

COURSE DESCRIPTION

Department and Course Number	CS 462	Course Coordinator	Thamar Solorio
Course Title	Natural Language Processing	Total Credits	3

Current Catalog Description

This course provides a broad introduction to Natural Language Processing (Computational Linguistics). Topics typically covered in this course include part-of-speech tagging, syntactic parsing, semantic analysis, speech recognition, machine translation, sequence labeling algorithms, n-gram language models, statistical parsing, grammar formalisms and treebanks.

Textbook

SPEECH and LANGUAGE PROCESSING, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Second Edition, by Daniel Jurafsky and James H. Martin, Prentice Hall, 2008

Additional Readings

Foundations of Statistical Natural Language Processing, by Christopher D. Manning and Hinrich Schütze, The MIT Press, 1999

Introduction to Information Retrieval by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Cambridge University Press. 2008 (online version available at <http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html>).

Course Goals

In this course, students will study algorithms and techniques for developing computational models for analyzing, understanding, and generating human language, including parsing techniques, semantic representations, discourse analysis, and statistical and corpus-based methods for text processing and knowledge acquisition. By the end of the course students will have a good understanding of and appreciation for natural language processing and the general issues related to syntax, semantics, and pragmatics of language. Throughout the course students will be exposed to applications that can benefit from automatic language processing, such as information extraction, question answering, machine translation and spoken language understanding, and the state-of-the-art-techniques involved in these applications.

Prerequisites by Topic

CS 350 Automata and Formal Language Theory
CS 455 Probability and Statistics for Computer Science or CS 460 Artificial Intelligence
(with grades of “C” or better)

Major Topics Covered in the Course

The lexicon and morphology
N-gram language models
Part of speech tagging
Syntactic parsing, grammar formalisms
Semantic analysis, word sense disambiguation
Information extraction
Automatic document summarization
Information retrieval
Machine translation

Laboratory projects (specify number of weeks on each)

Several programming assignments of various durations

Estimate CSAB Category Content

	CORE	ADVANCED		CORE	ADVANCED
Data Structures		X	Computer Organization and Architecture		
Algorithms		X	Concepts of Programming Languages		X
Software Design					

Oral and Written Communications

Every student is required to submit at least 2 written reports (not including exams, tests, quizzes, or commented programs) of typically 3-5 pages and to make 1 oral presentations of typically 10-20 minutes duration.

Social and Ethical Issues

None

Theoretical Content

Grammar formalisms (30%) – Probability and statistics (30%) – English morphology and syntax (5%)

Problem Analysis

Some course projects will require students to implement abstract algorithms from scratch and apply them to specific NLP tasks. These projects will allow students to experiment and acquire hands-on experience of the systems discussed in class. Other projects will include implemented versions of algorithms discussed in class and they will give students the opportunity for extensions and experimentation.

Solution Design

Course projects, as described above, will give students the opportunity to improve their problem analysis and solution design skills, as they will require the implementation of functional computer programs, as well as performance evaluation of their solutions.