

NLTK vs SpaCy Report:

By: Codie Munos

Natural Language Processing libraries are commonly used nowadays. NLTK and SpaCy provide open-source information to build models either from scratch or with pre-existing models. Researching both of these led me to the conclusion that one is not necessarily better than the other, rather they both work really well based on what the model is used for. In the sense of scholarly needs, it is best to go with NLTK as it is more suitable for customization and research. While it is better to use SpaCy when needing to analyze larger text data in a faster, efficient way. Another factor of choosing between the two is how advanced the user is in using NLP libraries. If it is someone just starting out, it is suggested to start with SpaCy as it is more user friendly and offers great pre-existing models. Something to also consider is the language that the user plans on using in the model. NLTK offers more compatibility with languages around the globe which gives it an advantage of being used more in specific regions. This research project gave me great insight for future uses and provides me with a guideline on how to pick an NLP library that caters to my needs.