Lecture 01: Introduction

Julia Hockenmaier
juliahmr@illinois.edu
Welcome to CS447!
Welcome to CS447!

Professor:
Julia Hockenmaier (Siebel 3324)
juliahmr@illinois.edu
http://juliahmr.cs.illinois.edu

TAs:
Marc Canby
Fred Choi
Rajarshi Haldar
Prashant Jayannavar
What is CS447?

This class is a broad introduction to NLP

Target audience:
  Advanced undergraduates
  Graduate students

What’s new this semester?
  We’re back in person for the first time since Fall 2019
  We’re trying to incorporate new elements that we’ve developed for the online/Coursera version of this class
  We’re switching to Canvas
  We’re updating some of the class content
Why should you take this class?

NLP is an (increasingly) important area

- NLP is now good enough for real-world applications.
- There is a huge growth in NLP companies and NLP jobs (in many industries)

NLP is far from solved (despite much recent progress)

- There is still a lot that remains to be done!

Doing NLP well requires a broad mix of knowledge:

- What is natural language?
- What about natural language is challenging for computers?
- What kind of data, algorithms, machine learning approaches can we use (and which ones do we need to develop)?
What will you learn in this class?

- **What is NLP?**
  
The core **tasks** (as well as **data sets** and **evaluation metrics**) that people work on in NLP

- **How does NLP work?**
  
The fundamental **models, algorithms** and **representations** that have been developed for these tasks

- **Why is NLP hard?**
  
The relevant **linguistic concepts and phenomena** that have to be handled to do well at these tasks
NLP is necessary to...

... **analyze** text automatically at scale
   (text = news, documents, social media, search queries, ...)

... **translate** automatically between languages
   (language = English, Chinese, Arabic, Hindi, etc.)

... **communicate** naturally with systems/devices
   (systems/devices = robots, computers, costumer support, digital assistants, smart devices, navigation systems, ...)

CS447 Natural Language Processing (J. Hockenmaier)  https://courses.grainger.illinois.edu/cs447/
The focus of this class

We want to identify the structure and meaning of words, sentences, texts and conversations

– N.B.: we do not deal with speech/audio (no signal processing)

We mainly deal with language analysis/understanding, and somewhat less with language generation/production

We focus on fundamental concepts, methods, models, tasks and algorithms, not so much on current research:

– Data (natural language): Linguistic concepts and phenomena
– Representations: Grammars, automata, embeddings/vectors, …
– Tasks: Analysis, generation, translation, …
– Models: Neural models, statistical models, …
What you should learn

You should be able to answer the following questions:

- What makes natural language difficult for computers?
- What are the core NLP tasks?
- What are the main modeling techniques used in NLP?

We won’t be able to cover all of the latest research…

(this requires more time, and a much stronger background in machine learning than we can assume for this class)

… but I would still like you to get an understanding of:

- How well does current NLP technology work (or not)?
- What NLP software and datasets are available?
- How to read NLP research papers [4 credits section]
Our Syllabus (tentative)

Week 1: Introduction
Week 2: The Structure and Distribution of Words
Week 3: Classification for NLP
Week 4: The Meaning of Words
Week 5: Introduction to Neural Networks for NLP
Week 6: POS Tagging and Sequence Labeling
Week 7: Neural Sequence Models
Week 8: Machine Translation and Large Language Models
Week 9: The Structure of Sentences
Week 10: The Structure and Meaning of Sentences
Week 11: Relations, Events, Times
Week 12: Semantic Roles, Referring Expressions
Week 13: Discourse Coherence, Question Answering
Week 14: Dialogue and Grounded NLP
Week 15: Ethics in NLP
Our Syllabus and Textbook

You can find brief descriptions of our syllabus at
https://courses.grainger.illinois.edu/cs447/

Our Textbook:
We loosely follow
Jurafsky and Martin, Speech and Language
Processing 3rd ed. (January 2023 version)
https://web.stanford.edu/~jurafsky/slp3/
What is NLP?
In Science Fiction (Kubricks’ 1968 movie 2001: A Space Odyssey)

A conversation onboard the Discovery One spacecraft between HAL 9000 (a sentient computer developed in Urbana, IL) and Dave, a human astronaut: https://en.wikipedia.org/wiki/HAL_9000

Dave: Open the pod bay doors, please, HAL. Open the pod bay doors, please, HAL. Hello, HAL, do you read me? Hello, HAL, do you read me? Do you read me, HAL? Do you read me, HAL? Hello, HAL, do you read me? Hello, HAL, do you read me? Do you read me, HAL?
HAL: Affirmative, Dave. I read you.
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.
Dave: What are you talking about, HAL?
HAL: This mission is too important for me to allow you to jeopardize it.
Dave: I don't know what you're talking about, HAL.
HAL: I know that you and Frank were planning to disconnect me. And I'm afraid that's something I cannot allow to happen.
Dave: Where the hell did you get that idea, HAL?
HAL: Dave, although you took very thorough precautions in the pod against my hearing you, I could see your lips move.
Dave: All right, HAL. I'll go in through the emergency airlock.
HAL: Without your space helmet, Dave, you're going to find that rather difficult.
Dave: [sternly] HAL, I won't argue with you anymore. Open the doors.
HAL: [monotone voice] Dave, this conversation can serve no purpose anymore. Good-bye.

In Science Fiction (Kubricks’ 1968)

What are HAL’s capabilities?

- HAL speaks **fluent, natural English**
- HAL’s **responses are appropriate in the context**
  - “Affirmative, I read you”
  - “I think you know what the problems is just as well as I do”
- HAL understands **references to the physical world**:
  - “Open the pod bay doors”
- HAL **talks about what HAL sees**:
  - “I could see your lips move”
  - “without your space helmet”
- HAL **reasons**
  - “I can’t do that”
  - “this mission is too important”
  - “I know that you and Frank were planning to disconnect me. And I’m afraid that’s something I can’t allow”
A real human-computer dialogue from the 1960s: ELIZA (Weizenbaum, 1966)

**Computer:** TELL ME MORE ABOUT YOUR FAMILY  
**Human:** My mother takes care of me.  
**Computer:** WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU  
**Human:** My father.  
**Computer:** YOUR FATHER  
**Human:** You are like my father in some ways.  
**Computer:** WHAT RESEMBLANCE DO YOU SEE  
**Human:** You are not very aggressive but I think you don't want me to notice that.  
**Computer:** WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE  
**Human:** You don't argue with me.  
**Computer:** WHY DO YOU THINK I DON'T ARGUE WITH YOU  
**Human:** You are afraid of me.
The Eliza Effect

People *really* trusted this system:

Weizenbaum: “I had not realized ... that extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people”

ELIZA was based on a cascade of finite-state transducers:

*Human:* You don't argue with me.

1. Replace *you* with *I* and *me* with *you*:
   
   *I* don't argue with *you*.

2. Replace <...> with *Why do you think* <...>:
   
   *Why do you think* I don't argue with you.

*Computer:* WHY DO YOU THINK I DON'T ARGUE WITH YOU

— How could you break a system like ELIZA?
— Where do you see the ELIZA effect today?
In the 21st century, NLP remains in the news

Facebook AI Creates Its Own Language In Creepy Preview Of Our Potential Future

Computers can now describe images using language you’d understand

The AI Text Generator That’s Too Dangerous to Make Public
Researchers at OpenAI decided that a system that scores well at understanding language could too easily be manipulated for malicious intent.

How AI Can Create And Detect Fake News

A.I. breakthroughs in natural-language processing are big for business

Barbie Wants to Get to Know Your Child
With the help of A.I., America’s most famous doll tries to fulfill a timeless dream — convincing little girls that she’s a real friend. What will happen if they believe her?
IBM’s Watson wins at Jeopardy!

https://www.youtube.com/watch?v=P18EdAKuC1U

https://dl.acm.org/doi/10.1147/JRD.2012.2184356
Have chatbots developed their own language?

Facebook AI Creates Its Own Language In Creepy Preview Of Our Potential Future

https://code.facebook.com/posts/1686672014972296/deal-or-no-deal-training-ai-bots-to-negotiate/

Bob: “I can can I I everything else.”
Alice: “Balls have zero to me to me to me to me to me to me to me to me to.”

No, not really…
http://www.snopes.com/facebook-ai-developed-own-language/
Barbie: “Do you have any sisters?”
Child: “Yeah, I only have one.”
Barbie: “What’s something nice that your sister does for you?”
Child: “She does nothing nice to me”
Barbie: “Well, what is the last nice thing your sister did?”
Child: “She helped me with my project — and then she destroyed it.”
Barbie: “Oh, yeah, tell me more!”
Child: “That’s it, Barbie,”
Barbie: “Have you told your sister lately how cool she is?”
Child: “No. She is not cool,”
Barbie: “You never know, she might appreciate hearing it”

(Thanks to Barbara Grosz for pointing out the example from the NYT)
Dialog systems, chatbots, digital assistants
Grandpa feels sorry for his granddaughter and let the baby sleep in

Uncle Zhong (a pseudonym), 70, fell into "summer anxiety": This summer, he would shuttle his granddaughter to and from training classes every week. With high temperatures and scorching heat, every day I come back, I feel as weak as stepping on cotton.

In addition to physical discomfort, there is also mental tension.

Feeling that his son and daughter-in-law were reporting too much for their children, Uncle Zhong also had a conflict with them: "It's a hot day, adults and children suffer."

During this period of time, Uncle Zhong came to the Mental Health Department of Zhejiang Provincial People's Hospital because he was prone to get angry, insomnia, poor appetite, and unstable blood pressure.

http://education.news.cn/2020-08/25/c_1210768533.htm
Examples of NLP applications
(What can NLP be used for?)

Natural language (and speech) interfaces
- Search/IR, database access, image search, image description
- Dialog systems (e.g. customer service, robots, cars, tutoring), chatbots

Information extraction, summarization, translation:
- Process (large amounts of) text automatically to obtain meaning/knowledge contained in the text
- Identify/analyze trends, opinions, etc. (e.g. in social media)
- Translate text automatically from one language to another

Convenience:
- Grammar/style checking, automate email filing, autograding
Examples of NLP tasks
(What capabilities do NLP systems need?)

Natural language understanding
Extract information (e.g. about entities, events or relations between them) from text
Translate raw text into a meaning representation
Reason about information given in text
Execute NL instructions

Natural language generation and summarization
Translate database entries or meaning representations to raw natural language text
Produce (appropriate) utterances/responses in a dialog
Summarize (newspaper or scientific) articles, describe images

Natural language translation
Translate one natural language to another
Generating text with huge language models…

Language models can be used to generate (produce) text

Massive neural language models trained on vast amounts of text have been developed in the last few years. Most recent incarnation: GPT-3 (175B parameters, trained on 300B tokens).

But these models have no access to meaning, and don’t “understand” much.

See also Bender & Koller ‘20 for a critique


Human Prompt (given to GPT-3)

At the party, I poured myself a glass of lemonade, but it turned out to be too sour, so I added a little sugar.

I didn’t see a spoon handy, so I stirred it with a cigarette. But that turned out to be a bad idea because

[GPT-3’s generated continuation]

it kept falling on the floor. That’s when he decided to start the Cremation Association of North America, which has become a major cremation provider with 145 locations.

from Marcus & Davis ‘20

What is the current state of NLP?

Lots of commercial applications and interest.
   Some applications are working pretty well already, others not so much.

A lot of hype around “deep learning” and “AI”
   Neural nets are powerful classifiers and sequence models
   Public libraries and datasets make it easy for anybody to get a model up and running
   “End-to-end” models put into question whether we still need the traditional NLP pipeline.
   This paradigm shift is well underway.
   But many of the fundamental problems haven’t been solved.
How do NLP systems work?
The traditional NLP pipeline

A (traditional) NLP system may use some or all of the following steps:

Tokenizer/Segmenter
to identify words and sentences

Morphological analyzer/POS-tagger
to identify the part of speech and structure of words

Word sense disambiguation
to identify the meaning of words

Syntactic/semantic Parser
to obtain the structure and meaning of sentences

Coreference resolution
to keep track of the various entities mentioned
What does it take to understand text?

- Death Valley measured 54.4 degrees Celsius, a new world record for the United States.
- Çavuşoğlu's warning to Athens: Some countries are filling up, do not put yourself at risk.
- Thai government warns against traveling: Do not enter areas.
- Qabiyyeen xalayaa dhimma Obbo Lidatu Ayyaaloorratti MM Abiyyiiif barraa'e maali?
- 'Dim angen cau tafarndai a bwytai i ailagor ysgolion'
Task: Tokenization/segmentation

We need to split text into words and sentences.

- Languages like Chinese or Thai don’t have spaces between words.
- Even in English, this cannot be done deterministically: *There was an earthquake near D.C. You could even feel it in Philadelphia, New York, etc.*

NLP task:
What is the *most likely* segmentation/tokenization?
Task: Part-of-speech-tagging

*Open the pod door, Hal.*

**open:**
Verb, adjective, or noun?
- Verb: *open* the door
- Adjective: the *open* door
- Noun: in the *open*
How do we decide?

We want to know the most likely tags $T$ for the sentence $S$

$$\arg\max_{T} P(T|S)$$

We need to define a statistical model of $P(T \mid S)$, e.g.:

$$\arg\max_{T} P(T|S) = \arg\max_{T} P(T) P(S|T)$$

$$P(T) =_{\text{def}} \prod_{i} P(t_{i} \mid t_{i-1}) P(w_{i} \mid t_{i})$$

$$P(S|T) =_{\text{def}} \prod_{i} P(w_{i} \mid t_{i})$$

We need to estimate the parameters of $P(T \mid S)$, e.g.:

$$P( t_{i} = V \mid t_{i-1} = N ) = 0.312$$
Disambiguation requires statistical models

**Ambiguity** is a core problem for any NLP task

Statistical models* are one of the main tools to deal with ambiguity.

*More generally: a lot of the models (classifiers, structured prediction models) you learn about in your machine learning classes can be used for this purpose.

We won’t assume you have taken a machine learning class.

These models need to be trained (estimated, learned) before they can be used (tested, evaluated).
“I made her duck”

What does this sentence mean?

“I made her crouch”,
“I cooked duck for her”,
“I cooked her [pet] duck (perhaps just for myself)”, …

“duck”: noun or verb?
“make”: “cook X” or “cause X to do Y”?
“her”: “for her” or “belonging to her”?
Ambiguity in natural language

Language has different kinds of ambiguity, e.g.:

**Structural ambiguity**

“I eat sushi with tuna” vs. “I eat sushi with chopsticks”
“I saw the man with the telescope on the hill”

**Lexical (word sense) ambiguity**

“I went to the bank”: financial institution or river bank?

**Referential ambiguity**

“John saw Jim. He was drinking coffee.”
Who was drinking coffee?
Dealing with ambiguity

- Search Algorithm
- Structural Representation
- Scoring Function
“I made her duck cassoulet”

(Cassoulet = a French bean casserole)

The second major problem in NLP is coverage:
We will always encounter unfamiliar words and constructions.

Our models need to be able to deal with this.

This means that our models need to be able to generalize from what they have been trained on to what they will be used on.
Task: Syntactic parsing

Open the pod door, Hal.
Observation:
Structure corresponds to meaning

Correct analysis

Incorrect analysis
Task: Semantic Analysis

\[ \exists x \exists y (\text{pod\_door}(x) \& \text{Hal}(y) \& \text{request}(\text{open}(x, y))) \]

\[ \text{Open the pod door, } \text{Hal}. \]
Representing meaning

If a natural language understanding system needs to return a symbolic representation (or data structure) of the meaning of text, it needs a pre-defined meaning representation language.

“Deep” semantic analysis: (Variants of) formal logic

\[ \exists x \exists y (\text{pod\_door}(x) \& \ \text{Hal}(y) \ & \ \text{request} (\text{open}(x, y))) \]

“Shallow” semantic analysis: Template-filling

(Often used in information extraction)

Named-Entity Recognition: identify all organizations, locations, dates,...

Event Extraction:

We also distinguish between
Understanding texts

On Monday, John went to Einstein’s. He wanted to buy lunch. But the store was closed. That made him angry, so the next day he went to Green Street instead.

Can you answer the following questions?

Was Einstein’s open for lunch on Monday? [No]
This requires the ability to identify that “Einstein’s” and “the store” refer to the same entity. (coreference resolution).

On which day did John go to Green Street? [On Tuesday].
This requires the ability to understand the implicit information that “the next day” means really “the next day after Monday” (and the knowledge that that is a Tuesday).
NLP Pipeline: Assumptions

Each step in the NLP pipeline embellishes the input with **explicit information** about its linguistic structure

- POS tagging: Parts of speech of word,
- Syntactic parsing: Grammatical structure of sentence,....

Each step in the NLP pipeline requires **its own explicit ("symbolic") output representation**:

- POS tagging requires a **POS tag set**
  (e.g. NN=common noun singular, NNS = common noun plural, …)
- Syntactic parsing requires **constituent or dependency labels**
  (e.g. NP = noun phrase, or nsubj = nominal subject)

These representations should capture **linguistically appropriate generalizations/abstractions**

- Designing these representations requires linguistic expertise
NLP Pipeline: Shortcomings

Each step in the pipeline relies on a learned model that will return the most likely representations

- This requires a lot of annotated training data for each step
- Annotation is expensive and sometimes difficult (people are not 100% accurate)
- These models are never 100% accurate
- Models make more mistakes if their input contains mistakes

How do we know that we have captured the “right” generalizations when designing representations?

- Some representations are easier to predict than others
- Some representations are more useful for the next steps in the pipeline than others
- But we won’t know how easy/useful a representation is until we have a model that we can plug into a particular pipeline
Sidestepping the NLU pipeline

Many current neural approaches for natural language understanding and generation go directly from the raw input to the desired final output.

With large amounts of training data, this often works better than the traditional approach.
   — We will soon discuss why this may be the case.

But these models don’t solve everything:
   — How do we incorporate knowledge, reasoning, etc. into these models?
   — What do we do when don’t have much training data?
      (e.g. when we work with a low-resource language)
How will we teach this class?
Class Structure and Platforms

Lectures: WF 2:00pm–3:15pm, Siebel 1404
  Recordings will be put on Mediaspace
Website (slides, syllabus, deadlines, policies, links):
  https://courses.grainger.illinois.edu/cs447

Assignments and Grades:
  Canvas (peer-graded assignments, grade book)
  Gradescope (quizzes, programming assignments)
Class Discussions: Campuswire
Office Hours (starting next week):
  Julia Hockenmaier: WF, 3:30pm—4:00pm, Siebel 3324
  TAs: TBA
Assignment Types

For everybody:

12 Quizzes (two weeks per quiz)
8 Peer-Graded Assignments (two weeks per assignment)
4 Programming Assignments (three weeks per assignment)

For 4th Credit Hour students, additionally:

Literature Review (final report due at the end of the semester)

Late policy?

No late assignments will be accepted (except for medical/religious exemptions)
Assessment

If you’re taking this class for 3 credit hours:

- 25% quizzes (all equally weighted)
- 25% peer-graded assignments (all equally weighted)
- 50% programming assignments (all equally weighted)

If you’re taking this class for 4 credit hours:

- 18.75% quizzes (all equally weighted)
- 18.75% peer-graded assignments (all equally weighted)
- 37.50% programming assignments (all equally weighted)
- 25.00% literature review
Programming Assignments

What?
4 programming assignments in Python/PyTorch

Why?
To make sure you can put what you’ve learned to practice.

How?
Released on Fridays in Weeks 2, 5, 8, 11
You will have three weeks for each assignment
Submit your assignments on Gradescope.
Quizzes

What?
Short questions (typically multiple choice)

Why?
We want to make sure you follow along during the semester
We want to evaluate that you understand the material

How?
We will use Gradescope
Released on Fridays in Weeks 01—12
You have two weeks per quiz
Peer-Graded Assignments

What?
You may be asked to write a short essay, or to analyze some text, based on what we cover in class.

Why?
To make sure you think about the material and try to apply what you have learned.

How?
Released on Fridays in Weeks 1, 3, 4, 6, 7, 9, 10, 12
You will have two weeks for each assignment
We will use Canvas
You will have to grade your peers’ responses
Grading is mostly based on effort.
4th Credit Hour: Literature Survey

What?
You need to read and describe several (5–7) NLP papers on a particular task or topic, and produce a written report that compares and critiques these approaches.

Why?
To make sure you get a deeper knowledge of NLP by reading a number of original papers in sufficient depth to discuss and compare them,
Grades

I don’t grade “on a curve”:
If everybody does really well in this class, everybody gets an A, not just the top X%.

I only assign letter grades at the end of the semester.
You should know what percent of the grade you have received so far, but I may not be able to tell you precisely what letter grade that may correspond to (although you should talk to me if you want to know whether you’re doing well or not so well).

For assignments, quizzes and peer-graded assignments, the undergrads’ performance will determine the grading scale for everybody.
Academic Integrity

You can talk to each other about the assignments, but what you submit needs to be your own work.

We may use tools such as MOSS/TurnItIn to detect plagiarism.

You are not allowed to use tools like ChatGPT, except if/when we ask you to analyze their output for an assignment.

If you’re taking this class for four credits, your literature review needs to be your own work, and you need to cite all sources.
DRES accommodations

If you need any disability related accommodations, talk to DRES (http://disability.illinois.edu, disability@illinois.edu, phone 333-4603)

If you are concerned you may have a disability-related condition that is impacting your academic progress, there are academic screening appointments available on campus that can help diagnosis a previously undiagnosed disability by visiting the DRES website and selecting “Sign-Up for an Academic Screening” at the bottom of the page.”

Come and talk to me as well, especially once you have a letter of accommodation from DRES.

Do this early enough so that we can take your requirements into account!