

Anna University  
B.E.Degree Examinations, November 2012  
(Computer Science and Engineering)  
Regulations 2008  
**CS 9040 LANGUAGE TECHNOLOGIES**

5

**Time : Three Hours**

**Maximum: 100 marks**

**Answer All Questions**

**PART A - (10 X 2 =20 marks)**

1. Explain the difference between Natural Language processing and Language Technologies.
2. Give examples of Semantic ambiguity.
3. How is Information Retrieval evaluated?
4. How is Information Extraction different from Natural Language Processing?
5. Compare and Contrast Document categorization and Document Clustering.
6. What are lexical chains?
7. What is the need for multilingualism?
8. Discuss one application where you need speech, text and image.
9. How is machine translation evaluated?
10. What is Discourse?

**PART B - (5 X 16 = 80 MARKS)**

11. (i) Take any one application of your choice and explain how language technology would enhance it's features. Ensure that you incorporate at least six language technology aspects. (10)
- (ii) Discuss in detail any two language technology functions you have used for the Application. (6)
- 12.(a) (i) Describe the various steps of the Porter Stemmer. (8)
- (ii) Give FST for the following (explain with examples the state transitions)
  - a. Plurals of nouns
  - b. Verb Morphologyfor a language of your choice. (8)

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 cooked dinner in the evening" (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) Explain in detail the various modules of a Web Search engine (8)  
(iii) Explain the PageRank algorithm used by Google. (4)

OR

- 13 (b) (i) Discuss the use of categories and clusters for organizing retrieval results. (8)  
(ii) Explain how relations are extracted from plain text using the Snowball system. (8)
- 14 (a) (i) Explain how multilingualism and multimodality can be used to enhance a web search engine. Discuss the methods used for the integration (8)  
(ii) Discuss the Bayes Theorem and some applications of Bayesian Learning (4)  
(iii) Explain how Naïve Bayes Classifier is used to classify text. (4)

OR

- 14(b) (i) Discuss the SVM algorithm in-detail. (4)  
(ii) Explain how SVM algorithm is used for document classification explaining clearly how the documents are represented using the vector space model (8)  
(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  
(i) Commercial search engine features  
(ii) Ambiguity Resolution  
(iii) Discourse Processing