

Roll No.

--	--	--	--	--	--	--	--	--	--

32

B.E / B.Tech (Full Time) DEGREE END SEMESTER EXAMINATIONS, APRIL / MAY 2013

Computer Science and Engineering

V Semester

CS 9040 Language Technology

(Regulation 2008)

Time : 3 Hours

Answer ALL Questions

Max. Marks 100

PART-A (10 x 2 = 20 Marks)

1. Explain the difference between Natural Language processing and Language Technologies
2. What is Word Sense Disambiguation? Explain with suitable examples.
3. Explain Morphographemics with illustrative examples.
4. Explain the mechanism used by Web Search Engine to provide better quality search results
5. Distinguish between Information Retrieval and Information Extraction.
6. Differentiate between Document categorization and Document Clustering.
7. What is Hierarchical Clustering? Explain.
8. How is machine translation evaluated?
9. What is speech coding? Explain.
10. What are the issues associated with Surface Realization?

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 the Earley algorithm. (4)
(ii) Simulate the Earley algorithm giving the grammar used for the sentence explaining clearly the procedures used
"The girl sang the song in the evening" (6)
(iii) Explain how natural language can be used to improve the performance of Search Engines (6)

(OR)

12. (b) (i) Describe the various steps of the Porter Stemmer (6)
(ii) Discuss the various important types of ambiguities that occur in Natural Language. (4)
(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) 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 : Chennai was formerly known as Madras. Chennai is located on the Coromandal coast of the Bay of Bengal.

Doc2: Chennai is the capital of Tamilnadu. The other names of Chennai are Madras and Chennaipatinam. Chennai is the base for Tamil Film Industry.

Doc3: Chennai is one of the most visited cities in India. This city boasts of two popular beaches Marina and Elliots. The snake park in Chennai is located in Guindy. The snake park has gained statutory recognition from Central zoo authority.

Doc4: Chennai is visited by many tourists from all over the world. They visit Marina and the zoo and the snake park. They also visit the film shooting taking place in Chennai (6)

- (ii) Discuss in detail the differences between Information Retrieval and Web Search. (6)

- (iii) Explain the PageRank algorithm used by Google. (4)

(OR)

13. (b) (i) Explain in detail the various modules of a Web Search engine (8)
(ii) Discuss the various steps in a typical Information Extraction System. (8)

14. (a) (i) Explain how Naïve Bayes Classifier is used to classify text. (6)
(ii) Explain how multilingualism and multimodality can be used to enhance a web search engine. Discuss the methods used for the integration (6)
(iii) Discuss some typical applications of Text Mining (4)

(OR)

14. (b) (i) Write a detailed note on Multimodal Information Access and Synthesis. (8)
(ii) Explain how relations are extracted from plain text using the **Snowball** system. (8)

- 15 (a) (i) 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)
(ii) What is meant by interlingua approach to Machine Translation? Discuss in detail any one approach. (8)

- 15 (b) Write Short Notes on the following: 2X8
(i) Discourse Processing
(ii) Dialog and Conversational Systems.