Course Introduction

CS 485, Fall 2023
Applications of Natural Language Processing
https://people.cs.umass.edu/~brenocon/cs485_f23/

Brendan O'Connor
College of Information and Computer Sciences
University of Massachusetts Amherst
What

• Learn fundamental principles in natural language processing, and how to practically use them
• Hands-on experience: data collection and implementation
• Appreciation of basic linguistic issues
• Know when NLP works and when it doesn’t
• Course topics may include: text processing, corpus data, probabilistic language models, syntax models, word embeddings, text classification, topic analysis, sentiment analysis, question answering, deep learning methods, language generation...
Requirements

- (10%) Quizzes / exercises (approx. weekly)
- (30%) Problem sets
  - Written: math and concepts
  - Programs: in Python
- (20%) In-class Midterm
- (40%) Final projects (groups of 2-3)
  [Choose a topic, or select a suggested topic]
  - Project Proposal
  - Progress Report
  - In-class presentations
  - Final Report
TA: Erica Cai

• We will hold weekly office hours - to be announced
Resources

• Webpage will have lectures/readings and links to other sites
  • Reading for Thursday: J&M Ch 2 (except 2.5)
• Piazza for communication
• Gradescope for turning in assignments
• Video recordings in Echo360 (will try to link from webpage)
Upcoming

• Exercise #1: Class survey, on Gradescope. Due tomorrow
• First homework coming next week
• Let Erica know if you can't access Gradescope or Piazza
Ambiguity in language

Why NLP is hard...

- Red Tape Holds Up Bridges
- March Planned for Next August
- At the drop of a hat
A Huge Threat to the U.S. Budget Has Receded. No One Is Sure Why.

Instead of growing and growing, as it always had, spending per Medicare beneficiary has nearly leveled off over more than a decade.
Levels of Linguistic Structure

This is a simple sentence.
Levels of Linguistic Structure: Characters

This is a simple sentence.
Levels of Linguistic Structure: Morphology

<table>
<thead>
<tr>
<th>This</th>
<th>is</th>
<th>a</th>
<th>simple</th>
<th>sentence</th>
<th>.</th>
</tr>
</thead>
<tbody>
<tr>
<td>be</td>
<td>3sg</td>
<td>present</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Levels of Linguistic Structure: Parts of Speech

This is a simple sentence.

<table>
<thead>
<tr>
<th>DT</th>
<th>VBZ</th>
<th>DT</th>
<th>JJ</th>
<th>NN</th>
<th>PUNCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>This</td>
<td>is</td>
<td>a</td>
<td>simple</td>
<td>sentence</td>
<td>.</td>
</tr>
</tbody>
</table>

be
3sg
present
Levels of Linguistic Structure: Syntax

This is a simple sentence.

be 3sg present
Levels of Linguistic Structure: Semantics

This is a simple sentence.

be 3sg present

be SIMPLE1 having few parts

sentence SENTENCE1 string of words satisfying the grammatical rules of a language
Levels of Linguistic Structure: Discourse

This is a simple sentence.

But it is instructive.

CONTRAST

be 3sg present

SIMPLE1 having few parts

SENTENCE1 string of words satisfying the grammatical rules of a language
NLP Today: Speech Interfaces & Voice Assistants
NLP Today: Speech Interfaces & Voice Assistants

Ophelia, call the police

Sure. Playing F*** tha Police by N.W.A.
NLP Today: Search & Summarization

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

Jul 2, 2020

https://www.ibm.com - Cloud - Cloud Learn
NLP Today: Search & Summarization

- Have technology (thanks to R6) – for English, Arabic and Chinese
- Allow queries like:
  - Show me all the word documents with references to IAEO
  - Show me all documents that reference Osama Bin Laden
  - Will allow a ‘show me more like this’ capability
NLP Today: Machine Translation

《三体》(Novel) Wikipedia, the free encyclopedia

"Three Body" is a full-length science fiction novel serialized by the Chinese writer Liu Cixin in the "World of Science Fiction" magazine from May to December 2006. After publication, it became one of the best-selling science fiction novels in Mainland China (1). In 2008, the pamphlet of the book was published by Chongqing Publishing House. This book is the first part of the three-body series (the series was originally called the Trilogy on Earth). The second part of the series "Three-body II: Dark Forest" was published in May 2008. In November 2010, the third "Three-body III: Death Eternal Life" was published. In 2011, the "Earth Past Trilogy" was successively published in Taiwan. The English version of the novel was nominated for the 2014 "Nebula Award" by the American Association of Science Fiction and Fantasy Writers (2) and won the 2015 Hugo Award for Best Novel. (3)

author  Liu Cixin
series  China Science Fiction
structure  Cornerstone Series

Three-body
NLP Today: Machine Translation

By the sword we seek peace, but peace only under liberty
NLP Today: Sentiment Analysis

Harry Potter and the Deathly Hallows
by J.K. Rowling

A spectre is haunting Europe — the spectre of communism. All the powers of old Europe have entered into a holy alliance to exorcise this spectre: Pope and Tsar, Metternich and Guizot, French Radicals and German police-spies.

From other samples

- [...] with some seriously big names. We're talking the Pope and the Tsar, Metternich and Guizot, and even those fiery French Radicals and crafty German police-spies.

- Picture this: the Pope and the Tsar, Metternich and Guizot, everyone on board: the Pope and the Tsar, Metternich and Guizot, even the French Radicals and those sneaky German police-spies. It's a real clash of the titans, and they're ready to wage a battle against this spooky ideology. Hold on tight, because things are about to get interesting!