Knowledge Graphs: A Practical Introduction across Disciplines

Mayank Kejriwal
University of Southern California

December, 2020

About me

E-Commerce

E-Commerce Knowledge Graphs and Representation Learning

Common Sense Reasoning

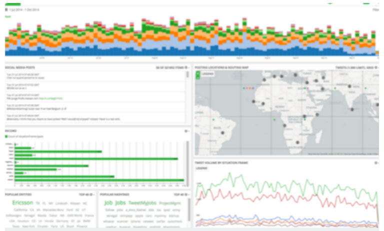
Multi-modal Open World Grounded Learning and Inference

The Human Trafficking Project

The Human Trafficking Project

GNOME

Generating Novelties in Open-world Multi-agent Environments



AI for Crisis Response

Text-enabled Humanitarian Operations in Real-time

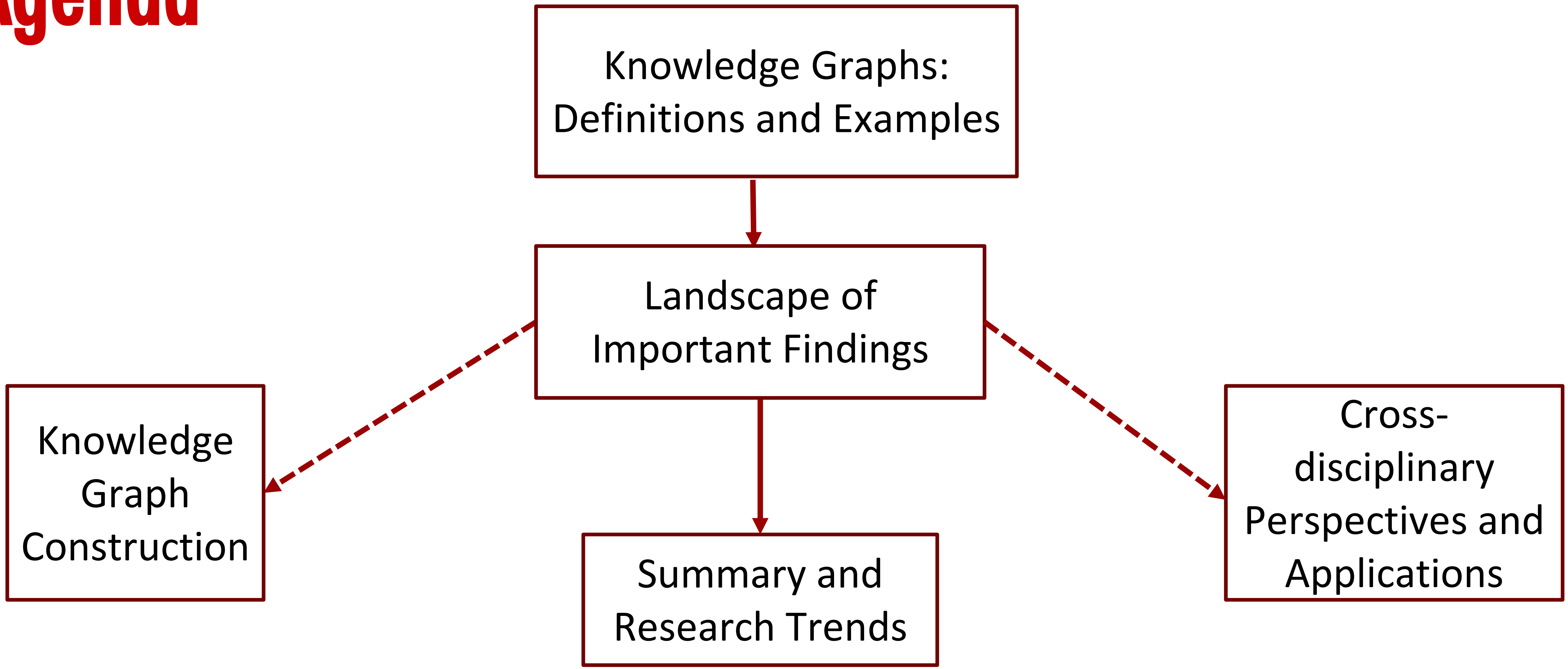
AI, Networks and Society

AI, Networks and Society

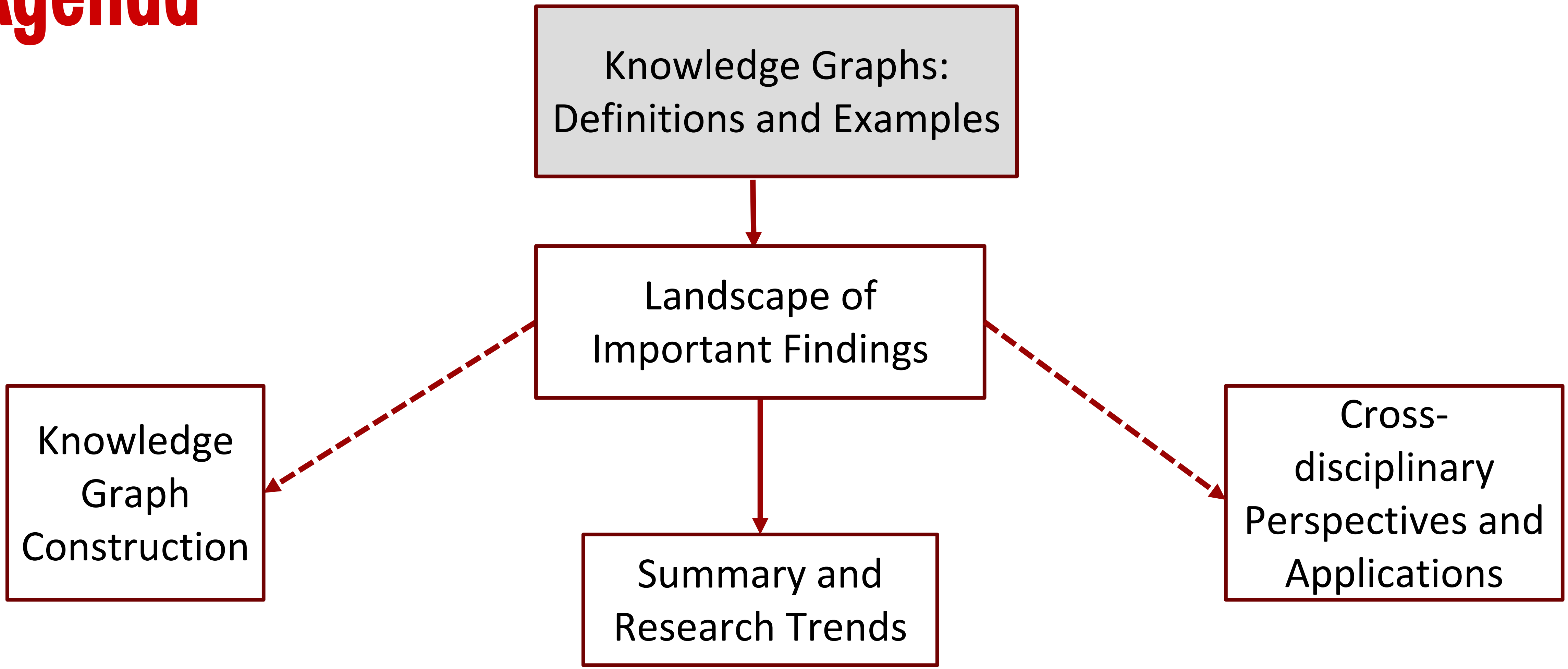
USC Viterbi
School of Engineering
Daniel J. Epstein Department of Industrial and Systems Engineering

USC Viterbi
School of Engineering
Information Sciences Institute

Agenda


Knowledge Graphs: Definitions and Examples
Landscape of Important Findings
Knowledge Graph Construction
Summary and Research Trends
Cross-disciplinary Perspectives and Applications

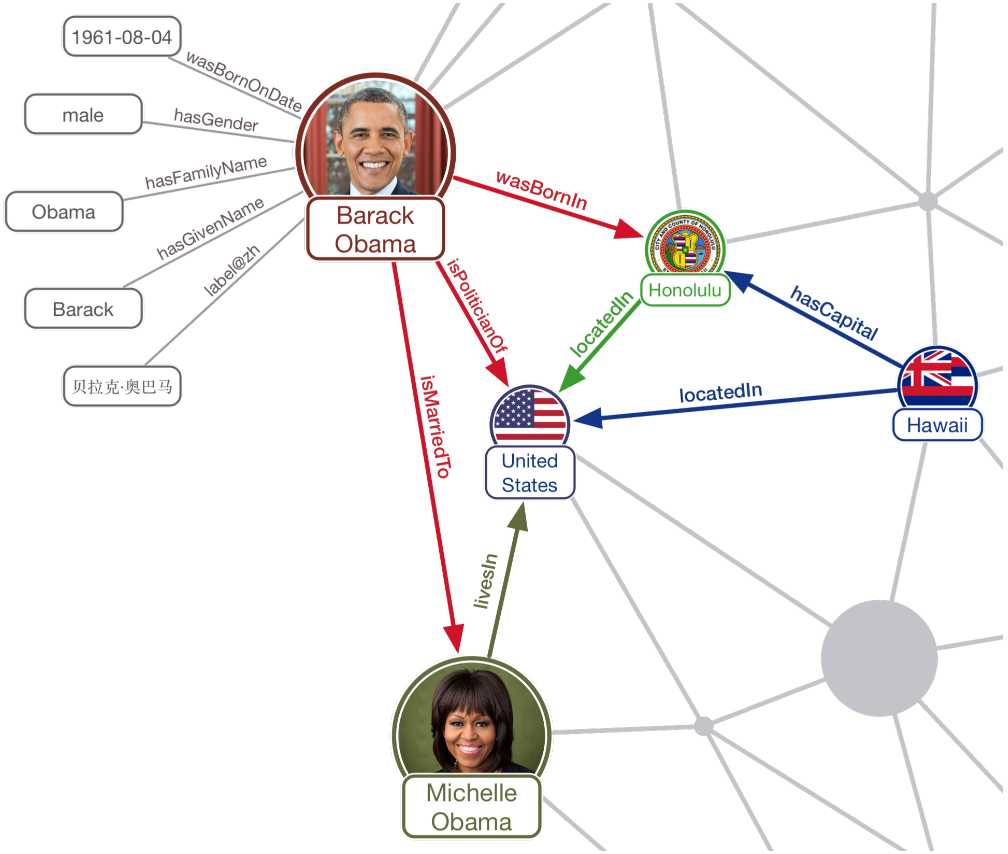
Agenda


Knowledge Graphs: Definitions and Examples
Landscape of Important Findings
Knowledge Graph Construction
Summary and Research Trends
Cross-disciplinary Perspectives and Applications

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

(Barack Obama, wasBornOnDate, 1961-08-04),
(Barack Obama, hasGender, male),
…
(Hawaii, hasCapital, Honolulu),
…
(Michelle Obama, livesIn, United States)



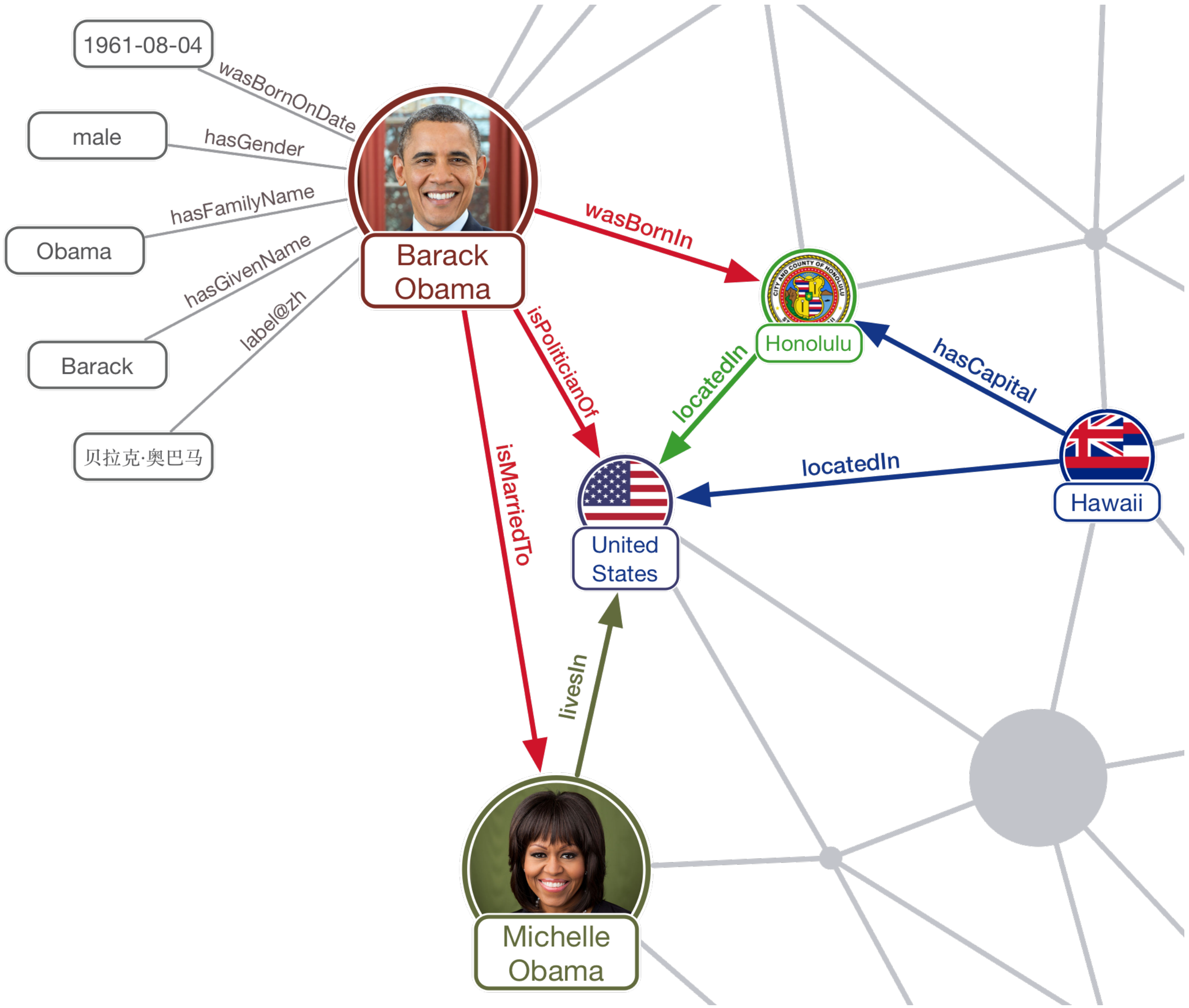
What is a Knowledge Graph?

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

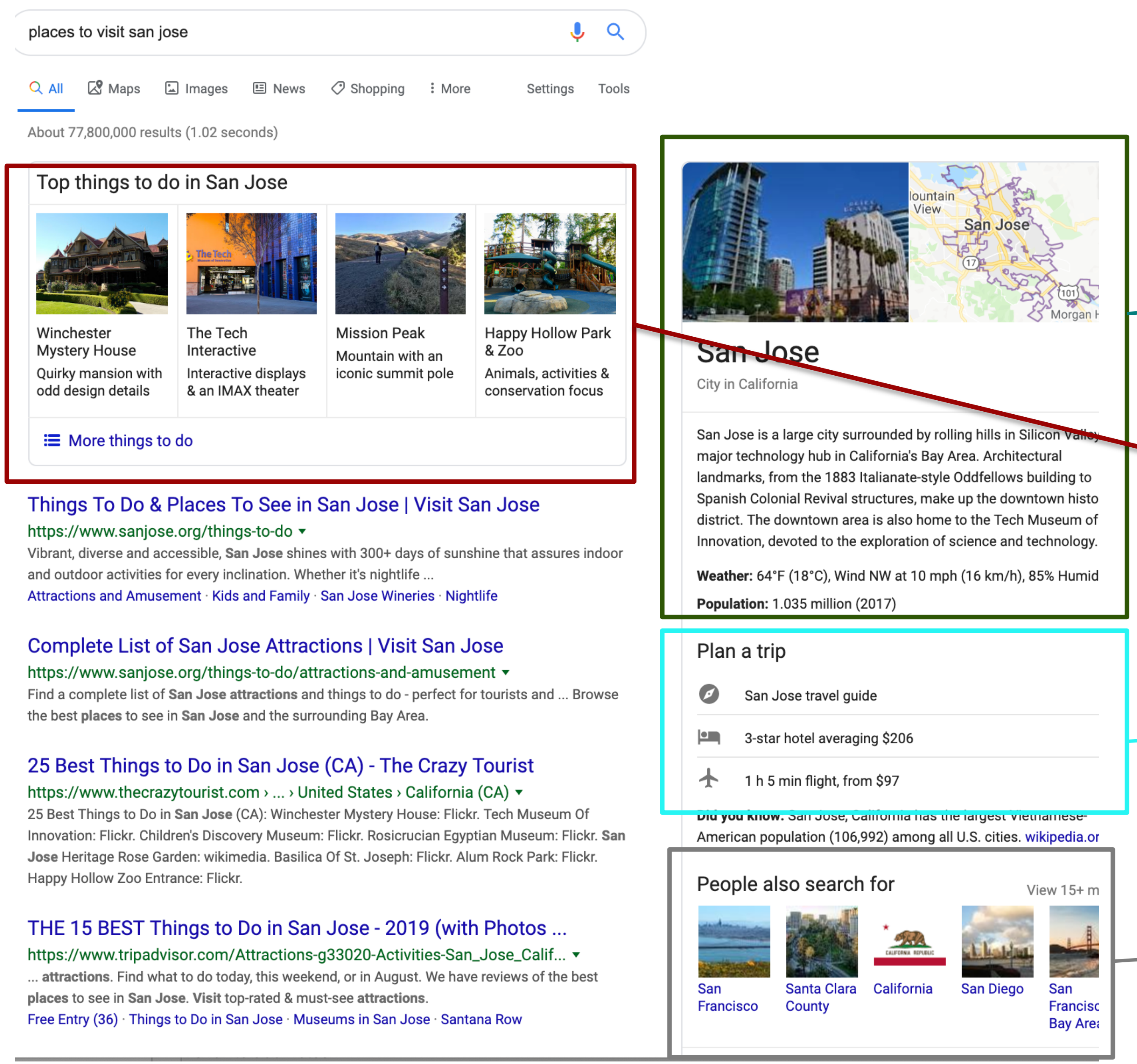
Traditionally, semantics are believed to come from ontology

- 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

Have I seen this before?



Knowledge panel

Recognition of user intent

Recommendations

Exploration-suggestions

Agenda

Knowledge Graphs: Definitions and Examples

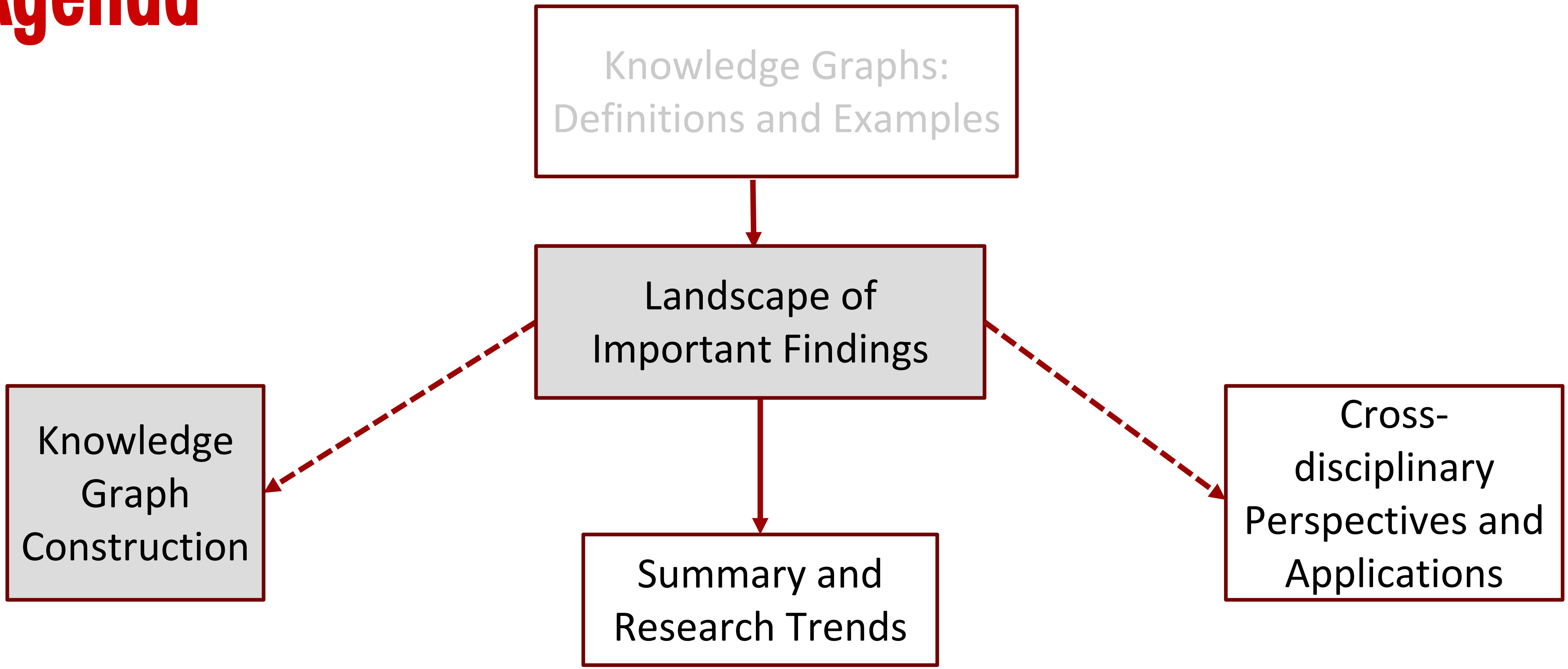
Landscape of Important Findings

Knowledge Graph Construction

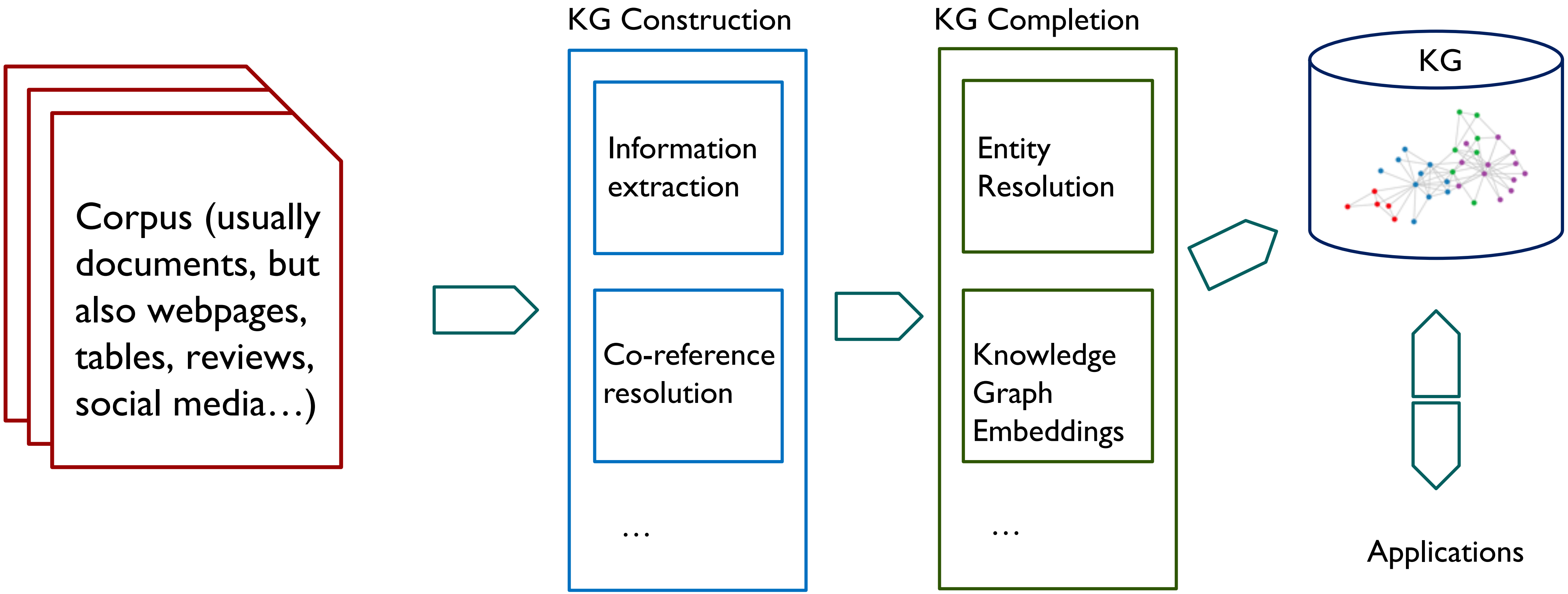
Summary and Research Trends

Cross-disciplinary Perspectives and Applications

Agenda


Knowledge Graphs: Definitions and Examples
Landscape of Important Findings
Knowledge Graph Construction
Summary and Research Trends
Cross-disciplinary Perspectives and Applications

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

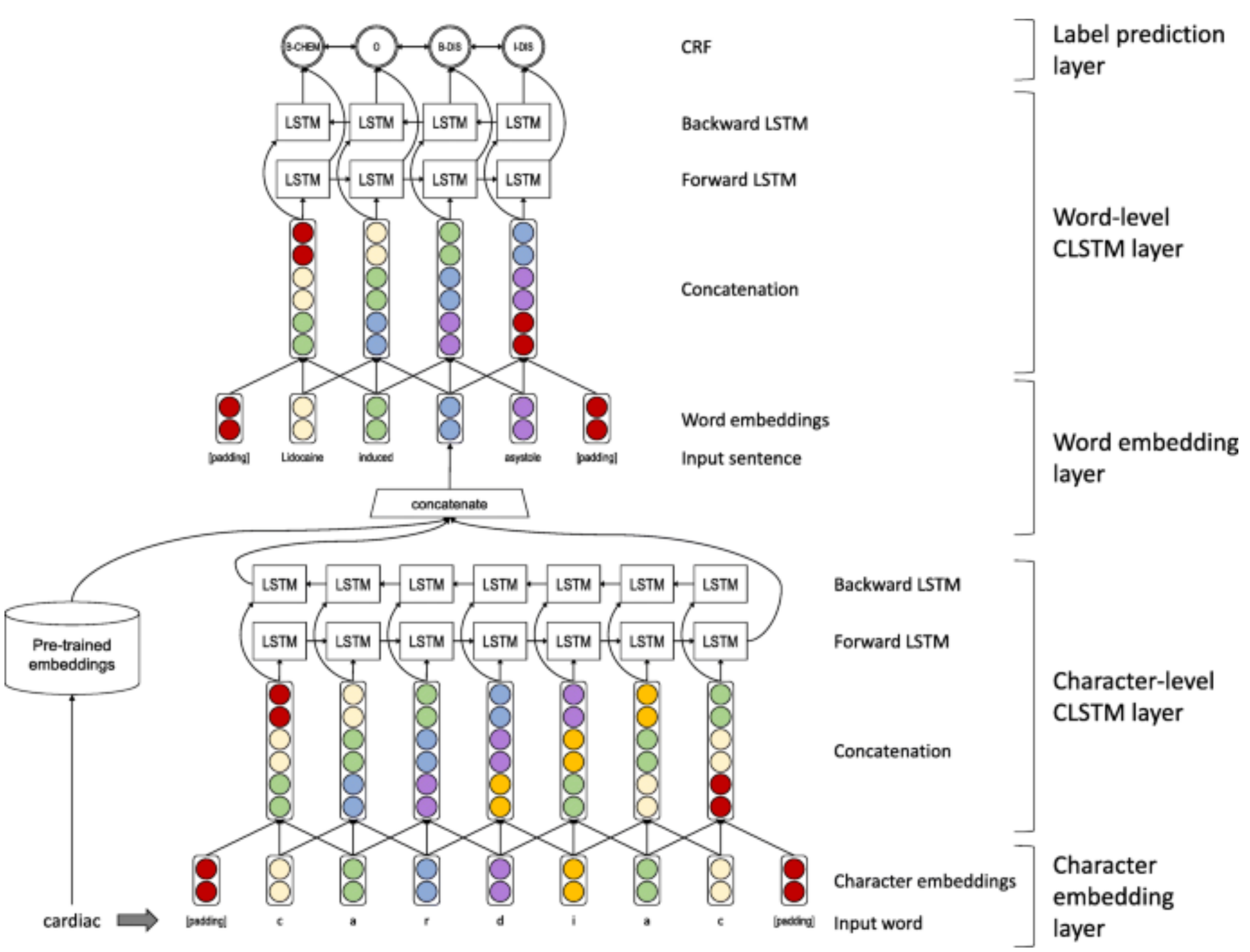
NER workflows

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)

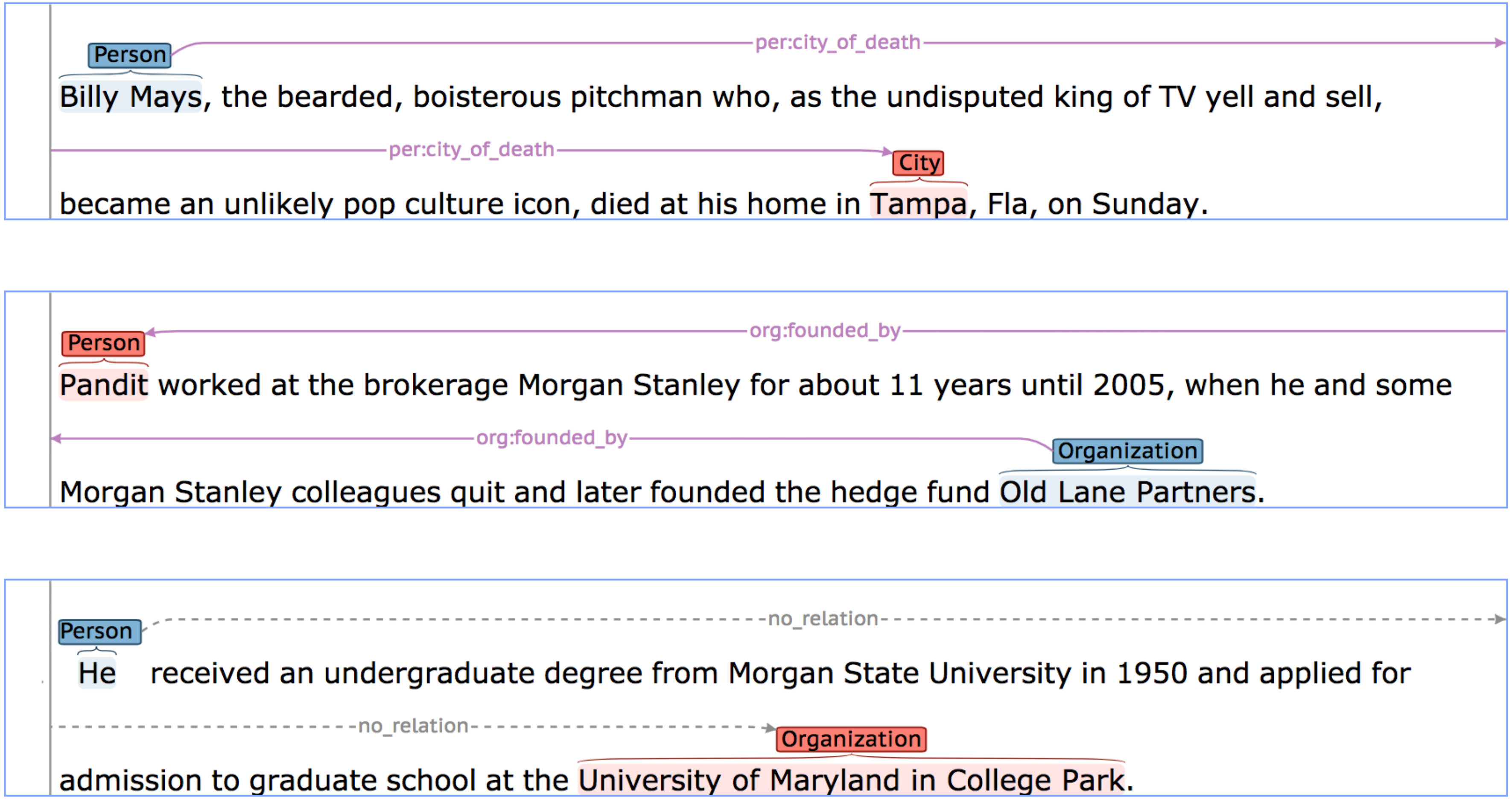
Today, deep learning methods designed for sequences (such as RNNs and, more recently, transformers) are state-of-the-art

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	*set out from*	eastern Queens	✓
your dry cleaner	*set out from_on*	foot	✓
your dry cleaner	*is near*	Maspeth	✓
North Korea	*has begun to allow*	tourist	✓
South Korean tourists	*to go to*	Kaesong	✓
...	...	...	
Queens	*on*	foot	✗
Kaesong	*on*	a limited basis	✗

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

Is IE a solved problem?

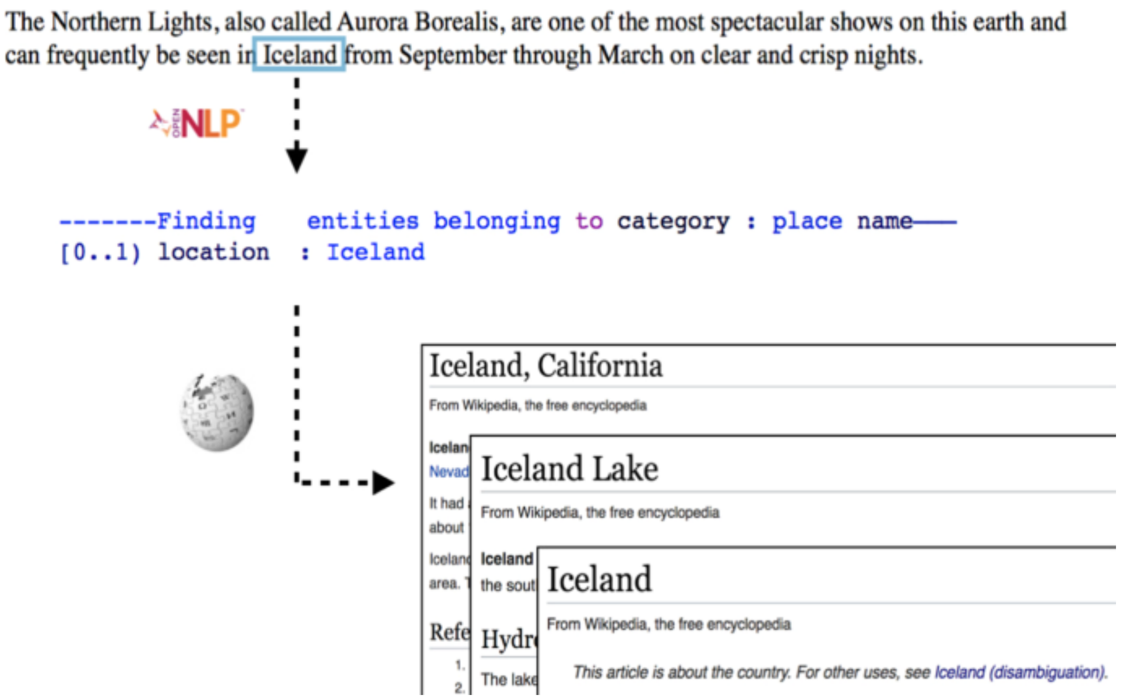
Table 2: Main Results on Testing Set: F_1 Score (Precision/Recall) (in %)

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)

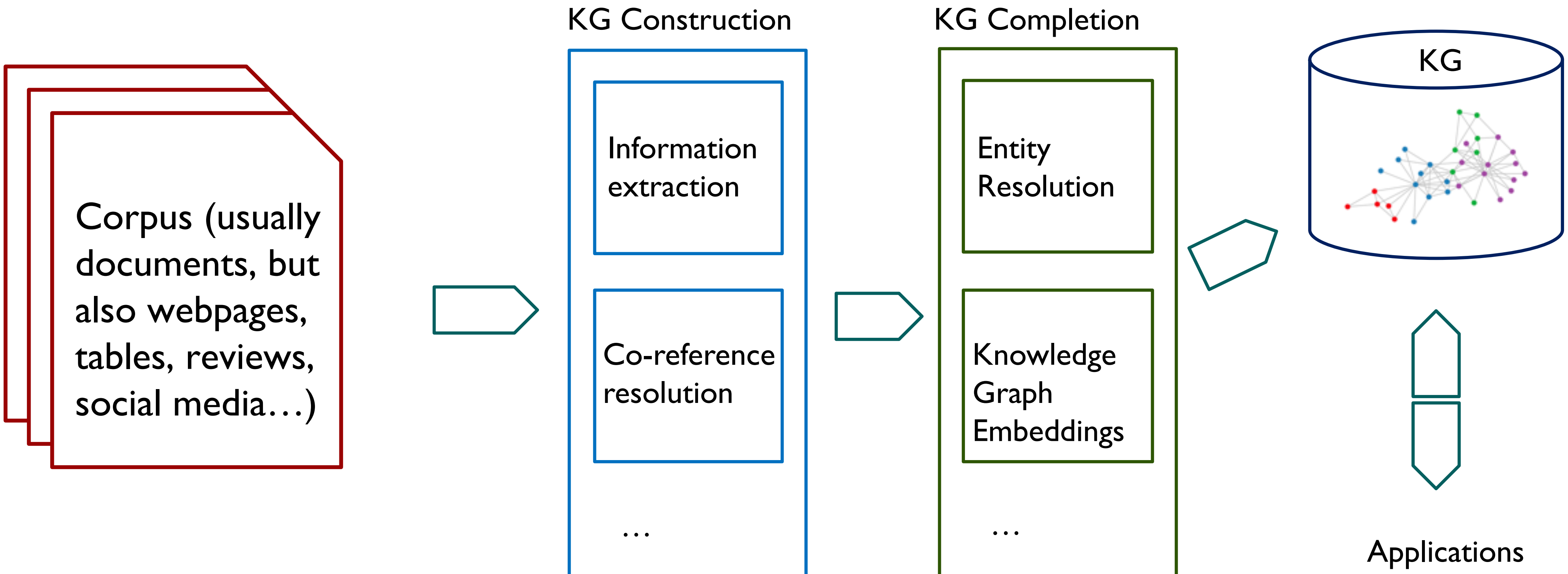


Sources:
https://aryamccarthy.github.io/wiseman2016learning/

(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

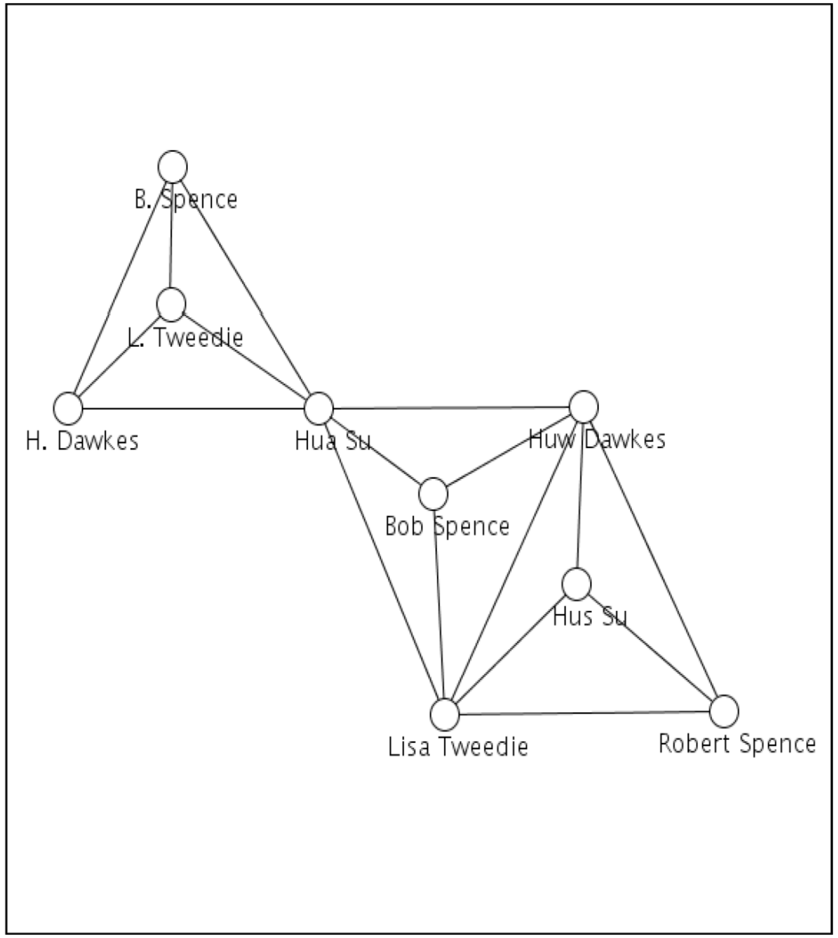


KNOWLEDGE GRAPH COMPLETION

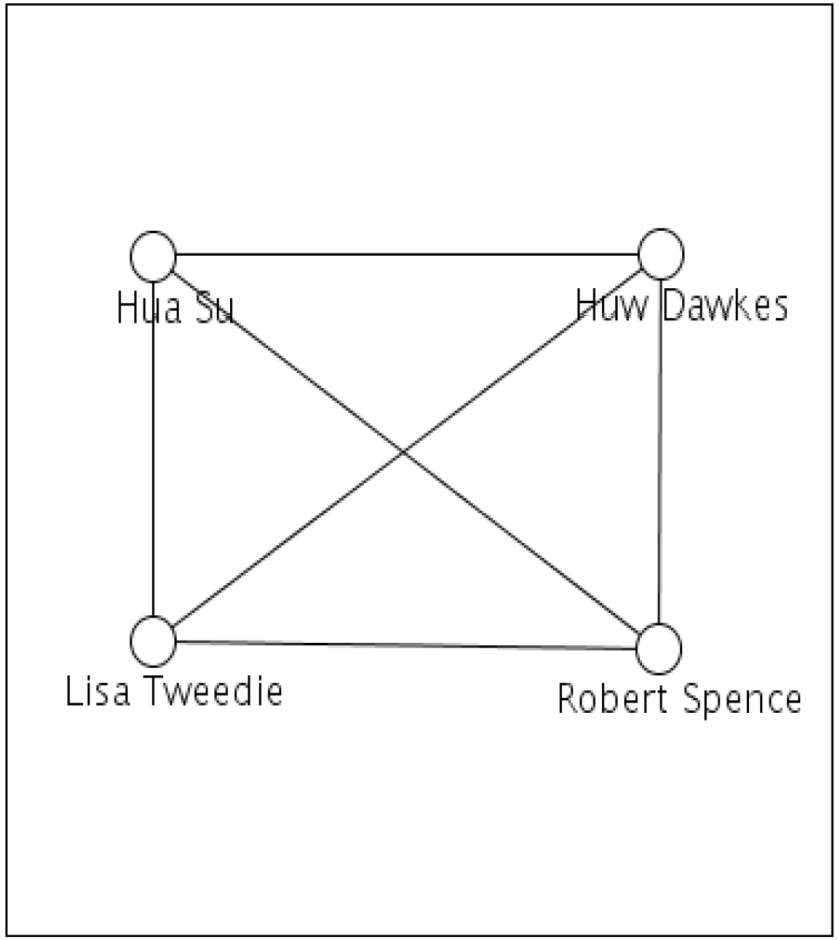
Entity Resolution

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…)



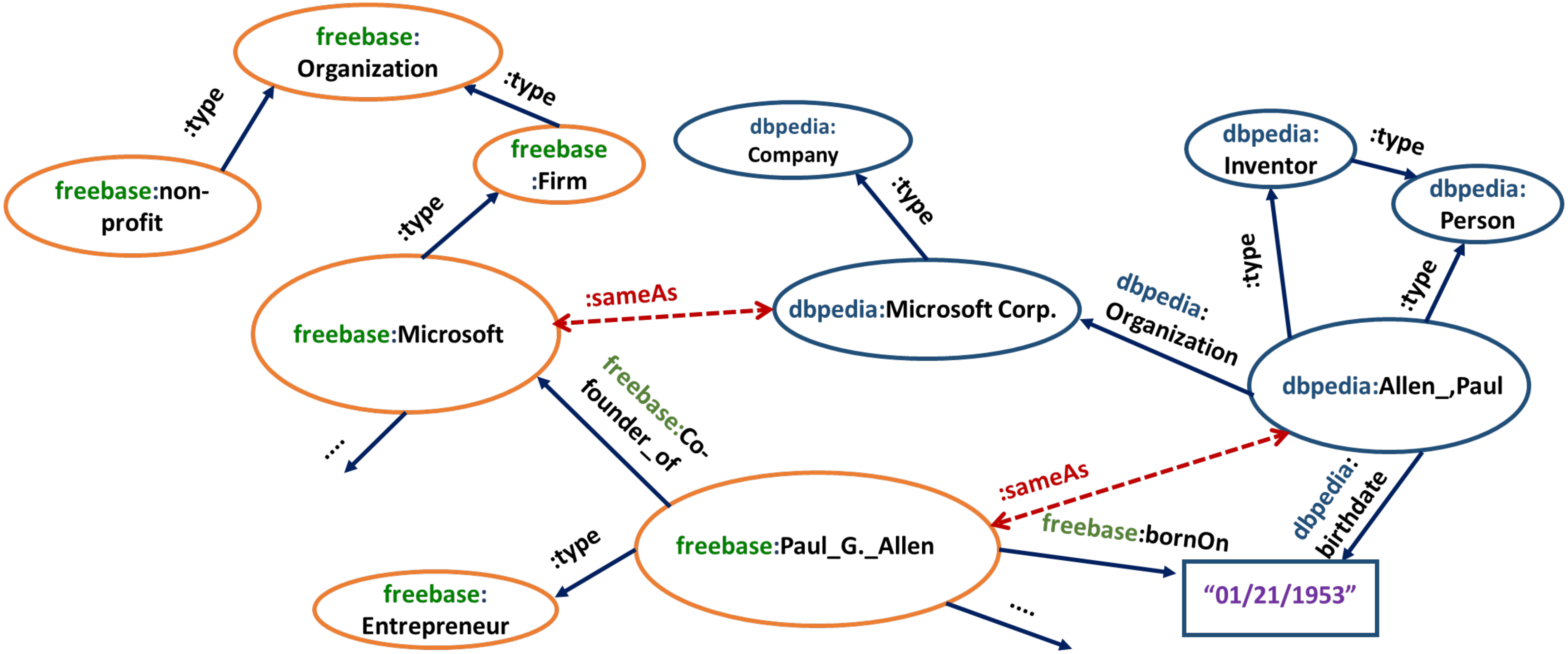
before



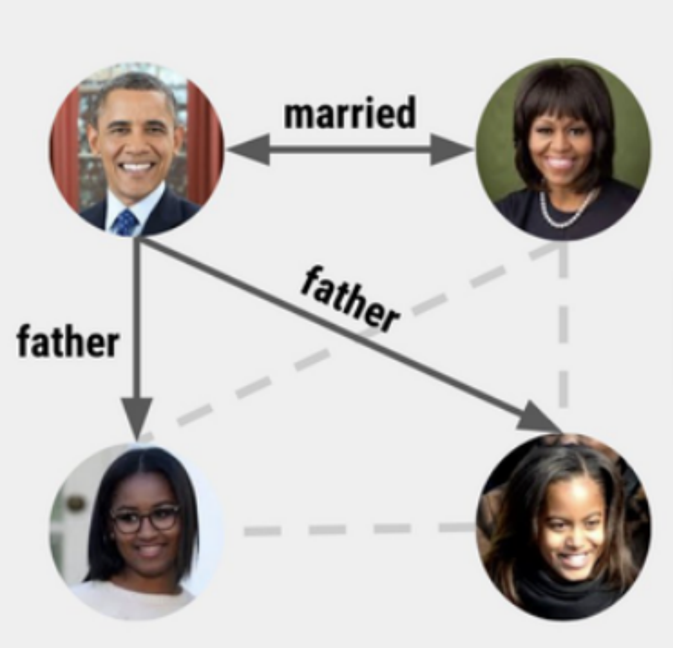
after

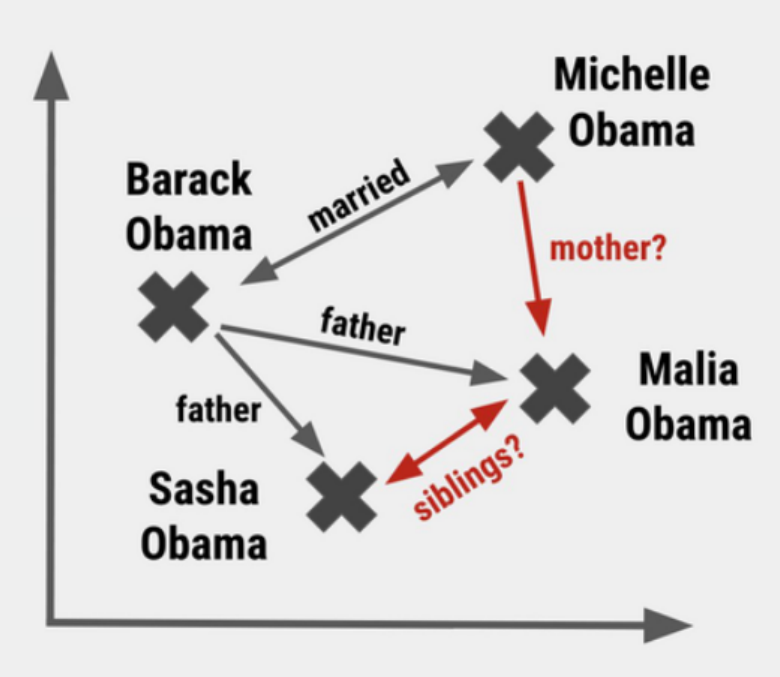
Source: Entity Resolution: Tutorial. Getoor and Machanavajjhala. VLDB, 2012

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**
HolE (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

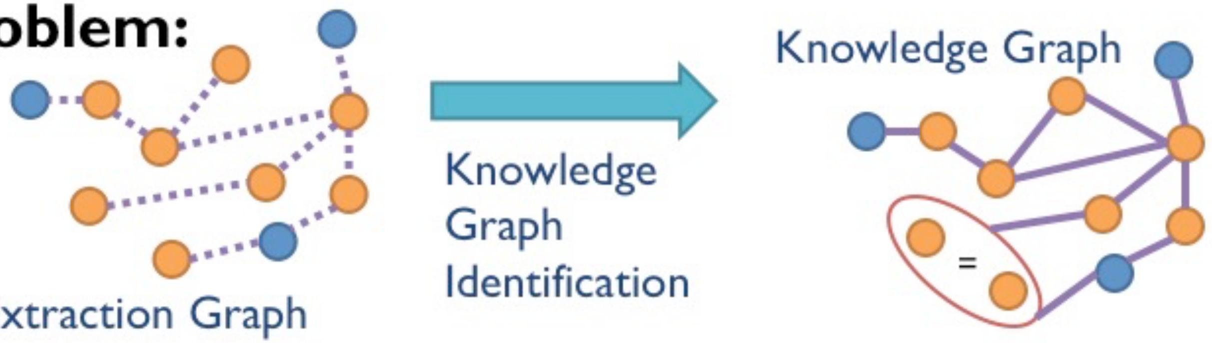
Examples of ontological constraints

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

(Pujara et al., ISWC13)

Knowledge Graph Identification

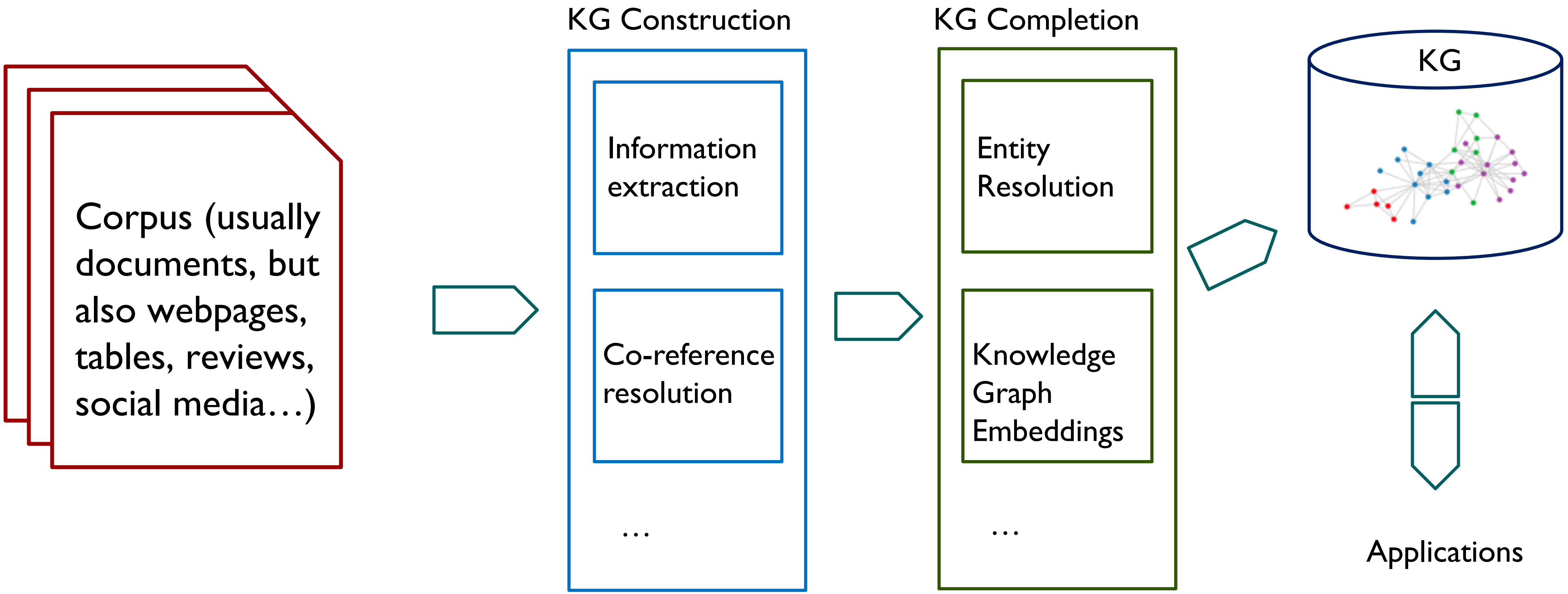
Problem:



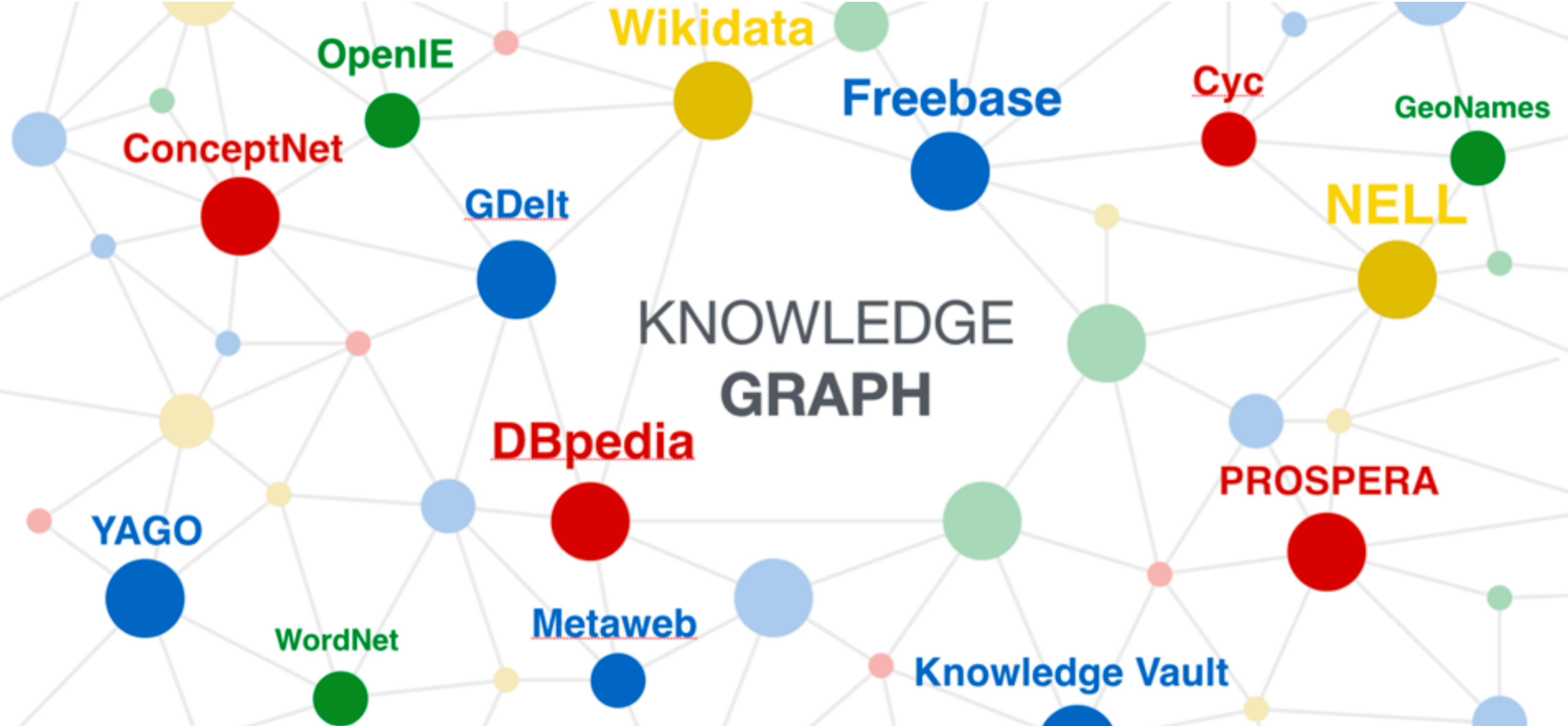
Solution: *Knowledge Graph Identification* (KGI)

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

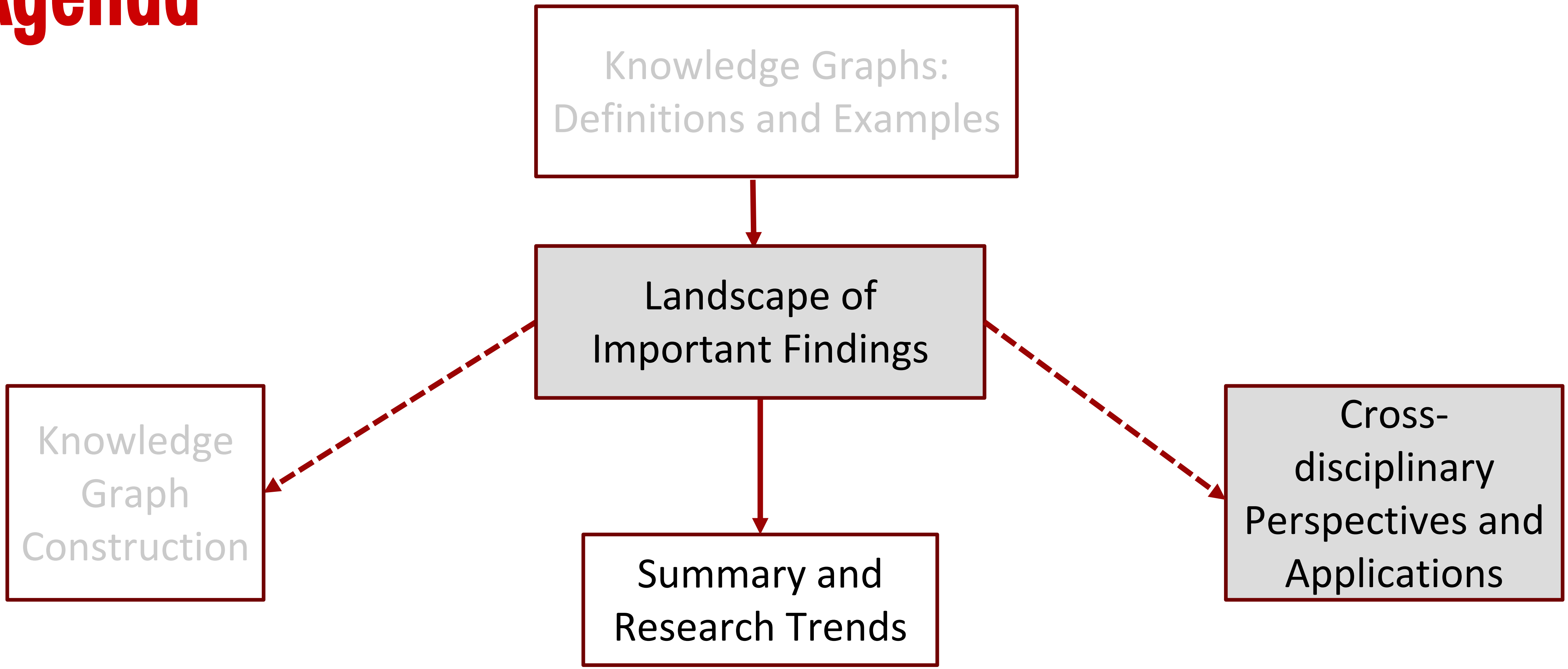
A typical KGC workflow starts from corpus acquisition and ends with applications



Open-source KGs that have been built

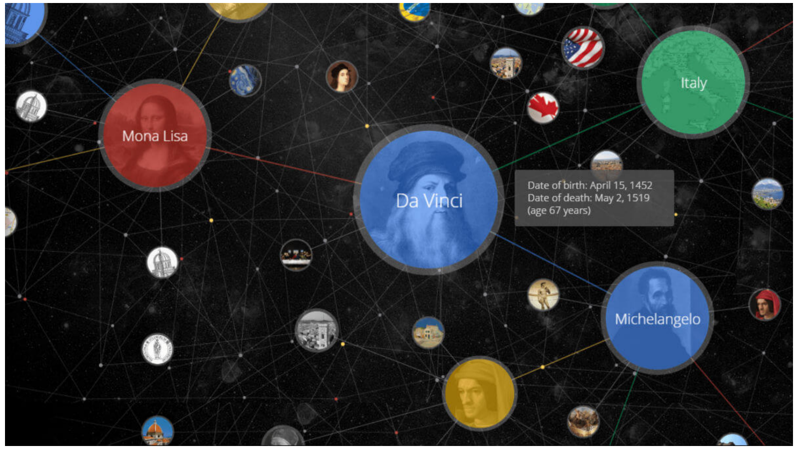


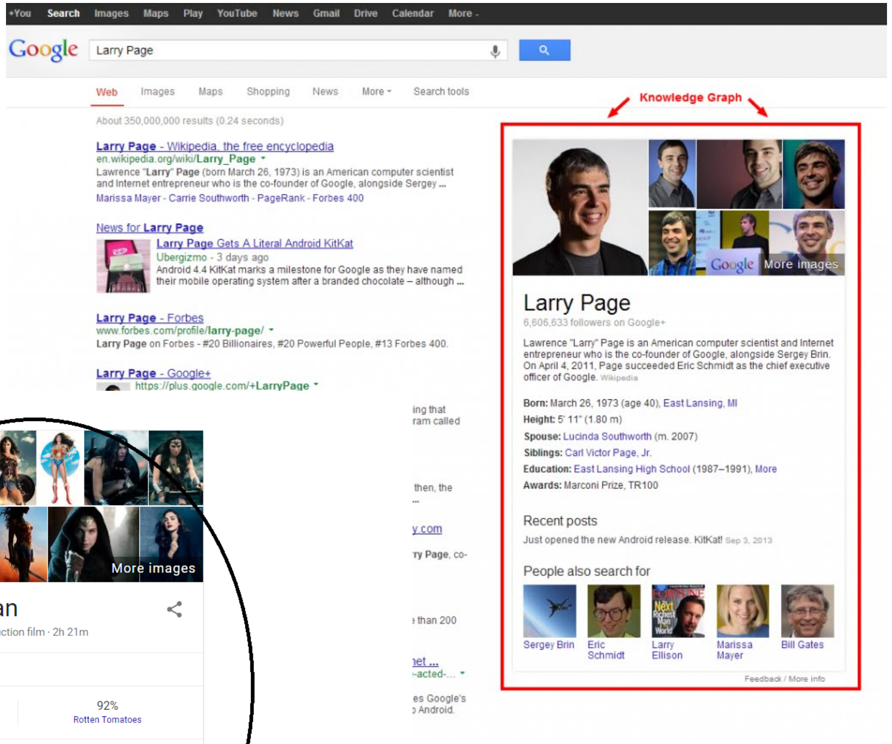
Agenda

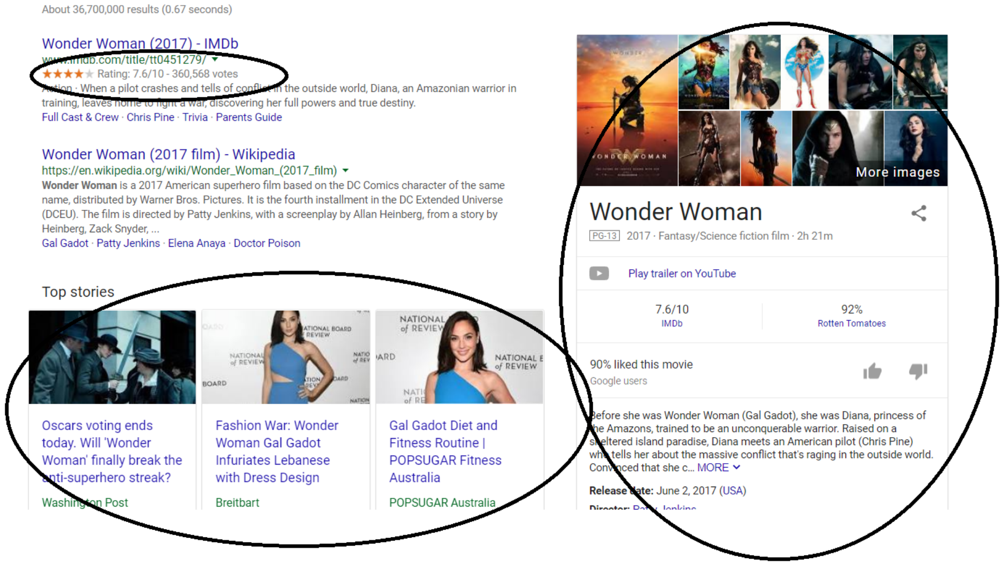

Knowledge Graphs: Definitions and Examples
Landscape of Important Findings
Knowledge Graph Construction
Summary and Research Trends
Cross-disciplinary Perspectives and Applications

CROSS-DISCIPLINARY PERSPECTIVES: WEB AND INFORMATION RETRIEVAL

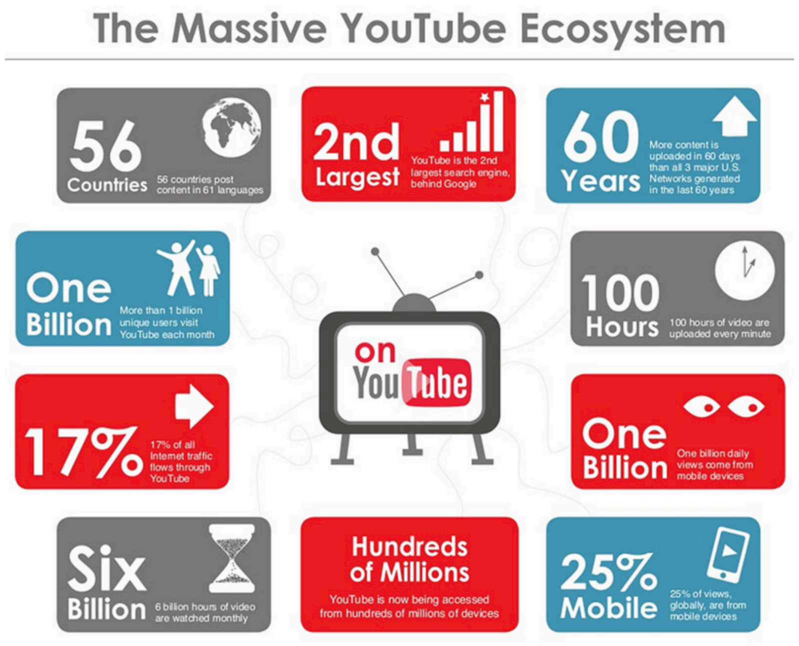
Google Knowledge Graph

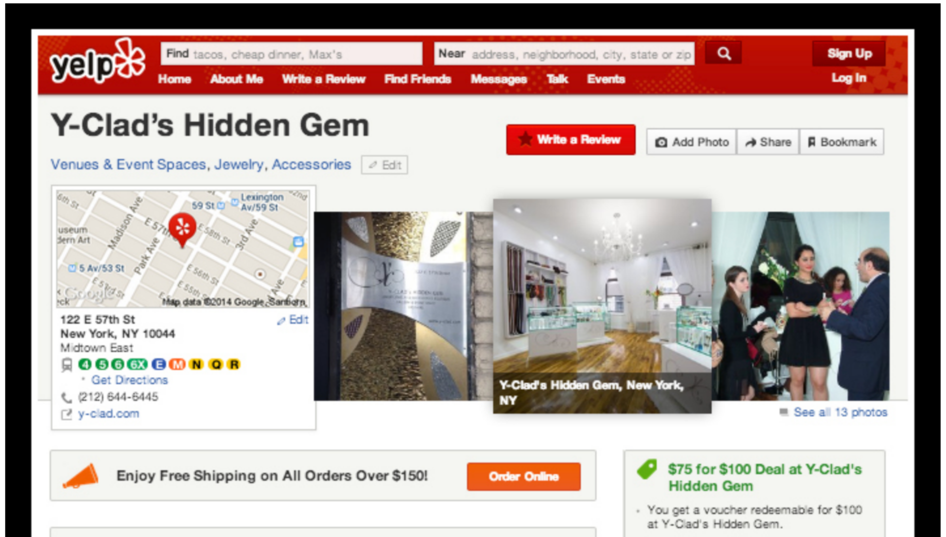






Domain-specific search (DSS)

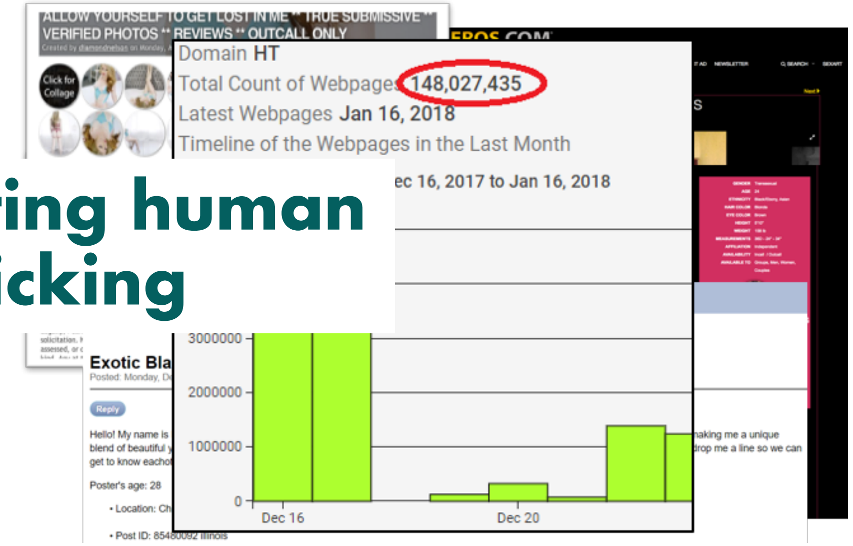




source: https://photos.prnewswire.com/prnfull/20151006/274273-INFO

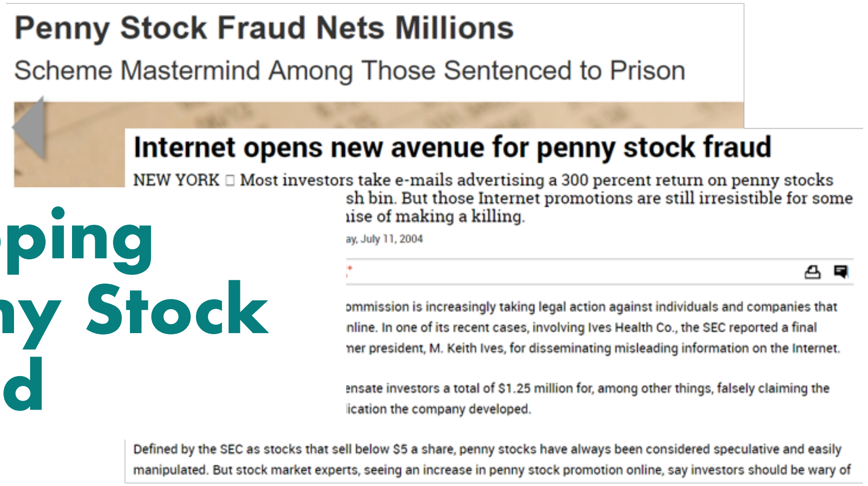
Emerging opportunities for DSS

Fighting human trafficking



Predicting cyberattacks

Stopping Penny Stock Fraud



Accurate geopolitical forecasting

DARPA/IARPA programs

DARPA Memex

IARPA Hybrid Forecasting Competition

DARPA AIDA

DARPA Causal Exploration

DARPA LORELEI

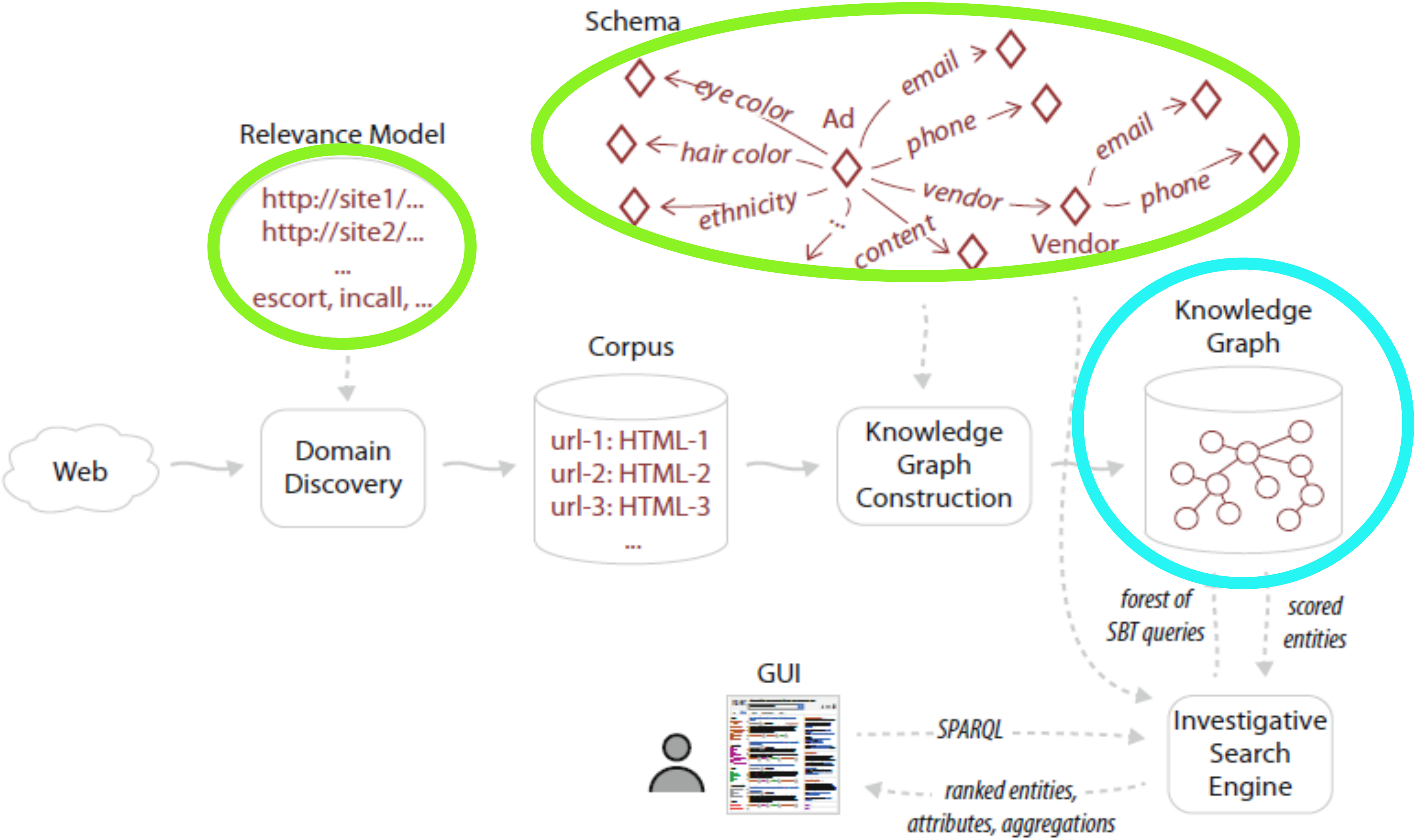
IARPA CAUSE

Research Question

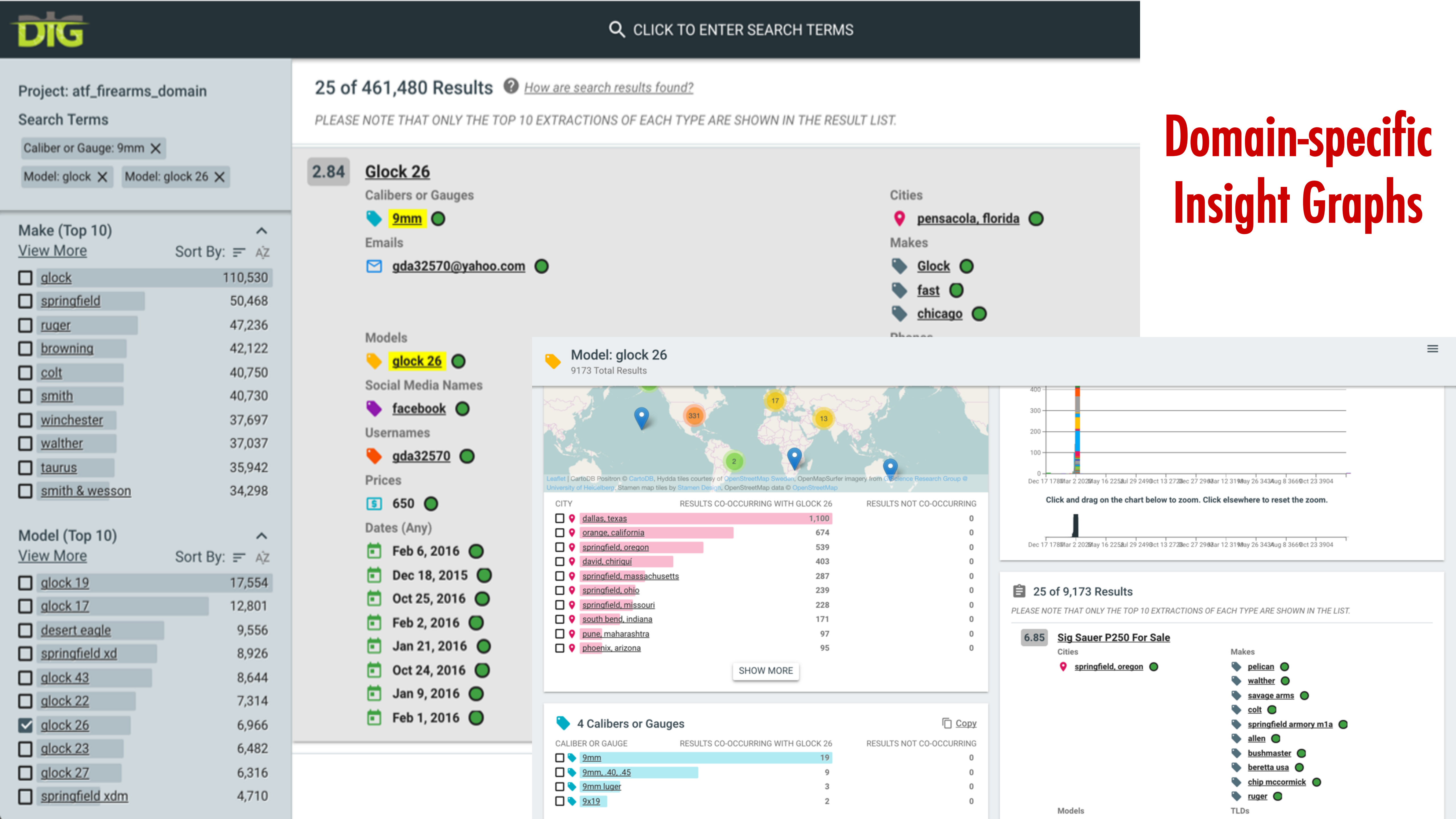
General Search	Google Knowledge Graph
DSS	Domain-Specific Knowledge Graphs

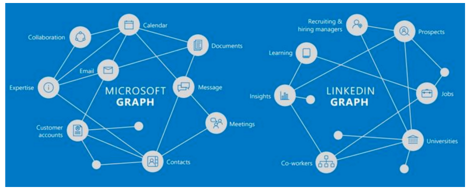
How do we construct domain specific knowledge graphs over web data for powerful DSS applications?

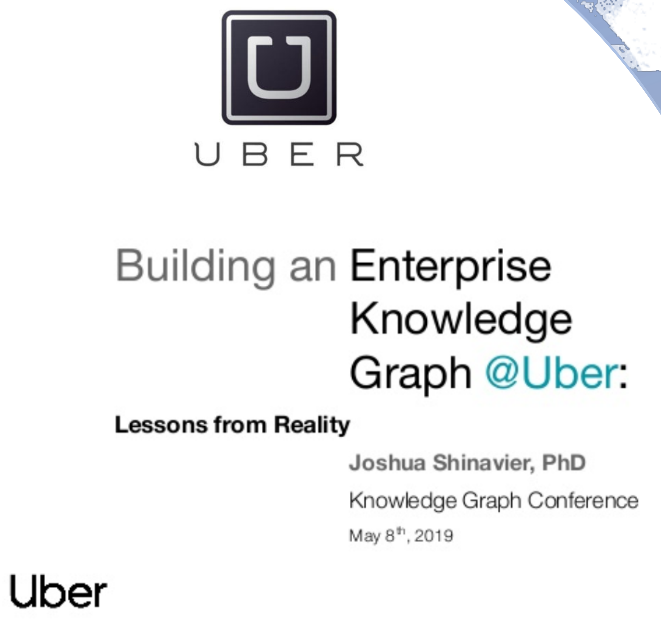
Knowledge Graphs for DSS



Domain-specific Insight Graphs



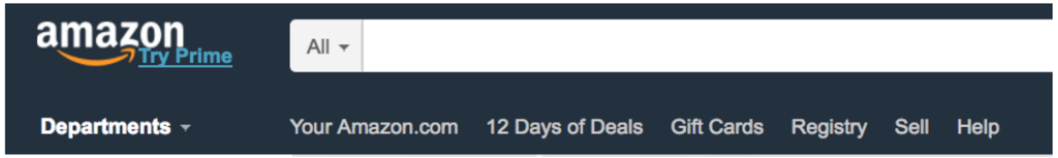




Many examples in industry and non-profit

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

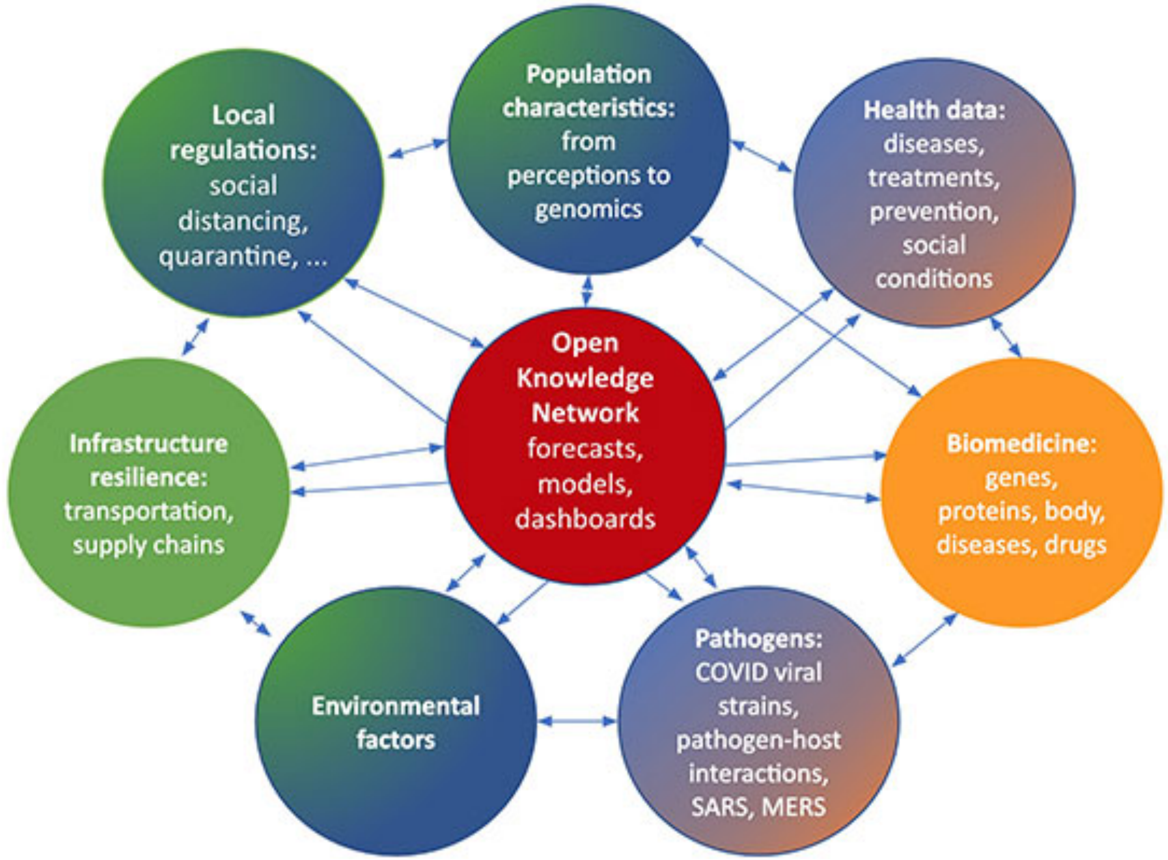
Another example: COVID-19

June 01, 2020 | By Jan Zverina

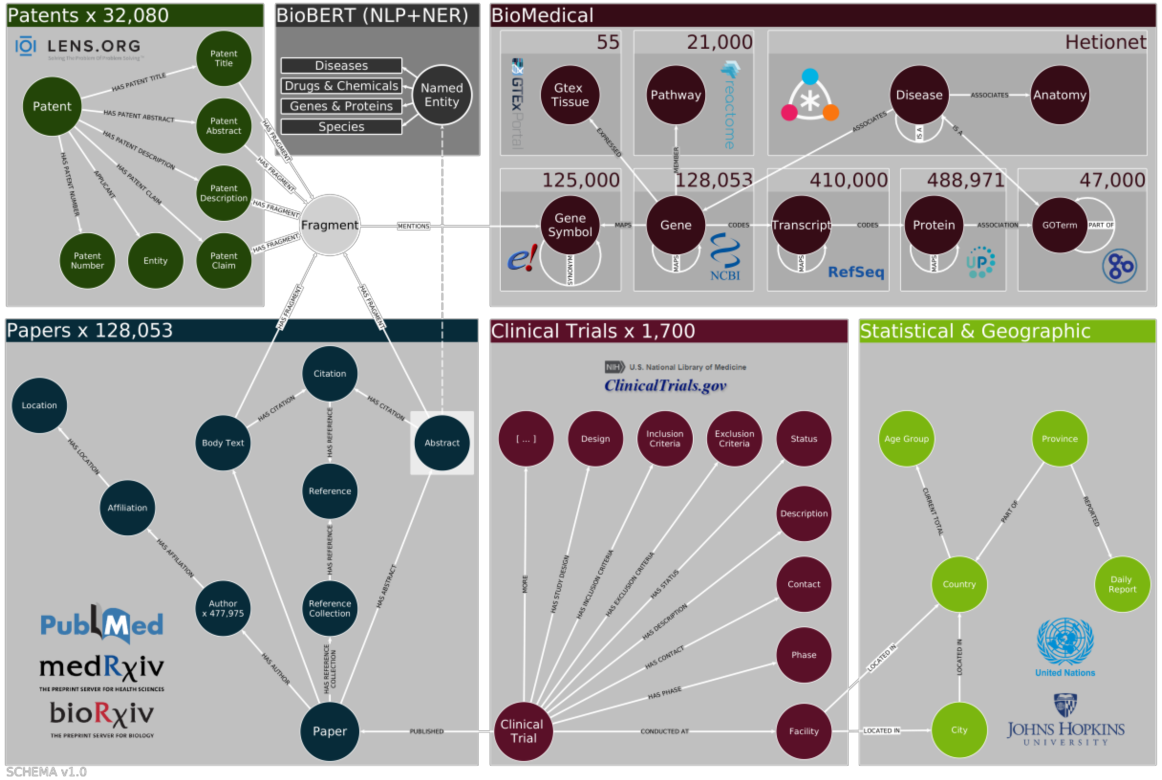
Fighting COVID-19 with Knowledge Graphs

National Science Foundation awards funding for a semantic integration platform

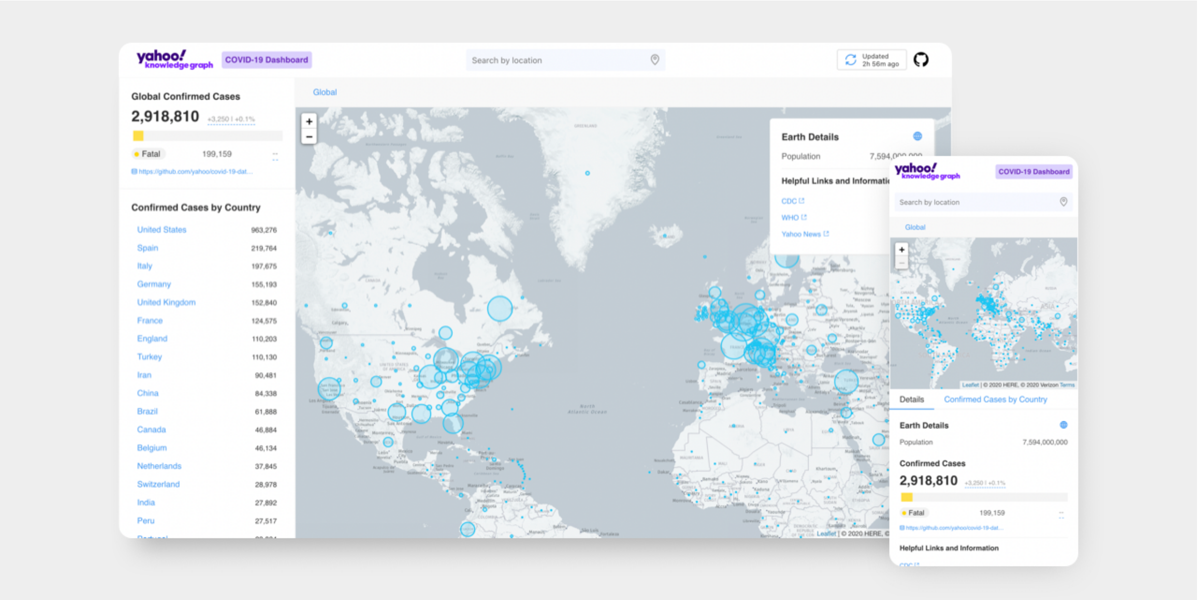
“The project will be based on our knowledge graph prototype linking information about pathogens, health data, and environmental indicators and enabling cross-domain inferencing,” said Peter Rose, director of SDSC’s Structural Bioinformatics Laboratory and principal investigator (PI) for the project, 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

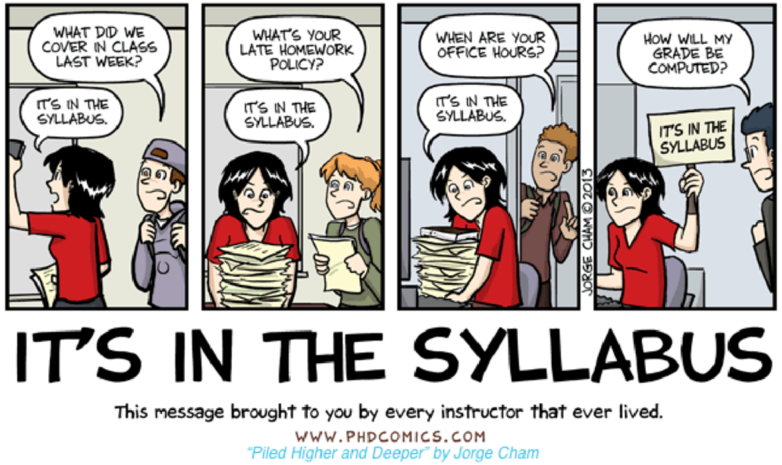


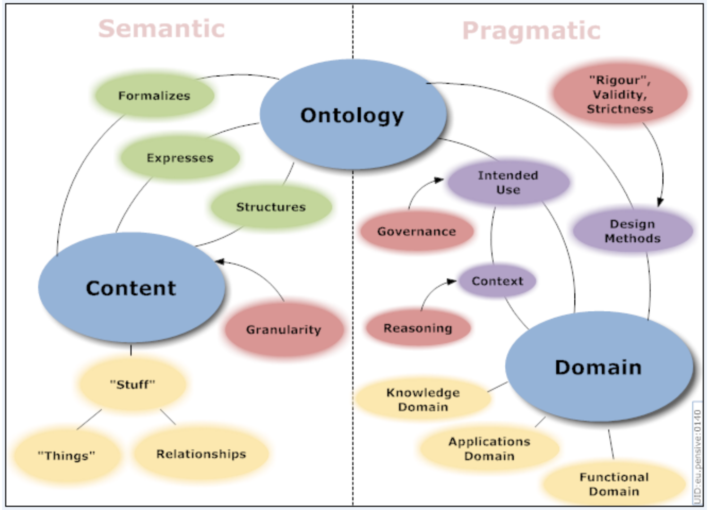
Source: Verizon Media
https://github.com/yahoo/covid-19-dashboard

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

CROSS-DISCIPLINARY PERSPECTIVES: SEMANTIC WEB

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





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)

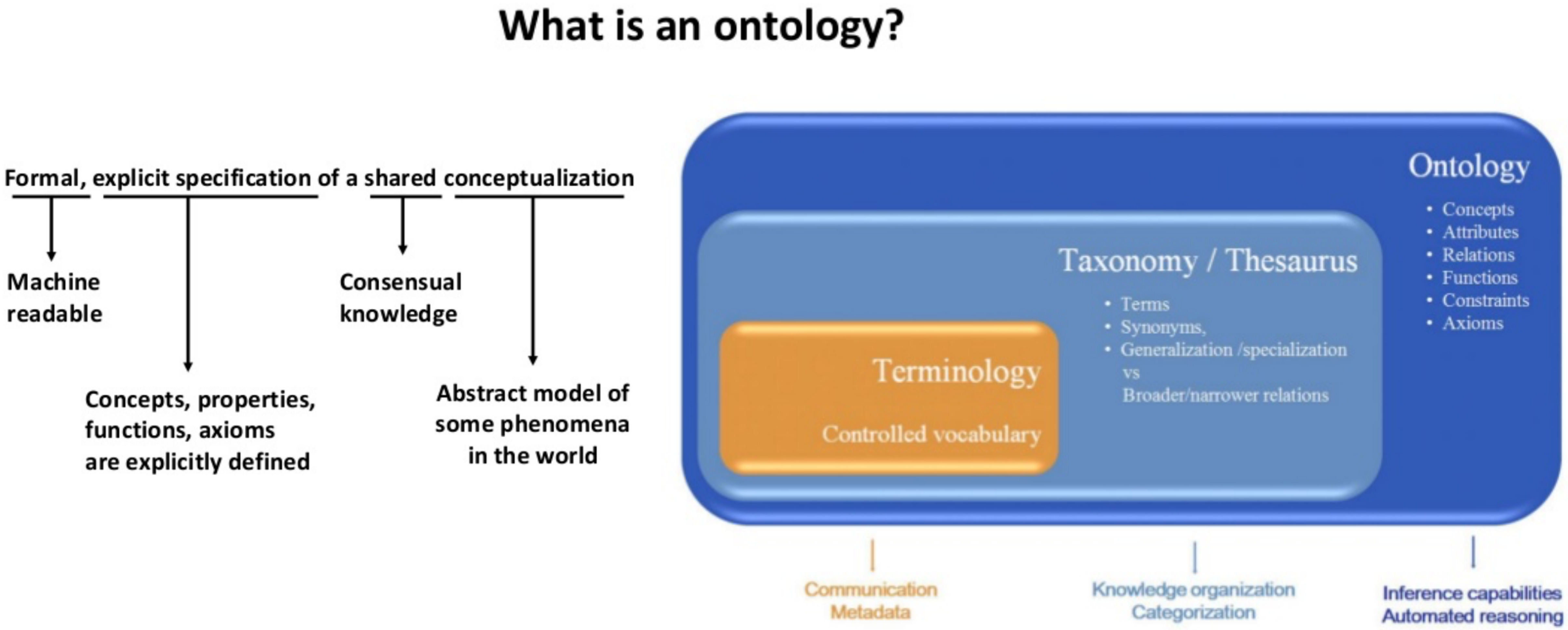
Examples

How do domain experts specify the sphere?

Examples

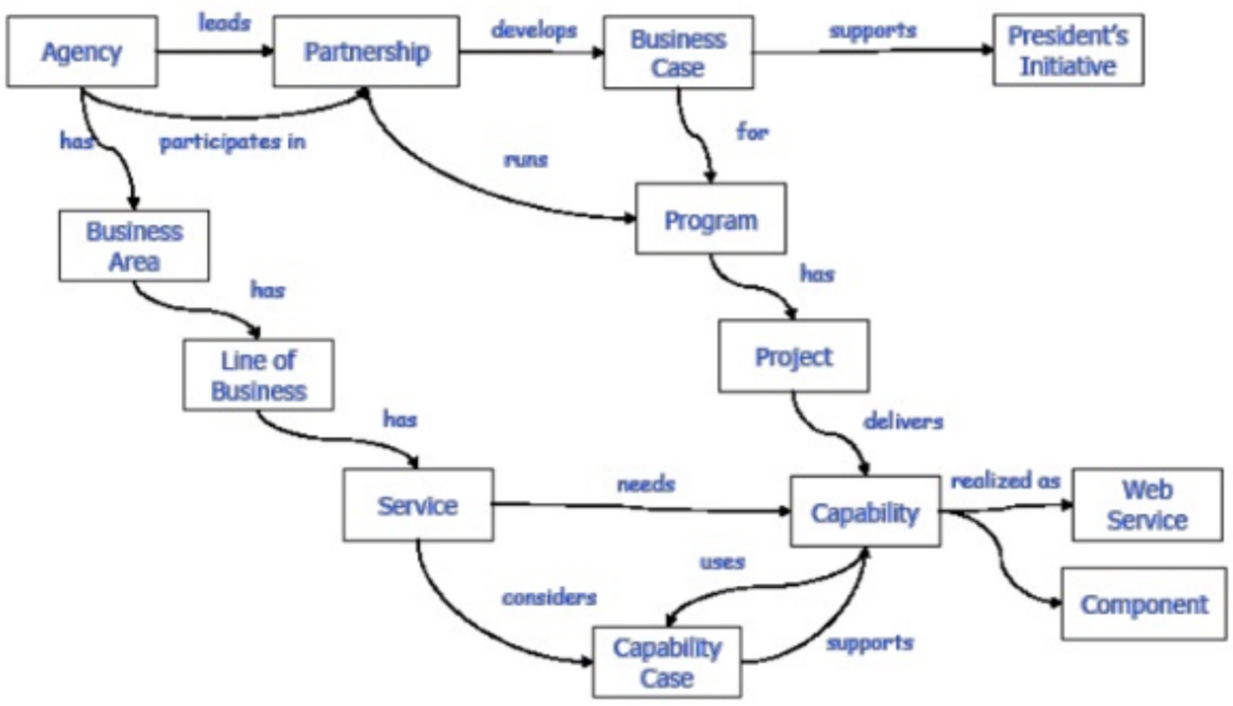
Ontology

Modeling domains: Ontologies

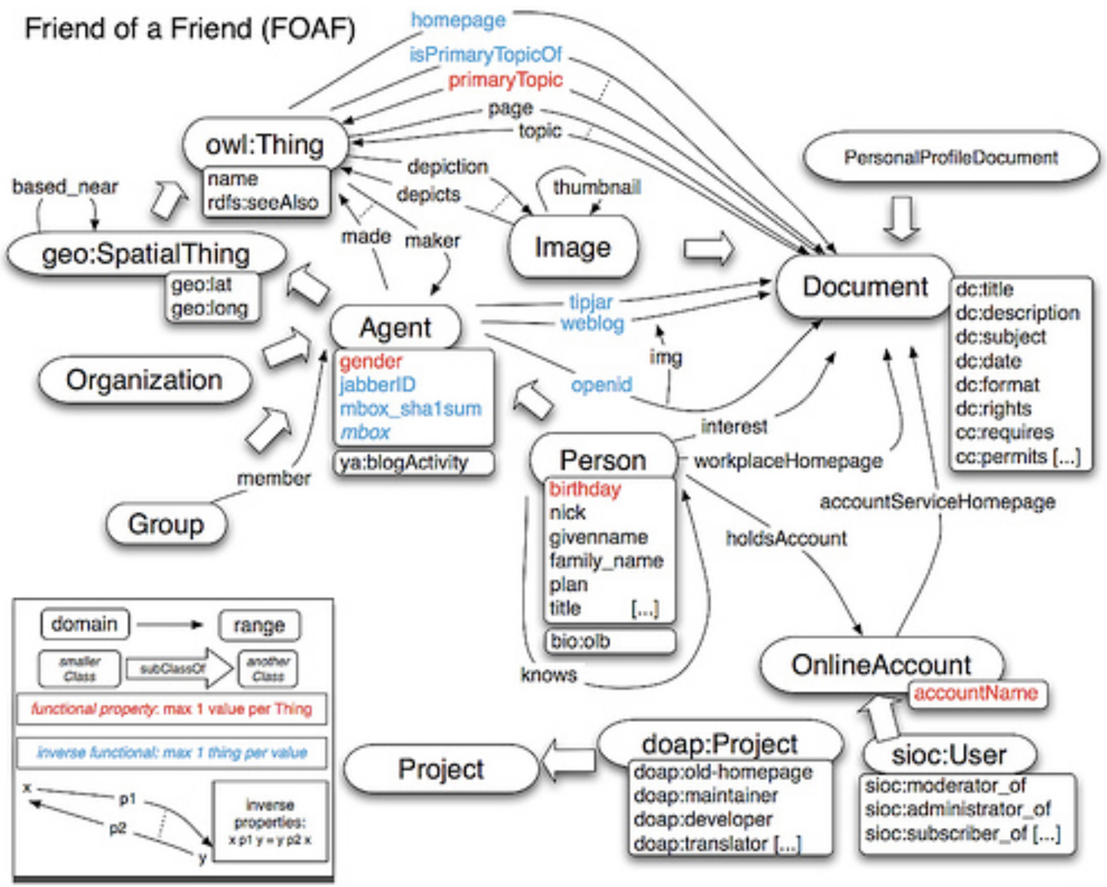


Source: "Ontologies and semantic web." Stanley Wang. https://www.slideshare.net/stanleywanguni/ontologies-and-semantic-web

Examples of ontologies

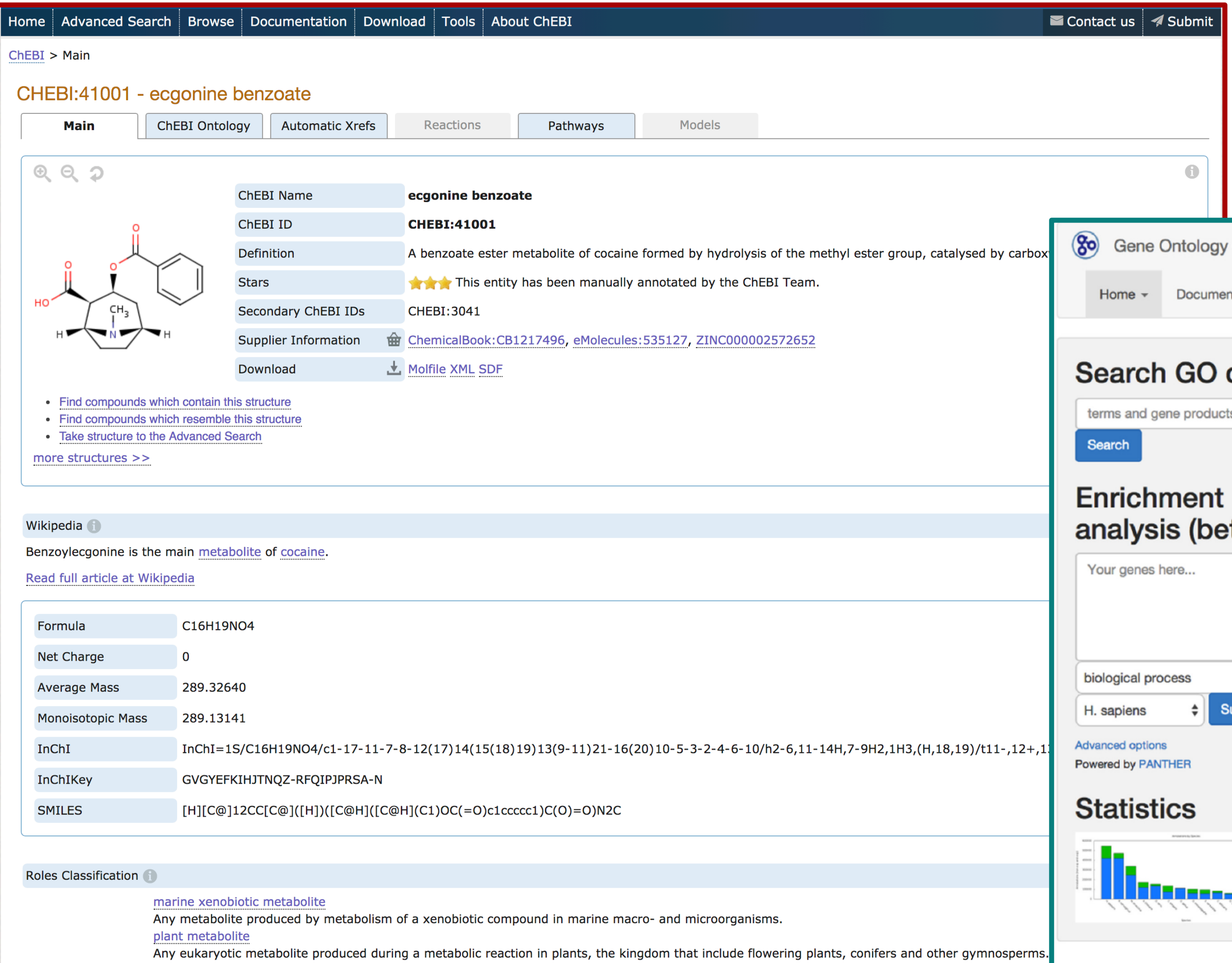


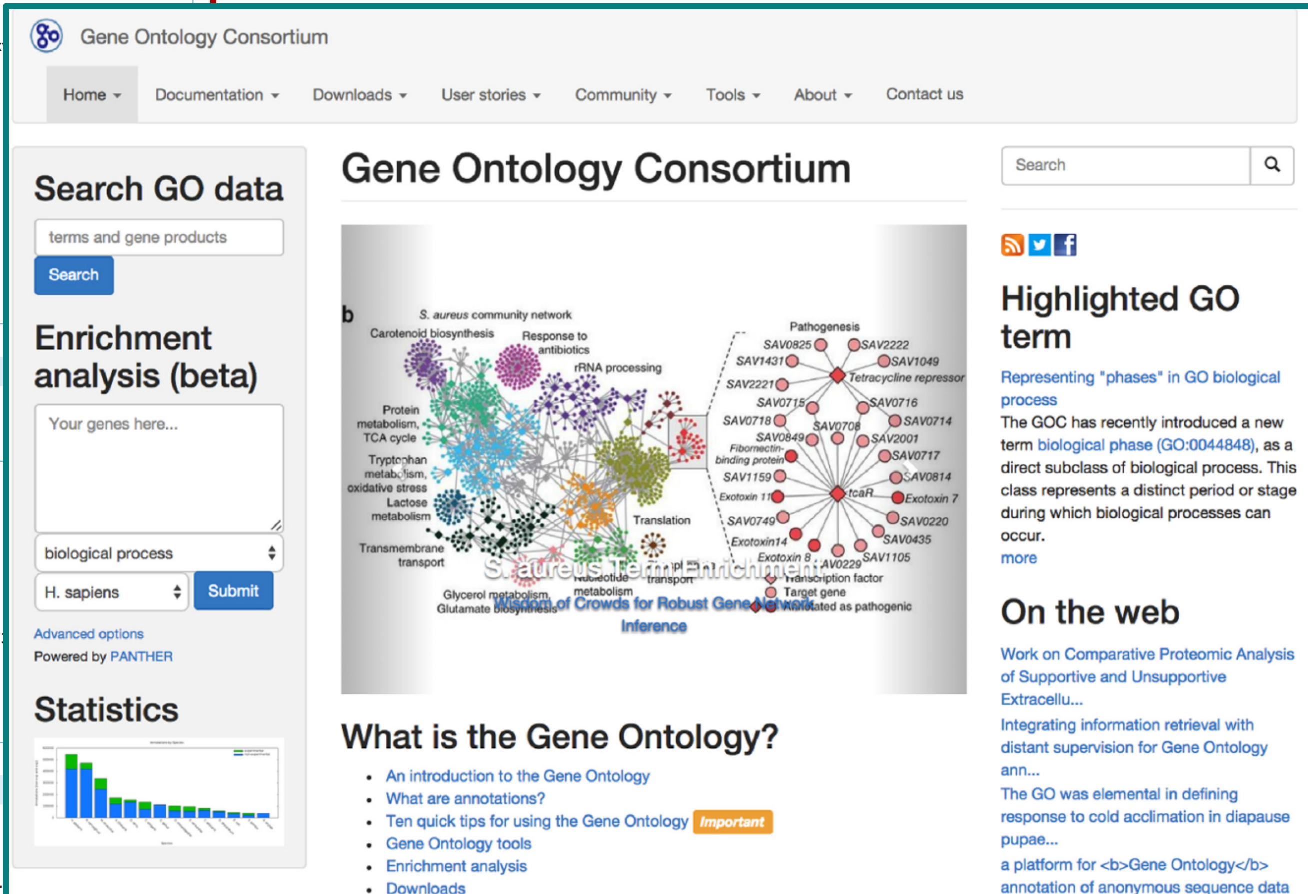
Agency domain



Friend-of-a-friend

Ontologies are big in Science



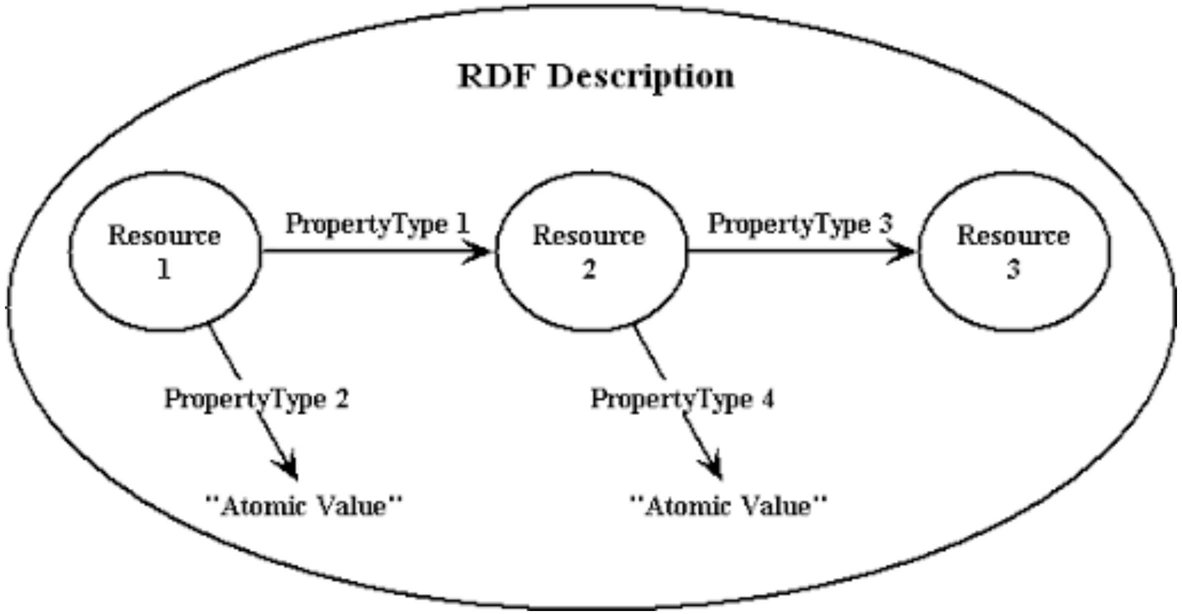


Representation of knowledge graphs (and ontologies)

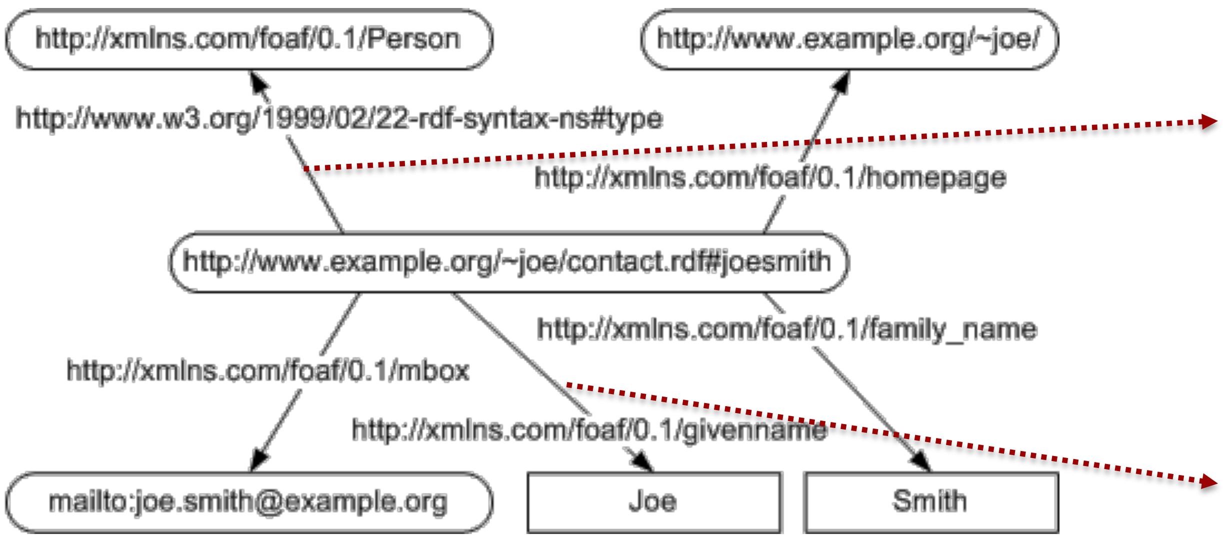
An RDF graph is a set of triples, where each triple is of the form (subject, predicate, object):

- 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 of 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



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 graph

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

(In)Equality:

- equivalentClass
- equivalentProperty
- sameAs
- differentFrom
- AllDifferent
- distinctMembers

Property Characteristics:

- ObjectProperty
- DatatypeProperty
- inverseOf
- TransitiveProperty
- SymmetricProperty
- FunctionalProperty
- InverseFunctionalProperty

Property Restrictions:

- Restriction
- onProperty
- allValuesFrom
- someValuesFrom

Restricted Cardinality:

- minCardinality (only 0 or 1)
- maxCardinality (only 0 or 1)
- cardinality (only 0 or 1)

Header Information:

- Ontology
- imports

Class Intersection:

- intersectionOf

Datatypes

- xsd datatypes

Versioning:

- versionInfo
- priorVersion
- backwardCompatibleWith
- incompatibleWith
- DeprecatedClass
- DeprecatedProperty

Annotation Properties:

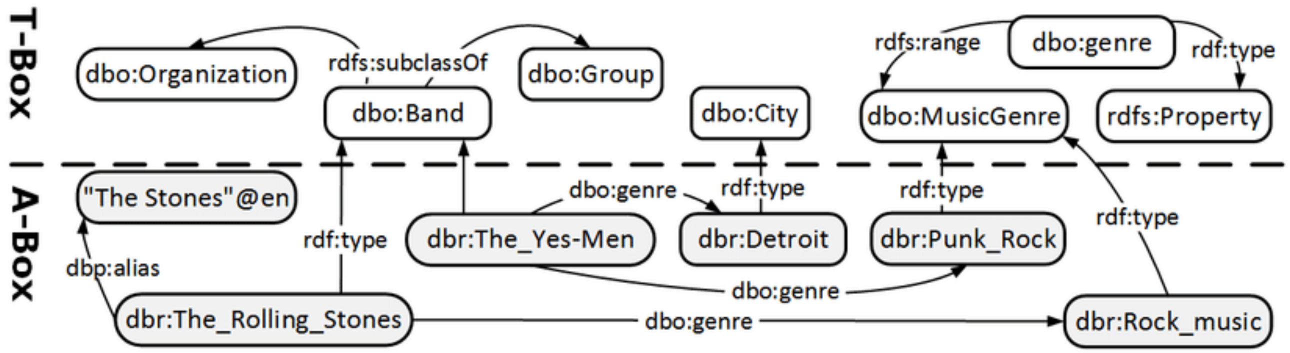
- rdfs:label
- rdfs:comment
- rdfs:seeAlso
- rdfs:isDefinedBy
- AnnotationProperty
- OntologyProperty

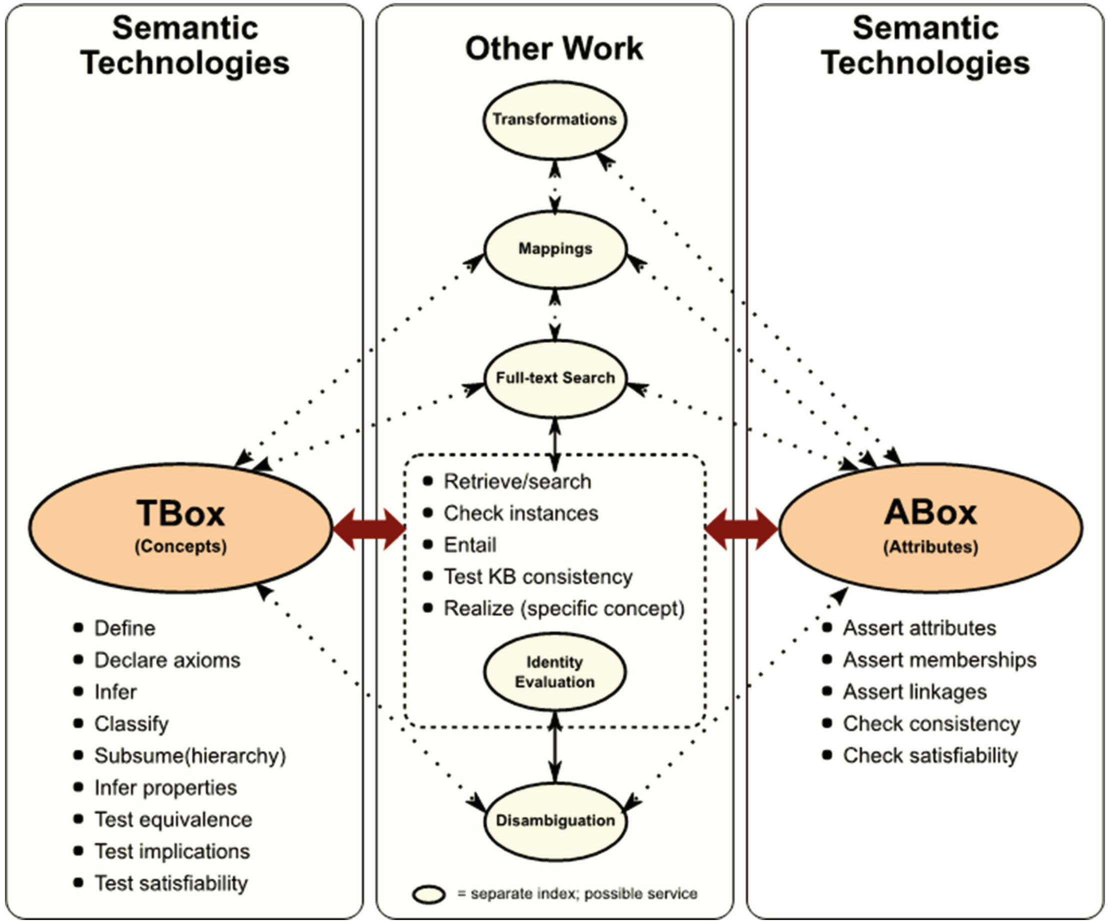
https://www.w3.org/TR/owl-features/

Reasoning over knowledge graphs

Assertions	DL-axioms
Every animal which can fly has wings	φ_1: $Fly \sqsubseteq HasWing$
Every animal which eats fish is a piscivore	φ_2: $\exists\ Eat.Fish \sqsubseteq Piscivore$
tweety is not an abnormal bird or cannot fly	φ_3: $(\neg AbnBird \sqcup \neg Fly)(tweety)$
ursidae eats *salmon*	φ_4: $Eat(ursidae, salmon)$
salmon is some fish	φ_5: $Fish(salmon)$
ursidae is not a piscivore	φ_6: $\neg Piscivore(ursidae)$

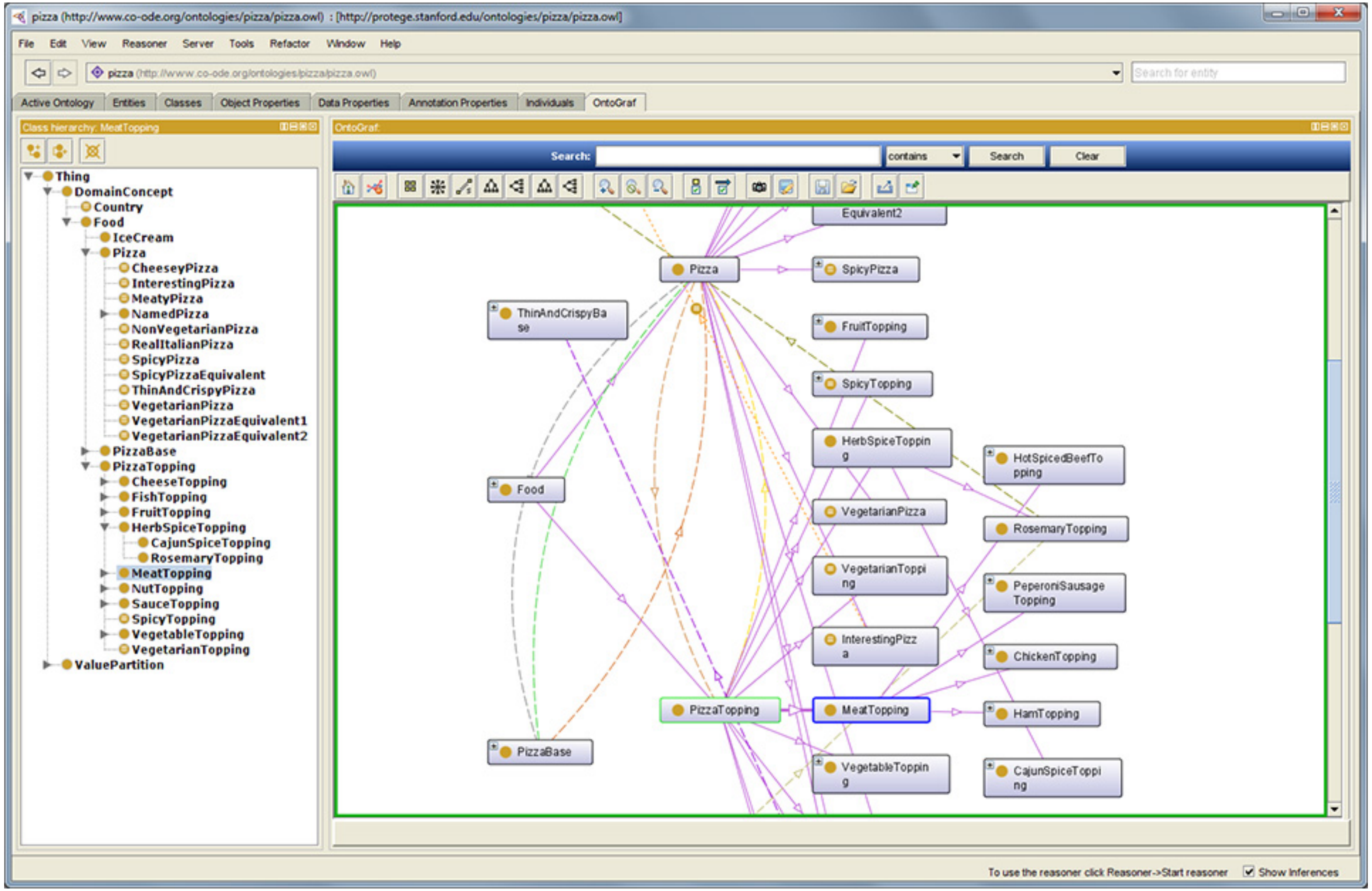
https://doi.org/10.1371/journal.pone.0181056.t001





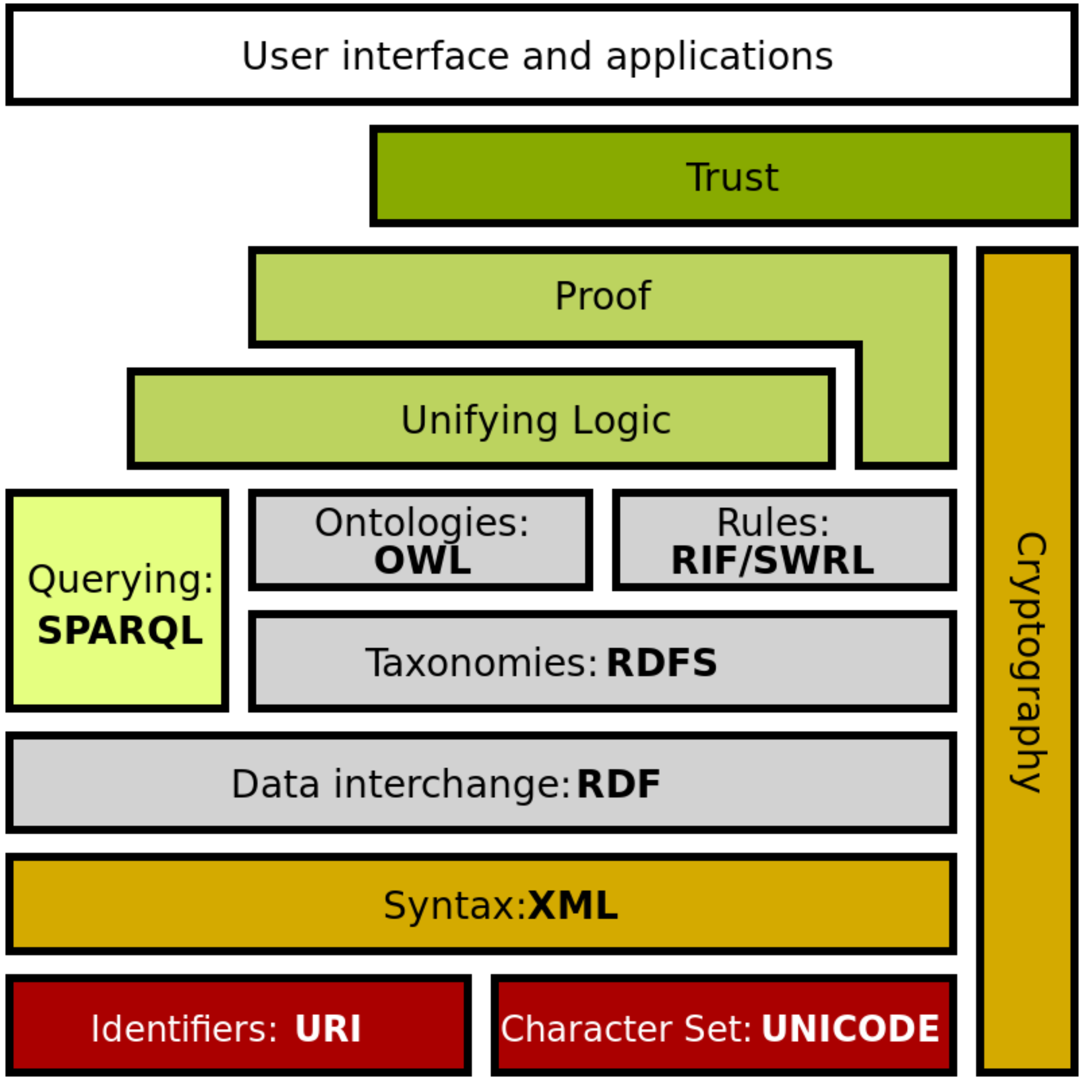
Source: Bergman. Platforms and Knowledge Management. 2018

Example tool for reasoning and ontologies: Protege



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

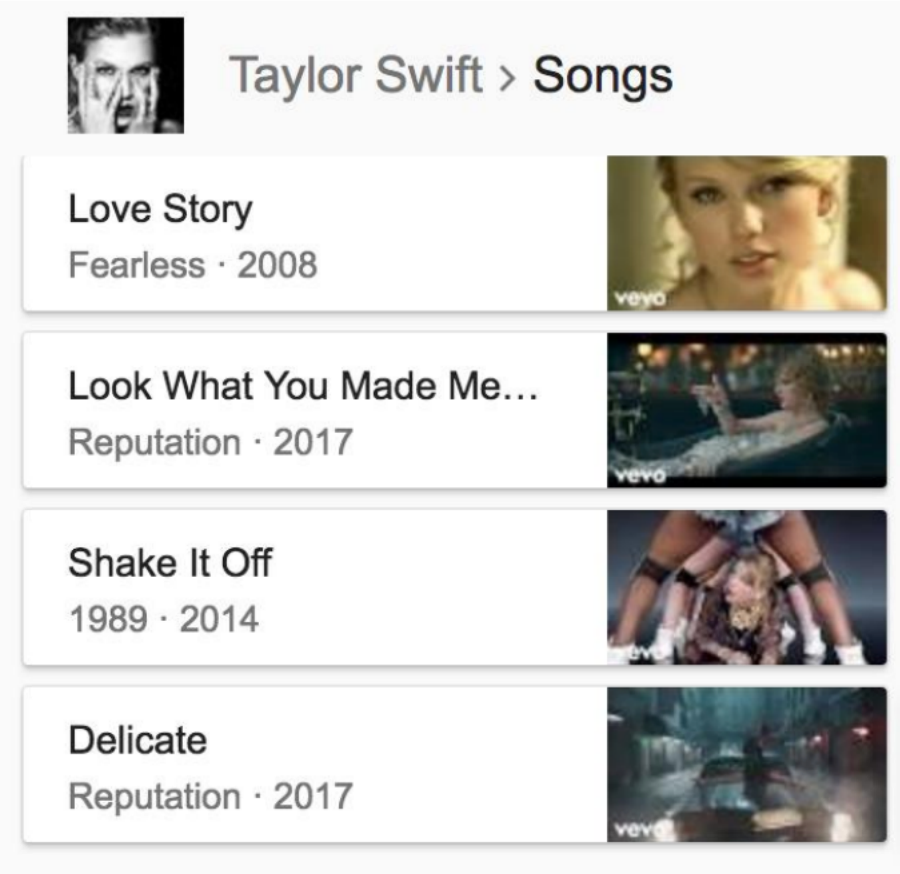
Putting it all together: Semantic Web Layer Cake



CROSS-DISCIPLINARY PERSPECTIVES: KNOWLEDGE DISCOVERY & DATA MINING

Knowledge Graph in Personal Assistant

Alexa, play Taylor Swift in the past year



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.

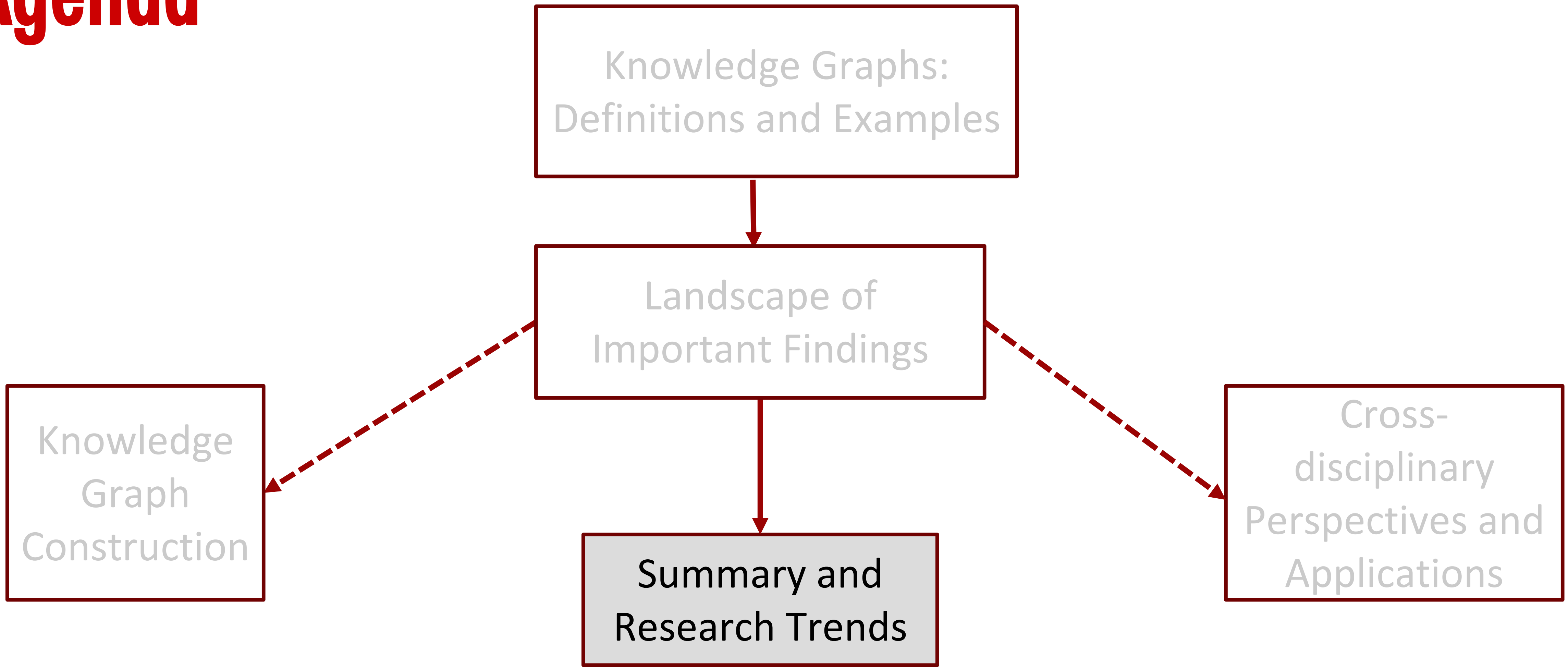
Summarization

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

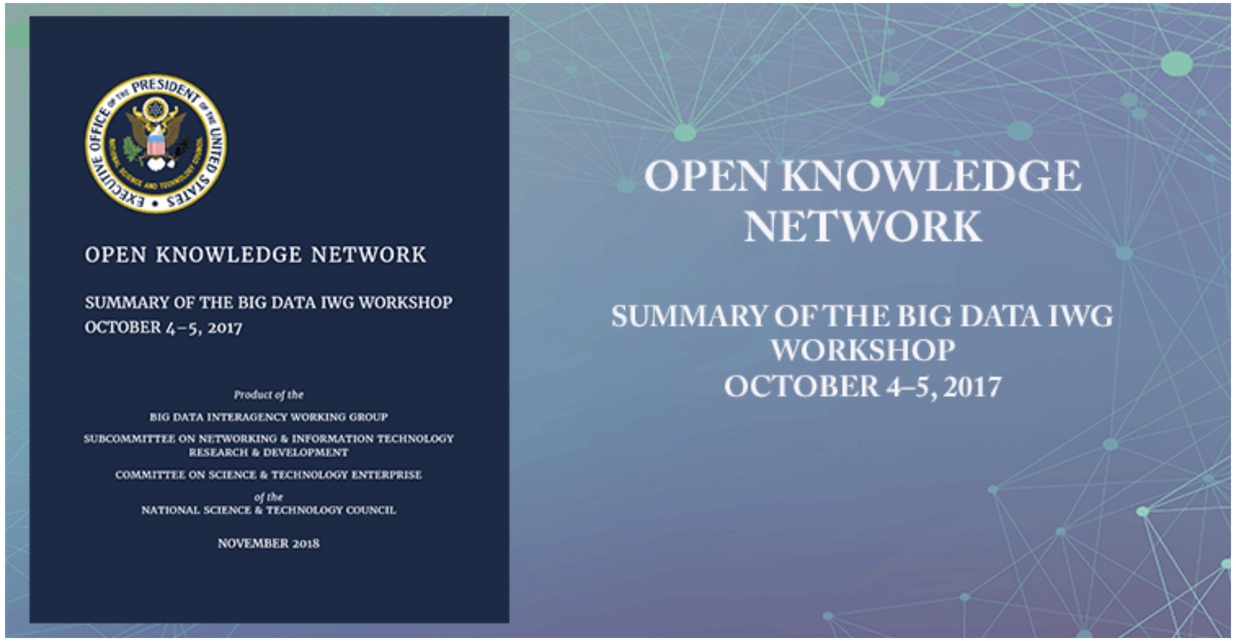
Shiralkar, P., Flammini, A., Menczer, F., & Ciampaglia, G. L. (2017, November). Finding streams in knowledge graphs to support fact checking. In *2017 IEEE International Conference on Data Mining (ICDM)* (pp. 859-864). IEEE.

Agenda


Knowledge Graphs: Definitions and Examples
Landscape of Important Findings
Knowledge Graph Construction
Summary and Research Trends
Cross-disciplinary Perspectives and Applications

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

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

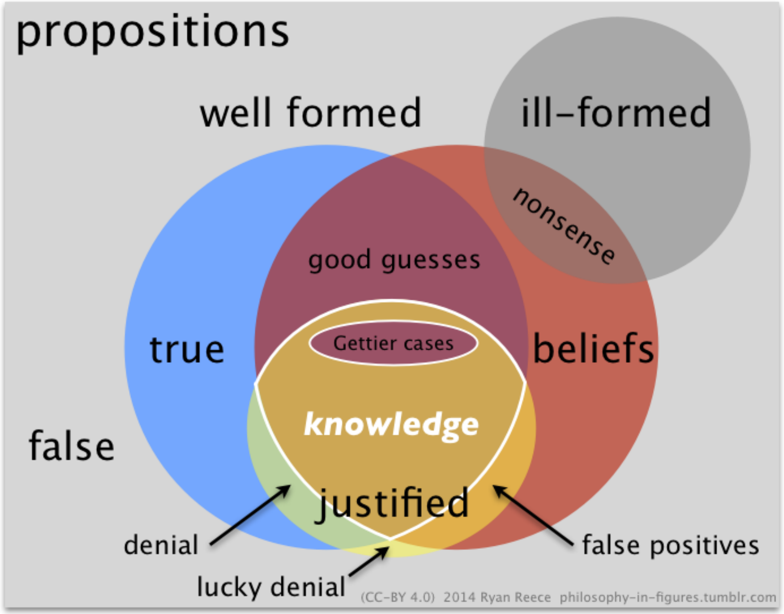
PDF

Mayank Kejriwal ; Pedro Szekely All Authors

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.
[1]Log base 10 scale. Deployment frequency capped at once per hour per year to prevent skewing; capping affected only a small number of use cases.
[2]Excluding sentiment analysis, speech to text, language understanding, and translation.

McKinsey&Company | Source: McKinsey Global Institute analysis

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 sl
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.

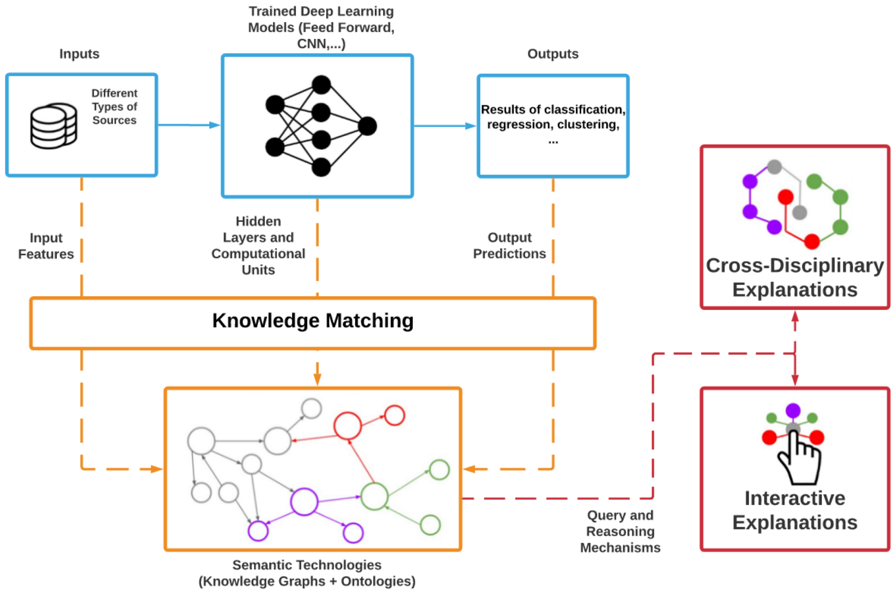
Quarks and Leptons are the building blocks which build up matter, i.e., they are seen as the "elementary particles". In the present standard model, there are six "flavors" of quarks. They can successfully account for all known mesons and baryons (over 200). The most familiar baryons are the proton and 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

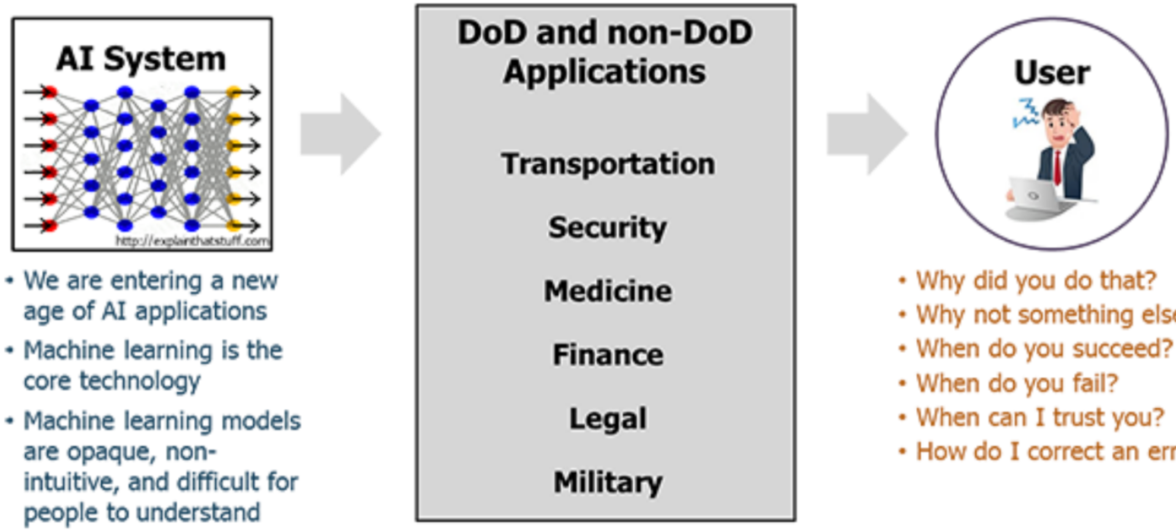
On The Role of Knowledge Graphs in Explainable AI

Freddy Lecue [a,b]
[a] CortAIx, Thales, Montreal, Canada
E-mail: freddy.lecue@inria.fr
[b] WIMMICS, Inria, Sophia Antipolis, France
E-mail: freddy.lecue@thalesgroup.fr



Explainable Artificial Intelligence (XAI)

Dr. Matt Turek



- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

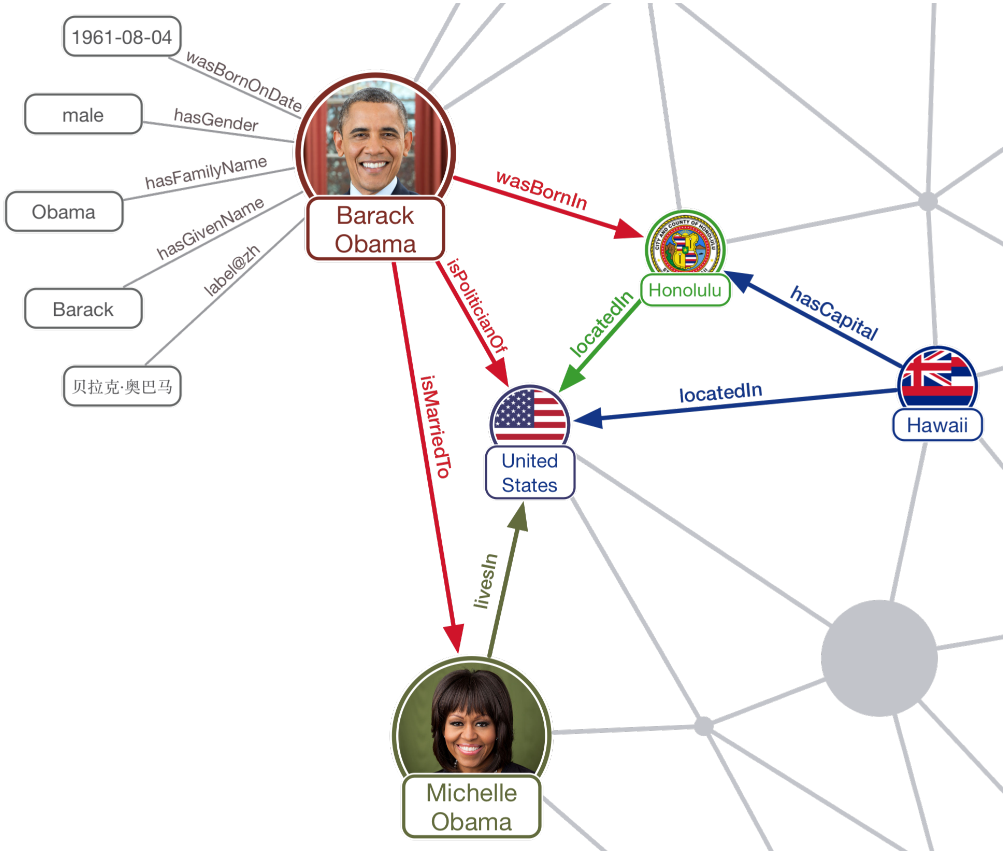
Source: Knowledge Graphs For eXplainable AI. On the Integration of Semantic Technologies and Symbolic Systems into Deep Learning Models for a More Comprehensible Artificial Intelligence.
https://towardsdatascience.com/knowledge-graphs-for-explainable-ai-dcd73c5c016

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

(Barack Obama, wasBornOnDate, 1961-08-04),
(Barack Obama, hasGender, male),
…
(Hawaii, hasCapital, Honolulu),
…
(Michelle Obama, livesIn, United States)



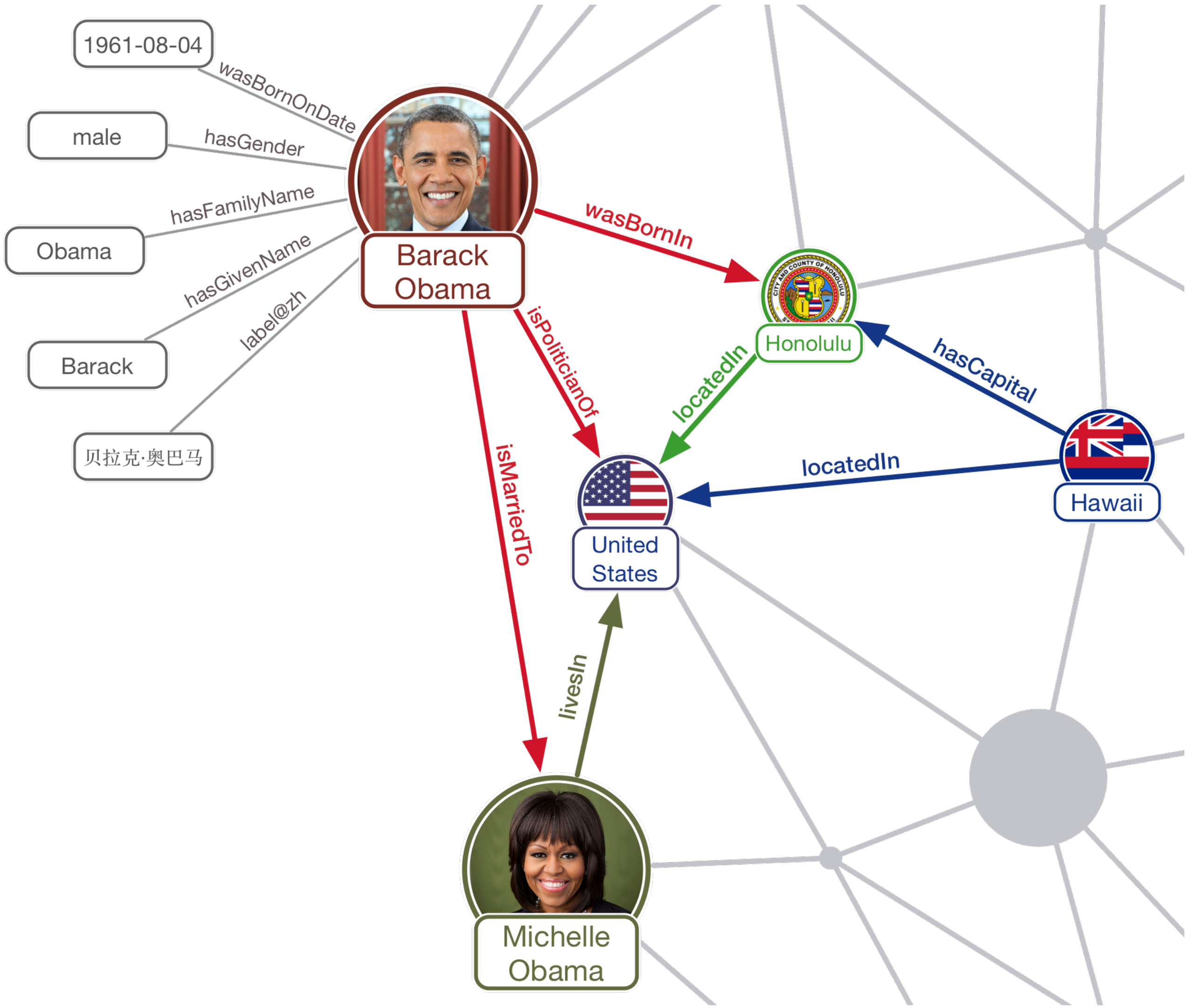
What is a Knowledge Graph?

Technically, a multi-relational directed labeled graph <u>with semantics</u>

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

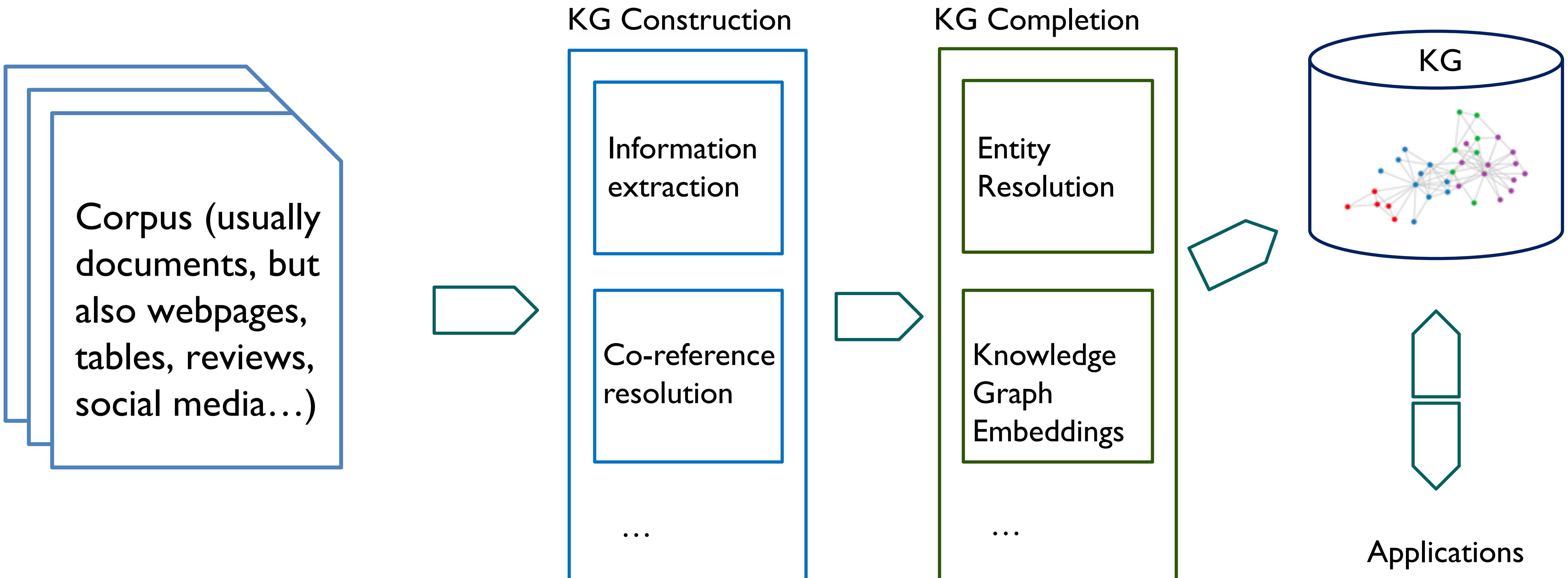
Traditionally, semantics are believed to come from ontology

- 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

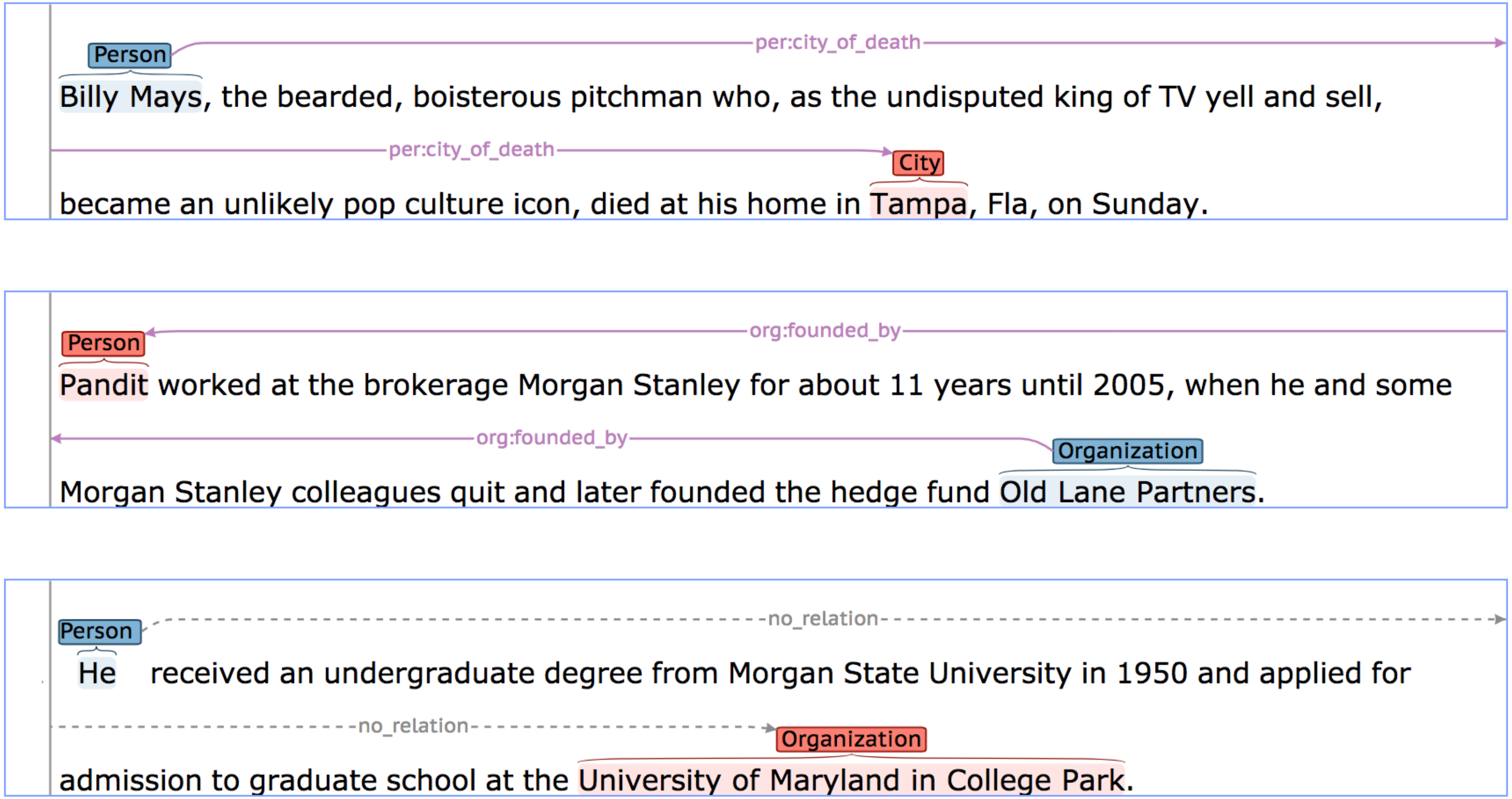


Named Entity Recognition (NER)

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

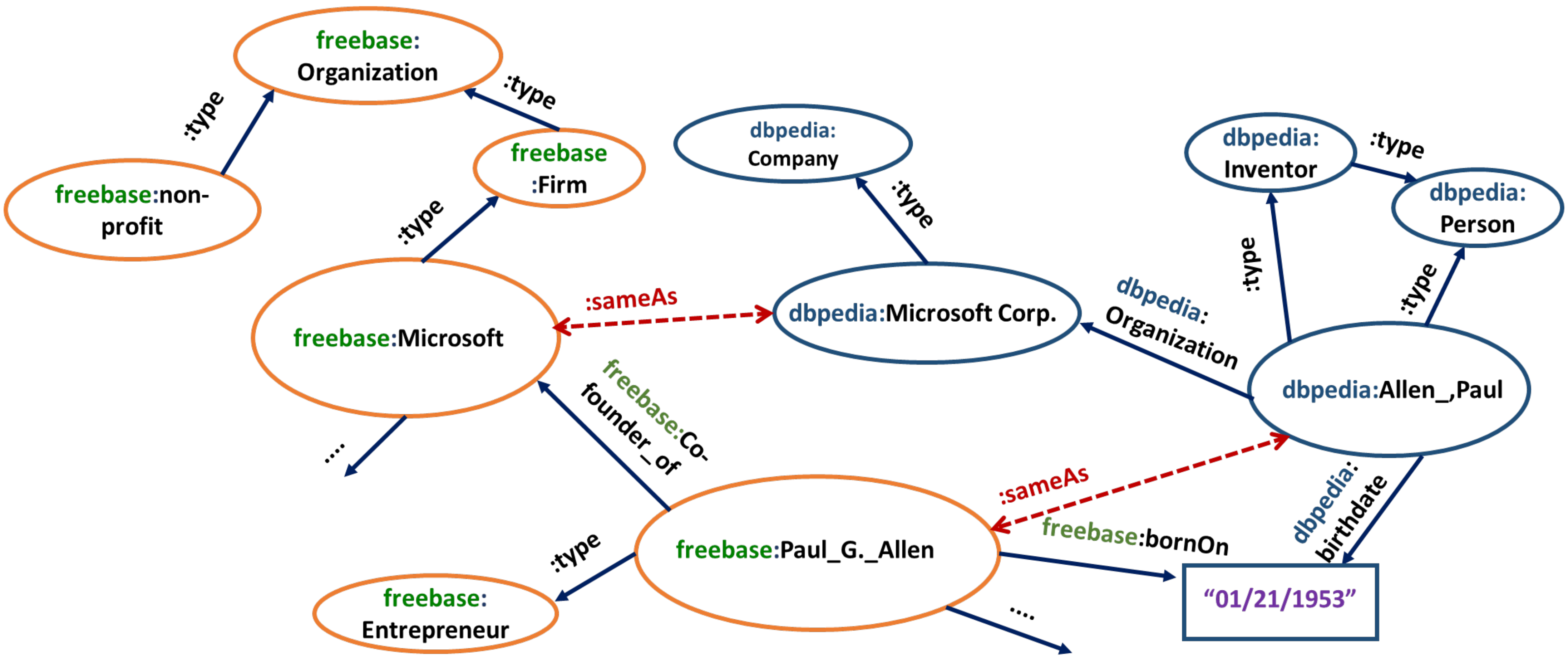
Source: Named Entity Recognition and Classification with Scikit-Learn. https://towardsdatascience.com/named-entity-recognition-and-classification-with-scikit-learn-f05372f07ba2

Other kinds of IE: Relation Extraction



Source: Stanford TACRED

In the world of knowledge graphs



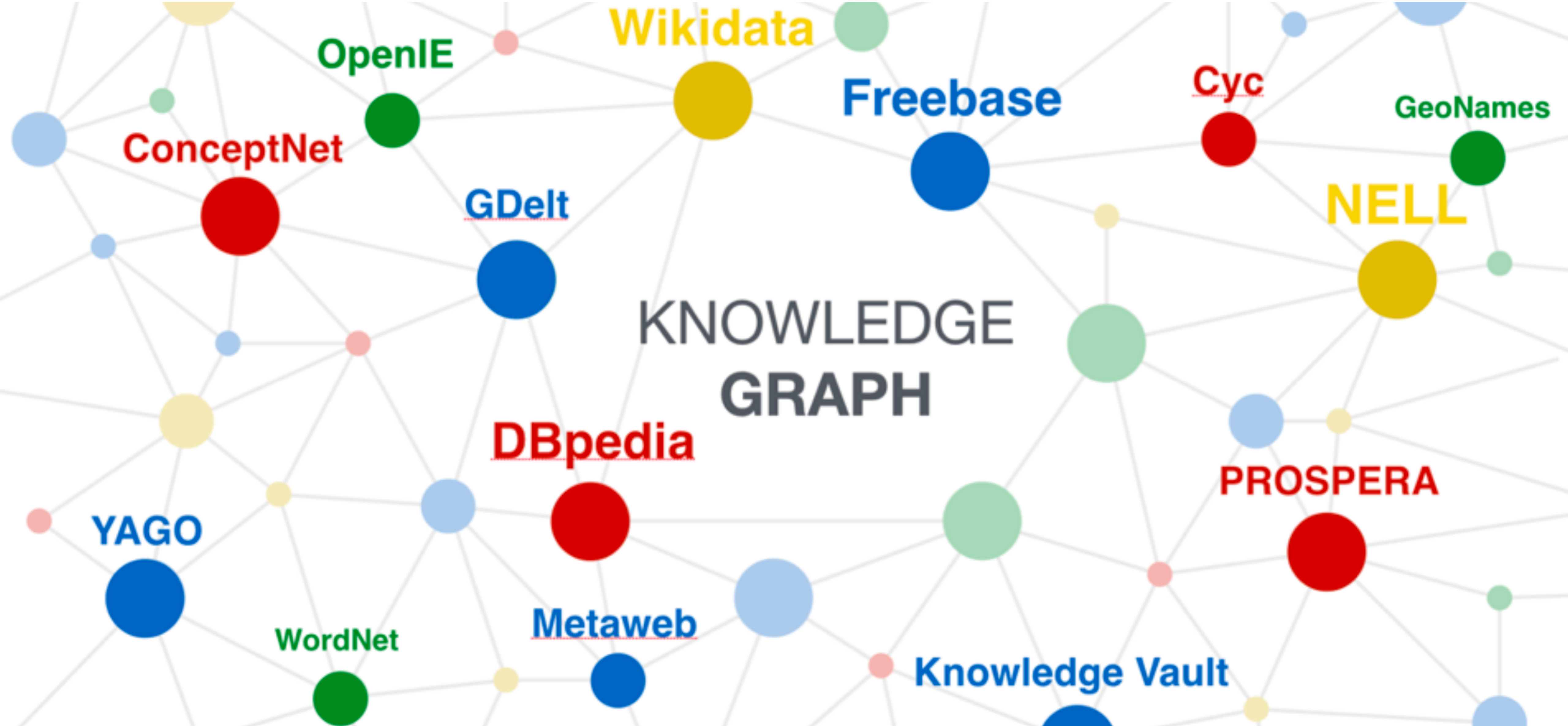
KGEs (results)

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**
HolE (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

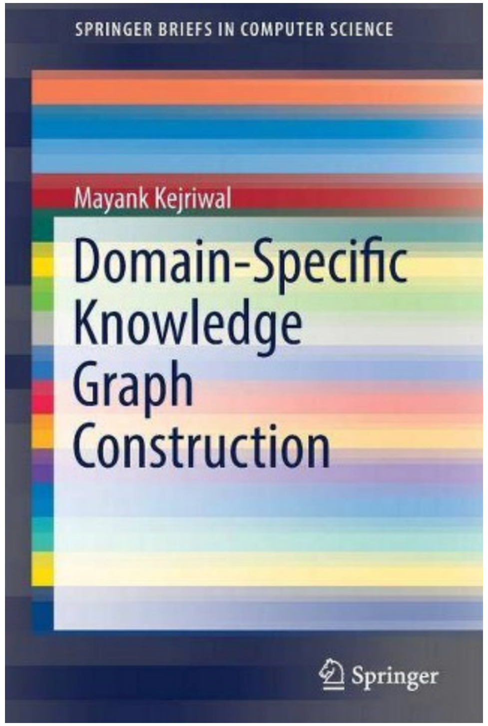
Recommender systems

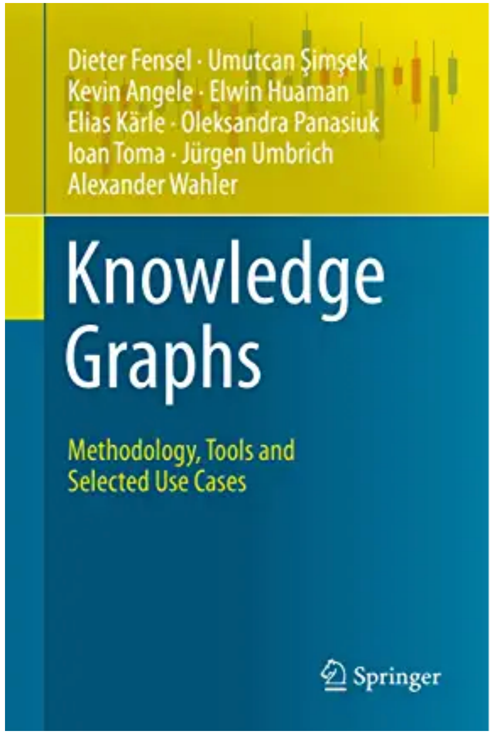
Knowledge discovery/data mining

?

Numerous surveys, some more technical/field-specific

- Ehrlinger, L., & Wöß, W. (2016). Towards a Definition of Knowledge Graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, *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

USC University of Southern California

Q & A

kejriwal@isi.edu