

Human Language Data Processing using NLTK

Venkata Mahesh Babu Batta

<https://orcid.org/0000-0002-1029-6402>

M.Tech, Department of CSE

University College of Engineering, Osmania University, Hyderabad, Telangana, India

Abstract: *Natural Language Toolkit (NLTK) is a comprehensive Python library designed to facilitate the exploration, analysis, and processing of human language data. With its extensive collection of tools, NLTK provides researchers, developers, and educators with a powerful platform for tasks ranging from basic text processing to advanced natural language understanding and machine learning. The toolkit includes modules for tokenization, stemming, lemmatization, part-of-speech tagging, named entity recognition, syntactic parsing, semantic analysis, and more. Furthermore, NLTK offers access to numerous linguistic resources such as corpora, lexicons, and treebanks, making it an invaluable resource for both learning and research in the field of natural language processing (NLP). NLTK serves as an indispensable tool for unlocking the complexities of human language*

Keywords: Natural Language Processing(NLP), Natural Language Toolkit(NLTK), Python

REFERENCES

- [1]. Bird, Steven, Edward Loper, and Ewan Klein. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.* O'Reilly Media, 2009.
- [2]. Manning, Christopher D., and Hinrich Schütze. *Foundations of Statistical Natural Language Processing.* MIT Press, 1999.
- [3]. NLTK Project. Natural Language Toolkit (NLTK) Documentation. Available online: <https://www.nltk.org/>.
- [4]. NLTK Project. Natural Language Toolkit (NLTK) GitHub Repository. Available online: <https://github.com/nltk/nltk>.
- [5]. WordNet: A Lexical Database for English. Available online: <https://wordnet.princeton.edu/>.