U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

February / March 2023 Semester End Main Examinations

Programme: B.E.
Branch: Artificial Intelligence and Machine Learning
Course Code: 22AM5PENLP
Course: Natural Language Processing

Semester:V
Duration: 3 hrs.
Max Marks: 100
Date: 09.03.2023

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

UNIT - I

1 a) Discuss the challenges in NLP. justify "why NLP is hard?" by analysing all possible meanings for the sentence "I made her a duck". **10**

b) Explain bigram language model and chain rule. Consider the following toy example:
Training corpus:
<s>I am from Vellore </s>
<s>I am a teacher </s>
<s>students are good and are from various cities</s>
<s>students from Vellore do engineering</s>

Calculate the probability of the given Test data below:
<s>students are from Vellore </s> **10**

UNIT - II

2 a) Write the Minimum Distance algorithm. Analyze whether "Duck" is closer to "Ducks" or to "Dust" by hand computing the minimum edit distance cost using the algorithm. **10**

b) Construct the parse tree for the sentences/phrases using the grammar below.
S -> NP VP
... PP -> P NP
... NP -> Det N | Det N PP | 'I'
... VP -> V NP | VP PP
... Det -> 'an' | 'my'
... N -> 'elephant' | 'pajamas'
... V -> 'shot'
... P -> 'in'

Check the given sentence "I shot an elephant in my pajamas" is ambiguous. Justify your answer. **5**

c) Define Hidden Markov Model (HMM). Derive its components with assumptions. **5**

OR

3   a)   Consider the sentences in dataset given below:
         Mary Jane can see will
         spot will see Mary
         Will Jane spot Mary?
         Mary will pat Spot
         Tag the sentence "Will jane spot mary?". Find the probabilities of all the
         combinations. **10**

    b)   List and explain grammar rules with respect to English helps in constructing
         phrase structures. **10**

UNIT - III

4   a)   With the help of a diagram explain the working of transition based dependency
         parsing. Show how the word list "[book, me, the, morning, flight]" would be
         parsed using transition based dependency parsing. **10**

    b)   Explain the following with suitable examples:
         1. Vectors and documents: term-document matrix
         2. Words as vectors: document dimensions
         3. Words as vectors: word dimensions **10**

OR

5   a)   Derive Word2vec skip gram model classifier and learning embeddings with
         suitable mathematical equations. **10**

    b)   Elucidate CYK algorithm. Consider the given CFG, parse the sentence "A pilot
         likes flying planes" using CYK algorithm.
         S -> NP VP
         VP -> VBG NNS
         VP -> VBZ VP
         VP -> VBZ NP
         NP -> DT NN
         NP-> JJ NNS
         DT -> a
         NN -> pilot
         VBZ -> likes
         VBG -> flying
         JJ -> flying
         NNS -> planes **10**

UNIT -IV

6   a)   With the help of a diagram explain key terms of co reference resolution. **6**

    b)   Assume a collection of related documents contain 10,000 documents. If 100
         documents out of 10,000 documents contain the term t. Imagine the term
         appears 20 times in a document that contains a total of 100 words. Calculate tf-
         idf. **6**

    c)   Infer and explain the steps involved in unsupervised approach to word sense
         disambiguation. **8**

UNIT -V

7   a)   Define machine translation. Discuss why machine translation is hard? **10**

    b)   Explain the following:
         i.   Information extraction
         ii.   Automatic text categorisation **10**

*****