

Natural Language Processing

EECS 348 Spring 2014

Doug Downey

(some examples from Dan Jurafsky's
NLP MOOC)

Natural Language Processing

- 1) What we can do, what we can't
- 2) What makes NLP difficult
- 3) Some key building blocks of NLP systems
 - A. Machine Learning, graphical models
 - B. Massive text corpora
 - C. Language Models (with neural nets)

What we can do

- Pretty well:
 - Factoid question answering
 - Part-of-Speech tagging
 - Named Entity Recognition
 - Document Classification
 - Focused-domain Information Extraction

IBM's Watson



“I for one welcome our new robot overlords.”
-- Ken Jennings

Computers bested humans in Jeopardy in 2011

Part of Speech Tagging

Named Entity Recognition

Document Classification

- *Local Ukrainians get angry with pro-Russia separatists...A day after President Vladimir Putin issued a pullout order in an ...*

=> Sports or Politics?

Restricted-domain Information Extraction

Join us on June 5
for a Semantic
Web seminar in
Room 302. . .



DATE:
TOPIC:
ROOM: 302
...

June 5
Semantic Web

- Small scale
- Restricted Domain
- Hand-labeled examples provided

Train

e.g. [Sundheim, 1989; Jacobs & Rau, 1990; Freitag 1996]

What we can sort of do

- Parsing
- Sentiment analysis
- Prepositional phrase attachment
- Speech recognition
- Word sense disambiguation
- Machine Translation

What we can't do (1 of 4)

- Harder versions of all the previous tasks, e.g.:
 - Tough entity names:
Intel and Microsoft
vs.
Pride and Prejudice
 - Obscure document classes or word senses, e.g.:
“documents about Chicago the typeface”
 - Speech with multiple speakers, noise, large vocab of named entities
 - General QA/Information Extraction
“How effective is ibuprofen at reducing fever in patients with acute febrile illness?”

What we can't do (2 of 4)

- Paraphrasing
 - “I can’t believe you did that”
 - “You did that? Unbelievable.”
- Dialog
 - “Where is Back to the Future playing in Chicago?”
“Century Evanston on Sunday at 2PM. Do you want to buy tickets?”

What we can't do (3 of 4)

- Summarization

“The Dow is up”

“Housing prices rise”

“unemployment is down”

=> the economy is good

What we can't do (4 of 4)

- Final Jeopardy

Category: **US Cities**

*Its largest airport is named for a World War II hero,
its second largest for a World War II battle.*

- Answer: Chicago!
- Watson's answer: Toronto 

Natural Language Processing

- 1) What we can do, what we can't
- 2) What makes NLP difficult**
- 3) Some key building blocks of NLP systems
 - A. Machine Learning, graphical models
 - B. Massive text corpora
 - C. Language Models (with neural nets)

What makes NLP difficult

- Ambiguity
 - Violinist Linked to JAL Crash Blossoms
 - Teacher Strikes Idle Kids
 - Red Tape Holds Up New Bridges
 - Hospitals Are Sued by 7 Foot Doctors
 - Juvenile Court to Try Shooting Defendant
 - Local High School Dropouts Cut in Half

(All REAL headlines, from Dan Jurafsky)

What makes NLP difficult

Neologism!



- Ambiguity
 - Violinist Linked to JAL **Crash Blossoms**
 - Teacher Strikes Idle Kids
 - Red Tape Holds Up New Bridges
 - Hospitals Are Sued by 7 Foot Doctors
 - Juvenile Court to Try Shooting Defendant
 - Local High School Dropouts Cut in Half

(All REAL headlines, from Dan Jurafsky)

Ambiguity is everywhere

- Time flies like an arrow
- Fruit flies like a banana
- Mary and Sue are sisters.
- Mary and Sue are mothers.
- Kick the bucket
- Throw in the towel

Generics

- Birds lay eggs
 - Birds are female
-
- Mosquitoes carry the West Nile Virus
 - Humans have blue eyes and brown hair

What do we need?

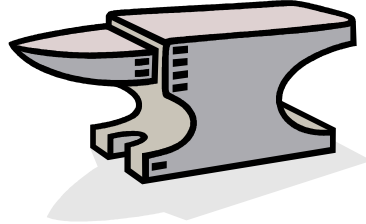
- Knowledge about the world
- Knowledge about language
- Key building block: probabilistic models
 $P(\text{"maison"} \rightarrow \text{"house"})$ **high**
 $P(\text{"L'avocat general"} \rightarrow \text{"the general avocado"})$ **low**
- Fact remains: human-level NLP is "AI-Complete"

Natural Language Processing

- 1) What we can do, what we can't
- 2) What makes NLP difficult
- 3) Some key building blocks of NLP systems
 - A. Machine Learning, graphical models**
 - B. Massive text corpora
 - C. Language Models (with neural nets)

Natural Language Processing

- 1) What we can do, what we can't
- 2) What makes NLP difficult
- 3) Some key building blocks of NLP systems
 - A. Machine Learning, graphical models
 - B. Massive text corpora**
 - C. Language Models (with neural nets)



Throw(person, x) ?



Weight(x) < 50lbs ^
Max_dim(x) < 20ft ^ ... ^
 $\Rightarrow$ Throw(person, x)

Weight(baseball) = 5oz ^.... $\Rightarrow$
Throw(person, baseball)

"throwable objects such as"

Web

Images

Maps

Shopping

Books

More ▼

Search tools

About 5,050 results (0.19 seconds)

[Patent US5984812 - Grippable surface for throwable object - Google ...](#)
www.google.com/patents/US5984812

This invention relates to a grippable surface for **throwable objects such as** a football, baseball, etc. which enhances the ease with which the object may be ...

[\[PDF\] Name Juggle.pdf - GOAL Consulting](#)

www.goalconsulting.org/page3/files/Name%20Juggle.pdf ▼

Materials: Many soft **throwable objects such as** fleece balls, wadded up pieces of paper, Nerf™ balls. Level: Grades K and higher. Suggested Procedure. 1.

Cities such as **X**

Y, mayor of **X**

- The Web makes hard AI problems easier
- ...but

Understanding Language is Hard

Michael Bloomberg, mayor of New York City, was born in ...

Eric Johnson, CEO of Texas Instruments, mayor of Dallas from 1964-1971, and ...

Redundancy enables Information Extraction

Michael Bloomberg, mayor of New York City, was born in ...

Announced by Michael Bloomberg, the mayor of New York City...

Michael Bloomberg, mayor of New York City, stated that...

...+10,000s more

Eric Johnson, CEO of Texas Instruments, mayor of Dallas from 1964-1971, and ...

(only once)



Mayor	City	p
Michael Bloomberg	New York City	0.99
Texas Instruments	Dallas	0.05
...	...	...

[Downey et al., AIJ 2010]