

COSC 6397 Natural Language Processing

Fall 2014

MW 1:00 - 2:30 PM, SW 221

Instructor: [Arjun Mukherjee](#)

Course webpage: <http://www.cs.uh.edu/~arjun/courses/nlp>

1. Overview

This is a senior level natural language processing course (NLP). The course is intended for developing foundations in NLP and text mining. The broader goal is to understand how NLP tasks are carried out in the real world (e.g., Web) and how to build tools for solving practical text mining problems. Throughout the course, large emphasis will be placed on tying NLP techniques to specific real-world applications through hands-on experience. The course covers fundamental topics in statistical machine learning and touches upon topics in sentiment analysis and psycholinguistics.

2. Prerequisites

The course requires basic background in mathematics and sufficient programming skills. If you have taken and did well in one or more of the equivalent courses/topics such as Algorithms, Artificial Intelligence, Numerical methods, or have some background in probability/statistics, it will be helpful. The course however reviews and covers required mathematical and statistical foundations. Sufficient experience for building projects in a high level programming language (e.g., Java) will prove beneficial.

3. Course Contents

NLP Resources, Language as a probabilistic phenomenon, Zipf's law, Word collocations, NLP and text retrieval basics, Probability theory, Conditional probability and independence, Random variables, density and mass functions, Mean, Variance, Common families of distributions, Multiple random variables: joints and marginal, Collocations, Hypothesis testing, statistical tests, p-values, N-gram language models, Hidden markov model (HMM), Part of speech tagging, Decision trees, Naive Bayes, Support Vector Machines, Evaluation metrics, Significance testing, Feature selection schemes, Latent semantics and clustering problem, Introduction to Bayes nets and PGMs, Latent Dirichlet Allocation, Aspect extraction, Deception and Opinion spam.

4. Grading

- Homework: 50%
- Mini project: 25%
- Final Exam: 25%