UPES

End Semester Examination, May 2024

Course: Natural Language Processing   Semester: 6th
Program: B.Tech CSE(All)   Time : 03 hrs.
Course Code: CSEG 3043P   Max. Marks: 100

Instructions:

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	
Q1	Describe five recent Platforms/applications of Question Answering.	4	CO5
Q2	Explain the applications of Natural Language Processing?	4	CO1
Q3	Explain the terms morphological analysis, pragmatic analysis, semantic analysis.	4	CO4
Q4	Explain with an example. How cosine similarity is used in recommendation systems	4	CO2
Q5	What is the underflow problem and how Laplace smoothing is used to overcome this.	4	CO3

SECTION B
(4Qx10M= 40 Marks)

Q6	Explain the approaches to natural language processing with examples.	10	CO1
Q7	What is Machine Translation and What are the key approaches to Machine Translation?	10	CO5
Q8	Explain the problem of Word Sense Disambiguation in NLP? Discuss Knowledge based method to remove the Word Sense ambiguity?	10	CO3
Q9	What is the Information retrieval system. Explain the diagram of classical search model. OR Explain the concept of Term Document Indices Matrix with the help of example in context of IR Systems.	10	CO4

	SECTION-C **(2Qx20M=40 Marks)**		
Q10	Explain PCFG, How is the probability of every rule added in PCFG. Consider a sentence "**The Flight includes a Meal**" and the PCFG rules as: 1. S → NP VP (0.8) 2. NP →Det N (0.3) 3. VP →V NP (0.2) 4. V→ includes (0.05) 5. Det → a (0.4) 6. Det → the (0.4) 7. N → Meal (0.01) 8. N → Flight (0.02) Calculate the highest probability parse tree using CKY Algorithm.	**20**	**CO3**
Q11	Explain Viterbi algorithm, how this algorithm can be used to calculate part of speech tag to sentence given below: **"Justin will spot Will."** Training data: Martin Justin can watch Will. Spot Will watch Martin. will Justin spot Martin? Martin will pat spot. Assume three part of speech tags are used: Noun, modal, verb, OR Document 1 - data science is best course. Document 2 - nlp is the best data science course. Document 3 - data in nlp is the text. Document 4 - science is good. Using the above documents as corpus. Build the bigram probabilities chart and write calculation as well. Using the bigram model which word ('is' or 'course') should follow the sentence… "learn data science…"? Show the calculations.	**20**	**CO2**