**ACTIVITY #3**

1. **Fill in the blanks with the appropriate NLP key terms.**

1. \_\_\_\_\_\_\_\_\_\_\_\_ is a discipline concerned with the interaction between natural human languages and computing devices, falling within the realms of computational linguistics, computer science, and artificial intelligence.

2. \_\_\_\_\_\_\_\_\_\_\_\_ is an early step in the NLP process that involves splitting longer strings of text into smaller pieces or tokens, which can be sentences or words.

3. Before further processing, text needs to be \_\_\_\_\_\_\_\_\_\_\_\_, which involves tasks to put all text on a level playing field, such as converting all text to the same case, removing punctuation, and expanding contractions.

4. \_\_\_\_\_\_\_\_\_\_\_\_ is the process of eliminating affixes from a word to obtain its stem, while \_\_\_\_\_\_\_\_\_\_\_\_ is related but captures canonical forms based on a word's lemma.

5. In linguistics and NLP, a \_\_\_\_\_\_\_\_\_\_\_\_ refers to a collection of texts used for statistical linguistic analysis and hypothesis testing.