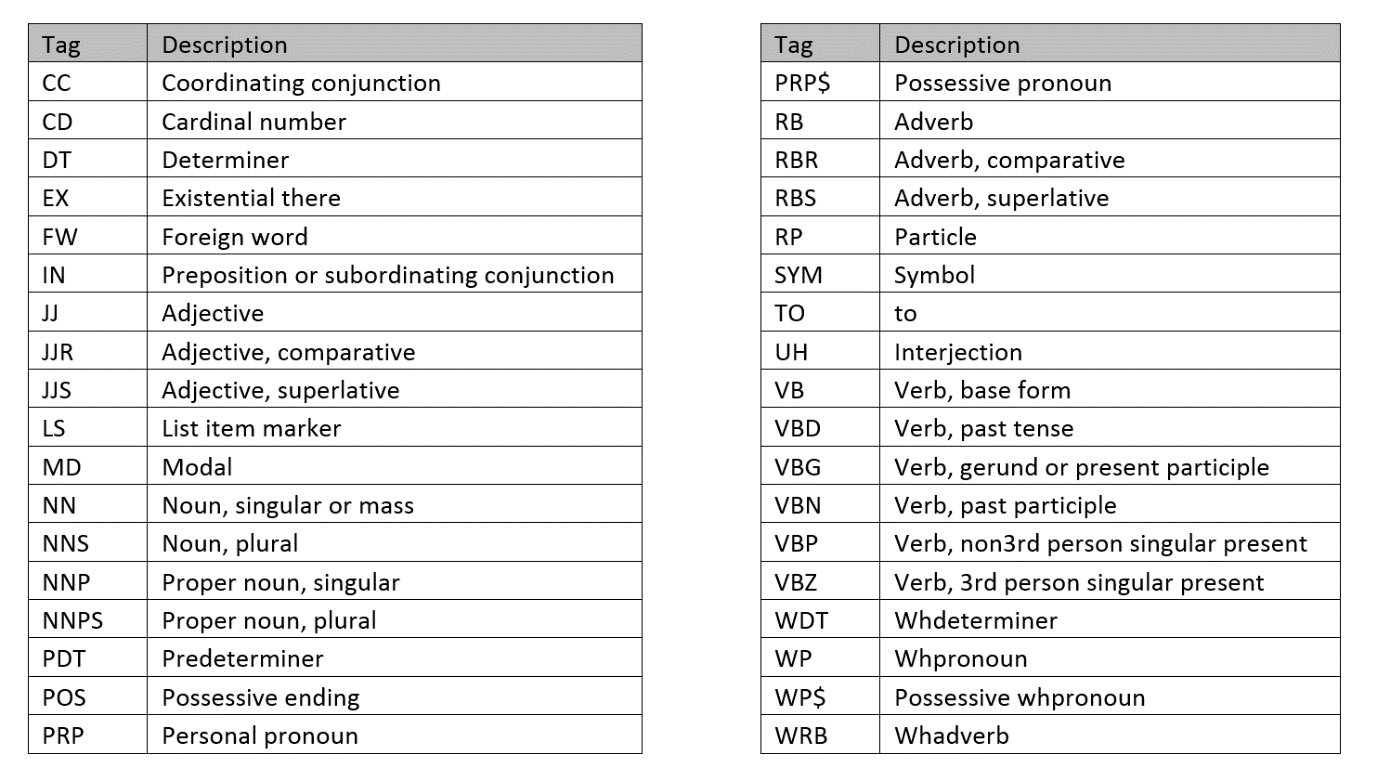
OpenNLP vs coreNLP :

1. Named Entity Recognition: I think CoreNLP clearly wins here, both on accuracy and ease-of-use. For one, OpenNLP has a model per NER tag, whereas CoreNLP detects all tags with a single Annotator. Furthermore, temporal resolution with SUTime is a nice perk in CoreNLP. Accuracy-wise, my anecdotal experience is that CoreNLP does better on general-purpose text.
2. Gender identification. I think both tools are kind of poorly documented on this front. OpenNLP seems to have a GenderModel class; CoreNLP has a gender Annotator.
3. Training API. I suspect the OpenNLP training API is easier-to-use for not off-the-shelf training. But, if all you want to do is, e.g., train a model from a CoNLL file, both should be straightforward. Training speed tends to be faster with CoreNLP than other tools I've tried, but I haven't benchmarked it formally, so take that with a grain of salt.

NLP pipeline - <https://www.analyticsvidhya.com/blog/2022/06/an-end-to-end-guide-on-nlp-pipeline/#:~:text=NLP%20Pipeline%20is%20a%20set,and%20Pipeline%20is%20non%2Dlinear>.

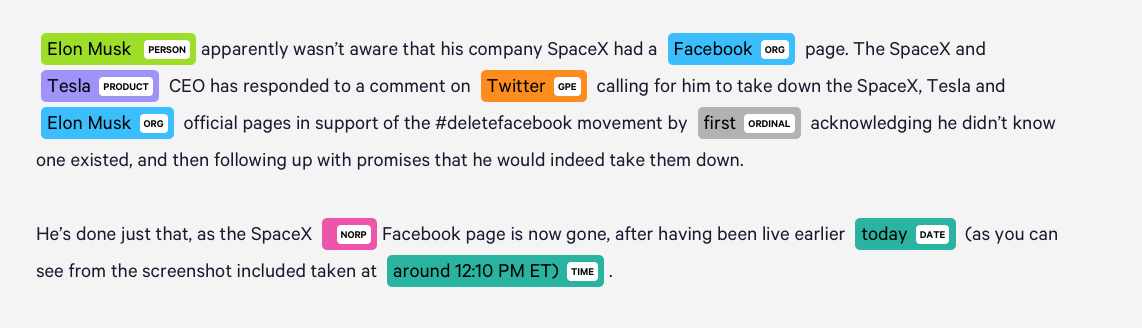
Tokenization - Tokenization is the process of tokenizing or splitting a string, text into a list of tokens. <https://www.geeksforgeeks.org/nlp-how-tokenizing-text-sentence-words-works/#:~:text=Tokenization%20is%20the%20process%20of,a%20token%20in%20a%20paragraph>.

Parts of Speech –



Lemmatization - Lemmatization is a text normalization technique used in Natural Language Processing (NLP), that switches any kind of a word to its base root mode. Lemmatization is responsible for grouping different inflected forms of words into the root form, having the same meaning. <https://www.engati.com/glossary/lemmatization#:~:text=REQUEST%20A%20DEMO-,What%20is%20Lemmatization%20in%20NLP%3F,form%2C%20having%20the%20same%20meaning>.

Named Entity Recognition - It involves the identification of key information in the text and classification into a set of predefined categories eg. Name, location, city etc.



Parse - is used to draw exact meaning or dictionary meaning from the text. It is also called Syntactic analysis or syntax analysis. Comparing the rules of formal grammar, syntax analysis checks the text for meaningfulness. The sentence like “Give me hot ice-cream”, for example, would be rejected by parser or syntactic analyser. <https://www.tutorialspoint.com/natural_language_toolkit/natural_language_toolkit_parsing.htm>

Sentiment - Sentiment analysis (or opinion mining) is a natural language processing (NLP) technique used to determine whether data is positive, negative or neutral. Sentiment analysis is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand customer needs.