

Information Extraction

Adapted from: Chris Callison-Burch, Tom Mitchell

Information Extraction (IE)

- Information extraction (IE), turns the unstructured text information into structured data
- Populate a relational database to enable further processing, support queries

Template Filling

Citing high fuel prices, United Airlines said Friday it has increased fares by \$6 per round trip on flights to some cities also served by lower- cost carriers. American Airlines, a unit of AMR Corp., immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL Corp., said the increase took effect Thursday and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Denver to San Francisco.

FARE-RAISE ATTEMPT:	[LEAD AIRLINE:	UNITED AIRLINES]
		AMOUNT:	\$6	
		EFFECTIVE DATE:	2006-10-26	
		FOLLOWER:	AMERICAN AIRLINES	

Steps in IE

- NER and co-reference resolution
- **relation extraction**
 - *spouse-of, child-of, employer-of, part-of, membership-in, located-in*
- event extraction
- temporal expression normalization
- template filling

Named Entity Recognition

Citing high fuel prices, [ORG United Airlines] said [TIME Friday] it has increased fares by [MONEY \$6] per round trip on flights to some cities also served by lower-cost carriers. [ORG American Airlines], a unit of [ORG AMR Corp.], immediately matched the move, spokesman [PER Tim Wagner] said. [ORG United], a unit of [ORG UAL Corp.], said the increase took effect [TIME Thursday] and applies to most routes where it competes against discount carriers, such as [LOC Chicago] to [LOC Dallas] and [LOC Denver] to [LOC San Francisco].

Type	Tag	Sample Categories	Example sentences
People	PER	people, characters	Turing is a giant of computer science.
Organization	ORG	companies, sports teams	The IPCC warned about the cyclone.
Location	LOC	regions, mountains, seas	The Mt. Sanitas loop is in Sunshine Canyon.
Geo-Political Entity	GPE	countries, states, provinces	Palo Alto is raising the fees for parking.
Facility	FAC	bridges, buildings, airports	Consider the Tappan Zee Bridge.
Vehicles	VEH	planes, trains, automobiles	It was a classic Ford Falcon.

Category Ambiguity in NER

Name	Possible Categories
<i>Washington</i>	Person, Location, Political Entity, Organization, Vehicle
<i>Downing St.</i>	Location, Organization
<i>IRA</i>	Person, Organization, Monetary Instrument
<i>Louis Vuitton</i>	Person, Organization, Commercial Product

Examples of type ambiguities in the use of the name *Washington*:

[PER Washington] was born into slavery on the farm of James Burroughs.

[ORG Washington] went up 2 games to 1 in the four-game series.

Blair arrived in [LOC Washington] for what may well be his last state visit.

In June, [GPE Washington] passed a primary seatbelt law.

The [VEH Washington] had proved to be a leaky ship, every passage I made...

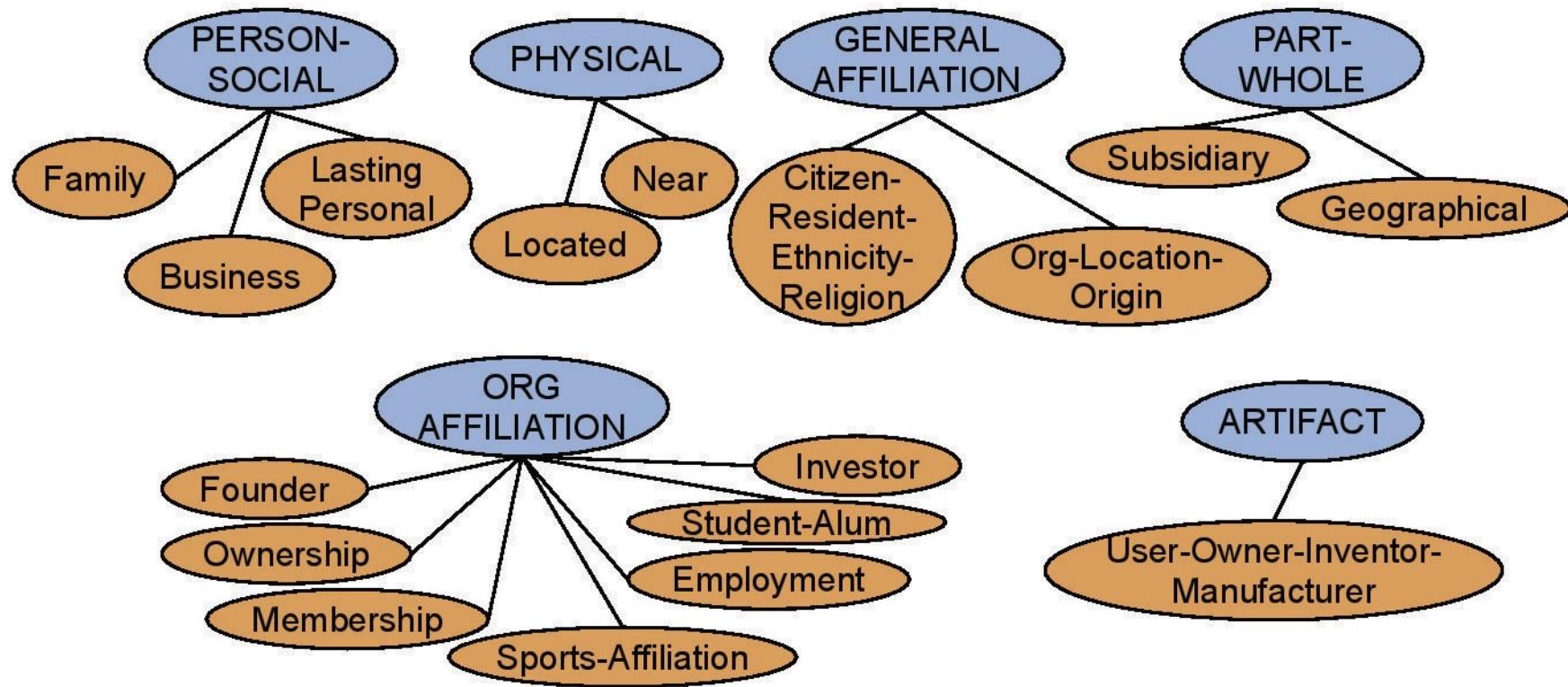
Relation Extraction

Citing high fuel prices, [ORG United Airlines] said [TIME Friday] it has increased fares by [MONEY \$6] per round trip on flights to some cities also served by lower-cost carriers. [ORG American Airlines], a unit of [ORG AMR Corp.], immediately matched the move, spokesman [PER Tim Wagner] said. [ORG United], a unit of [ORG UAL Corp.], said the increase took effect [TIME Thursday] and applies to most routes where it competes against discount carriers, such as [LOC Chicago] to [LOC Dallas] and [LOC Denver] to [LOC San Francisco].

Tim Wagner **is a spokesman for** American Airlines

United **is a unit of** UAL Corp

American **is a unit of** AMR



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

Unified Medical Language System

Entity	Relation	Entity
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

Given a medical sentence like:

Doppler echocardiography can be used to diagnose left anterior descending artery stenosis in patients with type 2 diabetes

Extract the UMLS relation:

Echocardiography (Doppler) Diagnoses Acquired stenosis

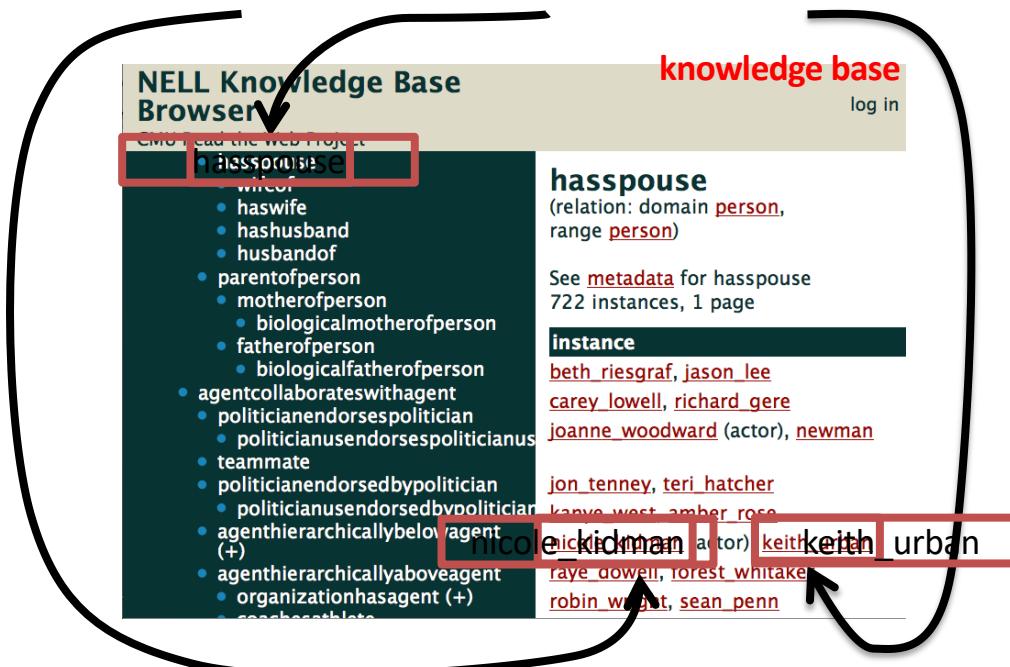
Wikipedia info boxes

Stephen Hawking CH CBE FRS FRSA	
	
Hawking at NASA's StarChild Learning Center, ca. 1999	
Born 8 January 1942 Oxford, England	Education St Albans School, Hertfordshire
Died 14 March 2018 (aged 76) Cambridge, England	Alma mater University of Oxford (BA) University of Cambridge (MA, PhD)
	Known for Hawking radiation Penrose–Hawking theorems Bekenstein–Hawking formula Hawking energy Gibbons–Hawking ansatz Gibbons–Hawking effect Gibbons–Hawking space Gibbons–Hawking–York boundary term Thorne–Hawking–Preskill bet
	Spouse(s) Jane Wilde (m. 1965; div. 1995) Elaine Mason (m. 1995; div. 2006)
	Children 3, including Lucy
	Awards Adams Prize (1966) Eddington Medal (1975) Maxwell Medal and Prize

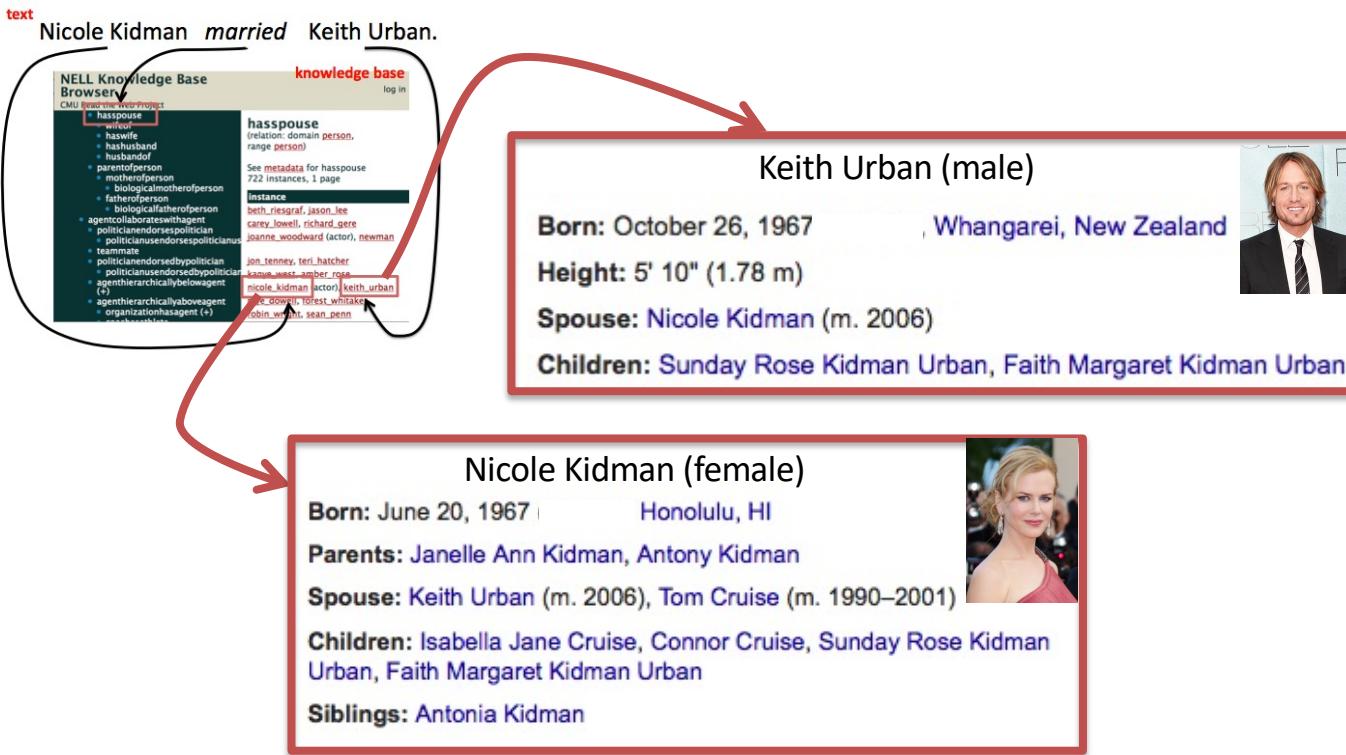
Natural Language Understanding (NLU)

text

Nicole Kidman *married* Keith Urban.

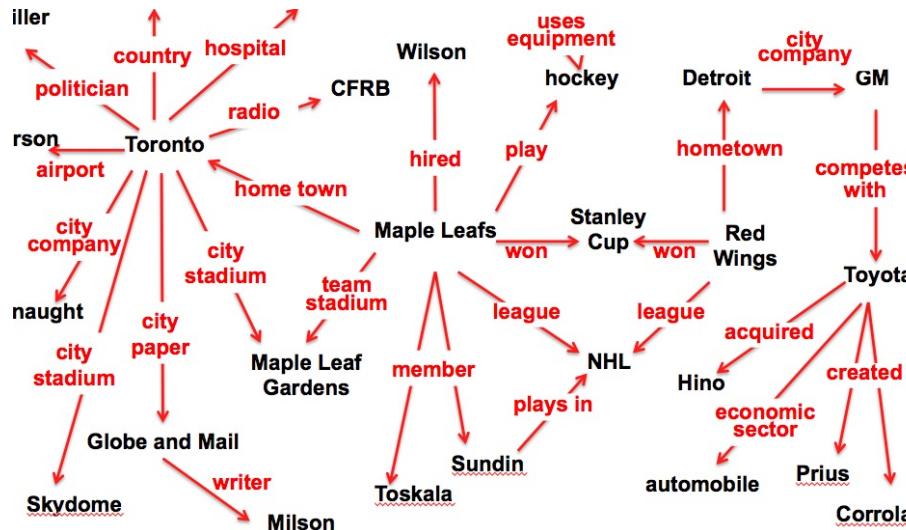


Incorporating KBs for NLU



Nicole Kidman married Keith Urban. She has two daughters with him.

Inference over KB



If: x_1 — competes with (x_1, x_2) \rightarrow x_2 — economic sector (x_2, x_3) \rightarrow x_3

Then: economic sector (x_1, x_3)

Inference over KB by Random Walks

KB:



Target Relation R:

economic sector (x1, x3)

Random walk
path type:

$x_1 \xrightarrow{\text{competes with}} x_2 \xrightarrow{\text{economic sector}} x_3$

Infer $P(R(x, y))$: Trained logistic function for R , where i^{th} feature is **probability** of arriving at node y when starting at node x , and taking a random walk along path type i

CityLocatedInCountry(Pittsburgh) = ?

Pittsburgh

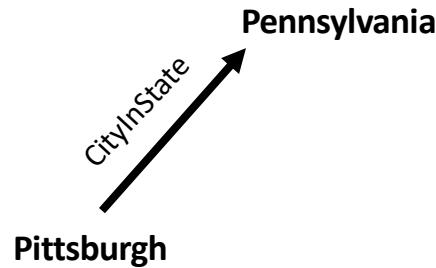
Feature = Typed Path

CityInState, CityInstate^{-1} , CityLocatedInCountry

Feature Value

Logistic
Regression
Weight
0.32

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



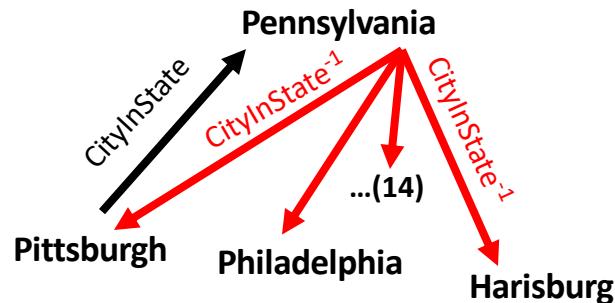
Feature = Typed Path

CityInState, CityInstate^{-1} , CityLocatedInCountry

Feature Value

**Logistic
Regression**
Weight
0.32

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



Feature = Typed Path

CityInState, CityInstate⁻¹, CityLocatedInCountry

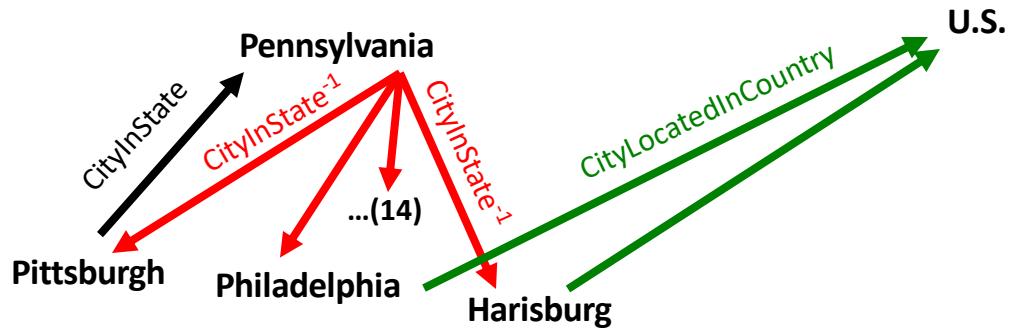
Feature Value

Logistic
Regression

Weight

0.32

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



Feature = Typed Path

CityInState , CityInstate^{-1} , $\text{CityLocatedInCountry}$

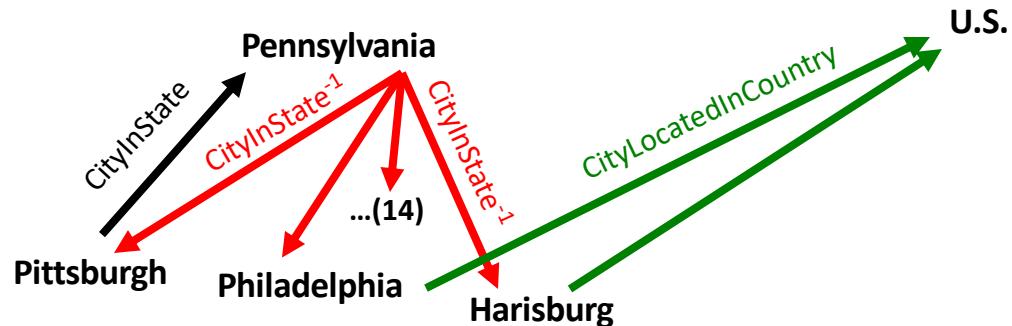
Feature Value

Logistic
Regression

Weight

0.32

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



$$\Pr(\text{U.S.} \mid \text{Pittsburgh}, \text{TypedPath})$$

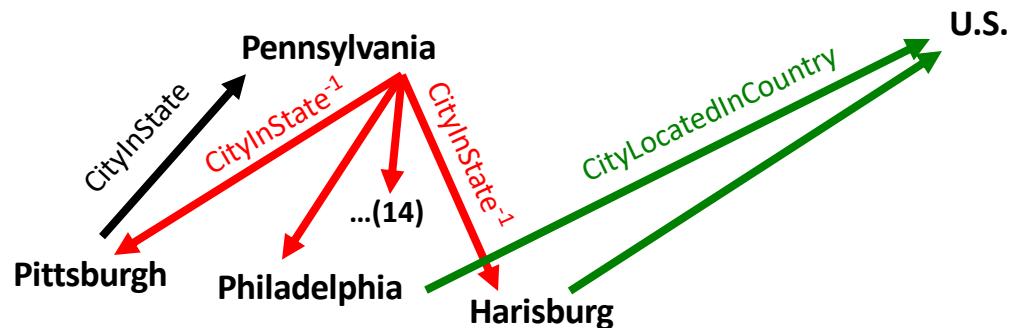
Feature = Typed Path

CityInState, CityInstate⁻¹, CityLocatedInCountry

Feature Value
0.8

Logistic
Regression
Weight
0.32

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



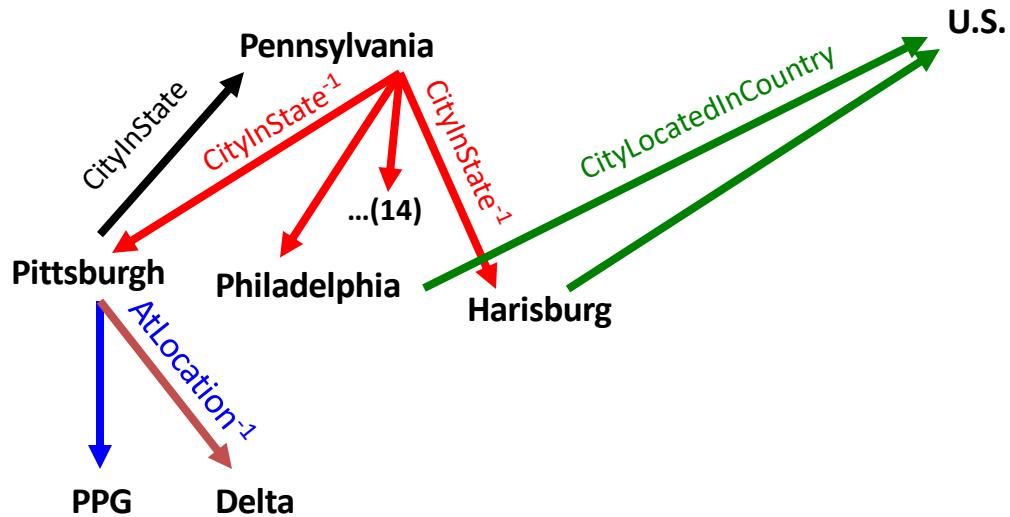
Feature = Typed Path

CityInState, CityInstate⁻¹, CityLocatedInCountry
AtLocation⁻¹, AtLocation, CityLocatedInCountry

Feature Value
0.8

Logistic Regression
Weight
0.32
0.20

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



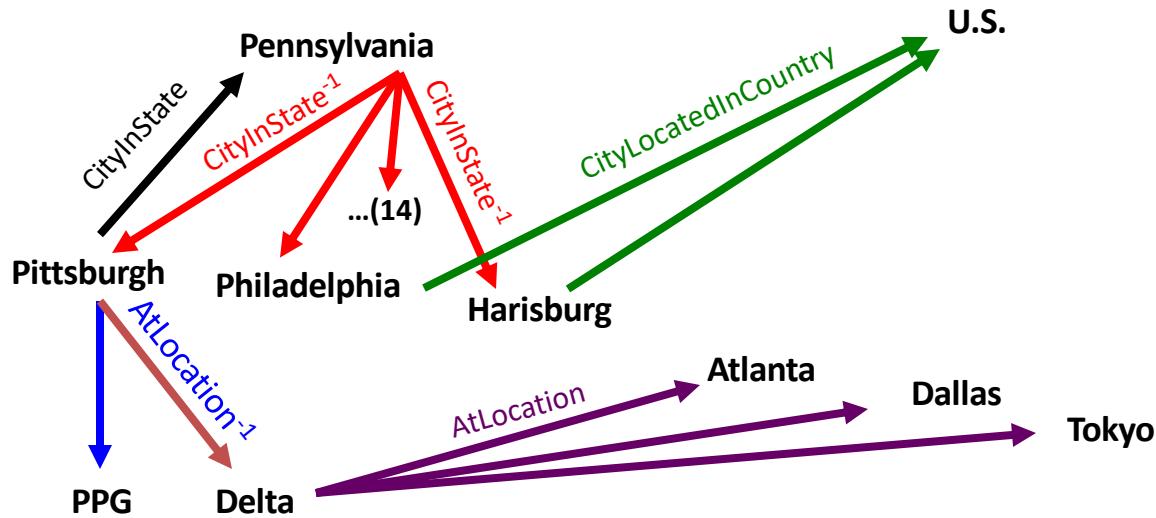
Feature = Typed Path

CityInState, CityInstate⁻¹, CityLocatedInCountry
AtLocation⁻¹, AtLocation, CityLocatedInCountry

Feature Value
0.8

Logistic Regression
Weight
0.32
0.20

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



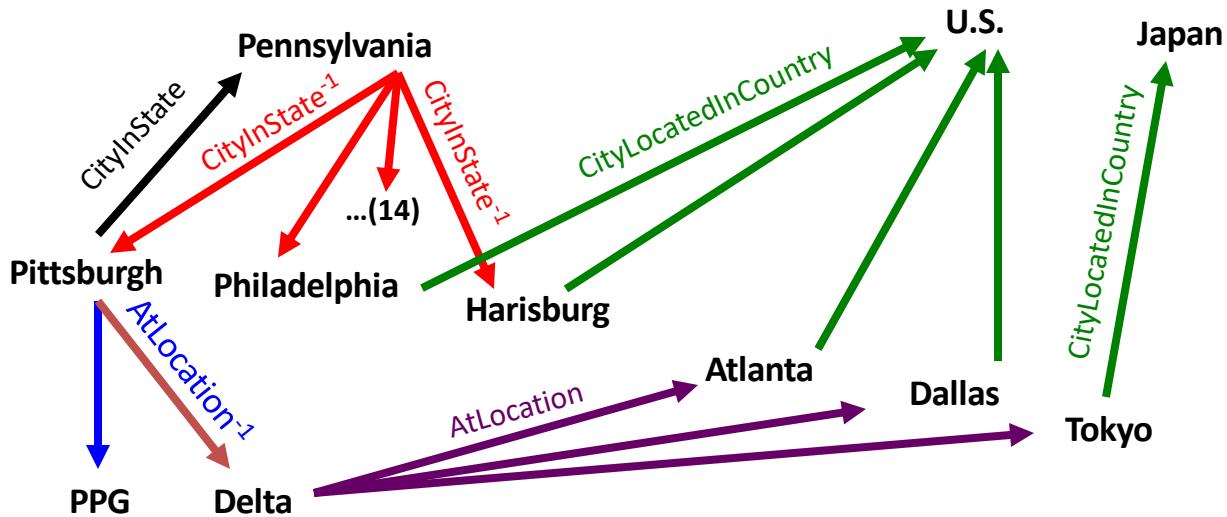
Feature = Typed Path

CityInState, **CityInstate⁻¹**, CityLocatedInCountry
AtLocation⁻¹, AtLocation, CityLocatedInCountry

Feature Value
0.8

Logistic Regression
Weight
0.32
0.20

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



Feature = Typed Path

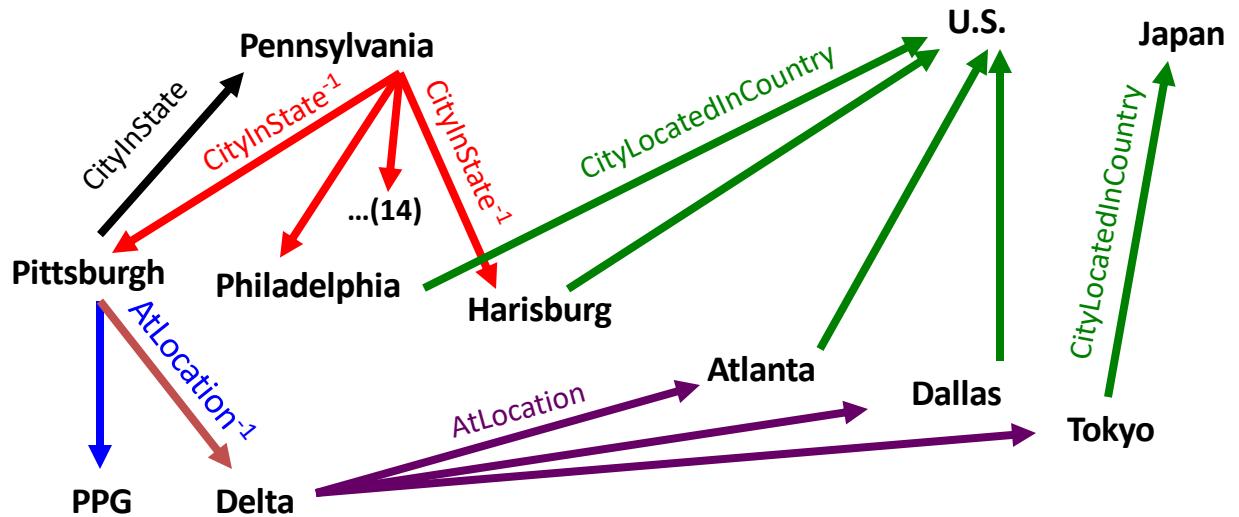
CityInState, CityInstate^{-1} , CityLocatedInCountry
 AtLocation^{-1} , AtLocation, CityLocatedInCountry

<u>Feature Value</u>
0.8
0.6

Logistic Regression
Weight

0.32
0.20

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = ?$



Feature = Typed Path

CityInState, CityInstate^{-1} , CityLocatedInCountry
 AtLocation^{-1} , AtLocation, CityLocatedInCountry

...

Feature Value	Logistic Regression Weight
0.8	0.32
0.6	0.20
...	...

$\text{CityLocatedInCountry}(\text{Pittsburgh}) = \text{U.S.}$ p=0.58

Ni Lao et al, EMNLP 2011

Random walk inference: learned path types

CityLocatedInCountry(*city, country*):

8.04 cityliesonriver, cityliesonriver⁻¹, citylocatedincountry
5.42 hasofficeincity⁻¹, hasofficeincity, citylocatedincountry
4.98 cityalsoknownas, cityalsoknownas, citylocatedincountry
2.85 citycapitalofcountry,citylocatedincountry⁻¹,citylocatedincountry
2.29 agentactsinlocation⁻¹, agentactsinlocation, citylocatedincountry
1.22 statehascapital⁻¹, statelocatedincountry
0.66 citycapitalofcountry
:
:

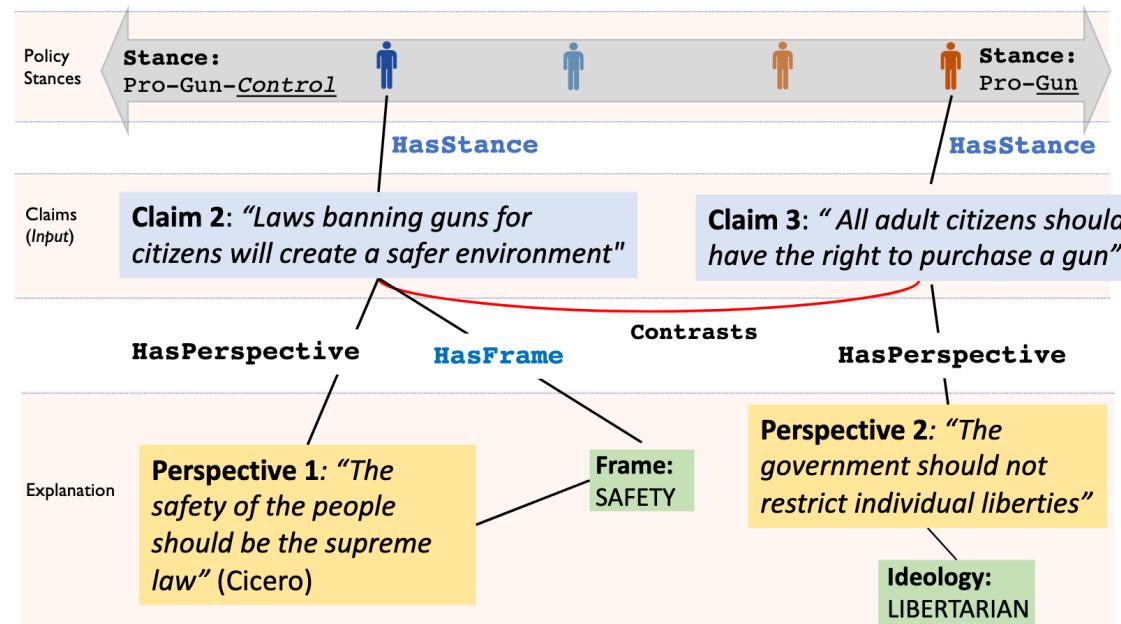
7 of the 2985 learned paths for CityLocatedInCountry

Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Explanations: *Parsing the Landscape of Opinions and Perspectives*



Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Declarative Modeling

- Often easier to think about structure in a declarative way
- Define **entities** and **relations** and **probabilistic rules**

$\text{Agree}(\text{doc1}, \text{doc2}) \wedge \text{Label}(\text{doc1}, \text{Pos}) \rightarrow \text{Label}(\text{doc2}, \text{Pos})$

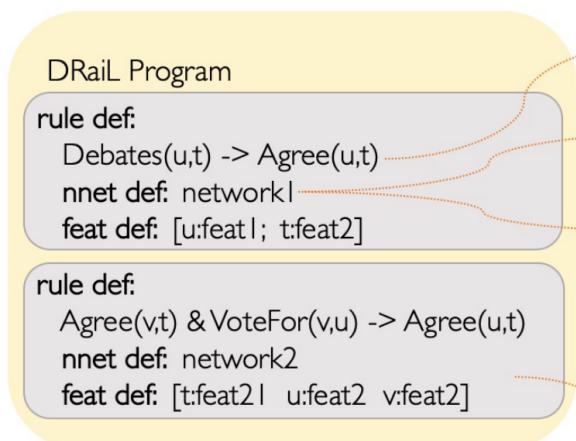
Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Deep Relational Learning

- **DRail**, **Declarative** framework for Deep Relational Learning
 - Rules as context, using a graphical model
 - Representation as context, using neural architectures



Each rule:

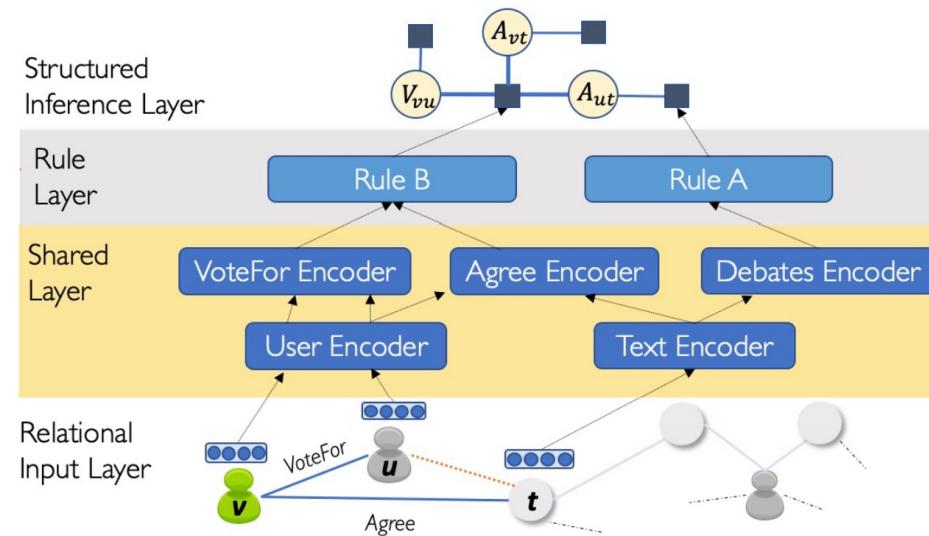
- (1) Defines a factor template to capture dependencies between decisions contained in it
- (2) Defines a sub-graph for compositional representation learning

Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Deep Relational Learning



DRaiL compiles logic programs into a neural factor graph and learns its parameters given data

Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Summary: DRaiL

- **A general framework for combining symbolic and neural representations**
- **Neural:** captures “implicit” interactions between entities in the embedding space.
- **Symbolic:** explicit interactions between entities, forced to provide a consistent view
- **Neuro-Symbolic:** consistency constraints are propagated to the embedding space
- **Provides a convenient way to compile symbolic explanations into neural classifiers**

Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Political Issue Stance and Framing



Inhofe Press Office @InhofePress

Follow

Six years later, health care costs have skyrocketed and millions have lost access to their doctors. #RepealObamacare

Stance:
Clearly, not a fan.

Framing: what are the right abstractions of the tweet,
capturing the arguments supporting the stance?

“*Obamacare should be repealed since it is too expensive*”



*Healthcare is framed as an **economic** issue*

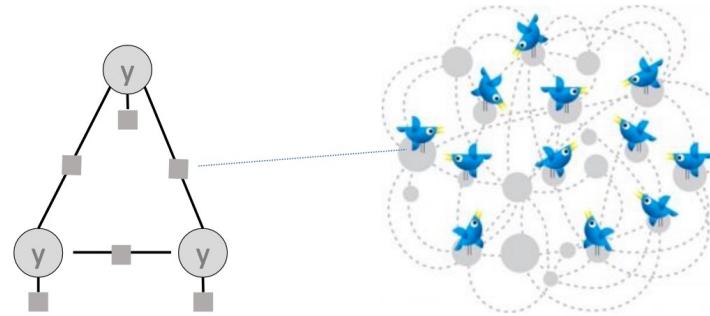
Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Modeling Social and Behavioral Information

$\text{Retweets}(T1, T2) \wedge \text{Frame}(T1, F) \rightarrow \text{Frame}(T2, F)$
 $\text{Follows}(T1, T2) \wedge \text{Frame}(T1, F) \rightarrow \text{Frame}(T2, F)$



Use twitter social activity (e.g., follows, retweet) as an indication of agreement

Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Morality Frames

- Moral Foundation Theory was repeatedly used to explain behaviors.
 - Liberals emphasize Fairness, Conservatives emphasize Loyalty and Authority
- **But.. Everybody CAREs ... but not about the same things!**

[@SenThadCochran and I]**CARING** are working to protect
[MS small businesses]**CARE-FOR** from more expensive
[#Obamacare mandates]**HARMING**.

[The ACA]**CARING** was a life saver for the more than
[130 million Americans]**CARE-FOR** with a preexisting
condition – including covid now. [Republicans]**HARMING**
want to take us back to coverage denials.

Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Morality Frames

- We define a morality frame structure to capture differences in the **targets** of moral sentiment

If the target of CARE is “illegal immigrants”, author more likely to be a...

If the causer of HARM is “illegal immigrants”, author more likely to be a...

MORAL FOUNDATIONS	MORAL ROLES
CARE/HARM: Care for others, generosity, compassion, ability to feel pain of others, sensitivity to suffering of others, prohibiting actions that harm others.	1. Target of care/harm 2. Entity causing harm 3. Entity providing care
FAIRNESS/CHEATING: Fairness, justice, reciprocity, reciprocal altruism, rights, autonomy, equality, proportionality, prohibiting cheating.	1. Target of fairness/cheating 2. Entity ensuring fairness 3. Entity doing cheating
LOYALTY/BETRAYAL: Group affiliation and solidarity, virtues of patriotism, self-sacrifice for the group, prohibiting betrayal of one's group.	1. Target of loyalty/betrayal 2. Entity being loyal 3. Entity doing betrayal
AUTHORITY/SUBVERSION: Fulfilling social roles, submitting to authority, respect for social hierarchy/traditions, leadership, prohibiting rebellion against authority.	1. Justified authority 2. Justified authority over 3. Failing authority 4. Failing authority over
PURITY/DEGRADATION: Associations with the sacred and holy, disgust, contamination, religious notions which guide how to live, prohibiting violating the sacred.	1. Target of purity/degradation 2. Entity preserving purity 3. Entity causing degradation

Deep Relational Learning

Pacheco and Goldwaser, 2021

Neuro-symbolic approaches in NLP

Morality Frames

- DRaiL Model

Basic Classifiers

$$r_1 : \text{Tweet}(t) \Rightarrow \text{MF}(t, m)$$

$$r_2 : \text{Tweet}(t) \wedge \text{Ent}(t, e) \Rightarrow \text{Role}(t, e, r)$$

Party's messaging preferences

$$r_3 : \text{Tweet}(t) \wedge \text{Ideo}(t, i) \wedge \text{Topic}(t, k) \Rightarrow \text{MF}(t, m)$$

$$r_4 : \text{Tweet}(t) \wedge \text{Ideo}(t, i) \wedge \text{Topic}(t, k) \wedge \text{Ent}(t, e) \Rightarrow \text{Role}(t, e, r)$$

Party's consistency preferences

$$\begin{aligned} c_3 : \text{SameIdeo}(t_1, t_2) \wedge \text{SameTopic}(t_1, t_2) \wedge \text{Ent}(t_1, e) \wedge \text{Ent}(t_2, e) \\ \wedge \text{Role}(t_1, e, r_1) \wedge \text{Role}(t_2, e, r_2) \Rightarrow \text{SamePolarity}(r_1, r_2) \end{aligned}$$

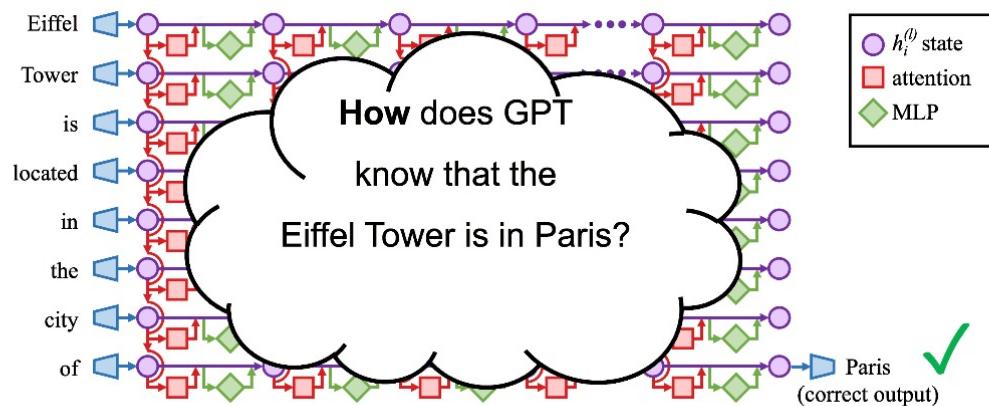
Locating Knowledge in LM

Meng et al., 2022

Locating and Editing Factual Knowledge in GPT

Where is the Knowledge Inside a Language Model?

Knowing differs from **saying**: uttering words by rote is different from knowledge, because *knowledge of a fact generalizes across contexts*. In this project, we show that factual knowledge within GPT **also corresponds to a localized computation that can be directly edited**. For example, we can make a small change to a small set of the weights of GPT-J to teach it the counterfactual "Eiffel Tower is located in the city of Rome." Rather than merely regurgitating the new sentence, it will generalize that specific counterfactual knowledge and apply it in very different linguistic contexts.



Locating Knowledge in LM

Meng et al., 2022

Locating and Editing Factual Knowledge in GPT

Why Locate Knowledge?

We are interested **how** and **where** a model stores its factual knowledge, for two reasons:

1. **To understand huge opaque neural networks.** The internal computations of large language models are obscure. Clarifying the processing of facts is one step in understanding massive transformer networks.
2. **Fixing mistakes.** Knowledge is sometimes incorrect, biased, or private, and we would like to develop methods that will enable debugging and fixing of specific factual errors.

The facts we study take the form of knowledge tuples $\mathbf{t} = (\mathbf{s}, \mathbf{r}, \mathbf{o})$, where \mathbf{s} and \mathbf{o} are subject and object entities, respectively, and \mathbf{r} is the relation connecting the two. For example, ($\mathbf{s} = \text{Megan Rapinoe}$, $\mathbf{r} = \text{plays sport professionally}$, $\mathbf{o} = \text{soccer}$) indicates that Rapinoe plays soccer for a living. Each variable represents an entity or relation that can be found in a knowledge graph, and that can be written as a natural language string.

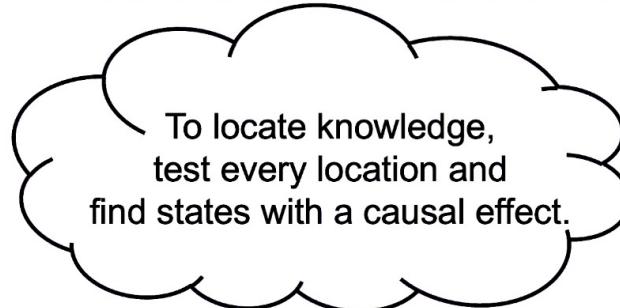
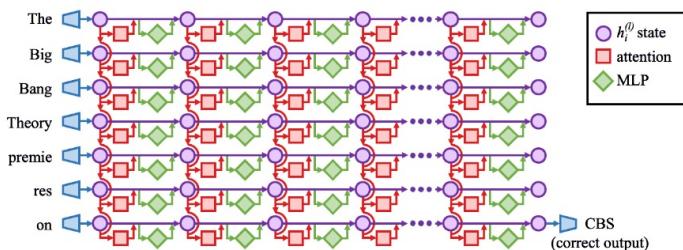
To query GPT for knowledge of a fact, we express (\mathbf{s}, \mathbf{r}) as a text prompt (by expanding a template from the **COUNTERFACT** data set), and check whether the generated continuation matches \mathbf{o} .

Locating Knowledge in LM

Meng et al., 2022

How to Locate Knowledge

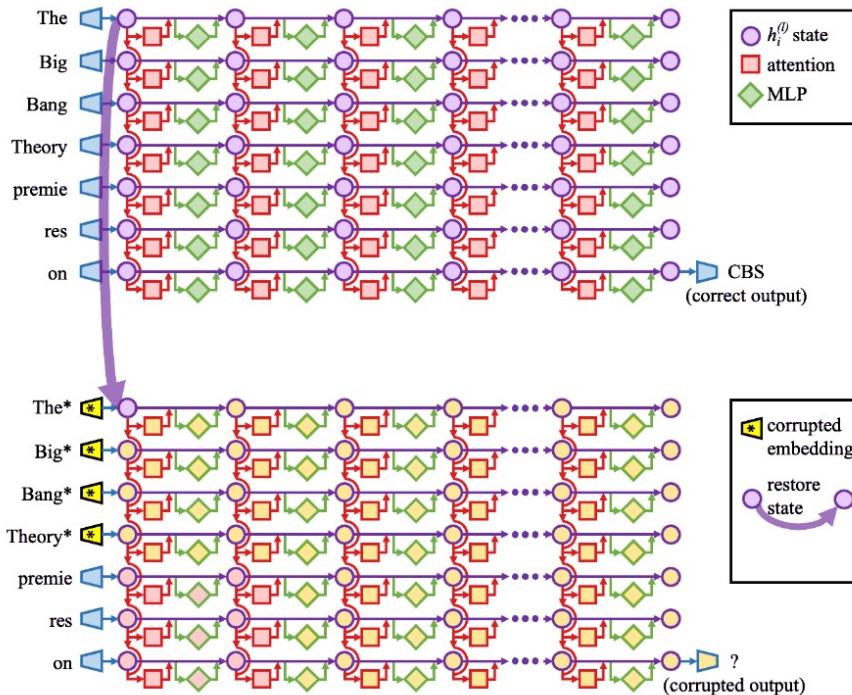
To identify decisive computations, we introduce a method called *Causal Tracing*. By isolating the causal effect of individual states within the network while processing a factual statement, we can trace the path followed by information through the network.



Causal traces work by running a network multiple times, introducing corruptions to frustrate the computation, and then restoring individual states in order to identify the information that restores the results. Tracing can be used to test any individual state or combinations of states. We use carefully-designed traces to identify a specific small set of MLP module computations that mediate knowledge retrieval.

Locating Knowledge in LM

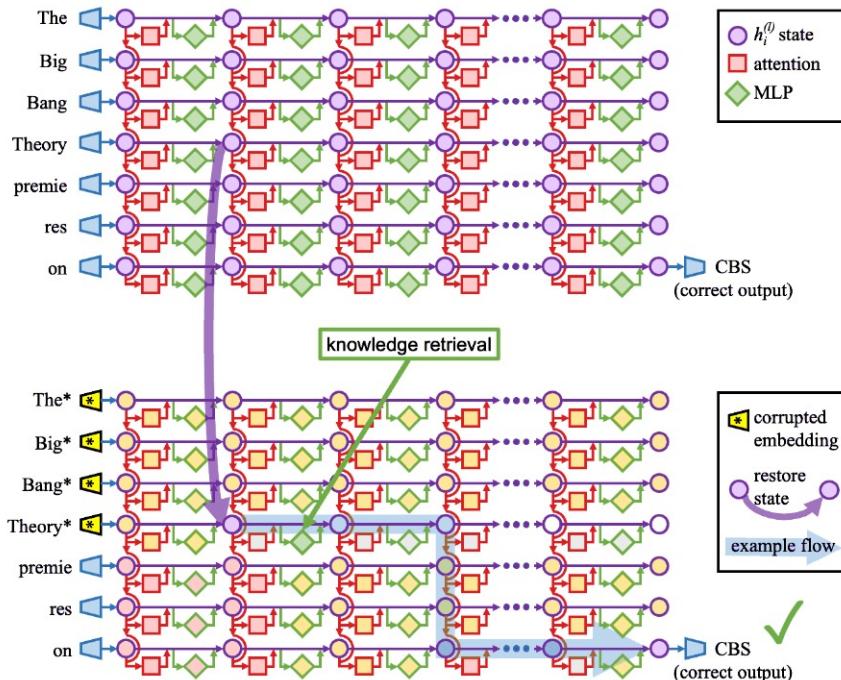
Meng et al., 2022



Causal traces work by running a network multiple times, introducing corruptions to frustrate the computation, and then restoring individual states in order to identify the information that restores the results. Tracing can be used to test any individual state or combinations of states. We use carefully-designed traces to identify a specific small set of MLP module computations that mediate knowledge retrieval.

Locating Knowledge in LM

Meng et al., 2022



Causal traces work by running a network multiple times, introducing corruptions to frustrate the computation, and then restoring individual states in order to identify the information that restores the results. Tracing can be used to test any individual state or combinations of states. We use carefully-designed traces to identify a specific small set of MLP module computations that mediate knowledge retrieval.

RDF Triples

- Resource Description Framework
- **RDF triple** is a tuple of
 - entity-relation-entity, aka
 - subject-predicate-object expression
- **subject** University of Pennsylvania
predicate location
object Philadelphia, PA



Zachary G. Ives

FOLLOW

Professor of Computer and Information Science, University of Pennsylvania
Verified email at cis.upenn.edu - [Homepage](#)

Databases data integration distributed systems web data management

TITLE	CITED BY	YEAR
DBpedia: A nucleus for a web of open data S Auer, C Bizer, G Kobilarov, J Lehmann, R Cyganiak, Z Ives The semantic web, 722-735	2950	2007

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available on the Web. DBpedia allows you to ask sophisticated queries against datasets derived from Wikipedia and to link other datasets on the Web to Wikipedia data. We describe the extraction of the DBpedia datasets, and how the resulting information is published on the Web for human-and machine-consumption. We describe some emerging applications from the DBpedia community and show how website authors can facilitate DBpedia content within their sites. Finally, we present the current status of interlinking DBpedia with other open datasets on the Web and outline how DBpedia could serve as a nucleus for an emerging Web of open data.

Freebase, WordNet, other ontologies

- **Freebase** relations:
people/person/nationality
location/location/contains
people/person/place-of-birth
biology/organism classification
- **WordNet** relations:
is-a, instance-of
hypernyms/hyponyms
*Giraffe is-a ruminant is-a ungulate is-a mammal
is-a vertebrate is-a animal...*

Strategies for relation extraction

- hand-written patterns
- supervised machine learning
- semi-supervised machine learning
- unsupervised machine learning

Hearst Patterns

Agar is a substance prepared from a mixture of red algae, such as Gelidium, for laboratory or industrial use.

She suggests that the following **lexico-syntactic pattern**

$$NP_0 \text{ such as } NP_1 \{, NP_2 \dots, (\text{and}|\text{or}) NP_i\}, i \geq 1$$

implies the following semantics

$$\forall NP_i, i \geq 1, \text{hyponym}(NP_i, NP_0)$$

allowing us to infer

$$\text{hyponym}(\text{Gelidium}, \text{red algae})$$

NP {, NP}* {,} (and or) other NP _H	temples, treasuries, and other important civic buildings
NP _H such as {NP,}* { (or and) } NP	red algae such as Gelidium
such NP _H as {NP,}* { (or and) } NP	such authors as Herrick, Goldsmith, and Shakespeare
NP _H {,} including {NP,}* { (or and) } NP	common-law countries , including Canada and England
NP _H {,} especially {NP,}* { (or and) } NP	European countries, especially France, England, and Spain

Machine learning techniques

- Supervised: training corpus annotated with manually annotated with fixed set of relations and entities
- Semi-supervised: high-precision seed patterns, or seed tuples, are used to bootstrap more examples
- Distant supervision: start with a huge number of seeds, learn noisy pattern fields (e.g. 100k examples of birth-place-of from infoboxes, help learn the corresponding text patterns)

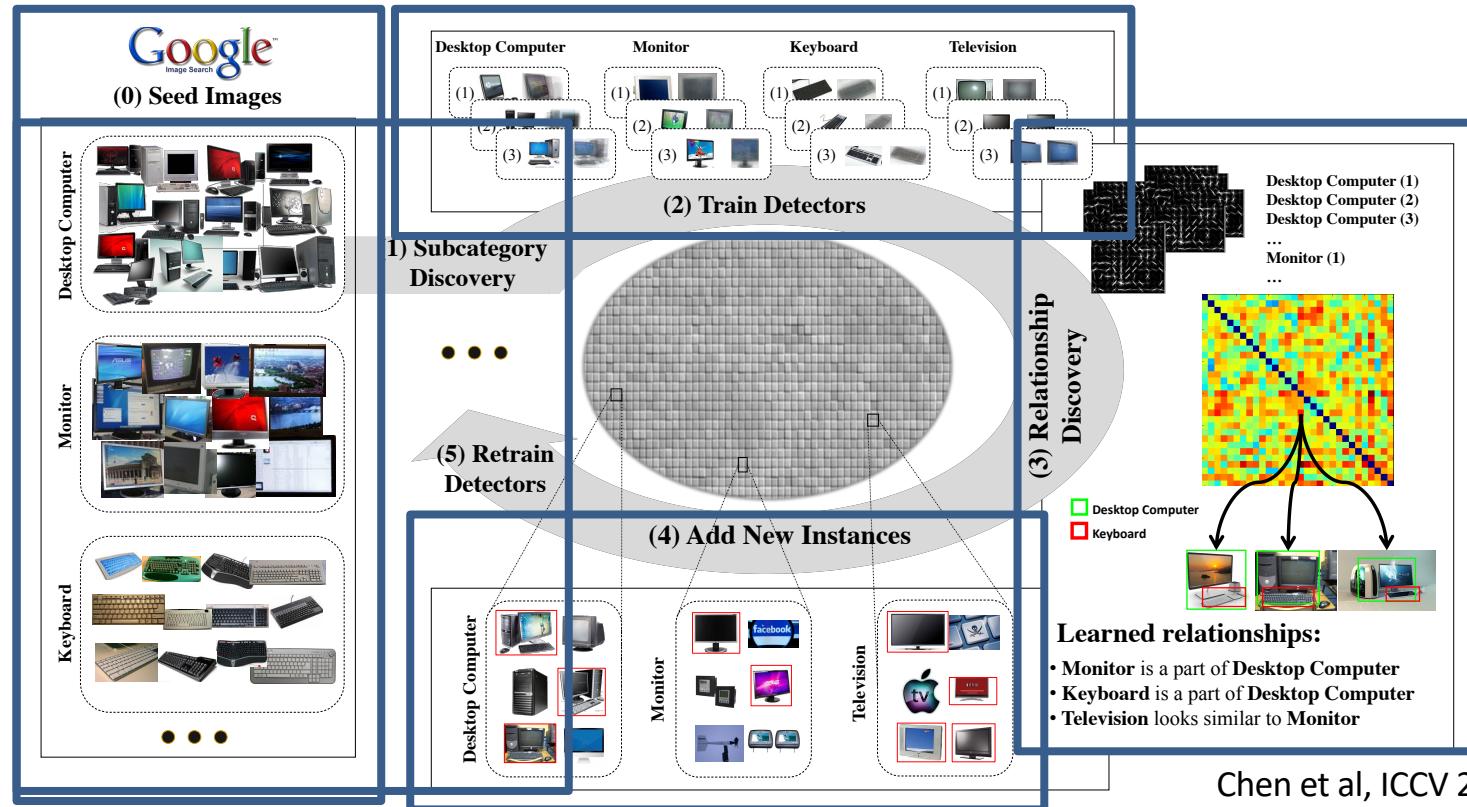
Never Ending Language Learner (NELL) rtw.ml.cmu.edu/



Carlson et al, WSDM 2010; Mitchell et al, AAAI 2015

Never Ending Image Learner (NEIL)

www.neil-kb.com/



Chen et al, ICCV 2013

Open IE

- Unsupervised relation extraction
- Find all *strings of words* that satisfy the triple relation.

United has a hub in Chicago, which is the headquarters of United Continental Holdings.

r1: <United, has a hub in, Chicago>

r2: <Chicago, is the headquarters of, United Continental Holdings>

Temporal Expression Extraction

Absolute	Relative	Durations
April 24, 1916	yesterday	four hours
The summer of '77	next semester	three weeks
10:15 AM	two weeks from yesterday	six days
The 3rd quarter of 2006	last quarter	the last three quarters

Lexical triggers for temporal expressions:

Category	Examples
Noun	<i>morning, noon, night, winter, dusk, dawn</i>
Proper Noun	<i>January, Monday, Ides, Easter, Rosh Hashana, Ramadan, Tet</i>
Adjective	<i>recent, past, annual, former</i>
Adverb	<i>hourly, daily, monthly, yearly</i>

Event Extraction

[EVENT Citing] high fuel prices, United Airlines [EVENT said] Friday it has [EVENT increased] fares by \$6 per round trip on flights to some cities also served by lower-cost carriers. American Airlines, a unit of AMR Corp., immediately [EVENT matched] [EVENT the move], spokesman Tim Wagner [EVENT said]. United, a unit of UAL Corp., [EVENT said] [EVENT the increase] took effect Thursday and [EVENT applies] to most routes where it [EVENT competes] against discount carriers, such as Chicago to Dallas and Denver to San Francisco.

In English events are typically expressed verbs

Temporal ordering of events

Delta Air Lines earnings soared 33% to a record in the fiscal first quarter, bucking the industry trend toward declining profits.

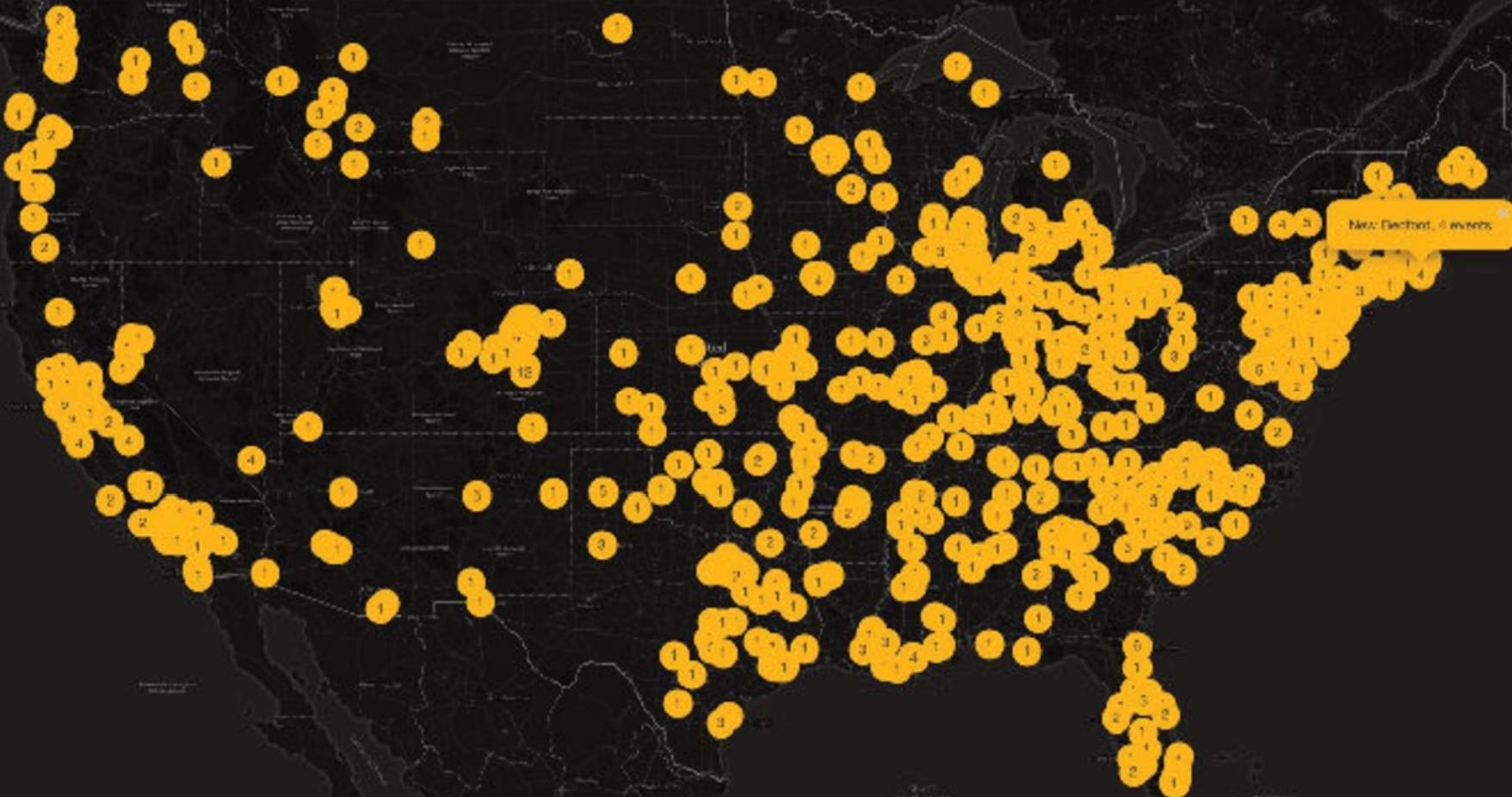
- Soaring_{e1} is **included** in the fiscal first quarter_{t58}
- Soaring_{e1} is **simultaneous** with the bucking_{e3}
- Declining_{e4} **includes** soaring_{e1}

Template Filling

- Scripts are common, stereotypical situations
- Templates represent scripts with a fixed set of slots that take slot-filler values

FARE-RAISE ATTEMPT:	[LEAD AIRLINE:	UNITED AIRLINES
		AMOUNT:	\$6
		EFFECTIVE DATE:	2006-10-26
		FOLLOWER:	AMERICAN AIRLINES

The Gun Violence Database



Ellie Pavlick and Chris Callison-Burch
University of Pennsylvania

Goals of the GVDB

Collect data about gun violence in the US to facilitate **public health research**

Draw sample from local newspapers and television stations that publish online

Use machine learning and crowdsourcing to **extract structured data from text**

Chicago shooting victims

The map below shows where people were shot in Chicago, broken down by community area. The darker the shade of blue, the larger the number of victims. [Read our special report on shootings](#). This data was last updated July 2.



1,043 shooting victims in 2013

VICTIMS BY NEIGHBORHOOD - 2013 - 2017

Fatal Encounters

A 501(c)(3) non-profit organization, Fatal Encounters maintains a comprehensive and transparent database of people killed during interactions with law enforcement.

Slate

How Many People Have Been Killed by Guns Since Newtown?

Slate partners with @GunDeaths for an interactive, crowdsourced tally of the toll firearms have taken since Dec. 14.



By Chris Korn and Dan Korn

The answer to the simple question in that headline is frustratingly hard to come by. So Slate is releasing this the most comprehensive interactive. Our data is necessarily incomplete (click here to contribute), and to learn more about @GunDeaths, the Twitter user who helped us create this interactive, but the more people who are paying attention, the better the outcome will be. You can help by drawing a more complete picture of gun violence in the U.S. If you know about a gun death in your community that isn't represented here, please email a link to a news report to [slatedat@googlemail.com](#). And if you'd like to use this data for your own projects, it's open. You can download it [here](#).

Update, Dec. 18, 2013: After a year-plus, @GunDeaths is closing its project. The count is being picked up by Michael Korn's Gun Violence Archive project, launching soon. Thank you to all who volunteered to make the data as comprehensive and accurate as possible.



SEARCH	REPORT	ABOUT	FEEDBACK	LOG OUT
SEARCH	REPORT	ABOUT	FEEDBACK	LOG OUT
SEARCH	REPORT	ABOUT	FEEDBACK	LOG OUT
SEARCH	REPORT	ABOUT	FEEDBACK	LOG OUT
SEARCH	REPORT	ABOUT	FEEDBACK	LOG OUT

793
4

2016 | 2015 | 1st | Map

Jan 1 - Oct 14, 2016

(40 states)

Deadspin Police-Shooting Database Update: We're Still Going

LOG IN | CONTACT US | Search Database

GUN VIOLENCE Archive



GUN VIOLENCE Archive 2016

Total Number of Incidents	41,309
Number of Deaths ¹	10,595
Number of Injuries ¹	21,023
Number of Children (Age 0-17) Killed or Injured ¹	477
Number of Teens (Age 18-21) Killed or Injured ¹	2,252
Mass Shooting ²	285
Officer Involved Incident	233
Officer Shot or Killed ²	233
Officer Involved Incident Subject Impact Shot or Killed ²	1,359
Home Invasion ²	1,724
Defensive Use ²	1,288
Other Use ²	1,252

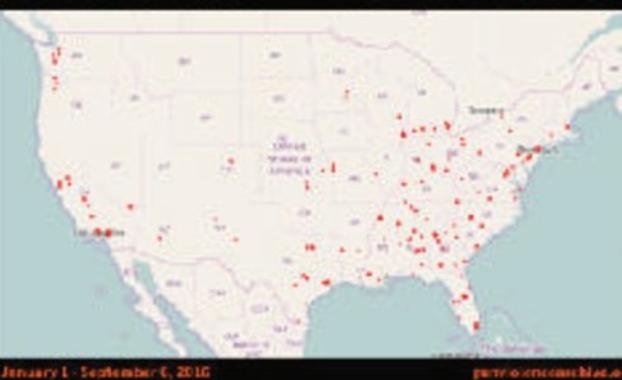
MISSION

Our mission at GVA is to keep the public informed. From 1999 to 2013, we provided free, live public access to accurate information about gun-related violence in the United States. Now we collect and check for accuracy comprehensive information about gun-related violence in 2014, and then send out official reports online.

Read More

CHARTS AND MAPS

GUN VIOLENCE Archive



MASS SHOOTINGS - 2016

Congressman Who Restricted Gun Violence Research Has Regrets

H

The politics of gun control were as divisive in the 1990s as they are today. Republicans had won big in the '94 elections by campaigning against President Bill Clinton's gun control legislation. And in the spring of 1996, the National Rifle Association and its allies set their sights on the Centers for Disease Control and Prevention for funding increasingly assertive studies on firearms ownership and the effects on public health. The gun rights advocates claimed the research veered toward advocacy and covered such logical ground as to be effectively useless.

At first, the House tried to close down the CDC's entire, \$46 million National Center for Injury Prevention. When that failed, Dickey stepped in with an alternative: strip \$2.6 million that the agency had spent on gun studies that year. The money would eventually be re-appropriated for studies unrelated to guns. But the far more damaging inclusion was language that stated, "None of the funds made available for injury prevention and control at the Centers for Disease Control and Prevention may be used to advocate or promote gun control."

Dickey proclaimed victory — an end, he said at the time, to the CDC's attempts "to raise emotional sympathy" around gun violence. But the agency spent the subsequent years petrified of doing *any* research on gun violence, making the costs of the amendment clear even to Dickey himself.

Why the CDC Hasn't Launched a Comprehensive Gun Study in 15 Years

By JULIE BARZILAY, DR. LAURA JOHNSON and GILLIAN MOHNEY · Jun 16, 2016, 4:37 PM ET

[Share with Facebook](#)

[Share with Twitter](#)



Scott Olson/Getty

WATCH | American Medical Association Calls Gun Violence a 'Public Health Crisis'

1K
SHARES

The U.S. [Centers for Disease Control and Prevention](#) studies a variety of public health threats every year, from infectious diseases to automobile safety. But for 15 years, the CDC has avoided comprehensive research on one of the top causes of death in the U.S.: firearms.



Why the
CDC Launched
a Gun Study in 15 Years



Feds Inv
to 'Conso
Pipeline R



What Co
Know Ab
Security I



Teen Wh
Sprayed



Dentist A
Hamming
Making M



I Did the
Orlando



SOME EXPERTS SAY
FIREARMS RESEARCH COULD
SAVE LIVES, NOTING A
DECREASE IN AUTOMOBILE
RELATED DEATHS CAN BE
ATTRIBUTED TO SAFETY
MEASURES DISCOVERED
THROUGH RESEARCH.



National Institutes of Health

Table 1 Major NIH research awards and cumulative morbidity for select conditions in the US, 1973–2002

Condition	Total cases	NIH research awards
Cholera	373	101
Diphtheria	1337	54
Polio	266	106
Rabies	55	59
Total of four diseases	2031	320
Firearm injuries	>3000000	3

Branas, Wiebe, Schwab, Richmond. Getting past the “f” word in federally funded public health research. Injury Prevention 2005; 11:191-192

Want to answer questions like

Gun control - a Guardian investigation



America's gun problem is so
much bigger than mass
shootings

Too much emphasis on mass shootings has a cost

How many deaths result
from mass shootings
compared to other gun
crimes? How has this
changed over time?

America's gun control debate continues to revolve around the exact circumstances of the shooting that is currently on the news. Is a new gun law worth it, or not? That depends on whether it might have prevented this particular shooting. While this is an understandable, human response, it is a terrible way to go about saving lives.

The shock and horror that follows mass shootings has led to a obsessive focus on the dangers of military-style rifles – even though rifles of any kind were used in less than 3% of gun murders in 2014, according to FBI data.

A tunnel focus on mass shootings has also fueled the public

Want to answer questions like



THE UPSHOT | The Science Behind Suicide Contagion

PUBLICITY AND PUBLIC HEALTH

The Science Behind Suicide Contagion



Margot Sanger-Katz @sangerkatz AUG. 13, 2014

When Marilyn Monroe died in August 1962, with the cause listed as probable suicide, the nation reacted. In the month that followed, there was extensive news coverage, widespread sorrow and a spike in suicides. According to one study, the suicide rate in the United States [jumped by 12 percent](#) compared with the same month the previous year.

Mental illness is not a communicable disease, but there is a growing body of evidence that [suicide is still contagious](#). Publicized cases of suicide have been repeatedly and definitively linked to an increase in suicide, especially among young people. Analysis suggests that at least 5 percent of youth suicides are influenced by contagion.

How strong is the effect of suicide contagion? Does it change with age, gender? Is it effected by the style of reporting?

Want to answer questions like

The screenshot shows a dark-themed podcast player interface. At the top left is the 'ON [WE] MEDIA' logo. To its right, the text 'Published in On The Media'. Below this is the title 'Racial Bias in Crime Reporting'. Underneath the title are three buttons: a blue 'Listen 4 min' button, a grey '+ Queue' button, and a grey '...' button. To the right of these are three social media sharing icons: Facebook, Twitter, and Email. The overall background is dark grey.

Jun 5, 2015

Summary Transcript

Does the media portray African-Americans differently than Whites in reporting on gun violence?

Research shows the media disproportionately depict African-Americans as criminals, and whites as victims. Brooke speaks with [Nazgol Ghandoosh](#), research analyst at [The Sentencing Project](#), about her study, "[Race and Punishment: Racial Perceptions of Crime and Support for Punitive Policies](#)," which details how media distortions feed our own implicit biases. (And you can take Harvard's Implicit Association Test yourself [here](#).)

Want to answer questions like



U.S. • Trayvon Martin Shooting Fast Facts

Live TV

U.S. Edition

menu

Trayvon Martin Shooting Fast Facts

CNN Library

Updated 4:25 PM ET, Sun February 7, 2016



Can we predict events that will become politically relevant touchstone events?

The New York Times

Freddie Gray Case Ends With No Convictions of Any Police Officers

Require detailed, local data

Time and Place

City
State
Other details (home, school, etc.)
Date
Clock Time
Time of day

Alleged Shooter(s)

Name
Gender
Age
Race

Victim(s)

Name
Gender
Age
Race
Was the victim injured?
Was the victim hospitalized?
Was the victim killed?

Circumstances of shooting

Type of gun
Number of shots fired
Answer Yes/No/Not able to determine
The shooter and the victim knew each other.
The incident was a case of domestic violence.
The firearm was used during another crime.
The firearm was used in self defense.
Alcohol was involved.
Drugs (other than alcohol) were involved.
The shooting was self-directed.
The shooting was a suicide or suicide attempt.
The shooting was unintentional.
The shooting was by a police officer.
The shooting was directed at a police officer.
The firearm was stolen.
The firearm was owned by the victim/victim's family

The End of the Gun Report

▶ Listen 10 min

+ Queue

...



Jennifer Mascia described how she wrote the Gun Report in an NPR interview

JENNIFER MASCIA: Well, I would google “shooting,” “man shot,” “woman shot,” “child shot,” “teen shot” and “accidentally shot”. You know, this was all day one coverage of shootings, so a lot of times the details aren’t flushed out. If there was no name and scant details, I had to skip over those. So each day, there’d be about 35 to 40 shootings that I would present.

Title	Description	Url	Source	Phrase
Man Shot Near Pierce Park in Coral Gables: Police	A man is in the hospital recovering after police say he was shot multiple times near a park in Coral Gables. The incident happened just before 3 a.m. Thursday outside Pierce Park, located at 101 Oak Avenue. Coral Gables police say the victim, a man in his ...	http://www.nbciami.com/news/local/Man-Shot-Near-Pierce-Park-in-Coral-Gables-Police-360424171.html	NBC Universal Media	man shot
I'm Voting for Hillary Clinton Because She's a Woman	"New drinking game: take a shot every time Hillary says 'as a woman' or 'as the first woman president,'" quips a straight white male on Facebook. This comment was part of a larger thread of young male Democrats discussing why Bernie Sanders is a better ...	http://www.huffingtonpost.com/jillian-gutewitz/im-voting-for-hillary-clinton-because-shes-a-woman_b_8634910.html	The Huffington Post	women shot
Child attends school despite allegedly being stabbed by guardian - FOX10 News WALA	... police said a 12-year-old child ran away from home after originally saying the child had been abandoned. A suspect was killed by officers after police said he shot at them while trying escape authorities in downtown Atlanta. A suspect was killed by ...	http://www.fox10tv.com/story/30649388/child-attends-school-despite-allegedly-being-stabbed-by-guardian-mother	WALA-TV FOX10	child shot
Palestinian who attacked soldier shot dead	The organisation has not claimed responsibility for killing the Henkins, who were shot in front of their young children as they drove on a West Bank road between the northern colonies of Itamar and Elon Moreh. The October 1 shooting was followed two days ...	http://gulfnews.com/news/mana/palestine/palestinian-who-attacked-soldier-shot-dead-1.1631076	Gulf News	child shot
12-year-old child abandoned at Atlanta police station	A 6-year-old child is hospitalized after allegedly being stabbed in the face by his mother Wednesday morning. A suspect was killed by officers after police said he shot at them while trying escape authorities in downtown Atlanta. More >> A suspect was ...	http://www.walb.com/story/30654553/12-year-old-child-abandoned-at-atlanta-police-station	WALB 10 News	child shot

The Gun Report: Training data

http://www.mlive.com/news/kalamazoo/index.ssf/2014/04/kalamazoo_teenager_13_shot_and.html



<http://www.jsonline.com/news/crime/new-developments-in-playground-shooting-to-be-announced-at-430-pm-b99278118z1-260682381.html>



http://www.mlive.com/news/kalamazoo/index.ssf/2014/05/fighting_led_up_to_fatal_shoot.html



http://www.mlive.com/news/kalamazoo/index.ssf/2014/05/michael_day_kalamazoo.html



http://www.mlive.com/news/kalamazoo/index.ssf/2014/05/15-year-old_charged_with_murde.html



<http://www.jsonline.com/news/crime/girl-10-on-life-support-after-being-hit-in-playground-shootout-b99275748z1-260251491.html>



gun-violence.org/hit/3?name=Elon Musk X gun-violence.org/groups X +

gun-violence.org/groups Search

Gun Violence DATABASE BROWSE MAP ABOUT PROJECT WORK ON TASKS DOWNLOAD DATA Derry W ▾

Work on tasks

Read more about our tasks and select how you want to contribute.

Extract information

We want data that is easy for researchers to search and study. We need your help reading articles about gun violence and extracting key pieces of information (such as the location of the shooting or the name and age of the victim).

Total submitted tasks: 21839
Total available tasks: 41937

[Go to task](#)

Scan the Headlines

Read headlines and tell us which ones describe incidents of gun violence.

Total submitted tasks: 6496
Total available tasks: 5144

[Go to task](#)

Combine records

We want to have as complete a database as possible. When there are multiple reports of the same incident, we want to combine the information from all the articles so nothing is left out. You can help by comparing two records and deciding which information is best to keep in the database.

Total submitted tasks: 305
Total available tasks: 0

[Go to task](#)

gun-violence.org/hit/3?name=Elaine X gun-violence.org/hit/1?name=Scan the Headlines +

gun-violence.org/hit/1?name=Scan the Headlines Search

Gun Violence DATABASE BROWSE MAP ABOUT PROJECT WORK ON TASKS DOWNLOAD DATA Derry W ▾

10-year-old Girl Shot in Cassia County Hunting Accident

Show Article

Yes No Unclear

Humans manually verify the predictions of the classifier.

Show Article

Yes No Unclear

KAKE.com | Wichita, Kansas News, Weather, Sports - Two people hit by vehicle near downtown Wichita

Show Article

Yes No Unclear

Parks & Recreation – City of Dunsmuir

Show Article

Yes No Unclear

gun-violence.org/hit/3?name=Elaine X gun-violence.org/hit/1?name=Scan the Headlines +

gun-violence.org/hit/1?name=Scan the Headlines Search

Gun Violence DATABASE BROWSE MAP ABOUT PROJECT WORK ON TASKS DOWNLOAD DATA Derry W ▾

10-year-old Girl Shot in Cassia County Hunting Accident

Yes No Unclear

Show Article

2 killed, 12 wounded in hunting accident in Wednesday night

Unclear

Show Article

Community News CentreDaily.com

Unclear

Show Article

KAKE.com | Wichita Falls Sports - Two teenagers shot downtown Wichita Falls

Unclear

Show Article

Parks & Recreation – City of Dunsmuir

Yes No Unclear

Show Article

10-year-old Girl Shot in Cassia County Hunting Accident

[www.kjonline.com/news
/Reward_offered_in_Pa_shooting_death_of_Maine_boy](http://www.kjonline.com/news/Reward_offered_in_Pa_shooting_death_of_Maine_boy)

A 10-year-old girl who was hunting with family members Sunday near Lake Walcott was shot in the face when she fell while carrying a .30-06 bolt-action 410 shotgun. Cedar Glaesemann was taken by ambulance to Minidoka Memorial Hospital, said Cassia County Undersheriff George Warrell. She later was transported to Primary Children's Medical Center in Salt Lake City, Utah. Medical center spokeswoman Bonnie Midget said she couldn't reach the family Tuesday to get permission to release Glaesemann's medical condition. "We haven't heard anything about her condition since she was taken to Primary's," said Warrell. The 911 call came in shortly after 10 a.m. Sunday, said a sheriff's news release. Deputies investigated and ruled it an accidental shooting. The sheriff's office urges everyone this hunting season to be safe when handling firearms, the statement said.

gun-violence.org/hit/3?name=... X gun-violence.org/groups X +

gun-violence.org/hit/3?name=Extract information

... ⌂ ⌂ Search

Gun Violence DATABASE BROWSE MAP ABOUT PROJECT WORK ON TASKS DOWNLOAD DATA Derry W ▾

PUBLICATION DATE: 02-28-2015

Greenly shooting victim in critical condition

Greeley, Colorado- Greeley police are investigating a shooting after an adult man showed up at North Colorado Medical Center with a gunshot wound Thursday morning, February 28. Lt. Carl Alm of the Greeley police said the man, 34 year old John Taylor, arrived at NCMC at 7:20 a.m. Doctors said he had been shot only two hours earlier, around 5:10 a.m. on Thursday. Taylor was in surgery Thursday afternoon, and his injuries did not appear life-threatening. Police say they received a call on suspicious activity in which people were loading an unconscious man into the back of a pickup at an apartment complex in the 1100 block of 26th Avenue just after 5 a.m. Thursday morning. We had the call at the shooting scene about the same time, and put two and two together, and said OK, this is the guy who got loaded into the truck, Alm said. We got officers on the ground, started knocking on doors, and traced it to one specific apartment. When he arrived at NCMC, the man had a gunshot

Please read the text carefully, and then select an answer for all questions. Please base your answers only on information that is explicitly stated or can be confidently inferred from the text of the article.

The shooter and the victim knew each other.
(Answer yes if the shooting was self-directed.)

Yes No Not Mentioned

The incident was a case of domestic violence.

Yes No Not Mentioned

The firearm was used during another crime.

Yes No Not Mentioned

The firearm was used in self defense.

Yes No Not Mentioned

Back Next

First answer a series of binary questions about the circumstances of the shooting....

gun-violence.org/hit/3?name=Extract information

gun-violence.org/groups

... Search

Gun Violence
DATABASE

BROWSE MAP ABOUT PROJECT WORK ON TASKS DOWNLOAD DATA

Derry W ▾

Please read the instructions before you start working on this HIT

PUBLICATION DATE: 02-28-2015

Greenly shooting victim in critical condition

Greeley, Colorado- Greeley police are investigating a shooting after an
a Clear this highlight
w
s
D
o
su

City

Clock Time (1p.m., 2:37a.m)

Time Of Day (Morning, Evening)

First, try to figure out the date of the described event, and select it by clicking on the calendar icon. The publication date and the day of week mentioned in article are helpful in determining the date of the shooting. Next, click on parts of the text that correspond to the other information listed below, if that information is present in the article. When you highlight a passage of text in the article, you will get a dropdown menu that lets you select which question it answers.

Date

02-28-2015

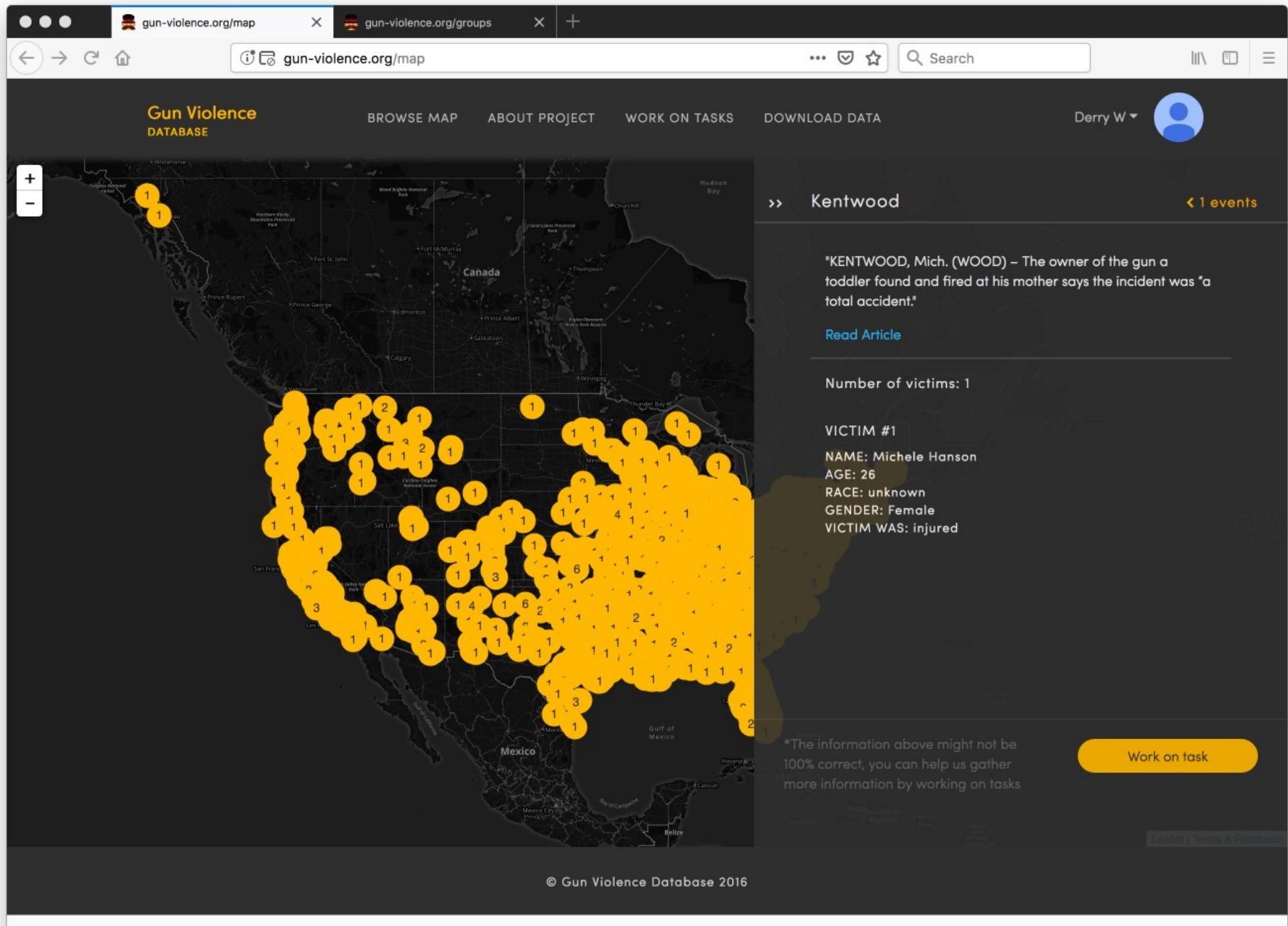
State

CO - Colorado

City

Clock Time (1p.m., 2:37a.m)

...then extract structured information from text to populate the database.



Automating The Pipeline

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and the point when Laquan McDonald falls to the ground. The video, released to the public for the first time late Tuesday, is a key piece of evidence in a case that's sparked protests in Chicago and has landed an officer behind bars. The 17-year-old McDonald was shot 16 times on that day the video shows in October 2014. Chicago police Officer Jason Van Dyke was charged Tuesday with first-degree murder....

Incident #1053

City	
Date	
Shooter	
Victim	
Victim Killed	

Person #1014

Name	
Gender	
Age	
Race	

Automating The Pipeline

Named Entity

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and the point where Laquan McDonald falls to the ground. The video, released to the public for the first time late Tuesday, is a key piece of evidence in a case that's sparked protests in Chicago and has landed an officer behind bars. The 17-year-old McDonald was shot 16 times on that day the video shows in October 2014. Chicago police Officer Jason Van Dyke was charged Tuesday with first-degree murder....

Recognition

Person #1014

Name	Laquan McDonald
Gender	
Age	
Race	

- Named Entity Recognition without Gazetteers.
Mikheev et al (1999).
- Neural Architectures for Named Entity Recognition. Lample et al (2016).

Automating The Pipeline

Event

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and the point when Laquan McDonald falls to the ground. The video, released to the public for the first time late Tuesday, is a key piece of evidence in a case that's sparked protests in Chicago and has landed an officer behind bars. The 17-year-old McDonald was shot 16 times on that day the video shows in October 2014. Chicago police Officer Jason Van Dyke was charged Tuesday with first-degree murder....

Detection

Person #1014

Name	Laquan McDonald
Gender	
Age	
Race	

Incident #1053

City	
Date	
Shooter	
Victim	
Victim Killed	

- Joint event extraction via structured prediction. Li et al (2013).
- Event Detection and Domain Adaptation with Convolutional Neural Networks. Nguyen and Grishman (2015).

Automating The Pipeline

Semantic Role

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and the point when Laquan McDonald falls to the ground. The video, released to the public for the first time late Tuesday, is a key piece of evidence in a case that's sparked protests in Chicago and has landed an officer behind bars. The 17-year-old McDonald was shot 16 times on that day the video shows in October 2014. Chicago police Officer Jason Van Dyke was charged Tuesday with first-degree murder....

Labeling

Person #1014

Name	Laquan McDonald
Gender	
Age	
Race	

Incident #1053

City	
Date	
Shooter	
Victim	McDonald
Victim Killed	

- Semantic Role Labeling. Palmer et al (2010).
- Joint event extraction via structured prediction with global features. Li et al (2013).

Automating The Pipeline

Entity Coreference

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and the point where **Laquan McDonald** falls to the ground. The video, released to the public for the first time late Tuesday, is a key piece of evidence in a case that's sparked protests in Chicago and has landed an officer behind bars. The 17-year-old **McDonald** was shot 16 times on that day the video shows in October 2014. Chicago police Officer Jason Van Dyke was charged Tuesday with first-degree murder....

Person #1014

Name	Laquan McDonald
Gender	
Age	
Race	

Incident #1053

City	
Date	
Shooter	
Victim	Person #1014
Victim Killed	

- Coreference Resolution in a Modular, Entity-Centered Model. Haghghi and Klein (2010).
- Using Wikitology for Cross-Document Entity Coreference Resolution. Finin et al (2009).

Automating The Pipeline

Event Coreference

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and

Police release video of officer shooting teen | Oklahoma City

Protesters took to the streets of Chicago late Tuesday after police released a video showing an officer shooting 17-year-old Laquan McDonald... McDonald was a black teenager. The officer who shot him, Jason Van Dyke, is white. He was charged Tuesday with first-degree murder in McDonald's death...

Person #1014

Name	Laquan McDonald
Gender	
Age	
Race	

Incident #1053

City	
Date	
Shooter	
Victim	Person #1014
Victim Killed	

- Cross-document Event Coreference Resolution based on Cross-media Features. Zhang et al (2015).

Automating The Pipeline

Cross-Document

Entity Coref.

Person #1014

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and

Police release video of officer shooting teen | Oklahoma City

Protesters took to the streets of Chicago late Tuesday after police released a video showing an officer shooting 17-year-old Laquan McDonald. McDonald was a black teenager. The officer who shot him, Jason Van Dyke, is white. He was charged Tuesday with first-degree murder in McDonald's death...

- Cross-document Event Coreference Resolution based on Cross-media Features. Zhang et al (2015).

Name	Laquan McDonald
Gender	
Age	
Race	

Incident #1053

City	
Date	
Shooter	
Victim	Person #1014
Victim Killed	

Automating The Pipeline

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and

Police release video of officer shooting teen | Oklahoma City

Protesters took to the streets of Chicago late Tuesday after police released the video showing an officer shooting a 17-year-old, Laquan McDonald. McDonald was a black teenager. The officer who shot him, Jason van Dyke, is white. He was charged Tuesday with first-degree murder in McDonald's death...

Semantic Parsing

Person #1014

Name	Laquan McDonald
Gender	
Age	17
Race	Black

Incident #1053

City	
Date	
Shooter	
Victim	Person #1014
Victim Killed	

- Cross-document Event Coreference Resolution based on Cross-media Features. Zhang et al (2015).

Automating The Pipeline

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and

Police release video of officer shooting teen | Oklahoma City

Protesters took to the streets of Chicago late Tuesday after police released the video showing an officer shooting a 17-year-old, Laquan McDonald... McDonald was a black teenager. The officer who shot him, Jason van Dyke, is white. He was charged Tuesday with first-degree murder in McDonald's death...

Semantic Parsing

Person #1014

Name	Laquan McDonald
Gender	
Age	17
Race	Black

Incident #1053

City	
Date	
Shooter	
Victim	Person #1014
Victim Killed	TRUE

- Cross-document Event Coreference Resolution based on Cross-media Features. Zhang et al (2015).

NLP Systems Require Training Data

- Most existing training datasets consist of factoid and pop-culture questions

NLP Systems Require Training Data

- Most existing training datasets consist of factoid and pop-culture questions
 - *What is the cover price of the X-Men comic book?* (Berant et al., 2013)
 - *How are Kim Kardashian and Kanye West related?* (Wijaya, et al., 2015)

NLP Systems Require Training Data

- Most existing training datasets consist of factoid and pop-culture questions
 - *What is the cover price of the X-Men comic book?* (Berant et al., 2013)
 - *How are Kim Kardashian and Kanye West related?* (Wijaya, et al., 2015)
- Usually assume entities are celebrities who have Wikipedia pages. (Singh et al., 2012)

NLP Systems Require Training Data

- Most existing training datasets consist of factoid and pop-culture questions
 - *What is the cover price of the X-Men comic book?* (Berant et al., 2013)
 - *How are Kim Kardashian and Kanye West related?* (Wijaya, et al., 2015)
- Usually assume entities are celebrities who have Wikipedia pages. (Singh et al., 2012)
- This leads makes it hard to extract the necessary data for in-depth social science research

gun-violence.org/hit/3?name=Extract information

gun-violence.org/groups

... Search

Gun Violence
DATABASE

BROWSE MAP ABOUT PROJECT WORK ON TASKS DOWNLOAD DATA Derry W ▾

Please read the instructions before you start working on this HIT

PUBLICATION DATE: 02-28-2015

Greenly shooting victim in critical condition

Greeley, Colorado- Greeley police are investigating a shooting after an ad Clear this highlight w sda. City D on su At shi. Clock Time (1p.m., 2:37a.m) di su Time Of Day (Morning, Evening) th At Additional Location Details shi. shooting scene about the same time, and put two and two together, and

First, try to figure out the date of the described event, and select it by clicking on the calendar icon. The publication date and the day of week mentioned in article are helpful in determining the date of the shooting. Next, click on parts of the text that correspond to the other information listed below, if that information is present in the article. When you highlight a passage of text in the article, you will get a dropdown menu that lets you select which question it answers.

Date

02-28-2015

State

CO - Colorado

City

Clock Time (1p.m., 2:37a.m)

Extracted data can provide training data for semantic parsing systems...

gun-violence.org/hit/4?name=... X gun-violence.org/groups X +

gun-violence.org/hit/4?name=Compare Articles ... ⌂ ⌂ ⌂ Search

Gun Violence DATABASE BROWSE MAP ABOUT PROJECT WORK ON TASKS DOWNLOAD DATA Derry W ⌂

Read the texts below. Are both related to the same event?

Are both texts related to the same event?

?

Yes No

Submit

Chief: Slain Officer Interrupted Drug Deal | Oklahoma City - OKC - KOCO.com

CRAWLED DATE: 08-03-2015

It started as a regular traffic stop on a weekend night in Tennessee. Officer Sean Bolton saw an illegally parked car on the streets of Memphis, pulled up in front of it and shone his spotlight. When he approached the vehicle, a passenger confronted him. A struggle ensued, and the passenger shot Bolton multiple times, authorities said. The car drove off, leaving the officer for dead. "Officer Bolton apparently interrupted some sort of drug transaction," Memphis Police Director Toney Armstrong said. "A digital scale and a small baggie of marijuana ...

Memphis Officer Shot During Traffic Stop | National News - KETV Home

CRAWLED DATE: 08-03-2015

Community mourns and honors Omaha man killed in motorcycle crash Authorities say Officer Sean Bolton died after he was shot during a traffic stop in Memphis Saturday night. Authorities say Officer Sean Bolton died after he was shot during a traffic stop in Memphis Saturday night.

....event coreference...

gun-violence.org/hit/5?name=Id X gun-violence.org/groups X +

gun-violence.org/hit/5?name=Identify People

... ⌂ ⌂ Search

Gun Violence DATABASE BROWSE MAP ABOUT PROJECT WORK ON TASKS DOWNLOAD DATA Derry W ⌄

Below you can see the list of victims/shooters that we got from two different articles. Please match people from the first article with people from the second article.

?

Read Article ↗

Victim #1a

NAME: unknown
AGE: unknown
RACE: unknown
GENDER: unknown
VICTIM WAS: hospitalized

Read Article ↗

Victim #2a

NAME: unknown
AGE: 23
RACE: unknown
GENDER: Male
VICTIM WAS: injured, hospitalized

Victim #1b

NAME: unknown
AGE: unknown
RACE: unknown
GENDER: unknown
VICTIM WAS: hospitalized

Victim #2b

NAME: unknown
AGE: 24
RACE: unknown

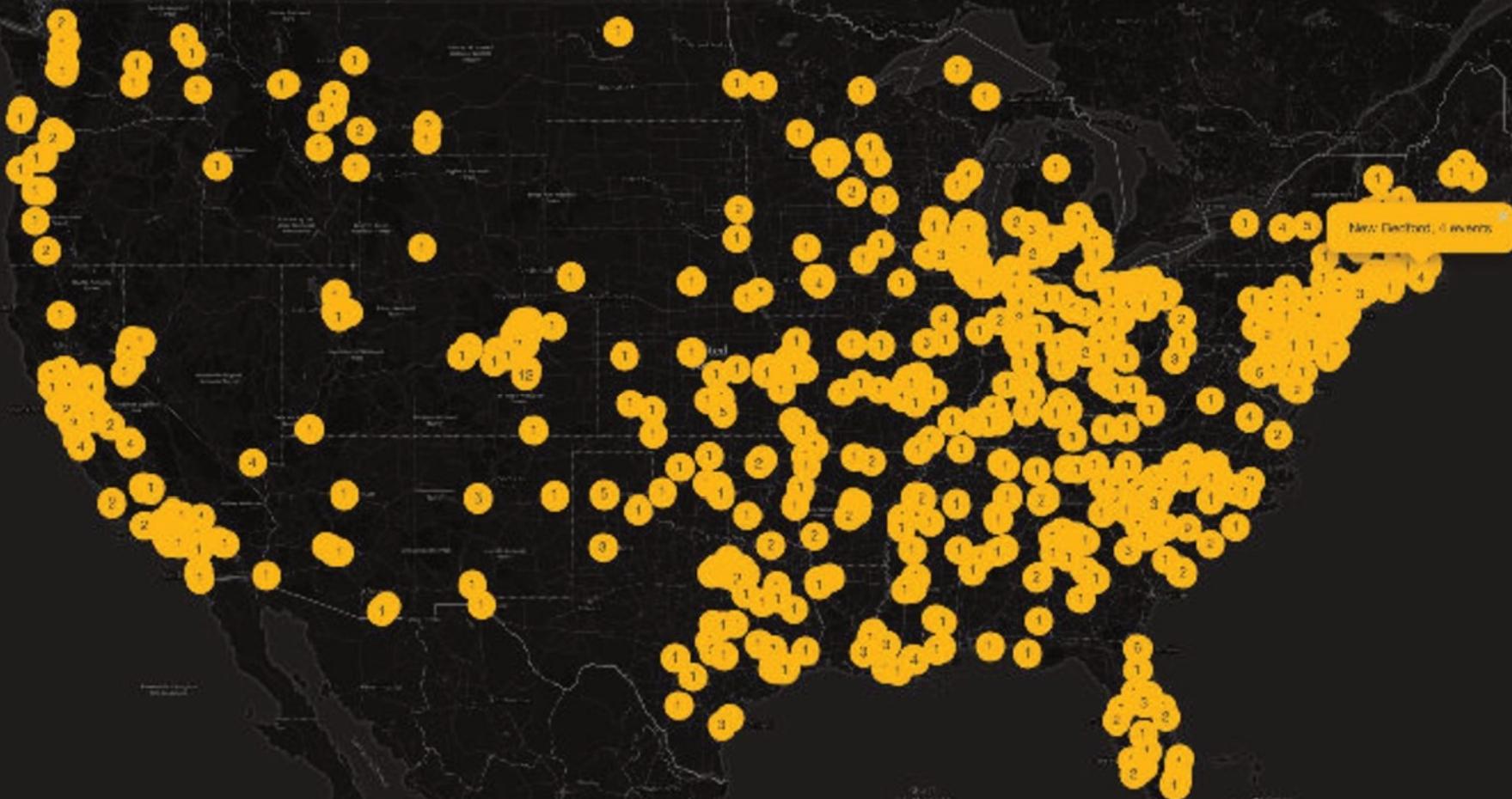
...and entity coreference.

Submit 1/2

The Gun Violence Database

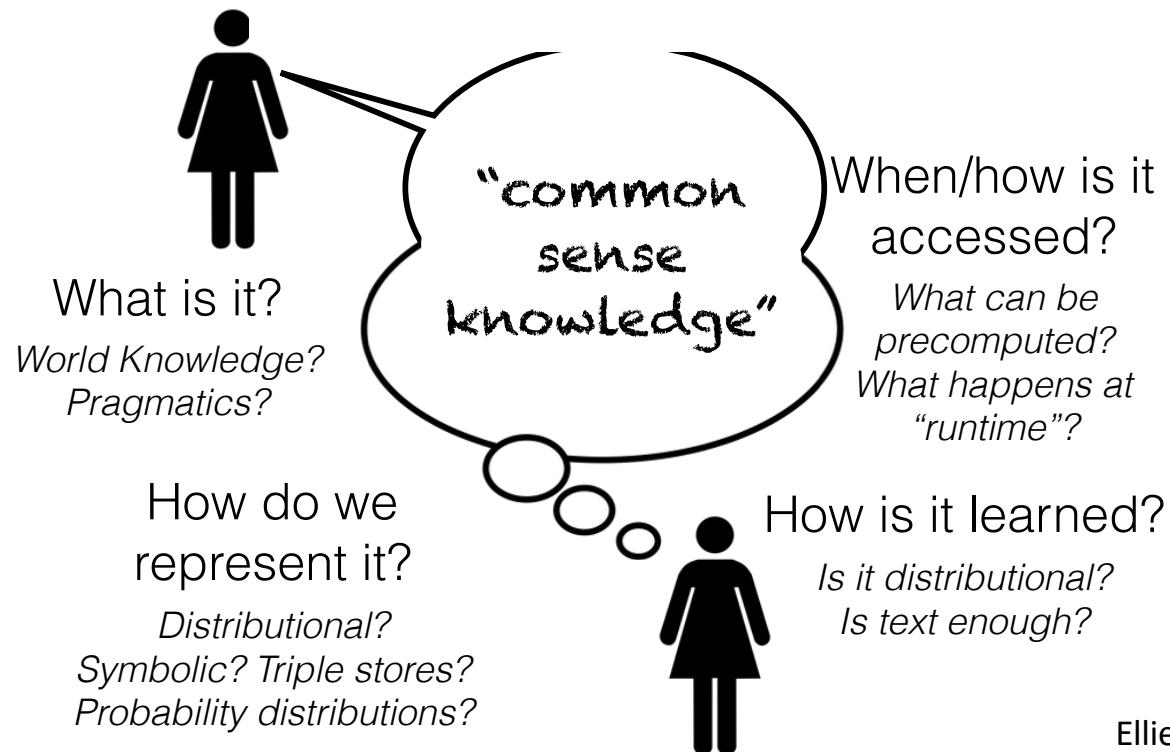
- Use crowdsourcing to collect detailed local data to support research about gun violence in the United States
- Collect data that can be used to train state-of-the-art NLP systems to automate the pipeline, enabling databases to be updated in real time

<http://gun-violence.org>



Ellie Pavlick and Chris Callison-Burch
University of Pennsylvania

Incorporating Knowledge in Understanding



Ellie Pavlick’s Thesis, 2017

Conclusion

- Knowledge is crucial
- Symbolic and/or distributional knowledge
- Inference over knowledge
- Many remaining open research questions
 - What knowledge to represent?
 - How to represent it?
 - How to incorporate knowledge in understanding?

