CS 6740 / INFO 6300
Advanced Language Technologies

Graduate-level introduction to technologies for the computational treatment of information in human-language form,

covering **natural-language processing** (NLP) and/or **information retrieval** (IR).

Possible topics include text categorization and clustering, information extraction, latent semantic analysis (LSI), click-through data for web search, language modeling, computational syntactic and semantic formalisms, grammar induction, and machine translation.

Focus: RESEARCH

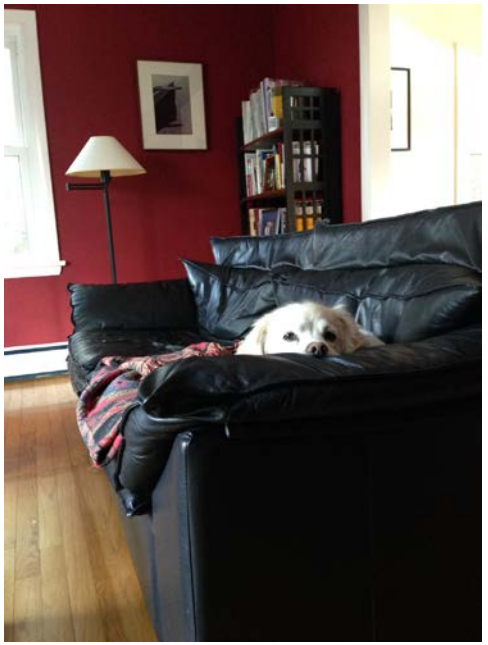
- The goal of the course is to teach you how to **evaluate** and **conduct** leading-edge research in natural language processing (NLP).
  - Tailor the course to your specific interests

Plan for Today

- Introductions
- Course structure and requirements

Introductions

- I'll start…
  - Who you are
  - Why you are interested in the course
  - If you could be any animal, what would it be?



What topics will we cover?

Language modeling
Lexical semantics and word-sense disambiguation
Part-of-speech tagging and HMMs
Parsing
Semantic analysis
Discourse processing
Coreference analysis
NL Generation
Machine translation

Information extraction
Information retrieval models
Sentiment analysis
Text categorization
Question answering
Summarization
…

It depends on you!!!!!

Start as in conference reviewing

- Bidding on papers to review
  - Essentially a survey of your interests
  - NAACL 2015
    - » http://naacl.org/naacl-hlt-2015/schedule.html
  - ACL 2015
    - » http://acl2015.org/complete_program.html
  - EMNLP 2015
    - » http://www.emnlp2015.org/program.html#session-1A

https://aclweb.org/anthology/

Classes

- Research paper presentations and discussions
  - Two per class
- Project-related presentations, exercises and discussions
  - Project proposals
  - Final project presentations

Classes

- Research paper presentations/ discussions
  - First 5-6 weeks
- Project-relat… …ions, exercis… …ssions
  - …sals
  - … weeks
  - …ossibly project-related exercises
  - Final project presentations
    - » During reading period

This is all subject to change depending on how the class evolves!!!

Requirements

- Research paper presentations and discussions
  - Paper 1: review
    - Graded (✓, ✓+, ✓- and written comments) ***by the presenter***
  - Paper 2: research idea generation
    - Graded (✓, ✓+, ✓- and written comments) ***by the presenter*** and Claire
- Project
  - Proposal, related work, pilot studies, preliminary results, final paper, presentation

Prereqs

- Prerequisites
  - Permission of instructor. Neither INFO/CS4300 nor CS4740 are prerequisites.
  - You will have trouble if you do have not at least SOME background in: machine learning or stats or probability or experiment design
  - Useful references:
    - Jurafsky and Martin, *Speech and Language Processing*, Prentice-Hall, **2nd edition**.
    - Manning and Schutze. *Foundations of Statistical NLP*, MIT Press, 1999.

Grading

I am most interested in productive research-oriented discussion participation (in class and on Piazza), interesting research proposals, and a good-faith final research project.

For the near term…

- Sign up for Piazza – watch for my email on that…
- Paper bidding --- by Saturday
- Thursday
  - Overview of Claire's current research
- Next week
  - Claire presents research papers