Avinashilingam Institute for Home Science and Higher Education for Women

(Deemed to be University under Category 'A' by MHRD, Estd. u/s 3 of UGC Act 1956)
Re-accredited with 'A+' Grade by NAAC.Recognised by UGC under Section 12 B
Coimbatore - 641 043, Tamil Nadu, India

Bachelor of Vocation Degree Examination – June 2021
IV Semester

Class : II B.Voc **Time: 3 Hrs**
Major : Artificial Intelligence and Machine Learning **Max. Marks: 100**

19VAIC13 Natural Language Processing

Part A **10 x 1 = 10**

Choose the Correct Answer

1. The process used for reducing the inflected words to their root form is _____ CO2K1
   a. Rooting b. Stemming c. Text-Proofing d. Rooting and Stemming

2. Which of the following doesn't require the usage of Natural Language Processing algorithms? CO1K2
   a. Classifying spam emails from good ones
   b. Classifying images of scanned documents as 'hand-written' or 'printed' documents
   c. Automatically generating captions for images
   d. Building sentiment analyzer for tweets on Twitter

3. Ambiguity can appear in which of the following steps/tasks? CO1 K1
   a. Tokenization b. Language Understanding
   c. Sentence Segmentation d. All of these

4. Word segmentation is mostly used when _____ CO5 K2
   a. Hyphens are present b. No space between words
   c. Long sentences d. Multiple alphabets intermingled

5. A 4-gram model is a _____ order Markov Model CO3 K2
   a. Constant b. Five c. Four d. Three

6. Consider the CFG given below: CO3 K3
   S->ASA | aB
   A->B |S
   B->b |ε
   How many non-terminals need to be added to convert the above grammar into CNF?
   a. 3 b. 4 c. 2 d. 1

7. Parts-of-Speech tagging determines _____ CO2 K1
   a. part-of-speech for each word dynamically as per meaning of the sentence
   b. part-of-speech for each word dynamically as per sentence structure
   c. all part-of-speech for a specific word given as input
   d. all of the mentioned

8. One of the main challenge/s of NLP is _____ CO1 K2
   a. Handling Ambiguity of Sentences b. Handling Tokenization
   c. Handling POS-Tagging d. All of the mentioned

9. Case folding is used for _____ CO5 K1
   a. Tokenization b. Normalization c. Stemming d. Lemmatization

10. What is the term frequency of a term which is used a maximum number of times in that document? CO4 K2
    a. t6 – 2/5 b. t4 – 2/6 c. t3 – 3/6 d. t1 – 2/6

Part B **5 x 6 = 30**

Answer the following

Answer should not exceed 400 words or two pages

11.a. Discuss the role of regular expression and automata in building an NLP system. CO1 K2

(or)

11.b. Explain the procedure involved in detection and correction of spelling errors in english morphology. CO1 K2

12.a. Discuss the smoothing techniques used for handling data sparseness issue in n-gram model. CO2 K2

(or)

12.b. Discuss stochastic part-of-speech tagging technique with an example. CO2 K2

13.a. Explain the process involved in Shallow Parsing with suitable example. CO3 K3

(or)

13.b. Discuss the key characteristics of Dependency Grammars that play important role in language processing systems. CO3 K3

14.a. Distinguish between semantics, pragmatics and discourse. CO4 K4

(or)

14.b Illustrate with an example the working of nearest neighbour algorithm for Word Sense Disambiguation (WSD). CO4 K4

15.a. Discuss the features and functionalities of WordNet, FrameNet and Stemmer. CO5 K3

(or)

15.b. Explain the tasks involved in Coreference resolution that influences understanding of natural language by computer systems. CO5 K3

Part C **5 x 12 = 60**

Answer the following

Answer should not exceed 800 words or four pages

16.a. Design a finite-state automaton that facilitates checking of account balance and withdrawal of money at an automated teller machine. CO1 K3

(or)

16.b. Discuss the key features and characteristics of Statistical Language Modelling with suitable examples. CO1 K3

17.a. Discuss the usage of Hidden Markov Model for Part-of-Speech Tagging. Illustrate with a use-case the step-by-step process involved in the model deployment and execution. CO2 K3

(or)

17.b. Consider the following corpus: CO2 K4

<s> I am Sam </s>
<s> Sam I am </s>
<s> I am Sam </s>
<s> I do not like green eggs and Sam </s>

If linear interpolation smoothing is used between a maximum-likelihood bigrammodel and a maximum-likelihood unigram model with λ_1 = 1/2 and λ_2 =1/2, what is P(Sam|am)? Include <s> and </s> in the counts just like any other token.

18.a. Derive an algorithm for converting an arbitrary Context-Free Grammar into Chomsky Normal Form. CO3K6

(or)

18.b. Draw tree structures for the following ATIS sentences: CO3 K5

i) I would like to fly on Indian Airlines
ii) Please repeat that
iii) What is the fare from Delhi to Chennai?

19.a.Discuss various atomic elements of First-Order Logic (FOL). For the given sentences, perform the FOL translations: CO4 K3

i) Vegetarians do not eat meat.
ii) Not all vegetarians eat eggs.

(or)

19.b. Discuss the usage of Thesaurus and distributional methods for computing word similarity. Illustrate with suitable use-cases. CO4 K3

20.a..Deliberate the usage of Penn Treebank and Propbank in facilitating effective sentence annotation for NLP tasks CO5K4

(or)

20. b)Discuss the usage of Lemmatizer, Tagger and Brown Corpus in facilitating effective sentence annotation for NLP tasks. CO5K3

****