

Chunking is a natural language processing (NLP) technique used to group words or tokens into meaningful phrases or chunks.

It often involves dividing a text into syntactically correlated parts such as noun phrases (NPs), verb phrases (VPs), and prepositional phrases (PPs).

Unlike full parsing, which provides complete syntactic trees, chunking identifies flat structures without deep hierarchical relations.

One of the most common applications of chunking is in shallow parsing, where the main constituents of a sentence are extracted.

For example, in the sentence "The quick brown fox jumps over the lazy dog," chunking would identify "The quick brown fox" as a noun phrase.

Chunking typically requires a part-of-speech (POS) tagged input so that patterns like "DT JJ NN" (determiner, adjective, noun) can be recognized.

Machine learning algorithms such as Conditional Random Fields (CRF) and Hidden Markov Models (HMM) are often used to implement chunkers.

In rule-based chunking, regular expressions are used over POS tags to define chunk boundaries.

Chunking helps in named entity recognition (NER), where named entities like person names, locations, and organizations are identified as chunks.

Chunking is a technique used to break down unstructured text into meaningful units, improving the performance of NLP tasks.

It helps simplify language processing by converting sentences into semantically rich segments.