

PG / INTEGRATED (CBCS) ODD SEMESTER EXAMINATION, 2021
Held in April 2022

COMPUTER SCIENCE
9th Semester / 3rd Semester

COURSE NO. MCSCC - 904 / MS - 304
(Natural Language Processing)

Full Marks : 70
Pass Marks : 28

Time : 3 hours

The figures in the margin indicate full marks for the questions

(Answer any five questions, taking one from each unit)

UNIT-I

1. a) What is corpus? Give some examples of corpus. 2+1=3
- b) Why do we study NLP? What are the goals of NLP? 2+2=4
- c) Explain document similarity measure. 3
- d) An aeroplane is flying above with a probability of 5%. There is a radar which registers a blip if the plane is above with 99% accuracy. The radar also registers a blip with a 10% probability when there is no plane flying above. What is the probability that a plane is flying above given that the radar registers a blip? 4

4

2. a) Explain a spell checker in the light of Bayes rule. 5
- b) Explain argmax computation with a suitable example. 3
- c) What is total probability the theorem? Explain. 3
- d) A patient gets tested for a disease which affects 1% of the population. Suppose the test is advertised as 95% accurate. What is the probability that the patient has the disease? 3

UNIT - II

3. a) What is an N-gram language model? 3
- b) What is a top down chart para? What are its disadvantages? 4
- c) For the CFG shown below, show the complete trace of a top down bottom up chart parser for the following sentence 7
"The large can hold water"

S → NP VP	Lexicon
NP → ART ADJ N	the : ART
NP → ART N	large : ADJ
NP → ADJ N	can : N, AUX, V
VP → AUX VP	hold : N, Y
VP → VNP	water : N, V

4. a) Give the formal definition of a PCFG. What are the consumption of a PCFG model? 2+3=5
- b) Explain probability of parse tree and how it can be determined? 4
- c) For the PCFG shown below calculate the probability of the parse trees for the sentence "The painter sprayed the building with blue colour" 5

S → NP VP 1.0	DT → the 1.0
NP → DT NN 0.5	NN → pain for 0.33
NP → ADJ NN 0.4	NN → building 0.34
NP → NP PP 0.2	NN → colour 0.33
PP → P NP 1.0	VBD → sprayed 1.0
VP → VP PP 0.6	P → with 1.0
VP → VBD NP 0.4	ADJ → blue 1.0

UNIT - III

5. a) Explain the HMM model. What is a trigram Markov assumption? How is a trigram Markov assumption represented? 3+2+2=7
- b) How can we interpret POS tagging as a sequence labelling problem through argmax based computation? 3
- c) Explain lexical probability assumption with an example. 4

6. a) Explain Forward and Backward probability. Why are they needed? 4+2=6
- b) What is Bayesian Decision theory principle? What are some of the issues in Bayesian Decision theory? 1+3=4
- c) Elucidate the challenges of POS Tagging in the context of some Indian languages. 4

UNIT - IV

7. a) Explain perception learning process. 4
- b) Explain the difference between perceptron activation & sigmoid activation functions. What is max and min sigmoid function? 2+1=3
- c) Explain Gradient descent with a suitable example. Why is the movement in gradient descent always opposite to the gradient? 5+2=7
8. a) What are contour plots? Why are they needed? 2+1=3
- b) What is a partial derivative of a function? Find the partial derivatives of the function $f(x, y) = x^2 e^{xy}$. 2
- d) Explain the process of back propagation with a suitable example. 5

UNIT - V

9. a) What is machine translation? 2
- b) Discuss sentiment analysis. 6
- c) Define the following: 6
- i) Word embedding
- ii) Word 2 Vec
- iii) Glove
10. a) Explain the encoder and decoder model of machine translation. 7
- b) Discuss RNN in the context of NLP. 7