NATURAL LANGUAGE PROCESSING
(CSEN 4161)

Time Allotted : 3 hrs **Full Marks : 70**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 5 (five) from Group B to E, taking at least one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternative for the following: **10 × 1 = 10**

(i) Phonology is study of organizing
(a) sound systematically (b) signal processing
(c) speech and Language processing (d) none of these.

(ii) Semantics is concerned with _____.
(a) measure the information (b) meaning of words
(c) correct the information (d) handle the noise.

(iii) Analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words is done through.
(a) Lexical Analysis (b) Syntactic Analysis
(c) Symentic Analysis (d) None of these.

(iv) The grammar that consists rules with a single symbol on the left hand side of thr rewrite rules
(a) CFG (b) CNG (c) any context (d) HMM.

(v) Classifiers used for decidind whether a word is spelled correctly are
(a) decision tree (b) support vector machine
(c) logistic regression (d) all of these.

(vi) To set up part of speech problem as a sequence labelling task we use
(a) Argmax computation (b) Applying Bayes Theorem
(c) Markov Assumption (d) All of these.

(vii) Subcategorize of verbs is classified into
(a) transitive (b) intransitive
(c) both (a) and (b) (d) None of these.

(viii) Smoothing is, is called smoothing.
(a) reevaluating some of the zero-probability
(b) low-probability N-grams intransitive
(c) assigning them non-zero values
(d) all the above.

(ix) The parsing problem for PCFGs is to
(a) disjunction of character
(b) produce the most-likely parse for a given sentence
(c) compute
(d) both (b) & (c).

(x) Machine learning approaches to sense disambiguation make it possible
(a) to automatically create robust sense disambiguation systems
(b) find ambiguity
(c) apply Baye's Rule
(d) none of these.

Group – B

2. (a) Explain Regular Expression.

(b) Differentiate between Inflectional Morphology and Derivational Morphology with example.

6 + 6 = 12

3. (a) Explain the Chomsky hierarchy of languages. Construct the parse tree for the sentence 'The man took the book from me'.

(b) Compare 'Top-down' and 'Bottom-up' approaches to NLP.

(4 + 3) + 5 = 12

Group – C

4. (a) Explain different models of computational phonology that use finite automata in various ways to realize phonological rules.

(b) What is the need of POS (Part-of Speech) Tagging in NLP?

7 + 5 = 12

5. (a) Why is smoothing necessary for NLP? Explain Witten-Bell smoothing.

(b) Give the equation of trigram probability estimation and provide suitable examples to explain them.

6 + 6 = 12

Group – D

6. (a) Explain the semantics of First Order Predicate Calculus and its role in representing meaning.
restaurant (AyCaramba) ∧ serves(AYCaramba, mexican-food)∧
near (location(AyCaramba), location (ICSI)
Explain when the above statement may be true.

   (b) How are Transformation Based Learning Rules applied in NLP?

(5 + 2) + 5 = 12

7. (a) What is meant by 'word sense disambiguation'? Briefly outline a selectional association based word sense disambiguation algorithm.

   (b) Write a short note to explain "Two Level Morphology".

(3 + 4) + 5 = 12

Group – E

8. (a) Explain the term "the entropy of a random variable X, with a suitable example.

   (b) Explain how, under Information Theory, the concept of "entropy" can be used as a metric to evaluate an N-gram system.

6 + 6 = 12

9. (a) What is a "confusion matrix"? Why and were is a confusion matrix necessary? Define "likelihood probability" using Bayes method.

   (b) Write a short note on rhetorical relation in discourse planning.

6 + 6 = 12