Graphite Knowledge Studio:
Empowering the human-in-the-loop

Delivering the Future – Together!

Knowledge Graph Forum – Synaptica LLC

Partnership for customer solutions: Autocategorization of enterprise content

1   **Graphite Knowledge Studio:**
Bringing together information science and data science

2   **Bootstrapping enterprise taxonomies for autocategorization**

3   **Empowering and scaling the human-in-the-loop for improved text analytics performance**

4   **Building a Content-Aware Knowledge Graph to power insight and further functionality**

ontotext

Human tagging today: costly and low quality

AI Art generated with Adobe Firefly

Tagging
identifies the **many** taxonomy concepts and named entities that are **mentioned** within the full text of a document.
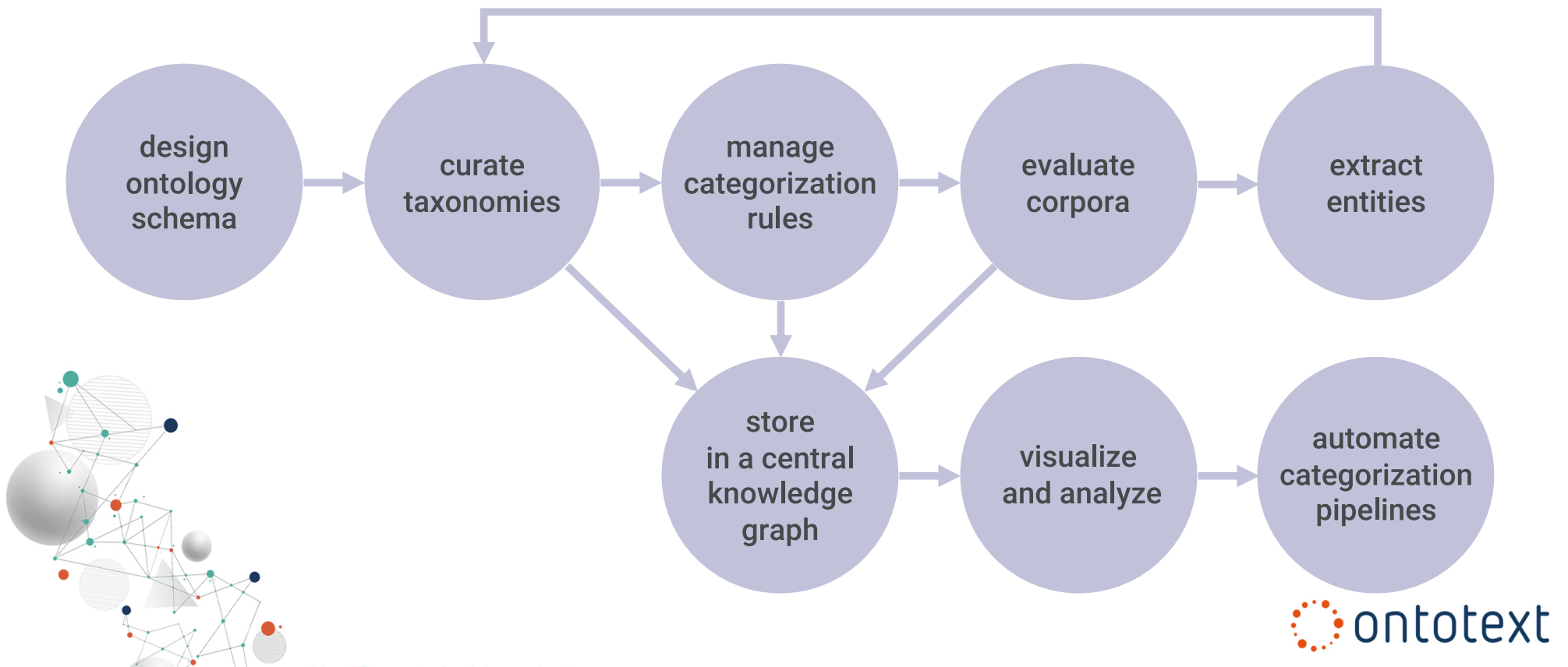
Categorization identifies the **few** concepts and named entities that best describe the **aboutness** of a whole document.

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

ontotext

A taxonomy-driven, autocategorizing alternative

Graphite Knowledge Studio — **synaptica**

Powered by ontotext Text Analytics

design ontology schema → **curate taxonomies** → **manage categorization rules** → **evaluate corpora** → **extract entities**

store in a central knowledge graph → **visualize and analyze** → **automate categorization pipelines**

ontotext

Graphite Knowledge Studio: a partnership

ontotext

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*

synaptica

Graphite Knowledge Studio

Content-aware Knowledge graph   **synaptica** 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.

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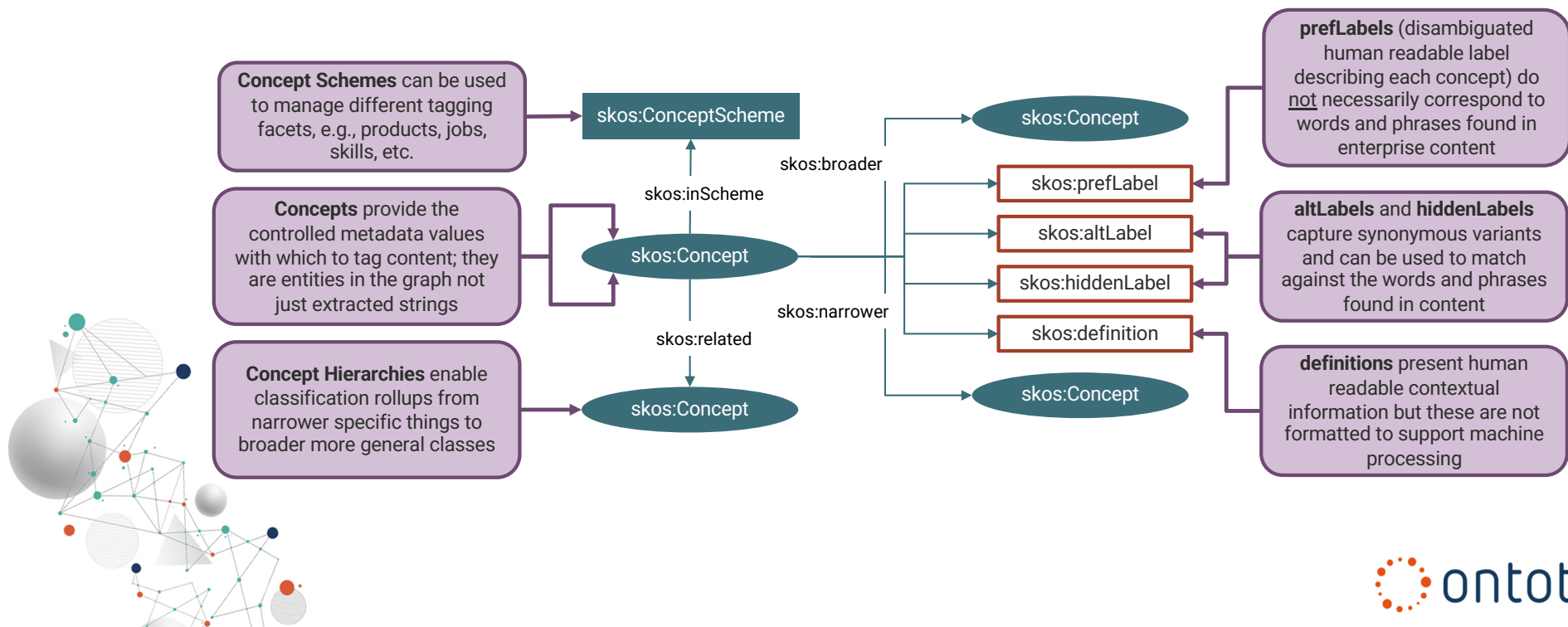
ontotext

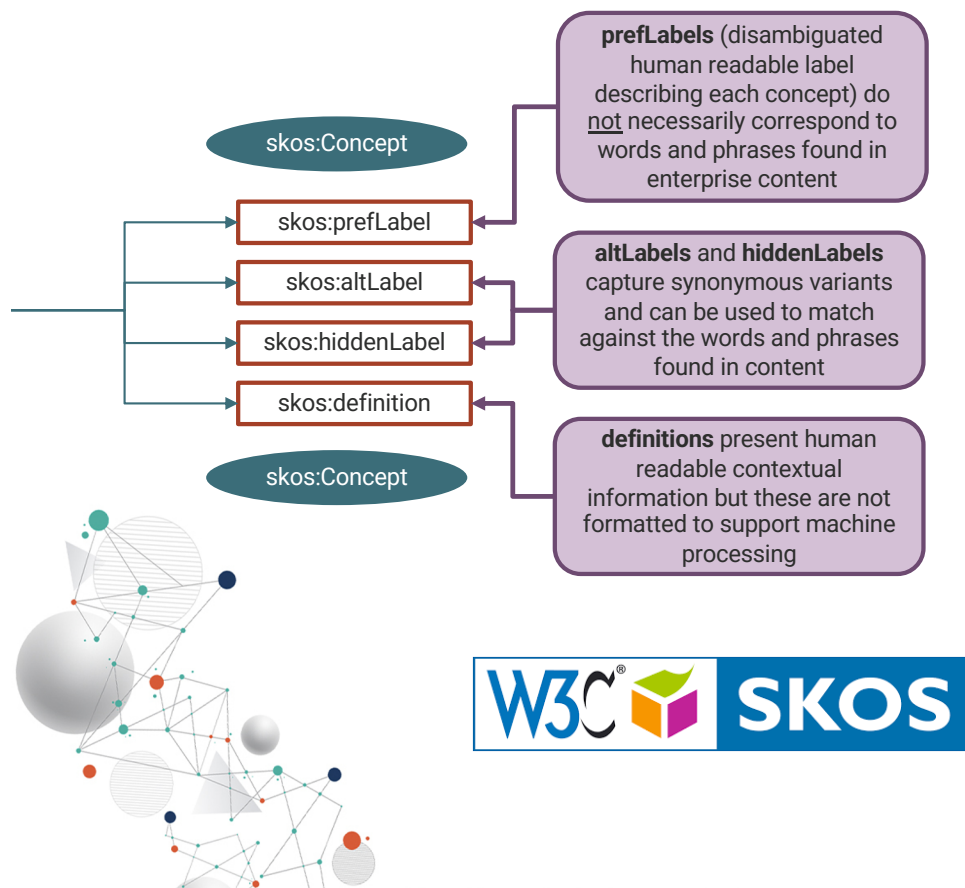# Bootstrapping enterprise taxonomies for autocategorization

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

W3C SKOS

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

Concept Schemes can be used to manage different tagging facets, e.g., products, jobs, skills, etc.

Concepts provide the controlled metadata values with which to tag content; they are entities in the graph not just extracted strings

Concept Hierarchies enable classification rollups from narrower specific things to broader more general classes

skos:ConceptScheme

skos:inScheme

skos:Concept

skos:related

skos:Concept

skos:broader

skos:narrower

skos:Concept

skos:prefLabel

skos:altLabel

skos:hiddenLabel

skos:definition

skos:Concept

prefLabels (disambiguated human readable label describing each concept) do <u>not</u> necessarily correspond to words and phrases found in enterprise content

altLabels and **hiddenLabels** capture synonymous variants and can be used to match against the words and phrases found in content

definitions present human readable contextual information but these are not formatted to support machine processing

ontotext

Bootstrapping enterprise taxonomies for autocategorization

skos:Concept

skos:prefLabel

skos:altLabel

skos:hiddenLabel

skos:definition

skos:Concept

prefLabels (disambiguated human readable label describing each concept) do <u>not</u> necessarily correspond to words and phrases found in enterprise content

altLabels and **hiddenLabels** capture synonymous variants and can be used to match against the words and phrases found in content

definitions present human readable contextual information but these are not formatted to support machine processing

✓

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

W3C® SKOS

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

ontotext

Bootstrapping enterprise taxonomies for autocategorization

skos:Concept

skos:Concept

skos:prefLabel

skos:altLabel

skos:hiddenLabel

skos:definition

skos:Concept

prefLabels (disambiguated human readable label describing each concept) do <u>not</u> necessarily correspond to words and phrases found in enterprise content

altLabels and **hiddenLabels** capture synonymous variants and can be used to match against the words and phrases found in content

definitions present human readable contextual information but these are not formatted to support machine processing

skos:Concept

contentKeywords

mustMatch

mustNotMatch

Set of properties for natural language matching and contextualization

contentKeywords regex

mustMatch regex

mustNotMatch regex

Set of properties for regex pattern-based matching and contextualization

contextualProximity

classificationSpecificity

relevanceRank

Set of properties for tagging relevancy 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 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

ontotext

Human-in-the-loop iterative workflow
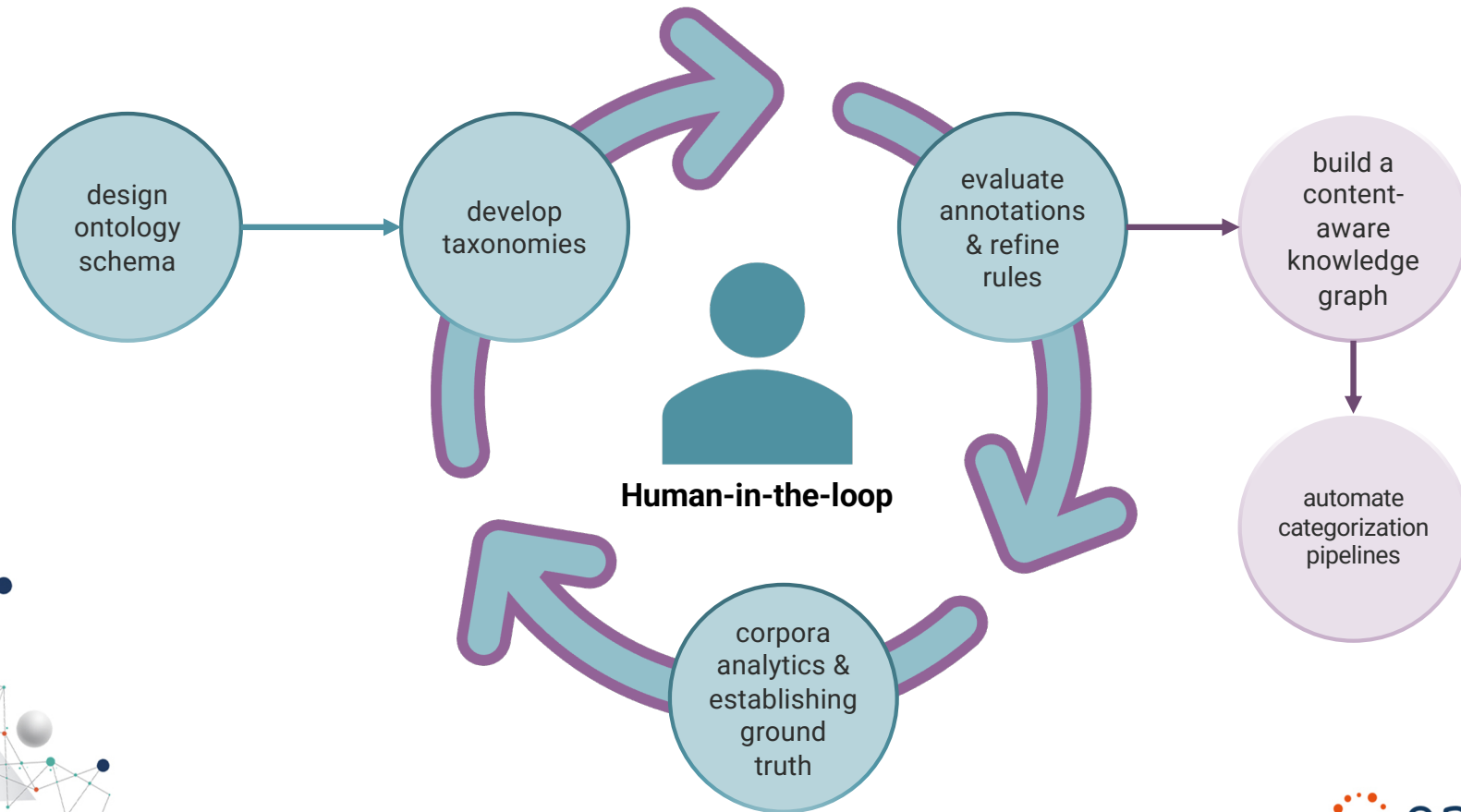
Human-in-the-loop iterative workflow

- **Usable & accessible to non-technical customers**

- **Speed of implementation and iteration**

- **Transparent and explainable**

- **Content-aware knowledge graph powers further insight and functionality**

Human-in-the-loop iterative workflow

Rule-writing: An illustrative example

Rule-writing: An illustrative example

Corpora Analytics

Graphite Knowledge Studio
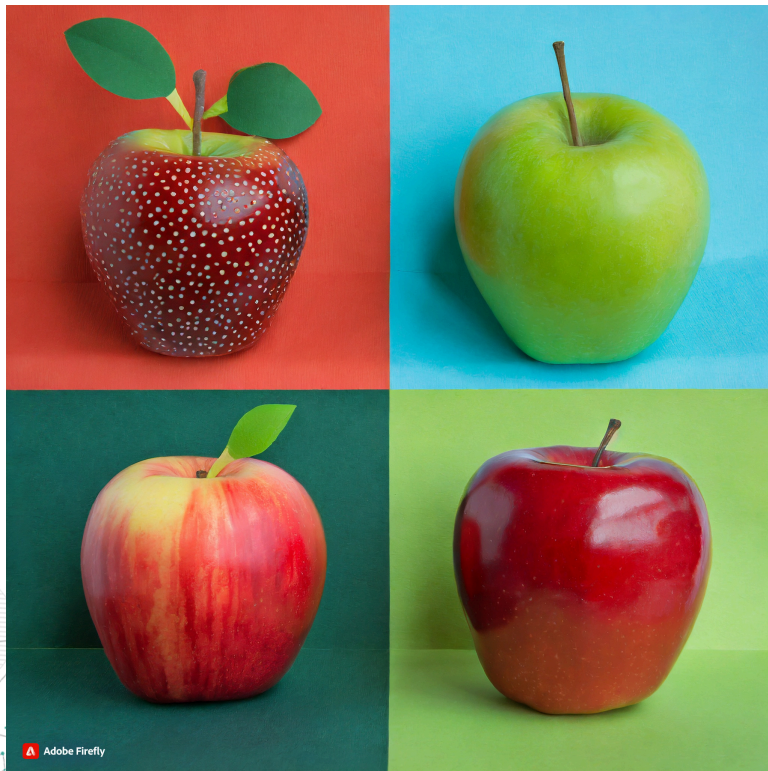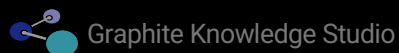
News Corpus Report

Annotation class	True positives	False positives	False negatives
Location	3	2	0
Total	3	2	0
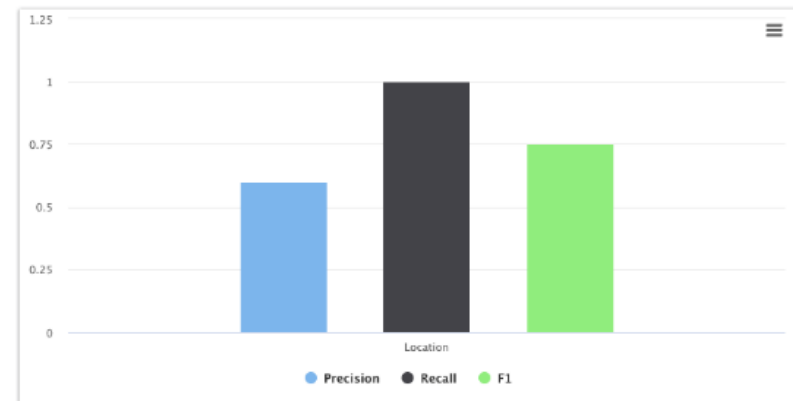
Annotation class	Precision micro	Recall micro	F1 micro	Jaccard Index
Location	0.6	1	0.75	0.6
Total	0.6	1	0.75	0.6

ontotext
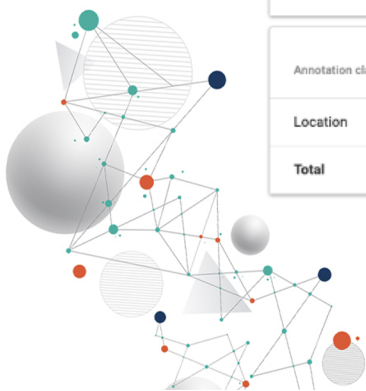
RESTful API Integrations

Graphite Knowledge Studio

Graphite REST API Web Services

Auto-categorization Semantic Tagging API Services

Read-Write

RDF4J API
Read-Write

RDF4J API
Read-Write

Controlled Vocabulary I/O Data Flow

Read-Write

REST

Enterprise Knowledge Graph

Enterprise Business Applications

Inferencing and Analytics Endpoints

GraphQL

SPARQL

Read

GraphDB

RDF Graph

Database

REST / RDF4J / SPARQL / GraphQL

Read-Write

Non-Controlled Vocabulary I/O Data Flow

Eco-system of Enterprise applications

Write

Write

Read-Write

Search Indexes

Lucene
Solr
elastic

Other Databases

File-based CMS

Confluence
SharePoint

ontotext

Add Optional: LLM Overlays

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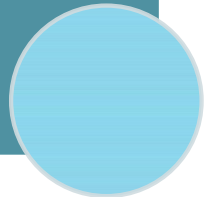- Work with Proprietary Documents & Metadata

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

ontotext

Recap

Through enrichment of SKOS, controlled vocabularies can power autocategorization of enterprise content
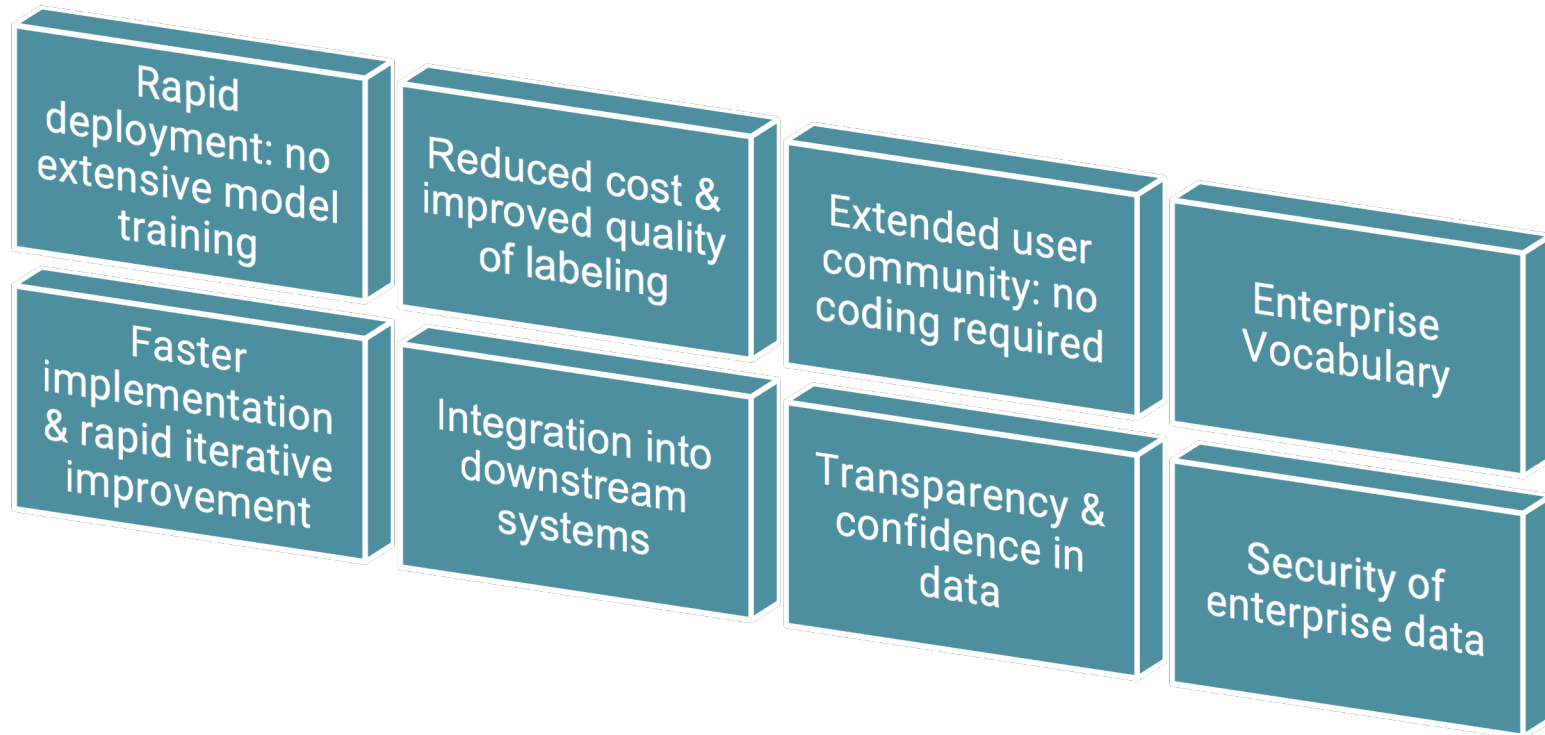
Graphite Knowledge Studio can support non-Data Science users in evaluating, training, and improving NLP text analytics services

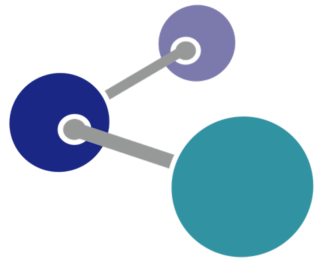Transparency and iterative speed rapidly improve text analytics performance

Enriching the knowledge graph with annotations ("content-aware knowledge graph") fuels further functionality and insight

ontotext

Benefits of Taxonomies + Explainability Taggers

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

ontotext

About Us – Partner description

synaptica

Synaptica provides award-winning taxonomy, ontology, knowledge graph software combined with professional service solutions

Synaptica helps people to organize, categorize and discover enterprise knowledge.

ontotext

![ontotext]

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 for your time!