

Course-Plan
Autumn 2020

| | |
|--------------------|---|
| School | Engineering |
| Department | Computer Science and Engineering |
| Course Code | CO 504 |
| Course Name | Natural Language Processing |
| Instructor | Utpal Sharma |

1. **Abstract:** CS 504 Natural Language Processing (NLP) covers how computers can deal with human languages. Human languages, referred to as natural languages, can be the ultimate language for human-machine communication. Due to the nature of these languages, the challenges in making computers deal with natural languages make it a topic of artificial intelligence. This course covers the issues involved, some possible approaches to tackling these issues, and various applications that can be built based on NLP.

2. Objective:

| Module | Topic | Learning Objectives |
|---------------|-------------------------------|--|
| 1 | Introduction | To get an overall idea of the characteristics of natural languages and the phases in NLP. |
| 2 | Text Representation | To understand the the issues in representation of texts in a computer, and the standards. |
| 3. | Linguistic Resources | To learn what different linguistic resources can be used for NLP. |
| 4. | Structure of words | To understand how words are formed in a natural language, and how to analyse words for NLP. |
| 5. | Part-of-Speech (POS) Tagging | To understand how the role of words or groups of words in a sentence have to be determined so as to recognise the syntactic structures of sentences. |
| 6. | Structural analysis | To understand the structural aspects of correctness of sentences, and formal grammars used for this task. To learn about techniques for this task- parsing. |
| 7. | Semantic Analysis | To understand how meaning of natural language expressions can be determined, firstly from structure of the expressions, then from groups of expressions (discourse) and then from extra knowledge about the context. The issues due to ambiguity are also discussed. |
| 8. | Applications | To understand various applications of NLP, from simple ones to complicated ones, upto that of machine translation. |

3. Prerequisites of the course:

Knowledge of programming, and familiarity with basic linguistic concepts.

4. Course outline (See Syllabus)

Suggested reading:

Text Book:

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

Reference Books:

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

5. (a)Time-Plan

| Tentative Lectures | Topics |
|--------------------|---|
| 1-2 | Introduction Human languages, models, ambiguity, processing paradigms Phases in Natural Language Processing |
| 3-4 | Text representation in computers, encoding schemes |
| 5 | Regular expressions, FSA, word recognition, Lexicon |
| 6-7 | Morphology, Acquisition models, FST |
| 8-10 | N-grams, smoothing Entropy |
| 11-14 | POS tagging, Stochastic POS tagging, HMM Transformation based tagging (TBL) Issues |
| 15-16 | CFG, spoken language syntax, word order |
| 17-22 | Parsing, Unification, Probabilistic parsing, Treebank |
| 23-24 | Semantics, Meaning representation |
| 25-29 | Semantic Analysis, Lexical semantics, WordNet, Summarization |
| 30-32 | WSD Selectional restriction Machine learning approaches, dictionary based approaches |
| 33-35 | IR Vector space model, term weighting, Homonymy, Polysemy, synonymy Improving user queries |

| | |
|-------|---|
| 36-39 | Discourse Reference resolution, constraints on coreference, algorithm for pronoun resolution Text Coherence Discourse structure |
| 40 | Generation – Overview |
| 41-42 | Machine Translation - Overview |

Term Assignment:

Students will be assigned an activity (via the course web-page) related to the theme of the course. They will explore literature, do work-out, and prepare a write-up. They shall have to submit the write-up and give a presentation on their findings, by end of October 2018.

(b) Evaluation plan

| | |
|-----------------------|------------|
| Test I | 25 |
| Mid Term (Major-I) | 40 |
| Test III (Assignment) | 25 |
| End Term | 60 |
| Total | 150 |

6. Pedagogy :

Teaching-learning methods to be used-
Lecture and Discussion
Term assignment
Class assignments

7. Expected outcome: Towards the end of the course the student would understand the utility of natural language processing, the issues involved, methods to deal with the issues, visualise NLP applications, and how to build resources useful for NLP.