|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Course Code**CSE 528** | NATURAL LANGUAGE PROCESSING | L T P C | 3 | 0 | 0 | 3 |
| Version No. |  |
| Course Prerequisite / Exposure: | Higher Mathematics, Advanced Database Systems |
| ObjectiveTo introduce the fundamental techniques of natural language processing, to develop an understanding of the limits of those techniques and of current research issues, and to evaluate some current and potential applications. |
| **Expected Outcome**At the end of the course students should* be able to discuss the current and likely future performance of several NLP applications, such as machine translation and email response
* Be able to describe briefly a fundamental technique for processing language for several subtasks, such as morphological analysis, parsing, word sense disambiguation etc.
* understand how these techniques draw on and relate to other areas of (theoretical) computer science, such as formal language theory, formal semantics of programming languages, or theorem proving
 |
| Unit No. I: | **INTRODUCTION** |   |
| Introduction to the Morphology, Syntax, Semantics by linking the “linguistics view” (computational linguistics) with the “artificial intelligence view” (natural language processing), Text Processing: Character Encoding, Word Segmentation, Sentence Segmentation, Introduction to Corpora, Corpora Analysis |
| Unit No. II: | **MORPHOLOGY**  |   |
| Analysis and generation of language on word level: Inflectional and Derivation Morphology, Morphological Analysis using finite state transducers, Introduction to word types, POS Tagging, Multi-word Expressions  |
| Unit No. III: | **SYNTAX** |   |
| Analysis and generation of language on phrasal and sentence level: Introduction to phrases, clauses and sentence structure, Constituency tests, Phase Structure Grammars, Unification Based Grammar Formalisms, Earley Algorithm for Parsing with CFG as well as Unification based grammars, Shallow Parsing and Chunking |
| Unit No. IV: | **SEMANTICS** |   |
| Language ambiguities on the level of “meaning”: represented by case structures and conceptual dependency structures, Lexical Semantics, Word Sense Disambiguation, WordNet, Introduction to Thematic roles, Semantic Role labelling |
| Unit No. V: | **APPLICATIONS OF NLP** |   |
| Grammar Checkers Dictation, Automatic Document Generation, NL Interfaces, Text Summarization, Sentiment Mining, Machine Translation |
| **Text / Reference Books** 1.Daniel Jurafsky and James H. Martin “Speech and Language Processing”, 3rd edition, Prentice Hall, 2009.2.Chris Manning and Hinrich Schütze, “Foundations of Statistical Natural Language Processing”, 2nd edition, MITPress Cambridge, MA, 2003.3. Nitin Indurkhya, Fred J. Damerau “Handbook of Natural Language Processing”, Second Edition, CRC Press, 2010 |
| Mode of Evaluation  | Tests, Assignments and Seminar  |
| Recommended by the Board of Studies on |  |
| Date of Approval by the Academic Council |  |