VIT-AP
UNIVERSITY

QUESTION PAPER

Name of the Examination: WINTER 2023-2024–CAT-1

Course Code: CSE3015

Course Title: NLP

Set number: 2

Date of Exam: 10/02/2024

Duration: 90 Minutes

Total Marks: 50 (FN) (FI)

Instructions:

1. Assume data wherever necessary.
2. Any assumptions made should be clearly stated.

Q1. a. Differentiate among Semantic Ambiguity, Syntactic Ambiguity, and Pragmatic Discourse. . **(5M)**

b. For the given sentence, identify whether the different meanings arise from structural ambiguity, semantic ambiguity or pragmatic ambiguity? **(5M)**

i. She told him she saved him.

ii. They saw her duck.

iii. Do you have the time.

iv. They planted roses around the garden.

v. She told him she liked the cat in the hat.

Q2. a. How can sentence tokenization be performed using the NLTK package? **(4M)**

b. Enumerate a variety of scenarios where Natural Language Processing (NLP) techniques can be applied to address specific challenges **(3M)**

c. How is the evaluation of an NLP model conducted, and which metrics are typically employed for assessment? **(3M)**

Q3. What does the term "Stemming" refer to? Provide an in-depth explanation of the types of stemmers, along with illustrative examples. Write a python program to perform stemming and find the result of the tokens listed as: **(10M)**

{thinking, understandable, aiming, probably, happily, being}

Q4. Describe the operational process of the Brill tagger and elucidate two specific rules utilized for tagging a token. Give clear explanations, snippet representations, and supporting reasons for each rule. Write a program to implement Brill Tagger.**(10M)**

Q5. Solve the tagging problem using stochastic tagging-based approach. Evaluate the probability of sequence of string **"Can Michael write Spray"**that best matches the corpus tagging shown below, where N: Noun, M:Modal, and V:verb.([10 Marks]

Training Sentences:

Michael /N *will/M slide/V spray/N*

spray/N will/M write/V Michael /N

will/M brown/N spray/V Michael /N

QP MAPPING

Q. No.	Module Number	COMapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	1	CO1	2,3,5	PEO2	PSO1	10
Q2	1	CO1	2,3,5	PEO2	PSO1	10
Q3	1	CO2	2,3,5	PEO2	PSO1	10
Q4	2	CO2	2,3,5	PEO2	PSO1	10
Q5	2	CO3	4,5,6	PEO3	PSO3	10

VIT-AP UNIVERSITY

QUESTION PAPER

Name of the Examination: Win 2023-24 Semester – CAT-1

Course Code: CSE 3015 **Course Title: Natural Language Processing**

Set number: 4 **Date of Exam:** 08/02/2024 (FN) (D1)

Duration: 90 Min **Total Marks: 50**

Instructions:

1. Assume data wherever necessary.
2. Any assumptions made should be clearly stated.

Q1. Identify the sources of multiple meanings in the provided sentence and determine whether these meanings arise from structural ambiguity, semantic ambiguity, or pragmatic ambiguity **[10M]**

(i) He drew one card.
(ii) Mr Spock was charged with illegal alien recruitment.
(iii) He crushed the key to my heart.
(iv) Time flies like an arrow
(v) John said he would come

Q2. (a) Design a program to remove rareword from a given string "Coca-Cola and Pepsi are fierce competitors, superior to Sprite" using Natural Language Tool Kit platform. Also, print the tokens before and after the removal of rareword. **[5M]**

(b) Differentiate between stemming and lemmatization. Provide proper theoretical and practical explanation to support the assertions. Based on Lancaster stemming find the result of the tokens listed as: ***{cats, walked, swimming, better, quickest}*** **[5M]**

Q3. (a) Explain the working procedure of Affix tagger and illustrate any two rules that is employed to tag a token. Provide proper explanation, snippet representation and reasons to support the explanation. **[5M]**

(b) Correct the following words using edit distance and Jaccard from nltk library compare the results and explain each step involved. **[5M]**

[soe comon challnges i natral language procesing, an hw can thy b adressed]

Q4. (a) Identify and classify various types of named entities within the provided content, placing them into one of five distinct categories: person, organization, time, location, and/or work of art. **[5M]**

India Independence Day is celebrated on August 15th each year to commemorate the nation's independence from British rule. On this day in 1947, after a prolonged struggle for freedom, India finally gained independence from British colonial rule, ending nearly 200 years of British domination.

(b) Compare and contrast the Brill and N-gram taggers. Then, outline a program for implementing one of these taggers. **[5M]**

5. Find the best sequence label using Viterbi Algorithm for the following phrase

"Nustin will stop will" **[10M]**

Nustin/noun Justin/ noun can/ model watch//verb will/ noun

Spot/ noun will/ model watch /verb Nustin / noun

Will/ model Justin/ noun spot/ verb Nustin / noun

Nustin/ noun will mode pat/ verb Spot/ noun

QP MAPPING

Q. No.	Module Number	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	1	CO1	2,3,5	PEO2	PSO1	10
Q2	1	CO1	2,3,5	PEO2	PSO1	10
Q3	2	CO2	2,3,5	PEO2	PSO1	10
Q4	2	CO2	2,3,5	PEO2	PSO1	10
Q5	2	CO3	4,5,6	PEO3	PSO3	10

VIT-AP
UNIVERSITY

QUESTION PAPER

Name of the Examination: Win 2023-24 Semester – CAT-1

Course Code: CSE 3015	**Course Title: Natural Language Processing**
Set number: 5	**Date of Exam:** 09/02/2024 (AN) (E2)
Duration: 90 Min	**Total Marks: 50**

Instructions:

1. Assume data wherever necessary.
2. Any assumptions made should be clearly stated.

Q1. Identify the sources of multiple meanings in the provided sentence and determine whether these meanings arise from structural ambiguity, semantic ambiguity, or pragmatic ambiguity **[10M]**

(i) She told him the story sitting on the bench.
(ii) They saw the man with the telescope.
(iii) You have to fly to Delhi to attend that interview.
(iv) Raja saw Ravi with his binocular.
(v) The old house had leaks.

Q2. Find the root word of each token for the below sentence using lemmatization with their respective POS of each token using NLTK library and explain each step involved. **[10M]**

"Natural language processing (NLP) is a subfield of artificial intelligence (AI) that focuses on the interaction between computers and humans through natural language. NLP techniques enable computers to understand, interpret, and generate human language in a way that is both meaningful and useful."

Q3. Design the regular expression tagger to tag the following tokens using the necessary POS rules. Use the POS tags **[10M]**

[“Running”,”Car”,”Is”,”Beautiful”,”Dog”,”yellow”,”Quickly”,”Jumping”,”over”,”Lazy”,Fox”]

Q4. (a) Identify and classify various types of named entities within the provided content, placing them into one of five distinct categories: person, organization, time, location, and/or work of art. **[5M]**

"Google, founded by Larry Page and Sergey Brin, is a multinational technology company specializing in internet-related services and products. It was incorporated on September 4, 1998, in Menlo Park, California."

(b) Compare and contrast the Affix and N-gram taggers. Then, outline a program for implementing one of these taggers. **[5M]**

Q5. Develop an HMM tagger by estimating the necessary transition and emission probabilities from a provided set of training sentences. Then, utilize the trained HMM tagger to perform POS tagging on the sentence "Will Will google CampusX" **[10M]**

Training sentences:
can/model Laasya/noun google/verb CampusX/noun
will/model Mahira/noun google/verb Campusx/noun
Mahira/noun loves/verb Will /noun
Will /noun loves/verb google/noun

QP MAPPING

Q. No.	Module Number	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	1	CO1	2,3,5	PEO2	PSO1	10
Q2	1	CO1	2,3,5	PEO2	PSO1	10
Q3	2	CO2	2,3,5	PEO2	PSO1	10
Q4	2	CO2	2,3,5	PEO2	PSO1	10
Q5	2	CO3	4,5,6	PEO3	PSO3	10

VIT-AP
UNIVERSITY
Apply Knowledge. Improve Life!®

QUESTION PAPER

Name of the Examination: WINTER(2023-2024)- CAT-1

Course Code: CSE3015 **Course Title: Natural language Processing**

Set number: 8 **Date of Exam:** 07/02/2024 (FN)(4)

Duration: 90 minutes **Total Marks: 50**

Instructions:

1. Assume data wherever necessary.
2. Any assumptions made should be clearly stated.

Q1. a) What are the challenges and applications of NLP?

b) Explain the distinctions between stemming and lemmatization, offering both theoretical and practical explanations to justify the differences. **(5+5)M**

Q2. a. Perform stemming the following tokens using appropriate regular expression patterns from NLTK library and explain the problems in stemmed words.
[Advisable, Drinking, Eating, Swimming, Comparable, Computers, Eats, Considered, Informed, Asked]

b. Calculate the number of operations required to convert the string from **str1="GEEXSFRGEEKKS" and str2="GEEKSFORGEEKS"** using editDistance()? **(5+5)M**

Q3. a. Compare the working procedure of Unigrams Tagger, Bigrams Tagger, and Trigram Tagger to tag given string token?

"Mary was scared because of the terrifying noise emitted by Chupacabra"

b. Design a program to illustrate pos_tag() method endorsed around Natural Language Tool Kit.Display the token and its output for the string ? **(5+5M)**

"I parked my car in the garage"

Q4. Explain briefly Named Entity Recorgination and implement NER on text given below. **(10M)**

"Trinamool Congress leader Mahua Moitra has moved the Supreme Court against her expulsion from the Lok Sabha over the cash-for-query allegations against her. Moitra was ousted from the Parliament last week after the Ethics Committee of the Lok Sabha found her guilty of jeopardising national security by sharing her parliamentary portal's login credentials with businessman Darshan Hiranandani"

Q5 Design the HMM tagger using required transitional and emission probabilities from the given set of training sentences. And find the POS tagging for the sentence. (10M)

" She quickly learn NLP"

Training sentence:

R/noun is/verb quickly/adverb.

AI/noun with/adposition she/preposition is/verb great/adverb.

I/noun learn/verb AI/noun.

So/conjunction, I/noun learn/verb NLP/noun.

QP MAPPING

Q. No.	Module Number	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	1	1, 2	1,2	1		10
Q2	1	1, 2	1,2	1,3		10
Q3	2	2, 2	1,2	1,3		10
Q4	2	3	1,2,3	1,3		10
Q5	2	1,2,3	1,2,3	1,3		10

VIT-AP
UNIVERSITY
Apply Knowledge. Improve Life!

QUESTION PAPER

Name of the Examination: WINTER 2023-2024 – CAT-1

Course Code: CSE3015 **Course Title: Natural language Processing**

Set number: 10 **Date of Exam:** 07/02/2024 (AN) (C2)

Duration: 90 minutes **Total Marks: 50**

Instructions:

1. Assume data wherever necessary.
2. Any assumptions made should be clearly stated.

Q1. a) Explain the data preprocessing techniques in NLP with examples. **[10M]**

b) Write a program remove stop words and lemanazatio below the paragraph**[5M]**

> "I have three visions for India. In 3000 years of our history, people from all over the world have come and invaded us, captured our lands, conquered our minds. From Alexander onwards, the Greeks, the Turks, the Moguls, the Portuguese, the British, the French, the Dutch, all of them came and looted us, took over what was ours. Yet we have not done this to any other nation. We have not conquered anyone. We have not grabbed their land, their culture, their history and tried to enforce our way of life on them. "

Q2. a) Write a program to remove the rare words from the below list: **[5M]**

tokens=['hi','i','am','am','whatever','this','is','just','a','test','test','java','python','java'].

b) Write a Python program for the list below using regex retrieve only emails.
"Hello, please contact support@example.com for assistance. You can also reach out to john.doe123@gmail.com for further inquiries." **[5M]**

Q3. a) Write a NLP Program for finding unigram for below text?. **[5M]**

"You will face many defeats in life, but never let yourself be defeated"

b) Design a program to illustrate the pos_tag() method endorsed around Natural Language Tool Kit. Display the token and its output for the string : **[5M]**

“ THE DOG SAW A MAN IN THE PARK”

Q4. a.) Explain about NER method and Implement NER on the below text **[5M]**

Text= [WASHINGTON -- In the wake of a string of abuses by New York police officers in the 1990s, Loretta E. Lynch, the top federal prosecutor in Brooklyn, spoke forcefully about the pain of a broken trust that African-Americans felt and said the responsibility for repairing generations of miscommunication and mistrust fell to law enforcement]

b.) Solve the tagging problem using Hidden Markov approach. Evaluate the probability of Sequence of string **“*Ravi Venkat playing model*”** that best matches the corpus tagging shown below, where N: Noun, M: Modal, and V:verb. **[10M]**

Training Sentences:

ravi/N model/M playing/V venkat/N

model/M lee/N ravi/V Walter/N

walter/N lee/N model/M write/V model/N

walter/N model/M slide/V ravi/N

QP MAPPING

Q. No.	Module Number	CO Mapped	PO Mapped	PEO Mapped	PSO Mapped	Marks
Q1	1	1, 2	1,2	1		15
Q2	1	1, 2	1,2	1,3		10
Q3	2	2, 2	1,2	1,3		10
Q4	2	3	1,2,3	1,3		15