**ACTIVITY #1**

**Type of activity: Matching Activity**

1. **Warm up activity: Match key terms and definitions.**

Match the Term with its Definition:

| **Keywords** | **Answer** | **Definitions** |
| --- | --- | --- |
| 1. Lemmatization |  | A. Assigning category tags to tokenized parts of a sentence, such as identifying words as nouns, verbs, adjectives, etc. |
| 2. Parts-of-speech (POS) Tagging |  | B. A collection of texts used for statistical linguistic analysis and hypothesis testing in linguistics and NLP. |
| 3. Corpus |  | C. The process of eliminating affixes (suffixes, prefixes, infixes, circumfixes) from a word to obtain a word stem. |
| 4. Normalization |  | D. Concerned with the interaction between natural human languages and computing devices, falling within the realms of computational linguistics, computer science, and artificial intelligence. |
| 5. Stop Words |  | E. An early step in the NLP process, splitting longer strings of text into smaller pieces or tokens, such as sentences into words. |
| 6. Bag of Words |  | F. The process of building a statistical language model to provide an estimate of natural language, assigning probabilities to sequences of input words. |
| 7. Tokenization |  | G. Words filtered out before further processing, as they contribute little to the overall meaning of a passage. |
| 8. Natural Language Processing (NLP) |  | H. A representation model simplifying text contents, focusing on the number of occurrences of words within the text without considering grammar and word order |
| 9. Statistical Language Modeling |  | I. Related to stemming, capturing canonical forms based on a word's lemma, providing the citation form. |
| 10. Stemming |  | J. Tasks meant to put all text on a level playing field before further processing, including converting text to the same case, removing punctuation, and expanding contractions. |