

Knowledge Graphs: A Practical Introduction across Disciplines

Mayank Kejriwal

University of Southern California

December, 2020

About me

USC Viterbi

School of Engineering

*Daniel J. Epstein Department of
Industrial and Systems Engineering*

USC
Viterbi
School of Engineering
Information
Sciences Institute



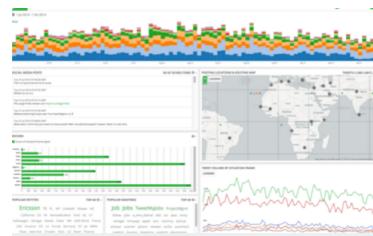
E-Commerce

E-Commerce Knowledge Graphs and
Representation Learning



The Human Trafficking Project

The Human Trafficking Project



AI for Crisis Response

Text-enabled Humanitarian Operations
in Real-time



Common Sense Reasoning

Multi-modal Open World Grounded
Learning and Inference



GNOME

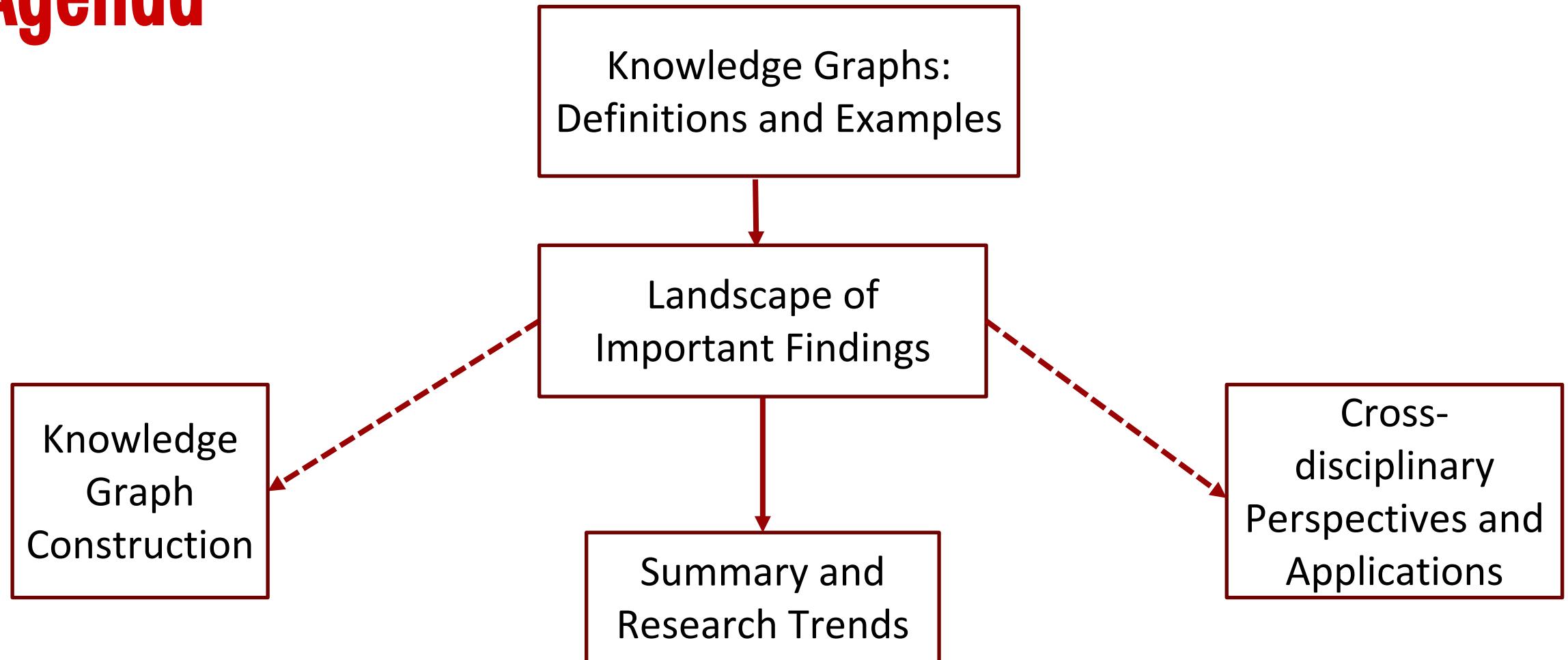
Generating Novelties in Open-world
Multi-agent Environments



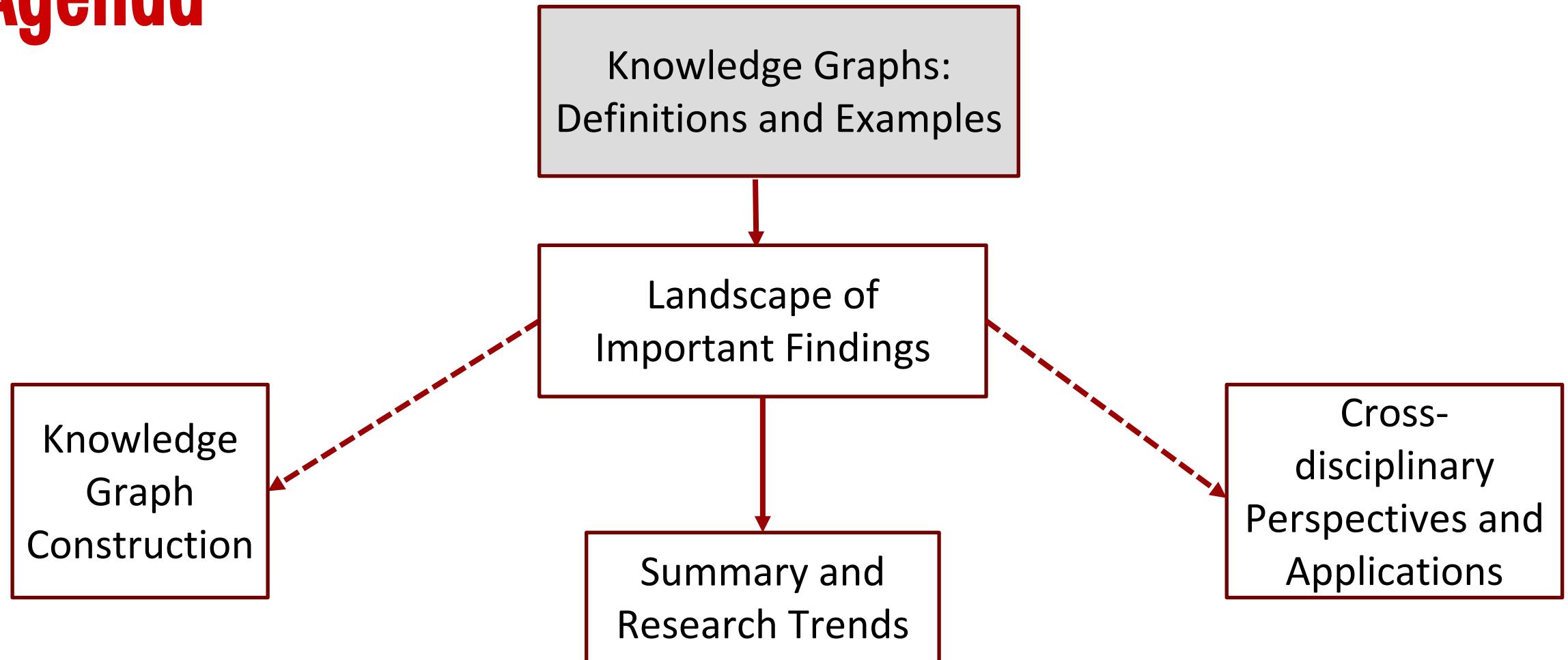
AI, Networks and Society

AI, Networks and Society

Agenda



Agenda



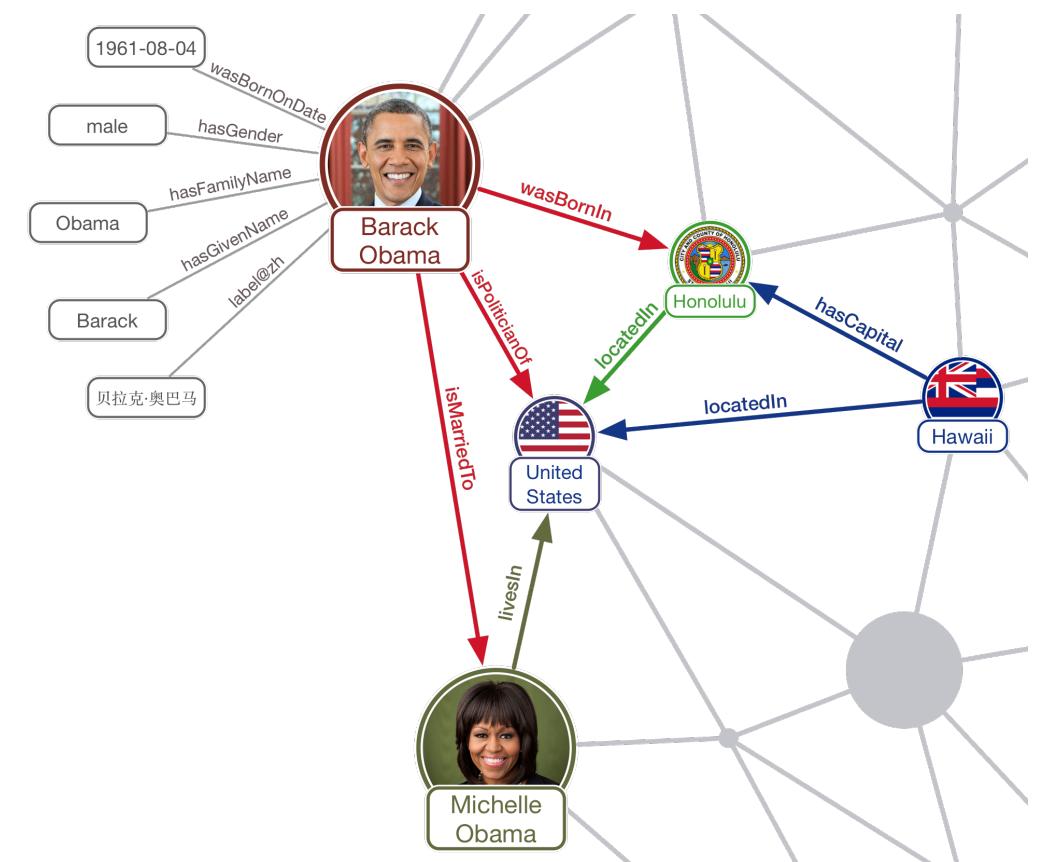
What is a Knowledge Graph?

Set of triples, where each triple (h, r, t) represents a relationship r between head entity h and tail entity t

$(\text{Barack Obama}, \text{wasBornOnDate}, 1961-08-04)$,
 $(\text{Barack Obama}, \text{hasGender}, \text{male})$,

...
 $(\text{Hawaii}, \text{hasCapital}, \text{Honolulu})$,

...
 $(\text{Michelle Obama}, \text{livesIn}, \text{United States})$



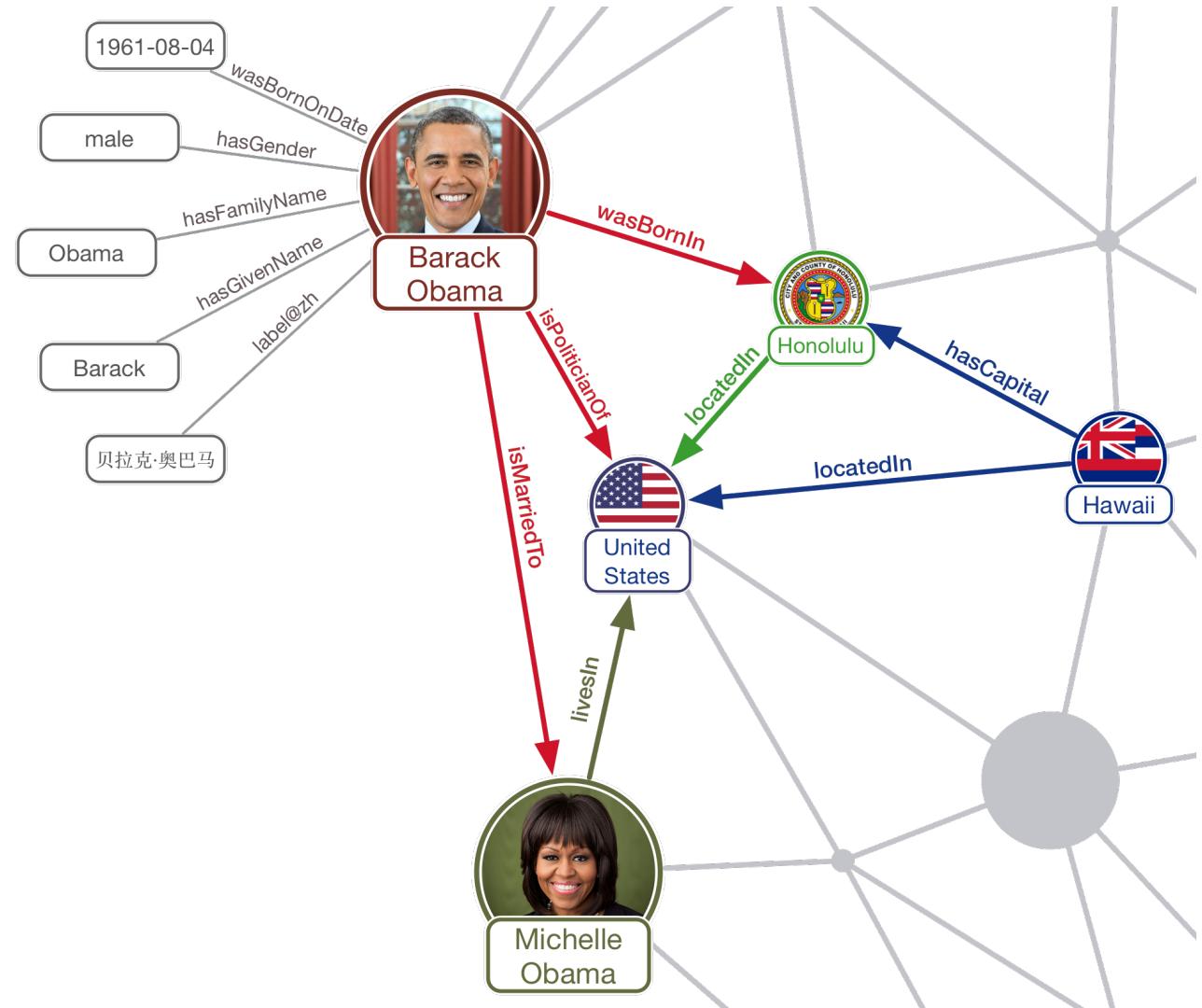
What is a Knowledge Graph?

Technically, a multi-relational directed labeled graph with semantics

Both edges and nodes have labels, but not all labels are equal (literals vs. identifiers)

Where do the semantics come from?

- Complex question, only starting to be understood



More on semantics

Traditionally, semantics are believed to come from ontology

- An ontology is a ‘formal, explicit specification of a shared conceptualization’ (we will go deeper into this in a while)
- In philosophy, an ontology is a ‘study of what there is’ including the study of the ‘most general features of what there is, and how the things there are relate to each other in the metaphysically most general ways’

Source: <https://plato.stanford.edu/entries/logic-ontology/>

More recently, in AI, we have started to recognize a more commonsense view of semantics guided by findings in linguistics and distributional semantics

places to visit san jose

All Maps Images News Shopping More Settings Tools

About 77,800,000 results (1.02 seconds)

Top things to do in San Jose

Winchester Mystery House Quirky mansion with odd design details	The Tech Interactive Interactive displays & an IMAX theater	Mission Peak Mountain with an iconic summit pole	Happy Hollow Park & Zoo Animals, activities & conservation focus
More things to do			

Things To Do & Places To See in San Jose | Visit San Jose

<https://www.sanjose.org/things-to-do> ▾
 Vibrant, diverse and accessible, San Jose shines with 300+ days of sunshine that assures indoor and outdoor activities for every inclination. Whether it's nightlife ...
 Attractions and Amusement · Kids and Family · San Jose Wineries · Nightlife

Complete List of San Jose Attractions | Visit San Jose

<https://www.sanjose.org/things-to-do/attractions-and-amusement> ▾
 Find a complete list of San Jose attractions and things to do - perfect for tourists and ... Browse the best places to see in San Jose and the surrounding Bay Area.

25 Best Things to Do in San Jose (CA) - The Crazy Tourist

<https://www.thecrazytourist.com> › ... › United States › California (CA) ▾
 25 Best Things to Do in San Jose (CA): Winchester Mystery House: Flickr. Tech Museum Of Innovation: Flickr. Children's Discovery Museum: Flickr. Rosicrucian Egyptian Museum: Flickr. San Jose Heritage Rose Garden: wikimedia. Basilica Of St. Joseph: Flickr. Alum Rock Park: Flickr. Happy Hollow Zoo Entrance: Flickr.

THE 15 BEST Things to Do in San Jose - 2019 (with Photos ...

https://www.tripadvisor.com/Attractions-g33020-Activities-San_Jose_Calif... ▾
 ... attractions. Find what to do today, this weekend, or in August. We have reviews of the best places to see in San Jose. Visit top-rated & must-see attractions.
 Free Entry (36) · Things to Do in San Jose · Museums in San Jose · Santana Row

San Jose
City in California

San Jose is a large city surrounded by rolling hills in Silicon Valley, major technology hub in California's Bay Area. Architectural landmarks, from the 1883 Italianate-style Oddfellows building to Spanish Colonial Revival structures, make up the downtown historic district. The downtown area is also home to the Tech Museum of Innovation, devoted to the exploration of science and technology.

Weather: 64°F (18°C), Wind NW at 10 mph (16 km/h), 85% Humid
Population: 1.035 million (2017)

Plan a trip

- San Jose travel guide
- 3-star hotel averaging \$206
- 1 h 5 min flight, from \$97

Did you know: San Jose, California has the largest Vietnamese-American population (106,992) among all U.S. cities. [wikipedia.org](https://en.wikipedia.org)

People also search for

San Francisco	Santa Clara County	California	San Diego	San Francisco Bay Area

[View 15+ m](#)

Have I seen this before?

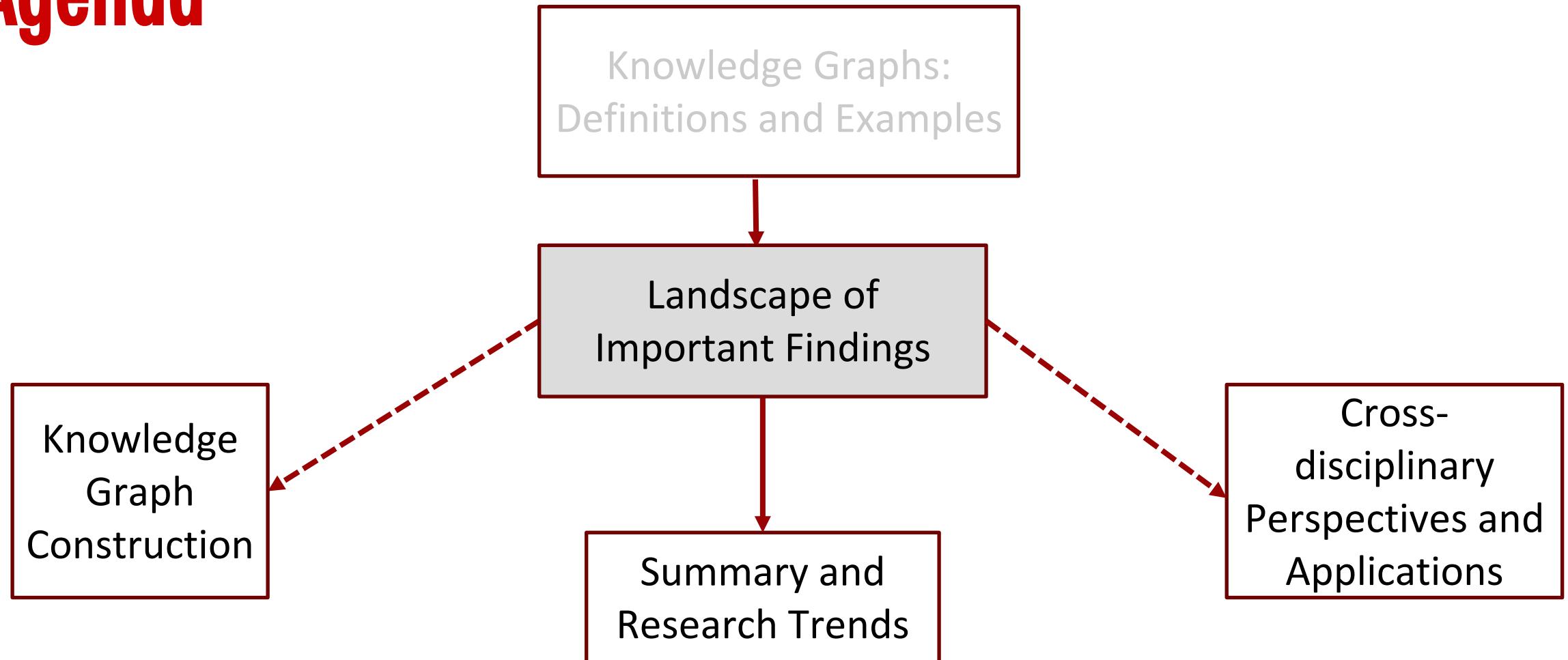
Knowledge panel

Recognition of user intent

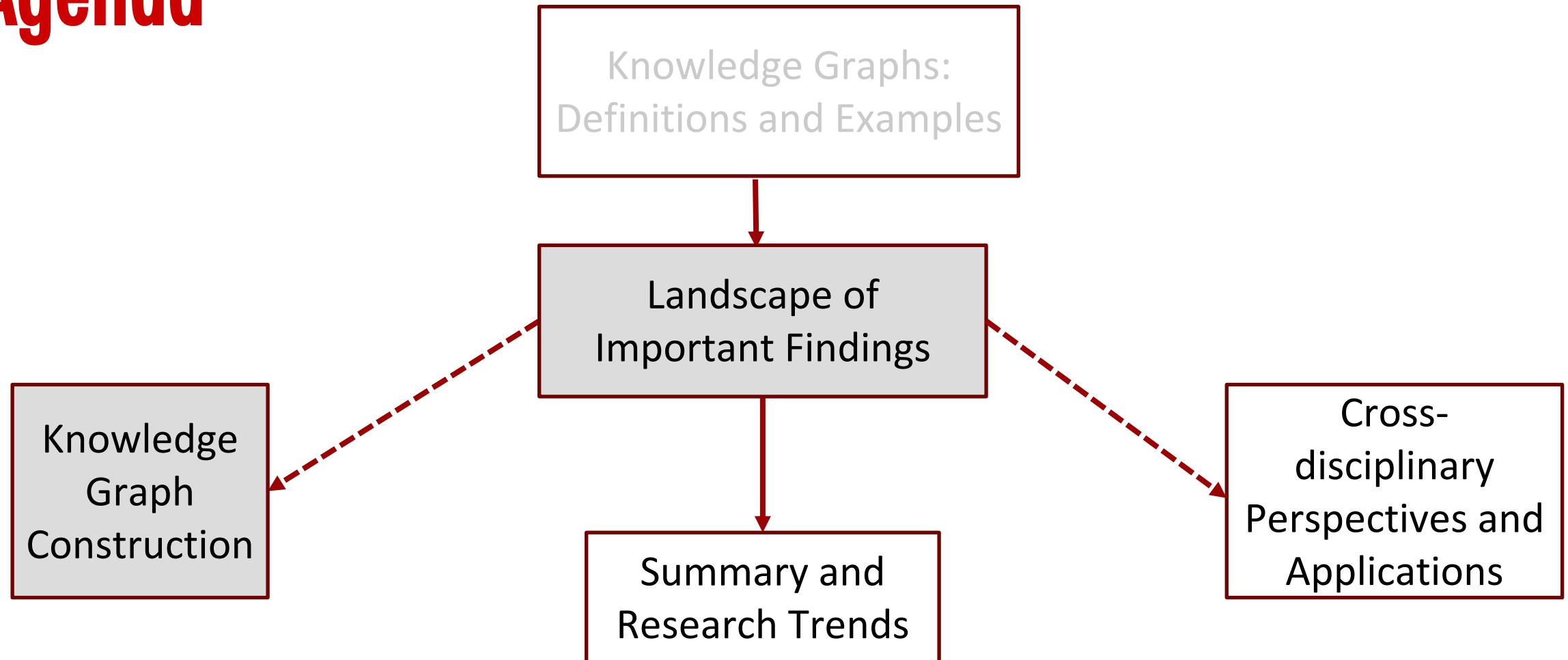
Recommendations

Exploration-suggestions

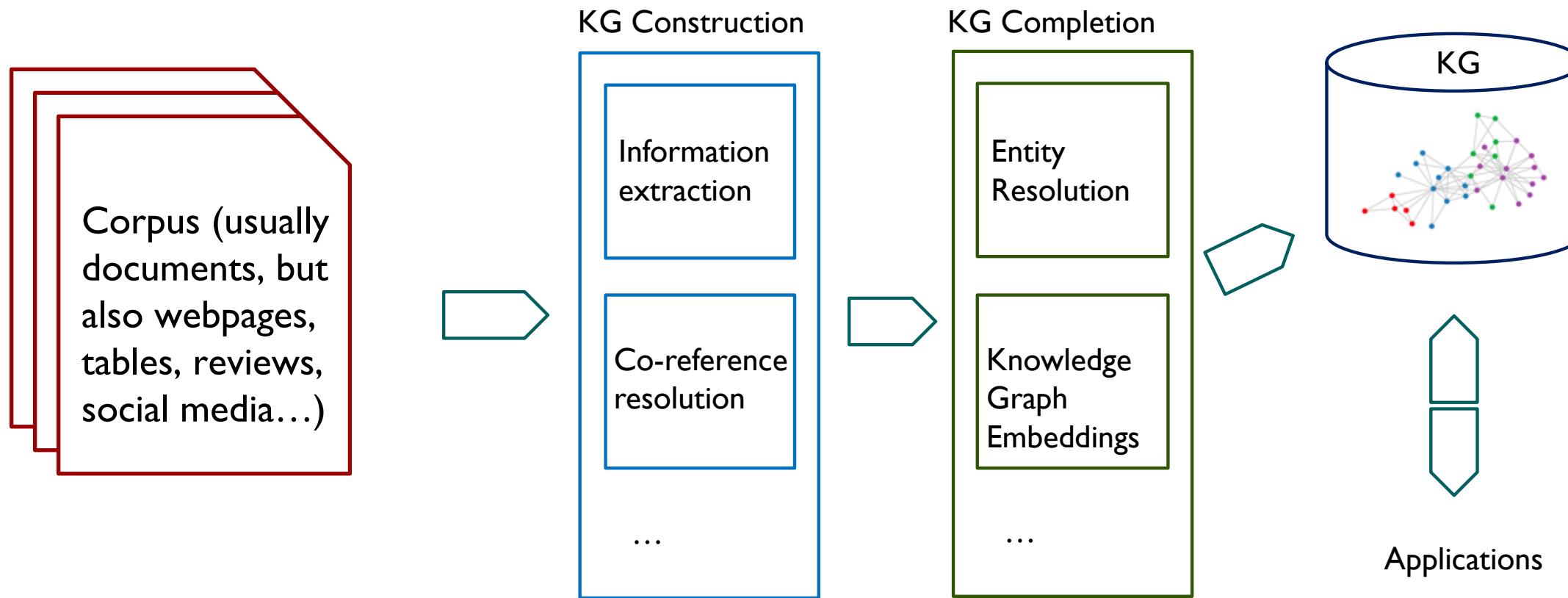
Agenda



Agenda



A typical KGC workflow starts from corpus acquisition and ends with applications



INFORMATION EXTRACTION (IE)

Named Entity Recognition (NER)

contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday's PaperAdvertisementSupported ORG byF.B.I. Agent Peter Strzok PERSON , Who Criticized Trump PERSON in Texts, Is FiredImagePeter Strzok, a top F.B.I. GPE counterintelligence agent who was taken off the special counsel investigation after his disparaging texts about President Trump PERSON were uncovered, was fired. CreditT.J. Kirkpatrick PERSON for The New York TimesBy Adam Goldman ORG and Michael S. SchmidtAug PERSON . 13 CARDINAL , 2018WASHINGTON CARDINAL — Peter Strzok PERSON , the F.B.I. GPE senior counterintelligence agent who disparaged President Trump PERSON in inflammatory text messages and helped oversee the Hillary Clinton PERSON email and Russia GPE investigations, has been fired for violating bureau policies, Mr. Strzok PERSON 's lawyer said Monday DATE .Mr. Trump and his allies seized on the texts — exchanged during the 2016 DATE campaign with a former F.B.I. GPE lawyer, Lisa Page — in PERSON assailing the Russia GPE investigation as an illegitimate "witch hunt." Mr. Strzok PERSON , who rose over 20 years DATE at the F.B.I. GPE to become one of its most experienced counterintelligence agents, was a key figure in the early months DATE of the inquiry.Along with writing the texts, Mr. Strzok PERSON was accused of sending a highly sensitive search warrant to his personal email account.The F.B.I. GPE had been under immense political pressure by Mr. Trump PERSON to dismiss Mr. Strzok PERSON , who was removed last summer DATE from the staff of the special counsel, Robert S. Mueller III PERSON . The president has repeatedly denounced Mr. Strzok PERSON in posts on

Source: Named Entity Recognition and Classification with Scikit-Learn. <https://towardsdatascience.com/named-entity-recognition-and-classification-with-scikit-learn-f05372f07ba2>

Demo: displaCy

<https://explosion.ai/demos/displacy-ent>

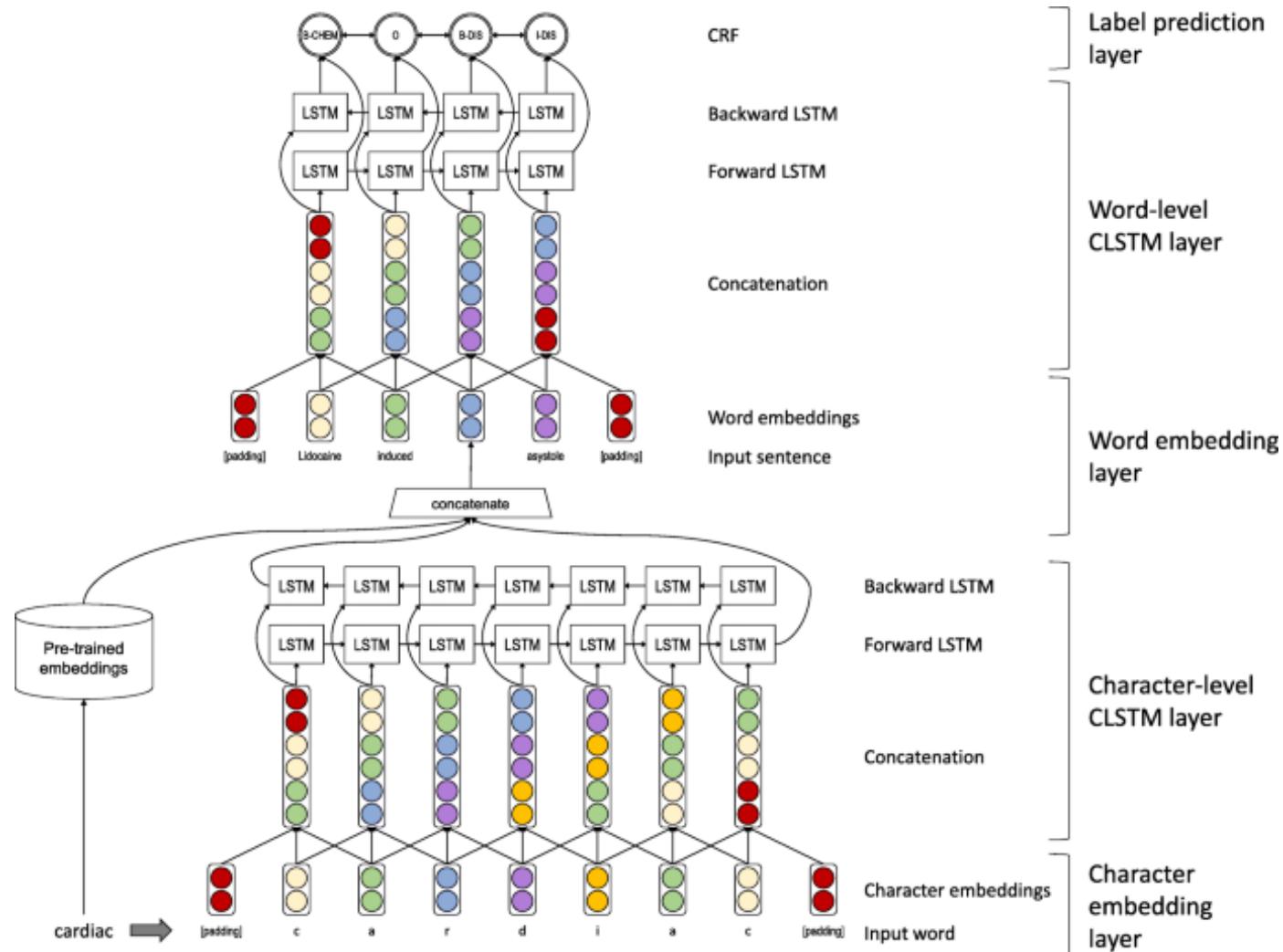
NER workflows

Many methods proposed over the previous 3-4 decades:

- Rule-based
- Dictionary-based
- Simple machine learning
- Sequence labeling (e.g., using conditional random fields or, before that, hidden Markov models)

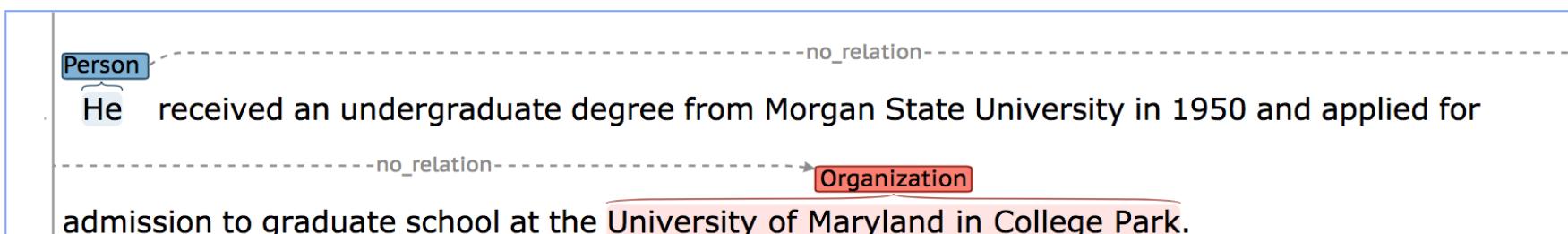
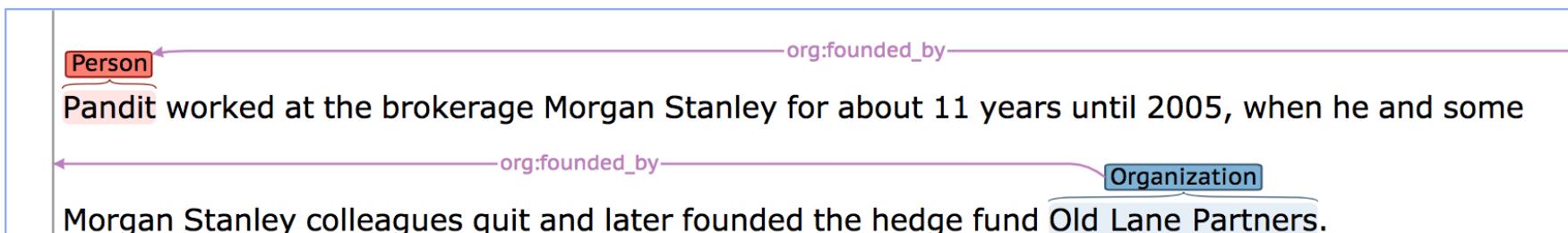
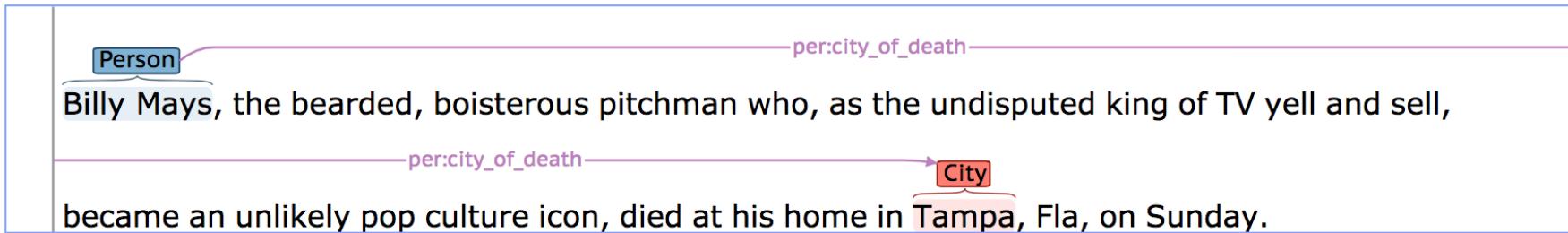
Today, deep learning methods designed for sequences (such as RNNs and, more recently, transformers) are state-of-the-art

Much research still remains (especially for social media!)



Source: Cho, H., and H. Lee. *Biomedical named entity recognition using deep neural networks with contextual information*. BMC Bioinformatics. 2019.

Other kinds of IE: Relation Extraction



Source: Stanford TACRED

Other kinds of IE: Open Information Extraction

ID	Document
1	Your dry cleaner set out from eastern Queens on foot Tuesday morning and now somewhere near Maspeth.
2	Recently, North Korea has begun to allow tourists, including Americans, ..., and South Korean tourists have been able to go to Kaesong on a limited basis .
...	...

Entity 1	Relation Phrase	Entity 2	Human Evaluation
your dry cleaner	<i>set out from</i>	eastern Queens	✓
your dry cleaner	<i>set out from_on</i>	foot	✓
your dry cleaner	<i>is near</i>	Maspeth	✓
North Korea	<i>has begun to allow</i>	tourist	✓
South Korean tourists	<i>to go to</i>	Kaesong	✓
...	
Queens	<i>on</i>	foot	✗
Kaesong	<i>on</i>	a limited basis	✗

Source: Zhu et al. Open Information Extraction with Global Structure Constraints.
ACM WWW Conference. 2018.

Is IE a solved problem?

Table 2: Main Results on Testing Set: F_1 Score (Precision/Recall) (in %)

Method	CoNLL03	Tweet	OntoNote5.0	Webpage	Wikigold
Entity Types	4	10	18	4	4
KB Matching	71.40(81.13/63.75)	35.83(40.34/32.22)	59.51(63.86/55.71)	52.45(62.59/45.14)	47.76(47.90/47.63)
Fully-Supervised (Our implementation)					
RoBERTa	90.11(89.14/91.10)	52.19(51.76/52.63)	86.20(84.59/87.88)	72.39(66.29/79.73)	86.43(85.33/87.56)
BiLSTM-CRF	91.21(91.35/91.06)	52.18(60.01/46.16)	86.17(85.99/86.36)	52.34(50.07/54.76)	54.90(55.40/54.30)
Baseline (Our implementation)					
BiLSTM-CRF	59.50(75.50/49.10)	21.77(46.91/14.18)	66.41(68.44/64.50)	43.34(58.05/34.59)	42.92(47.55/39.11)
AutoNER	67.00(75.21/60.40)	26.10(43.26/18.69)	67.18(64.63/69.95)	51.39(48.82/54.23)	47.54(43.54/52.35)
LRNT	69.74(79.91/61.87)	23.84(46.94/15.98)	67.69(67.36/68.02)	47.74(46.70/48.83)	46.21(45.60/46.84)
Other Baseline (Reported Results)					
KALM [†]	76.00(- / -)	-	-	-	-
ConNET [◦]	75.57(84.11/68.61)	-	-	-	-
Our BOND Framework					
Stage I	75.61(83.76/68.90)	46.61(53.11/41.52)	68.11(66.71/69.56)	59.11(60.14/58.11)	51.55(49.17/54.50)
BOND	81.48(82.05/80.92)	48.01(53.16/43.76)	68.35(67.14/69.61)	65.74(67.37/64.19)	60.07(53.44/68.58)

Source: Liang et al. BOND: BERT-Assisted Open-Domain Named Entity Recognition with Distant Supervision. KDD Conference. 2020.

Other NLP steps: Coreference Resolution, Entity Linking...

"I had no idea I was getting in so deep," says Mr. Kaye, who founded Justin in 1982. Mr. Kaye had sold Capetronic Inc., a Taiwan electronics Maker, and retired, only to find he was bored. With Justin, he began selling toys and electronics made mostly in Hong Kong, beginning with Mickey Mouse radios. The company has grown -- to about 40 employees, from four initially, Mr. Kaye says. Justin has been profitable since 1986, adds the official, who shares [his] office... (nw/wsJ/2418)

The Northern Lights, also called Aurora Borealis, are one of the most spectacular shows on this earth and can frequently be seen in Iceland from September through March on clear and crisp nights.



Sources:

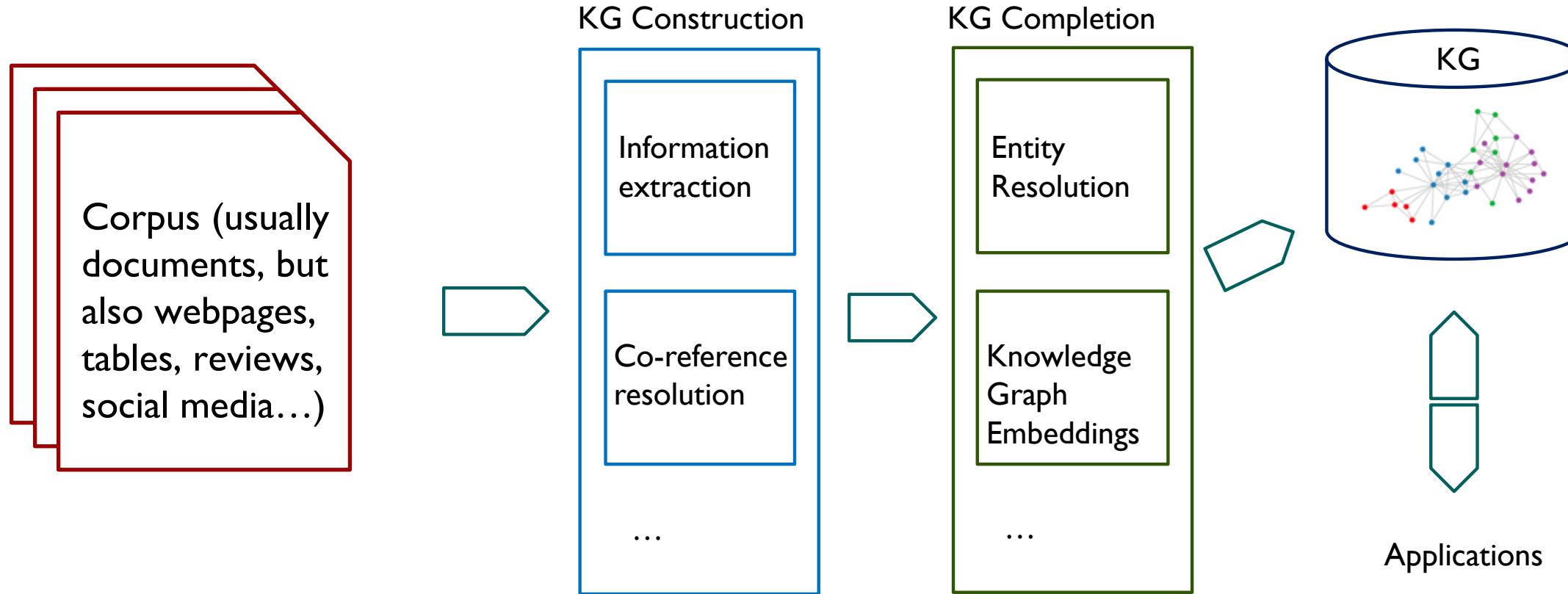
<https://aryamccarthy.github.io/wiseman2016learning/>

(Wiseman, Rush, and Shieber, 2016) at NAACL

Source:

Alokaili and Menai. SVM ensembles for named entity disambiguation. Computing. 2019.

A typical KGC workflow starts from corpus acquisition and ends with applications

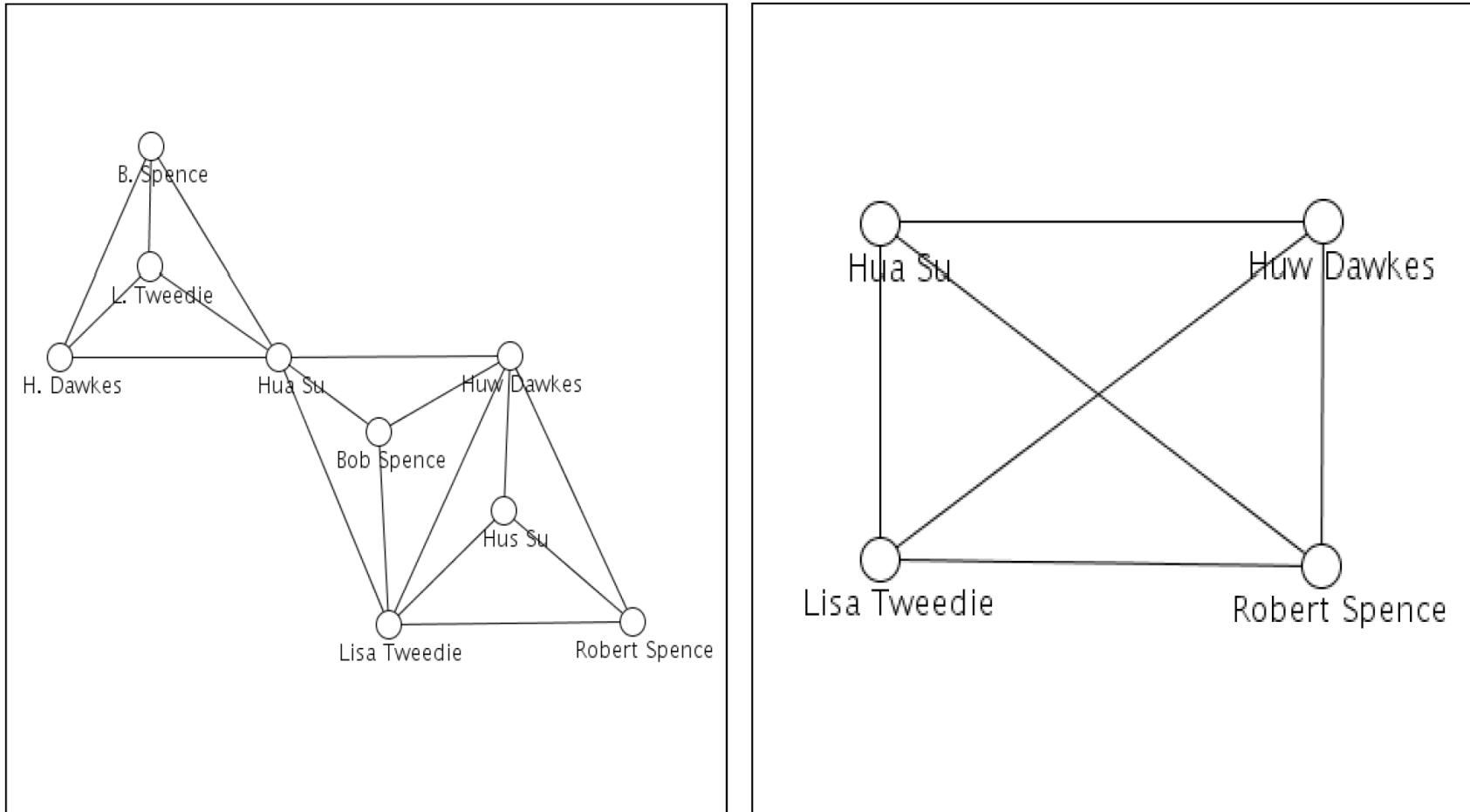


KNOWLEDGE GRAPH COMPLETION

Entity Resolution

Algorithmically identifying and linking/grouping different manifestations of the same real-world object

Problem has existed for 50 years in many communities (databases, graphs, networks, tables...)

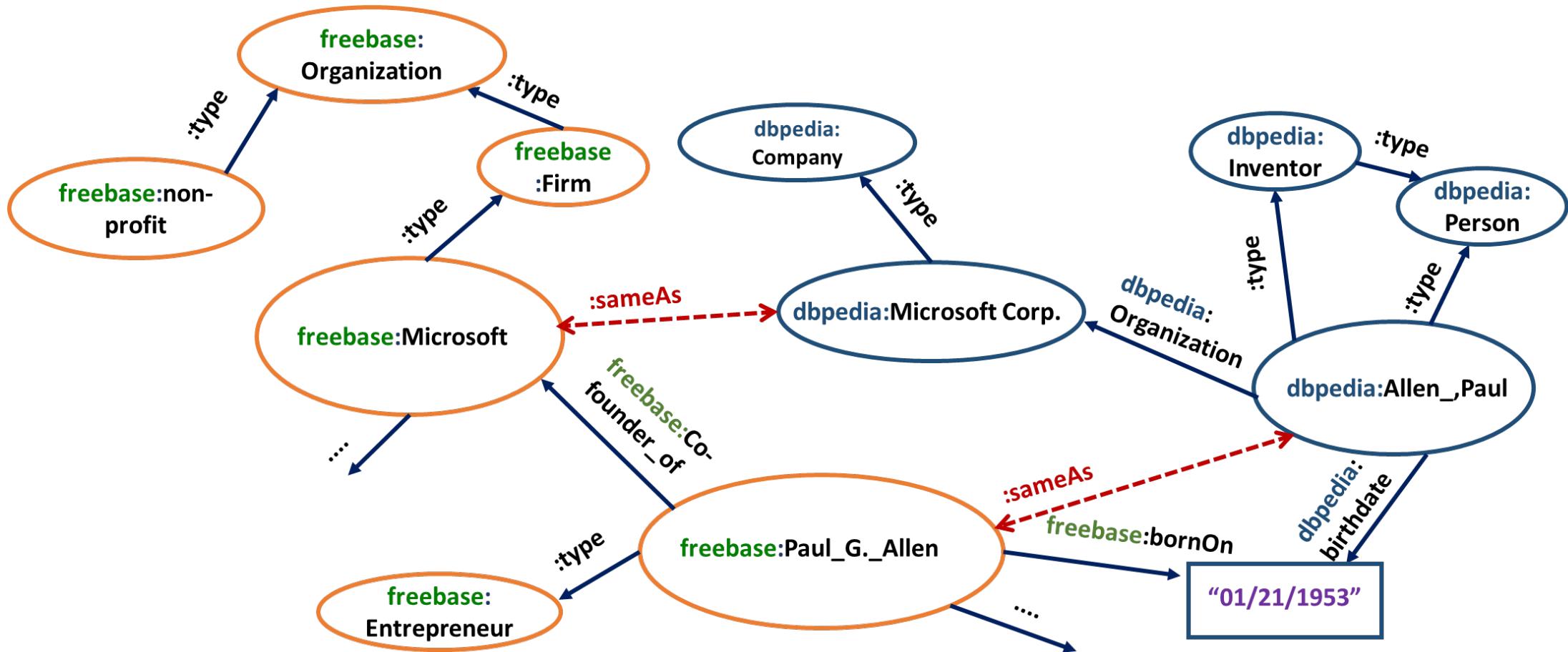


before

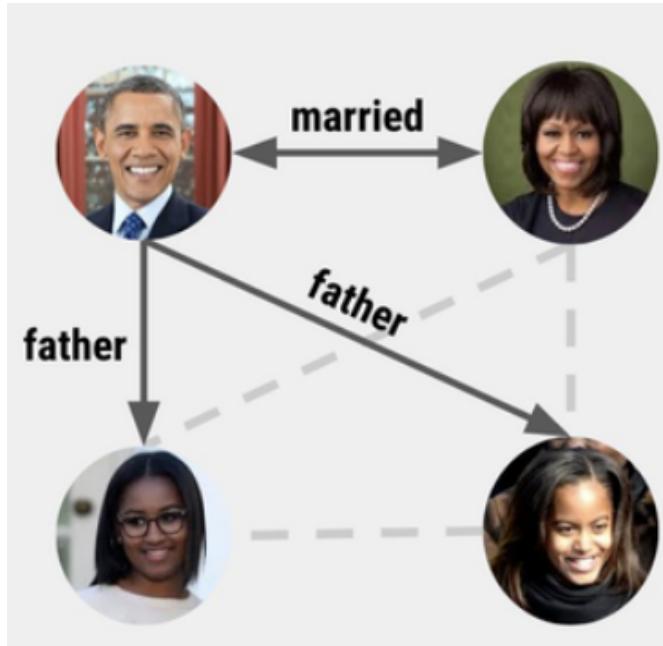
after

Source: Entity Resolution: Tutorial. Getoor and Machanavajjhala. VLDB, 2012

In the world of knowledge graphs

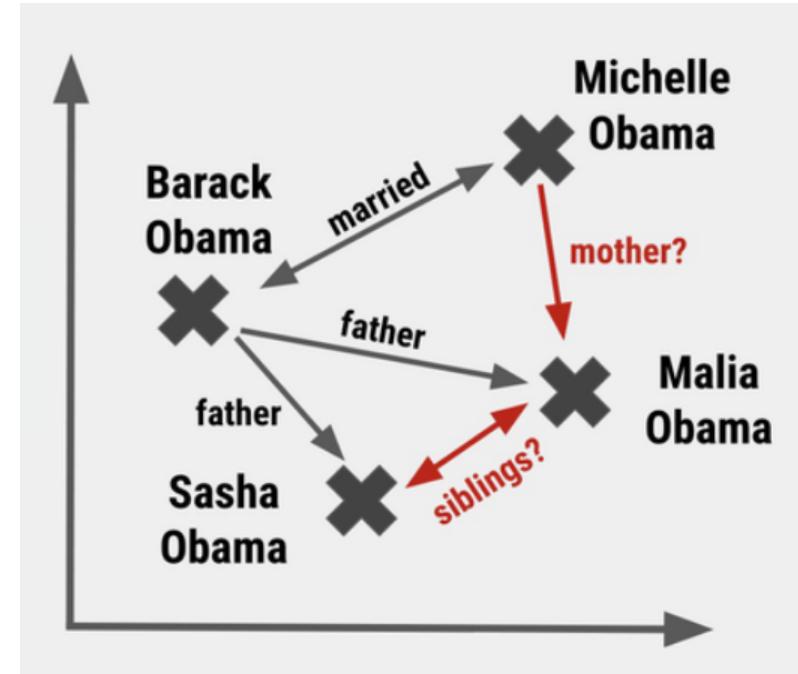


Representation Learning on Knowledge Graphs aka Knowledge Graph Embeddings



Knowledge graph embeddings:

- TransE, H...
- Neural tensor networks
- Graph convolutional networks (or their variants)
- Matrix factorization
- ...



KGEs (results)

Useful resources:

- OpenKE:
<http://139.129.163.161//index/toolkits#pretrained-embeddings>
- StarSpace:
<https://github.com/facebookresearch/StarSpace>
- Recent transformer-based models could potentially be adapted, including BERT and RoBERTa:
<https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems/>

Method	Raw						Filtered					
	WN18			FB15k			WN18			FB15k		
	MR	H@10	MRR	MR	H@10	MRR	MR	H@10	MRR	MR	H@10	MRR
SE (Bordes et al., 2011)	1011	68.5	-	273	28.8	-	985	80.5	-	162	39.8	-
Unstructured (Bordes et al., 2012)	315	35.3	-	1074	4.5	-	304	38.2	-	979	6.3	-
SME (Bordes et al., 2012)	545	65.1	-	274	30.7	-	533	74.1	-	154	40.8	-
TransH (Wang et al., 2014)	401	73.0	-	212	45.7	-	303	86.7	-	87	64.4	-
TransR (Lin et al., 2015b)	238	79.8	-	198	48.2	-	225	92.0	-	77	68.7	-
CTransR (Lin et al., 2015b)	231	79.4	-	199	48.4	-	218	92.3	-	75	70.2	-
KG2E (He et al., 2015)	342	80.2	-	174	48.9	-	331	92.8	-	59	74.0	-
TransD (Ji et al., 2015)	224	79.6	-	194	53.4	-	212	92.2	-	91	77.3	-
lppTransD (Yoon et al., 2016)	283	80.5	-	195	53.0	-	270	94.3	-	78	78.7	-
TranSparse (Ji et al., 2016)	223	80.1	-	187	53.5	-	211	93.2	-	82	79.5	-
TATEC (García-Durán et al., 2016)	-	-	-	-	-	-	-	-	-	58	76.7	-
NTN (Socher et al., 2013)	-	-	-	-	-	-	-	66.1	0.53	-	41.4	0.25
DISTMULT (Yang et al., 2015)	-	-	-	-	-	-	-	94.2	0.83	-	57.7	0.35
ComplEx (Trouillon et al., 2016)	-	-	0.587	-	-	0.242	-	94.7	0.941	-	84.0	0.692
HoIE (Nickel et al., 2016b)	-	-	0.616	-	-	0.232	-	94.9	0.938	-	73.9	0.524
RESCAL (Nickel et al., 2011) [*]	-	-	0.603	-	-	0.189	-	92.8	0.890	-	58.7	0.354
TransE (Bordes et al., 2013) [*]	-	-	0.351	-	-	0.222	-	94.3	0.495	-	74.9	0.463
STransE (Nguyen et al., 2016b)	217	80.9	0.469	219	51.6	0.252	206	93.4	0.657	69	79.7	0.543
RTransE (García-Durán et al., 2015)	-	-	-	-	-	-	-	-	-	50	76.2	-
PTransE (Lin et al., 2015a)	-	-	-	207	51.4	-	-	-	-	58	84.6	-
GAKE (Feng et al., 2016b)	-	-	-	228	44.5	-	-	-	-	119	64.8	-
Gaifman (Niepert, 2016)	-	-	-	-	-	-	352	93.9	-	75	84.2	-
Hiri (Liu et al., 2016)	-	-	-	-	-	-	-	90.8	0.691	-	70.3	0.603
NLFeat (Toutanova and Chen, 2015)	-	-	-	-	-	-	-	94.3	0.940	-	87.0	0.822
TEKE_H (Wang and Li, 2016)	127	80.3	-	212	51.2	-	114	92.9	-	108	73.0	-
SSP (Xiao et al., 2017)	168	81.2	-	163	57.2	-	156	93.2	-	82	79.0	-

Other proposals: knowledge graph identification using probabilistic soft logic

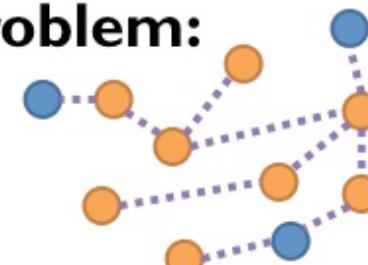
(Pujara et al., ISWC13)

Examples of ontological constraints

1. $DOM(R, L) \wedge REL(E_1, E_2, R) \xrightarrow{wo} LBL(E_1, L)$
2. $RNG(R, L) \wedge REL(E_1, E_2, R) \xrightarrow{wo} LBL(E_2, L)$
3. $INV(R, S) \wedge REL(E_1, E_2, R) \xrightarrow{wo} REL(E_2, E_1, S)$
4. $SUB(L, P) \wedge LBL(E, L) \xrightarrow{wo} LBL(E, P)$
5. $RSUB(R, S) \wedge REL(E_1, E_2, R) \xrightarrow{wo} REL(E_1, E_2, S)$
6. $MUT(L_1, L_2) \wedge LBL(E, L_1) \xrightarrow{wo} \neg LBL(E, L_2)$
7. $RMUT(R, S) \wedge REL(E_1, E_2, R) \xrightarrow{wo} \neg REL(E_1, E_2, S)$

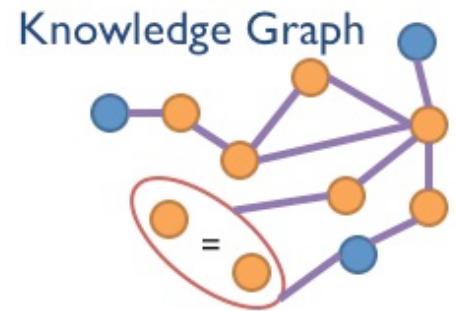
Knowledge Graph Identification

Problem:



Extraction Graph

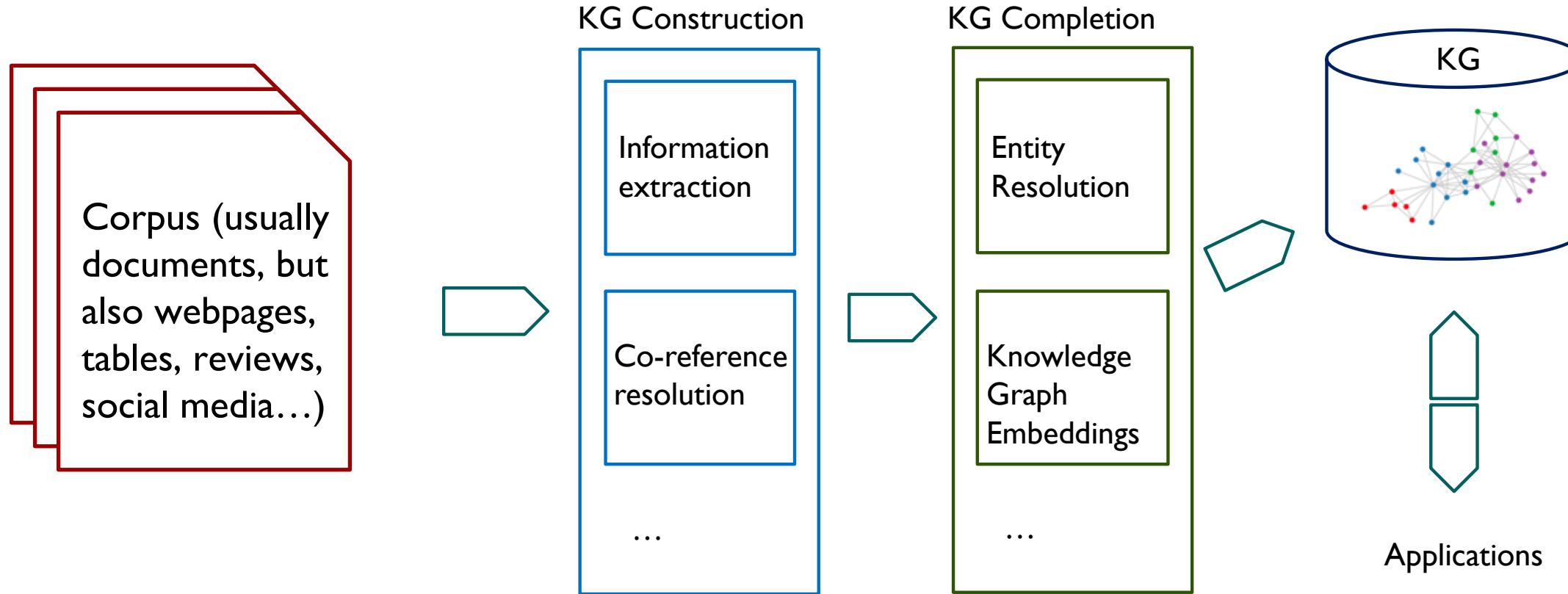
Knowledge
Graph
Identification



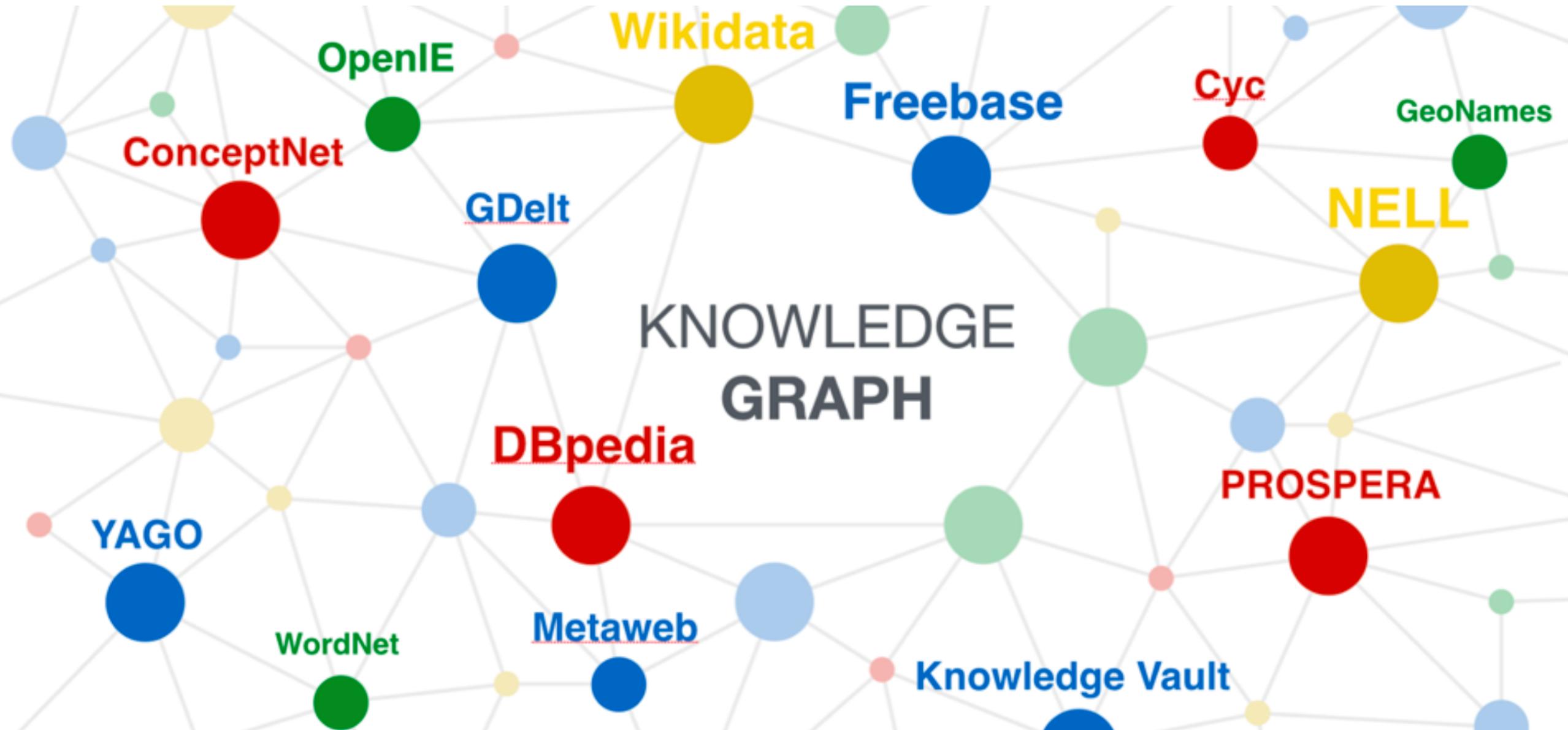
Solution: Knowledge Graph Identification (KGI)

- Performs *graph identification*:
 - entity resolution
 - node labeling
 - link prediction
- Enforces *ontological constraints*
- Incorporates *multiple uncertain sources*

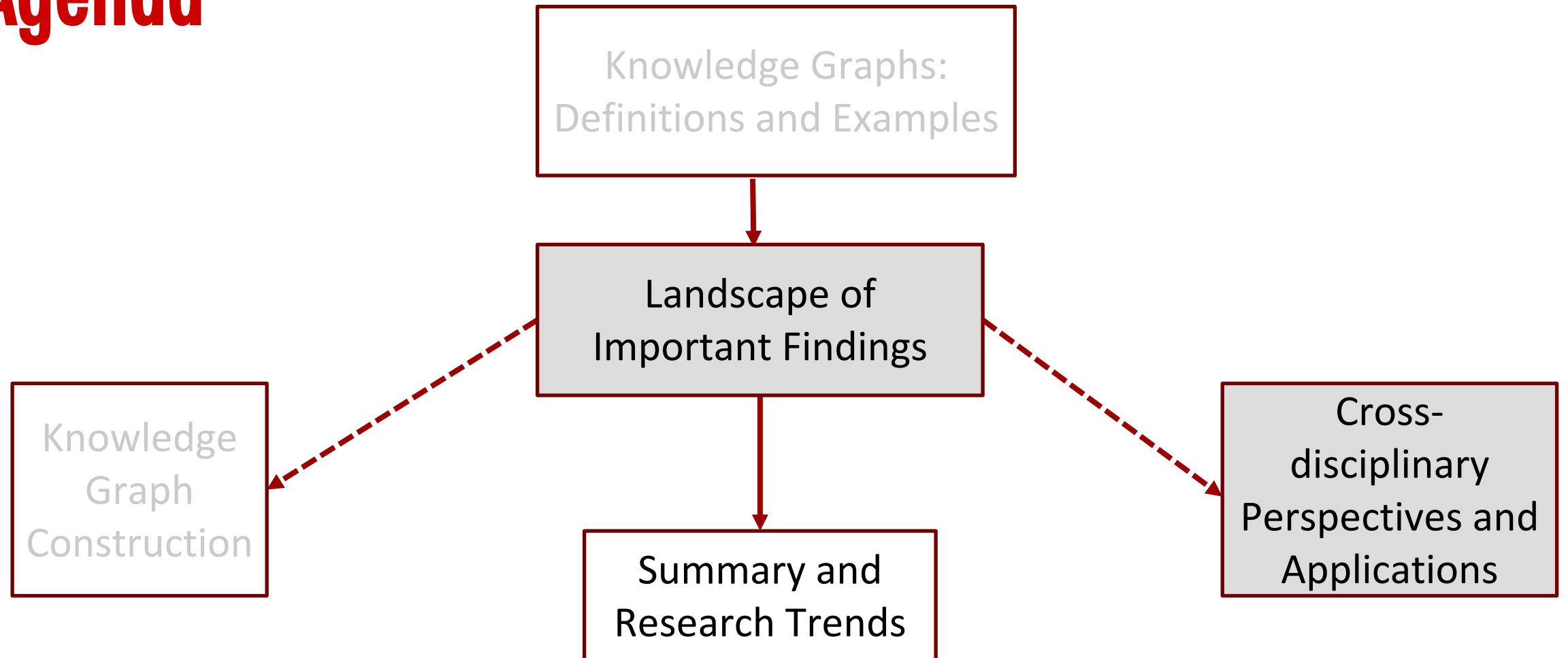
A typical KGC workflow starts from corpus acquisition and ends with applications



Open-source KGs that have been built



Agenda



CROSS-DISCIPLINARY PERSPECTIVES: WEB AND INFORMATION RETRIEVAL

Google Knowledge Graph

The screenshot shows a Google search results page for the query "Larry Page". A red box highlights the "Knowledge Graph" panel on the right side of the results. The panel contains a large photo of Larry Page, several smaller profile pictures of him and his colleagues, and a summary of his biography, including his birth date (March 26, 1973), height (5' 11" / 1.80 m), and spouse (Lucinda Southworth). It also lists his education (East Lansing High School), awards (Marconi Prize, TR100), and recent posts about the new Android release, KitKat. Below the main summary, there's a section titled "People also search for" with links to profiles of Sergey Brin, Eric Schmidt, Larry Ellison, Marissa Mayer, and Bill Gates.

Knowledge Graph

Larry Page

Born: March 26, 1973 (age 40), East Lansing, MI
Height: 5' 11" (1.80 m)
Spouse: Lucinda Southworth (m. 2007)
Siblings: Carl Victor Page, Jr.
Education: East Lansing High School (1987–1991), More
Awards: Marconi Prize, TR100

Recent posts
Just opened the new Android release, KitKat! Sep 3, 2013

People also search for

Sergey Brin Eric Schmidt Larry Ellison Marissa Mayer Bill Gates

Wonder Woman (2017) - IMDb

Wonder Woman (2017 film) - Wikipedia

Wonder Woman (2017 film) - Wikipedia

Top stories

Oscars voting ends today. Will 'Wonder Woman' finally break the anti-superhero streak? Washington Post

Fashion War: Wonder Woman Gal Gadot Infuriates Lebanese with Dress Design Breitbart

Gal Gadot Diet and Fitness Routine | POPSUGAR Fitness Australia

POPSUGAR Australia

Da Vinci

Mona Lisa

Michelangelo

Italy

About 36,700,000 results (0.67 seconds)

Domain-specific search (DSS)

The Massive YouTube Ecosystem



Find tacos, cheap dinner, Max's

Near address, neighborhood, city, state or zip

Sign Up Log In

Y-Clad's Hidden Gem

Venues & Event Spaces, Jewelry, Accessories

Write a Review Add Photo Share Bookmark

Map data ©2014 Google, Sanborn

122 E 57th St New York, NY 10044 Midtown East

Get Directions (212) 644-8445 y-clad.com

Enjoy Free Shipping on All Orders Over \$150! Order Online

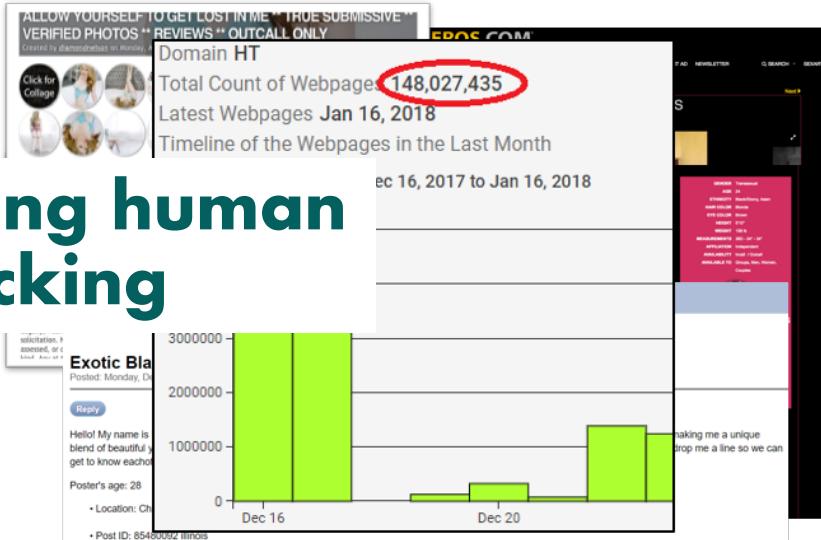
\$75 for \$100 Deal at Y-Clad's Hidden Gem

You get a voucher redeemable for \$100 at Y-Clad's Hidden Gem.

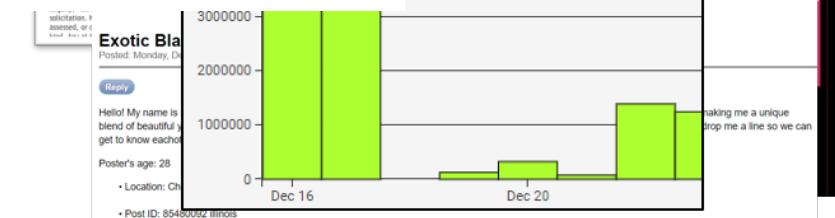


source: <https://photos.prnewswire.com/prnfull/20151006/274273-INFO>

Emerging opportunities for DSS



Fighting human trafficking



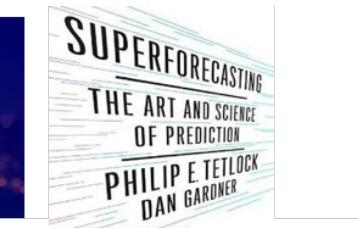
Stopping Penny Stock Fraud

Defined by the SEC as stocks that sell below \$5 a share, penny stocks have always been considered speculative and easily manipulated. But stock market experts, seeing an increase in penny stock promotion online, say investors should be wary of

Predicting cyberattacks



Accurate geopolitical forecasting



forecasters begin by gathering as much information possible.

forecasters nurture and develop the habit of thinking terms of probabilities when exploring the likelihood of specific events.

forecasting improves when individuals work in teams. forecasters ensure that they are regularly keeping score of their projections.

5. The most successful forecasters are willing to admit error and quickly change course on their projections.



DARPA/IARPA programs

DARPA Memex

IARPA Hybrid Forecasting
Competition

DARPA AIDA

DARPA Causal Exploration

DARPA LORELEI

IARPA CAUSE



Research Question

General Search

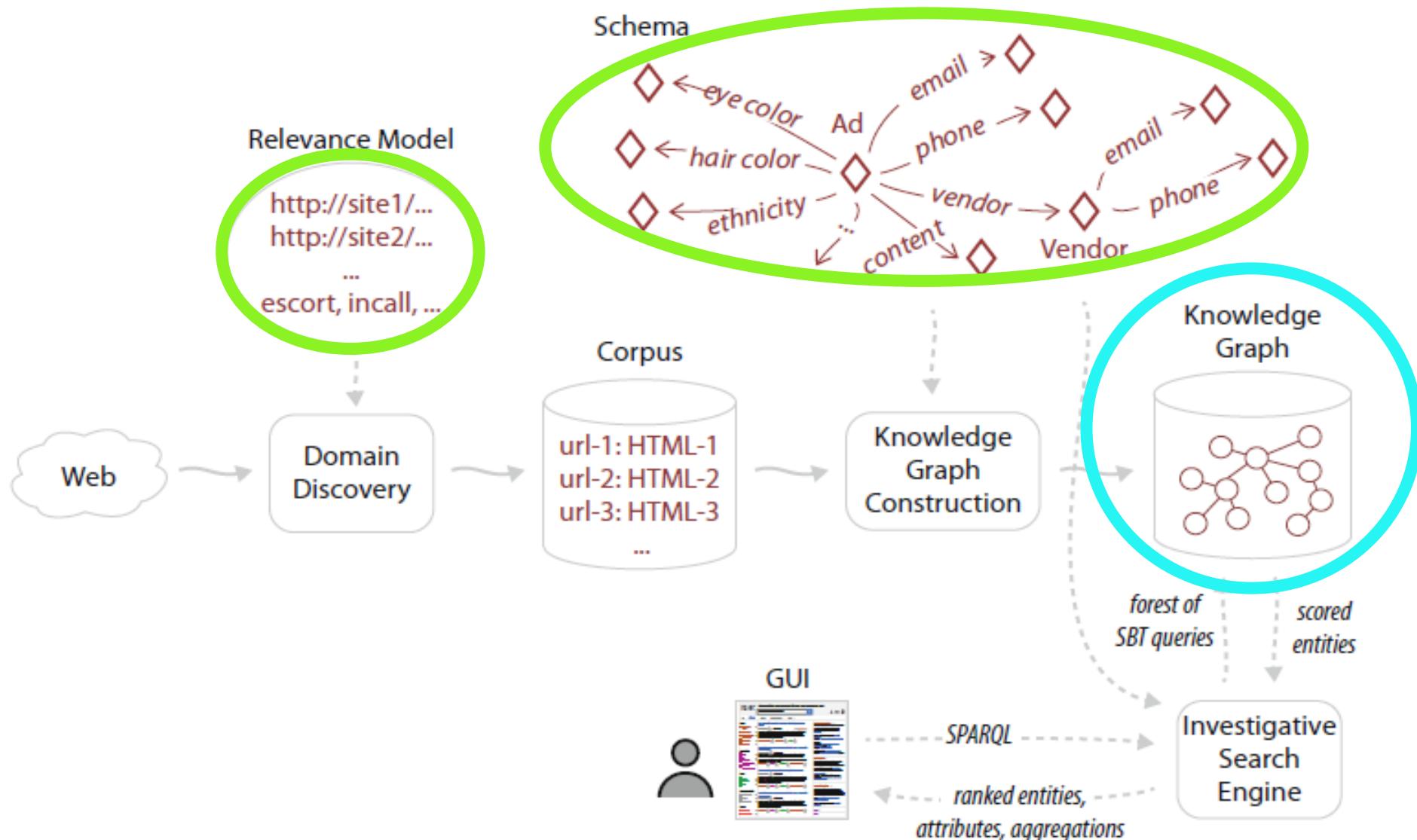
Google Knowledge Graph

DSS

Domain-Specific Knowledge Graphs

How do we construct domain specific knowledge graphs over web data for powerful DSS applications?

Knowledge Graphs for DSS



Project: atf_firearms_domain

Search Terms

Caliber or Gauge: 9mm X

Model: glock X Model: glock 26 X

Make (Top 10)

[View More](#)

Sort By: A-Z

glock	110,530
springfield	50,468
ruger	47,236
browning	42,122
colt	40,750
smith	40,730
winchester	37,697
walther	37,037
taurus	35,942
smith & wesson	34,298

Model (Top 10)

[View More](#)

Sort By: A-Z

glock 19	17,554
glock 17	12,801
desert eagle	9,556
springfield xd	8,926
glock 43	8,644
glock 22	7,314
<input checked="" type="checkbox"/> glock 26	6,966
glock 23	6,482
glock 27	6,316
springfield xdm	4,710

25 of 461,480 Results [How are search results found?](#)

PLEASE NOTE THAT ONLY THE TOP 10 EXTRactions OF EACH TYPE ARE SHOWN IN THE RESULT LIST.

2.84

[Glock 26](#)

Calibers or Gauges

[9mm](#)

Emails

[gda32570@yahoo.com](#)

Cities

[pensacola, florida](#)

Models

[glock 26](#)

Social Media Names

[facebook](#)

Usernames

[gda32570](#)

Prices

[650](#)

Dates (Any)

[Feb 6, 2016](#)[Dec 18, 2015](#)[Oct 25, 2016](#)[Feb 2, 2016](#)[Jan 21, 2016](#)[Oct 24, 2016](#)[Jan 9, 2016](#)[Feb 1, 2016](#)

Makes

[Glock](#)[fast](#)[chicago](#)

Model: glock 26

9173 Total Results



CITY	RESULTS CO-OCCURRING WITH GLOCK 26	RESULTS NOT CO-OCCURRING
dallas, texas	1,100	0
orange, california	674	0
springfield, oregon	539	0
david, chiriquí	403	0
springfield, massachusetts	287	0
springfield, ohio	239	0
springfield, missouri	228	0
south bend, indiana	171	0
pune, maharashtra	97	0
phoenix, arizona	95	0

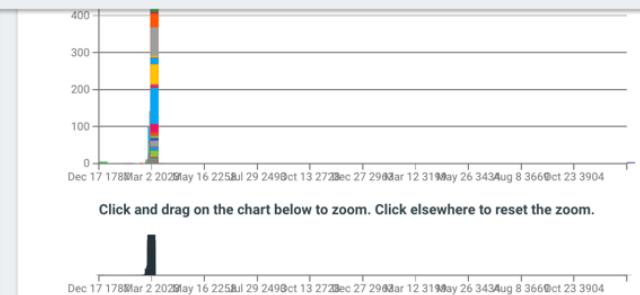
[SHOW MORE](#)

4 Calibers or Gauges

[Copy](#)

CALIBER OR GAUGE	RESULTS CO-OCCURRING WITH GLOCK 26	RESULTS NOT CO-OCCURRING
9mm	19	0
9mm, .40, .45	9	0
9mm luger	3	0
9x19	2	0

Domain-specific Insight Graphs



25 of 9,173 Results

PLEASE NOTE THAT ONLY THE TOP 10 EXTRactions OF EACH TYPE ARE SHOWN IN THE LIST.

6.85 Sig Sauer P250 For Sale

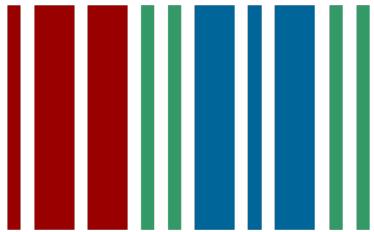
Cities

[springfield, oregon](#)

Makes
pelican
walther
savage arms
colt
springfield armory m1a
allen
bushmaster
beretta usa
chip mccormick
ruger

Models

TLDs



WIKIDATA



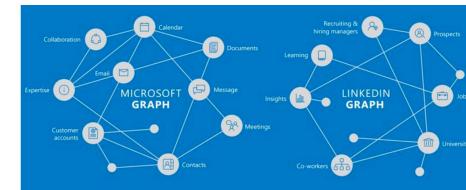
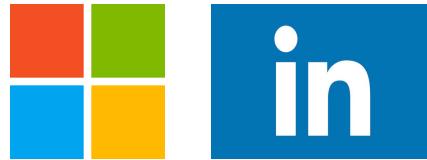
U B E R

Building an Enterprise
Knowledge
Graph @Uber:
[@Uber](#)

Lessons from Reality

Joshua Shinavier, PhD
Knowledge Graph Conference
May 8th, 2019

Uber



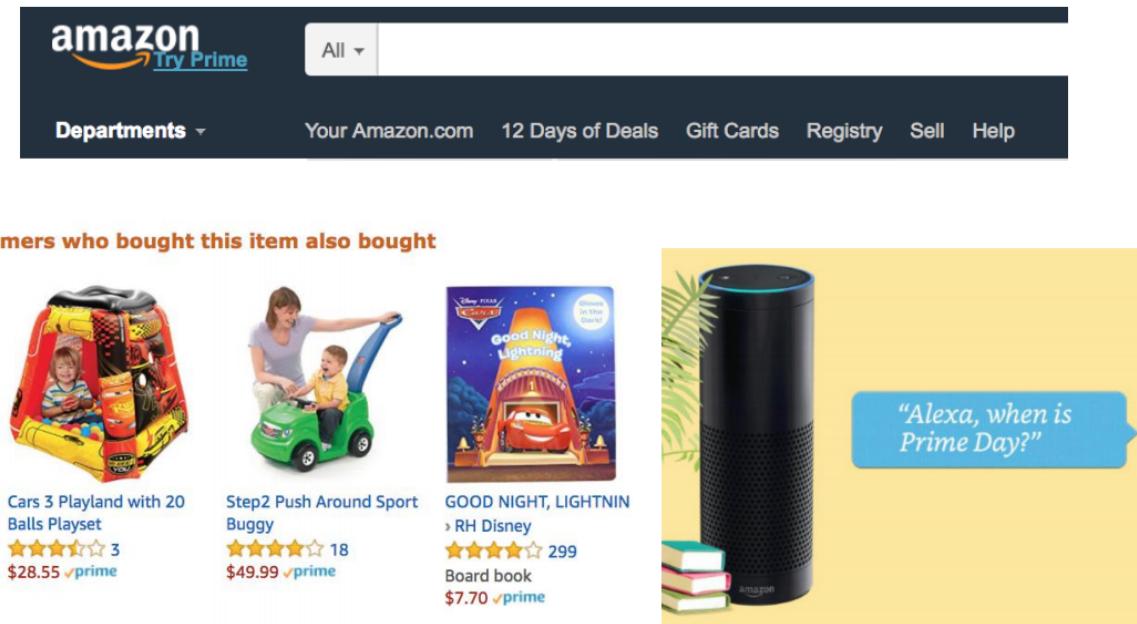
amazon



Many examples in industry
and non-profit

Commercial domains: Amazon Product Graph

- ❑ Mission: To answer any question about products and related knowledge in the world



Source: Dong, Luna. Building a Broad Knowledge Graph for Products. Keynote at ICDE. 2019

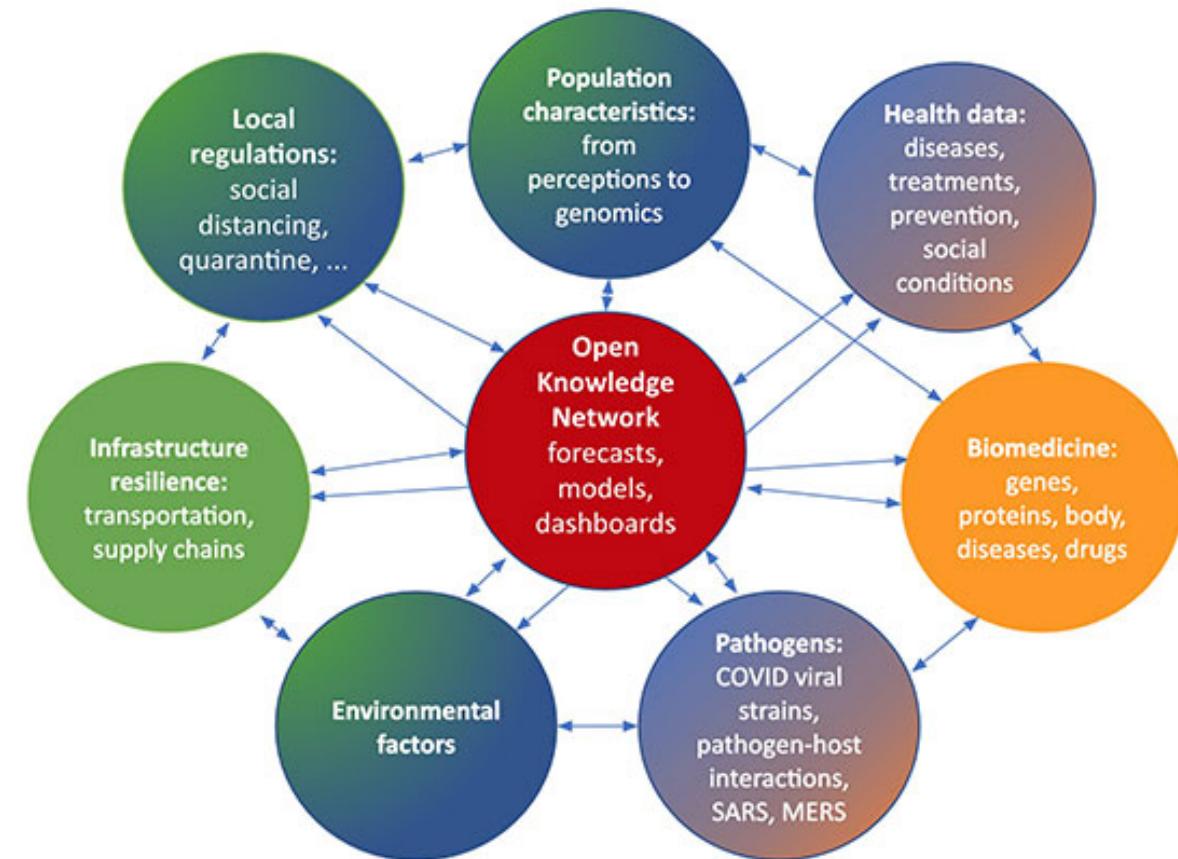
Another example: COVID-19

June 01, 2020 | By Jan Zverina

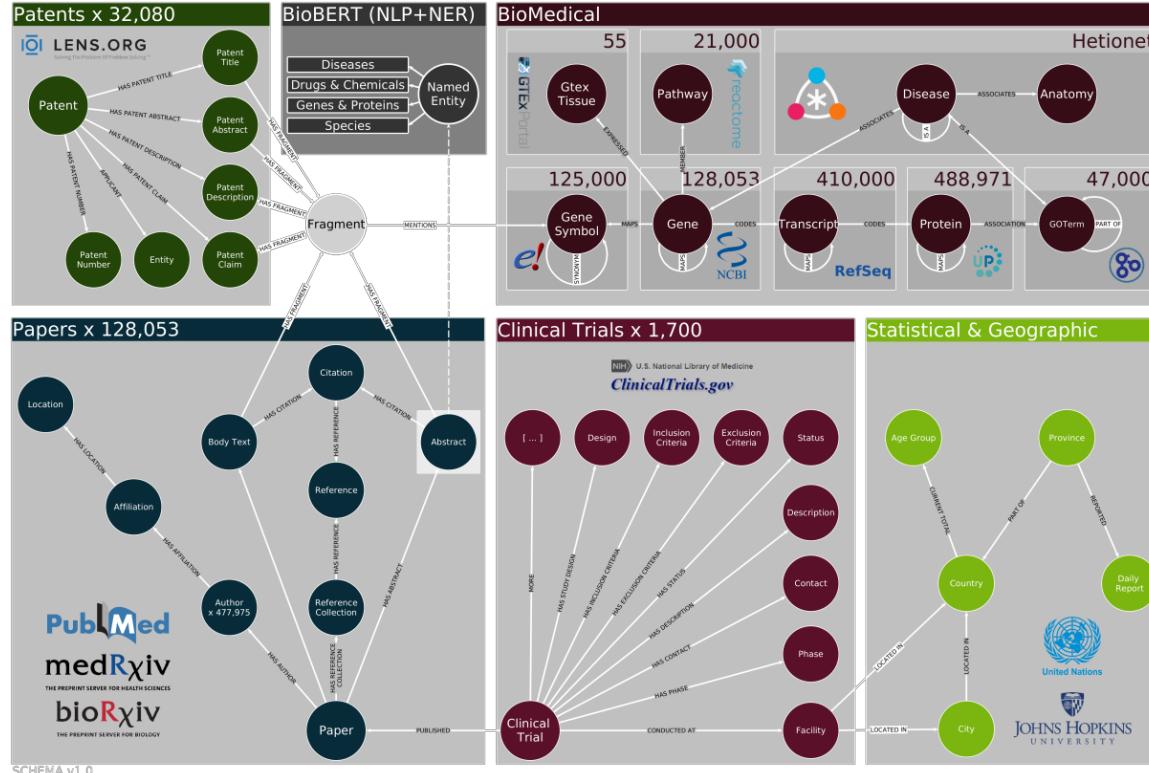
Fighting COVID-19 with Knowledge Graphs

National Science Foundation awards funding for a semantic integration platform

"The project will be based on our knowledge graph prototype linking information about pathogens, health data, and environmental indicators and enabling cross-domain inferencing," said Peter Rose, director of SDSC's Structural Bioinformatics Laboratory and principal investigator (PI) for the project, called 'COVID-19-Net: Integrating Health, Pathogen and Environmental Data into a Knowledge Graph for Case Tracking, Analysis, and Forecasting.' "Such a graph lets researchers trace the spread of the coronavirus in different geographic conditions, focusing on specific virus strains and transmissions."

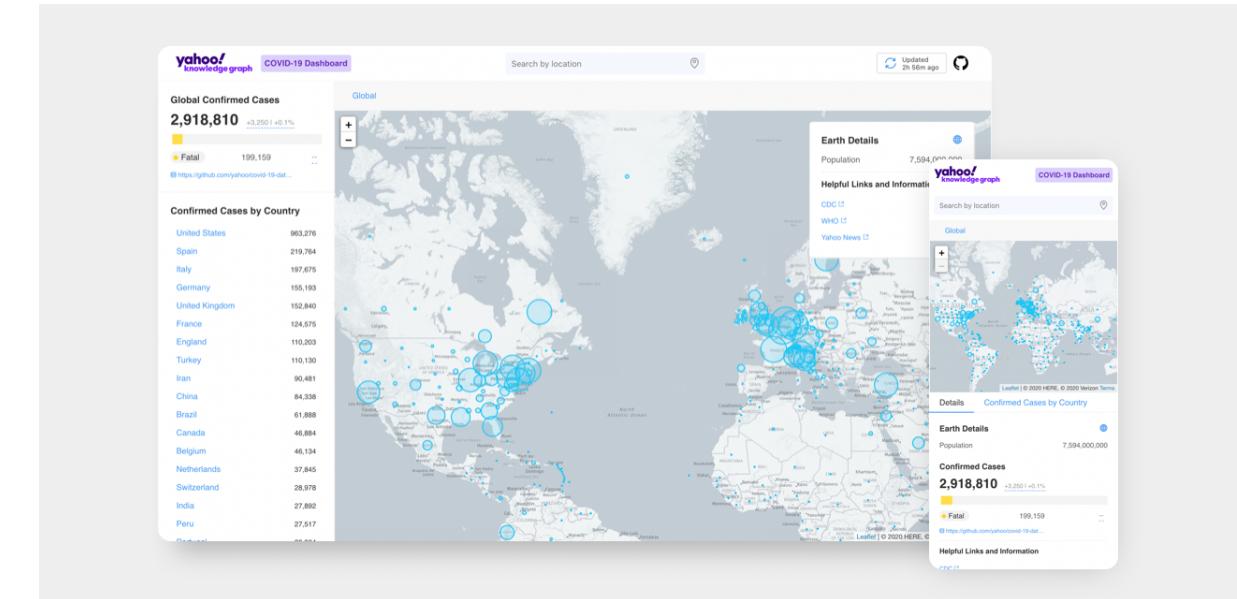


Other COVID-19 KG examples



Source: CovidGraph

Further reading: Kejriwal, M. (2020). Knowledge Graphs and COVID-19: Opportunities, Challenges, and Implementation. *Harvard Data Science Review*.

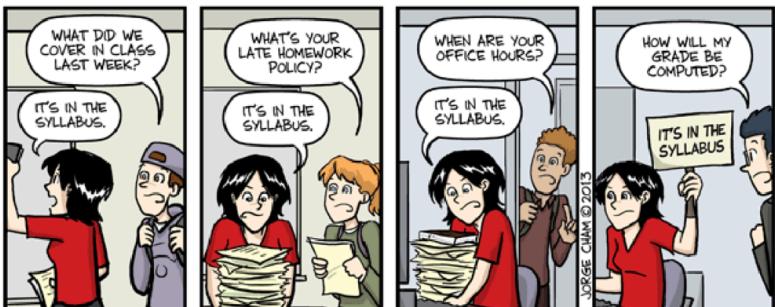


Source: Verizon Media

<https://github.com/yahoo/covid-19-dashboard>

CROSS-DISCIPLINARY PERSPECTIVES: SEMANTIC WEB

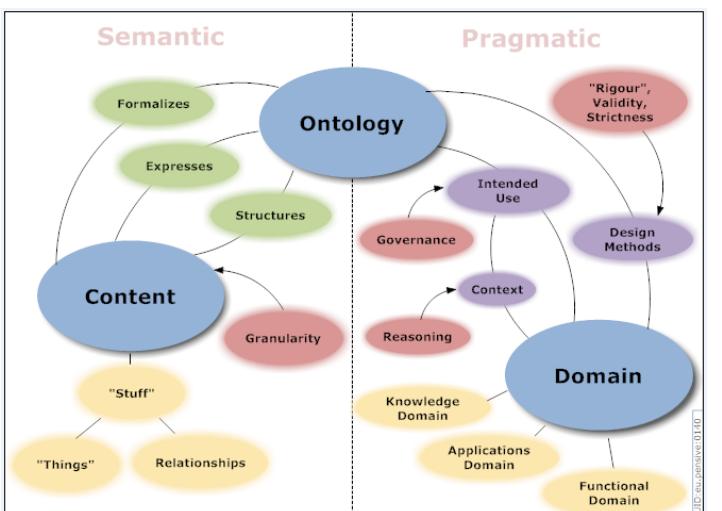
What is (or even isn't) a domain?



IT'S IN THE SYLLABUS

This message brought to you by every instructor that ever lived.

WWW.PHDCOMICS.COM
"Piled Higher and Deeper" by Jorge Cham



Some dictionary definitions

(Merriam Webster) A sphere of **knowledge, influence or activity**

(Oxford) A **specified** sphere of activity or knowledge

Specifying the sphere

Rules

Scope (e.g., the legal system)

Syllabi (for classrooms)

Examples

How do domain experts specify the sphere?

Examples

Ontology

Modeling domains: Ontologies

What is an ontology?

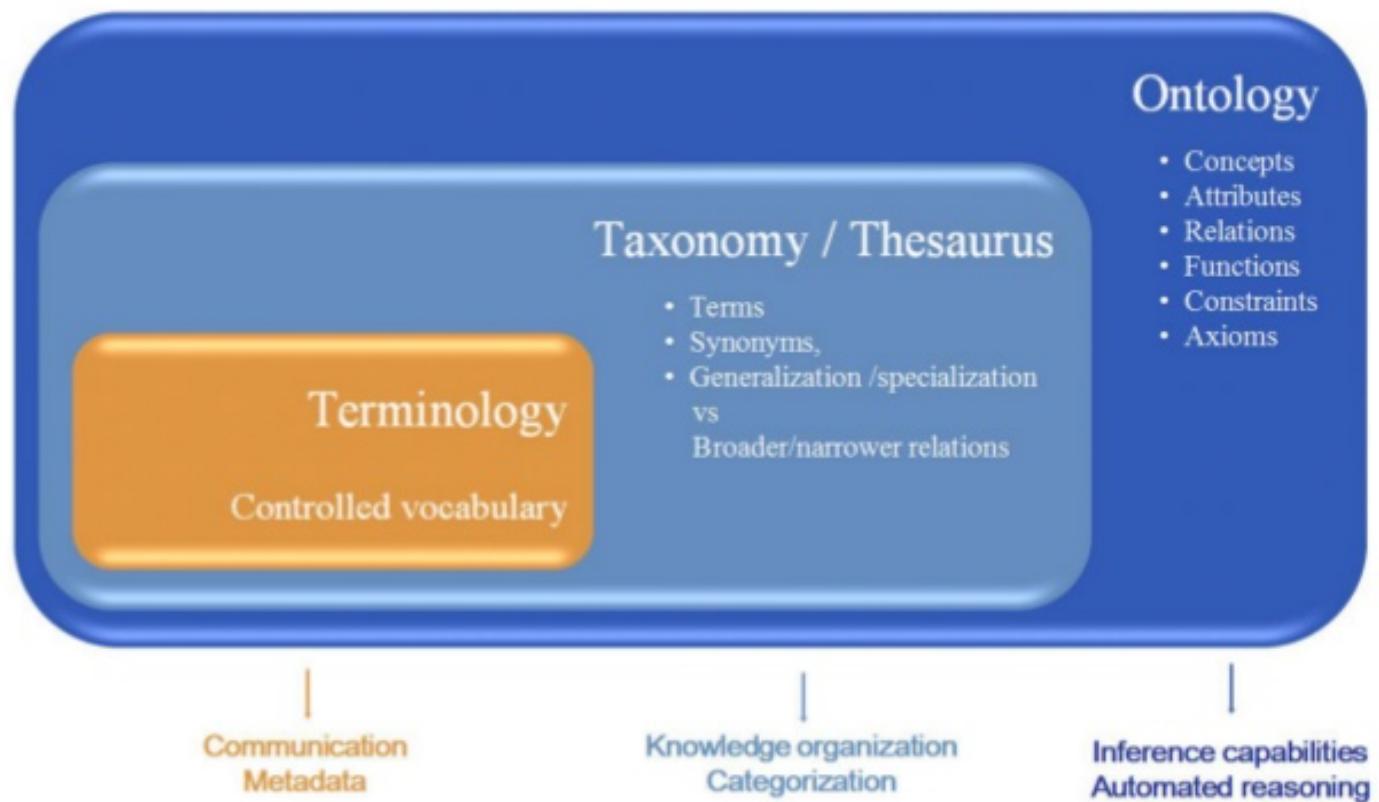
Formal, explicit specification of a shared conceptualization

Machine readable

Consensual knowledge

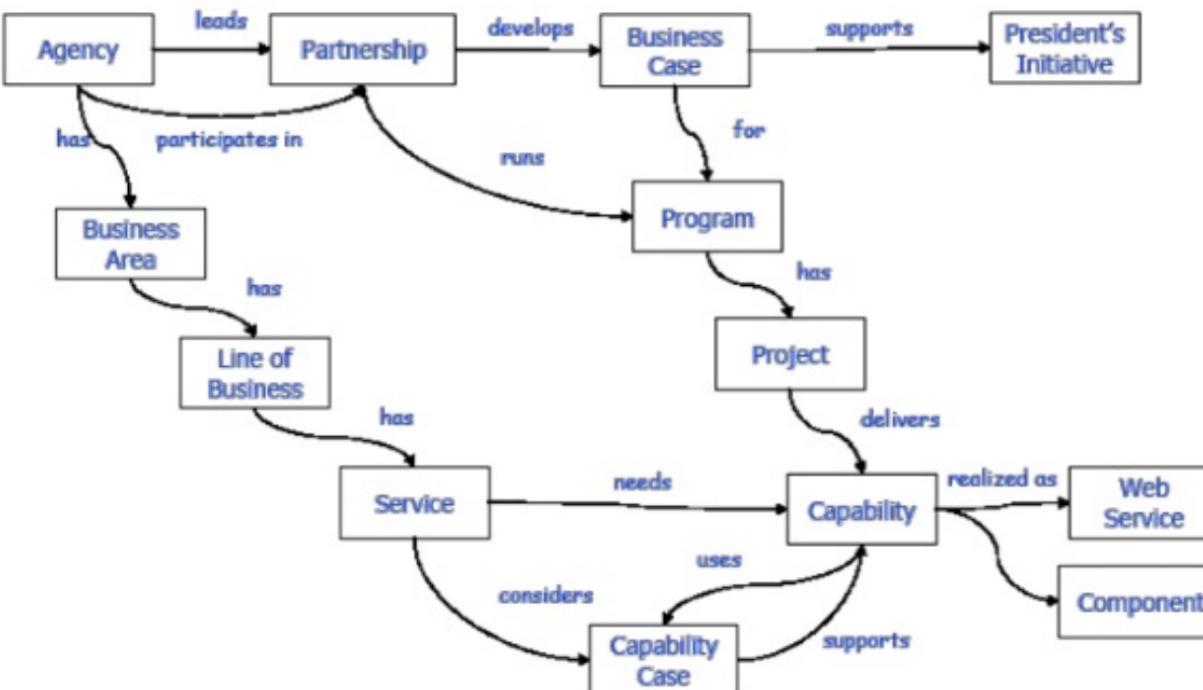
Abstract model of some phenomena in the world

Concepts, properties, functions, axioms are explicitly defined

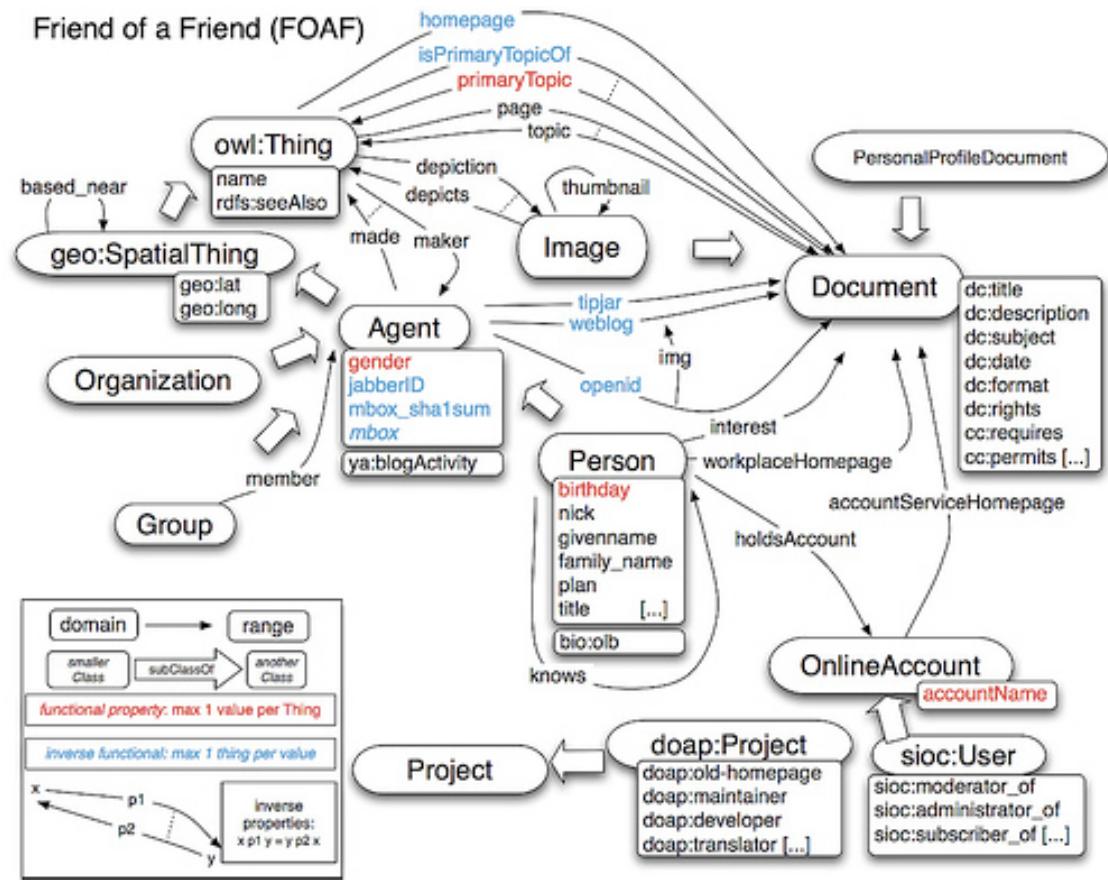


Source: “Ontologies and semantic web.” Stanley Wang. <https://www.slideshare.net/stanleywanguni/ontologies-and-semantic-web>

Examples of ontologies



Agency domain



Friend-of-a-friend

Ontologies are big in Science

Home | Advanced Search | Browse | Documentation | Download | Tools | About ChEBI

ChEBI > Main

CHEBI:41001 - ecgonine benzoate

Main ChEBI Ontology Automatic Xrefs Reactions Pathways Models

ChEBI Name: ecgonine benzoate
ChEBI ID: CHEBI:41001
Definition: A benzoate ester metabolite of cocaine formed by hydrolysis of the methyl ester group, catalysed by carboxylesterases.
Stars: ★★★ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs: CHEBI:3041
Supplier Information: ChemicalBook:CB1217496, eMolecules:535127, ZINC000002572652
Download: Molfile XML SDF

• [Find compounds which contain this structure](#)
• [Find compounds which resemble this structure](#)
• [Take structure to the Advanced Search](#)

[more structures >>](#)

Wikipedia ⓘ
Benzoylecgonine is the main [metabolite](#) of cocaine.

[Read full article at Wikipedia](#)

Formula	C16H19NO4
Net Charge	0
Average Mass	289.32640
Monoisotopic Mass	289.13141
InChI	InChI=1S/C16H19NO4/c1-17-11-7-8-12(17)14(15(18)19)13(9-11)21-16(20)10-5-3-2-4-6-10/h2-6,11-14H,7-9H2,1H3,(H,18,19)/t11-,12+,1-
InChIKey	GVGYEFKIHJTNQZ-RFQIPJPRSA-N
SMILES	[H][C@]12CC[C@]([H])([C@H](C1)OC(=O)c1ccccc1)C(O)=O)N2C

Roles Classification ⓘ

[marine xenobiotic metabolite](#)
Any metabolite produced by metabolism of a xenobiotic compound in marine macro- and microorganisms.

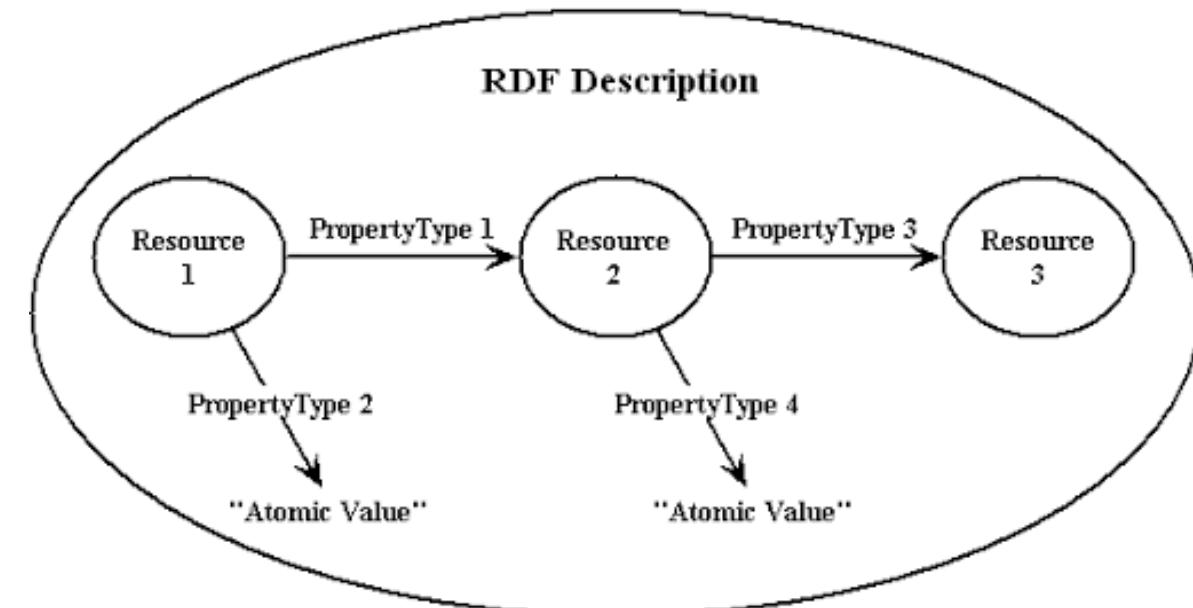
[plant metabolite](#)
Any eukaryotic metabolite produced during a metabolic reaction in plants, the kingdom that include flowering plants, conifers and other gymnosperms, ferns, mosses, algae, fungi, and bacteria.

Representation of knowledge graphs (and ontologies)

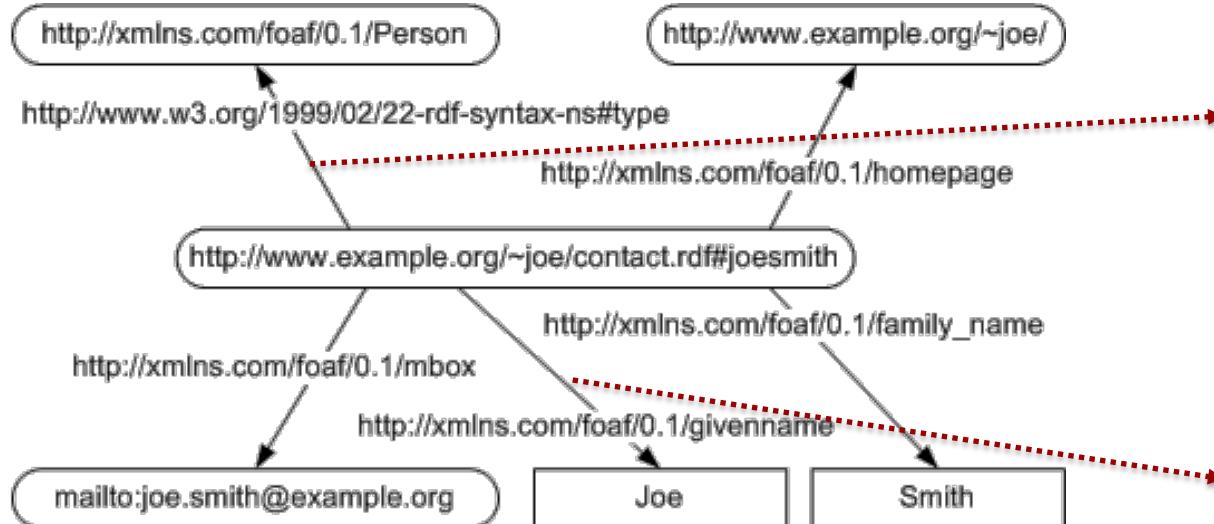
An RDF graph is a set of triples, where each triple is of the form (subject, predicate, object):

- Subjects must be URIs (technically, *internationalized resource identifiers*, in practice, just Uniform Resource Locators)
- Predicates (also called ‘properties’) must be URIs
- Objects can be either URIs or literals (strings, numbers, dates...)

In the Semantic Web, RDF is the ‘building block’ of higher order vocabularies (such as RDF Schema and OWL) that can be used to represent ontologies



Example of RDF KG



As a graph

<http://www.example.org/~joe/contact.rdf#joesmith>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person>

...

<http://www.example.org/~joe/contact.rdf#joesmith>
<http://xmlns.com/foaf/0.1/givenname> "Joe"

As a set of triples

Web Ontology Language (OWL)

OWL builds on RDF
(and another layer
called RDF Schema or
RDFS) to provide a
systematic vocabulary
for defining ontologies

Because OWL builds on
RDF, every OWL
ontology is **also** an RDF
graph, but not
necessarily vice-versa

RDF Schema Features:

- [Class \(Thing, Nothing\)](#)
- [rdfs:subClassOf](#)
- [rdf:Property](#)
- [rdfs:subPropertyOf](#)
- [rdfs:domain](#)
- [rdfs:range](#)
- [Individual](#)

Property Restrictions:

- [Restriction](#)
- [onProperty](#)
- [allValuesFrom](#)
- [someValuesFrom](#)

Class Intersection:

- [intersectionOf](#)

Datatypes

- [xsd.datatypes](#)

(In)Equality:

- [equivalentClass](#)
- [equivalentProperty](#)
- [sameAs](#)
- [differentFrom](#)
- [AllDifferent](#)
- [distinctMembers](#)

Restricted Cardinality:

- [minCardinality](#) (only 0 or 1)
- [maxCardinality](#) (only 0 or 1)
- [cardinality](#) (only 0 or 1)

Versioning:

- [versionInfo](#)
- [priorVersion](#)
- [backwardCompatibleWith](#)
- [incompatibleWith](#)
- [DeprecatedClass](#)
- [DeprecatedProperty](#)

Property Characteristics:

- [ObjectProperty](#)
- [DatatypeProperty](#)
- [inverseOf](#)
- [TransitiveProperty](#)
- [SymmetricProperty](#)
- [FunctionalProperty](#)
- [InverseFunctionalProperty](#)

Header Information:

- [Ontology](#)
- [imports](#)

Annotation Properties:

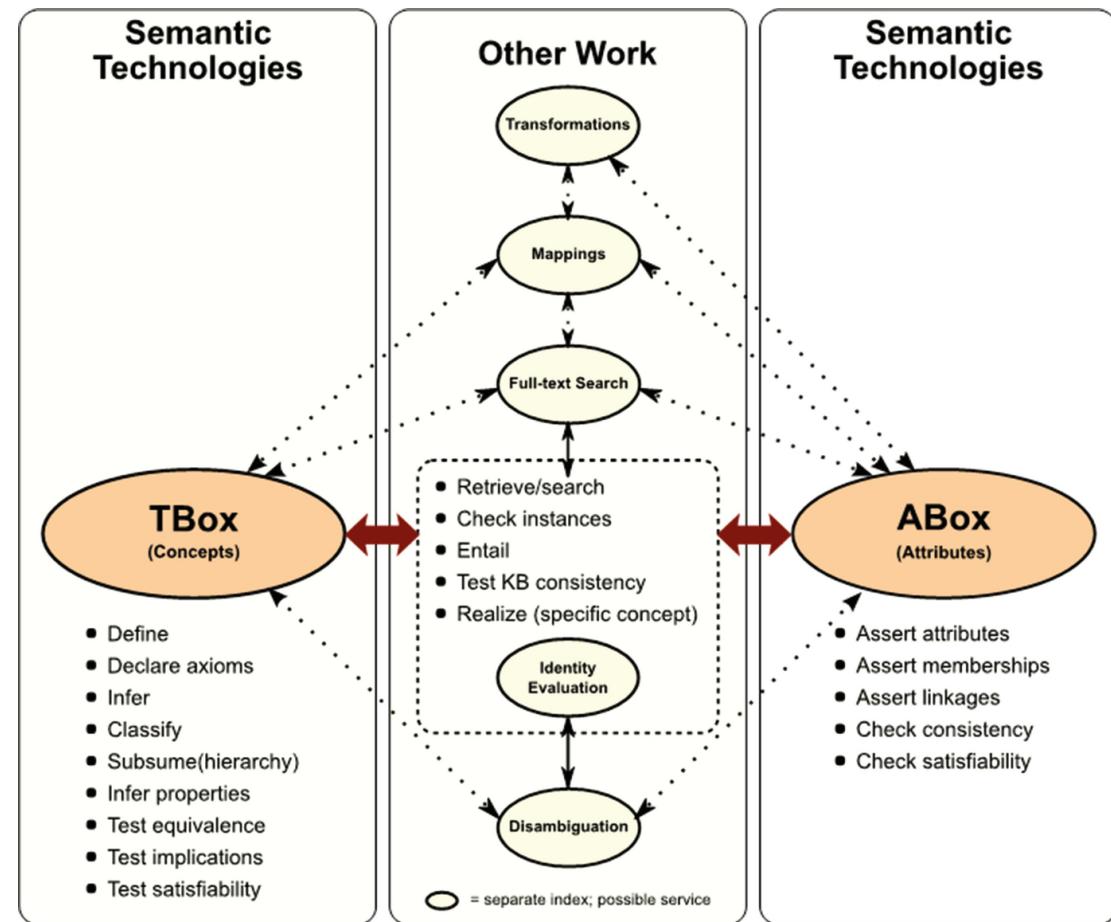
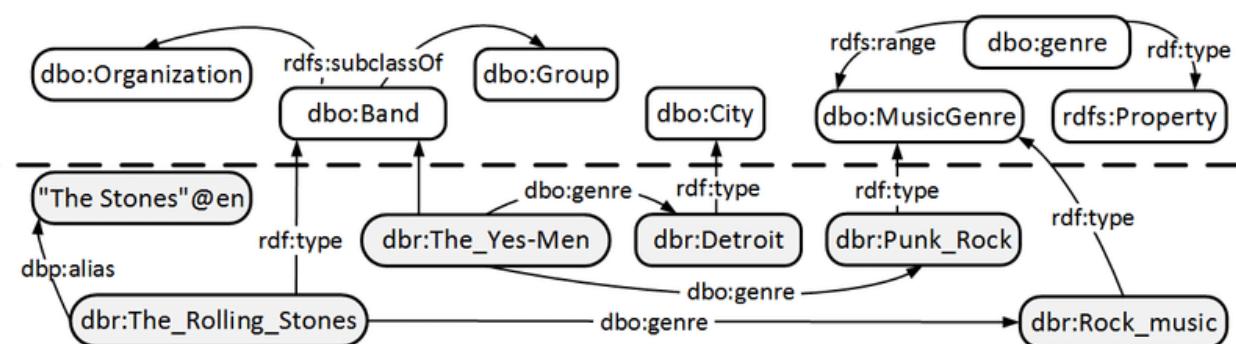
- [rdfs:label](#)
- [rdfs:comment](#)
- [rdfs:seeAlso](#)
- [rdfs:isDefinedBy](#)
- [AnnotationProperty](#)
- [OntologyProperty](#)

<https://www.w3.org/TR/owl-features/>

Reasoning over knowledge graphs

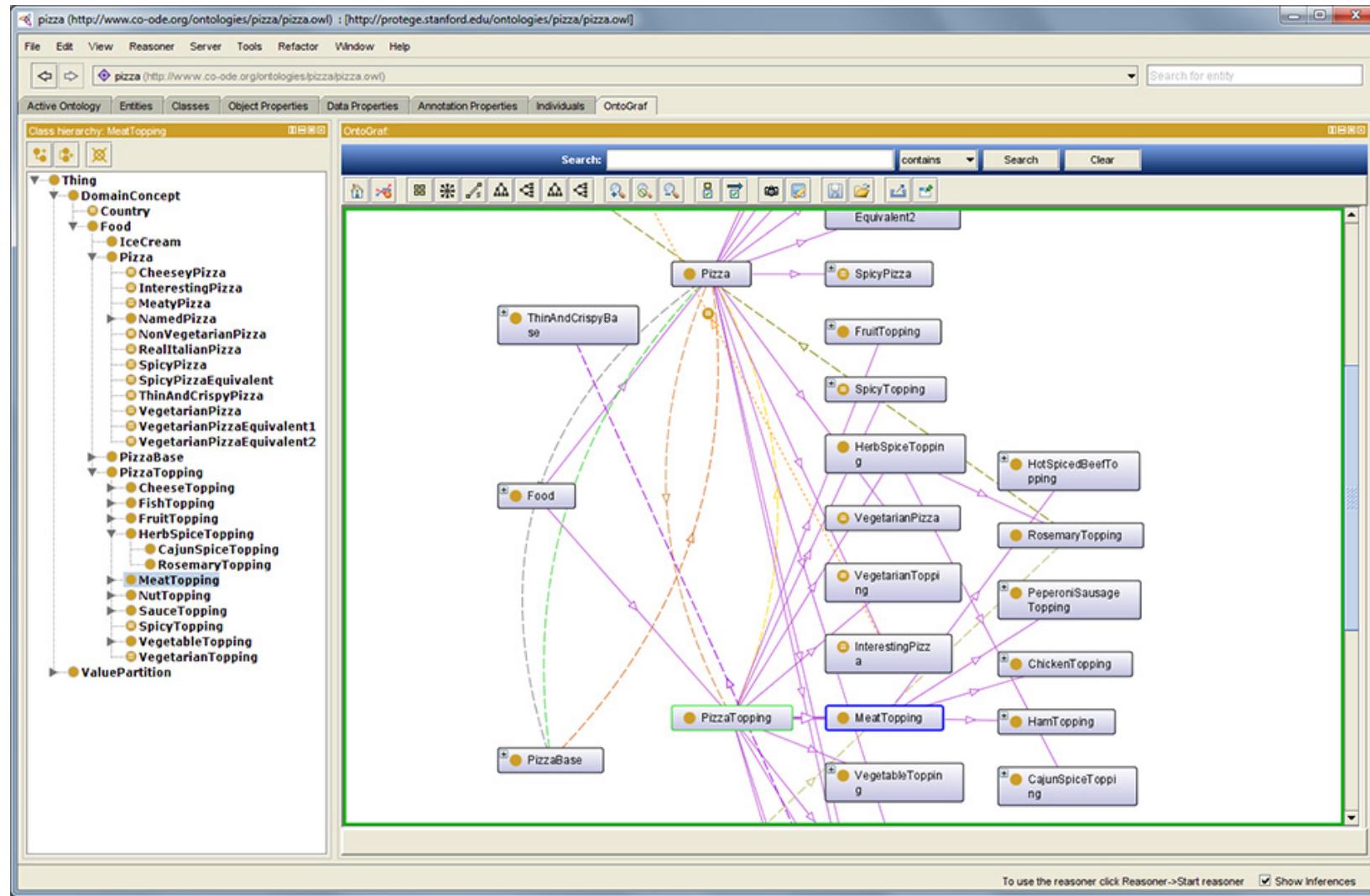
Assertions	DL-axioms
Every animal which can fly has wings	$\varphi_1: Fly \sqsubseteq HasWing$
Every animal which eats fish is a piscivore	$\varphi_2: \exists Eat.Fish \sqsubseteq Piscivore$
tweety is not an abnormal bird or cannot fly	$\varphi_3: (\neg AbnBird \sqcup \neg Fly)(tweety)$
ursidae eats salmon	$\varphi_4: Eat(ursidae, salmon)$
salmon is some fish	$\varphi_5: Fish(salmon)$
ursidae is not a piscivore	$\varphi_6: \neg Piscivore(ursidae)$

<https://doi.org/10.1371/journal.pone.0181056.t001>



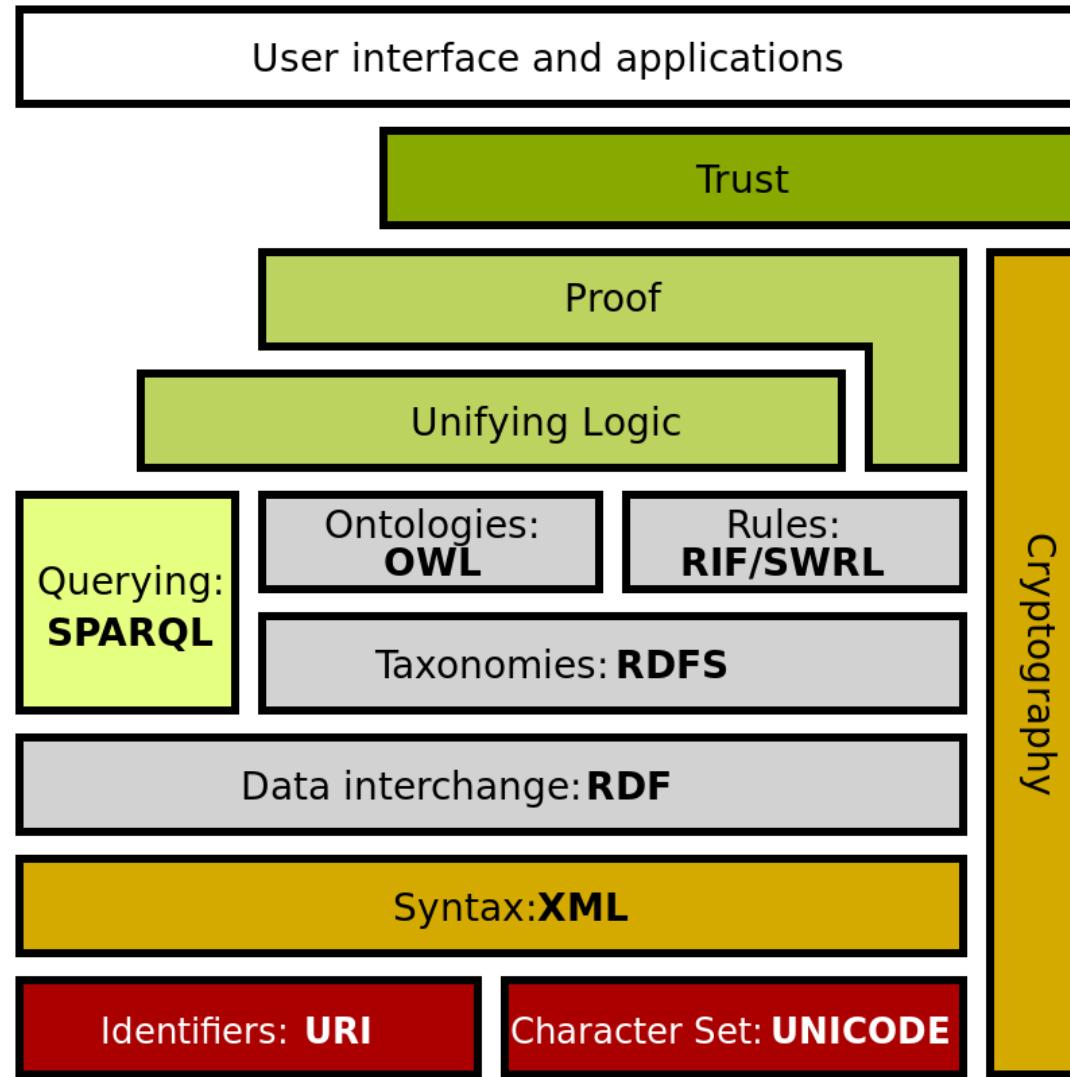
Source: Bergman. Platforms and Knowledge Management. 2018

Example tool for reasoning and ontologies: Protege



Source: <https://protege.stanford.edu/>

Putting it all together: Semantic Web Layer Cake



CROSS-DISCIPLINARY PERSPECTIVES: KNOWLEDGE DISCOVERY & DATA MINING

Knowledge Graph in Personal Assistant



Taylor Swift > Songs

Love Story Fearless · 2008	vevo
Look What You Made Me... Reputation · 2017	vevo
Shake It Off 1989 · 2014	vevo
Delicate Reputation · 2017	vevo

Source: Dong, Luna. Building a Broad Knowledge Graph for Products. Keynote at ICDE. 2019

Others

Scientific Text Mining

Jiang, M., & Shang, J. (2020, August). Scientific Text Mining and Knowledge Graphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 3537-3538).

Question Answering

Hixon, B., Clark, P., & Hajishirzi, H. (2015). Learning knowledge graphs for question answering through conversational dialog. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 851-861).

Recommendation Systems

Oramas, S., Ostuni, V. C., Noia, T. D., Serra, X., & Sciascio, E. D. (2016). Sound and music recommendation with knowledge graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(2), 1-21.

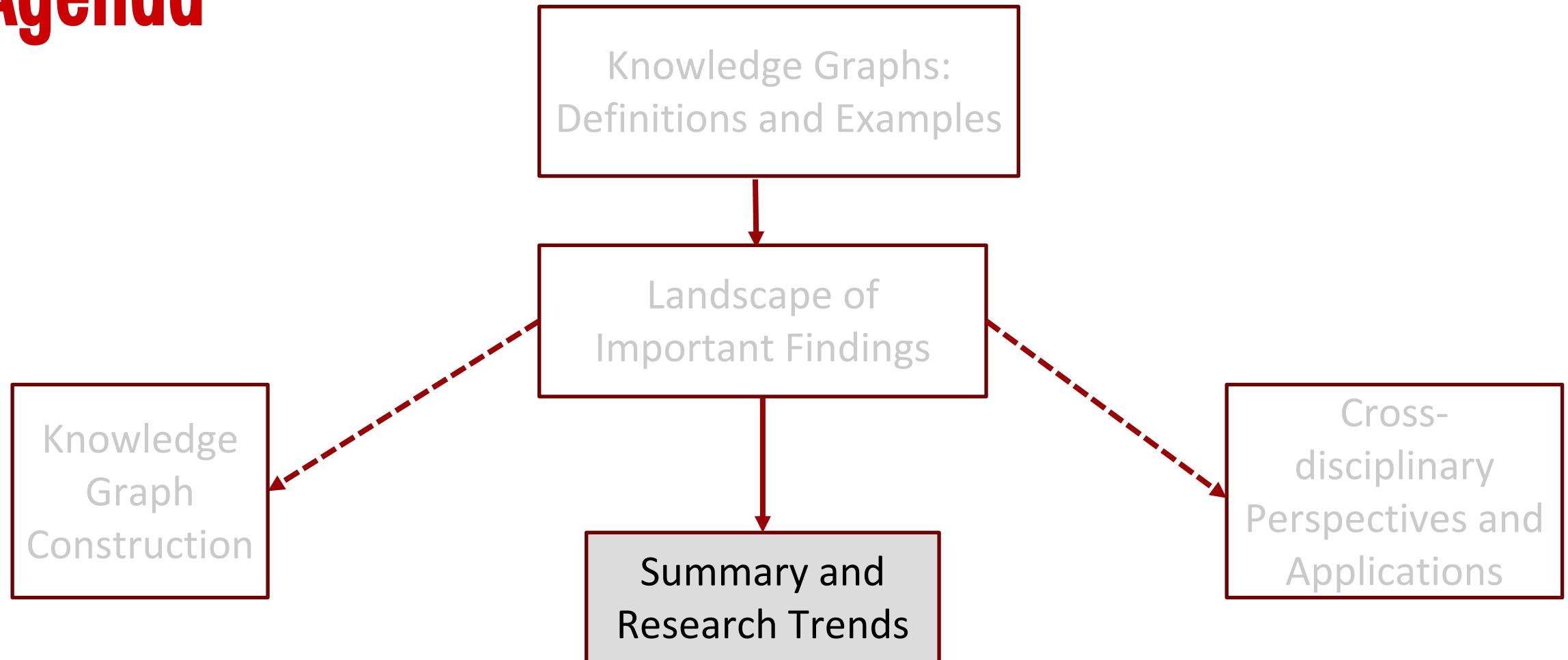
Summarization

Gunaratna, K., Yazdavar, A. H., Thirunarayan, K., Sheth, A., & Cheng, G. (2017, August). Relatedness-based multi-entity summarization. In *IJCAI: proceedings of the conference* (Vol. 2017, p. 1060). NIH Public Access.

Truth/fact-checking

Shiralkar, P., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2017, November). Finding streams in knowledge graphs to support fact checking. In *2017 IEEE International Conference on Data Mining (ICDM)* (pp. 859-864). IEEE.

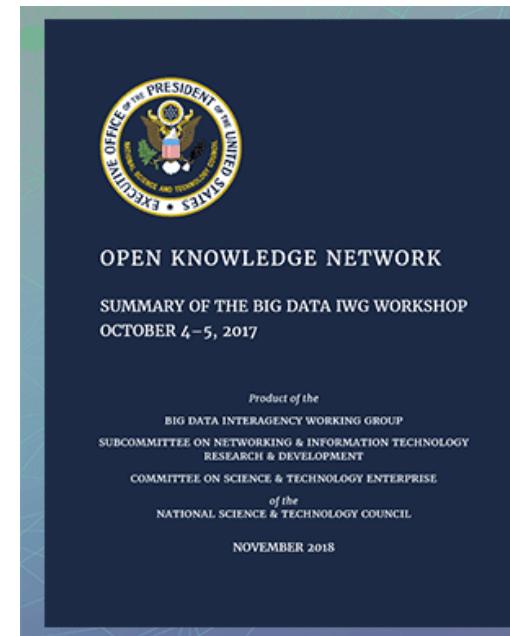
Agenda



Open Knowledge Network (OKN)

Technology companies develop **proprietary knowledge networks** as key business technologies today. However, because these networks are proprietary and expensive to construct, government, academia, small businesses, and nonprofits do not have access to them. In contrast, an open knowledge network (OKN) would be available to all stakeholders, including the researchers who will help push this technology further. An OKN requires a nonproprietary, public-private development effort that spans the entire data science community and will result in an **open, shared infrastructure**.

<https://www.nitrd.gov/pubs/Open-Knowledge-Network-Workshop-Report-2018.pdf>



<https://www.nitrd.gov/news/Open-Knowledge-Network-Workshop-Report-2018.aspx>

Knowledge Graphs for Social Good (KGSG)

Best Practices, Methods, and Challenges - Held May 4th, 2020 at KGC 2020

accenture

Technology Innovation

Applying knowledge graphs for social good

JUNE 26, 2020

Knowledge Graphs for Social Good: An Entity-centric Search Engine for the Human Trafficking Domain

Publisher: IEEE

Cite This

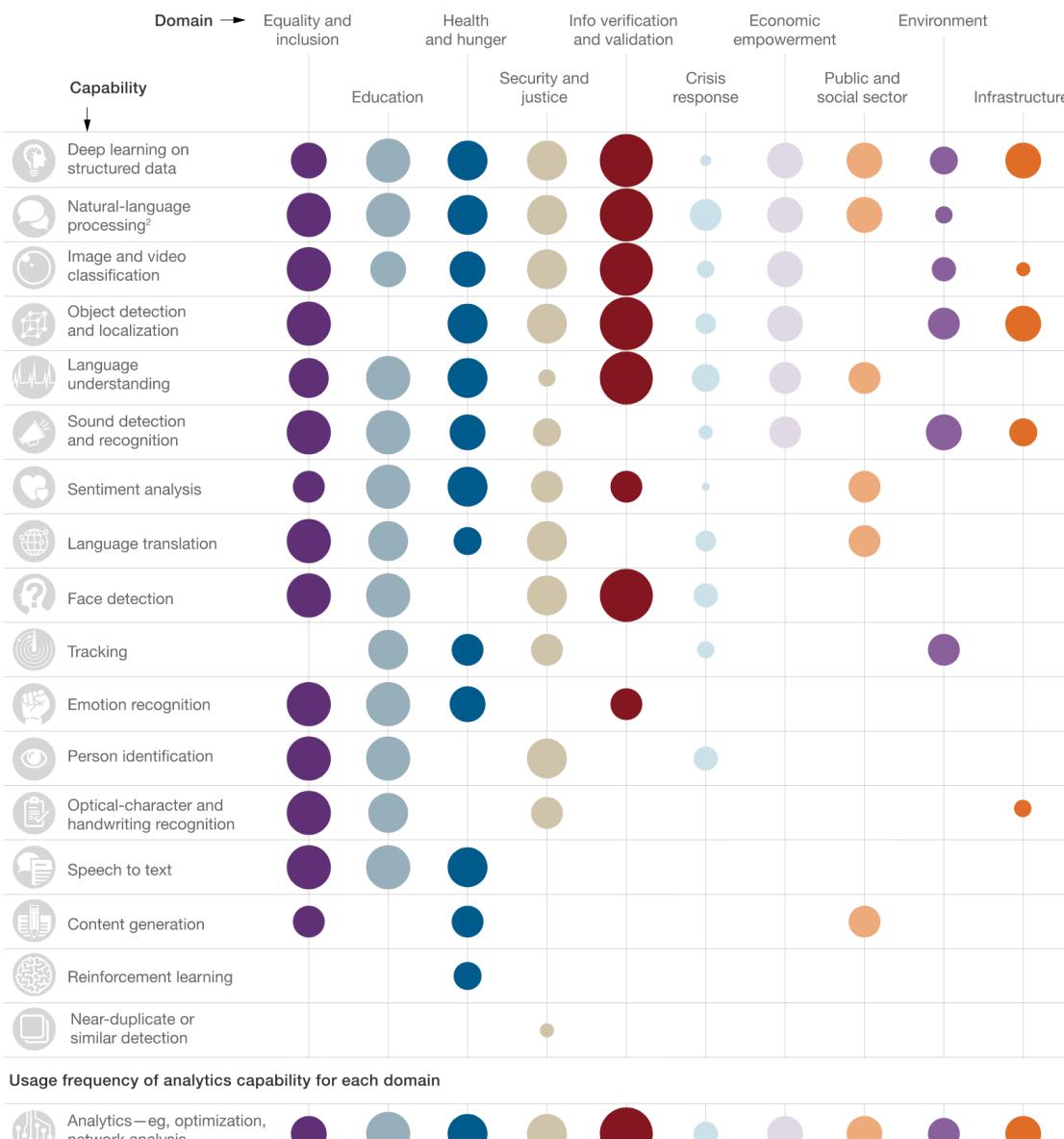
PDF

Mayank Kejriwal ; Pedro Szekely All Authors

Information Sciences Institute

Usage frequency of AI capability for each domain¹

Lower Higher

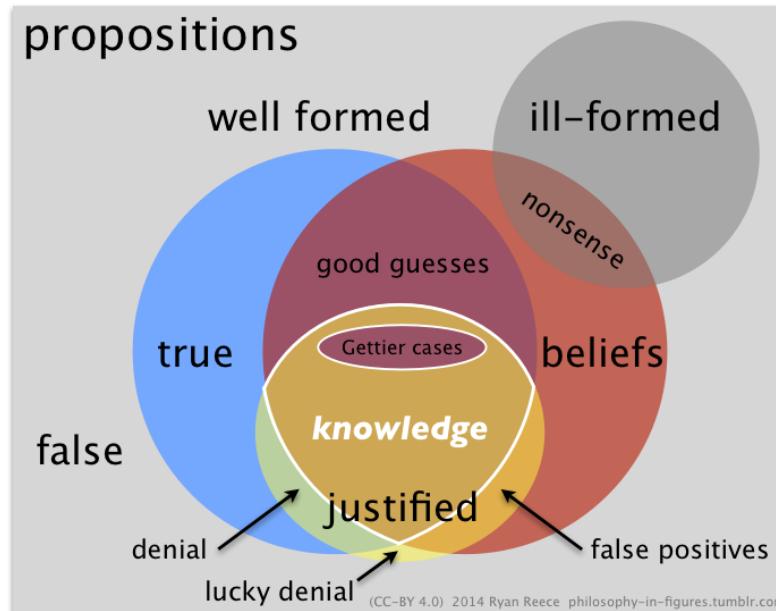


Note: Our library of about 160 use cases with societal impact is evolving and this heat map should not be read as a comprehensive gauge of the potential application of AI or analytics capabilities. Usage frequency estimates the number of times that models trained using AI would be used in a year to predict an outcome.

¹Log base 10 scale. Deployment frequency capped at once per hour per year to prevent skewing; capping affected only a small number of use cases.

²Excluding sentiment analysis, speech to text, language understanding, and translation.

Knowledge, semantics and context: what are they and how do we better define/represent them?



— Three quarks for Muster Mark!
Sure he hasn't got much of a bark
And sure any he has it's all beside the mark.
But O, Wreneagle Almighty, wouldn't un be a sky
To see that old buzzard whooping about for uns s/
And he hunting round for uns speckled trousers a/
stown Park?

Quark, one of the most influential of modern Ferengi thanks to his location at Deep Space Nine when the Bajoran wormhole was discovered, owns Quark's Bar on DS9's Promenade, but hates being called a "barkeep," preferring "host" instead as he fancies himself an empathetic dispenser of advice as well as a goodwill ambassador and legitimate entrepreneur extrordinaire.

Quarks and [Leptons](#) are the building blocks which build up matter, i.e., they are seen as the "elementary particles". In the present standard model, there are six "flavors" of quarks. They can successfully account for all known [mesons](#) and [baryons](#) (over 200). The most familiar baryons are the [proton](#) and [neutron](#), which are each constructed from up and down quarks. Quarks are observed to occur only in combinations of two quarks (mesons), three quarks (baryons). There was a recent claim of observation of particles with five quarks ([pentaquark](#)), but further experimentation has not borne it out.

Quark is similar to French [fromage blanc](#), Indian [paneer](#), and the [queso fresco/queijo fresco](#) made in the Iberian Peninsula and in some Latin American countries. It is distinct from Italian [ricotta](#) because ricotta (Italian "recooked") is made from scalded [whey](#). Quark is somewhat similar to [yogurt cheeses](#) such as the South Asian [chak\(k\)a](#), the Arabic [labneh](#), and the Central Asian [suzma](#) or [kashk](#), but while these products are obtained by straining [yogurt](#) (milk fermented with [thermophile](#) bacteria),

Explainable AI

On The Role of Knowledge Graphs in Explainable AI

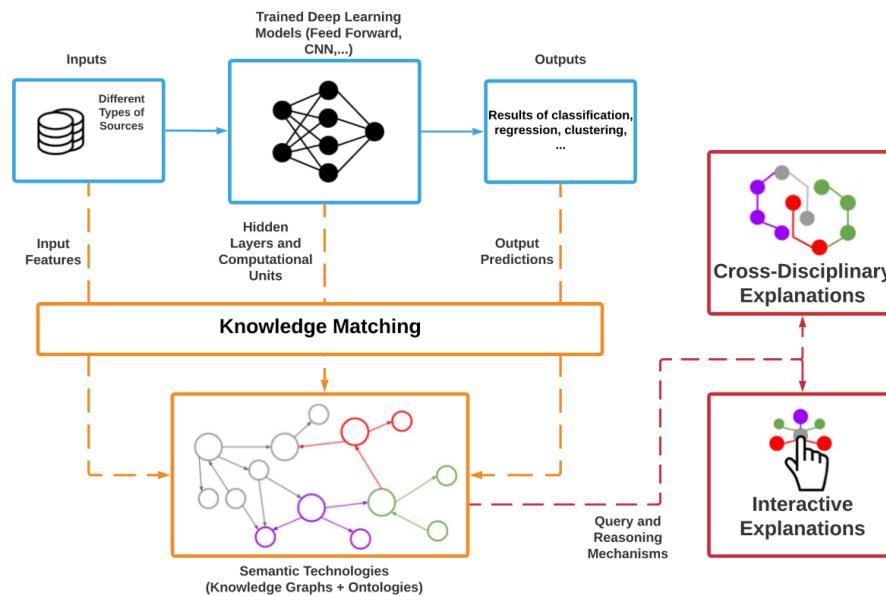
Freddy Lecue^{a,b}

^a CortAIx, Thales, Montreal, Canada

E-mail: freddy.lecure@inria.fr

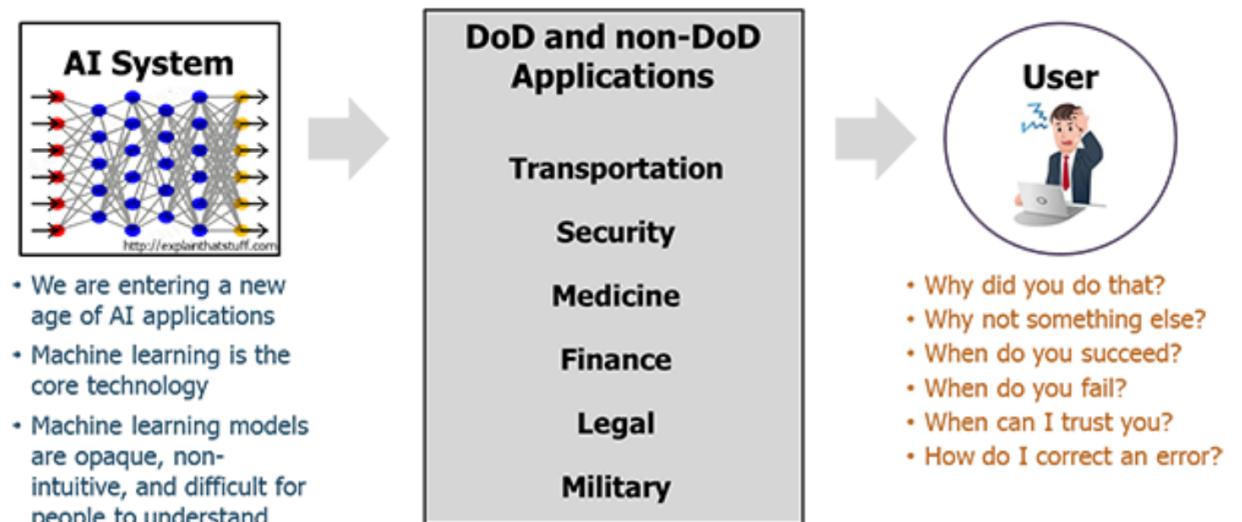
^b WIMMICS, Inria, Sophia Antipolis, France

E-mail: freddy.lecue@thalesgroup.fr



Explainable Artificial Intelligence (XAI)

Dr. Matt Turek



Source: Knowledge Graphs For eXplainable AI. On the Integration of Semantic Technologies and Symbolic Systems into Deep Learning Models for a More Comprehensible Artificial Intelligence.

<https://towardsdatascience.com/knowledge-graphs-for-explainable-ai-dcd73c5c016>

WRAPUP

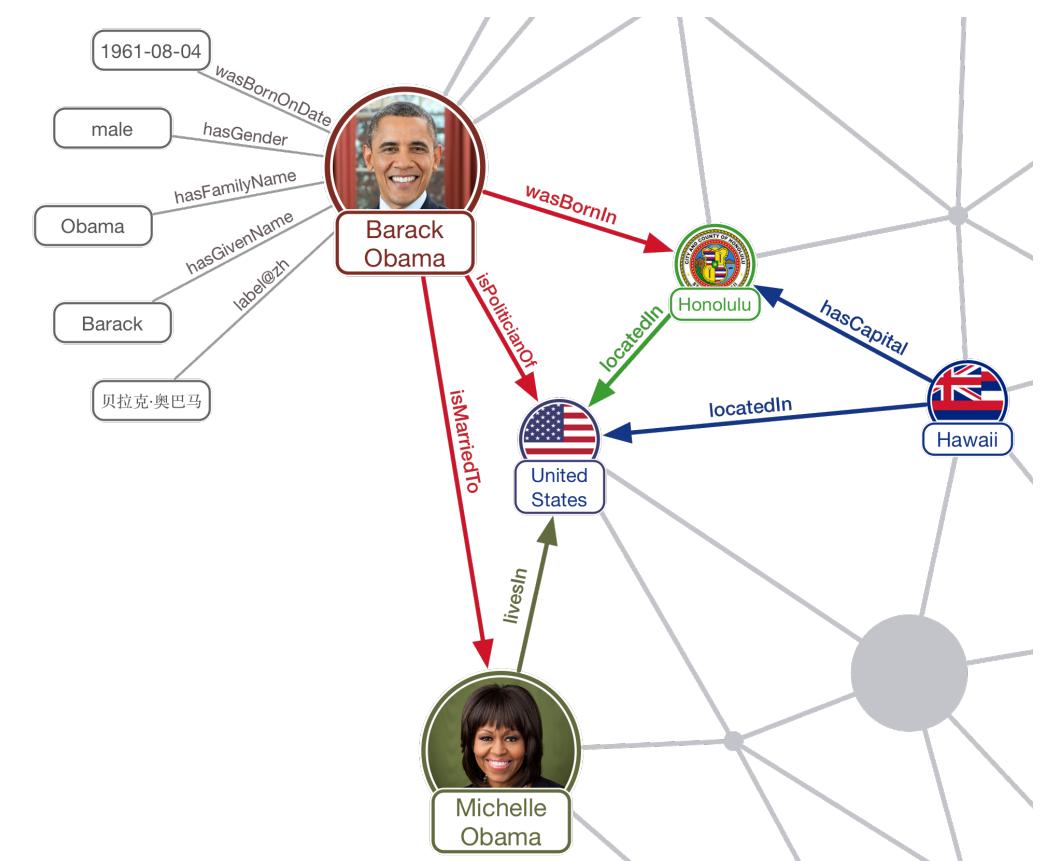
What is a Knowledge Graph?

- Set of triples, where each triple (h, r, t) represents a **relationship r** between **head entity h** and **tail entity t**

$(\text{Barack Obama}, \text{wasBornOnDate}, 1961-08-04)$,
 $(\text{Barack Obama}, \text{hasGender}, \text{male})$,

...
 $(\text{Hawaii}, \text{hasCapital}, \text{Honolulu})$,

...
 $(\text{Michelle Obama}, \text{livesIn}, \text{United States})$



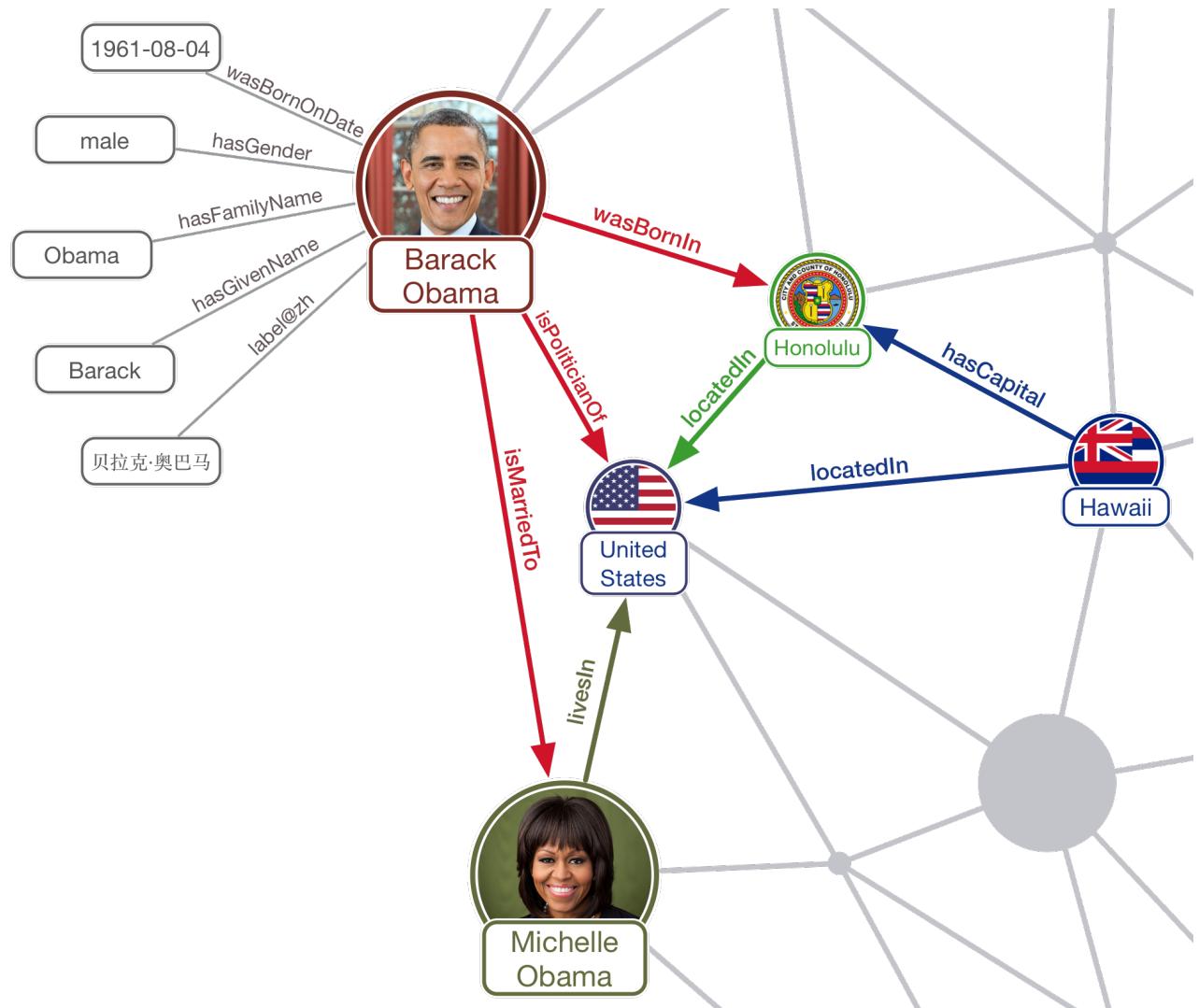
What is a Knowledge Graph?

Technically, a multi-relational directed labeled graph with semantics

Both edges and nodes have labels, but not all labels are equal (literals vs. identifiers)

Where do the semantics come from?

- Complex question, only starting to be understood



More on semantics

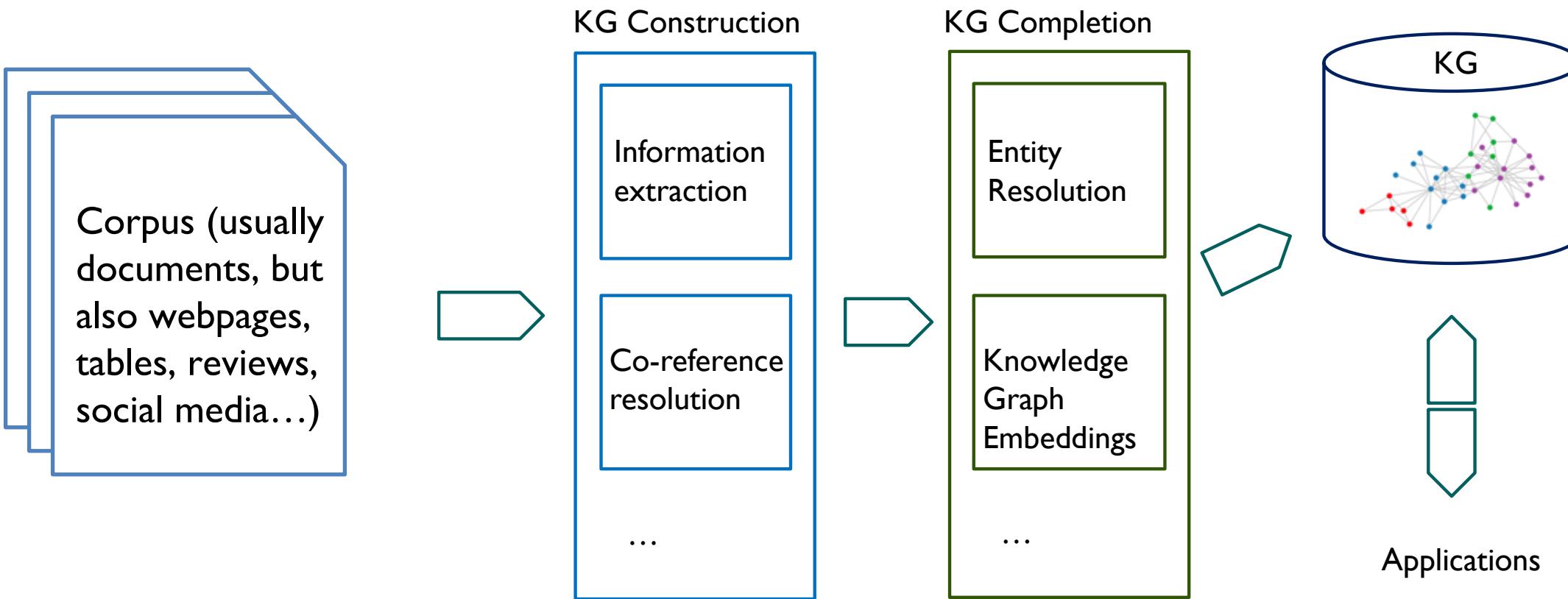
Traditionally, semantics are believed to come from ontology

- An ontology is a ‘formal, explicit specification of a shared conceptualization’ (we will go deeper into this in a while)
- In philosophy, an ontology is a ‘study of what there is’ including the study of the ‘most general features of what there is, and how the things there are relate to each other in the metaphysically most general ways’

Source: <https://plato.stanford.edu/entries/logic-ontology/>

More recently, in AI, we have started to recognize a more commonsense view of semantics guided by findings in linguistics and distributional semantics

A typical KGC workflow starts from corpus acquisition and ends with applications

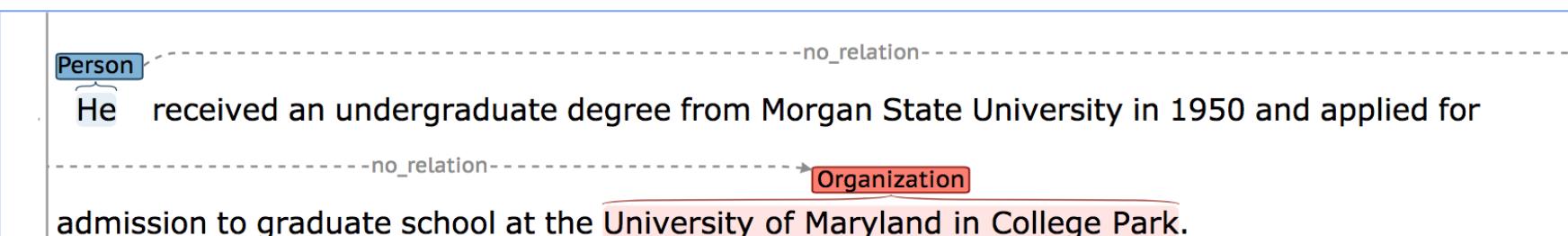
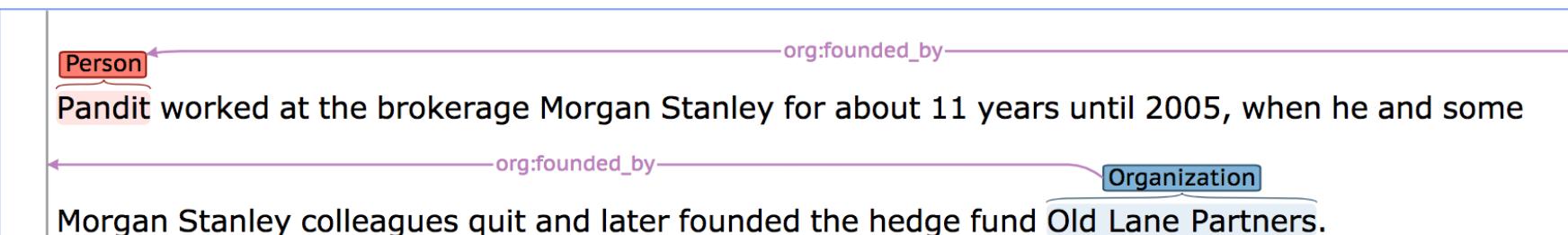
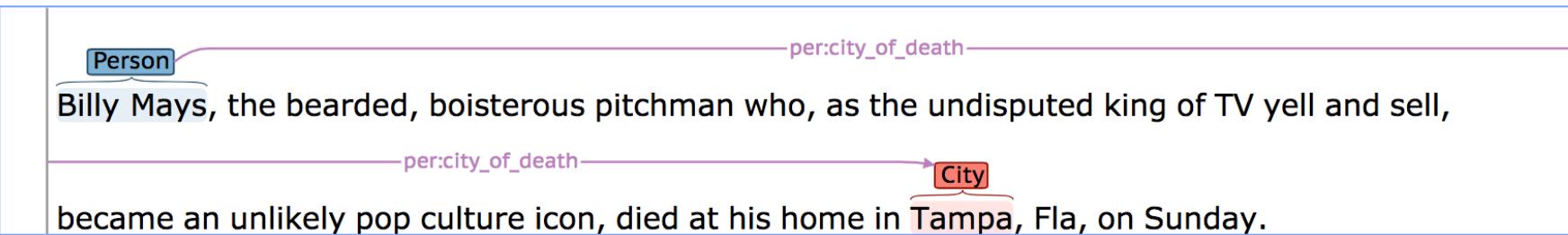


Named Entity Recognition (NER)

contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday's PaperAdvertisementSupported ORG byF.B.I. Agent Peter Strzok PERSON , Who Criticized Trump PERSON in Texts, Is FiredImagePeter Strzok, a top F.B.I. GPE counterintelligence agent who was taken off the special counsel investigation after his disparaging texts about President Trump PERSON were uncovered, was fired. CreditT.J. Kirkpatrick PERSON for The New York TimesBy Adam Goldman ORG and Michael S. SchmidtAug PERSON . 13 CARDINAL , 2018WASHINGTON CARDINAL — Peter Strzok PERSON , the F.B.I. GPE senior counterintelligence agent who disparaged President Trump PERSON in inflammatory text messages and helped oversee the Hillary Clinton PERSON email and Russia GPE investigations, has been fired for violating bureau policies, Mr. Strzok PERSON 's lawyer said Monday DATE .Mr. Trump and his allies seized on the texts — exchanged during the 2016 DATE campaign with a former F.B.I. GPE lawyer, Lisa Page — in PERSON assailing the Russia GPE investigation as an illegitimate "witch hunt." Mr. Strzok PERSON , who rose over 20 years DATE at the F.B.I. GPE to become one of its most experienced counterintelligence agents, was a key figure in the early months DATE of the inquiry.Along with writing the texts, Mr. Strzok PERSON was accused of sending a highly sensitive search warrant to his personal email account.The F.B.I. GPE had been under immense political pressure by Mr. Trump PERSON to dismiss Mr. Strzok PERSON , who was removed last summer DATE from the staff of the special counsel, Robert S. Mueller III PERSON . The president has repeatedly denounced Mr. Strzok PERSON in posts on

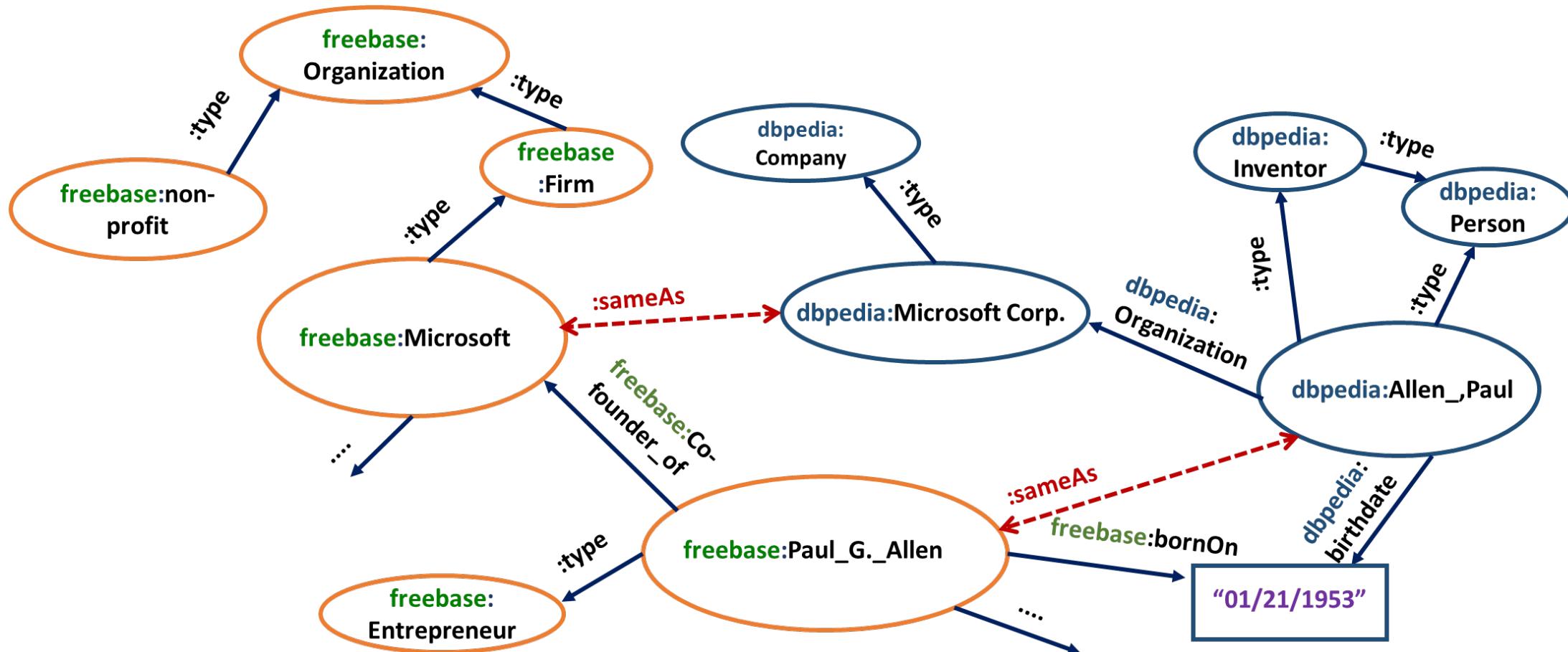
Source: Named Entity Recognition and Classification with Scikit-Learn. <https://towardsdatascience.com/named-entity-recognition-and-classification-with-scikit-learn-f05372f07ba2>

Other kinds of IE: Relation Extraction



Source: Stanford TACRED

In the world of knowledge graphs



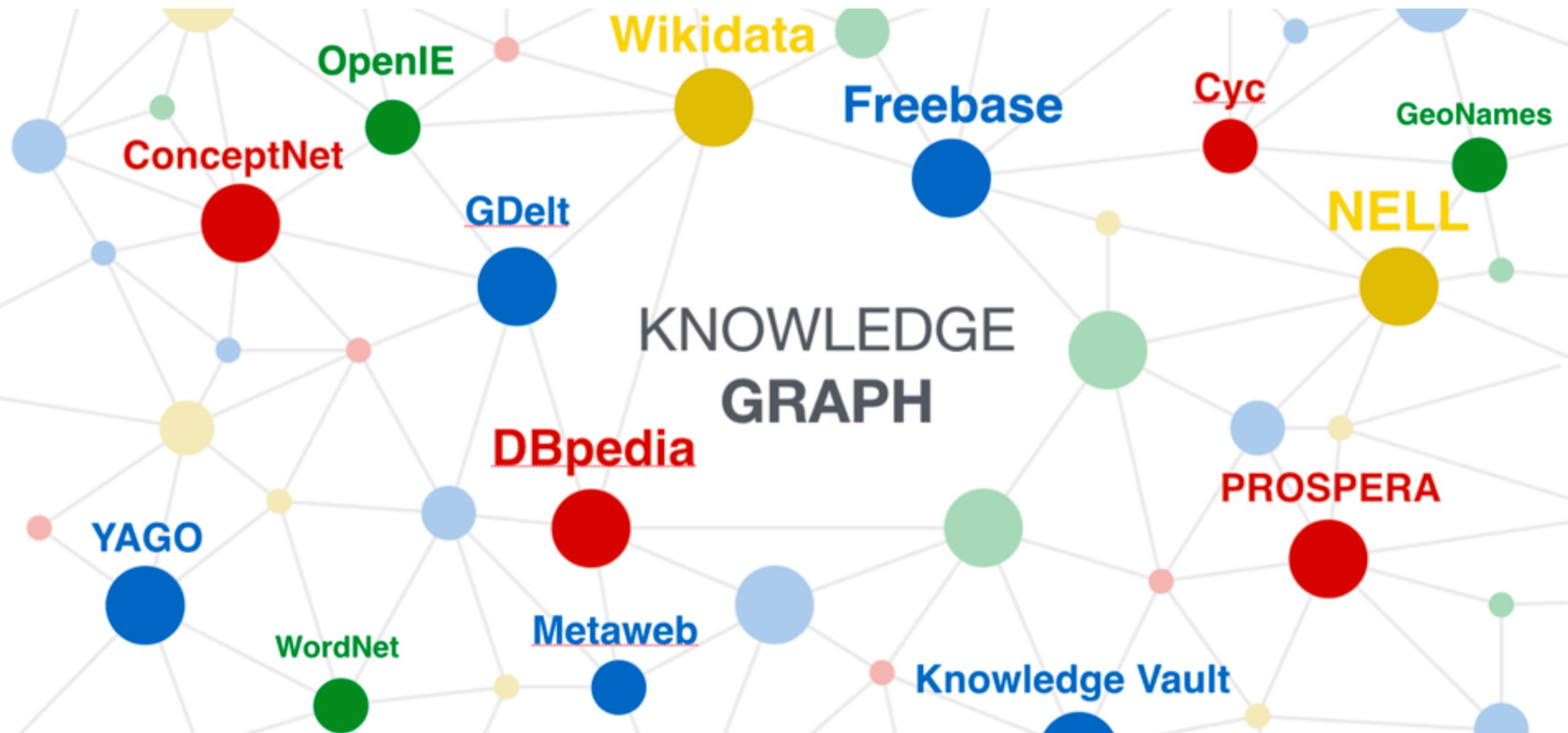
KGEs (results)

Useful resources:

- OpenKE:
<http://139.129.163.161//index/toolkits#pretrained-embeddings>
- StarSpace:
<https://github.com/facebookresearch/StarSpace>
- Recent transformer-based models could potentially be adapted, including BERT and RoBERTa:
<https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems/>

Method	Raw						Filtered					
	WN18			FB15k			WN18			FB15k		
	MR	H@10	MRR	MR	H@10	MRR	MR	H@10	MRR	MR	H@10	MRR
SE (Bordes et al., 2011)	1011	68.5	-	273	28.8	-	985	80.5	-	162	39.8	-
Unstructured (Bordes et al., 2012)	315	35.3	-	1074	4.5	-	304	38.2	-	979	6.3	-
SME (Bordes et al., 2012)	545	65.1	-	274	30.7	-	533	74.1	-	154	40.8	-
TransH (Wang et al., 2014)	401	73.0	-	212	45.7	-	303	86.7	-	87	64.4	-
TransR (Lin et al., 2015b)	238	79.8	-	198	48.2	-	225	92.0	-	77	68.7	-
CTransR (Lin et al., 2015b)	231	79.4	-	199	48.4	-	218	92.3	-	75	70.2	-
KG2E (He et al., 2015)	342	80.2	-	174	48.9	-	331	92.8	-	59	74.0	-
TransD (Ji et al., 2015)	224	79.6	-	194	53.4	-	212	92.2	-	91	77.3	-
lppTransD (Yoon et al., 2016)	283	80.5	-	195	53.0	-	270	94.3	-	78	78.7	-
TranSparse (Ji et al., 2016)	223	80.1	-	187	53.5	-	211	93.2	-	82	79.5	-
TATEC (García-Durán et al., 2016)	-	-	-	-	-	-	-	-	-	58	76.7	-
NTN (Socher et al., 2013)	-	-	-	-	-	-	-	66.1	0.53	-	41.4	0.25
DISTMULT (Yang et al., 2015)	-	-	-	-	-	-	-	94.2	0.83	-	57.7	0.35
ComplEx (Trouillon et al., 2016)	-	-	0.587	-	-	0.242	-	94.7	0.941	-	84.0	0.692
HoIE (Nickel et al., 2016b)	-	-	0.616	-	-	0.232	-	94.9	0.938	-	73.9	0.524
RESCAL (Nickel et al., 2011) [*]	-	-	0.603	-	-	0.189	-	92.8	0.890	-	58.7	0.354
TransE (Bordes et al., 2013) [*]	-	-	0.351	-	-	0.222	-	94.3	0.495	-	74.9	0.463
STransE (Nguyen et al., 2016b)	217	80.9	0.469	219	51.6	0.252	206	93.4	0.657	69	79.7	0.543
RTransE (García-Durán et al., 2015)	-	-	-	-	-	-	-	-	-	50	76.2	-
PTransE (Lin et al., 2015a)	-	-	-	207	51.4	-	-	-	-	58	84.6	-
GAKE (Feng et al., 2016b)	-	-	-	228	44.5	-	-	-	-	119	64.8	-
Gaifman (Niepert, 2016)	-	-	-	-	-	-	352	93.9	-	75	84.2	-
Hiri (Liu et al., 2016)	-	-	-	-	-	-	-	90.8	0.691	-	70.3	0.603
NLFeat (Toutanova and Chen, 2015)	-	-	-	-	-	-	-	94.3	0.940	-	87.0	0.822
TEKE_H (Wang and Li, 2016)	127	80.3	-	212	51.2	-	114	92.9	-	108	73.0	-
SSP (Xiao et al., 2017)	168	81.2	-	163	57.2	-	156	93.2	-	82	79.0	-

Open-source KGs that have been built



Many applications and open research areas!

Information retrieval

Semantic Web

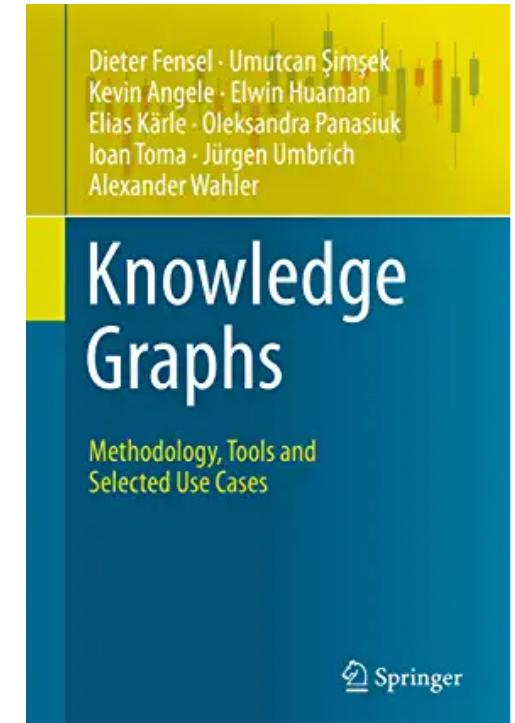
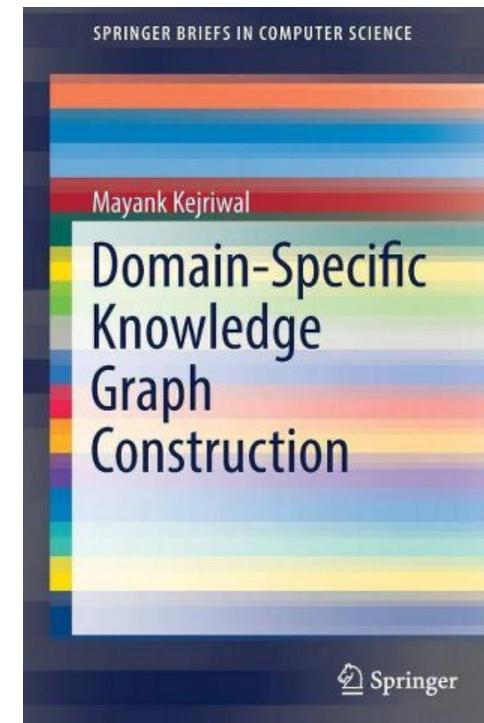
Recommender systems

Knowledge discovery/data mining

?

Numerous surveys, some more technical/field-specific

- Ehrlinger, L., & Wöß, W. (2016). Towards a Definition of Knowledge Graphs. *SEMANTiCS (Posters, Demos, SuCESS)*, 48, 1-4
- Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., & Taylor, J. (2019). Industry-scale knowledge graphs: lessons and challenges. *Queue*, 17(2), 48-75
- Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2015). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1), 11-33.
- Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2020). A survey on knowledge graphs: Representation, acquisition and applications. *arXiv preprint arXiv:2002.00388*.
- Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3), 489-508.



Upcoming:

Knowledge Graphs: Fundamentals, Techniques, and Applications (Adaptive Computation and Machine Learning series). *Kejriwal, Knoblock and Szekely*.



Center on Knowledge Graphs

Information Sciences Institute



Q & A

kejriwal@isi.edu

