

ECE8813 Special Topic on Statistical Language Processing (C.-H. Lee)

Course Description:

With the availability of a large collection of text corpora in electronic forms in libraries and on the web, statistical language processing is becoming an important tool for understanding words and their interactions. Many signal processing tools are applicable to language analysis. Language engineering also emerges as a new area for research and applications. In this course, foundations of statistical language processing will be covered, and many applications will be addressed. Plenty of text examples will be used for hands-on homework exercises. An individual course project is the main outcome of this special topic course.

Intended Audience: students interested in learning about language engineering

Course Outline:

- Introduction
- Mathematical Foundations
- Linguistics Essentials
- Corpus-Based Linguistic Analysis
- Word Collocations
- Statistical n -Gram
- Word Sense Disambiguation
- Markov Models
- Part-of-Speech Tagging
- Probabilistic Parsing
- Clustering
- Information Retrieval
- Text Categorization
- Statistical Alignment and Machine Translation

Grading Policy: Students are graded based on the following:

- **Homework (25%)**
- **Project (35%): a project presentation during the Final Week is required**
- **Examinations (40%)**
 - Midterm (15%):
 - Final (25%):

Prerequisites: ECE3075 or equivalent (ISYE 3770 or MATH 3770 or CEE 3770)

Estimated Student Size: 12-15 (with potential attendees from CoC)

Main Text: C. D. Manning and H. Schutze, *Foundations of Statistical Natural Language Processing*, MIT Press, 2001

Supplemental Reading: will be furnished during the semester