

Hellooo Computer? When Large Language Models meet RDF Knowledge Graphs

1



Hellooo computer?



MOVIECLIPS.COM

2

 zazuko

Scotty and the Computer in Star Trek IV

In this memorable scene, Montgomery "Scotty" Scott, the chief engineer from the Starship Enterprise, finds himself in the past, facing a 20th-century computer. Mistaking the mouse for a microphone, he humorously attempts to communicate verbally with it, illustrating the vast differences in technological interaction across eras. This scene playfully highlights the evolution of human-computer interaction and the ever-changing expectations of technology.

This was the Mac model used.

NOTE: There is an error in this paragraph above about the scene, typical for LLMs. Spot the error and explain why it might have happened.

UNDERSTANDING THE BASICS

LLMs and KGs Unveiled

4



Understanding the individual strengths and applications of LLMs and KGs sets the stage for their combined potential.

- LLMs: Neural networks for human-like text.
- KGs: Structured knowledge networks.
- Emphasis on RDF KGs and SPARQL.

Adrian Gschwend

Founder, CEO
Zazuko
Biel, Switzerland

5



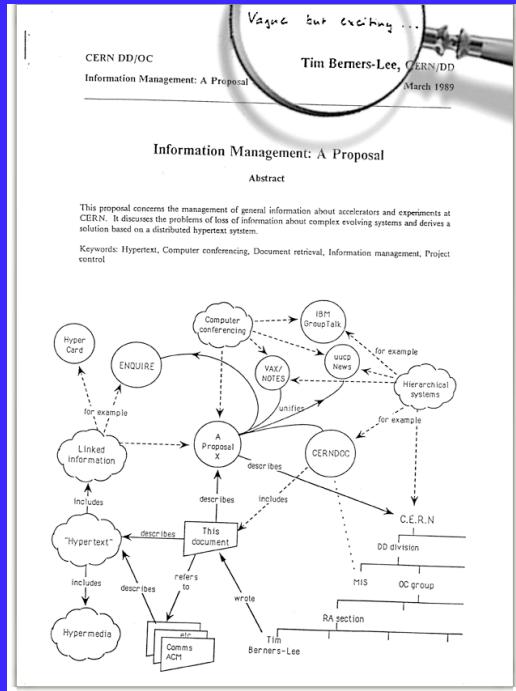
- LinkedIn:
<https://www.linkedin.com/in/adriangschwend/>
- Mastodon/Fediverse: @ktk@social.netlabs.org
- adrian.gschwend@zazuko.com

About Zazuko:

- We are a Knowledge Graph & Linked Data consulting company
- Open Source focus
- Remote-first
- Based in Europe & Southeast Asia (Indonesia)
- If you enjoy this tutorial, contact us!

THE SEMANTIC WEB VISION

By Tim Berners-Lee



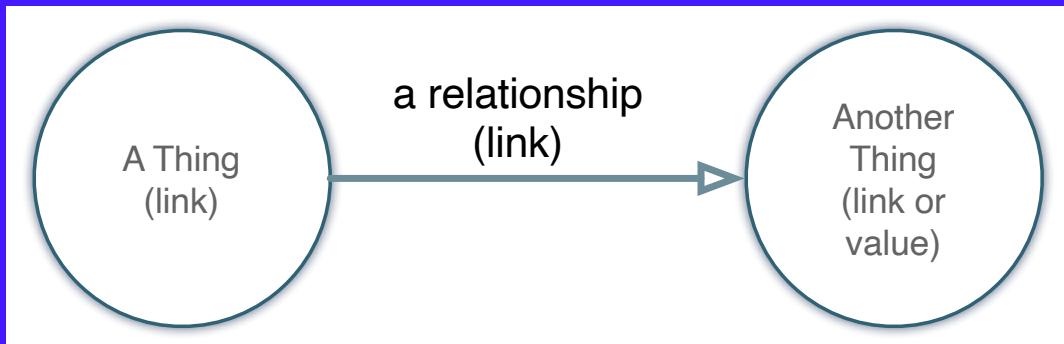
Envisioning a web of data that can be processed by machines and humans alike, aiming for a universal data exchange medium.

- Probably always part of Tims vision, see "the proposal"

RDF

THE BUILDING BLOCK

Resource Description Framework



7



RDF as the foundational model for data interchange on the web, with SPARQL as its standardized query language.



VARIOUS FORMATS

RDF Serializations

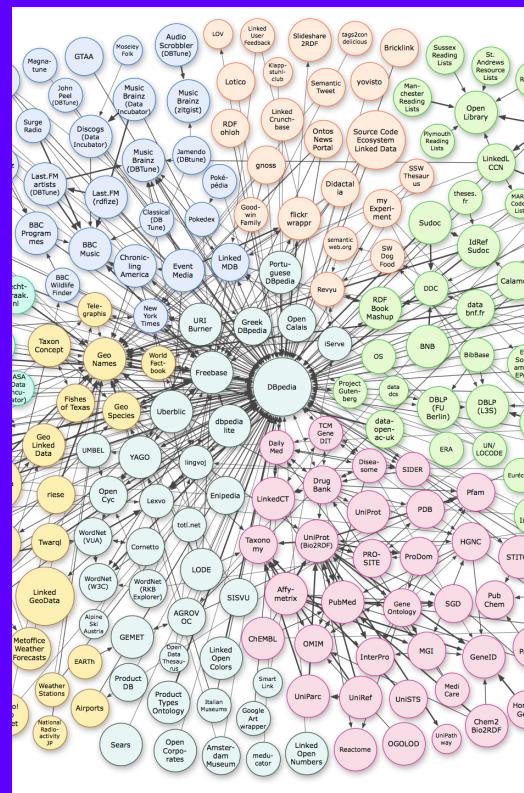
8

 zazuko

Exploring different formats like Turtle and JSON-LD, designed to make RDF data comprehensible for both humans and machines.

LINKED DATA PRINCIPLES

Connecting the Web, using 5- Star data



9



The idea of interlinked data sets across the web, enhancing access and utility of information globally.

SCHEMA.ORG INITIATIVE

Standardizing Structured Data

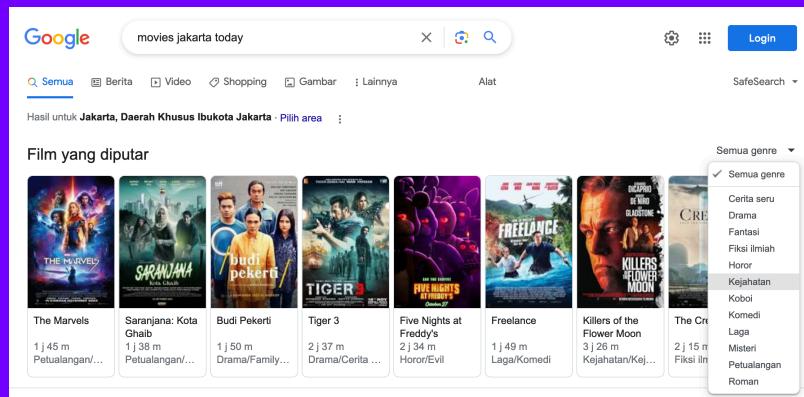
10



Initiated in 2011 by leading search engines, schema.org provides a common vocabulary for structured data on the internet.

GOOGLE'S KNOWLEDGE GRAPH

Revolutionizing Search

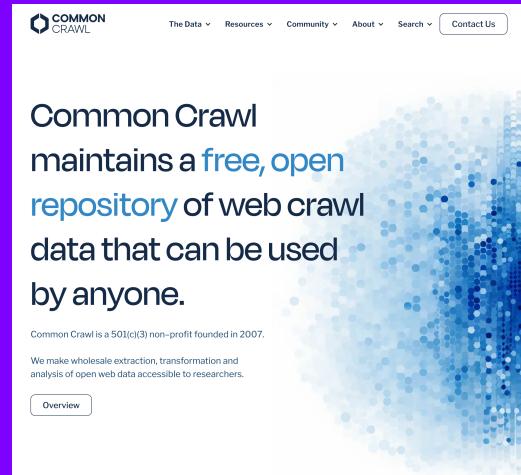


11



Google's introduction of the term "Knowledge Graph" in 2012 marked a new era in search technology, leveraging interconnected data.

EMBEDDED STRUCTURED DATA



12



With half of all websites embedding some form of RDF triples, the web is rich with structured, machine-readable data.

- We can't say for sure what the effect of embedded, structured data is for Large Language Models
- But we can assume that is very big for common knowledge

Source

**... we found
structured data
within 1.7 billion
HTML pages
out of the 3.4
billion pages
contained in
the crawl
(50.0%)**

Commons - RDFa, Microdata, Embedded JSON-LD, and Microformats data sets which have been released by the Common Crawl Foundation.

statistics about the [Web Data Commons](#) RDFa, Microdata and Microformats data sets which have been released by the Common Crawl Foundation.

structured data within 1.7 billion HTML pages out of the 3.4 billion pages contained in the crawl (40.0% of the crawl).

on different pay-level-domains out of the 34.5 million pay-level-domains covered by the crawl (40.0% of the crawl).

list of 863.5 billion [RDF quads](#).

download the RDFa, Microdata, Embedded JSON-LD and Microformats data sets are given on the [data page](#).

acted [schema.org class-specific datasets](#) from the Microdata and JSON-LD corpora.

following term *Domains* refers to pay-level-domains. Subdomains are not counted as separate domains.

September 2020

81.8 Terabyte (compressed)

3,410,268,379

1,701,573,394

34,596,585

15,316,527

21,636,494,250

86,381,005,124

1.9 Terabyte (compressed)

at

Domains	URLs	Typed Entities	Triples
---------	------	----------------	---------

7,809,978	892,571,812	7,374,224,188	35,612,247,646
-----------	-------------	---------------	----------------

7,659,585	767,626,882	6,684,258,935	32,078,019,309
-----------	-------------	---------------	----------------

4,309,621	371,420,305	6,462,214,971	12,102,799,203
-----------	-------------	---------------	----------------



Source

RISE OF ENTERPRISE KGS

Beyond Public Data

14



The growing trend of implementing Knowledge Graphs in enterprise environments for enhanced internal data analysis and decision-making.

LLMs AND RDF

AI's Comprehension of Structured Data

15



Large Language Models demonstrate an increasing understanding of RDF and its various serializations, thanks to widespread data embedding.

THE EVOLUTION OF AI

Tracing the Journey

16



Explore the historical development of AI, leading to the creation of LLMs and KGs.

- Key moments in AI history.
- Emergence of LLMs and KGs.

EARLY AI CONCEPTS

Beginnings in the 1950s

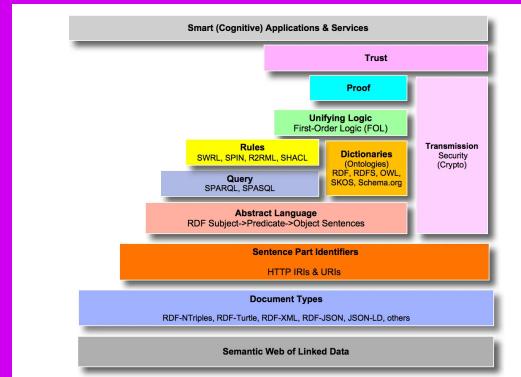
17



The birth of AI as a scientific discipline, exploring concepts of human-like intelligence in machines.

RULE-BASED AI ERA

Logic and Expert Systems



18



Prominence of rule-based systems in early AI. Systems like Prolog and Datalog exemplified this approach, focusing on logic and structured rules.

- Notation3 is an RDF approach to that. Those with Prolog and Datalog background will feel at home.

Image credit: [Kingsley Uyi Idehen, OpenLink Software](#)

EVOLUTION OF MACHINE LEARNING

Shifting Paradigms

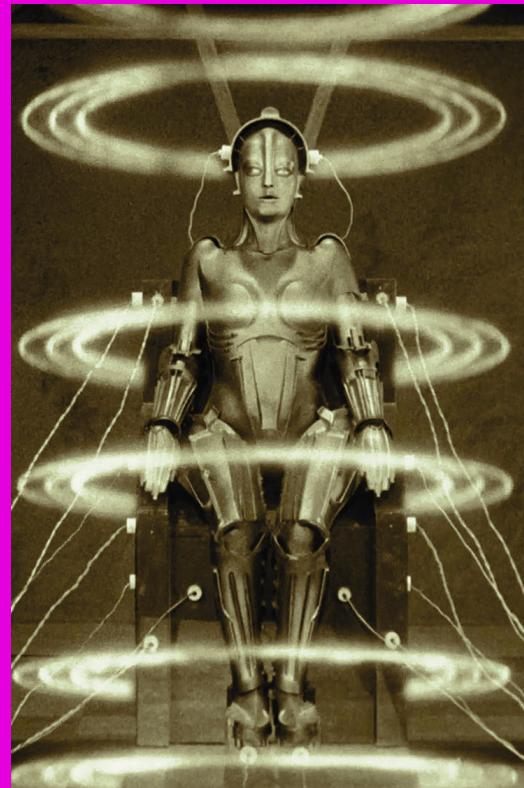
19



Gradual shift from rule-based systems to machine learning, emphasizing learning from data rather than following hardcoded rules.

RISE OF NEURAL NETWORKS

The Deep Learning Breakthrough



20

 zazuko

1980s–2015ish: Advancement of neural networks, leading to the current deep learning era, vastly improving AI's capabilities. (Yes, there was a "AI winter" in between).

LARGE LANGUAGE MODELS (LLMs)

Deep Learning in Natural Language Processing

21



LLMs: The Frontier of NLP

- Advanced application of deep learning for processing language.

Core Functionality:

- Deep neural networks analyzing vast datasets to understand and generate text.

Impact and Capabilities:

- Breakthroughs in text completion, translation, and more.
- Demonstrating a new level of AI's linguistic ability.

Addressing Challenges:

- Ongoing efforts to handle context, reduce biases, and improve reliability.
- Future focus on enhancing transparency and ethical AI.

VECTOR MODELS IN AI

Transforming Language into Mathematics

22



Essence of Vector Models:

- Turning text data into numerical form using vectors for AI processing.

Importance in NLP:

- Essential for tasks like semantic analysis, where understanding context and meaning is crucial.

How They Work:

- Words, phrases, and even entire documents are mapped to vectors in multi-dimensional space.

Applications:

- From basic similarity checking to complex operations like machine translation and sentiment analysis.

Advancements:

- Continuous evolution with techniques like word embeddings (Word2Vec, GloVe) and transformers.

MODERN AI LANDSCAPE

Diverse Applications

23



Today's AI: A blend of machine learning, deep learning, and traditional AI methods, applied from natural language processing to complex decision-making.

THE SYNERGY LLMs and KGs in Collaboration

24



Uniting LLMs' rapid linguistic analysis with KGs' deep, structured reasoning for enhanced AI capabilities.

LLMs: Masters of Context and Nuance

- Excelling in understanding and generating natural language.
- Quick, intuitive responses akin to human 'System 1' thinking.

KGs: Powerhouses of Depth and Structure

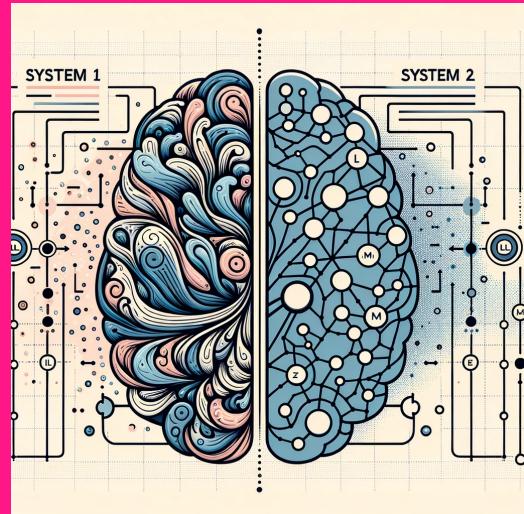
- Providing a rich, interconnected framework for data.
- Methodical and logical, paralleling human 'System 2' thought processes.

Synergistic Strengths and Complementary Weaknesses:

- LLMs bring speed and adaptability, while KGs contribute precision and reliability.
- Addressing the limitations of each through their combined use.

A DUAL MINDSET

Quick Thoughts & Deep Insights



25

 zazuko

Delving into how LLMs and KGs complement each other like the two systems of thought.

- System 1: LLMs for intuition.
- System 2: KGs for analysis.

From the book *Thinking, Fast and Slow* by Daniel Kahneman

Credits for KG meets LLM: Tony Seale, KG Engineer at UBS

LLM LIMITATIONS

Addressing the Challenges

26



Acknowledging and discussing the current limitations and challenges faced by LLMs.

- Issues with reliability and speed.
- High training and operational costs.
- You cannot really query facts.
- Not trained to your particular problem.

KG ADVANTAGES

Emphasizing Strengths

27



- Speed and efficiency of KG queries.
- Cost-effectiveness.
- But: No understanding of natural language.



HANDS-ON EXPERIENCE

Crafting SPARQL Queries with LLMs

29



Engage in interactive exercises to transform natural language into SPARQL queries.

- Creating queries from scratch.
- Analyzing and refining results.

QUERY 1: INDIVIDUAL FACTS

**Create a SPARQL query
to find the birthplace of
Albert Einstein**

Explanation:

- This query tests the AI's ability to extract specific factual data about a well-known individual.

QUERY 2: AUTHOR'S WORKS

Generate a SPARQL query to list all books written by Agatha Christie.

Explanation:

- Challenges the AI to understand relationships between an author and their works.

QUERY 3: FILM GENRES BY DECADE

What's the SPARQL query to find all science fiction films released in the 1990s?

Explanation:

- This query introduces filtering results by genre and time period.

QUERY 4: GEOGRAPHICAL DATA
I need a SPARQL query
to list all the capital
cities in Europe.

Explanation:

- Geographical queries test the knowledge graph's data richness and the AI's capability in handling spatial information.

QUERY 5: HISTORICAL EVENTS

Can you create a SPARQL query to find the date and location of the signing of the Treaty of Versailles?

Explanation:

- This query combines a historical event with specific details, testing the AI's ability in handling complex historical data queries.

LANGCHAIN LIVE

Demonstrating

Integration

35



Experience Langchain in action, revealing the integration of LLMs with KGs.

- Live Jupyter notebook interaction.
- RDF cube queries and dialogue.

IMAGINING THE FUTURE

Open Source Innovation

36



Engage in a collaborative session to envision the future of open-source LLM and KG tools.

- Current state of tools.
- Envisioning future enhancements.
- {Interactive brainstorming session.}

THE ROAD AHEAD

From Non-RDF to RDF

37



Discuss the potential for converting non-RDF data into RDF format using LLMs.

- Challenges in data conversion.
- Role of LLMs in simplification.
- {Talk about automating data transformation.}

WRAP-UP AND Q&A

Reflections and

Discussions

38



Concluding the workshop with reflections on key learnings and an open Q&A session.

- Recap of key points.
- Open floor for questions.