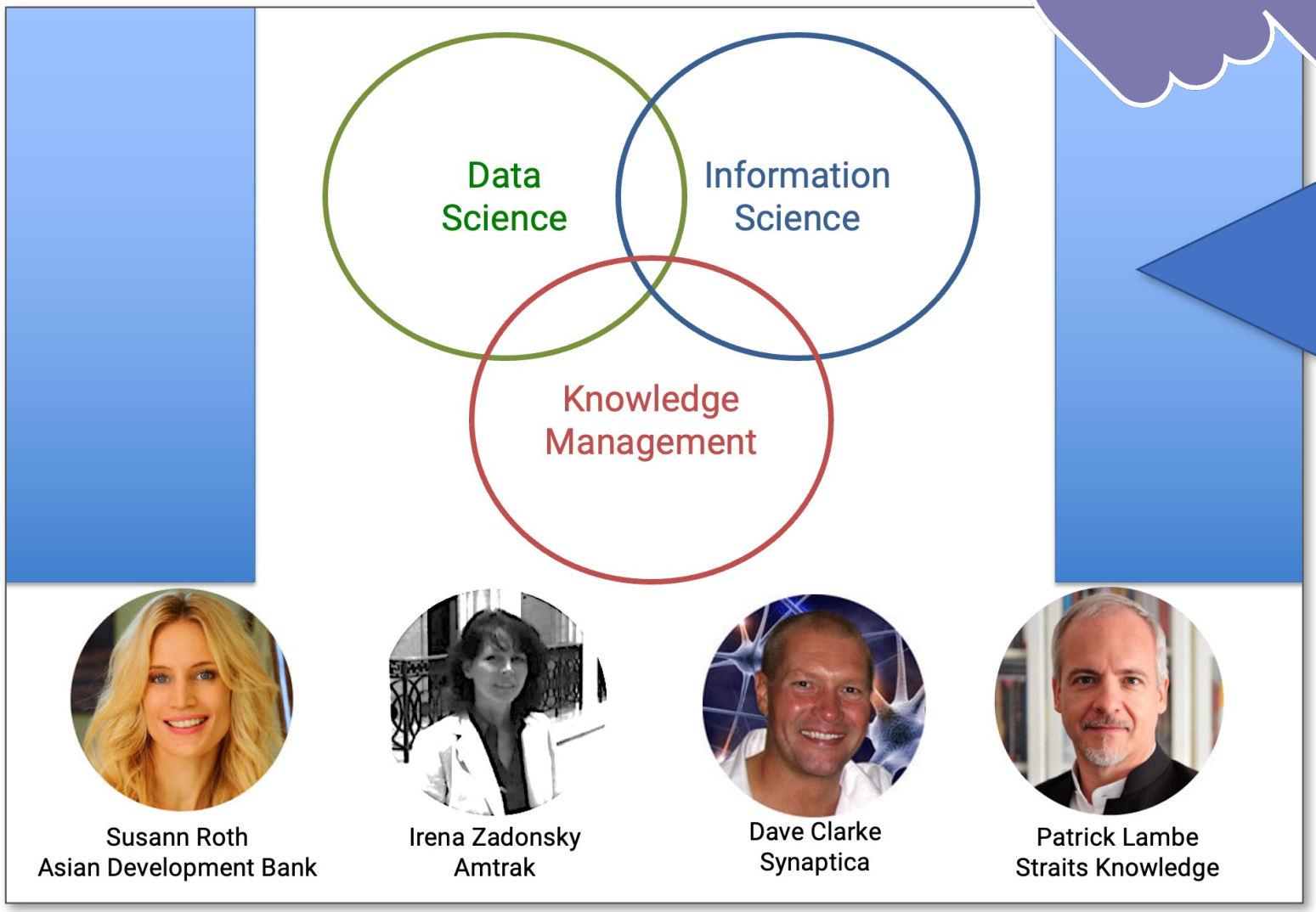
Synaptica

Cross-Team Collaboration for Knowledge Discovery

Dave Clarke
Founder, Synaptica

KM World Keynote 2022



Wednesday 9th
8:30 AM
Opening Keynote
Grand Ballroom

Deeper look into the human aspects of cross-team collaboration moderated by Patrick Lambe

Synaptica

Cross-Team Collaboration for Knowledge Discovery

Dave Clarke
Founder, Synaptica

1. Differences between data science and information science
2. How Synaptica is developing software tools to help bridge both disciplines
3. Look at what successful collaboration looks like

Take away 1

Information Science
cf.
Data Science

What's the Diff?

Information Science

Library and Information Science 1

Information science

From Wikipedia, the free encyclopedia

Information science (also known as **information studies**) is an academic field which is primarily concerned with analysis, collection, classification, manipulation, storage, retrieval, movement, dissemination, and protection of information.[1] Practitioners within and outside the field study the application and the usage of knowledge in organizations in addition to the interaction between people, organizations, and any existing information systems with the aim of creating, replacing, improving, or understanding information systems.

Historically, information science is associated with computer science, data science, psychology, technology, library science, healthcare, and intelligence agencies.[2] However, information science also incorporates aspects of diverse fields such as archival science, cognitive science, commerce, law, linguistics, museology, management, mathematics, philosophy, public policy, and social sciences.

Classification and Retrieval 2

Taxonomy and Knowledge Organization Systems 3

4 Organizational and human knowledge … overlap with KM

Data Science

1 Not Information Science

Data science

From Wikipedia, the free encyclopedia

Not to be confused with information science.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data,[1][2] and apply knowledge from data across a broad range of application domains. Data science is related to data mining, machine learning and big data.[3]

Data science is a "concept to unify statistics, data analysis, informatics, and their related methods" in order to "understand and analyse actual phenomena" with data.[4] It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, information science, and domain knowledge.[3] However, data science is different from computer science and information science. Turing Award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational, and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge.[5][6]

2 Extraction of information

3 From noisy data

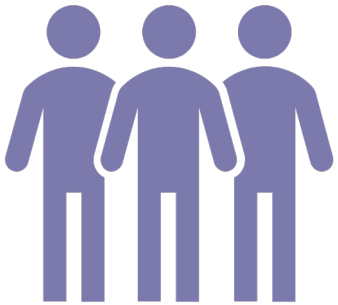
4 Maths and statistics based

5 Use of big knowledge graphs like Wikidata

6 IT capabilities driving a changing landscape

https://en.wikipedia.org/wiki/Data_science

Information Science



Human Curated

Top-Down Description

Human Readable

Defines Knowledge

Data Science

Computational

Bottom-Up Extraction

Machine Readable

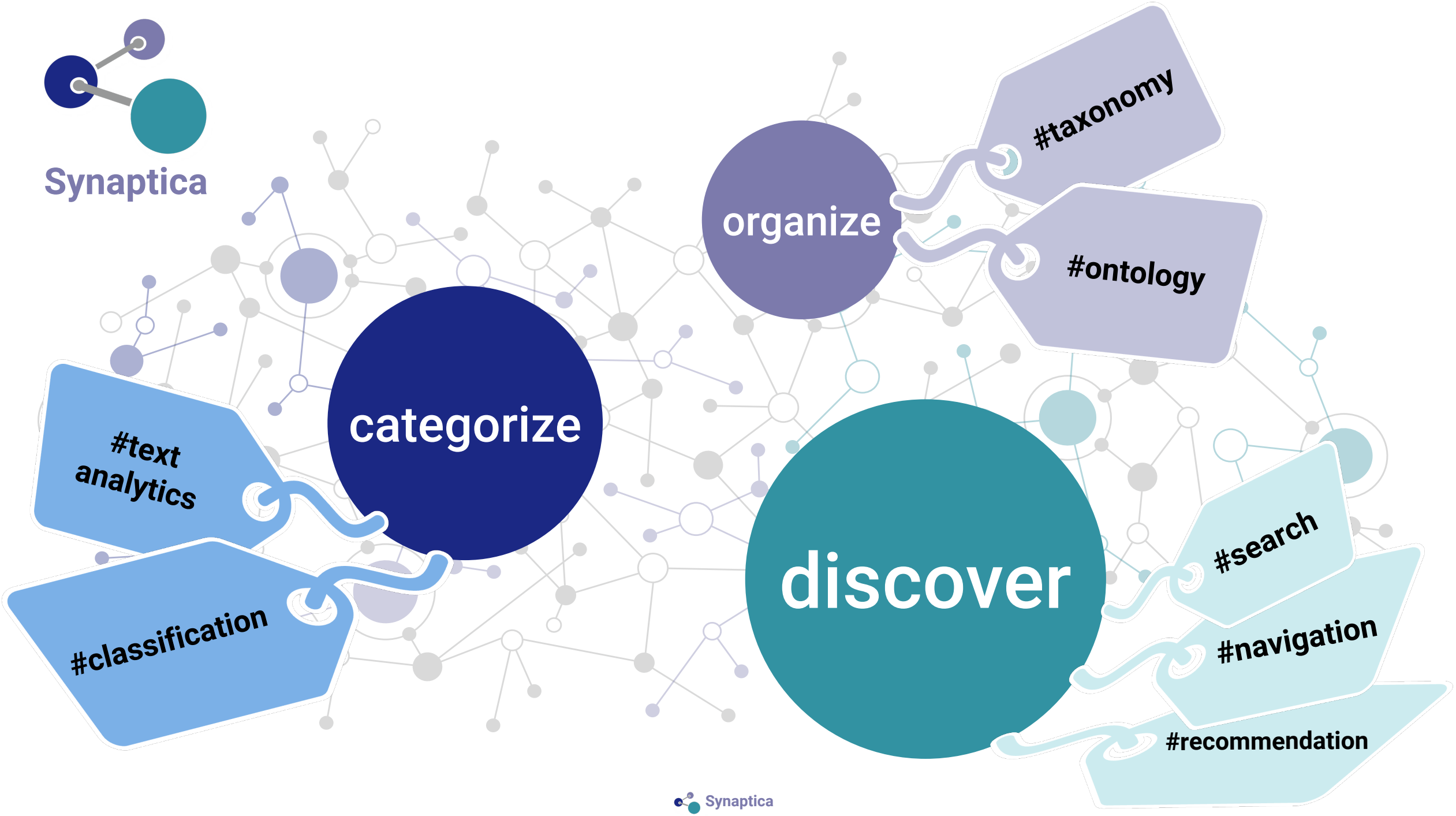
Infers Knowledge


Taxonomists & Ontologists
Data Scientists
Information Architects
Content & Metadata Managers
cross-team collaboration

Take Away 2

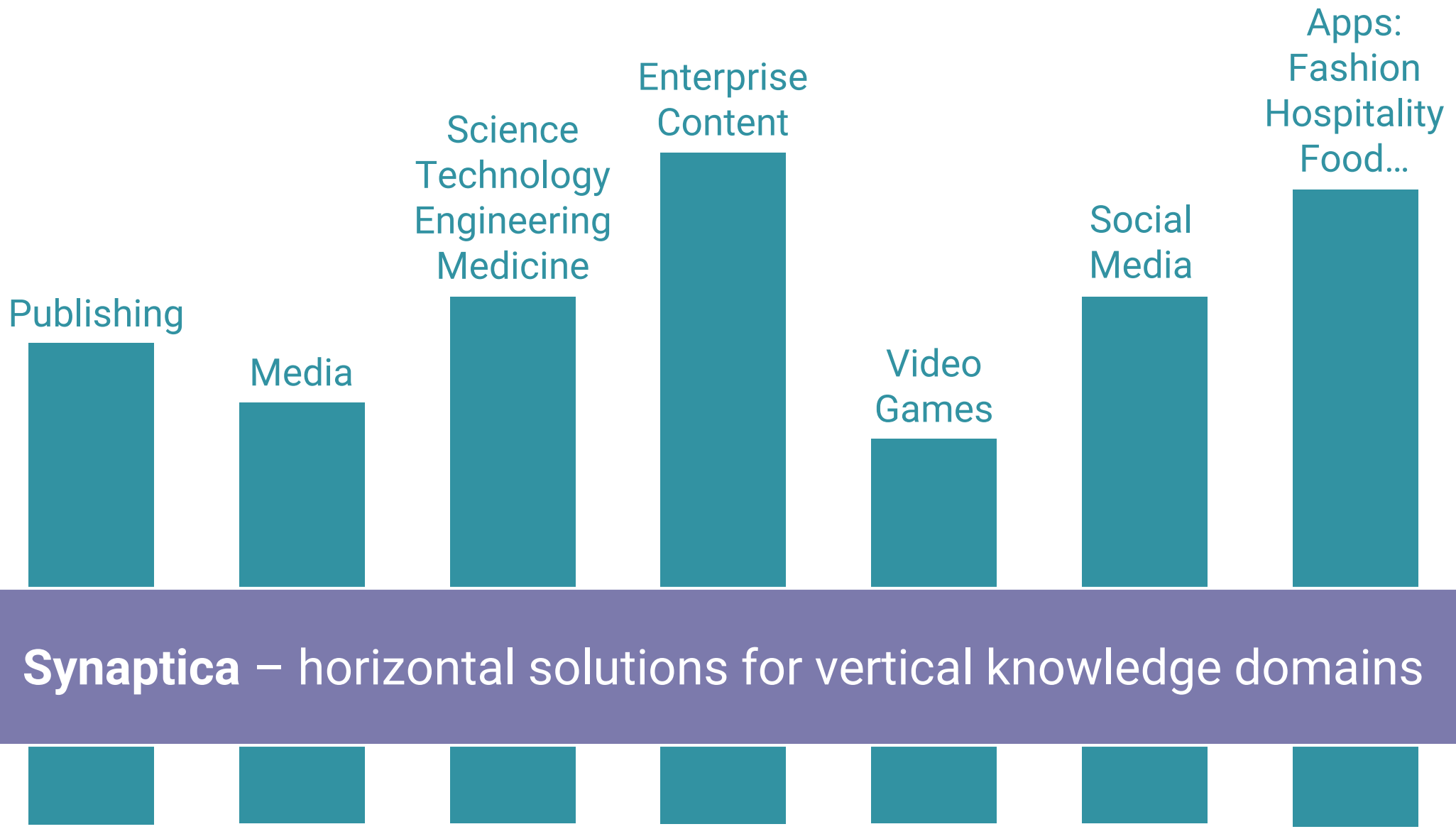
How does Synaptica help these stakeholders

We develop tools to help knowledge workers


Synaptica
organize
#taxonomy
#ontology
categorize
#text analytics
#classification
discover
#search
#navigation
#recommendation
Synaptica

Synaptica

1995 Just Info Science → Info Science + Data Science 2022



Synaptica – horizontal solutions for vertical knowledge domains

Data Science

Information Science

ontotext

synaptica

20+ years	25+ years
RDF graph database	Controlled vocabulary
NLP text analytics	Taxonomy & ontology

synaptica

Introducing

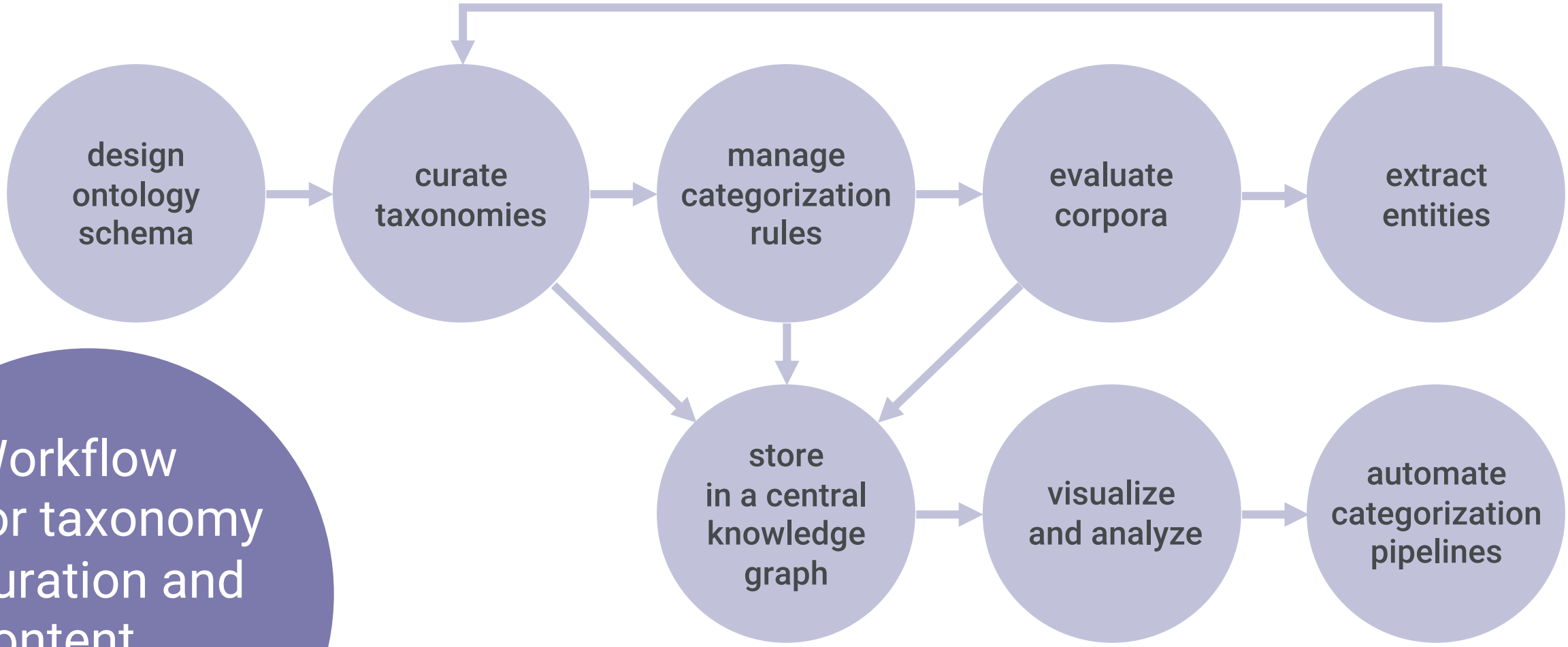
Graphite Knowledge Studio

Powered by ontotext Text Analytics

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics

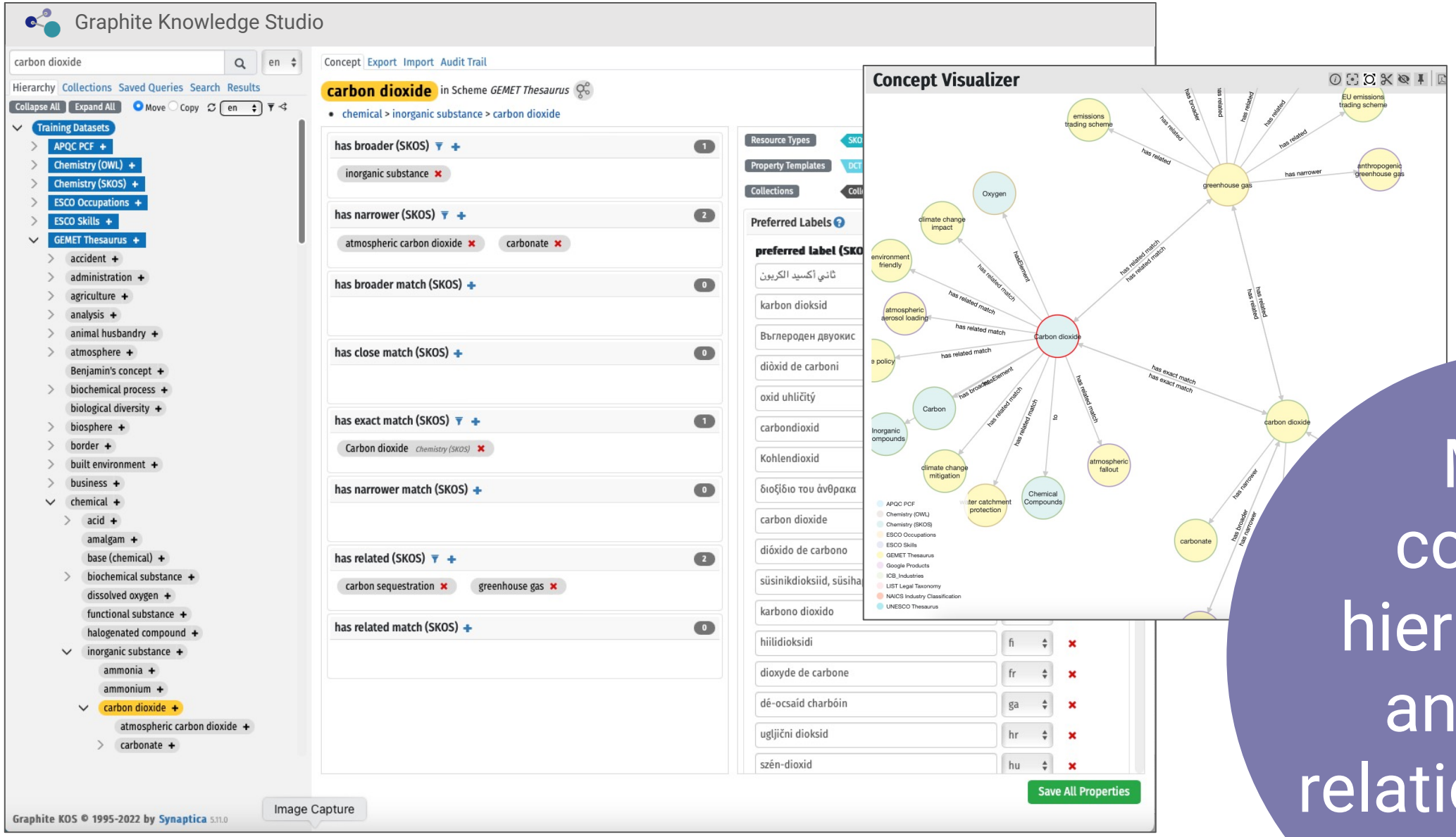


Workflow for taxonomy curation and content classification

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics



Manage concepts, hierarchies, and graph relationships

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics

Graphite Knowledge Studio

Projects > News > News July 2022

News July 2022 ACTIVE

Annotate Corpus | Reports | Create document | Archive corpus | Edit corpus

8 columns selected

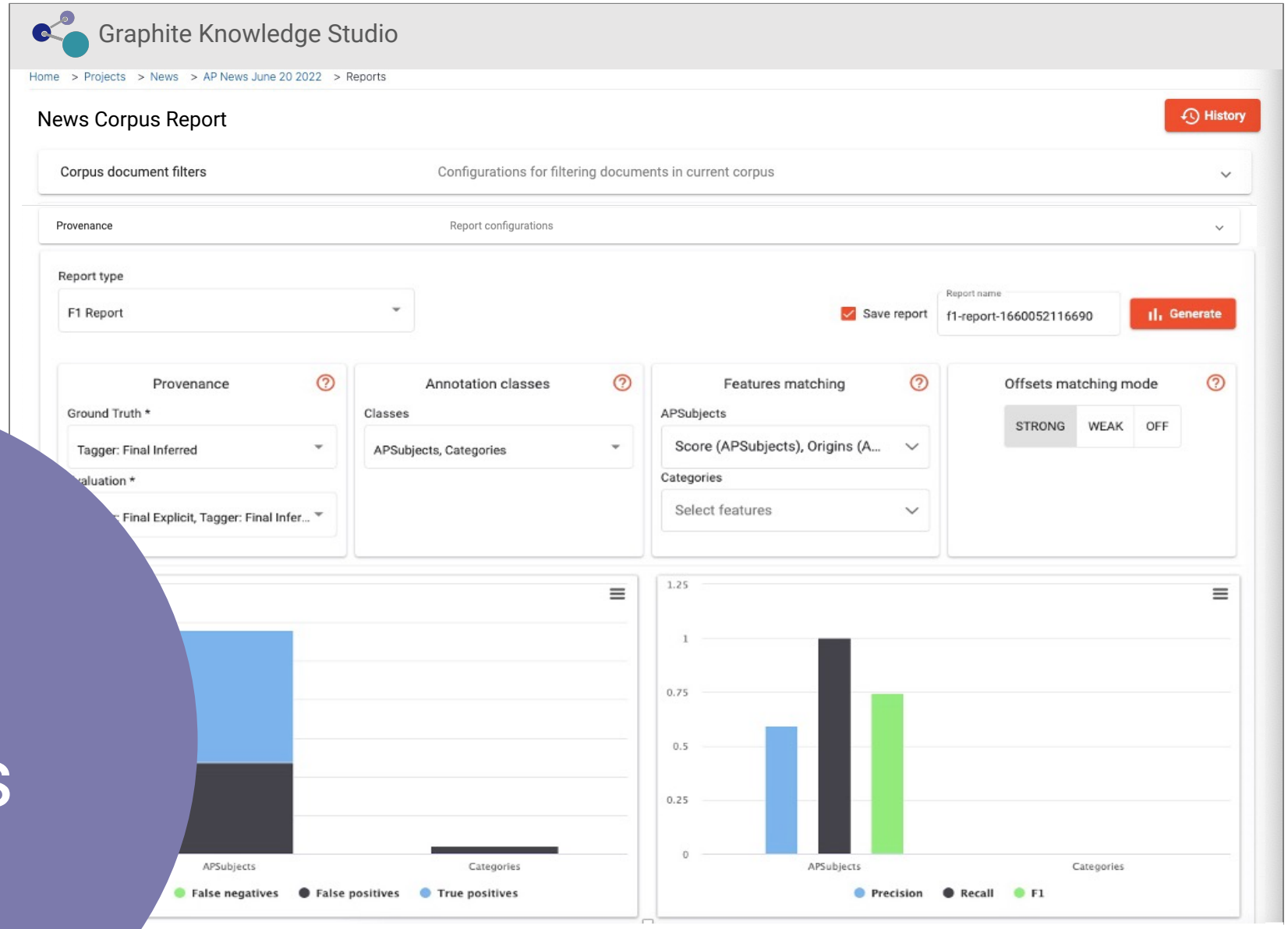
Label	Created at	Created by	Modified at	Modified by	Annotations count	Annotations modified at	Document type
Le Pen: Huge gains in French parliament a 'seismic event'	Sep 21, 2022, 3:26:55 PM	admin	Sep 21, 2022, 3:26:55 PM	admin	791	Oct 28, 2022, 11:27:56 AM	SimpleDocument
South Asia floods hampering access to food, clean water	Sep 21, 2022, 3:26:28 PM	admin	Sep 21, 2022, 3:26:28 PM	admin	786	Oct 28, 2022, 11:27:56 AM	SimpleDocument
Biden's optimism collides with mounting political challenges	Sep 21, 2022, 3:26:02 PM	admin	Sep 21, 2022, 3:26:02 PM	admin	1519	Oct 28, 2022, 11:27:29 AM	Simple
Bitcoin inches up above psychological threshold of $20,000	Sep 21, 2022, 3:25:36 PM	admin	Sep 21, 2022, 3:25:36 PM	admin	250	Oct 28, 2022, 11:27:29 AM	
New body armor rules in NY miss vest worn by Buffalo killer	Sep 21, 2022, 3:25:07 PM	admin	Sep 21, 2022, 3:25:07 PM	admin	1385	Oct 28, 2022, 11:27:29 A	
Colombia picks 1st leftist president in tight runoff contest	Sep 21, 2022, 3:24:39 PM	admin	Sep 21, 2022, 3:24:39 PM	admin	930	Oct 28, 2022, 11:27:29	
How much for gas? Around the world, pain is felt at the pump	Sep 21, 2022, 3:24:10 PM	admin	Sep 21, 2022, 3:24:10 PM	admin	1193	Oct 28, 2022, 11:27:29	
'It's just hell there': Russia still pounds eastern Ukraine	Sep 21, 2022, 3:23:40 PM	admin	Sep 21, 2022, 3:23:40 PM	admin	824	Oct 28, 2022, 11:27:29 AM	
The moments resonating							

Manage corpora

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics



Corpora analytics

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics

Graphite Knowledge Studio

Projects > News > News Headlines 2022 > http://example.com/9b81e918-3eb7-5e2b-9e16-8d8fc3d7d10b

BRAC, a non-governmental nonprofit organization in Bangladesh, opened a center Monday to prepare food items as part of a plan to feed 5,000 families in one of the affected districts, but the arrangement was not enough, said senior director Arinjoy Dhar.

In a video posted online, Dhar asked for help ensuring food for the flood-affected, saying they alone were trying to reach out to about 52,000 families with emergency supplies.

Last month, a pre-monsoon flash flood triggered by a rush of water from upstream in India's northeastern states hit Bangladesh's northern and northeastern regions, destroying crops and damaging homes and roads.

Bangladesh, a nation of 160 million people, is low-lying and faces threats from natural disasters such as floods and cyclones, made worse by climate change. According to the United Nations' Intergovernmental Panel on Climate Change, about 17% of people in Bangladesh would need to be relocated over the next decade or so if global warming persists at the present rate.

Selected document annotation: APSubjects Delete Save Highlight

origins List of APSubjectsConcept

Open search

id	label	description
https://[redacted]/concept/ac1043543	Floods	Unforeseen and damaging flooding due to events such as natural disasters, heavy rains, dam or levee breaches or water-main breaks, usually resulting in evacuation,

Annotation filters | Concepts menti

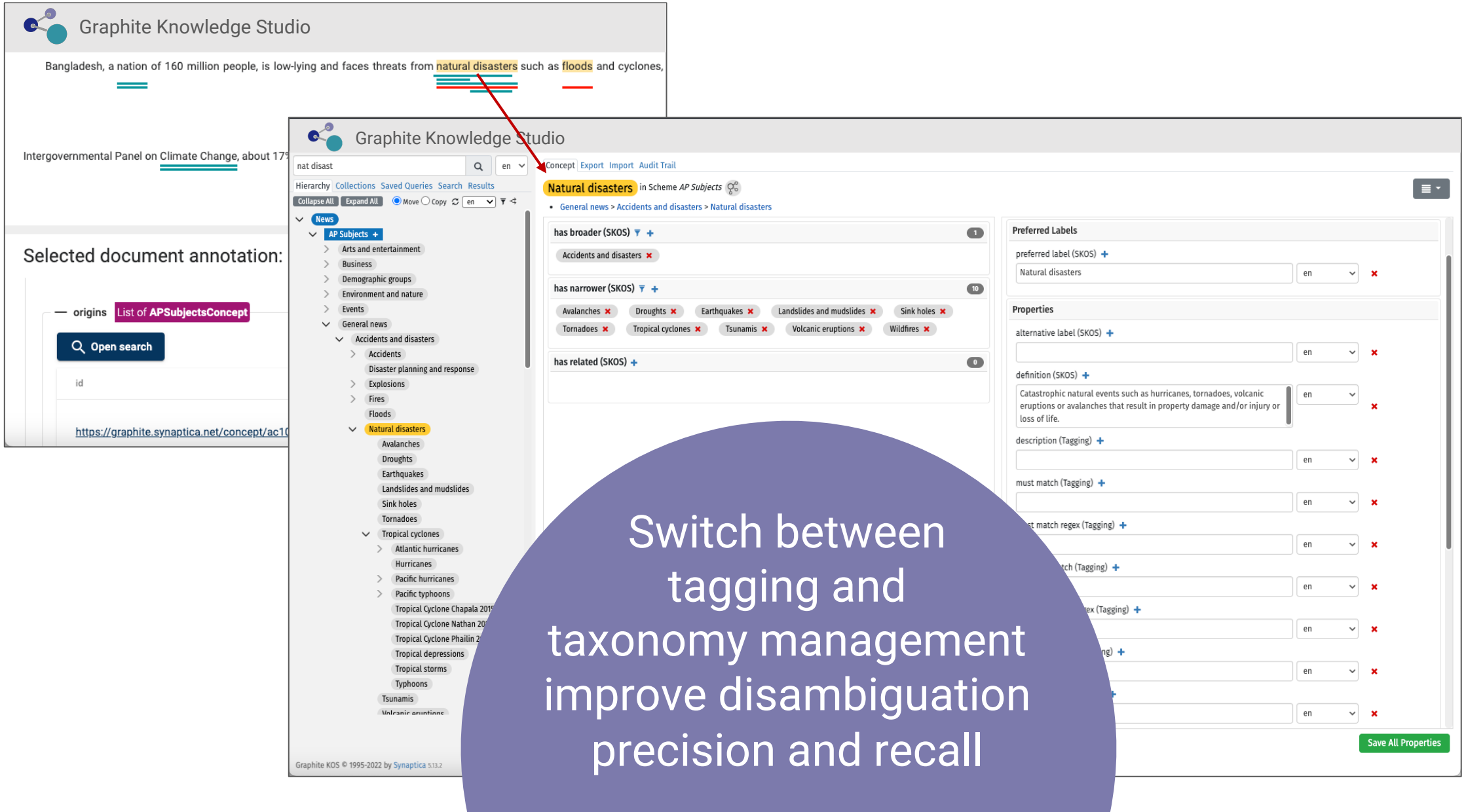
Label	Occurrence
water	21
Floods	17
Nationalization	9
Climate change	
disaster	
Monsoons	
Nationalism	
Food	
climate ch	
Camping	
Capitalism	
Natural disaster	

Review inline tagging and highlight graph-inferred classification

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics



Switch between tagging and taxonomy management improve disambiguation precision and recall

Take Away 3

What does successful collaboration look like

inferred knowledge
speed of automation
customer relevance

Empowering Tagging Food Items and Merchants with the Clustering Taxonomy Candidates and Release Pipeline

by Zhiyang Zhang, October 2022

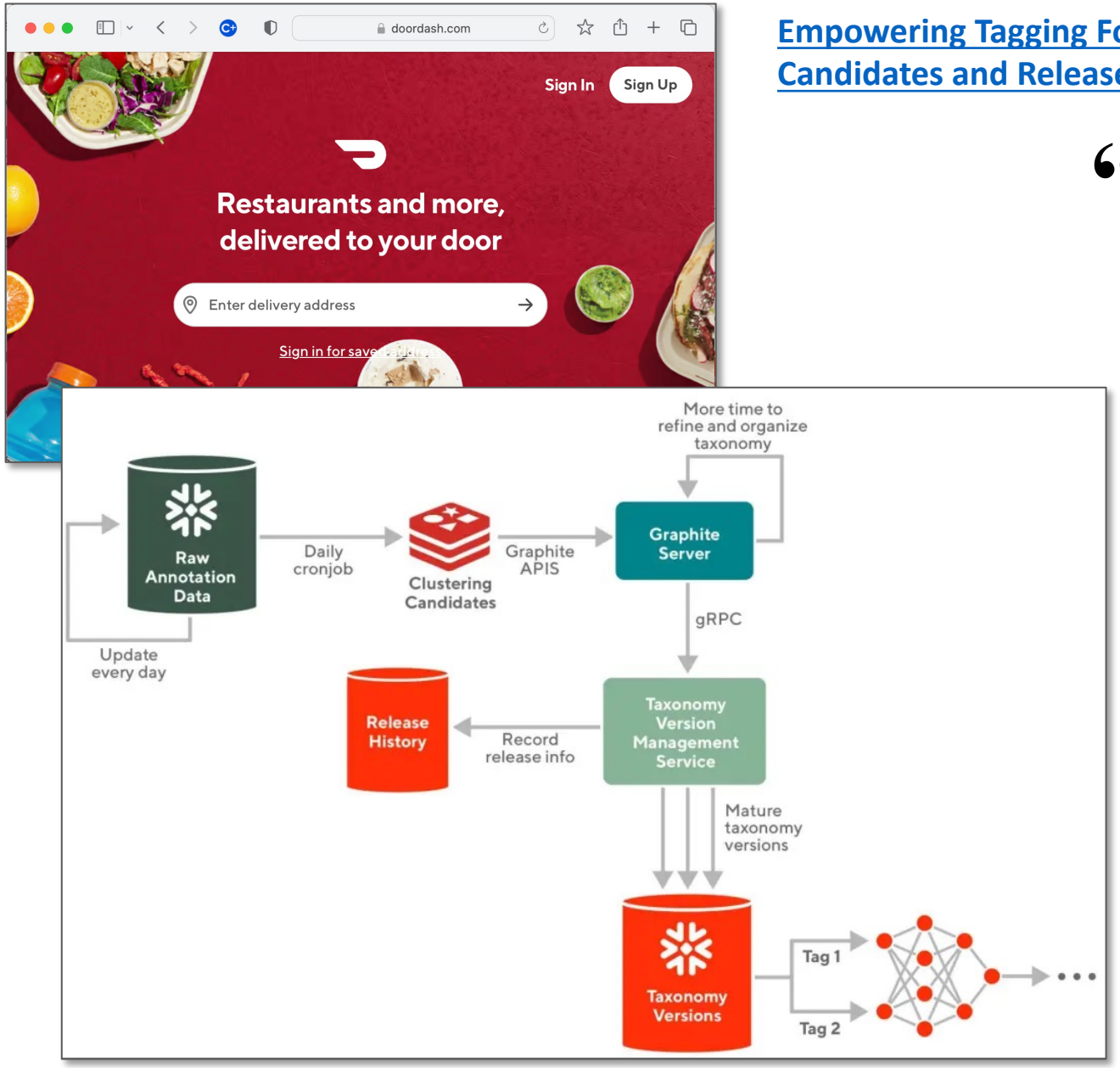


Figure 2: The automated pipeline of clustering candidates and releasing taxonomy versions

“Our pipeline can automatically handle tens of thousands of raw data and import the top 50 frequent candidates into Graphite. It saves the taxonomy team hours every day filtering taxonomy candidates and exporting mature taxonomy versions. This means that they can be more efficient and focus more on expanding the taxonomy efficiently. Also, it provides the ML team with the single source of truth tables for taxonomy versions and they can easily consume and train models.

Overall, we could achieve eliminating hours of manpower per week with this effort and replace repetitive, error prone manual processes with automated processes and proper channels. Thus, our taxonomy team and ML team can focus on their own dedicated area to improve our user experience and meet consumers' needs of ease and convenience.”

Thank you

Visit with us at
KM World
Expo Booth 101

Connect with 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

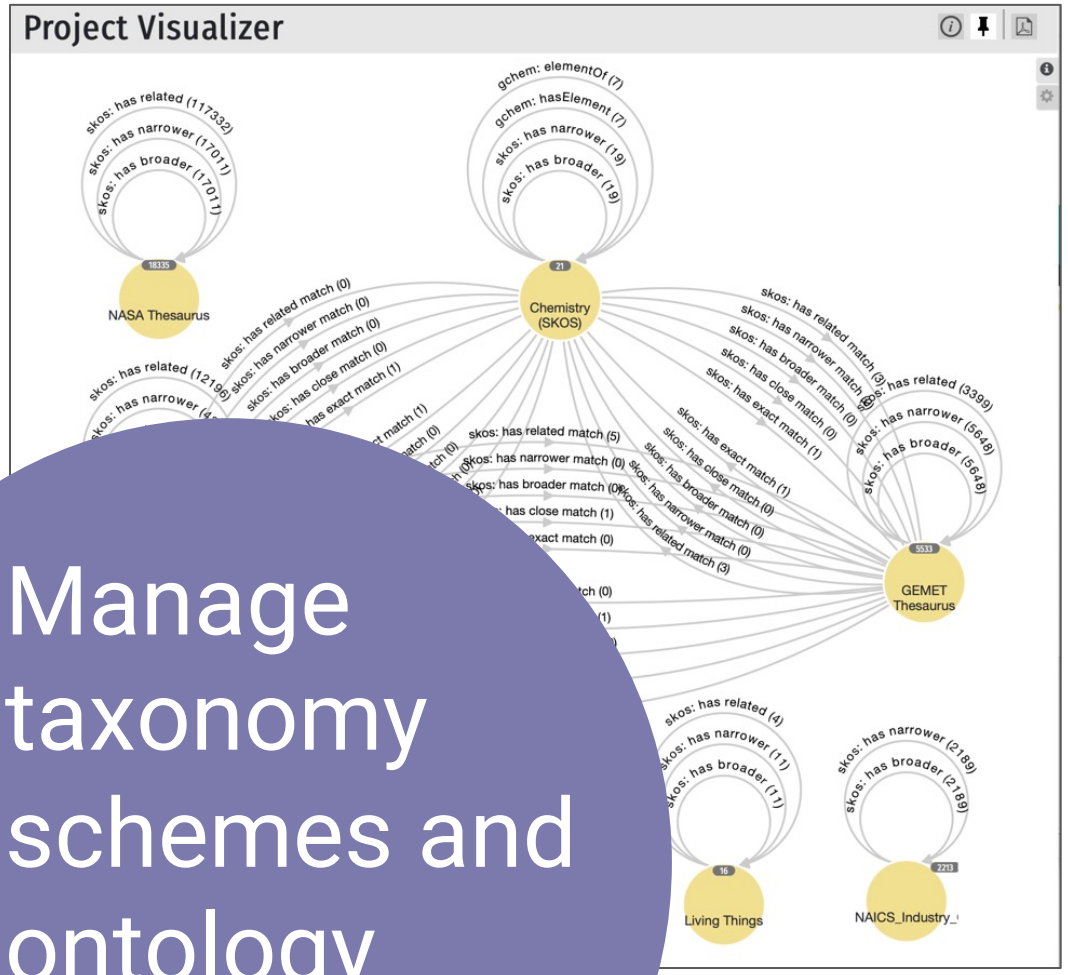
scan to access downloads

Reserve slides

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics



Manage taxonomy schemes and ontology schema

Ontologies comprise semantic *Schema* of class, property and relationship types, plus *Taxonomies* of specific named concepts, classes and individuals.

Ontology

Schema

Class Types
Property Types
Relationship Types

+

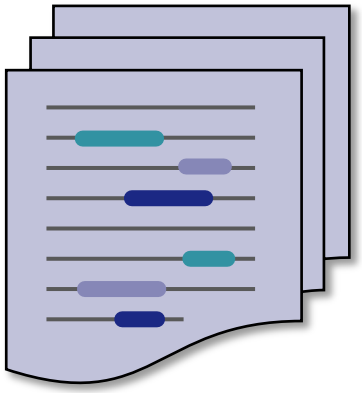
Taxonomy

Specific Concepts, Classes & Named Individuals

synaptica

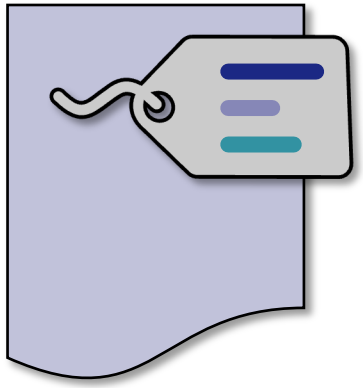
Graphite Knowledge Studio

Powered by ontotext Text Analytics



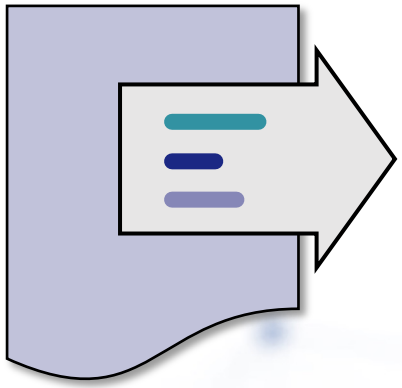
Tagging

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



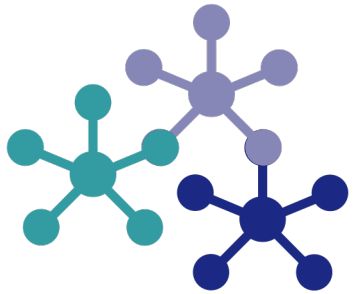
Categorization

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.



Big Graphs

such as DBpedia, GeoNames, and Wikidata, can be used for disambiguation and to validate extraction.

Tagging – Classification – Extraction

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics

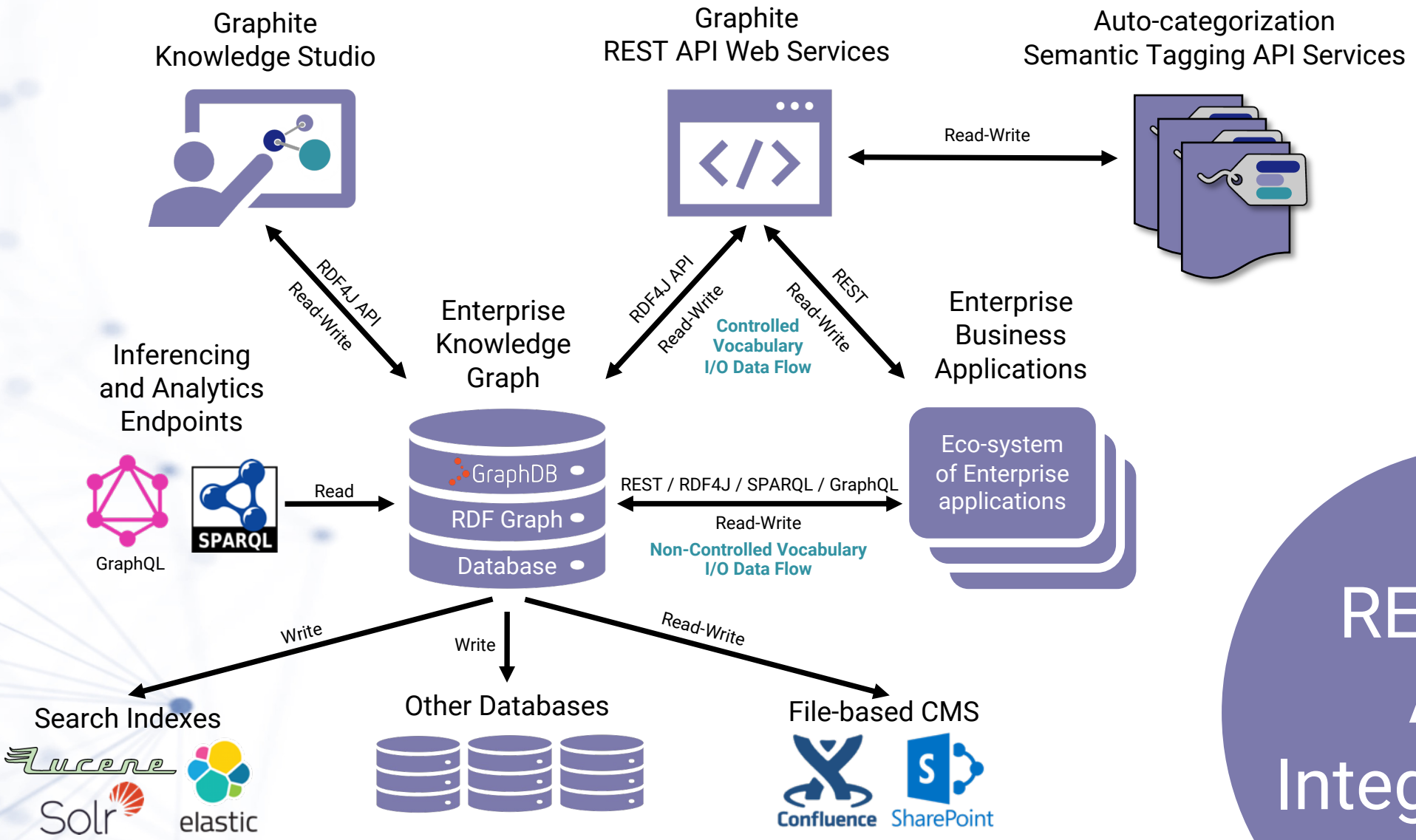
Human in the loop

regardless of whether rules-based or ML methods are employed, the process of developing and refining tagging and categorization pipelines is iterative and incorporates human feedback

synaptica

Graphite Knowledge Studio

Powered by ontotext Text Analytics



RESTful
API
Integrations

Graphite Knowledge Studio

Highlights, USPs, and benefits

Highlight Features & USPs	Business Benefits
Tightly integrated OOTB toolsets	Rapid deployment; lower cost of entry; horizontal solution for all industry sectors
Ease-of-use tools and user experiences	Extends user community; compresses learning curve; innovative time-saving workflows
Transparent and editable tagging and classification rules	Faster implementation and agile iterations; reduce or eliminate time and cost to develop training sets; avoid unmodifiable 'black box' components
Standards-based controlled vocabulary management	Immediate import and use of existing enterprise taxonomies, keyword/phrase lists, and/or adopt public domain datasets; data portability
Scalability; high performance; multi-tier architecture	Start small and economically; scale rapidly as and when needed; guarantee server resources for target performance
REST, SPARQL and GraphQL APIs	Ease and speed of integration with downstream systems including search, content management, and metadata management