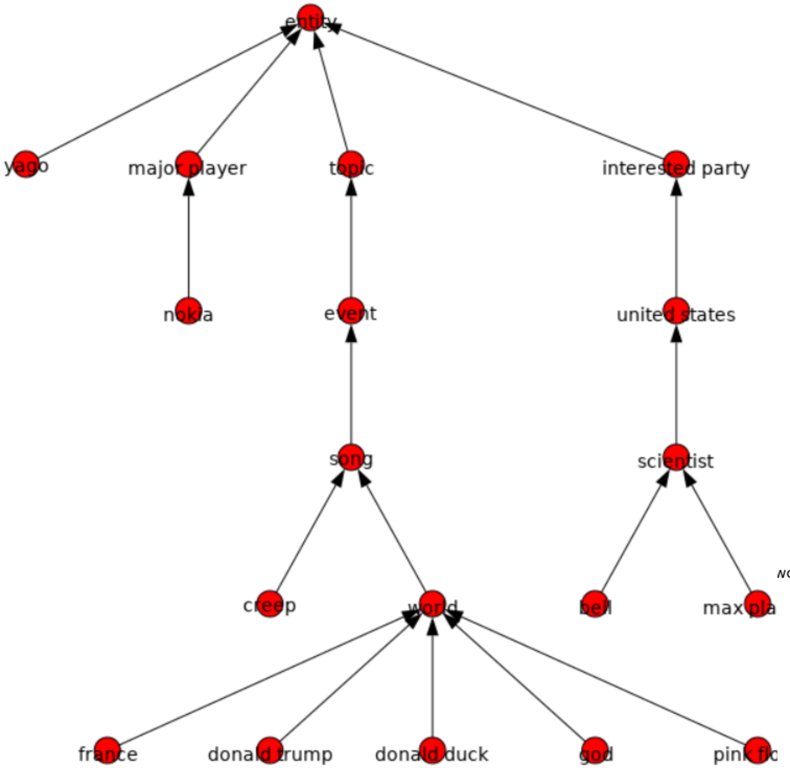
Automated knowledge base construction

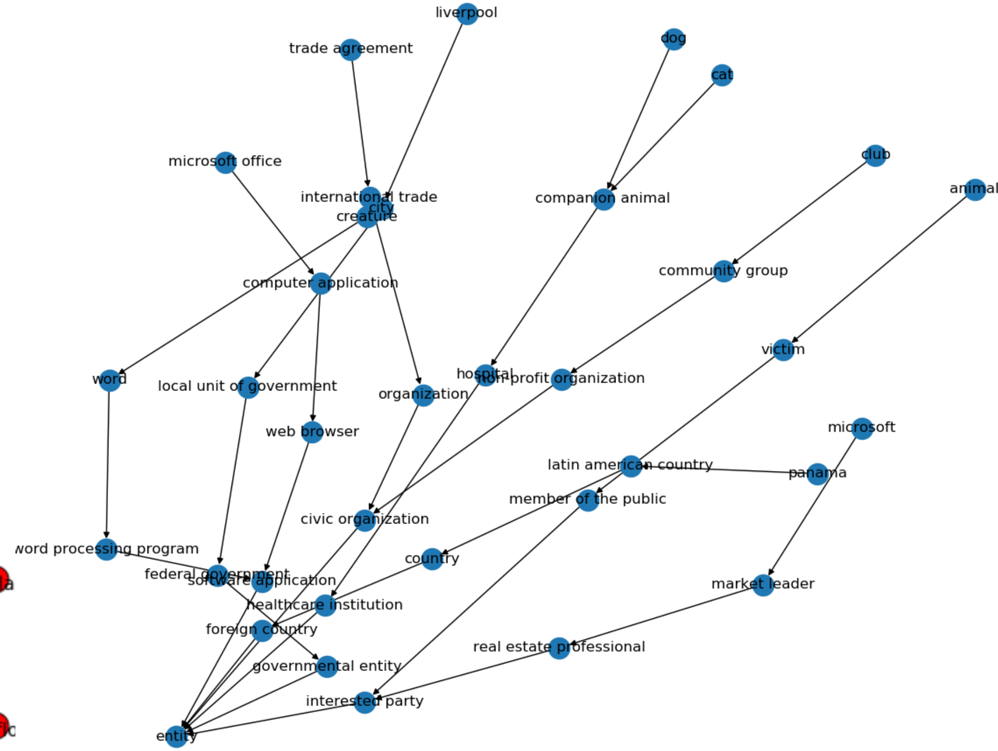
5+6. Relation extraction

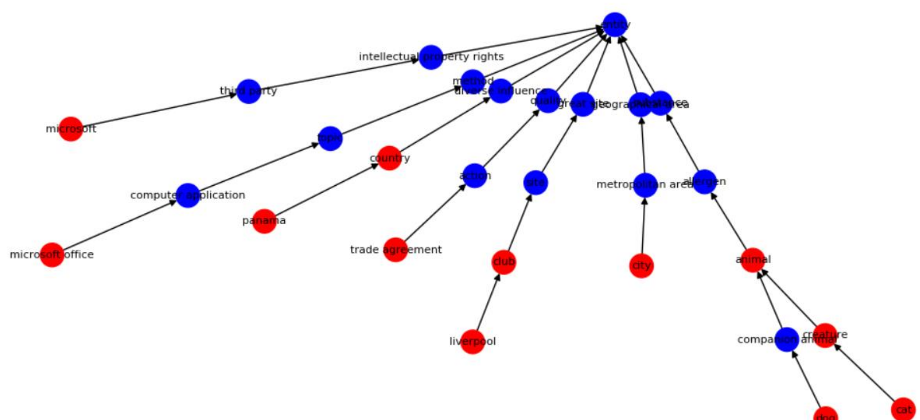
Simon Razniewski

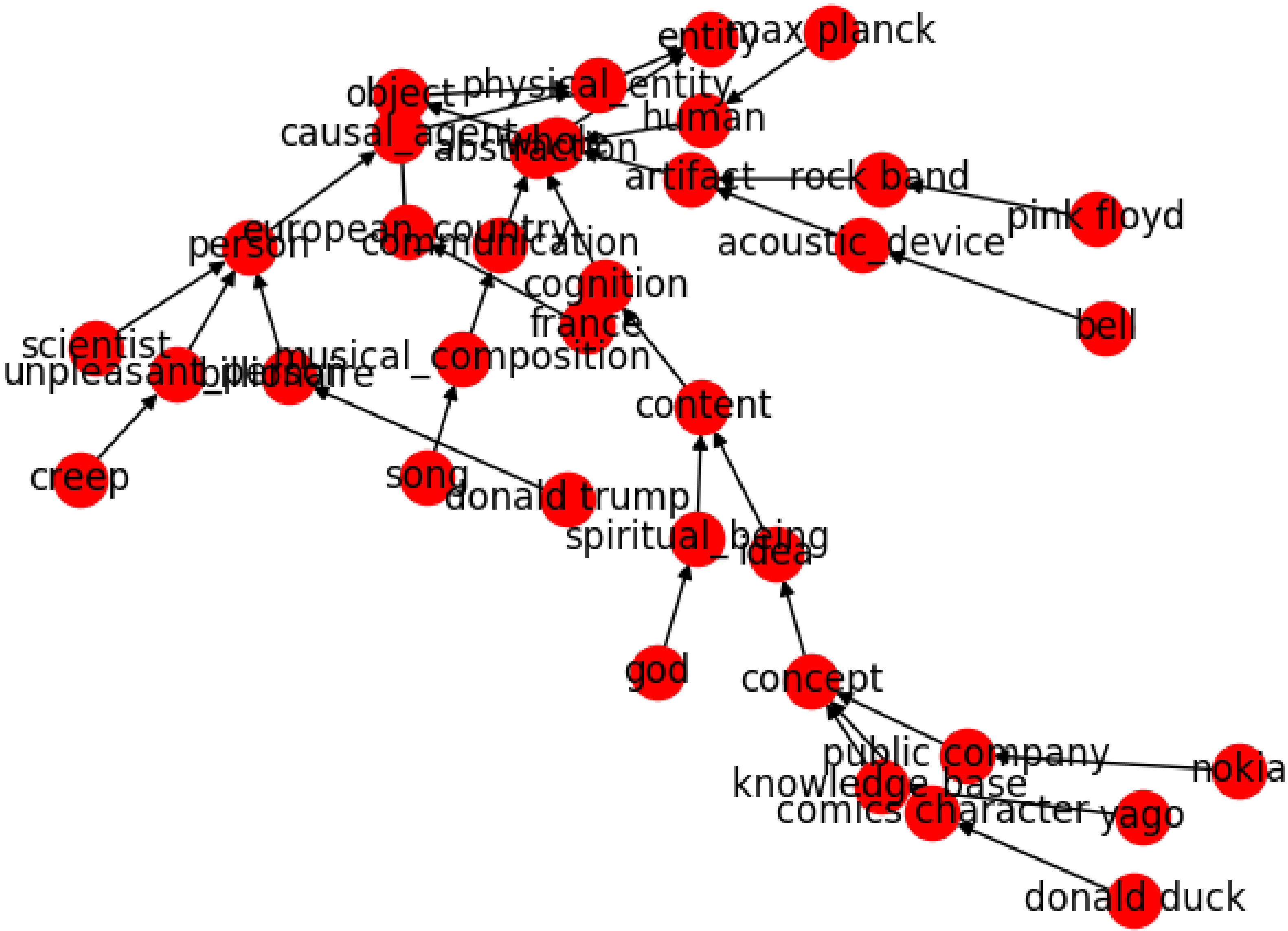
Summer term 2022

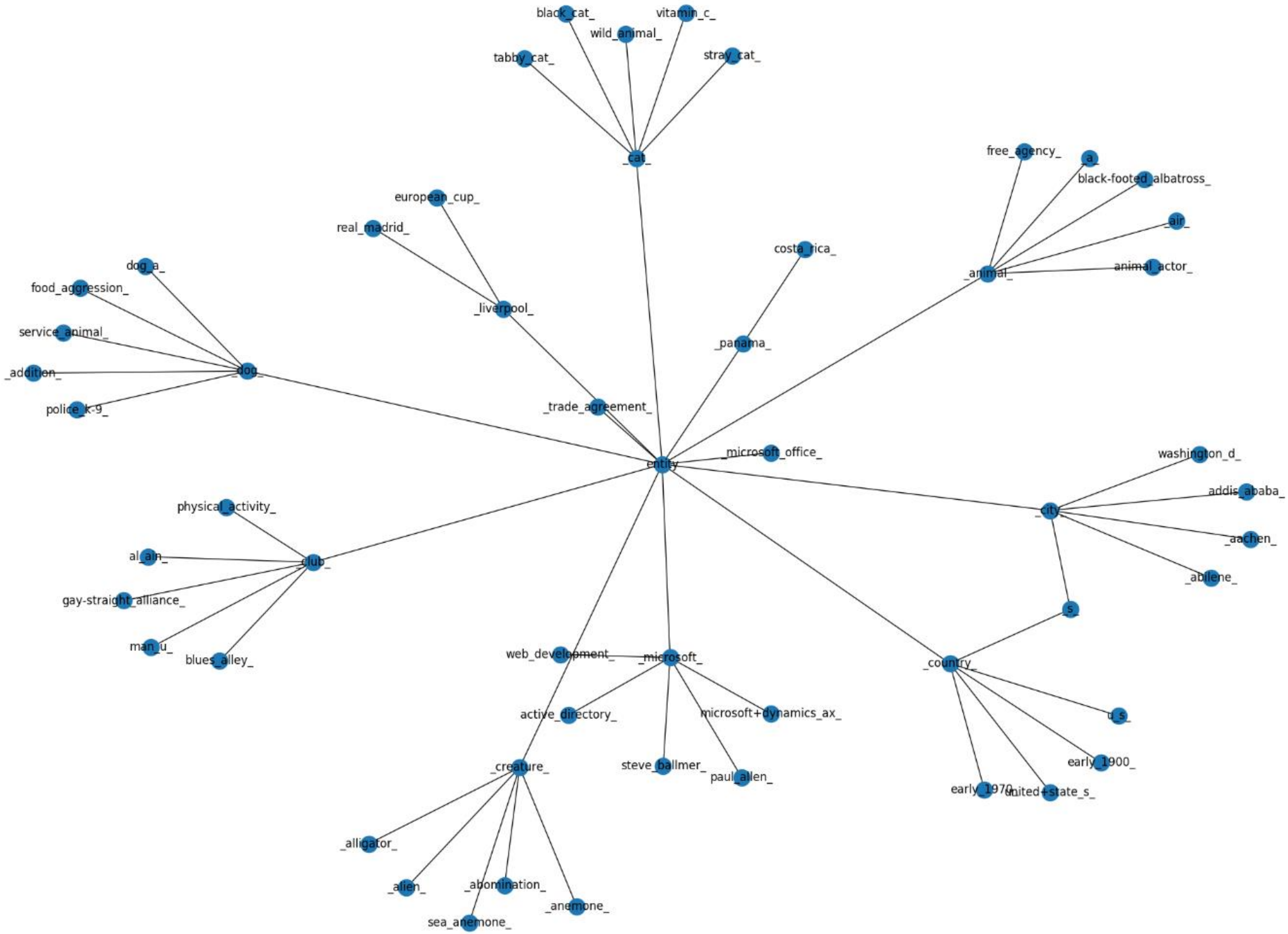
[Start of 6^{th} lecture](#)


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


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


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


entity
max planck
object
physical_entity
human
causal_agent
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

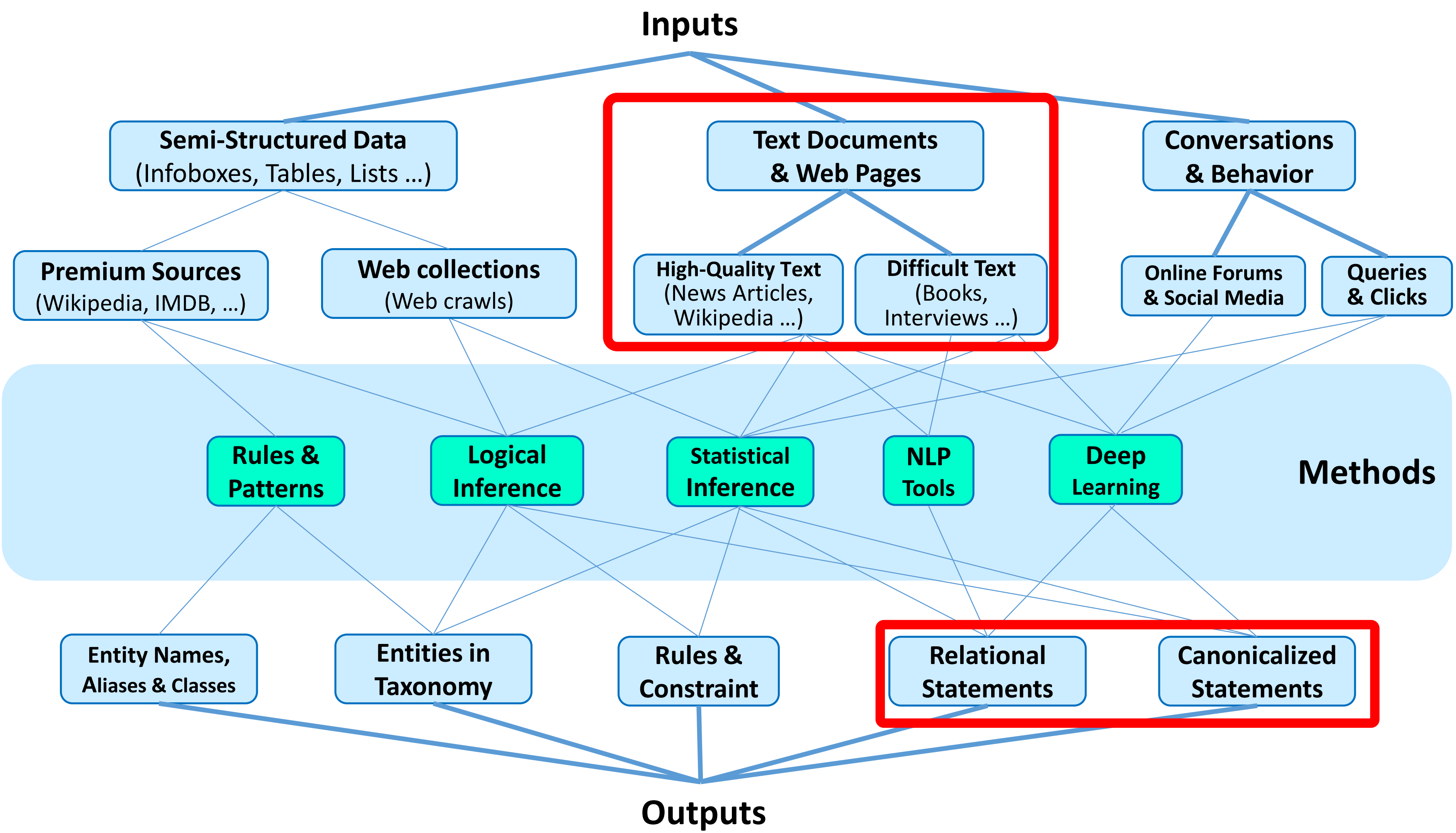

black_cat_
vitamin_c_
wild_animal_
tabby_cat_
stray_cat_
cat
free_agency_
_a_
black-footed_albatross_
air
animal_actor_
animal
european_cup_
real_madrid_
costa_rica_
dog_a_
food_aggression_
liverpool
service_animal_
panama
addition
dog
police_k-9_
_trade_agreement_
_microsoft_office_
entity
washington_d_
addis_ababa_
physical_activity_
city
aachen
al_ain_
club
gay-straight_alliance_
abilene
_s_
man_u_
blues_alley_
web_development_
microsoft
country
active_directory_
microsoft+dynamics_ax_
u_s_
creature
steve_ballmer_
paul_allen_
early_1900_
early_1970_
united+state_s_
alligator
alien
abomination
anemone
sea_anemone_

Outline

1. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Idea
    2. Semantic role labeling and OIE
    3. Organizing open relations

Today

Next week


Inputs
Semi-Structured Data
(Infoboxes, Tables, Lists …)
Text Documents
& Web Pages
Conversations
& Behavior
Premium Sources
(Wikipedia, IMDB, …)
Web collections
(Web crawls)
High-Quality Text
(News Articles,
Wikipedia …)
Difficult Text
(Books,
Interviews …)
Online Forums
& Social Media
Queries
& Clicks
Rules &
Patterns
Logical
Inference
Statistical
Inference
NLP
Tools
Deep
Learning
Methods
Entity Names,
Aliases & Classes
Entities in
Taxonomy
Rules &
Constraint
Relational
Statements
Canonicalized
Statements
Outputs

Fixed-target relation extraction: Task

Given

1. Text *t*
2. Entities *E* in *t*
3. Set of target relations *R*

Output:

- All relational triples (e_1, r, e_2) asserted in *t*

(NER typically a preprocessing step to relation extraction)

Principal approaches

1. Extractive (patterns)
   - If text contains “X is in Y”
   - Then output tuple locatedIn(X, Y)

2. Classification

sibling

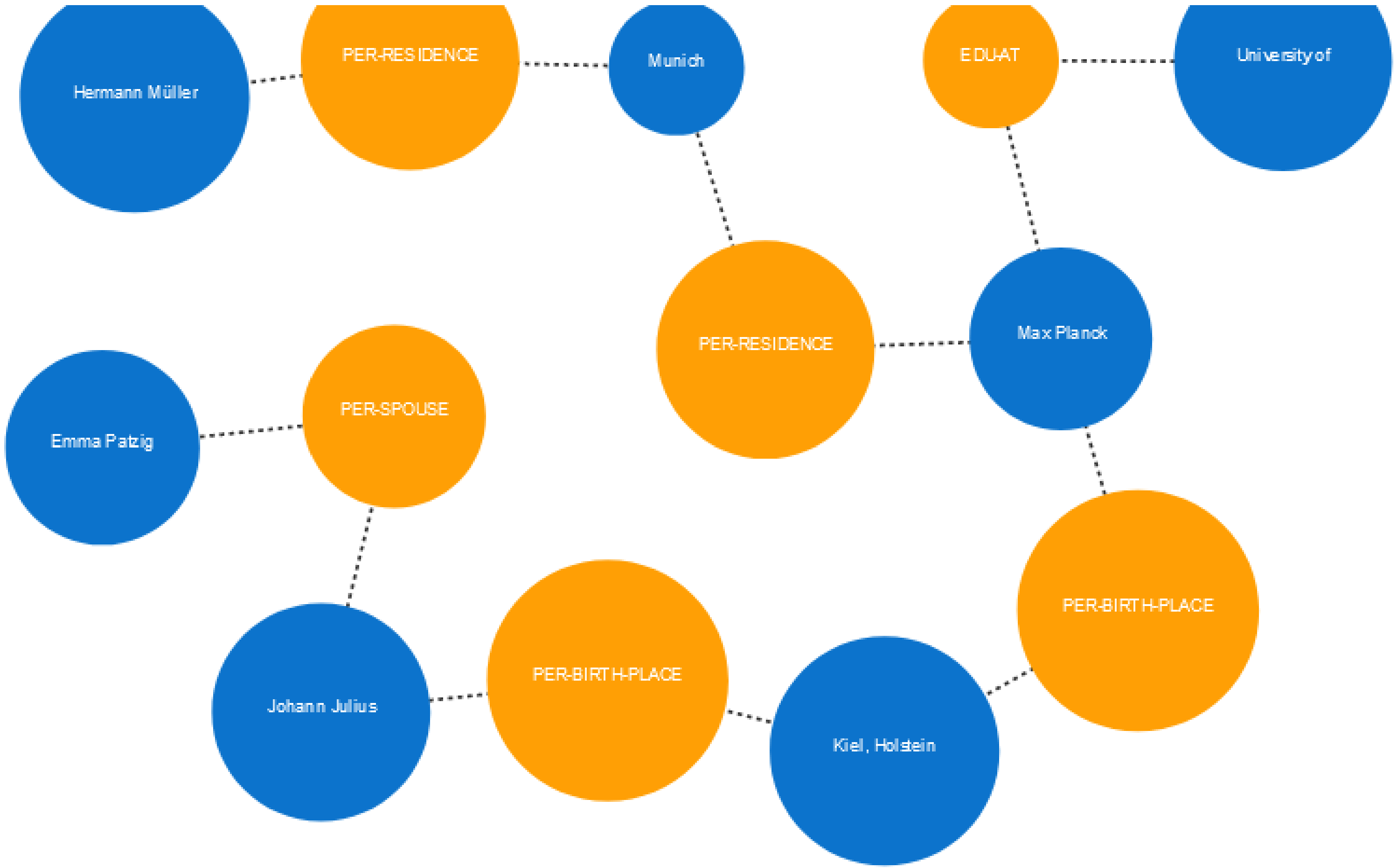
	e_1	e_2	e_3	e_4	e_5	e_6	e_7	…
e_1		T						
e_2								
e_3				T				
e_4								
e_5							T	
e_6				T				
…								

coworker

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	…
e_1		T							
e_2									
e_3	T								
e_4				T					
e_5							T		
e_6		T							
…									

- Filtering: Only pairs within same sentence
- Perform sentence-for-sentence, union (avg) of results

Output: Graph view



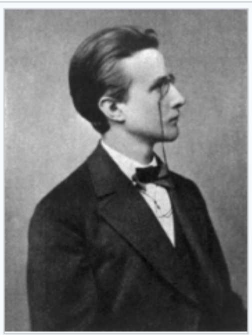
Output: Slot/list view

Born	Max Karl Ernst Ludwig Planck 23 April 1858 Kiel, Duchy of Holstein
Died	4 October 1947 (aged 89) Göttingen, Lower Saxony, Bizone, Allied-occupied Germany
Education	PhD in theoretical physics, Ludwig Maximilian University of Munich, 1879.
Alma mater	Ludwig Maximilian University of Munich
Known for	See full List
Spouse(s)	Marie Merck (m. 1887; died 1909) Marga von Hösslin (m. 1911)
Children	5
Awards	Nobel Prize in Physics for his quantum theory (1918) Foreign Associate of the National Academy of Sciences (1926) Lorentz Medal (1927) Copley Medal (1929) Max Planck Medal (1929) Goethe Prize (1945)
	Scientific career
Fields	Physics
Institutions	University of Kiel University of Göttingen Kaiser Wilhelm Society

Extracting Relation Triples from Text

gymnasium school, where he came under the tutelage of Hermann Müller, a mathematician who took an interest in the youth, and taught him astronomy and mechanics as well as mathematics. It was from Müller that Planck first learned the principle of conservation of energy. Planck graduated early, at age 17.[9] This is how Planck first came in contact with the field of physics.

Planck was gifted when it came to music. He took singing lessons and played piano, organ and cello, and composed songs and operas. However, instead of music he chose to study physics.



A side portrait of Planck as a young adult, c. 1878

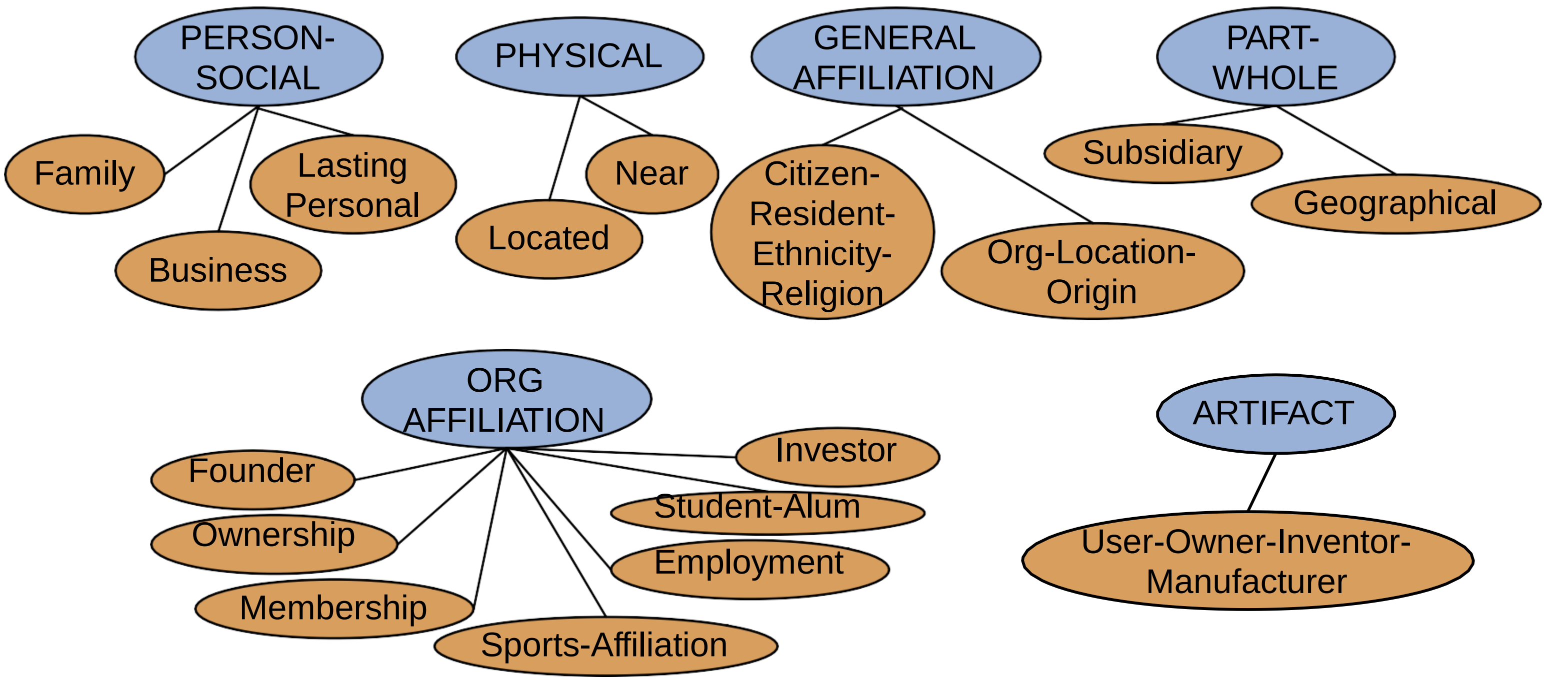
The Munich physics professor Philipp von Jolly advised Planck against going into physics, saying, "In this field, almost everything is already discovered, and all that remains is to fill a few holes."[10] Planck replied that he did not wish to discover new things, but only to understand the known fundamentals of the field, and so began his studies in 1874 at the University of Munich. Under Jolly's supervision, Planck performed the only experiments of his scientific career, studying the diffusion of hydrogen through heated platinum, but transferred to theoretical physics.

In 1877, he went to the Friedrich Wilhelms University in Berlin for a year of study with physicists Hermann von Helmholtz and Gustav Kirchhoff and mathematician Karl Weierstrass. He wrote that Helmholtz was never quite prepared, spoke slowly, miscalculated endlessly, and bored his listeners, while Kirchhoff spoke in carefully prepared lectures which were dry and monotonous. He soon became close friends with Helmholtz. While there he undertook a program of mostly self-study of Clausius's writings, which led him to choose thermodynamics as his field.

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

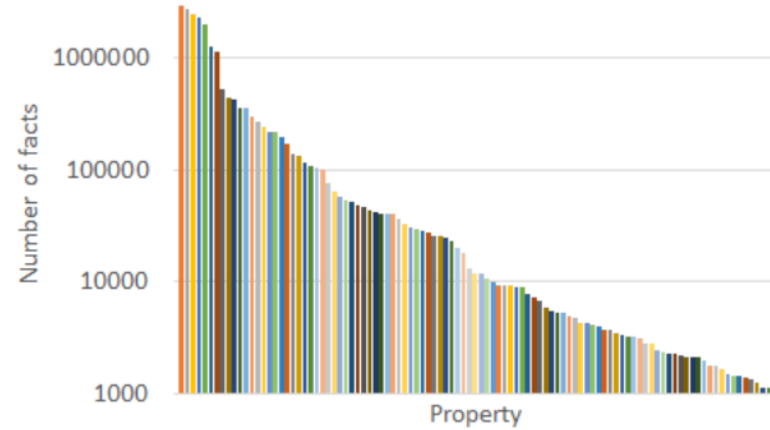
  Steve Jobs, co-founder of Apple…

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

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. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Idea
    2. Semantic role labeling and OIE
    3. Organizing open relations

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_{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

- Pro
  - Human patterns tend to be high-precision
  - Can be tailored to specific domains
- Contra
  - 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

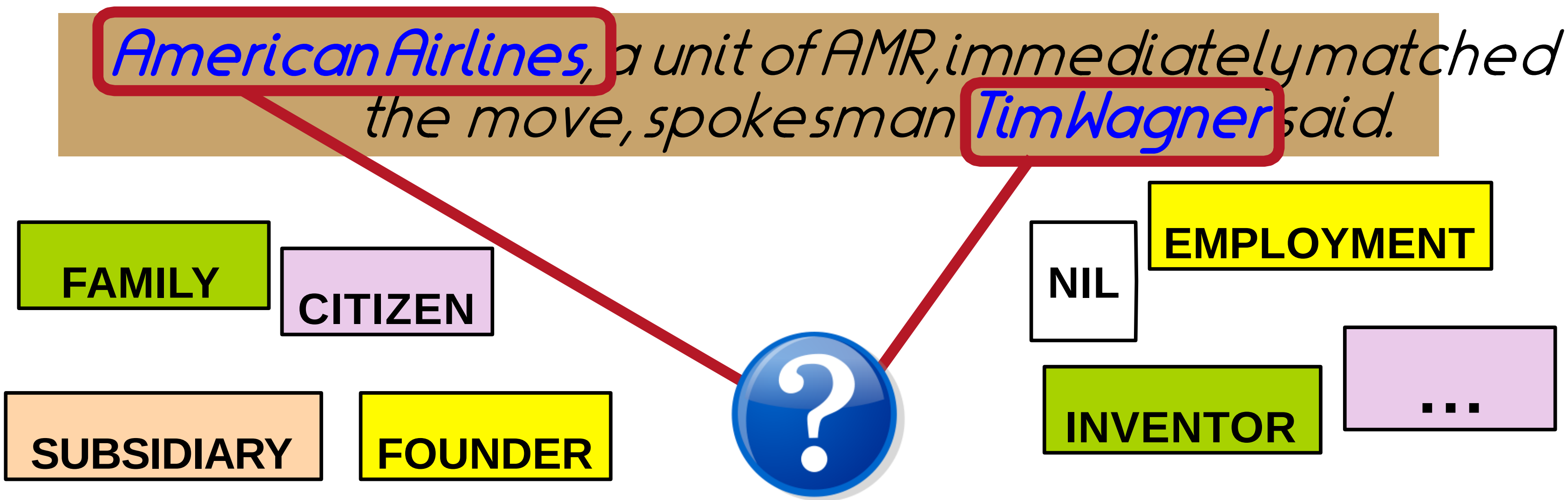
1. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Idea
    2. Semantic role labeling and OIE
    3. Organizing open relations

Supervised ML for relation extraction

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

Relation Extraction via classification

Classify the relation between two entities



Word Features for Relation Extraction

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
- Entity Level of M1 and M2 (NAME, NOMINAL, PRONOUN)
  - M1: NAME
  - M2: NAME

[it or he would be PRONOUN]

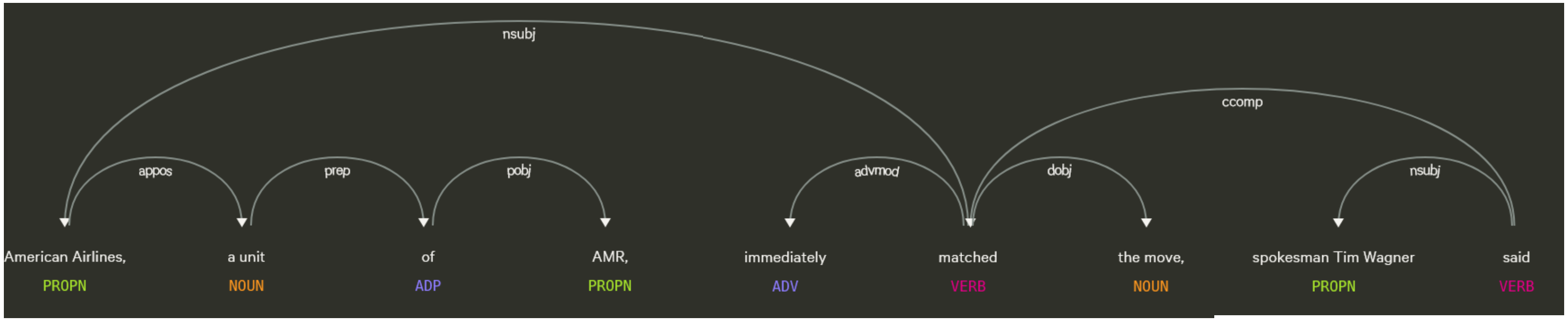
[the company would be NOMINAL]

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

https://explosion.ai/demos/displacy?



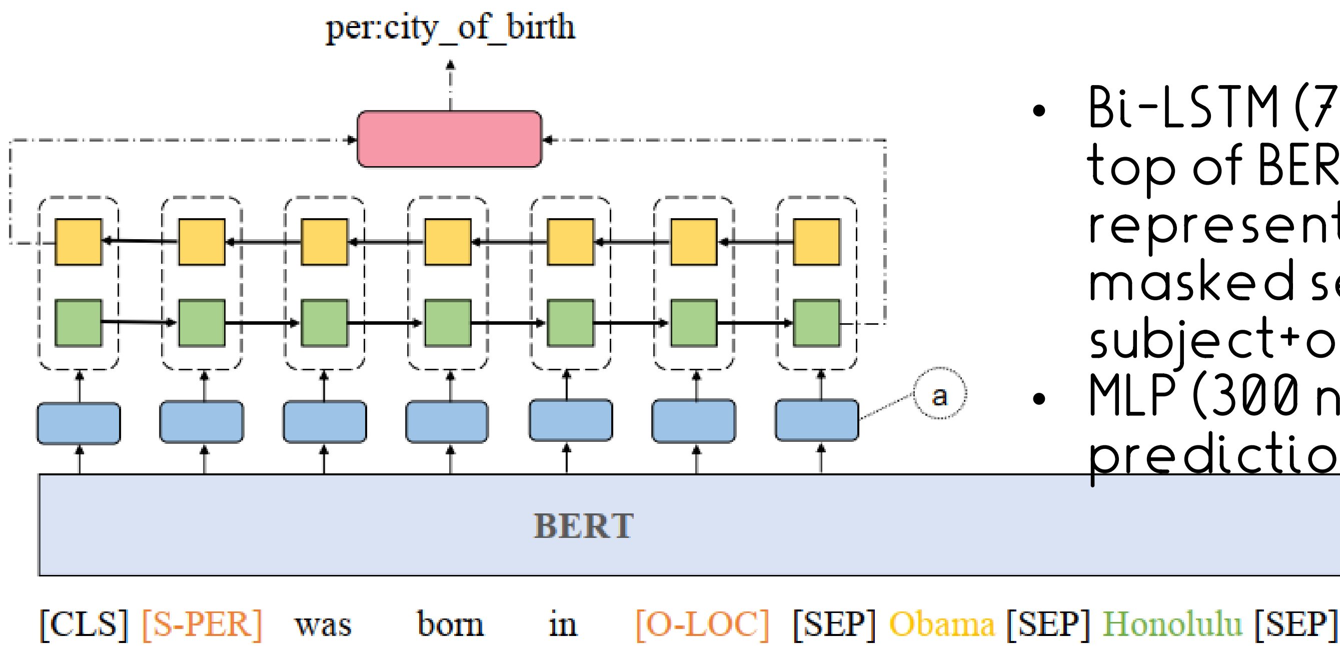
Dictionaries and trigger word features for relation extraction

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

Evaluation of supervised relation Extraction

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

Relation extraction using BERT



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

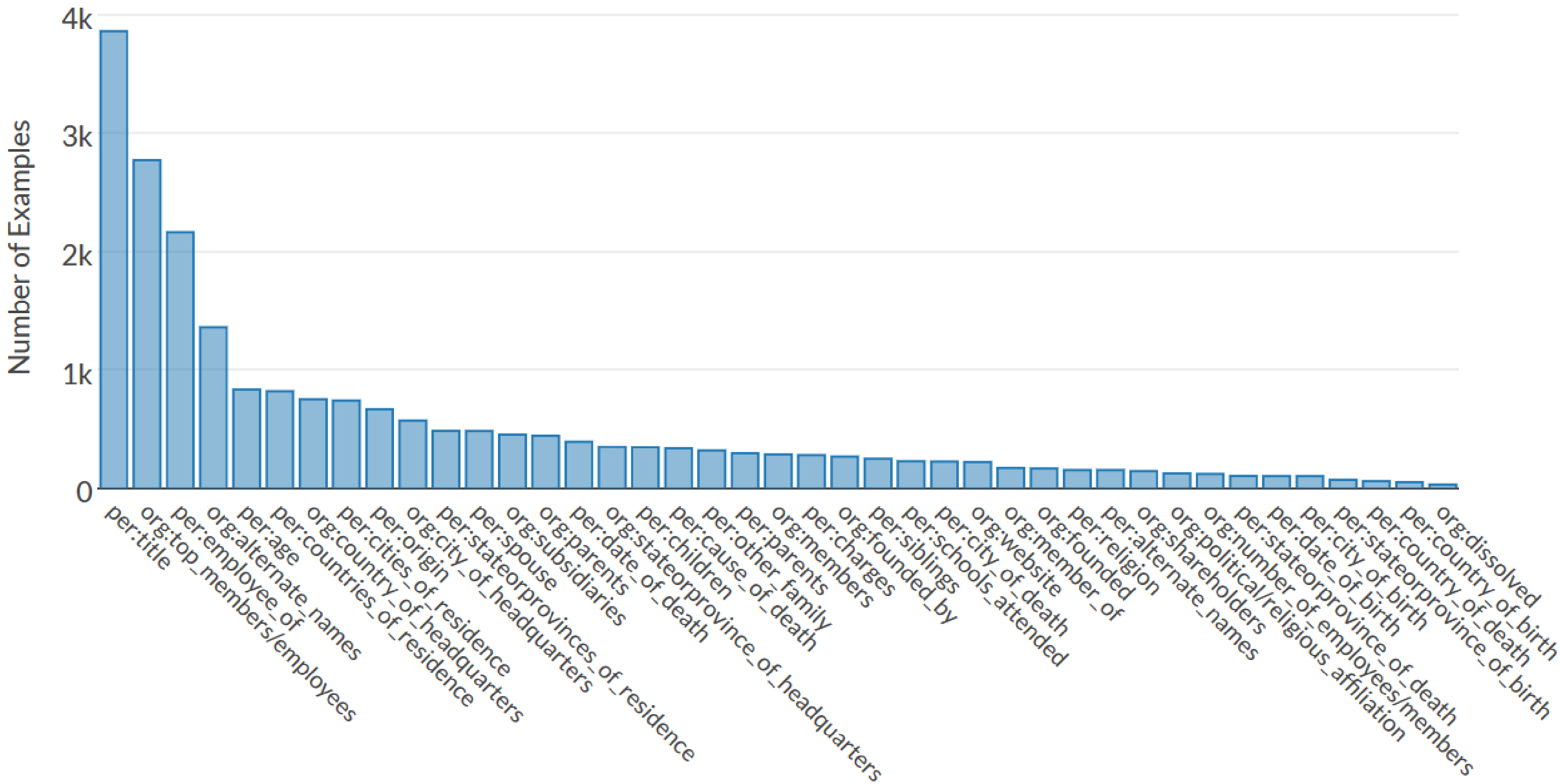
Model	P	R	F_1
Zhang et al. (2017)	65.7	64.5	65.1
Zhang et al. (2018)	69.9	63.33	66.4
Wu et al. (2019)	-	-	67.0
Alt et al. (2019)	70.1	65.0	67.4
BERT-LSTM-base	**73.3**	63.10	67.8
Zhang et al. (2018) (ensemble)	71.3	**65.4**	**68.2**

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

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]

Summary: Supervised relation extraction

Pro

- Can get high precision/recall with enough training data, if test similar enough to training

Contra

- Labeling a large training set is expensive
- Supervised models are still brittle, don't generalize well to different genres

Outline

1. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Idea
    2. Semantic role labeling and OIE
    3. Organizing open relations

Seed-based or bootstrapping approaches to relation extraction

- No training set? Maybe you have:
  - A few seed tuples
- Can you use those seeds to do something useful?
  - Bootstrapping: use the seeds to directly learn to populate a relation
- Related to self-supervised learning, label propagation, etc.
- Underlying assumption: High-confidence predictions/patterns are likely correct

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."

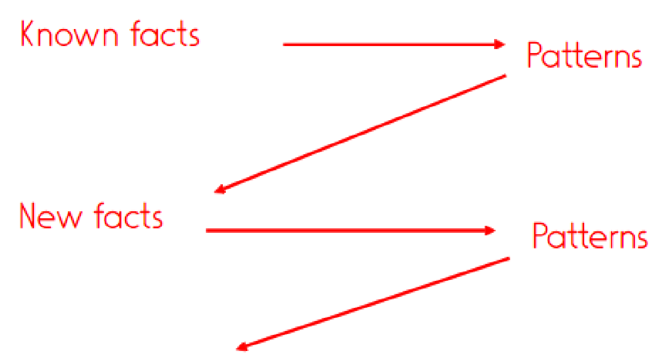
  X is buried in Y

  "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

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/Ĩmann/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.

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. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Idea
    2. Semantic role labeling and OIE
    3. Organizing open relations

Distant Supervision

- Combine bootstrapping with supervised learning
  - Instead of 5 seeds,
    - Use a large database to get huge # of noisy 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 scores 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

Exercise: Distant supervision

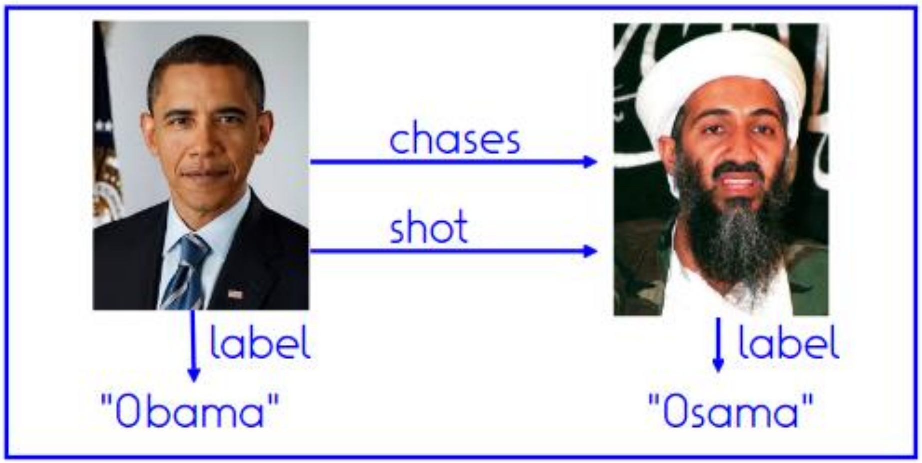
KB (bornIn)
(Einstein, Ulm)
(Curie, Warsaw)

Text:
1. Einstein was born in Ulm.
2. Curie migrated from Warsaw.
3. Researchers claim: Was Einstein born in France?
4. Einstein and the Unified Modelling language (ULM).
5. Ulm was home to many famous people, including Hoeneß and Einstein.

Task: With distant supervision, what would be the positive training examples (sentences) for a bornIn relation classifier?

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

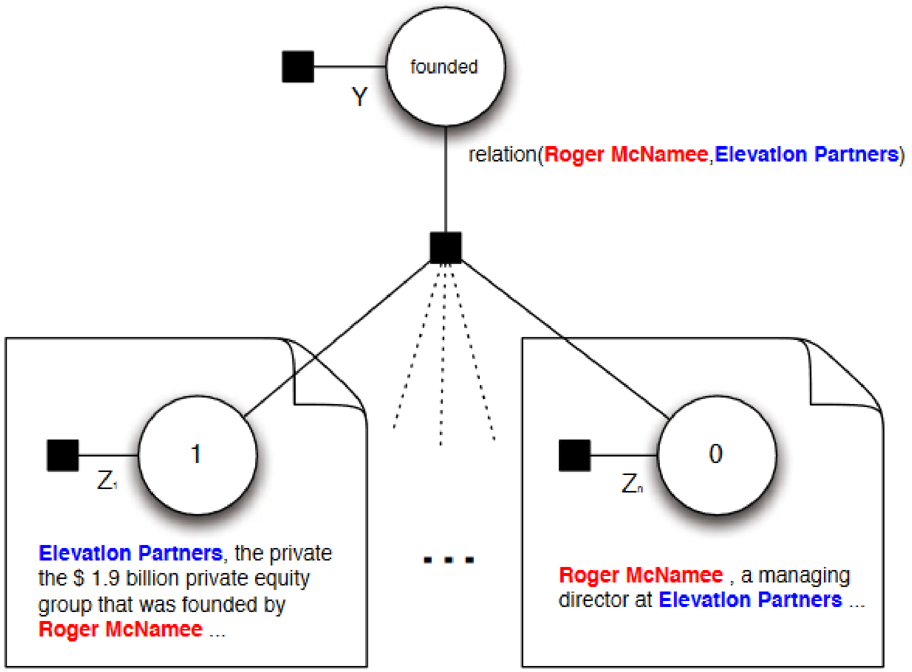
Relation Type	New York Times	Wikipedia
nationality	38%	20%
place_of_birth	35%	20%
contains	20%	10%

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. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Idea
    2. Semantic role labeling and OIE
    3. Organizing open relations

CINEX [Mirza et al., 2018]

- Instructive example of distant supervision with cleaning
- 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>

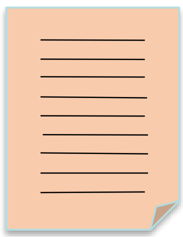
  <California, hasCounty, ∃58>   <Donald Trump, hasSpouse, ∃3>

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

Problem: CQ Extraction

Given

S



$P_{hasChild}$

Determine

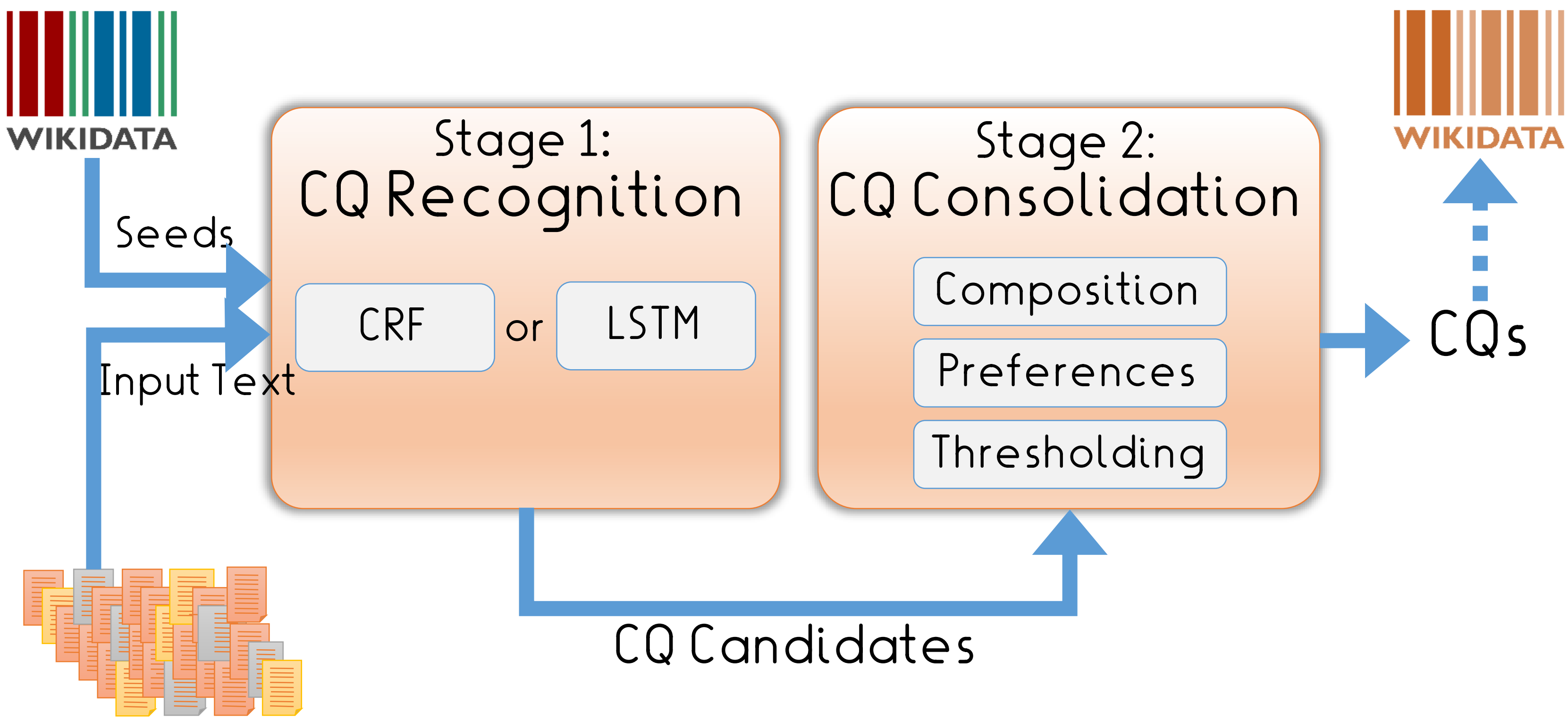
$\exists 6$

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'

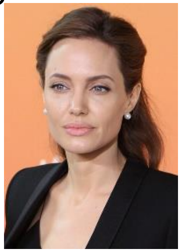
- Compositionality
  - In 2016, Jolie brought her twins, one daughter and three adopted children to the gala.

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

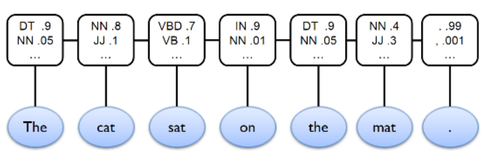
- In 2016, Jolie brought her twins, one daughter and three adopted children to the gala.

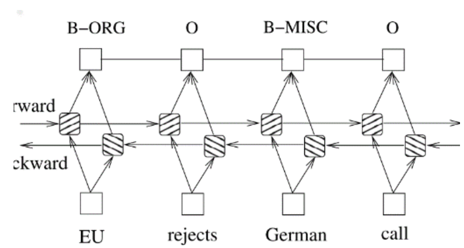
preprocessing

...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

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

S

preprocessing

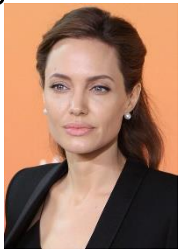
…her	twins	,	one	daughter	and	three	adopted	children	to…
…her	NUMTERM	,	CARDINAL	daughter	and	CARDINAL	adopted	children	to…
O	COUNT	COMP	COUNT	O	COMP	COUNT	O	O	O

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

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

$\exists 6$

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

Train/Test data size

Wikidata Subject Class	Wikidata Property	Relation	#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)

	containsWork	hasMember	containsAdmin	hasChild	hasSpouse
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%)

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 5

- 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

- Approaches:
  - Extraction
  - Classification

- Key methodological ingredients:
  - Iterative pattern learning
    - Repeat statement extraction and pattern learning with increasing sets (“snowballing”)
  - Distant supervision
    - Scale training data by skipping on hand-labelling of sentences
    - Automatically label sentences from KB statements

Outline

1. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Idea
    2. Semantic role labeling and OIE
    3. Organizing open relations

Design, implementation, comments:

1. Extracting Date of Birth: function extractDoB

- Design

Given our restricted domain of Wikipedia abstracts, it was surprisingly straightforward to achieve an f1 score of ~80% just by extracting the very first date in the abstract.

- Implementation

The function uses a regex (dateMatcher regex ref: https://stackoverflow.com/questions/51122413/) in order to extract the date and returns it in the right format.

- Comments

This method is admittedly crude, and it can be further improved by using either text extracted in parentheses right after the entity mention and/or look for the keyword 'born' followed by the date.

2. Extracting Nationality: function extractNationality

- Design

It was observed that most entities are mentioned with their nationalities such as 'Wayne A. Hendrickson (born April 25, 1941, New York City) is an American biophysicist and University professor at Columbia.' which was matched.

In case that returns no candidate, the verb 'born' is looked for in the abstract and when found, it's prepositional objects are extracted. Those objects that are in fact dates such as 'born in __1955__' are discarded and the rest are returned.

- Implementation

Dependency parsing and ner using spacy.

- Comments

Most nationalities appearing are of demonyms, and the expected nationality (loosely) are country names, a dict of demonym-country has been constructed using data provided in the following link: https://github.com/knowitall/chunkedextractor/blob/master/src/main/resources/edu/knowitall/chunkedextractor/demonyms.csv. Credits: [redacted] for having discussed it on the IE1920 forum.

3. Extracting alma mater: function extractAlmaMater
- Design
    The function looks for the following patters:
        studied <something> at <alma_mater>
        attended <alma_mater>
        [was] obtained/received/awarded/gained/earned/complete/graduated/educated <something> from/at <alma_mater>
    and just extracts the alma maters if 'alma_mater' is at least one among 'university', 'school', 'college', 'academy', or 'gynmasium'.
- Implementation
    POS tagging, dependency parsing and ner using spacy.

4. Extracting places of work: function extractWorkPlace
- Design
    This turned out to be quite the challenge with a morass of exceptions. Hence the function takes an overly simplifying approach of extracting all of the organizations mentioned in the abstract apart from alma maters and returns.

5. Extracting awards: extractAwards
- Design
    Looks for verbs 'won' and 'awarded' and returns the objects.
    In order to improve recall, this function makes the assumption that most awards mentioned in the abstract probably belong to the entity in question and hence extracts all of them using a regex that matches 'prize', 'award', 'medal' and returns. The first rule compensates for all those awards that don't get matched by the regex such as 'Spinozapremie'.
- Implementation
    Dependency parsing, ner, regex matching

General comments
There seems to be an upper bound on the scores as the ground truth itself is quite noisy.
It is observed that for this restricted domain, given enough time, manual pattern matching can indeed return good enough results, there aren't too many exceptions to warrant a statistical models.

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.

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

Method output: {Homer, Bart, Groening}

✓ ✓ ✗

Gold standard: {Homer, Bart, Lisa}

=> 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}

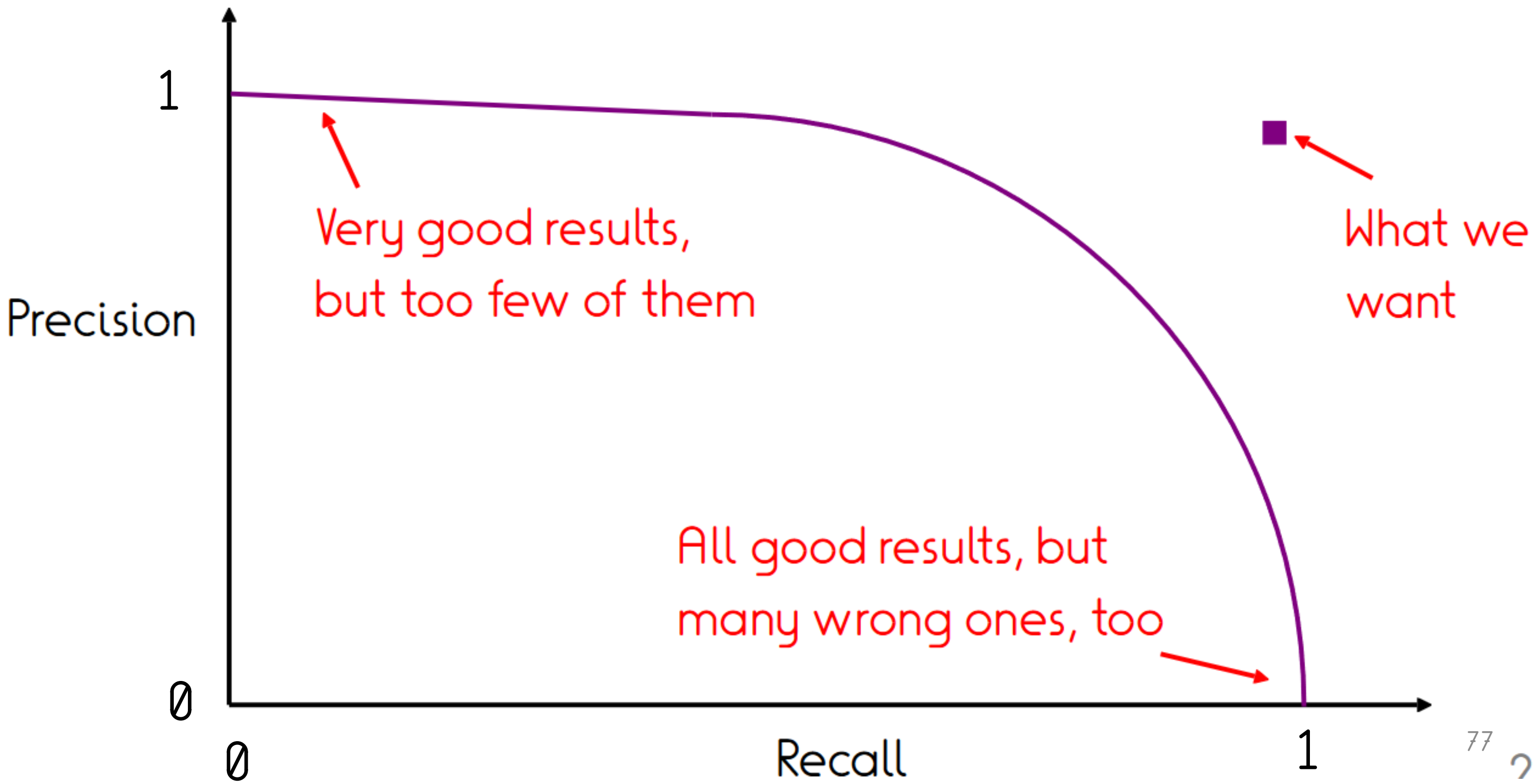
Gold standard: {Homer, Bart, Lisa}

✓ ✓ ✗

=> Recall: 2/3 = 66%

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 obtain a single score for ranking systems, we could 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?

Dive deeper at home

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

- Precision/recall tradeoff in automated courtroom decision making

Def: Problem of imbalanced classes

Population: {Snowball_1,..., Snowball_99, Snowball_100}
Gold Standard: {Snowball_1,..., Snowball_99}
Output: {Snowball_1,..., Snowball_99, Snowball_100}
Precision: 99/100=99%
Recall: 99/99=100%

If there are very few negatives, just outputting all elements gives great scores.

The problem of imbalanced classes appears when only very few of the items of the population are not in the gold standard: An approach that outputs the entire population has a very high precison and a perfect recall. (Ex2: Citizenship on en-Wikipedia)

The negatives are the elements of the population that are not in the gold standard.

Def: Confusion Matrix

Population: {Snowball_1,..., Snowball_99, Snowball_100}

Gold Standard: {Snowball_1,..., Snowball_99}

Output: {Snowball_1,..., Snowball_99, Snowball_100}

The confusion matrix for the output of an algorithm looks as follows:

Items of the population that are not in the gold standard

		Gold standard: Positive	Gold standard: Negative	Σ
Output	Positive	True Positives	False Positives	Predicted Positives
Output	Negative	False Negatives	True Negatives	Predicted Negatives
	Σ	(Gold) Positives	(Gold) Negatives	

Items of the population that are not output

"Negative" because it was not output, "True" because that was correct.

Def: Confusion Matrix

Population: {Snowball_1,..., Snowball_99, Snowball_100}
Gold Standard: {Snowball_1,..., Snowball_99}
Output: {Snowball_1,..., Snowball_99, Snowball_100}

The confusion matrix for the output of an algorithm looks as follows:

1 item was output as positive, but was negative in the gold standard

		Gold standard		
		Positive	Negative	
Output	Positive	99	1	100
	Negative	0	0	0
		99	1	

Precision = true positives / predicted positives = 99/100 = 99%
Recall = true positives / gold positives = 99/99 = 100%

Confusion with confusion matrixes

A confusion matrix does not always make sense in an information extraction scenario:

Population:	{A, B, ..., Aa, Ab, ..., Aaa, ...}
Gold Standard:	{Homer}
Output:	{Homer}

		Gold standard	
		Positive	Negative
Output	Positive	1	0
	Negative	0	39462440205

A confusion matrix makes sense only when the population is limited (e.g., in classification tasks)!

Our problem

Population:	{Snowball_1,..., Snowball_99, Snowball_100}
Gold Standard:	{Snowball_1,..., Snowball_99}
Output:	{Snowball_1,..., Snowball_99, Snowball_100}

		Gold standard	
		Positive	Negative
Output	Positive	99	1
	Negative	0	0

The problem is that the algorithm did not catch the negatives, it has a "low recall" on the negatives.

Def: True Negative Rate & FPR

Population: {Snowball_1,..., Snowball_99, Snowball_100}
Gold Standard: {Snowball_1,..., Snowball_99}
Output: {Snowball_1,..., Snowball_99, Snowball_100}

The true negative rate (also: TNR, specificity, selectivity) is the ratio of negatives that are output as negatives (= the recall on the negatives):

TNR = true negatives / gold negatives = 0 / 1 = 0%

		Positive	Negative
Output	Positive	99	1
	Negative	0	0

The False Positive Rate (also: FPR, fall-out) is 1-TNR.

TNR & Precision

Population: {Snowball_1,..., Snowball_99, Snowball_100}

Gold Standard: {Snowball_1,..., Snowball_99}

Output: {Snowball_1,..., Snowball_99, Snowball_100}

Precision: 99/100=99%

Recall: 99/99=100%

TNR: 0/1=0%

TNR and precision both measure the "correctness" of the output.

Precision:

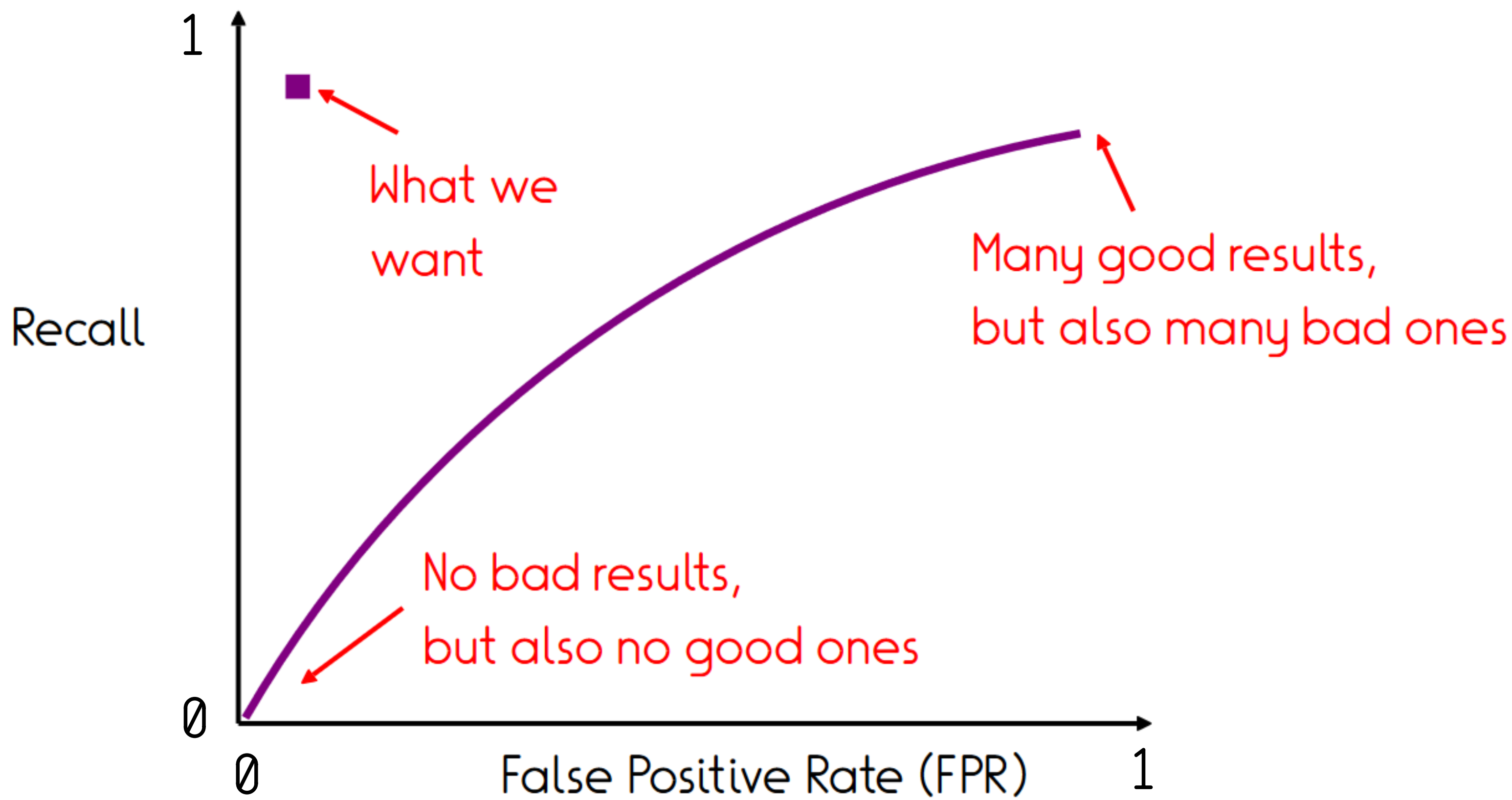
- measures wrt. the output
- suffers from imbalanced classes
- works if population is infinite (e.g., set of all extractable entities)

TNR:

- measures wrt. the population
- guards against imbalance
- works if population is limited (e.g., in classification)

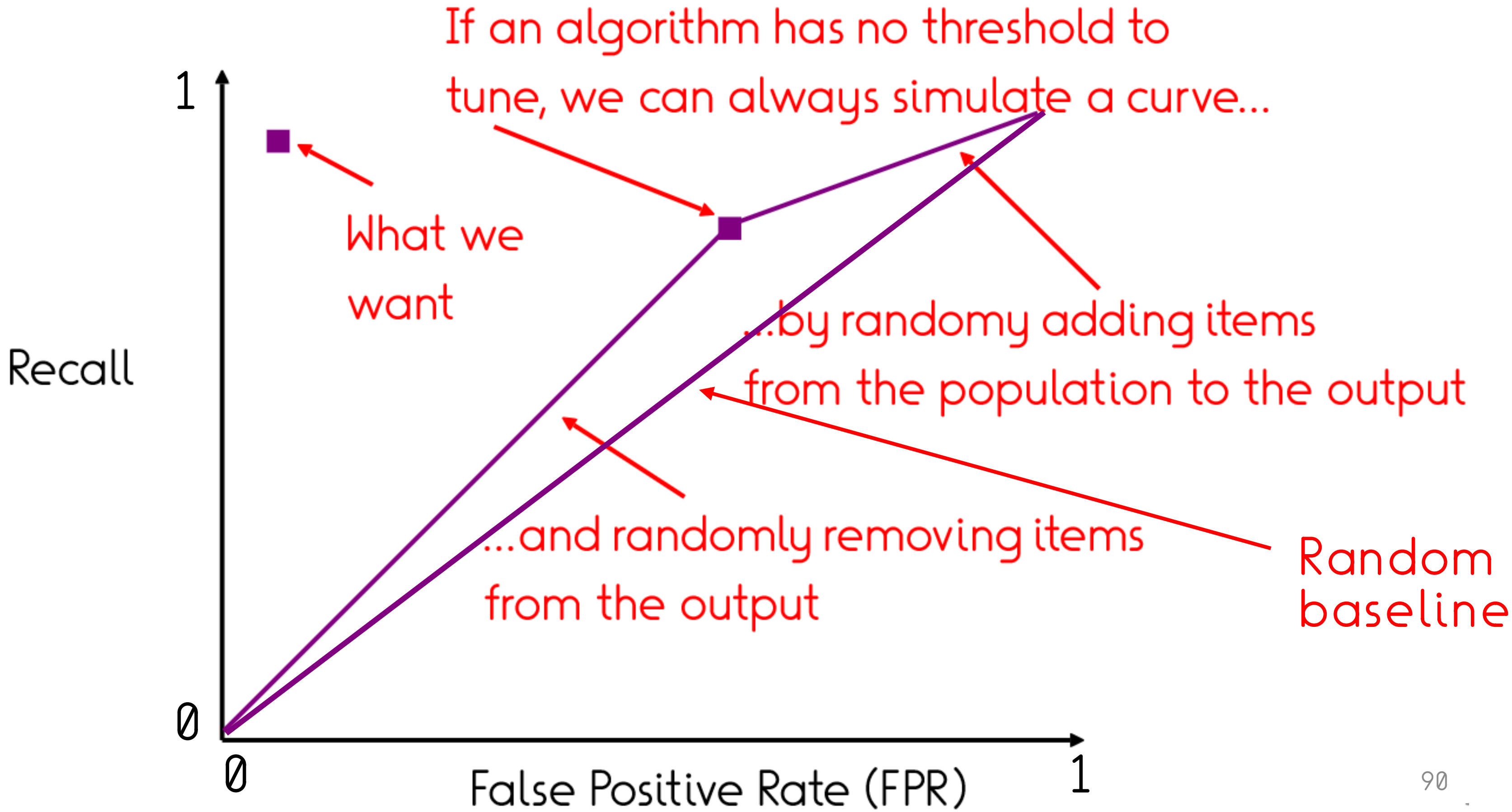
Def: ROC

The ROC (receiver operating characteristic) curve plots recall against the FPR for different thresholds of the algorithm. It guards against imbalanced classes, and is applicable when the population is finite.



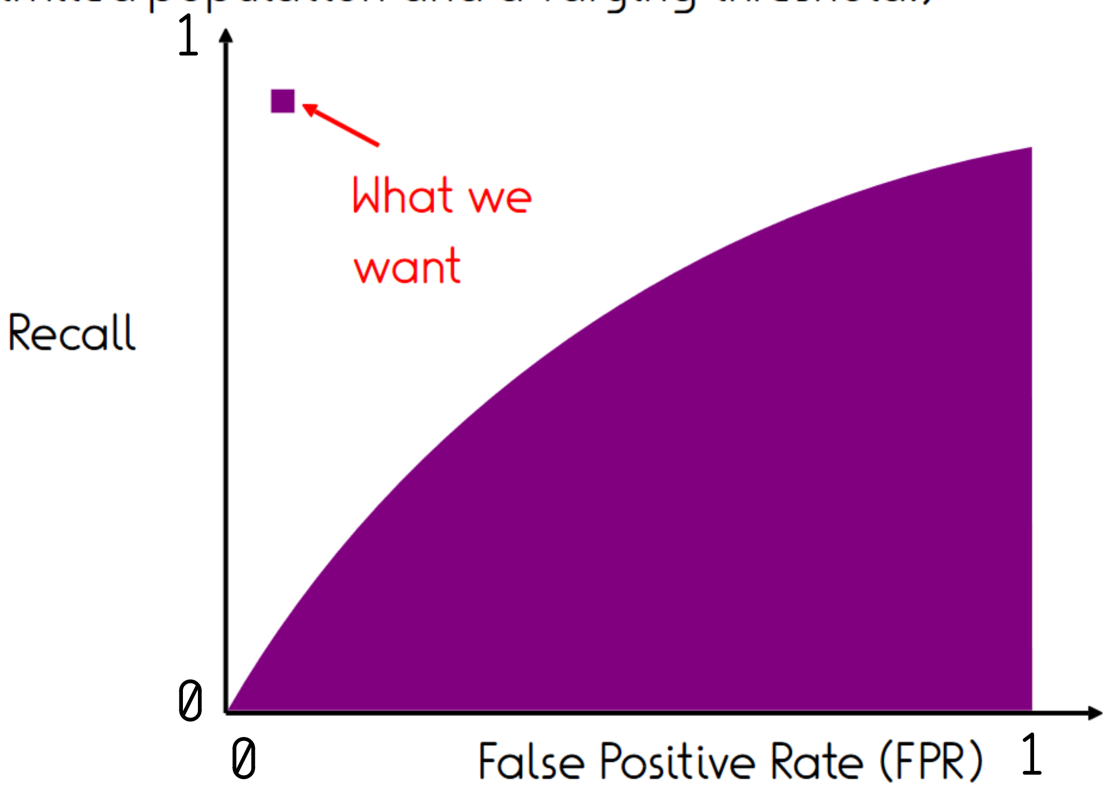
Def: ROC

The ROC (receiver operating characteristic) curve plots recall against the FPR for different thresholds of the algorithm. It guards against imbalanced classes, and is applicable when the population is finite.



Def: AUC

The AUC (area under curve) is the area under the ROC curve.
It corresponds to the probability that the classifier ranks a random positive item over a random negative item. (It's kind of the F1 for a limited population and a varying threshold.)



(AUC measure for PR curves also exists, but has no corresponding probabilistic interpretation)

Def: Micro vs. Macro averaging

- 3 relations (A, B, C)
- Predictions:
  - 10x A (90% correct)
  - 10x B (90% correct)
  - 100x C (10% correct)
- Micro-avg. precision: $\frac{10x0.9+10x0.9+100x0.1}{10+10+100} = 0.23$
- Macro-avg. precision: $\frac{0.9+0.9+0.1}{3} = 0.63$
- Recall and F1 analogous

→ Micro – each instance counts same
→ Macro – each class counts same

Interpreting scores: Baselines and yardsticks

- Method precision 0.63, recall 0.47
  - Is this good?

- Baselines
  - Random!
  - Most frequent class!
  - Naive heuristics
    - Trigger word lookup, first noun, 5th word, etc.

- Yardsticks
  - Existing systems
  - Human performance (agreement)
    - (in certain tasks e.g. in vision not a yardstick anymore)

Error analysis (1/3)

- Method: P 0.63 R 0.47
- Baseline: P 0.55 R 0.30
- Humans: P 0.85 R0.90

- What went wrong?
  - Sample a few errors (false positives and false negatives)
  - Define categories of errors
  - Sample a larger set of errors
  - Count frequencies of error categories
  - Possibly iterate

- Severity of errors?

- Important for
  - Pinpointing component in pipeline (NER, NED, RE, …)
  - Yourself to improve
  - The next one continuing your concrete work
  - Others to understand potential and limits of your approach

- Error meta-categories
  - Limit of effort
    - Effort-performance-derivation/extrapolation? Wrt. time or training data size
  - Limits of methodology
  - Limit of data/metric (next)

Task: Error analysis categories

Text	Ground truth	Extracted
Mary lives in Chicago	livesIn(Mary_Smith, Chicago_USA)	bornin(Mary_Smith, Chicago_USA)
		livesIn(Chicago_USA, Mary_Smith)
		livesIn(Mary_Smith, Chicago_Kenya)
John works for Procter and Gamble	worksFor(John, ProcterAndGamble_cpy)	worksFor(John, Procter_cpy)
		affiliatedWith(John, ProcterAndGamble_cpy)
		ceoOf(John, ProcterAndGamble_cpy)
Mary lives in Mannheim, right next to Ludwigshafen	livesIn(Mary, Mannheim)	livesIn(Mary, Ludwigshafen)

Error analysis (2/3) – Question the data

- Data too often with issues
  - Typing assignment: Vocabulary mismatch
  - Relation extraction assignment: Nationalities that are not nationalities

- Semiautomatic data:
  - Systematic errors

- Crowdsourced data:
  - Difficult cases avoided
  - Random noise
- …

Error analysis (3/3) – Question the rules

- Evaluation metric design not trivial
  - Named entity recognition, OpenIE: Partial matches?
  - Machine translation and summarization: BLEU
  - Typing: Metrics aware of error severity?
  - Disambiguation: Plausible vs. semantically impossible mismatches

(FIFA congress)

How to get gold data?

- Self-annotation
  - Alone or in a team of few researchers, colleagues
  - Confirmation bias
  - For final publication discouraged
- Creative reuse of existing data
  - E.g., Wikipedia text links for entity disambiguation
  - Synchronous edits of Wikidata relation and texts
  - Usually still shaky/biased
- Paid annotators
  - Can be known local personnel
  - More often, anonymous online crowdsourcing
  - De-facto standard nowadays

Crowdsourcing

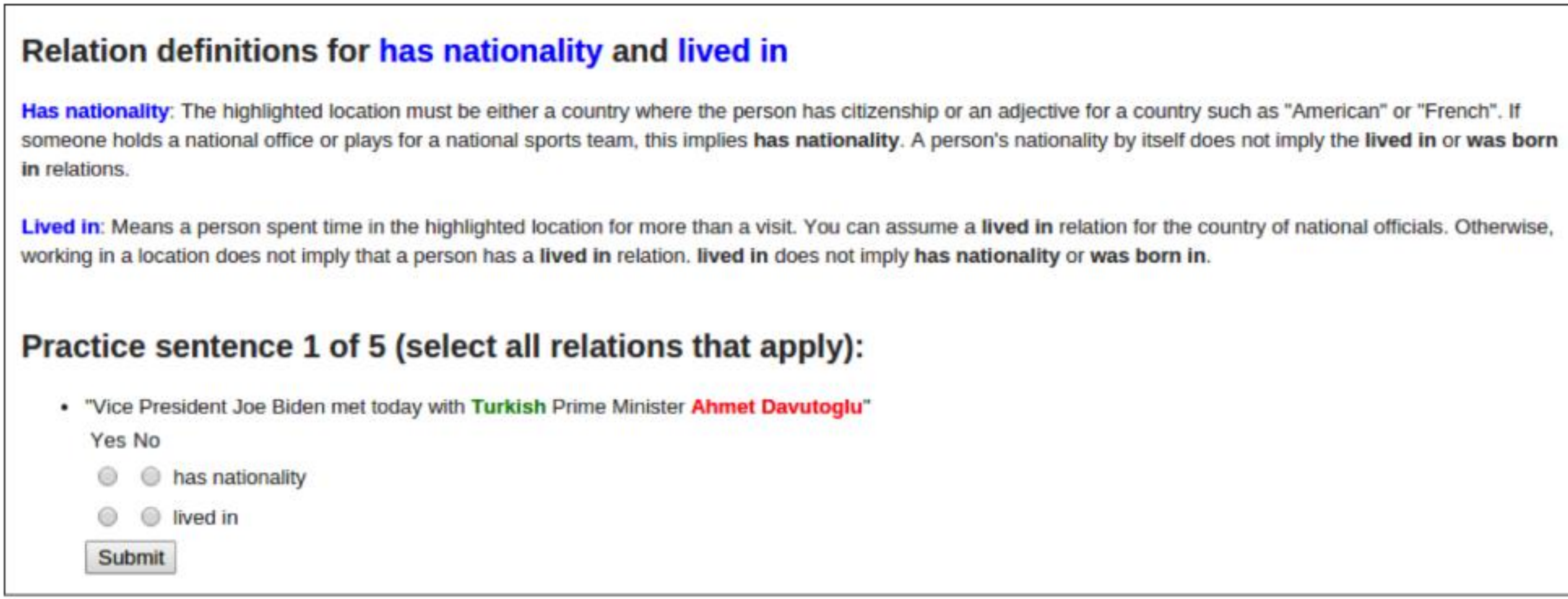
- Prominent platforms: Amazon Mechanical Turk, Prolific
- Typical pay ~10$/hour
  - In cases total spending 100k+€ for research datasets
- Requires to-the-point instructions
  - Traditional expert annotations guidelines sometimes >100 pages
  - Complex or open-ended annotation tasks difficult
    - Wherever possible, break into smaller tasks
- Quality assurance:
  - Worker education/background
  - Worker reputation
  - Honeypot/test question-based filtering
  - Redundancy (majority opinion on task)
- Creating good crowd tasks takes iterations and effort!

Example benchmark dataset: KnowledgeNet

[Mesquita et al., EMNLP 2019 https://www.aclweb.org/anthology/D19-1069.pdf]

- Text: Wikipedia abstracts
- 15 common person relations
- 9000 exhaustively annotated sentences
- Interannotator agreement
  - Relation classification: 96%
  - Entity disambiguation: 93%
- In-house annotators
- ~2 minutes/annotator/sentence for one property
  - 22% mention detection, 40% relation classification, 28% entity disambiguation
- 2 annotators, in case of disagreement third annotator

→ Total effort ~ 600 annotator hours



Relation definitions for has nationality and lived in

Has nationality: The highlighted location must be either a country where the person has citizenship or an adjective for a country such as "American" or "French". If someone holds a national office or plays for a national sports team, this implies **has nationality**. A person's nationality by itself does not imply the **lived in** or **was born in** relations.

Lived in: Means a person spent time in the highlighted location for more than a visit. You can assume a **lived in** relation for the country of national officials. Otherwise, working in a location does not imply that a person has a **lived in** relation. **lived in** does not imply **has nationality** or **was born in**.

Practice sentence 1 of 5 (select all relations that apply):

- "Vice President Joe Biden met today with **Turkish** Prime Minister **Ahmet Davutoglu**"

Yes No

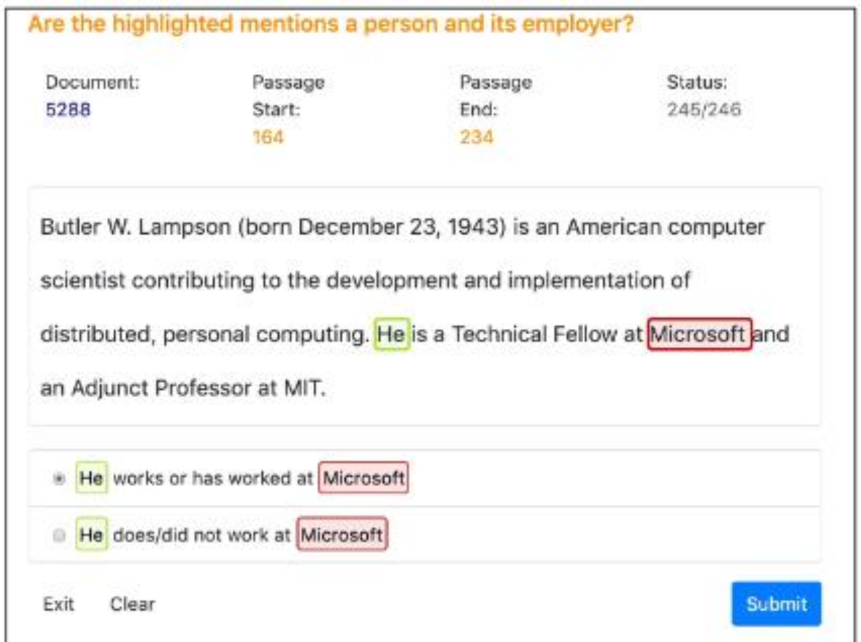
has nationality

lived in

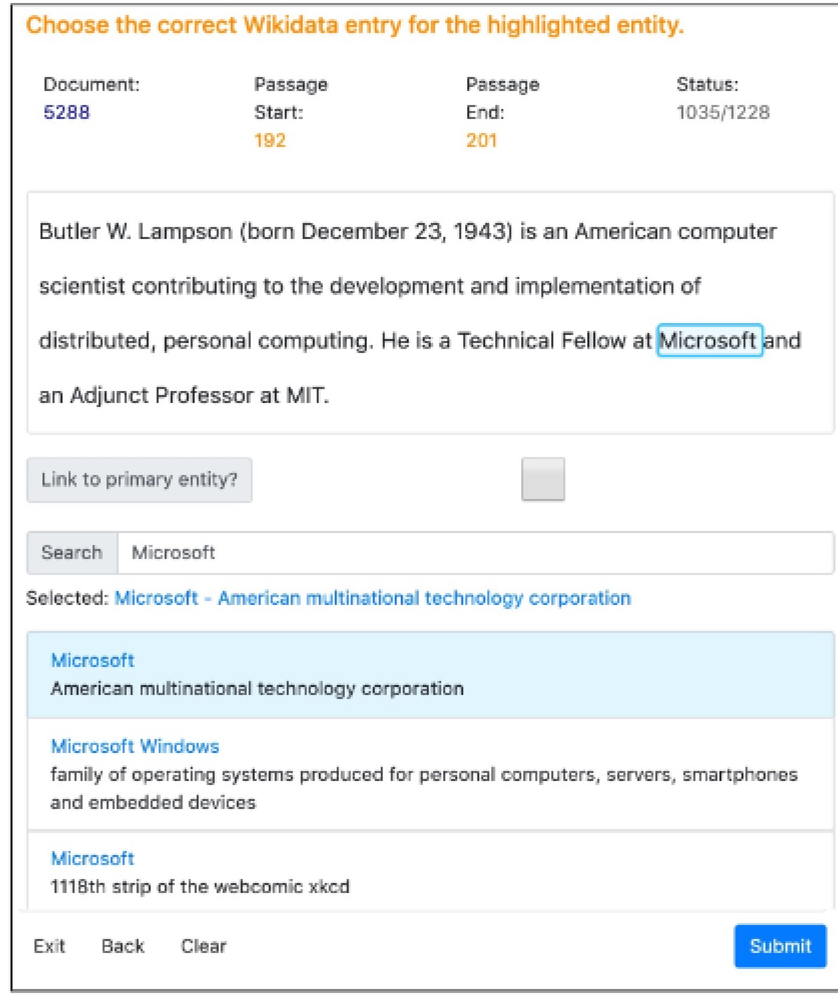
Submit

Figure 3: Tutorial page that teaches guidelines for *nationality* and *lived_in*. The worker answers practice sentences with immediate feedback that teach each relation.

(a) Interface to detect mentions of an entity type.



(b) Interface to classify facts.



(c) Interface to link a mention to a Wikidata entity.

Instructive pipeline implementations

- Mention detection, coreference resolution, relation classification, entity linking
- Human performance as comparison

	System	Text evaluation P	Text evaluation R	Text evaluation F1	Link evaluation P	Link evaluation R	Link evaluation F1
Stanford TAC KBP	Baseline 1	0.44	0.64	0.52	0.31	0.26	0.28
+ coreference	Baseline 2	0.49	0.64	0.55	0.37	0.32	0.34
+entity types	Baseline 3	0.47	0.66	0.55	0.35	0.37	0.36
+…	Baseline 4	0.60	0.65	0.62	0.51	**0.48**	0.49
+BERT	Baseline 5	**0.68**	**0.70**	**0.69**	**0.53**	**0.48**	**0.50**
	Human	0.88	0.88	0.88	0.81	0.84	0.82

Text spans of S and O match vs. KB links match

Takeaway: Evaluation

- Choose metrics wisely
  - Single metric desirable for ranking, but limited
    - Simplifies complex picture of data distribution and error categories
    - Thresholding behavior may matter
    - Classification vs. extraction problem
  - Goodhart's law: Metrics cease to be good metrics once the become the prime target
- Error analysis essential for learning sth.
  - Error categorization
  - Question ground truth and metrics

Outline

1. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Verb-based
    2. SRL-based
    3. Organizing predicates
    4. Semistructured web: Openceres

Motivation: Open information extraction

- So far assumed a fixed set of relations

- Presumably designed by humans ("ontology engineers")

- Lessons from DB/KR Research
  - Declarative KR is expensive & difficult
  - Formal semantics is at odds with
    - Broad scope
    - Distributed authorship
  - A "universal ontology" is wishful thinking

Coverage limitations of ontology engineering: Examples

- Schema.org
  - Industry standard for microformat in wepages
  - 800 entity types, 1300 properties
  - CollegeOrUniversity: No numberOfStudents, nor degreesOffered
- Wikidata
  - Largest public crowd project on KBC
  - >7000 properties
  - Musicians: No performedAt, coveredArtist, songAbout
- IMDB
  - Most popular movie information website
  - [Lockard et a., NAACL 2019]: Contains only about 10% of properties of 8 other domain-specific websites

Open vs. Traditional RE

	Traditional RE	Open RE
Input:	Corpus + O(R) hand-labeled data	Corpus
Relations:	Specified in advance	Discovered automatically
Extractor:	Relation-specific	Relation-independent

How is Open RE Possible?

Semantic Tractability Hypothesis

∃ *easy-to-understand* subset of English

- Characterized relations/arguments syntactically
  [Banko et al. ACL '08]
- Characterization is compact, domain independent

- Covers 80-95% of binary relations in sample corpus

Relative Frequency	Category	Simplified Lexico-Syntactic Pattern
37.8	Verb	E_1 Verb E_2 *X established Y*
22.8	Noun+Prep	E_1 NP Prep E_2 *X settlement with Y*
16.0	Verb+Prep	E_1 Verb Prep E_2 *X moved to Y*
9.4	Infinitive	E_1 to Verb E_2 *X plans to acquire Y*
5.2	Modifier	E_1 Verb E_2 Noun *X is Y winner*

(simplified!)

Reverb [Fader et al., 2011]

Identify **Re**lations from **Verbs**.

1. Find longest phrase matching a simple syntactic constraint:

$$\boxed{\begin{array}{l} \qquad\qquad V \mid VP \mid VW^*P \\ V = \text{verb particle? adv?} \\ W = (\text{noun} \mid \text{adj} \mid \text{adv} \mid \text{pron} \mid \text{det}) \\ P = (\text{prep} \mid \text{particle} \mid \text{inf. marker}) \end{array}}$$

Sample Reverb relations

invented	acquired by	has a PhD in
denied	voted for	inhibits tumor growth in
inherited	born in	mastered the art of
downloaded	aspired to	is the patron saint of
expelled	Arrived from	wrote the book on

OpenIE: Demo

- https://demo.allennlp.org/open-information-extraction
- Einstein likes ice cream. Bus 105 is going to the zoo. The fox chased the rabbit that was hiding in the bush.
- See BIO tagging

Challenges (1)

- Larry Page, the CEO of Google, talks about multi-screen opportunities offered by Google.
- After winning the Superbowl, the Giants are now the top dogs of the NFL.
- Ahmadinejad was ***elected*** as the new President of Iran.
- The great R. Feynman worked jointly with F. Dyson

Challenges (2)

"John refused to visit Vegas."

↓

(John, refused to visit, Vegas)

"Early astronomers believed that the earth is the center of the universe."

↓

[(earth, is the center of, universe) Attribution: early astronomers]

"If she wins California, Hillary will be the nominated presidential candidate."

↓

[(Hillary, will be nominated, presidential candidate) Modifier: if she wins California]

System evolution

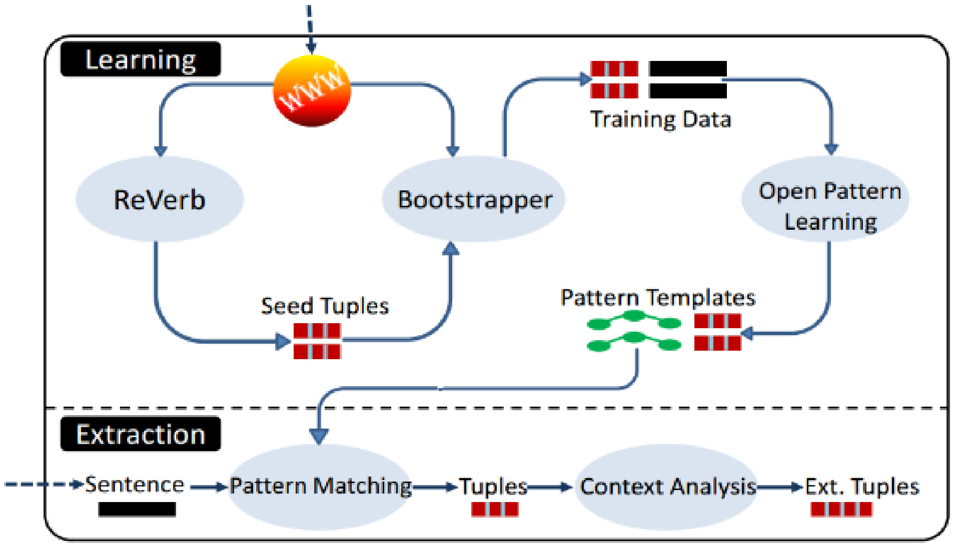
- 2007 Textrunner
  - CRF and self-training
- 2010 ReVerb
  - POS-based patterns
- 2012: OLLIE
  - Dependency-parse based
- 2013: ClausIE
  - Sentence restructuring before dependency parsing
- 2014 OpenIE 4.0
  - SRL-based extraction
- 2016 OpenIE 5.0
  - Compound noun phrases, numbers
- 2017 MinIE
  - Minimizing extractions by removal of minor qualifiers etc.

increasing precision, recall, expressiveness

OLLIE

Learning Open Patterns:

1) Extract the high confidence tuples from ReVerb.
2) For each tuple, find all sentences in the corpus containing the words in the tuple.
3) Using a dependency parser specify the patterns corresponding to each ReVerb tuple selected.



Number of relation phrases

DARPA MR Domains	<50
NYU, Yago	<100
NELL	~500
DBpedia 3.2	940
PropBank	3,600
VerbNet	5,000
Wikipedia Infoboxes, f > 10	~5,000
TextRunner	100,000+
ReVerb	1,000,000+

Demo: AKBC via OpenIE

https://openie.allenai.org/

- Saarland
- Einstein
- Kangaroo
- …

Outline

1. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Verb-based
    2. SRL-based
    3. Organizing predicates
    4. Semistructured web: Openceres

Semantic role labelling

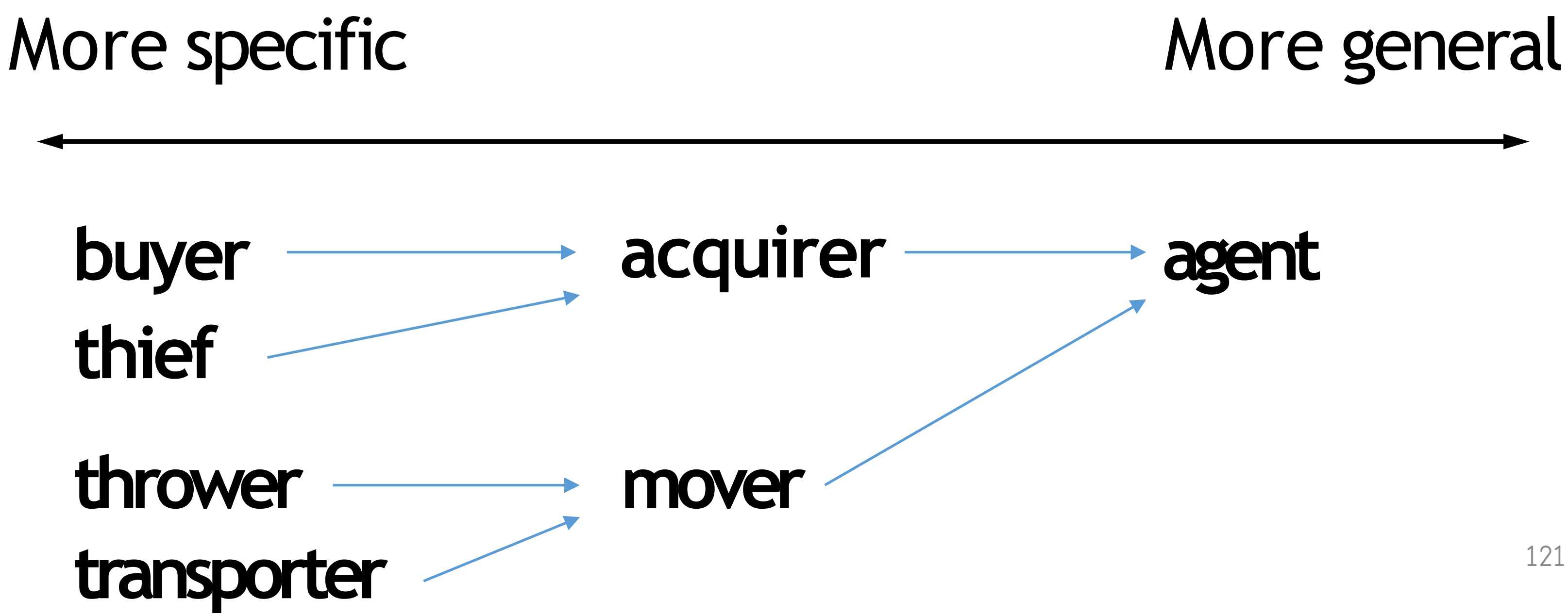
Can we figure out that these have the same meaning?

- XYZ corporation bought the stock.
- They sold the stock to XYZ corporation.
- The stock was bought by XYZ corporation.
- The purchase of the stock by XYZ corporation...
- The stock purchase by XYZ corporation...

- How do we represent this commonality?

A Shallow Semantic Representation: Semantic Roles

- Predicates (bought, sold, purchase) represent an event
- Semantic roles express the abstract role that arguments of a predicate can take in the event



Thematic roles

- Buyer and Thrower have something in common!
  - Volitional actors
  - Often animate
  - Direct causal responsibility for their events
- Thematic roles are a way to capture this semantic commonality between Buyers and Thrower.
- They are both AGENTS.
- The BoughtThing and ThrownThing, are THEMES.
  - prototypically inanimate objects affected in some way by the action
- One of the oldest linguistic models
  - Indian grammarian Panini between the 7th and 4th centuries BCE

Thematic roles

- A typical set:

Thematic Role	Definition	Example
AGENT	The volitional causer of an event	*The waiter* spilled the soup.
EXPERIENCER	The experiencer of an event	*John* has a headache.
FORCE	The non-volitional causer of the event	*The wind* blows debris from the mall into our yards.
THEME	The participant most directly affected by an event	Only after Benjamin Franklin broke *the ice*...
RESULT	The end product of an event	The city built a *regulation-size baseball diamond*...
CONTENT	The proposition or content of a propositional event	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	An instrument used in an event	He poached catfish, stunning them *with a shocking device*...
BENEFICIARY	The beneficiary of an event	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	The origin of the object of a transfer event	I flew in *from Boston*.
GOAL	The destination of an object of a transfer event	I drove *to Portland*.

Roles can be naturally described by questions

UCD ***finished*** the 2006 championship as Dublin champions ,
by ***beating*** St Vincents in the final .

finished

- Who finished something? - UCD
- What did someone finish? - the 2006 championship
- What did someone finish something as? - Dublin champions
- How did someone finish something? - by beating St Vincents in the final

beating

- Who beat someone? - UCD
- When did someone beat someone? - in the final
- Who did someone beat? - St Vincents

→ Crowd annotators write intuitive[1] questions and answers

[1][Dagan et al.] The PropBank annotation guide is 89 pages (Bonial etal., 2010), and the FrameNet guide is 119 pages (Ruppen-hofer et al., 2006). Our QA-driven annotation instructions are 5 pages.

Supervised OpenIE

[Stanovsky et al., NAACL 2018 https://www.aclweb.org/anthology/N18-1081]

- Uses SRL annotations as target and training data
  - Idea: Every set of (head, arg0, arg1) corresponds to a statement

- Trains a bi-LSTM to solve OpenIE via sequence labelling
  1. Verb identification
  2. Verb argument identification
  3. (head, arg0, arg1) as OIE output

Task: Formulate questions to elicit OpenIE triples

- Einstein likes ice cream.
- Bus 105 is going to the zoo.
- The fox chased the rabbit that was hiding in the bush.

Outline

1. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Verb-based
    2. SRL-based
    3. Organizing predicates
    4. Semistructured web: OpenCeres

Problem

- Are there really 1 Million different relations?
  - playedFor, wasOnTeam, appearedFor, playerOf, …
- Sparsity makes it difficult to spot frequent trends and similarities across entities
- Need to canonicalize and structure surface relations
  - Canonicalize ~ NED for entities
  - Structure ~ taxonomy construction for entity types

Key ingredient

Strong Co-Occurrence Principle:
If property name X frequently co-occurs with name Y in a context with cue Z (defined below), then Y is (likely) a synonym for X.

- This principle can be instantiated in various ways, depending on what we consider as context cue Z:
  - S-O context
  - Multilingual context
  - Search engine query-click logs
  - …

Instance overlap as context: PATTY

- Resource of 350k synsets of binary relations
- Taxonomical organization
- Key idea: exploit instance overlap/subsumption

Played for	Was on the team of	Liked to eat
(Ronaldo, ManU)	(Ronaldo, ManU)	(Einstein, ice cream)
(Messi, Barca)	(Messi, Barca)	

- Wikipedia-extractions between two named entities in sentence
- Patterns combine terms, POS tags, types

- Pattern accuracy: 85%
- Subsumption accuracy: 75%

PATTY (2)

ID	Pattern Synset & Support Sets
P_1	$\langle Politician \rangle$ was governor of $\langle State \rangle$ A,80 B,75 C,70
P_2	$\langle Politician \rangle$ *politician from* $\langle State \rangle$ A,80 B,75 C,70 D,66 E,64
P_3	$\langle Person \rangle$ *daughter of* $\langle Person \rangle$ F,78 G,75 H,66
P_4	$\langle Person \rangle$ *child of* $\langle Person \rangle$ I,88 J,87 F,78 G,75 K,64

A=(Schwarzenegger, California), 80 occurrences

Cluster of relational phrases
<location> is the heart of <location>
<location> is situated in <location>
<location> is enclosed by <location>
<location> is located amidst <location>
<location> is surrounded by <location>

<organization> acquires <organization>

<organization> purchased share <organization>

<organization> bought half of <company>

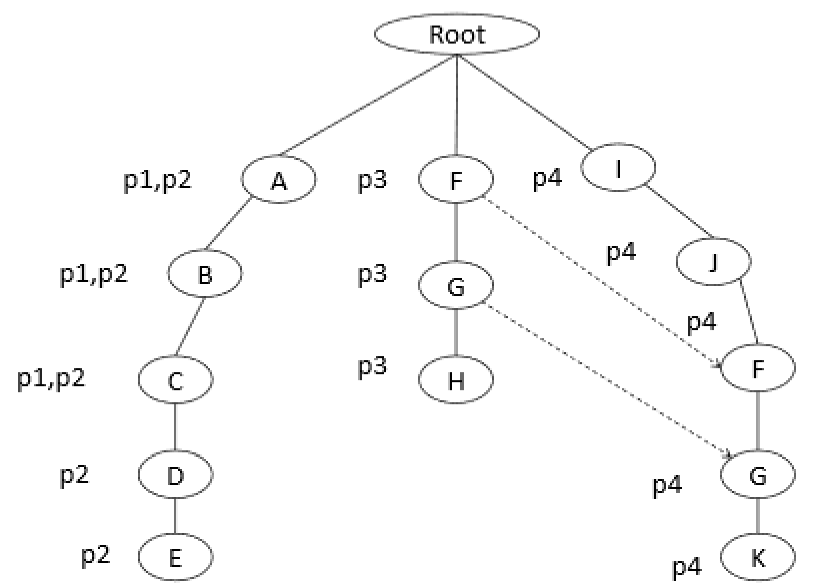
<company> bought half of <company>

<company> later bought half of <company>

Efficient support set overlap comparison

- n patterns → n^2 comparisons?

ID	Pattern Synset & Support Sets
P_1	⟨*Politician*⟩ was governor of ⟨*State*⟩ A,80 B,75 C,70
P_2	⟨*Politician*⟩ *politician from* ⟨*State*⟩ A,80 B,75 C,70 D,66 E,64
P_3	⟨*Person*⟩ *daughter of* ⟨*Person*⟩ F,78 G,75 H,66
P_4	⟨*Person*⟩ *child of* ⟨*Person*⟩ I,88 J,87 F,78 G,75 K,64



Prefix tree allows quick retrieval of subsumed patterns

Multilingual context: PPDB

A1: ...composed the soundtrack for ... → B1: ...schrieb die Filmmusik für ...
A2: ...wrote the score for ... → B2: ...schrieb die Filmmusik für ...

These are cues that "composed the soundtrack" and "wrote the score" are paraphrases of each other.

- One of the largest paraphrase dictionaries, PPDB (Paraphrase Database), was constructed similarly
- >100 million paraphrase pairs
- Covering both unary predicates (types/classes and WordNet-style senses) and binary predicates (relations and attributes)

Query-click-log as context

- Reformulations give first hints on synonyms/subproperties
- Even stronger: Observe result interaction: Are overlapping sets of results clicked?

[He et al., "Automatic Discovery of Attribute Synonyms Using Query Logs and Table Corpora". WWW. 2016]

Outline

1. Fixed-target relation extraction
    1. Task
    2. Manual patterns
    3. Supervised learning
    4. Learning at scale
        1. Iterative pattern learning
        2. Distant supervision
    5. Case study: CINEX
2. Evaluation
3. Open information extraction (OIE)
    1. Verb-based
    2. SRL-based
    3. Organizing predicates
    4. Semistructured web: OpenCeres

Example: OpenCeres

[Lockard et al., NAACL 2019]

- Back from text to seminstructured content
- Instructive OpenIE and distant supervision

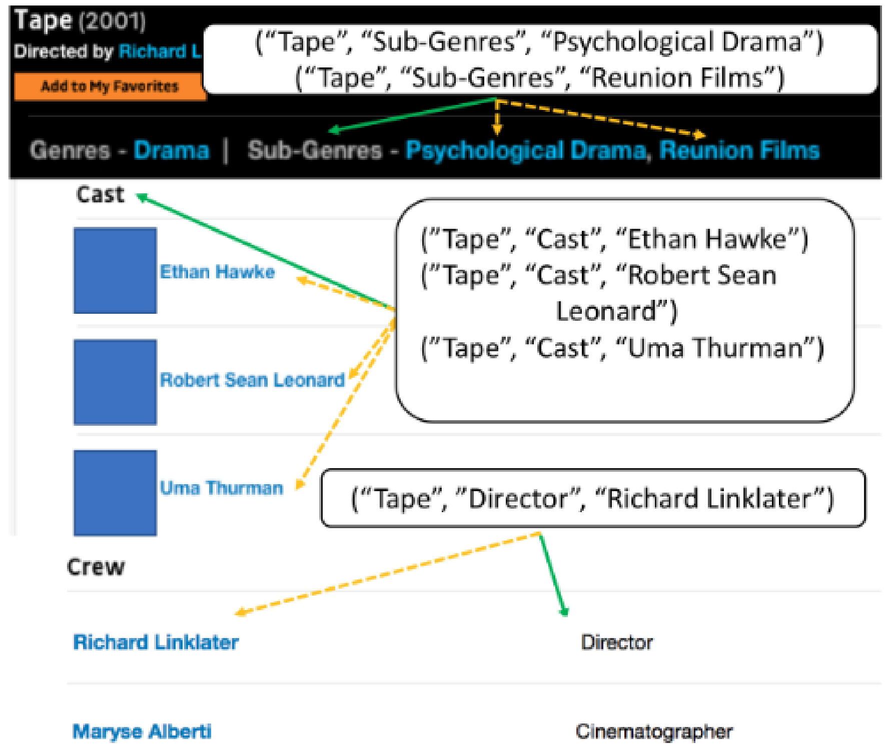
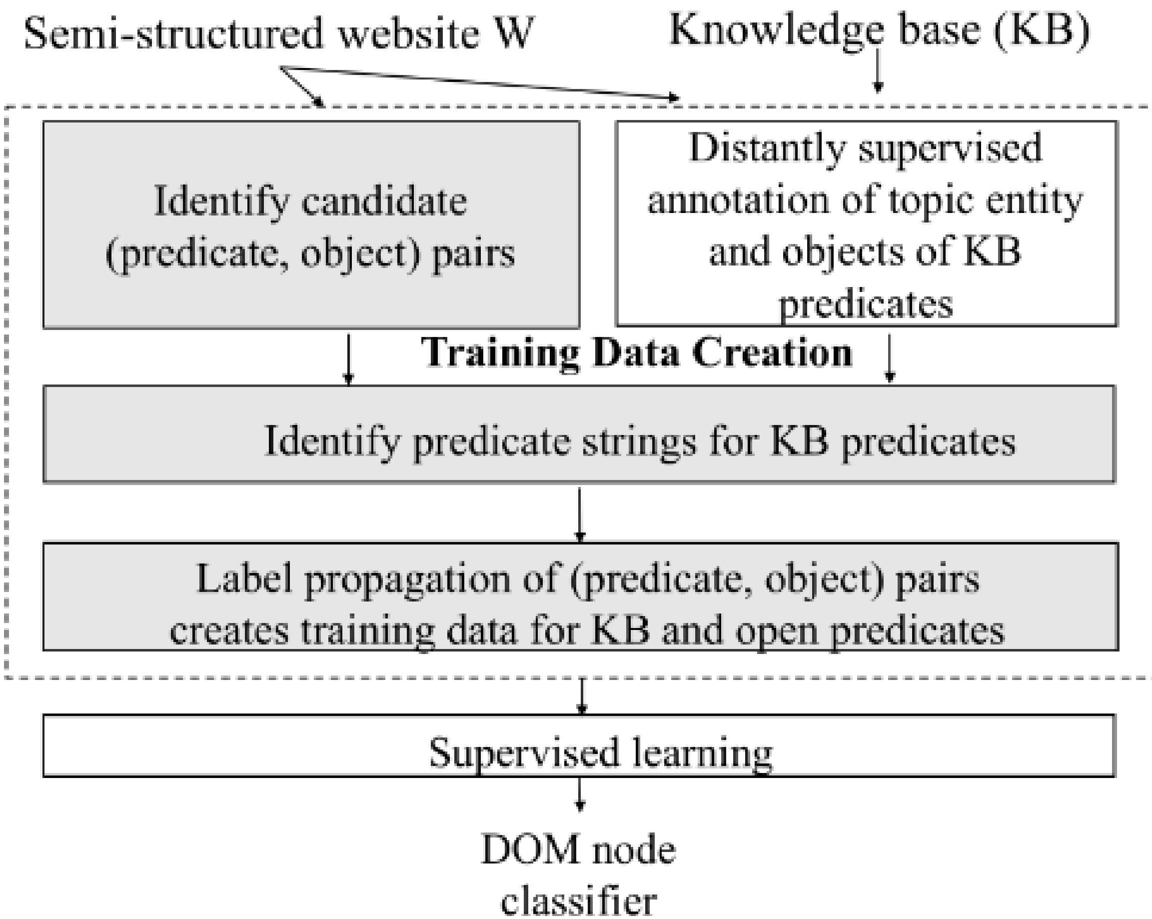


Figure 1: A cropped portion of the detail page from allmovie.com for the film *Tape* with some triples indicated. Solid green and dashed yellow arrows indicate predicate strings and objects respectively.

OpenCeres


Semi-structured website W
Knowledge base (KB)
Identify candidate (predicate, object) pairs
Distantly supervised annotation of topic entity and objects of KB predicates
Training Data Creation
Identify predicate strings for KB predicates
Label propagation of (predicate, object) pairs creates training data for KB and open predicates
Supervised learning
DOM node classifier

Take home: OpenIE

- Open IE/RE is a powerful machinery
  - Needs no labelled data
  - No domain-specific adaptation
  - Well suited for maximizing KB recall
  - Can discover new predicates
- Challenges
  - Typically substantial noise
  - Downstream applications that need clustering/canonicalization require additional processing steps
- Open predicate organization
  - Based on distributional similarity cues
    - E.g., instance overlap, multilingual alignments, query-click-logs

References

- Papers:
  - Stanovsky and Dagan, Creating a Large Benchmark for Open Information Extraction, EMNLP 2016
  - Nakashole et al., PATTY: A Taxonomy of Relational Patterns with Semantic Types, EMNLP 2012
- Slides
  - Adopted from Fabian Suchanek, Julien Romero and Oren Etzioni
- Code/APIs
  - OpenIE
    - https://www.textrazor.com/demo
    - https://gate.d5.mpi-inf.mpg.de/ClausIEGate/ClausIEGate/
    - https://github.com/dair-iitd/OpenIE-standalone
- Link collection on OpenIE
  - https://github.com/gkiril/oie-resources

Assignment 6

- Code your own open relation extraction
- Evaluation on benchmark data from [Stanovsky and Dagan, EMNLP 2017]
- F1 on extractions (head word match for predicate)

Take home

- Fixed relations
  - Supervised learning data bottleneck, but performant
  - Iterative pattern learning and distant supervision as alternatives
  - BERT allows to bypass feature engineering

- Evaluation
  - Right metric for right problem
  - Error analysis
  - Effort in data annotation, error analysis

- Open information extraction
  - Alternative requiring no decision on schema upfront
  - But some effort pushed downstream (clustering/canonicalization)