

Natural Language Processing - Video course

COURSE OUTLINE

Sound : Biology of Speech Processing; Place and Manner of Articulation; Word Boundary Detection; Argmax based computations; HMM and Speech Recognition.

Words and Word Forms : Morphology fundamentals; Morphological Diversity of Indian Languages; Morphology Paradigms; Finite State Machine Based Morphology; Automatic Morphology Learning; Shallow Parsing; Named Entities; Maximum Entropy Models; Random Fields.

Structures : Theories of Parsing, Parsing Algorithms; Robust and Scalable Parsing on Noisy Text as in Web documents; Hybrid of Rule Based and Probabilistic Parsing; Scope Ambiguity and Attachment Ambiguity resolution.

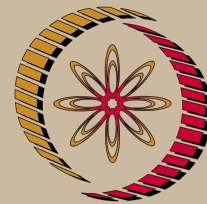
Meaning : Lexical Knowledge Networks, Wordnet Theory; Indian Language Wordnets and Multilingual Dictionaries; Semantic Roles; Word Sense Disambiguation; WSD and Multilinguality; Metaphors; Coreferences.

Web 2.0 Applications : Sentiment Analysis; Text Entailment; Robust and Scalable Machine Translation; Question Answering in Multilingual Setting; Cross Lingual Information Retrieval (CLIR).

COURSE DETAIL

A video course shall consist of 40 or more lectures with 1 hour duration per lecture.

Lecture No.	Topics
1	Introduction
2	Machine Learning and NLP
3	ArgMax Computation
4	WSD : WordNet
5	Wordnet; Application in Query Expansion
6	Wiktionary; semantic relatedness
7	Measures of WordNet Similarity
8	Similarity Measures (contd.)
9	Resnick's work on WordNet Similarity



NP-TEL

NPTEL

<http://nptel.iitm.ac.in>

Computer Science and Engineering

Pre-requisites:

1. A previous course on Artificial Intelligence will help.
2. Courses of Data Structures and Algorithms should have been done.
3. Exposure to Linguistics is useful, though not mandatory.

Additional Reading:

1. Radford, Andrew et. al., Linguistics, An Introduction, Cambridge University Press, 1999.
2. Journals : Computational Linguistics, Natural Language Engineering, Machine Learning, Machine Translation, Artificial Intelligence.
3. Conferences : Annual Meeting of the Association of Computational Linguistics (ACL), Computational Linguistics (COLING), European ACL (EACL), Empirical Methods in NLP (EMNLP), Annual Meeting of the Special Interest Group in Information Retrieval (SIGIR), Human Language Technology (HLT).

Hyperlinks:

<http://www.cse.iitb.ac.in/~cs626-449>

Coordinators:

Prof. Pushpak Bhattacharyya
Department of Computer Science and Engineering IIT Bombay

10	Parsing Algorithms
11	Evidence for Deeper Structure; Top Down Parsing Algorithms
12	Noun Structure; Top Down Parsing Algorithms- contd
13	Non-noun Structure and Parsing Algorithms
14	Probabilistic parsing; sequence labeling, PCFG
15	Probabilistic parsing; PCFG (contd.)
16	Probabilistic parsing: Training issues
17	Arguments and Adjuncts
18	Probabilistic parsing; inside-outside probabilities
19	Speech : Phonetics
20	HMM
21	Morphology
22	Graphical Models for Sequence Labelling in NLP
23	Graphical Models for Sequence Labelling in NLP (contd.)
24	Phonetics
25	Consonants (place and manner of articulation) and Vowels
26	Vowels (contd.)
27	Forward Backward probability; Viterbi Algorithm
28	Phonology
29	Sentiment Analysis and Opinions on the Web
30	Machine Translation and MT Tools - GIZA++ and Moses.

31	Text Entailment
32	POS Tagging.
33	Phonology; ASR, Speech Synthesis
34	HMM and Viterbi
35	HMM and Viterbi (contd)
36	Precision, Recall, F-score, Map
37	Semantic Relations; UNL; Towards Dependency Parsing.
38	Universal Networking Language
39	Semantic Role Extraction
40	Baum Welch Algorithm; HMM training
41	Baum Welch Algorithm; HMM training

References:

1. Allen, James, Natural Language Understanding, Second Edition, Benjamin/Cumming, 1995.
2. Charniack, Eugene, Statistical Language Learning, MIT Press, 1993.
3. Jurafsky, Dan and Martin, James, Speech and Language Processing, Second Edition, Prentice Hall, 2008.
4. Manning, Christopher and Heinrich, Schutze, Foundations of Statistical Natural Language Processing, MIT Press, 1999.