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:6

Duration: 90 Min

Date of Exam: 09/02/2024

Total Marks: 50   (FN) (CE)

Instructions:

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

Q1. Explain different phases of NLP and Identify the sources of multiple meanings in the provided sentence and determine the ambiguity type.   **(10M)**
   (i)   Annie bumped into a man with an umbrella.
   (ii)   The Chicken is ready to eat.// change sentence as it is given in notes
   (iii) A modern English textbook.

Q2. Consider the below sentence and tokenize it using necessary operations in the NLTK library. From the resultant tokens, remove the stop words using the wordnet dictionary, find frequency distribution and explain each step involved.   **(10M)**

> The voice that navigated was definitely that of a machine, and yet you could tell that the machine was a woman, which hurt my mind a little. How can machines have genders? The machine also had an American accent. How can machines have nationalities? This can't be a good idea, making machines talk like real people, can it? Giving machines humanoid identities?

Q3. Differentiate between Jaccard distance and Edit distance. Using the Jaccard distance, Edit distance calculations determine the replacement words for the following wrong terms.   **(10M)**

Mistake = 'fling'

Words = ['flight','drawing','mango','flying',floating','banana','forest',walking']

Q4. Explain Rule based tagging approach. Write a program using regular expression tagger to tag the tokens using the necessary POS rules.   **(10M)**

Q5. Construct an HMM tagger by deriving the necessary transition and emission Probabilities from a provided set of training sentences. Then, apply the trained HMM Tagger to perform POS tagging on the sentence.   **(10M)**
"I learn Python with Natural Language Processing (NLP)"

Training sentences:

Python/noun   is/verb   easier/adverb.
NLP/noun   with/ad position   Python/noun   is/verb   great/adverb.
I/noun   love/verb   NLP/noun.
So/conjunction,   I/noun   learn/ver b   Python/noun.

QUESTION PAPER

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

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

Set number: 9   **Date of Exam:** 10/02/2024 (AN)(F2)

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

Instructions:

1. Assume data wherever necessary.

2. Any assumptions made should be clearly stated.

Q1. a. Write a program snowball stemmer for the below list
   ["'generous','generate','generously','generation, friendships"]
b. Explain the different steps in NLP with examples   **(5+10)M**

Q2. a. Perform stemming on 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. Design HMM model for below training sentences and calculate transition and emission probability
   And assign appropriate POS Tagging for the below sentence.   **[15]M**

   "Will had spot Mary"
   Training sentence

   mary/noun can/model see/verb.

   will/model spot/verb she/preposition pat/verb great/adverb.

   spot/noun learn/verb jane/noun.

   So/conjunction, learn/verb, had\verb.

Q4. Explain about Brill and sequence Tagger in detail with an an example program
   (10) M

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

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

Set number: 11

Date of Exam: 08/02/2024 (AN) (D2)

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

Instructions:

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

Q1. Why do NLP models require data preprocessing before testing? Explain step by step with examples.   **(10M)**

Q2. a. Tag the following sentences using proper POS tags from Penn tree bank Tag set.   **(5M)**

"Look at the sky. We are not alone. The whole universe is friendly to us and conspires only to give the best to those who dream and work."

b. Explain POS Tagging Architecture. Write a python program to process parts of speech tagging.   **(5M)**

Q3. a. Differentiate between Regex and N-gram tagger. **(5M)**

b. Identifying and categorizing named entities such as persons, organizations, locations, dates, and more in a given text using Named Entity Recognition (NER), and write a Python code.   **(10M)**

"Abdul Kalam was born on October 15, 1931, in Rameswaram, Tamil Nadu. He served as the 11th President of India from 2002 to 2007. Before his presidency, Kalam was a renowned scientist and played a pivotal role in India's missile development programs. He authored several books, including "Wings of Fire," an autobiography detailing his life journey. Kalam received numerous awards and honors, including the Bharat Ratna, India's highest civilian award, in 1997. He passed away on July 27, 2015, leaving behind a legacy of scientific innovation and inspirational leadership."

Q4. Find the best sequence label using 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