

Sample PDF for Text Extraction

Section 1: Introduction

Natural Language Processing (NLP) is a field of Artificial Intelligence that focuses on the interaction between computers and humans through natural language. It involves reading, understanding, and deriving meaning from human language.

Section 2: Basic Concepts

Tokenization: The process of splitting text into individual words or phrases.

Stemming: Reducing words to their root forms.

Lemmatization: Grouping together different forms of a word to analyze them as a single item.

Section 3: Conclusion

NLP is a powerful tool that allows machines to understand and respond to human language effectively. Its applications range from sentiment analysis to machine translation.