

Assignment

Q1. What is Text similarity ? Explain cosine similarity and Jaccard similarity in detail with examples.

Q2. What is TF-IDF ? Calculate TF-IDF by giving an example.

Q3. Explain Term Document Incidence Matrix and Inverted Index by giving example.

Q4. What is n-gram ?. Write a python program for n-gram.

Q5. Explain the concept of bag of words and POS tagging.

Q6. Discuss different types of parsing techniques in detail by giving example.

Q7. Write a python program for word tokenization and sentence tokenization.

Q8. Explain the concept of phoneme , grapheme , morpheme .

Q9. What are stop words ? Write a python program to list all stop words.

Q10. Write a python program for stemming and lemmatization using nltk.