Name : .....

Roll No. : .....

Invigilator's Signature : .....

CS/B.Tech (CSE)/SEM-8/CS-802B/2010

2010

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 : $10 \times 1 = 10$

   i) Word probability is calculated by

      a) Likelihood probability b) Prior probability

      c) Baye's Rule d) None of these.

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

      a) any context b) any number

      c) any character d) none of these.

   iii) Minimum edit distance is computed by

      a) Phonology

      b) Dynamic programming

      c) Tautology

      d) Hidden Markov Model (HMM).

http://qpaper.wbut.ac.in

iv) The use of brackets [] is to specify

a) disjunction of characters

b) disjunction of numbers

c) word sequence

d) none of these.

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

a) γ b) λ

c) σ d) μ.

vi) Viterbi algorithm is used in

a) Speech processing

b) Language processing

c) Speech & Language processing

d) None of these.

vii) Entropy is used to

a) measure the information

b) correct the information

c) detect the information

d) handle the noise.

http://qpaper.wbut.ac.in

viii) Open class contains

a) verbs b) nouns

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

ix) Subcategorization of verbs is classified into

a) intransitive b) transitive

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

a) Column 1 Column 2 Column 3

b) The set of all alphabetic string

c) 4·3 Gb 2 + 3

http://qpaper.wbut.ac.in

3. Write down the differences between Inflectional Morphology and Derivational Morphology with suitable example. What is stem ? What are Morphemes ? (2 + 1) + 1 + 1

4. Define Two-level Morphology with suitable example. Briefly describe the different types of Error Handling mechanism.

(1 + 1) + 3

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

6. Write down the concept of Feature Structure. What is Unification ? What is Word Sense Disambiguation (WSD) ?

2 + 1 + 2

GROUP – C

(Long Answer Type Questions)

Answer any *three* of the following. 3 × 15 = 45

7. 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. Is the same thing used for Witten-Bell smoothing ? How do they differ ?

2 + 1 + 3 + 5 + 4

http://qpaperwbut.ac.in

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

   b) Briefly describe the roles of Finite State Transducer (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

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

http://qpaper.wbut.ac.in

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

I prefer 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

http://qpaper.wbut.ac.in

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

a) Regular Expression (R.E) Patterns.

b) Orthographic Rules.

c) Problems with the basic Top-Down Parser.

d) Stochastic Part-of-Speech Tagging.

e) HMM (Hidden Markov Model) Tagging.

f) Constituency & Agreement.

http://qpaper.wbut.ac.in