

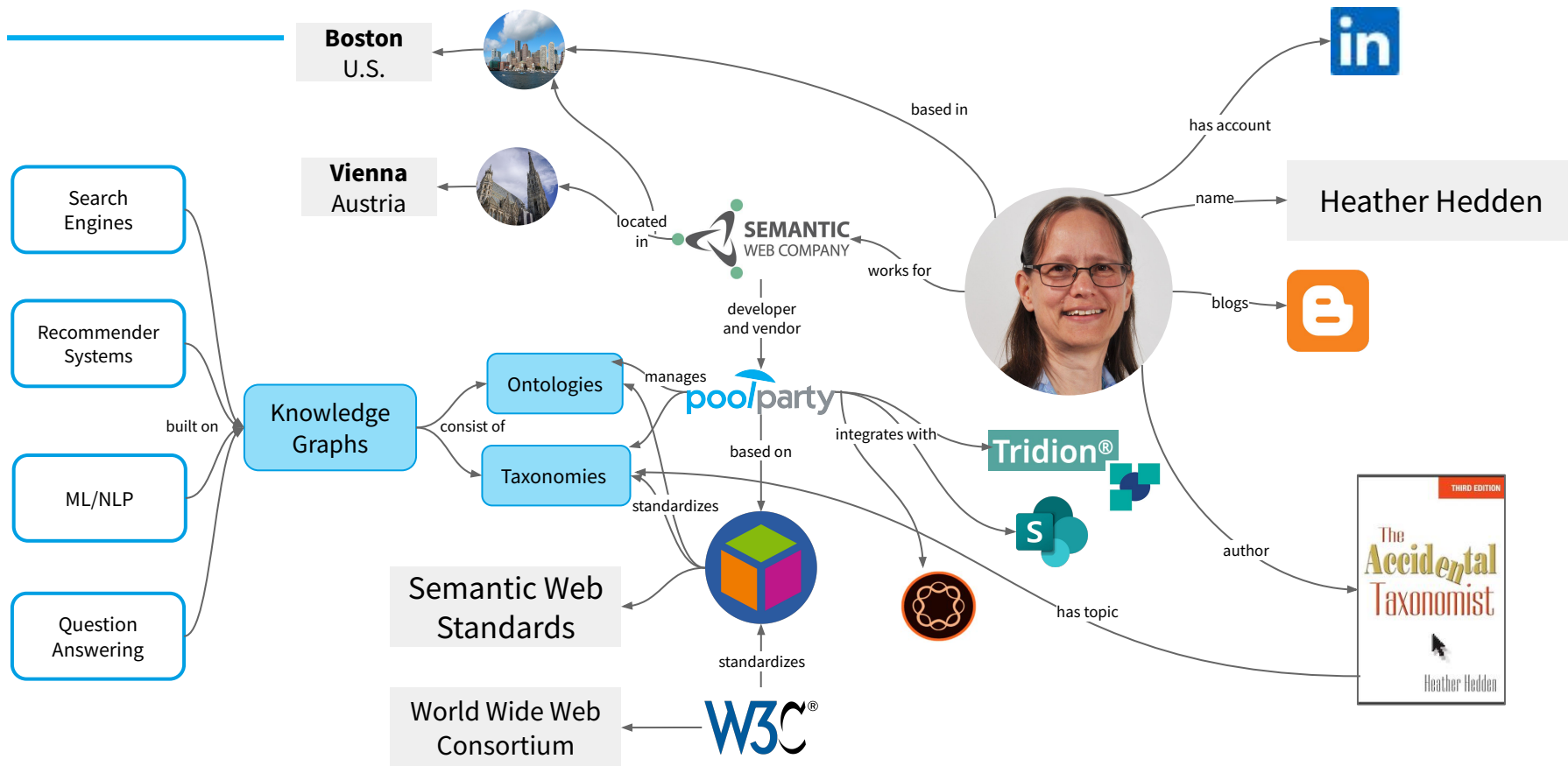


Generative AI, Context, & Knowledge Models

Enterprise AI World
Washington, DC - November 8, 2023

Heather Hedden
Communications Manager
Knowledge Manager
Semantic Web Company

Click the Graph—get in contact with us!



Semantic Web Company (SWC) and PoolParty

SWC is developer/vendor of
PoolParty Semantic Suite

Most complete **Semantic
AI platform** on the market

W3C standards compliant



ISO 27001:2013
certified (since 2019)

PoolParty stands out in the
market with a **user-
friendly** interface.



Current version **9.0**

On-premises or
cloud-based



Over **180** **customers**
world-wide

Integrations:



Microsoft 365, Tridion
Docs & Sites, AEM,
and many more



Gartner named SWC a Visionary
in their **Magic Quadrant** for
Metadata Management Systems
2019 and 2020



Semantic Web Company
has been recognized by
KMWorld as one of the
AI Top 100 companies 2023



Forrester listed SWC as sample
vendor in their **report** on *The
Document-Oriented Text
Analytics Platforms Landscape
2022*

- ▶ Introduction
 - ▷ Information Seeking: Web vs. the Enterprise, Searching vs. Questioning
 - ▷ Generative AI and LLMs
- ▶ Knowledge Models for Context
 - ▷ Knowledge Models
 - ▷ Knowledge Organization Systems
 - ▷ Knowledge Graphs
- ▶ Example Application with a Domain Knowledge Model: PoolParty Meets ChatGPT
 - ▷ Demo and Structure
 - ▷ ESG Knowledge Model Behind PoolParty Meets ChatGPT
- ▶ Other Benefits of LLM and Knowledge Graph Combinations

Introduction: Information Seeking, Generative AI, and LLMs

On the Web vs. Within the Enterprise

On the Web	Within the Enterprise
Find something useful	Find something specific and precise
Get some information on a subject	Get all the information on a subject
Expect some irrelevant results, which can quickly be dismissed	Expect better results and waste time reading through what turns out to be irrelevant
➡ Know and accept that it's a black box	Would like to have answers explainable
➡ Simple user interface	Same simple web user interface desired

Searching vs. Questioning/Information Requests

Searching	Questioning and Info Requests
Using combinations of keywords and noun phrases	Using full natural sentences: questions or “tell me about”
Results are a list of matching items	Result is is curated or generated content
Results refined by adding more keywords	Results refined with with additional statement/question prompts - “conversational”
Incorrect results are identified, dismissed, and tolerated	Incorrect results cannot be identified easily and can have negative repercussions
Enhanced in the enterprise with taxonomies and refinement filters	Enhanced in the enterprise with taxonomies, ontologies, and knowledge graphs



▶ Generative AI

- ▷ Artificial intelligence capable of generating content (text, images, etc.) using generative models.
- ▷ Models learn the patterns and structure of their input training data, by unsupervised or self-supervised deep machine learning.
- ▷ Generative AI systems trained on words or **word tokens** include **GPT-3**, **LaMDA**, **LLaMA**, **BLOOM**, **GPT-4**, etc.

▶ Large Language Models

- ▷ Deep learning algorithms that can perform various NLP tasks.
- ▷ Algorithms are trained on very large datasets, using transformer models
- ▷ They can recognize, translate, predict, or *generate* text or other content.

▶ ChatGPT

- ▷ A front-end chatbot to a generative AI system

- ▶ **General Generative AI and LLM problems:**
 - ▷ Lack of transparency - information sources not disclosed
 - ▷ Hallucinations - generating incorrect answers
 - ▷ Training time cutoff - lack of information after LLM training date
 - ▷ Generality - lack of knowledge of a specific subject domain
 - ▷ No long-term memory - getting different answers to the same question asked later
- ▶ In sum: Lack of effectiveness, reliability, precision, and trustworthiness
- ▶ **Bringing generative AI and LLMs into the enterprise has the potential to change:**
 - ▷ Transparency: information sources are the enterprise's internal sources - **Explainable**
 - ▷ Correctness: hallucinations rare with controlled content
 - ▷ Timeliness: training can continue and be recent
 - ▶ Domain specific: rich content for deep knowledge of a specific subject domain
 - ▶ Potential greater consistency: if combined with a knowledge model

What does the enterprise have?

- ▶ Large volumes of digital content - for predictive generative text modeling/training
- ▶ Large volumes of data - to be mined, along with content, and served as answers
- ▶ Specialized subject content - to support appropriate, fuller answers
- ▶ Structure to portions of the content/data - for context for better answers
- ▶ Knowledge organization systems (controlled vocabularies, glossaries, terminologies, taxonomies, etc.) - for more accurate and consistent answers that can also link to additional information

What could the enterprise develop?

- ▶ Knowledge models and knowledge graphs - to link content, data

Knowledge Models, Knowledge Organization Systems, and Knowledge Graphs

Context provides understanding

- ▶ Words have different meanings in different contexts.
 - ▷ Words in content have different meanings depending on the context of a document.
- ▶ Context needs to be provided in generative AI prompts.
 - ▷ Words in a query/prompt will have different meanings in different contexts.

Examples: [Discovery](#), [Index](#), [Interpretation](#), [Running](#)

Context of content

- ▶ Context of the subject of the content used for *training* the model
- ▶ Context of the subject of the content being *queried*

Context of a knowledge model

- ▶ Context of a taxonomy:
 - ▷ Terms have different meanings depending on what's broader, narrower, or related, and possible notes/definitions
Broader concepts, concept schemes, definitions, and notes
- ▶ Context of an ontology:
 - ▷ Classes provided context for their entity members and the semantic relations between classes

What is a knowledge model?


- ▶ A computer-readable representation of a knowledge domain
- ▶ A set of interrelated **knowledge organization systems (KOSs)**, controlled vocabularies, metadata properties, taxonomies, and ontologies
- ▶ Comprising entities/concepts and relationships between them
- ▶ Serving a shared purpose or related purposes in information management
- ▶ May include or link out to externally published knowledge organization systems (e.g classification schemes).
- ▶ A knowledge model can form the basis of a knowledge graph
- ▶ Managed in a tool, such as PoolParty, as a “project”

Knowledge Organization system (KOS)

- ▶ Any system of terms, terminology, classification, etc.
- ▶ to organize, define, manage, and/or retrieve information.
- ▶ Not any method to organize knowledge directly, but rather a *scheme* to organize concepts for organizing, classifying, defining, tagging, or retrieving information.

KOS types:

term lists
synonym rings
name authorities
taxonomies
thesauri
glossaries
dictionaries
gazetteers
terminologies
categorization schemes
classification systems
subject heading schemes
semantic networks
ontologies



Controlled
Vocabularies
for tagging and
information
retrieval

What is a taxonomy?

Controlled and *organized*

1. **Controlled:**

A kind of controlled vocabulary or knowledge organization system, based on unambiguous concepts, not just words:
things, not strings

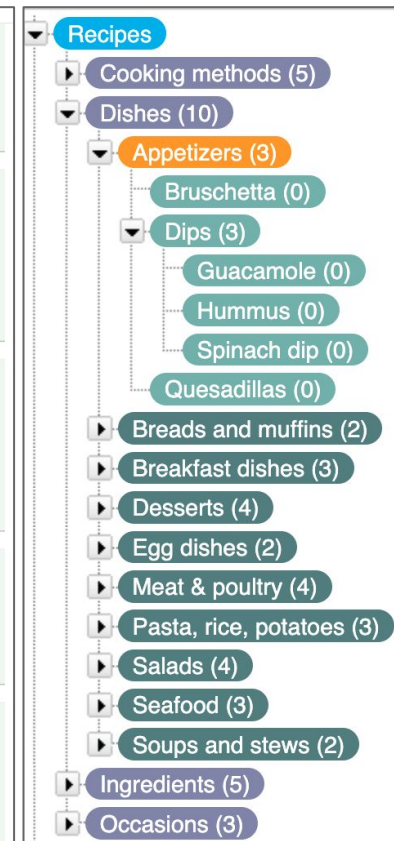
2. **Organized:**

Concepts are arranged in a structure of hierarchies, categories, or facets to organize them.

controlled

Preferred Label	
<input type="radio"/> Vorspeisen	de
<input type="radio"/> Appetizers	en
Alternative Labels	
<input type="radio"/> Hors d'oeuvres	en
<input type="radio"/> Starters	
+	
Hidden Labels	
<input type="radio"/> Appetisers	en
<input type="radio"/> Horderves	
+	
Scope Notes	
<input type="radio"/> Dishes usually served as appetizers	en
+	
Definitions	
<input type="radio"/> A small dish of food or a drink taken before a meal or the main course of a meal to stimulate one's appetite	en

organized

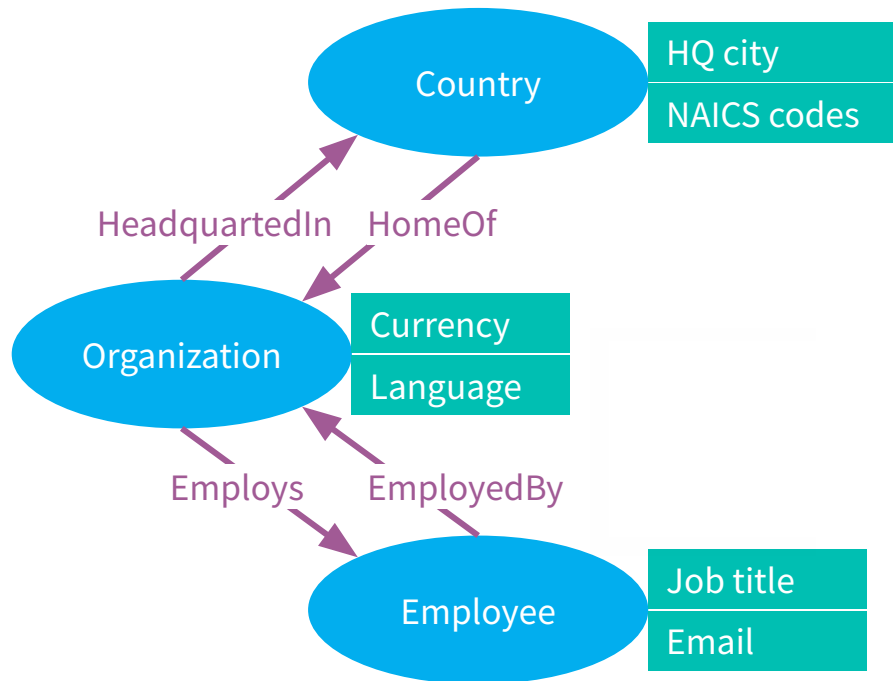


Ontology

- ▶ A model of a knowledge domain
- ▶ Similar to (most of) a knowledge graph, but excludes the actual instance data
- ▶ A formal naming and definition of the types (classes), properties and interrelationships of entities in a particular domain.
 - ▷ Relations contain meaning, are “semantic.”
 - ▷ Properties are customized attributes of entities.
- ▶ Common standards provided by W3C: Web Ontology Language (OWL) and RDF-Schema.
- ▶ Comprises classes, relations, and attributes, which are linked in statements of triples.



Ontology



Classes:

Employee, Country, Organization

Relations:

HeadquarteredIn < > HomeOf

EmployedBy < > Employs

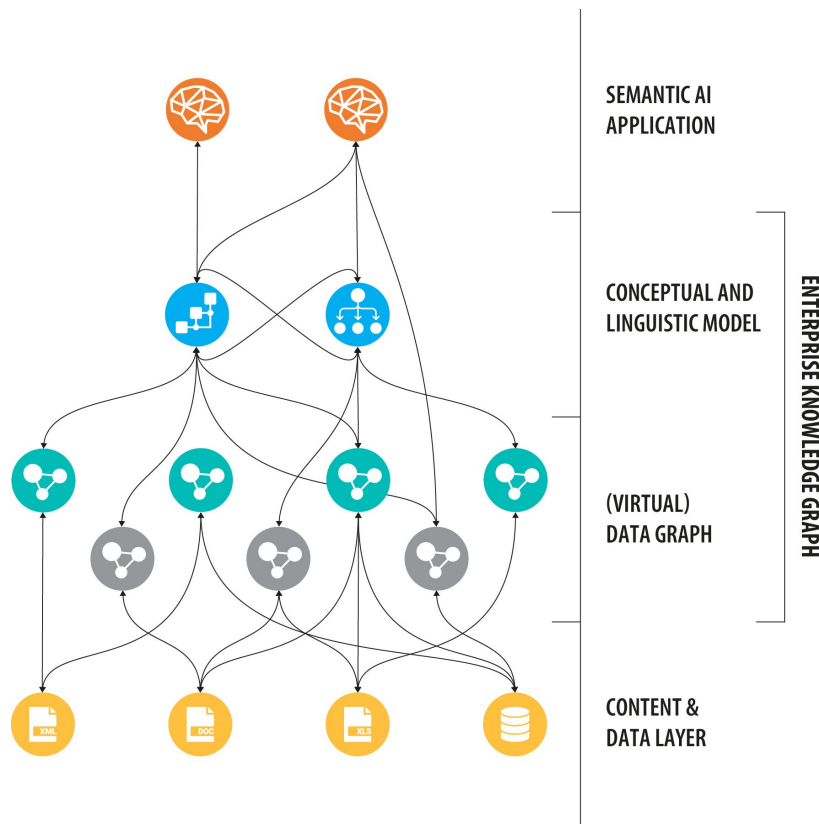
Attributes:

Email address, Job title, HQ city, NAICS codes, Currency, Language

Knowledge Graphs

What is a knowledge graph?

- ▶ A **knowledge model** (taxonomies, etc. + ontology) combined with instance **data** stored in a **graph database**
- ▶ Provides **unified information** across a domain/organization, enriched with semantics.
- ▶ Contains business objects and topics that are closely **linked**, classified, and connected to existing data and documents.
- ▶ A **layer** between the actual content and the querying layer
- ▶ Implemented with a **semantic AI front-end**, such as a knowledge hub or question-answering



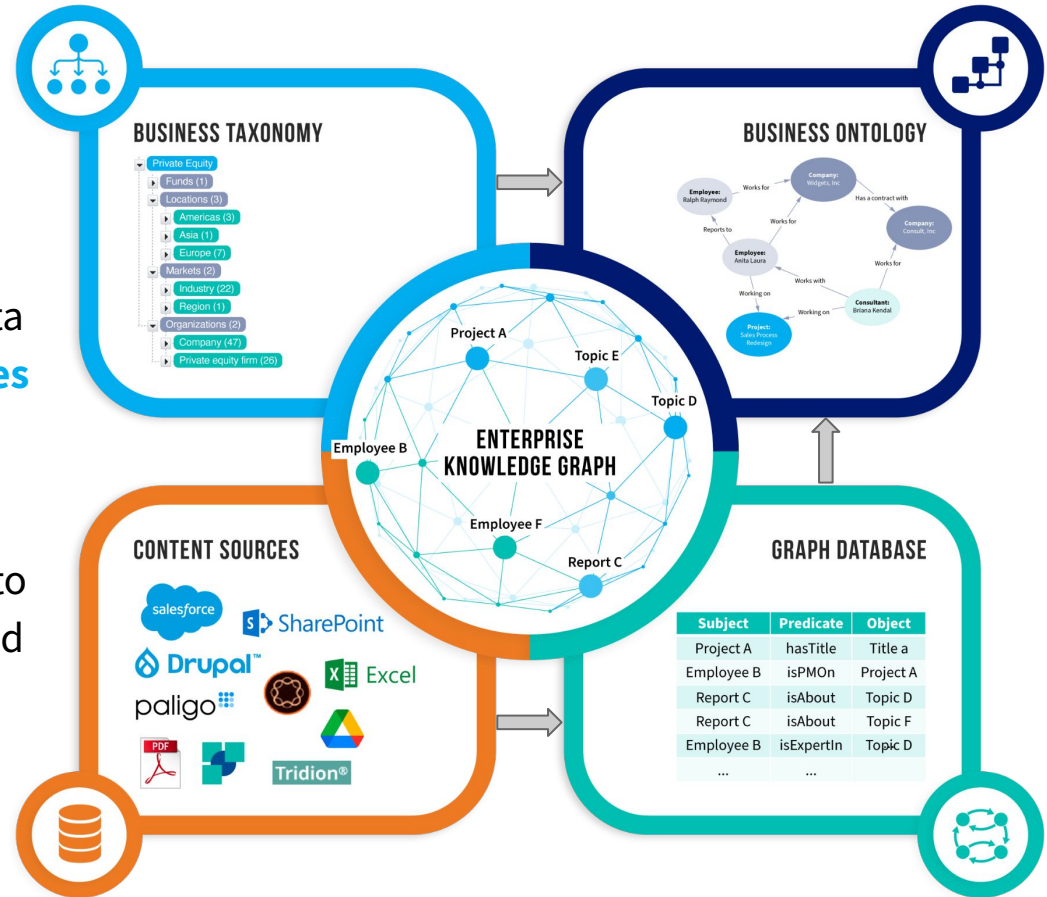
What knowledge graphs can do

- ▶ Integrate knowledge
- ▶ Serve data governance
- ▶ Provide semantic enrichment
- ▶ Bring structured and unstructured data together
- ▶ Provide unified view of different kinds of unconnected data sources
- ▶ Provide a semantic layer on top of the metadata layer
- ▶ Improve search results beyond machine learning and algorithms
- ▶ Answer complex user questions instead of merely returning documents on a topic
- ▶ Combine with deep text analytics, semantic AI, and machine learning



Knowledge Graph Components

- ▶ Extracted **data**, stored in a **graph database** (either an RDF-based triple store or a labeled property graph)
- ▶ which are tagged/linked with metadata or concepts in **controlled vocabularies** (especially **taxonomies**) to label and organize the data,
- ▶ which in turn are semantically linked to represent conceptual relationships and enriched with additional metadata attributes in an **ontology**.



Knowledge Graph Technology Components

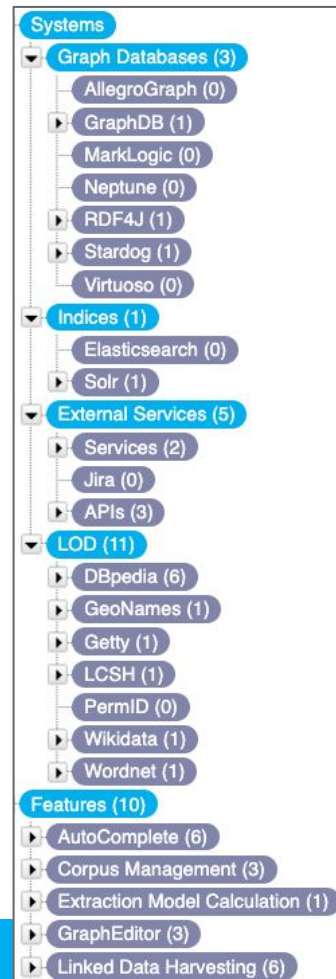
Core software and technology needed:

- ▶ Graph database management software
- ▶ Taxonomy/ontology management software based on W3C standards
- ▶ Search software (such as Solr or Elasticsearch)
- ▶ Front-end (web) application

Also important:

- ▶ Extract-Transform-Load tool to extract data
- ▶ Text mining/natural language processing/entity extraction tool
- ▶ Machine-learning auto-classification tool
- ▶ Capabilities (such as algorithms for weighting/scoring relations) specified in SPARQL query language for RDF

PoolParty Semantic Suite includes all of these, with an API to connect to or custom build a front-end applications

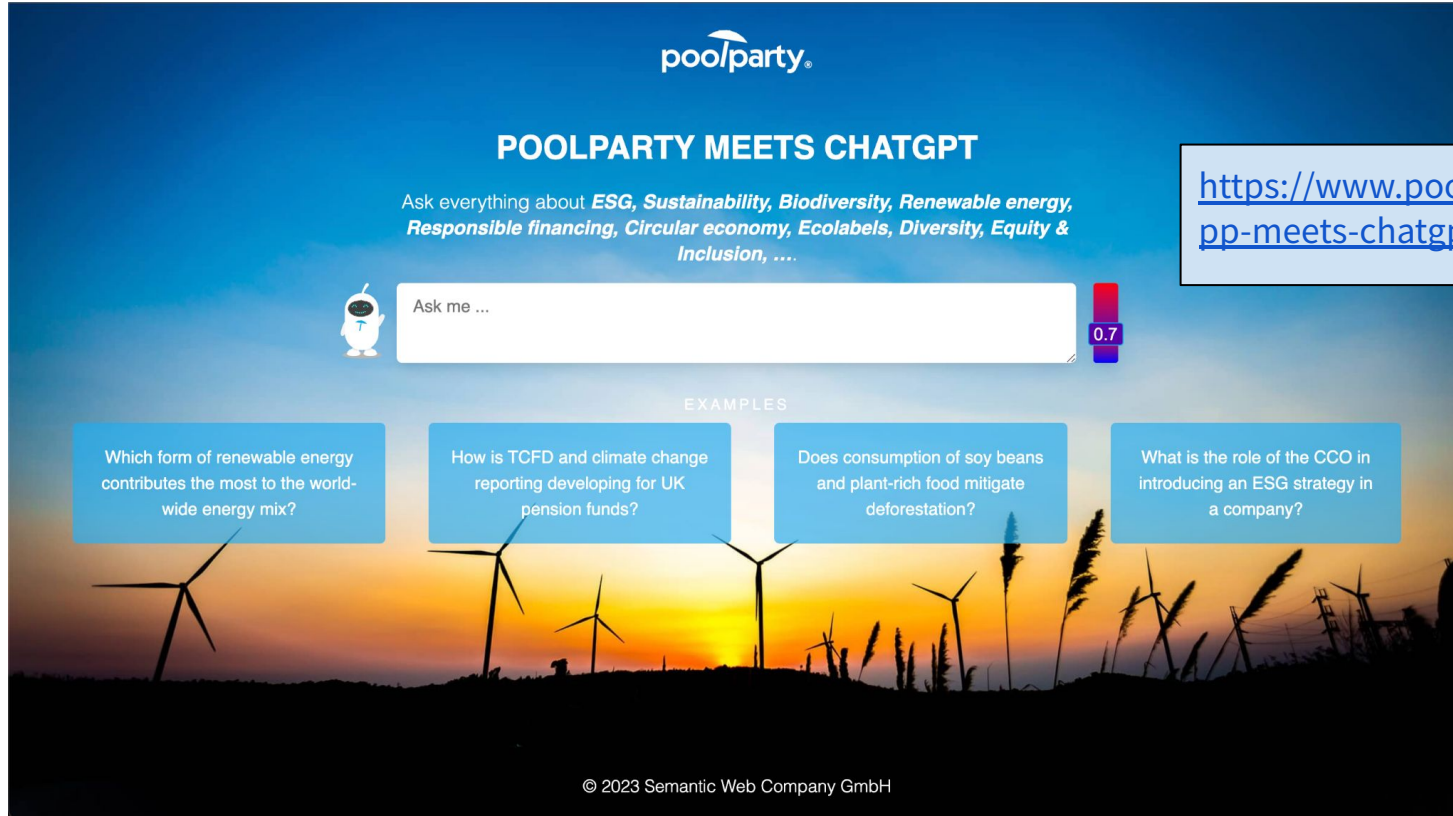


Example Application of Generative AI with a Domain Knowledge Model

To serve the information needs of an enterprise

- ▶ Knowledge model should be in the domain of the enterprise's interest.
- ▶ Content for training should be in the same domain of interest.
- ▶ Content for training ideally should be from within the enterprise, but could be external if restricted, and internal content is insufficient in size
- ▶ Content queried by generative AI may be within the enterprise or external, as needed.

Generative AI + Domain Knowledge Model



The screenshot shows a web interface for "POOLPARTY MEETS CHATGPT". At the top, the "poolparty" logo is displayed. Below it, the title "POOLPARTY MEETS CHATGPT" is centered. A subtitle reads: "Ask everything about **ESG, Sustainability, Biodiversity, Renewable energy, Responsible financing, Circular economy, Ecolabels, Diversity, Equity & Inclusion,**". To the left of a search bar is a small white robot icon. The search bar contains the placeholder text "Ask me ...". To the right of the search bar is a red and purple vertical bar with the number "0.7" in white. Below the search bar, the word "EXAMPLES" is centered. There are four blue boxes, each containing a question: "Which form of renewable energy contributes the most to the world-wide energy mix?", "How is TCFD and climate change reporting developing for UK pension funds?", "Does consumption of soy beans and plant-rich food mitigate deforestation?", and "What is the role of the CCO in introducing an ESG strategy in a company?". The background of the interface is a sunset scene with silhouettes of wind turbines and reeds. At the bottom, the copyright notice "© 2023 Semantic Web Company GmbH" is visible.

poolparty

POOLPARTY MEETS CHATGPT

Ask everything about **ESG, Sustainability, Biodiversity, Renewable energy, Responsible financing, Circular economy, Ecolabels, Diversity, Equity & Inclusion,**

Ask me ...

0.7

EXAMPLES

- Which form of renewable energy contributes the most to the world-wide energy mix?
- How is TCFD and climate change reporting developing for UK pension funds?
- Does consumption of soy beans and plant-rich food mitigate deforestation?
- What is the role of the CCO in introducing an ESG strategy in a company?

© 2023 Semantic Web Company GmbH

<https://www.poolparty.biz/pp-meets-chatgpt>

PoolParty Meets ChatGPT



How could open data and knowledge graphs help to prevent critical climate tipping points?

EXAMPLES

Which form of renewable energy contributes most to the world-wide energy mix?

How is TCFD and climate change reporting developing for UK pension funds?

Does consumption of soy beans and plant-rich food mitigate deforestation?

Which role plays the CCO when rolling out an ESG strategy in a company?

ChatGPT's direct answer

Open data and knowledge graphs can play a crucial role in helping to not reach critical climate tipping points by providing a better understanding of the complex interactions between different environmental factors and their impacts on climate change. Here are some ways in which they can help:

1. Improved Climate Modeling: Open data and knowledge graphs can enable the development of more accurate climate models by providing access to a vast array of environmental data, such as temperature, precipitation, and greenhouse gas emissions. This can help researchers to better understand how different factors interact with each other and predict the potential impacts of climate change.
2. Identifying Risks and Opportunities: Open data and knowledge graphs can help identify areas that are most at risk of climate change, such as areas prone to flooding, drought, or wildfires. This information can be used to develop better land-use planning strategies, improve disaster response plans, and identify opportunities for mitigation and adaptation.
3. Promoting Public Awareness: Open data and knowledge graphs can help raise public awareness of the impacts

ChatGPT's answer when background information from the ESG graph is added

[Open data](#) and [knowledge graphs](#) can play an essential role in preventing the Earth from reaching a critical [climate tipping point](#). Here are some ways in which they can help:

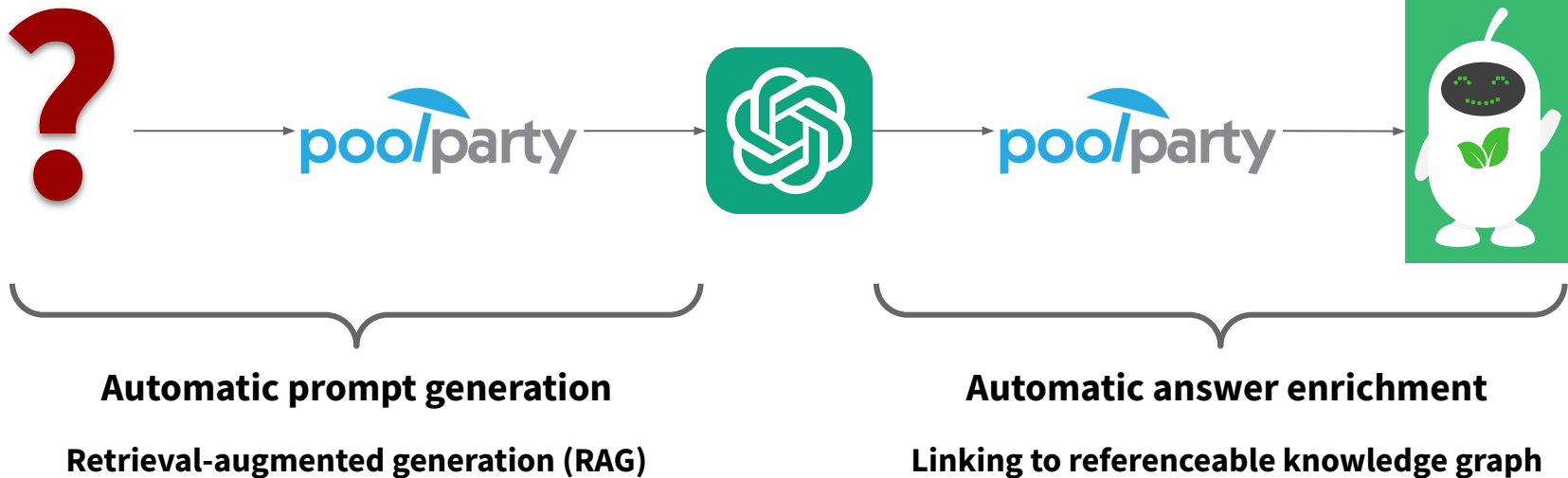
1. [Semantic Web: The Semantic Web](#) is a [technology](#) that enables data to be shared and reused across applications, enterprise, and community boundaries. By using [Semantic Web](#) technologies, we can make climate data more accessible, interoperable, and machine-readable. This will help to integrate data from various sources and build a comprehensive understanding of [climate change](#).
2. [Knowledge models: Knowledge models](#) are a formal representation of knowledge that can be used to reason about complex systems. By using [Knowledge models](#), we can simulate the effects of different climate scenarios, identify potential [tipping points](#), and develop [strategies](#) to prevent them from occurring.
3. [Vocabularies and identifiers](#): Vocabularies and identifiers are essential in ensuring that data is accurately described and can be easily integrated into [knowledge graphs](#). By using standardized vocabularies and identifiers, we can ensure that data is interoperable and can be linked to other [datasets](#).

PoolParty Meets ChatGPT

Benefit from more comprehensive answers backed by referenceable knowledge

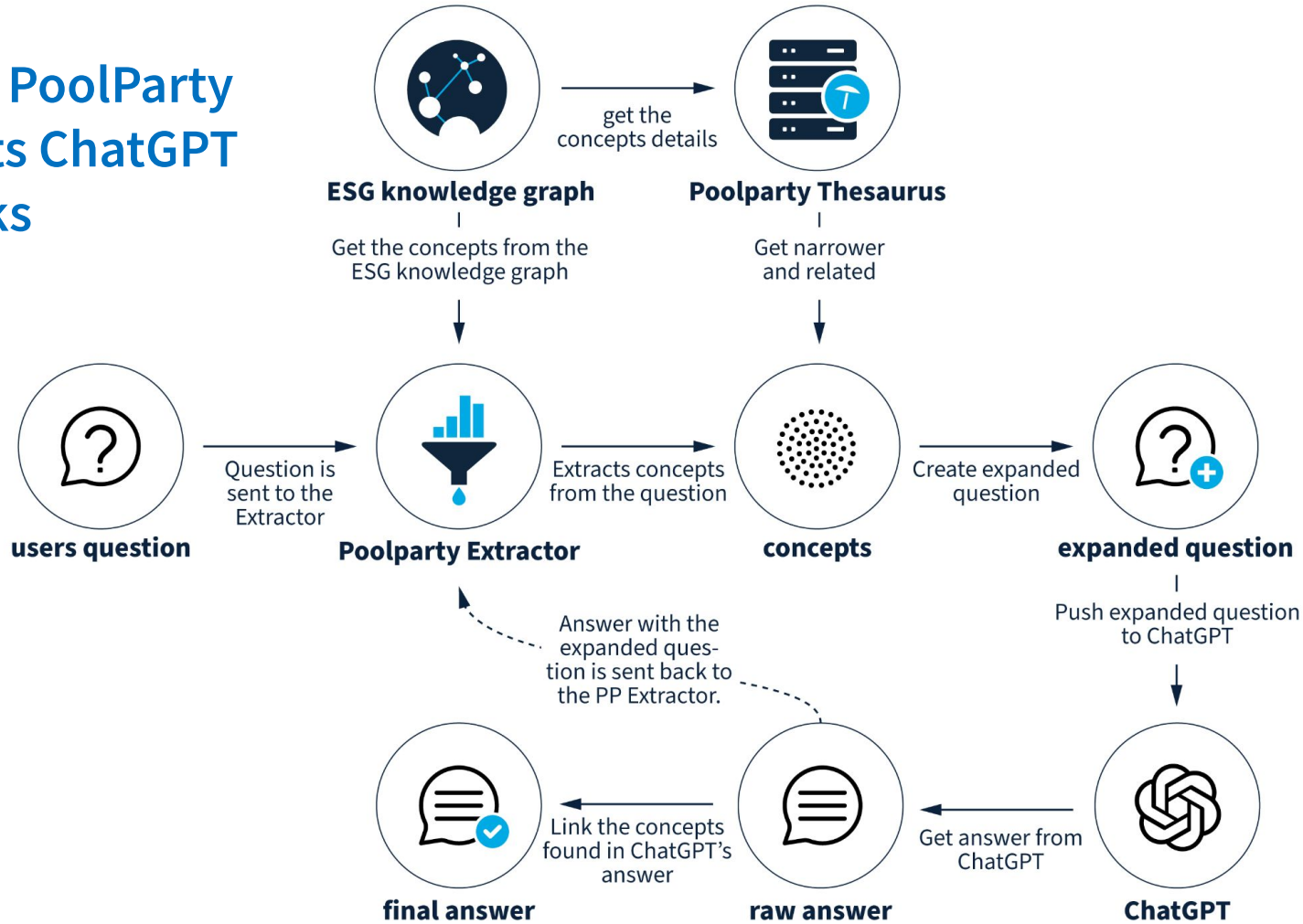
Question

Answer



[How can LLMs and knowledge graphs developed with software such as PoolParty be combined to serve as responsible AI overall?](#)

How PoolParty Meets ChatGPT works



How PoolParty Meets ChatGPT works

1. Expand the given question from user with data from ESG Knowledge Graph
 2. Get the answers from ChatGPT
 - a. with the given question from user
 - b. with the expanded question
 3. Locate concepts from the KG in the answer with the expanded question and link them to the ESG thesaurus.
 4. Get relevant links to the found concepts.
-
- ▶ Question must be at least 10 words with at least one concept in the taxonomy
 - ▶ The PP Extractor extracts the concepts from the question (checks for prefLabels, altLabels and hiddenLabels)
 - ▶ The process adds the **narrowers** (max. 5) to each found concept in brackets
 - ▶ The process adds **relateds** (max. 5 per concept) to the end of the question

Example Knowledge

The screenshot shows the PoolParty web interface. The top navigation bar includes 'PROJECT', 'TOOLS', and 'ADVANCED'. Below it, a sidebar contains a tree view of the 'ESG Core Knowledge Model' with categories like 'Core concepts (18)', 'ESG (5)', 'Geography (9)', 'GRI Standards (22)', 'Industries (37)', 'Models, data, and software (11)', 'Organizations (26)', 'Products, materials, and industrial processes (4)', 'Regulatory framework (8)', 'Risks (9)', 'Roles and stakeholders (21)', 'Sciences (3)', 'SDGs (17)', and 'Solutions (6)'. At the bottom of the sidebar are 'Lists', 'Collections', and 'GraphEditors'. The main content area displays the 'ESG Core Knowledge Model' metadata, including 'Class Statistics' (Number of Concepts, Number of Concepts, Number of Suggestions) and 'Label Statistics' (Number of Preferred, Number of Alternative, Number of Hidden).

The screenshot shows the PoolParty web interface. The top navigation bar includes 'PROJECT', 'TOOLS', and 'ADVANCED'. Below it, a sidebar contains a tree view of the 'ESG Core Knowledge Model' with categories like 'Core concepts (18)', 'ESG (5)', 'Geography (9)', 'GRI Standards (22)', 'Industries (37)', 'Models, data, and software (11)', 'Organizations (26)', 'Products, materials, and industrial processes (4)', 'Regulatory framework (8)', 'Risks (9)', 'Roles and stakeholders (21)', 'Sciences (3)', 'SDGs (17)', and 'Solutions (6)'. At the bottom of the sidebar are 'Lists', 'Collections', and 'GraphEditors'. The main content area displays the 'ESG Core Knowledge Model' details, including 'General' information (Total concepts: 4904, Last Modified: 12 August 2023 08:57 CET) and 'Concept Scheme' (Core concepts, ESG, Geography, GRI Standards, Industries, Models, data, and software, Organizations, Products, materials, and industrial processes, Regulatory framework, Risks, Roles and stakeholders, Sciences, SDGs). A right sidebar shows 'History' and '6245 11539'.

esg.poolparty.biz
<https://esg.poolparty.biz/esg-core.html>

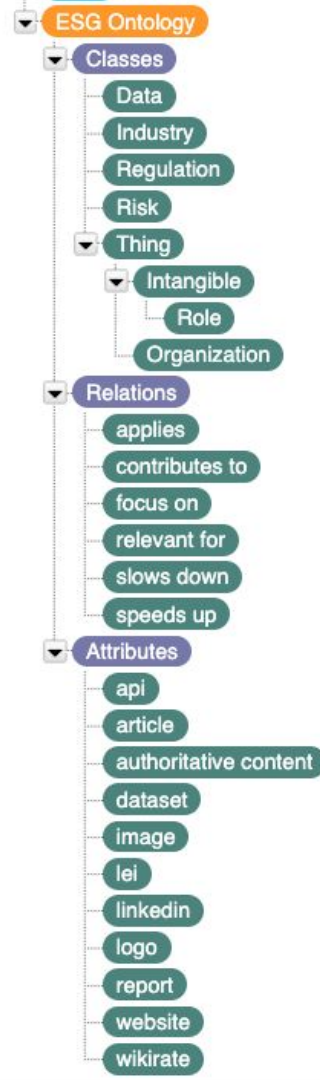
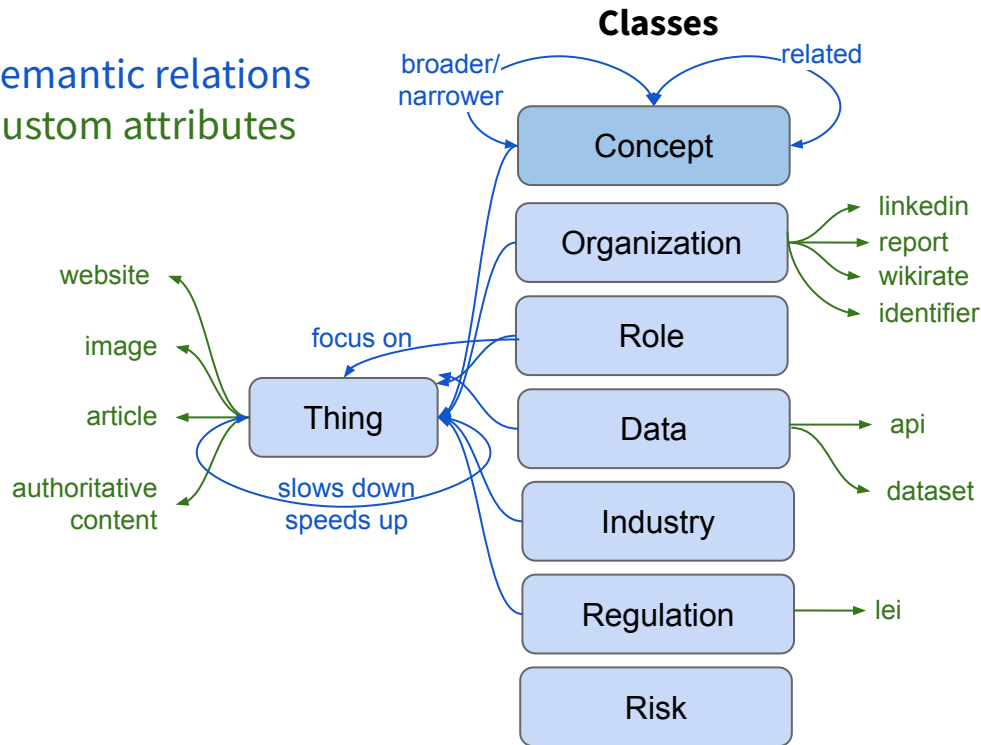
Features

- ▶ Definitions on most concepts
- ▶ Multiple synonyms (alternative labels) on concepts
- ▶ Associative relations (related concept) in addition to hierarchical relations
- ▶ 3,500 mappings to Wikidata / DBpedia
- ▶ 2,300 mappings to external content (LinkedIn, Wikirate, authoritative sources, etc.)

Example Knowledge Model: ESG

Ontology semantic layer

Semantic relations
Custom attributes



The ESG Core Knowledge Model

Ontology application as a “custom scheme”

- Custom relations
- Custom attributes, including linked content

ESG Core Knowledge Model

Core concepts (18)

ESG (5)

Geography (9)

GRI Standards (22)

Industries (37)

Models, data, and software (11)

Organizations (26)

Products, materials, and industrial processes

Regulatory framework (8)

Risks (9)

Business risk (9)

ESG risk (6)

Financial risk (6)

Investors' risk (4)

Non-financial risk (2)

Operational risk (5)

Strategic risk (5)

Systemic risk (1)

Technological risk (1)

Roles and stakeholders (21)

Sciences (3)

SDGs (17)

Solutions (6)

Financial risk

<https://esg.poolparty.biz/esg-core/66e68737-f3d0-425a-a8d6-3f334eb74dd2>

Classes: Risk

Details

Notes

Documents

Linked Data

Triples

Visualization

Quality Management

History

SKOS

ESG Schema

contributes to ⓘ

article ⓘ

URI

https://www.lazardassetmanagement.com/kr/en_us/references/sustainable-investing/demystifying-sustainability/financial-esg-risks

relevant for ⓘ

[Chief financial officer](#)

authoritative content ⓘ

URI

slows down ⓘ

image ⓘ

URI

speeds up ⓘ

linkedin ⓘ

URI

Example Knowledge Model: ESG

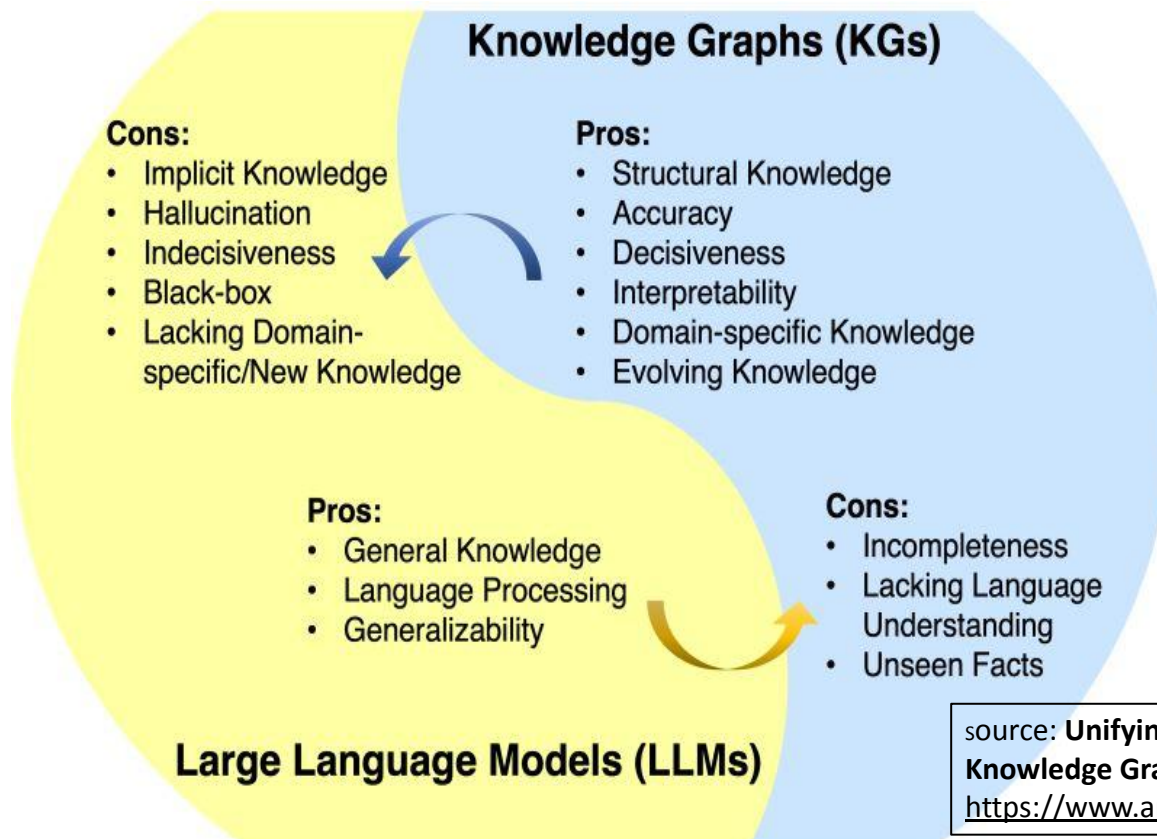
Sources for the ESG knowledge model concepts

- ▶ Vocabularies of agencies
 - ▷ US government agencies, e.g. NASA
 - ▷ UN agencies, e.g. ILO
 - ▷ EU agencies, e.g. ESA Copernicus
 - ▷ World Bank
- ▶ Published classification schemes for risk and finance
 - ▷ PRI - Principles for Responsible Investing
 - ▷ GRI - Global Reporting Initiative
 - ▷ IFRS - International Financial Reporting Standards
 - ▷ EFRAG - European Financial Reporting Advisory Group
 - ▷ SDGs - UN Sustainable Development Goals



LLMs and Knowledge Graphs

More Benefits of LLMs and Knowledge Graphs

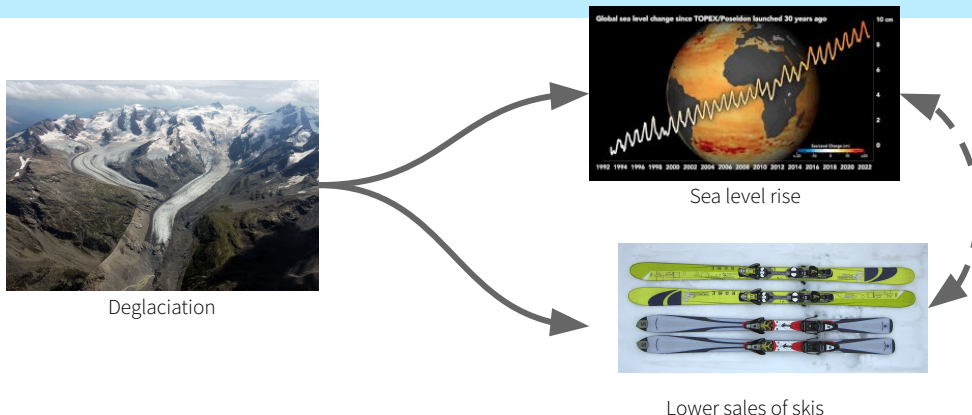


source: **Unifying Large Language Models and Knowledge Graphs: A Roadmap**
<https://www.arxiv-vanity.com/papers/2306.08302/>

LLMs and Knowledge Graphs

- ▶ Large Language Models (LLMs) are based on correlations
- ▶ LLMs are trained using unsupervised learning
- ▶ LLMs are black-box models
- ▶ LLMs fall short of capturing and accessing referenceable factual knowledge
- ▶ KGs provide semantic incl. causal relationships
- ▶ Knowledge models are built using supervised learning
- ▶ Knowledge Graphs (KGs) are based on explicit, structured knowledge models
- ▶ KGs provide referenceable rich factual knowledge

Can Large Language Models **Infer Causation from Correlation**?



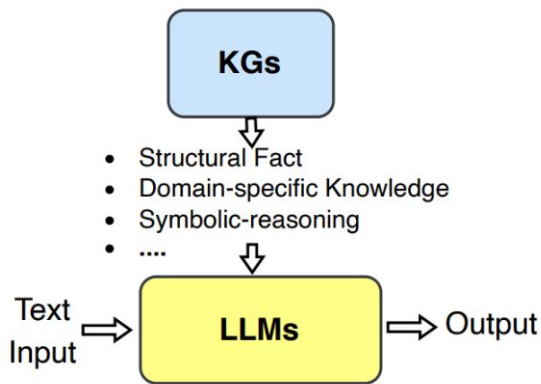
What is the relationship between deglaciation and sea level rise?

Could an increase of sales of skis help sea levels to fall again?

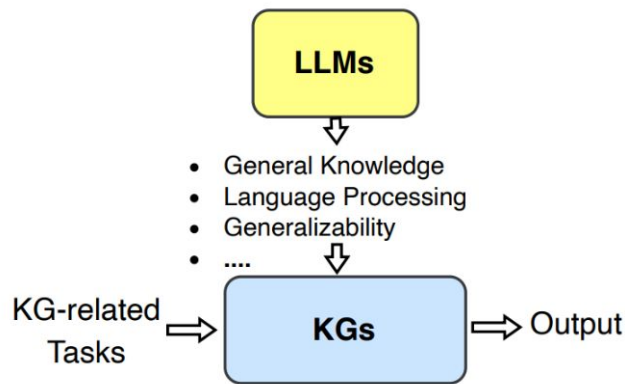
- A. **KG-enhanced LLMs**, which incorporate KGs during the pre-training and inference phases of LLMs, or for the purpose of enhancing understanding of the knowledge learned by LLMs;
- ➡ B. **LLM-augmented KGs**, that leverage LLMs for different KG tasks such as embedding, completion, construction, graph-to-text generation, and question answering; and
- C. **Synergized LLMs + KGs**, in which LLMs and KGs play equal roles and work in a mutually beneficial way to enhance both LLMs and KGs for bidirectional reasoning driven by both data and knowledge

PP Meets
ChatGPT

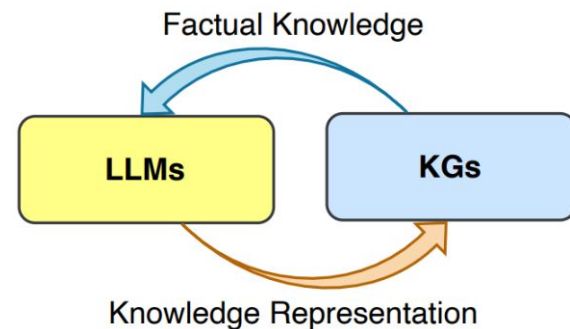
Combining LLMs with KGs



a. KG-enhanced LLMs



b. LLM-augmented KGs



c. Synergized LLMs + KGs

Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, Xindong Wu:
Unifying Large Language Models and Knowledge Graphs: A Roadmap (2023)

Knowledge Graphs can help to create training data with less effort.



Risks of LLMs without a Knowledge Graph

1. Enterprise have starting using LLMS as a KB, but that is not a good idea, because there are no reference points. It can be misused internally.
2. Enterprises will run risk of running errors., based on how LLMs are generated
3. LLM can eventually degenerate in results when it trains on its own content.
4. Governance is needed.

[“Intelligent Content - Bringing Together Knowledge Graphs and Large Language Models,”](#) by Helmut Nagy, Text Analytics Forum, November 8, 2023.

[“Responsible AI Based on LLMs,”](#) by Andreas Blumauer and Michael Iantosca, July 14, 2023, Webinar recording.

[“Augmenting large language models with knowledge graphs for effective, responsible, and explainable AI \(XAI\),”](#) by Michael Iantosca, June 2023.

[“Unifying Large Language Models and Knowledge Graphs: A Roadmap,”](#) by Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, Xindong Wu. *Journal of LATEX Class Files*, Vol. 14., No. 8, August 2021.

[“Knowledge Graphs and Taxonomies,”](#) by Heather Hedden, July 31, 2023, The Accidental Taxonomist Blog

[*The Knowledge Graph Cookbook*](#), by Andreas Blumauer and Helmut Nagy, 2020, free ebook

Questions/Contact

Heather Hedden

Communications Manager/Knowledge Manager

Semantic Web Company Inc.

One Boston Place, Suite 2600

Boston, MA 02108

857-400-0183

heather.hedden@semantic-web.com

www.linkedin.com/in/hedden

<http://accidental-taxonomist.blogspot.com>

Buy *The Accidental Taxonomist*, 3rd ed. In the KMWorld book store, near registration.

Visit the PoolParty booth in the KM World Showcase.

Semantic Web Company www.semantic-web.com

PoolParty software www.poolparty.biz

