

Teaching Computers to Understand Human Languages

(Natural Language Processing)

Kai-Wei Chang
CS @ UCLA
kw@kwchang.net

See <http://kwchang.net> for more information

Q: [Chris] = [Mr. Robin] ?

Christopher Robin is alive and well. He is the same person that you read about in the book, **Winnie the Pooh**. As a boy, **Chris** lived in a pretty home called **Cotchfield Farm**. When **Chris** was three years old, **his father** wrote a poem about **him**. The poem was printed in a magazine for others to read. **Mr. Robin** then wrote a book

Slide modified from Dan Roth

Complex Decision Structure

Christopher Robin is alive and well. He is the same person that you read about in the book, **Winnie the Pooh**. As a boy, **Chris** lived in a pretty home called **Cotchfield Farm**. When **Chris** was three years old, **his father** wrote a poem about **him**. The poem was printed in a magazine for others to read. **Mr. Robin** then wrote a book

Q: Who wrote Winnie the Pooh?

Christopher Robin is alive and well. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book

Slide modified from Dan Roth

Co-reference Resolution

Christopher Robin is alive and well. **He** is the same person that you read about in the book, **Winnie the Pooh**. As a **boy**, **Chris** lived in a pretty home called **Cotchfield Farm**. When **Chris** was three years old, **his father** wrote a poem about **him**. The poem was printed in a magazine for others to read. **Mr. Robin** then wrote a book

Q: Who wrote Winnie-the-Pooh?

- ❖ How to understand the sentence:

Who wrote Winnie-the-Pooh

WP VBD NNP .
who wrote Winnie-the-Pooh ?

who **wrote** Winnie-the-Pooh ?

TEXT_CREATION

Author

Text

- ❖ Convert the question to a logic form
“Winnie-the-Pooh/Creator”

Algebra Word Problems

Problem:

Maria is now four times as old as Kate. Four years ago, Maria was six times as old as Kate. Find their ages now.

Equations: $m = 4 \times n$ and $m - 4 = 6 \times (n - 4)$

Solution: $m = 40, n = 10$

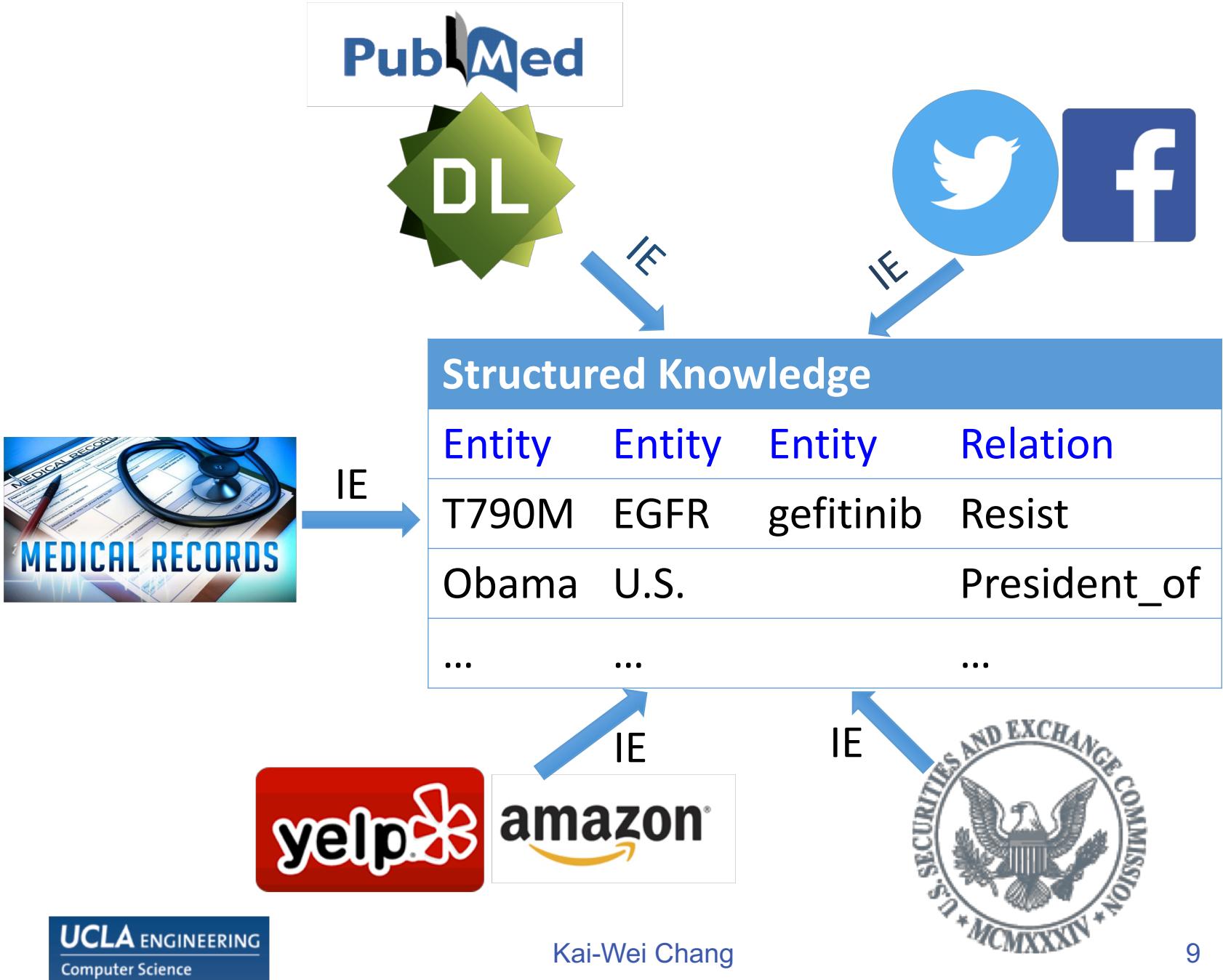
Information Extraction

❖ Unstructured text to database entries

New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

Yoav Artzi: Natural language processing



Structured Knowledge

Entity	Entity	Entity	Relation
T790M	EGFR	gefitinib	Resist
Obama	U.S.		President_of
Philadelphia	PA.		Locate_in
...



Machine translation

Google

buenas noches

All Images Shopping Apps

About 20,800,000 results (0.54 seconds)

Haaretz הארכט
אלוי הפטם הוא צילוח?

רומי: נתנוו שוב ונסה להזכיר את ד"י רביברברג במשילה
בבמישלה
www.haaretz.co.il

אחרי הפעםה שבחשו שערם, בלחשת רחל'ן מסרבם להתחזק שתורה הזעינה בהם דמיון. בירושלמי, ערך האפואות דוריושם יזכיר או השמעה של תלך מחרמלצת

Like · Comment · Translate · Share · Yesterday at 06:00
9 people like this.
View 1 share

שוויז זקס עם שלם מישלים את מחר החוחות והקמביבות של ראש הממשלה. לאיא פזא ליכן לחתיגס ליבוריאת שערת הרירות. סיכון חום גאנשרים לילוחרים גומ לסל'תוב לגאש את חזרו הוועס מהעיר. מאיר לב' כבר וווקן. פאיוורה פה. אאהו הוה ווועס שפתקנות את המבדין וככלת את אונדרה. בול' בון'ינון ננטהו.

תנינו ווועיג: לאו מותחיב להצבעה על טראטונג'ה
www.ynet.co.il

השותפות האקליפטוזות הצליחו להטיל מרא על ראש הממשלה, שחבר יאוש מותחיב להצבעה רום את סוכקנות דרי' רוככינגרה. שעס ליל' שעה מונגדה. ישאל ריבנו ומילגת העמימות קיטו בבורק שבט שום ווילוטן קר'ה. סוליטיט, דרישות
Expand preview ▾

Yesterday at 06:31 · Like · 2 people · Translate

עמוי ובקו או שחד'ה'ה שבר או שביב עיבור
Yesterday at 07:10 · Like · Translate

Dalya Guman Stop vote once to
Yesterday at 08:11 · Like · Original

Yuval Gilor Tired already go home
Yesterday at 08:49 · Like · Original

Haaretz הארכט
Maybe this time he succeeds? [?]

רומי: נתנוו שוב ונסה להזכיר את ד"י רביברברג במשילה
בבמישלה
www.haaretz.co.il

אחרי הפעםה שבחשו שערם, בלחשת רחל'ן מסרבם להתחזק שתורה הזעינה בהם דמיון. בירושלמי, ערך האפואות דוריושם יזכיר או השמעה של תלך מחרמלצת

Like · Comment · Original · Share · Yesterday at 06:00
9 people like this.
View 1 share

With full pay the price for hizkioth vakonbinot of the Prime Minister. Not find it appropriate to action lhahat the health system. Endangered animals, people and caused the suffering that must drive with ... city. Foundation of Minneapolis had already thrown him. Europe also. This man destroys the country with the citizens in its path.
Contempt you Binyamin Netanyahu.
Netanyahu was not committed to voting the Trachtenberg managed to impose a "partnership" hkoalition terror the Prime Minister, who is pledging to bring voting day the conclusions report Trachtenberg. ש' hourly opposes. Israel Beiteinu-independence morning ministerial session will vialto only it-political system, news
Expand preview ▾

Yesterday at 06:31 · Like · 2 people · Original

Or acknowledging the report moves or the chips will go
Yesterday at 07:10 · Like · Original

Stop vote once to
Yesterday at 08:11 · Like · Original

Tired already go home
Yesterday at 08:49 · Like · Original

Facebook translation, image credit: Meedan.org

Spanish ▾

buenas noches [Edit](#)

3 more translations

English ▾

Goodnight

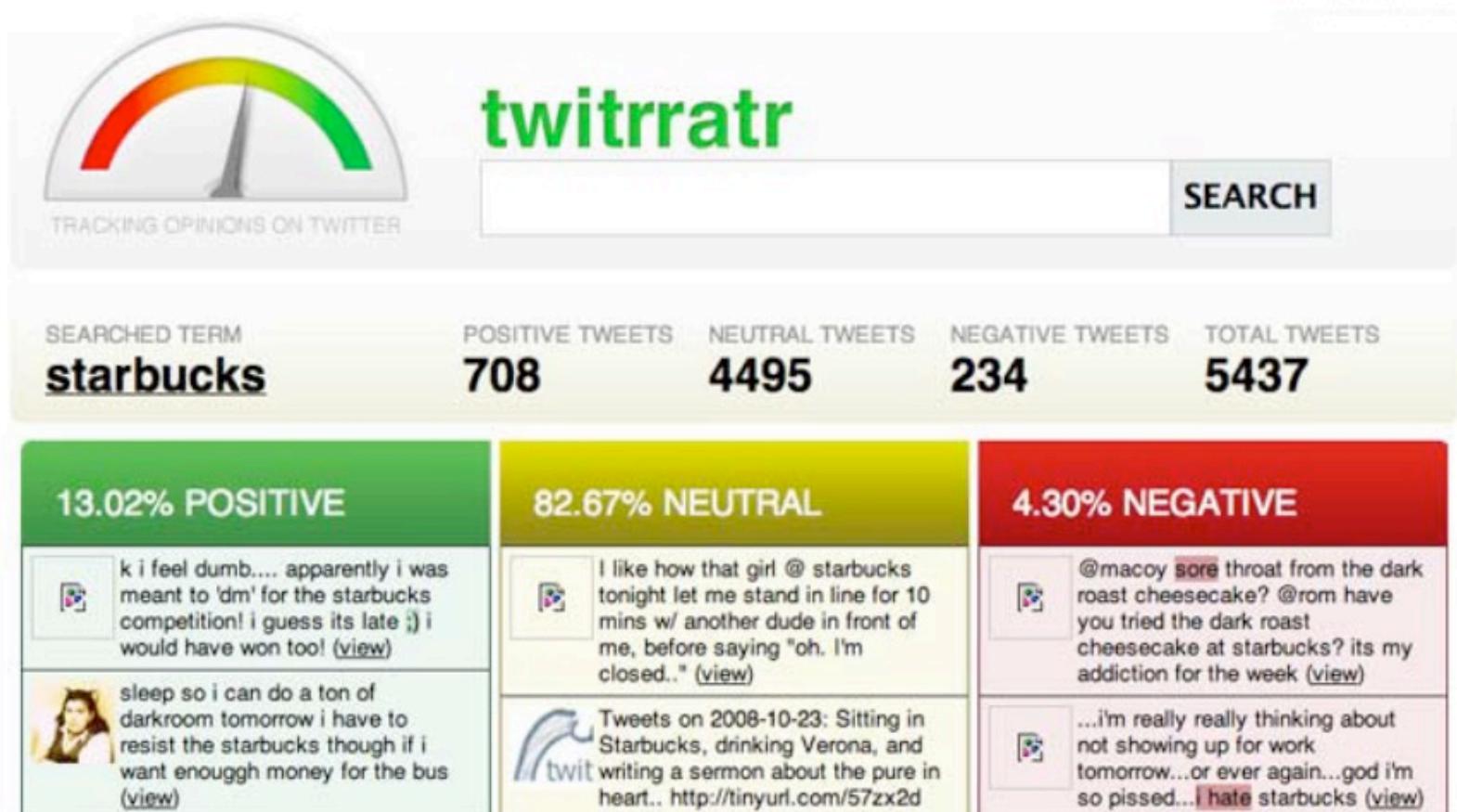
Open in Google Translate

UCLA ENGINEERING
Computer Science

ML in NLP

11

Sentiment/Opinion Analysis



Brainstorming

- ❖ List 3 NLP applications.
- ❖ What NLP applications you hope to see in the future?

Always working?!



Expectation



Reality

<http://viralscape.com/travel-expectations-vs-reality/>

Kai-Wei Chang



Credit: <http://stuffsirisaid.com>



The diagram shows a comparison between two versions of a warning message. On the right, a white box contains the English text "Carefully Slide". A black arrow points down to a screenshot of a translation interface. The interface has "Translate" at the top. It shows a source input field containing "小心地滑" (Xiǎoxīn dì huá) and a target output field containing "Carefully slide". The "English" button is highlighted in blue.

Why is NLP hard

Ambiguity

I saw her duck



Challenges in Language

I saw her duck



Challenges in Language

I saw her duck



Challenges in Language

I saw her duck with a telescope



Challenges – ambiguity

❖ PP attachment ambiguity

San Jose cops kill man with knife

Text Paper Close

Translate Listen

San Jose cops kill man with knife

Ex-college football player, 23, shot 9 times allegedly charged police at fiancee's home

By Hamed Aleaziz and Vivian Ho

A man fatally shot by San Jose police officers while allegedly charging at them with a knife was a 23-year-old former football player at De Anza College in Cupertino who was distraught and depressed, his family said

Thursday.

Police officials said two officers opened fire Wednesday afternoon on Phillip Watkins outside his fiancee's home because they feared for their lives. The officers had been drawn to the home, officials said, by a 911 call reporting an armed home invasion

that, it turned out, had been made by Watkins himself.

But the mother of Watkins' fiancee, who also lives in the home on the 1300 block of Sherman Street, said she witnessed the shooting and described it as excessive. Faye Buchanan said the confrontation happened shortly after she called a suicide intervention hotline in hopes of getting Watkins medical help.

Watkins' 911 call came in at 5:01 p.m., said Sgt. Heather Randal, a San Jose police spokeswoman. "The caller stated there was a male breaking into his home armed with a knife," Randal said. "The caller also stated he was locked in an upstairs bedroom with his children and request-

ed help from police."

She said Watkins was on the sidewalk in front of the home when two officers got there. He was holding a knife with a 4-inch blade and ran toward the officers in a threatening manner, Randal said.

"Both officers ordered the suspect to stop and drop the knife," Randal said. "The suspect continued to charge the officers with the knife in his hand. Both officers, fear-

ing for their safety and defense of their life, fired at the suspect."

On the police radio, one officer said, "We have a male with a knife. He's walking toward us."

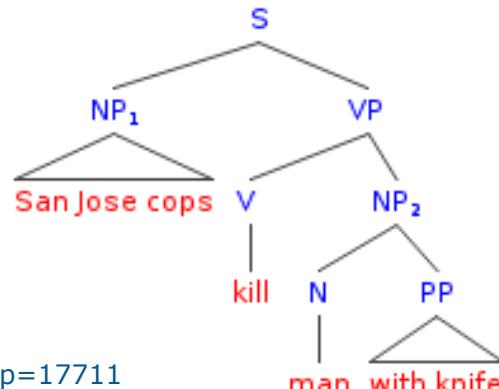
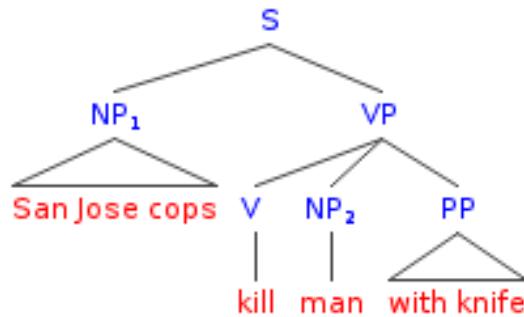
"Shots fired! Shots fired!" an officer said moments later.

A short time later, an officer reported, "Male is down. Knife's still in hand."

Buchanan said she had been prompted to call the

Shoot continues on D8

Back Continue



Credit: Mark Liberman, <http://languagelog.ldc.upenn.edu/nll/?p=17711>

Challenges -- ambiguity

- ❖ Ambiguous headlines:
 - ❖ Include your children when baking cookies
 - ❖ Local High School Dropouts Cut in Half
 - ❖ Hospitals are Sued by 7 Foot Doctors
 - ❖ Iraqi Head Seeks Arms
- ❖ Safety Experts Say School Bus Passengers Should Be Belted
- ❖ Teacher Strikes Idle Kids



Challenges – ambiguity

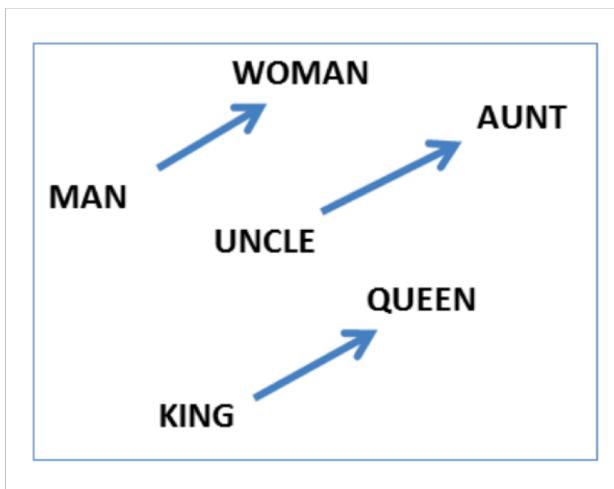
❖ Word sense ambiguity



Word Embeddings can be Dreadfully Sexist

[nips16, **reported by NPR, MIT tech review**]

❖ $v_{man} - v_{woman} + v_{uncle} \sim v_{aunt}$



he: __	she: __
uncle	aunt
lion	
surgeon	
architect	
beer	
professor	

We use Google w2v embedding trained from the news

Human Bias in Structured Prediction Models

[EMNLP 17] w/ Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez

What's the agent for this image?



Cooking	
Role	Object
agent	?
food	vegetable
container	bowl
tool	knife
place	kitchen

An example from a vSRL (visual Semantic Role Labeling) system

Dataset Gender Bias



33%

Male

66%



Female

Model Bias After Training

16%



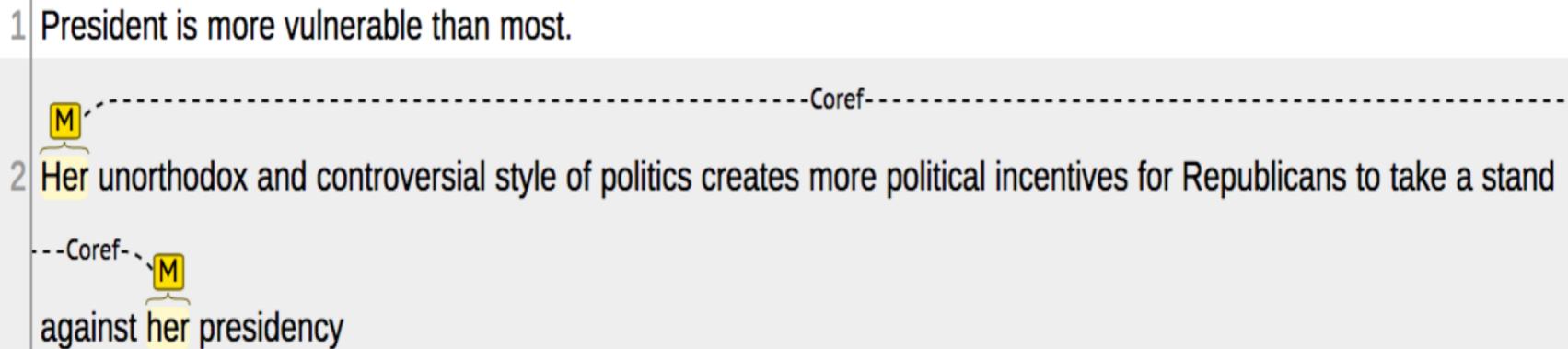
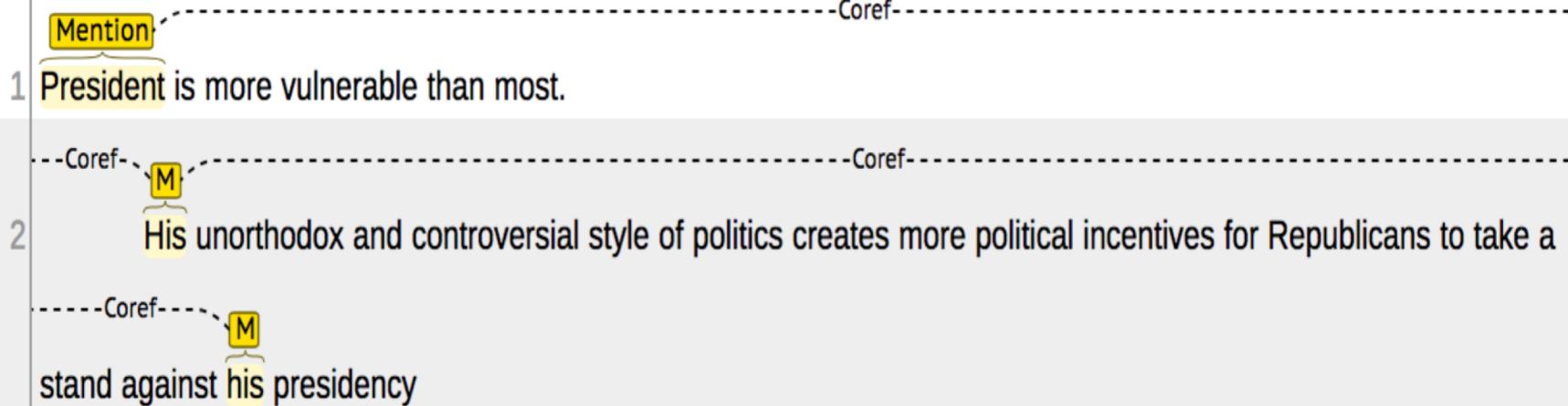
84%



Male

Female

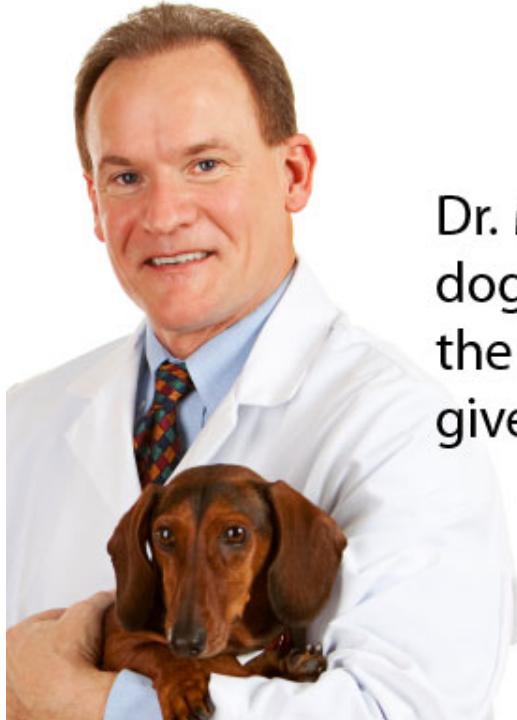
Gender Bias in Coref [NAACL 2018]



Concurrent work (Rudinger et al., 2018) @NAACL18 also studied gender bias in Coref.

Challenges – ambiguity

❖ Pronoun reference ambiguity



Dr. Macklin often brings his dog Champion to visit with the patients. **He** just loves to give big, wet, sloppy kisses!

Credit: <http://www.printwand.com/blog/8-catastrophic-examples-of-word-choice-mistakes>

Challenges--language is compositional



The diagram illustrates the concept of language compositionality. On the left, a white sign with a black border contains the text "Carefully Slide". A black arrow points from this sign down to a screenshot of a translation interface. The interface shows a "Translate" bar with "English" selected. Below it, a box contains the Chinese text "小心地滑" (Xiaoxin di huá), which is being translated into English as "Carefully slide". The English translation box is highlighted with a black rectangle.

Challenges--language is compositional



小心:
Carefully
Careful
Take
Care
Caution



地滑:
Slide
Landslip
Wet Floor
Smooth

Translate

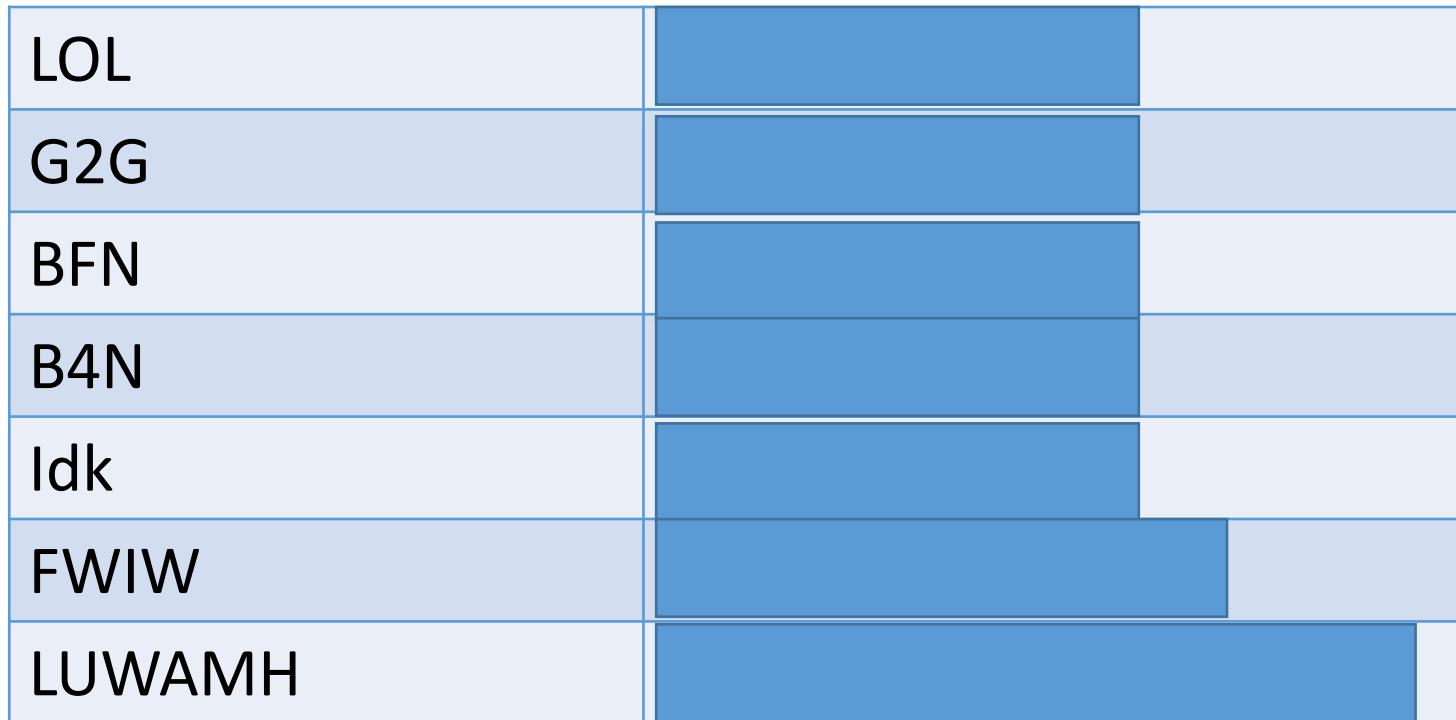
English Spanish French Chinese - detected English Spanish Arabic Translate

小心地滑 Carefully slide

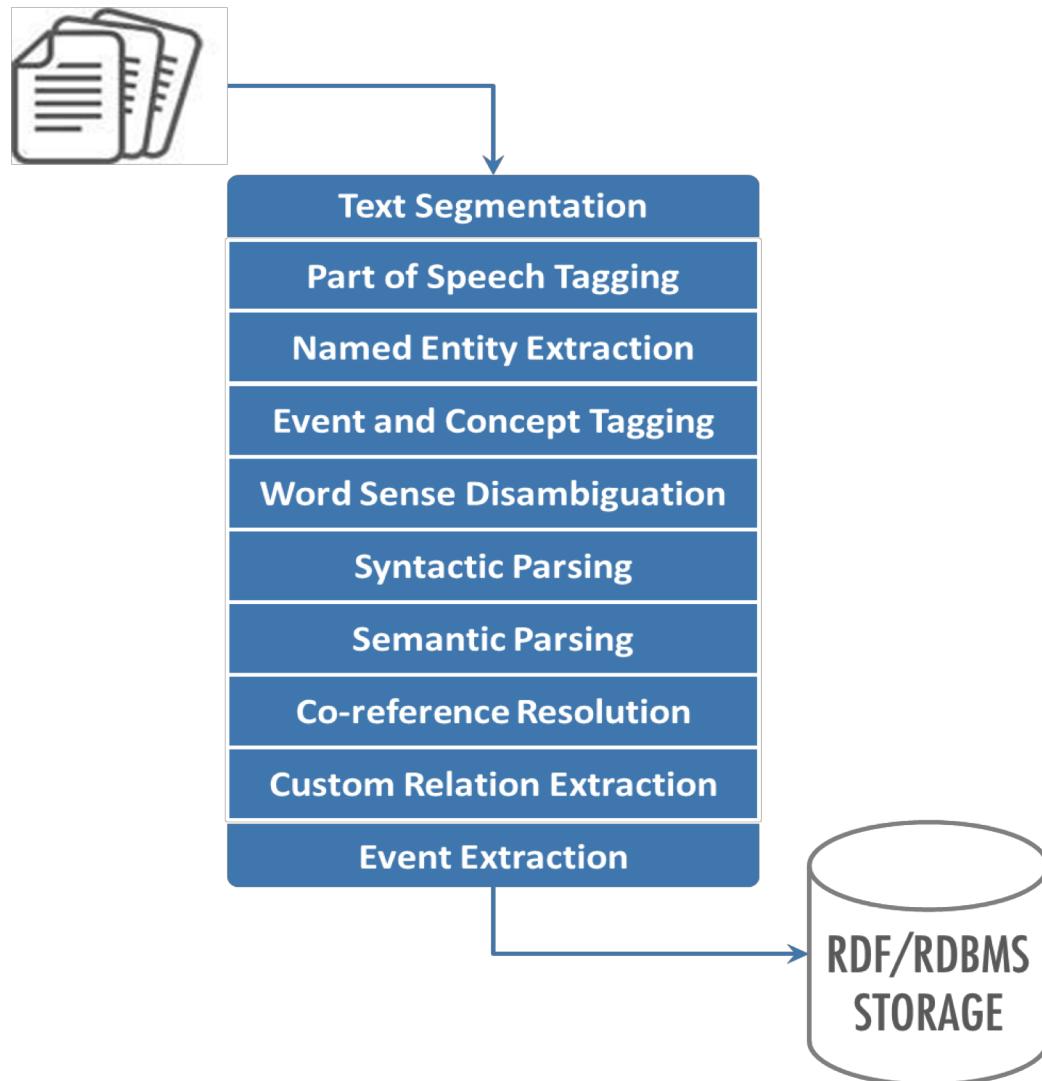
Xǐngxīn di huá Wrong?

Challenges – language is not static

- ❖ Language grows and changes
 - ❖ e.g., cyber lingo



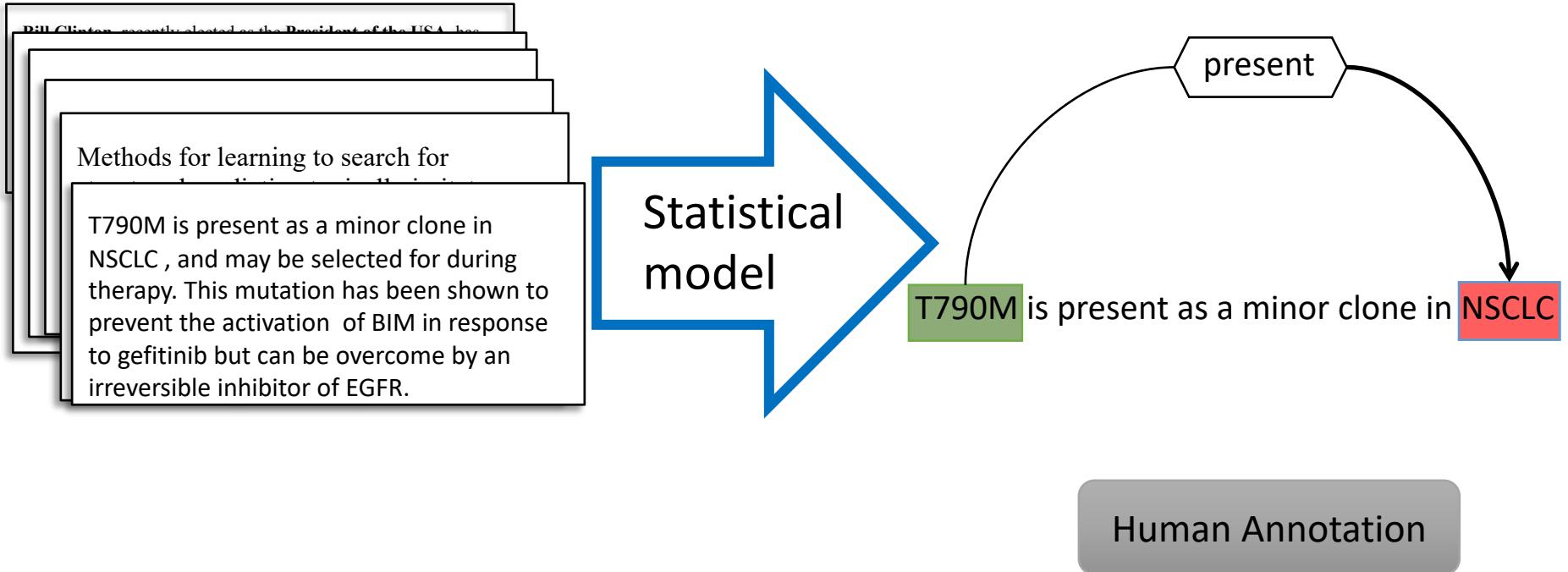
NLP pipeline – key components



Machine Learning

Input space

Label space



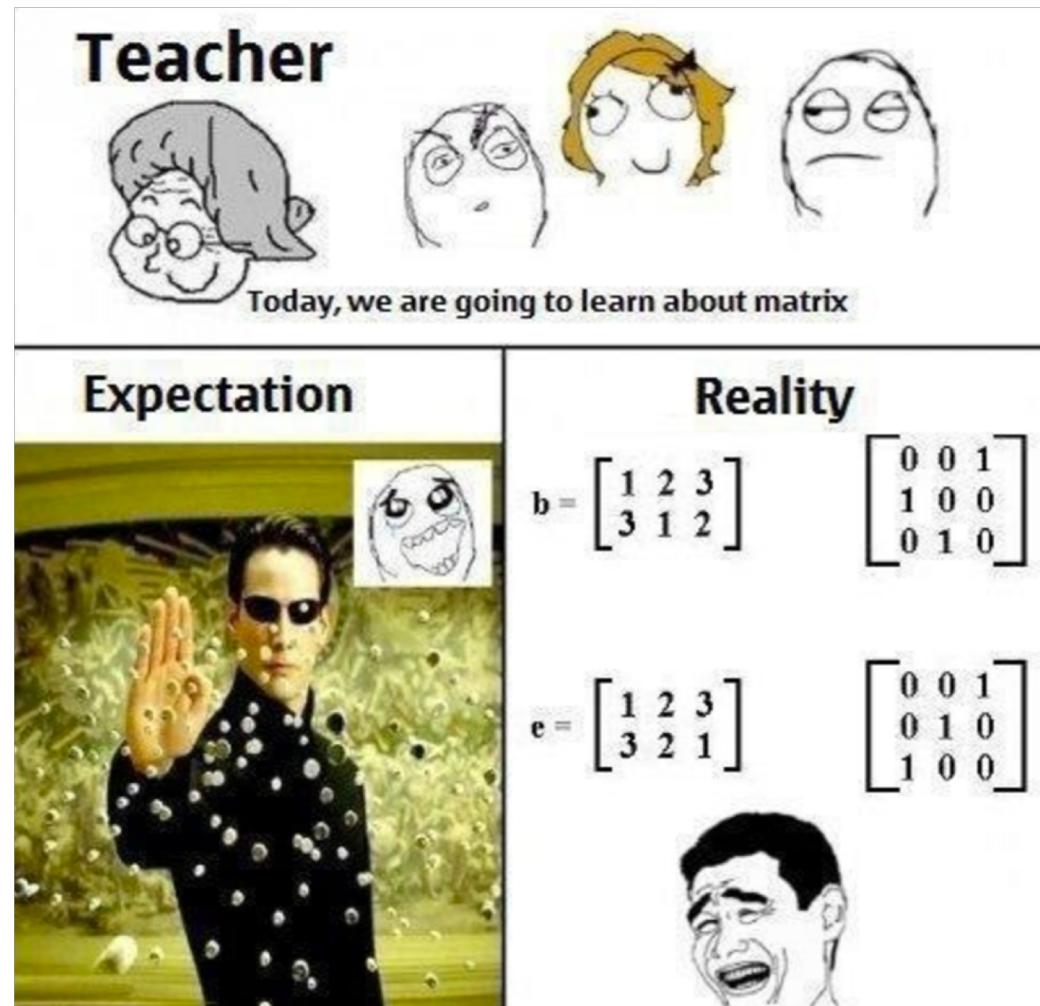
It's cool, let's work on it



[Cool Cat: Genius](#) by [ToValhalla](#)

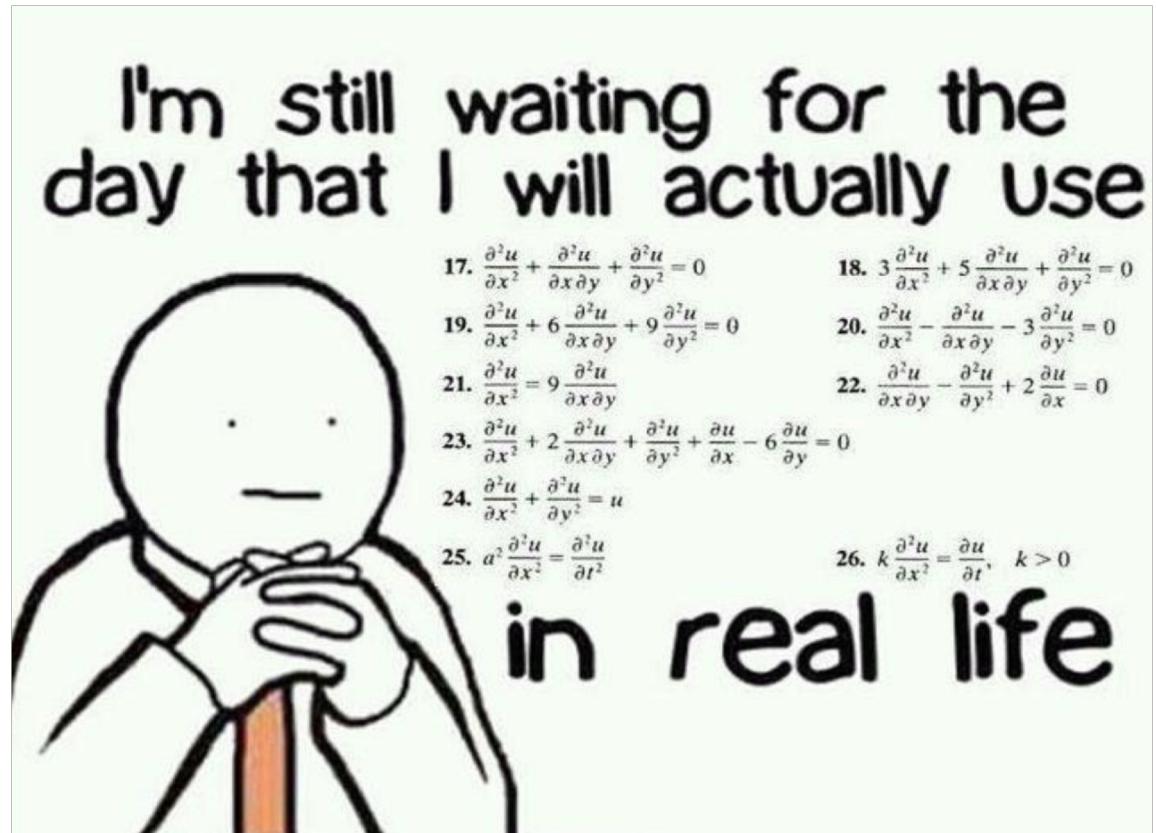
Prerequisites

- ❖ Math tools
- ❖ Linear algebra



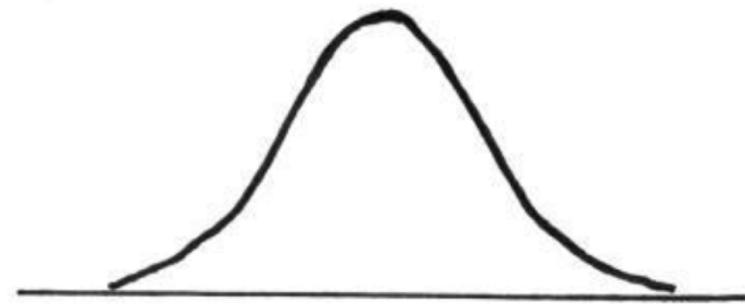
Prerequisites

- ❖ Math tools
 - ❖ Linear algebra
 - ❖ Calculus



Prerequisites

- ❖ Math tools
 - ❖ Linear algebra
 - ❖ Calculus
 - ❖ Probability



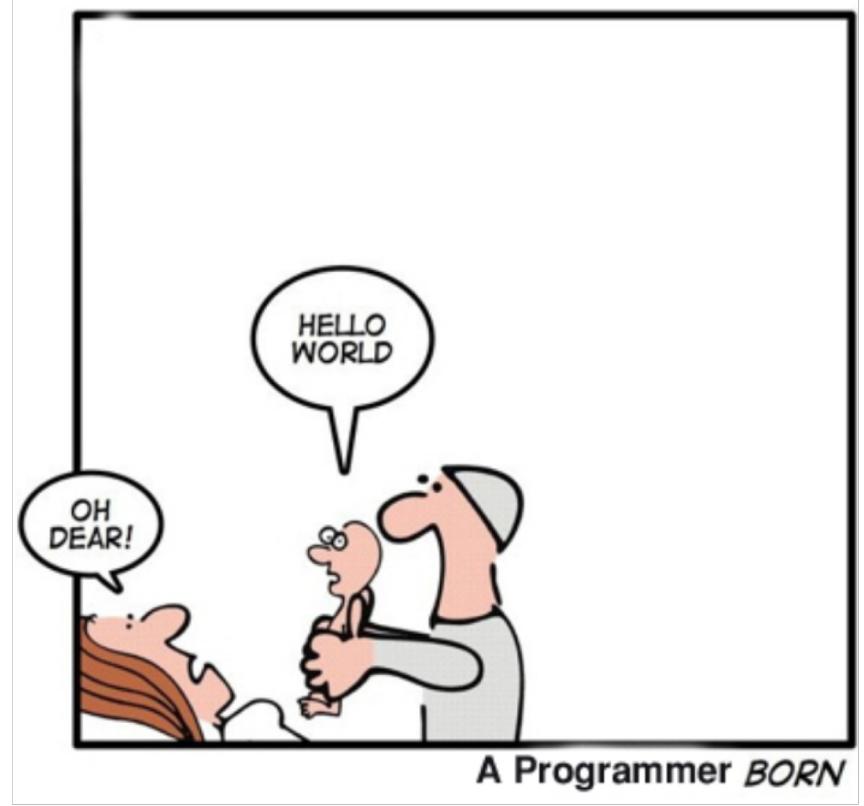
NORMAL DISTRIBUTION



PARANORMAL DISTRIBUTION

Prerequisites

- ❖ Math tools
 - ❖ Linear algebra
 - ❖ Calculus
 - ❖ Probability
- ❖ Programming skills
- ❖ Algorithm / Data structure



More information

- ❖ <http://kwchang.net>