R22

Code No: 185DT

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD

B. Tech III Year I Semester Examinations, January - 2025

NATURAL LANGUAGE PROCESSING

(Computer Science and Engineering)

Time: 3 Hours **Max. Marks: 60**

Note: This question paper contains two parts A and B.
i) Part- A for 10 marks, ii) Part - B for 50 marks.

- Part-A is a compulsory question which consists of ten sub-questions from all units carrying equal marks.
- Part-B consists of ten questions (numbered from 2 to 11) carrying 10 marks each. From each unit, there are two questions and the student should answer one of them. Hence, the student should answer five questions from Part-B.

PART- A

(10 Marks)

1.a)	Define morphological analysis in NLP.	[1]
b)	A document contains the words "Natural Language Processing is fun". How many unigrams, bigrams, and trigrams can be generated?	[1]
c)	What is the purpose of a Treebank in natural language parsing?	[1]
d)	Parse the sentence "The cat sleeps" using a simple Context-Free Grammar (CFG).	[1]
e)	What is ambiguity resolution in parsing?	[1]
f)	When a word has three possible senses, how many probabilities need to be computed in word sense disambiguation?	[1]
g)	Define predicate-argument structure in semantic parsing.	[1]
h)	Write a meaning representation for the sentence "John eats an apple."	[1]
i)	Define smoothing.	[1]
j)	Name one evaluation metric for language models.	[1]

PART - B

(50 Marks)

2.a)	Explain the challenges of morphological analysis for languages with rich inflectional systems. Provide examples.	
b)	A document has 15 words. How many possible bigrams and trigrams can be generated? Calculate the total and explain how these features are used in NLP.	[5+5]

OR

3.a)	Compare and contrast rule-based and statistical methods for document structure analysis. Which approach would you recommend for a large dataset?	
b)	Identify all the morphemes in the word "unbelievably" and classify them as prefixes, roots, or suffixes.	[5+5]

4.a) Construct a parse tree for the sentence "The quick brown fox jumps over the lazy dog" using a given CFG:

S → NP VP
NP → Det Adj N | Det N
VP → V NP | V NP PP
PP → P NP
Det → "The"
Adj → "quick" | "brown" | "lazy"
N → "fox" | "dog"
V → "jumps"
P → "over"

b) Describe the differences between top-down and bottom-up parsing strategies. Provide examples where each method is preferred. [5+5]

OR

5. Illustrate the mechanism of shift reduced parsing with neat sketch. [10]

6.a) A word has three possible senses. If the probabilities of each sense are 0.2, 0.5, and 0.3, calculate the entropy of the distribution. Explain how this entropy value reflects the uncertainty of word sense disambiguation.

b) Illustrate with examples how multilingual issues like word order and morphology affect parsing strategies. [5+5]

OR

7.a) Consider the sentence "He runs fast." Translate it into a language of your choice, highlighting challenges in retaining its syntactic and semantic structure.

b) Explain the role of probabilistic models in ambiguity resolution, with examples from word sense disambiguation and part-of-speech tagging. [5+5]

8.a) Explain the term 'logical form' and its role in representing meaning. Provide a simple example of a logical form expression.

b) Convert the sentence "Every student loves mathematics" into First-Order Predicate Logic. [5+5]

OR

9.a) Compare and contrast rule-based and statistical approaches to semantic parsing. Provide an example where each approach is advantageous.

b) Explain the concept of semantic roles in NLP and discuss their importance in tasks like machine translation and information extraction. [5+5]

10.a) Given the trigram probabilities:

P(Natural | <s>, <s>) = 0.4
P(language | <s>, Natural) = 0.3
P(processing | Natural, language) = 0.2

Calculate the probability of the sequence "<s> <s> Natural language processing."

b) Discuss the difference between statistical language models (e.g., N-grams) and neural language models. Provide a scenario where each model excels. [5+5]

OR

11.a) For a sentence with unigram probabilities:

P(I) = 0.2
P(love) = 0.3
P(NLP) = 0.4

Compute the unigram probability of "I love NLP."

b) Explain the process of evaluating a language model using perplexity. Why is perplexity a good measure? [5+5]

---ooOoo---