



Final Assessment Test – November 2024

Course: BCSE409L - Natural Language Processing

Class NBR(s): 1856/1861/1876/1881/1896/1899

Time: Three Hours

Slot: E2+TE2

Max. Marks: 100

- > KEEPING MOBILE PHONE/ANY ELECTRONIC GADGETS, EVEN IN 'OFF' POSITION IS TREATED AS EXAM MALPRACTICE
- > DON'T WRITE ANYTHING ON THE QUESTION PAPER

Answer ALL Questions

(10 X 10 = 100 Marks)

1.
 - a) Examine Natural Language Processing (NLP) and the various stages involved in the language processing with suitable example. [6]
 - b) Identify and analyse the type of ambiguity present in the following sentences. [4]
 - i. I heard his cell phone ring in my office.
 - ii. I like java.
2. Explain the mathematical representation of Finite State Transducer. How can Finite State Transducers (FSTs) handle ambiguity in language processing tasks? Explain with a suitable example and a neat transition diagram.
3. Let's consider the following as a training corpus:

<S>	John	Mohan	can	find	Will	</S>
	N	N	M	V	N	
<S>	Will	Mohan	spot	john	</S>	
	M	N	V	N		
<S>	John	Will	play	spot	</S>	
	N	M	V	N		
<S>	Spot	Will	find	John	</S>	
	N	M	V	N		

Create the emission and transition probability matrix. Employ the Viterbi algorithm to get the best Parts-of-speech tags and corresponding probabilities for the test sentence: <S> Will can spot John </S>.

4.
 - a) Examine the Condition Random Field (CRF) and its use in Semantic Role Labelling (SRL) in the sentence "VIT gave John a prize for his publication." [5]
 - b) Design a dependency tree with transition-based parsing for the sentence "Book me the morning flight" and justify your answer. The following are the predicted transitions for the given statement at each step: SHIFT, SHIFT, RIGHTARC, SHIFT, SHIFT, SHIFT, LEFTARC, LEFTARC, RIGHTARC, RIGHTARC. [5]

5. Apply the CYK algorithm to the input sentence "The really thick green book" and show the parse tree with the following Context-Free Grammar:

S → NP VP
NP → Det Nom
Nom → VP Nom
VP → Adv A
VP → thick green
A → thick green
Nom → book green
Adv → really
Det → The

6. Examine homonymy, hyponymy, and hypernym using appropriate examples to see how they pose challenges for Natural Language Processing and what potential solutions exist.
7. Explain Word2Vec and its two primary architectures for creating word embeddings. Explain the process of creating training datasets for the architectures using the statement "Students from VIT enrolled in computer science". Note: Assume the window size length is two.
8. Let's consider the following as a sample training corpus:

<s> I am from VIT </s>
 <s> I am a professor </s>
 <s> students are good and are from various cities </s>
 <s> students from VIT enrolled in computer science </s>

Apply the bi-gram and tri-gram models to predict the probability of the given test sentence "<s> students are from VIT </s>." Use the appropriate technique to deal with the zero-probability in the calculations. Using the perplexity measure, choose the best of these two models.

- 9.a) Consider the following traffic dataset with Attributes = {'Weather', 'Road', 'Traffic', 'Engine Problem'}. This dataset contains examples of the different conditions that are associated with accidents. Use the Laplace smoothing technique to forecast the occurrence of any event. If we give the following hypothesis for the above dataset, how likely is it that an accident will occur?

The task is to calculate the probability of an accident occurring or not. Classify the attribute 'Accident' (target variable) as Yes or No.

- i) If Road = 'good' ii) If Weather = 'Sunny'

SNo.	Weather condition	Road condition	Traffic condition	Engine problem	Accident
1	Rain	bad	high	no	yes
2	snow	average	normal	yes	yes
3	clear	bad	light	no	no
4	clear	good	light	yes	yes
5	snow	good	normal	no	no
6	rain	average	light	no	no
7	rain	good	normal	no	no
8	snow	bad	high	no	yes
9	clear	good	high	yes	no
10	clear	bad	high	yes	yes

OR

- 9.b) Analyse text summarization and its different types in natural language processing. Examine a case study and the challenges of automatic text summarization.
- 10.a) Assume you've gathered a large set of customer reviews from an e-commerce platform. Describe how you would conduct a sentiment analysis on this feedback. What are the key steps involved in the sentiment analysis process? Provide an in-depth examination of each step, including the techniques and tools used in natural language processing (NLP) to extract and analyse customer sentiment.

OR

- 10.b) Examine the significance of text processing in Natural Language Processing (NLP) through a case study of customer services to analyze customer complaints to improve service. Provide a detailed explanation of each step involved in text processing, and include diagrammatic illustrations to demonstrate how these steps contribute to effective NLP.

⇔⇔⇔F/L/TX⇔⇔⇔