



# Bootstrapping enterprise taxonomies for autocategorization



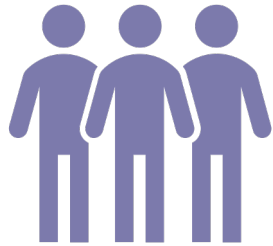
Sarah Downs



## Our topics today

- Introduction & Overview
- SKOS Taxonomies
  - Use in human tagging
  - Extensibility for machine tagging
- A human-in-the-loop workflow
  - Transparency and Explainability
  - Iterative refinement and enrichment
- Recap & Close

Information  
Science



Human Curated  
Top-Down Description  
Human Readable  
***Defines*** Knowledge

Data  
Science



Computational  
Bottom-Up Extraction  
Machine Readable  
***Infers*** Knowledge



# Marrying data science & information science



20+ years of helping customers understand *unstructured data* with *computational data science* approaches

RDF graph database  
**GraphDB & Metadata studio**



Taxonomy & Ontology  
**Graphite**

25+ years of helping customers manage *controlled vocabularies* through *human-curated information science*

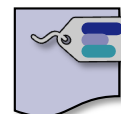
## Graphite Knowledge Studio

### Content-aware Knowledge graph



#### Tagging

identifies the **many** taxonomy concepts and named entities that are **mentioned** within the full text of a document.



#### Categorization

identifies the **few** concepts and named entities that best describe the **aboutness** of a whole document.



#### Extraction

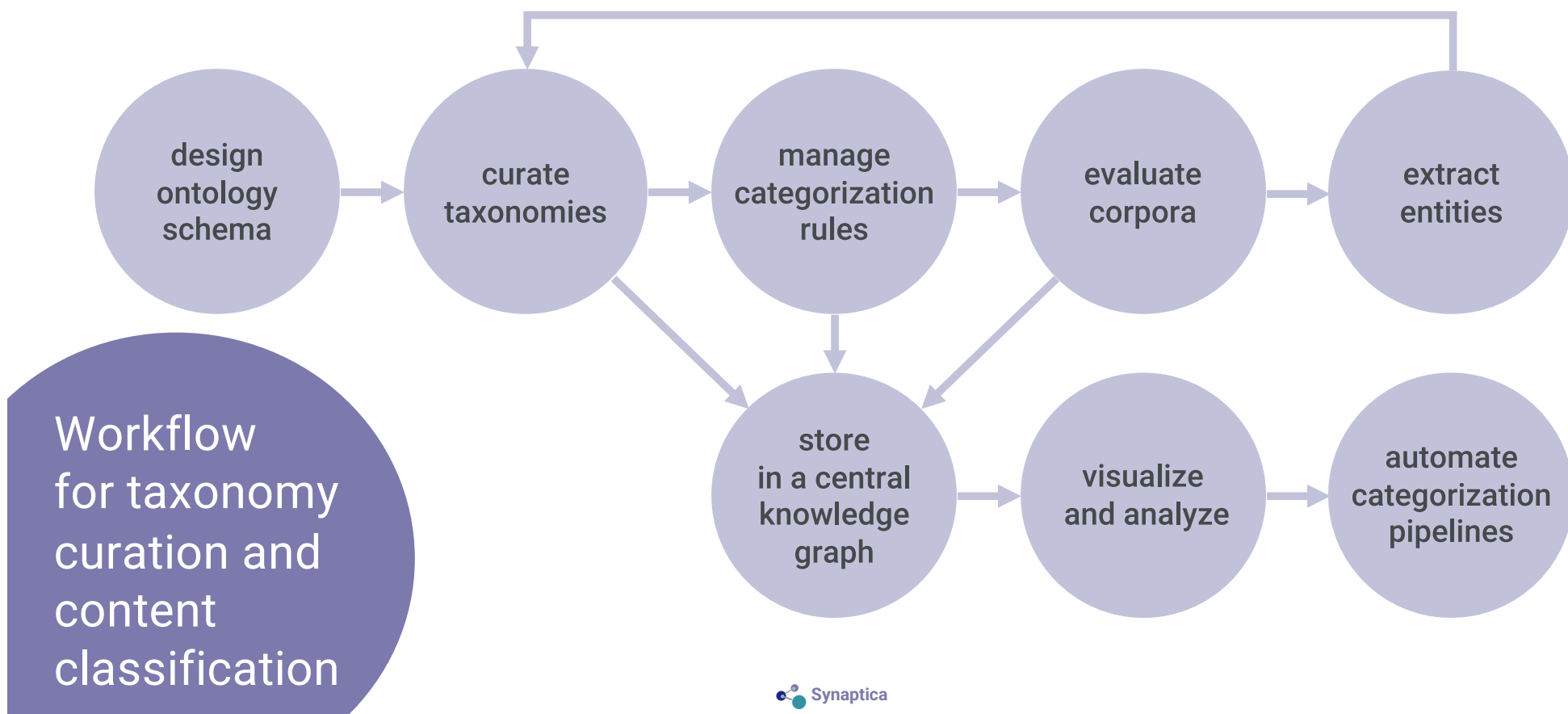
Identifies the **new** concepts and named entities that are found in the full text of a document but **not** in the taxonomy.



#### Similarity

supports **discovery** of content trends and connection through **similarity** matching across the knowledge graph.

# Taxonomies to Tags (and back again)





# Human tagging today



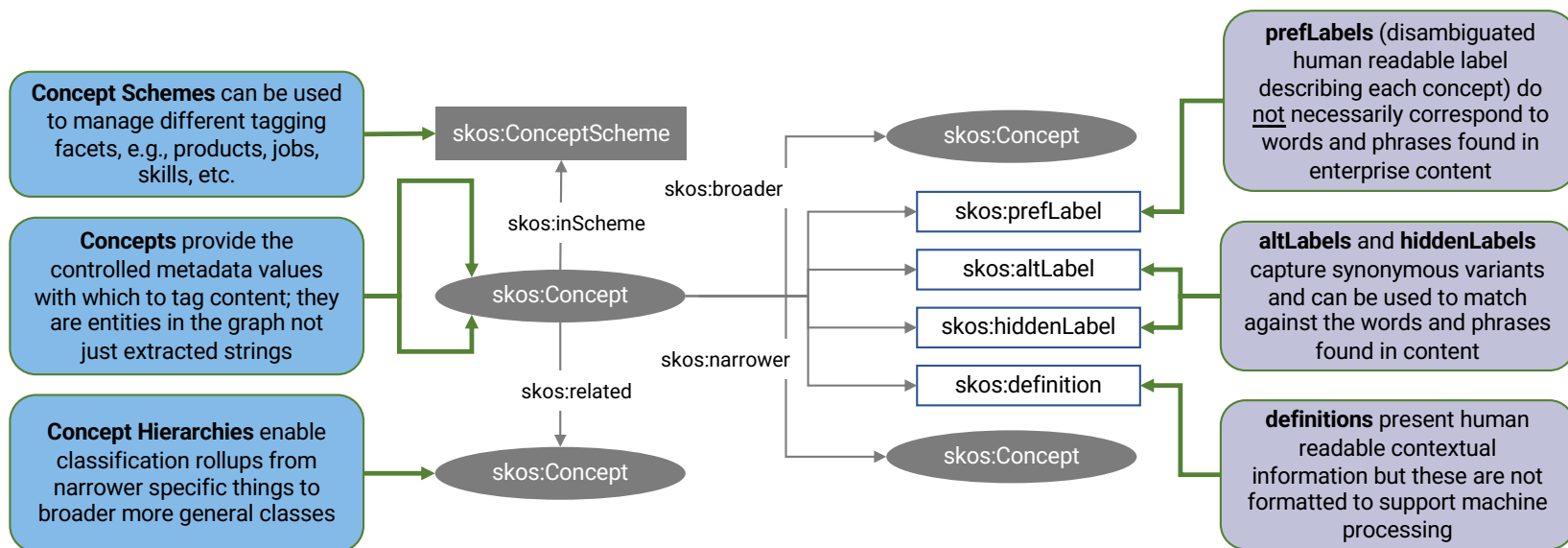
*AI Art generated with Adobe Firefly*

# Unmodified SKOS for autocategorization

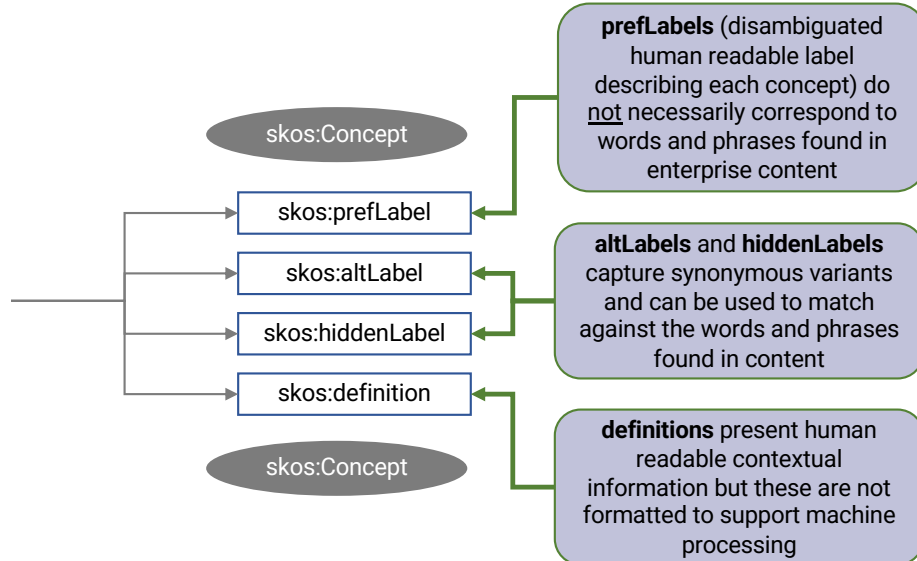
SKOS has become the go-to schema for managing enterprise taxonomies.



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



# Unmodified SKOS for autocategorization



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



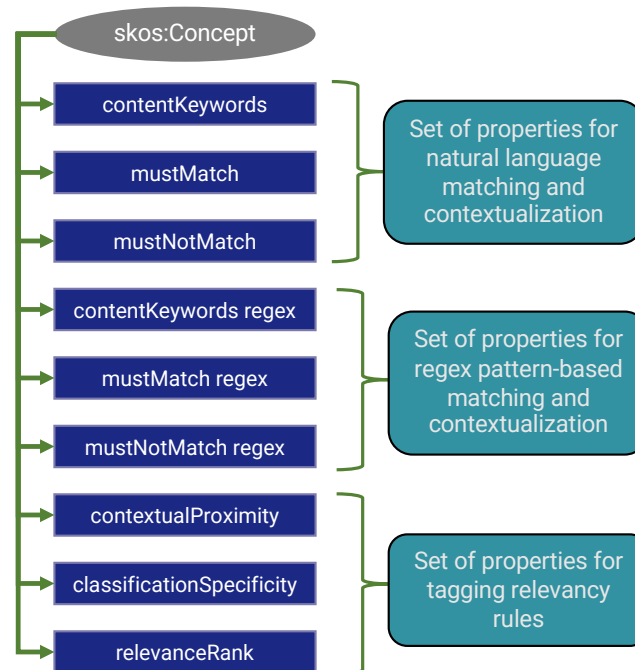
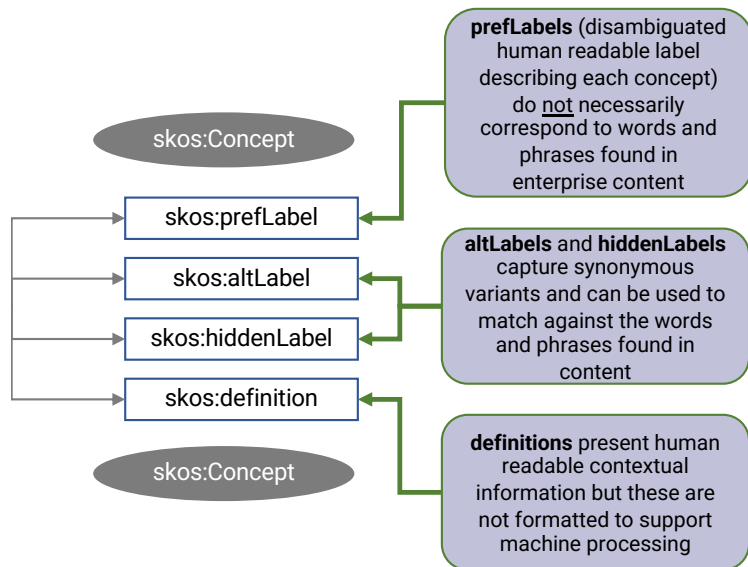
Can't support positive or negative contexts to help eliminate false matches

Can't support textual patterns to identify novel entities

Can't support proximity and relevancy ranking rules



# Enriched SKOS for autocategorization



Can manage multiple tagging facets using different concept schemes

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

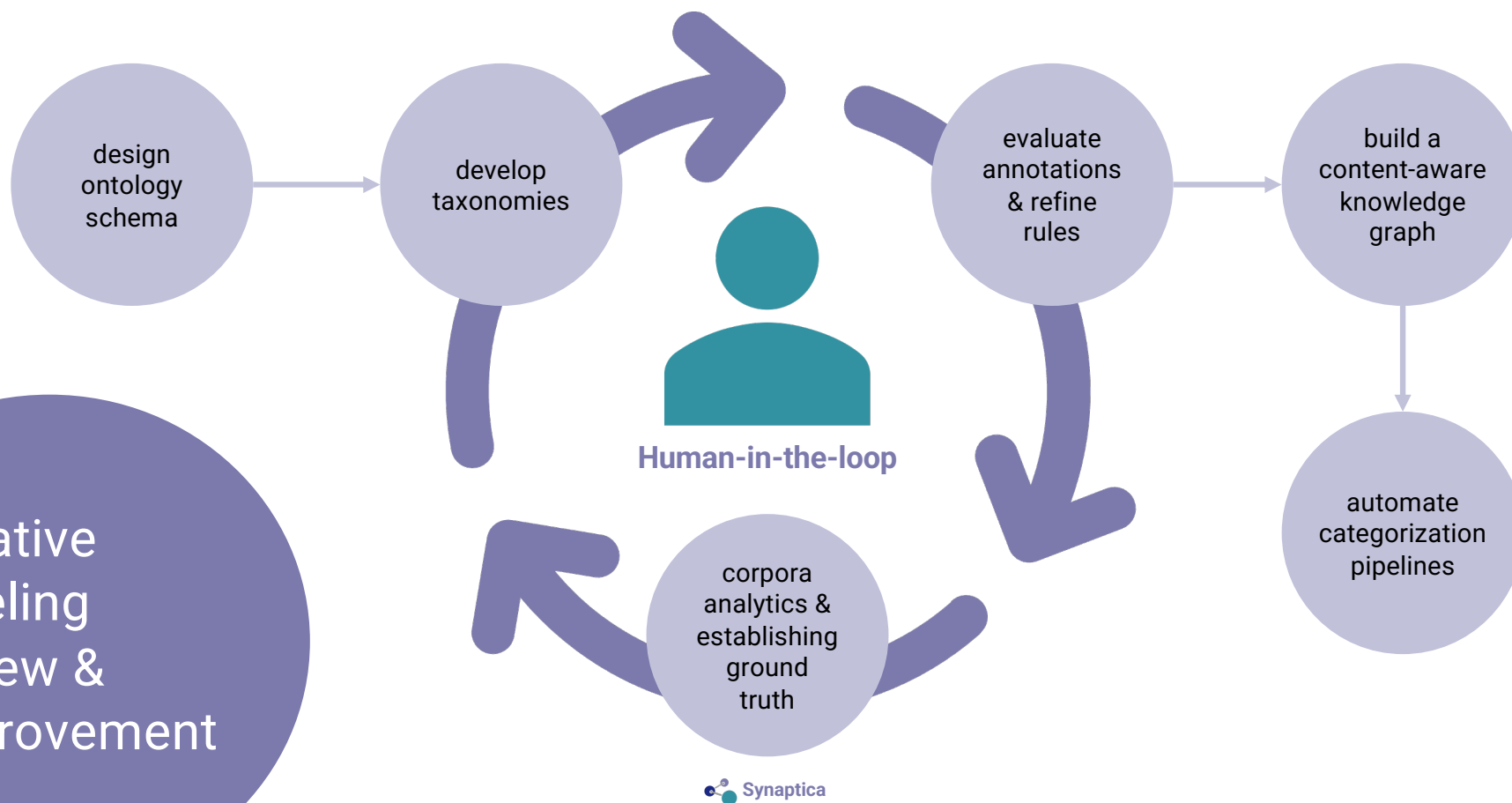
Can support classification rollups from specific to general things

Can support positive or negative contexts to help eliminate false matches

Can support textual patterns to identify novel entities

Can support proximity and relevancy ranking rules

# Taxonomies to Tags (and back again)





# The key: transparency & explainability

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 ☒

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:

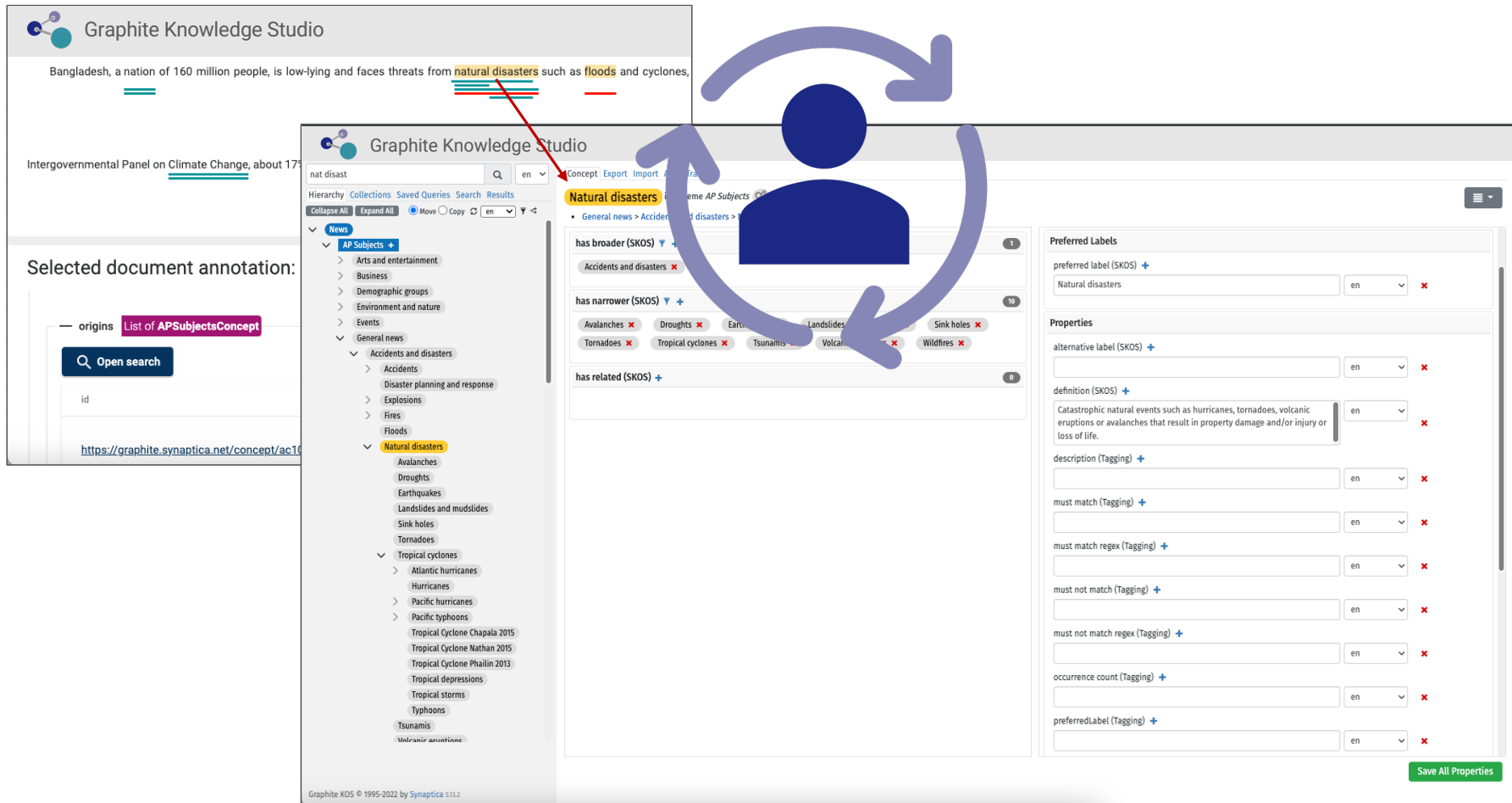
Ability to interact with internal and external stakeholders at all levels and able to lead technical discussions to influence decisions or come to resolution.

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 I 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

# The key: transparency & explainability



**Graphite Knowledge Studio**

Bangladesh, a nation of 160 million people, is low-lying and faces threats from **natural disasters** such as **floods** and cyclones,

Intergovernmental Panel on **Climate Change**, about 17

**Selected document annotation:**

— origins **List of APSubjectsConcept**

**Open search**

<https://graphite.synaptica.net/concept/ac1>

**Graphite Knowledge Studio**

nat disast

Hierarchy Collections Saved Queries Search Results

Expand All Collapse All Move Copy en

**News**

AP Subjects +

- Arts and entertainment
- Business
- Demographic groups
- Environment and nature
- Events
- General news
- Accidents and disasters
  - Accidents
  - Disaster planning and response
  - Explosions
  - Fires
  - Floods
  - Natural disasters**
    - Avalanches
    - Droughts
    - Earthquakes
    - Landslides and mudslides
    - Sink holes
    - Tornadoes
    - Tropical cyclones
      - Atlantic hurricanes
      - Hurricanes
      - Pacific hurricanes
      - Pacific typhoons
      - Tropical Cyclone Chapala 2015
      - Tropical Cyclone Nathan 2015
      - Tropical Cyclone Phailin 2013
      - Tropical depressions
      - Tropical storms
      - Typhoons
    - Tsunamis
    - Volcanic eruptions

**Natural disasters**

General news > Accidents and disasters >

has broader (SKOS) +

Accidents and disasters x

has narrower (SKOS) +

Avalanches x Droughts x Earthquakes x Landslides x Sink holes x

Tornadoes x Tropical cyclones x Tsunamis x Volcanic eruptions x Wildfires x

has related (SKOS) +

**Preferred Labels**

preferred label (SKOS) +

Natural disasters en x

**Properties**

alternative label (SKOS) +

definition (SKOS) +

Catastrophic natural events such as hurricanes, tornadoes, volcanic eruptions or avalanches that result in property damage and/or injury or loss of life. en x

description (Tagging) +

must match (Tagging) +

must match regex (Tagging) +

must not match (Tagging) +

must not match regex (Tagging) +

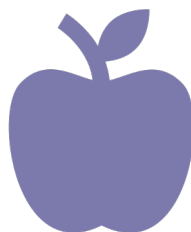
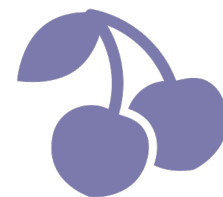
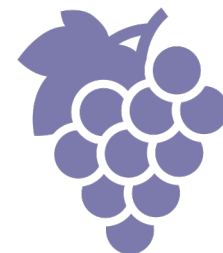
occurrence count (Tagging) +

preferredLabel (Tagging) +

**Save All Properties**

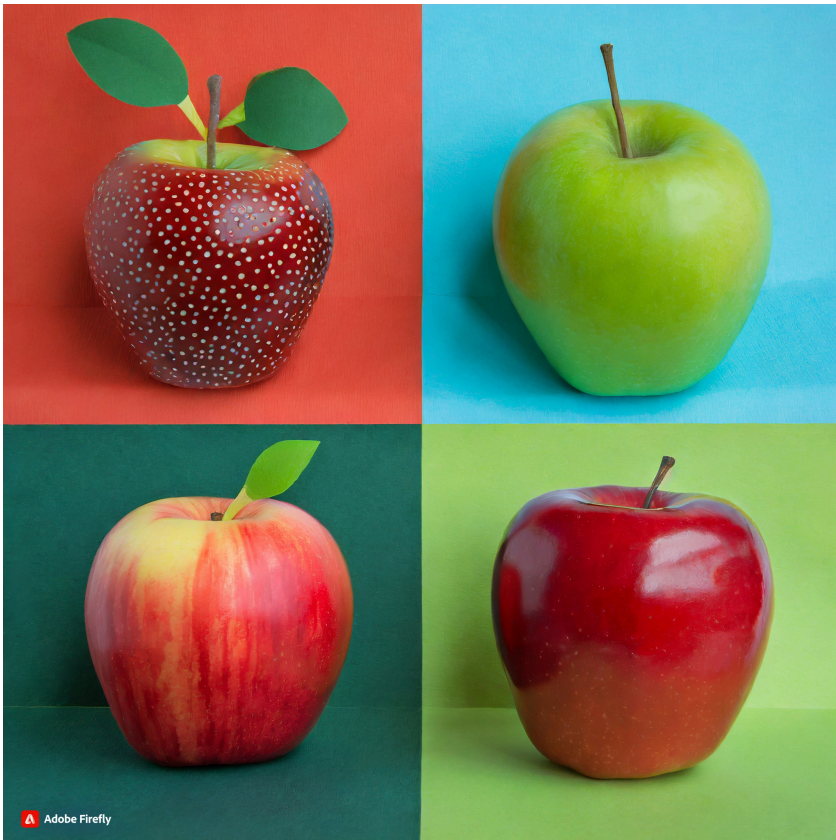
Graphite KOS © 1995-2022 by Synaptica 5.13.2

## Rule-writing: An illustrative example



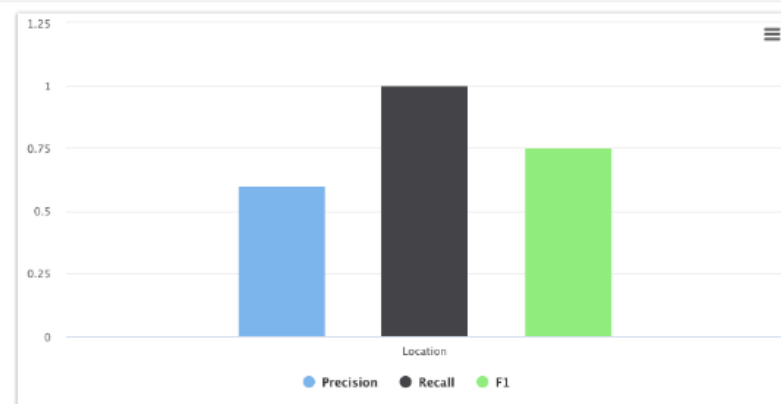


# Rule-writing: An illustrative example

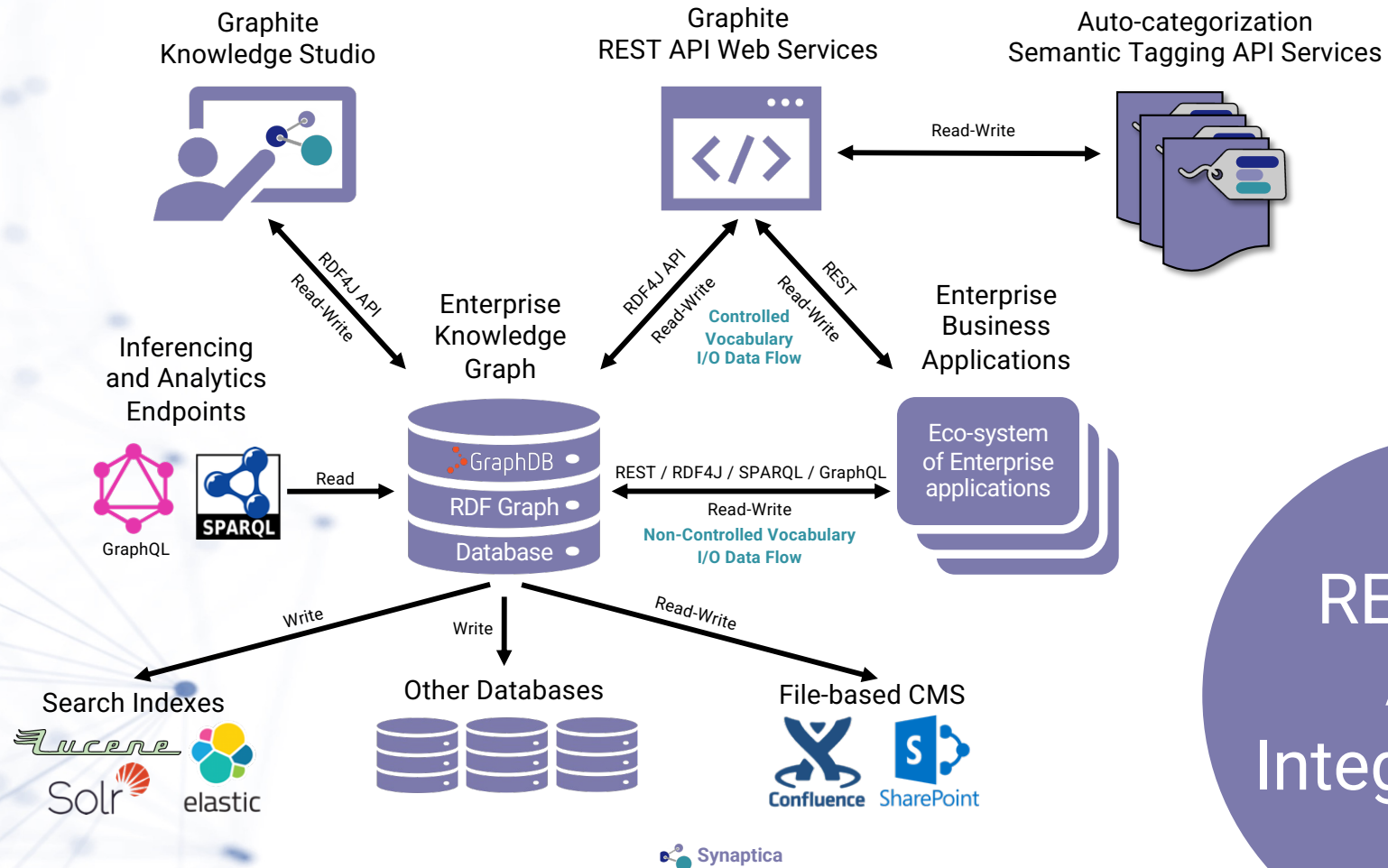


# Measuring quality

## News Corpus Report



Corpora  
analytics





# LLM or no LLM? That is the question.

- Boost LLM inputs with KGs
- Retrieval Augmented with KGs
- Interpretability of LLM Output
- Work with Proprietary Documents & Metadata

LLMs  
enhanced with  
KGs



- Generate KG Query
- Summarize KG Query results
- Explain Query & Schemas
- Add embedding from LLMs: Similarity search

KGs enhanced  
with LLMs

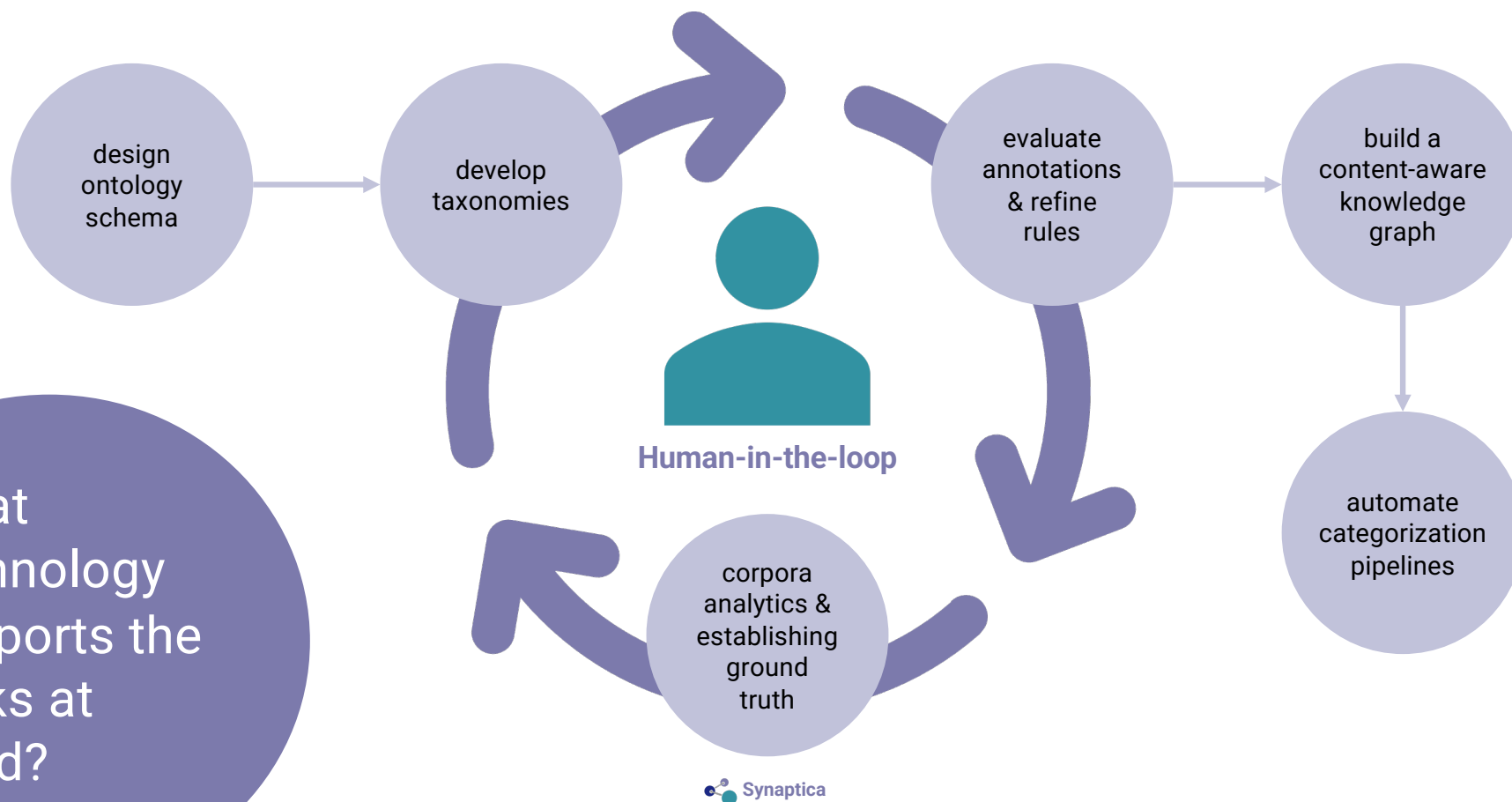


- Suggest alt labels
- Generate definitions
- Suggest missing concepts

Taxonomy  
workflows



# Supporting the human-in-the-loop





# The power of taxonomies + explainability

Rapid  
deployment: no  
extensive model  
training

Reduced cost &  
improved quality  
of labeling

Extended user  
community: no  
coding required

Enterprise  
Vocabulary

Faster  
implementation  
& rapid iterative  
improvement

Integration into  
downstream  
systems

Transparency &  
confidence in  
data

Security of  
enterprise data

# Thank You

## Synaptica



[info@synaptica.com](mailto:info@synaptica.com)



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



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



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



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