synaptica

Dave Clarke
Synaptica

Borislav (Borko) Ankov
Ontotext

Introducing

Graphite Knowledge Studio

how to bootstrap enterprise taxonomies for auto-categorization

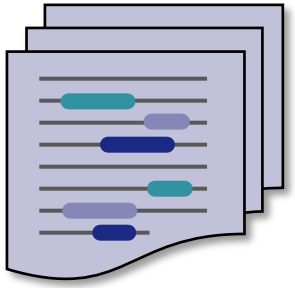
Powered by ontotext Text Analytics

synaptica
Powered by ontotext Text Analytics

Graphite Knowledge Studio

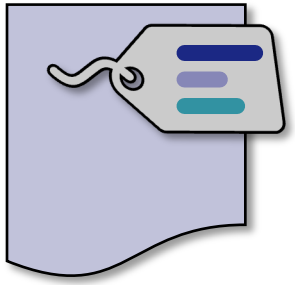
Up-coming Webinars about Auto-categorization

This Webinar March 1st, 2023



Tagging

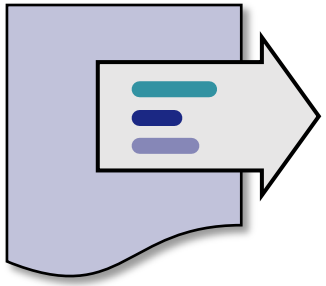
identifies the **many** taxonomy concepts and named entities that are **mentioned** within the full text of a document.



Categorization

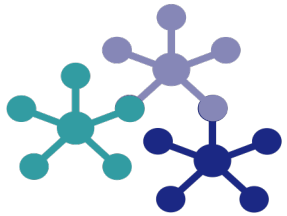
identifies the **few** concepts and named entities that best describe the **aboutness** of a whole document.

Next up Webinars



Extraction

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

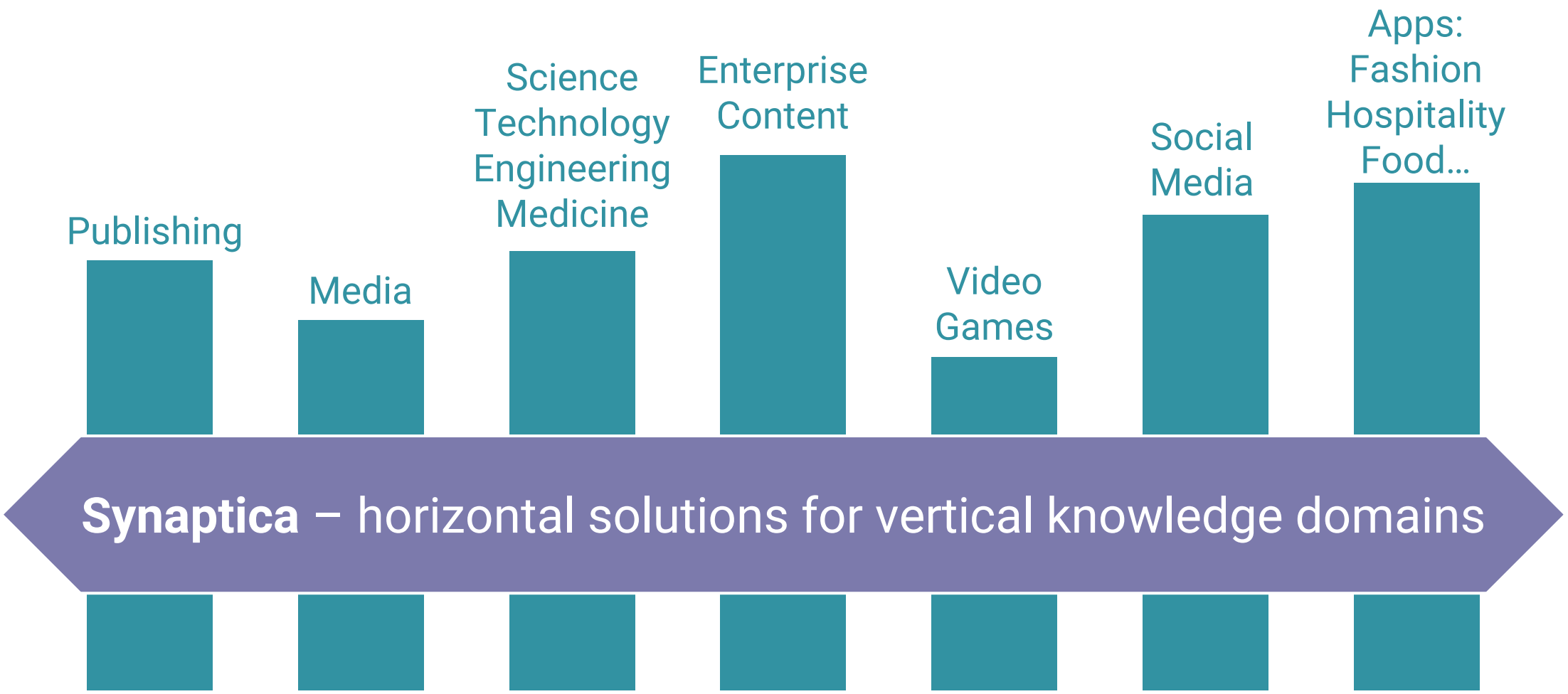


ML & Big Graphs

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

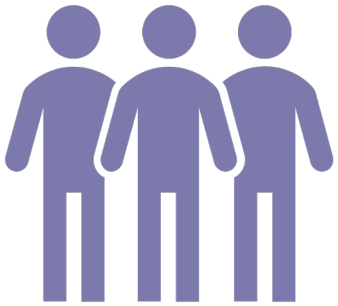
Tagging – Classification – Extraction

1995 Info Science → Info Science + Data Science **2023**



Characteristics of Information Science and Data Science

Information Science



Data Science

Human Curated

Top-Down Description

Human Readable

Defines Knowledge

Computational

Bottom-Up Extraction

Machine Readable

Infers Knowledge

Ontotext and Synaptica – Complementary Technologies – Now Integrated

ontotext

20+ years

RDF graph database

NLP text analytics

synaptica

25+ years

Controlled vocabulary

Taxonomy & ontology

Today's Webinar Agenda

Part 1: Creating a User Experience for Taxonomy-Based Tagging (20 mins)

1. Auto-categorization using un-modified SKOS taxonomies
2. Enhancing categorization precision by extending the SKOS ontology
3. Putting the taxonomist in the driving seat with transparent rules-management
4. Creating an Agile-like workflow for taxonomy development – rule management – corpora management – and rule refinement

Part 2: Showcase Examples (30 mins)

5. Showcase 2: **eCommerce** – tagging products by type and parametric attributes
6. Showcase 1: **job-listings** – tagging job listings by skills, location, industry, etc. AND performing similarity-based recommendations for matching résumés to jobs

Part 3: Conclusions and Q&A (10 mins)

1. Key Take-Aways

synaptica
Powered by ontotext Text Analytics

Graphite Knowledge Studio
Auto-categorization using unmodified SKOS

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

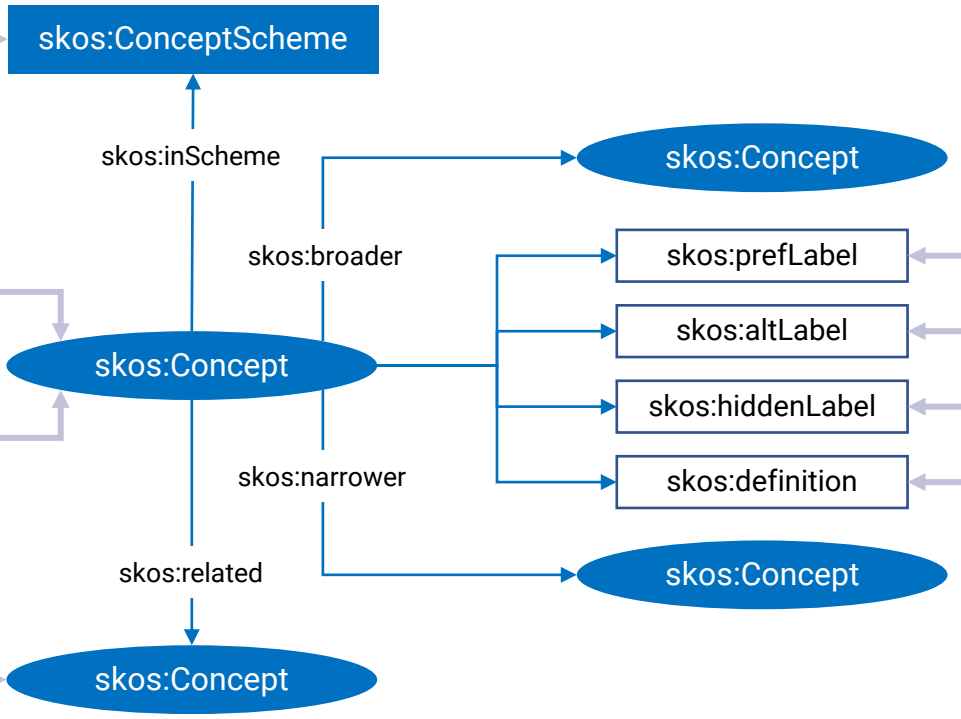
W3C SKOS

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



prefLabels present the disambiguated human readable label describing each concept, but these do not necessarily correspond to the words and phrases found in content

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

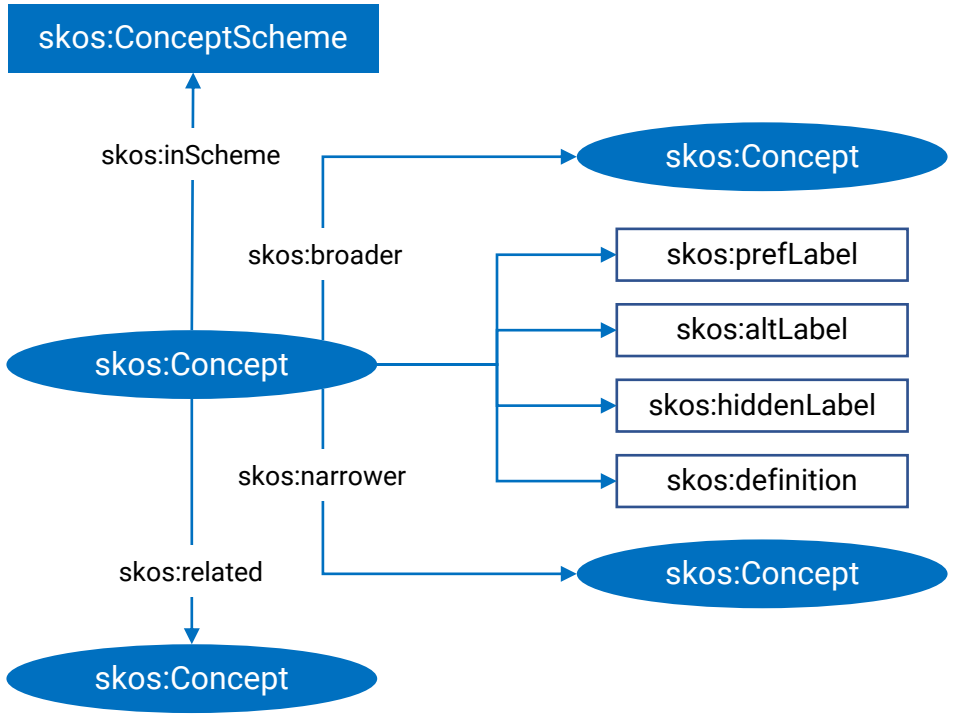
definitions present human readable contextual information but these are not formatted to support machine processing

Graphite Knowledge Studio
Auto-categorization using un-modified SKOS

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

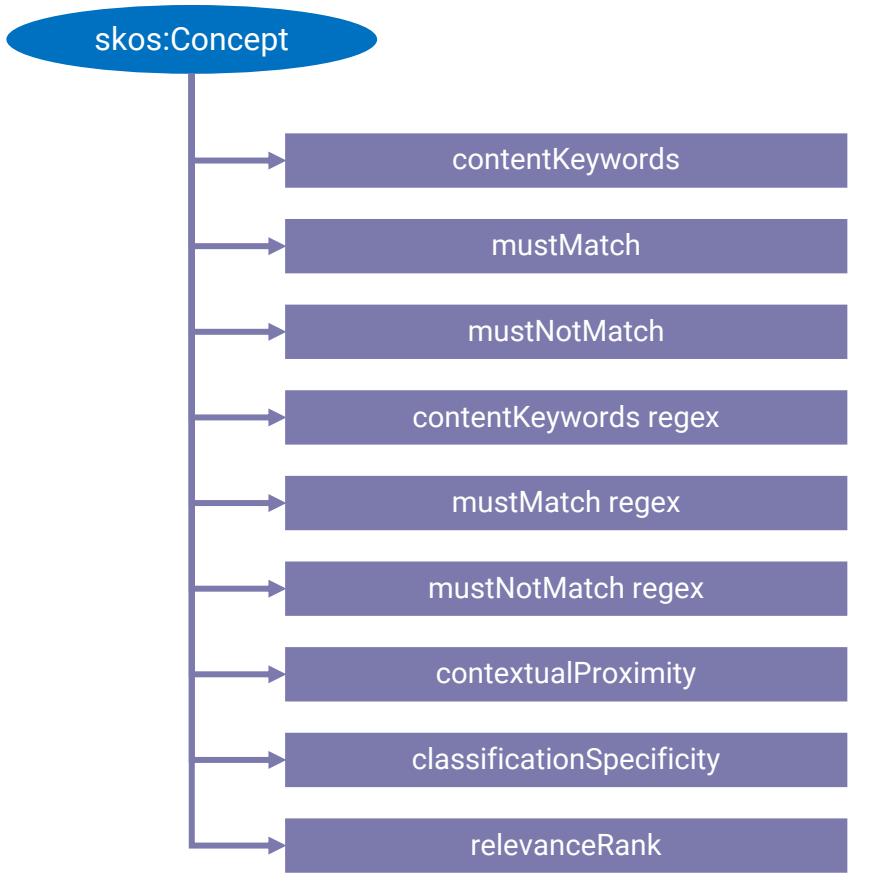


Can’t support positive or negative contexts to help eliminate false matches

Can’t support textual patterns to identify novel entities

Can’t support proximity and relevancy ranking rules

Graphite Knowledge Studio
Graphite's Extensions to SKOS to Support Precision Tagging



Set of properties for natural language matching and contextualization

Set of properties for regex pattern-based matching and contextualization

Set of properties for tagging relevancy rules

Can support positive or negative contexts to help eliminate false matches

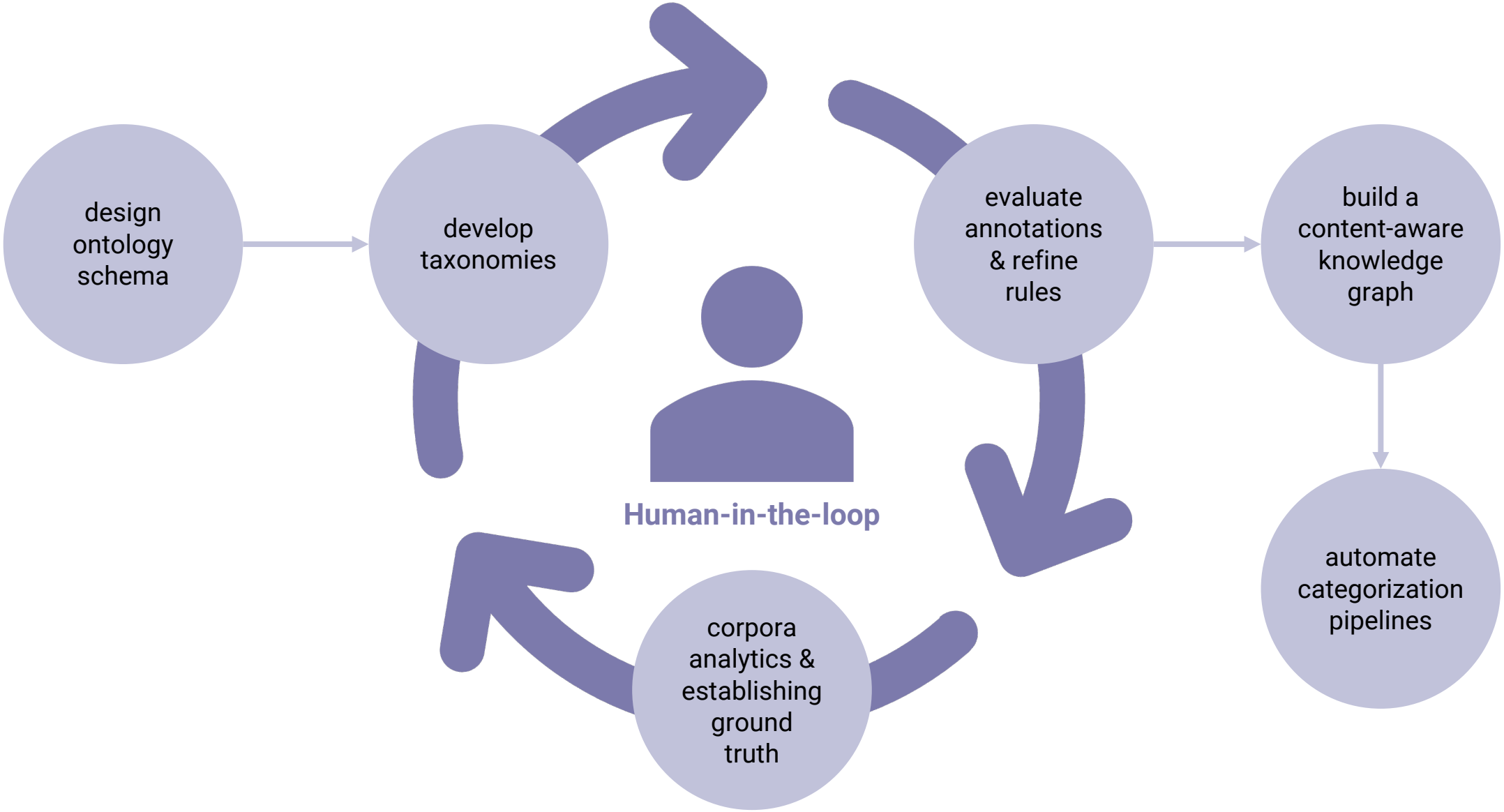
Can support textual patterns to identify novel entities

Can support proximity and relevancy ranking rules

synaptica
Powered by ontotext Text Analytics

Graphite Knowledge Studio

Workflows for taxonomy management and content classification



Key concepts that you are about to see demonstrated (jargon busting)

No-Code Solution: Graphite offers a GUI-driven no-code solution for creating and managing semantic schema, taxonomies, and iterative rules-based auto-categorization

Annotation: tagging content with named entities from taxonomies

Corpora: any set of content / documents used for a tagging process

Aboutness: identifying the few highest relevancy-ranked concepts that best describe the overall 'aboutness' of a document

Inline Tagging: identifying the many concepts mentioned anywhere within the body of a document

Explainability: the ease with which a taxonomist can identify why a concept was tagged to a document, and refine the taxonomy-tagging rules if required

Context: concept labels may match words and phrases found in documents, but not match their semantic context – to eliminate false matches and improve tagging precision taxonomists need to be able to add positive and negative context rules

Ground Truth: also referred to as gold standard training data, means a set of annotations that have been validated as correct after human review – Machine Learning (ML) models need Ground Truth data in order to validate machine tagging

Human-in-the-Loop: workflows that allow humans to review the result of machine processes and intervene to refine tagging precision, and establish Ground Truth data

Agile-like Development: borrowed from software development practices, Graphite supports Agile-like categorization workflows based on the early delivery of results and continual evolutionary improvement

synaptica
Powered by ontotext Text Analytics

Graphite Knowledge Studio
Live Demos

Borislav (Borko) Ankov
Ontotext

Demo 1

eCommerce: inline tagging product pages to taxonomies of products, features and attributes

Demo 2

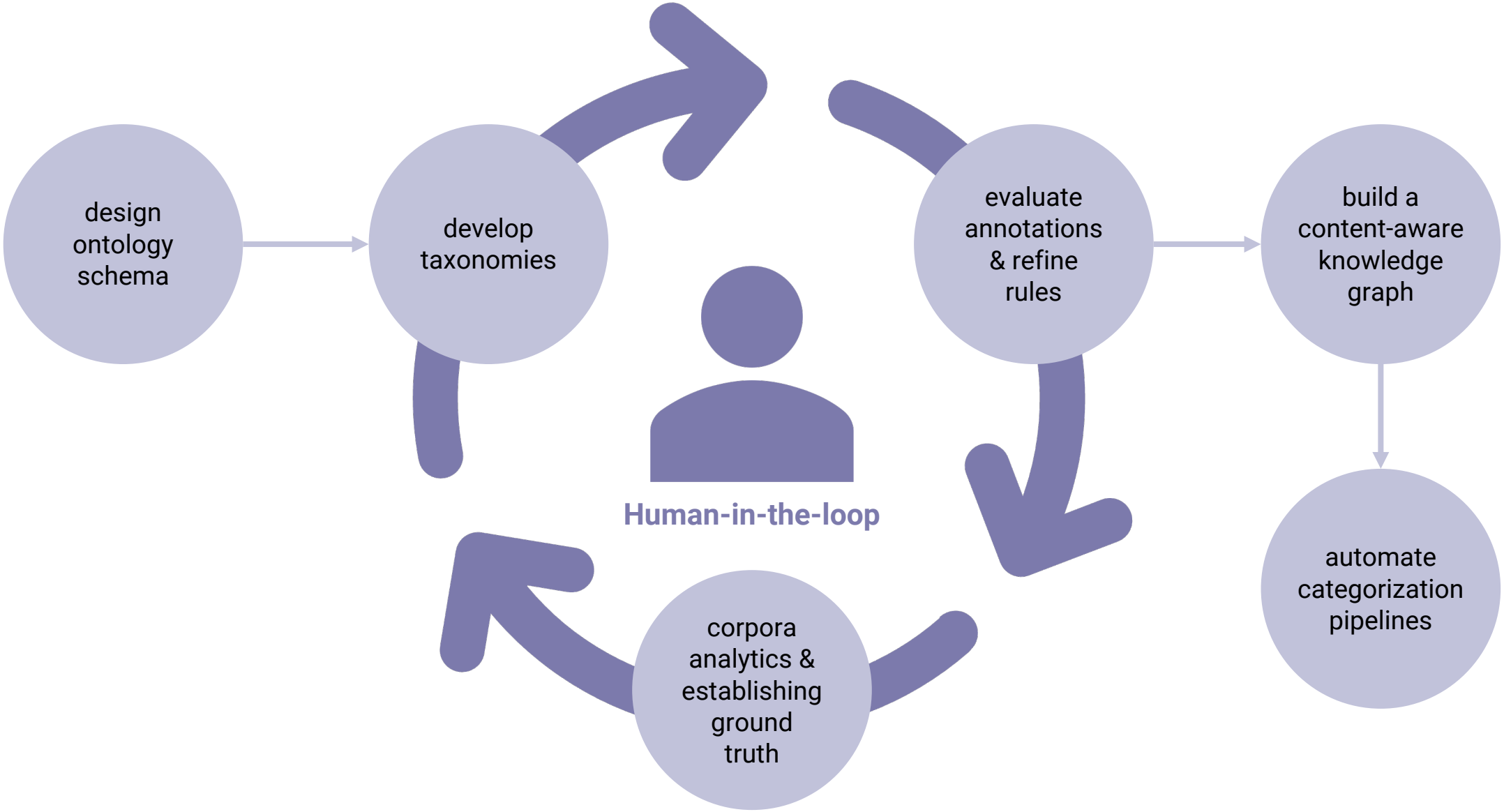
Jobs: document tagging job listings by occupation, skills, location, etc.

then similarity matching résumés to recommend the best fit job listings

synaptica
Powered by ontotext Text Analytics

Graphite Knowledge Studio

Workflows for taxonomy management and content classification



synaptica

Powered by ontotext Text Analytics

Graphite Knowledge Studio

Workflows for taxonomy management and content classification

design ontology schema

develop taxonomies

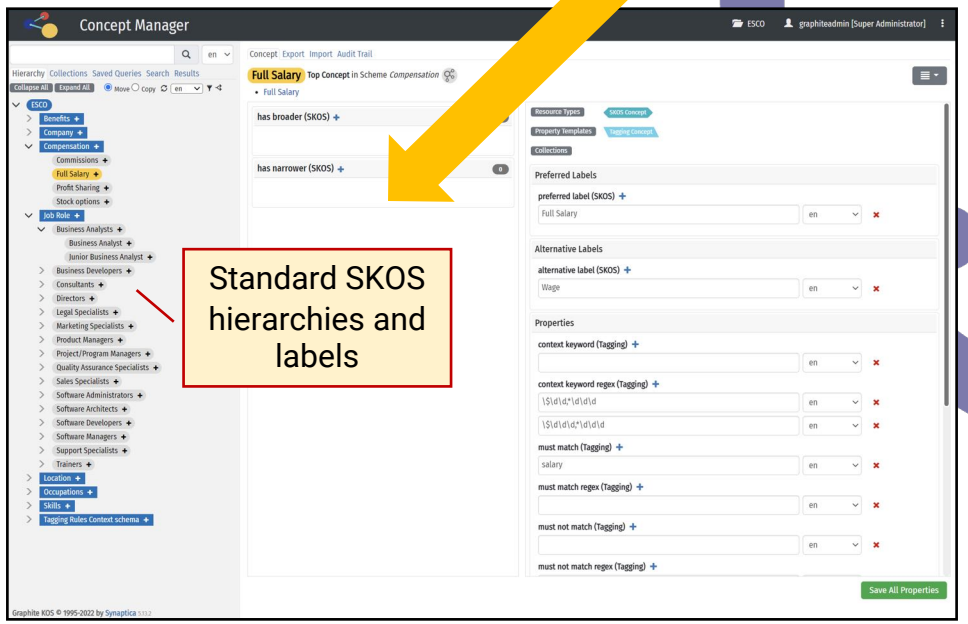
evaluate annotations & refine rules

build a content-aware knowledge graph

automate categorization pipelines

corpora analytics & establishing ground truth

Human-in-the-loop



Standard SKOS hierarchies and labels

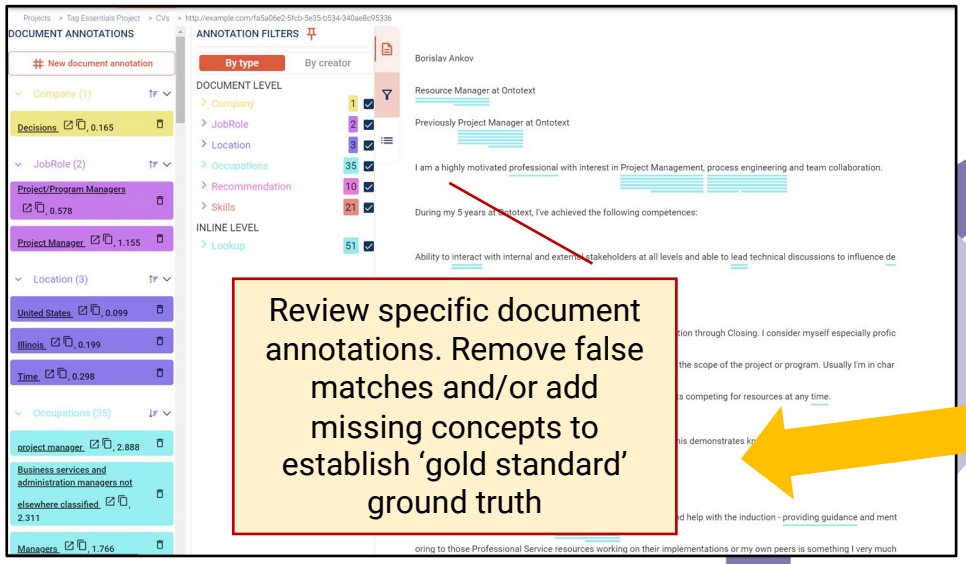
Synaptica

synaptica

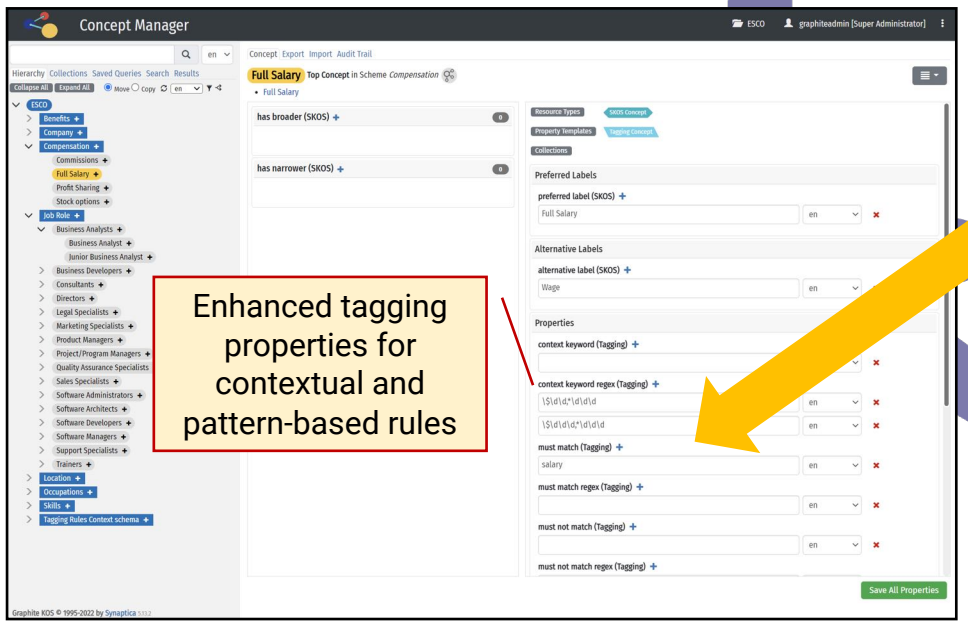
Powered by ontotext Text Analytics

Graphite Knowledge Studio

Workflows for taxonomy management and content classification



Review specific document annotations. Remove false matches and/or add missing concepts to establish ‘gold standard’ ground truth



Enhanced tagging properties for contextual and pattern-based rules

evaluate annotations & refine rules

build a content-aware knowledge graph

Human-in-the-loop

automate categorization pipelines

corpora analytics & establishing ground truth

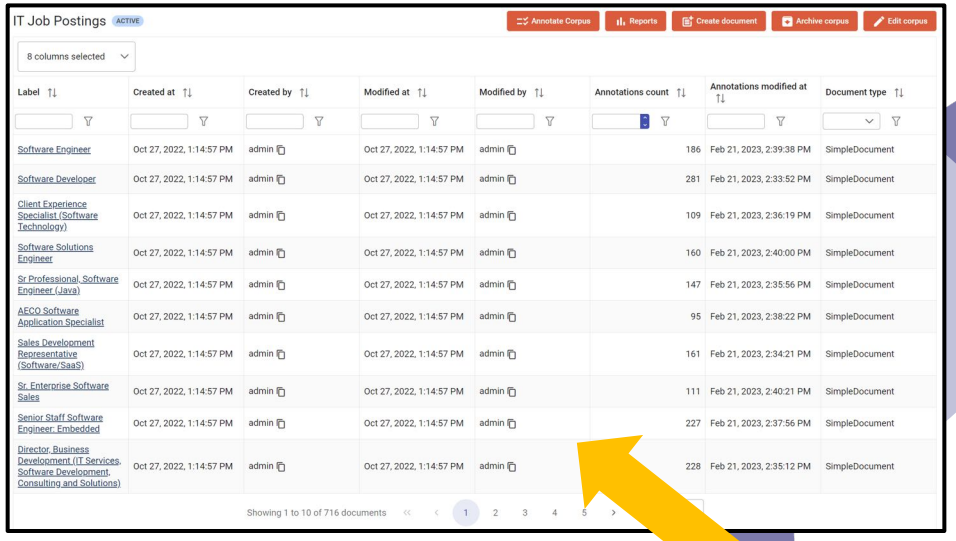
Synaptica

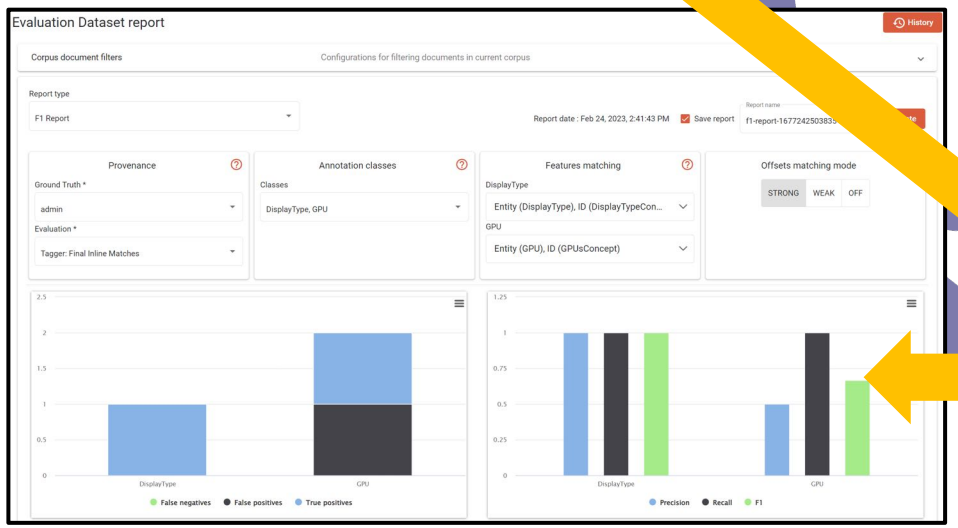
synaptica

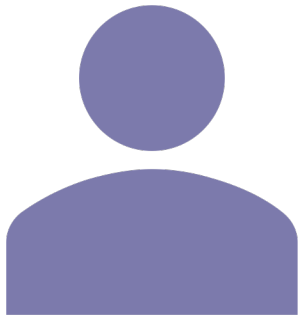
Powered by ontotext Text Analytics

Graphite Knowledge Studio

Workflows for taxonomy management and content classification







Human-in-the-loop

evaluate annotations & refine rules

build a content-aware knowledge graph

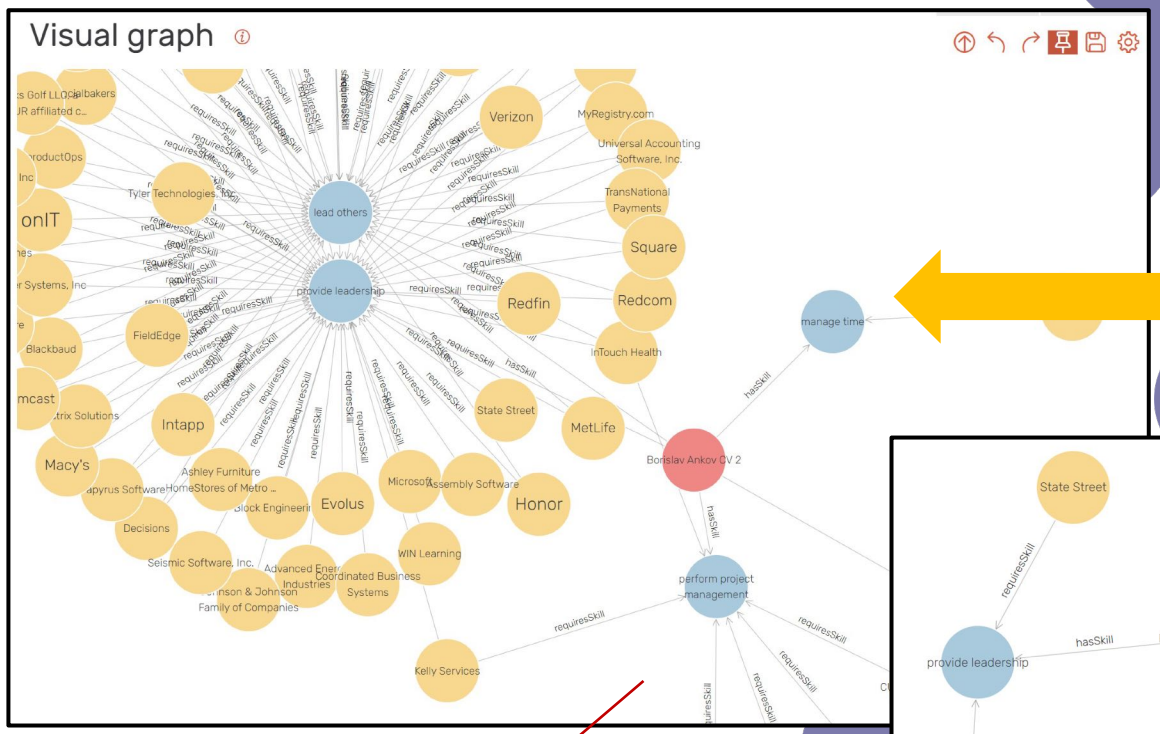
automate categorization pipelines

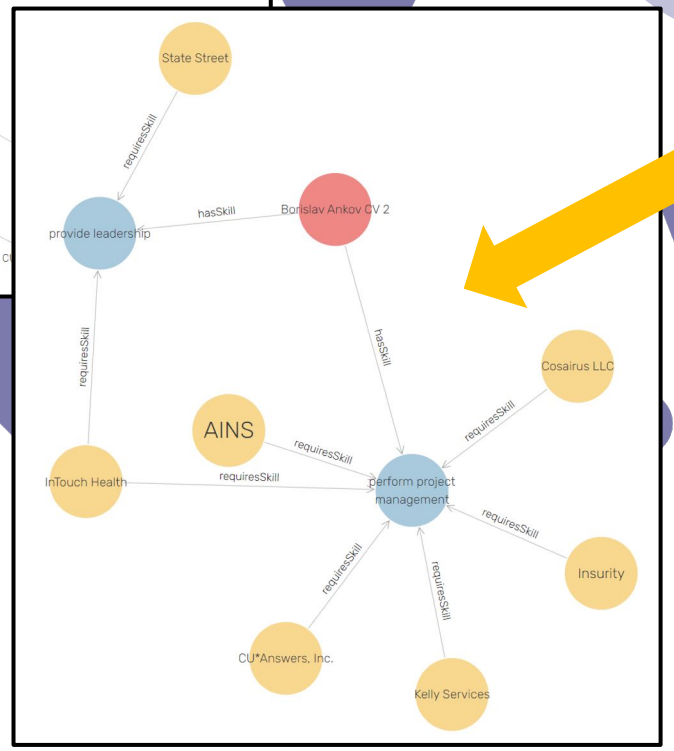
corpora analytics & establishing ground truth

synaptica
Powered by ontotext Text Analytics

Graphite Knowledge Studio

Workflows for taxonomy management and content classification





content-aware knowledge graphs can power semantic similarity search

evaluate annotations & refine rules

build a content-aware knowledge graph

automate categorization pipelines

synaptica
Powered by ontotext Text Analytics

Graphite Knowledge Studio

Enterprise Knowledge Graphs are 'Content-Aware'

If you do auto categorization and only store all your metadata tags in search indexes and content management systems

then you are missing out on building enterprise knowledge graphs

build content-aware knowledge graphs

Bring content metadata back into the graph

build enterprise knowledge graphs

empower business process automation and insights

synaptica
Powered by ontotext Text Analytics

Graphite Knowledge Studio
Tools for cross-team collaboration

Data Scientists & Knowledge Engineers

who use text analytics and graphs to analyze data and content, to develop business insights, and to infer new knowledge

Taxonomists & Ontologists

who model knowledge systems and design semantic schema, who control vocabularies, build hierarchies, and map associations

Information Architects

who design the user experiences by which people search, browse, associate, recommend, filter and discover information

Content & Metadata Managers

who know their content, understand their audience, and know how to bring them together

Tools for cross-team collaboration

Graphite Knowledge Studio
Review of Key Features and Benefits

Features	Benefits
No-code solution GUI-driven	Rapid bootstrap deployments Lower cost of entry Extends user community to embrace Taxonomists and Content Managers as well as Data Scientists Compresses learning curve
Standards-based controlled vocabulary and semantic schema management	Immediate import and utilization of existing enterprise taxonomies Extensibility of semantic schema for enterprise-specific domain knowledge
Explainability with transparent and editable tagging and classification rules	Faster implementation and Agile-like development iterations Early delivery of usable results without pre-requirement of large training sets Avoid unexplainable and unmodifiable 'black box' components
Human-in-the-loop	View inline tagging and document annotations, correct errors and omissions, refine tagging rules Establish Ground Truth datasets to compare with and improve machine learning and processes
Build Content-Aware Enterprise Knowledge Graphs	Manage taxonomies and ontologies in the graph and capture content annotations in the graph Content-aware graphs drive business analytics Generate similarity-based recommendations and insights

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

scan to access downloads