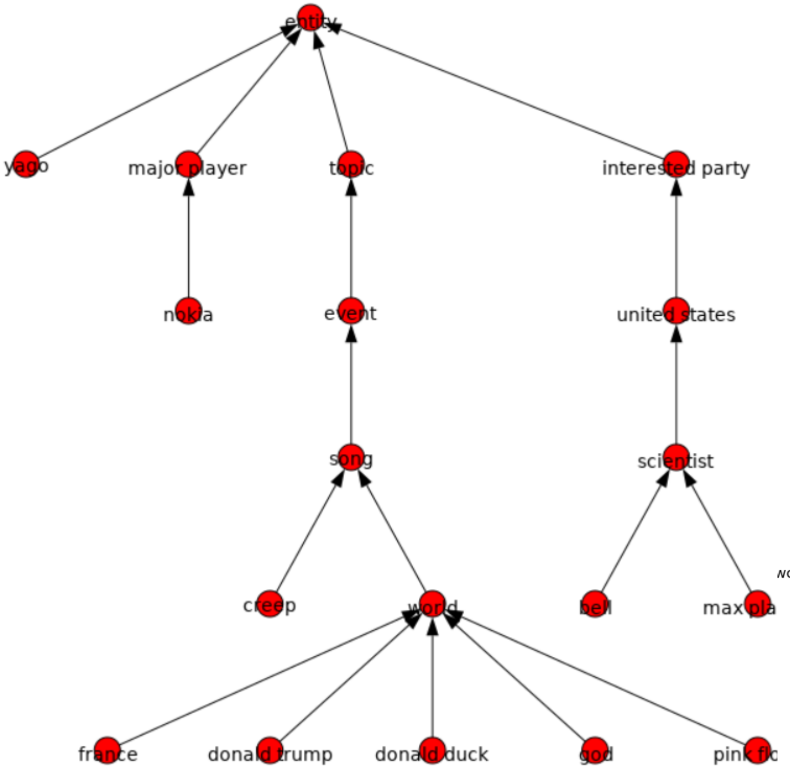
# Information extraction

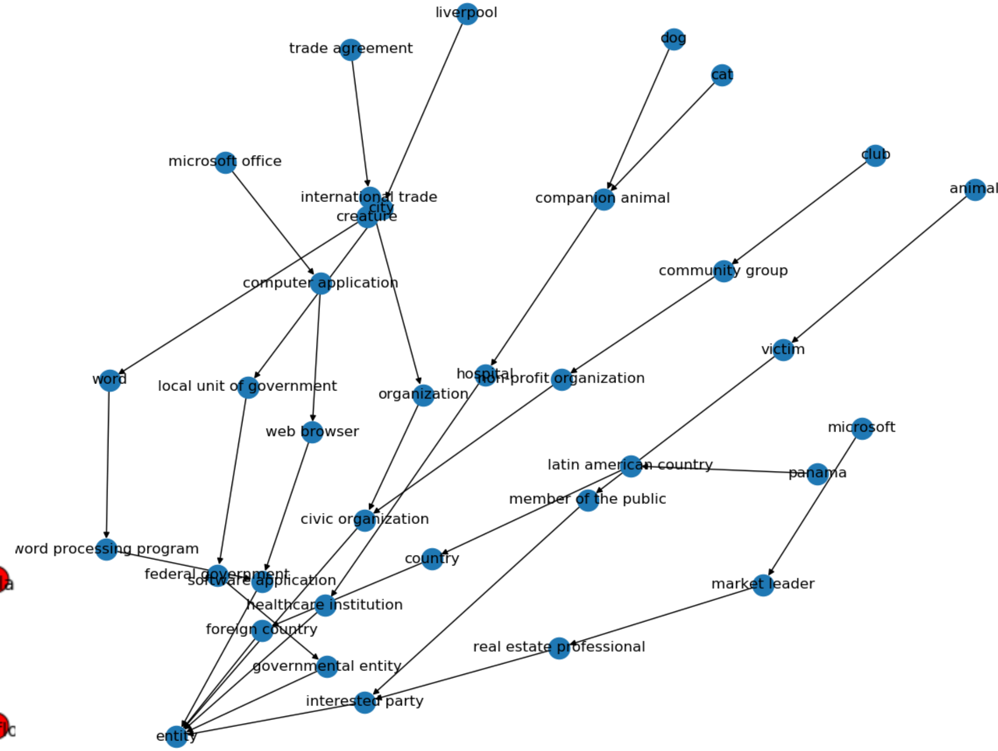
## 6.+7. Relation extraction

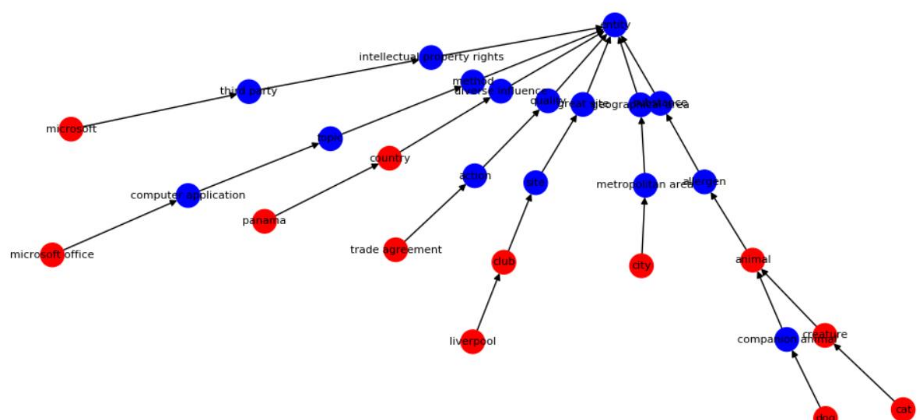
Simon Razniewski
Winter semester 2019/20

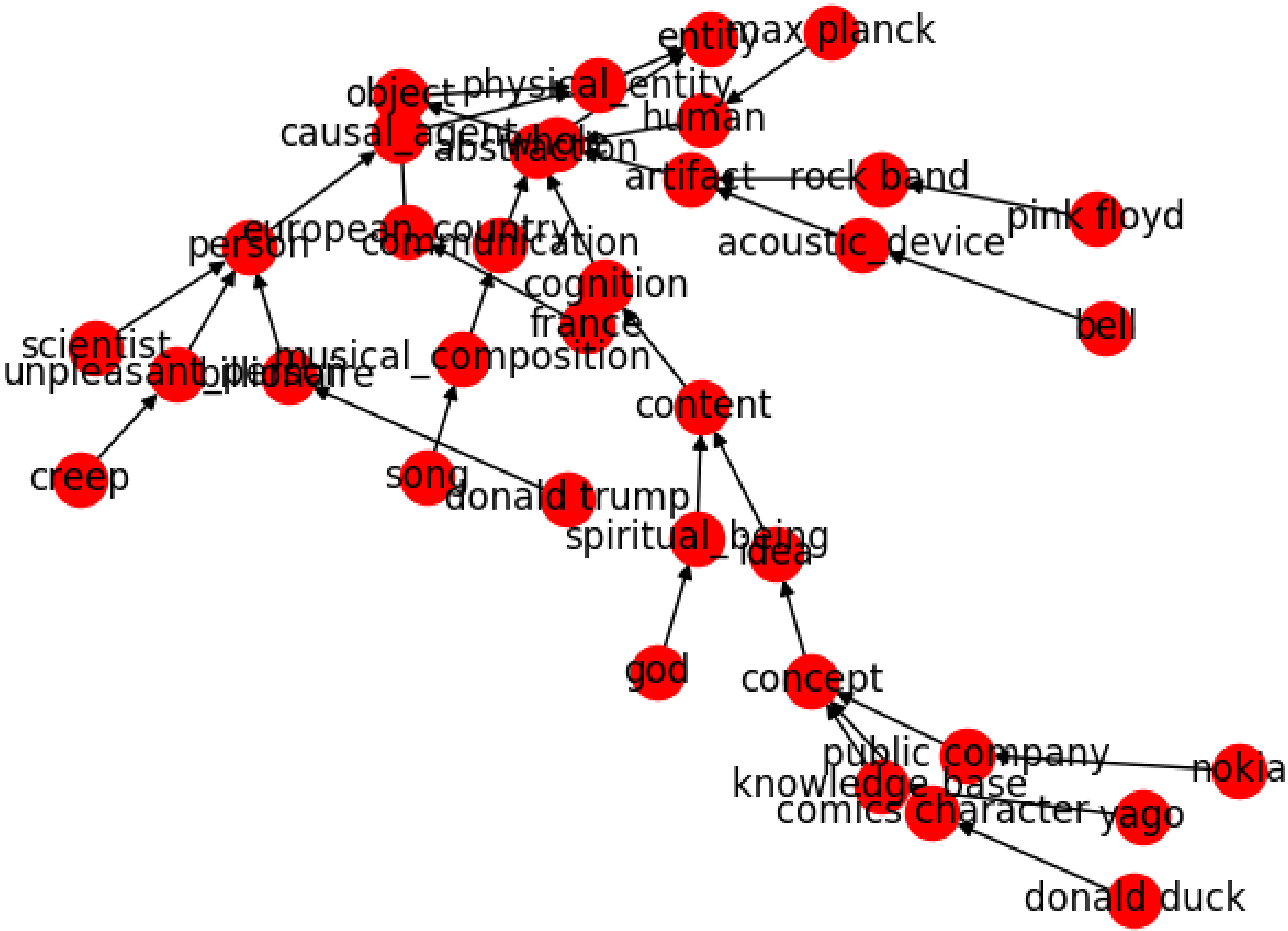
# Announcements

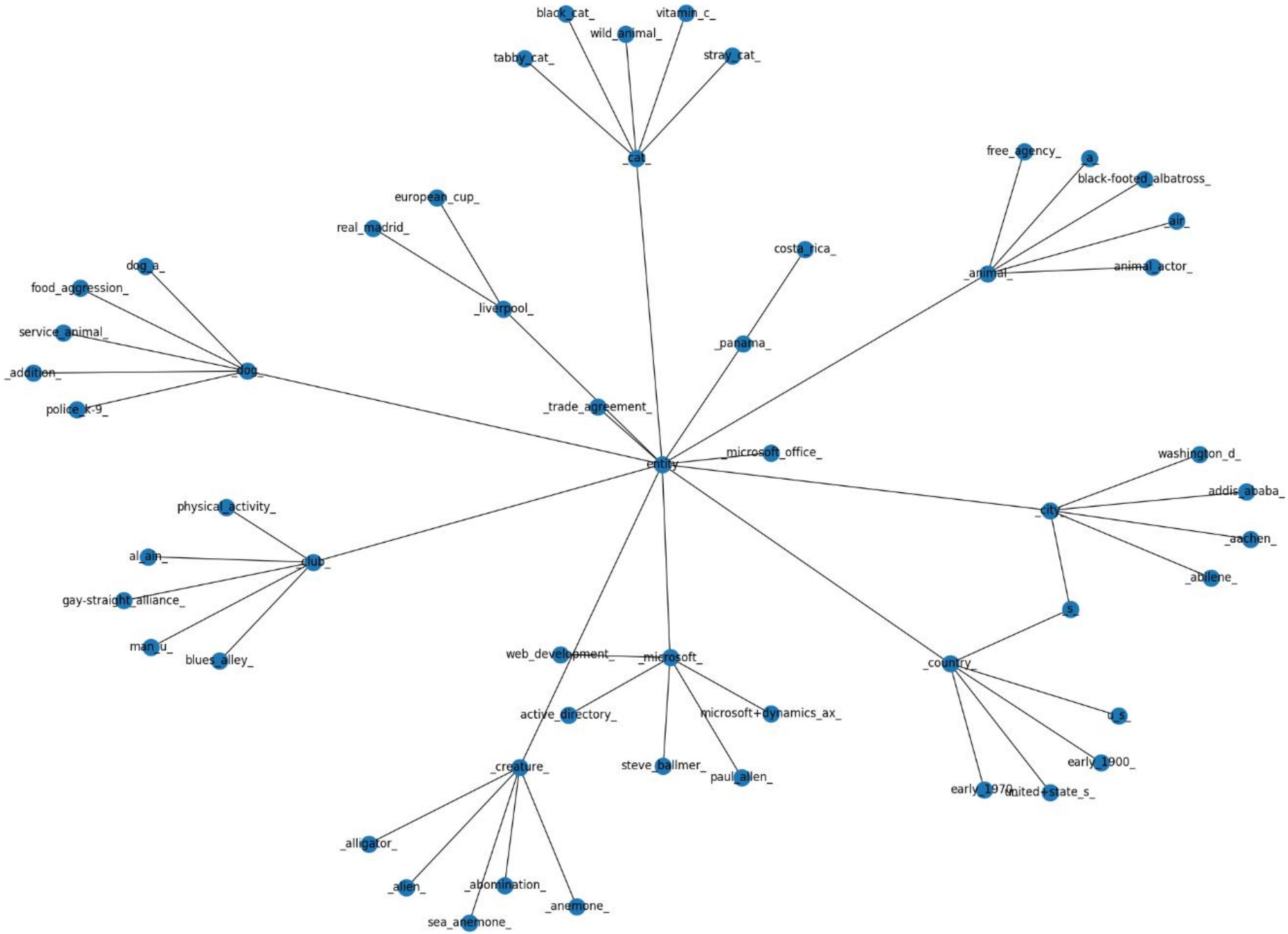
- Neural coref: Score below mentions is score for having no antecedents
- Videolectures will be considered next year
  - For now slides, and literature pointers (ask me for more)
- Results assignment 5
  - Sample solutions next slides


entity
yago
major player
topic
interested party
nokia
event
united states
song
scientist
creep
world
bell
max pla
france
donald trump
donald duck
god
pink fl


liverpool
trade agreement
dog
cat
microsoft office
club
international trade
city
creature
companion animal
animal
computer application
community group
victim
word
local unit of government
hospital
non-profit organization
organization
web browser
microsoft
latin american country
panama
member of the public
civic organization
word processing program
country
federal government
software application
healthcare institution
market leader
foreign country
real estate professional
governmental entity
interested party
entity


intellectual property rights
third party
microsoft
computer application
microsoft office
country
panama
trade agreement
city
club
liverpool
metropolitan area
animal
companion animal
creature
dog
cat


entity
max planck
object
physical_entity
human
causal_agent
abstraction
artifact
rock band
pink floyd
person
european_country
communication
acoustic_device
cognition
france
bell
scientist
musical_composition
content
creep
song
donald trump
spiritual_being
idea
god
concept
public company
nokia
knowledge base
comics character
yago
donald duck

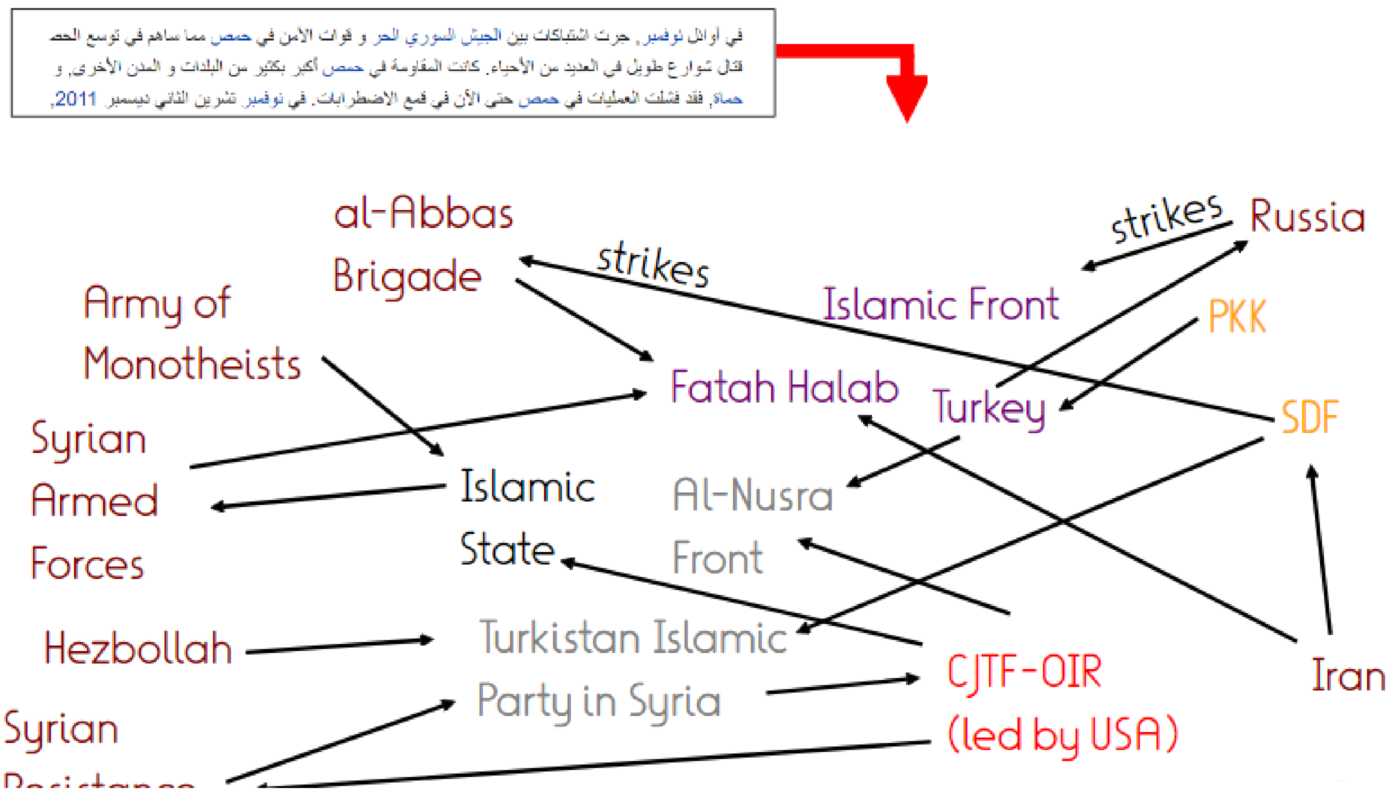

black_cat_
vitamin_c_
wild_animal_
tabby_cat_
stray_cat_
_cat_
free_agency_
_a_
black-footed_albatross_
european_cup_
real_madrid_
_air_
costa_rica_
animal_actor_
dog_a_
_animal_
food_aggression_
_liverpool_
service_animal_
_panama_
_addition_
_dog_
police_k-9_
_trade_agreement_
_microsoft_office_
washington_d_
entity
addis_ababa_
physical_activity_
_city_
_aachen_
al_ain_
_club_
_abilene_
gay-straight_alliance_
_s_
man_u_
blues_alley_
web_development_
_microsoft_
_country_
active_directory_
microsoft+dynamics_ax_
u_s_
steve_ballmer_
_creature_
paul_allen_
early_1900_
early_1970_
united+state_s_
_alligator_
_alien_
_abomination_
_anemone_
sea_anemone_

# 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.



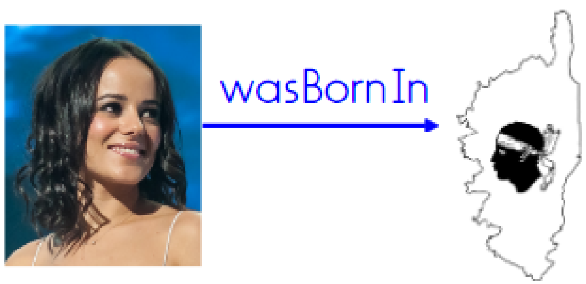
# Relation extraction, happier example

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

For now, we concentrate on facts with a single relation.

Alizée kommt aus Corsica.

For the computer, the corpus is completely incomprehensible – as if it were written in a foreign language!




The extracted facts, on the other hand, use well-defined relations.

# Extracting relations from text

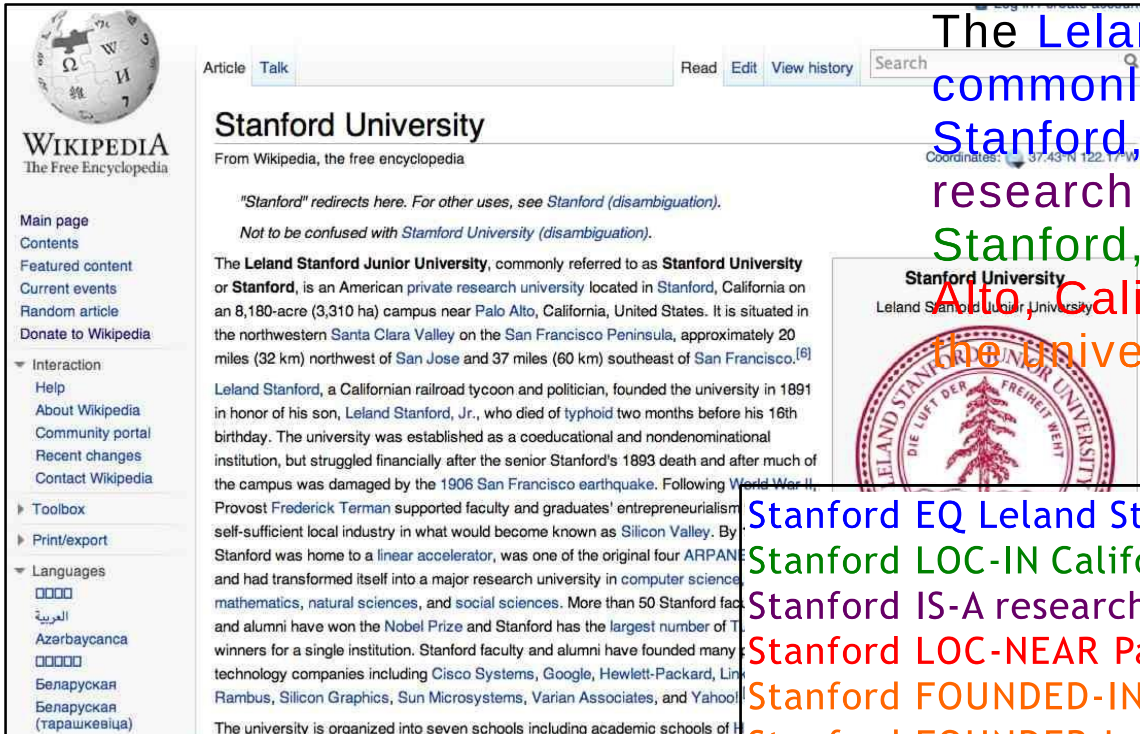
- Company report: "International Business Machines Corporation (IBM or the company) was incorporated in the State of New York on June 16, 1911, as the Computing-Tabulating-Recording Co. (C-T-R)…"
- Extracted Complex Relation:

  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


WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact Wikipedia

Toolbox
Print/export
Languages
العربية
Azerbaycanca
Беларуская
Беларуская (тарашкевіца)

Article Talk Read Edit View history Search

## Stanford University

From Wikipedia, the free encyclopedia

*"Stanford" redirects here. For other uses, see Stanford (disambiguation).*

*Not to be confused with Stamford University (disambiguation).*

The **Leland Stanford Junior University**, commonly referred to as **Stanford University** or **Stanford**, is an American private research university located in Stanford, California on an 8,180-acre (3,310 ha) campus near Palo Alto, California, United States. It is situated in the northwestern Santa Clara Valley on the San Francisco Peninsula, approximately 20 miles (32 km) northwest of San Jose and 37 miles (60 km) southeast of San Francisco.[6]

Leland Stanford, a Californian railroad tycoon and politician, founded the university in 1891 in honor of his son, Leland Stanford, Jr., who died of typhoid two months before his 16th birthday. The university was established as a coeducational and nondenominational institution, but struggled financially after the senior Stanford's 1893 death and after much of the campus was damaged by the 1906 San Francisco earthquake. Following World War II, Provost Frederick Terman supported faculty and graduates' entrepreneurialism self-sufficient local industry in what would become known as Silicon Valley. By Stanford was home to a linear accelerator, was one of the original four ARPAN and had transformed itself into a major research university in computer science mathematics, natural sciences, and social sciences. More than 50 Stanford fac and alumni have won the Nobel Prize and Stanford has the largest number of T winners for a single institution. Stanford faculty and alumni have founded many technology companies including Cisco Systems, Google, Hewlett-Packard, Lin Rambus, Silicon Graphics, Sun Microsystems, Varian Associates, and Yahoo!

The university is organized into seven schools including academic schools of H

Stanford University
Leland Stanford Junior University

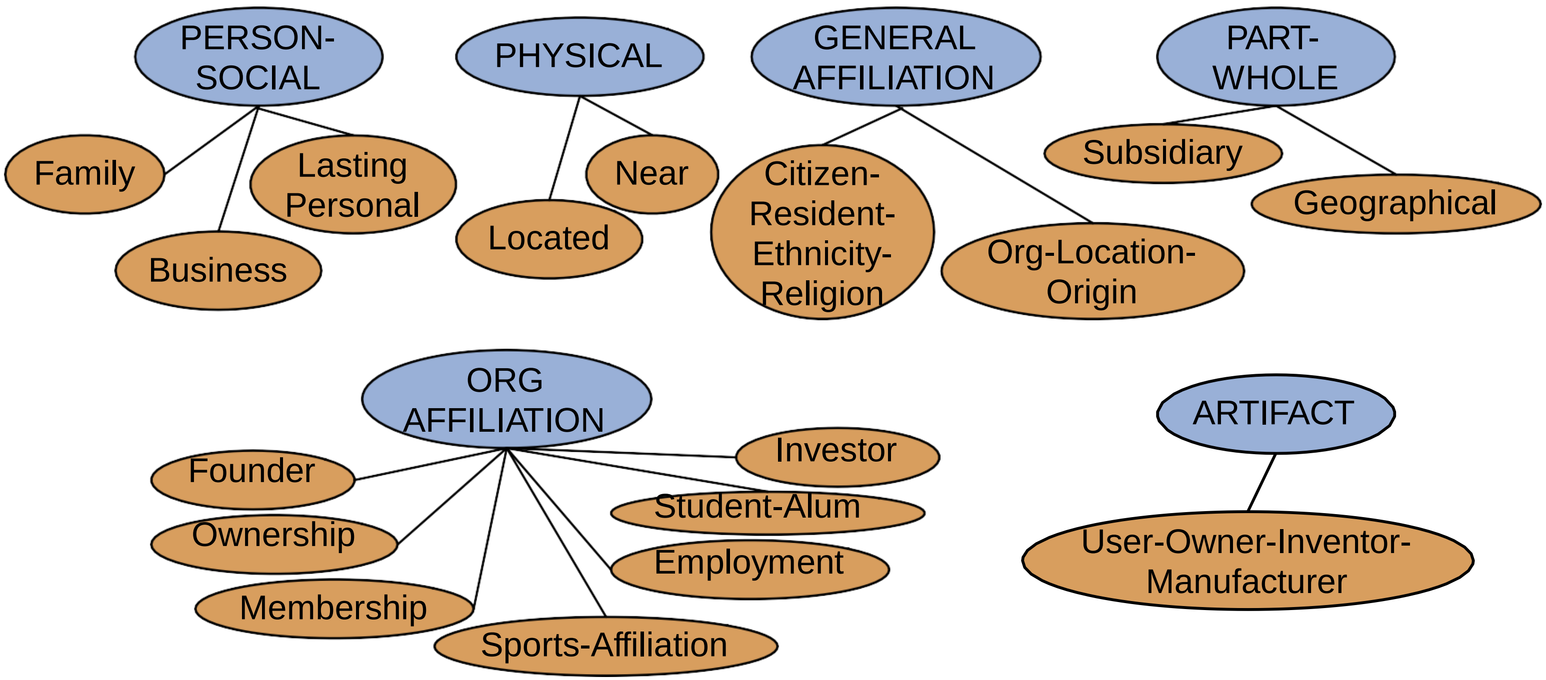
The Leland Stanford University, commonly referred to as Stanford, is an American private research university located in Stanford, California , near Palo Alto, California… Leland founed the university in 1891.

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)

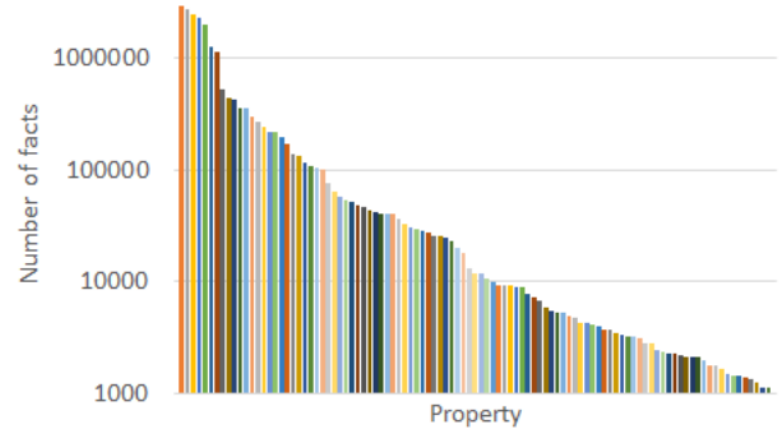
- Physical-Located PER-GPE
  
  He was in Tennessee
- Part-Whole-Subsidiary ORG-ORG
  
  XYZ, the parent company of ABC
- Person-Social-Family PER-PER
  
  John's wife Yoko
- Org-AFF-Founder PER-ORG
  
  Steve Jobs, co-founder of Apple…
- 

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

"$X_{PERS}$ ($Y_{LOC}$, DATE-)"
"Born in $Y_{LOC}$, $X_{PERS}$"
…

# 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

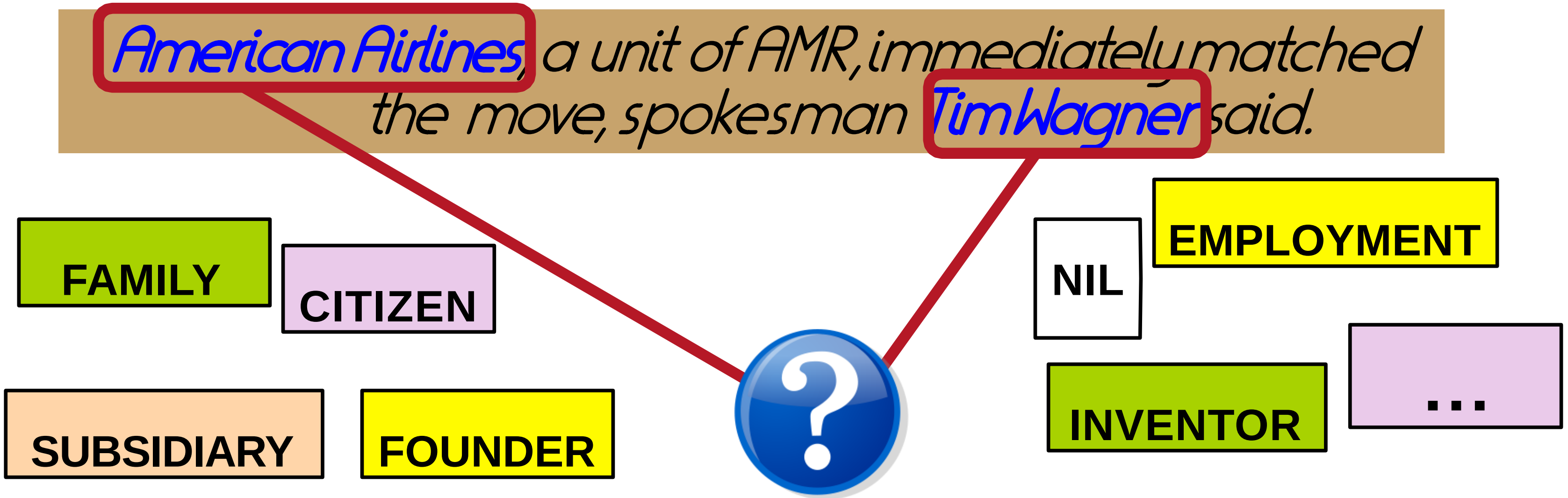
1. Problem
2. Manual patterns
3. Supervised learning
    1. Feature-based
    2. TACRED and BERT
4. Semi- and unsupervised extraction
    1. Iterative pattern learning (DIPRE)
    2. Distant supervision
        - CINEX
5. Evaluation
6. OpenIE
    1. PATTY
    2. Universal schema
    3. Quasimodo
7. Negation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



# 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 $\Uparrow$ NP $\Uparrow$ S $\Uparrow$ S $\Downarrow$ 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$_1$ type | ORG |
| Entity$_1$ head | *airlines* |
| Entity$_2$ type | PERS |
| Entity$_2$ head | *Wagner* |
| Concatenated types | ORGPERS |
| **Word-based features** | |
| Between-entity bag of words | { *a, unit, of, AMR, Inc., immediately, matched, the, move, spokesman* } |
| Word(s) before Entity$_1$ | NONE |
| Word(s) after Entity$_2$ | *said* |
| **Syntactic features** | |
| Constituent path | $NP \uparrow NP \uparrow S \uparrow S \downarrow NP$ |
| Base syntactic chunk path | $NP \rightarrow NP \rightarrow PP \rightarrow NP \rightarrow VP \rightarrow NP \rightarrow NP$ |
| Typed-dependency path | $Airlines \leftarrow_{subj} matched \leftarrow_{comp} said \rightarrow_{subj} Wagner$ |

# 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

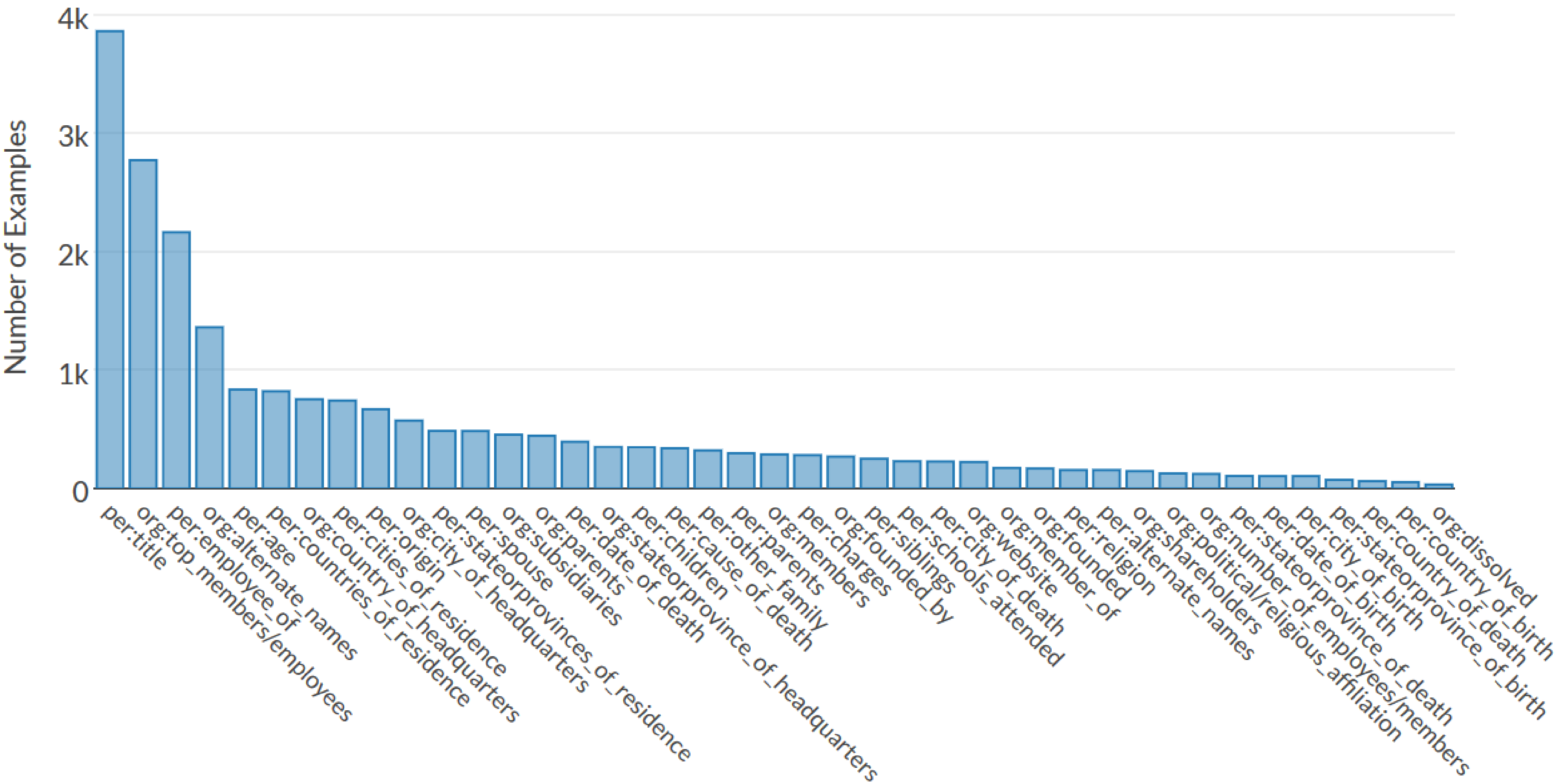
# Outline

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)

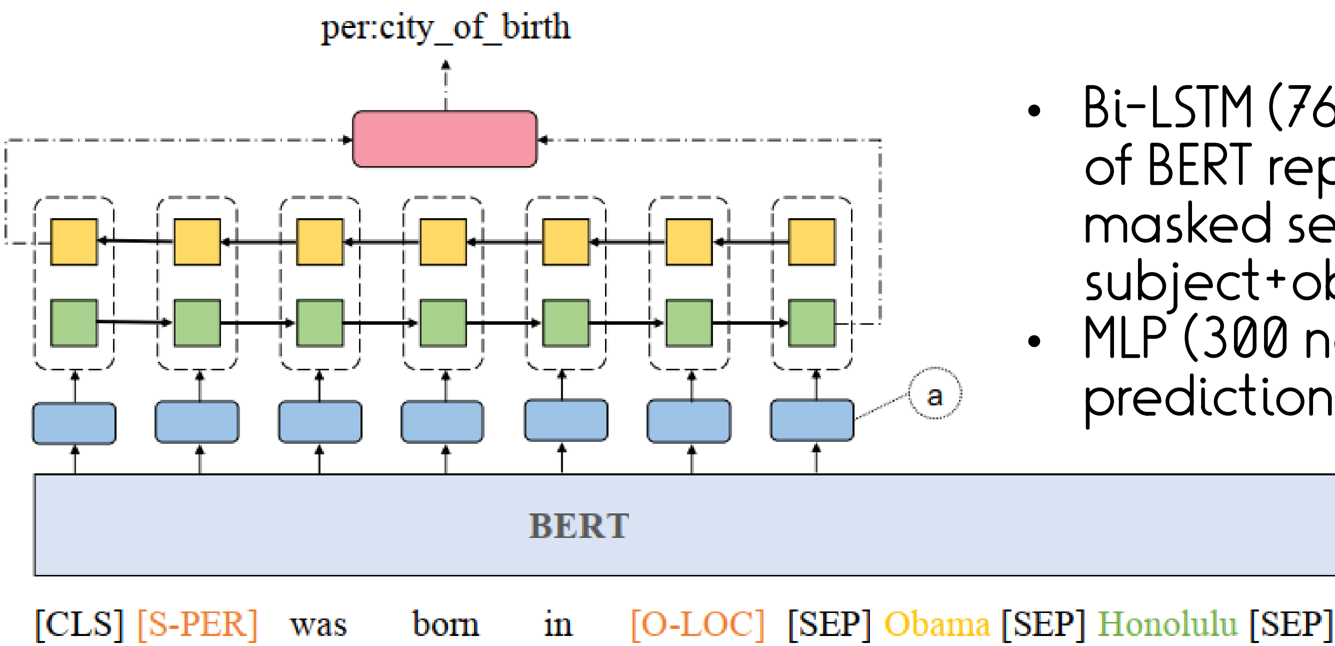

Number of Examples
4k
3k
2k
1k
0
per:title
org:top_members/employees
per:employee_of
org:alternate_names
per:age
per:countries_of_residence
org:country_of_headquarters
per:cities_of_residence
per:origin
org:city_of_headquarters
per:stateorprovinces_of_residence
per:spouse
org:subsidiaries
org:parents
per:date_of_death
org:stateorprovince_of_headquarters
per:children
per:cause_of_death
per:other_family
per:parents
org:members
per:charges
org:founded_by
per:siblings
per:schools_attended
per:city_of_death
org:website
org:member_of
org:founded
per:religion
per:alternate_names
org:shareholders
org:political/religious_affiliation
org:number_of_employees/members
per:stateorprovince_of_death
per:date_of_birth
per:city_of_birth
per:stateorprovince_of_birth
per:country_of_death
per:country_of_birth
org:dissolved

# TACRED (3)

| | Model | P | R | F1 |
|---|---|---|---|---|
| Traditional | Patterns | **86.9** | 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 | **64.5** | **65.1** |

[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 | $F_1$ |
|---|---|---|---|
| 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 | **73.3** | 63.10 | 67.8 |
| Zhang et al. (2018) (ensemble) | 71.3 | **65.4** | **68.2** |

[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."

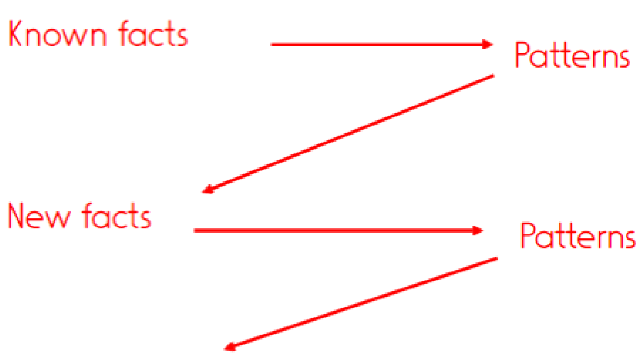
  X is buried in Y

  "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

KB

| | | |
|---|---|---|
| Obama | $\xrightarrow{\text{chases}}$ | Osama |
| Tom | $\xrightarrow{\text{chases}}$ | Jerry |

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

KB Merkel $\xrightarrow{\text{marriedTo}}$ Sauer

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 |
|---|---|
| `www.sff.net/locus/c.*` | `<LI><B>`*title*`</B> by` *author* `(` |
| `dns.city-net.com/l̃mann/awards/hugos/1984.html` | `<i>`*title*`</i> by` *author* `(` |
| `dolphin.upenn.edu/d̃cummins/texts/sf-award.htm` | *author* `\|\|` *title* `\|\| (` |

→ 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.

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

- 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

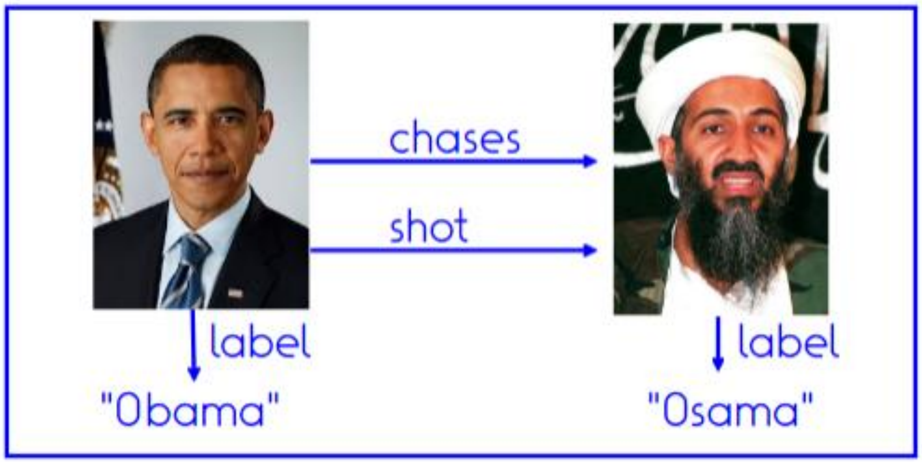
P(born-in | f1,f2,f3,…,f70000)

# 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

KB




Corpus

Obama verfolgt Osama.

=> "X verfolgt Y" is a pattern for chases(X,Y) for 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% |

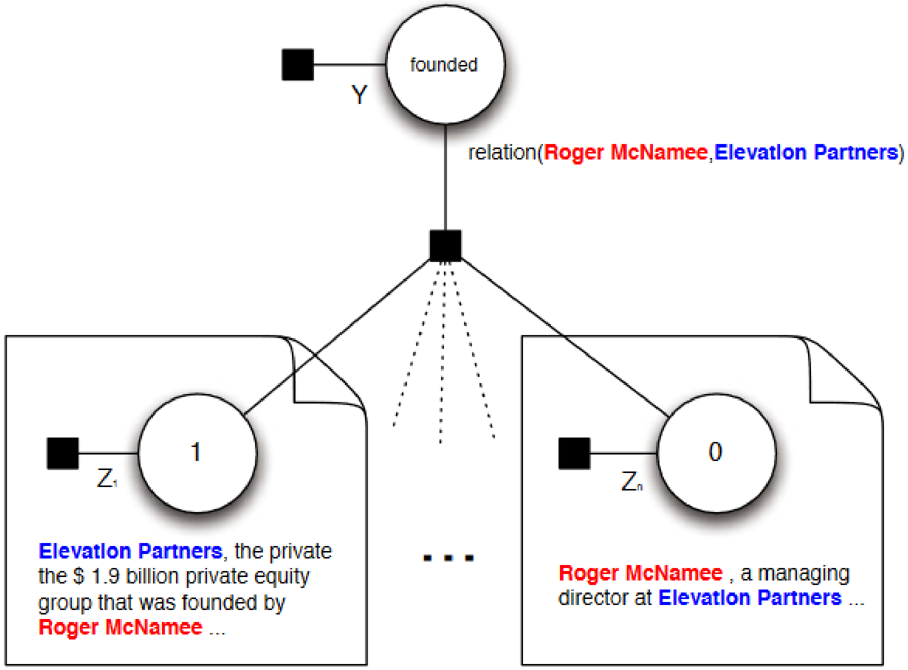
# Fixing the naive assumption

→ 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: <S, P, O>

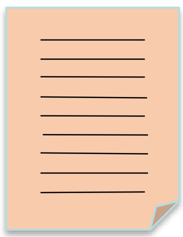
  <California, hasCounty, Monterey> <Donald Trump, hasSpouse, Melania Knauss>

- Counting information: <S, P, ∃O>

  <California, hasCounty, ∃58> <Donald Trump, hasSpouse, ∃3>

"There exists a specific number of O for a given SP pair"

# Problem: CQ Extraction

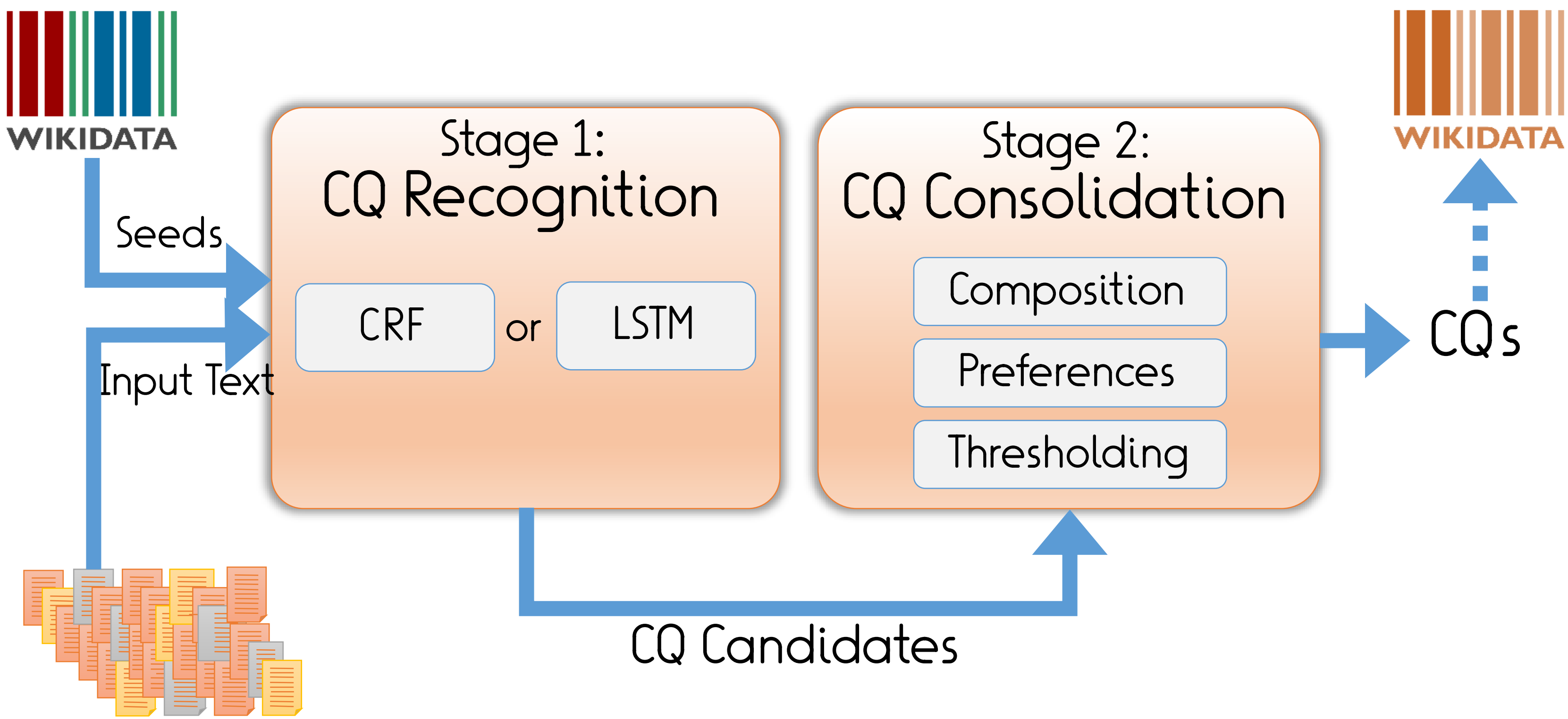
| Given | | Determine |
|---|---|---|
| $S$   | $P_{hasChild}$ | $\exists 6$ |

# 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'

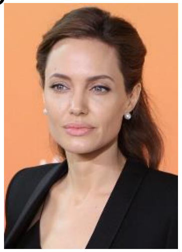
- Compositionality
  - In 2016, Jolie brought her twins, one daughter and three adopted children to the gala.

# CINEX: Counting INformation EXtraction



# Stage 1: CQ Recognition

S



P hasChild

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

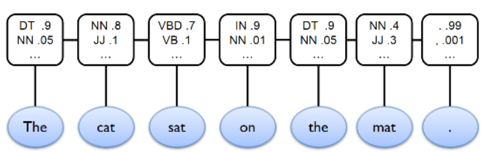
- In 2016, Jolie brought her twins, one daughter and three adopted children to the gala.

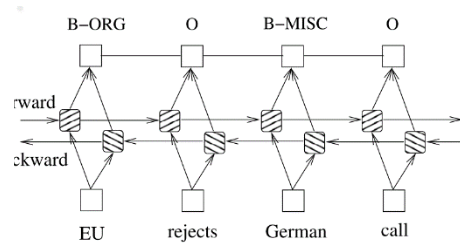
preprocessing

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ...her | twins | , | one | daughter | and | three | adopted | children | to... |
| ...her | NUMTERM | , | CARDINAL | daughter | and | CARDINAL | adopted | children | to... |
| O | COUNT | COMP | COUNT | O | COMP | COUNT | O | O | O |

$P_{hasChild}$

- 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 gala.

preprocessing

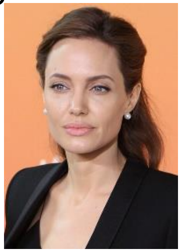
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| …her | twins | , | one | daughter | and | three | adopted | children | to… |
| …her | NUMTERM | , | CARDINAL | daughter | and | CARDINAL | adopted | children | to… |
| O | COUNT | COMP | COUNT | O | COMP | COUNT | O | O | O |

$P_{hasChild}$

- 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



P hasChild

∃6

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

| | |
|---|---|
| 1. cardinals | $6_{0.5}$ |
| 2. numterms | $2_{0.8}$ |
| 3. ordinals | $1_{0.5}$ |
| 4. articles | $2_{0.2}$ |

*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<br>#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, ∃6⟩
    - *"…Mazibuko (the eldest of the six brothers) joined Mambazo…"*
    - Confused with hasSibling
  - ⟨Ruth W. Khama, hasSpouse, ∃2⟩
    - *"…and twins Anthony and Tshekedi were born in…"*
    - Confused with hasChild
- Confusion of entity type granularity
  - ⟨Scandal (TV series), containsWork, ∃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}

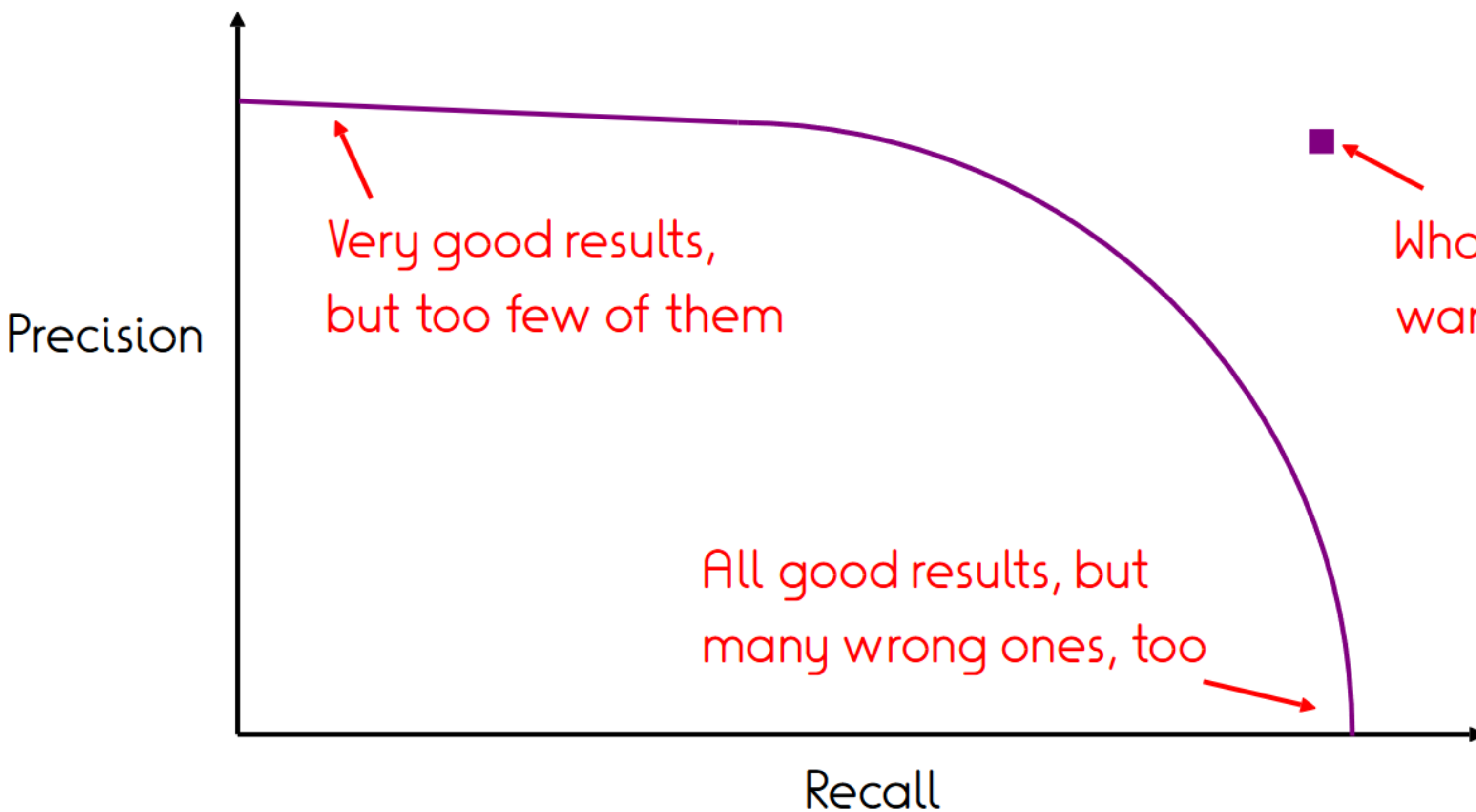
G.Standard: {Homer, Bart, Lisa, Marge}

✓ ✓ ✗ ✗

=> Recall: 2/4 = 50%

# Precision-Recall-Tradeoff

It is very hard to get both good precision and good recall.
Algorithms usually allowing 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.

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

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

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

# F1 given P and R

| P | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 0 | 0.1 | 0.18 | 0.26 | 0.33 | 0.4 | 0.46 | 0.52 | 0.57 | 0.62 | 0.67 | 0.71 | 0.75 | 0.79 | 0.82 | 0.86 | 0.89 | 0.92 | 0.95 | 0.97 | 1 |
| **0.95** | 0 | 0.1 | 0.18 | 0.26 | 0.33 | 0.4 | 0.46 | 0.51 | 0.56 | 0.61 | 0.66 | 0.7 | 0.74 | 0.77 | 0.81 | 0.84 | 0.87 | 0.9 | 0.92 | 0.95 | 0.97 |
| **0.9** | 0 | 0.09 | 0.18 | 0.26 | 0.33 | 0.39 | 0.45 | 0.5 | 0.55 | 0.6 | 0.64 | 0.68 | 0.72 | 0.75 | 0.79 | 0.82 | 0.85 | 0.87 | 0.9 | 0.92 | 0.95 |
| **0.85** | 0 | 0.09 | 0.18 | 0.26 | 0.32 | 0.39 | 0.44 | 0.5 | 0.54 | 0.59 | 0.63 | 0.67 | 0.7 | 0.74 | 0.77 | 0.8 | 0.82 | 0.85 | 0.87 | 0.9 | 0.92 |
| **0.8** | 0 | 0.09 | 0.18 | 0.25 | 0.32 | 0.38 | 0.44 | 0.49 | 0.53 | 0.58 | 0.62 | 0.65 | 0.69 | 0.72 | 0.75 | 0.77 | 0.8 | 0.82 | 0.85 | 0.87 | 0.89 |
| **0.75** | 0 | 0.09 | 0.18 | 0.25 | 0.32 | 0.38 | 0.43 | 0.48 | 0.52 | 0.56 | 0.6 | 0.63 | 0.67 | 0.7 | 0.72 | 0.75 | 0.77 | 0.8 | 0.82 | 0.84 | 0.86 |
| **0.7** | 0 | 0.09 | 0.18 | 0.25 | 0.31 | 0.37 | 0.42 | 0.47 | 0.51 | 0.55 | 0.58 | 0.62 | 0.65 | 0.67 | 0.7 | 0.72 | 0.75 | 0.77 | 0.79 | 0.81 | 0.82 |
| **0.65** | 0 | 0.09 | 0.17 | 0.24 | 0.31 | 0.36 | 0.41 | 0.46 | 0.5 | 0.53 | 0.57 | 0.6 | 0.62 | 0.65 | 0.67 | 0.7 | 0.72 | 0.74 | 0.75 | 0.77 | 0.79 |
| **0.6** | 0 | 0.09 | 0.17 | 0.24 | 0.3 | 0.35 | 0.4 | 0.44 | 0.48 | 0.51 | 0.55 | 0.57 | 0.6 | 0.62 | 0.65 | 0.67 | 0.69 | 0.7 | 0.72 | 0.74 | 0.75 |
| **0.55** | 0 | 0.09 | 0.17 | 0.24 | 0.29 | 0.34 | 0.39 | 0.43 | 0.46 | 0.5 | 0.52 | 0.55 | 0.57 | 0.6 | 0.62 | 0.63 | 0.65 | 0.67 | 0.68 | 0.7 | 0.71 |
| **0.5** | 0 | 0.09 | 0.17 | 0.23 | 0.29 | 0.33 | 0.38 | 0.41 | 0.44 | 0.47 | 0.5 | 0.52 | 0.55 | 0.57 | 0.58 | 0.6 | 0.62 | 0.63 | 0.64 | 0.66 | 0.67 |
| **0.45** | 0 | 0.09 | 0.16 | 0.23 | 0.28 | 0.32 | 0.36 | 0.39 | 0.42 | 0.45 | 0.47 | 0.5 | 0.51 | 0.53 | 0.55 | 0.56 | 0.58 | 0.59 | 0.6 | 0.61 | 0.62 |
| **0.4** | 0 | 0.09 | 0.16 | 0.22 | 0.27 | 0.31 | 0.34 | 0.37 | 0.4 | 0.42 | 0.44 | 0.46 | 0.48 | 0.5 | 0.51 | 0.52 | 0.53 | 0.54 | 0.55 | 0.56 | 0.57 |
| **0.35** | 0 | 0.09 | 0.16 | 0.21 | 0.25 | 0.29 | 0.32 | 0.35 | 0.37 | 0.39 | 0.41 | 0.43 | 0.44 | 0.45 | 0.47 | 0.48 | 0.49 | 0.5 | 0.5 | 0.51 | 0.52 |
| **0.3** | 0 | 0.09 | 0.15 | 0.2 | 0.24 | 0.27 | 0.3 | 0.32 | 0.34 | 0.36 | 0.37 | 0.39 | 0.4 | 0.41 | 0.42 | 0.43 | 0.44 | 0.44 | 0.45 | 0.46 | 0.46 |
| **0.25** | 0 | 0.08 | 0.14 | 0.19 | 0.22 | 0.25 | 0.27 | 0.29 | 0.31 | 0.32 | 0.33 | 0.34 | 0.35 | 0.36 | 0.37 | 0.37 | 0.38 | 0.39 | 0.39 | 0.4 | 0.4 |
| **0.2** | 0 | 0.08 | 0.13 | 0.17 | 0.2 | 0.22 | 0.24 | 0.25 | 0.27 | 0.28 | 0.29 | 0.29 | 0.3 | 0.31 | 0.31 | 0.32 | 0.32 | 0.32 | 0.33 | 0.33 | 0.33 |
| **0.15** | 0 | 0.07 | 0.12 | 0.15 | 0.17 | 0.19 | 0.2 | 0.21 | 0.22 | 0.22 | 0.23 | 0.24 | 0.24 | 0.24 | 0.25 | 0.25 | 0.25 | 0.25 | 0.26 | 0.26 | 0.26 |
| **0.1** | 0 | 0.07 | 0.1 | 0.12 | 0.13 | 0.14 | 0.15 | 0.16 | 0.16 | 0.16 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 |
| **0.05** | 0 | 0.05 | 0.07 | 0.07 | 0.08 | 0.08 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.1 |
| **0** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | **0** | **0.05** | **0.1** | **0.15** | **0.2** | **0.25** | **0.3** | **0.35** | **0.4** | **0.45** | **0.5** | **0.55** | **0.6** | **0.65** | **0.7** | **0.75** | **0.8** | **0.85** | **0.9** | **0.95** | **1** |
| | R | | | | | | | | | | | | | | | | | | | | |

# 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