CMSC 473/673
Natural Language Processing
Fall 2017
Frank Ferraro

ITE 358
ferraro@umbc.edu
Monday: 2:15-3
Tuesday: 11:30-12
by appointment

Natural language processing
Semantics
Vision & language processing
Learning with low-to-no supervision
Chi Zhang

ITE 353

chzhang1@umbc.edu

Thursday: 2-4

Part-of-speech tagging & parsing

Event detection on social media

Event classification
The Great A.I. Awakening

How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

BY GIDEON LEWIS-KRAUS  DEC. 14, 2016
Les talibans mènent des attaques-suicides dans Kaboul

Près de 17 personnes, dont un Français et un Italien, ont été tuées dans une série d'attaques revendiquées par les talibans. C'est l'attaque la plus meurtrière depuis janvier.

Air France s'attend à une perte historique pour l'exercice 2009-2010

La direction annonce que la compagnie va perdre 1,3 milliard d'euros sur l'exercice qui sera clos fin mars.
Hi all,
We wanted to invite you to join us for an early Thanksgiving on November 22nd, beginning around 2PM. Please bring your favorite dish! RSVP by next week.

Dave

Hi team,

The server appears to be dropping about 10% of requests (see attached dashboards). There hasn't been a new release since last night, so I'm not sure what's going on. Is anyone looking into this?

...
SPORTS
A massive climate change study is canceled ... because of climate change

By Doug Criss, CNN

1 Updated 10:37 AM ET, Tue June 20, 2017

Story highlights

Arctic sea ice has traveled farther south than normal along Newfoundland's northeast coast

An icebreaker has been repeatedly diverted to take part in rescue operations

(CNN) — A $17 million study of climate change in the Canadian Arctic has been nixed for now -- because of climate change.

A team of scientists from the University of Manitoba and four other schools were in the middle of the first leg of a four-year study of how climate change is affecting the areas around the Hudson Bay, the university said in statement. The study, named BaySys, started last month, and the scientists were traveling on the Canadian Research Icebreaker CCGS Amundsen.

But because of warmer temperatures in the Arctic, hazardous sea ice is traveling farther south than usual. The Amundsen, which is part of the Canadian Coast Guard fleet, has been diverted several times because its ice-breaking capabilities have been needed to help out in rescue efforts along Newfoundland's northeast coast. All of the delays and concerns about safety forced
Pat and Chandler agreed on a plan.

He said Pat would try the same tactic again.
Pat and Chandler agreed on a plan.

He said Pat would try the same tactic again.
Pat and Chandler agreed on a plan.

He said Pat would try the same tactic again.

is “he” the same person as “Chandler?”
Course Goals

Be introduced to some of the core problems and solutions of NLP (big picture)
Building the Next New York Times Recommendation Engine

When artificial intelligence makes a picture worth way more than a thousand words

Facebook AI Still Can’t Do Things Even A Baby Has Mastered

Shifting toward the Knowledge Economy

Google's AI can translate language pairs it has never seen
Course Goals

Be introduced to some of the core problems and solutions of NLP (big picture)

Learn different ways that success and progress can be measured in NLP
Deep Learning
Natural Language Processing

What society thinks I do

What my friends think I do

What other computer scientists think I do

What mathematicians think I do

What I think I do

What I actually do

from theano import import *
Course Goals

Be introduced to some of the core problems and solutions of NLP (big picture)
Learn different ways that success and progress can be measured in NLP
Relate to statistics, machine learning, and linguistics
Implement NLP programs
Course Goals

Be introduced to some of the core problems and solutions of NLP (big picture)
Learn different ways that success and progress can be measured in NLP
Relate to statistics, machine learning, and linguistics
Implement NLP programs
Read and analyze research papers
Practice your (written) communication skills
T-REX IN: "COMPUTATIONAL LINGUISTICS"
T-REX IN: "COMPUTATIONAL LINGUISTICS"

Computational linguistics is the study of computer-based language processing!
Natural Language Processing
≈
Computational Linguistics
Natural Language Processing
≈
Computational Linguistics

science focus

computational bio
computational chemistry
computational X
build a system to translate
create a QA system

engineering focus

Natural Language Processing

≈

Computational Linguistics

science focus

computational bio
computational chemistry
computational X
Natural Language Processing ≈ Computational Linguistics

Machine learning
Natural Language Processing ≈ Computational Linguistics

Machine learning

Information Theory
Natural Language Processing ≈ Computational Linguistics

Machine learning

Information Theory

Data Science
Natural Language Processing ≈ Computational Linguistics

Machine learning

Information Theory

Data Science

Systems Engineering
Natural Language Processing \approx\ Computational Linguistics

Machine learning

Information Theory

Data Science

Systems Engineering

Logic

Theory of Computation
Natural Language Processing ≈ Computational Linguistics

Machine learning
Information Theory
Data Science
Systems Engineering
Logic
Theory of Computation
Natural Language Processing ≈ Computational Linguistics

Machine learning
Information Theory
Data Science
Systems Engineering
Logic
Theory of Computation

Linguistics
Cognitive Science
Psychology
Natural Language Processing ≈ Computational Linguistics

Machine learning  Linguistics
Information Theory  Cognitive Science
Data Science  Psychology
Systems Engineering  Political Science
Logic  Digital Humanities
Theory of Computation  Education
T-REX IN: "COMPUTATIONAL LINGUISTICS"

Computational linguistics is the study of computer-based language processing!

A major area of computational linguistics is that of "ambiguity resolution". It turns out that many things people say in a language - English, for example - can have more than one meaning!

Consider the phrase "fruit flies like a banana". Is it describing the taste of fruit flies, or rather flying fruit? How can a computer hope to figure this out?
orthography
orthography

morphology

Adapted from Jason Eisner, Noah Smith
orthography

morphology

lexemes

syntax

semantics

pragmatics

Adapted from Jason Eisner, Noah Smith
Language is Productive
Language is Ambiguous
Ambiguity

Kids Make Nutritious Snacks
Kids Make Nutritious Snacks
Kids *Prepare* Nutritious Snacks
Kids *Are* Nutritious Snacks
Ambiguity

British Left Waffles on Falkland Islands
British Left Waffles on Falkland Islands
British Left Waffles on Falkland Islands
British Left Waffles on Falkland Islands

Ambiguity
Part of Speech Tagging

British Left Waffles on Falkland Islands

British Left Waffles on Falkland Islands

Adjective  Noun  Verb

British Left Waffles on Falkland Islands

Noun  Verb  Noun
Ambiguity

Pat saw Chris with the telescope on the hill.
Man Struck by Lightning Faces Battery Charge
Man Struck by Lightning Faces Battery Charge
Language Can Be Surprising
Garden Path Sentences
The Garden Path Sentences
Garden Path Sentences

The old
Garden Path Sentences

The old man
The old man the
The old man the boat
The old man the boat.
Garden Path Sentences

The old man the boat.
Garden Path Sentences

The complex houses married and single soldiers and their families.
The complex houses married and single soldiers and their families.
The rat the cat the dog chased killed ate the malt.
The rat *that* the cat the dog chased killed ate the malt.
Garden Path Sentences

The rat *that* the cat *that* the dog chased killed ate the malt.
The rat *that* the cat *that* the dog *chased* killed ate the malt.
The rat *that* the cat *that* the dog chased *killed* ate the malt.
The rat *that* the cat *that* the dog chased killed ate the malt.
Garden Path Sentences

[The rat [the cat [the dog chased] killed] ate the malt].
[The rat [the cat [the dog chased] killed] ate the malt].

center embedding
[The rat [the cat [the dog chased] killed] ate the malt].
Syntactic Parsing

[The rat [the cat [the dog chased] killed] ate the malt].
Discourse Processing through Coreference

I spread the cloth on the table to protect it.

I spread the cloth on the table to display it.
Discourse Processing through Coreference

I spread the cloth on the table to protect it.

I spread the cloth on the table to display it.
Discourse Processing through Coreference

I spread the cloth on the **table** to protect **it**.

I spread the **cloth** on the table to display **it**.

Courtesy Jason Eisner
Discourse Processing

John stopped at the donut store.
Discourse Processing

John stopped at the donut store.
Discourse Processing

John stopped at the donut store before work.
Discourse Processing

John stopped at the donut store *on his way home.*
Discourse Processing

John stopped at the donut shop.
Discourse Processing

John stopped at the *trucker shop*.  

*Courtesy Jason Eisner*
Discourse Processing

John stopped at the *mom & pop shop.*
Discourse Processing

John stopped at the red shop.
Discourse Processing

John thought a coffee was good every few hours.
Discourse Processing

John thought a coffee was good every few hours.
Discourse Processing

John thought a coffee was good every few hours.
Discourse Processing

John thought a coffee was good every few hours.

(coffee should be consumed very often)
Discourse Processing

John thought a coffee was good every few hours.

(coffee gets cold & stale after a while)
Computational linguistics is the study of computer-based language processing.

A major area of computational linguistics is that of "ambiguity resolution". It turns out that many things people say in a language - English, for example - can have more than one meaning!

Consider the phrase "fruit flies like a banana". Is it describing the taste of fruit flies, or rather flying fruit? How can a computer hope to figure this out?

Many have focused on statistical modelling of language, but this approach is approximate. I agree!
Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.
Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.
$p_\theta( )$

Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.
probabilistic model

\[ p_\theta(X) \]

objective

\[ F(\theta) \]
Gradient Ascent
\[ \arg \max_{\theta} F(\theta) \]
Gradient Ascent

\[ \text{arg max } F(\theta) \]
Gradient Ascent

\[
\arg \max_{\theta} \quad F(\theta)
\]
Gradient Ascent

$$\arg \max_\theta F(\theta)$$
T-REX IN: "COMPUTATIONAL LINGUISTICS"

Computational linguistics is the study of computer-based language processing!

A major area of computational linguistics is that of "ambiguity resolution". It turns out that many things people say in a language - English, for example - can have more than one meaning!

Consider the phrase "fruit flies like a banana". Is it describing the taste of fruit flies, or rather flying fruit? How can a computer hope to figure this out?

Many have focused on statistical modelling of language, but this approach is approximate. I agree!

What do YOU know about computational linguistics?

Ever read a little paper called "Non-Statistical Models for Unsupervised Prepositional Phrase Attachment?"

That was me!

It was some of my earliest work on head word tuples!

(C) 2003 Ryan North
Continuous Meaning

The paper reflected the truth.
Continuous Meaning

The paper reflected the truth.
Continuous Meaning

The paper reflected the truth.
Structured Meanings

The paper reflected the truth.
Structured Meanings

The paper reflected the truth.

<table>
<thead>
<tr>
<th>Role Filler</th>
<th>Role</th>
<th>Meaning/Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>The paper</td>
<td>COMMUNICATOR</td>
<td>The sentient entity that uses language in the written or spoken modality to convey a MESSAGE to the ADDRESSEE.</td>
</tr>
<tr>
<td>---</td>
<td>MEDIUM</td>
<td>The physical or abstract setting in which the MESSAGE is conveyed.</td>
</tr>
<tr>
<td>the truth</td>
<td>MESSAGE</td>
<td>MESSAGE is a proposition or set of propositions that the COMMUNICATOR wants the ADDRESSEE to believe or take for granted.</td>
</tr>
</tbody>
</table>
Featurized Meanings

The paper **reflected** the truth.

<table>
<thead>
<tr>
<th>Property</th>
<th>The paper</th>
<th>the truth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AWARENESS</strong></td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>MANIPULATED</strong></td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>SENTIENT</strong></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>VOLITIONAL</strong></td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
Semantic Projection

Syntactic Argument
- dobj
- nsubj
- nsubjpass
Semantic Projection
What are Related Words?

Learn grammatical inflections

<table>
<thead>
<tr>
<th>Anticipated</th>
<th>PropBank</th>
<th>sep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 anticipate</td>
<td>6 intimidated</td>
<td></td>
</tr>
<tr>
<td>2 anticipating</td>
<td>7 separating</td>
<td></td>
</tr>
<tr>
<td>3 anticipates</td>
<td>8 separates</td>
<td></td>
</tr>
<tr>
<td>4 stabbing</td>
<td>9 drag</td>
<td></td>
</tr>
<tr>
<td>5 separate</td>
<td>10 guarantee</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Invented</th>
<th>PropBank</th>
<th>sep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 invent</td>
<td>6 aspire</td>
<td></td>
</tr>
<tr>
<td>2 document</td>
<td>7 documenting</td>
<td></td>
</tr>
<tr>
<td>3 documented</td>
<td>8 aspires</td>
<td></td>
</tr>
<tr>
<td>4 invents</td>
<td>9 inventing</td>
<td></td>
</tr>
<tr>
<td>5 documents</td>
<td>10 swinging</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Producing</th>
<th>PropBank</th>
<th>sep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 produces</td>
<td>6 ridden</td>
<td></td>
</tr>
<tr>
<td>2 produce</td>
<td>7 improves</td>
<td></td>
</tr>
<tr>
<td>3 produced</td>
<td>8 surround</td>
<td></td>
</tr>
<tr>
<td>4 prized</td>
<td>9 surrounds</td>
<td></td>
</tr>
<tr>
<td>5 originates</td>
<td>10 originating</td>
<td></td>
</tr>
</tbody>
</table>
## What are Related Words?

Learn grammatical inflections

Identify paraphrastic and synonymous words

<table>
<thead>
<tr>
<th>anticipated</th>
<th>anticipated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filler</td>
<td>sep</td>
</tr>
<tr>
<td>foresaw</td>
<td>6 pondered</td>
</tr>
<tr>
<td>2 figuring</td>
<td>7 kidded</td>
</tr>
<tr>
<td>3 alleviated</td>
<td>8 constituted</td>
</tr>
<tr>
<td>4 craved</td>
<td>9 uttering</td>
</tr>
<tr>
<td>5 jeopardized</td>
<td>10 forgiven</td>
</tr>
<tr>
<td>anticipate</td>
<td>6 intimidated</td>
</tr>
<tr>
<td>2 anticipating</td>
<td>7 separating</td>
</tr>
<tr>
<td>3 anticipates</td>
<td>8 separates</td>
</tr>
<tr>
<td>4 stabbing</td>
<td>9 drag</td>
</tr>
<tr>
<td>5 separate</td>
<td>10 guarantee</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>invented</th>
<th>invented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filler</td>
<td>sep</td>
</tr>
<tr>
<td>pioneered</td>
<td>6 tolerated</td>
</tr>
<tr>
<td>2 scratch</td>
<td>7 resurrected</td>
</tr>
<tr>
<td>3 complemented</td>
<td>8 sweated</td>
</tr>
<tr>
<td>4 competed</td>
<td>9 fancies</td>
</tr>
<tr>
<td>5 consoled</td>
<td>10 concocted</td>
</tr>
<tr>
<td>invent</td>
<td>6 aspire</td>
</tr>
<tr>
<td>2 document</td>
<td>7 documenting</td>
</tr>
<tr>
<td>3 documented</td>
<td>8 aspires</td>
</tr>
<tr>
<td>4 invents</td>
<td>9 inventing</td>
</tr>
<tr>
<td>5 documents</td>
<td>10 swinging</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>producing</th>
<th>producing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filler</td>
<td>sep</td>
</tr>
<tr>
<td>containing</td>
<td>6 storing</td>
</tr>
<tr>
<td>2 contains</td>
<td>7 reproduce</td>
</tr>
<tr>
<td>3 manufactures</td>
<td>8 store</td>
</tr>
<tr>
<td>4 contain</td>
<td>9 exhibiting</td>
</tr>
<tr>
<td>5 consume</td>
<td>10 furnish</td>
</tr>
<tr>
<td>produces</td>
<td>6 ridden</td>
</tr>
<tr>
<td>2 produce</td>
<td>7 improves</td>
</tr>
<tr>
<td>3 produced</td>
<td>8 surround</td>
</tr>
<tr>
<td>4 prized</td>
<td>9 surrounds</td>
</tr>
<tr>
<td>5 originates</td>
<td>10 originating</td>
</tr>
</tbody>
</table>
T-REX IN: "COMPUTATIONAL LINGUISTICS"

Computational linguistics is the study of computer-based language processing!

A major area of computational linguistics is that of "ambiguity resolution". It turns out that many things people say in a language - English, for example - can have more than one meaning.

Consider the phrase "fruit flies like a banana". Is it describing the taste of fruit flies, or rather flying fruit? How can a computer hope to figure this out?

Many have focused on statistical modelling of language, but this approach is approximate. I agree!

What do YOU know about computational linguistics?

Ever read a little paper called "Non-Statistical Models for Unsupervised Prepositional Phrase Attachment"?

That was me!

It was some of my earliest work on head word tuples!

Shit man, you know more about this than I do!

You know what? You should be the one doing the talking here!
Data In, Knowledge Out?

Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.
Data In, Knowledge Out?

Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.

The death toll in the Los Angeles riots rose to 50 today. The Los Angeles County Coroner's Office is continuing to try to identify those killed, many of whom had no identification.
Data In, Knowledge Out?

Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.

The death toll in the Los Angeles riots rose to 50 today. The Los Angeles County Coroner's Office is continuing to try to identify those killed, many of whom had no identification.

The 150-year-old weeping Camperdown elm in Prospect Park in Brooklyn stands a mere 12 feet high, but its intricate pattern of branches etches a 25-foot circle.
Data In, Knowledge Out?

Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.

The death toll in the Los Angeles riots rose to 50 today. The Los Angeles County Coroner's Office is continuing to try to identify those killed, many of whom had no identification.

The 150-year-old weeping Camperdown elm in Prospect Park in Brooklyn stands a mere 12 feet high, but its intricate pattern of branches etches a 25-foot circle.

"What a glorious silhouette it is against the winter sky," said Robert Makla, a founder of the Friends of Prospect Park, which since 1966 has insured the tree's survival. "It's duplicated by no other tree or abstract work of art."
Data In, Knowledge Out?

Three people have been fatally shot, and five people, including a mayor, were seriously wounded as a result of a Shining Path attack today.

The death toll in the Los Angeles riots rose to 50 today. The Los Angeles County Coroner's Office is continuing to try to identify those killed, many of whom had no identification.

The 150-year-old weeping Camperdown elm in Prospect Park in Brooklyn stands a mere 12 feet high, but its intricate pattern of branches etches a 25-foot circle.

"What a glorious silhouette it is against the winter sky," said Robert Makla, a founder of the Friends of Prospect Park, which since 1966 has insured the tree's survival. "It's duplicated by no other tree or abstract work of art."

Weeping trees like this elm - plants whose limbs grow down, not up - are spectacular sculptures in winter and graceful accents in summer.
Administrivia
# Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>473</th>
<th>673</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midterm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Two Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Project</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>473</th>
<th>673</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>Midterm</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Final</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td><strong>Two Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Project</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>473</th>
<th>673</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>50%</td>
<td>30%</td>
</tr>
<tr>
<td>Midterm</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Final</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td><strong>Two Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Project</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>473</th>
<th>673</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>50%</td>
<td>30%</td>
</tr>
<tr>
<td>Midterm</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Final</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Two Papers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Project</td>
<td>30%</td>
<td>30%</td>
</tr>
</tbody>
</table>
## Grading

<table>
<thead>
<tr>
<th>Component</th>
<th>473</th>
<th>673</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four Assignments</td>
<td>50%</td>
<td>30%</td>
</tr>
<tr>
<td>Midterm</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Final</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td><strong>Two Papers</strong></td>
<td>***</td>
<td>20%</td>
</tr>
<tr>
<td>Course Project</td>
<td>30%</td>
<td>30%</td>
</tr>
</tbody>
</table>
# Final Grades

<table>
<thead>
<tr>
<th>Score</th>
<th>Letter</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 90</td>
<td>A</td>
</tr>
<tr>
<td>80</td>
<td>B</td>
</tr>
<tr>
<td>70</td>
<td>C</td>
</tr>
<tr>
<td>60</td>
<td>D</td>
</tr>
<tr>
<td>0</td>
<td>F</td>
</tr>
</tbody>
</table>
Natural Language Processing

Fall 2017 — CMSC 473/673

Announcements

- 8/30/17 [673] Graduate Paper 1 is available. It's due 10/11/17, by 11:59 AM.
- 8/30/17 [673] Assignment 1 is available. It's due 9/20/17, by 11:59 AM.
- 8/30/17 [473] Assignment 1 is available. It's due 9/20/17, by 11:59 AM.

Who, What, When, and Where

Check out the full syllabus for all this information, including policies on academic honesty, accomodations, and late assignments.

Meeting Times
Performing Arts and Humanities, 107
Monday & Wednesday, 1pm - 2:15pm

Instructor
Frank Ferraro
ferraro [at] umbc [dot] edu
ITE 385
Monday 2:15 - 3pm
Tuesday 11:30 - 12
by appointment

Topics
The topics covered will include

- probability, classification, and the efficacy of simple counting methods
- language modeling (n-gram models, smoothing heuristics, maxent/loopy, distributed/vector-valued representations)
- sequences of latent variables (e.g., hidden Markov models, some basic alignment)
- trees and graphs, as applied to syntax and semantics
- some discourse-related applications (coreference resolution, textual entailment)
- special and current topics (e.g., fairness and ethics in NLP)

Goals
After taking this course, you will

- be introduced to some of the core problems and solutions of NLP;
- learn different ways that success and progress can be measured in NLP;
- be exposed to how these problems relate to those in statistics, machine learning, and natural language generation and understanding;
- have experience implementing a number of NLP programs;
- read and analyze research papers; practice your (written) communication skills.
Online Discussions

https://piazza.com/umbc/fall2017/cmsc473673
## Important Dates

<table>
<thead>
<tr>
<th>Date</th>
<th>473</th>
<th>673</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, 9/20</td>
<td>Assignment 1</td>
<td></td>
</tr>
<tr>
<td>Wednesday, 10/11</td>
<td>---</td>
<td>Paper 1</td>
</tr>
<tr>
<td>Wednesday, 10/18</td>
<td>Assignment 2</td>
<td></td>
</tr>
<tr>
<td>Monday, 10/30</td>
<td>Midterm</td>
<td></td>
</tr>
<tr>
<td>Monday, 11/6</td>
<td>Project Update</td>
<td></td>
</tr>
<tr>
<td>Wednesday, 11/15</td>
<td>Assignment 3</td>
<td></td>
</tr>
<tr>
<td>Monday, 11/27</td>
<td>---</td>
<td>Paper 2</td>
</tr>
<tr>
<td>Monday, 12/11</td>
<td>Assignment 4</td>
<td></td>
</tr>
<tr>
<td>Wednesday, 12/20</td>
<td>Project Final Report</td>
<td></td>
</tr>
<tr>
<td>Wednesday, 12/20</td>
<td>Final Exam</td>
<td></td>
</tr>
</tbody>
</table>
Late Policy

Everyone has a budget of 10 late days
Late Policy

Everyone has a budget of 10 *late days*

If you have them left: assignments turned in after the deadline will be graded and recorded, no questions asked
Late Policy

Everyone has a budget of 10 *late days*

If you have them left: assignments turned in after the deadline will be graded and recorded, no questions asked

If you don’t have any left: still turn assignments in. They could count in your favor in borderline cases
Late Policy

Everyone has a budget of 10 *late days*

Use them as needed throughout the course

They’re meant for personal reasons and emergencies

Do not procrastinate
Late Policy

Everyone has a budget of 10 late days

Contact me privately if an extended absence will occur