U.S.N.										

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

Autonomous Institute Affiliated to VTU

August 2024 Supplementary Examinations

Programme: B.E. **Semester: V**
Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**
Course Code: 22AM5PENLP **Max Marks: 100**
Course: Natural Language Processing

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> **10**
<s>students from Vellore do engineering</s>

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

UNIT - II

2 a) Justify the need of regular expressions in NLP. Write regular expressions for the following languages.
1. the set of all alphabetic strings;
2. the set of all lower case alphabetic strings ending in a b;
3. the set of all strings from the alphabet a,b such that each a is immediately preceded by and immediately followed by a b; **10**
4. Email address
5. date format dd-mm-yyyy

b) Describe the working of porter stemmer. With suitable example explain following text normalization processes:
1. Tokenizing (segmenting) words
2. Normalizing word formats **5**
3. Segmenting sentences

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 **10**

Tag the sentence "Will jane spot mary?". Find the probabilities of all the combinations.

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

UNIT - III

4 a) Describe Ambiguous Grammar. Identify and justify if the given grammar is ambiguous with the phrase: "I ate a burger in the evening". Assume appropriate POS tags for the words, if missing. **10**

Grammar	Lexicon
$S \rightarrow NP\ VP$	$Det \rightarrow that \mid this \mid the \mid a$
$S \rightarrow Aux\ NP\ VP$	$Noun \rightarrow book \mid flight \mid meal \mid money$
$S \rightarrow VP$	$Verb \rightarrow book \mid include \mid prefer$
$NP \rightarrow Pronoun$	$Pronoun \rightarrow I \mid she \mid me$
$NP \rightarrow Proper\text{-}Noun$	$Proper\text{-}Noun \rightarrow Houston \mid NWA$
$NP \rightarrow Det\ Nominal$	$Aux \rightarrow does$
$Nominal \rightarrow Noun$	$Preposition \rightarrow from \mid to \mid on \mid near \mid through$
$Nominal \rightarrow Nominal\ Noun$	
$Nominal \rightarrow Nominal\ PP$	
$VP \rightarrow Verb$	
$VP \rightarrow Verb\ NP$	
$VP \rightarrow Verb\ NP\ PP$	
$VP \rightarrow Verb\ PP$	
$VP \rightarrow VP\ PP$	
$PP \rightarrow Preposition\ NP$	

b) Show how the Cosine computes which of the words 'college' or 'natural' is closer in meaning to 'campus'. The raw counts are given in the table below for your reference. **06**

	students	language	canteen
college	442	8	2
natural	5	1683	1670
campus	5	3982	3325

c) Describe Applications of TF-IDF vector model. **04**

OR

5 a) Discuss the semantic properties of embeddings also list out the possible ways to visualize embeddings. **10**

b) Describe the characteristics of dependency parsing. Parse the below sentences using dependency parsing. **10**

1. Deemed universities charge huge fees.
2. I prefer morning flights to Denver
3. I ate the fish with a fork

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)	Elucidate Information Retrieval and Information Extraction with neat diagrams for each.	**10**
	b)	Alex wants to automate the process of summarization for the text documents that he has. Analyze the various approaches that he could follow to summarize the text.	**10**

****