

Natural Language Processing (NLP) involves several key concepts and techniques:

1. **Tokenization:** Splitting text into individual words or tokens. For example, "I love NLP" becomes ["I", "love", "NLP"].
2. **Lemmatization:** Reducing words to their base or root form, ensuring the word represents a single item. For example, "running" becomes "run".
3. **Stemming:** Similar to lemmatization but less precise, it cuts off word endings to find the root. For example, "running" becomes "runn".
4. **Part-of-Speech Tagging (POS):** Identifying and labeling each word's part of speech (noun, verb, adjective, etc.) in a sentence.
5. **Named Entity Recognition (NER):** Identifying and classifying proper names in text into predefined categories such as person names, organizations, locations, etc.
6. **Parsing:** Analyzing the grammatical structure of a sentence to identify relationships between words, such as dependencies or constituency.
7. **Word Embeddings:** Representing words as vectors in a high-dimensional space to capture semantic meaning, such as Word2Vec or GloVe.
8. **Bag of Words (BoW):** Representing text by the frequency of words, ignoring grammar and word order.
9. **TF-IDF (Term Frequency-Inverse Document Frequency):** A statistical measure to evaluate the importance of a word in a document relative to a collection of documents.
10. **Stop Words Removal:** Removing common words (like "the", "is", "in") that are usually less informative for text analysis.

Understanding these concepts is foundational for applying NLP techniques effectively.