

CMPT 825

Natural Language Processing

Anoop Sarkar

<http://www.cs.sfu.ca/~anoop>

Goals of the Course

- Convince you that understanding language is a subtle, interesting, and tractable problem
- Give insight into various algorithms including statistical machine learning algos
- Hands on experience with non-trivial data
- Experience with formal models and their implementation

Course Webpage

- Office hours by appointment
- Check web page regularly (Announcements section for important dates)
- Textbook: Manning and Schutze
- Syllabus and readings are listed for each week
- Homework data and software also linked to webpage

Your Responsibilities

- One (short) homework every week before mid-term exam
- A homework every two weeks after midterm (20%)
- Midterm Exam (20%)
- Class presentation (20%)
- Final Project (40%)

Guidelines for Class Presentation

- 20 mins (no more than 10 slides). You can simply have a handout, slides are not necessary
- Source of software (motivation for existence)
- Technical Background (algorithms implemented, trade-offs in design)
- Programs in toolkit (invocation and input/output format, error checking)
- Potential NLP applications (demo)

Perl

- Used in several homeworks
- Still the most popular scripting language for NLP: quick prototyping and low-level OS support
- Powerful regular expression library
- Other popular languages include Python (in my opinion, better to combine Perl with C++)

Software and Datasets

- For some homeworks we will use standard software packages that are implementations of the models and algorithms we will study
- We will use Linux. Use the machines in the CSIL labs or the grad area
- You can also install Perl on your own machines (visit www.perl.com)