NATURAL LANGUAGE PROCESSING (SEMESTER - 8)

CS/B.TECH(CSE)/SEM-8/CS-802F/09

1. ..... *Signature of Invigilator*

2. ..... *Signature of the Officer-in-Charge*

Reg. No. | | | | | | | | | | | | | | |

Roll No. of the Candidate | | | | | | | | | | | | | | |

CS/B.TECH(CSE)/SEM-8/CS-802F/09

ENGINEERING & MANAGEMENT EXAMINATIONS, APRIL – 2009

NATURAL LANGUAGE PROCESSING (SEMESTER - 8)

Time : 3 Hours] [Full Marks : 70

INSTRUCTIONS TO THE CANDIDATES :

1. This Booklet is a Question-cum-Answer Booklet. The Booklet consists of **32 pages**. The questions of this concerned subject commence from Page No. 3.
2. a) In **Group – A**, Questions are of Multiple Choice type. You have to write the correct choice in the box provided **against each question**.
   b) For **Groups – B** & **C** you have to answer the questions in the space provided marked 'Answer Sheet'. Questions of **Group – B** are Short answer type. Questions of **Group – C** are Long answer type. Write on both sides of the paper.
3. **Fill in your Roll No. in the box** provided as in your Admit Card before answering the questions.
4. Read the instructions given inside carefully before answering.
5. You should not forget to write the corresponding question numbers while answering.
6. Do not write your name or put any special mark in the booklet that may disclose your identity, which will render you liable to disqualification. Any candidate found copying will be subject to Disciplinary Action under the relevant rules.
7. **Use of Mobile Phone and Programmable Calculator is totally prohibited in the examination hall.**
8. You should return the booklet to the invigilator at the end of the examination and should not take any page of this booklet with you outside the examination hall, **which will lead to disqualification**.
9. Rough work, if necessary is to be done in this booklet only and cross it through.

No additional sheets are to be used and no loose paper will be provided

---

FOR OFFICE USE / EVALUATION ONLY

Marks Obtained

	Group – A										Group – B			Group – C			***Total Marks***	***Examiner's Signature***
Question Number																		
Marks Obtained																		

.....

Head-Examiner/Co-Ordinator/Scrutineer

8876-F/F (27/04)

DO NOT WRITE ON THIS PAGE

ENGINEERING & MANAGEMENT EXAMINATIONS, APRIL – 2009

NATURAL LANGUAGE PROCESSING

SEMESTER - 8

Time : 3 Hours] [Full Marks : 70

GROUP – A

(Multiple Choice Type Questions)

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

i) / [0 1 2 3 4 5 6 7 8 9] / specifies

a) single digit b) multiple digits

c) any digit d) none of these. ☐

ii) Colou?r matches

a) color b) color or colour

c) colour d) none of these. ☐

iii) **Minimum edit distance** is computed by

a) phonology

b) dynamic programming

c) tautology

d) Hidden Markov Model (HMM). ☐

iv) **Word probability** is calculated by

a) Likelihood probability b) Prior probability

c) Baye's Rule d) None of these. ☐

v) **Viterbi algorithm** is used in

a) speech processing b) language processing

c) speech & language processing d) none of these. ☐

vi) In deleted interpolation algorithom, which symbol is used ?

a) γ b) λ

c) σ d) μ. ☐

vii) **Entropy** is used for

a) measure the information b) correct the information

c) detect the information d) handle the noise. ☐

viii) **Open class** contains

a) Nouns b) Verbs

c) both (a) & (b) d) none of these. ☐

ix) Phrase Structure Grammar is used in

a) Regular Grammar

b) Context–Free Grammar (CFG)

c) Context–Sensitive Grammar (CSG)

d) None of these. ☐

x) Subcategorize of verbs is classified into

a) transitive b) intransitive

c) both (a) & (b) d) none of these. ☐

GROUP – B
(Short Answer Type Questions)

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

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

a) The set of all alphabetic String

b) $ 199.99

c) 4.3 MHz.

3. Write down the differences between Inflectional Morphology and Derivational Morphology with suitable example. Briefly describe **Morpheme, stem, affix** and **suffix** with a suitable example.

4. Define **predeterminers, cardinal numbers, ordinal numbers** and **quantifiers** with suitable example. Briefly describe the different types of **Error Handling** mechanism.

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

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

GROUP – C

(Long Answer Type Questions)

Answer any *three* questions. 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.

   Do the same thing used for **Witten - Bell** smoothing. How do they differ ?

   2 + 1 + 3 + 5 + 4

8. a) Find one tagging error in each of the following sentence that are tagged with the Penn Treebank tagset :

      i) I/PRP need/VBP a/DT flight/NN from/IN Atlana/NN

      ii) Does/VBZ this/DT flight/NN serve/VB dinner/NNS

      iii) I/PRP have/VB a/DT friend/NN living/VBG in/IN Denver/NNP.

   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) ? 6 + 4 + 3 + 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 **p-th** 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) 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) How Transformation Based Learning (**TBL**) Rules are applied in **NLP** (Natural Language Processing) ? 15

11. Write short notes on any *three* of the following : 3 × 5 = 15

a) **Orthographic** Rules.

b) Stochastic Part-of-Speech Tagging (**POS**).

c) **HMM** Tagging.

d) **Constituency & Agreement.**

e) Problems with the basic **Top-Down Parser.**

f) Two level morphology.

7

END