VIT®
Vellore Institute of Technology

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

Winter 2024-2025   ***Continuous Assessment Test – II***   *Duration: 90 Mins*

Program: B.Tech   ***Course Code: CSI4001***   <u>*Slot: C1+TC1*</u>

Course Name: Natural Language processing & Computational Linguistics

Faculty: Dr.Biji & Dr.Sharmila Banu K

ANSWER ALL QUESTIONS   (5*10 = 50 Marks)

1. Consider the following Corpus:

   This is the Transformer model that Mr.Toby built.

   This is the Chatbot that Mr.Harry designed.

   This is the Entropy Model that Ms. Granger designed for Part-of-Speech Tagging.

   Now, **what will be your approach** to calculate the probability of the **following sentences** based on the given corpus?

   (a)   This is the Program.

   (b)   This is Ms. Granger

   Solve:

   (i) Identify the specific inadequacies in computing the probability of sentences (a) and (b).   (5 Marks, CO4, BL3)

   (ii) Design a solution that addresses the inadequacies and compute the probabilities of the sentences (a) and (b)?   (5 Marks, CO4, BL3)

---

2.   Answer the following precisely: (5 X 2.5 = 10 Marks, CO4, CO5, CO6, BL2)

a. What is the role of ***Argmax*** in N-Gram Hidden Markov Model?

b. Present the Chomsky's notation for E-Insertion Rule in English and explain the symbols.

c. Why is ENGTWOL challenging to use for Part-of-Speech Tagging? Give your argument based on its working and considering the text data sourced from Social Media.

d. Design the Mathematical Model for 3-Gram Language Modeling.

---

3. The transducer given below shows the nominal inflection for English. You can see that all plural forms are not captured correctly in this transducer. Design an automaton to fix this plural form error and mention where it has to be included in the given transducer. (10 Marks, CO6, BL2)
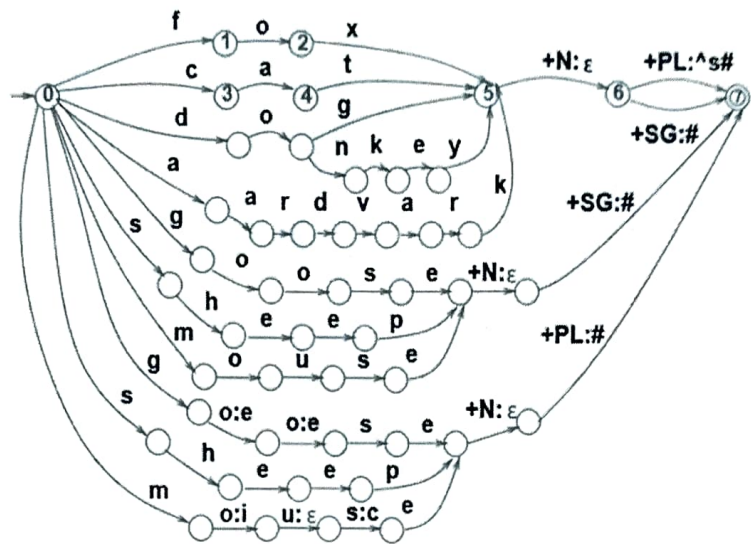
4. Using Viterbi Approximation, perform the Part-of-Tagging for the given sentence. Consider the Bi-gram tagging model to build your solution.

   She danced well (10 Marks, CO5, BL3)

Observed Data for your reference:

Starting Probability	Verb	Pronoun	Adjective
	0.4	0.5	0.1

Emission Probability		She	danced	well
	Verb	0.3	0.5	0.5
	Pronoun	0.5	0.1	0.02
	Adjective	0.3	0.1	0.6

Transition Matrix		Verb	Pronoun	Adjective
	Verb	0.3	0.4	0.3
	Pronoun	0.2	0.2	0.6
	Adjective	0.3	0.2	0.5

5. How does Maximum Entropy perform Part-of-Speech tagging? Present the Mathematical Model, explain the notations, working and its advantages over the HMM tagging? (10 Marks, CO5, BL4)

****