

Source: Generated with AI

What is a Knowledge Graph ?

Understanding Knowledge Graphs, their uses, progresses, and challenges

Florian Rascoussier - Double Master PhDTrack Program 2023 - 7th June 2023

This section is based on personal work

Let's imagine... ...different data sources

- Known rules <domain specific knowledge>
- Raw heap dump files <*.raw>
- Annotations files <*.json>

```
(base) [onyr@kenzael phdtrack_data]$ cat ./Training/Training/scp_V/7_8_P1/16/1010-1644391327.json | json_pp
{
    "ENCRYPTION_KEY_1_NAME" : "aes128-ctr",
    "ENCRYPTION_KEY_1_NAME_ADDR" : "558b967f7620",
    "ENCRYPTION_KEY_2_NAME" : "aes128-ctr",
    "ENCRYPTION_KEY_2_NAME_ADDR" : "558b967fb160",
    "HEAP_START" : "558b967e9000",
    "KEY_A" : "119bd34f49d27bbbc0f9af400d4edc39",
    "KEY_A_ADDR" : "558b967fefe0",
    "KEY_A_LEN" : "16",
    "KEY_A_REAL_LEN" : "16",
    "KEY_B" : "8a77835eb2007a46a776ae0c183253b9",
    "KEY_B_ADDR" : "558b967f5ce0",
    "KEY_B_LEN" : "16",
    "KEY_B_REAL_LEN" : "16",
    "KEY_C" : "539ff6dbd2007b3b4cfbd22fb22b8f2e7"
```

The screenshot shows a GitHub code search interface. The search bar at the top has 'master' selected. Below it, there is a search input field with a magnifying glass icon. A sidebar on the left lists several files: jffs2_fs_sb.h, malloc.c (which is currently selected), nodelist.c, nodelist.h, and nodemgmt.c. To the right of the sidebar, the main area displays the content of the selected file, malloc.c. The code is as follows:

```
116     ret = kmalloc(sizeof(struct jffs2_full_dirent) + namesize, G
117     dbg_memalloc("%p\n", ret);
118     return ret;
119 }
```

Below the code, there is a 'Code' tab and a 'Blame' tab. The 'Blame' tab shows statistics: 315 lines (270 loc) · 7.33 KB. To the right of the blame stats, there is a 'Memory Allocation Guide' section. This section provides information about memory allocation APIs in Linux, mentioning GFP flags and various allocation functions like kmalloc, kmalloc_cache, vmalloc, and alloc_pages. It also notes the diversity of allocation APIs and the complexity of determining how memory should be allocated.

^ rules

raw data >

< annotations

```
00000350:90957e968b550000b0957e968b550000...~..U...~..U...
00000360:d0957e968b5500000967e968b550000...~..U...~..U...
00000370:30967e968b55000080967e968b550000...~..U...~..U...
00000380:a0967e968b550000c0967e968b550000...~..U...~..U...
00000390:00000000000000002100000000000000.....!.....
000003a0:5348454c4c3d2f62696e2f26261736800SHELL=/bin/bash.
000003b0:00000000000000002100000000000000.....!.....
000003c0:4c414e47554147453d656e5f55533a65LLANGUAGE=en_US:e
000003d0:5e0000000000000021000000000000n.....!.....
000003e0:5057443d2f726f6f74000000000000PWD=/root.....
000003f0:00000000000000002100000000000000.....!.....
00000400:4c4f474e414d453d726f6f740000000LOGNAME=root....
00000410:00000000000000002100000000000000.....!.....
00000420:5844475f5345533494f4e5f54595045XDG_SESSION_TYPE
00000430:3d747479000000002100000000000000tty!.....
00000440:484f4d453d2f726f6f7400000000000HOME=/root.....
00000450:00000000000000002100000000000000.....!.....
00000460:4c414e473d656e5f55532e5554462d38LANG=en_US.UTF-8
00000470:00000000000000002100000000000000.....!.....
00000480:4c435f5445524d494e414c3d69546572LC_TERMINAL=iTer
00000490:6d320000000000001400000000000000m2.....A.....
000004a0:5353485f434f4e4e454354494f4e3d31SSH_CONNECTION=1
000004b0:302e34322e302e3220353831303020310.42.0.2581001
```

How to combine all that into a single-meaningful representation ?

It's your time to shine :)

use aKG



Plan

- 1. Introduction & History**
- 2. So, what is a KG ?**
- 3. KG construction introduction**
- 4. Critical Evaluation of the papers**
- 5. Conclusion**

Extended presentation:

- Advanced KG-concepts
(deduction, induction)
- Implementation
example



Sources & References

Main source:

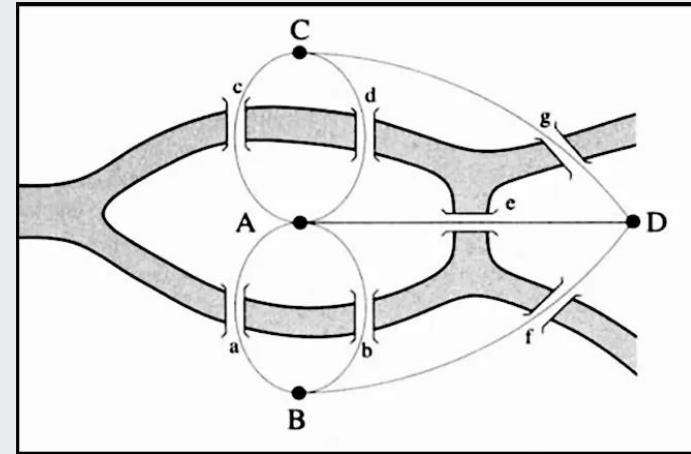
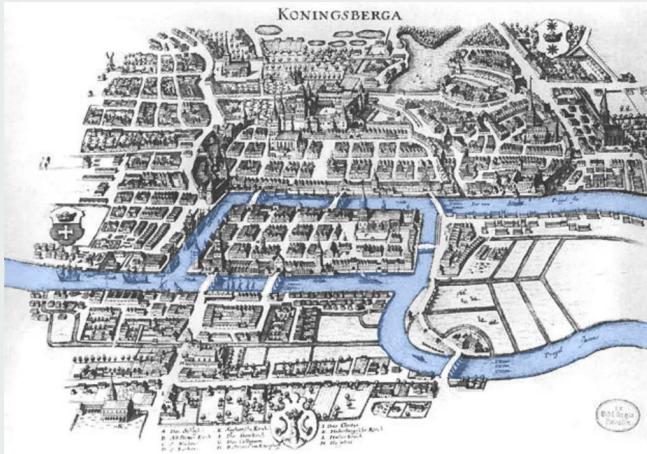
[KG21] <ACM Compt.> Aidan Hogan u. a. „Knowledge Graphs“. In: ACM Comput. Surv. 54.4 (Juli 2021). issn: 0360-0300. doi: 10.1145/3447772. url: <https://doi.org/10.1145/3447772.3>

Additional sources:

[KGKE22] <Dagstuhl Reports> Paul Groth u. a. „Knowledge Graphs and their Role in the Knowledge Engineering of the 21st Century“. In: Dagstuhl Reports 12.9 (2022). Report from Dagstuhl Seminar 22372, S. 60–120. doi: 10.4230/DagRep.12.9.60. Specific usage: pp. 60-72, Subsection "3.2 A Brief History of Knowledge Engineering: A Practitioner's Perspective", doi: 10.4230/DagRep.12.9.60.

[CKG23] <preprint> Marvin Hofer u. a. „Construction of Knowledge Graphs: State and Challenges“. In: arXiv preprint arXiv:2302.11509 (2023). url: <https://doi.org/10.48550/arXiv.2302.11509>.

Introduction & History



Source: [towardsdatascience.com]

This section is based on [KGKE22]

Introduction

Knowledge Graph:

A tool for **storing** and **organizing** information...

... with many real-world use cases.

Aim: Understanding knowledge graphs - its function, operation, and increasing relevance in businesses and organizations.

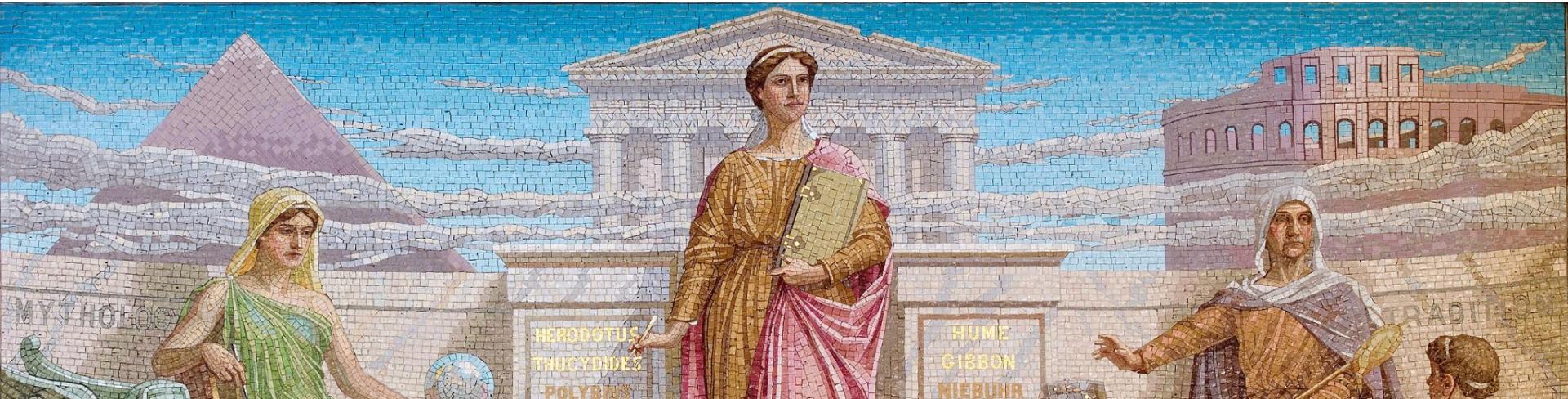


Source: [Wikipedia]

Historical Evolution of Knowledge Engineering (KE)

- From Expert Systems in the 1980s to the current Language Model Era
- [KGKE22] identifies 4 key periods for Knowledge engineering
- Each phase introduced new requirements for knowledge production

Source: [Wikipedia]



Source: [KGKE22]

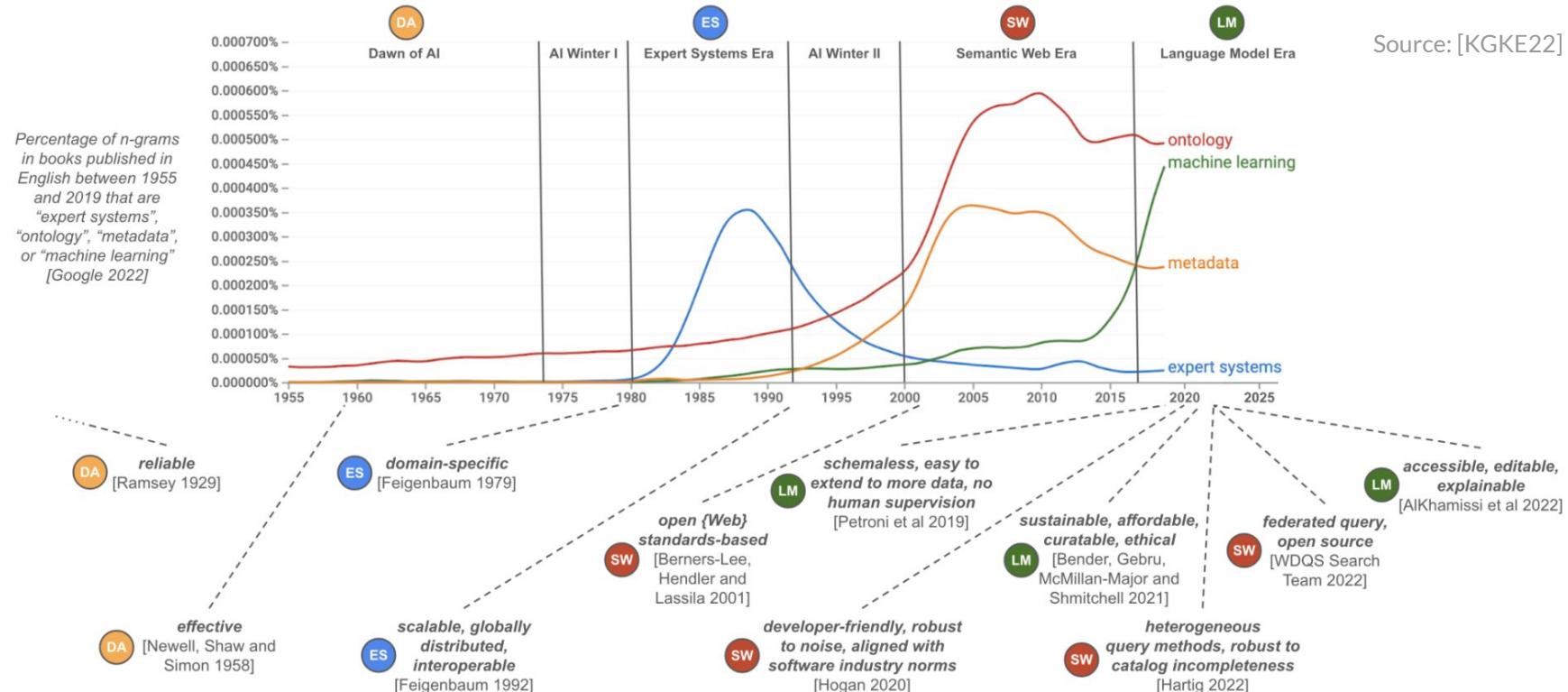


Figure 2 Seventy years of evolving requirements for knowledge production processes [1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13].

So, what *is* a KG ?

This section is based on [KG21]

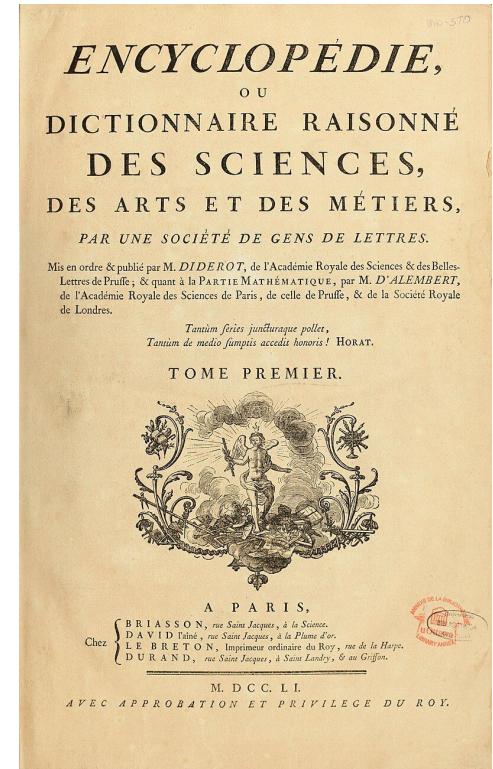
Knowledge ? Graph Theory ?

“Explicit” Knowledge:

- Known, **recordable** information
 - “something that is known and can be written down” [KG21 p4]
- Sequences of words establishing relationships between concepts and data

Graph Theory:

- Bridges computer science and mathematics
- Graphs: Data structures composed of **nodes** (vertices) and **edges** (arcs)
- Analyzes various relationship types and structures across multiple fields



Source: [Wikipedia]

KG

"At the foundation of any knowledge graph is the principle of first modelling data as a graph" [KG21 p4]

- Fundamental principle: Modeling data as a graph
 - A data graph intended to accumulate and convey real-world knowledge
 - **Nodes represent entities, edges represent relations** between these entities

Source: [KG21]

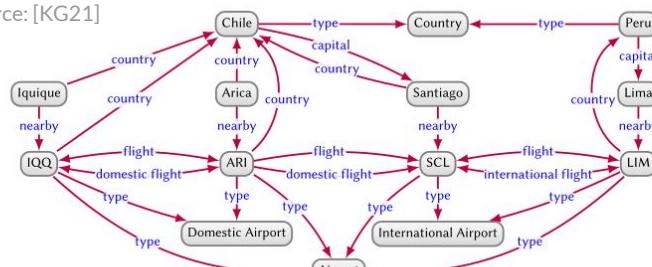


Fig. 21. An incomplete del graph describing flights between airports.

Benefits over relational or NoSQL models

- Flexibility
 - Tolerance for Incompleteness
 - Semantic Interlinking
 - Scalability
 - Advanced Queries
 - Knowledge Representation

This slide is based on [CKG23]

Many definitions

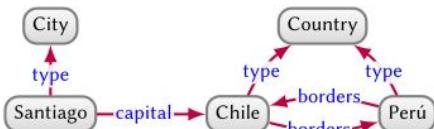


- Term dates back to **1973**. [ISIFSCD73]
 - Gained popularity through a **2012 blog post** about **Google's Knowledge Graph**.
 - **Several definitions** of knowledge graphs have been proposed in research papers and by companies using or supporting KGs
-
- "A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge." [TDKG16]
 - "A graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent potentially different relations between these entities." [KG21]
 - "A graph of data consisting of semantically described entities and relations of different types that are integrated from different sources. Entities have a unique identifier. KG entities and relations are semantically described using an ontology or, more clearly, an ontological representation." [CKG23]

Types of KG

- Directed Edge-labelled Graphs (DEL)
- Heterogeneous Graphs
- Property Graphs

NB: KGs can adopt any graph data model; conversion possible between models.

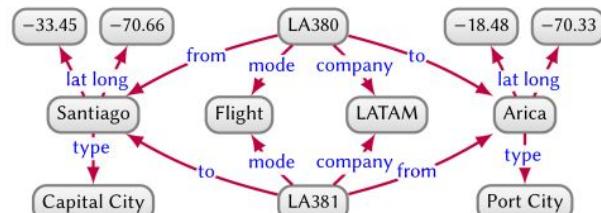


(a) Del graph

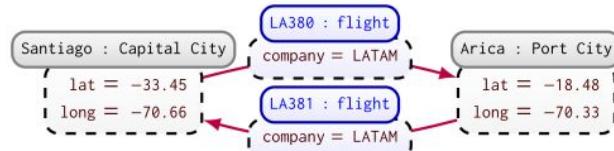


(b) Heterogeneous graph

Fig. 2. Data about capitals and countries in a del graph and a heterogeneous graph.



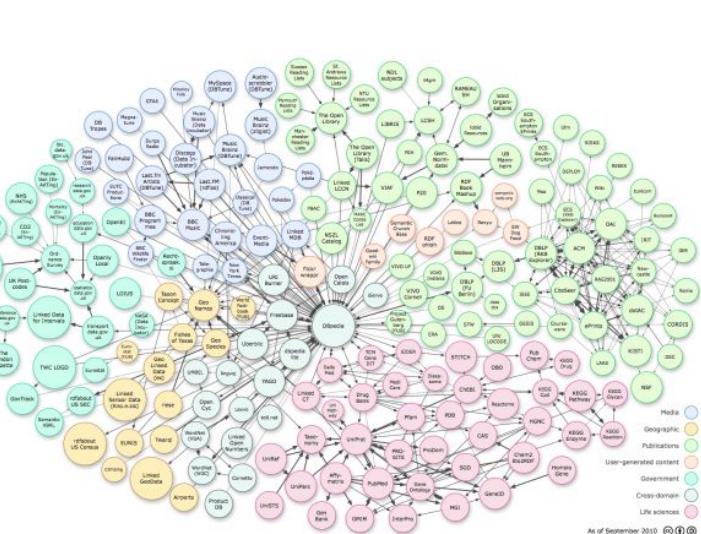
(a) Del graph



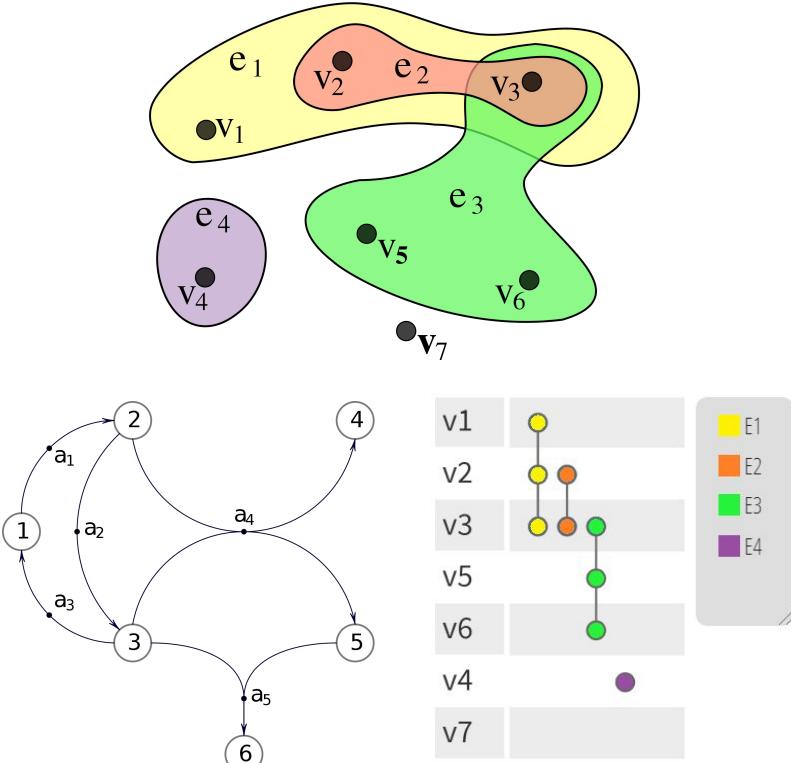
(b) Property graph

Fig. 3. Flight data in a del graph and a property graph.

- Graph Dataset
- Hypergraphs



Source: [Wikipedia]

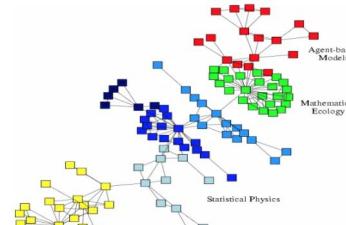


A wide range of use-case

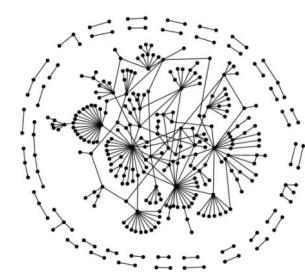
- Flexibility allows to model data in different fields
 - Web Search
 - Commerce
 - Social Networks
 - Finance
 - ...
- Other Applications like Information extraction, personal agents, advertising, automation, etc.



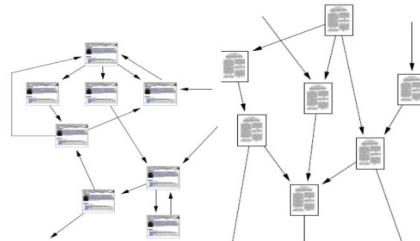
Social networks



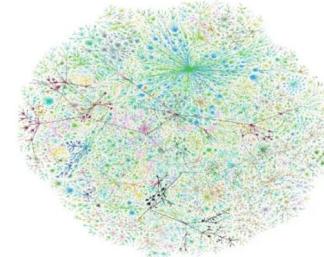
Economic networks



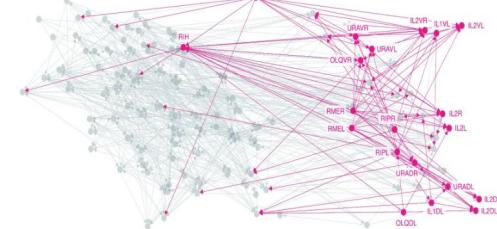
Biomedical networks



Information networks:
Web & citations



Internet



Networks of neurons

Source: [towardsdatascience.com]

Famous KGs

	Year	Domain	Srcs.	Model	Entities	Relations	Types	R-Types	Vers.	Update
<u>Closed KG</u>										
Google KG [195]	2012	Cross,MLang	>>>1	Custom,RDF	1B	>100B	?	?	?	?
Diffbot.com	2019	Cross	>>>1	RDF	5.9B	>1T	?	?	?	?
Amazon PG [196]	2020	Products	>1	Custom	30M	1B	19K	1K	?	?
<u>Open Access KG</u>										
*Freebase [197]	2007	Cross	>>1	RDF	22M	3.2B	53K	70K	>1	2016
DBpedia [198]	2007	Cross,MLang	140	RDF	50M	21B	1.3K	55K	>20	2023
YAGO [199] [200]	2007	Cross	2-3	RDF(-Star)	67M	2B	10K	157	5	2020
NELL [201]	2010	Cross	≥1	Custom,RDF	2M	2.8M	1.2K	834	>1100	2018
*Wikidata [202]	2012	Cross,MLang	>>>1	RDB/RDF	100M	14B	300K	10.3K	>100	2023
DBpedia-EN Live [203]	2012	Cross	1	RDF	7.6M	1.1B	800	1.3K	>>>1	2023
Artist-KG [204]	2016	Artists	4	Custom	161K	15M	>1	18	1	2016
*ORKG [205]	2019	Research	>>1	RDF	130K	870K	1.3K	6.3K	>1	2023
AI-KG [206]	2020	AI Science	3	RDF	820K	1.2M	5	27	2	2020
CovidGraph [207]	2020	COVID-19	17	PGM	36M	59M	128	171	>1	2020
DRKG [208]	2020	BioMedicine	>7	CSV	97K	5.8M	17	107	1	2020
VisualSem [209]	2020	Cross,MLang	2	Custom	90k	1.5M	(49K)	13	2	2020
WorldKG [210]	2021	Geographic	1	RDF	113M	829M	1176	1820	1	2021

Enterprise Knowledge Graphs:

- **Google Knowledge Graph:** Enhances Google Search's results with semantic-search information gathered from various sources
- **Amazon Product Knowledge Graph (PKG):** is a large-scale, semi-structured knowledge graph that organizes information about products sold on Amazon and relationships between them.

Open Knowledge Graphs:

- **BabelNet:** Several resources: Wikipedia, WordNet for multilingual lexical knowledge
- **DBpedia:** Structured content from Wikipedia
- **Freebase:** Crowdsourced database of well-known people, places, and things
- **Wikidata:** Central storage for the structured data of Wikimedia projects
- **YAGO:** Automatically extracts and integrates knowledge from Wikipedia and other sources

KG construction introduction

This section is based on [CKG21]

What's the issue ?

- Goal: having a “useful” and “usable” organisation of information

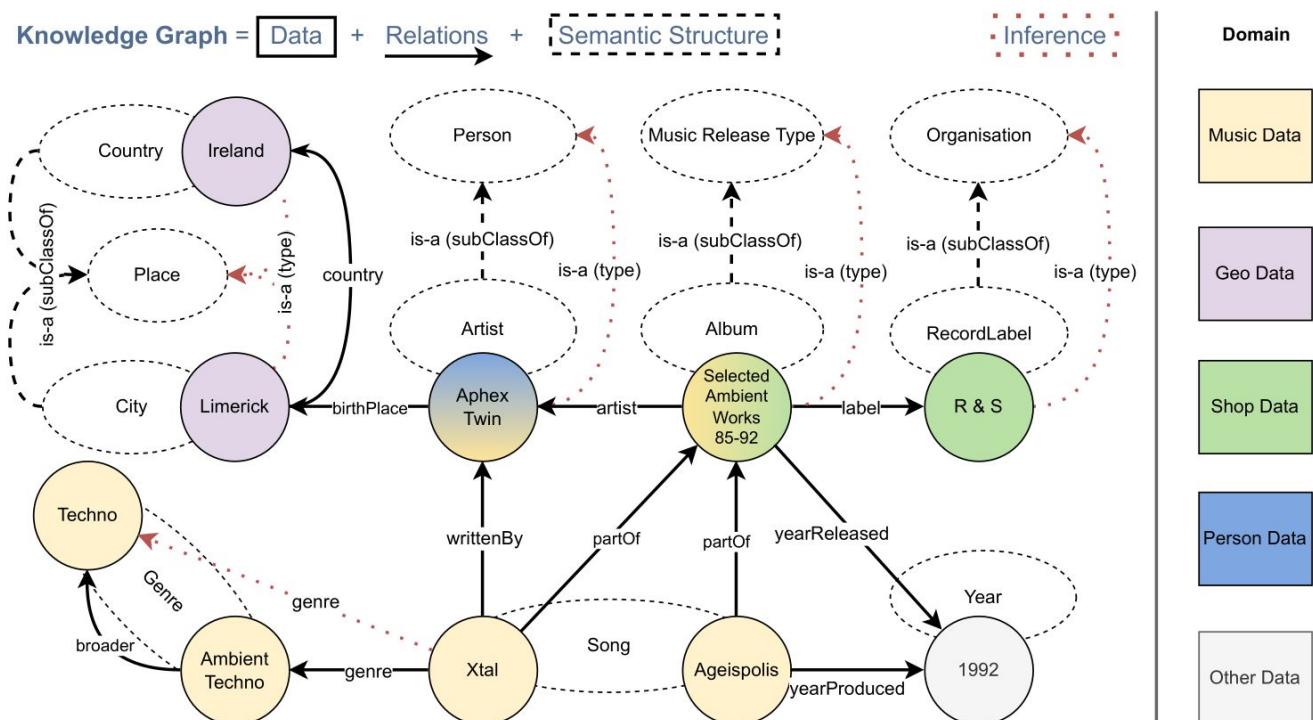


Fig. 1. **Simplified KG example** with integrated data from several domains. Entities and relations are described by an underlying ontology that allows the inference of additional relations (dashed red lines).

Source: [CKG23]

From data to KG

- Need to integrate consistently different sources
- Metadata can be quite diverse
- Temporality, redundancies, missing or incorrect information

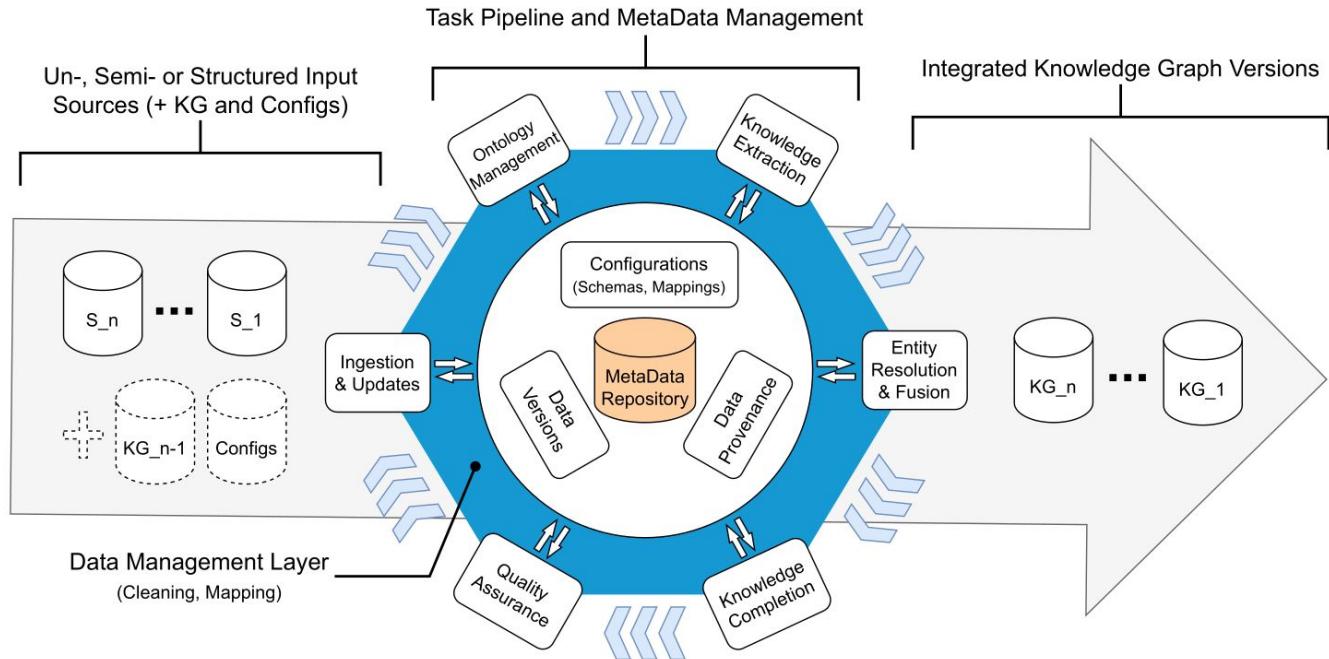


Fig. 2. Incremental Knowledge Graph Construction Pipeline

Source: [CKG23]

Overview of KG Construction Tasks and Approaches

- Data Acquisition & Preprocessing
- Metadata Management
- Ontology Management
- Knowledge Extraction
- Entity Resolution (ER) and Fusion
- Quality Assurance (QA)
- Knowledge Completion

Wait for more...

topic	student	Monitor
What is a KG?	Rascoussier, Florian Guillaume Pierre	Prof. Algergawy
KG construction from text	Boudemagh, Houria-Chiraz	Prof. Algergawy
KG construction from tabular	Tayebi, Ilnaz	Prof. Algergawy
KG construction from both	Natesh Vijay, Chirag	Asha

Critical Evaluation of the papers

This section is based on [KG21]

Paper Overview

- **Overview** on structures, applications, and related concepts of KGs,
- Presents **advanced concepts** in depth (inference, embedding, GNN, ML, ...)
- **Meta-analysis** of 13 external papers and books

Quality / Readability of the Paper

- **Extended online version** available with **concrete examples** on GitHub
- **Complexity** of topics may challenge readers new to the field
- Some small **mistakes**, missing KG construction

Overall Assessment

- **Significant** contribution to the field, **both intro and overview**
- Thorough and detailed examination of KGs
- Useful for researchers and practitioners alike

Paper Overview

- Need for more **open-source toolsets** for KG development
- Importance of **data / metadata** management
- Discusses KG construction **pipelines**

Quality / Readability of the Paper

- **Strength:** Clear, well illustrated, complete
- **Weakness:** Lack practical examples / case studies
- **WARN:** Still a preprint ! No guaranty of peer review

Overall Assessment

- **Valuable contribution** to the field of KG construction
- Successfully identifies the current state of KG construction and Valuable resource for researchers and
- highlights areas for **future research** and improvement

[KGKE22] (3.2 only)

Paper Overview

- Evolution of KE since the 1980s.
- **4 distinct periods** with their consequences summarized

Quality / Readability

- **Clarity:** well-structured, presents a clear progression of ideas
- **Figures:** Good visualisation, simplicity
- **No applications,** more for general knowledge about the field

Overall Assessment

- **Relevance:** valuable insights into the history and evolution of KE
- **Contribution:** contributes to the understanding of the changing requirements and challenges in KE over time.
- **Implications:** Raises questions about the future of KE

Conclusion

Knowledge Graphs

- Powerful tool for **modeling**, storing, **organizing**, and accessing complex information
- Represents data as a network of interconnected **nodes** and **edges**
- **Applications** across many industries and domains
- Lot's of **research** and techniques has been develop to **exploit its properties**
- Some challenges still remain

Source: [KG21]

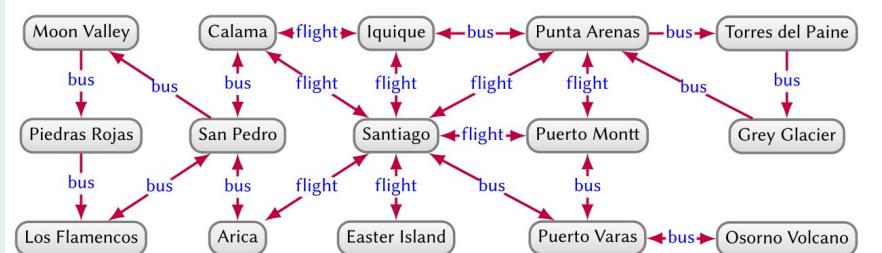


Fig. 15. Data graph representing transport routes in Chile.

Sources & References

Main source:

[KG21] <ACM Compt.> Aidan Hogan u. a. „Knowledge Graphs“. In: ACM Comput. Surv. 54.4 (Juli 2021). issn: 0360-0300. doi: 10.1145/3447772. url: <https://doi.org/10.1145/3447772.3>

Additional sources:

[KGKE22] <Dagstuhl Reports> Paul Groth u. a. „Knowledge Graphs and their Role in the Knowledge Engineering of the 21st Century“. In: Dagstuhl Reports 12.9 (2022). Report from Dagstuhl Seminar 22372, S. 60–120. doi: 10.4230/DagRep.12.9.60. Specific usage: pp. 60-72, Subsection "3.2 A Brief History of Knowledge Engineering: A Practitioner's Perspective", doi: 10.4230/DagRep.12.9.60.

[CKG23] <preprint> Marvin Hofer u. a. „Construction of Knowledge Graphs: State and Challenges“. In: arXiv preprint arXiv:2302.11509 (2023). url: <https://doi.org/10.48550/arXiv.2302.11509>.

NB: Images with no source are personal ones

Additional References

[TDKG16] L. Ehrlinger and W. Wöß, Towards a Definition of Knowledge Graphs., in: SEMANTiCS (Posters, Demos, SuCESS), 2016.

[ISIFSCD73] E.W. Schneider, Course Modularization Applied: The Interface System and Its Implications For Sequence Control and Data Analysis. (1973).

Online resources for additional images

<https://math.stackexchange.com/questions/1173328/eulers-solution-of-seven-bridges-of-k%C3%B6nigsberg-in-layman-terms>

https://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

<https://www.learner.org/courses/mathilluminated/units/11/textbook/02.php>

https://www.storyofmathematics.com/16th_tartaglia.html

<https://www.aboutamazon.com/news/innovation-at-amazon/making-search-easier>

And Wikipedia for images

<https://cs.mcgill.ca/~wlh/comp551/slides/25-gnns.pdf>

<https://en.wikipedia.org/wiki/Encyclop%C3%A9die>

<https://towardsdatascience.com/graph-theory-and-data-science-ec95fe2f31d8>

<https://www.stateofdigital.com/search-in-the-knowledge-graph-era/>

Remark: Wikipedia has been used **only** as a source for illustration and images



Q&A

Thanks for your attention!