



What autocategorization can teach you
about your enterprise content and taxonomies



Sarah Downs





Our topics today

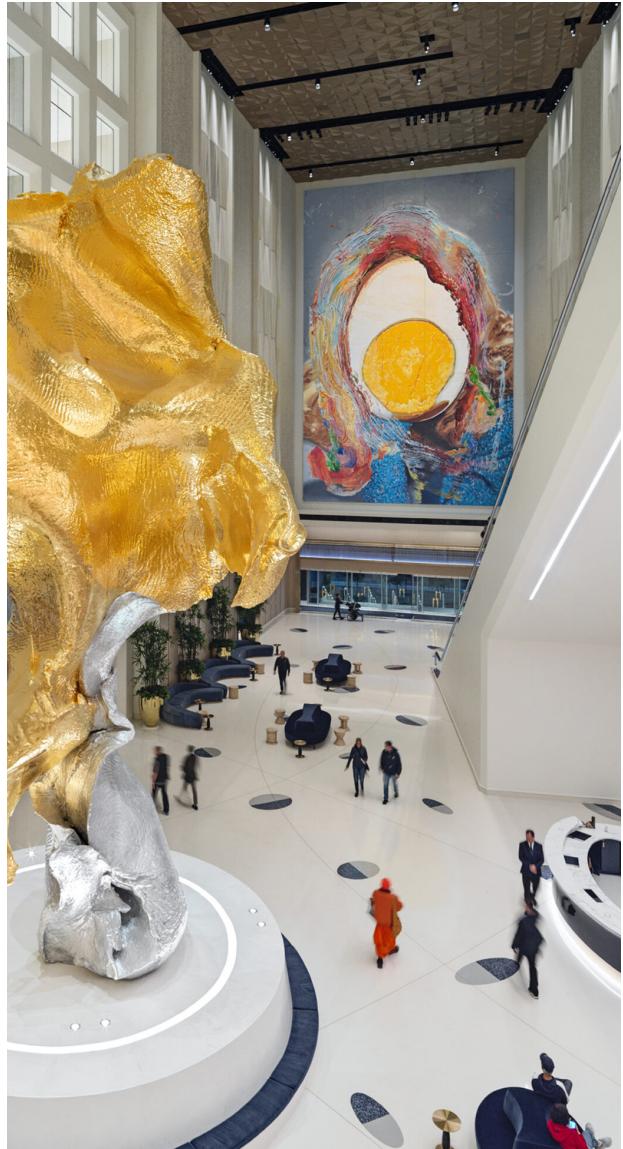
- What is autocategorization?
- SKOS Taxonomies
 - Use in human tagging
 - Extensibility for machine tagging
- A human-in-the-loop workflow for autocategorization
- Learning from the “content aware knowledge graph”
- Recap & Close



Tagging enterprise content



AI Art generated with Adobe Firefly



Can Big Art Make It in Las Vegas? Urs Fischer Weighs In.

The desert entertainment mecca is “artificial, but in a good way,” says the Swiss artist. His 46-foot-tall, gold-leaf sculpture is the city’s latest addition.

<https://www.nytimes.com/2024/06/08/arts/design/urs-fischer-las-vegas-fontainebleau.html>

- Urs Fischer (artist)
- Lovers #3 (work of art)
- Fountainebleau Hotel, Las Vegas
- Art in Las Vegas

Document tags

“The Few”

- Urs Fischer
- Las Vegas, NV
- Lovers #3 (work of art)
- Gold Leaf
- Cast Aluminum
- “The Eye” (work of art)
- Fountainebleau Hotel, Las Vegas
- Manhattan
- Greenwich Village
- Red Hook, Brooklyn
- Long Island City, Queens
- Cannolis
- New York-New York Hotel and Casino
- Mannequin
- Julian Schnabel
- Dasha Zhukova
- Los Angeles
- Dodgers Stadium

Novel concepts

Inline tags

“The Many”

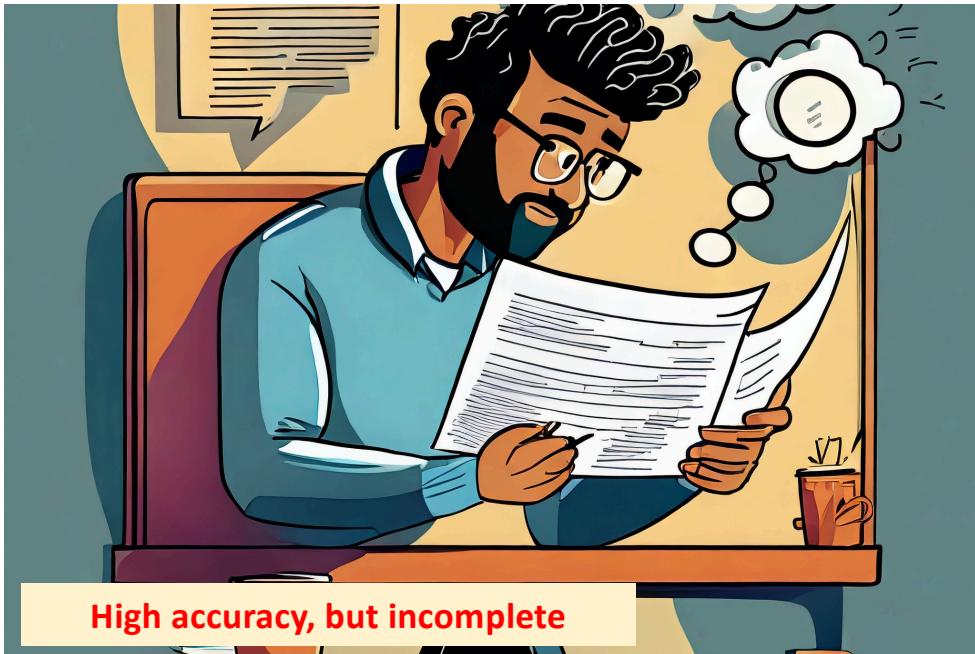


Our topics today

- What is autocategorization?
- SKOS Taxonomies
 - Use in human tagging
 - Extensibility for machine tagging
- A human-in-the-loop workflow for autocategorization
- Learning from the “content aware knowledge graph”
- Recap & Close



Human tagging today



High accuracy, but incomplete



Medium to low accuracy, but incomplete

AI Art generated with Adobe Firefly



Tagging enterprise content



AI Art generated with Adobe Firefly



SKOS Concept record

Graphite Knowledge Studio

Concept Export Import Audit Trail Merge Concept

ecosystem-based management in Scheme GEMET Thesaurus

environmental management > ecosystem-based management

View Populated Panels View All Panels

has broader (SKOS)

environmental management

has narrower (SKOS)

has narrower match (SKOS)

has related (SKOS)

ecosystem-based approach

management of natural resources

has WikipediaArticle (GEMET)

sameEAGlossary (GEMET)

seeAlso (GEMET)

subGroupOf (GEMET)

Resource Types SKOS Concept

Property Templates GEMET SKOS Complete

Collections

Preferred Labels

preferred label (SKOS)

ecosystem-based management

Alternative Labels

alternative label (SKOS)

Properties

acronymLabel (GEMET)

change note (SKOS)

definition (SKOS)

editorial note (SKOS)

example (SKOS)

hidden label (SKOS)

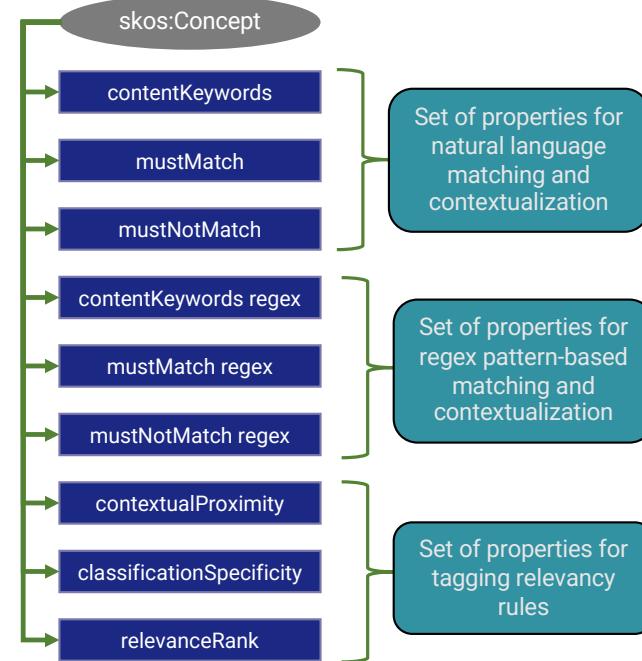
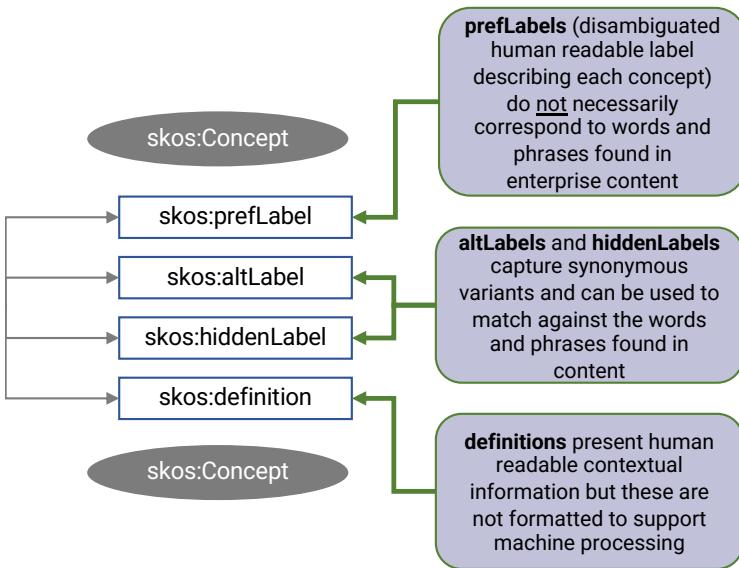
history note (SKOS)

License (DCT)

notation (SKOS)



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



Concept record >> machine tags

The screenshot shows two instances of the Graphite Knowledge Studio interface. The top instance is a search results page for 'computer science, 0.057'. It includes a sidebar with categories like EDUCATION and COMPENSATION AND BENEFITS, and a main panel showing annotations such as 'Bootstrap v3+' and 'Bachelor/Masters in Computer Science / equivalent deg'. A red arrow points from the 'Bachelor/Masters in Computer Science / equivalent deg' annotation to the bottom instance. The bottom instance is a detailed view of the 'computer science' concept record. It displays various properties and their values, including:

- type (RDF)**: hasIndividual (Graphite)
- subClassOf (RDFS)**: (empty)
- superClassOf (Graphite)**: (empty)
- has broader (SKOS)**: (empty)
- has narrower (SKOS)**: (empty)
- has related (SKOS)**: acoustical engineer, application engineer, clinical informatics manager, component engineer, computer science lecturer, ICT teacher secondary school, linguist, mathematician, picture archiving and communication systems administrator, robotics engineering technician, secondary school teacher

The right side of the bottom instance shows a table of preferred labels for 'computer science' in various languages, with entries for Arabic (علوم الحاسوب), Bulgarian (компютърни науки), Czech (informatika), Danish (datalogi), German (Informatik), Greek (πληροφορική), English (computer science), English (en-US), Spanish (informática), Estonian (arvutiteadus), Finnish (tietojenkäsittelytiede), French (informatique), and Irish (eolaiocht ríomhaireachta). A green 'Save All Properties' button is at the bottom right.

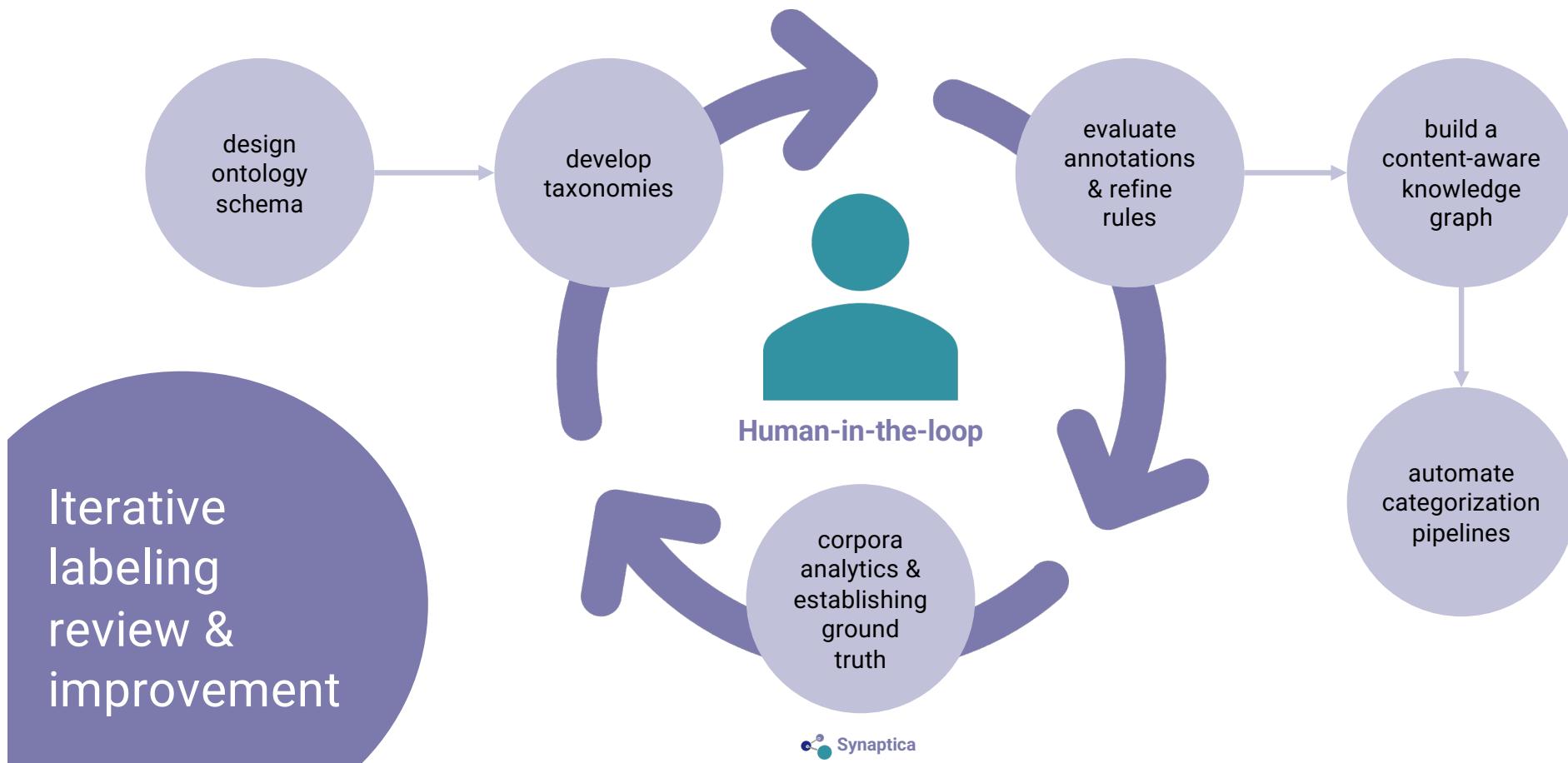


Our topics today

- What is autocategorization?
- SKOS Taxonomies
 - Use in human tagging
 - Extensibility for machine tagging
- A human-in-the-loop workflow for autocategorization
- Learning from the “content aware knowledge graph”
- Recap & Close



Taxonomies to Tags (and back again)





The key: transparency & explainability

Projects > IT Job Search > IT Job Postings > Eaton Internship Program: Software Engineer

Java (computer programming) ✖

Cycle
Annotations

In Document Annotations In text

Object oriented design and programming skills experience

Development experience with Java and/or C# and/or Python and/or Scala

Development experience with Angular and/or React and/or iOS and/or Android

Development experience in Microsoft technologies, development experiences with Microsoft Azure a plus

Demonstrate and document solutions by using flowcharts, diagrams, code comments, code snippets, and performing instru

ments

Knowledge of CI/CD concepts, tools, and technologies

Skills

Tagger: Final Classes Explicit

Type: Skills

ID

https://annotations.omds.com/tag/45 ✖

Entity

Lookup a concept ✖

Java (computer programming) ✖

Created by

Tagger: Final Classes Explicit ✖

Save ✖

Delete ✖

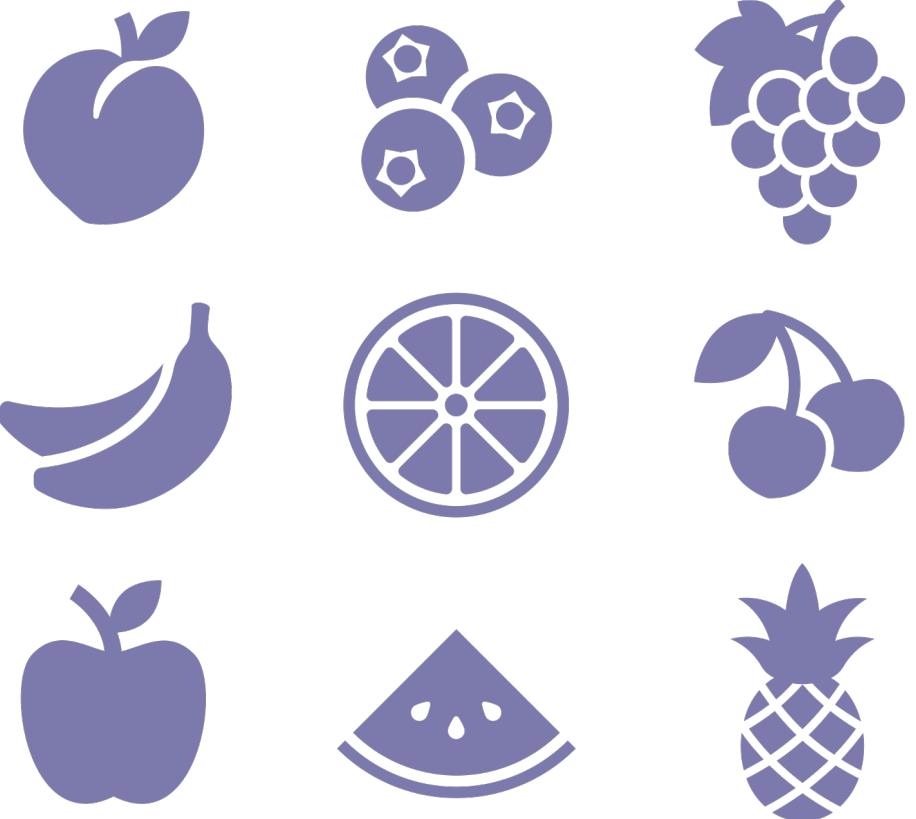
Cancel ✖

Selected

Highlight ✖



Rule-writing: An illustrative example

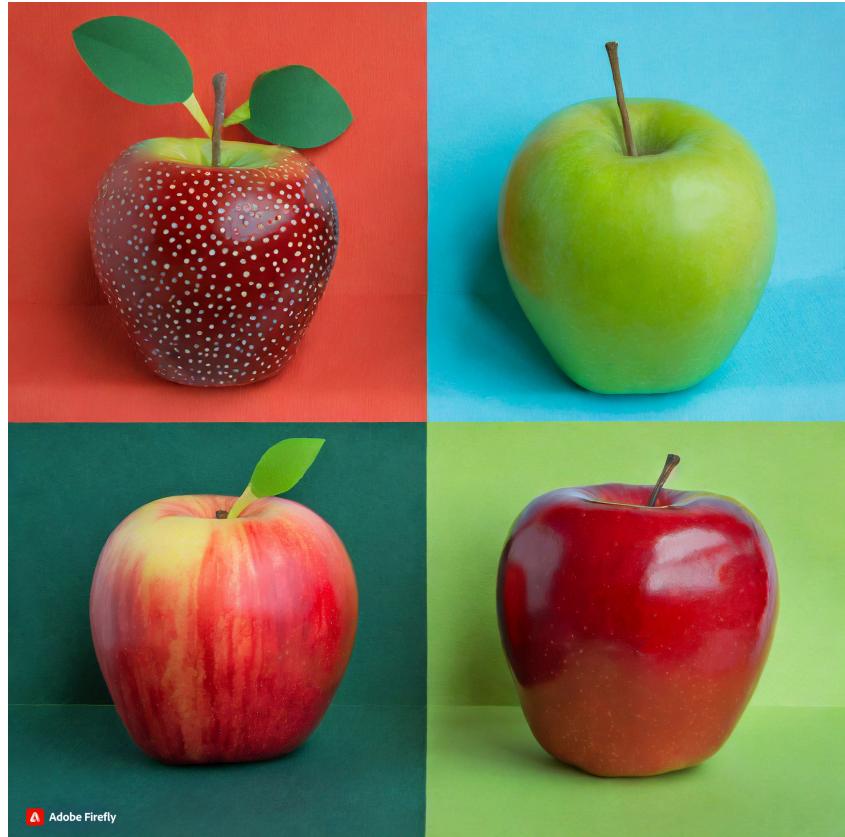


AI Art generated with Adobe Firefly

 Synaptica



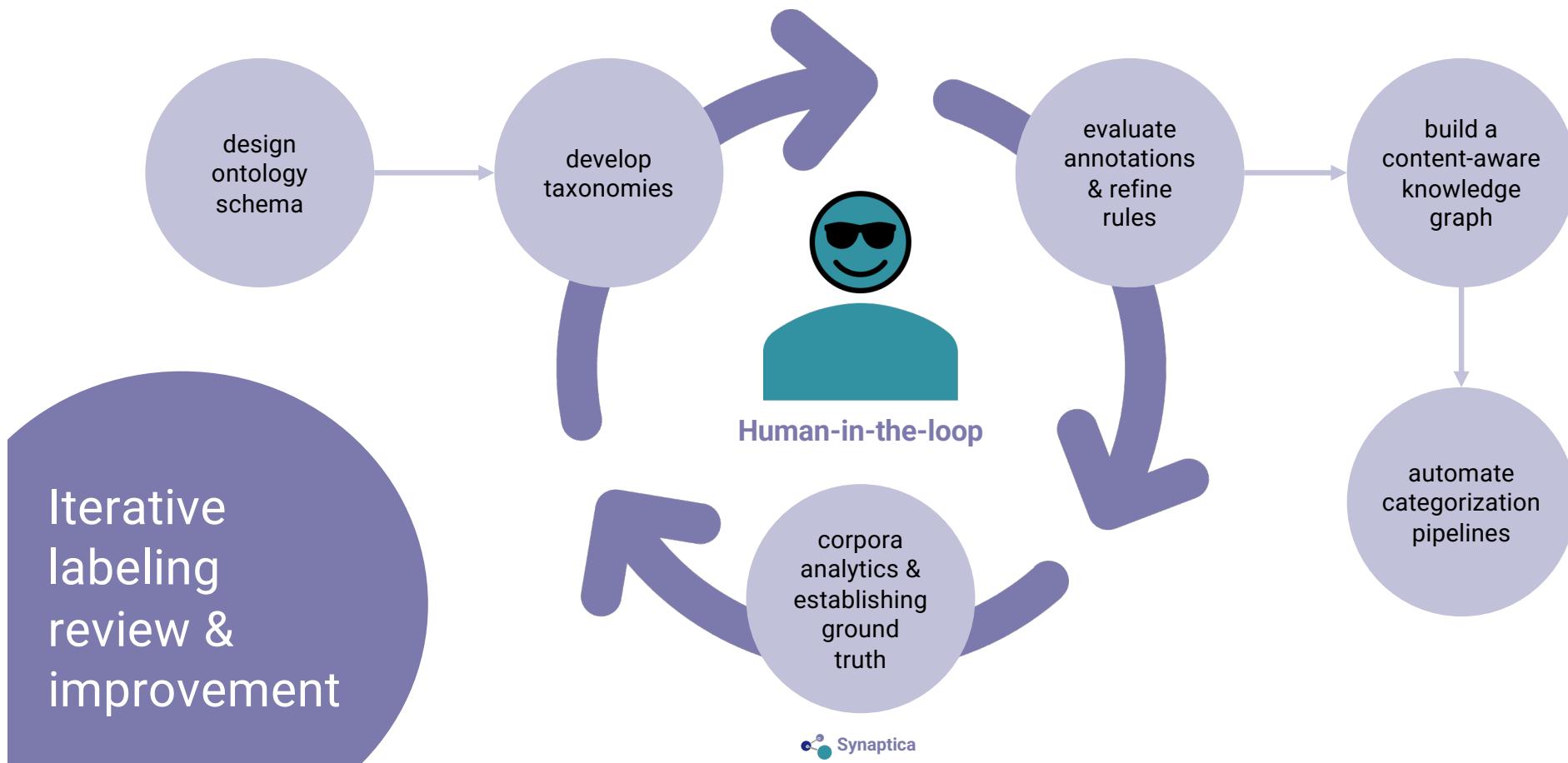
Rule-writing: An illustrative example



AI Art generated with Adobe Firefly

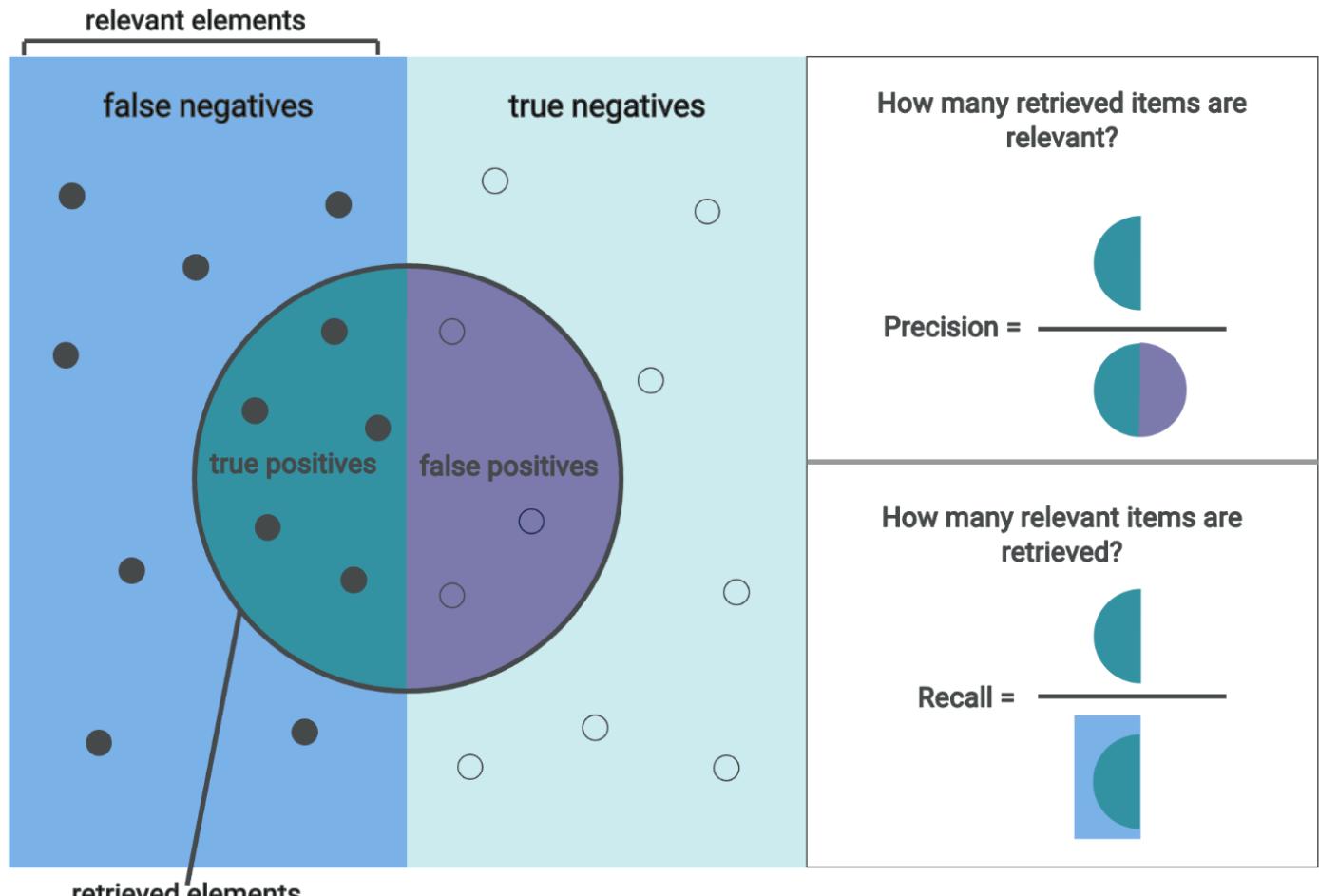


Taxonomies to Tags (and back again)



Measuring quality

Corpora
analytics





Machine tagging performance

General, subject
taxonomies

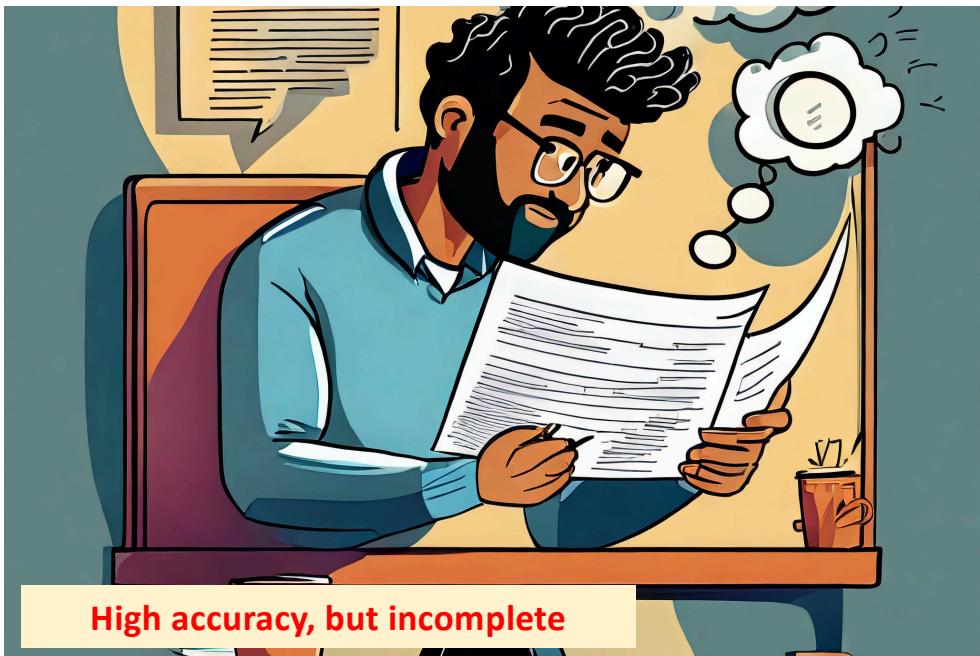
Document-level
annotation precision

Specific, named entity
taxonomies

Inline annotation recall



Comparison to human tagging

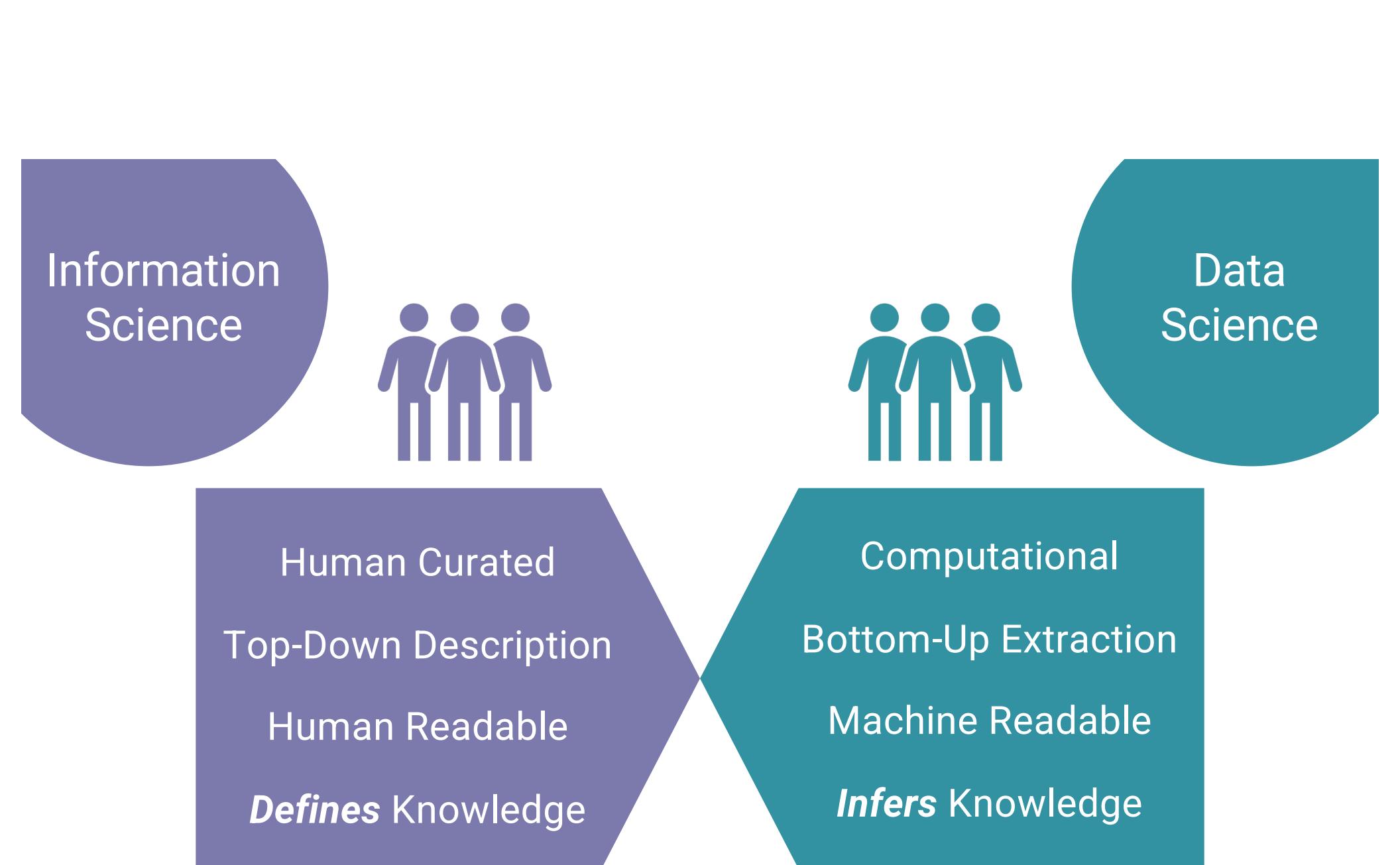


High accuracy, but incomplete

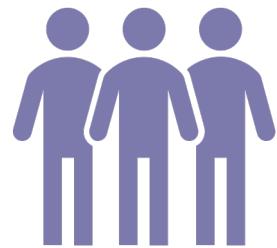


Medium to low accuracy, but incomplete

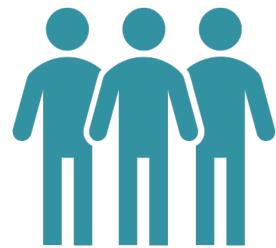
AI Art generated with Adobe Firefly



Information
Science



Data
Science



Human Curated
Top-Down Description
Human Readable
Defines Knowledge

Computational
Bottom-Up Extraction
Machine Readable
Infers Knowledge

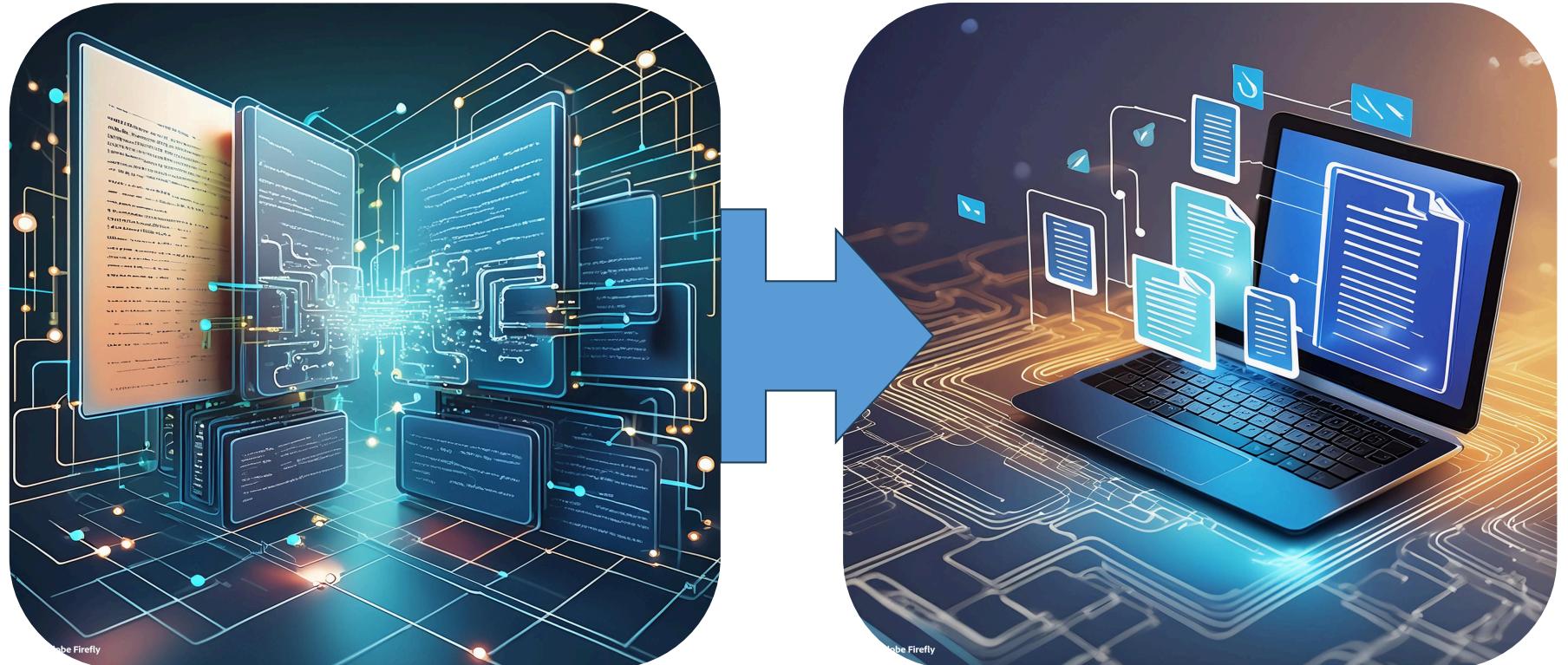


Our topics today

- What is autocategorization?
- SKOS Taxonomies
 - Use in human tagging
 - Extensibility for machine tagging
- A human-in-the-loop workflow for autocategorization
- Learning from the “content aware knowledge graph”
- Recap & Close



Similarity indexing informs recommendations



AI Art generated with Adobe Firefly



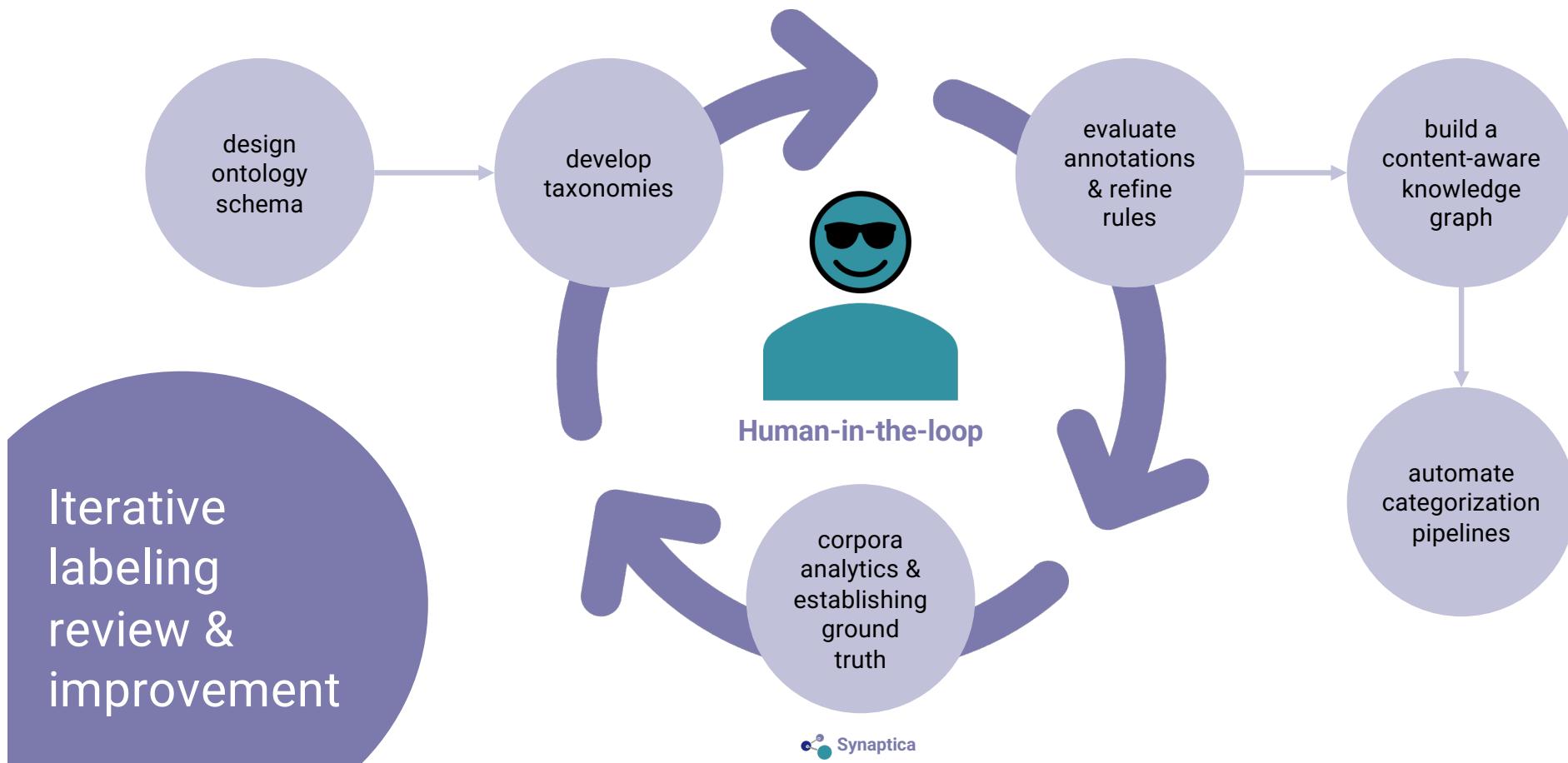


Content trends and analysis





Taxonomies to Tags (and back again)





Our topics today

- What is autocategorization?
- SKOS Taxonomies
 - Use in human tagging
 - Extensibility for machine tagging
- A human-in-the-loop workflow for autocategorization
- Learning from the “content aware knowledge graph”
- Recap & Close



The power of taxonomies + explainability

Rapid deployment: no extensive model training

Reduced cost & improved quality of labeling

Extended user community: no coding required

Leverage proprietary enterprise vocabulary

Faster implementation & rapid iterative improvement

Integration into downstream systems

Transparency & confidence in data

Security of enterprise data

Questions?

Synaptica



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

Synaptica



info@synaptica.com



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



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



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



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