Common Natural Language Processing (NLP) terms along with their definitions:

Tokenization:
The process of breaking text into individual units such as words or phrases, known as tokens.

Part-of-Speech (POS) Tagging:
Definition: Assigning a grammatical category (e.g., noun, verb, adjective) to each word in a sentence.

Named Entity Recognition (NER):
Identifying and classifying entities (such as names of people, organizations, locations) in text.

Stemming:
Reducing words to their base or root form, often by removing suffixes, to simplify analysis.

Lemmatization:
The process of reducing a word to its base or root form while considering the context and meaning.

Stop Words:
Common words (e.g., "the," "and," "is") that are often removed from text during preprocessing as they carry little semantic meaning.

Corpus:
A collection of text documents used for training and testing NLP models.

TF-IDF (Term Frequency-Inverse Document Frequency):
A numerical statistic used to reflect the importance of a word in a document relative to a collection of documents, considering both its frequency and rarity.

Word Embedding:
Representing words as vectors in a multi-dimensional space, capturing semantic relationships between words.

N-gram:
A contiguous sequence of n items (words or characters) in a given text or speech.

Syntax:
The arrangement of words and phrases to create well-formed sentences in a language.
Semantic Analysis:

Understanding the meaning of words, phrases, and sentences in context.

Parsing:
Analyzing the grammatical structure of a sentence to determine the relationships between its components.

Text Classification:
Assigning predefined categories or labels to text based on its content.

Sentiment Analysis:
Determining the emotional tone or sentiment expressed in a piece of text, often categorized as positive, negative, or neutral.

Machine Translation:
The automatic translation of text or speech from one language to another using computational methods.

Dependency Parsing:
Identifying the grammatical relationships between words in a sentence and representing them as a dependency tree.

Bag of Words (BoW):
A simple representation of text that counts the frequency of words in a document without considering the order.

Recurrent Neural Network (RNN):
A type of neural network architecture designed for processing sequential data, often used in NLP tasks.

Attention Mechanism:
A mechanism in neural networks that allows the model to focus on specific parts of the input sequence when making predictions, improving performance in tasks like machine translation.