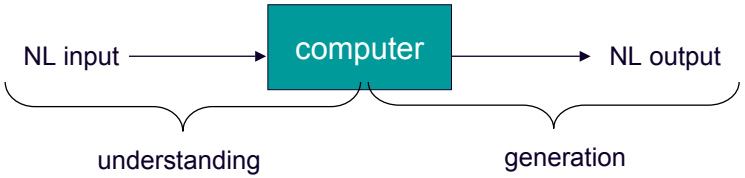
CS674 Natural Language Processing

- 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

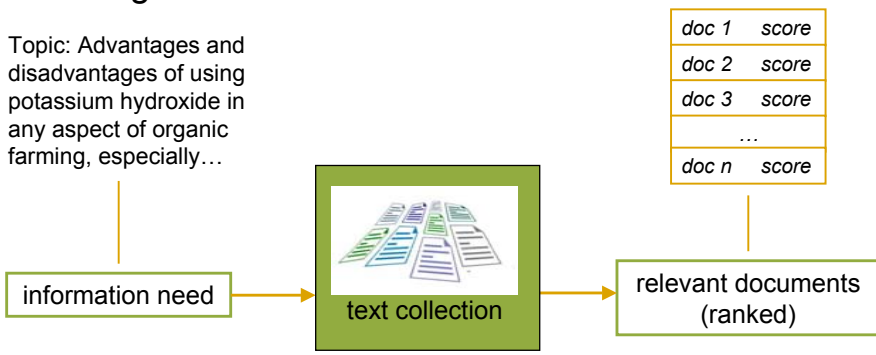


Dialogue systems

- 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

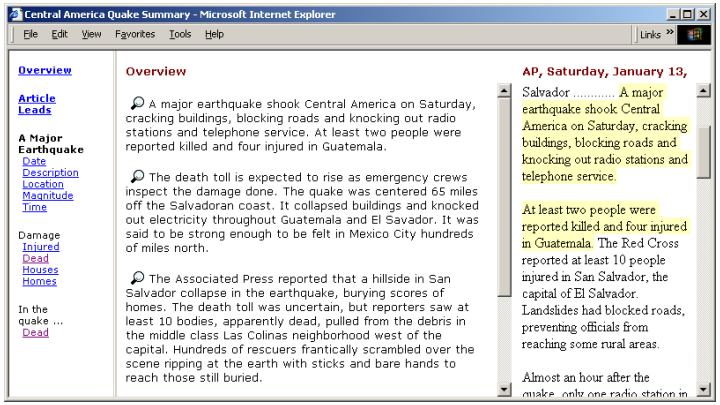


Why study NLP?

- 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

Why study NLP?

- Useful applications…
  - E.g. summarization



[White et al., 2002]

Why study NLP?

- Useful applications…
  - E.g. machine translation
    - » Would clearly facilitate human-human communication
    - » Certainly see a need for it…
      - The extension of the coverage of the health services to the underserved or not served population of the countries of the region was the central goal of the Ten-Year Plan and probably that of greater scope and transcendence.
      - Welcome to Chinese Restaurant. Please try your Nice chinese Food With chopsticks. the traditional and typical of Chinese glorious history and cultual. PRODUCT OF CHINA

Bill Gates, 1997 "…now we're betting the company on thse natural interface technologies"

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
  - Other Handouts
  - Lillian Lee's list of on-line NLP resources

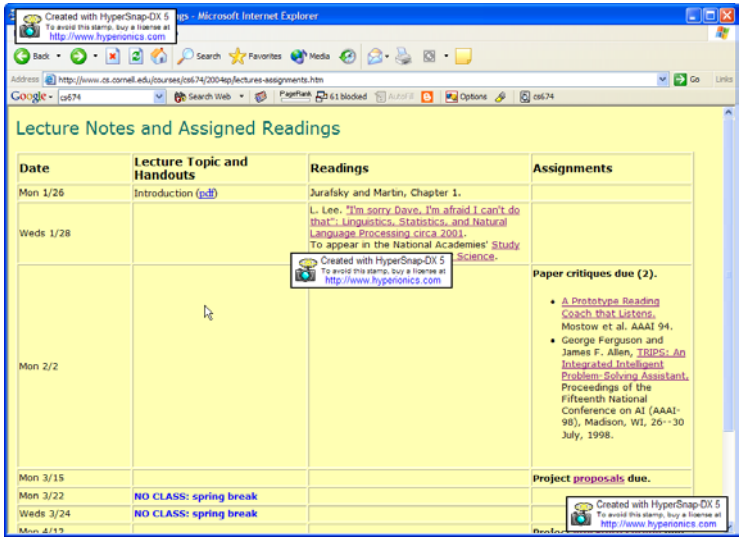
Reference Material

- Required text book:
  - Jurafsky and Martin, *Speech and Language Processing*, Prentice-Hall, 2000.
- Other useful references:
  - Manning and Schutze. *Foundations of Statistical NLP*, MIT Press, 1999.
  - James Allen. *Natural Language Understanding*, 2nd edition.
  - Eugene Charniak. *Statistical Language Learning*, MIT Press, 1996.
  - Frederick Jelinek. *Statistical Methods for Speech Recognition*, MIT Press, 1998.
  - 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
  - 30%: critiques of selected readings and research papers
  - 60%: final project. Grade based on
    - » (1) preliminary project proposal (Mon 3/15),
    - » (2) project literature survey (Mon 4/12),
    - » (3) project presentation (study week: Mon 5/10, Weds 5/12),
    - » (4) final write-up (Mon 5/17).
  - 10%: participation

Readings and Critiques



Information Sheet

- Please fill one out…