Name : .....

Roll No. : .....

Invigilator's Signature : .....

CS/B.Tech/CSE/SEM-8/CS-802F/2013

2013

NATURAL LANGUAGE PROCESSING

Time Allotted : 3 Hours *Full Marks* : 70

The figures in the margin indicate full marks.

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

GROUP – A

(Multiple Choice Type Questions)

1. Choose the correct alternatives for the following : 10 × 1 = 10

i) Minimum edit distance is computed by

a) Phonology

b) Dynamic programming

c) Tautology

d) Hidden Markov Model (HMM).

ii) Word probability is calculated by

a) Likelihood probability

b) Prior probability

c) Baye's rule

d) none of these.

iii) Viterbi algorithm is used in

a) Speech processing

b) Language processing

c) Speech & Language processing

d) none of these.

iv) The use of the period (.) is to specify

a) any context b) any number

c) any character d) none of these.

v) The use of | is to specify

a) disjunction of characters

b) disjunction of numbers

c) words sequence

d) none of these.

vi) Open class contains

a) Nouns b) Verbs

c) both (a) & (b) d) none of these.

vii) In deleted interpolation algorithm, which symbol is used ?

a) γ b) λ

c) σ d) μ .

viii) Entropy is used to

a) measure the information

b) correct the information

c) detect the information

d) handle the noise.

ix) Subcategorize of verbs is classified into

a) Transitive b) Intransitive

c) both (a) & (b) d) none of these.

x) Phrase Structure Grammar is used in

a) Regular Grammar

b) Context–Free Grammar (CFG)

c) Context–Sensitive Grammar (CSG)

d) None of these.

GROUP – B

(Short Answer Type Questions)

Answer any *three* of the following. 3 × 5 = 15

2. Define two level Morphology with suitable example. Briefly describe the different types of Error Handling mechanism. 3 + 2

3. Why POS (Part – of – Speech) Tagging is required in NLP (Natural Language Processing) ? Briefly compare the Top – Down & Bottom – Up Parsing techniques. 3 + 2

4. What is Regular Expression ? Write down the Regular Expression for the following languages :

   a) The set of all alphabetic strings

   b) Column 1 Column 2 Column 3

   c) 5.7 Gb. 2 + 3

5. Write down the concept of feature structure. What is unification ? What is Word Sense Disambiguation (WSD) ? 2 + 1 + 2

6. Write down the differences between Inflectional Morphology and Derivational Morphology with suitable example. What is stem ? What is morpheme ? 3 + 1 + 1

GROUP – C
(Long Answer Type Questions)
Answer any *three* of the following. 3 × 15 = 45

7. a) Define wordform, lemma, type, token.

   b) Briefly describe the role of Finite State Tranducer (FST) with suitable example.

   c) Define prior probability and likelihood probability using Bayesian Method.

   d) What is Confusion Matrix ? Why is it required in NLP (Natural Language Processing) ? 4 + 5 + 4 + 2

8. a) Draw tree structure for the following ATIS sentences :

   I perfer a morning flight

   I want a morning flight

   Using S → NP VP

   NP → Pronoun

   |Pronoun-Noun

   |Det Nominal

   Nominal →|Noun Nominal

   |Noun

   VP → verb

   |Verb NP

   |Verb NP PP

   | Verb PP

b) Write rules expressing the verbal subcategory of English auxiliaries with example.

c) Define predeterminers, cardinal numbers, ordinal numbers and quantifiers with suitable examples.

d) How are Transformation Based Learning (TBL) Rules applied in NLP (Natural Language Processing) ?

5 + 3 + 4 + 3

9. a) Compute Minimum edit by hand. Figure out whether the word intention is closer to the word execution and calculate a minimum edit distance.

b) Estimate p (t / c) as follows (where c_p is the pth character of the word c) using Kernigham *et al.* four confusion matrices, one for each type of single error.

c) Briefly describe Hidden Markov Model (HMM).

d) Compare open class & closed class word groups with suitable examples. 6 + 3 + 4 + 2

10. a) What is Smoothing ? Why is it required ?

b) Write down the equation for trigram probability estimation.

c) Write down the equation for the discount $d = c^*/c$ for add-one smoothing. Do the same thing used for Written Bell smoothing. How do they differ ?

2 + 1 + 3 + 5 + 4

11. Write short notes on any *three* of the following : 3 × 5

a) Regular Expression Patterns

b) Orthographic Rules

c) Stochastic Part-of-Speech Tagging

d) HMM Tagging

e) Constituency & Agreement.