Synaptica smart solutions for managing enterprise knowledge

Ontologies, Knowledge Graphs & Semantic Classification

Knowledge Graph Forum - September 26th 2022 Dave Clarke Founder, Synaptica

Bringing together data science and information science Library and Information Science

1

Information Science

4

Classification and Retrieval

2

Taxonomy and 3 Knowledge Organization Systems

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.

Organizational and human knowledge ... overlap with KM

Data Science



IT capabilities driving a changing landscape

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 <u>mathematics</u>, <u>statistics</u>, 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]}

Maths and statistics

Noisy data

Information

extraction

2

3

4

Not Information Science

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

Messy Internal Content

APIs Articles

Blogs Courses

Podcasts

Series

Tutorials

Videos

Events

Messy External Data



https://developer.ibm.com/tutorials/ba-cleanse-process-visualize-data-set-1/



https://lod-cloud.net

Information Science

Data Science

Curated Top-Down Defines Knowledge Human Readable Computational Bottom-Up Infers Knowledge Machine Readable

Introducing Graphite Knowledge Studio Graphite Knowledge Studio

ontotext

1. Suite of enterprise taxonomy and ontology management tools

synaptica

- 2. Semantic tagging API services with transparent and explainable rules allowing taxonomists to directly control auto-categorization outcomes
- 3. Taxonomies, annotations, and corpora managed in an RDF graphdatabase with corpora analytics

Managing ontologies and taxonomies



Managing Taxonomies and Ontologies

An Ontology comprises a semantic Schema of class, property and relationship types, plus a Taxonomy of specific concepts, classes and named individuals.





Graphite taxonomy management system

Simple user interface

Security & permissions

Workflow & Governance

Namespace and Predicate Admin

	Create Namespace Namespace and P
✓ Graphite Namespace and Predicates	Cranhite No.
> DCT	Graphite Na
> Graphite Chemistry	Garrenter
> Graphite Knowledge Model	Namespaces
Graphite SharePoint	DCT Graphit
> Graphite System	Graphite Syste
> OWL	
> PROV	Predicates not
> schemaOrg	100.000.000
✓ SKOS	Search
Properties +	Abstract (DCT)
alternative label	Accrual Period
change note	Active (Graphit
definition	alternative lab
editorial note	approverComn
example	Audience (DCT
hidden label	Autocomplete
history note	Conforme To /
notation	Conforms to (E
note	Coverage (DCI)
preferred label	Date Available
scope note	Date Submitte
 Associative Relationships + 	Day Of Week Te
has broader match	Delete Concep
has close match	disjointWith (O
has exact match	editorCommen
has narrower match	Enable Logging
has related	Error Report Er
has related match	Favicon (Graph
Is in mapping relation with	generatedAtTir
has broader / has parrower	hadMember (P
has broader (has harrower has broader transitive / has n	hadlisage (PRC
> SKOS-XL	has narrower to
Graphite KOS © 1995-2022 by Synaptica 5.13.2	r trainerter e

Namespace and Predicates Import

Graphite Namespace and Predicates

lamespaces	•)
DCT Graphite Chemistry Graphite Knowledge Model Graphite	SharePoint	
Graphite System OWL PROV schemaOrg SKOS SKOS-XL	Tagging	
redicates not in Use	100 of 155)
Search	Q Reset	
Abstract (DCT) Access Rights (DCT) Accrual Method (DCT)		
Accrual Periodicity (DCT) Accrual Policy (DCT) actedOnBehalfOf	(PROV)	
Active (Graphite SharePoint) activity (PROV) agent (PROV) alt	ternateOf (PROV)	
alternative label (SKOS-XL) Alternative Title (DCT)		
approverComment (Graphite Knowledge Model) atLocation (PROV)) atTime (PROV)	
Audience (DCT) Audience Education Level (DCT)		
Autocomplete Dropdown Items (Graphite System) Bibliographic Ci	itation (DCT)	
Conforms To (DCT) contentKeyword (Graphite Knowledge Model)	Contributor (DCT)	
Coverage (DCT) Creator (DCT) Date (DCT) Date Accepted (DCT	r)	
Date Available (DCT) Date Copyrighted (DCT) Date Issued (DCT)	1	
Date Submitted (DCT) Date Valid (DCT) Day Of Month To Run (Gr	raphite SharePoint)	
Day Of Week To Run (Graphite SharePoint)		
Delete Concepts In Collection (Graphite SharePoint) differentFrom	(OWL)	
disjointWith (OWL) doNotEdit (Graphite Knowledge Model)		
editorComment (Graphite Knowledge Model) Email (Graphite Systematics Systemate	em)	
Enable Logging (Graphite SharePoint) endedAtTime (PROV) equ	uivalentClass (OWL)	
Error Report Email Recipients (Graphite SharePoint) example (SKO	S) Extent (DCT)	
Favicon (Graphite System) Format (DCT) generated (PROV)		
generatedAtTime (PROV) hadActivity (PROV) hadGeneration (PR	(VO)	
hadMember (PROV) hadPlan (PROV) hadPrimarySource (PROV)	hadRole (PROV)	
hadUsage (PROV) Has Annotation (Graphite System) Has Format	t (DCT)	
has narrower transitive (SKOS) Has Part (DCT) Has Version (DCT))	



🗁 News 👤 graphiteadmin [Super Administrator] 🚦

Metadata

Total Namespaces: 11 Total Properties: 134 Total Associative Relationships: 73 Total Hierarchical Relationships: 4 Total Predicates: 211

Manage Schema

Visualizing the Schema of a Project





Concept Manager			🗁 Training Datasets 👤 I	DaveClarke (Super Administrator)
Q	en 🛊 Concept Export Import Audit Trail			
Hierarchy Collections Saved Queries Search Results Collapse All Expand All O Move Copy C (en a	Canines in Scheme Living Things			≡
✓ (Training Datasets)	has broader (SKOS) 🔻 🛨	0	Resource Types SKOS Concept	
Chemistry (SKOS) + ESCO Occupations +	Animals 🗙		Property Templates Governance Taxonomy Collections	
ESCO SKILS + GEMET Thesaurus +	has narrower (SKOS) 🔻 💠	0	Preferred Labels	
> IPTC +	Borzoi 🗙 Golden Retriever 🗙 Pointer 🗙		preferred label (SKOS) +	
Animals +	has related (SKOS) 🔻 🕂	O	Canines	en 💠 🗙
Cannes + Borzoi +	Vets ×		Alternative Labels	
Pointer +			alternative label (SKOS) +	
✓ Felines +			Dogs	en 💠 🗙
bengal + Oak trees +			Hounds	en 💠 🗙
V Plants +			Properties 🕤	
Grasses +			definition (SKOS) +	
Deciduous Trees +			A domesticated carnivorous mammal that typically ha a long snout, an acute sense of smell, non-retractable claws, and a barking, howling, or whining voice.	en 🛊 🗙
Vets + Industry_Classification +			readyForApprover (Graphite Knowledge Model) Yes No ×	
urus + urus +			rejectedByApprover (Graphite Knowledge Mode	0
			approverComment (Graphite Knowledge Model)	+
			smeComment (Granhite Knowledge Model) 🛨	en 💠 🗙
Ivianage				en 🛟 🗙
			Metadata	
Concente			Status: Candidate \$	View \$
ooncepto				

Semantic tagging and classification



Semantic Tagging identifies the many

taxonomy concepts and named entities that are **mentioned** within the full text of a document. It uses concept labels, disambiguators and contextual rules.

Categorization identifies the few concepts

and named entities that best describe the **aboutness** of a whole document. It uses term frequency, document zone relevancy, semantic proximity, inferencing, and TF-IDF.



/

Entity Extraction identifies new named entities (people, places, organizations, etc.) and/or conceptual entities, that are found in the full text of a document and **not** present in the taxonomy. It uses Natural Language Processing (NLP).

Big knowledge graphs, such as DBpedia, GeoNames and Wikidata, can be used for disambiguation and to identify content **recommendations**. They use inferencing and similarity matching to identify connections between concepts and content.





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.

Graphite Knowledge Studio

What you can do in Knowledge Studio



0

Graphite Knowledge Studio

Taxonomists & Ontologists

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

Data Scientists & Knowledge Engineers

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

Information Architects

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

> Common tools for different stakeholder communities

Content & Metadata Managers

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

Graphite Knowledge Studio

Taxonomy and Tagging Rules Managed as Ontological Schema

- In Studio one or more concept schemes facilitate facetted tagging for different metadata elements
- Tagger recognizes standard SKOS or OWL labels and hierarchies
- Other properties may be used to extend the schema
- Inter-scheme hierarchies are supported
- Broad categories may be managed in one scheme and used for document-level categorization
- While more granular taxonomies of specific concepts may be managed in other schemes and used for inline tagging
- One application for inter-scheme hierarchies is to link broad categories to hierarchy branches containing relevant specific concepts

Taxonomy

Management

Thereby, through the process of inline tagging the specific concepts mentioned within content, it is possible to aggregate pathways to parent categories and determine document-level aboutness

- Studio gives the taxonomist direct control of a transparent set of tagging rules
- Natural language keywords or phrases may be added concepts to provide contextual must/mustNot rules for tagging disambiguation
- Optionally, advanced regular expression must/mustNot contextual rules may be added
- Boost rules
- Occurrence count rules
- Description



tagging:contextDefinition

Context for the tagging must not rules. tagging:contextDefinition

Tagging **Rules**

- System-wide contextual proximity rules may be defined by sentences or tokens
- Sentence *n* or Token *n* where n=1 means within 1 sentence and *n*=0 means within same sentence
- Contextual proximity rules set the effective distance between a concept's set of labels and boost rules, and its set of must/mustNot rules

Contextual Proximity

Bridging the Human v.	Concept Manager Q en Concept Export Import Audit Trail rarchy Collections Saved Queries Search Results ollapse All Expand All Move Copy Copy T	mi •	inister (government) politics > government > min	in Scheme <i>IPTC </i> %
Machine-	IPTC + > arts, culture, entertainment and media + > trac (DDT) :	0	(not state to be	
	> conflict, war and peace +		Preferred Labels	
Readable Gap	<pre>disaster, accident a economy, business description (Tagging) +</pre>	Ĩ	minister (government)	Explicit disambiguator in
	> environment + A politician who hoads a dovorpmont		Alternative Labels	parentheses
	> health + A politician who neads a government		alternative label (SKOS) +	
Natural language	> tabour + > lifestyle and leisure > politics + > election + > election +		Properties description (Tagging) +	Hierarchy provides category context
definitions are readable by humans	government +	0	A politician who heads a government department and works in conjunction with other ministers to make and implement decisions on policy	Both are human and machine
and by NLP ML	> defence + espionage and intelligence +		must match (Tagging) +	readable for
semantic similarity	executive (government) + has narrower (SKOS) +	0	en 🗸 🗙	semantic
algorithms	> government budget + government department +		must match regex (Tagging) +	contextualization
	heads of government +		[^\p[L]]minister ~ X	
To disambiguate	impeachment + must not match regex (Tagging) +		government ~ X	
ambiguous concepts	local government and authority + [^\p[L]]church	~	ministry 🗸 🗙	
simple NOT words	national government +	_	must not match (Tagging) +	
and phrases can be	> parliament + Christian	~	en 🗸 🗙	
and prinases call be	political convention + liturgy	~	must not match regex (Tagging) +	
machino oliminato	authority			
false matches	regulatory authority +			
	S development policy 1		uturgy 💙 🗙	

Novel Information Extraction

NLP coupled with big knowledge graphs can help with the identification and disambiguation of entities

news.google.com · 12 hours ago

home for all to see.

This demo example uses NLP and Wikidata

Best ever image of the universe revealed by Nasa's James Webb telescope



aboutness Precision vs recall bias Many links NLP most suited to extraction of novel Nasa and the James Webb Space Telescope just peered billions of years further into the past than ever before, and brought photos Annotation types named entities (people, places, Around 6.20pm EDT, US President Joe Biden revealed the very first public image taken by the Webb telescope, a sea of distant Topic 48 organizations, etc.) galaxies and brilliant spiky stars, the most distant view yet of the Cosmos . The image was presented on a screen at the White House Cocation 16 along with comments by Nasa Administrator Bill Nelson . Extraction of Organization 11 "Mr President, if you held a grain of sand on the tip of your finger at arms length, that is the part of the universe you are seeing," Mr abstract topical Nelson said, a tiny portion of the sky magnified by Webb to reveal thousands of galaxies. "That light you are seeing on one of those Person 10 concepts is harder little specks, has been traveling for over 13 billion years." to achieve Software 6 Nasa's Twitter account published a higher resolution version of the image than was projected at the White House.

This article is about

Telescope

Präsident der Vereinigten

Twitter

James Webb Space

Bill Nelson

White Hous

space observatory launched in 2021

Finding specific

things mentioned in

text is easier than determining overall

document

Stars in the foreground appear as brilliant blue white spikes while the most distant galaxies appear as grange-re-

Graph database and corpora analytics



			e ac	imii
Home > Projects > News > AP News June 20 2022 > http://example.com/54d344d1-e553-543d-b677-f88000e60136				
0	Document a	nnotations A	nnotation filters	2
Le Pen: Huge gains in French parliament a 'seismic event'				
PARIS (AP) - Far-right leader Marine Le Pen said on Monday that her party's extraordinary surge in the country's parliamentary election is a	APSubjects	Laws,	ticon	-
historic victory" and a "seismic event" in French politics.		cept/ac104321	6	٦
Many voters in Sunday's poll opted for far-right or far-left candidates, denying President Emmanuel Macron's centrist alliance a straight majority n the National Assembly.		Foreign policy,	t/con	đ
Le Pen's National Rally got 89 seats in the 577-member parliament, up from a previous total of eight. On the other side of the political spectrum, he leftist Nupes coalition, led by hardliner Jean-Luc Melenchon, won 131 seats to become the main opposition force.		Censures,	t/con	Ō
Macron's centrist alliance Together! won the most seats – 245 – but fell 44 seats short of a straight majority in the National Assembly, France's most powerful house of parliament.		cept/ac104435 Constitutions,	4 !/con	ñ
The outcome of the legislative election is highly unusual in France and the strong performance of both Le Pen's National Rally and Melenchon's		cept/ac104394	1	Ŭ
coalition — composed of his own hard-left party, France Unbowed, the Socialists, Greens and Communists — will make it harder for Macron to mplement the agenda he was reelected on in May, including tax cuts and raising France's retirement age from 62 to 65.		Voting,	t/con	Ō
"Macron is a minority president now. () His retirement reform plan is buried," a beaming Le Pen declared on Monday in Hénin-Beaumont, her stronghold in northern France, where she was reelected for another five-year term in the parliament. "It's a historic victory () a seismic event."		cept/ac104390 Elections,		_
She told reporters: "We are entering the parliament as a very strong group and as such we will claim every post that belongs to us." As the biggest single party in the parliament – Macron and Melenchon both lead coalitions – she said National Rally will seek to chair the parliament's		cept/ac104389	3 Jections	U
powerful finance committee, one of the eight commissions that oversee the national budget.		cept/ac104521	l/con	Ō
Le Pen's far-right party now has a sufficient number of legislators to constitute a formal group at the National Assembly and request seats in other committees, including a parliamentary investigation committee and those focusing on defense and foreign policy.		Events,	t/con	Ō
In addition, the National Rally party now has enough seats – more than 58 – to trigger a censure motion against the government that can lead to a no-confidence vote.		cept/ac1041792 National budget	2 ts,	
Prime Minister Elisabeth Borne suggested Sunday evening that Macron's alliance will seek to find "good compromises" with lawmakers from diverse political forces.	Decele	cept/ac104412	8 8	0
	People	Marine Le Pen,		-
Selected: Marine Le Pen 📋 Delete 🕞 Save 🍟 Highlight similar		Re 56	esolved: 100% 5 targets, 0 conflic	teo
Type: Lookup				

Document level inline tagging review



Synaptica smart solutions for managing enterprise knowledge

Corpora level analytics



Graph database – the foundation of the enterprise knowledge graph



Highly scalable RDF graph database supporting high performance semantic search, data reconciliation, analytics and visualization.

The foundation of the enterprise knowledge graph.

Take aways



Graphite Knowledge Studio

- 1. Managing enterprise knowledge graphs combines data science and information science
- 2. Stakeholders need a common platform of tools to design ontology schema; curate taxonomies; and manage corpora tagging and analytics
- 3. Graphite Knowledge Studio provides a deep integration of Synaptica's taxonomy and ontology tools with Ontotext's semantic tagging classification tools
- 4. Tier one of this implementation emphasizes transparency and explainability giving taxonomists direct control over the tagging process and outcomes

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

Thank you

info@synaptica.com