Natural Language Processing

Introduction

Knowledge Language Processing, Ambiguity, Models and Algorithms, Language, Thought, and Understanding

Regular expressions and automata

Regular Expressions, Basic Regular Expression Patterns**,** Disjunction, Grouping, and Precedence, Advanced Operators, Regular Expression Substitution, Finite-State Automata for morphological analysis

Words & Transducers

English and Indian Languages Morphology, Inflectional Morphology, Derivational Morphology, Finite-State Morphological Parsing, The Lexicon and Morphotactics, Morphological Parsing with Finite-State Transducers, Orthographic Rules and Finite-State Transducers, Combining FST Lexicon and Rules, Lexicon-Free FSTs, Stemming.
(Assignment : Analysis of Porter Stemmer and modify to work as a Morph Analyzer)

Syntex and Lexical Semantics

Syntex and different issues, relations among lexemes and their senses, Word Classes, Relations of words as Homonymy, Polysemy, Synonymy, Hyponymy, Antonymy, Hypernomy and Meronomy, Wordnet: A Database of Lexical relations, Internal structure of words
(Assignment: Study of Wordnet)

N-Grams

Counting Words in Corpora, Simple (Unsmoothed) N-grams, N-grams and Their Sensitivity to the Training Corpus, Smoothing, Add-Qne Smoothing, Witten-Bell Discounting, Good-Turing method

Parts-of-Speech Tagging

English word Classes, Tagsets, HMM POS tagging, Formalizing HMM taggers, Forward Algorithm, Viterbi Algorithm for HMM tagging
(Assignment: Analysis of POS tagger and Chunker).

Context-Free Grammars for English

Constituency, Context-Free Rules and Trees, Sentence-Level Constructions, Noun Phrase, Before the Head Noun, After the Noun, Coordination Agreement, The Verb Phrase and Subcategorizatkm, Auxiliaries, Spoken language Syntax, Disfluencies, Grammar Equivalence and Normal Form, Finite-State and Context-Free Grammars

Semantic Analysis

Syntax-Driven Semantic Analysis, Semantic Augmentations to Context Free Grammar Rules, Attachments for a Fragment of English Sentences, Noun Phrases, Verb Phrases, Prepositional Phrases, Idioms and Multi word expressions

Syntactic Parsing

Parsing as Search, Top-Down Parsing, Bottom-Up Parsing, Comparing Top-Down and Bottom-Up Parsing, Structural Ambiguity, A Basic Top-Down Parser, Problems with the Basic Top-Down Parser, Left-Recursion, Ambiguity, Repeated Parsing of Subtrees, The Earley Algorithm.
(Assignment: Analysis of Backtracking algorithm and Earley Algorithm)

Word Sense Disambiguation

Ambiguities (words – category, senses), Knowledge based approach (Lesk Algorithm and Simplified Lesk Algorithm), Machine Learning based approaches (Supervised approaches and unsupervised approaches)
(Assignment: Lesk Algorithm)

Named Entity Recognition

Issues, rules to extract NER and method
(Assignment: Analysis of NER tool)

Discourse

Reference Phenomena, Syntactic and Semantic Constraints on dereference, Preferences in Pronoun Interpretation, An Algorithm for Pronoun Resolution, The Phenomenon, An Inference Based Resolution Algorithm, Discourse Structure, Lappin and Lease Algorithm and Hobb's algorithm.
(Assignment: implementation of Lappin and Lease Algorithm and Hobb's algorithm

Applications with work flow: Co-reference Resolution, Machine Translation, Document Summarization

Text Book:

1. Speech and Language Processing, Daniel Jurafsky and James H. Martin, Pearson Education International

Ref. Book:

1. Natural Language understanding, James Allen, second edition, Pearson