CS674 Natural Language Processing

Topics for today
- General introduction to NLP
  » Why study NLP?
- Handouts
  » Class description and syllabus
  » Student info sheet

Natural language and NLP
- "natural" language
  » Languages that people use to communicate with one another
- Ultimate goal
  » To build computer systems that perform as well at using natural language as humans do
- Immediate goal
  » To build computer systems that can process text and speech more intelligently

Dialogue systems
- Require both understanding and generation
  - Dave: Open the pod bay doors, HAL.
  - HAL: I'm sorry Dave, I'm afraid I can't do that.
  - Dave: What's the problem?
  - HAL: I think you know what the problem is just as well as I do.

Why study NLP?
- Useful applications...
  - E.g. information retrieval

Topic: Advantages and disadvantages of using potassium hydroxide in any aspect of organic farming, especially...
Why study NLP?

- **Useful applications…**
  - E.g. question answering systems
    - How many calories are there in a Big Mac?
    - Who is the voice of Miss Piggy?
    - Who was the first American in space?
  - Retrieve not just relevant documents, but return the answer
    - answer + supporting text

Why study NLP?

- **Useful applications…**
  - E.g. summarization

```
AP, Saturday, January 19, 1997
SALVADOR — A major earthquake struck Central America on Saturday, cracking buildings, buckling roads and knocking out water systems and telephone service. At least two people were reported killed and four injured in Guatemala.

The death toll is expected to rise as emergency crews inspect the damage done. The quake was centered 56 miles off the Salvadoran coast. It collapsed buildings and knocked out electricity throughout Guateamala and El Salvador. It was said to be strong enough to be felt in Mexico City hundreds of miles north.

The Associated Press reported that a worker in San Salvador was injured in the earthquake, injuring scores of homes. The death toll was uncertain, but reports of at least 100 bodies, apparently dead, pulled from the debris in the middle class Los Cerritos neighborhood west of the capital. Hundreds of rescuers frantically scrambled over the scene trying to save those still buried.
```
[White et al., 2002]

Why study NLP?

- **Interdisciplinary…**
  - Linguistics
    - models for language
  - Psychology and psycholinguistics
    - models of cognitive processes/language
  - Mathematics
    - studies properties of formal models, methods of inference from these models
  - vs. NLP
    - Computational study of language use
      - Definite engineering aspect in addition to a scientific one
        - Engineering: to enable effective human-machine communication
        - Scientific: to explore the nature of linguistic communication
      - Emphasis on computational, not cognitive plausibility
      - Models of language: optional
Why study NLP?

- Challenging…
  - AI-complete
    » To solve NLP, you’d need to solve all of the problems in AI
  - Turing test
    » Posits that engaging effectively in linguistic behavior is a sufficient condition for having achieved intelligence.
- …But little kids can “do” NLP…
  - Next time: Why is NLP hard?

Syllabus (tentative)

- Introduction
- History and state-of-the-art
- Morphology
- N-grams
- Context-sensitive spelling correction
- Part-of-speech tagging and HMMs
- Parsing
- Partial parsing
- Semantic analysis
- Inference and world knowledge
- Information extraction
- Lexical semantics and word-sense disambiguation
- Discourse processing
- Generation
- Machine translation

Additional Course Info

- Time: Mondays and Wednesdays, 11:15-12:05
  - Possibly occasional Fridays
- Office hours: Monday 1-2, Thursdays 3-4
- Course Materials:
  - Lecture Notes, Readings, Assignments
  - Critique and Project Information
  - Lillian Lee's list of on-line NLP resources

Reference Material

- Required text book:
- Other useful references:
  - Others listed on course web page…
Prereqs and Grading

- **Prerequisites**
  - Elementary computer science background, elementary knowledge of probability, familiarity with context-free grammars, some background in machine learning.

- **Grading**
  - 20%: critiques of selected readings and research papers
  - 20%: presentations of critique papers
  - 50%: final project. Grade based on
    - (1) preliminary project proposal (Mon 3/7),
    - (2) project literature survey (Mon 4/6),
    - (3) project presentation (last week of classes),
    - (4) final write-up (Mon 5/9).
  - 10%: participation

Readings and Critiques