CS 537 Natural Language Processing

Syllabus

L-T-P: 3-0-0 Credits 3

- Introduction- Human languages, models, ambiguity, processing paradigms; Phases in natural language processing, applications.
- Text representation in computers, encoding schemes.
- Linguistics resources- Introduction to corpus, elements in balanced corpus, TreeBank, PropBank, WordNet, VerbNet etc. Resource management with XML, Management of linguistic data with the help of GATE, NLTK.
- Regular expressions, Finite State Automata, word recognition, lexicon.
- Morphology, acquisition models, Finite State Transducer.
- N-grams, smoothing, entropy, HMM, ME, SVM, CRF.
- Part of Speech tagging- Stochastic POS tagging, HMM, Transformation based tagging (TBL), Handling of unknown words, named entities, multi word expressions.
- A survey on natural language grammars, lexeme, phonemes, phrases and idioms, word order, agreement, tense, aspect and mood and agreement, Context Free Grammar, spoken language syntax.
- Parsing- Unification, probabilistic parsing, TreeBank.
- Semantics- Meaning representation, semantic analysis, lexical semantics, WordNet
- Word Sense Disambiguation- Selectional restriction, machine learning approaches, dictionary based approaches.
- Discourse- Reference resolution, constraints on co-reference, algorithm for pronoun resolution, text coherence, discourse structure.
- Applications of NLP- Spell-checking, Summarization
- Information Retrieval- Vector space model, term weighting, homonymy, polysemy, synonymy, improving user queries.

Machine Translation– Overview.

Textbook:

1. Daniel Jurafsky and James H Martin. *Speech and Language Processing, 2e*, Pearson Education, 2009

Reference Books:

- 1. James A. Natural language Understanding 2e, Pearson Education, 1994
- 2. Bharati A., Sangal R., Chaitanya V.. *Natural language processing: a Paninian perspective*, PHI, 2000
- 3. Siddiqui T., Tiwary U. S.. *Natural language processing and Information retrieval*, OUP, 2008