

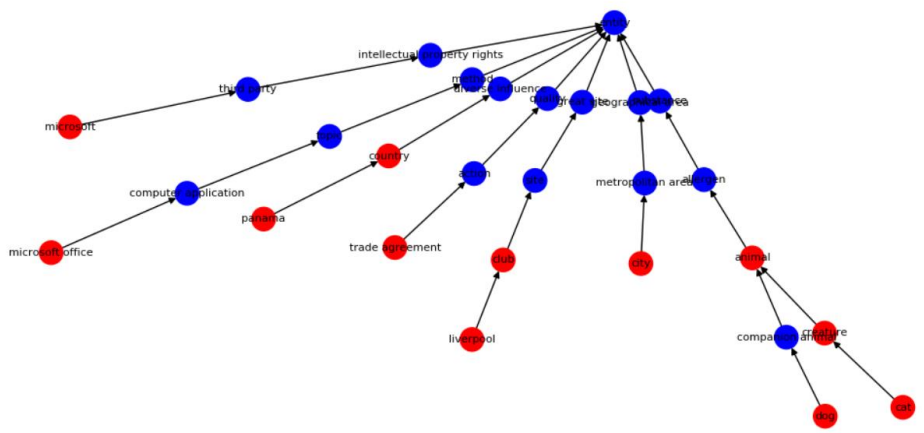
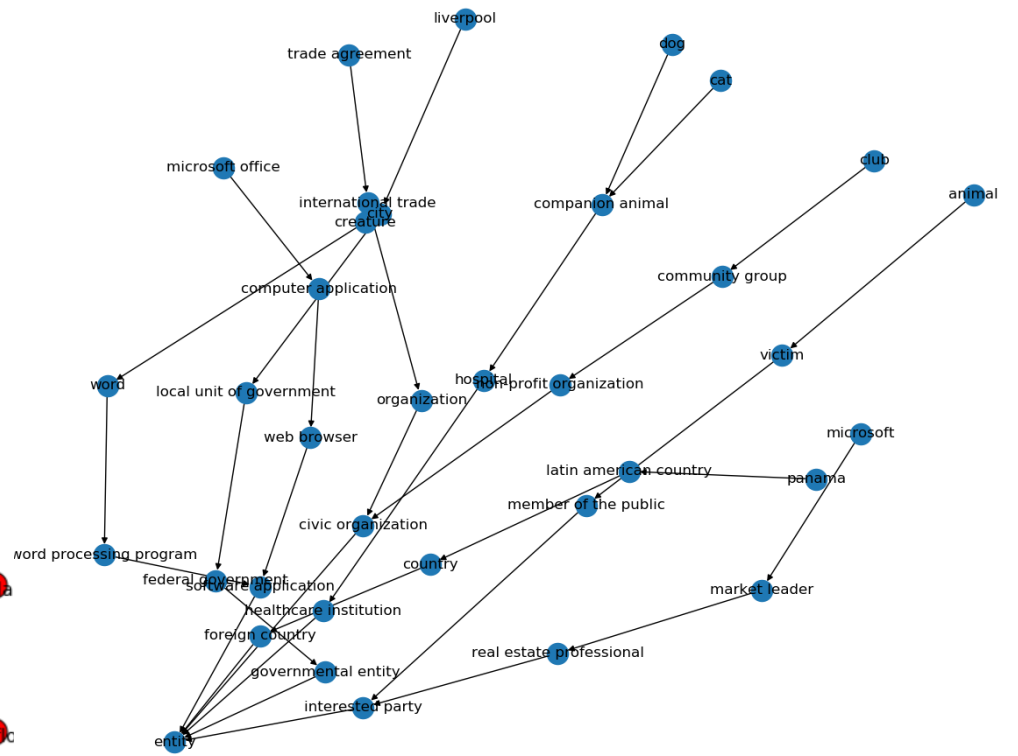
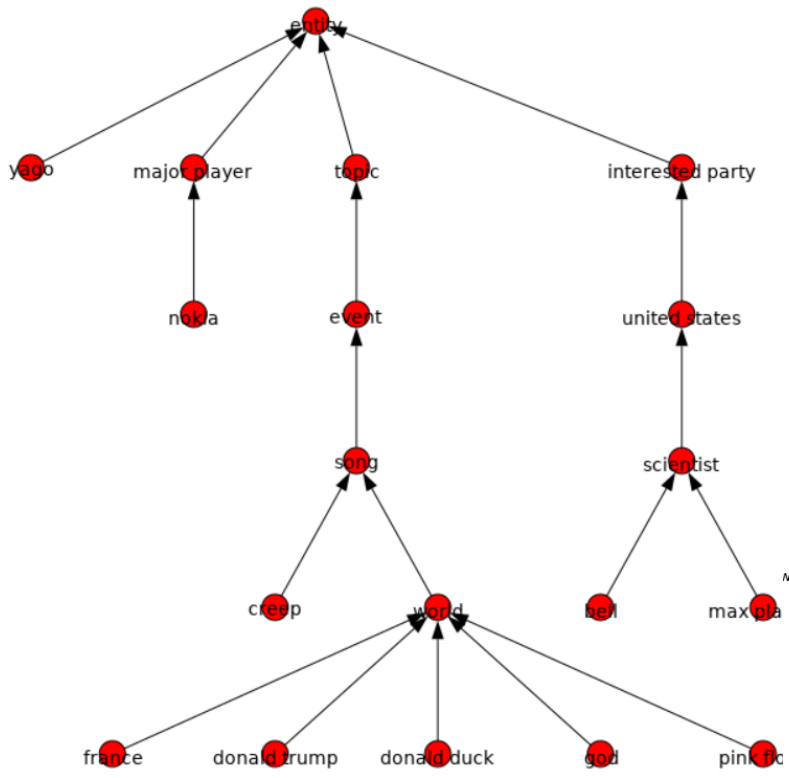
# Information extraction

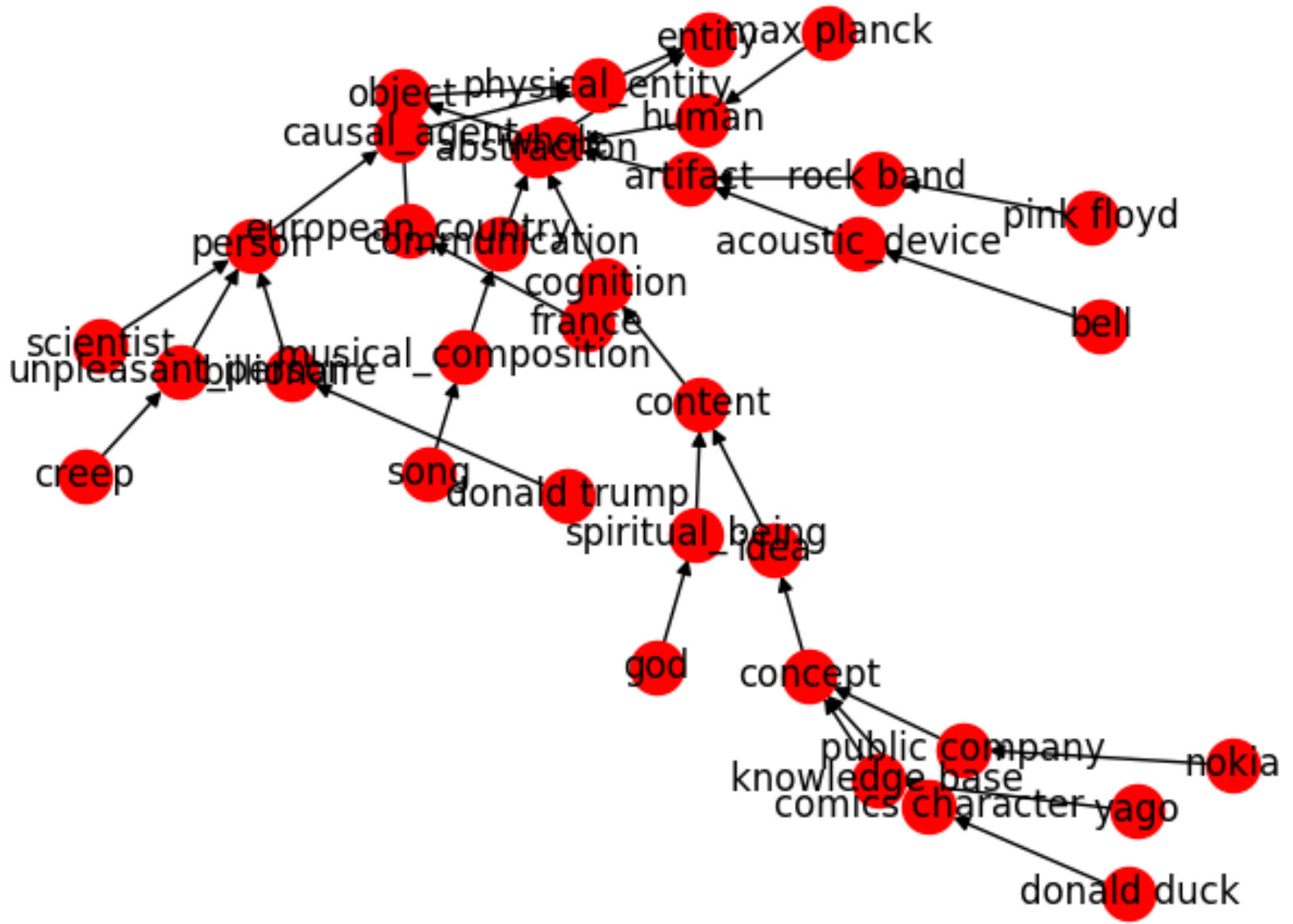
## 6.+7. Relation extraction

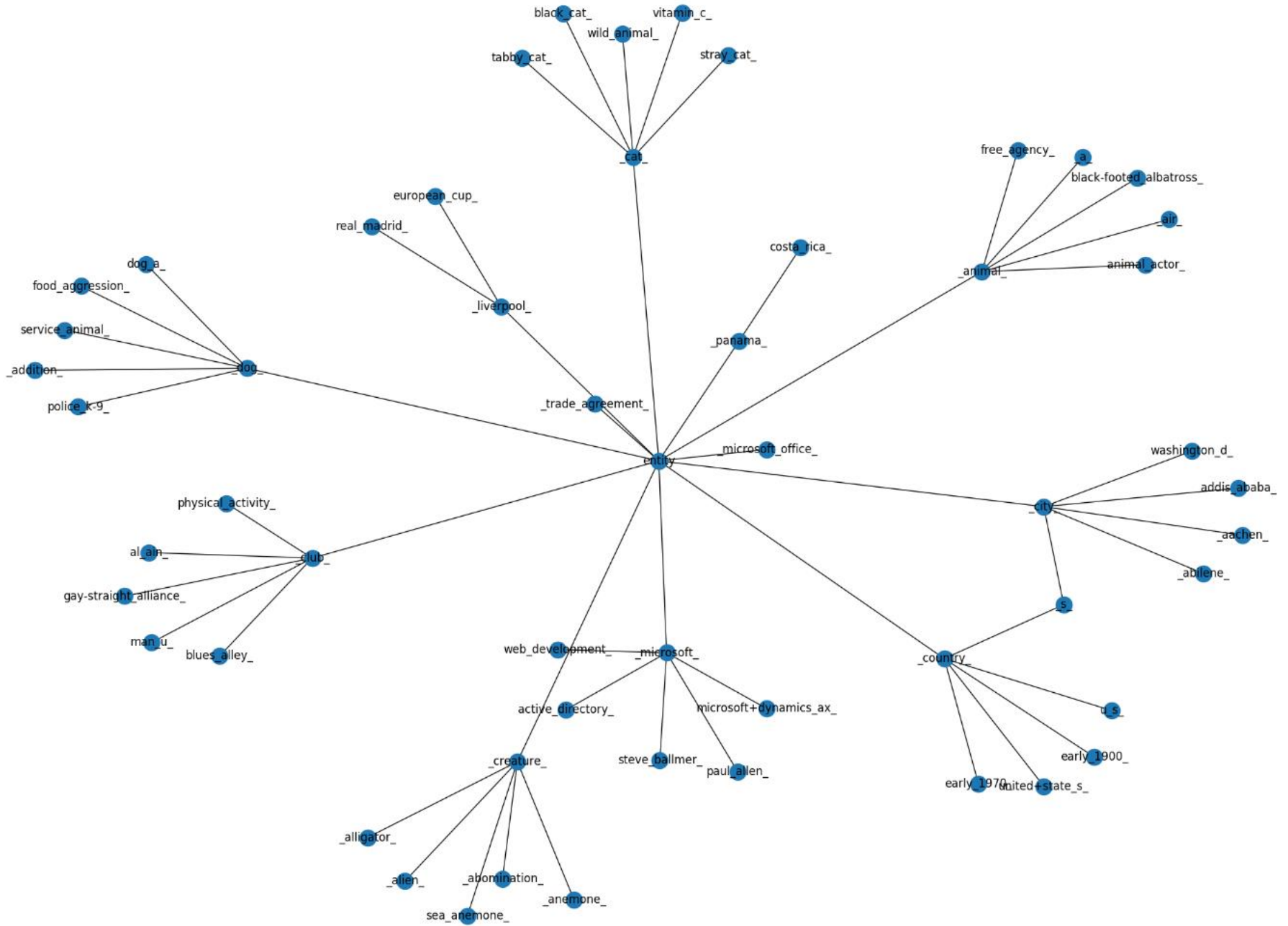
Simon Razniewski  
Winter semester 2019/20

# Announcements

- Neural coref: Score below mentions is score for having no antecedents
- Videlectures will be considered next year
  - For now slides, and literature pointers (ask me for more)
- Results assignment 5
  - Sample solutions next slides







# Outline

## 1. Problem

## 2. Manual patterns

## 3. Supervised learning

1. Feature-based
2. TACRED and BERT

## 4. Semi- and unsupervised extraction

1. Iterative pattern learning (DIPRE)
2. Distant supervision
  - CINEX

## 5. Evaluation

## 6. OpenIE

1. PATTY
2. Universal schema
3. Quasimodo

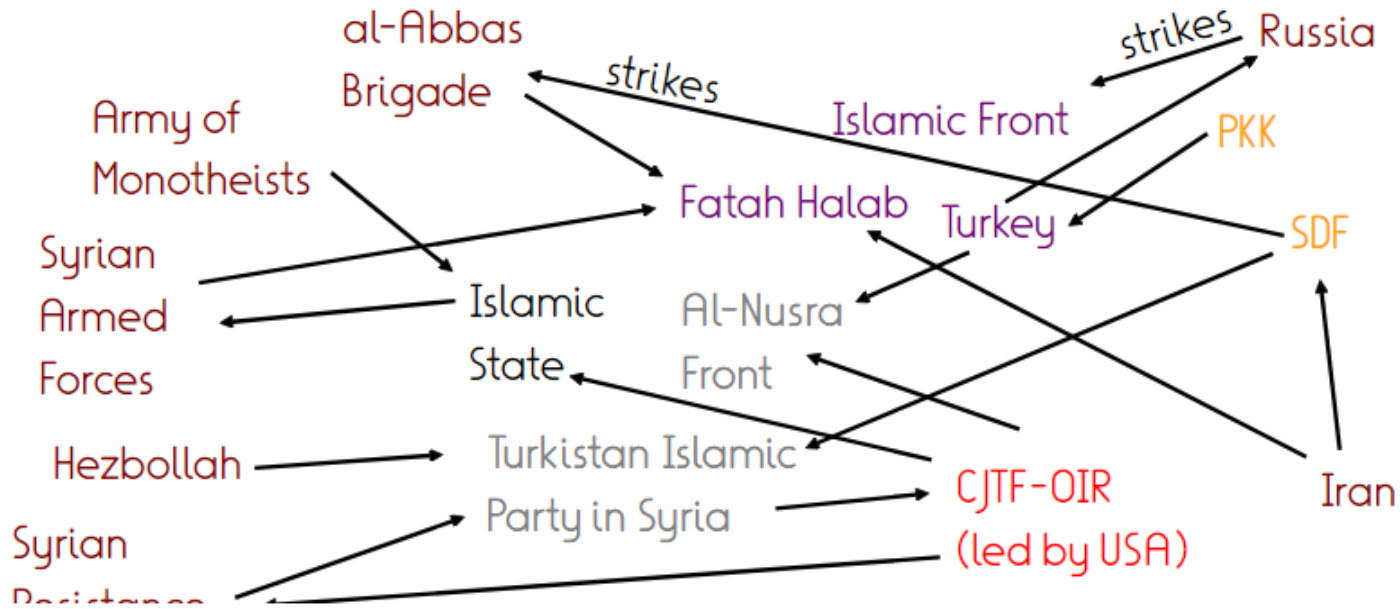
## 7. Negation

## 8. Extraction/prediction from latent representations

# Problem: Relation extraction

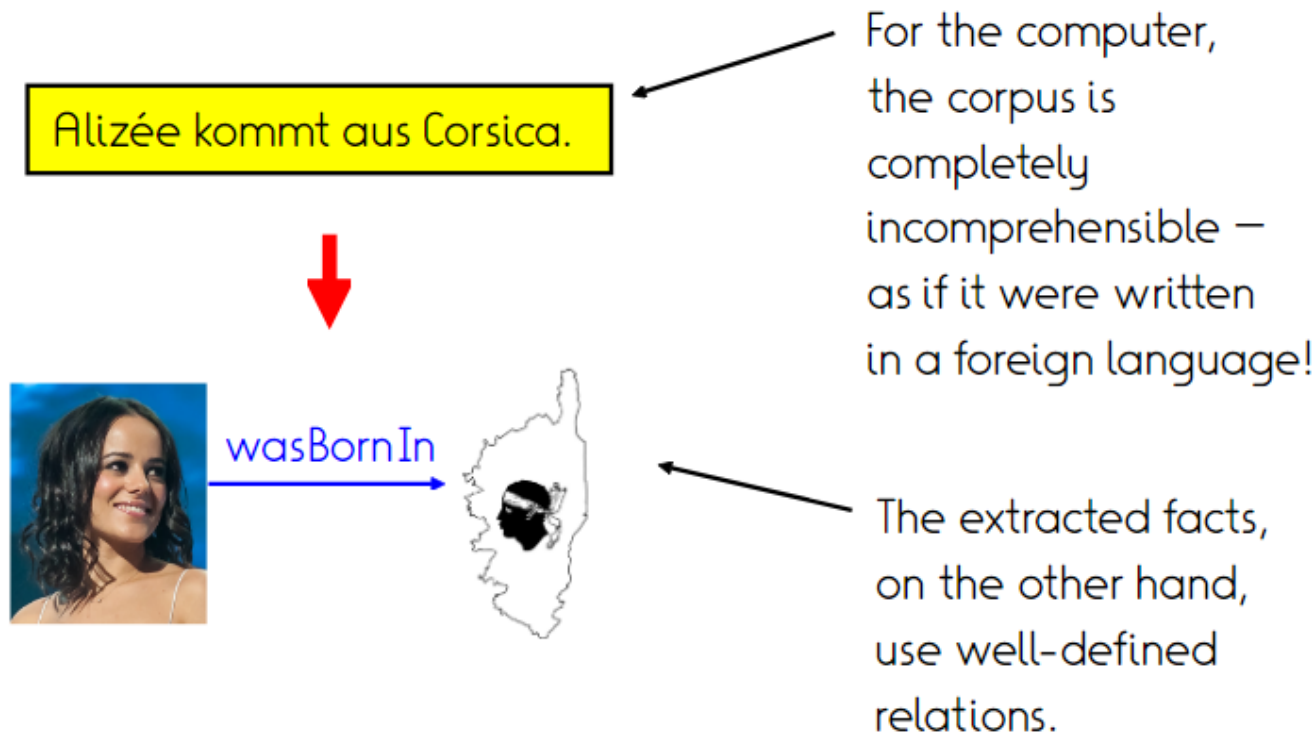
Fact extraction is the extraction of facts about entities from a corpus.

في اوائل نوفمبر، جرت اشتباكات بين الجيش السوري الحر و قوات الأمن في حمص مما ساهم في توسع الحاص  
قتال شوارع طويل في العديد من الأحياء. كانت المقاومة في حمص أكبر بكثير من البلدات و المدن الأخرى، و  
حمص، فقد قُتل العمليات في حمص حتى الآن في قمع الاضطرابات. في نوفمبر تشرين الثاني ديسمبر 2011.



# Relation extraction, happier example

Fact extraction is the extraction of facts about entities from a corpus.  
For now, we concentrate on facts with a single relation.





# Extracting relations from text

- Company report: "International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)..."
- Extracted Complex Relation:
  - companyFounding
    - company: IBM
    - location: New York
    - date: June 16, 1911
    - originalName: Computing-Tabulating-Recording Co.
- But we will focus on the simpler task of extracting relation triples

foundingYear(IBM, 1911)

foundingLocation(IBM, New York)

# Extracting Relation Triples from Text



The screenshot shows the Wikipedia article for Stanford University. The text is annotated with relation triples in various colors. A black arrow points from the text 'Leland founded the university in 1891.' to a list of extracted triples.

**Text from image:**

The Leland Stanford University, commonly referred to as Stanford, is an American private research university located in Stanford, California, near Palo Alto, California... Leland founded the university in 1891.

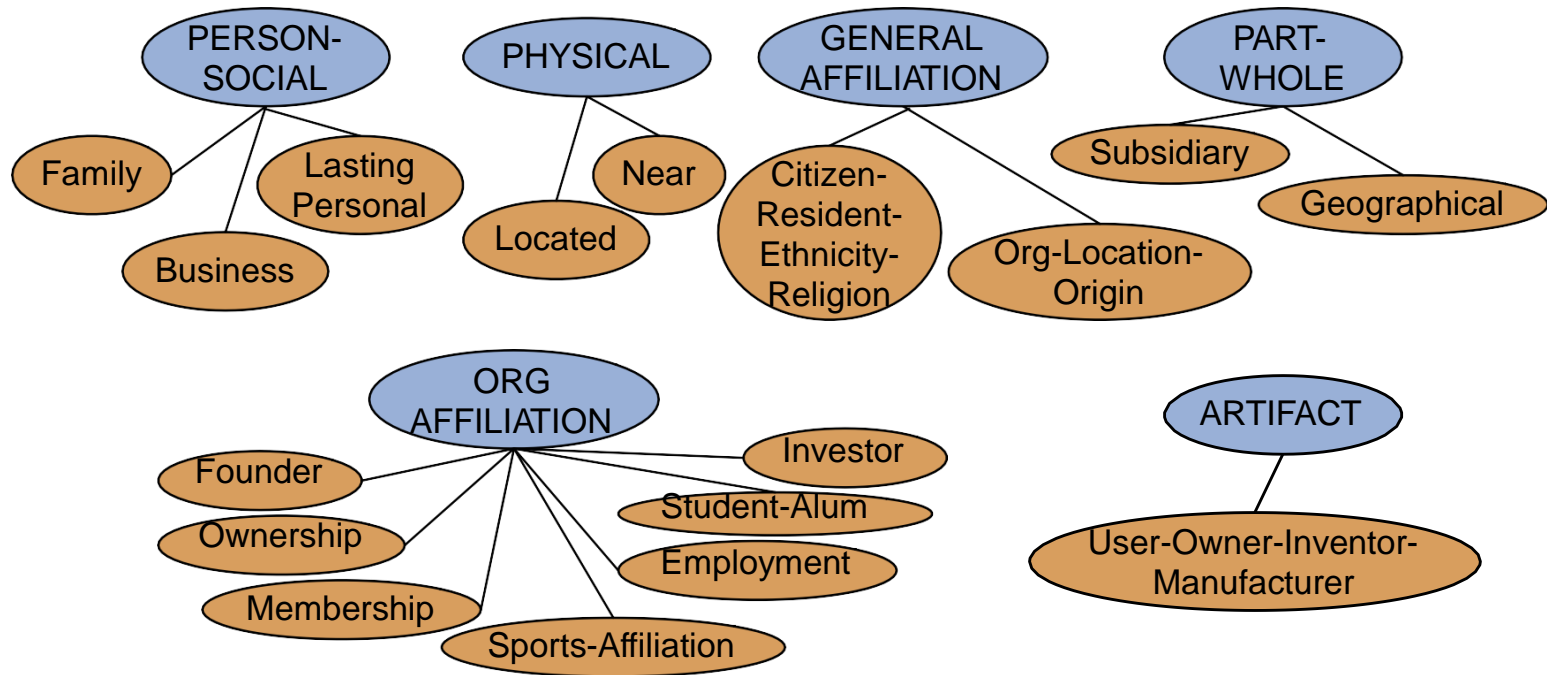
**Extracted Triples:**

- Stanford EQ Leland Stanford Junior University
- Stanford LOC-IN California
- Stanford IS-A research university
- Stanford LOC-NEAR Palo Alto
- Stanford FOUNDED-IN 1891
- Stanford FOUNDER Leland Stanford

Which relations should we extract?

# Automated Content Extraction (ACE)

17 relations from 2008 "Relation Extraction Task"



# Automated Content Extraction (ACE)

- Physical-Located            PER-GPE  
    He was in Tennessee
- Part-Whole-Subsidiary    ORG-ORG  
    XYZ, the parent company of ABC
- Person-Social-Family    PER-PER  
    John's wife Yoko
- Org-AFF-Founder            PER-ORG  
    Steve Jobs, co-founder of Apple...
-

# UMLS: Unified Medical Language System

- 134 entity types, 54 relations

Injury	disrupts	Physiological Function
Bodily Location	location-of	Biologic Function
Anatomical Structure	part-of	Organism
Pharmacologic Substance	causes	Pathological Function
Pharmacologic Substance	treats	Pathologic Function

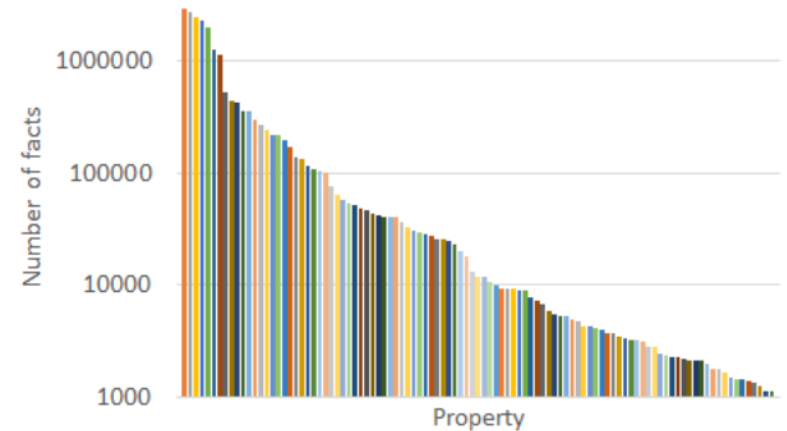
# Wikidata relations

> 5000 relations

Most frequent relations for humans:

- Gender (89%)
- Occupation (77%)
- Date of birth ( 69%)
- Given name (59%)
- Citizenship (58%)
- ...
- Languages spoke (13%)
- Position held (10%)
- ...

11/2019: 67 human properties used at least 100k times



# Ontological relations



Remember  
lecture 4

Examples from WordNet

- isA (hypernym): subsumption between classes
  - Giraffe isA ruminant isA ungulate isA mammal isA vertebrate isA animal...
- instanceOf: relation between individual and class
  - San Francisco instanceOf city
- Synonym: Same meaning
- Antonym: Opposite meaning
- Meronym: Part of another concept
- ...

# Outline

1. Problem

2. Manual patterns

3. Supervised learning

1. Feature-based
2. TACRED and BERT

4. Semi- and unsupervised extraction

1. Iterative pattern learning (DIPRE)
2. Distant supervision
  - CINEX

5. Evaluation

6. OpenIE

1. PATTY
2. Universal schema
3. Quasimodo

7. Negation

8. Extraction/prediction from latent representations



# Hearst Patterns++ for extracting relations

"such Y as X"

"X or other Y"

"X and other Y"

"Y including X"

"Y, especially X"



"X was born in Y"

"Born in Y, X"

...

# Extracting richer relations using rules and named entities

- Intuition: relations often hold between specific entities
  - located-in (ORGANIZATION, LOCATION)
  - founded (PERSON, ORGANIZATION)
  - cures (DRUG, DISEASE)
- Utilize NERC tags to help extract relation!

" $X_{\text{PERS}}$  ( $Y_{\text{LOC}}$ , DATE-)"

"Born in  $Y_{\text{LOC}}$ ,  $X_{\text{PERS}}$ "

...

# Extracting richer relations using rules and named entities

Who holds what office in what organization?

PERSON, POSITION of ORG

- George Marshall, Secretary of State of the United States

PERSON (named | appointed | chose | *etc.*) PERSON Prep? POSITION

- Truman appointed Marshall Secretary of State

PERSON [be]? (named | appointed | *etc.*) Prep? ORG POSITION

- George Marshall was named US Secretary of State

# Hand-built patterns for relations

- Plus
  - Human patterns tend to be high-precision
  - Can be tailored to specific domains
- Minus
  - Human patterns are often low-recall
  - A lot of work to think of all possible patterns!
  - Don't want to have to do this for every relation!

# Outline

1. Problem
2. Manual patterns
3. Supervised learning
  1. Feature-based
  2. TACRED and BERT
4. Semi- and unsupervised extraction
  1. Iterative pattern learning (DIPRE)
  2. Distant supervision
    - CINEX
5. Evaluation
6. OpenIE
  1. PATTY
  2. Universal schema
  3. Quasimodo
7. Negation
8. Extraction/prediction from latent representations

# Supervised ML for relation extraction

- Choose a set of relations we'd like to extract
- Choose a set of relevant named entities
- Find and label data
  - Choose a representative corpus
  - Label the named entities in the corpus
  - Hand-label the relations between these entities
  - Break into training, development, and test
- Train a classifier on the training set

# Relation Extraction

Classify the relation between two entities

*American Airlines*, a unit of AMR, immediately matched the move, spokesman *Tim Wagner* said.



# Word Features for Relation Extraction

(Remember lec 5 on coreference)

*American Airlines*, a unit of AMR, immediately matched the move, spokesman *Tim Wagner* said  
Mention 1 Mention 2

- Headwords of M1 and M2  
*Airlines* *Wagner*
- Bag of words and bigrams in M1 and M2  
{*American, Airlines, Tim, Wagner, American Airlines, Tim Wagner*}
- Words or bigrams in particular positions left and right of M1/M2  
*M2: -1 spokesman*  
*M2: +1 said*
- Bag of words or bigrams between the two entities  
{*a, AMR, of, immediately, matched, move, spokesman, the, unit*}



# Named Entity Type and Mention Level Features for Relation Extraction

*American Airlines*, a unit of AMR, immediately matched the move, spokesman *Tim Wagner* said  
Mention 1 Mention 2

- Named-entity types
  - M1: **ORG**
  - M2: **PERSON**
- Concatenation of the two named-entity types
  - **ORG-PERSON**
- Entity Level of M1 and M2 (NAME, NOMINAL, PRONOUN)
  - M1: **NAME** [it or he would be **PRONOUN**]
  - M2: **NAME** [the company would be **NOMINAL**]

# Parse Features for Relation Extraction

*American Airlines*, a unit of AMR, immediately matched the move, spokesman *Tim Wagner* said

Mention 1

Mention 2

- Base syntactic chunk sequence from one to the other  
NP NP PP VP NP NP
- Constituent path through the tree from one to the other  
NP ↑↑ NP ↑↑ S ↑↑ S ↓↓ NP
- Dependency path  
Airlines matched Wagner said

# Gazetteer and trigger word features for relation extraction

- Trigger list for family: kinship terms
  - parent, wife, husband, grandparent, etc.
- Gazetteer:
  - Lists of useful geo or geopolitical words
    - Country name list
    - Other sub-entities

*American Airlines, a unit of AMR, immediately matched the move, spokesman Tim Wagner said.*

### Entity-based features

Entity <sub>1</sub> type	ORG
Entity <sub>1</sub> head	<i>airlines</i>
Entity <sub>2</sub> type	PERS
Entity <sub>2</sub> head	<i>Wagner</i>
Concatenated types	ORGPERS

### Word-based features

Between-entity bag of words	{ <i>a, unit, of, AMR, Inc., immediately, matched, the, move, spokesman</i> }
Word(s) before Entity <sub>1</sub>	NONE
Word(s) after Entity <sub>2</sub>	<i>said</i>

### Syntactic features

Constituent path	<i>NP ↑ NP ↑ S ↑ S ↓ NP</i>
Base syntactic chunk path	<i>NP → NP → PP → NP → VP → NP → NP</i>
Typed-dependency path	<i>Airlines ←<sub>subj</sub> matched ←<sub>comp</sub> said →<sub>subj</sub> Wagner</i>

# Evaluation of Supervised Relation Extraction

- Now you can use any standard supervised classifier
- Evaluate on withheld annotated data (more later)

# Summary: Supervised Relation Extraction

- + Can get high precision/recall with enough training data, if test similar enough to training
- Labeling a large training set is expensive
- Supervised models are still brittle, don't generalize well to different genres

# Outline

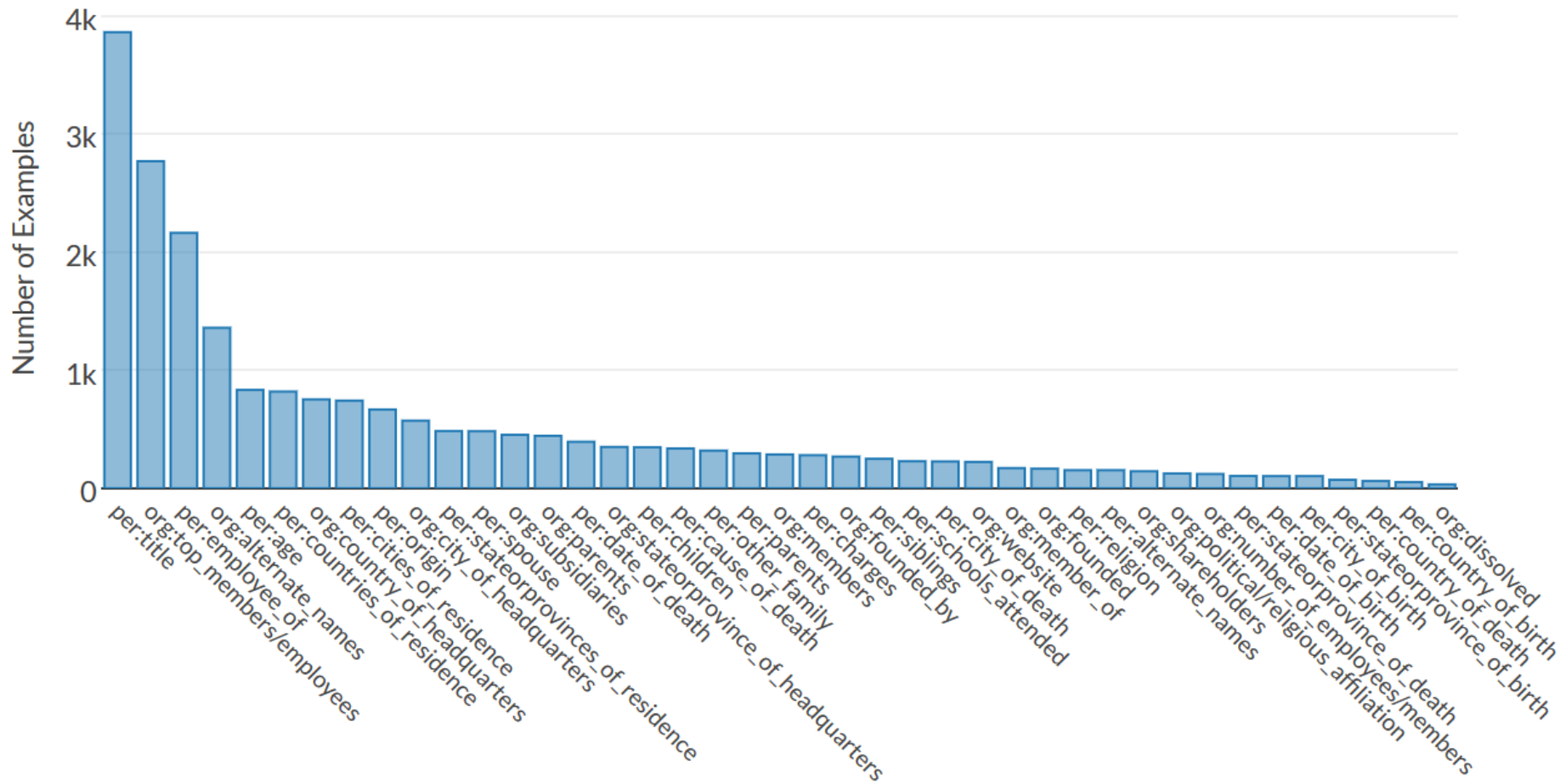
1. Problem
2. Manual patterns
3. Supervised learning
  1. Feature-based
  2. TACRED and BERT
4. Semi- and unsupervised extraction
  1. Iterative pattern learning (DIPRE)
  2. Distant supervision
    - CINEX
5. Evaluation
6. OpenIE
  1. PATTY
  2. Universal schema
  3. Quasimodo
7. Negation
8. Extraction/prediction from latent representations

# TACRED [Zhang et al., EMNLP 2017]

- TAC: Text analysis conference, at national institute for standards (NIST), USA
- Annual competitions around information extraction, retrieval, question answering, etc.
- <https://tac.nist.gov/>
- TACRED:
  - Relation extraction dataset, competition since 2014
  - 106,264 human-labelled entity pairs in a sentence sampled from newswire and web forum discussions
  - 41 common relation types
  - 23 entity types
  - *no\_relation* if no defined relation holds



# TACRED (2)

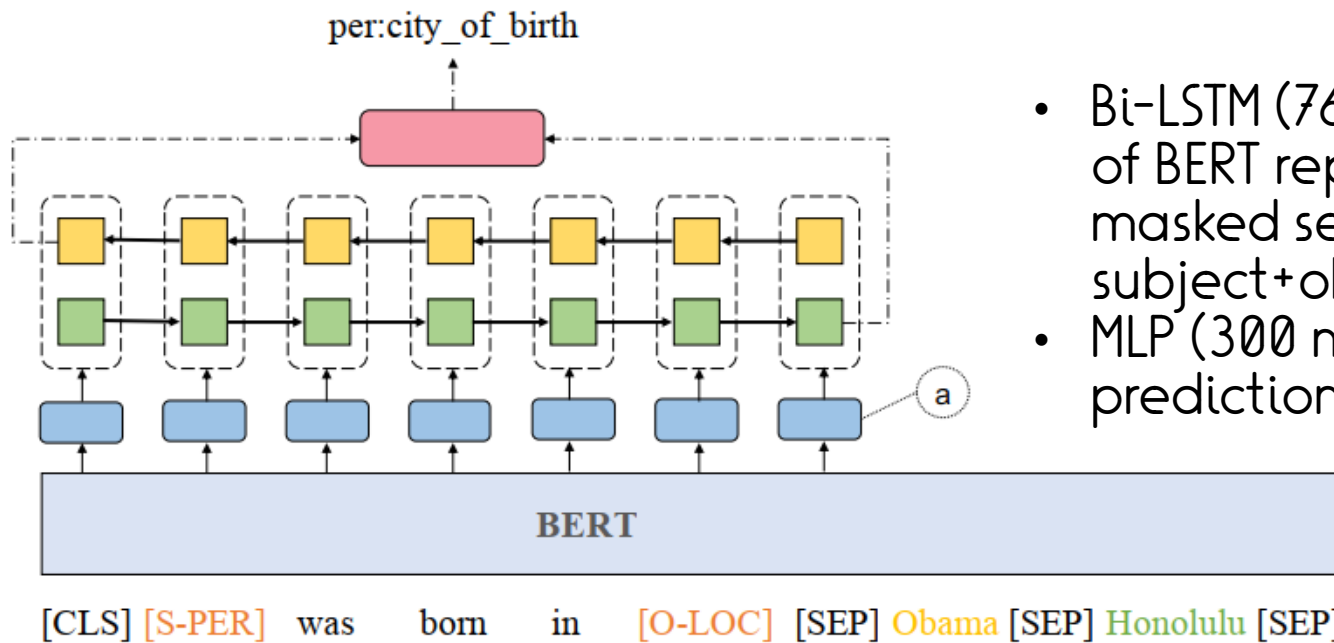


# TACRED (3)

	Model	P	R	F1
Traditional	Patterns	<b>86.9</b>	23.2	36.6
	Logistic Regression (LR)	73.5	49.9	59.4
	LR + Patterns	72.9	51.8	60.5
Neural	CNN	75.6	47.5	58.3
	LSTM	65.7	59.9	62.7
	LSTM + Position-aware attention	65.7	<b>64.5</b>	<b>65.1</b>

[TACRED website]

# Relation extraction using BERT



- Bi-LSTM (768 nodes) on top of BERT representation of masked sentence + subject + object
- MLP (300 nodes) for final prediction

Model	P	R	F <sub>1</sub>
Zhang et al. (2017)	65.7	64.5	65.1
Zhang et al. (2018)	69.9	63.33	66.4
Wu et al. (2019)	-	-	67.0
Alt et al. (2019)	70.1	65.0	67.4
BERT-LSTM-base	<b>73.3</b>	63.10	67.8
Zhang et al. (2018) (ensemble)	71.3	<b>65.4</b>	<b>68.2</b>

[Simple BERT Models for Relation Extraction and Semantic Role Labeling, Peng Shi and Jimmy Lin, ArXiv, 2019]

# Outline

1. Problem
2. Manual patterns
3. Supervised learning
  1. Feature-based
  2. TACRED and BERT
4. Semi- and unsupervised extraction
  1. Iterative pattern learning (DIPRE)
  2. Distant supervision
    - CINEX
5. Evaluation
6. OpenIE
  1. PATTY
  2. Universal schema
  3. Quasimodo
7. Negation
8. Extraction/prediction from latent representations

# Seed-based or bootstrapping approaches to relation extraction

- No training set? Maybe you have:
  - A few **seed tuples** or
  - A few **high-precision patterns**
- Can you use those seeds to do something useful?
  - **Bootstrapping**: use the seeds to directly learn to populate a relation

# Relation Bootstrapping (Hearst 1992)

- Gather a set of seed pairs that have relation R
- Iterate:
  1. Find sentences with these pairs
  2. Look at the context between or around the pair and generalize the context to create patterns
  3. Use the patterns for grep for more pairs

# Bootstrapping/Pattern iteration

- buriedIn(Mark Twain, Elmira) - Seed tuple
  - Grep (google) for the environments of the seed tuple

"Mark Twain is buried in Elmira, NY."

X is buried in Y

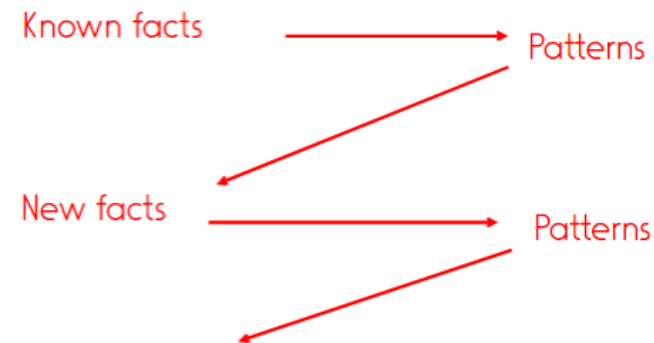
"The grave of Mark Twain is in Elmira"

The grave of X is in Y

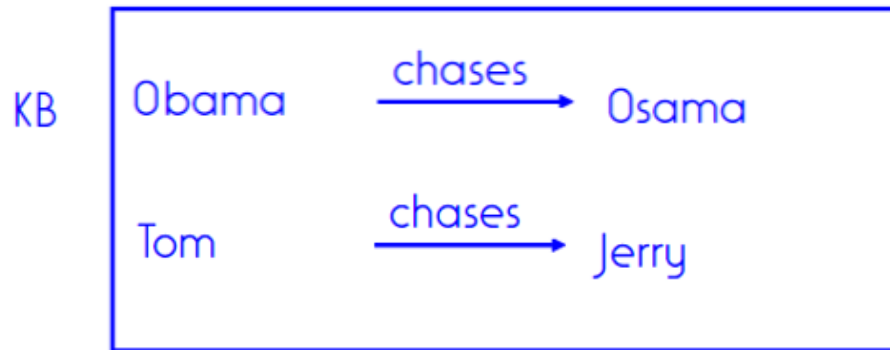
"Elmira is Mark Twain's final resting place"

Y is X's final resting place.

- Use those patterns to grep for new tuples
- Iterate



# Example: Pattern iteration



Obama hetzt Osama. Tom jagt Jerry. Tom hetzt Jerry.

=> "X hetzt Y" is a pattern for chases(X, Y)

=> "X jagt Y" is a pattern for chases(X, Y)



# Task: Pattern iteration



Michelle ist verheiratet mit Barack.  
Merkel ist die Frau von Sauer.  
Michelle ist die Frau von Barack.  
Priscilla ist verheiratet mit Elvis.

# DIPRE: Extracting <author,book> pairs

- Dual iterative pattern relation extraction

Brin, Sergei. 1998. Extracting Patterns and Relations from the World Wide Web.

- Start with 5 seeds:

Author	Book
Isaac Asimov	The Robots of Dawn
David Brin	Startide Rising
James Gleick	Chaos: Making a New Science
Charles Dickens	Great Expectations
William Shakespeare	The Comedy of Errors

- Find Instances:

The Comedy of Errors, by William Shakespeare, was

The Comedy of Errors, by William Shakespeare, is

The Comedy of Errors, one of William Shakespeare's earliest attempts

The Comedy of Errors, one of William Shakespeare's most

- Extract patterns (group by middle, take longest common prefix/suffix)

?x , by ?y ,

?x , one of ?y 's

- Now iterate, finding new seeds that match the pattern

# DIPRE

- 5 seeds
- 199 occurrences

- 3 patterns

URL Pattern	Text Pattern
<code>www.sff.net/locus/c.*</code>	<code>&lt;LI&gt;&lt;B&gt;title&lt;/B&gt; by author (</code>
<code>dns.city-net.com/Imann/awards/hugos/1984.html</code>	<code>&lt;i&gt;title&lt;/i&gt; by author (</code>
<code>dolphin.upenn.edu/dcummings/texts/sf-award.htm</code>	<code>author    title    (</code>

→ 4047 pairs

- 3972 occurrences in first 5 million websites
- 25 patterns

→ 9369 pairs

- 9938 occurrences in documents containing "book" term
- 346 patterns
- 15k pairs
  - Starting from 5!
  - Precision 95% (n=20..)

# Snowball

E. Agichtein and L. Gravano 2000. Snowball: Extracting Relations from Large Plain-Text Collections. ICDL

- Similar iterative algorithm

Organization	Location of Headquarters
Microsoft	Redmond
Exxon	Irving
IBM	Armonk

- Group instances w/similar prefix, middle, suffix, extract patterns
  - But require that X and Y be named entities
  - And compute a confidence for each pattern

.69    **ORGANIZATION**    {'s, in, headquarters}    **LOCATION**

.75    **LOCATION**    {in, based}    **ORGANIZATION**

# Example: Patterns in NELL

NELL (Never Ending Language Learner) is an information extraction project at Carnegie Mellon University.

Apple  $\xrightarrow{\text{produced}}$  MacBook

- CPL @851 (100.0%) on 28-jun-2014 [ "arg1 claims the new arg2" "arg1 were to release arg2" "arg2 are trademarks of arg1" "arg1 Store to get arg2" "arg1 AppleCare Protection Plan for arg2" "arg1 will announce a new arg2" "arg1 would release a new arg2" "arg2 Pro now includes arg1" "arg2 nano at arg1" "arg1 will release a new arg2" "arg1 announced their new arg2" "arg1 releases a new version of arg2" "arg1 already sells arg2" "arg1 announced that the new arg2" "arg1 recently switched their arg2" "arg2 and iPod are trademarks of arg1" "arg1 TV and arg2" "arg2 Pro from arg1" "arg1 says the new arg2" "arg1 unveils new arg2" "arg1 iMac and arg2" "arg1 has now released arg2" ] using (apple, macbook)

# Outline

1. Problem
2. Manual patterns
3. Supervised learning
  1. Feature-based
  2. TACRED and BERT
4. Semi- and unsupervised extraction
  1. Iterative pattern learning (DIPRE)
  2. Distant supervision
    - CINEX
5. Evaluation
6. OpenIE
  1. PATTY
  2. Universal schema
  3. Quasimodo
7. Negation
8. Extraction/prediction from latent representations

# Distant Supervision

- Combine bootstrapping with supervised learning
  - Instead of 5 seeds,
    - Use a large database to get huge # of seed examples
  - Create lots of features from all these examples
  - Combine in a supervised classifier

# Distantly supervised learning of relation extraction patterns

- 1 For each relation
- 2 For each tuple in a KB
- 3 Find sentences in large corpus with both entities
- 4 Extract frequent features (parse, words, etc)
- 5 Train supervised classifier using thousands of instances  
(negatives random entity pairs not in relation)

## Born-In

<Edwin Hubble, Marshfield>  
<Albert Einstein, Ulm>

*Hubble was born in Marshfield*  
*Einstein, born (1879), Ulm*  
*Hubble's birthplace in Marshfield*

PER was born in LOC  
PER, born (XXXX), LOC  
PER's birthplace in LOC

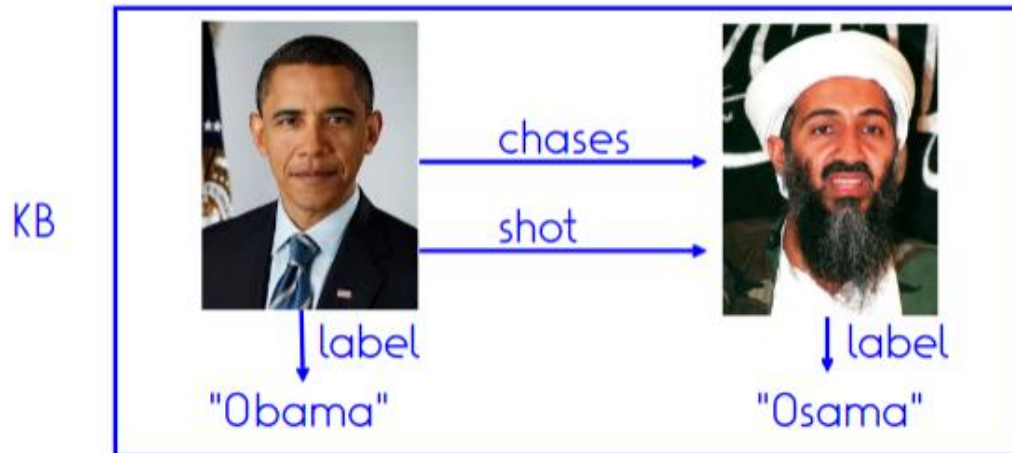
$P(\text{born-in} \mid f_1, f_2, f_3, \dots, f_{70000})$



# Distant supervision paradigm

- Like supervised classification:
  - Uses a classifier with lots of features
  - Supervised by detailed hand-created knowledge
  - Doesn't require iteratively expanding patterns
- Like unsupervised pattern iteration:
  - Uses very large amounts of unlabeled data

# Challenge 1: Overlapping relations



Corpus

Obama verfolgt Osama.

=> "X verfolgt Y" is a pattern for  $\text{chases}(X,Y)$  for  $\text{shot}(X,Y)$ ?

# Challenge 2: Irrelevant contexts

- capitalOf(Paris, France)
- *Paris is the capital of France.*
- *French authorities tightened security measures after the Paris attacks.*
- *Paris is a popular tourist destination in France.*

→ May lead to learning of **wrong patterns**

→ May lead to not extracting relations if few relevant contexts are **overshadowed** by many irrelevant ones

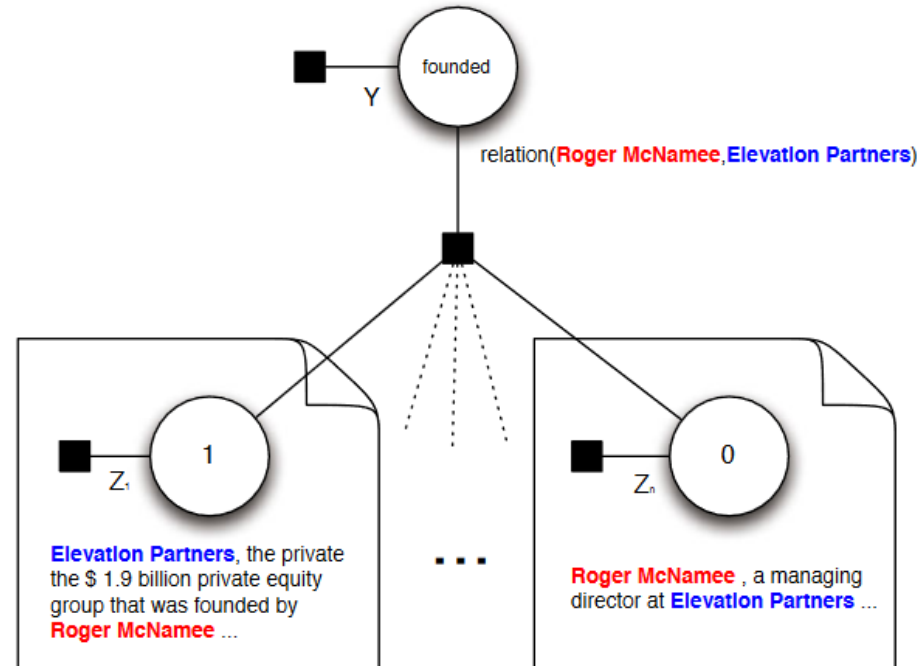
**Table 1.** Percentage of times a related pair of entities is mentioned in the same sentence, but where the sentence does not express the corresponding relation

Relation Type	New York Times	Wikipedia
nationality	38%	20%
place_of_birth	35%	20%
contains	20%	10%

# Fixing the naive assumption

→ At-least-one assumption  
[Riedel et al., 2010]

- "If two entities participate in a relation, at least one sentence that mentions these two entities might express that relation."
- Probabilistic model that simultaneously estimates whether relations hold, and which sentences express them.
  - Binary variables for contexts per entity pair
  - Contexts grouped for relation prediction
- Precision jumps from 87% to 91% (=31% reduction in error)



# Outline

1. Problem
2. Manual patterns
3. Supervised learning
  1. Feature-based
  2. TACRED and BERT
4. Semi- and unsupervised extraction
  1. Iterative pattern learning (DIPRE)
  2. Distant supervision
    - CINEX
5. Evaluation
6. OpenIE
  1. PATTY
  2. Universal schema
  3. Quasimodo
7. Negation
8. Extraction/prediction from latent representations

# CINEX [Mirza et al., 2018]

- Instructive example of (doubly) distant supervision
- Common twin of Wikipedia, Wikidata
- Focused on relation between entities and quantity expressions (counting quantifiers)

# Counting Quantifiers (CQs)

- Fully qualified facts:  $\langle S, P, O \rangle$

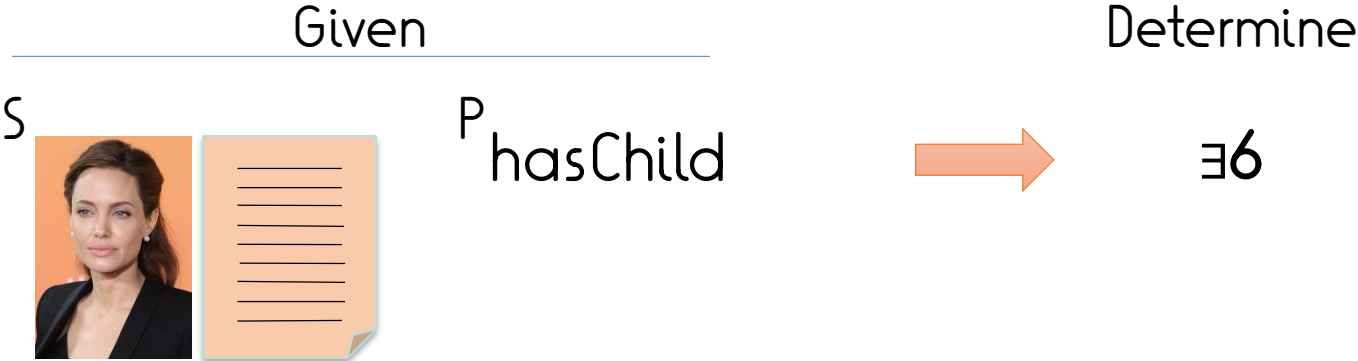
$\langle \text{California, hasCounty, Monterey} \rangle$        $\langle \text{Donald Trump, hasSpouse, Melania Knauss} \rangle$

- Counting information:  $\langle S, P, \exists O \rangle$

$\langle \text{California, hasCounty, } \exists 58 \rangle$        $\langle \text{Donald Trump, hasSpouse, } \exists 3 \rangle$

“There exists a specific number of  $O$  for a given  $SP$  pair”

# Problem: CQ Extraction





# Problem hardness

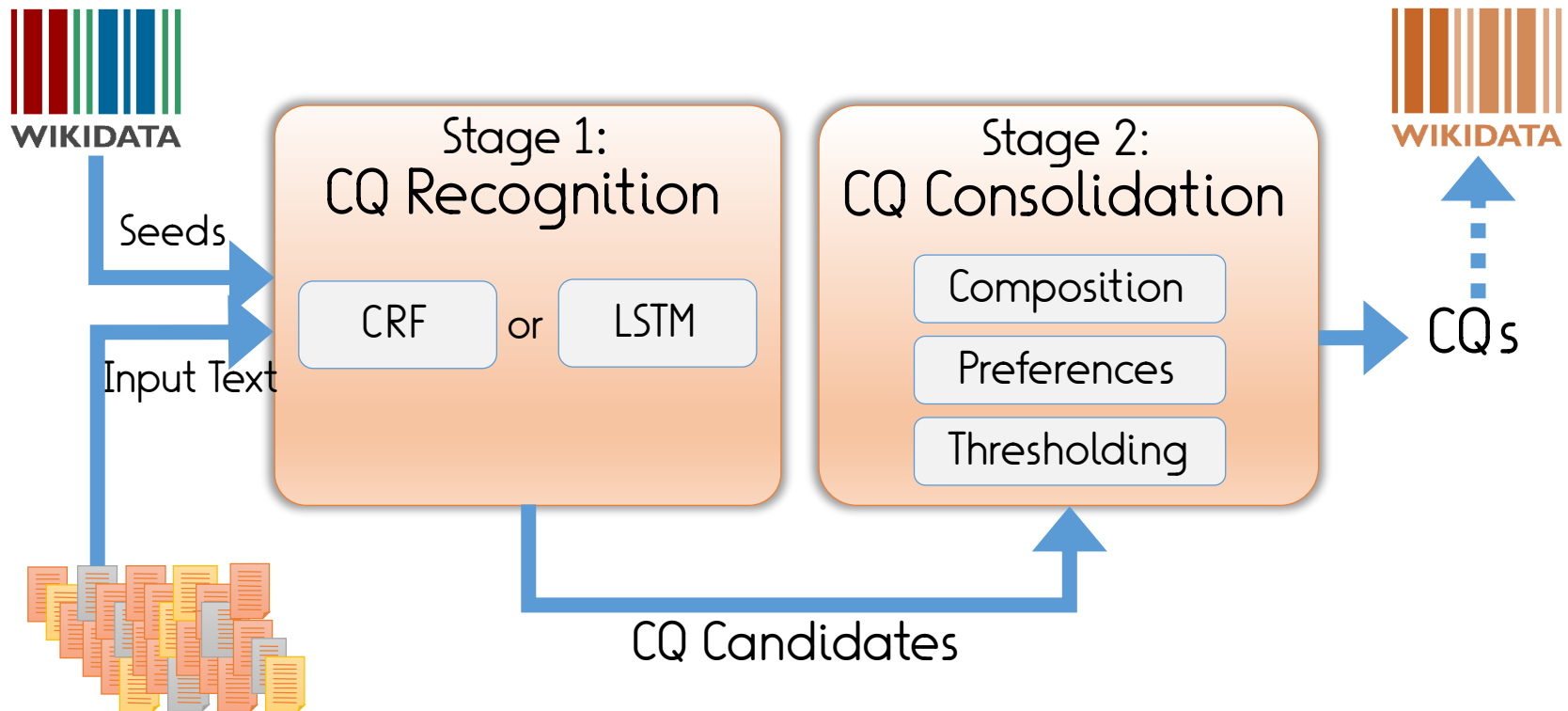
- Various expressions

1. Explicit numerals (cardinal numbers) "has five children"
2. Lower bounds (ordinal numbers) "his third wife"
3. Number-related noun phrases 'twins' or 'quartet'
4. Existence-proving articles "has a brother"
5. Non-existence adverbs 'never' or 'without'

- Compositionality

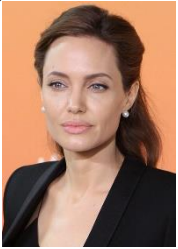
- In 2016, Jolie brought her twins, one daughter and three adopted children to the gala.

# CINEX: Counting Information Extraction



# Stage 1: CQ Recognition

S



P<sup>hasChild</sup>

cardinals

ordinals

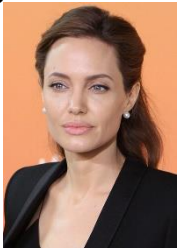
numterms

articles

1. She has a grand total of **six** children together: **three** biological and **three** adopted.
2. Angelina Jolie and **four** of her kids soaked up the last few days of summer over Labor Day.
3. She has received an Academy Award, two Screen Actors Guild Awards, and three Golden Globe Awards, and has been cited as Hollywood's highest-paid actress.
4. Divorced from actors Jonny Lee Miller and Billy Bob Thornton, she separated from her third husband, actor Brad Pitt, in September 2016.
5. The arrival of the **first** biological child Jolie and Pitt caused an excited flurry with fans.
6. On July 12, 2008, she gave birth to **twins**: a son, Knox Leon, and a daughter, Vivienne Marcheline.
7. In 2016, Jolie brought her **twins**, **one** daughter and **three** adopted children to the gala.

# Stage 1: CQ Recognition

S



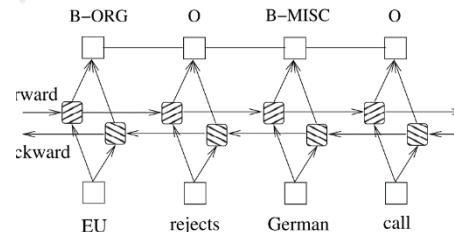
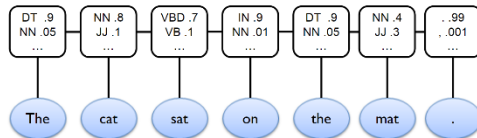
- In 2016, Jolie brought her twins, one daughter and three adopted children to the gala

preprocessing

...her	twins	,	one	daughter	and	three	adopted	children	to...
...her	NUMTERM	,	CARDINAL	daughter	and	CARDINAL	adopted	children	to...
0	COUNT	COMP	COUNT	0	COMP	COUNT	0	0	0

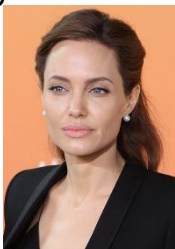
P<sub>hasChild</sub>

- Sequence labelling task
  - One model learned per predicate
  - Feature-based model (CRF) vs Neural model (bi-LSTM-CRF)



# Stage 1: CQ Recognition

S



- In 2016, Jolie brought her twins, one daughter and three adopted children to the gap **preprocessing**

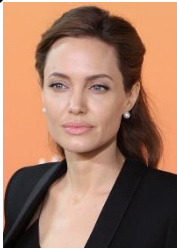
...her	twins	,	one	daughter	and	three	adopted	children	to...
...her	NUMTERM	,	CARDINAL	daughter	and	CARDINAL	adopted	children	to...
0	COUNT	COMP	COUNT	0	COMP	COUNT	0	0	0

P<sub>hasChild</sub>

- Incompleteness-aware distant supervision
  - COUNT DISTINCT <Angelina Jolie, hasChild, \*> as seed counts
  - Filtering training data based on subject popularity
  - Ignoring higher counts, unless > upper bound (count at 99th percentile)
    - e.g., 2016 cannot be number of children
  - Ignoring counts with low entropy
    - Count '1' appears abundantly in the text
  - Label the tokens with COUNT (and COMP) when
    - the token itself, OR
    - the sum of several tokens match the seed count

# Stage 2: CQ Consolidation

S



- She has a grand total of six<sub>0.4</sub> children together: three<sub>0.5</sub> biological [and] three<sub>0.3</sub> adopted. → 6<sub>0.4</sub>, 6<sub>0.5</sub>
- Angelina Jolie and four<sub>0.3</sub> of her kids soaked up the last few days of summer over Labor Day. → 4<sub>0.3</sub>
- The arrival of the first<sub>0.5</sub> biological child Jolie and Pitt caused an excited flurry with fans. → 1<sub>0.5</sub>
- On July 12, 2008, she gave birth to twins<sub>0.8</sub>: a<sub>0.2</sub> son, Knox Leon, [and] a<sub>0.1</sub> daughter, Vivienne Marcheline. → 2<sub>0.8</sub>, 2<sub>0.2</sub>

P  
hasChild

∃6

1. cardinals      6<sub>0.5</sub>

2. numterms      2<sub>0.8</sub>

3. ordinals      1<sub>0.5</sub>

4. articles      2<sub>0.2</sub>

*threshold = 0.5*

# Training data setup

- Wikidata as source KB, Wikipedia pages of subject S as input texts
- 5 relation/predicate P

Wikidata Subject Class	Wikidata Property	Relation	Train/Test data size	
			#Subjects	#Sentences
series of creative works	has part	containsWork	642	7,984
musical ensemble	has part	hasMember	8,901	96,056
admin. territorial entity	contains admin...	containsAdmin	6,266	13,199
human	child	hasChild	40,145	319,807
human	spouse	hasSpouse	45,261	408,974

At least one object

- Training set: Wikidata object counts as seed counts
- Test set: manually annotated CQs

# Evaluation

- Stage 1: CQ recognition
  - CRF models more robust than bi-LSTMs (57% vs 40% avg F1-score)
    - Neural models much more prone to **overfitting to noisy training data**

	containsWork	hasMember	containsAdmin	hasChild	hasSpouse
CINEX-CRF	39.8	56.1	77.3	49.0	62.4

F1-scores

- Stage 2: CQ consolidation

	containsWork	hasMember	containsAdmin	hasChild	hasSpouse
CINEX-CRF	49.2	64.3	78.6	50.0	58.1

Precision

(Contribution)

CARDINAL	55.0 (33.9)	62.5 (28.6)	85.7 (87.5)	67.3 (70.5)	75.0 (18.6)
NUMT.+ART.	62.5 (40.7)	65.0 (71.4)	33.3 (10.7)	6.3 (20.5)	43.8 (37.2)
ORDINAL	20.0 (25.4)	0 (0)	0 (1.8)	14.3 (9.0)	63.2 (44.2)
ORDINAL (as lower bound)	86.7	0	0	85.7	89.5



# Evaluation: Error Analysis

- Confusion of relations having similar CQs
  - <Ladysmith Black Mambazo, hasMember,  $\exists 6$ >
    - "...Mazibuko (the eldest of the six brothers) joined Mambazo..."
    - Confused with hasSibling
  - <Ruth W. Khama, hasSpouse,  $\exists 2$ >
    - "...and twins Anthony and Tshekedi were born in..."
    - Confused with hasChild
- Confusion of entity type granularity
  - <Scandal (TV series), containsWork,  $\exists 10$ >
    - "...the first season consisting of ten episodes."
    - TV series contains seasons
    - seasons contains episodes

# KB Enrichment Potential

- Enrich KB with knowledge that facts exist
- Apply CINEX on all Wikidata relations:
  - Filter out functional properties
  - Relations → properties paired with 10 most frequent subject classes
  - Per relation → Evaluate CINEX on 10% (up to 200) most popular subjects as test set
    - CINEX yields >50% precision → 110 relations → having good extracted CQs
  - Apply 110 CINEX models on all subject entities of corresponding classes
- CINEX enrich KB (for 110 relations) with existence of 28.3% more facts

property	class	KB facts	CQ facts
has part	rock band	1,147	1,516 (+32.2%)

# Outline

1. Problem
2. Manual patterns
3. Supervised learning
  1. Feature-based
  2. TACRED and BERT
4. Semi- and unsupervised extraction
  1. Iterative pattern learning (DIPRE)
  2. Distant supervision
    - CINEX
5. Evaluation
6. OpenIE
  1. PATTY
  2. Universal schema
  3. Quasimodo
7. Negation
8. Extraction/prediction from latent representations

# Detect members of the Simpsons

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.



# Def: Gold Standard

The gold standard (also: ground truth) for an IE task is the set of desired results of the task on a given corpus.

Task: Detect Simpson members

Corpus:

in The Simpsons, Homer Simpson is the father of Bart Simpson and Lisa Simpson. The M above his ear is for Matt Groening.

Gold Standard:

{Homer Simpson, Bart Simpson, Lisa Simpson}

# Def: Precision

The **precision** of an IE algorithm is the ratio of its outputs that are in the respective gold standard.

$$prec = \frac{|Output \cap GStandard|}{|Output|}$$

Output: {Homer, Bart, Groening}

✓   ✓   ✗

G.Standard: {Homer, Bart, Lisa, Marge}

=> Precision:  $2/3 = 66\%$

# Def: Recall

The **recall** (also: sensitivity, true positive rate, hit rate) of an IE algorithm is the ratio of the gold standard that is output.

$$rec = \frac{|Output \cap GStandard|}{|GStandard|}$$

Output: {Homer, Bart, Groening}

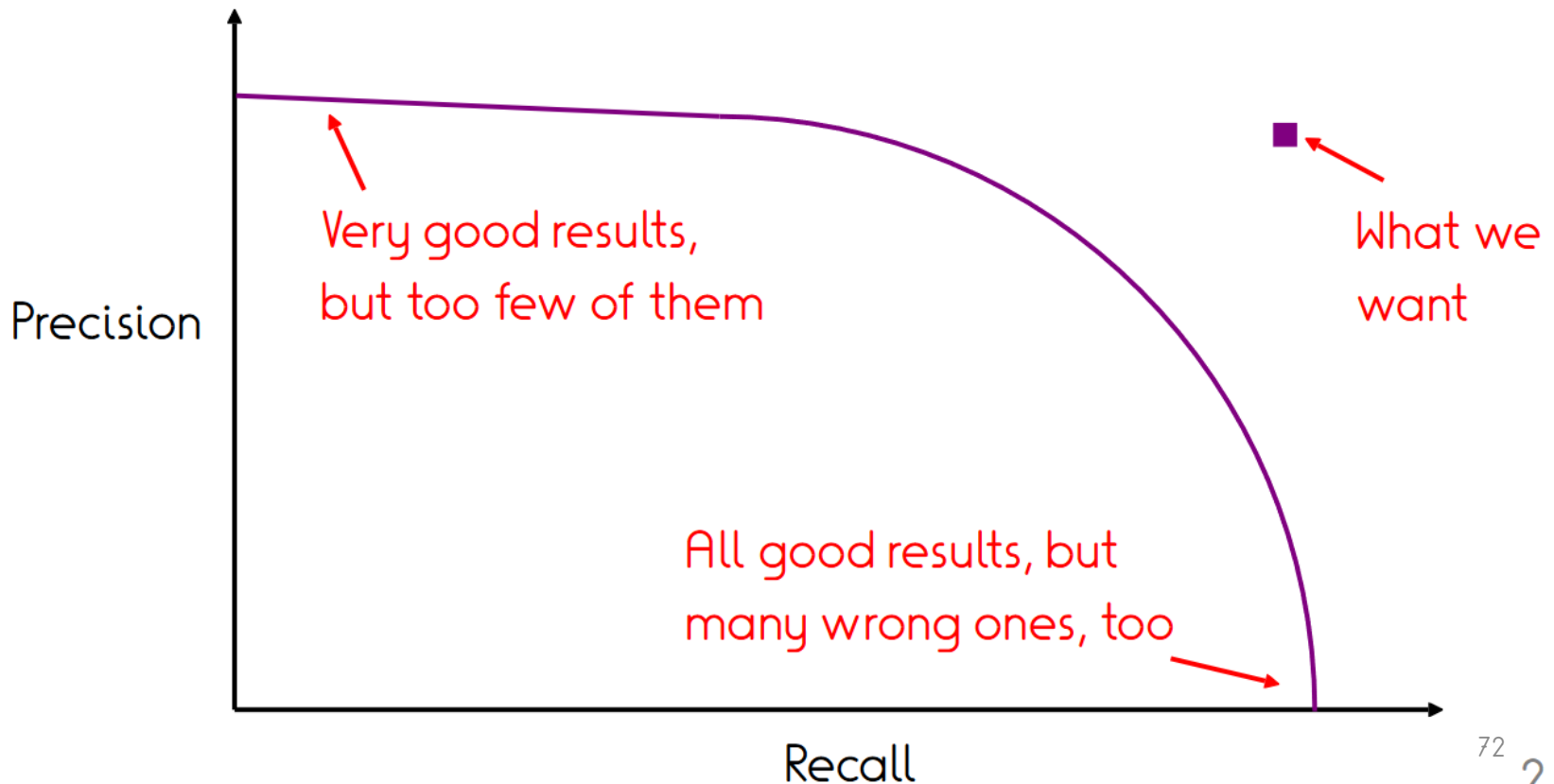
G.Standard: {Homer, Bart, Lisa, Marge}

✓   ✓   ✗   ✗

=> Recall: 2/4 = 50%

# Precision-Recall-Tradeoff

It is very hard to get both good precision and good recall. Algorithms usually allow varying one at the expense of the other (e.g., by choosing different threshold values). This usually yields:





# Def: F1

To trade off precision and recall, we could use the average:

Gold Standard: {Homer, Bart, Lisa, Snowball\_4, ..., Snowball\_100}

Output: {Homer Simpson}

Precision:  $1/1=100\%$ , Recall:  $1/100=1\%$

Average:  $(100\%+1\%)/2=50\%$

Outputting just  
a single result  
already gives a  
score of 50%!



The **F1 measure** is the harmonic mean of precision and recall.

$$F1 = 2 \times \frac{\textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}$$

Precision:  $1/1=100\%$ , Recall:  $1/100=1\%$

F1:  $2 \times 100\% \times 1\% / (100\% + 1\%) = 2\%$



# Task: Precision & Recall

What is the algorithm output, the gold standard, the precision and the recall in the following cases?

1. Nostradamus predicts a trip to the moon for every century from the 15th to the 20th incl.
2. The weather forecast predicts that the next 3 days will be sunny. It does not say anything about the 2 days that follow. In reality, it is sunny during all 5 days.
3. On Elvis Radio™, 90% of the songs are by Elvis. An algorithm learns to detect Elvis songs. Out of 100 songs on Elvis Radio, the algorithm says that 20 are by Elvis (and says nothing about the other 80). Out of these 20 songs, 15 were by Elvis and 5 were not.
4. How can you improve the algorithm?

# Precision-recall tradeoff - Example

- <https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm>

# References

- Papers:
  - Sergey Brin, Extracting Patterns and Relations from the World Wide Web, WebDB 1998
  - Mintz et al., Distant supervision for relation extraction without labeled data, 2009
  - Riedel et al., Modeling Relations and Their Mentions without Labeled Text, ECML 2010
  - Mirza et al., Enriching Knowledge Bases with Counting Quantifiers, ISWC 2018
- Slides
  - Fabian Suchanek, Paramita Mirza and Dan Jurafsky
- Code/APIs
  - No off-the shelf solutions (training needed)
  - Extensive code on Github etc.
  - Rosette API <https://www.rosette.com/capability/relationship-extraction/#try-the-demo> (commercial)

# Assignment 6

- Pattern-based relation extraction
- Similar to type extraction, but now longer text
- Suggestion: Pattern-based extraction using spaCy NER tags
- Evaluation using micro F1

# Take home

- Supervised learning data bottleneck, but performant
- Iterative pattern learning and distant supervision as alternatives