CSE 4011 NATURAL LANGUAGE PROCESSING [3 0 0 3]

Course Objectives:

- To get introduced to the concepts in Natural Language Processing
- To understand the variety of issues in processing various Natural Languages and mechanisms
- To familiarize with the Applications of Natural Language Processing

Course Outcomes:

At the end of the course, students will be able to

- Describe the basic concepts and techniques of Natural Language Processing
- Develop skills of using Natural Language Processing Techniques in various applications
- Analyze the various types of approaches to Natural Language Processing Problems and their merits

1. INTRODUCTION TO NATURAL LANGUAGE PROCESSING

   Knowledge in Speech and Language Processing, Ambiguity, Models and Algorithm.

   (Chapter 1.1-1.3 of Text Book 1) (2 hrs)

2. WORDS AND TRANSDUCERS

   Survey of English Morphology, Finite-State Morphological Parsing, Building a Finite-State Lexicon, Finite-State Transducers, FSTs for Morphological Parsing, Lexicon-Free FSTs: The Porter Stemmer, Detecting and Correcting Spelling Errors, Minimum Edit Distance.

   (Chapter 3.1-3.5, 3.8, 3.10 of Text Book 1) (5 hrs)

3. N-GRAMS

   Counting words in Corpora, Simple(Unsmoothed) N-Grams, Training and Test Data, Evaluating N-Grams : Perplexity, Smoothing, Interpolation, Backoff, Advanced: Information theory background.

   (Chapter 4.1-4.10 of Text Book 1) (6 hrs)

4. WORD CLASSES AND PART-OF-SPEECH TAGGING

   English Word Classes, Tagsets for English, Part-of-Speech Tagging, Rule-based Part-of-Speech Tagging, HMM Part-of-Speech Tagging, Transformation-based Tagging, Evaluation and Error Analysis, Advance: The Noisy Channel Model for Spelling.

   (Chapter 5.1-5.7, 5.9 of Text Book 1) (7 hrs)

5. FORMAL GRAMMARS OF ENGLISH

   Constituency, Some Grammar Rules for English, Treebanks, Dependency Grammar.

   (Chapter 12.1-12.3, 12.7 of Text Book 1) (4 hrs)

6. PARSING WITH CONTEXT-FREE GRAMMARS

   Parsing as Search, Ambiguity, Dynamic Programming Parsing Methods.

(Chapter 13.1, 13.2, 13.4 of Text Book 1) (6 hrs)

7. STATISTICAL PARSING
   Probabilistic Context-Free Grammars, Evaluating Parsers.
   (Chapter 14.1, 14.7 of Text Book 1) (2 hrs)

8. APPLICATIONS OF NATURAL LANGUAGE PROCESSING
   Information Extraction, Question Answering and Summarization, Classical MT & The Vauquois Triangle, Statistical MT.
   (Chapter 22.1, 23.1, 25.2, 25.3 of Text Book 1) (4 hrs)

Text Books :

1. Daniel Jurafsky & James H. Martin, *Speech and Language Processing*, Second Edition, 2000.

References :

1. Akshar Bharati, Rajeev Sangal and Vineet Chaitanya, *Natural Language Processing: A Paninian Perspective on* , Prentice-Hall of India, New Delhi, 1995.
2. Steven Bird, Ewan Klein, Edward Loper, *Natural Language Processing with Python – Analysing Text with natural language toolkit* , O'Reilly Media, 2009.
3. Chris Manning, Hinrich Schutze , *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, 1999.