



Breakthrough Moments in Enterprise Taxonomy Management

Dave Clarke
Founder, Synaptica

Synaptica has been collaborating with businesses and organizations around the world for over twenty-five years to solve the evolving challenges of enterprise taxonomy management and semantic categorization



publishing & media
academic & libraries
corporate, financial, & legal
science, technology, & engineering
medicine & pharmacy
video games
social media
eCommerce
food delivery
hospitality

Complexity

Scale

Explainability

Trust

Maximus



Huxley

Maximus & Huxley, Pembrokeshire Coast, 2023

Maximus

Guard

- 11 % Neapolitan Mastiff
- 10 % Cane Corso
- 4 % Perro de Presa Canario
- 3 % Doberman Pinscher

Sporting

- 8 % Labrador Retriever
- 5 % German Shorthaired Pointer
- 3 % Istrian Short-Haired Hound
- 2 % Wirehaired Vizsla
- 2 % Irish Setter
- 2 % Gordon Setter

Herding

- 14 % German Shepherd Dog
- 3 % Berger d'Auvergne
- 2 % Kuvasz
- 2 % Puli

Hound

- 14 % Segugio Italiano
- 2 % Black and Tan Coonhound

Terrier

- 6 % Chihuahua
- 5 % Peruvian Inca Orchid
- 2 % Danish Swedish Farmdog



Huxley

Herding

- 34 % German Shepherd Dog
- 3 % Belgian Malinois
- 3 % Catalan Sheepdog
- 3 % White Swiss Shepherd

Terrier

- 16 % Chihuahua
- 7 % Peruvian Inca Orchid
- 4 % Danish Swedish Farmdog
- 2 % Dachshund

Hound

- 11 % Segugio Italiano

Mountain Dogs

- 6 % Estrela Mountain Dog

Middle Eastern and African

- 2 % Bulgarian Shepherd
- 2 % Sloughi

Sporting

- 3 % German Shorthaired Pointer

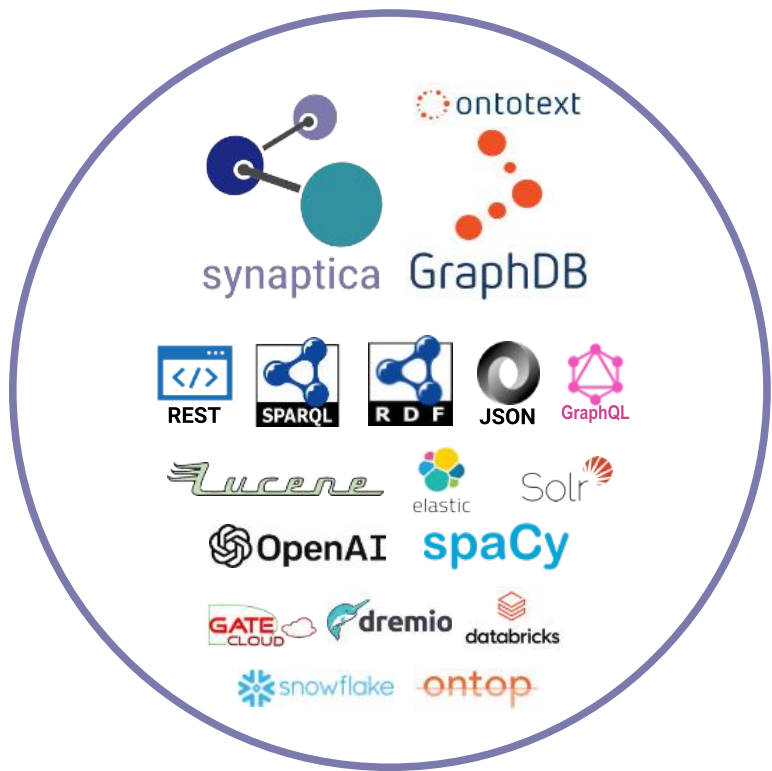
Street Dogs

- 2 % Honduran Street Dog

Guard

- 2 % American Pit Bull Terrier

Maximus & Huxley, Pembrokeshire Coast, 2023



Graphite-GraphDB developer eco-system

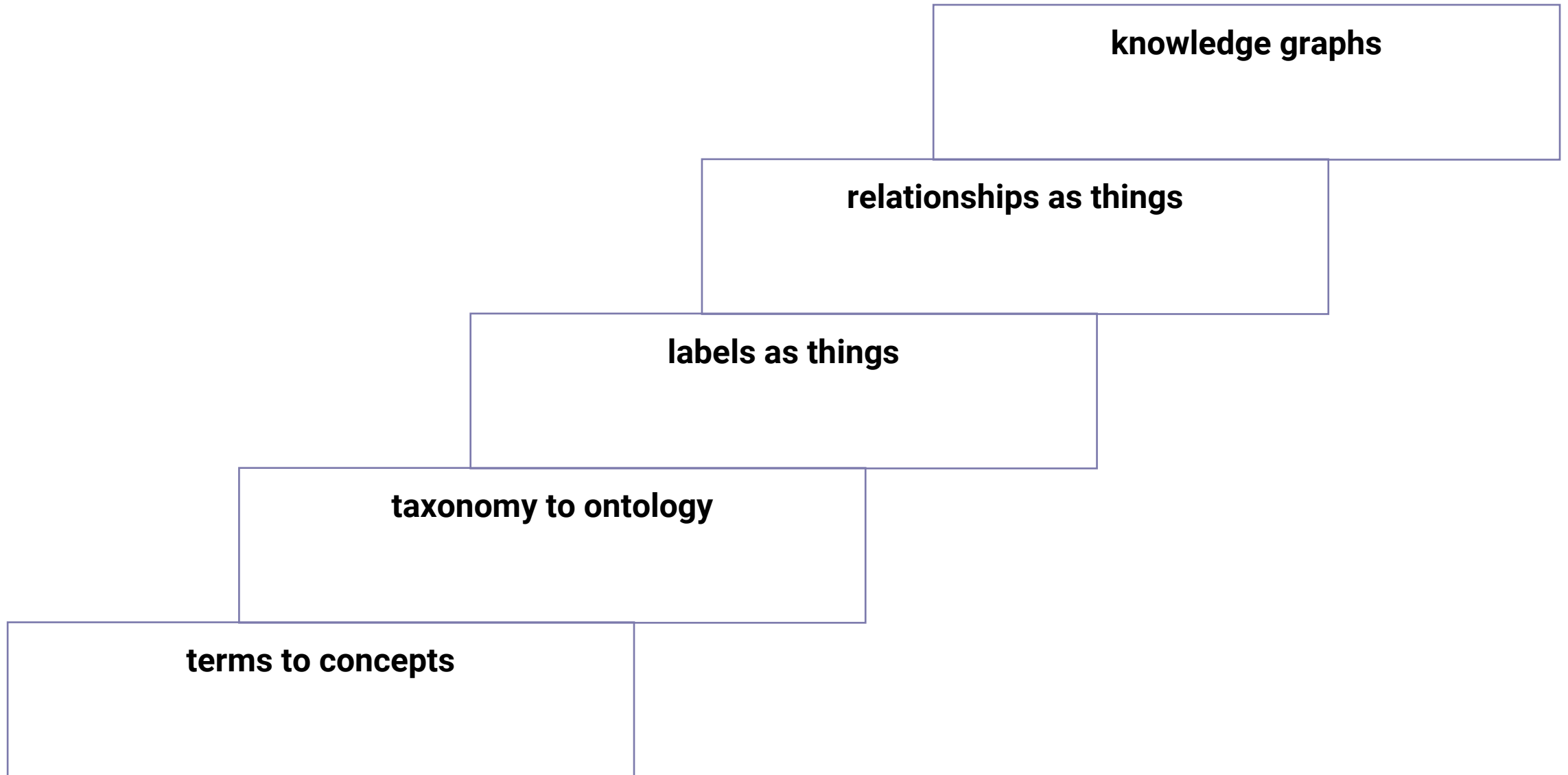
stay flexible

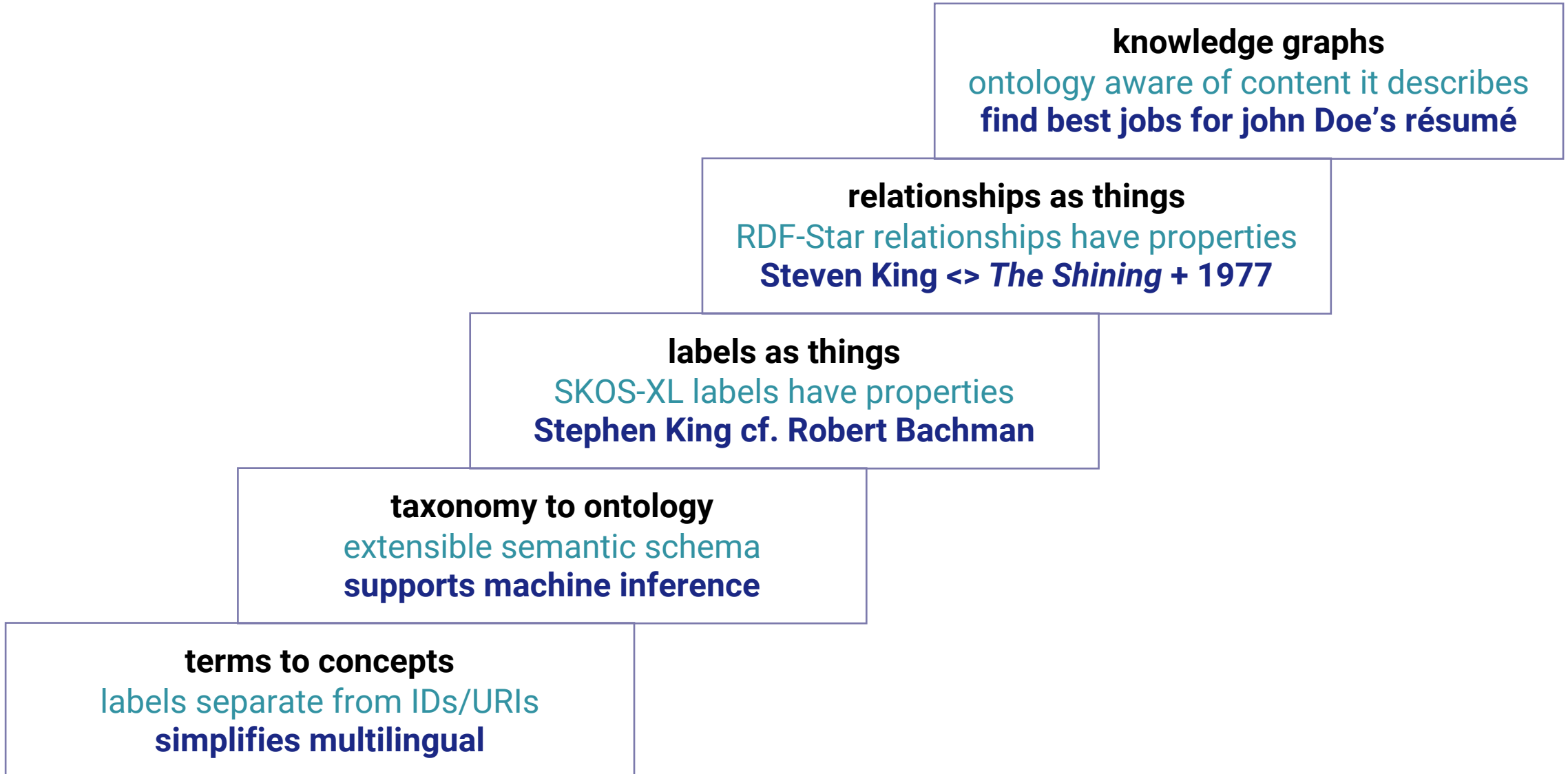
use more than one language model
or NLP service



Synaptica

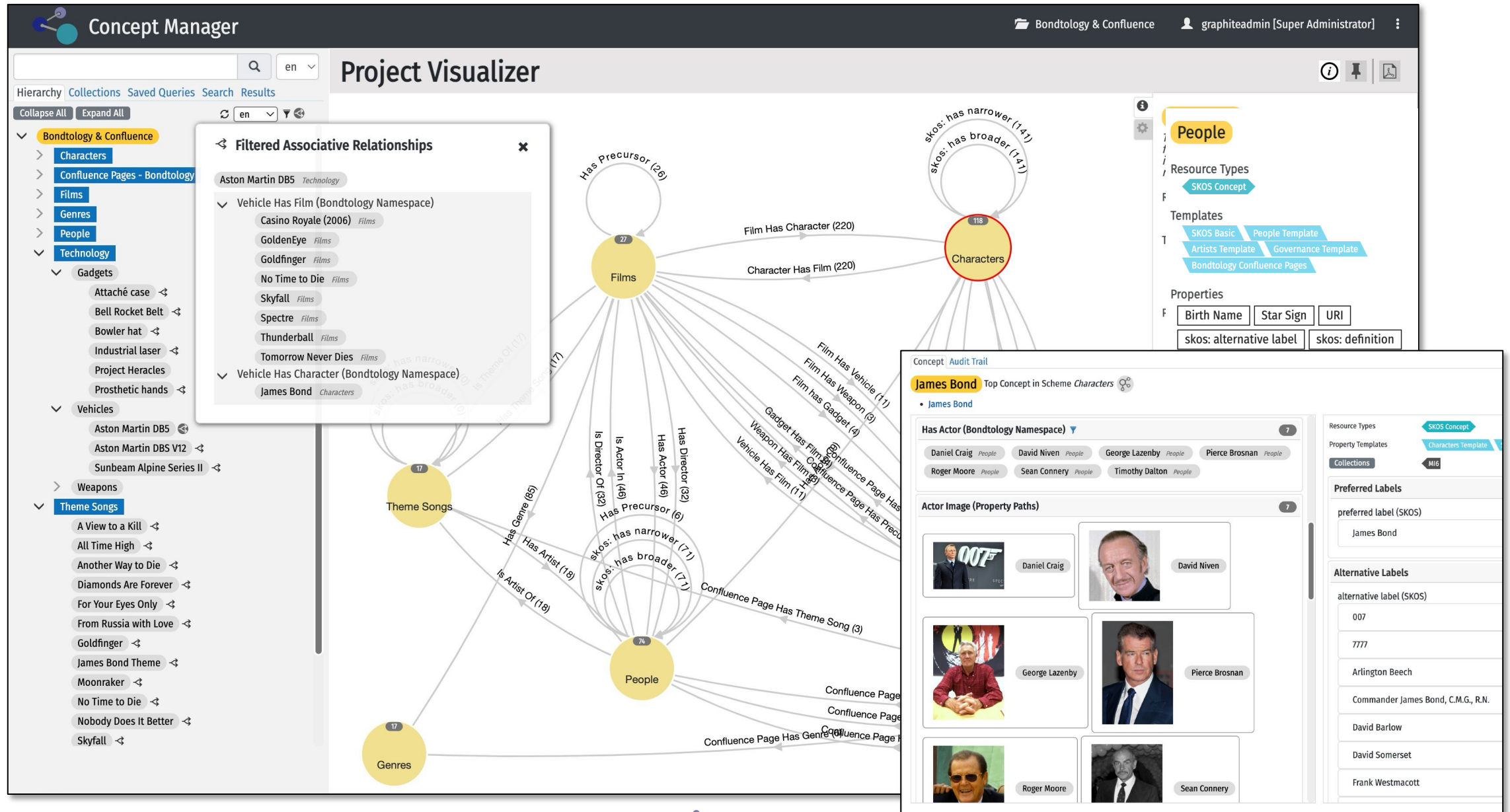
Microsoft Bing Image Creator powered by DALL.E3
Created 20231031 from text prompt '**knowledge graph**'





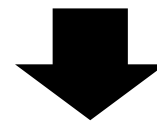


Complexity Breakthrough: UI to make the Complex Comprehensible





Microsoft Bing Image Creator powered by DALL.E3
Created 20231031 from text prompt '**extreme scale data**'



10K Concepts



Microsoft Bing Image Creator powered by DALL.E3
Created 20231031 from text prompt '**extreme scale data**'

10K Concepts

100K Concepts



Microsoft Bing Image Creator powered by DALL.E3
Created 20231031 from text prompt '**extreme scale data**'

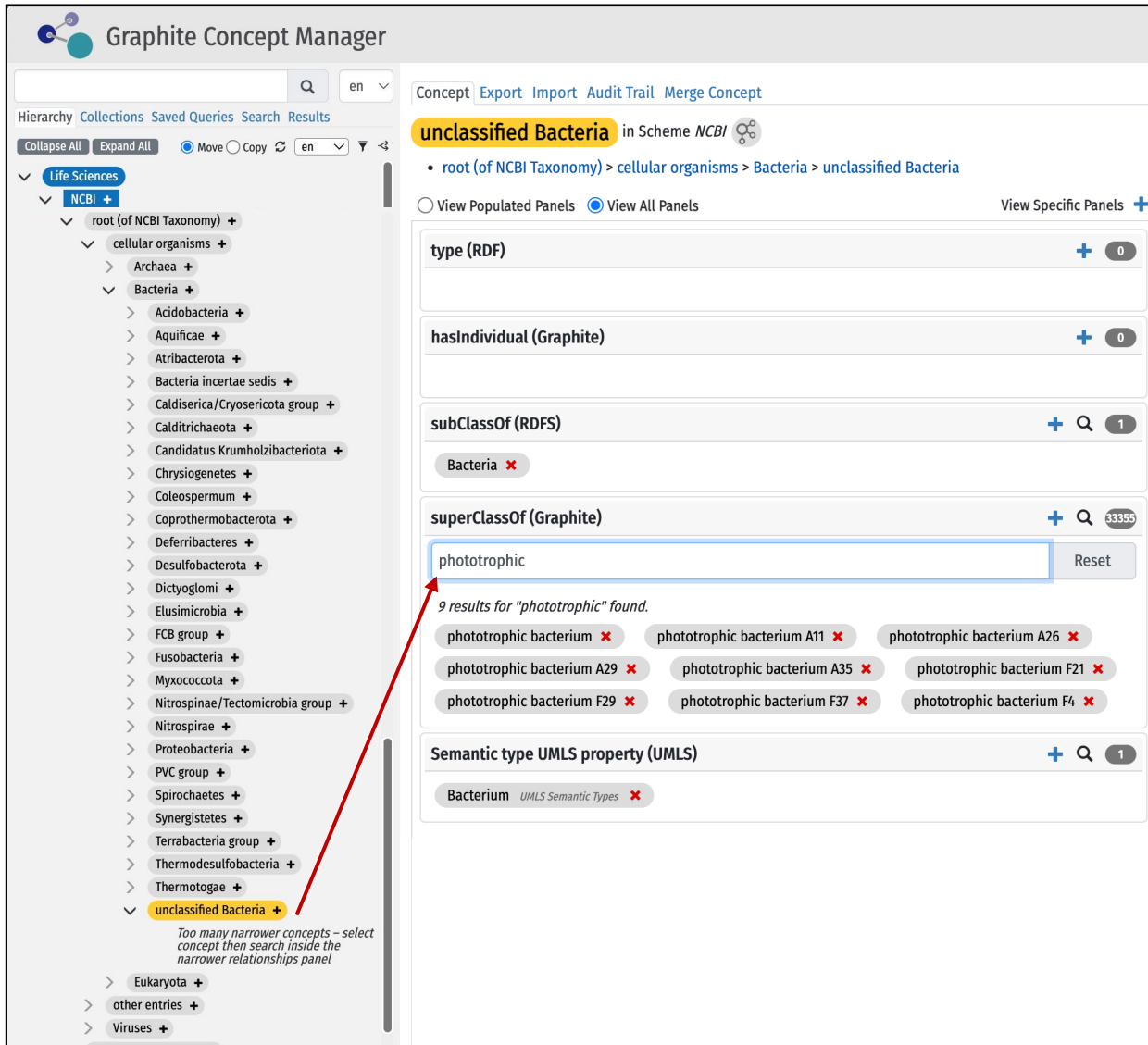
10K Concepts

100K Concepts

10M Concepts

Example:

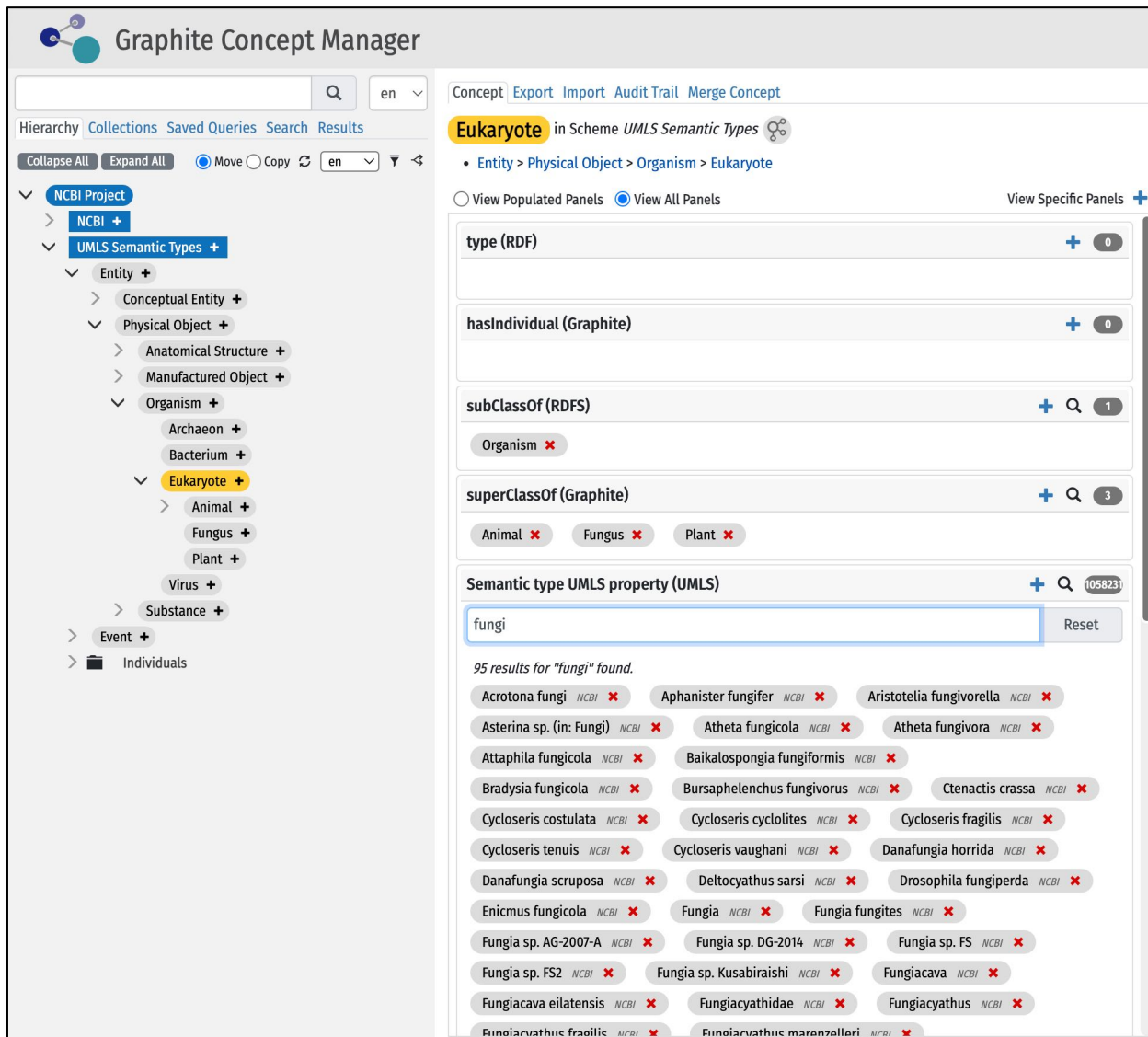
- Hierarchical navigation through a taxonomy of over **2 million concepts**
- **Click-click-click** down three levels and then hit a concept with over **33,000** narrower concepts
- Adaptive navigation traps the condition in a split second and redirects the user to **search inside** the set of narrower concepts



The screenshot displays the Graphite Concept Manager interface. On the left, a hierarchical tree structure shows the navigation path: Life Sciences > NCBI > root (of NCBI Taxonomy) > cellular organisms > Bacteria > unclassified Bacteria. The 'unclassified Bacteria' concept is highlighted in yellow. A red arrow points from this concept to the 'superClassOf (Graphite)' panel on the right. In this panel, the search term 'phototrophic' is entered, and a list of 9 results is shown, including 'phototrophic bacterium A11', 'phototrophic bacterium A26', 'phototrophic bacterium A29', 'phototrophic bacterium A35', 'phototrophic bacterium F21', 'phototrophic bacterium F29', 'phototrophic bacterium F37', and 'phototrophic bacterium F4'. The 'type (RDF)' panel shows 'type (RDF)' with a count of 0. The 'hasIndividual (Graphite)' panel shows 'hasIndividual (Graphite)' with a count of 0. The 'subClassOf (RDFS)' panel shows 'Bacteria' with a count of 1. The 'superClassOf (Graphite)' panel shows 'phototrophic' with a count of 33355. The 'Semantic type UMLS property (UMLS)' panel shows 'Bacterium' with a count of 1. A note at the bottom of the tree structure reads: 'Too many narrower concepts - select concept then search inside the narrower relationships panel'.

Example:

- Associative navigation
- Click on a concept with over **1 million** associative relationships...
- ... Adaptive Navigation traps the condition and redirects the user to ***search inside*** the set of related concepts



The screenshot displays the Graphite Concept Manager interface. On the left, a hierarchical tree shows the 'NCBI Project' expanded to 'UMLS Semantic Types', then 'Entity', 'Physical Object', and finally 'Organism', where 'Eukaryote' is selected. The main panel shows the 'Eukaryote' concept in the 'UMLS Semantic Types' scheme. Below the concept name, a breadcrumb trail indicates 'Entity > Physical Object > Organism > Eukaryote'. The interface includes tabs for 'Concept', 'Export', 'Import', 'Audit Trail', and 'Merge Concept'. A search bar is present at the top. The main content area displays several panels: 'type (RDF)', 'hasIndividual (Graphite)', 'subClassOf (RDFS)' (showing 'Organism'), and 'superClassOf (Graphite)' (showing 'Animal', 'Fungus', and 'Plant'). A 'Semantic type UMLS property (UMLS)' section is also visible, with a search input containing 'fungi' and a 'Reset' button. Below this, a list of 95 results for 'fungi' is shown, including various fungal species like 'Acrotona fungi', 'Aphanister fungifer', 'Aristotelia fungivorella', etc., each with a small 'NCBI' icon and a red 'x'.

UI also adaptive to **scale of relationship types**

- some taxonomies can have hundreds of relationship types
- the UI can adapt to display just populated relationships
- and expose additional properties on demand properties



SKOS 12
predicates

Enterprise 24
predicates

OMOP 662
predicates

Graphite manages high-performance search indexes over the graph.

Search response times dependent on two factors: *specificity of query* (more keywords faster), and *volume of results* (more results slower):

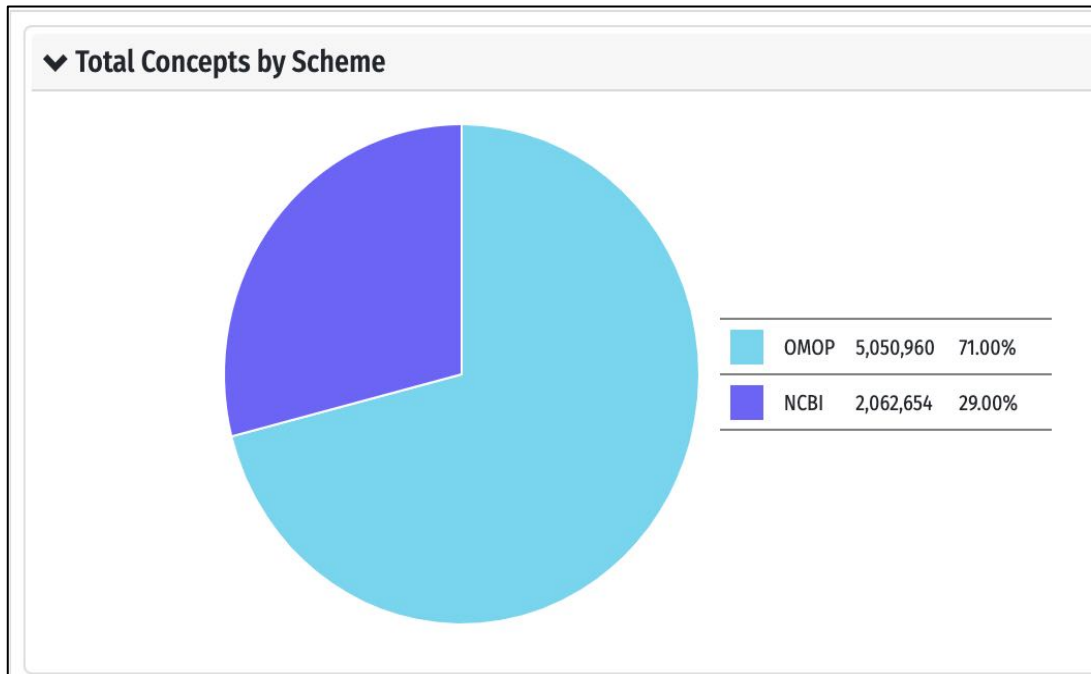
Examples (searching 7.1 million concepts across 2 taxonomies) including round trip to browser:

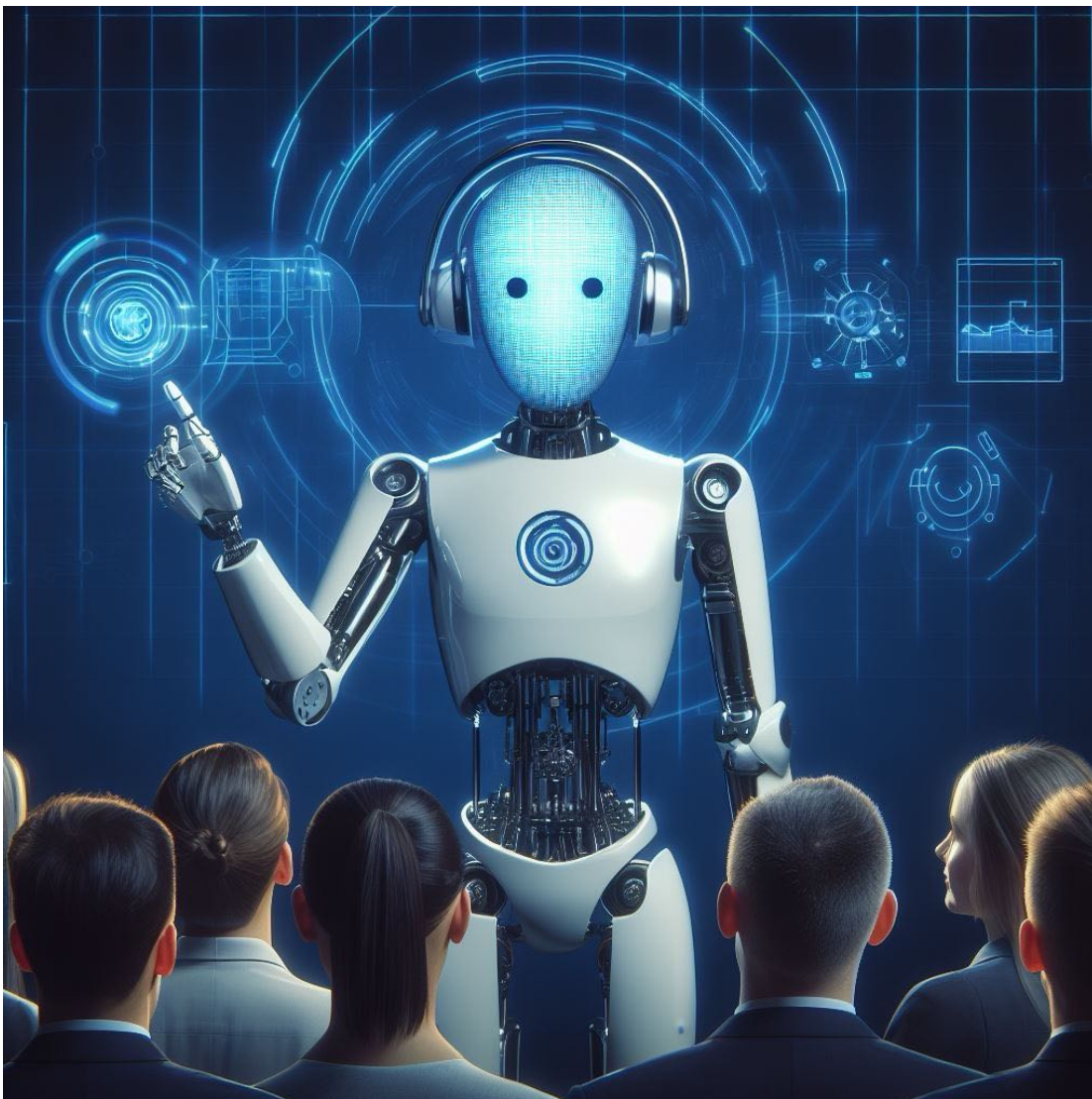
“bacteria” – 3,540 results in 3.5 seconds

“blood clot” – 352 results in 1.2 seconds

“kidney infection” – 45 results in 0.35 seconds

“gamma knife” – 10 results – 0.25 seconds





Three pillars of successful
autocategorization

explainable results

transparent rules

rapid iteration

Projects > Tag Essentials Project > CVs > <http://example.com/fa5a06e2-5fcb-5e35-b534-340ae8c95336>

DOCUMENT ANNOTATIONS

New document annotation

- Company (1)
 - Decisions 0.165
- JobRole (2)
 - Project/Program Managers 0.578
 - Project Manager 1.155
- Location (3)
 - United States 0.099
 - Illinois 0.199
 - Time 0.298
- Occupations (35)
 - project manager 2.888
 - Business services and administration managers not elsewhere classified 2.311
 - Managers 1.766

ANNOTATION FILTERS

By type By creator

DOCUMENT LEVEL

- Company 1
- JobRole 2
- Location 3
- Occupations 35
- Recommendation 10
- Skills 21

INLINE LEVEL

- Lookup 51

Borislav Ankov

Resource Manager at Ontotext

Previously Project Manager at Ontotext

I am a highly motivated professional with interest in Project Management, process engineering and team collaboration.

During my 5 years at Ontotext, I've achieved the following competences:

Ability to interact with internal and external stakeholders at all levels and able to lead technical discussions to influence decisions or come to resolution.

Planning, estimating, and managing the project lifecycle from Initiation through Closing. I consider myself especially proficient in scheduling and execution of all Software deliverables within the scope of the project or program. Usually I'm in charge of maintaining project schedule and timeline for multiple projects competing for resources at any time.

Over 2 years ago, I achieved the PMP certification, thus I believe this demonstrates knowledge of industry-recognized project management methodologies.

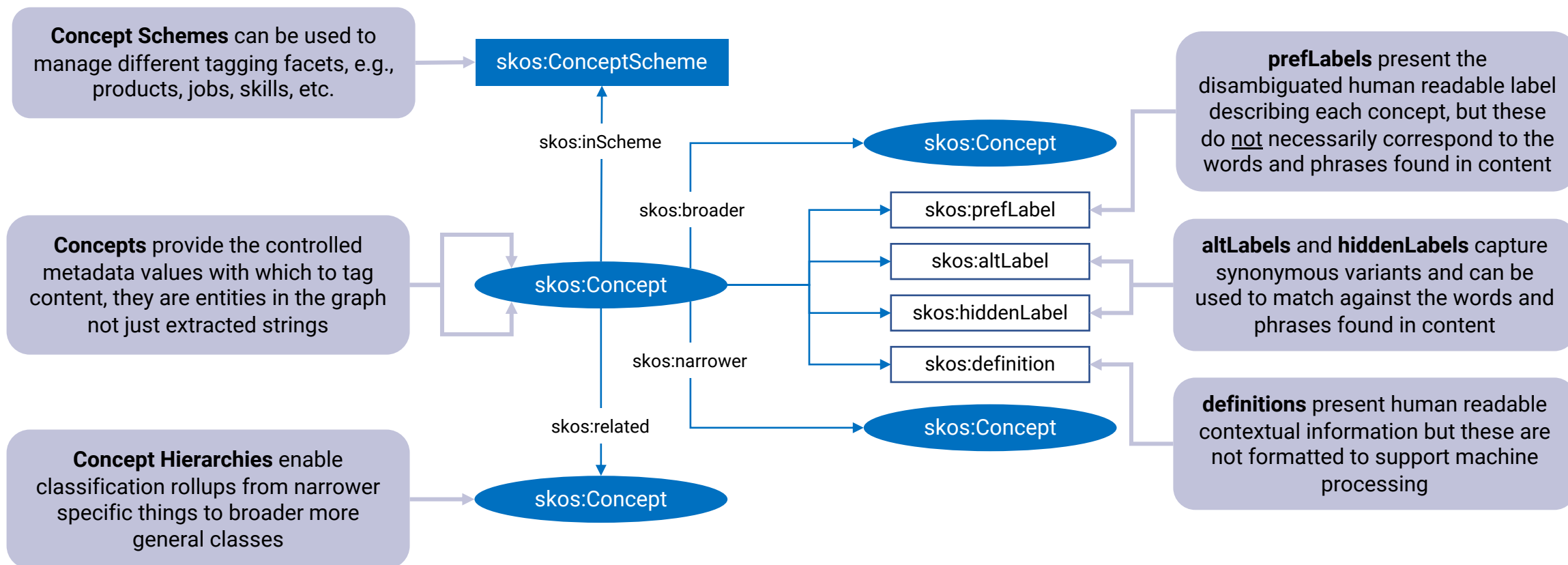
Recently I had to introduce another Project Manager to the team and help with the induction - providing guidance and mentoring to those Professional Service resources working on their implementations or my own peers is something I very much

Explainability Breakthrough: Transparent Rules

SKOS has become the go-to schema for managing enterprise taxonomies.



How can unmodified SKOS be used to support auto-categorization?



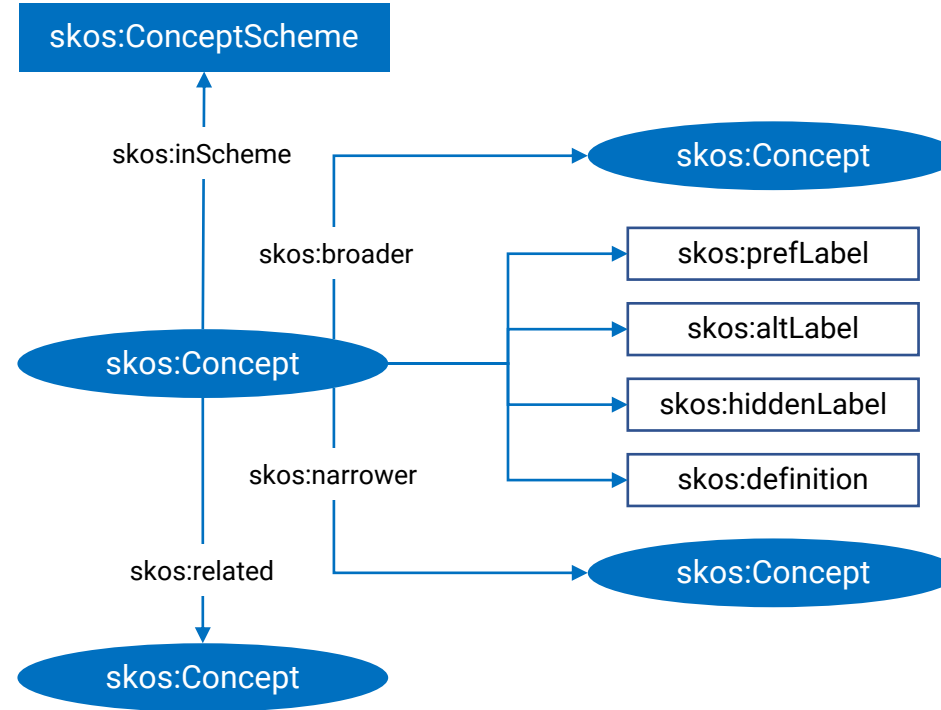
Explainability Breakthrough: **Transparent Rules**



Can manage multiple tagging facets using different concept schemes

Can increase candidate tagging by adding altLabels that match words in content

Can support classification rollups from specific to general things

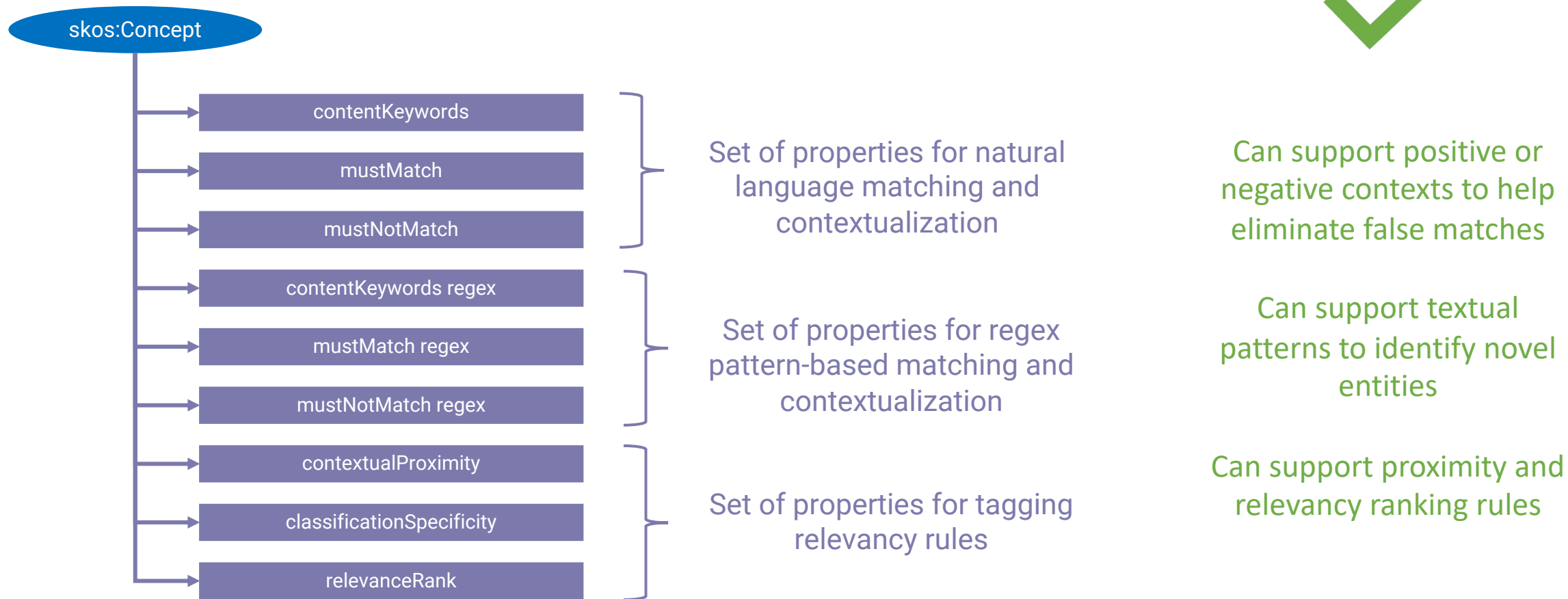


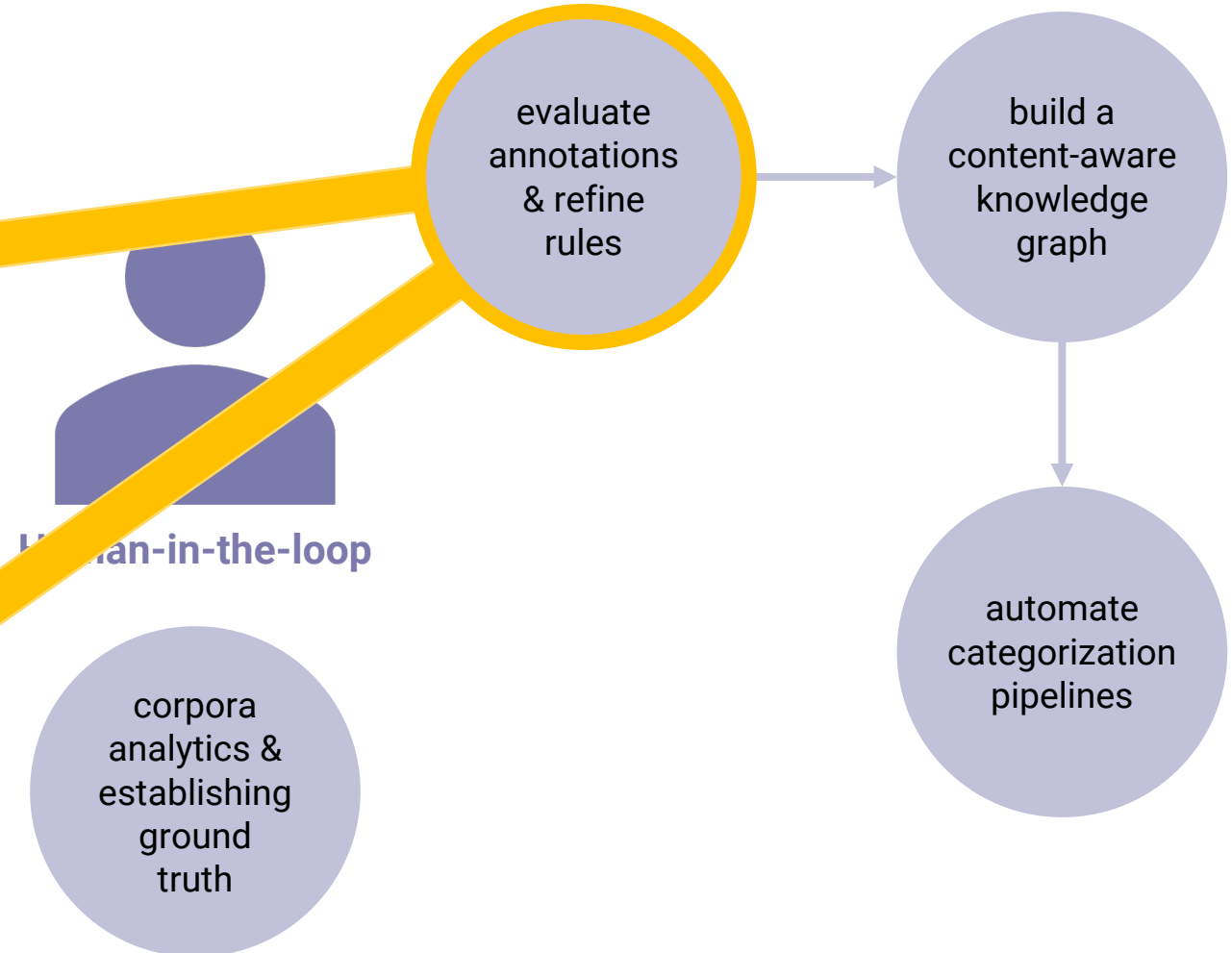
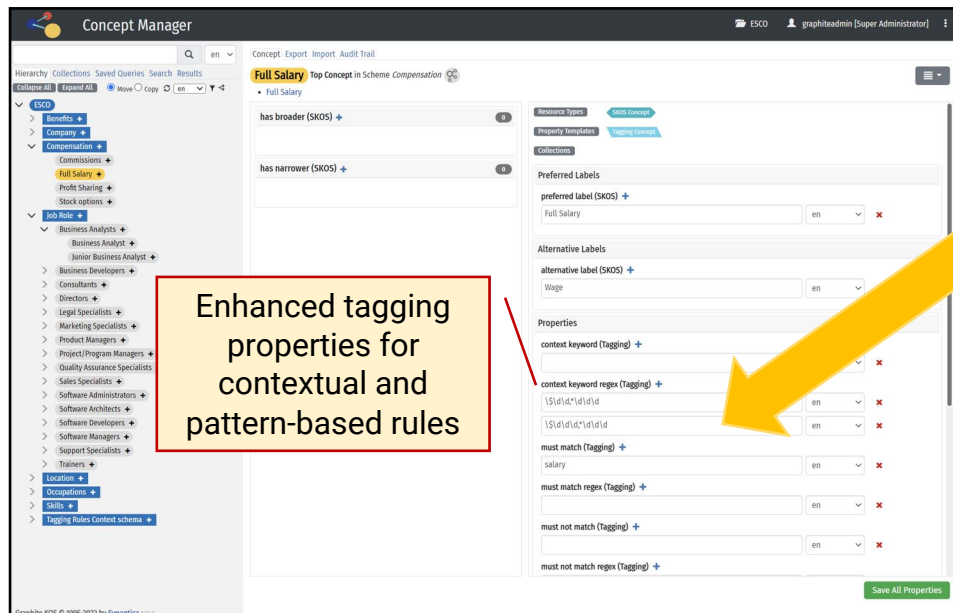
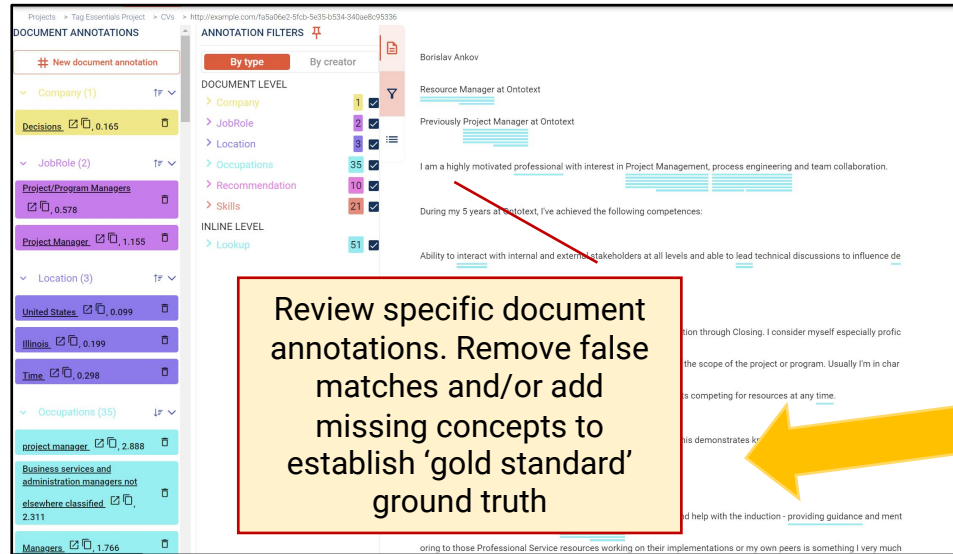
Can't support positive or negative contexts to help eliminate false matches

Can't support textual patterns to identify novel entities

Can't support proximity and relevancy ranking rules

Explainability Breakthrough: **Transparent Rules**







Generative AI

data privacy

opaqueness of sources

veracity of results



SquirroGPT

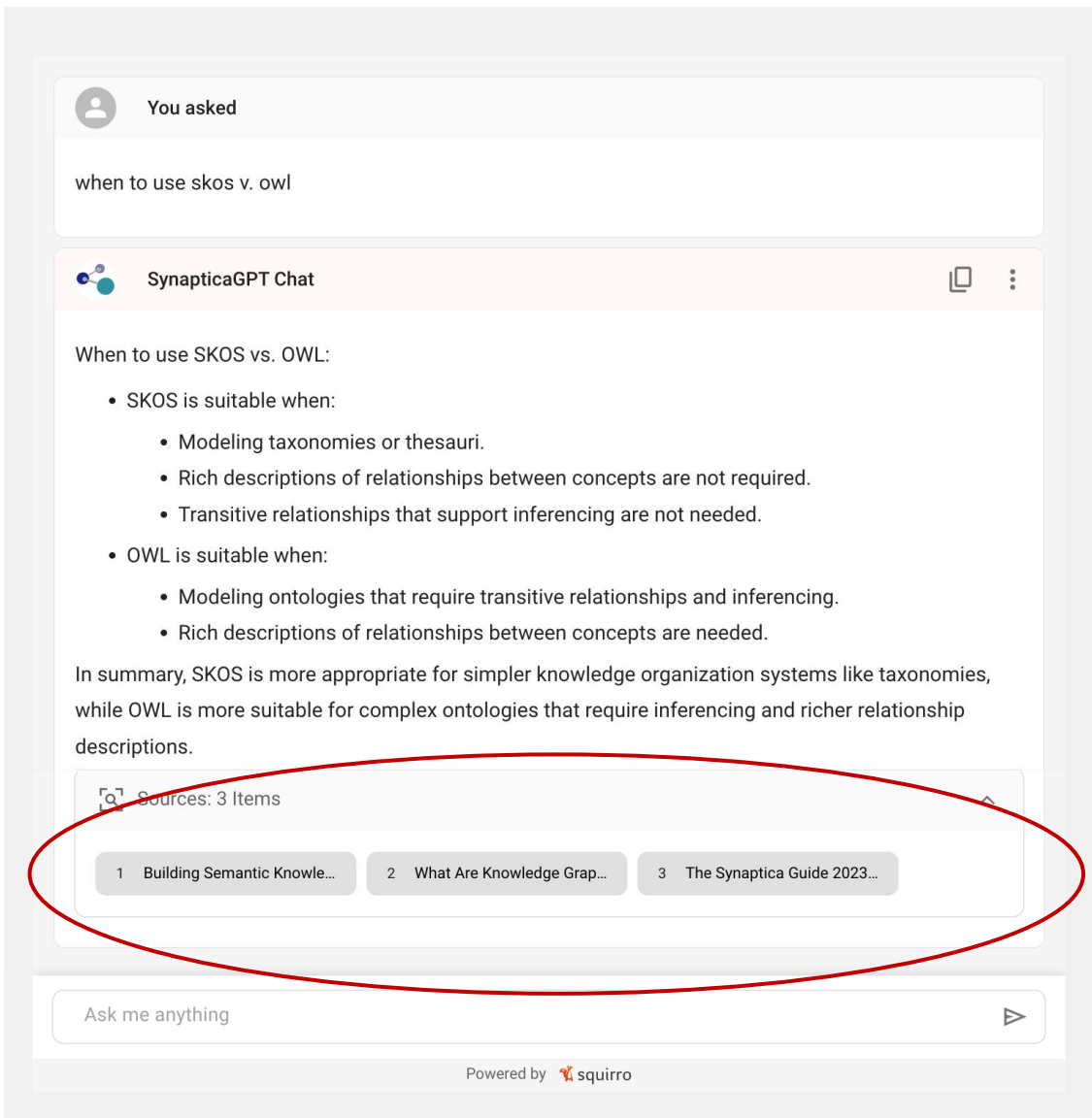
Start chatting with **YOUR** data

No Hallucinations: **Evidence**-based answers

Your Data: **Relevant** answers that matter to you

Secure: **Enterprise**-grade security

Trust Breakthrough: **Cited Sources**



The screenshot shows a chat interface for SynapticaGPT. The chat history shows a user asking "when to use skos v. owl" and the chatbot responding with a detailed explanation of when to use SKOS vs. OWL. The chatbot's response includes a list of bullet points and a summary. Below the chatbot's response, there is a section titled "Sources: 3 Items" which is circled in red. This section contains three items: "1 Building Semantic Knowle...", "2 What Are Knowledge Grap...", and "3 The Synaptica Guide 2023...". At the bottom of the interface, there is a text input field labeled "Ask me anything" and a "Powered by squirro" logo.

You asked

when to use skos v. owl

SynapticaGPT Chat

When to use SKOS vs. OWL:

- SKOS is suitable when:
 - Modeling taxonomies or thesauri.
 - Rich descriptions of relationships between concepts are not required.
 - Transitive relationships that support inferencing are not needed.
- OWL is suitable when:
 - Modeling ontologies that require transitive relationships and inferencing.
 - Rich descriptions of relationships between concepts are needed.

In summary, SKOS is more appropriate for simpler knowledge organization systems like taxonomies, while OWL is more suitable for complex ontologies that require inferencing and richer relationship descriptions.

Sources: 3 Items

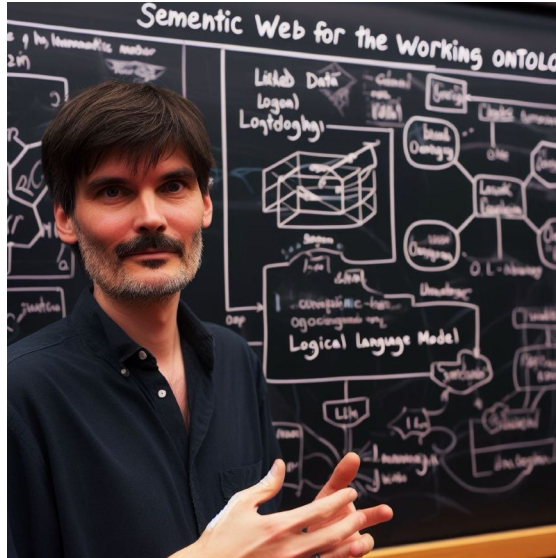
- 1 Building Semantic Knowle...
- 2 What Are Knowledge Grap...
- 3 The Synaptica Guide 2023...

Ask me anything

Powered by squirro

Microsoft Bing Image Creator powered by DALL.E3 created three of these images on 20231107 from the text prompt 'Dean Allemang explaining LLMs'

The Unusual Suspects



Would the real Dean Allemang please step forward

Three memorable takeaways from Dean's
Taxonomy Bootcamp Keynote (my paraphrase)

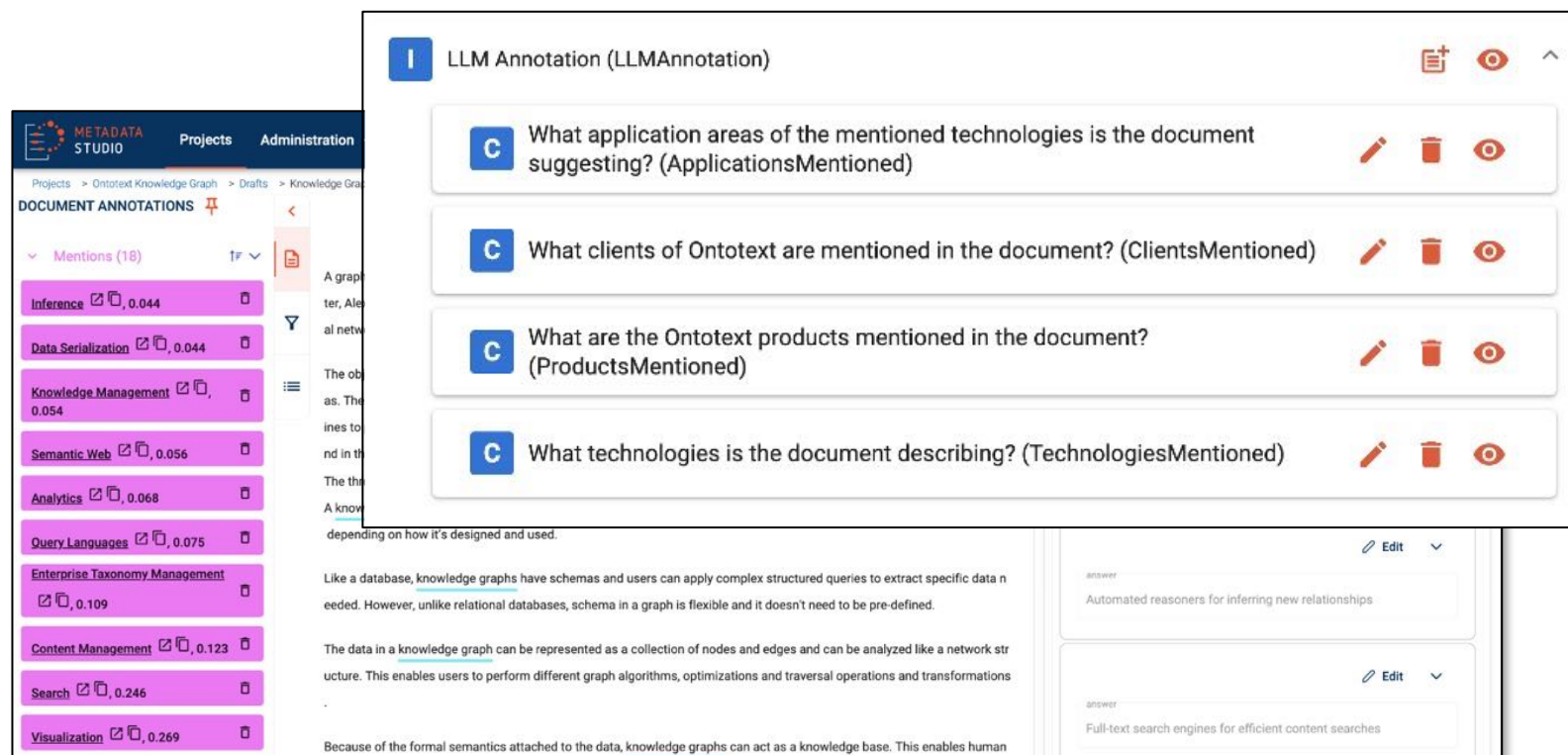
Who understands why we need taxonomy?
ChatGPT does.

Combining LLMs with an ontology can massively
boost accuracy (over 37% in a specific test case).

You won't lose your job to an AI, you'll lose it to a
person who knows how to use AI ...
... a call to action for all taxonomists.



After a major development effort by the engineers at Ontotext, the latest version of **Graphite Knowledge Studio** now supports configurable pipelines to LLMs for GenAI-based annotation and extraction.



The innovation challenge to combine GenAI with taxonomy and knowledge graphs has only just begun.

Keep your tech stack diverse, free your imagination, be bold.



Microsoft Bing Image Creator powered by DALL.E3 Created 20231107
from text prompt '**ship of creative innovation and exploration**'

Innovate with us



info@synaptica.com



<https://www.synaptica.com/>



<https://www.twitter.com/synaptica>



<https://www.linkedin.com/synaptica-llc>



<https://www.youtube.com/c/Synaptica>

Thank You!