

TOO MUCH INFORMATION

CAN AI COPE WITH MODERN KNOWLEDGE GRAPHS?

Markus Krötzsch[†]

reporting on joint work with David Carral[†], Irina Dragoste[†], Maximilian Marx[†], Ana Ozaki^{†*}, Sebastian Rudolph, Veronika Thost^{†*}, and Denny Vrandečić

[†] Knowledge-Based Systems (* former member)

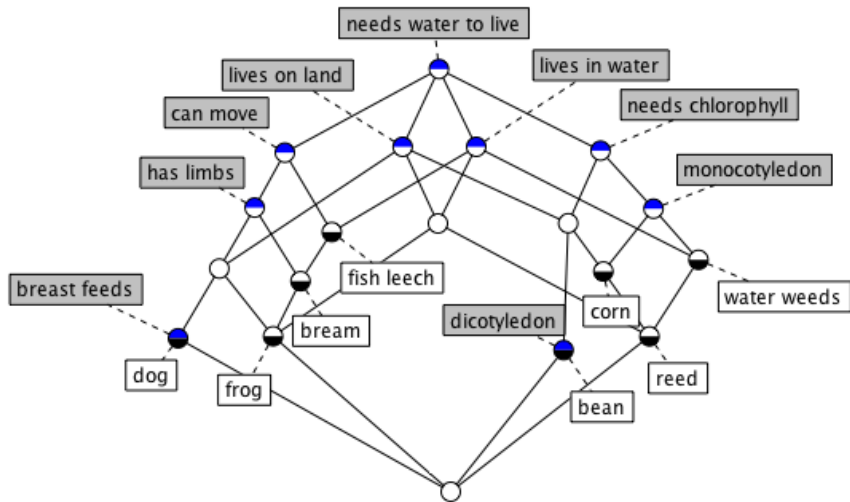
TU Dresden

Full paper: <https://iccl.inf.tu-dresden.de/web/Inproceedings3217/en>

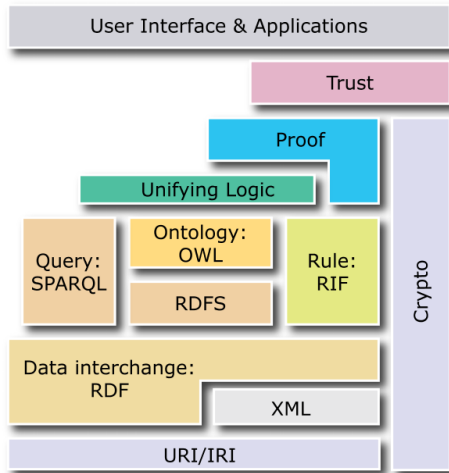
Invited Keynote

International Conference on Formal Concept Analysis (ICFCA 2019)

Formal Concept Analysis



2001: The Semantic Web



2012: The Knowledge Graph

Google Inside Search

Home How Search Works Tips & Tricks **Features** Search Stories Playground Blog Help



The Knowledge Graph

Learn more about one of the key breakthroughs behind the future of search.



See it in action

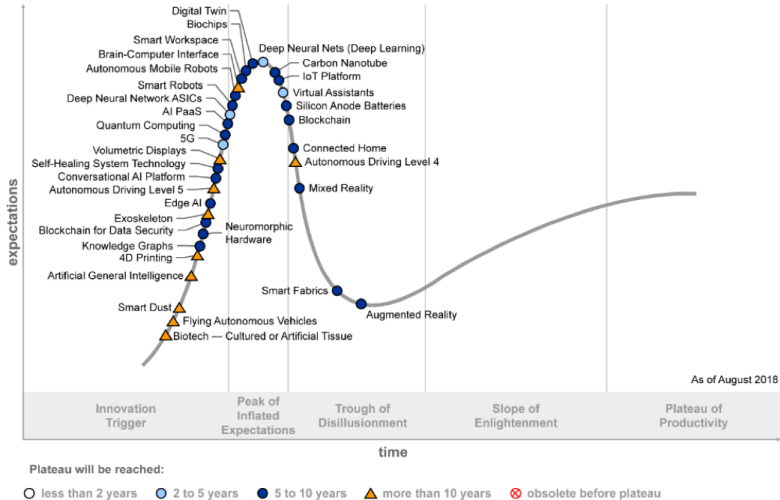
Discover answers to questions you never thought to ask, and explore collections and lists.



2012: Wikidata



2018: The Hype



(c) 2018 Gartner, Inc. All rights reserved.

2019: Knowledge Graphs Everywhere



EDITION: EU

CENTRAL EUROPE MIDDLE EAST SCANDINAVIA AFRICA UK ITALY SPAIN MORE

Knowledge graphs beyond the hype: Getting knowledge in and out of graphs and databases

What exactly are knowledge graphs, and what's with all the hype about them? Learning to tell apart hype from reality, defining different types of graphs, and picking the right tools and database for you want to be like the Airbnbs, Amazons, Googles, and LinkedIns of the world.

TechRepublic

BIG DATA

Amazon Neptune is here: 6 ways customers use the AWS graph database

Customers including Samsung, Intuit, and Pearson previewed the database, building new graph applications and testing production workloads.

By Alison DeNisco Rayome | May 31, 2018, 7:44 AM PST

TE

- Startups
- Apps
- Gadgets
- Events
- Videos
-
- Crunchbase
- More

acquire Lattice Data over the weekend. The startup was working to transform the way businesses deal with paragraphs of text and other information that lives outside neatly structured databases. These engineers are uniquely prepared to assist Apple with building a next-generation internal knowledge graph to power Siri and its next generation of intelligent products and services.

Broadly speaking, the Lattice Data deal was an acquire. Apple paid roughly \$10 million for each of Lattice's 20 engineers. This is generally considered to be fair market value. Google paid

Billionaires Innovation Leadership Money Consumer Industry

Is The Enterprise Knowledge Graph Finally Going To Make All Data Usable?



All company logos subject to copyrights. All rights reserved.

Many knowledge graphs, many technologies

There are a number of widely used publicly available knowledge graphs:



... and a variety of technologies for working with them:



What is a Knowledge Graph?

More than “a database used in an AI application”?

What is a Knowledge Graph?

More than “a database used in an AI application”?

Charateristics of today's KGs:

Normalised: Data decomposed into small units (“edges”)

Connected: Knowledge represented by relationships between these units

Annotated: Enriched with context information, meta-data, and auxiliary details

What is a Knowledge Graph?

More than “a database used in an AI application”?

Charateristics of today's KGs:

Normalised: Data decomposed into small units (“edges”)

Connected: Knowledge represented by relationships between these units

Annotated: Enriched with context information, meta-data, and auxiliary details

- Typical for many KG applications
- Often comes with a promise of declarative processing



A Free Knowledge Graph

Wikidata

- Wikipedia's knowledge graph
- Free, community-built database
- Large graph
(March 2018: >680M statements on >55M entities)
- Large, active community
(250,000 logged-in human editors)
- Many applications

Freely available, relevant, and active knowledge graph



[Vrandečić & K; Comm. ACM 2014]

I'm in ur phone ...

Who is Grover Cleveland

Tap to Edit >

OK. Check it out:



KNOWLEDGE

Grover Cleveland

22nd and 24th president of the United States



Stephen Grover Cleveland was an American politician and lawyer who was the 22nd and 24th President of the United States. He won the popular vote for three presidential elections – in 1884, 1888, and 1892 – and was one of two Democrats to be elected president during the era of Republican political domination dating from 1861 to 1933. He was also the first and to date only President in American history to serve two non-consecutive terms in office.

See More on Wikipedia



Date of birth

March 18, 1837

Tim Berners-Lee (Q80)

British computer scientist

 [edit](#)

[TimBL](#) | [Sir Tim Berners-Lee](#) | [Timothy John Berners-Lee](#) | [TBL](#) | [Tim Berners Lee](#) | [T. Berners-Lee](#) | [T Berners-Lee](#) | [Tim Berners-Lee](#) | [T.J. Berners-Lee](#)

Tim Berners-Lee (Q80)

British computer scientist

 [edit](#)

[TimBL](#) | [Sir Tim Berners-Lee](#) | [Timothy John Berners-Lee](#) | [TBL](#) | [Tim Berners Lee](#) | [T. Berners-Lee](#) | [T Berners-Lee](#) | [Tim Berners-Lee](#) | [T.J. Berners-Lee](#)

⋮

instance of



human

 [edit](#)

[▶ 1 reference](#)

Tim Berners-Lee (Q80)

British computer scientist

 [edit](#)

[TimBL](#) | [Sir Tim Berners-Lee](#) | [Timothy John Berners-Lee](#) | [TBL](#) | [Tim Berners Lee](#) | [T. Berners-Lee](#) | [T Berners-Lee](#) | [Tim Berners-Lee](#) | [T.J. Berners-Lee](#)

⋮

instance of



human

 [edit](#)

[▶ 1 reference](#)

⋮

employer



CERN

 [edit](#)

start time 1984

end time 1994

position held Fellow

[▼ 0 references](#)

[+ add reference](#)

Tim Berners-Lee (Q80)

British computer scientist

 edit

TimBL | Sir Tim Berners-Lee | Timothy John Berners-Lee | TBL | Tim Berners Lee | T. Berners-Lee | T Berners-Lee | Tim Berners-Lee | T.J. Berners-Lee

⋮

instance of



human

 edit

▶ 1 reference

⋮

employer



CERN

 edit

start time 1984

end time 1994

position held Fellow

▼ 0 references

+ add reference

⋮

award received



Queen Elizabeth Prize for Engineering

 edit

point in time 2013

together with Robert Kahn

Vint Cerf

Louis Pouzin

Marc Andreessen

▶ 1 reference

Statements in Wikidata

Wikidata's basic information units

- Built from **Wikidata items** (“CERN”, “Vint Cerf”), **Wikidata properties** (“award received”, “end time”), and **data values** (“2013”)
- Based on **directed edges**
 (“Tim Berners-Lee –employer→ CERN”)
- **Annotated with property-value pairs** (“end time: 1994”)
 - same property can have multiple annotation values
 (“together with: Robert Kahn, Vint Cerf, . . .”)
 - same properties/values used in directed edges and annotations
- Items and properties can be subjects/values in statements
- **Multi-graph**

Elizabeth Taylor (Q34851)

Elizabeth Rosemond Taylor | Liz Taylor | Dame Elizabeth Rosemond Taylor

British-American actress

instance of: Elizabeth Taylor is a(n) human

Human relationships

Own statements

spouse

8 statements

Larry Fortensky (construction worker and seventh husband of Elizabeth Taylor)

end time : 1996-10-31

start time : 1991-10-06

John Warner (Republican politician and Secretary of the Navy from the United States)

end time : 1982-11-07

start time : 1976-12-04

Richard Burton (Welsh actor)

start time : 1975-10-10

end time : 1976-07-29

Richard Burton (Welsh actor)

start time : 1964-03-15

end time : 1974-06-26

Eddie Fisher (American entertainer and singer)

end time : 1964-03-06

start time : 1959-05-12

Mike Todd (American theatre and film producer)

end time : 1958-03-22

start time : 1957-02-02

Michael Wilding (English television and film actor)

end time : 1957-01-30

start time : 1952-02-21

Conrad Hilton, Jr. (American hotelier)

end time : 1951-01-29

start time : 1950-05-06

From related entities

edit label



Links

[Wikidata page](#)

[Official website](#)

[Wikipedia article](#)

[Reasonator](#)

Identifiers

SFDb person ID 75200 [↗](#)

Elonet person ID 224907 [↗](#)

PORT person ID 7869 [↗](#)

AllMovie artist ID p70015 [↗](#)

Wikidata Statements in Terms of Graphs

Elizabeth Taylor (Q34851)

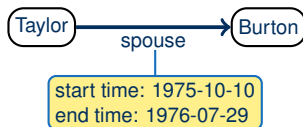
spouse	 Richard Burton  edit
	start time 10 October 1975
	end time 29 July 1976

Wikidata Statements in Terms of Graphs

Elizabeth Taylor (Q34851)

spouse	 Richard Burton  edit
start time	10 October 1975
end time	29 July 1976

“Property Graph”:

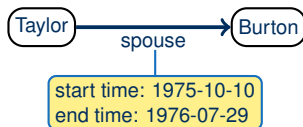


Wikidata Statements in Terms of Graphs

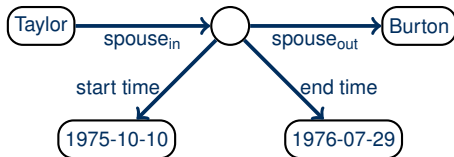
Elizabeth Taylor (Q34851)

spouse	 Richard Burton  edit
start time	10 October 1975
end time	29 July 1976

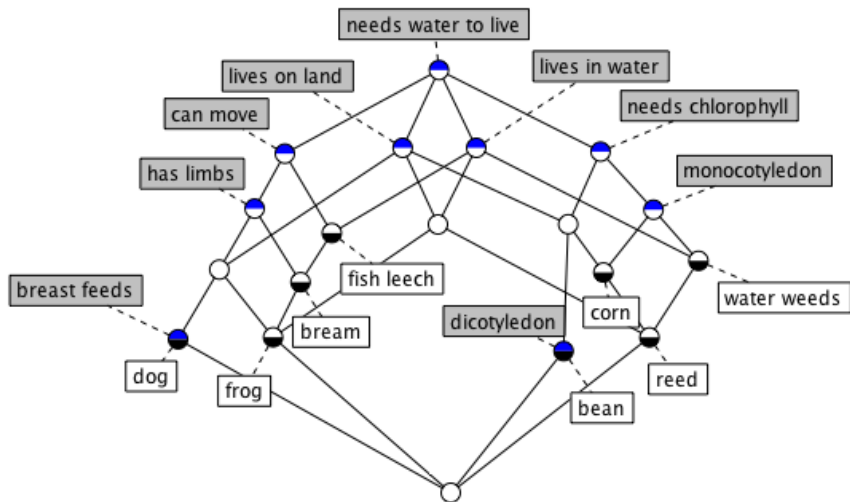
“Property Graph”:

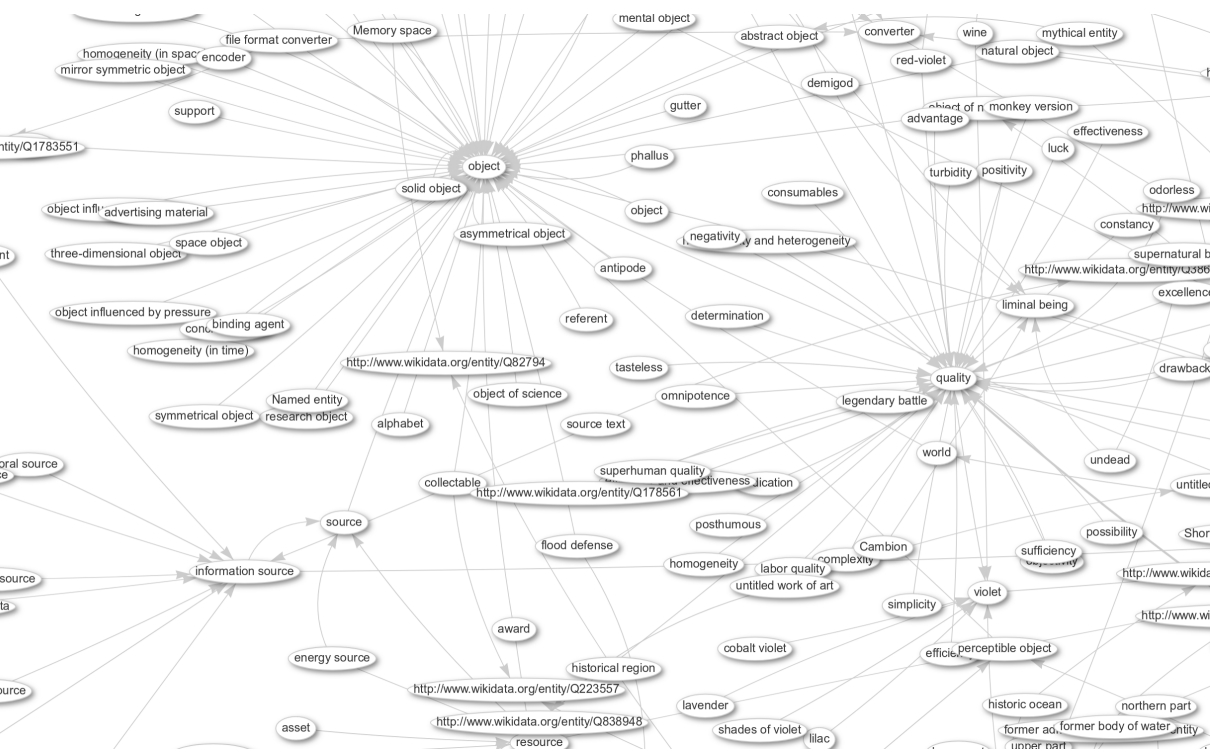


“RDF”:



Formal Concept Analysis





Wikidata's main challenge: conceptual modelling

Wikidata's main challenge: conceptual modelling

One or two items for one bridge in two locations ?

Hi all,

With @Jura1: we have a discussion on the french Project chat about should we have one or two items about a bridge that has been moved but it's going in circle and other point of view could help (@Fralambert, El Caro: also take part in the discussion.).

Wikidata's main challenge: conceptual modelling

One or two items for one bridge in two locations ?

Hi all,

With [@Jura1](#): we have a discussion on the french PR chat about should we have one or two items about a bridge that has been moved but it's going in circle and other view could help ([@Fralambert](#), [El Caro](#): also take part in the discussion.).

Egg yolk

Hi, I'm trying to sort the entries on the egg yolk ... [egg yolk \(Q181409\)](#), [yolk \(Q16336079\)](#) and the [egg yolk \(Q1302994\)](#).

If I understand correctly, there is one for yolk, one for egg yolk and one for chicken egg yolk ... but interwikis links are also big mess. [Mikani \(talk\)](#) 15:37, 9 August 2018 (UTC)

Wikidata's main challenge: conceptual modelling

One or two items for one bridge in two locations ?

Hi all,

With @Jura1: we have a discussion on the french Pr chat about should we have one or two items about ?

Goal as a criterion

I have added to [goal \(Q18530\)](#): subclass of (P279) --> [criterion \(Q1789452\)](#) because [goal \(Q18530\)](#) can be used with [criterion used \(P1013\)](#) (value type constraint appear). But I am not sure if it is correct... [Xaris333 \(talk\)](#) 20:11, 30 April 2018 (UTC)

Egg yolk

Hi, I'm trying to sort the entries on the egg yolk ... [egg yolk \(Q181409\)](#), [yolk \(Q16336079\)](#) and the [egg yolk \(Q1302994\)](#).

If I understand correctly, there is one for yolk, one for egg yolk and one for chicken egg yolk ... but interwikis links are [ni \(talk\)](#) 15:37, 9 August 2018 (UTC)

Wikidata's main challenge: conceptual modelling

One or two items for one bridge in two locations ?

Hi all,

With @Jura1: we have a discussion on the french PR chat about should we have one or two items about ?

Goal as a criterion

I have added to [goal \(Q18530\)](#): [subclass \(Q1789452\)](#) because [goal \(Q18530\)](#) can be [described by source \(P1343\)](#) (value type constraint applied) correct... [Xaris333 \(talk\)](#) 20:11, 30

Egg yolk

Hi, I'm trying to sort the entries on the egg yolk ... [egg yolk \(Q181409\)](#), [yolk \(Q16336079\)](#) and the [egg yolk \(Q1302994\)](#).

If I understand correctly, there is one for yolk, one for egg yolk and one for [chicken egg yolk ... but interwikis list](#)

Death in episode number...

How to better model that a character was killed in specific episode/book/film of series? I use [manner of death \(P1196\)](#) with *qualifier* [described by source \(P1343\)](#). Or better as *reference*? Or some other property? And how to distinguish from the case when a death of the character *was described* in episode (as a flashback, not a main storyline)? --[Infovarius \(talk\)](#) 11:20, 19 June 2018 (UTC)

Wikidata's main challenge: conceptual modelling

One or two items for one bridge in two locations ?

Hi all,

With @Jura1: we have a discussion on the french Pr chat about should we have one or two items about ?

Goal as a criterion

I have added to [goal \(Q18530\)](#): [subcl](#) ([Q1789452](#)) because [goal \(Q18530\)](#) ca

([P1013](#)) (valu correct... X

Egg yolk

Hi, I'm trying to sort the entries on the egg yolk ... [egg yolk \(Q181409\)](#), [yolk \(Q16336079\)](#) and the [egg yolk \(Q1302994\)](#).

If I understand correctly, there is one for yolk, one for egg yolk and one for chicken egg yolk ... but [intewiki](#) list

Death in episode number...

How to better model that a character was killed in specific episode/book/film of series? I use [manner of death \(P1196\)](#) with *qualifier* [described by](#)

Edit war at [Jesus Christ \(Q302\)](#) concerning [father \(P22\)](#)

An edit war has been going on at [Jesus Christ \(Q302\)](#) concerning a use of the property [father \(P22\)](#). Would be nice if we could have it resolved.

--[Njardarlogar \(talk\)](#) 15:28, 2 July 2018 (UTC)

Or some other n a death of the not a main TC)

AI

Possibilities?

Wikidata is the largest public knowledge graph ever created.

Now, finally, we can apply all our methods to real data!

- Logical reasoning!
- Data mining!
- Machine learning!
- ...

Or can't we?

Mining and learning

There are many techniques for mining and learning from discrete and graph-based data:

- FCA
- Network analysis
- Knowledge graph embeddings
- Rule mining
- ...

... but none of them works on Wikidata as it is

Why?

Mining and learning

There are many techniques for mining and learning from discrete and graph-based data:

- FCA
- Network analysis
- Knowledge graph embeddings
- Rule mining
- ...

... but none of them works on Wikidata as it is

Why?

- **Scalability:** Wikidata is huge
- **Complexity:** Known methods require simpler data

Example: FCA

Required:

- formal context (Boolean matrix)
- of moderate size (in at least one of the two dimensions)

Example: FCA

Required:

- formal context (Boolean matrix)
- of moderate size (in at least one of the two dimensions)

~> extraction via custom pre-processing

Example: Hanika, Marx, Stumme [ICFCA 2019] extract contexts with <100 attributes and up to 430K items (<0.8%) from Wikidata.

Example: Knowledge graph embeddings

Required:

- directed labelled graph of plain “triplets”
- of moderate size (more or less, depending on method)

Example: Knowledge graph embeddings

Required:

- directed labelled graph of plain “triplets”
- of moderate size (more or less, depending on method)

~> extraction via custom pre-processing

Example: A still often used “benchmark” for link prediction is FB15k, which is based on less than 0.035% of the Freebase KG (discontinued in 2014). Annotations of edges (“Compound Value Types” in Freebase) are not included.

Example: Machine Learning

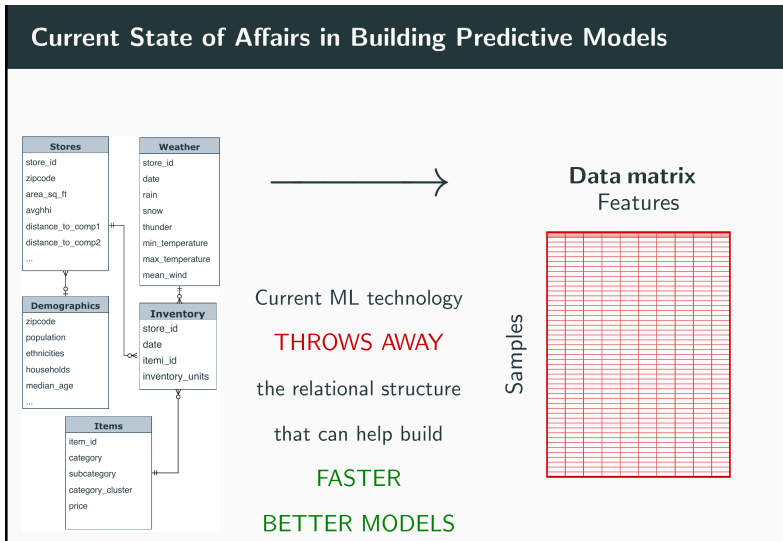


Fig.: Slide by Dan Olteanu: “Learning Models over Relational Databases ” (ICDT/EDBT 2019)

Big deal?

“Isn’t this **custom pre-processing** just a small syntactic adjustment, maybe with some application-specific sampling?”

Big deal?






“Isn't this custom pre-processing just a small syntactic adjustment, maybe with some application-specific sampling?”

No, much more interpretation is needed in this step!

London (Q84)

capital and largest city of the United Kingdom

London, UK | London, United Kingdom | London, England

located in the administrative territorial entity	 England	
	end time	1 January 1707 <i>Gregorian</i>
	start time	927
	▼ 0 references	
	 England	
end time	1889	
start time	1 January 1707 <i>Gregorian</i>	
▼ 0 references		
 County of London		
start time	1889	
end time	1 April 1965	
▼ 0 references		
 Greater London		
start time	1 April 1965	
▼ 0 references		
 Kingdom of Wessex		
end time	927	
▼ 0 references		

Instance of (P31) and Subclass of (P279)

Some classes that [Frankfurt am Main \(Q1794\)](#) is (indirectly) an instance of:

- big city, independent city of Germany, financial centre, college town

Instance of (P31) and Subclass of (P279)

Some classes that **Frankfurt am Main** (Q1794) is (indirectly) an instance of:

- big city, independent city of Germany, financial centre, college town
- city, city/town, place with town rights and privileges, municipality, human settlement, artificial geographic entity, artificial physical object, artificial entity, physical object

Instance of (P31) and Subclass of (P279)

Some classes that **Frankfurt am Main** (Q1794) is (indirectly) an instance of:

- **big city, independent city of Germany, financial centre, college town**
- **city, city/town, place with town rights and privileges, municipality**, human settlement, artificial geographic entity, artificial physical object, artificial entity, physical object
- **urban municipality of Germany**, municipality of Germany, districts and some cities of Germany, administrative territorial entity of Germany, administrative territorial entity of a single country, administrative territorial entity of a specific level, fourth-level administrative country subdivision, local administrative unit in the NUTS system, LAU 2, human-geographic territorial entity

Instance of (P31) and Subclass of (P279)

Some classes that **Frankfurt am Main** (Q1794) is (indirectly) an instance of:

- **big city, independent city of Germany, financial centre, college town**
- **city, city/town, place with town rights and privileges, municipality**, human settlement, artificial geographic entity, artificial physical object, artificial entity, physical object
- **urban municipality of Germany**, municipality of Germany, districts and some cities of Germany, administrative territorial entity of Germany, administrative territorial entity of a single country, administrative territorial entity of a specific level, fourth-level administrative country subdivision, local administrative unit in the NUTS system, LAU 2, human-geographic territorial entity
- **urban area**, geographic region, geographical object, geographic location, geographic entity, territorial entity, location, locality, spatial entity, spacio-temporal entity

Instance of (P31) and Subclass of (P279)

Some classes that **Frankfurt am Main** (Q1794) is (indirectly) an instance of:

- **big city, independent city of Germany, financial centre, college town**
- **city, city/town, place with town rights and privileges, municipality**, human settlement, artificial geographic entity, artificial physical object, artificial entity, physical object
- **urban municipality of Germany**, municipality of Germany, districts and some cities of Germany, administrative territorial entity of Germany, administrative territorial entity of a single country, administrative territorial entity of a specific level, fourth-level administrative country subdivision, local administrative unit in the NUTS system, LAU 2, human-geographic territorial entity
- **urban area**, geographic region, geographical object, geographic location, geographic entity, territorial entity, location, locality, spatial entity, spacio-temporal entity
- **community, social group, group of humans, group of living things, group of physical objects, group**

Instance of (P31) and Subclass of (P279)

Some classes that **Frankfurt am Main** (Q1794) is (indirectly) an instance of:

- **big city, independent city of Germany, financial centre, college town**
- **city, city/town, place with town rights and privileges, municipality**, human settlement, artificial geographic entity, artificial physical object, artificial entity, physical object
- **urban municipality of Germany**, municipality of Germany, districts and some cities of Germany, administrative territorial entity of Germany, administrative territorial entity of a single country, administrative territorial entity of a specific level, fourth-level administrative country subdivision, local administrative unit in the NUTS system, LAU 2, human-geographic territorial entity
- **urban area**, geographic region, geographical object, geographic location, geographic entity, territorial entity, location, locality, spatial entity, spacio-temporal entity
- **community, social group, group of humans, group of living things, group of physical objects, group**
- **local government**, municipal corporation, government, executive branch, government organization, political organisation, organization, political entity, public body, political territorial entity
- **public authority, organ, authority, legal person, juridical person, agent, individual, state power, political power, power (Q25107), power (Q18340964)**

Instance of (P31) and Subclass of (P279)

Some classes that **Frankfurt am Main** (Q1794) is (indirectly) an instance of:

- **big city, independent city of Germany, financial centre, college town**
- **city, city/town, place with town rights and privileges, municipality**, human settlement, artificial geographic entity, artificial physical object, artificial entity, physical object
- **urban municipality of Germany**, municipality of Germany, districts and some cities of Germany, administrative territorial entity of Germany, administrative territorial entity of a single country, administrative territorial entity of a specific level, fourth-level administrative country subdivision, local administrative unit in the NUTS system, LAU 2, human-geographic territorial entity
- **urban area**, geographic region, geographical object, geographic location, geographic entity, territorial entity, location, locality, spatial entity, spacio-temporal entity
- **community, social group, group of humans, group of living things, group of physical objects, group**
- **local government**, municipal corporation, government, executive branch, government organization, political organisation, organization, political entity, public body, political territorial entity
- **public authority, organ, authority, legal person, juridical person, agent, individual, state power, political power, power (Q25107), power (Q18340964)**
- **system, unit, structure (Q6671777), structure (Q517966), concrete object, abstract object, object (Q488383), object (Q17553950), subject, entity**

Instance of (P31) and Subclass of (P279)

Some classes that **Frankfurt am Main** (Q1794) is (indirectly) an instance of:

- **big city, independent city of Germany, financial centre, college town**
- **city, city/town, place with town rights and privileges, municipality**, human settlement, artificial geographic entity, artificial physical object, artificial entity, physical object
- **urban municipality of Germany**, municipality of Germany, districts and some cities of Germany, administrative territorial entity of Germany, administrative territorial entity of a single country, administrative territorial entity of a specific level, fourth-level administrative country subdivision, local administrative unit in the NUTS system, LAU 2, human-geographic territorial entity
- **urban area**, geographic region, geographical object, geographic location, geographic entity, territorial entity, location, locality, spatial entity, spacio-temporal entity
- **community, social group, group of humans, group of living things, group of physical objects, group**
- **local government**, municipal corporation, government, executive branch, government organization, political organisation, organization, political entity, public body, political territorial entity
- **public authority, organ, authority, legal person, juridical person, agent, individual, state power, political power, power (Q25107), power (Q18340964)**
- **system, unit, structure (Q6671777), structure (Q517966), concrete object, abstract object, object (Q488383), object (Q17553950), subject, entity**
- **statistical territorial entity, unit of analysis, research object**

Conceptual modelling in Wikidata

Conceptual models are an important part of Wikidata's content

However, Wikidata has no built-in ontology language:

- Schema information stored with special properties (e.g., P279 “subclass of”)
- Classes (and metaclasses) are just regular items
- No clear distinction between instance and schema knowledge
- No fixed formal semantics

Conceptual modelling in Wikidata

Conceptual models are an important part of Wikidata's content

However, Wikidata has no built-in ontology language:

- Schema information stored with special properties (e.g., P279 “subclass of”)
- Classes (and metaclasses) are just regular items
- No clear distinction between instance and schema knowledge
- No fixed formal semantics

↪ interpretation is necessary to make sense of data

- Domain-specific background knowledge can be required
- Interpretation might depend on context

Implicit schema knowledge in Wikidata

Airbus A380 (Q5830)

double-deck aircraft made by Airbus

 [edit](#)

A380 | Airbus Jumbo Jet | A380 Jumbo Jet

powered by



Rolls-Royce Trent 700

 [edit](#)

quantity

4

 [0 references](#)

[+ add reference](#)

Implicit schema knowledge in Wikidata

Airbus A380 (Q5830)

double-deck aircraft made by Airbus

 edit

A380 | Airbus Jumbo Jet | A380 Jumbo Jet

powered by



Rolls-Royce Trent 700

 edit

quantity

4

▼ 0 references

+ add reference

Intended meaning: “Every aircraft of this type has engines of that type.”

In description logics: “A380 \sqsubseteq =4 poweredBy.Trent700”

Ontology-based views

Ontology-based views

How can we capture the background knowledge used to interpret KGs?

Ontology-based views

How can we capture the background knowledge used to interpret KGs?

Ontology-based view definitions

- Describe mappings/queries with logical axioms
- Extract data (for mining and learning) by reasoning
- Draw inferences about KG by inverting view

Example: Description logics have been used to define attributes for FCA [Rudolph, ICCS 2004]:

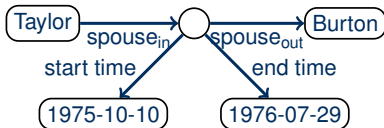
$$\text{Attribute}_{\text{hasMother}} \equiv \exists \text{mother. T}$$

This view is invertible: if objects have $\text{Attribute}_{\text{hasMother}}$, they must be in class $\exists \text{mother. T}$. Rule mining lets us learn ontologies [Borchmann, ICFCA 2013].

The limits of description logics

DLs are not ideal for interpreting KG data:

- No support for annotations – RDF-style pre-transformation of data needed:



- Can only express tree-like structures

Example: attribute “child of married parents” is not expressible.

- No closed-world reasoning

Example: attribute “currently married” (without end date) is not expressible.

- Very limited modelling of binary relations (needed for graph-like views)

Logics for annotations

Annotations as in Wikidata or Property Graph are not supported by standard relational first-order logic (and any of its fragments).

Idea: extend first-order logic with sort for annotation sets

Logics for annotations

Annotations as in Wikidata or Property Graph are not supported by standard relational first-order logic (and any of its fragments).

Idea: extend first-order logic with sort for annotation sets

MAPL: Multi-attributed predicate logic supports axioms with annotations [Marx, K, Thost; IJCAI 2017]:

- Annotated fact:
spouse(taylor, burton)@{start : 1975, end : 1976}
- Sentence with object and set variables:
 $\forall x, y, Z. \text{spouse}(x, y)@Z \rightarrow \text{spouse}(y, x)@Z$

Related approaches:

- Attributed description logics [K, Marx, Ozaki, Thost; ISWC 2017 & IJCAI 2018]
- Relational algebra with complex values; see [Abiteboul, Hull, Vianu; 1994]

Reasoning in attributed logics?

[Marx, K, Thost; IJCAI 2017]

Unrestricted quantification over finite annotation sets is extremely powerful:

Theorem: Entailment in attributed first-order logic captures entailment in Weak Second-Order Logic (and in particular is undecidable).

The problem becomes simpler when restricting to rules with “safe” quantification over annotation sets \rightsquigarrow rule language **MARPL** (“Datalog with annotation sets”):

Theorem: Entailment in MARPL is ExpTime-complete in data complexity and in combined complexity.

\rightsquigarrow overall similar to DLs and Datalog, but greater expressivity with respect to data

Practical reasoning in attributed logics?

Status quo:

- There is no dedicated reasoner for any annotated logic
- Modern rule engines for known decidable fragments of first-order Horn logic rules can handle large inputs
- These fragments mostly have PTime data complexity (too weak for MARPL)

Practical reasoning in attributed logics?

Status quo:

- There is no dedicated reasoner for any annotated logic
- Modern rule engines for known decidable fragments of first-order Horn logic rules can handle large inputs
- These fragments mostly have PTime data complexity (too weak for MARPL)

A recent insight changes the picture:

Theorem [K, Marx, Rudolph; ICDT 2019]: Algorithms implemented by modern rule engines solve problems of non-elementary data complexity.

Practical reasoning in attributed logics?

Status quo:

- There is no dedicated reasoner for any annotated logic
- Modern rule engines for known decidable fragments of first-order Horn logic rules can handle large inputs
- These fragments mostly have PTime data complexity (too weak for MARPL)

A recent insight changes the picture:

Theorem [K, Marx, Rudolph; ICDT 2019]: Algorithms implemented by modern rule engines solve problems of non-elementary data complexity.

In particular, such rule engines can be used to solve ExpTime-complete problems [Carral, Dragoste, K; IJCAI 2019]:

- Expressive DL reasoning
- Fact entailment for guarded Horn logic rules

↪ practical reasoning with annotations seems within reach

Conclusions

A research programme

Goal: Close the gap between large scale knowledge representation and intelligent data analysis

- (1) Model declarative conceptual views over knowledge graphs using a suitable ontology language
- (2) Design scalable reasoning algorithms for exchanging data through these views – forwards and backwards
- (3) Integrate data mining and machine learning methods with this reasoning process for efficiency and explainability

Conclusions

Summary

- Knowledge Graphs are enriched graphs
- Wikidata: large dataset + conceptual world model
- Mining KGs: struggling with size, but mainly with structural complexity
- Ontology-based views: declarative, invertible data excerpts
- Attributed logics: towards ontology support for KGs

What next?

Integrate ontological reasoning, data mining, and learning
in a clean and coherent way.

References

- [Abiteboul, Hull, Vianu; 1994] Serge Abiteboul, Richard Hull, Victor Vianu: **Foundations of Databases**. Addison Wesley, 1994.
- [Borchmann, ICFA 2013] Daniel Borchmann: **Towards an error-tolerant construction of \mathcal{EL}^+ -ontologies from data using formal concept analysis**. ICFA 2013.
- [Carral, Dragoste, K; IJCAI 2019] David Carral, Irina Dragoste, Markus Krötzsch: **Chasing Sets: How to Use Existential Rules for Expressive Reasoning**. IJCAI 2019 (to appear).
- [Hanika, Marx, Stumme [ICFA 2019]] Tom Hanika, Maximilian Marx, Gerd Stumme: **Discovering implicational knowledge in Wikidata**. ICFA 2019.
- [K, Marx, Ozaki, Thost; ISWC 2017] Markus Krötzsch, Maximilian Marx, Ana Ozaki, Veronika Thost: **Attributed Description Logics: Ontologies for Knowledge Graphs**. ISWC 2017.
- [K, Marx, Rudolph; ICDT 2019] Markus Krötzsch, Maximilian Marx, Sebastian Rudolph: **The power of the terminating chase**. ICDT 2019.
- [Marx, K, Thost, IJCAI 2017] Maximilian Marx, Markus Krötzsch, Veronika Thost: **Logic on MARS: Ontologies for Generalised Property Graphs**. IJCAI 2017.
- [Rudolph, ICCS 2004] Sebastian Rudolph: **Exploring relational structures via FLE**. ICCS 2004.
- [Vrandečić & K; Comm. ACM 2014] Denny Vrandečić, Markus Krötzsch: **Wikidata: A Free Collaborative Knowledgebase**. Comm. ACM, 57:10, 2014.