

# Coreference & Entity Linking

Prof. Sameer Singh

---

CS 295: STATISTICAL NLP

WINTER 2017

March 9, 2017

# Upcoming...


---

## Homework

- Homework 4 is due on **March 13**
- Lowest grade of the homeworks will be dropped

## Summaries

- Paper summaries: **March 14**
- Summary 2 graded



TA/Instructor  
Evaluations  
are available!

## Project

- Final report due: **March 20, 2017**
- Instructions coming soon, only **5 pages**

# Outline

---

Coreference Resolution

Entity Linking

Question Answering

# Outline

---

Coreference Resolution

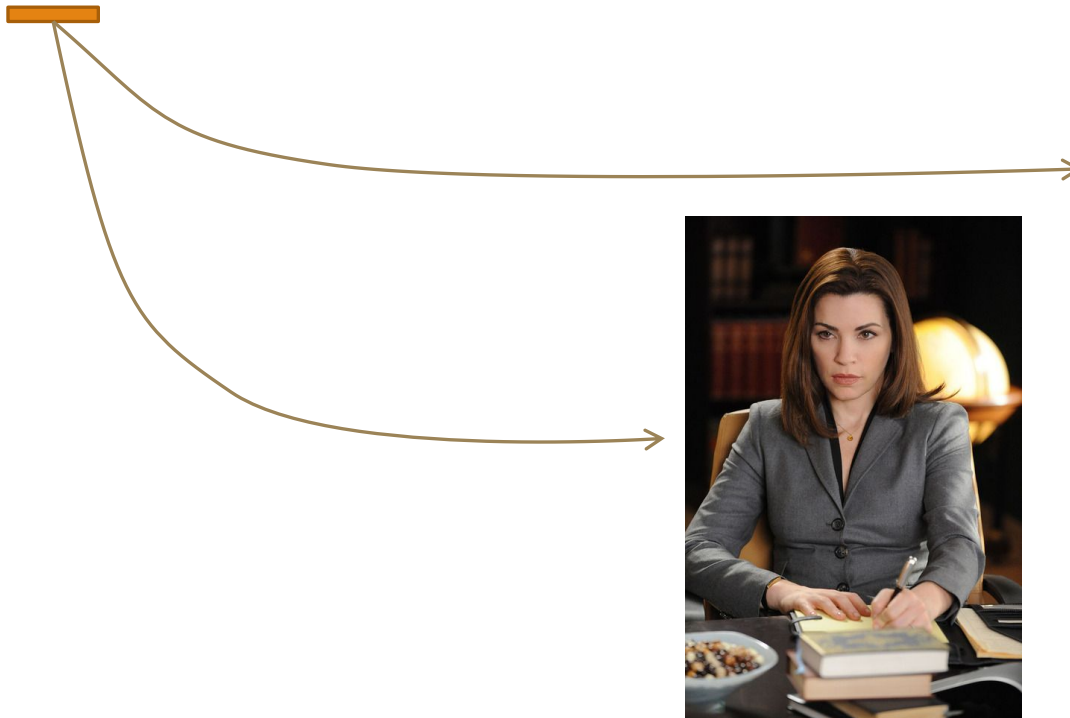
Entity Linking

Question Answering

# Coreference Resolution

---

My girlfriend and I met my lawyer for a drink,  
but she became ill and had to leave.



# Winograd Schema

---

The city councilmen refused the demonstrators a permit because they feared violence.

Randomly 2016 Humans  
44% 32-58% >90%

The city councilmen refused the demonstrators a permit because they advocated violence.

# Coreference Ambiguities

---



# At a Document Level

---

President Barack Obama received the Serve America Act after Congress's vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.



# At a Document Level

---

President Barack Obama received the Serve America Act after Congress's vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.

# At a Document Level

---

President Barack Obama received the Serve America Act after Congress's vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.



# Mentions and Entities

---

## Input: text (and mentions)

President Barack Obama received the Serve America Act after Congress's vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.

## Output: clustering of the mentions in text

President Barack Obama received the Serve America Act after Congress's vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.

# Applications

---

Relation Extraction

He married her in 1927.

Sentiment Analysis

I really loved the movie...  
I liked how evil the villain's plan was.  
It was abhorrent!

QA/Dialog Systems

Find me a flight from LA to London,  
make sure it is not too long.

Summarization

A shooting took place in Wallingford  
this morning... The shooter targeted  
two women.. He is a 34-year old ...

# Semantics vs Pragmatics

---

Semantics

What does the sentence **mean**?

Pragmatics

What does the sentence **imply**?

(+ world knowledge  
common sense)

One tries to be as informative as one possibly can,  
and gives as much information as is needed, **and no more.**

- *Grice's Maxim of Quantity*

# Example: Semantic/Pragmatic

BIG WEST

CONFERENCE TOURNAMENTS

## Big West Tournament primer: It's UC Irvine's to lose

### Semantics

- “UC Irvine” is a team/player
- It has not lost yet
- The event in question might be the “Big West Tournament”

### Pragmatics

- UC Irvine is likely to win
- A team has to win for UC Irvine to lose
- Other team is not good?
- BW Tournament is a college sport
- Must be a big deal, all of west coast?
- Must be happening soon?
- “Primer”: Must be important!

<http://www.midmajormadness.com/2017/3/8/14837458/big-west-tournament-primer-uc-irvine-anteaters-how-to-watch-prediction-bracket-champ-week>

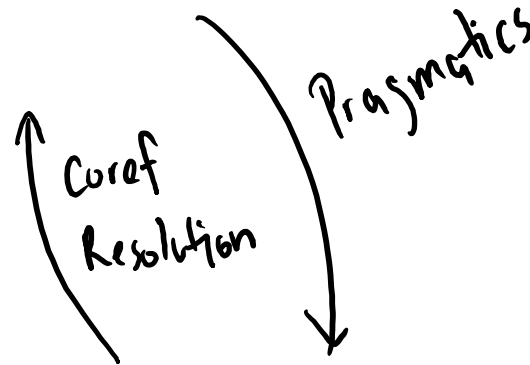
# Reverse Pragmatics

---

President Barack Obama received the Serve America Act after Congress's vote.

President Barack Obama signed the Serve America Act last Thursday.

President Barack Obama said ...



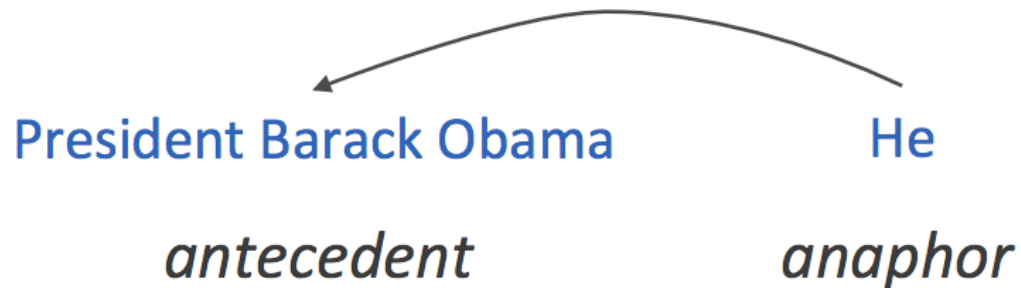
President Barack Obama received the Serve America Act after Congress's vote.

He signed the bill last Thursday.

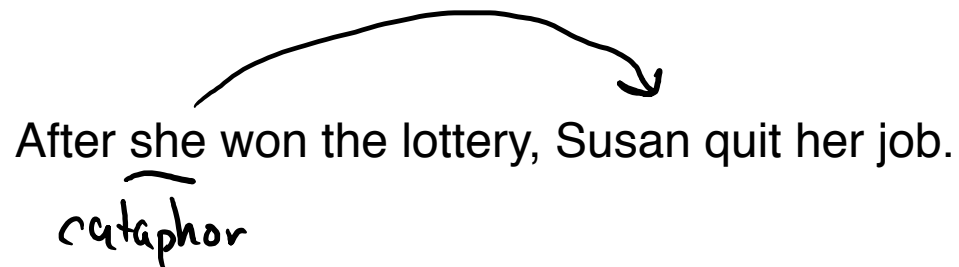
The president said ...

# Antecedents / Anaphor

---



Cataphor





# Types: Proper Names

---

President Barack Obama, 44th president of  
the United States, ...

President Obama

Obama



Lexical  
similarity

# Types: Pronouns

---

President Barack Obama received the Serve America Act after Congress's vote.

He ...

President Barack Obama met with Chancellor Merkel.

He ...

President Barack Obama met with President Hollande after ...

→ he signed the bill.

→ he flew in from Paris.



"agreement",  
salience

# Types: Nominals

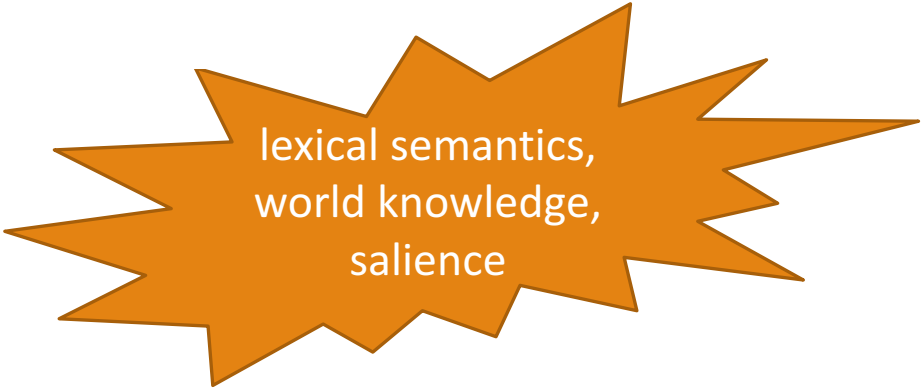
---

President Obama ... The president

Serve America Act ... The bill

Barack Obama and Angela Merkel ... The leaders

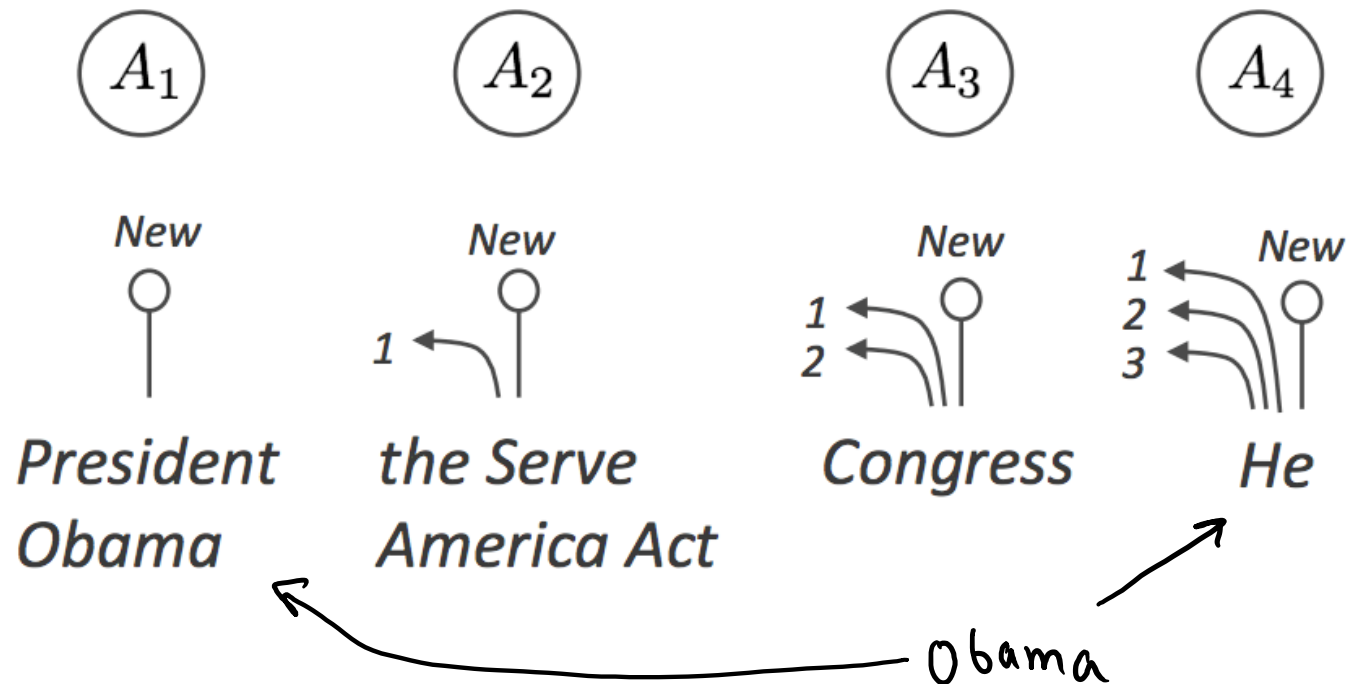
NBC ... The network



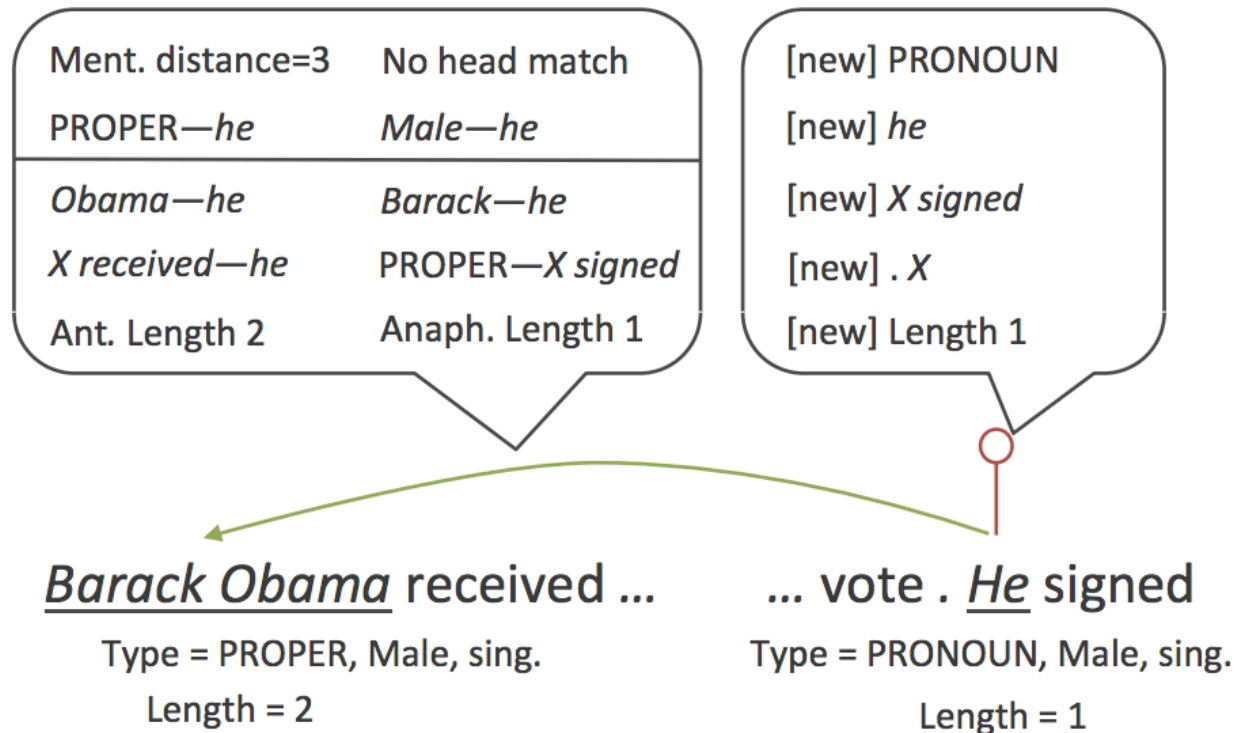
lexical semantics,  
world knowledge,  
salience

# Learning-based Methods

$$Pr(A_i = a|x) \propto \exp(w^\top f(i, a, x))$$



# Learning-based Methods

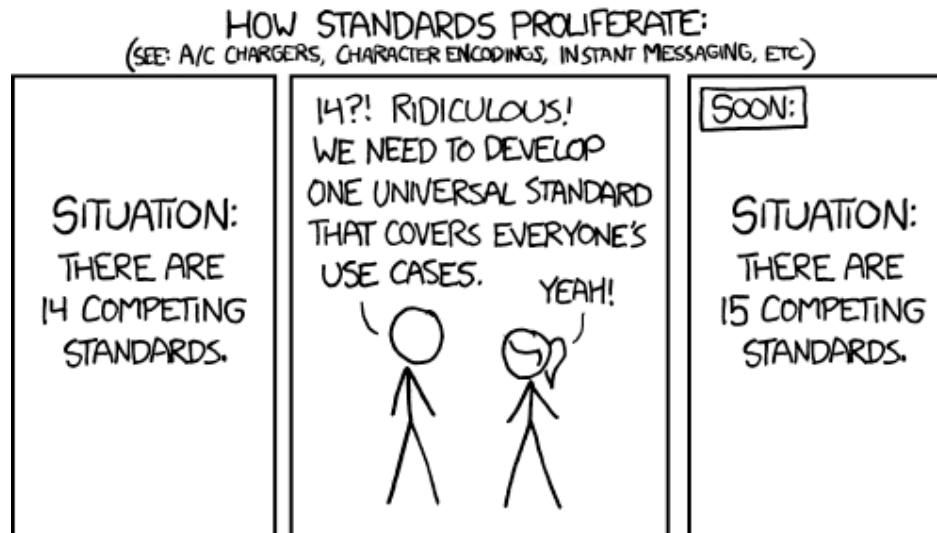


# Evaluation

$m_1$   
 $\vdots$   
 $m_M$

$\hat{c}_1$   
 $\hat{c}_2$   
 $\vdots$   
 $\hat{c}_P$

$c_1$   
 $c_2$   
 $\vdots$   
 $c_T$



# Evaluation Metrics

$\left. \begin{array}{l} P \\ R \\ \text{acc.} \end{array} \right\} F_{1/\beta}$

Pairwise

“How many total edges did you get right?”

MUC

“How many antecedents did you get right?”

B<sup>3</sup> Metric

“How many edges in predicted clusters did you get right?”

CEAF

“Do a maximum matching between predicted and gold entities;  
how close are they?”  $\hookrightarrow$  NP-Hard

CONLL

2011

$$\frac{(B^3 + MUC + CEAF)}{3}$$

# Outline

---

Coreference Resolution

Entity Linking

Question Answering



# Entity Resolution & Linking

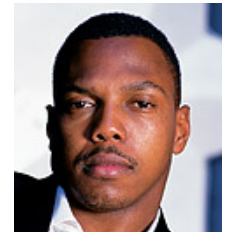
...during the late 60's and early 70's, **Kevin Smith** worked with several local...

...the term hip-hop is attributed to **Lovebug Starski**. What does it actually mean...



Like Back in 2008, the Lions drafted **Kevin Smith**, even though Smith was badly...

... backfield in the wake of **Kevin Smith**'s knee injury, and the addition of Haynesworth...



The filmmaker **Kevin Smith** returns to the role of Silent Bob...

Nothing could be more irrelevant to **Kevin Smith**'s audacious ''Dogma'' than ticking off...



... The Physiological Basis of Politics," by **Kevin Smith**, Douglas Oxley, Matthew Hibbing...



# World Knowledge

---

America Online announced on Monday that the company plans to update its instant messaging service.

Prediction:

America Online announced on Monday that the company plans to update its instant messaging service.

# World Knowledge

---

**America Online** ← → **company**

**America Online, LLC** (commonly known as **AOL**) is an American global Internet services and media company operated by Time Warner. It is headquartered at 770 Broadway in Midtown Manhattan, New York City.<sup>[2][3]</sup> Founded in 1983 as **Quantum Computer Services**, it has franchised its services to companies in several nations around the world or set up international versions of its services.<sup>[4]</sup>

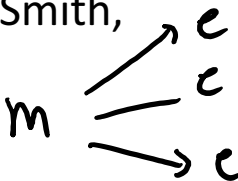
America Online	
	
Type	Subsidiary of Time Warner
Founded	1983 as <i>Quantum Computer Services</i>

# Entity Names: Two Problems

## Entities with Same Name

### Same type of entities share names

Kevin Smith, John Smith,  
Springfield, ...



```
graph LR; m --> e1[e]; m --> e2[e]; m --> e3[e];
```

### Things named after each other

Clinton, Washington, Paris,  
Amazon, Princeton, Kingston, ...

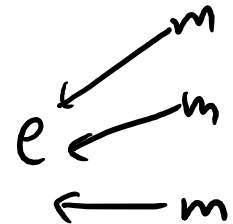
### Partial Reference

First names of people, Location  
instead of team name, Nick names

## Different Names for Entities

### Nick Names

Bam Bam, Drumpf, ...



### Typos/Misspellings

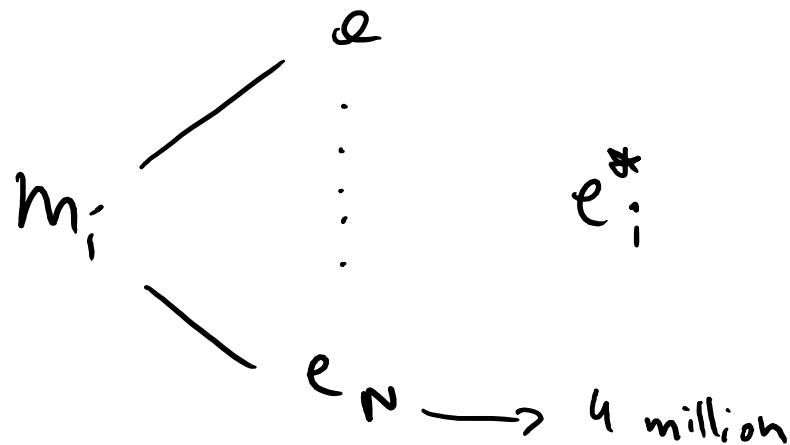
Baarak, Barak, Barrack, ...

### Inconsistent References

MSFT, APPL, GOOG...

# Evaluating Entity Linking

---



# Baseline: Link Probabilities

## Washington drops 10 points after game with UCLA Bruins.

*Washington*

Washington DC,  
George Washington,  
Washington state,  
Lake Washington,  
Washington Huskies,  
Denzel Washington,  
University of Washington,  
Washington High School,  
...

100s

$$P(e|m)$$

$$P(\text{"Washington DC"} | \text{"Washington"})$$

He was born in Washington.

e

# Entity Linking Approach

Washington drops 10 points after game with UCLA Bruins.

## Candidate Generation

Washington DC, George Washington, Washington state, Lake Washington, Washington Huskies, Denzel Washington, University of Washington, Washington High School, ...

## Entity Types

LOC/ORG

Washington DC, ~~George Washington~~, Washington state, Lake Washington, Washington Huskies, ~~Denzel Washington~~, University of Washington, Washington High School, ...

## Coreference

UWashington,  
Huskies

~~Washington DC~~, ~~George Washington~~, ~~Washington state~~, ~~Lake Washington~~, Washington Huskies, ~~Denzel Washington~~, University of Washington, ~~Washington High School~~, ...

## Coherence

UCLA Bruins,  
USC Trojans

~~Washington DC~~, ~~George Washington~~, ~~Washington state~~, ~~Lake Washington~~, Washington Huskies, ~~Denzel Washington~~, ~~University of Washington~~, ~~Washington High School~~, ...

# Global Inference

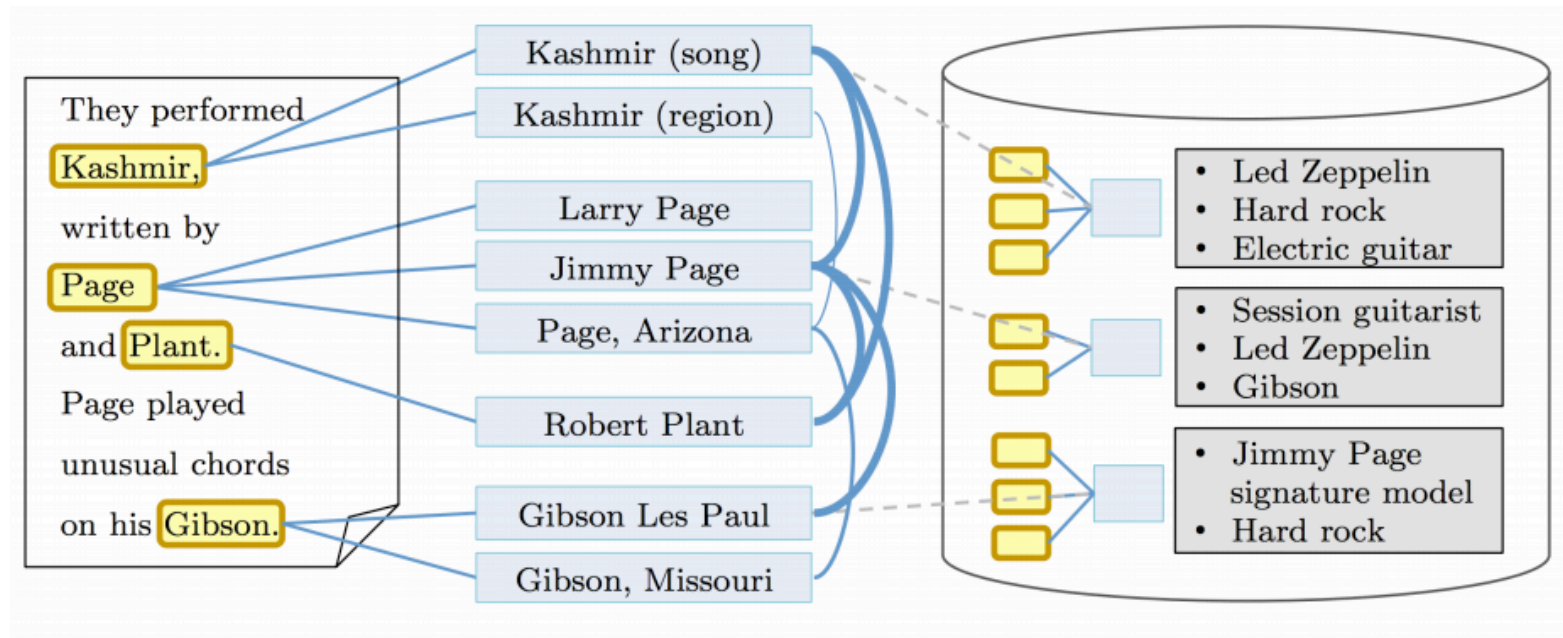


Figure from Hoffart et al. (2011)



# Outline

---

Coreference Resolution

Entity Linking

Question Answering

# Questions are very common

---

who invented surf music?  
how to make stink bombs  
where are the snowdens of yesteryear?  
which english translation of the bible is used in official catholic liturgies?  
how to do clayart  
how to copy psx  
how tall is the sears tower?  
how can i find someone in texas  
where can i find information on puritan religion?  
what are the 7 wonders of the world  
how can i eliminate stress  
What vacuum cleaner does Consumers Guide recommend

Around 10-15% of  
search queries

(  
underestimation

# Applications of QA

---

Natural Language  
Database Systems

Schema-specific matching of text to SQL queries

“List the authors who have written books about business”

```
SELECT firstname, lastname FROM authors, titleauthor,  
titles WHERE authors.id = titleauthor.authors_id AND  
titleauthor.title_id = titles.id
```

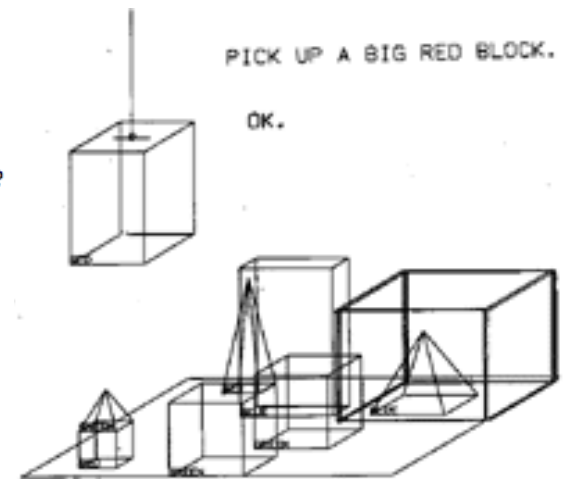
**Early Systems:** BASESBALL (1961) and LUNAR (1977)

# Applications of QA

## Spoken Dialog Systems

Domain-specific dialogs from an environment

Person: WHAT DOES THE BOX CONTAIN?  
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.  
Person: WHAT IS THE PYRAMID SUPPORTED BY?  
Computer: THE BOX.  
Person: HOW MANY BLOCKS ARE NOT IN THE BOX?  
Computer: FOUR OF THEM.  
Person: IS AT LEAST ONE OF THEM NARROWER THAN THE ONE WHICH I TOLD YOU TO PICK UP?  
Computer: YES, THE RED CUBE.  
Person: IS IT SUPPORTED?  
Computer: YES, BY THE TABLE.  
Person: CAN THE TABLE PICK UP BLOCKS?  
Computer: NO.  
Person: CAN A PYRAMID BE SUPPORTED BY A BLOCK?  
Computer: YES.  
Person: CAN A PYRAMID SUPPORT A PYRAMID?  
Computer: I DON'T KNOW.  
Person: STACK UP TWO PYRAMIDS.  
Computer: I CAN'T.



Early Work: SHRDLU, Winograd (1972)

# Applications of QA

---

## Reading Comprehension

Questions from a paragraph, answers in them.

### How Maple Syrup is Made

Maple syrup comes from **sugar maple trees**. At one time, maple syrup was used to make sugar. This is why the tree is called a "sugar" maple tree. Sugar maple trees make sap. **Farmers** collect the sap. The best time to collect sap is in **February and March**. The nights must be cold and the days warm. The farmer drills a few small holes in each tree. He puts a spout in each hole. Then he hangs a **bucket** on the end of each spout. The bucket has a cover **to keep rain and snow out**. The sap drips into the bucket. About 10 gallons of sap come from each hole.

- Who collects maple sap? (**Farmers**)
- What does the farmer hang from a spout? (**A bucket**)
- When is sap collected? (**February and March**)
- Where does the maple sap come from? (**Sugar maple trees**)
- Why is the bucket covered? (**to keep rain and snow out**)

**Early Work:** QUALM, Lehnert (1977)

# Applications of QA

---

## Open-domain QA

Questions about anything, have huge corpus available!

Last 15 years or so, the whole web!

*“When was Mozart born?”*

- TREC**
- Annual competition of open-ended question answering
  - Provides a text corpus, and a collection of *factoid* questions
  - Made more difficult every year
    - Questions more “realistic”
    - Answer may not be in the corpus
    - Submit only one answer, not a ranked list, ...

# TREC Example Questions

---

- Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"?
- What was the monetary value of the Nobel Peace Prize in 1989?
- What does the Peugeot company manufacture?
- How much did Mercury spend on advertising in 1993?
- What is the name of the managing director of Apricot Computer?
- Why did David Koresh ask the FBI for a word processor?
- What debts did Quintex group leave?
- What is the name of the rare neurological disease with symptoms such as: involuntary movements (tics), swearing, and incoherent vocalizations (grunts, shouts, etc.)?