Seminar: Knowledge Graphs

Alsayed Algergawy

Module: 5369S

Welcome!!

- Prof. Dr. Ing. Alsayed Algergawy
  - Data and knowledge engineering
- Senior researcher and data manager at Heinz-Nixdorf Endowed Chair for Distributed Information Systems
  - Friedrich-Schiller-University Jena
  - Prof. Dr. Birgitta König-Ries

Welcome!!

- Research Interests:
  - Schema/data integration
  - Schema/ontology matching
  - Semantic data integration
  - Data and knowledge engineering
- Room: HK30 209
- alsayed.algergawy@uni-passau.de

https://www.fim.uni-passau.de/dke/team

Welcome!!

Name: Vishvapalsinhji Parmar

Education: M.Sc. Computer Science

Occupation: Research Assistant at the Chair of Data and Knowledge Engineering

Office: HK30 – 211

E-Mail: Vishvapalsinhji.Parmar@uni-passau.de

Research interests:

- **Big Data Management**
- **Natural Language Processing**
- **Information Retrieval**
- **Deep Learning**

Welcome!!

Asha Mannarapotta Venugopal (M.Sc.)

asha.mannarapottavenugopal@uni-passau.de

Research Areas:

- Synthetic Data Generation
- Generative Adversarial Network (GAN)
- Synthetic Minority Oversampling Technique (SMOTE)
- Locality Sensitive Hashing (LSH)
- Data Anonymization: Masking & Generalization

And you....??

Agenda

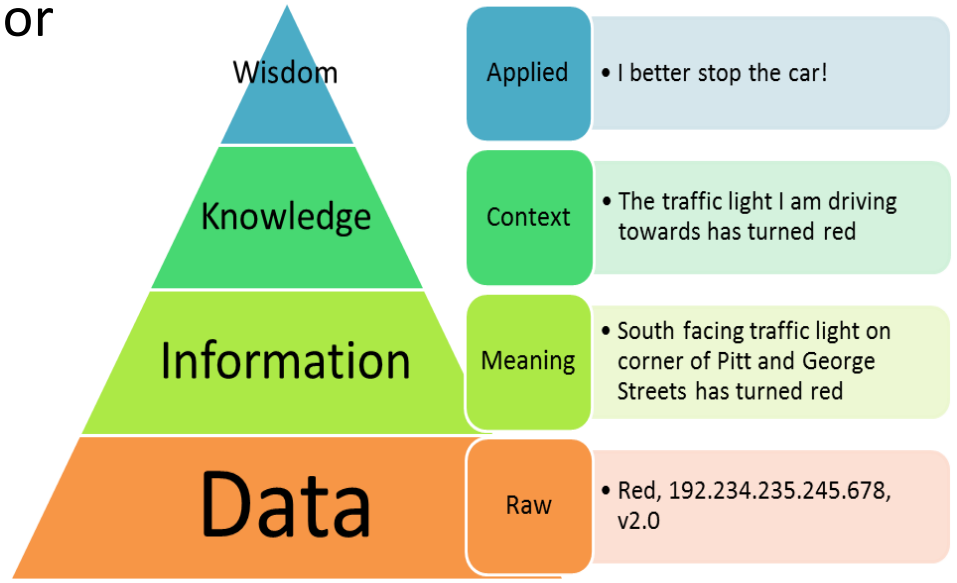
- Introduction
- Seminar rules and requirements
- Topics
- Topic Registration

Agenda

- Introduction
- Seminar rules and requirements
- Topics
- Topic Registration

Knowledge

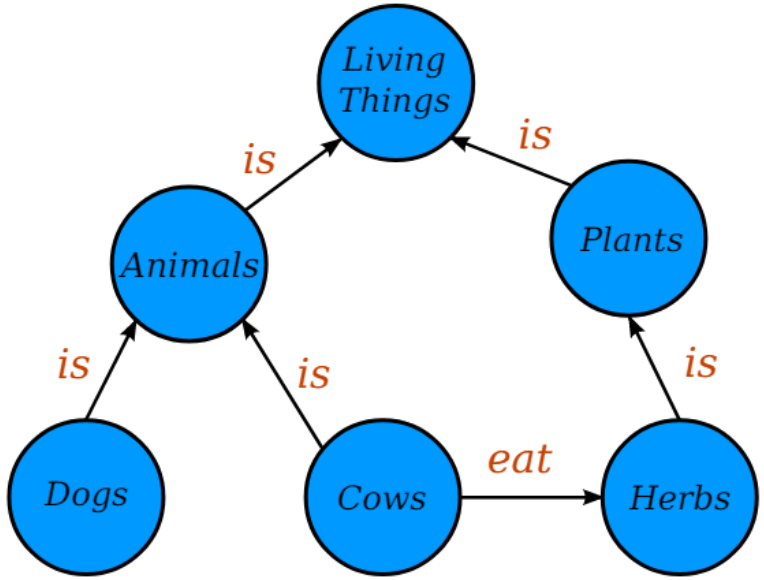
- What is knowledge?
  - "Knowledge is a familiarity, awareness or understanding of someone or something, such as facts, information, descriptions, or skills, which is acquired through experience or education by perceiving, discovering, or learning" (Wikipedia)
- The Breadth Hypothesis: "To behave intelligently in unexpected situations, an agent must be capable of falling back on increasingly general knowledge."[1]
- Data, information, knowledge



[1]Taken from: Lenat & Feigenbaum, Artificial Intelligence 47 (1991), \On the Threshold of Knowledge"

Graphs

- Data is increasingly connected (graph data)
- Relationships within data are often themselves an integral part of the data
- The world is structured: we are surrounded by **entities** connected by **relations**
- Graphs are a natural abstraction that captures the relationships between entities

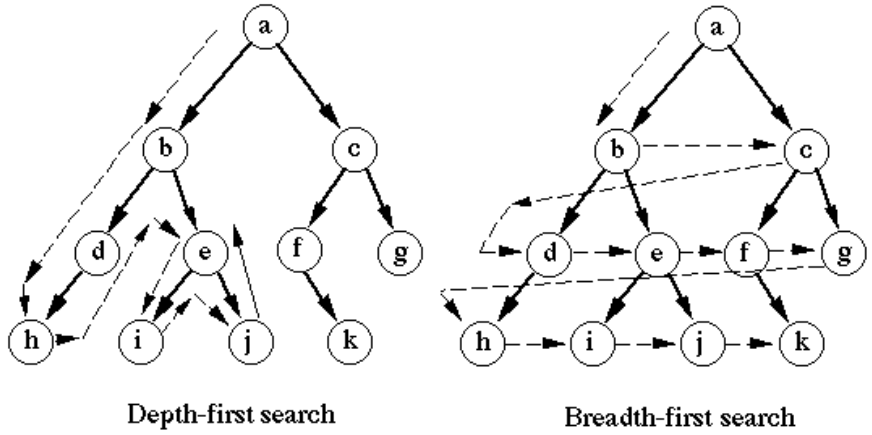


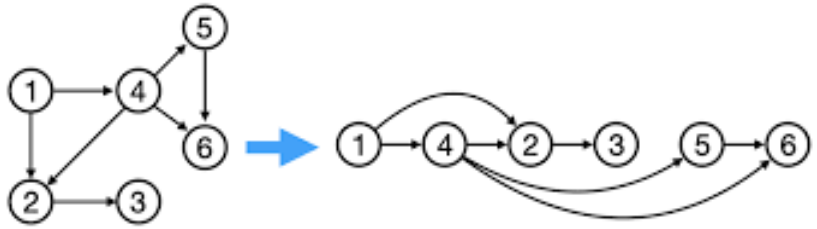
Graphs

- Data is increasingly connected (graph data)
- Relationships within data are often themselves an integral part of the data
- The world is structured: we are surrounded by **entities** connected by **relations**
- Graphs are a natural abstraction that captures the relationships between entities
- Web Graphs: trillions of nodes and edges
  - 4.7 billion web pages and 8 billion links
- Social Graphs
  - **Facebook**: 2.963 billion monthly active users with hundreds of billions of relationships (as of January 2023)
  - **Twitter**: 450 million monthly active users

Graph Algorithms

- Traversal
  - Depth first search
  - Breadth first search
- Topological sort
- Minimum spanning trees
- Shortest paths
- Connected components (strong/weak)
- Graph matching





http://faculty.simpson.edu/lydia.sinapova/www/cmsc250/LN250_Weiss/L27-GraphSummary.htm
https://web.stanford.edu/class/cs97si/06-basic-graph-algorithms.pdf
https://www-users.cs.umn.edu/~karypis/parbook/Lectures/AG/chap10_slides.pdf

Knowledge Graphs (KG)

- Set of entities

Knowledge Graphs (KG)

Type: city
Area : 114.30 km^2
Pop: 109.527
Name: Jena

ID4

- Set of entities, labeled with attributes/ properties

Type: Person
Name: B. König-Ries
Title: Prof

ID2

Type: Person
Name: A. Algergawy
Title: Dr. Ing

ID1

Type: University
Founded : Feb. 2, 1558
Students: 18.916

ID3

Type: research group
Name: fusion
Spec: distributed IS

ID5

Knowledge Graphs (KG)

- Set of entities, labeled with attributes

Type: city
Area : 114.30 km^2
Pop: 109.527
Name: Jena

ID4:
(Q3150)

Type: Person
Name: B. König-Ries
Title: Prof

ID2:
(Q56476045)

Type: Person
Name: A. Algergawy
Title: Dr. Ing

ID1:
(Q107504296)

Type: University
Founded : Feb. 2, 1558
Students: 18.916

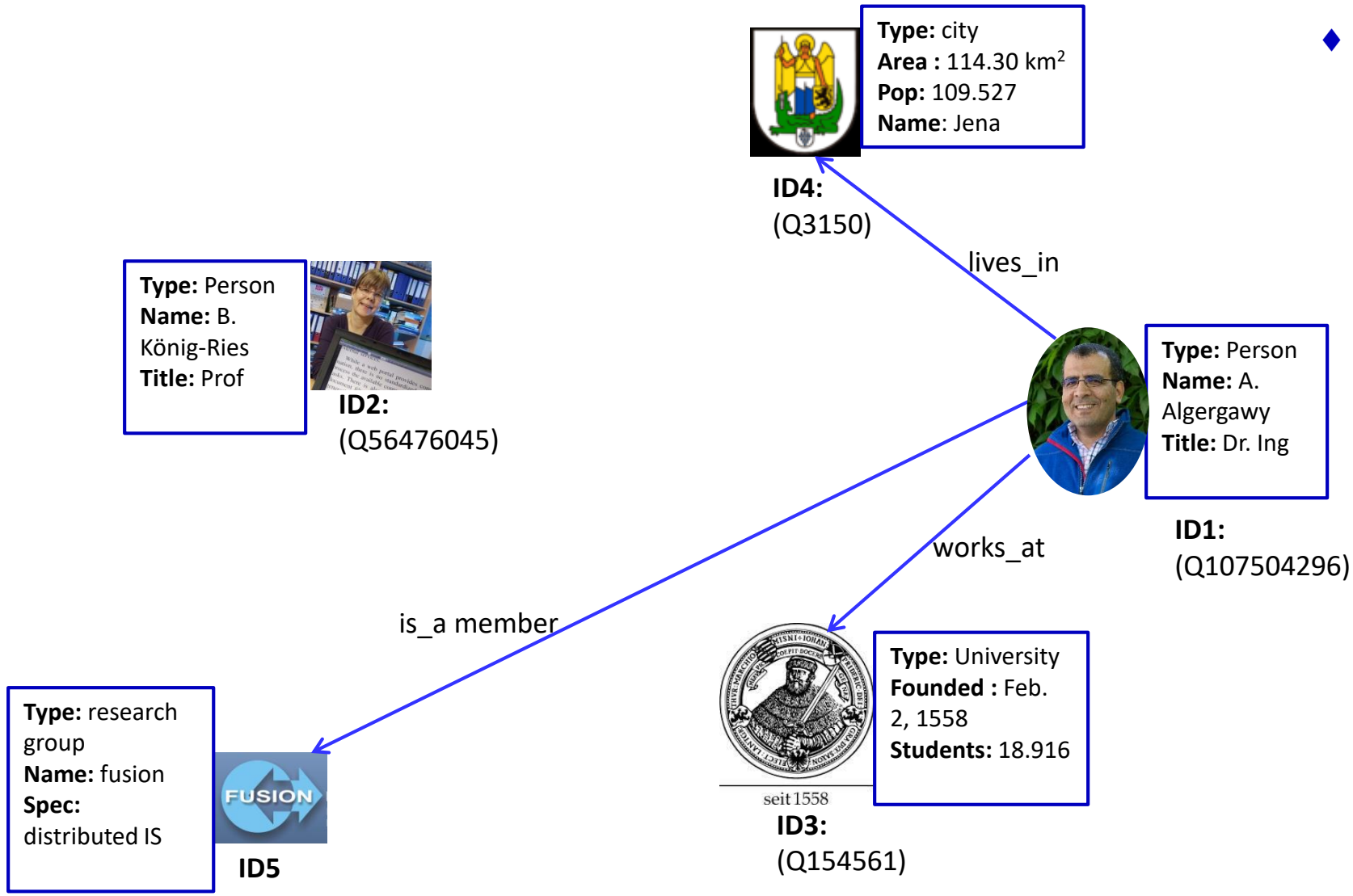
ID3:
(Q154561)

Type: research group
Name: fusion
Spec: distributed IS

ID5

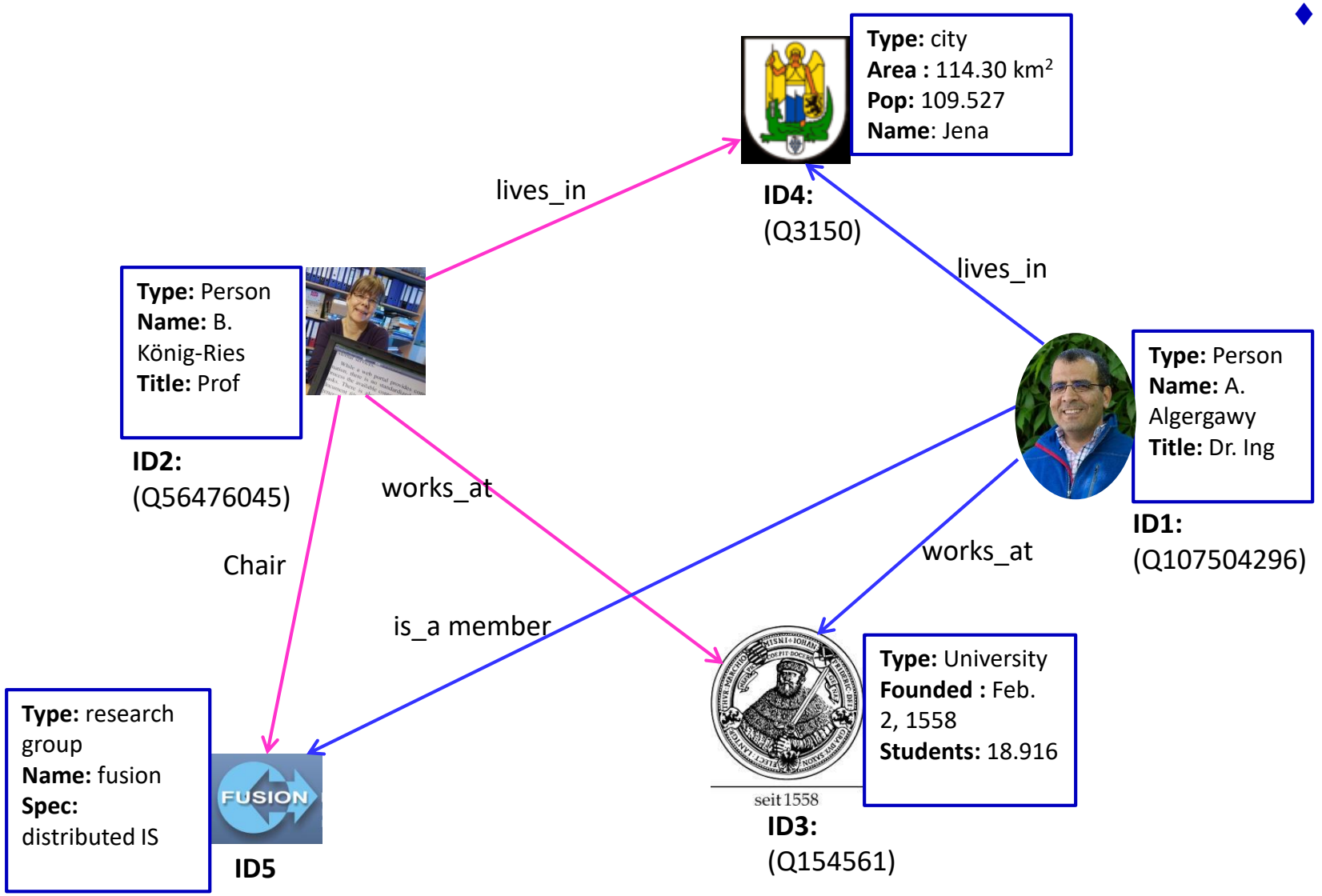
Knowledge Graphs (KG)

- Set of entities + relationships

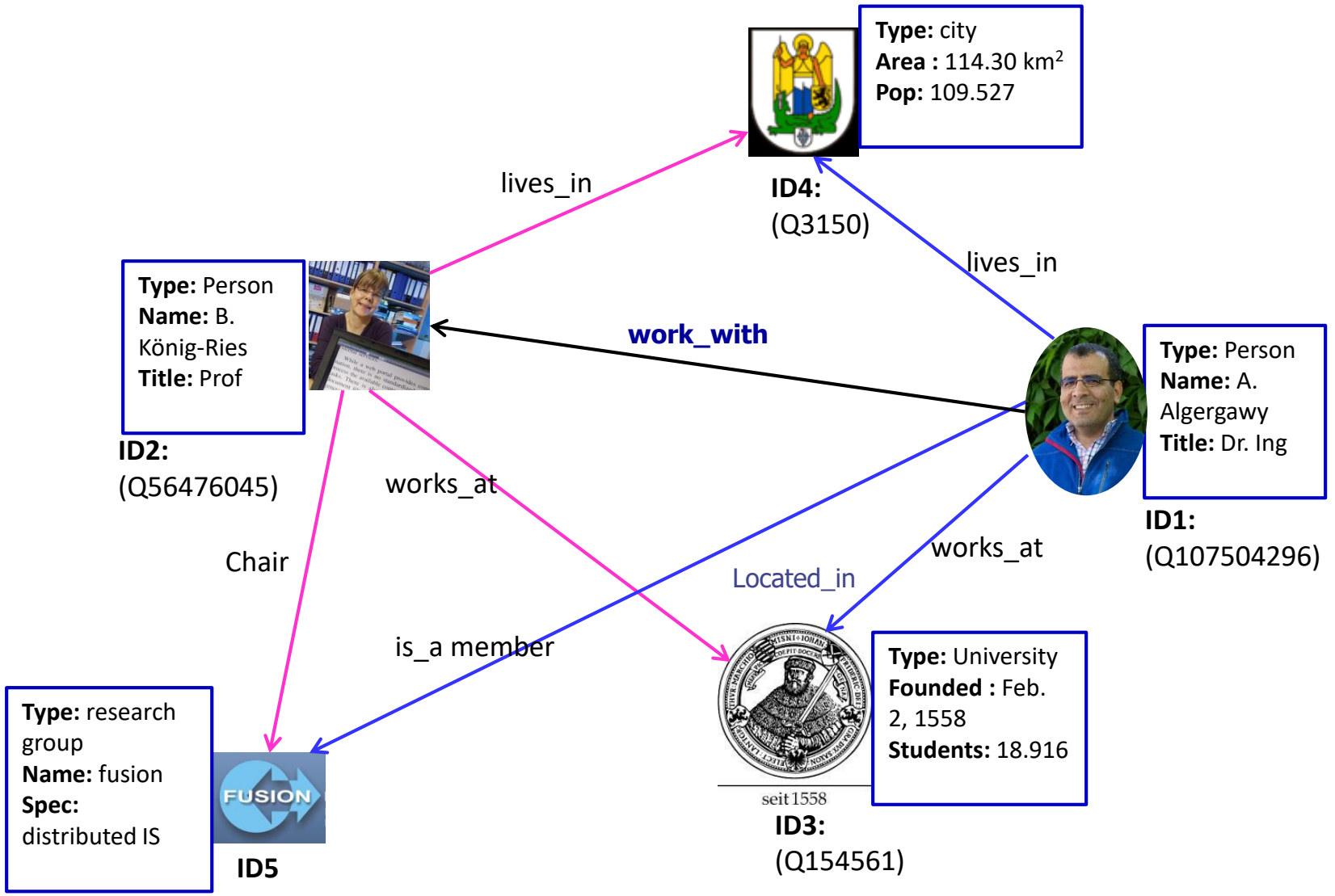


Knowledge Graphs (KG)

- Set of entities + relationships

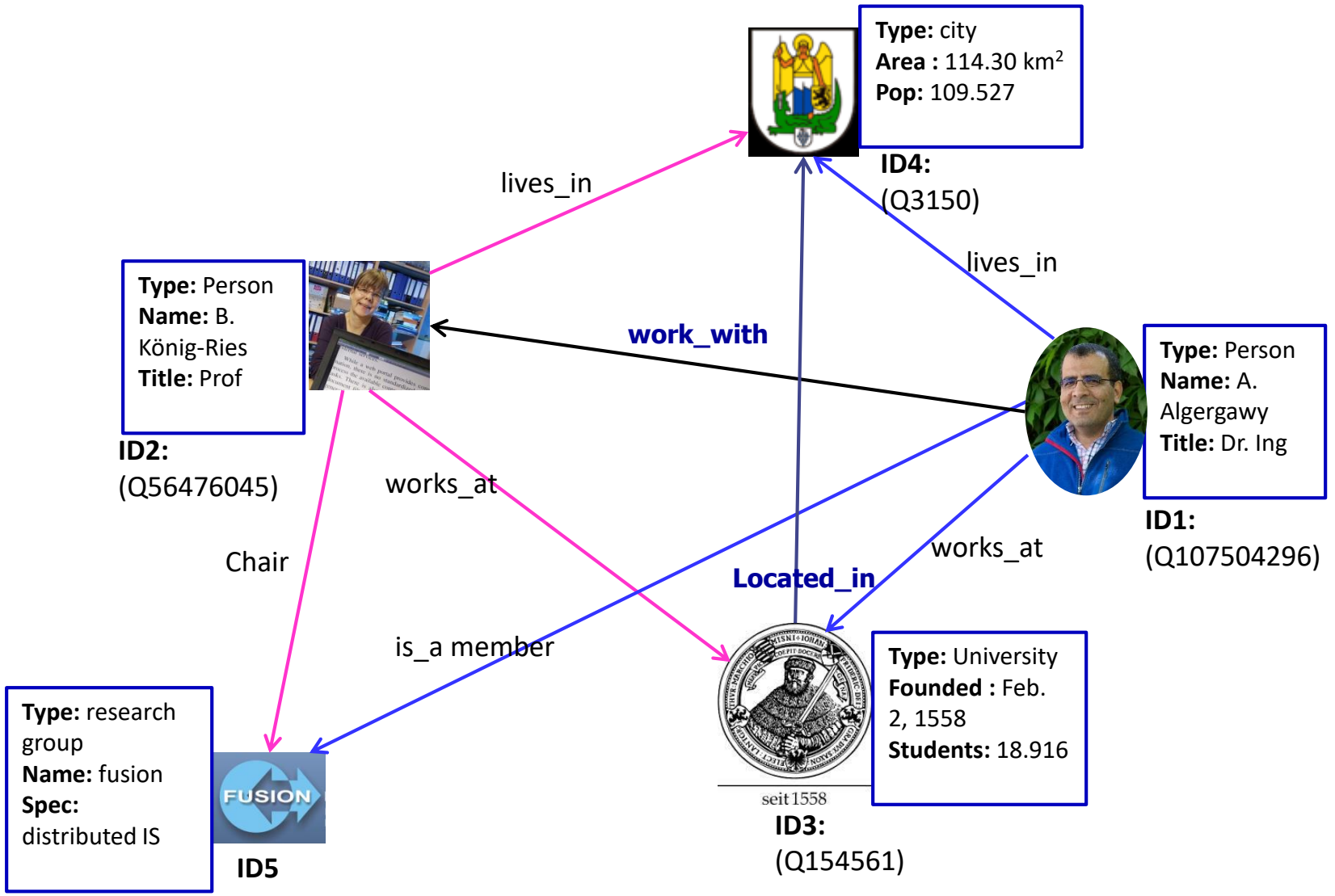


Knowledge Graphs (KG)



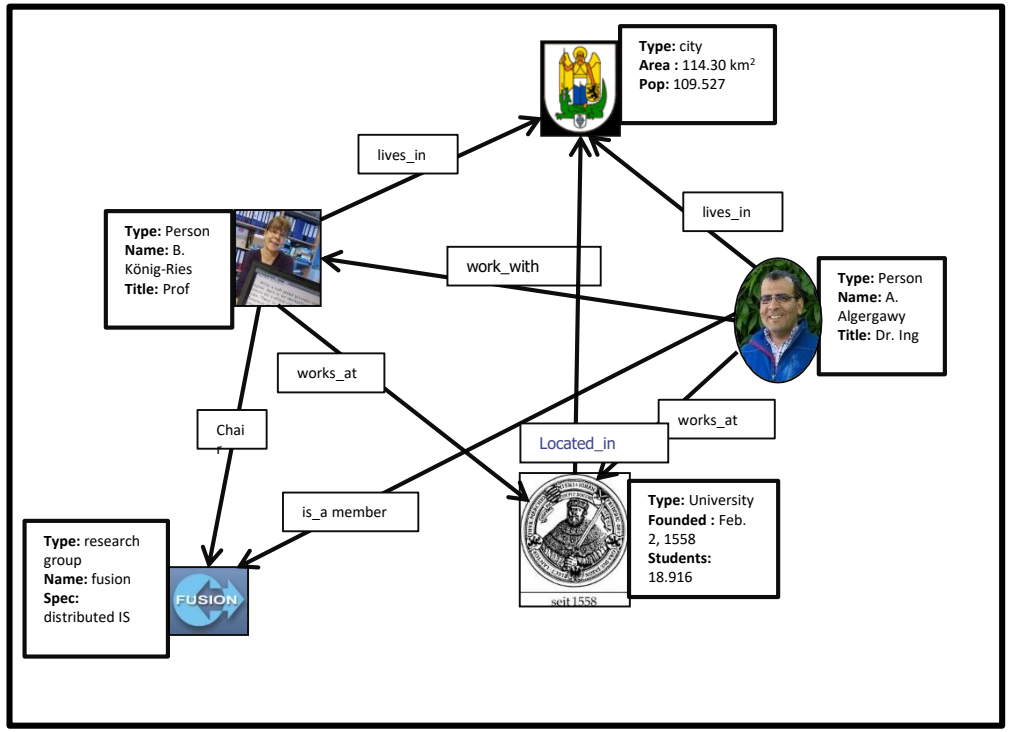
- Set of entities + relationships

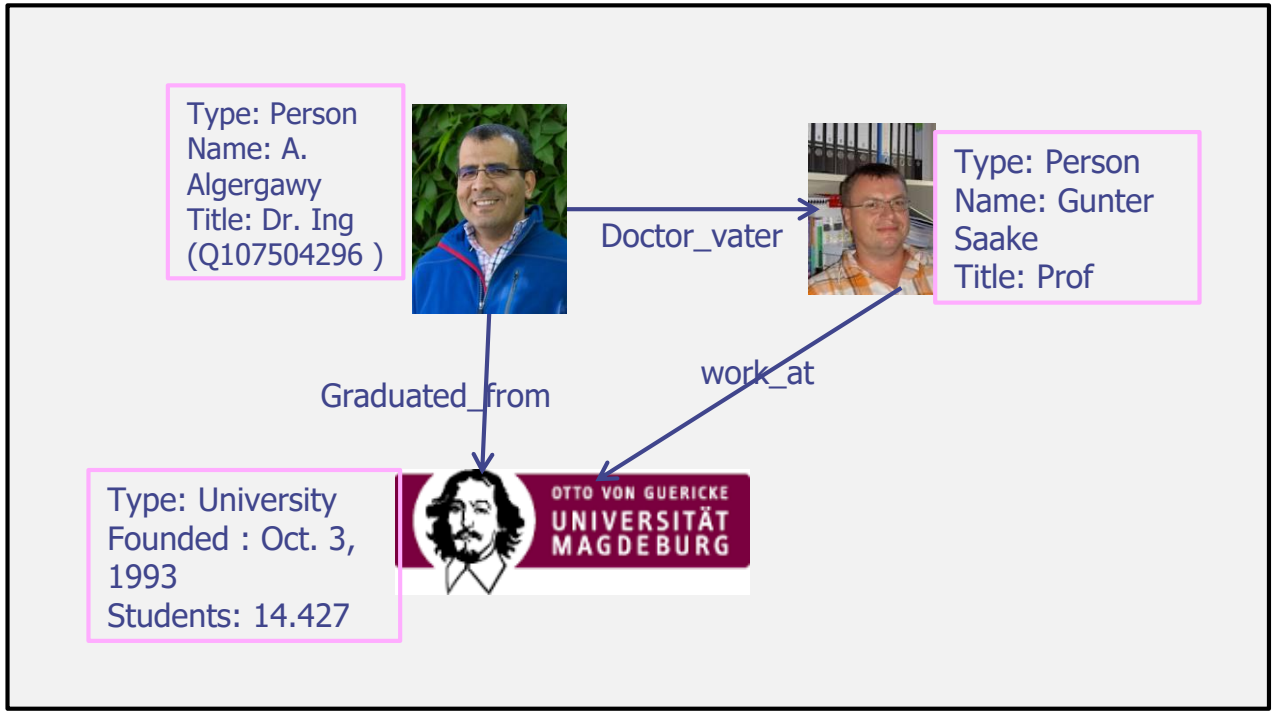
Knowledge Graphs (KG)

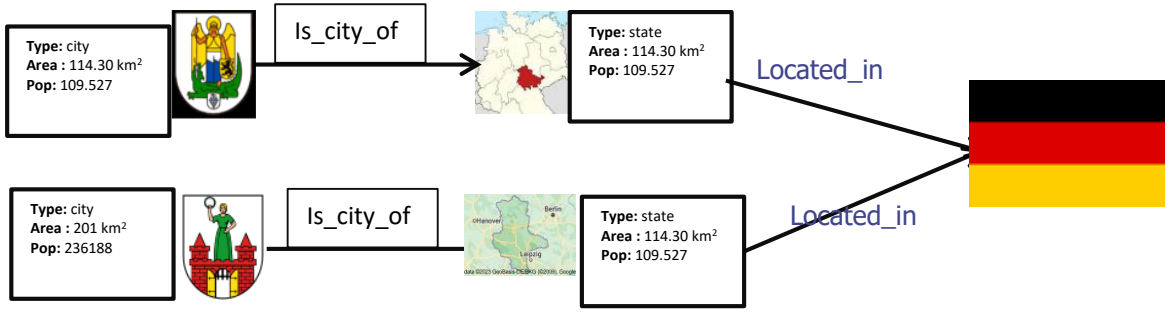


- Set of entities + relationships

Knowledge Graphs (KG)

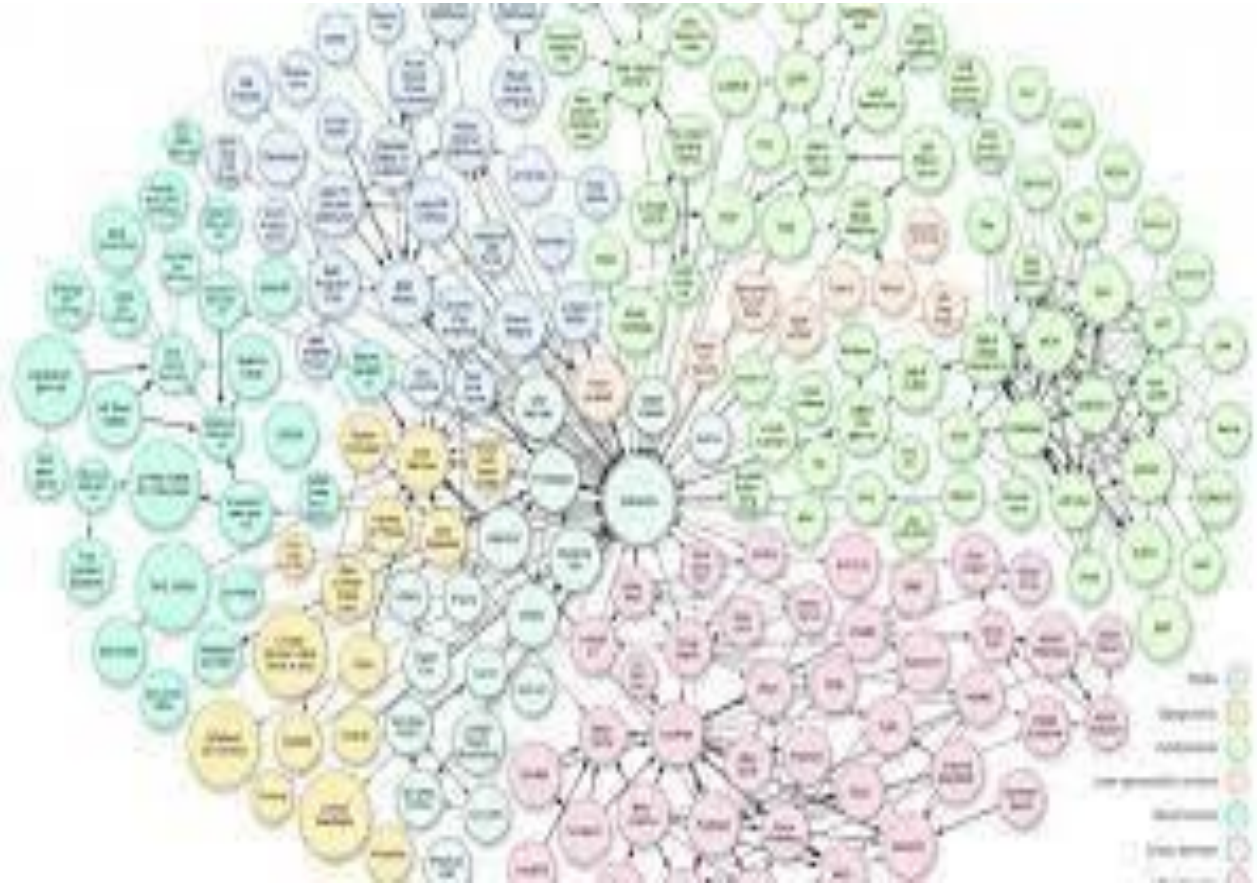






What is a Knowledge Graph?

- No clear (or generic) definition...
- Structured representation of knowledge
- Collection of *facts* about real-world concepts
- Represented as ***< s; p; o >*** triples
  - <FSU, *located_in*, Jena>
  - *S and o:* ***entities****, represented as nodes in the graph*
  - *p:* ***relation*** *between entities, represented as an edge in the graph*



What is a Knowledge Graph?

<ID1, has_name, “A. Algergawy”>

<ID1, is-a, Person>

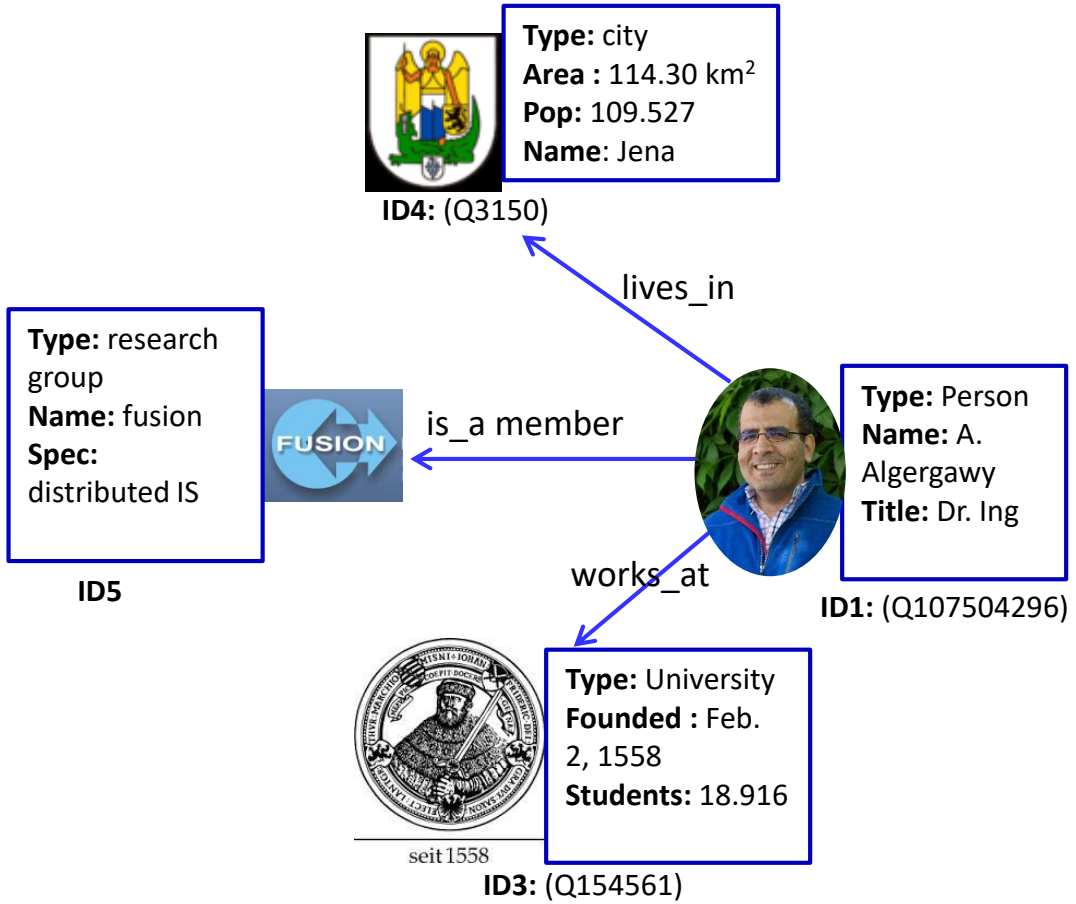
<ID4, is-a, City>

<ID4, has_name, “Jena”>

<ID1, lives_in, ID4>

<ID1, work_at, ID3>

<ID3, locate_in, ID4>



What is a Knowledge Graph?

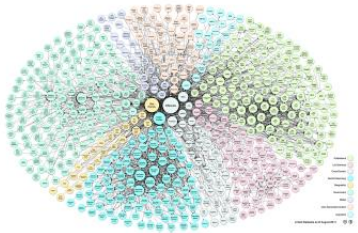
LinkedIn
Knowledge Graph

Freebase

Google

amazon

DBpedia



yago
select knowledge

Amazon
Product Graph

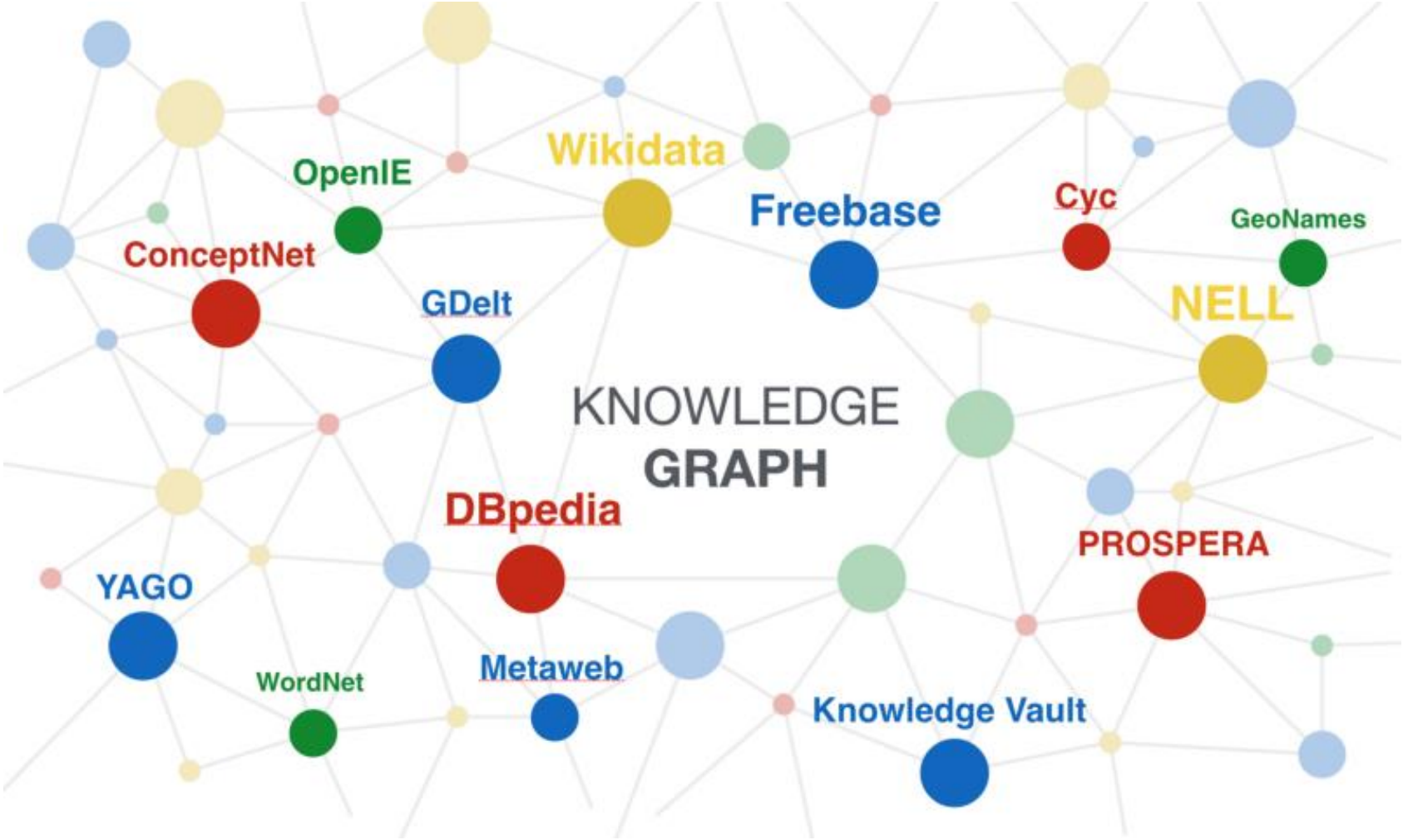
WIKIDATA

ELECTRONIC ARTS™

Facebook
Entity Graph

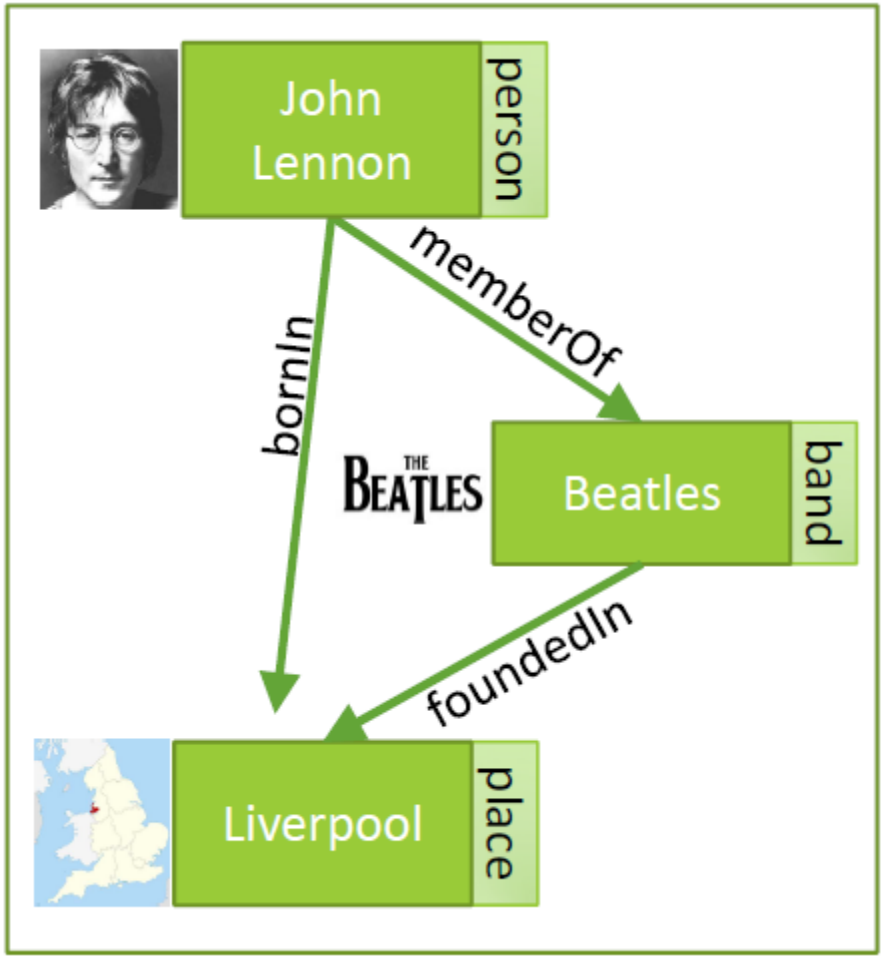
Microsoft
Satori

What is a Knowledge Graph?

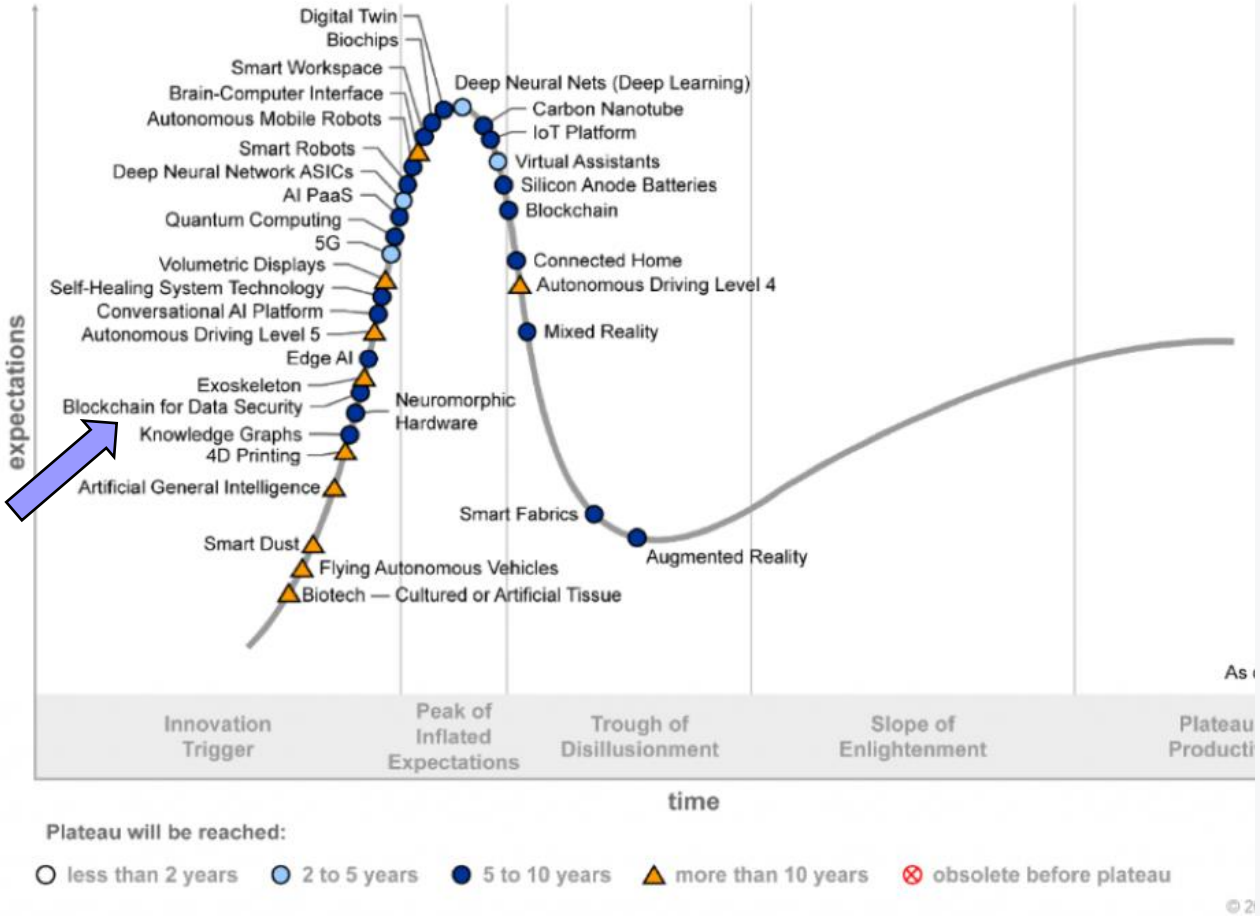


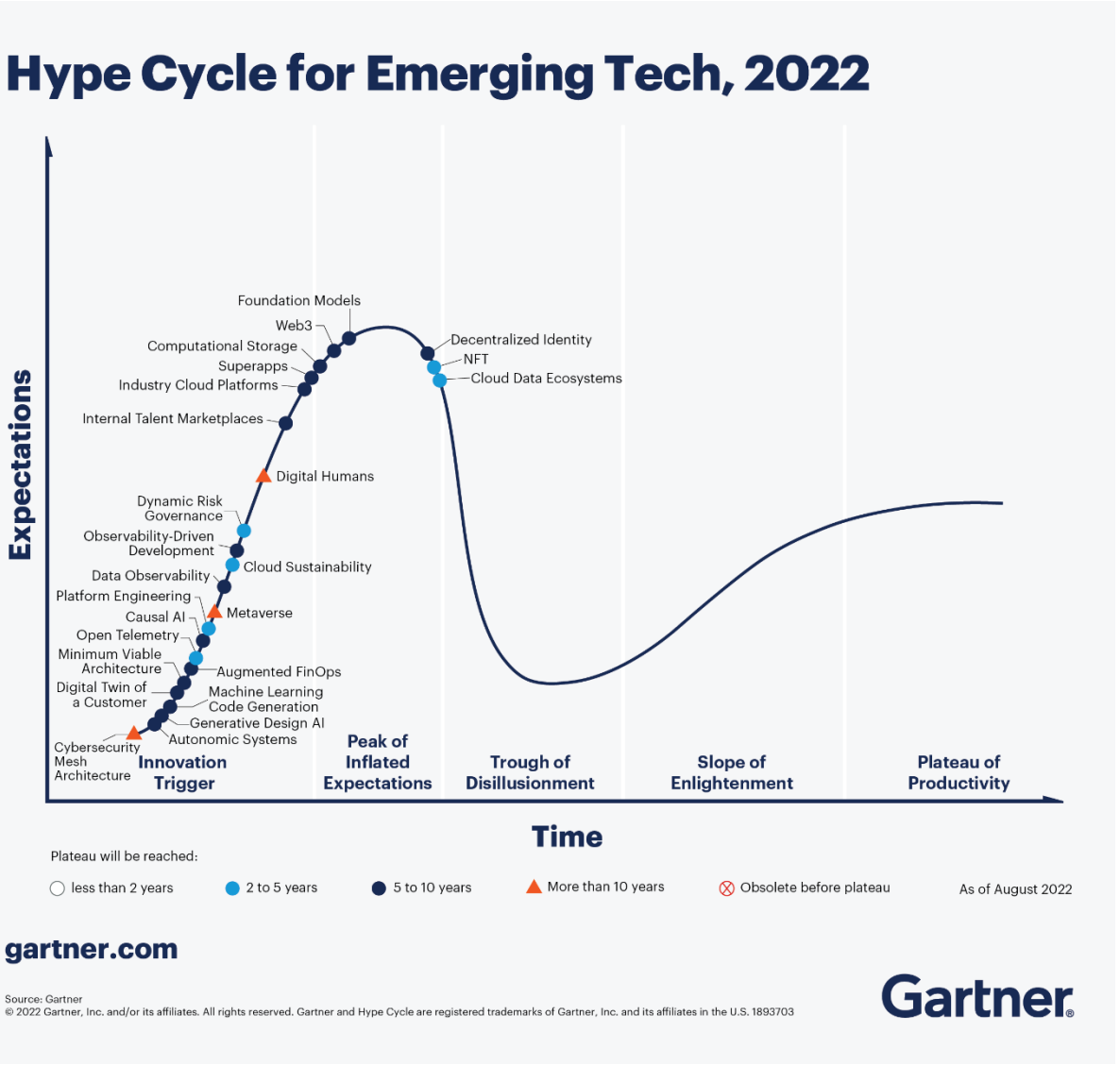
So, what is a Knowledge Graph?

- Knowledge in graph form!
- Captures entities, attributes, and relationships
- Nodes are entities
- Nodes are labeled with attributes (e.g., types)
- Typed edges between two nodes capture a relationship between entities



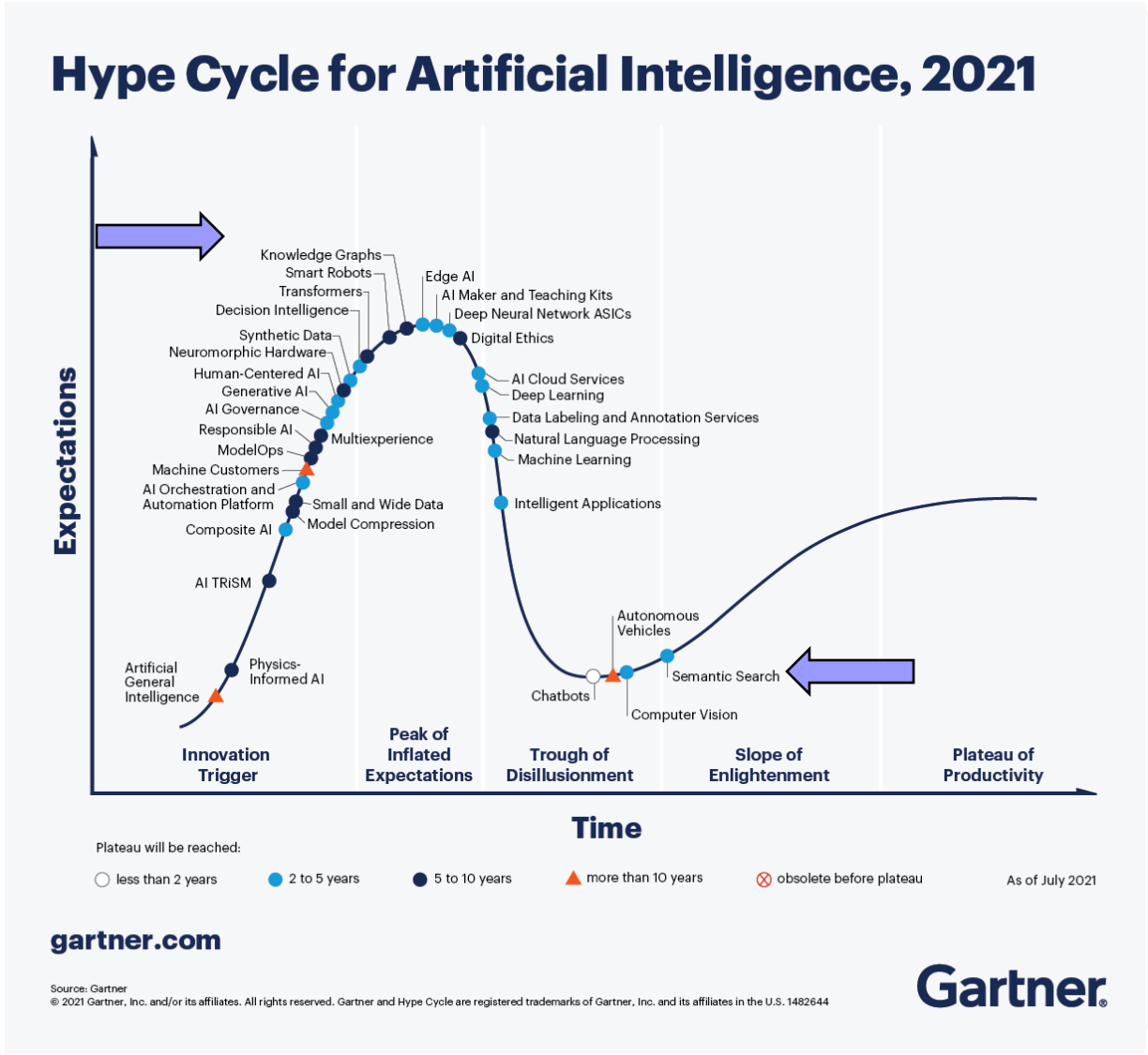
Why KGs?

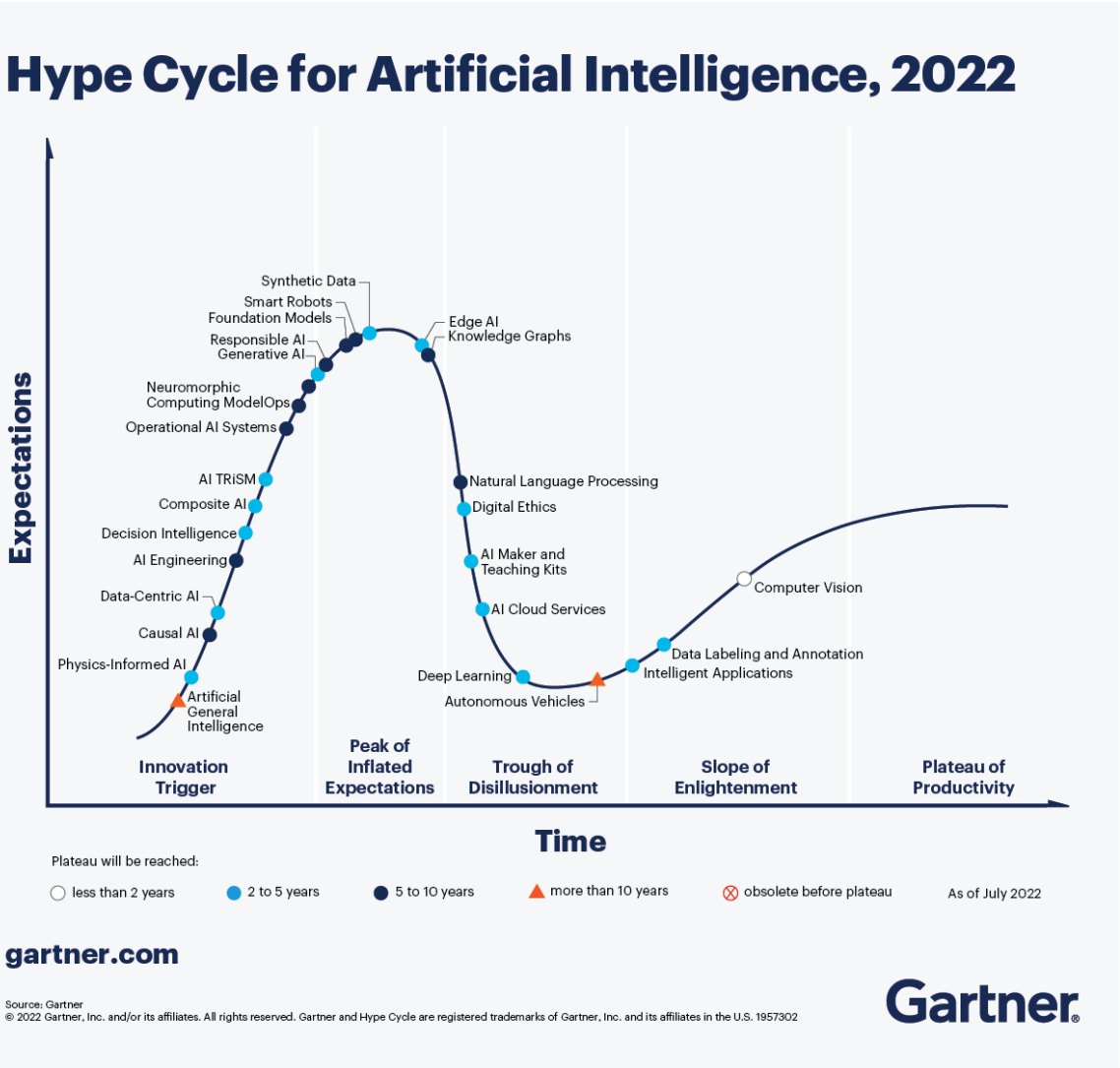




https://www.criticallink.com/2018/11/gartners-latest-emerging-technologies-hype-cycle/

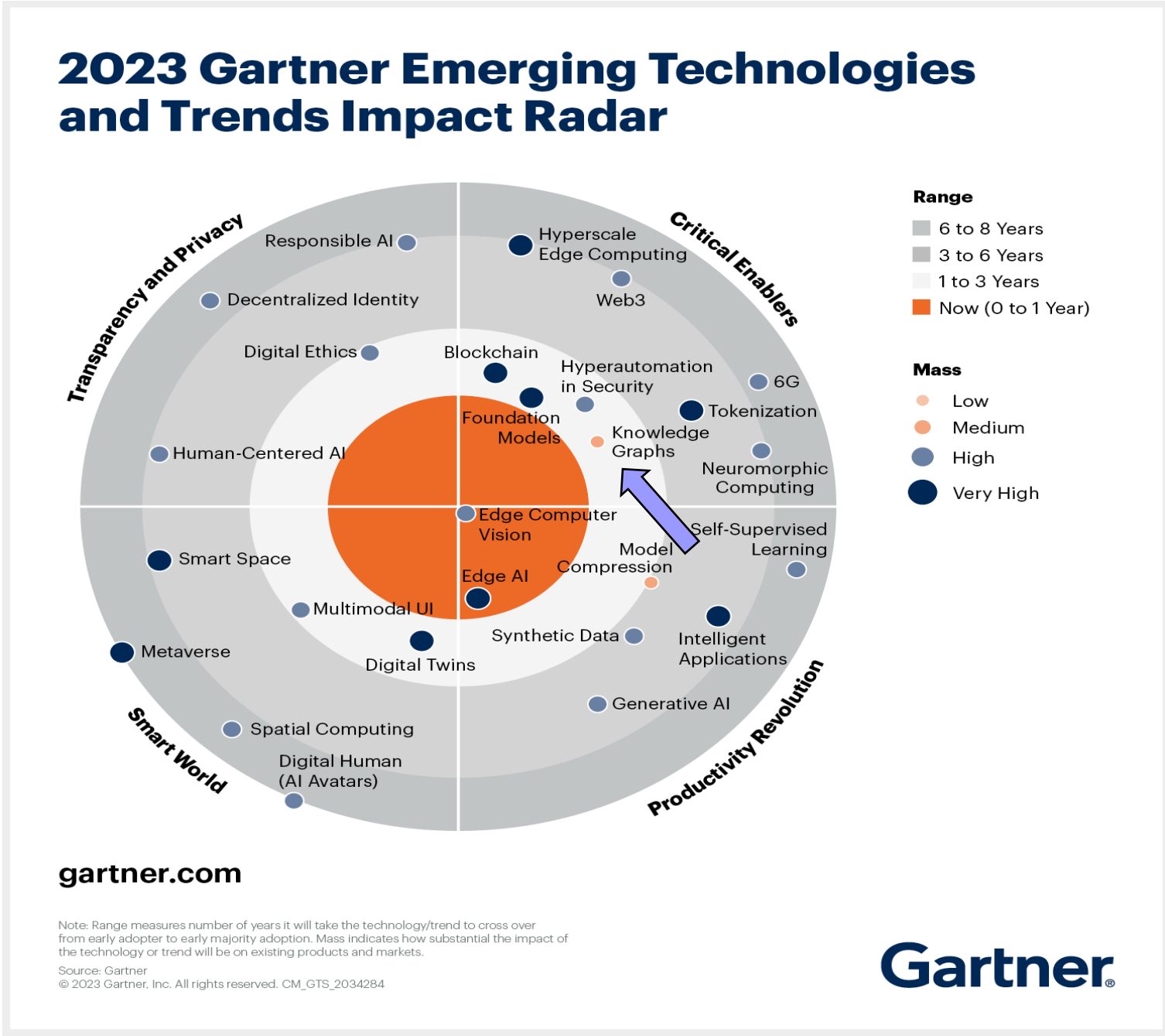
Why KGs?





https://www.gartner.de/de/artikel/die-4-wichtigsten-trends-im-gartner-hype-cycle-fuer-ki-2021

Why KGs?

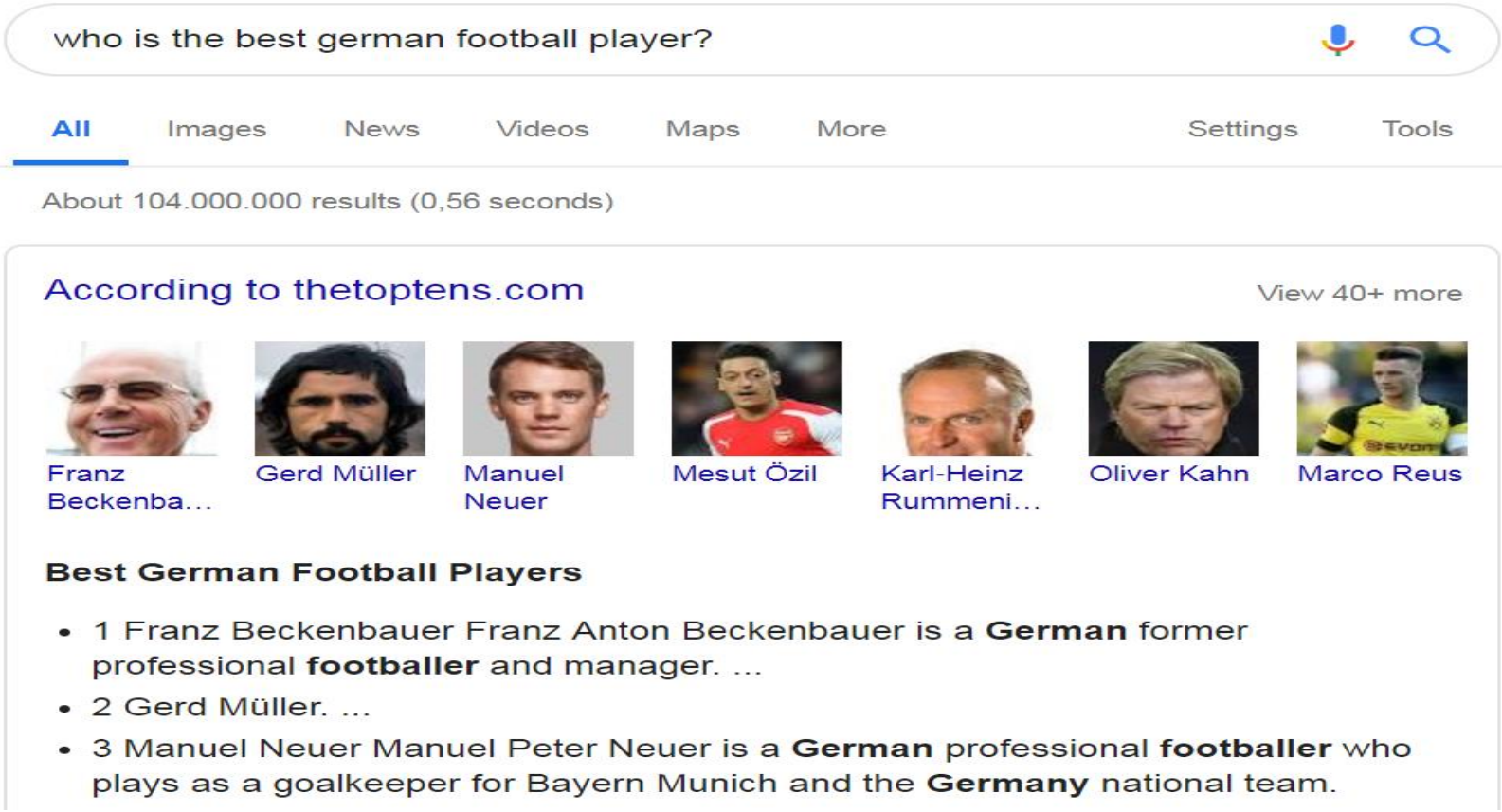


https://www.gartner.com/en/articles/4-emerging-technologies-you-need-to-know-about

Have you used KGs?

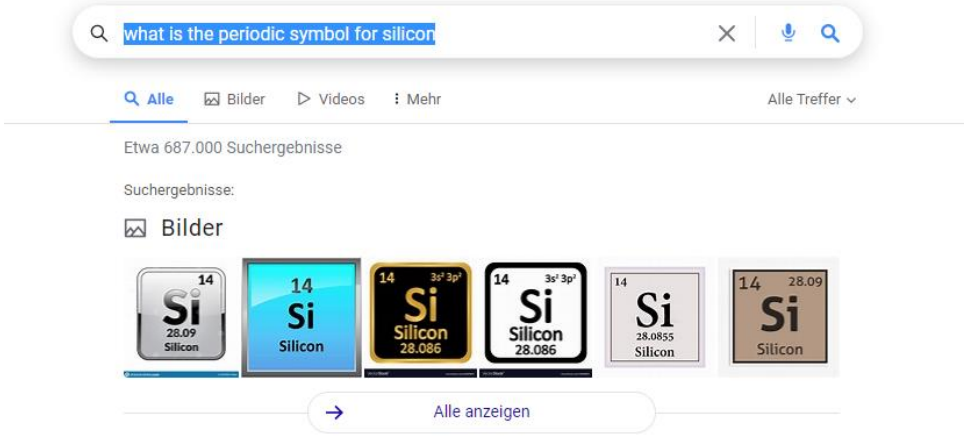
What are KGs used for?

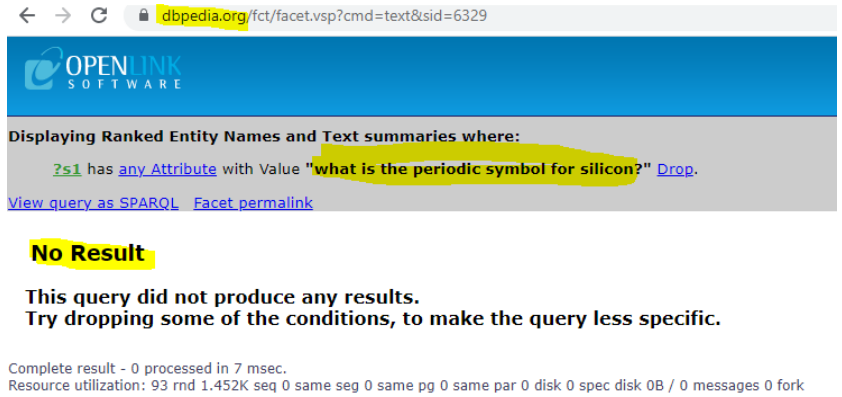
- Query Answering

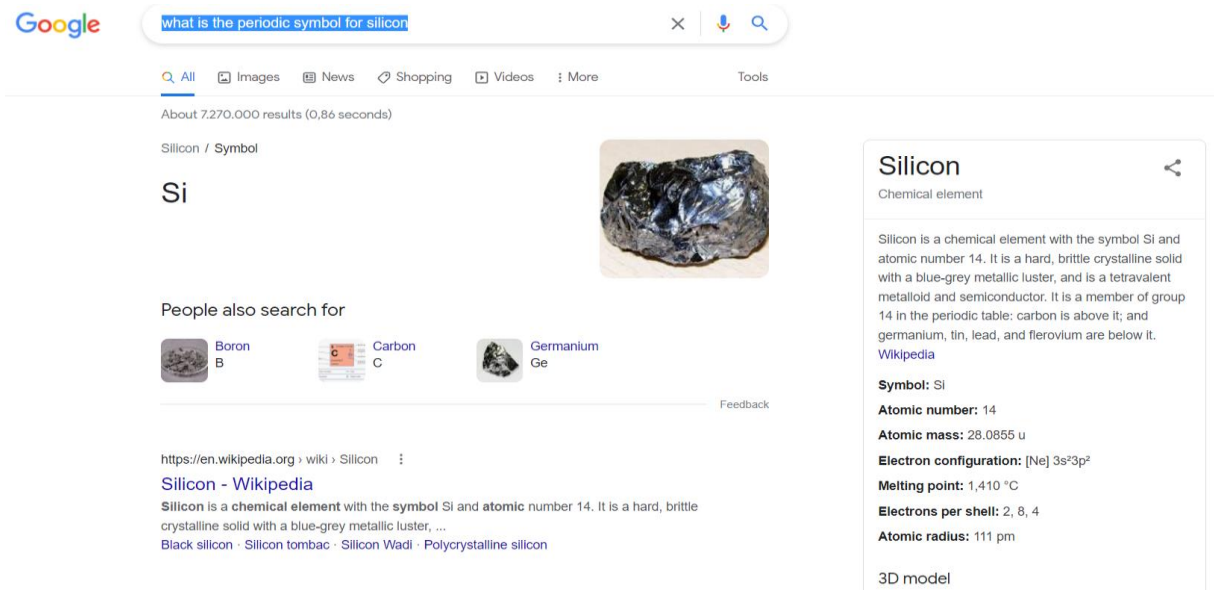


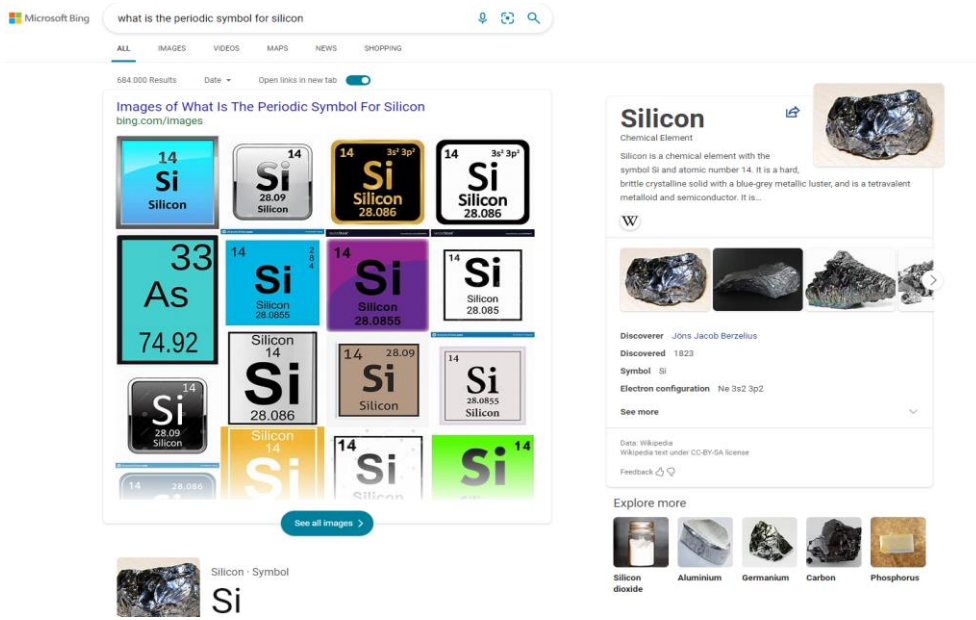
What are KGs used for?

- what is the periodic symbol for silicon?



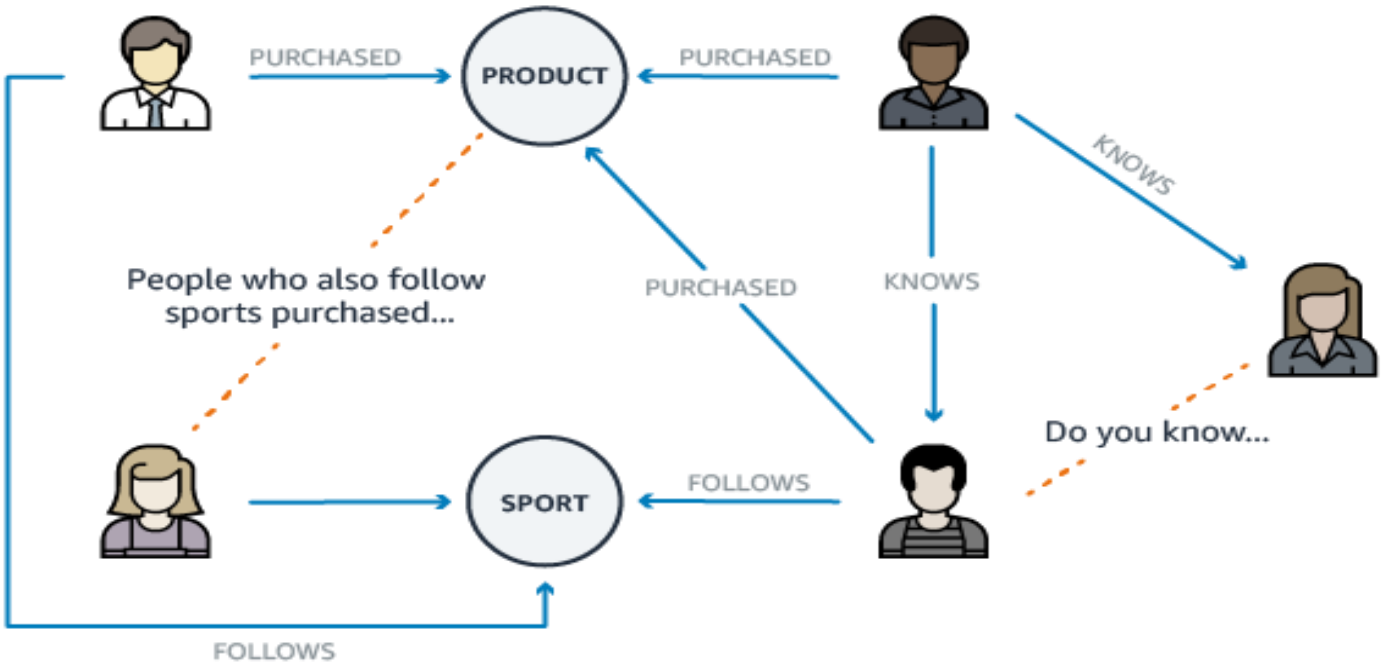






What are KGs used for?

- Recommendation



Source: https://geomarketing.com

Agenda

- Introduction
- Seminar organization and rules
- Topics
- Topic Registration

Seminar organization

- Module number: 5369S/ SoSe23
- 2 SWS/ 5 ECTS points
- When/where: Wed., 19.04.2023 08:00 - 10:00 Uhr,
  - Room: (HK 28) SR 101

Seminar organization: schedule

Week	Topic
Wed. 19.04.2023	Introduction
Wed. 26.04.2023	Topic selection/Monitoring
Wed. 03.05.- 31.05.2023	Monitoring/presentation/report preparing
Wed. 07.06.2023	Student presentations (3 talks)
Wed. 14.06.2023	Student presentations (3 talks)
Wed. 21.06.2023	Student presentations (3 talks)
Wed. 28.06.2023	Student presentations (3 talks)
Wed. 05.07.2023	Student presentations (3 talks)
Wed. 12.07.2023	Student presentations (3 talks)
Wed. 19.07.2023	Student presentations (3 talks)
Wed. 26.07.2023	Final report submission

Seminar rules

- Please pick One topic (for example):
  - What are Knowledge graphs (KGs)?
  - How to construct KGs (text, semi-structure, table, csv,…)
  - How to evaluate KGs? KG quality?
  - KGs in different applications, e.g. agriculture, medicine, education, generic
  - Examples of KGs
  - KGs embedding
  - KGs alignment

Seminar rules

- For the selected topic, use 2/3 papers (published in highly ranked conference, such as ISWC, ESWC, VLDB, WWW, CIKM or journals)
- The selected topic as well as the selected set of papers should be registered and sent through the email to your monitor
- In a case having questions, simply send and email or ask for a meeting

Seminar rules

- The seminar contains two main parts: presentation and technical report (implementation/deployment one of the related system is plus)
- Presentation
  - 20 minutes talk in English
  - Around 10 minutes of Q&A
  - Talk regularly to your instructor/tutor, especially before your talk
  - **Prepare your own slides** (power point or latex code must be sent to your tutor)

Guidelines for presentation

- It is important to **clearly introduce** the problem and the idea presented in the papers
- In the papers look for:
  - **Contributions** of the paper
  - Improvements to the **state-of-the-art**
  - Main **results**
  - **Conclusions** and **future work**
- Discuss the **insights** that are provided in the papers
- Identify **strengths** and **weaknesses**, question the **assumptions**, criticize the bad **decisions** in the papers

Report

- Up to **3-5 pages** using your own template
- Maximum of **one week** after the end of presentations
- Contents of the report:
  - include the **basic idea** presented in the papers
  - **summarize** the papers
  - include the points raised in the seminar by the moderator (collaborate with the opponents if necessary)
  - Identify **strengths** and **weaknesses**, question the **assumptions**, criticize the bad **decisions** in the papers
  - include the **important results** and **conclusions**
  - **bonus points** if you include the content from the papers outside of your assigned papers (such as any follow up works, new results etc)

Moderators

- One **moderator** per talk
  - **introduces** the speaker and topic
  - **moderates** the Q&A session
  - must **read the papers thoroughly**
    - prepare **questions**
    - **challenge** the ideas and results in the papers (if there are any weaknesses)

Grading

- Presentation 40%
- Report 30%
- Your performance as a moderator 15%
- Active participation in the seminars 15%

Agenda

- Introduction
- Seminar organization and rules
- Topics
- Topic Registration

Topics

- What is a KG? -- 1
- KG construction from
  - Text --2
  - Table --3
  - Mixed --4
- KG search
  - Keyword over KG --5
  - Question answering over KG--6
  - ChatGPT for KG –7
- KG schema
  - Ontology design for KG --8
- KG alignment---9
- KG embedding ---10
- KG for ML ---11
- KG applications:
  - KG for health --- 12
  - KG for agriculture --- 13
  - KG for energy ---14
  - KG for biodiversity ---15
- KG frameworks:
  - Google KG --- 16
  - Amazon PKG --- 17
  - Dbpedia ---- 18
  - Wikidata --- 19

Agenda

- Introduction
- Seminar organization and rules
- Topics
- Topic Registration

Topic registration

- Select your topic from the poll
- Next week will revise selection and monitoring

Reference

- Hogan, Aidan, et al. "Knowledge graphs." ACM Computing Surveys (CSUR) 54.4 (2021): 1-37
- Ji, Shaoxiong, et al. "A survey on knowledge graphs: Representation, acquisition, and applications." IEEE transactions on neural networks and learning systems 33.2 (2021): 494-514.
- Zeng, Xiangxiang, et al. "Toward better drug discovery with knowledge graph." Current opinion in structural biology 72 (2022): 114-126.
- Tiddi, Ilaria, and Stefan Schlobach. "Knowledge graphs as tools for explainable machine learning: A survey." Artificial Intelligence 302 (2022): 103627.
- Bonner, Stephen, et al. "A review of biomedical datasets relating to drug discovery: a knowledge graph perspective." Briefings in Bioinformatics 23.6 (2022).
- Ryen, Vetle, Ahmet Soylu, and Dumitru Roman. "Building semantic knowledge graphs from (semi-) structured data: a review." Future Internet 14.5 (2022): 129.
- Li, Haotian, et al. "KG4Vis: A knowledge graph-based approach for visualization recommendation." IEEE Transactions on Visualization and Computer Graphics 28.1 (2021): 195-205.
- Qin, Hongchen, and Yiheng Yao. "Agriculture knowledge graph construction and application." Journal of Physics: Conference Series. Vol. 1756. No. 1. IOP Publishing, 2021.
- Shi, Longxiang, et al. "Semantic health knowledge graph: semantic integration of heterogeneous medical knowledge and services." BioMed research international 2017 (2017).
- Rossanez, Anderson, et al. "KGen: a knowledge graph generator from biomedical scientific literature." BMC medical informatics and decision making 20.4 (2020): 1-24.

Reference

- Piero Andrea Bonatti, Stefan Decker, Axel Polleres, Valentina Presutti: Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). Dagstuhl Reports 8(9): 29-111 (2018)
- Jay Pujara, Sameer Singh: Mining Knowledge Graphs From Text. WSDM 2018: 789-790 (https://kgtutorial.github.io/)
- Knowledge Graphs: In Theory and Practice. (http://sumitbhatia.net/source/knowledge-graph-tutorial.html)
- Enterprise Knowledge Graphs for Large Scale Analytics. (https://cci.drexel.edu/bigdata/bigdata2017/files/Tutorial1-1.pdf)
- Utilizing Knowledge Graphs in Text-centric Information Retrieval. (https://github.com/laura-dietz/tutorial-utilizing-kg)
- Natthawut Kertkeidkachorn, Ryutaro Ichise:
  T2KG: An End-to-End System for Creating Knowledge Graph from Unstructured Text. AAAI Workshops 2017
- Michael Färber, Achim Rettinger: Which Knowledge Graph Is Best for Me? CoRR abs/1809.11099 (2018)
- Michael Färber, Frederic Bartscherer, Carsten Menne, Achim Rettinger: Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. Semantic Web 9(1): 77-129 (2018)

Reference

- Heiko Paulheim: Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web 8(3): 489-508 (2017)
- Completeness and consistency analysis for evolving knowledge bases. Web Semantics: Science, Services and Agents on the World Wide Web 54 (2019) 48–71
- A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, and S. Auer: Quality Assessment for Linked Open Data: A Survey. SWJ 7(1), 2016
- Jay Pujara, Hui Miao, Lise Getoor, William W. Cohen: Knowledge Graph Identification. International Semantic Web Conference (1) 2013: 542-557
- Knowledge graph visualization: https://allegrograph.com/products/gruff/
- Ganggao Zhu, Carlos Angel Iglesias: Exploiting semantic similarity for named entity disambiguation in knowledge graphs. Expert Syst. Appl. 101: 8-24 (2018)
- Shreyansh P. Bhatt, Swati Padhee, Amit P. Sheth, Keke Chen, Valerie L. Shalin, Derek Doran, Brandon S. Minnery: Knowledge Graph Enhanced Community Detection and Characterization. WSDM 2019: 51-59
- Sahar Vahdati, Guillermo Palma, Rahul Jyoti Nath, Christoph Lange, Sören Auer, Maria-Esther Vidal: Unveiling Scholarly Communities over Knowledge Graphs. TPDL 2018: 103-115
- Angen Luo, Sheng Gao, Yajing Xu: Deep Semantic Match Model for Entity Linking Using Knowledge Graph and Text. IIKI 2017: 110-114