Roll No.: _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

Amrita Vishwa Vidyapeetham

Amrita School of Engineering, Coimbatore

B.Tech. Degree Examinations – November 2021

Seventh Semester

Computer Science and Engineering

15CSE358 Natural Language Processing

Time : Three hours   Maximum : 100 Marks

Course Outcomes

CO	Course Outcomes
CO01	Understand and represent the linguistic phenomena at morphological, word, syntactic and semantic levels of Natural language processing
CO02	Design the linguistic phenomena using formal language and probabilistic model
CO03	Understand the ambiguity in language at morphological, word, syntactic and semantic levels and apply appropriate methods to avoid them.
CO04	Understand the probabilistic parsing and statistical alignment..

Answer all questions

1) Differentiate types and tokens. Compute the Type-to-Token-Ratio (TTR) for the following text (assume words are tokenized by spaces only).
   "Mr. Bennet was among the earliest of those who waited on Mr. Bingley. He had always intended to visit him, though to the last always assuring his wife that he should not go; and till the evening after the visit was paid she had no knowledge of it.   [3]   [CO01][L2]

2) List 3 different interpretations for the sentence "I made her duck"   [3]   [CO01] [L2]

3) Differentiate the following words inflectional/derivational
          a) Happiest   b) Computational   c) Googling   [3]   [CO02][L2]

4) Explain how the word "foxes" create ambiguity in morphological analysis   [3] [CO03][L2]

5) Consider the following corpus of 3 sentences, compute the total number of unique bigrams?
   Assume we do not perform any pre-processing , and use the corpus as given
          a) Alice went to the cafe
          b) Bob was waiting for Alice
          c) Alice and Bob went to the museum   [3][CO02][L3]

6) Consider the PCFG given below:

S -> A B 0.5   B -> CC 0.4   S -> B C 0.5   B -> b 0.6
A -> B A 0.3   C -> A B 0.2   A -> a 0.7   C -> a 0.8

R

Compute the probability of any one of the parse for the string "aaab" [3] [CO04][L3]

7) What happens to the cosine similarity between two word vectors if [3][CO03] [L2]

    a. anagle between vectors is 0

    b. anagle between vectors is increasing from 0 to 90 degrees

8) Consider the sentence: "Although the buttons did not look good on the shirt, the dress fit well". Identify the meronym and hypernym relations from the above sentence. [3] [CO03][L3]

9) The term Java occurs in Wikipedia 400,657 times, among which 30,057 times it links to the Wikipedia pages of Java Island and 370,600 times it has referred to the page of Java programming language. Where as another term C++, which appears 51,332 times in Wikipedia, always referred to the web-page of the programming language C++. Compare the keyphraseness of the words Java and C++. [3] [CO03][L3]

10) Which of the properties of dependency graph is not satisfied in the graph given below?

root   John   saw   a   dog   yesterday   which   was   a   Yorkshire   Terrier

[3][CO04][L2]

11) Consider the following corpus C2 consisting of 5 sentences. [10] [CO02][L3]

    <s> the school was open </s>

    <s> michael and zack went to the school </s>

    <s> the playground at the school was huge </s>

    <s> bob and zack played at the playground </s>

    <s> bob , michael and zack were friends </s>

    Consider a Tri-gram Language model. Let w-1 be a non-existent empty word, w_0 = <s> and w_{n+1} = </s>. Probability estimate of a sentence S = {$w_{0,n+1}$} in such a model is given by:

    $P(w_{0,n+1}) = P(w_1|w_{-1,0}) P(w_2|w_{0,1}) P(w_3|w_{1,2}) ... P(w_{n+1}|w_{n-1,n})$

    For simplicity, $P(w_{0,n+1}) = P(w_1|w_0) P(w_2|w_{0,1}) P(w_3|w_{1,2}) ... P(w_{n+1}|w_{n-1,n})$.

    Calculate P(<s> michael and zack played at the playground </s>)

R

12) Assuming that POS tagging follow a first order Hidden Markov Model with the following emission and transition matrices for its states, calculate P(x1 = "man", x2 = "devil", x3 = "god", y1 = "T2", y2 = "T3", y3 = "T3").

[Hint: all POS tags are equally likely to be at the starting of a sequence.]   [10] [CO2] [L3]

Emission probabilities:

	man	devil	god
T1	0.1	0.1	0.8
T2	0.8	0.1	0.1
T3	0.2	0.2	0.6
T4	0.8	0.1	0.1

Transition probabilities

	T1	T2	T3	T4
T1	0.18	0.01	0.8	0.01
T2	0.9	0	0.05	0.05
T3	0.4	0.5	0.05	0.05
T4	0.4	0.5	0.05	0.05

13) Verify the given dependency parser using a transition based parser (Shift-reduce parser). Write down the sequence of Shift/LA/RA operations.   [10] [CO01][L2]

1⁴

root   he   sent   her   a   funny   meme   today

R
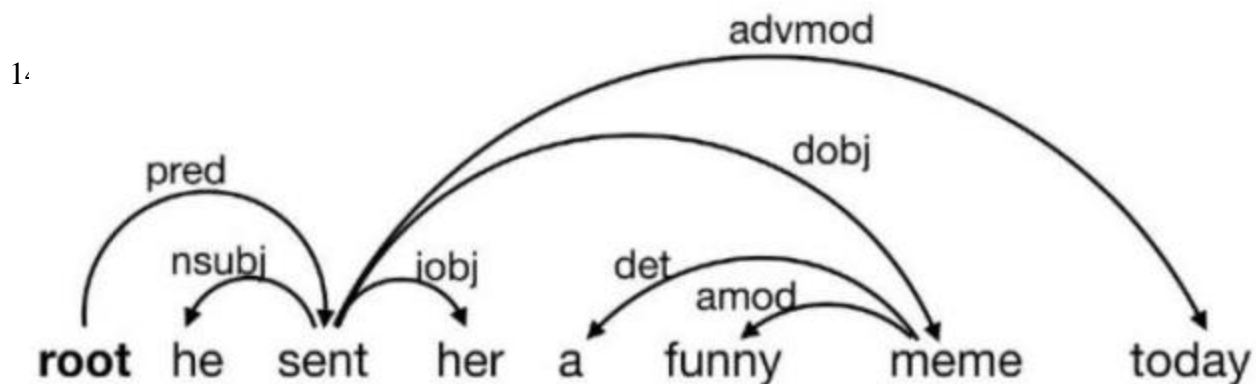
der the hypothetical wordnet given in the figure below:
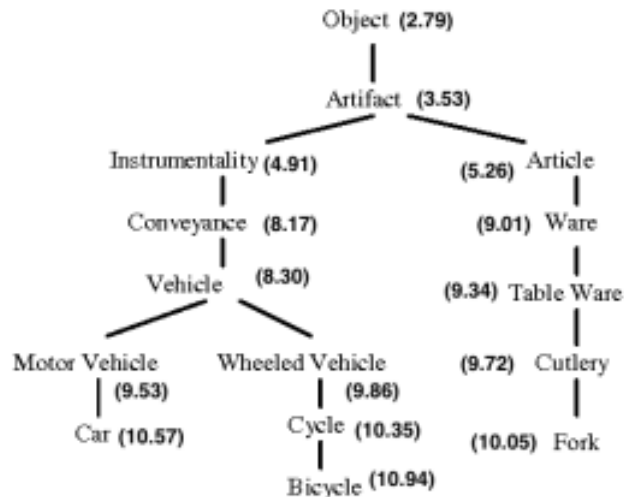
a. What is the Resnik similarity between car and Motor vehicle? [3] [CO03] [L3]

b. What is the Lin similarity between Cycle and Article? [3] [CO03] [L3]

c. What is the Jiang-Conrath distance between Object and Cutlery? [4] [CO03][L3]

15) (a) What are the challenges in abstractive summary? [3] [CO01] [L2]

(b) Consider the system generated summary (S) and the reference summaries as follows:

S : the cat was sleeping.

R1 : the cat was sleeping under the bed

R2 : the cat was found under the bed

R3 : the cat was under the bed, sleeping.

What are the ROUGE-1 and ROUGE-2 recall values for the give summary with respect to the references? [7] [CO03][L4]

16) Consider the sample movie reviews in the Table below. Assume that the three feature words are *predictable, no and fun*.

R

Text	Category
Just plain boring	Neg
Entirely predictable and lack energy	Neg
No surprises and very few laugh	Neg
Very powerful	Pos
The most fun film of the summer	Pos

    a. What will be likelyhood of predictable being negative?   [3] [CO02][L3]

    b. Using Naive Bayes classifier, predict the category of the review "predictable but fun"
                                            [7] [CO03][L3]

17)

    a. Compare and contrast the approaches for distributional semantics [5]   [CO04] [L2]

    b. Propose an algorithm to find detect the analogy relation (W1:W2 :: W3:?) between words, given word vectors as input   [5] [CO04][L4]

.   .   .

Course Outcome / Bloom's Taxonomy Level (BTL) Mark Distribution Table

CO	Marks	BTL	Marks
CO1	18	L1	0
CO2	29	L2	36

R

CO3	30	L3	52
CO4	23	L4	12
		L5	0
		L6	0

R