Houston Community College

**NLTK vs spaCy Report**

NLP tools

Author: Trevon Woods

Deep Learning - ITAI - 2376

Anna Devarakonda

06/12/2025

For this assignment I have decided to research and conduct analysis of the NLTK and spaCy libraries. Both libraries have their own strengths but each have significant drawbacks as well depending on how you use them. Through this report I will convey these strengths and weaknesses and also provide a broadened understanding of the application and history behind each library.

Firstly, let's begin with the background of NLTK. NLTK is a popular library used for natural language processing. It was initially created in 2001 at the University of Pennsylvania by Steven Bird and Edward Loper. The library's purpose was mainly to support the teaching and research of NLP concepts. Some of this library’s key features are syntax analysis, natural language understanding, and Named Entity Recognition. It is also a platform that provides a way to experiment with different NLP concepts and techniques. NLTK is open-source so its development is driven by the collaborative effort of researchers and developers.

Now let’s move on to the background of spaCy. The library was first released in 2015 and developed by Explosion AI, a software company founded by Mathew Honnibal and Ines Montani. Their initial motivation for creating the library was to provide an efficient and production-ready NLP library. It has many key features such as Part-of-speech tagging, rule based matching, and word vectors. SpaCy is also an open-source project so its development is driven by the collaborative effort of its community. It has many purposes such as Industrial-Strength NLP, speed and efficiency, production readiness, accuracy, and a streamlined workflow.

For my comparative analysis I'll start with NLTK. It offers fine-grained control for diverse NLP tasks giving rise to many different customized approaches. From personal experience I’ve noticed that It struggles with things like stopword removal. SpaCy on the other hand, is very robust and seems to cover most edge cases. It excels in efficiency and speed which is ideal for production environments. It does a great job with stop word removal, even catching and removing contractions. The library boasts many pre-trained models for various NLP tasks which saves a lot of development time.

Overall both of these frameworks have many strengths and weaknesses that make them viable for different preprocessing scenarios. These libraries are used in everyday AI technology like chatbots and for NLP tasks such as sentiment analysis and text summarization. They each have extensive history and libraries in the NLP space. Both are open-source projects that are fueled by the developers and researchers in their respective community. Overall, spaCy does a better job with being thorough when processing text data while NLTK seems to be more useful for tokenization.

**Sources**

[**https://github.com/nltk/nltk**](https://github.com/nltk/nltk)

[**https://aclanthology.org/P04-3031.pdf**](https://aclanthology.org/P04-3031.pdf)

[**https://spacy.io/usage/spacy-101**](https://spacy.io/usage/spacy-101)

[**https://www.dsstream.com/post/the-grand-tour-of-nlp-spacy-vs-nltk#:~:text=Model%20Size%20Comparison,-%E2%80%A2%20spaCy:&text=The%20difference%20in%20memory%20usage,specific%20tasks%2C%20optimized%20for%20efficiency.**](https://www.dsstream.com/post/the-grand-tour-of-nlp-spacy-vs-nltk#:~:text=Model%20Size%20Comparison,-%E2%80%A2%20spaCy:&text=The%20difference%20in%20memory%20usage,specific%20tasks%2C%20optimized%20for%20efficiency.)

[**https://www.seaflux.tech/blogs/nltk-vs-spacy-nlp-libraries-comparison**](https://www.seaflux.tech/blogs/nltk-vs-spacy-nlp-libraries-comparison)

[**https://www.cybrosys.com/blog/what-are-the-key-features-of-the-natural-language-toolkit-nltk-in-python**](https://www.cybrosys.com/blog/what-are-the-key-features-of-the-natural-language-toolkit-nltk-in-python)

[**https://www.cybrosys.com/blog/what-is-spacy-and-its-key-features**](https://www.cybrosys.com/blog/what-is-spacy-and-its-key-features)