

# Natural Language Processing and Cloud Computing

1. What is tokenization?
2. What is a bag-of-words model? Mention one advantage and limitation.
3. Which is more time-efficient - stemming or lemmatization?
4. Mention all the steps that need to be taken to compute text similarity from scratch.
5. In TF-IDF, why do we normalize the term-frequency (TF) component?
6. Synonymy and Polysemy are two concepts that may affect NLP. Briefly describe these concepts along with an example for each and mention why they may cause issues in NLP.
7. What is Amazon IAM. Why is it important?
8. How do you plan to use cloud computing for your project (if applicable)?