

Here We Go!

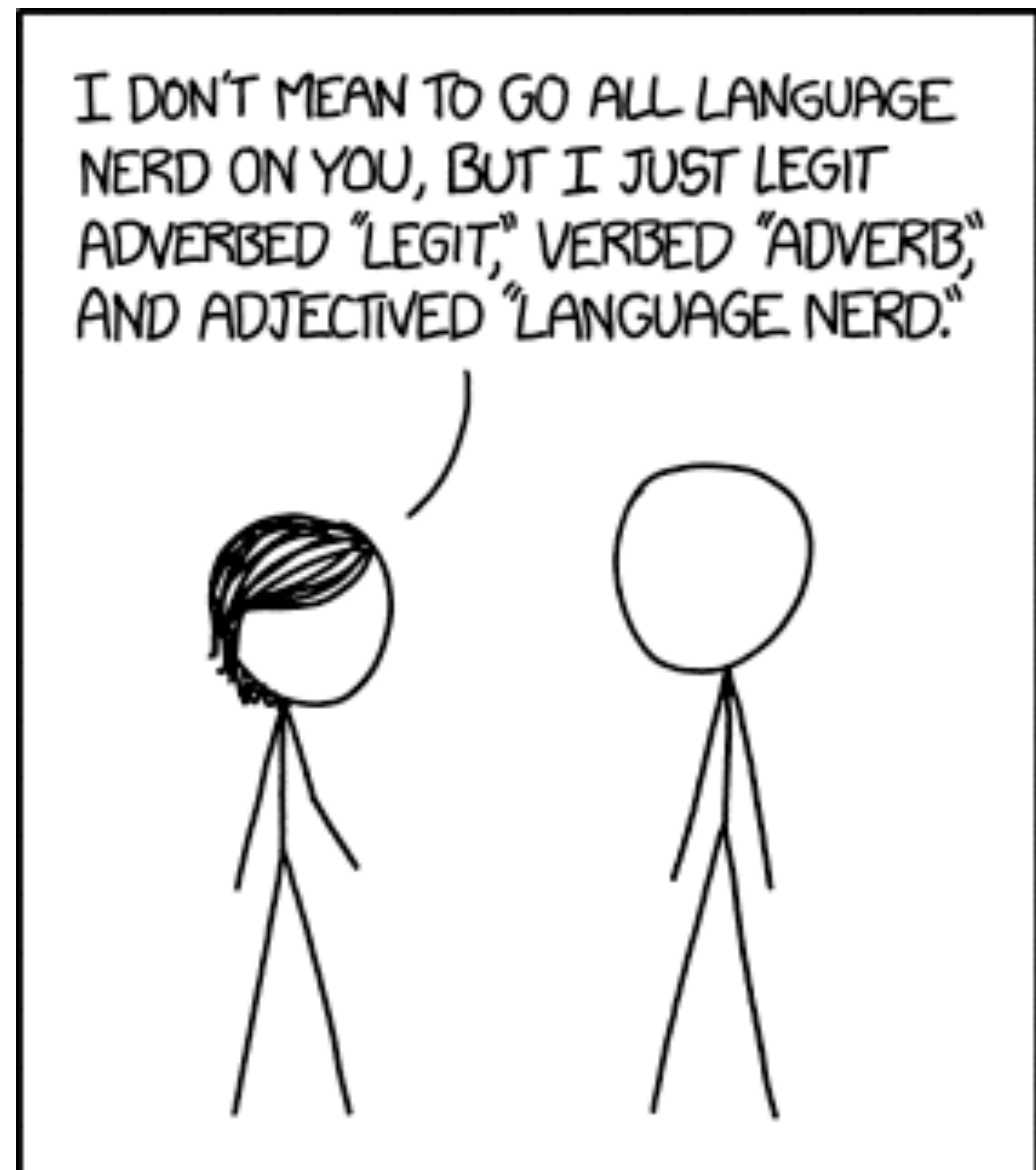
Wednesday, January 21, 2015

Plan for Today:

- Intro to CS159:
The Whos, Whys,
Whats, and Hows

Announcements:

- Research colloquium
tomorrow 4:15



<http://xkcd.com/1443/>

Who am I?

Julie Medero

julie@cs.hmc.edu

<http://cs.hmc.edu/~julie>

Olin 1269

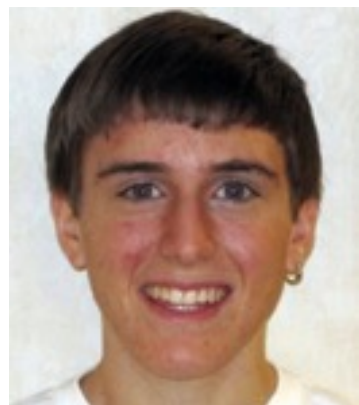
Office Hours:

M 3:30-5:00

Th 2:30-4:00



Who are You?



Why are you here?

This class will help me...

The one thing I really want to learn is...

I hope this class will...

What is NLP?

(NOT neuro-linguistic
programming!)

Computer Science
(Modeling, ML)

Natural Language
Processing

Human Languages
(Linguistics)

Signal Processing
(Electrical Engineering)

NLP TASKS

English Spanish French English - detected

English Hindi Zulu

Translate

Provide your full legal last name, first name, and middle initial.

अपना पूरा कानूनी अंतिम नाम, प्रथम नाम, और मध्य प्रारंभिक प्रदान करें.

Language Transformation

Diabetes mellitus

Consult a doctor if you have a medical concern.

Diabetes is usually a lifelong (chronic) disease in which there are high levels of sugar in the blood.

Causes - Symptoms - Tests - Treatment - Prevention

Language Understanding

Language Generation

Obesity

Natural Language is Hard!

- Differences between languages
- Real-world grounding
- Non-literal language
("kick the bucket")
- Plus some special challenges with "real" language data...

We are creative with language...



Sparseness
problem for
machine
learning

Language changes



Keri Ford @KeriFord

4h

zomg. seriously. if you're a big 5 publisher, the covers for ur digi line shld be just as good as your print! [#whywhywhy](#) [#SaveTheEyes](#)

Collapse

← Reply ↕ Retweet ★ Favorite ... More

4

RETWEETS

2

FAVORITES



8:30 AM - 31 Jan 2014 · Details

Generalization
challenges for
machine
learning

Language is Everywhere!

Diabetes mellitus type 1

From Wikipedia, the free encyclopedia

It has been suggested that *Honeymoon period (diabetes)* be merged into this article. (Discuss) Proposed since December 2013.

Los Angeles Times

Tuesday, October 14, 2008

U.S. to invest \$250 billion



facebook

Search

Mark Zuckerberg

Works at Facebook
California
Knows English
1984

Wall



Beast

Now living in my new home in Palo Alto with Mark Zuckerberg and Priscilla Chan.

11 hours ago · View Post

RECENT ACTIVITY

Mark likes Beast.

DM; formerly insulin dependent diabetes or juvenile
immune destruction of the insulin-producing beta cells in
blood and urine glucose. The classical symptoms are

Diabetes type 1

Classification and external resources



2013 Free and Reduced Priced School Meals Application (One Form per Household)

Incomplete applications will be denied

For reduced price meals, complete this application, sign your
name on the application to food service. Please call the following
number: 467-0947

If you have been notified this school year that your child
is approved for free meals, do not complete this form.

Information	Check Box if Foster Child*	No Income	Grade	Name of School
	<input type="checkbox"/>	<input type="checkbox"/>		
	<input type="checkbox"/>	<input type="checkbox"/>		
	<input type="checkbox"/>	<input type="checkbox"/>		
	<input type="checkbox"/>	<input type="checkbox"/>		
	<input type="checkbox"/>	<input type="checkbox"/>		
	<input type="checkbox"/>	<input type="checkbox"/>		

*Foster child must be legal responsibility of welfare agency or court

2. Food Stamp, TAFI, or FDIPIR Number

Name of person receiving benefit
(can be ANY member in household)

Case Number (EBT or Quest Card # Not Allowed)

Check one box below

☐ Food Stamp ☐ TAFI ☐ FDIPIR

3. Household Members and Income (not required if Food Stamp, TAFI, or FDIPIR Case number has been provided, or if all students are foster children)

List the names of everyone in your household and their gross income (only list students from #1 again if they have income). If your household member listed below has no income, you must check the "No Income" box. If they are a student already listed above, you must check the "Student" box. The "How Often" box must be answered if there is income in a category.			Earnings from Work Before Deductions ("How Often?" must be answered if income)		Welfare, Child Support, Alimony Received ("How Often?" must be answered if income)		Pensions, Retirement, Social Security ("How Often?" must be answered if income)		All Other Income ("How Often?" must be answered if income)	
Name	No Income	Student	How Much?	How Often?	How Much?	How Often?	How Much?	How Often?	How Much?	How Often?
	<input type="checkbox"/>	<input type="checkbox"/>								
	<input type="checkbox"/>	<input type="checkbox"/>								
	<input type="checkbox"/>	<input type="checkbox"/>								
	<input type="checkbox"/>	<input type="checkbox"/>								

Why are these funny?

- Iraqi Head Seeks Arms
- Ban on Nude Dancing on Governor's Desk
- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- Stolen Painting Found by Tree
- Kids Make Nutritious Snacks
- Local HS Dropout Cut in Half
- Hospitals Sued by 7 Foot Doctors

Ambiguity

- Lexical Ambiguity

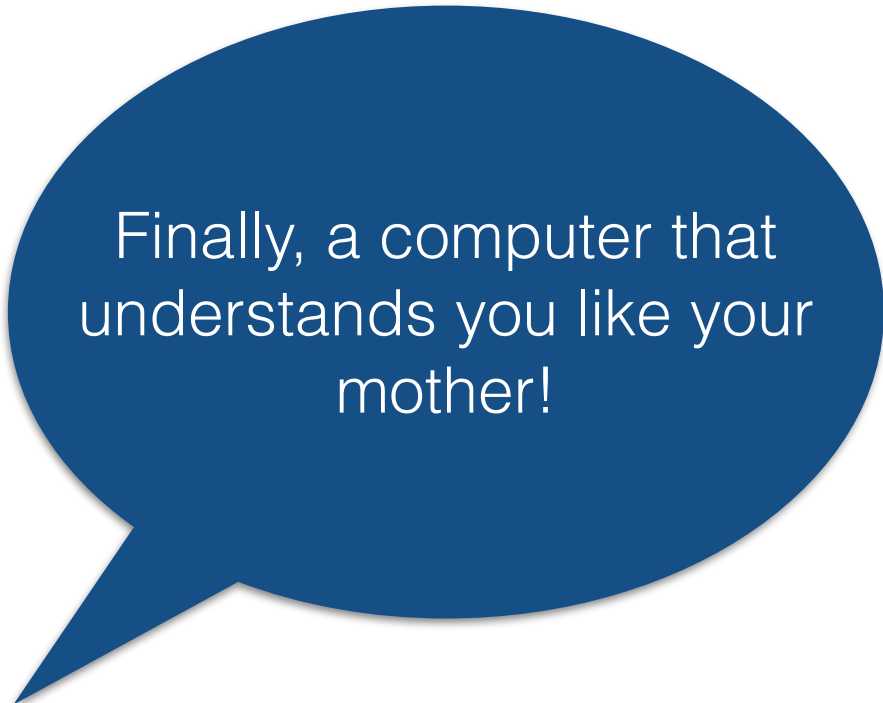
- I cooked waterfowl for her
- I cooked the waterfowl that is on the plate in front of her.
- I created a toy waterfowl for her.
- I caused her to quickly lower her head.

- Ambiguity we need auto to resolve:

- I made HER duck vs. I made her DUCK
- I made her duck vs. I made her duck?
- Ai made her duck. (Ai is a person's name)
- In some dialects: A maid heard "uck."



(example from Jurafsky & Martin)



Finally, a computer that
understands you like your
mother!

- Syntactic ambiguity
 - understands [(that) you like your mother]
 - understands [you] [like your mother (does)]
- Need world knowledge to interpret fully
 - Does your mother understand you well?

(example from Lillian Lee)

What do NLP researchers do?

- Linguistic analysis of sub-problems: Part of Speech tagging, automatic speech recognition, syntactic parsing
- Applications: (multi-document) summarization, question answering, translation, dialog systems, forensic linguistics, medical record processing, ...
- Complex systems: e.g. Analyst queries a system in English, gets back a summary of multi-lingual, multi-modal sources related to the query.

This class...

What will we learn?

Emphasize depth over breadth

Partially following an open-source, grad-level course on machine translation developed at JHU, UPenn, UIUC, U of Edinburgh: mt-class.org

Fill in some more background (linguistics, corpus processing)

Syllabus Overview

Part 1: Background (linguistics, computational tools)

Part 2: Machine Translation

Part 3: TBD (advanced topics OR breadth)

How will this all work?

Grading

Weekly “PReview” homework

- Short (1 hour, including reading)
- Due in class on Monday morning

Three big group projects

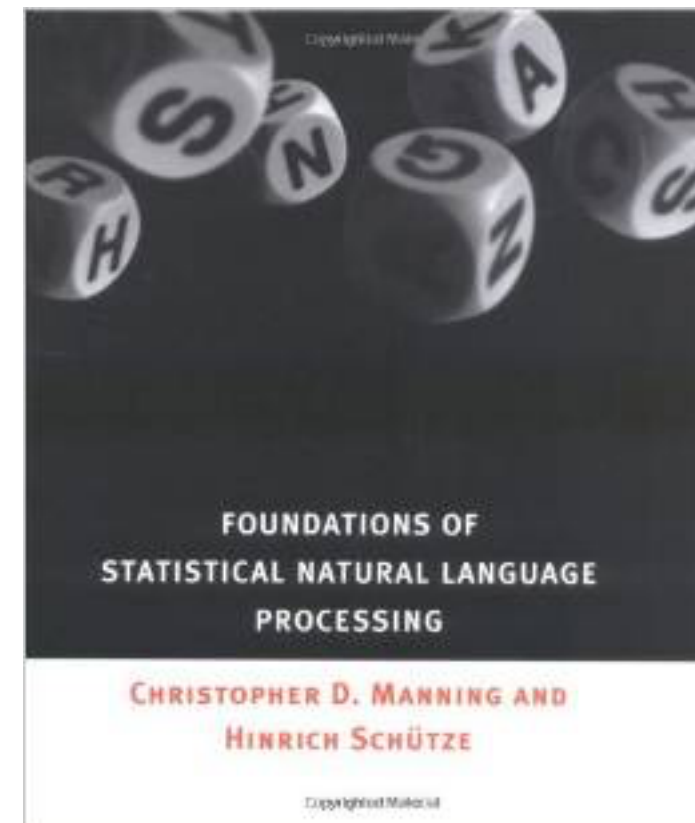
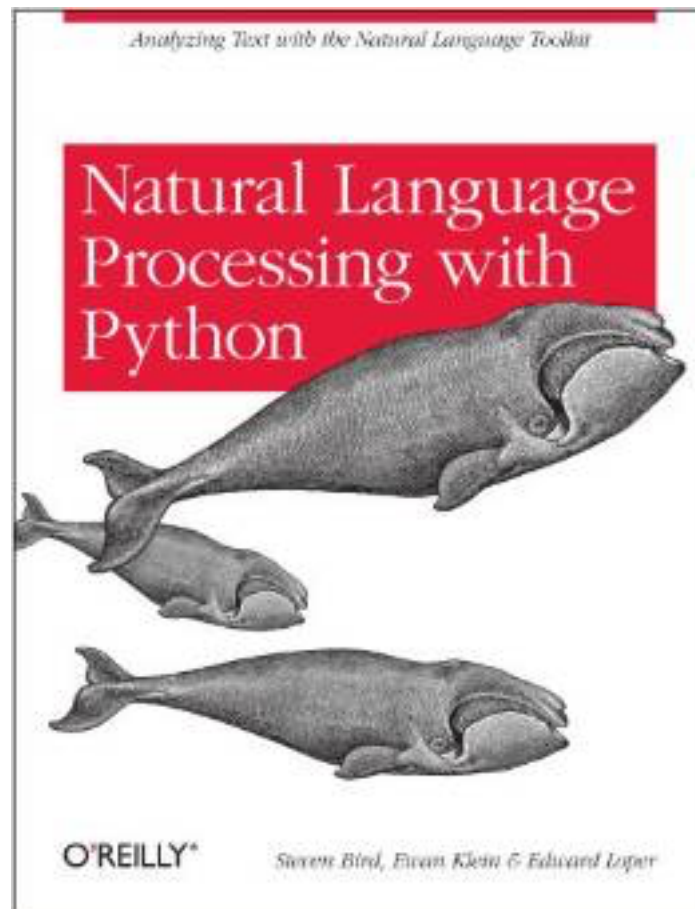
- All groups do the first part
- Each group does a different extension, then presents

“Language in 10” Presentation

- Present a language you don’t speak natively
- Pick a google apt. slot for a presentation day

Take-home midterm and final

Textbook(s)



nltk.org/book

Resources

- Piazza
- Course website:
<http://hmc-cs159-spring2015.github.io/site/>
- Shared github organization (see above)
- Office hours
- Grutor
 - Mike Bartoli
 - Grutoring hours: 6-8 Sunday, Olin 1261

Setting up OUR community

What expectations do we have

- of ourselves?
- of each other?

How can we work together to make this semester work for all of us?