

Natural Language Processing

- 1.1 Course Number: CS466
- 1.2 Contact Hours: 3-0-0] Credits: 9
- 1.3 Semester-offered: 4th Year-Odd
- 1.4 Prerequisite: Linear algebra, probability and statistics, artificial intelligence and neural networks, basic programming.
- 1.5 Syllabus Committee Member: Dr. Sushum Biswas, Dr. Daya Sagar Gupta & Dr. Gargi Srivastava

2. **Objective:** Upon successful completion of this subject students should be able to:
- i) Explain the advantages and disadvantages of different NLP technologies and their applicability in different business situations.
 - ii) Use NLP technologies to explore and gain a broad understanding of text data.
 - iii) Use NLP methods to analyse sentiment of a text document.
 - iv) Use NLP methods to perform topic modelling.
 - v) Organise and implement a NLP project in a business environment.
 - vi) Interpret the results of a NLP project.

3. **Course Content:**

Unit-wise distribution of content and number of lectures

Unit	Topics	Sub-topic	Lectures
1	Introduction	NLP tasks in syntax, semantics, and pragmatics. Applications such as information extraction, question answering, and machine translation. The problem of ambiguity. The role of machine learning. Brief history of the field.	9
2	N-gram Language Models	The role of language models. Simple N-gram models. Estimating parameters and smoothing. Evaluating language models.	5
3	Part Of Speech Tagging and Sequence Labeling	Lexical syntax. Hidden Markov Models (Forward and Viterbi algorithms and EM training).	3
4	Basic Neural Networks	Any basic introduction to perceptron and backpropagation	3
5	LSTM Recurrent Neural Networks	LSTM Recurrent Neural Networks	2
6	Syntactic parsing	Grammar formalisms and treebanks. Efficient parsing for context-free grammars (CFGs). Statistical parsing and	7

		probabilistic CFGs (PCFGs). Lexicalized PCFGs. Neural shift-reduce dependency parsing.	
7	Semantic Analysis	Lexical semantics and word-sense disambiguation. Compositional semantics. Semantic Role Labeling and Semantic Parsing.	5
8	Information Extraction (IE)	Named entity recognition and relation extraction. IE using sequence labeling.	3
9	Machine Translation (MT)	Basic issues in MT. Statistical translation, word alignment, phrase-based translation, and synchronous grammars.	3
		Total	40

4. Readings

4.1 Textbook:

An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Second Edition by Daniel Jurafsky and James H. Martin

Manning, Christopher D., and Hinrich Schütze. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press, 1999. ISBN: 0262133601.

4.2 Reference books:

NLP tutorial: <https://www.upf.edu/web/mtg/nlp-tutorial>

Foundations of Statistical Natural Language Processing: <https://nlp.stanford.edu/fsnlp>

The structure of modern English: <https://muse.jhu.edu/article/19425>

5 Outcome of the Course: Upon completion of this course, the students will be able to:

1. Understand AI's fundamental concepts and methods
2. Acquire knowledge of modern AI tools, including Deep Learning framework Tensorflow and Deep Learning capabilities of RapidMiner.
3. Learn how to apply AI-based methods to solving practical business problems
4. Understand implications of AI for business strategies
5. Examine where the AI technologies are heading within the next few years.