VIT-AP UNIVERSITY	Final Assessment Test – Winter 2023-24 Semester - May 2024	
	Max Marks: **100**	Duration: **3 Hours**
Course Code: **CSE3015**	Course Title: **Natural Language Processing**	
Question Paper ID: **009905**	Exam Type: **CLOSED BOOK**	School: **SCOPE**
Date: **14/05/2024**	Slot: **C1+TC1**	Session: **AN**

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice.

General Instructions if any:

1. "*fx* series" - non Programmable calculator is permitted: No
2. Reference tables permitted: No

Section - 1: Answer any 10 questions. (10 × 10 = 100 Marks)

		Marks	CO	BL
Q1.	How NLP is importance in our daily life? Discuss any five real time applications of NLP with their limitations?	**10**	**1**	**1**
Q2.	Briefly explain stemming and lemmatization in NLP with Python code.	**10**	**1**	**2**
Q3.	Solve the tagging problem using a stochastic tagging-based approach. Evaluate the probability of a sequence of string "**Can Rahul write Story that** best matches the corpus tagging shown below, where N: Noun, M:Modal, and V:verb. **Training Sentences:** ***Rahul /N will/M slide/V story/N*** ***story/N will/M write/V Rahul /N*** ***will/M brown/N story/V Rahul /N***	**10**	**2**	**3**
Q4.	Find the best sequence label using the Viterbi Algorithm for the following phrase **"Book is heavy weight"** **(15M)** **Training Sentences:** The/Det book/Noun looks/V less/Ad weight/Ad The/Det light/N is/DET in/DET the/DET street/N The/Det heavy/Ad book/Noun weight/Ad	**10**	**2**	**4**
Q5.	Design the **Shift Reduce Parser** and **Recursive Descent Parser** for the following context-free language with the input string "**a man is running**". S-> NP VP NP-> Det NN Det->a/an NN->dog/man VP -> AUX VRB AUX->is/was VRB->running	**10**	**3**	**4**

Q6. Write nltk code to parse the POS tree of the following sentences using shallow parsing(chucking) technique with the given chuck pattern and draw the resultant parse tree of each input sentence. **10 3 1**

Input sentences:

"Babasaheb Ambedkar, widely recognized as the chief architect of the Indian Constitution, was affectionately referred to as Dr. B. R. Ambedkar by all."
"Rahul Dravid is one of the best batsmen".

Chuck Details:
Chunk name: Noun phrase, pattern=DET*NN+
Chunk name: Verb Phrase, pattern=AUX*VRB+
Chunk name: Preposition Phrase, pattern= PREP*ADJ*DET*NN+
Note: PREP-preposition, DET- determiner, NN- Noun, AUX- auxiliary verb, VRB-verb, ADJ-Adjective

Q7. Design a Naive Bayes model using sklearn python library from the following training dataset. And **(X={age=senior, sex=female, BP=low, cholesterol=high}) Note: play golf is class label.** **10 4 5**

Temperature	Humidity	Windy	Play Golf
Hot	High	False	No
Hot	High	True	No
Hot	High	False	Yes
Mild	High	False	Yes
Cool	Normal	False	Yes
Cool	Normal	True	No
Cool	Normal	False	Yes
Mild	High	False	No
Mild	Normal	True	Yes

Q8. Design a Decision tree classifier from the following training dataset. And also find which drug will be suggested for the given data. **(X={age=Young, sex=female, BP=low, cholesterol=high}) Note: Drug is class label.** **10 4 5**

Patient ID	Age	Sex	BP	Cholesterol	Drug
p1	Young	F	High	Normal	Drug A
p2	Young	F	High	High	Drug A
p3	Middle-age	F	Hiigh	Normal	Drug B
p4	Senior	F	Normal	Normal	Drug B
p5	Senior	M	Low	Normal	Drug B
p6	Senior	M	Low	High	Drug A
p7	Middle-age	M	Low	High	Drug B
p8	Young	F	Normal	Normal	Drug A
p9	Young	M	Low	Normal	Drug B
p10	Senior	M	Normal	Normal	Drug B

Q9.	Write a spider program to extract a list of URLs from "HTTP://books.toscrapy.com" and discuss the spider installation steps in your local system.	**10**	**5**	**5**
Q10.	What steps are required to create a web scraping program (Spider), and can you showcase the process by developing a Spider program named 'Product Details' to extract product details from the URL "https://www.snapdeal.com// CT Scan"? Ensure the utilization of CSS selectors to scrape product title, description, price, and specifications from each product on the page, and subsequently, save the extracted data in respective JSON file formats. Provided below are the CSS class details for the product details.	**10**	**5**	**5**
Q11.	How can we extract tweets from Twitter and conduct sentiment analysis based on geographical location? Discuss all the steps involved in this process.	**10**	**6**	**6**
Q12.	How can chatbots integrated with Natural Language Processing (NLP) revolutionize the medical field?	**10**	**6**	**6**

*****End of Question Paper*****

VIT-AP UNIVERSITY	Final Assessment Test – Winter 2023-24 Semester - May 2024	
	Max Marks: **100**	Duration: **3 Hours**
Course Code: **CSE3015**	Course Title: **Natural Language Processing**	
Question Paper ID: **009735**	Exam Type: **CLOSED BOOK**	School: **SCOPE**
Date: **15/05/2024**	Slot: **C2+TC2**	Session: **AN**

Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice.

General Instructions if any:

1. "*fx* series" - non Programmable calculator is permitted: No
2. Reference tables permitted: No

Section - 1: Answer any 10 questions. (10 × 10 = 100 Marks)

		Marks	CO	BL
Q1.	Explain the core concept of text pre-processing (text wrangling) in natural language processing (NLP) and discuss its different techniques. Provide a practical demonstration of text pre-processing using the Natural Language Toolkit (NLTK) with a raw text input.	**10**	**1**	**1**
Q2.	**a.**Identify the morphological type (Noun Phrase, Verb Phrase, Adjective phase) of the following sentences **(5M)** **Digital marketing has become a cornerstone for businesses, providing a cost-effective and efficient way to reach a global audience.** **b.** For the given sentence, identify whether the different meanings arise from structural Ambiguity, semantic ambiguity or pragmatic ambiguity. **(5M)** i. They listened to the music play ii. Bought one ticket iii. She made him laugh iv. He saw that she meant well	**10**	**1**	**3**
Q3.	Differentiate between Regex and N-gram tagger. Design a program for both Regex and N-gram tagger and explain the reason for similarity or difference in the output	**10**	**2**	**2**
Q4.	Design HMM tagger using required transitional and emission probabilities from the given set of training sentences. And find the POS tagging for the sentence "**Laasya will see Mahira**" using trained HMM tagger. **Training sentences:** Laasya/noun can/model see/verb flight/noun Ramu/noun will/model see/verb Laasya/noun Will/model Mahira/noun spot/verb Laasya/noun Laasya/noun will/model pat/verb Spot/noun	**10**	**3**	**4**
Q5.	Write nltk code to parse the pos tree of the following sentences using Regexp parser with given grammar and also draw the resultant parse tree of each input sentences **Input sentences:** 1. Microsoft Corporation is an American multinational corporation and technology 2. Amazon has a reputation as a disruptor of industries 3. Sachin Bansal and Binny Bansal founded Flipkart whose valuation in 2018 NP: DT? JJ* NN* P: IN V: V.* PP: P NP VP: V NP|PP	**10**	**5**	**3**

Q6. Write nltk code to parse the pos tree of the following sentences using shallow parsing (chunking) Technique with given chunk pattern and also draw the resultant parse tree of each input sentences **10 6 2**

a.India dominated international field hockey in the mid-20th century,

b.The Indian men's field hockey team won six gold medals

c. Dhyan Chand was father of Hockey

Chuck Details:

Chunk name: Noun phrase, pattern=DET*NN+

Chunk name: Verb Phrase, pattern=AUX*VRB+

Chunk name: Preposition Phrase, pattern= PREP*ADJ*DET*NN+

Note: PREP-preposition, DET- determiner, NN- Noun, AUX- auxiliary verb, VRB-

Q7. a. How does text classification play a role in spam email detection and filtering **(5M)** **10 6 2**

b.What is the purpose of sampling in NLP tasks such as language modeling or text generation **(5M)**

Q8. Design a Naive Bayes classifier from the following training dataset. And also find the class label for the given data. **(X= {Color= Red, Type= SUV and Origin=Domestic})(note: Stolen is the class label)** **10 6 2**

Example No	Color	Type	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes

Q9. a.Explain the scenario and type of data requirement to consider before writing Web Crawlers. Provide proper explanation and reasons to support the assertions. **(7M)** **10 6 2**

b.Design a snippet to illustrate bar plot and sub plot with six data/images to be plotted with 2 column and 3 rows. **(3M)**

Q10. a. Explain the Key Components of an Item Pipeline class in Scarpy **(4M)** **10 6 2**

b.How are multiple pipelines organized and executed in a Scarpy project **(3M)**

c.How can item pipelines be used to filter or drop certain items during processing **(3M)**

Q11. Outline the detailed process of information retrieval used to match documents from a corpus to a given query, explaining the essential natural language processing (NLP) steps involved. Provide a practical example using a document corpus to illustrate each step. **10 6 4**

Q12. Outline the steps for collecting applicable social media data, pre-processing it, and analyzing sentiment. Discuss potential challenges you may encounter and how you would resolve them. Explain your approach to conducting sentiment analysis using Natural Language Processing (NLP) techniques. You're working for a social media monitoring company tasked with analyzing public sentiment about a new product release. **10 6 4**

*****End of Question Paper*****

VIT-AP UNIVERSITY	Final Assessment Test – Long Summer (2023-24) - July 2024	
	Maximum Marks: 100	Duration: 3 Hours
Course Code:CSE3015	Course Title: Natural Language Processing	
Set No: 2	Exam Type : **Closed Book**	School: SCOPE
Date: 22-07-2024	Slot: E	Session: AN
Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice		
General Instructions if any: 1. **"*fx* series" - non Programmable calculator are permitted: YES** 2. **Reference tables permitted : NO**		

PART – A: Answer any TEN Questions, Each Question Carries 10 Marks (10×10=100 Marks)

1. Explain the core concept of text pre-processing (text wrangling) in natural language processing (NLP) and discuss its different techniques. Provide a practical demonstration of text pre-processing using the Natural Language Toolkit (NLTK) with a raw text input. **(10 M)**

2. a. Identify the morphological type (Noun Phrase, Verb Phrase. Adjective phase) of the following sentences **(5M)**
   Digital marketing has become a cornerstone for businesses, providing a cost-effective and efficient way to reach a global audience.
   b. Lemmatize the given sentence and determine the root words of token with their corresponding tags using NLTK library and explain each step involved **(5M)**
   The research team conducted extensive experiments, validating theoretical hypotheses and generating insightful findings

3. Design HMM tagger using required transitional and emission probabilities from the given set of training sentences. And find the POS tagging for the sentence "Laasya will see Mahira" using trained HMM tagger. **(10 M)**
   Training sentences:
   Laasya/noun can/model see/verb flight/noun
   Ramu/noun will/model see/verb Laasya/noun
   Will/model Mahira/noun spot/verb Laasya/noun
   Laasya/noun will/model pat/verb Spot/noun

4. Explain the concept of regular expression tagger in NLP. How does it use pattern matching to assign part of speech (POS) tags to tokens what are the key factors influencing the effectiveness of regular expression based POS tagging in any application. **(10 M)**

5. Explain Dependency Parsing with example. Implement a python code for Dependency parser and generate parse tree for the following Input sentence:
   (This/DT is/VBZ the/DT Indian/JJ traditional/NN dress/NN) **(10 M)**

6. What is Parsing? Differentiate Top Down and Bottom-Up Parsing with example. Create your own productions and draw the parse tree for the sentence: "I saw the boy playing in the garden". **(10 M)**

7. Outline the essential steps for developing a text classification system using Naive Bayes methods. Use your own datasets and apply the classifiers to a specific application. **(10 M)**

8. Design a Naive Bayes classifier from the following training dataset. And find the class label for the given data. (X= {Color= Red, Type= SUV, Origin=Domestic}). Note:Stolen is the class label **(10 M)**

Example No	Color	Type	Origin	Stolen
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Red	SUV	Imported	No

9. a. Explain the scenario and type of data requirement to consider before writing Web Crawlers. Provide proper explanation and reasons to support the assertions. **(7M)**

   b. Design a snippet to illustrate bar plot and sub plot with six data/images to be plotted with 2 columns and 3 rows. **(3M)**

10. Build spider program using with NLTK Libraries to perform given operations
    1. Extract title of the Web Page
    2. Extract hyperlinks
    3. Extract images
    4. Extract content/data from any site. **(10 M)**

11. What does Sentiment Analysis involve in the realm of NLP? What are the main hurdles encountered in analysing sentiment? What are some common applications of sentiment analysis beyond social media. How does sentiment analysis contribute to customer feedback analysis and product improvement **(10 M)**

12. Explain the architecture of Question and Answering in NLP with example. **(10 M)**

VIT-AP UNIVERSITY	Regular Arrear Examinations (2023-24) - July 2024	
	Maximum Marks: 100	Duration: 3 Hours
Course Code: CSE3015	Course Title: Natural Language Processing	
Set No: 2	Exam Type : Closed Book	School: Scope
Date: 06/08/2024	Slot: C	Session: AN
Keeping mobile phone/smart watch, even in 'off' position is treated as exam malpractice		
General Instructions if any: **1. "*fx* series" - non Programmable calculator are permitted : /NO** **2. Reference tables permitted : NO (if Yes, Please specify:)**		

PART – A: Answer any TEN Questions, Each Question Carries 10 Marks (10×10=100 Marks)

Q1 What is stemming in the context of natural language processing (NLP)? Describe the process of stemming. Use the given example sentences to demonstrate stemming using the NLTK library. **(10M)**

Q2 Explain the similarities and differences among Semantic Ambiguity, Syntactic Ambiguity, and Pragmatic Discourse in the context of Natural Language Processing with example. **(10M)**

Q3 This implementation demonstrates the operational process of the Brill Tagger and provides an example. **(10M)**

Q4 Find the best sequence label using Viterbi Algorithm for the following phrase **10M**

"Book is heavy weight".

Training Sentences

The/Det book/Noun looks/V less/Ad weight/Ad
The/Det light/N is/Det in/Det The/Det street/N
The/Det heavy/Ad book/Noun weight/Ad

Q5 . Explain Dependency parsing with example. Implement a python code for Dependency parser and generate parse tree for the following **10M**

Input sentence:
(The/DT fastest/JJS land/NN animal/NN in/IN the/DT world/NN is/VBZ the/DT cheetah/NN ./)

Q6 Considers the following context free grammar of English language, check whether the string "The tall man wore a hat"is acceptable by the given grammar or not using shift reduce parser. And also write the python code using Nltk library to design the given grammar and generate all possible acceptable strings. **10M**

Sentence: "The tall man wore a hat"

Grammar:

S → NP VP
NP → Det N | Det N AdjP
AdjP → Adj
VP → V NP
Det → 'a' | 'the'
N → 'man' | 'hat'
Adj → 'tall'
V → 'wore'

Q7 Please outline the essential steps for developing a text classification system using Naive Bayes methods. Use your own datasets and apply the classifiers to a specific application. **10M**

Q8 Design a Decision tree classifier from the following training dataset. And also find the drug will be suggested for the given data. (X={age=senior, sex=female, BP=low, cholesterol=high}) **10M**

Patient ID	Age	Sex	BP	Cholesterol	Drug
P1	Young	F	High	Normal	Drug A
P2	Young	F	High	Normal	Drug B
P3	Middle-Age	F	Normal	Low	Drug A
P4	Senior	M	Low	Low	Drug B
P5	Senior	M	Low	High	Drug B

Q9 Write the commands based for a crawler and display the following contents from the webpage. **10M**

I. Extracting the title

II. Extract all the titles in the page with this selector

III. Extract all the hyperlinks in the web page

IV. Command to list the names of all the images used in the page.

Q10 Explain the scenario and type of data requirement to consider before writing Web Crawlers. **10M**
Also, demonstrate under which scenario depth first manner is more useful than breadth first manner. Provide proper explanation and reasons to support the assertions.

Q11 Explain how Information retrieval systems can act as a bridge between people and data repositories.Illustrate the solution with major emphasis on components and feature extraction of information retrieval system. Provide proper explanation, overview diagram and reasons to support the assertions. **10M**

Q12 Describe the design of both single-document text summarization and multiple-document text summarization systems, accompanied by clear and concise diagrams illustrating their architectures." **10M**

QP MAPPING

Q. No.	E/A/T	Module Number	Marks	BL	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped
Q1	E	1	10	BL1,2	CO1	2,3,5	1,2,3	PSO1
Q2	A	1	10	BL1,2	CO1	2,3,5	1,2,3	PSO1
Q3	T	2	10	BL4	CO2,3	4,5,6	1,2,3	PSO3
Q4	A	2	10	BL1,2	CO2,3	4,5,6	1,2,3	PSO3
Q5	E	3	10	BL1,2	CO3, CO4	1,2,3	1,2,3,4,9	PSO1,02, 03
Q6	T	3	10	BL4	CO5, CO6	1,2,3	1,2,3,4,9	PSO1,02, 03
Q7	A	4	10	BL3	CO5, CO6	1,2,3	1,2,3,4,9	PSO1,02, 03
Q8	A	4	10	BL3	CO5, CO6	1,2,3	1,2,3,4,9	PSO1,02, 03
Q9	T	5	10	BL4	CO4,CO5,CO6	1,2,3,4,5	1,2,3,4,9, 10,11,12	PSO1,02, 03
Q10	A	5	10	BL3	CO4,CO5,CO6	1,2,3,4,5	1,2,3,4,9, 10,11,12	PSO1,02, 03
Q11	A	6	10	BL4	CO4,CO5,CO6	1,2,3,4,5	1,2,3,4,9, 10,11,12	PSO1,02, 03
Q12	E	6	10	BL1,2	CO4,CO5,CO6	1,2,3,4,5	1,2,3,4,9, 10,11,12	PSO1,02, 03