# NLP

# Introduction to Natural Language Processing

*Introduction*

# Introduction

- Instructor:
  - Dragomir Radev (PhD, Computer Science, Columbia University)
- Course duration:
  - 20 hours: 12 weeks x 1.5–2 hours per week
- Intended audience:
  - Primarily undergraduate students in Computer Science and Linguistics, Informatics,
  - Possibly also Mathematics, Statistics, Management, and Engineering
- Related courses
  - More introductory than Collins on Coursera (2013)
  - More focused on linguistics and resources than Jurafsky and Manning on Coursera (2012)

# What is Natural Language Processing

- Natural Language Processing (NLP) is the study of the computational treatment of natural (human) language.
- In other words, teaching computers how to understand (and generate) human language.

# Quiz

- Where is this quote from?

> **Dave Bowman**: Open the pod bay doors, HAL.
> **HAL**: I'm sorry Dave. I'm afraid I can't do that.

# Quiz answer

- "2001: A Space Odyssey"
  - 1968 film by Stanley Kubrick
  - based on a joint screenplay with Arthur C. Clarke.

# Modern Applications

- Search engines (Google, Yahoo!, Bing, Baidu)
- Question answering (IBM's Watson)
- Natural language assistants (Apple's Siri)
- Translation systems (Google Translate)
- News digest (Yahoo!)
- Automatic earthquake reports (LA Times)

# Notes

- Computers are confused by (human) language
  - Specific techniques are needed
  - NLP draws on research in Linguistics, Theoretical Computer Science, Mathematics, Statistics, Artificial Intelligence, Psychology, Databases, etc.
- Goals of this class
  - Understand that language processing is hard (and why)
  - Understand the key problems in NLP
  - Learn about the methods used to address these problems
  - Understand the limitations of these methods

# Language and Communication

- Speaker
  - Intention (goals, shared knowledge and beliefs)
  - Generation (tactical)
  - Synthesis (text or speech)
- Listener
  - Perception
  - Interpretation (syntactic, semantic, pragmatic)
  - Incorporation (internalization, understanding)
- Both
  - Context (grounding)

# Basic NLP Pipeline

- (U)nderstanding and (G)eneration

# NLP