Roll No. | | | | | | | | |

B.E / B.Tech (Full Time) DEGREE END SEMESTER EXAMINATIONS, NOV / DEC 2013

Computer Science and Engineering

VII Semester

CS 9040 Language Technology

(Regulation 2008)

Time : 3 Hours | Answer ALL Questions | Max. Marks 100

PART-A (10 x 2 = 20 Marks)

1. What is the difference between Natural Language Processing and Language Technology? Explain
2. How is probability used in syntax analysis?
3. How is Information Extraction different from Information retrieval?
4. Compare and Contrast Document categorization and Document Clustering.
5. Differentiate between Generative and Discriminative Models.
6. What is Word Sense Disambiguation? Illustrate using examples.
7. How is Information Retrieval evaluated?
8. How is machine translation evaluated?
9. Outline Grice's Maxims regarding Discourse?
10. Discuss one application where you need speech, text and image.

PART B - (5 X 16 = 80 MARKS)

11. Imagine that you are a personal assistant to the Managing Director of a Multinational Company. You are required to handle all documents for the Company. In case you are replaced by semi-automatic system with Language Technology skills, list out and explain with a block diagram All the skills required and the corresponding Language Technology issues. (16)

12.(a) (i)Explain in detail two level Morphological Analysis used for natural language. Discuss the use of this technique for an Indian Language of your choice. Clearly explain the morphological rules used. (10)

(ii) Describe a typical Morphographemic Transducer with an example. (6)

OR

12 (b) (i) Explain the Earley algorithm in detail. (3)

ii) Simulate the Earley algorithm for the grammar given below:

S→ NP VP	NP → Ram
S → VP NP	N → spoon
NP→Det N	N→payasam
NP→ NP PP	V → ate
VP→ V NP	N → dish
VP→ V NP NP	P → with
VP→ VP PP	P → in
PP→ P NP	Det → the
	Det → a

The sentence is "Ram ate the payasam in the dish with a spoon" (7)

(iii) Give a detailed account of Thematic roles and Case Frames with suitable examples from English and an Indian Language of your choice (6)

13.(a) (i) Compare and contrast Information Retrieval and Web Search. (4)
(ii) Explain the Vector Space Model used for Information Retrieval. (6)
(iii)Explain the PageRank algorithm used by Google. (6)

OR

13 (b) (i) Discuss the various steps in a typical Information Extraction System. (8)
(ii) Explain how relations are extracted from plain text using the **Snowball** system. (8)

14 (a) (i) Explain how Naïve Bayes Classifier is used to classify text. (8)
(ii) Explain how multilingualism and multimodality can be used to enhance a web search engine. Discuss the methods used for the integration (8)

OR

14(b) (i) Discuss the SVM algorithm in detail. (6)
(ii) Explain how SVM algorithm is used for document classification explaining The various issues (6)
(iii) Write a short note on speech coding. (4)

15 (a) (i) Explain the different approaches to machine translation. (4)
(ii) We need to translate an Indian Language of your choice to English. Discuss the various stages of statistical machine translation required for the task. (8)
(iii) Explain how speech acts are generally used to describe illocutionary acts.(4)

OR

15 (b) Write Short Notes on **any two** of the following: 2X8

i. Natural Language Generation system
ii. Question Answering System
iii. Discourse Processing