BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)

CLASS: BTech **SEMESTER : V**
BRANCH: CS/IT **SESSION : MO/2023**

SUBJECT: IT351 NATURAL LANGUAGE PROCESSING

TIME: 3 Hours **FULL MARKS: 50**

INSTRUCTIONS:
1. The question paper contains 5 questions each of 10 marks and total 50 marks.
2. Attempt all questions.
3. The missing data, if any, may be assumed suitably.
***4*. Before attempting the question paper, be sure that you have got the correct question paper.**
5. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.

			CO
Q.1(a)	Discuss the importance of NLP in various domains. Provide examples of how NLP is used in fields such as healthcare, finance, and customer service.	[5]	1,1,1
Q.1(b)	Explain the role of spelling errors in NLP and their impact on language processing tasks. How can NLP systems handle and correct spelling errors?	[5]	1,2,1
Q.2(a)	Consider a bigram (2-gram) language model trained on a small corpus of text: i) Given the following bigram counts: • count("the quick")=5 • count("quick brown")=3 • count("brown fox")=2 • count("fox jumps")=4 • count("jumps over")=3 • count("over the")=6 • count("the lazy")=2 • count("lazy dog")=4 ii) Calculate the bigram probabilities using maximum likelihood estimation (MLE) for each bigram. iii) Assume you encounter the bigram "the quick" in a test set, but it was not present in the training data. Apply add-one (Laplace) smoothing and calculate the smoothed probability. iv) Compute the perplexity of the language model for the test set containing the bigrams "the quick brown fox jumps over the lazy dog."	[5]	2,3,2
Q.2(b)	Discuss the concept of a neural language model. How does it differ from traditional N-gram models, and what advantages does it offer in capturing complex language patterns?	[5]	2,2,2
Q.3(a)	Consider a simple language with three parts of speech: Noun (N), Verb (V), and Adjective (Adj). We have the following transition probabilities and emission probabilities for the HMM: Transition probabilities: P(N | start) = 0.4 P(V | start) = 0.3 P(Adj | start) = 0.3 P(N | N) = 0.2 P(V | N) = 0.6 P(Adj | N) = 0.2 P(N | V) = 0.3 P(V | V) = 0.4 P(Adj | V) = 0.3 P(N | Adj) = 0.4 P(V | Adj) = 0.1 P(Adj | Adj) = 0.5	[5]	3,3,3

PTO

Emission probabilities:

P("book" | N) = 0.7
P("book" | V) = 0.2
P("book" | Adj) = 0.1
P("read" | N) = 0.1
P("read" | V) = 0.6
P("read" | Adj) = 0.3
P("interesting" | N) = 0.3
P("interesting" | V) = 0.1
P("interesting" | Adj) = 0.6

Suppose we have the observation sequence "book read interesting." Using the Viterbi algorithm, calculate the most likely sequence of parts of speech for this observation sequence and the corresponding probability of the sequence.

Q.3(b) Consider a small corpus with the following sentences and their corresponding POS tags: [5] 3,3,3

"The cat is on the mat." (DET NOUN VERB PREP DET NOUN PUNCT)
"A quick brown fox jumps over the lazy dog." (DET ADJ ADJ NOUN VERB PREP DET ADJ NOUN PUNCT)
Build a Maximum Entropy Model for POS tagging using the following features:

Assume you have annotated data for training, and the model has been trained. Now, given the test sentence "A cat jumps," apply the Maximum Entropy Model to predict the POS tags for each word.

Q.4(a) Consider the following context-free grammar (CFG): [5] 4,4,4

S -> NP VP
NP -> Det N
VP -> V NP
Det -> 'the' | 'a'
N -> 'cat' | 'dog'
V -> 'chased' | 'caught'

Apply the CKY parsing algorithm to parse the sentence "the cat chased the dog."

Q.4(b) Consider the following Probabilistic Context-Free Grammar (PCFG): [5] 4,4,4

S -> NP VP [0.6]
NP -> Det N [0.4]
VP -> V NP [0.5]
Det -> 'the' [0.8]
Det -> 'a' [0.2]
N -> 'cat' [0.3]
N -> 'dog' [0.7]
V -> 'chased' [0.6]
V -> 'caught' [0.4]

Given the PCFG rules with their probabilities, calculate the probability of generating the sentence "the cat chased a dog."

Q.5(a) Consider a vector space model for words where words are represented as dense vectors. The vectors for three words, 'cat,' 'dog,' and 'bird,' are given below: [5] 5,3,3

'cat': [0.5, 0.3, 0.8]
'dog': [0.7, 0.6, 0.2]
'bird': [0.1, 0.9, 0.4]

Calculate the cosine similarity between the vectors for 'cat' and 'dog.'
Determine which pair of words, out of 'cat' and 'dog,' 'cat' and 'bird,' and 'dog' and 'bird,' is the most similar based on cosine similarity.
Suppose you have a new word 'kitten' with the vector representation [0.4, 0.2, 0.7]. Calculate the cosine similarity between 'kitten' and each of the existing words ('cat,' 'dog,' 'bird').

Q.5(b) What is Latent Semantic Analysis in NLP? Write its advantages and disadvantages [5] 5,2,5