

~~21/11/13~~ 21/11/13

Roll No.

								(35)
--	--	--	--	--	--	--	--	------

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. Discuss what is meant by Language Technologies.
2. "Ambiguity make natural language processing complex". Discuss
3. How is probability used in syntactic analysis?
4. How is Information Extraction different from Information retrieval?
5. Compare and Contrast Document categorization and Document Clustering.
6. Differentiate between Generative and Discriminative Models.
7. What is the need for multimodality?
8. Discuss one application where you Language Technology.
9. How is machine translation evaluated?
10. Outline Grice's Maxims regarding Discourse?

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) Describe the various steps of the Porter Stemmer. (10)
(ii) Explain how natural language can be used to improve the performance of Search Engines (6)

(OR)

- 12 (b) (i) Explain the Earley algorithm. (4)
(ii) Simulate the Earley algorithm giving the grammar used for the sentence explaining clearly the procedures used "The girl ate payasam with a spoon" (6)
(iii) Explain how Universal Networking Language (UNL) is used to represent semantics of Natural language sentences. (6)

- 13.(a) (i) Compare and contrast Information Retrieval and Web Search. (4)
(ii) Given the following documents, create the vector space model, compute the inverse document frequency, compute the weight of the words and normalize all documents to unit length.
Doc1 : Text Mining is an important area of research. Research in Text Mining involves machine learning.
Doc2: Text classification and text clustering are important areas of text mining. Other areas include feature extraction and dimensionality reduction.
Doc3: There are many algorithms for text classification. Supervised techniques of machine learning are generally used for text classification.
Doc4: There are many algorithms for text clustering. Unsupervised techniques of machine learning are generally used for text clustering. (8)
(iii) Explain the PageRank algorithm used by Google. (4)

(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. Bayes Theorem for POS tagging
iv. Discourse Processing

{red, brown, green}. The part number ranges from 10020301 to 10020322.
Design black box test cases using boundary value analysis.

15. a) Explain SCM Process and achieving version control over configuration items with a neat diagram.

(OR)

b) What are risks? Explain risk matrix development, risk prevention and risk mitigation measures in detail.