U.S.N.										

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

Autonomous Institute Affiliated to VTU

September / October 2023 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 Date: 15.09.2023

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 different phases of NLP by giving suitable example at each phase. **10**

b) Consider the following toy example:
Training data:
<s> I am Sam </s>
<s> Sam I am </s>
<s> Sam I like </s>
<s> Sam I do like </s>
<s> do I like Sam </s>
Assume that we use a bigram language model based on the above training data.
1. What is the most probable next word predicted by the model for the following word sequences?
(1) <s> Sam . . .
(2) <s> Sam I do . . .
(3) <s> Sam I am Sam . . .
(4) <s> do I like . . .
2. Which of the following sentences is better, i.e., gets a higher probability with this model?
(5) <s> Sam I do I like </s>
(6) <s> Sam I am </s>
(7) <s> I do like Sam I am </s> **10**

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;
4. Email address
5. date format dd-mm-yyyy **10**

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

10

OR

3 a) Define HMM. Derive the equations for the following:
    i.   The components of an HMM tagger
    ii.   HMM tagging as decoding

10

b) Explain how the CCG limitations can be overcome by operations like compose and type raising. Apply the same for the phrase/sentences below:
    i.   The company that IBM bought
    ii.   I enjoy books

10

UNIT - III

4 a) Consider the grammar below,
S -> NP VP
S -> NN VP
VP -> VB NP
VP -> VB NN
NP -> DT NN
PP -> IN NP
PP -> IN NN
NP -> NP PP
NP -> NN PP
VP -> VP PP

IN -> with
NN -> dog
NN -> cat
VB -> ate
NN -> mouse
DT -> the

Parse the sentence "The cat ate the dog with the cat" using cky algorithm and write the algorithm.

10

b) Explain the following:
1. Cosine for measuring similarity
2. TF-IDF: Weighing terms in the vector
3. Pointwise Mutual Information (PMI)

10

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

10

UNIT - IV

6 a) Illustrate the architectures for Coreference Algorithms.

10

b) Justify the need of Word Sense Disambiguation. Elucidate WSD algorithms.

10

UNIT - V

7 a) Elucidate the Challenges in Information Retrieval. Describe the following Information retrieval tasks:
   1. Documents as Bags-of-Words **10**
   2. The Vector Space Model

   b) Justify the need of automatic text summarization and discuss the approaches to Automatic Summarization. **10**

******