

# Statistical NLP

# Introduction

Tim Rocktäschel, Sebastian Riedel

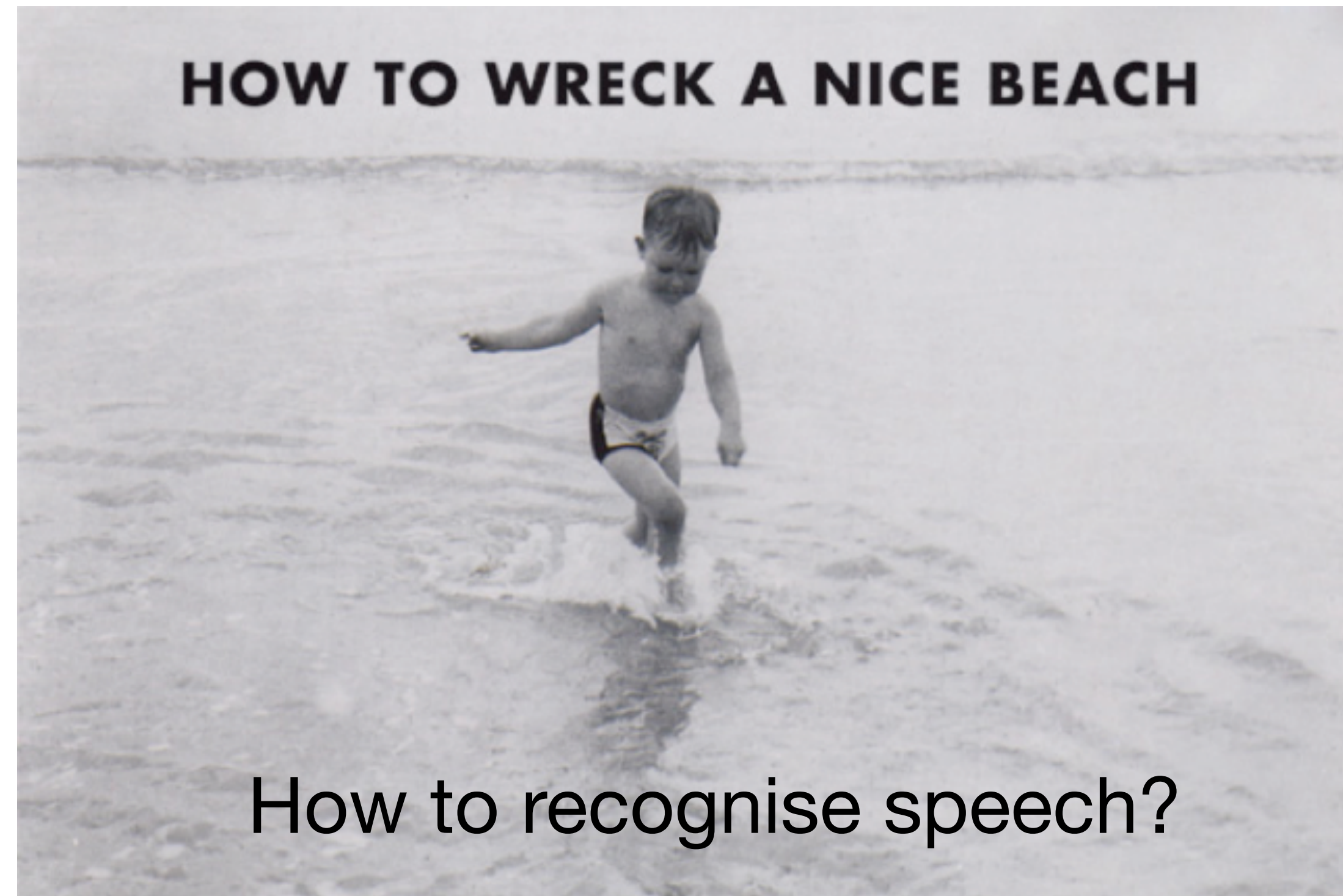
# What is NLP?



- Building computer systems that understand and generate natural languages.
  - Deep understanding of broad language
  - not just string processing or keyword matching
- Can you think of NLP application?




# Speech Recognition




- Not "NLP" -- not covered here!


# Machine Translation

[Web](#) [Images](#) [Videos](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▼




From: English - detected ▼  To: Vietnamese ▼ Translate

Will Justin Bieber ever hit puberty


 Listen


English to Vietnamese translation

Justin Bieber sẽ bao giờ đến tuổi dậy thì


 Listen

[Web](#) [Images](#) [Videos](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▼




From: Vietnamese - detected ▼  To: English ▼ Translate

Justin Bieber sẽ bao giờ đến tuổi dậy thì

 Listen

Vietnamese to English translation

Justin will never reach puberty


 Listen

# Personal Assistants





# Information Extraction



[Web](#) [Images](#) [Maps](#) [Shopping](#) [More ▾](#) [Search tools](#)

About 4,380,000 results (0.20 seconds)

Cookies help us deliver our services. By using our services, you agree to our use of cookies.

[OK](#) [Learn more](#)

**Andrew McCallum Homepage**  
[www.cs.umass.edu/~mccallum/](http://www.cs.umass.edu/~mccallum/) ▾  
Machine learning, text and information retrieval and extraction, reinforcement learning.  
[Andrew McCallum Publications](#) - [Andrew McCallum Bio](#) - [People](#) - [Teaching](#)


**Andrew McCallum - London Metropolitan University**  
[www.londonmet.ac.uk/faculties/faculty-of...k.../andrew-mccallum/](http://www.londonmet.ac.uk/faculties/faculty-of...k.../andrew-mccallum/) ▾  
**Andrew** taught English in London secondary schools for 15 years before coming to London Met in 2008. He is course tutor for the PGCE in Secondary English ...

**Andrew McCallum - Wikipedia, the free encyclopedia**  
[en.wikipedia.org/wiki/Andrew\\_McCallum](http://en.wikipedia.org/wiki/Andrew_McCallum) ▾  
**Andrew McCallum** is a professor and researcher in the computer science department at University of Massachusetts Amherst. His primary specialties are in ...

**Andrew Mccallum - United Kingdom profiles | LinkedIn**  
[uk.linkedin.com/pub/dir/Andrew/Mccallum](http://uk.linkedin.com/pub/dir/Andrew/Mccallum) ▾  
View the profiles of professionals on LinkedIn named **Andrew Mccallum** located in the United Kingdom. There are 25 professionals named **Andrew Mccallum** in ...






**Andrew McCallum**  
Software Developer

Andrew McCallum is a professor and researcher in the computer science department at University of Massachusetts Amherst. [Wikipedia](#)



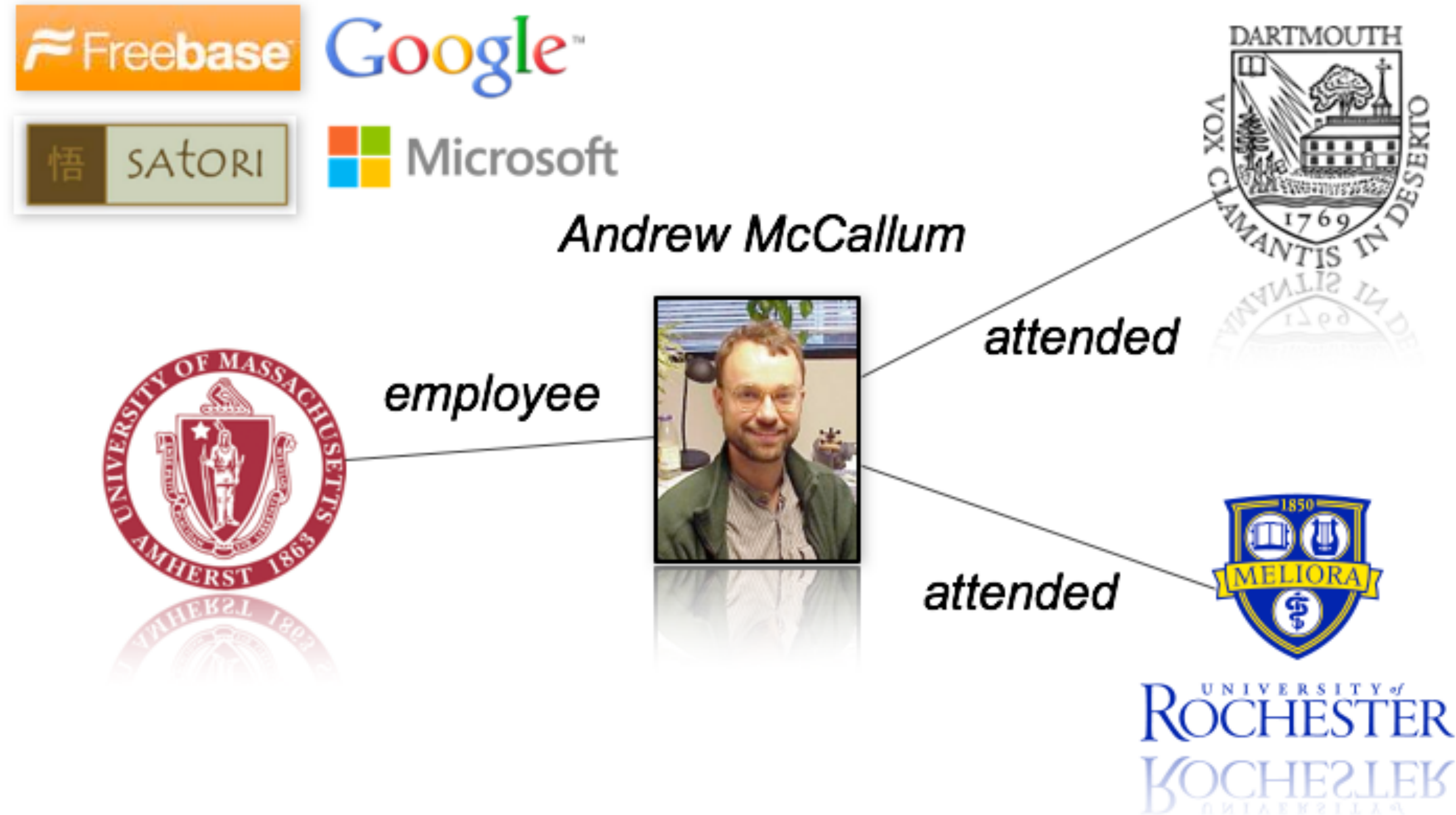
**Education:** [Dartmouth College](#), [University of Rochester](#)  
**Awards:** Best 10-year Paper Award of the ICML

People also search for

[Tom M. Mitchell](#)[Lee Giles](#)[David M. Blei](#)[Michael Collins](#)[Robert Schapire](#)

[Feedback/More info](#)

# Information Extraction





# Summarization



Japan dispatched four military ships to help Russia rescue seven crew members aboard a small submarine trapped on the seabed in the Far East. The Russian Pacific Fleet said the crew had 120 hours of oxygen reserves *on board when the submarine submerged at midday Thursday (2300 GMT Wednesday) off the Kamchatka peninsula, the stretch of Far Eastern Russia facing the Bering Sea.* The submarine, *used in rescue, research and intelligence-gathering missions,* became stuck at the bottom of the Bay of Berezovaya off Russia's Far East coast when its propeller was caught *in a fishing net.* The Russian submarine had been tending an underwater antenna mounted to the sea floor *when it became snagged on a wire helping to stabilize a ventilation cable attached to the antenna.* Rescue crews lowered a British remote-controlled underwater vehicle to a Russian mini-submarine trapped *deep* under the Pacific Ocean, hoping to free the vessel and its seven trapped crewmen *before their air supply ran out.*



# Generation



**Generated story about image**

Model: Romantic Novels

*“He was a shirtless man in the back of his mind, and I let out a curse as he leaned over to kiss me on the shoulder.”*

*He wanted to strangle me, considering the beautiful boy I’d become wearing his boxers.”*



# Question Answering

## IBM's Watson Gets Its First Piece Of Business In Healthcare



34 comments, 33 called-out

+ Comment Now

+ Follow Comments

[IBM](#)'s Watson, the Jeopardy!-playing supercomputer that [scored one for Team Robot Overlord](#) two years ago, just put out its shingle as a doctor or, more specifically, as a combination lung cancer specialist and expert in the arcane branch of health insurance known as utilization management. Thanks to a business partnership among IBM, Memorial Sloan-Kettering and [WellPoint](#), health care providers will now be able to tap Watson's expertise in deciding how to treat patients.



The old Watson that beat Ken Jennings. Now it can fit into a desk drawer. (Credit: Getty Images via @daylife)



# Sentiment Analysis

Sentiment140

NLP

English

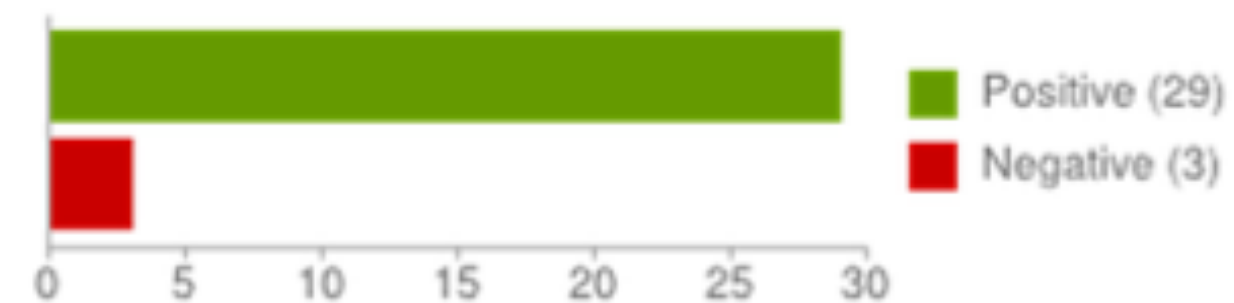
Search

## Sentiment analysis for NLP

Sentiment by Percent



Sentiment by Count



JordanBone1: Recently enrolled on to a **NLP** life coaching course & now an online wedding planning course! ???? I love groupon & wowcher lol

Posted: 6 hours ago

BluffMasterPUA: this guy on the bus next to me was breaking some chick down with logic. Crazy **NLP** but he looked like a

Posted: 9 hours ago

# Machine Comprehension

"The rock was still wet. The animal was glistening, like it was still swimming," recalls Hou Xianguang. Hou discovered the unusual fossil while surveying rocks as a paleontology graduate student in 1984, near the Chinese town of Chengjiang. "My teachers always talked about the Burgess Shale animals. It looked like one of them. My hands began to shake." Hou had indeed found a *Naraoia* like those from Canada. However, Hou's animal was 15 million years older than its Canadian relatives.

It can be inferred that Hou Xianguang's "hands began to shake", because he was:

- (A) afraid that he might lose the fossil
- (B) worried about the implications of his finding
- (C) concerned that he might not get credit for his work
- (D) uncertain about the authenticity of the fossil
- (E) excited about the magnitude of his discovery



# Cognitive Science and Psycholinguistics



# Why is it Difficult?





# Why is it Difficult?



# Why is it Difficult?

- Lexical Ambiguity
- Structural Ambiguity
- Synonymy
- Anaphora
- Language Change
- Long Distance Phenomena
- Sarcasm, Figurative Language
- ...



# Lexical Ambiguity

- **Words** with **identical** forms have **different** meanings
  - Bank vs Bank (en: bench vs financial institute)
  - Steuer vs Steuer (de: steering wheel vs tax)
- Fool a Machine Translator using lexical ambiguity

# Structural Ambiguity

- **Sentences** with **identical** forms have **different** meanings
  - The professor said on Monday he would give an exam.
  - The burglar threatened the student with the knife.
- Play the Structural Ambiguity Game
- Fool a Machine Translator using structural ambiguity



# Synonymy

- **Words** with **different** forms have **the same** meanings
  - Instance vs Example (en)
  - Lenkrad vs Steuer (de)
- Fool a Machine Reading model using synonyms

# Anaphora

- Use of expressions that **depend** specifically upon **previous expressions**
  - Susan dropped the plate. **It** shattered loudly.
- Fool a Machine Translator using anaphora



# Language Change

- Language and the way it is used is constantly changing
  - New words coming in: bigly, rona, bothsideism
  - Old words on the way out: betwixt
- Count ngrams with the [Google N-gram viewer](#)

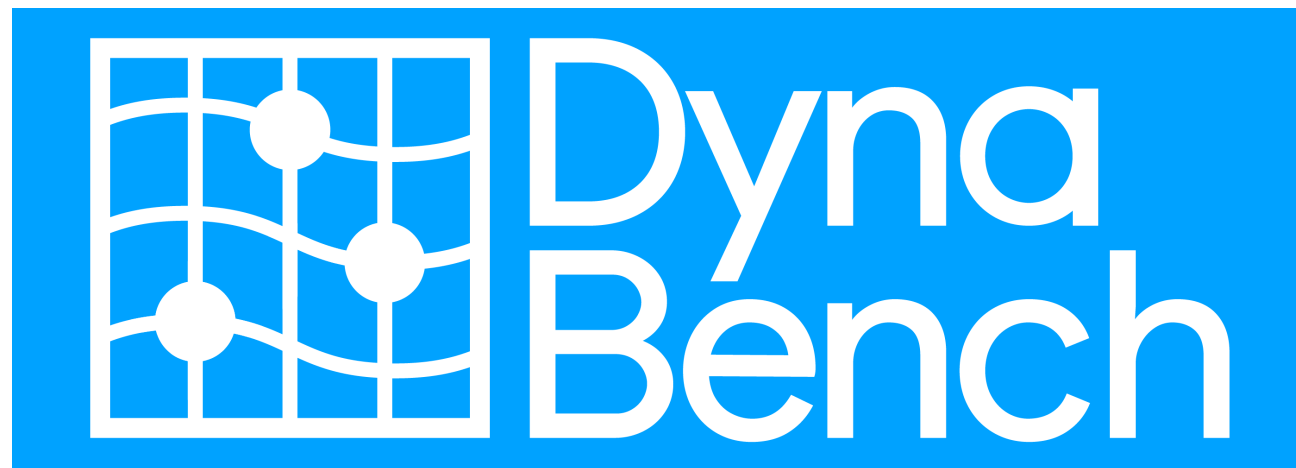
# “Long Distance” Phenomena

- Words and their dependent words might be far from each other (and have complex content in between);
  - That's the elephant Chris said Ben believes you like.
  - The man who recently read a book by a woman that likes food likes wine (linguists won't really call this long distance)
- Fool a Machine Reading model using long distance phenomena



# Figurative Speech

- Ordinary sentences refer to something without directly stating it.
  - My roommate is going through a rollercoaster of emotions.
  - I have told you a million times to wash the dishes.
- Fool a Sentiment Analyser using figurative speech



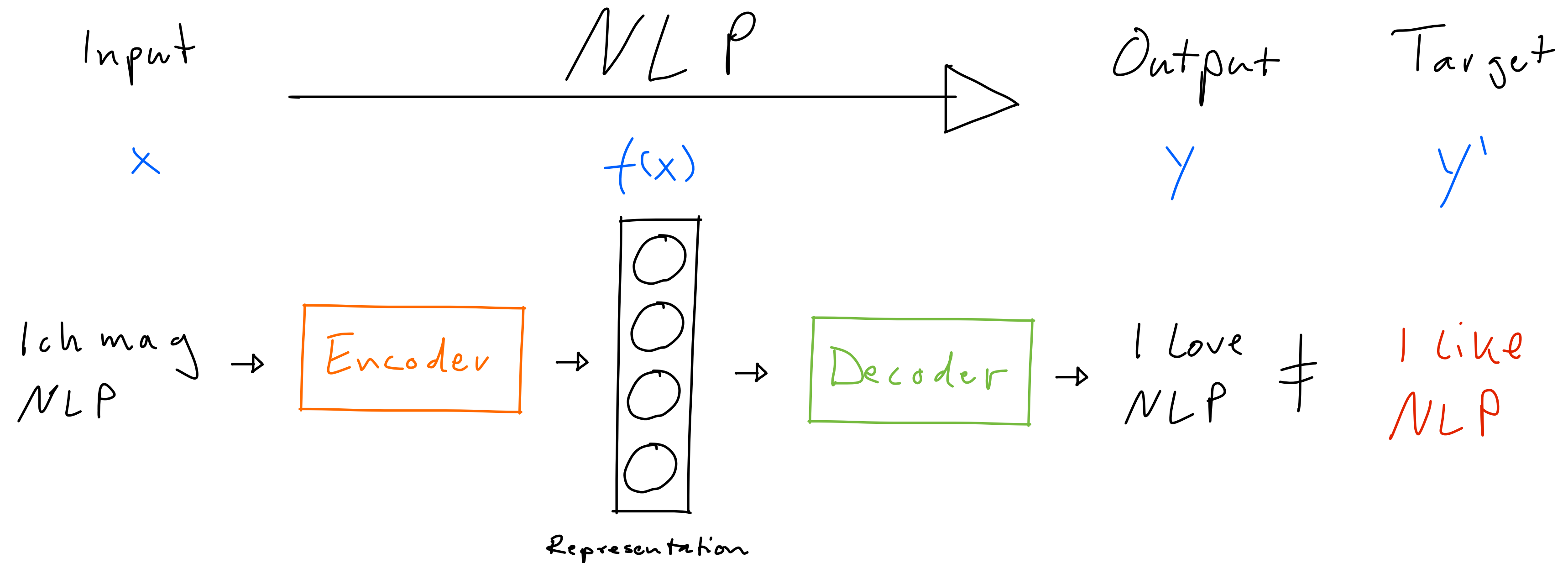
# Dynabench



- In groups (assuming blackboard works out for this):
  - Visit <https://dynabench.org>
  - Choose a task and develop/find an adversarial example
  - Take a screenshot
  - Find “explanation” for this error
  - Volunteer to share (or get “volunteered” if no group wants to)
  - Spend around 10 minutes



# NLP in a Nutshell



- |             |           |             |            |             |              |
|-------------|-----------|-------------|------------|-------------|--------------|
| · Sentences | · Manual  | · symbolic  | · softmax  | · labels    | · supervised |
| · Documents | · Learnt  | · sparse    | · sequence | · sentences | · weak       |
| · Languages | · Pooling | · dense     | · decoder  | · trees     | · unsup.     |
| · Domains   | · RNN     | · attention | · tree     | · graphs    | · semi-sup.  |
| · Databases | · CNN     |             | · decoder  |             |              |