

B.Tech Computer Engineering 8th Semester 2021
Natural Language Processing and Information Extraction [CEN-807]

Sessional 2

Max Time : 1 hr

Max Marks: 15

Note: Attempt all questions.

1. What are word vectors? Illustrate with the help of a suitable diagram. Define the cosine similarity between two word vectors. [9]
2. A *char-context matrix* (also called character co-occurrence matrix) is a char x char matrix which captures the number of times a character appears in the context of another character. Choose your full name as the corpus and consider the context of window +-4, i.e. 4 characters left and 4 characters right of any character. Generate the simple char-context matrix for the corpus. [6]

Note: If the corpus of any two students is found to be the same, it will be deemed as copied and therefore will be liable for deduction of marks.