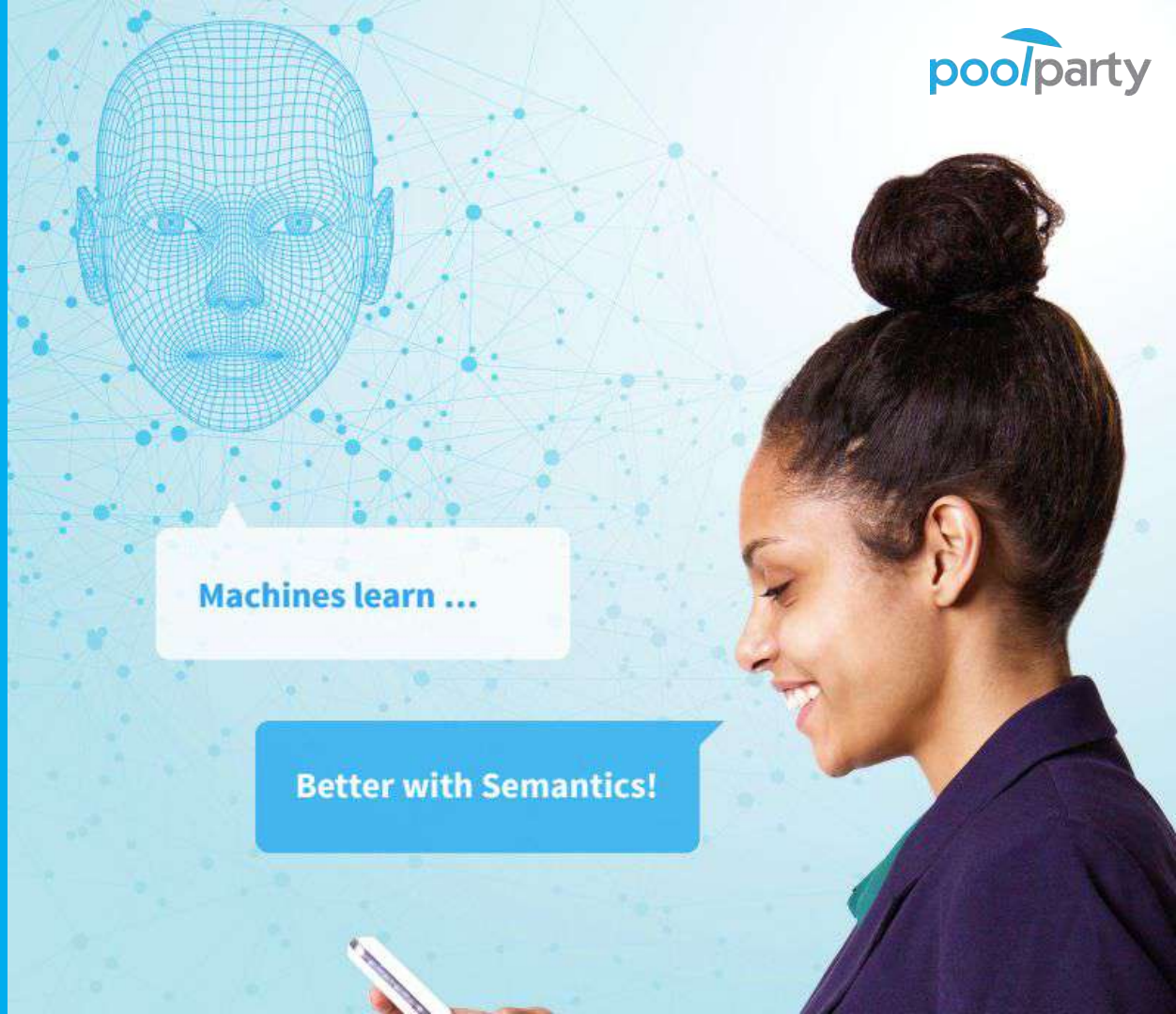
SEMANTICS
Vienna 2018

The Fast Track to Knowledge Engineering

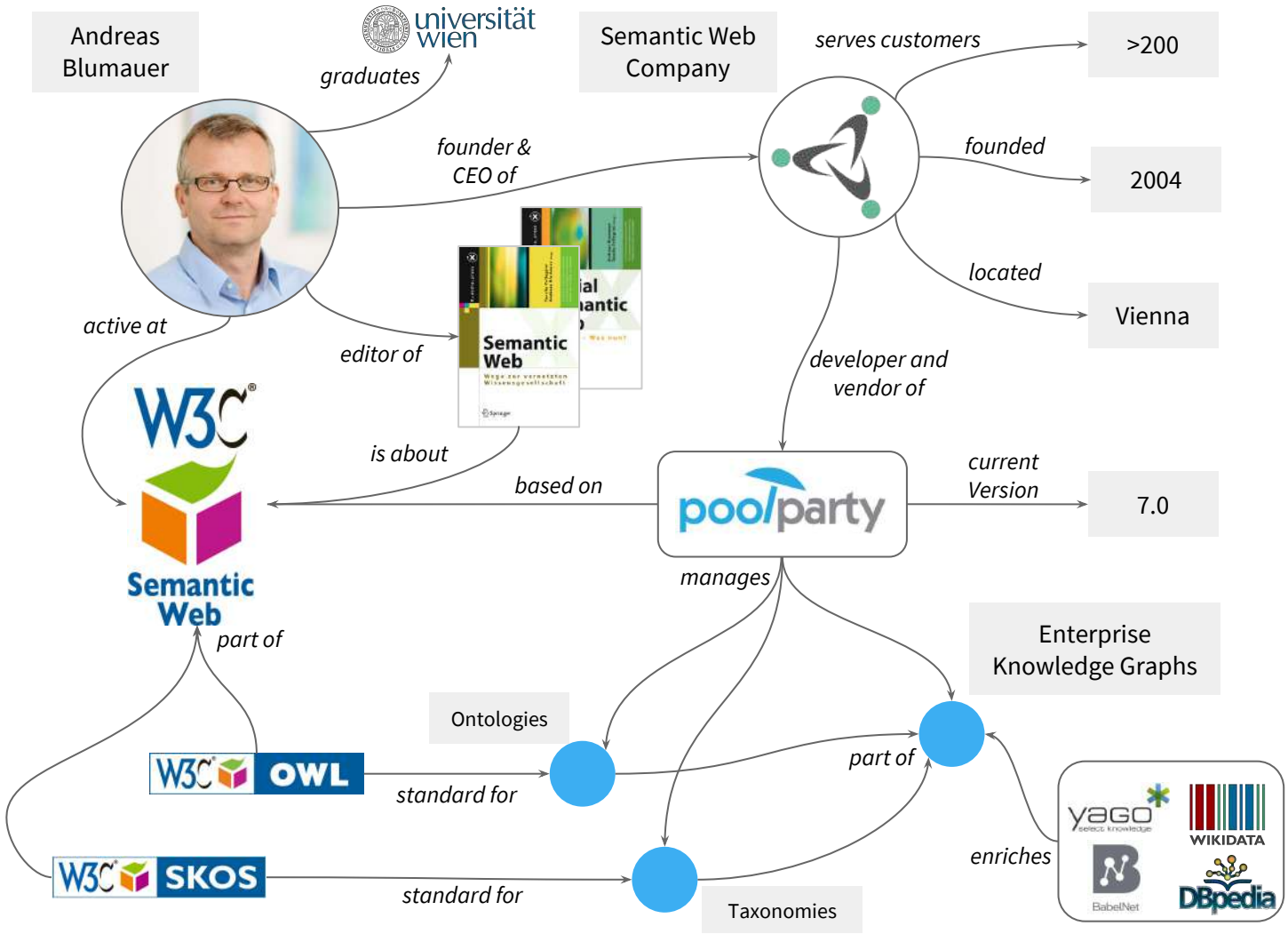
Andreas Blumauer
CEO & Managing Partner

Semantic Web Company /
PoolParty Semantic Suite



2

Introduction



3

PoolParty Academy

Three Training Programs

Technology oriented

Product related

- Initial Launch: **September 2016**
- E-learning tracks: **4** (incl. Partner Track)
- Learners enrolled: **>700**
- Certifications: **330**

4

PoolParty Academy

Get certified!

poolparty

Course

Semantic Technologies Training

Start

0% Complete

Course Outline

Overview of the course

- 1.1 How to use Semantic Technologies for business and technology challenges
- 1.2 Introduction to Knowledge Graphs
- 1.3 The value of consistent Metadata
- 1.4 When and how to use Linked Data
- 1.5 Semantic Linguistics
- 1.6 Semantic Data Integration
- 1.7 SPARQL: The query language of the Semantic Web
- 1.8 A corporate Semantic Web Architecture

About the Course

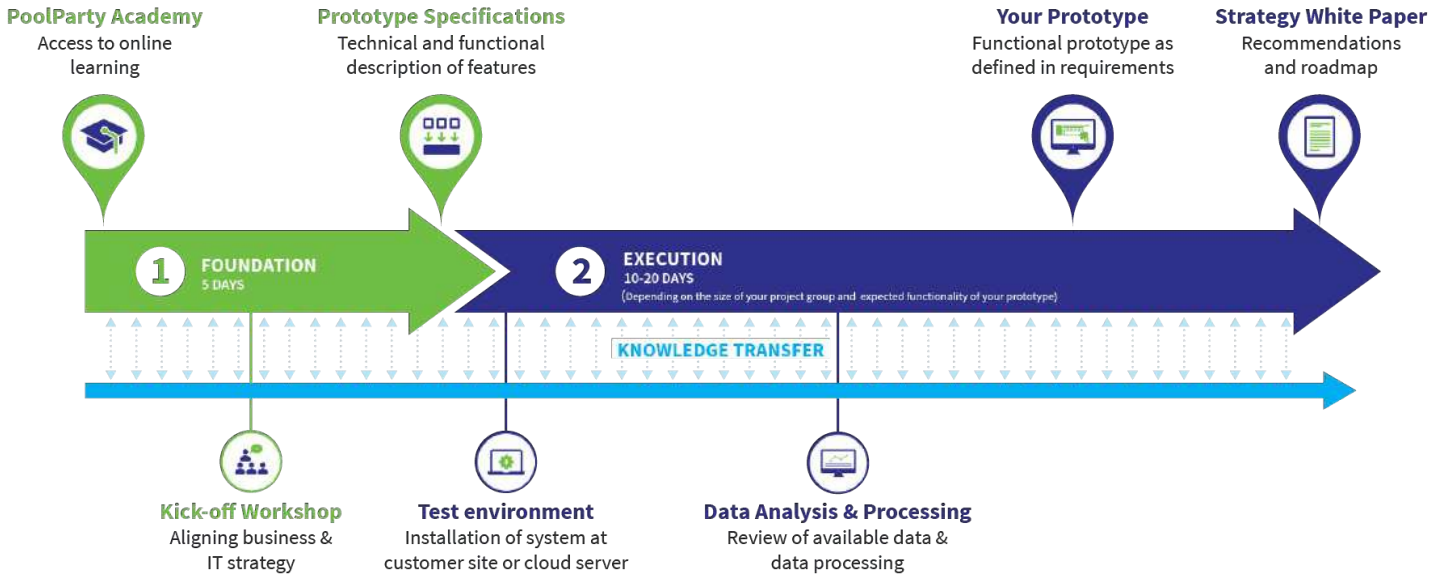
The Semantic Technologies Training provides you with an overview of the most relevant Semantic Web concepts and methodologies. You develop an understanding for which use cases this technology approach provides substantial advantages. You get acquainted with specialised terminology, which enables you to dive deeper into the semantic technologies field on your own.

After successful completion of the course, you become certified as Semantic Web Associate.

https://www.poolparty.biz/academy/

5

Semantic Web Starter Kit



6

GET STARTED

Get your test account at
www.poolparty.biz

7

poolparty

Semantics in a nutshell

Knowledge graphs
Linked Data Management
Text Analytics
Use Cases

Different shades of metadata

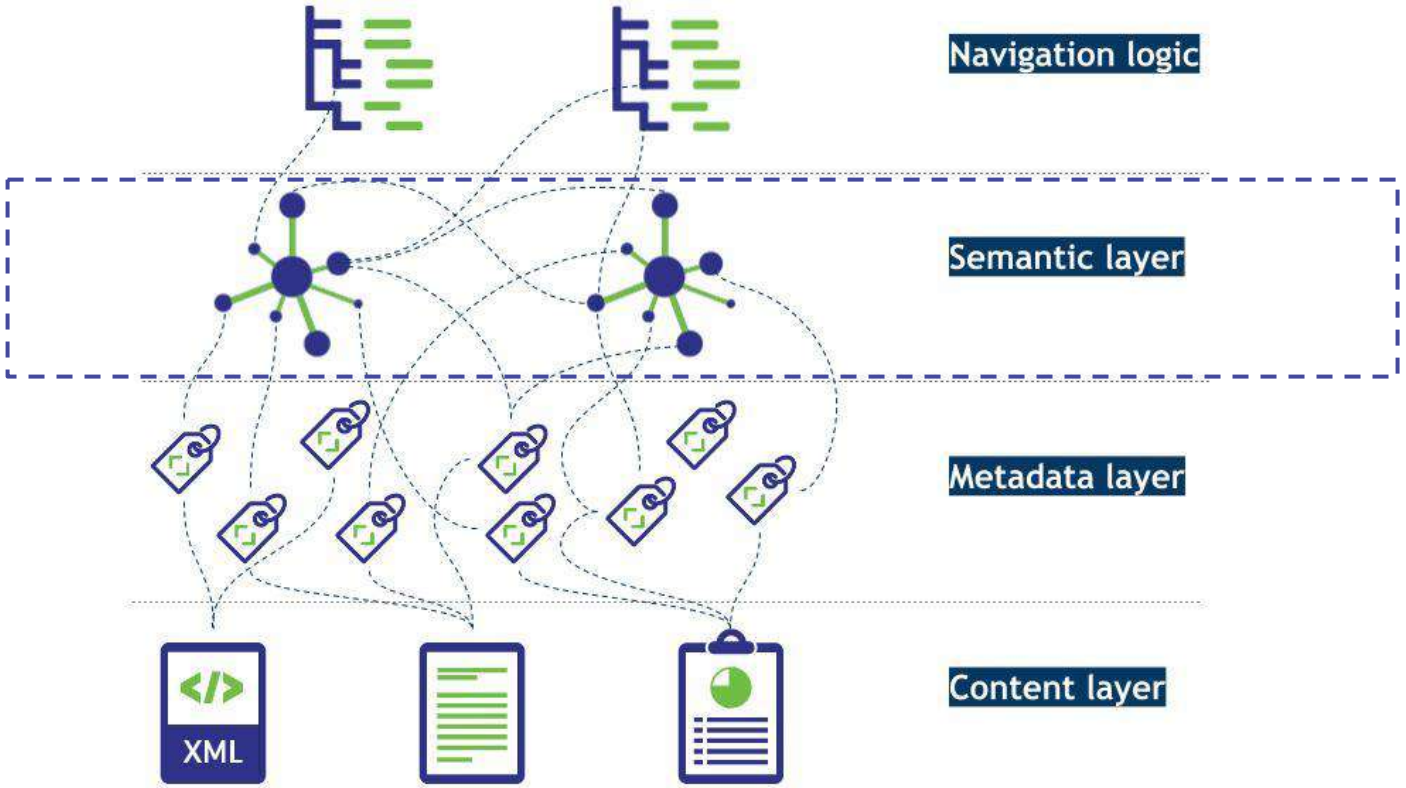
> [Read more!]

9

Core Principle

The Semantic Layer completes the Four-layered Data & Content Architecture

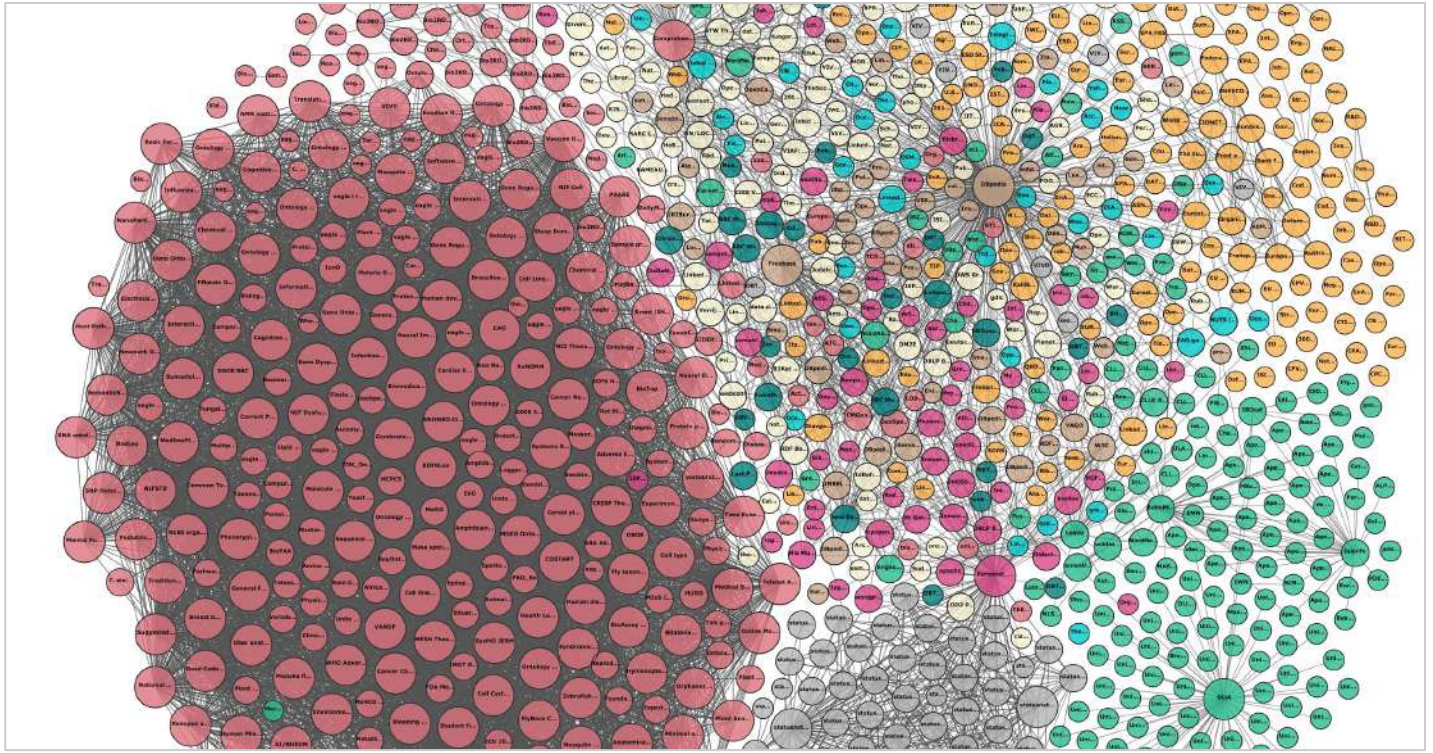
> Read more!



10

The Semantic Web

A standards-based graph of knowledge graphs

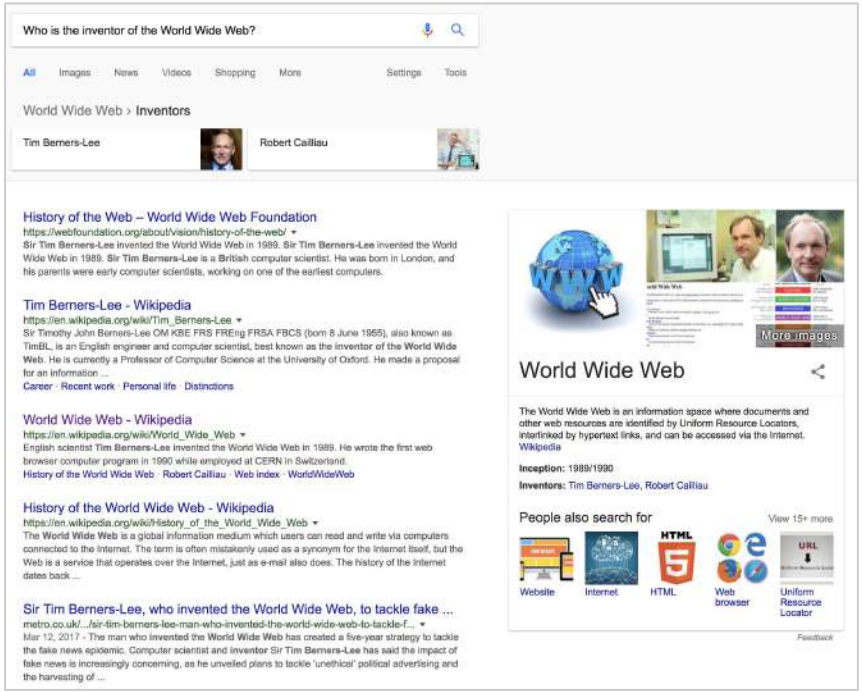


Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

11

Google Knowledge Graph

Who is the inventor of the World Wide Web?



Knowledge Graphs (KG) can cover **general knowledge** (often also called cross-domain or encyclopedic knowledge), or provide **knowledge about special domains** such as biomedicine.

In most cases KGs are based on Semantic Web standards, and have been generated by a mixture of **automatic extraction** from text or structured data, and **manual curation work**.

Examples:

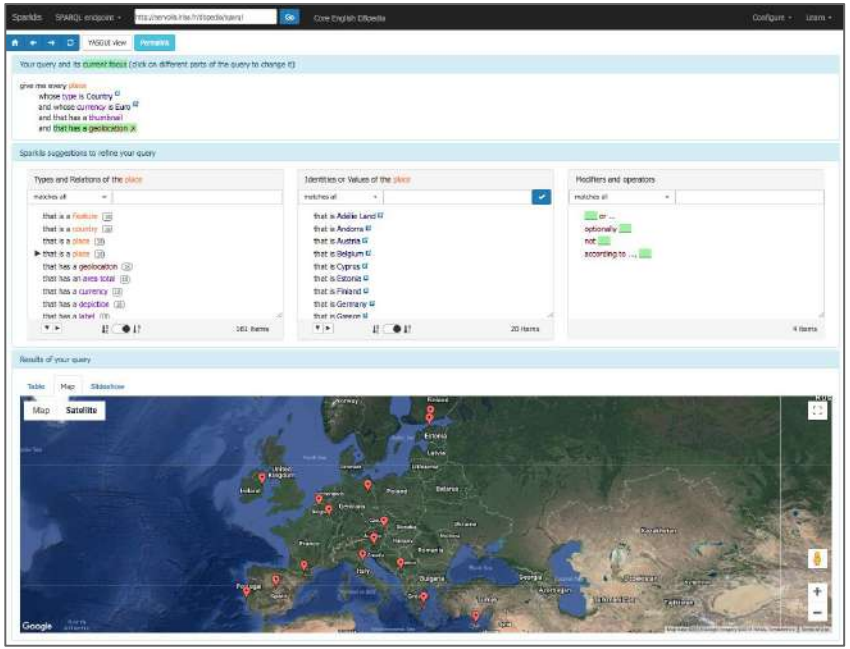
- DBpedia
- Google Knowledge Graph
- YAGO
- OpenCyc
- Wikidata

12

DBpedia, YAGO & Co.

Use weakly structured sources like Wikipedia as they were a structured data source

Show me all the countries using Euro on a map



[Link to Online Demo]

DBpedia is a frequently used Knowledge Graph covering **general knowledge** (often also called cross-domain or encyclopedic knowledge), also used as a central hub for entity linking to other graphs in the LOD cloud.

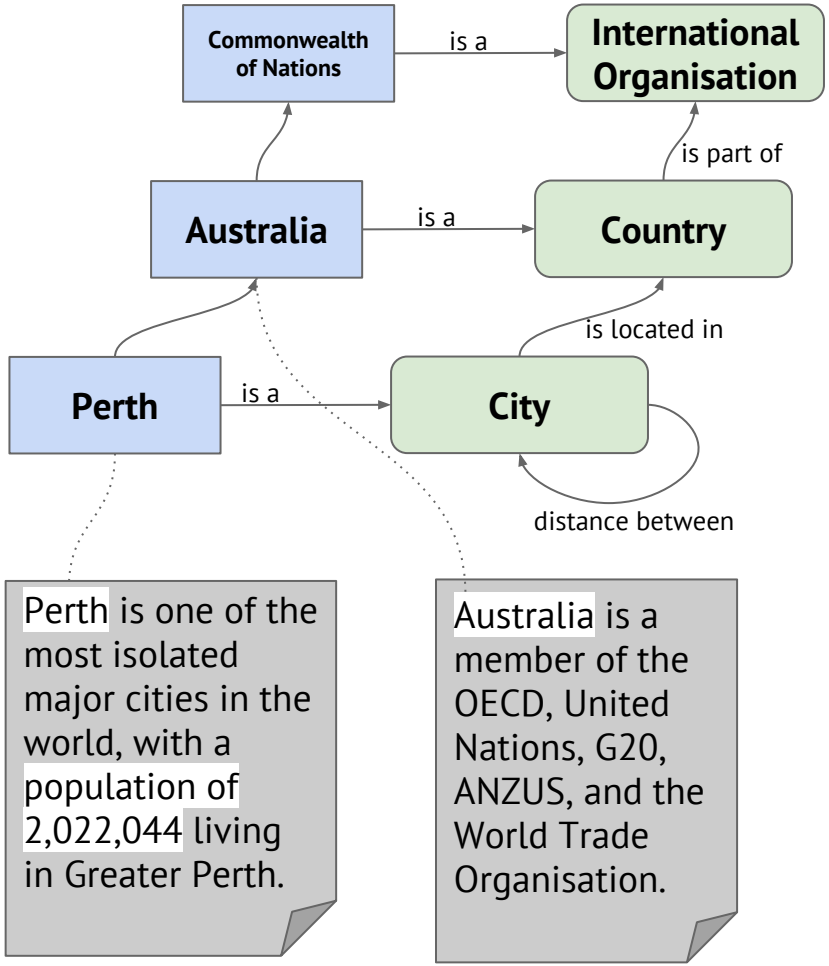
Like **YAGO**, it is harvested from **Wikipedia** and structured by its own ontology.

Just recently, new large-scale knowledge graphs like **Diffbot** or **Unigraph** were introduced.

13

Semantic Search & Assistants

Based on Knowledge Graphs & Knowledge Extraction



Support complex Q&A:

Which cities located in the Commonwealth of Nations have a population of more than 2 mio. people?

Avoid illogical answers:

poolparty

Natural languages

- Ambiguity versus Universality
- Context information and background knowledge needed

Susan observes Mike on a tower with a telescope.

15

Context is King

Natural languages

- Ambiguity versus Universality
- Context information and background knowledge needed

Susan observes Mike on a tower with a telescope.

- Susan and Mike are persons.
- Yesterday Michael bought a Celestron.
- If one buys something, (s)he owns it and can use it.
- Mike and Michael is the same person.
- A Celestron is a telescope.

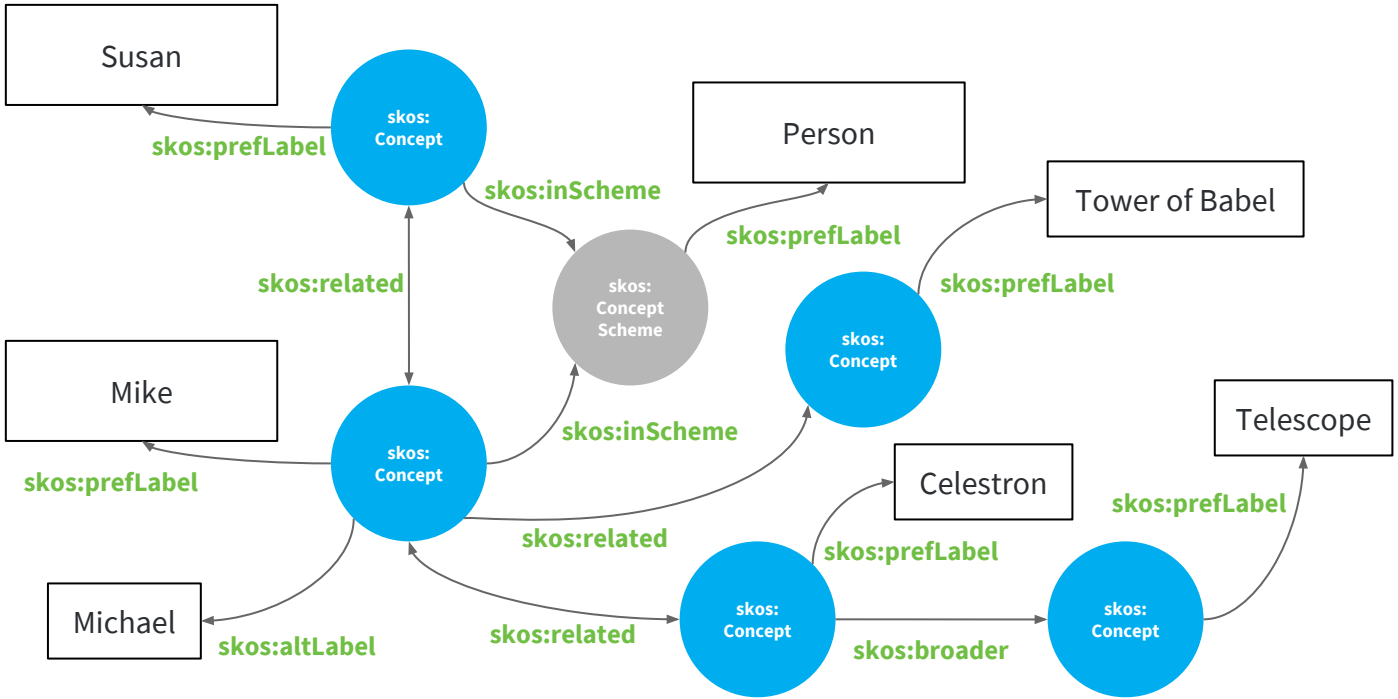
16

Taxonomies

Introducing some explicit semantics

Taxonomies

- SKOS taxonomies are concept- and resource-based knowledge models
- SKOS stands for Simple Knowledge Organization System and is part of the Semantic Web Standard



17

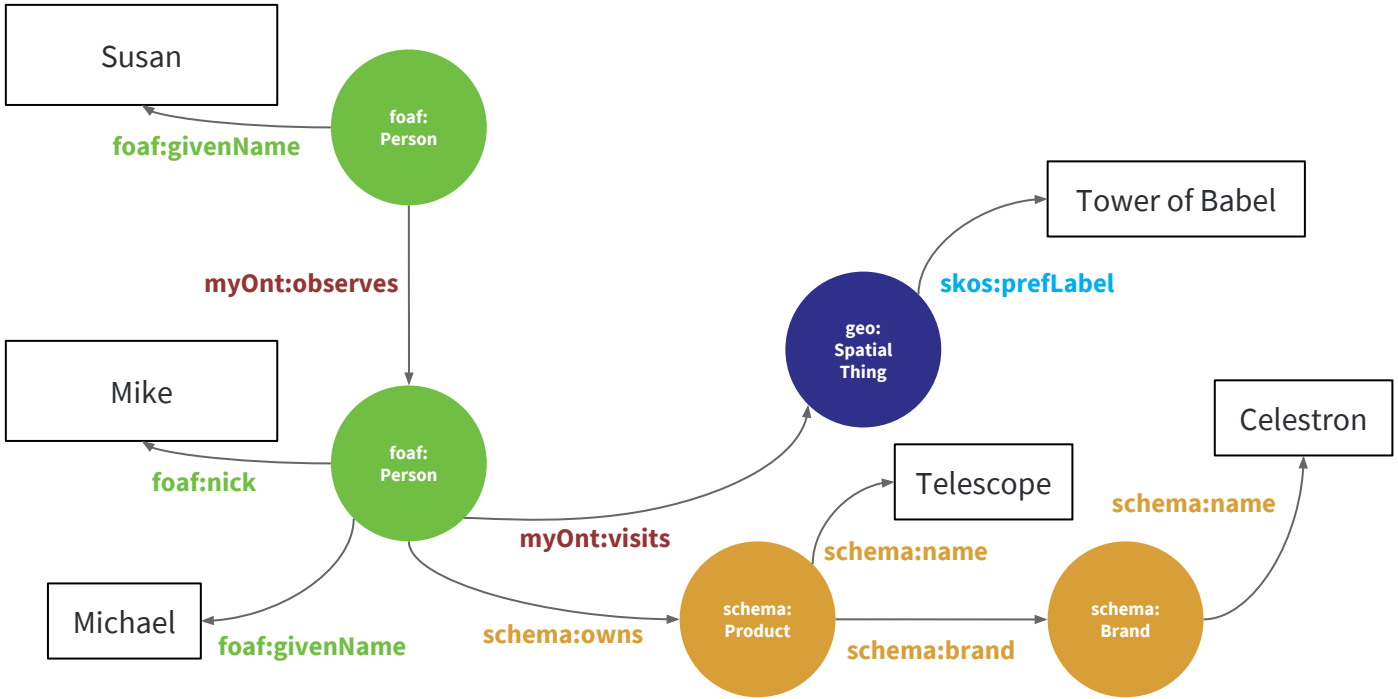
Ontologies

Some more explicit semantics

> Read more!

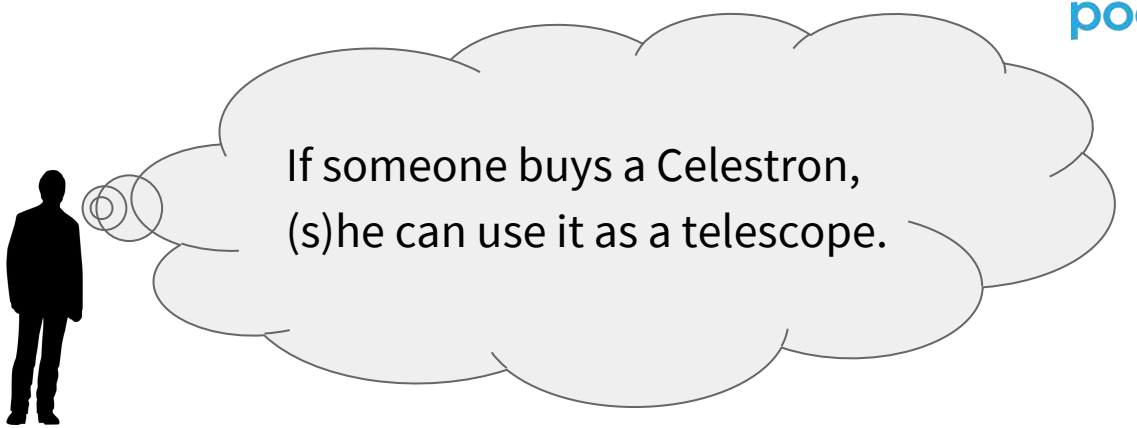
Ontologies

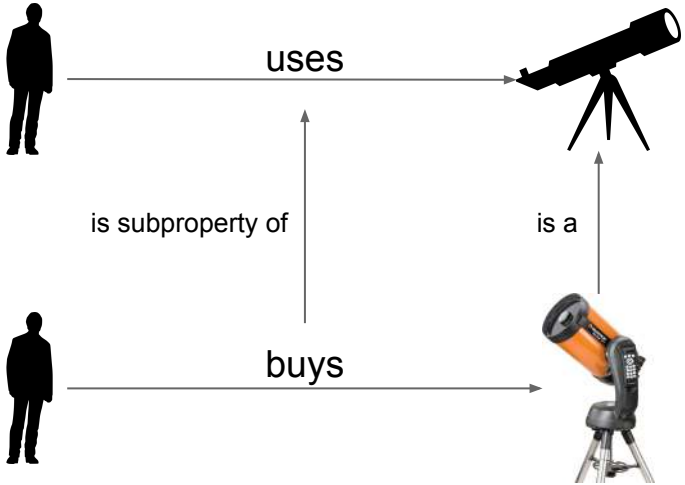
- Ontologies classify things and define more specific relations and attributes
- Locally and globally recognised ontologies can be combined
- Ontologies can have various levels of expressivity (RDFS, OWL)



18

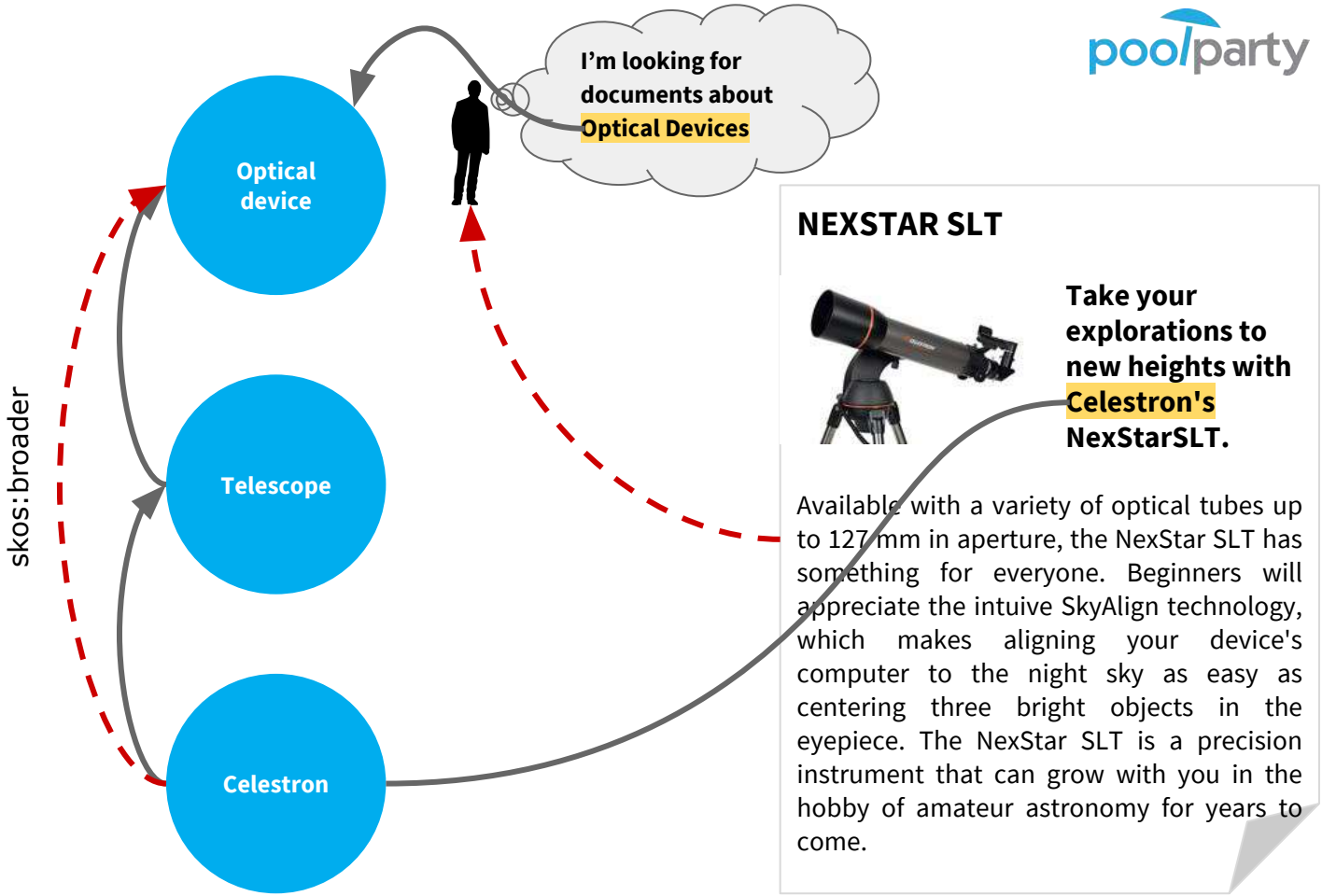
Reasoning





19

Reasoning over taxonomies

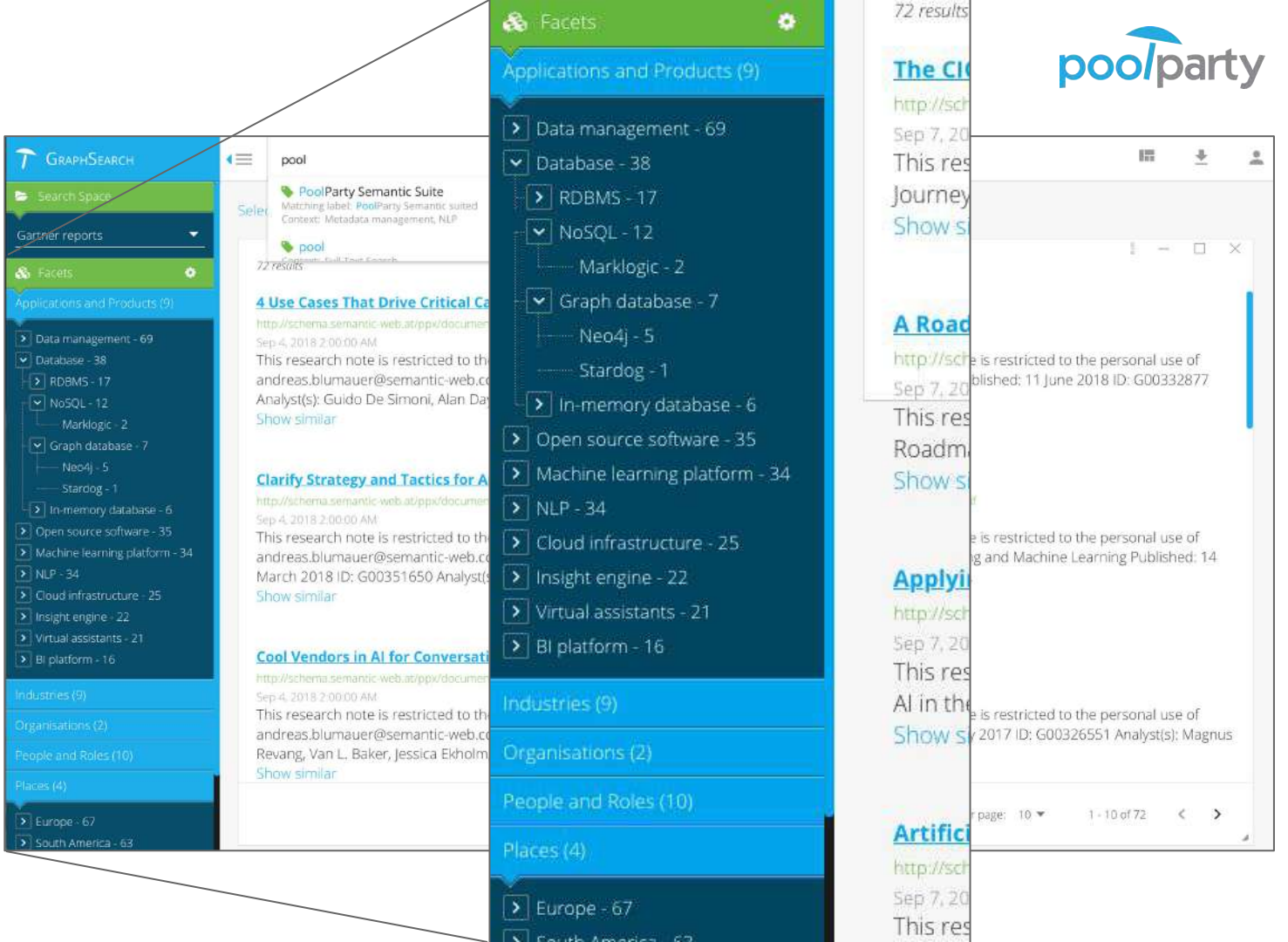


skos:broader *is a* owl:TransitiveProperty

20

Use Case

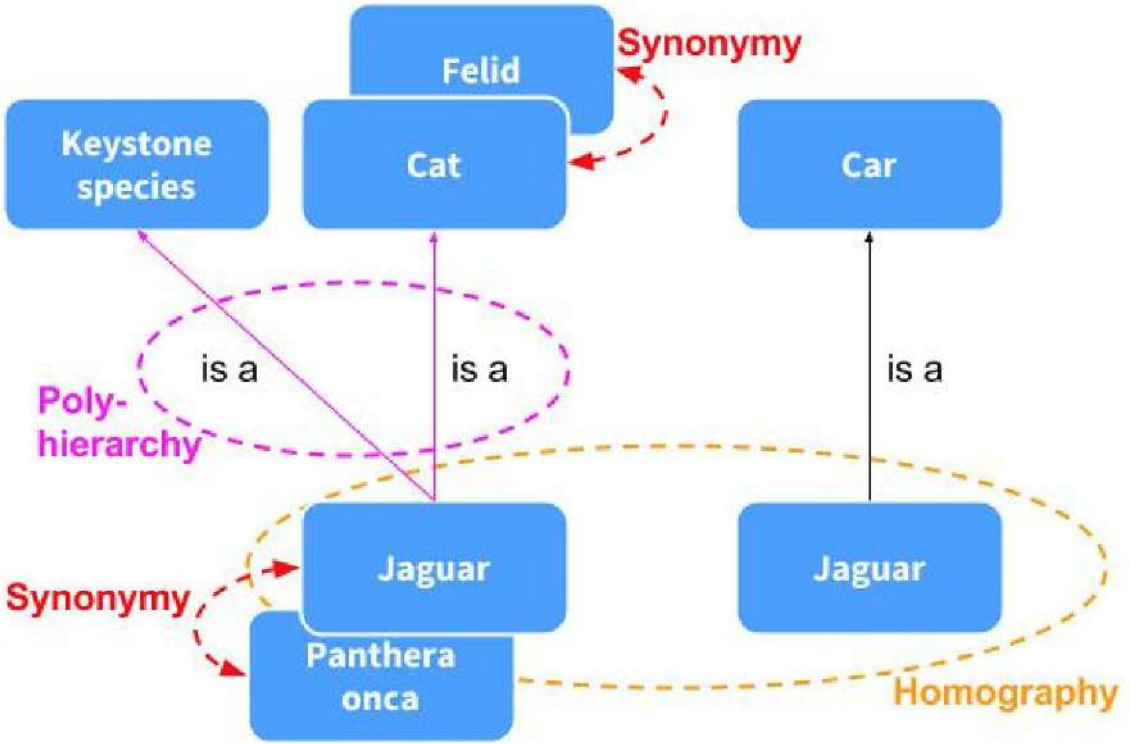
Faceted Search based on Taxonomies



21

Resolving Language Problems

> Read more!



"While most people can deal with linguistic features as synonyms, homographs, polyhierarchies, and even with far more peculiar characteristics of natural languages, **machines often struggle with automatic sense-making** because of the lack of a semantic knowledge model that can be used programmatically."

‘Things’ but not Strings:
Using a ‘Semantic Knowledge Graph’

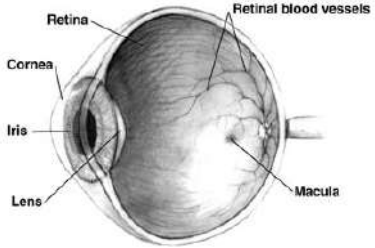
Retina

prefLabel

http://www.my.com/
taxonomy/62346723

image

http://www.my.com/
images/90546089



Home → Medical Encyclopedia → Retina

Retina

The retina is the light-sensitive layer of tissue at the back of the eyeball. Images that come through the eye's lens are focused on the retina. The retina then converts these images to electric signals and sends them along the optic nerve to the brain.

The retina usually looks red or orange because there are many blood vessels right behind it. An ophthalmoscope allows a health care provider to see through your pupil and lens to the retina. Sometimes photos or special scans of the retina can show things that the provider cannot see just by looking at the retina through the ophthalmoscope. If other eye problems block the provider's view of the retina, ultrasound can be used.

Anyone who experiences these vision problems should get a retinal examination:

- Changes in sharpness of vision
- Loss of color perception
- Flashes of light or floaters
- Distorted vision (straight lines look wavy)

Watch this video about:
Retina

http://www.my.com/
taxonomy/
97345854

Funduscope

prefLabel

Ophthalmoscope

altLabel

has broader

http://www.mycom.com
/taxonomy/4543567

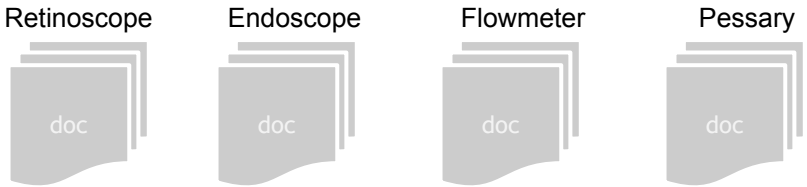
prefLabel

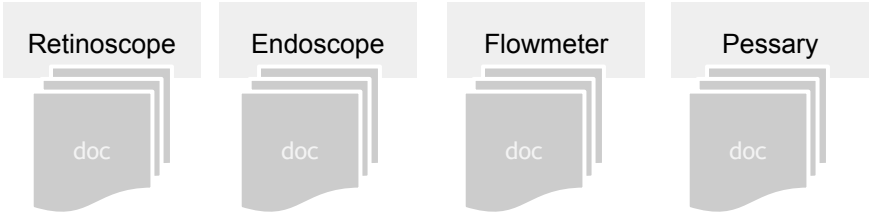
Diagnostic Equipment

The power of knowledge graphs: Agility, extensibility, precision

Traditional approach

Graph-based approach





The power of knowledge graphs: Agility, extensibility, precision

Traditional approach

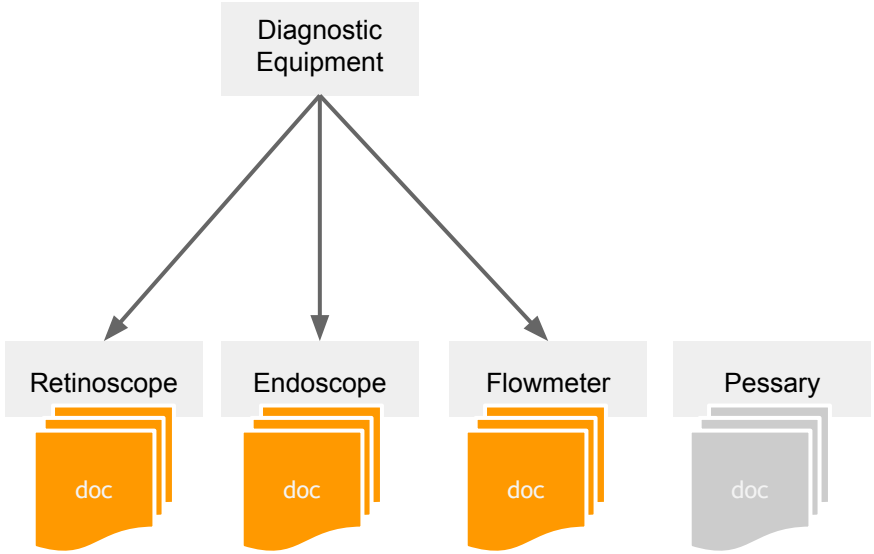
Diagnostic Equipment, Retinoscope
doc

Diagnostic Equipment, Endoscope
doc

Diagnostic Equipment, Flowmeter
doc

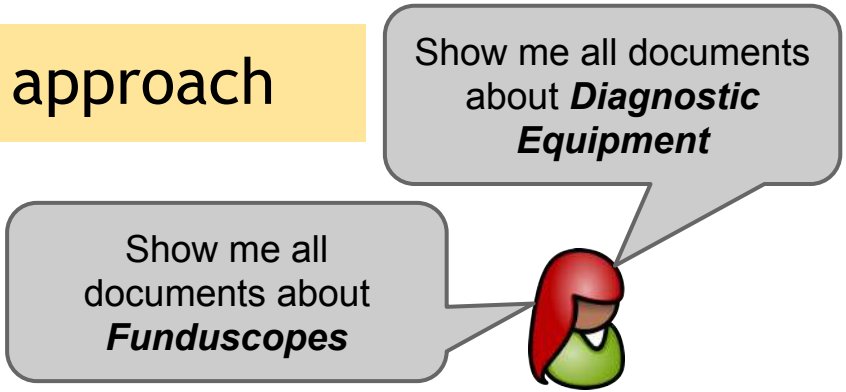
Surgical Equipment, Pessary
doc

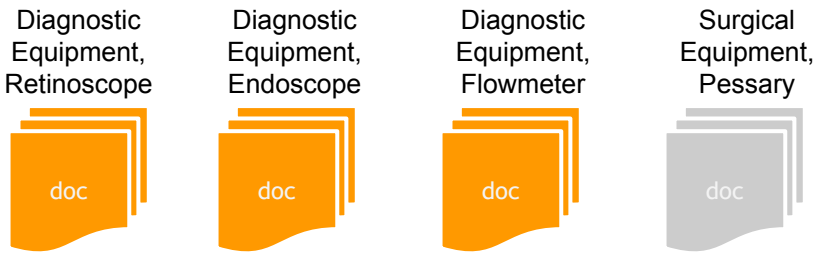
Graph-based approach



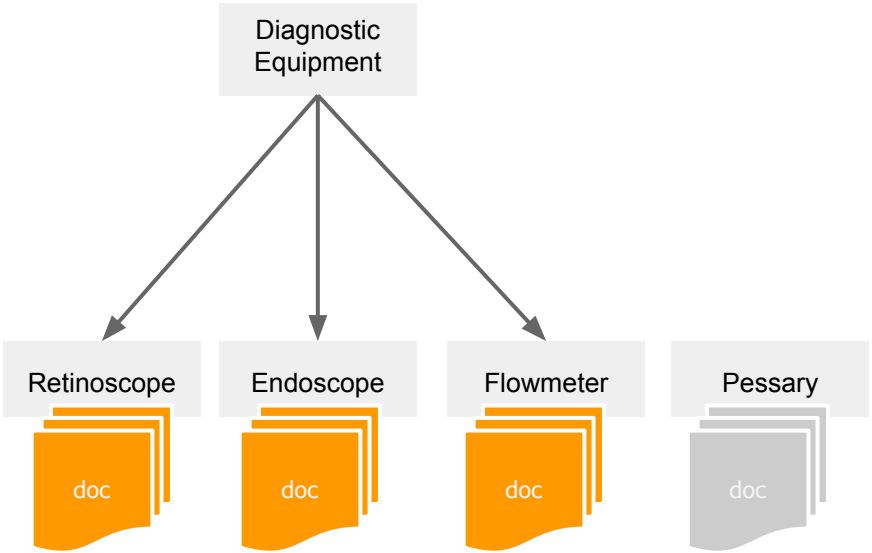
The power of knowledge graphs: Agility, extensibility, precision

Traditional approach



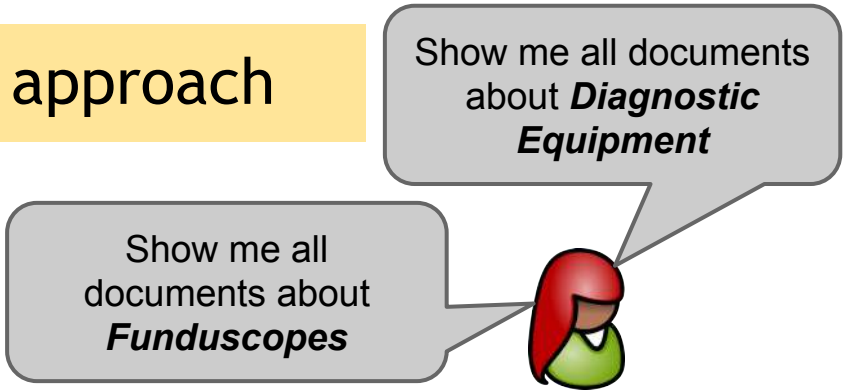


Graph-based approach



The power of knowledge graphs: Agility, extensibility, precision

Traditional approach



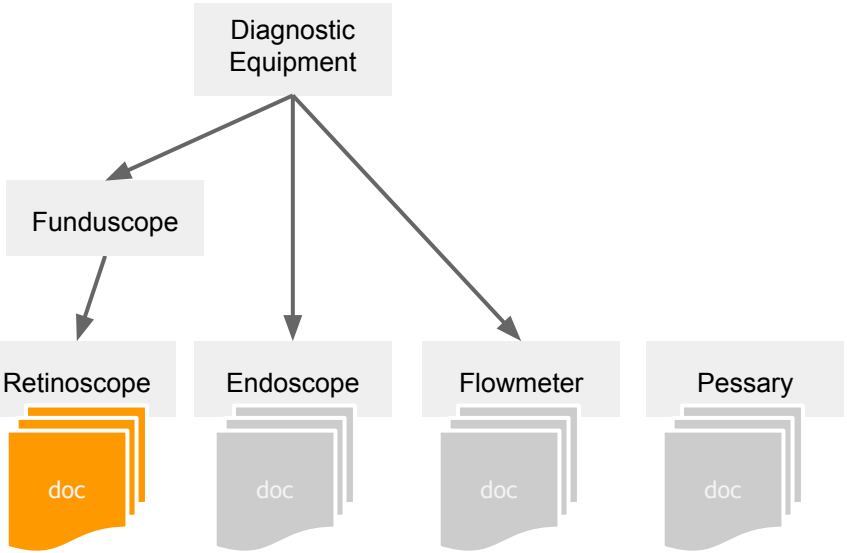
Funduscope, Diagnostic Equipment, Retinoscope

Diagnostic Equipment, Endoscope

Diagnostic Equipment, Flowmeter

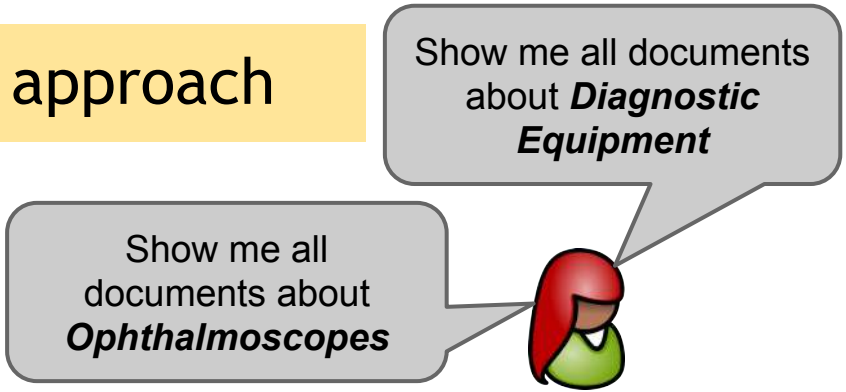
Surgical Equipment, Pessary

Graph-based approach



The power of knowledge graphs: Agility, extensibility, precision

Traditional approach



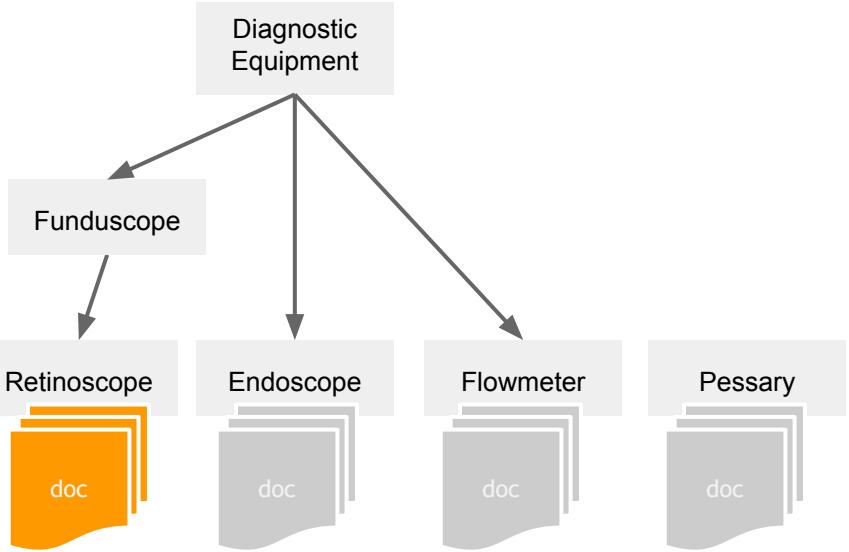
Ophthalmoscope, Funduscope, Diagnostic Equipment, Retinoscope

Diagnostic Equipment, Endoscope

Diagnostic Equipment, Flowmeter

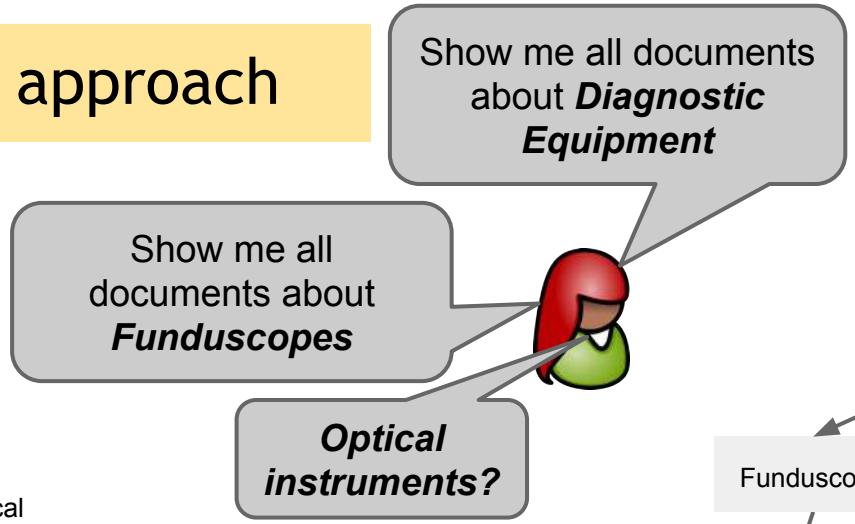
Surgical Equipment, Pessary

Graph-based approach



The power of knowledge graphs: Agility, extensibility, precision

Traditional approach



Optical Instruments, Ophthalmoscope Funduscope, Diagnostic Equipment, Retinoscope

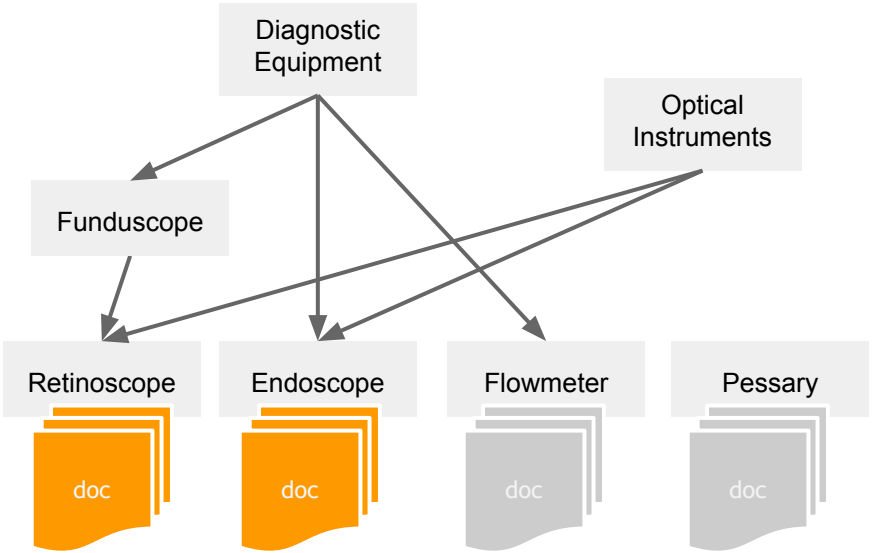
Optical Instruments, Diagnostic Equipment, Endoscope

Diagnostic Equipment, Flowmeter

Surgical Equipment, Pessary

doc doc doc doc

Graph-based approach

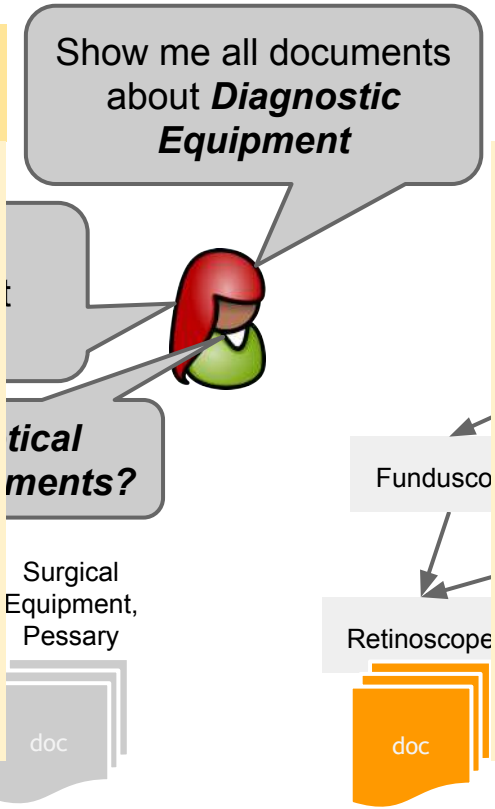


The power of knowledge graphs: Agility, extensibility, precision

Traditional approach

Metadata per document

1. No or little network effects
2. No reuse of metadata
3. Metadata resides in silos
4. Data quality hard to measure
5. Not machine-readable



Graph-based approach

Knowledge about metadata

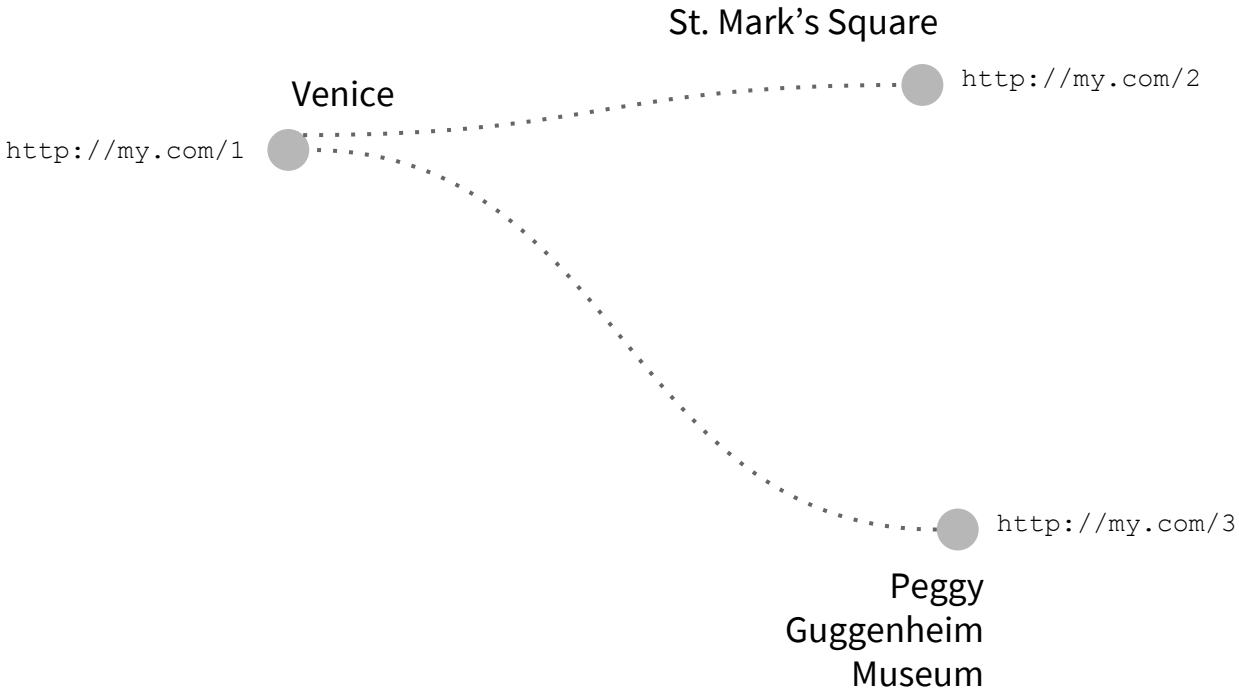
1. Explicit knowledge models
2. Reusable and measurable
3. Metadata is machine-processable
4. Standards-based metadata
5. Linkable metadata opens silos

30

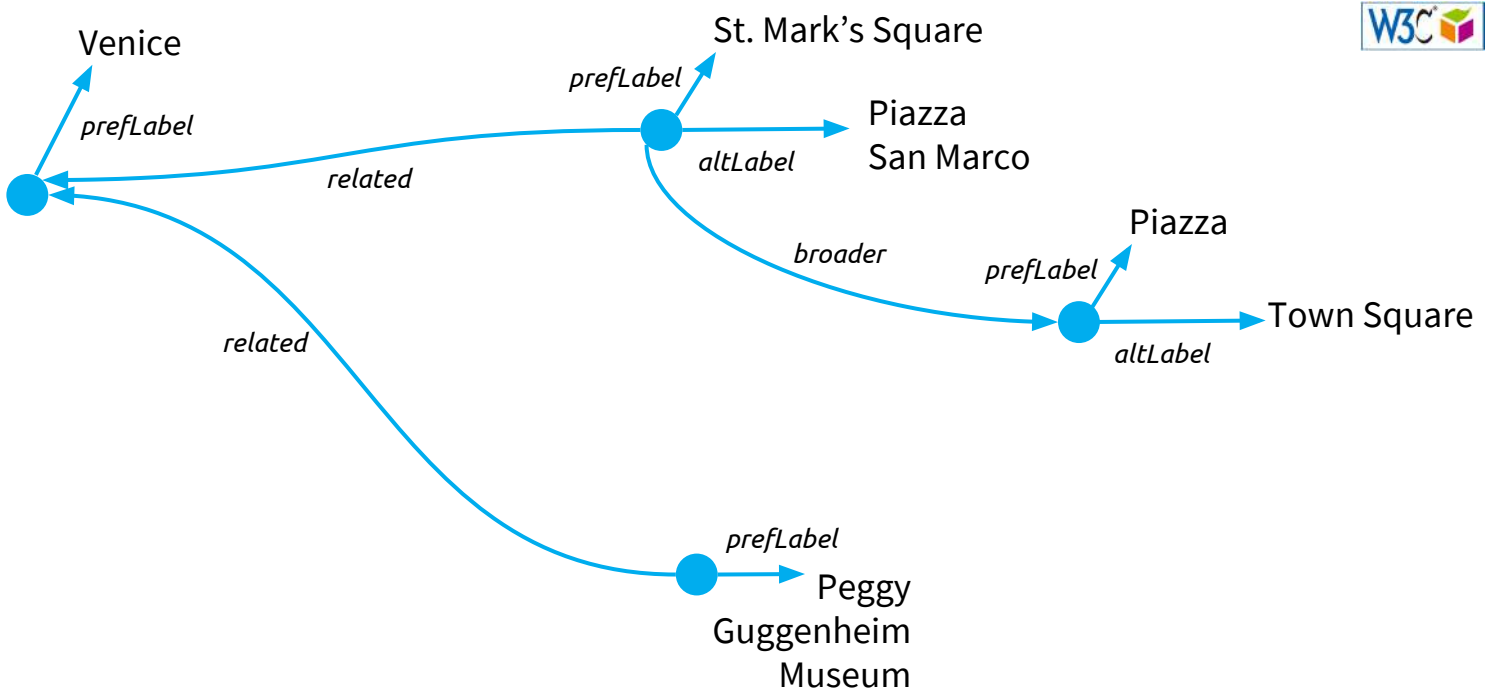
How to build a Knowledge Graph?

Anatomy of an Enterprise Knowledge Graph

Things and URIs



Labels and basic relations: Taxonomies and Thesauri



> **Read more!**

Classes, specific relations, restrictions: Ontologies and Custom Schemas

Venice

prefLabel

containedInPlace

St. Mark's Square

prefLabel

altLabel

Piazza San Marco

image

broader

Piazza

prefLabel

Town Square

altLabel

containedInPlace

Monday through Sunday, all day

opening Hours

prefLabel

Peggy Guggenheim Museum

schema.org

- http://schema.org/City
- http://schema.org/TouristAttraction
- http://schema.org/ArtGallery
- http://schema.org/containedInPlace

W3C SKOS

W3C OWL2

W3C RDF

Metadata and Graph annotations

creative commons

CC BY-SA 3.0

Venice

prefLabel

containedInPlace

St. Mark's Square

prefLabel

altLabel

Piazza San Marco

W3C SKOS

W3C OWL2

W3C RDF

image

broader

Piazza

prefLabel

Town Square

altLabel

containedInPlace

Monday through Sunday, all day

opening Hours

prefLabel

Peggy Guggenheim Museum

schema.org

http://schema.org/City

http://schema.org/TouristAttraction

http://schema.org/ArtGallery

http://schema.org/containedInPlace

Entity linking and schema mappings: Links to other graphs

creative commons

CC BY-SA 3.0

Venice

prefLabel

image

containedInPlace

St. Mark's Square

prefLabel

altLabel

Piazza San Marco

broader

Piazza

prefLabel

Town Square

altLabel

containedInPlace

Monday through Sunday, all day

opening Hours

prefLabel

Peggy Guggenheim Museum

W3C SKOS

W3C OWL2

W3C RDF

schema.org

http://schema.org/City

http://schema.org/TouristAttraction

http://schema.org/ArtGallery

http://schema.org/containedInPlace

yago

WIKIDATA

BabelNet

DBpedia

Linking to data and documents stored in other systems

creative commons

CC BY-SA 3.0

Venice

prefLabel

containedInPlace

St. Mark's Square

prefLabel

altLabel

Piazza San Marco

W3C SKOS

W3C OWL2

W3C RDF

image

broader

Piazza

prefLabel

Town Square

altLabel

containedInPlace

Monday through Sunday, all day

opening Hours

prefLabel

Peggy Guggenheim Museum

The Peggy Guggenheim Collection is a modern art museum on the Grand Canal in the Dorsoduro sestiere of Venice, Italy.

schema.org

http://schema.org/City

http://schema.org/TouristAttraction

http://schema.org/ArtGallery

http://schema.org/containedInPlace

yago

WIKIDATA

BabelNet

DBpedia

37

Linked Data Life Cycle

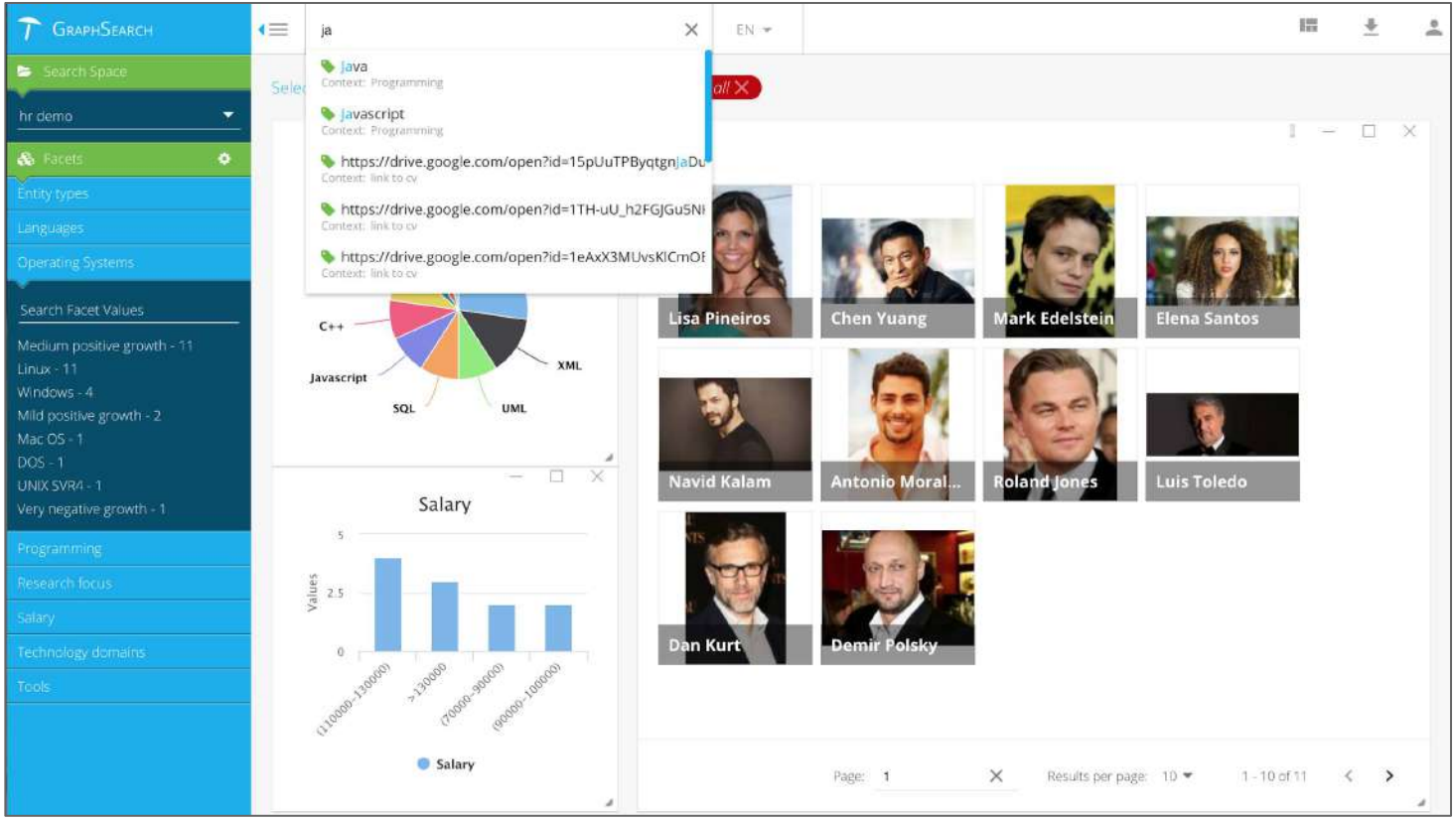
How to build Enterprise Knowledge Graphs?

> **Read more!**

38

Use Case: Integrated Views on Business Data

> [Try it out!]

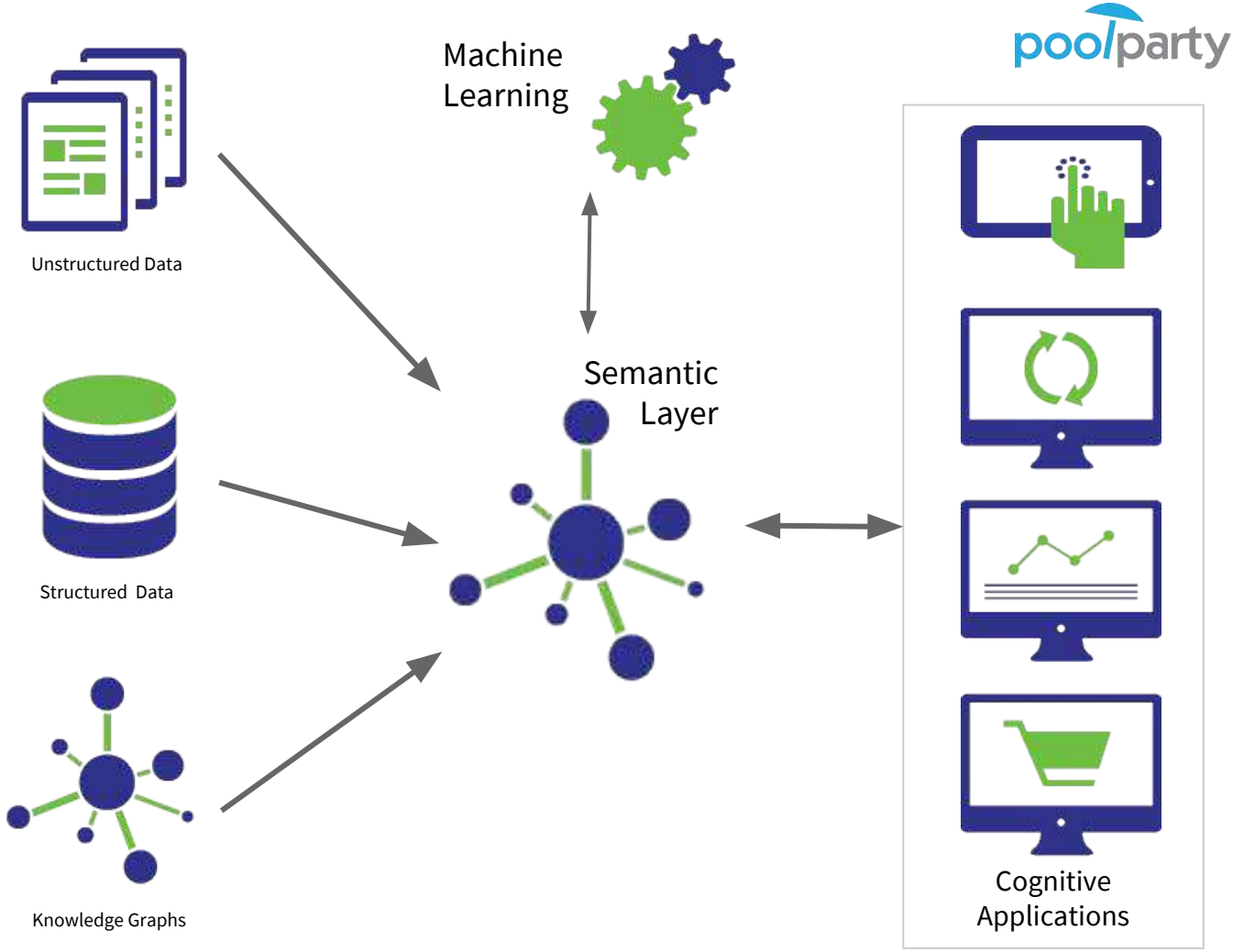


Learn more about it: [Structured and Unstructured Data Linked]

39

Knowledge Graphs as input for Machine Learning

> [Read more!]

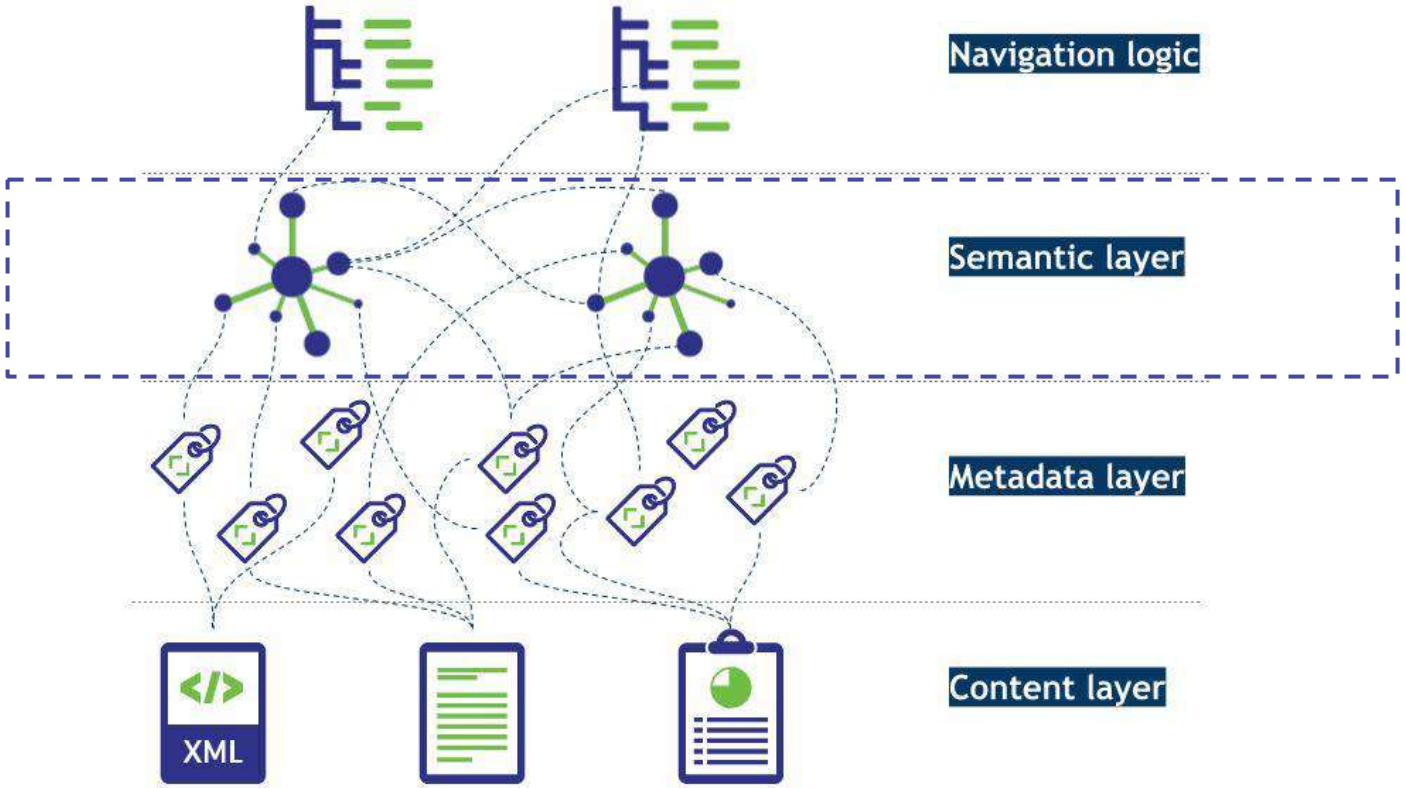


40

Core Principle

The Semantic Layer completes the Four-layered Data & Content Architecture

> Read more!



PoolParty Hands-on Session

Try it out!

42

Fact sheet: PoolParty

PoolParty Semantic Suite

- Most complete Semantic Middleware on the Global Market
- *Semantic AI*: Fusion of Knowledge Graphs, NLP, and Machine Learning
- Linked Data Management along the whole Data Life Cycle
- W3C standards compliant

- First release in 2009
- Current version 7.0
- Over 200 installations world-wide

- On-premises or cloud-based
- KMWorld listed PoolParty as Trend-Setting Product 2015, 2016 and 2017
- www.poolparty.biz

PoolParty supports the management of semantic knowledge graphs and linked data governance along the entire data lifecycle.

KMWorld Trend-Setting PRODUCT OF 2017

43

Solutions based on PoolParty

poolparty

Process Automation

- Enhanced Machine Learning
- Text Mining & NLP
- Document Classification

Enhanced Customer Experience

- Recommender Systems
- SEO
- Smart Helpdesk Solutions
- Chatbots and Q&A engines

Knowledge Management

- Semantic Search
- Personalization
- Knowledge Discovery Portals

Information Management

- Semantic Content Management
- Metadata Management
- Masterdata Management

Knowledge Engineering

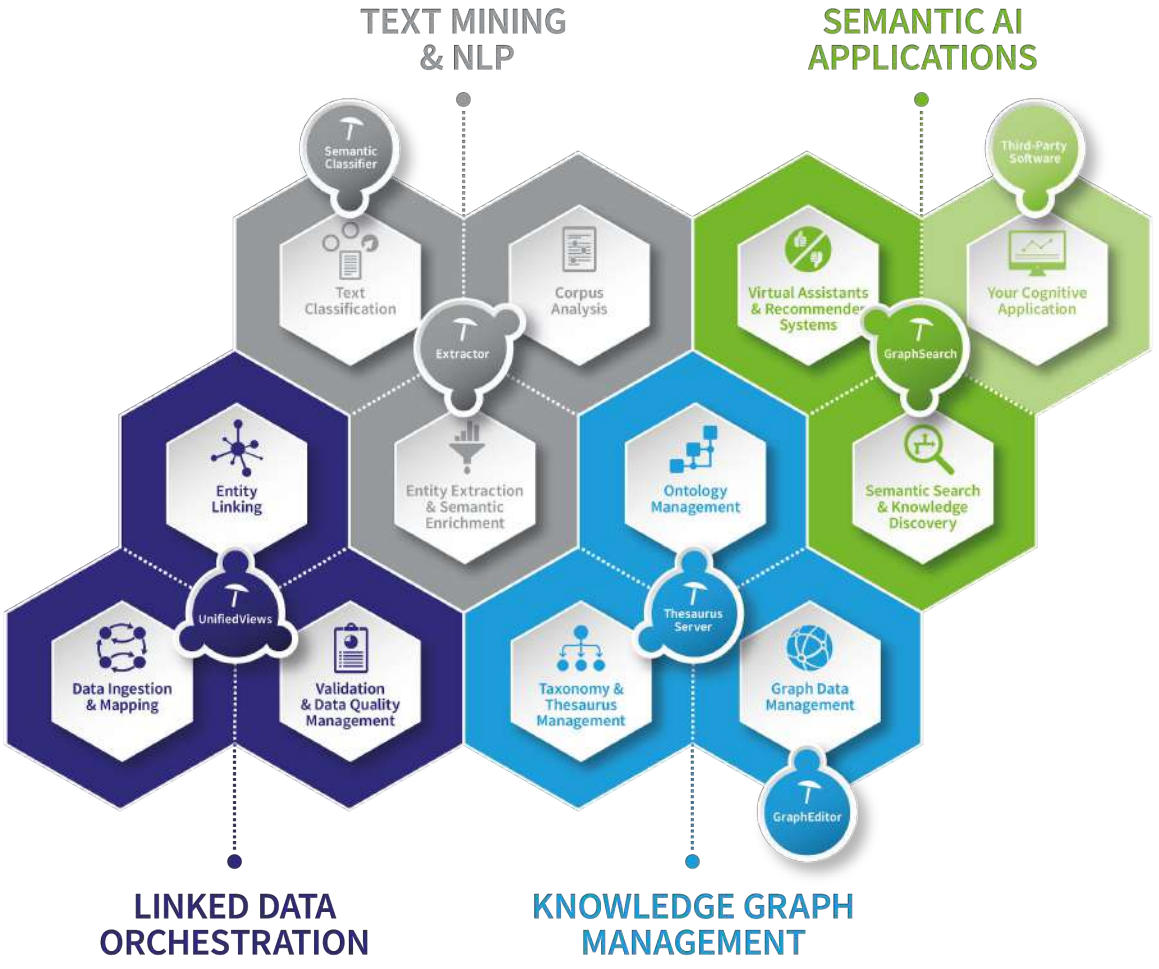
- Taxonomy Management
- Ontology Management
- Knowledge Graph Management
- Data Visualization

Agile Data Integration

- Linked Data
- Integrating heterogeneous data
- Entity Linking

44

Functions and Components



45

Selected Customer References and Partners

We work with Global Fortune Companies, and with some of the largest GOs and NGOs from over 20 countries.

Selected Customer References

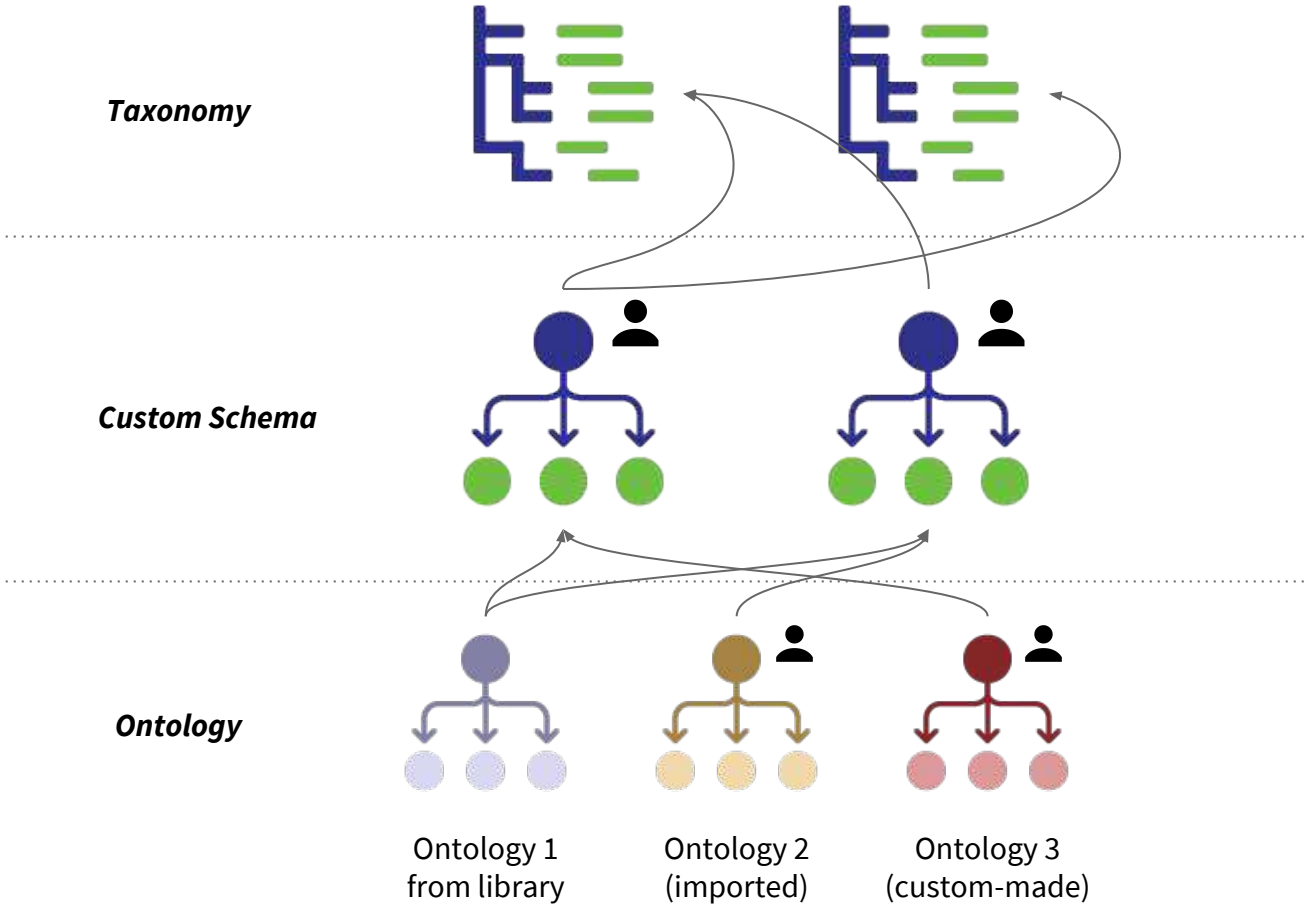
- Credit Suisse
- Boehringer Ingelheim
- Roche
- adidas
- The Pokémon Company
- Fluor
- Harvard Business School
- Wolters Kluwer
- Philips
- Nestlé
- Electronic Arts
- Springer Nature
- Pearson - Always Learning
- Healthdirect Australia
- World Bank Group
- Canadian Broadcasting Corporation
- Oxford University Press
- International Atomic Energy Agency
- Siemens
- Singapore Academy of Law
- Inter-American Development Bank
- Council of the E.U.
- AT&T

Selected Partners

- Enterprise Knowledge
- Mekon Intelligent Content Solutions
- Soitron
- Accenture
- EPAM Systems
- BAON Enterprises
- Findwise
- Tellura Semantics
- HPC
- Minerva Intelligence
- Make it a Triple

46

How PoolParty's ontology and custom schema management plays together with taxonomies



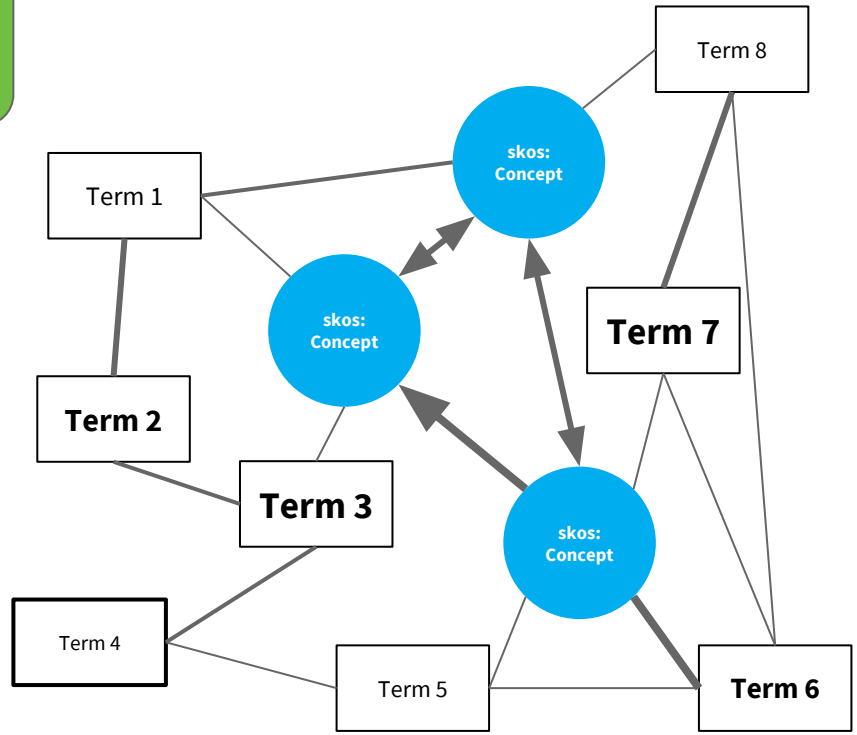
47

Corpus analysis results in a network of concepts and terms

I need support to continuously extend our taxonomy / controlled vocabulary!

Reference Corpus

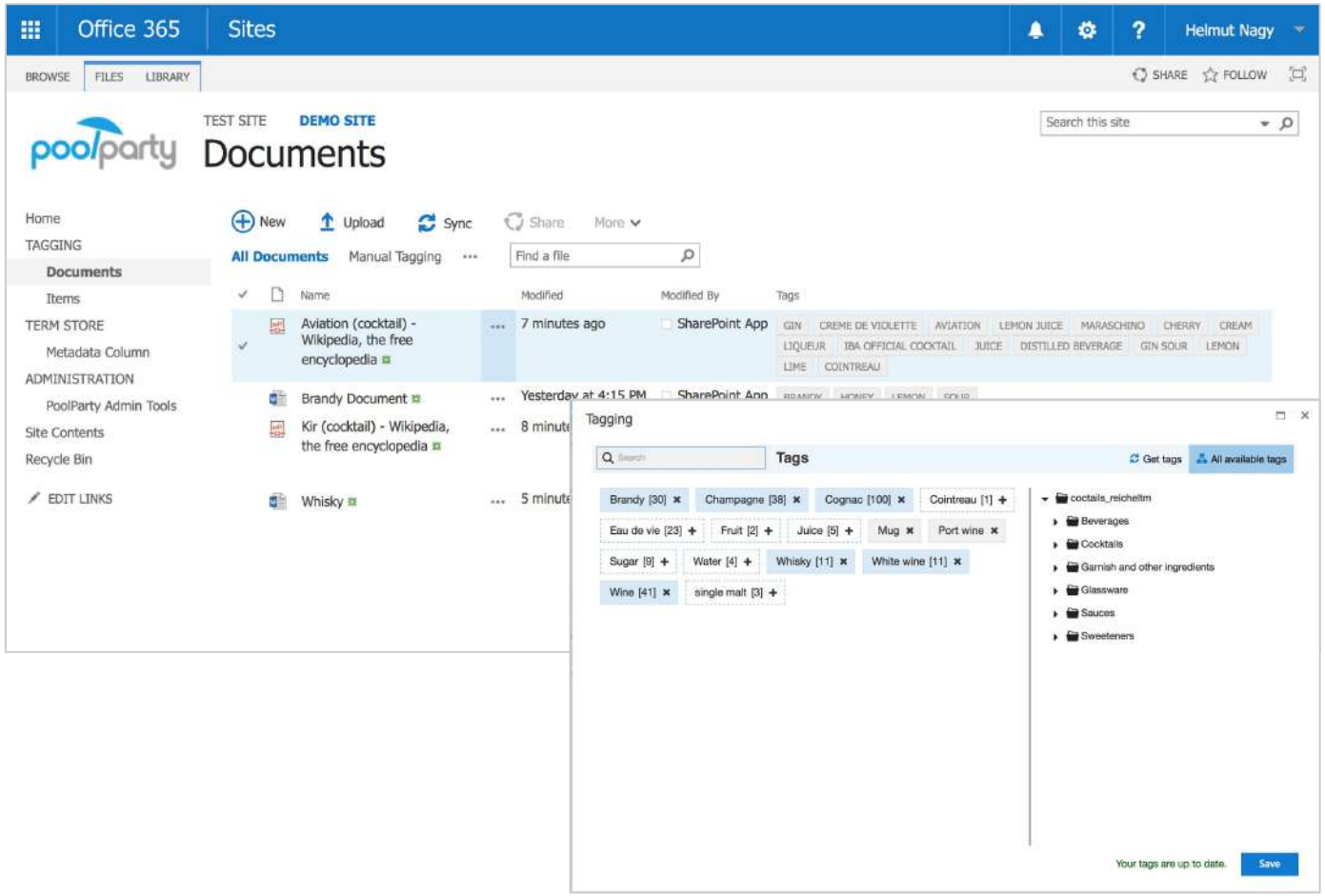
- Websites
- PDF, Word, …
- Abstracts from DBpedia
- RSS Feeds



- Relevant terms and phrases
- Relevancy of concepts
- co-occurence between concepts and terms
- co-occurence between terms and terms

48

Autotagging & Consistent Tagging based on controlled vocabularies

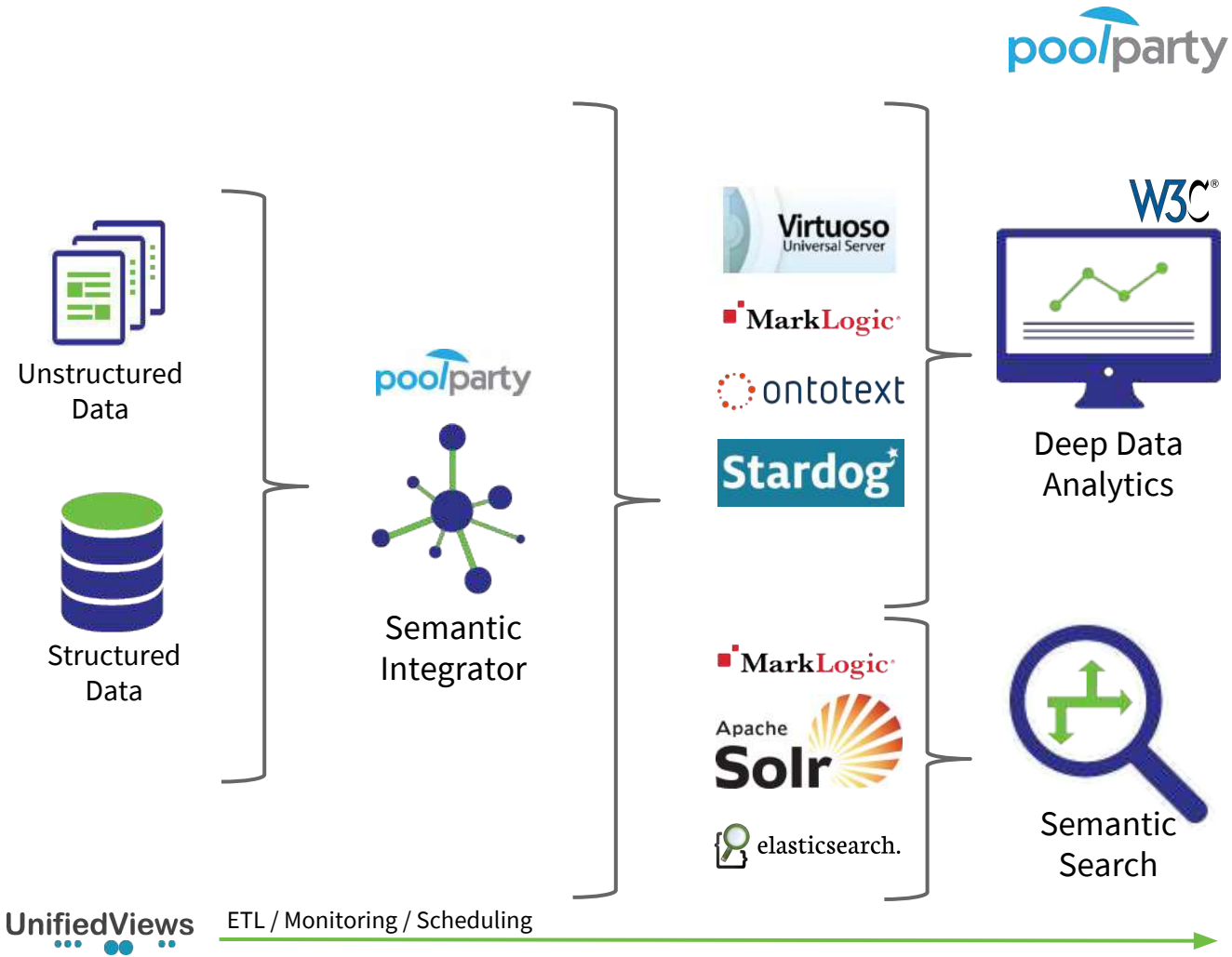


49

PoolParty Semantic Integrator - at a glance

YouTube

Watch Tutorial



50

BASIC FUNCTIONALITIES

PoolParty's core competencies at a glance

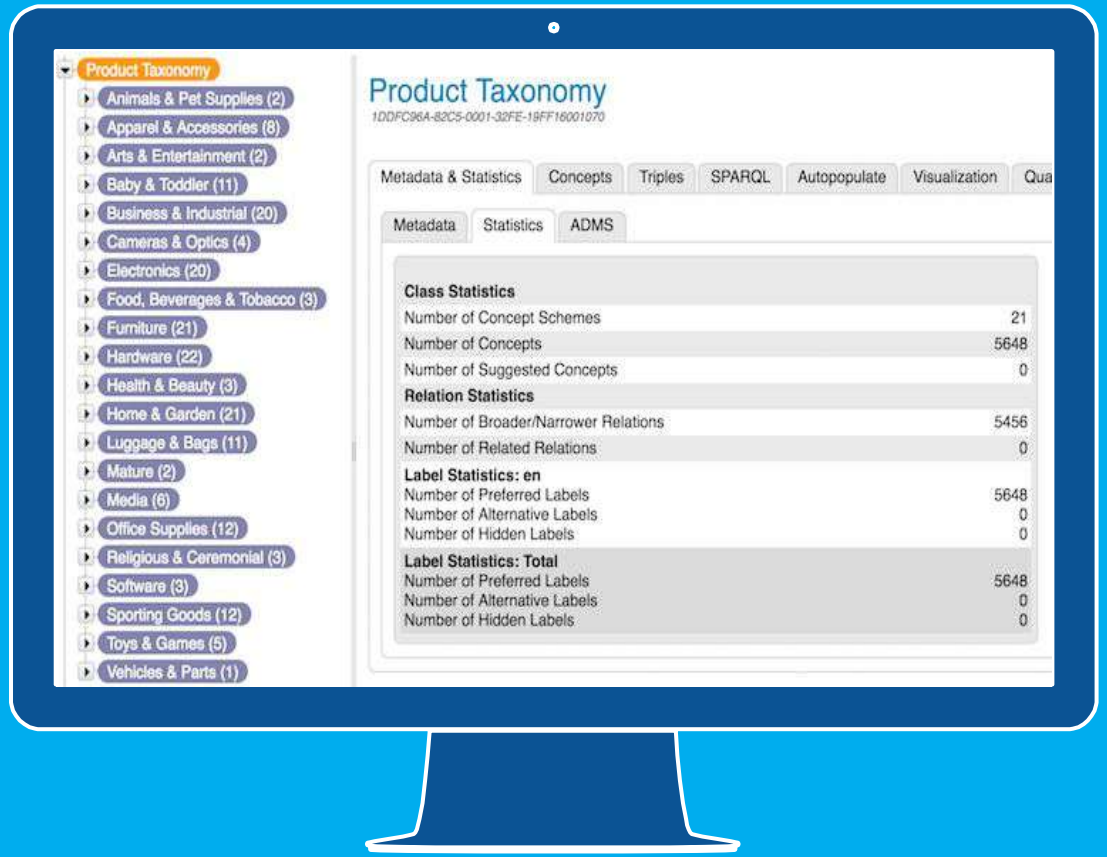
51

Maintaining Vocabularies

Taxonomies and controlled vocabularies are maintained by using the SKOS standard of W3C.

The intuitive user interface provides comfortable control elements like drag & drop or autocomplete.

A tree view on the taxonomy plays a central part in navigation and orientation.



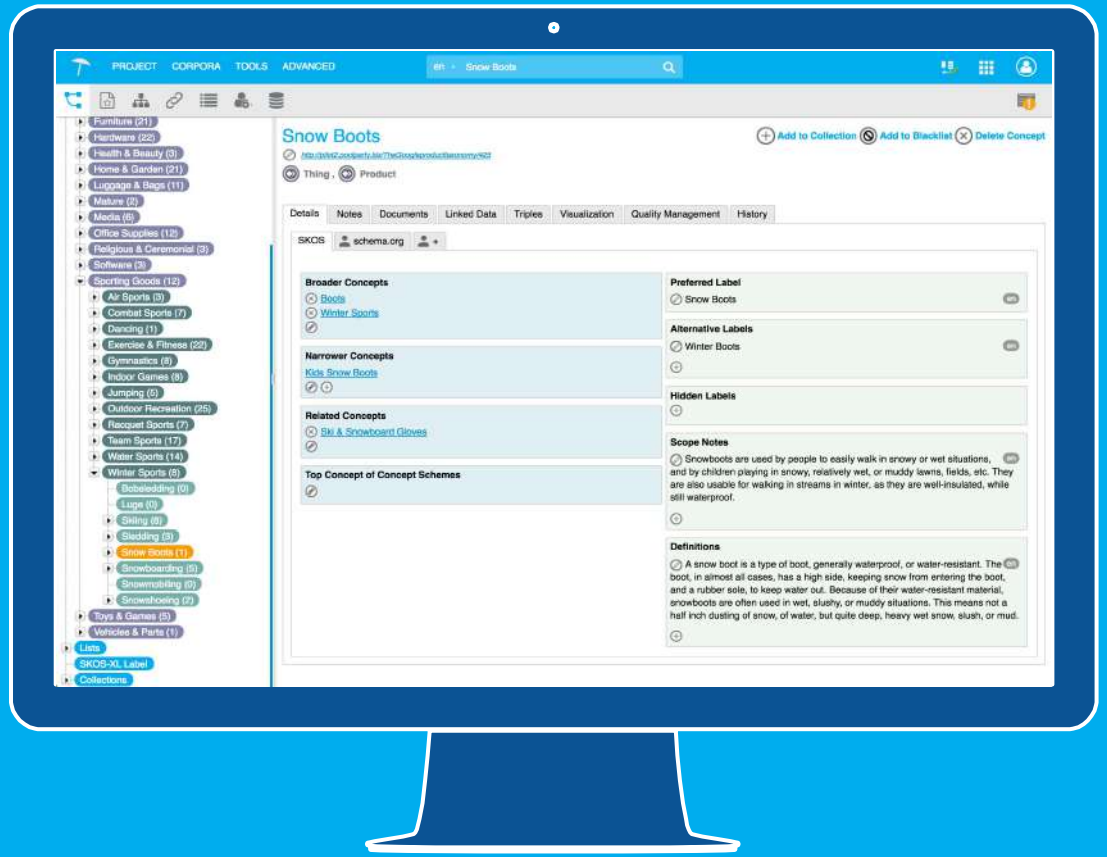
52

SKOS Editor

The SKOS View on a concept allows the management of labels (e.g. synonyms), hierarchies and non-hierarchical relations, and mappings to other vocabularies.

Also more complex actions like merging of concepts, moving of subtrees or the creation of poly-hierarchies are supported.

PoolParty fully covers the SKOS standard of W3C incl. SKOS-XL and SKOS Collections.



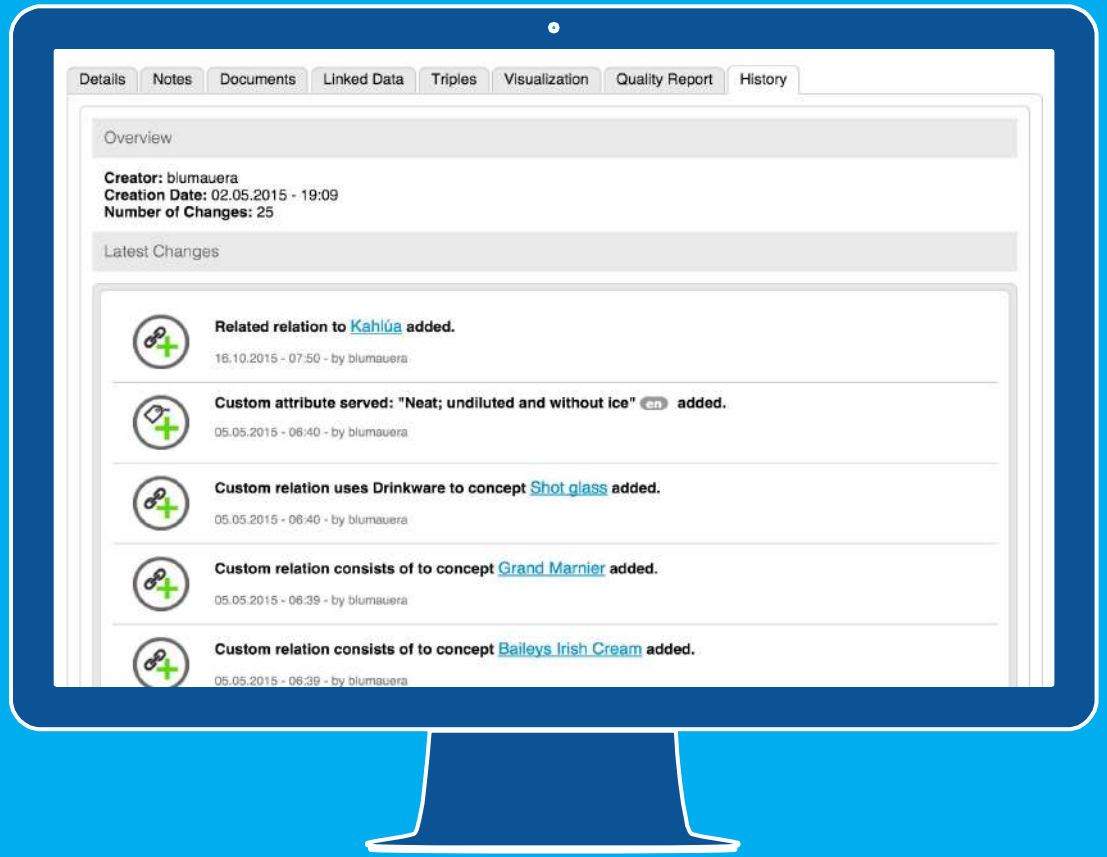
53

History & Audit Trails

Every change being made on a concept of a thesaurus is stored and can be tracked.

A full history containing the author, timestamp and action being taken can be displayed for each concept and for the whole project.

Recovery and rollback can be managed by PoolParty's snapshot mechanism.



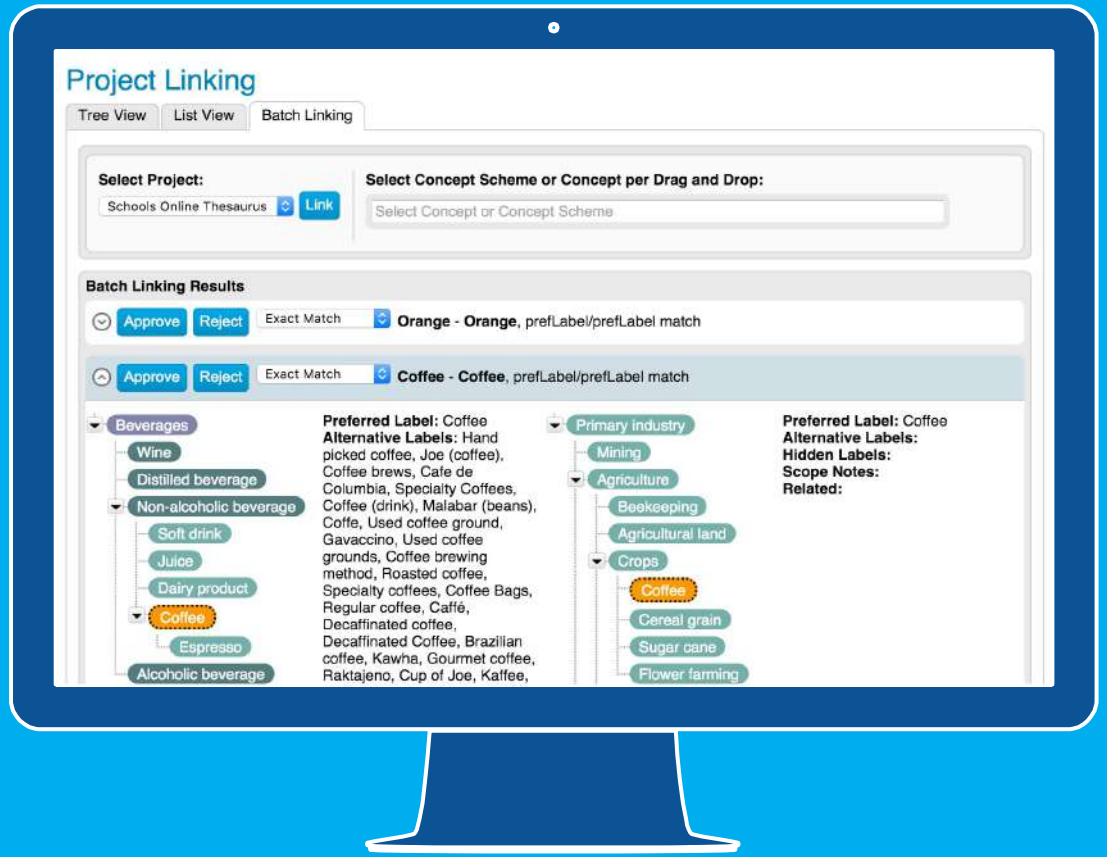
54

Linking & Mapping

The same concept can occur in several taxonomies and can be put in different contexts.

PoolParty provides a comfortable dialogue for the semi-automatic linking between concepts from several thesauri.

Additionally, concepts can also be mapped to linked data sources like DBpedia or Geonames, or even to non-RDF sources provided by you.



poolparty

55

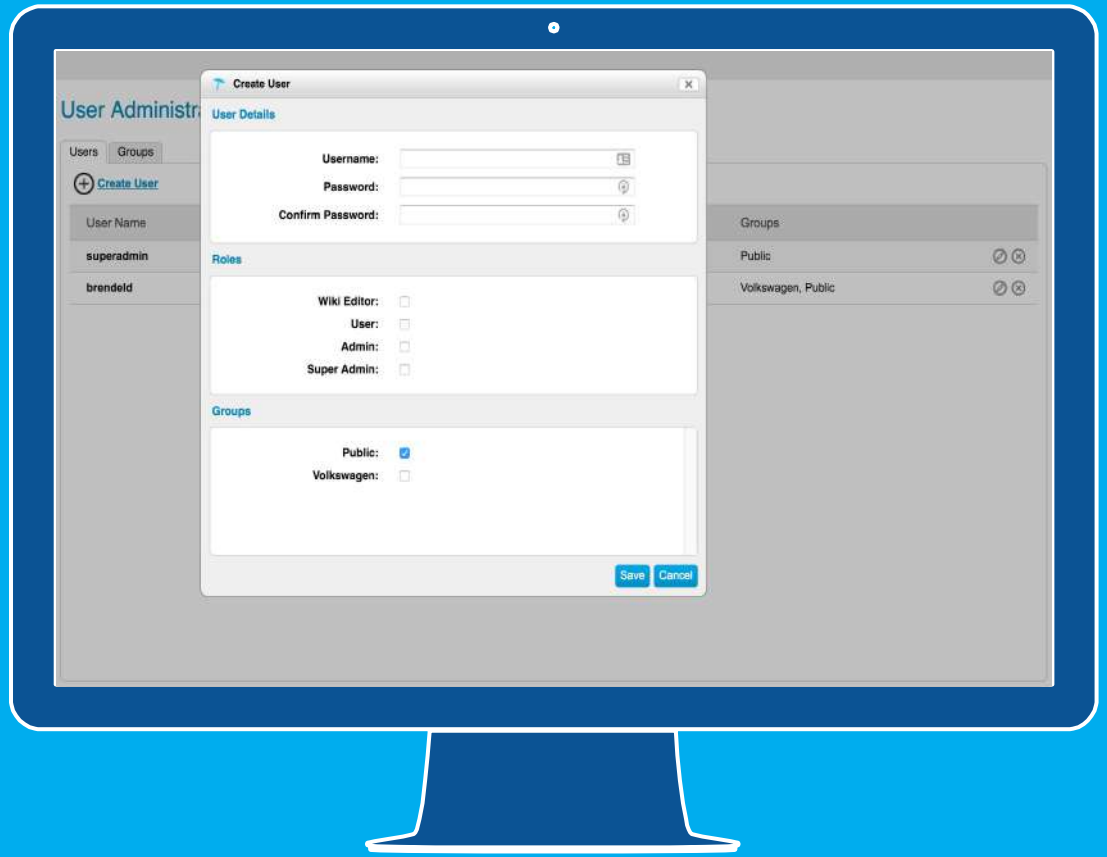
User Management & Roles

User Management is based on user accounts, roles, and groups.

User authentication can be integrated with LDAP.

PoolParty's security layer is based on Spring Security.

PoolParty's API is fully integrated with the security layer.



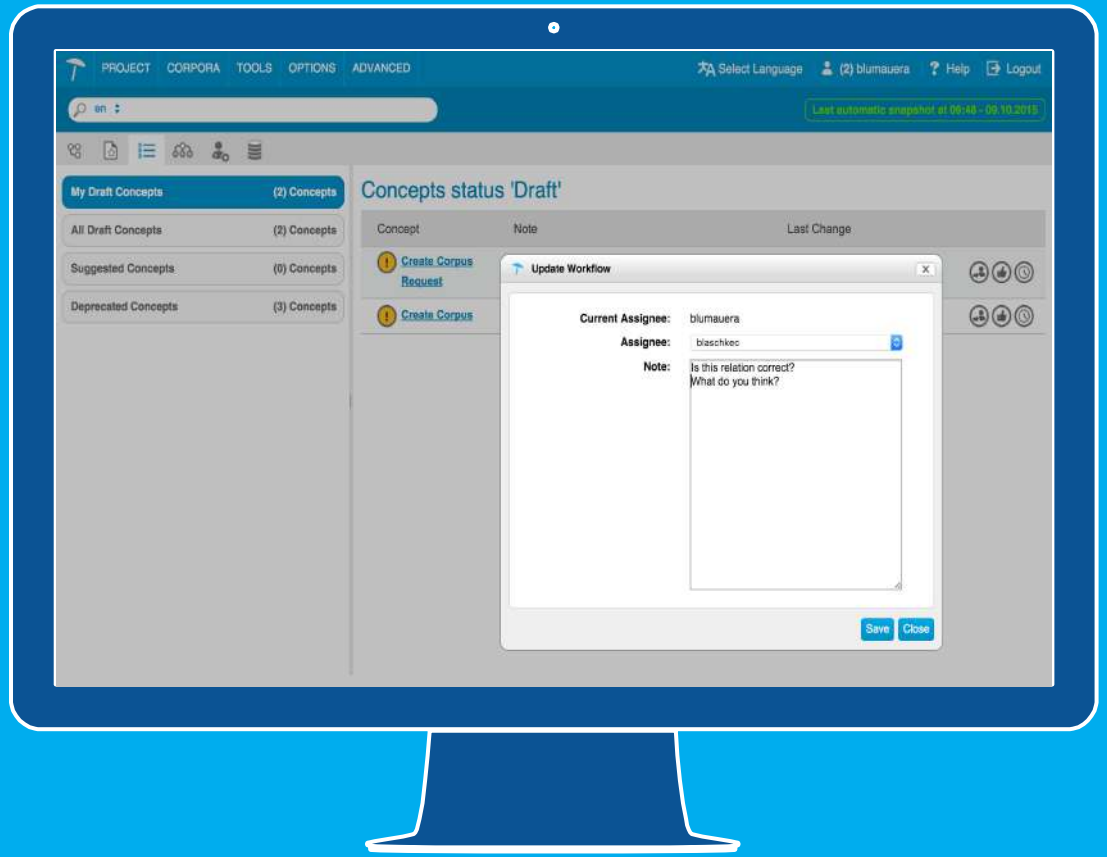
56

Workflows

Approval (or rejection) of changes on a thesaurus can be governed by workflows.

Several roles in the PoolParty system have different rights to apply changes, reject or approve those.

A clearly structured dashboard helps taxonomists not to loose track of all the tasks that need to be performed.



57

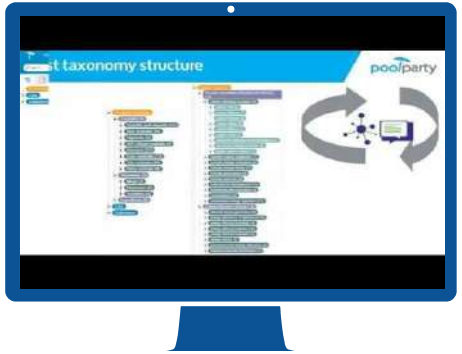
SELECTED VIDEOS

> [PoolParty on YouTube]

SKOS based Taxonomy Management

Workflows

Taxonomy Linking



Import Excel

poolparty

ADVANCED FUNCTIONALITIES

Efficient taxonomy management and text mining based on PoolParty

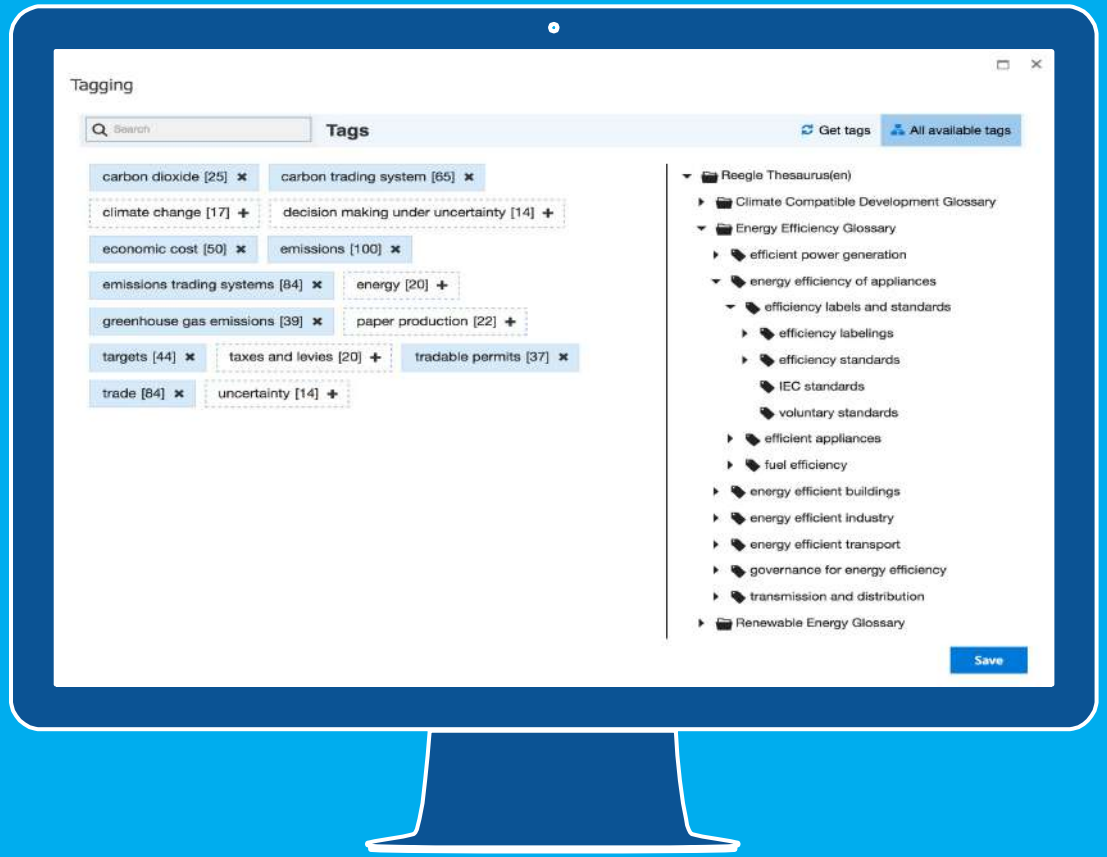
59

Entity Extraction

PoolParty's API provides a rich set of methods for text mining and entity extraction.

This ultra-fast service makes use of your controlled vocabularies, therefore it is highly accurate for your specific domain.

The service will improve over time and learns from reference text corpora. It supports over 40 languages and comes with a powerful disambiguation algorithm.



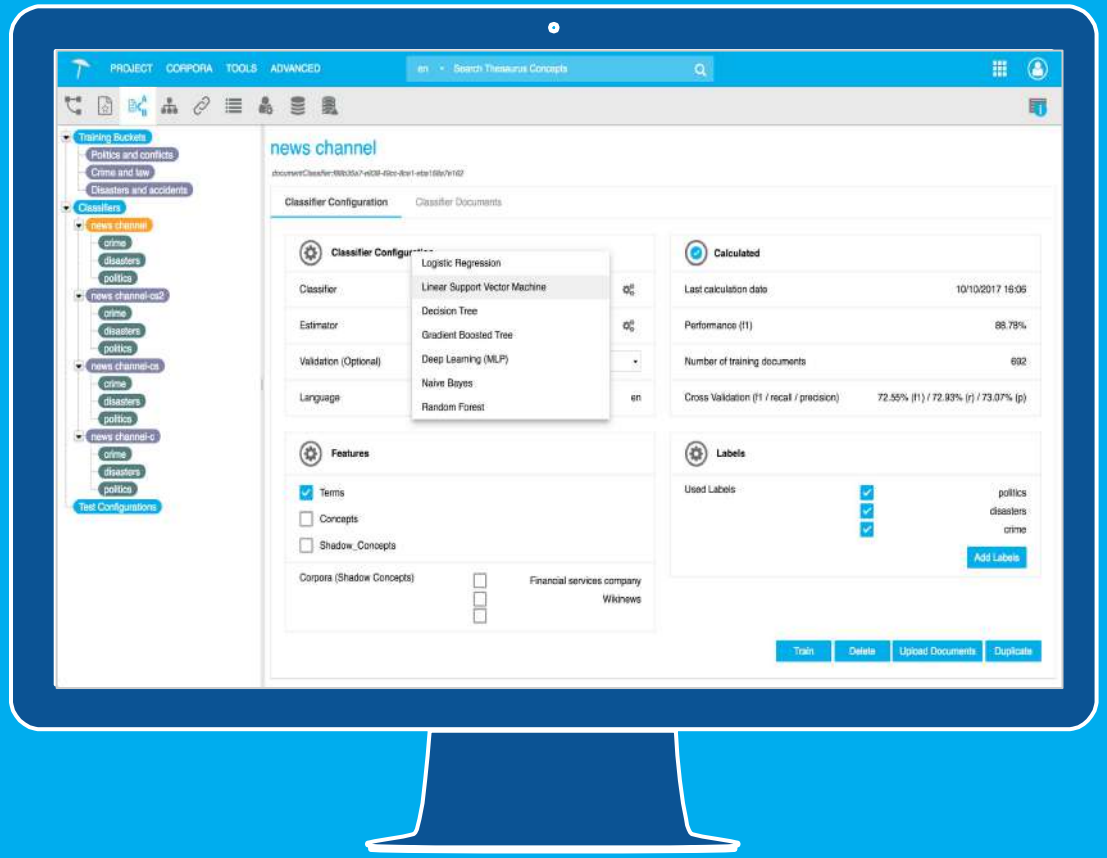
60

Semantic Classifier

Text Classification based on Machine Learning and Semantic Knowledge Models.

PoolParty Semantic Classifier combines machine learning algorithms (SVM, Deep Learning, Naive Bayes, etc.) with Semantic Knowledge Graphs.

The combined approach improves the classification results by up to 3% as compared to traditional term-based approaches.



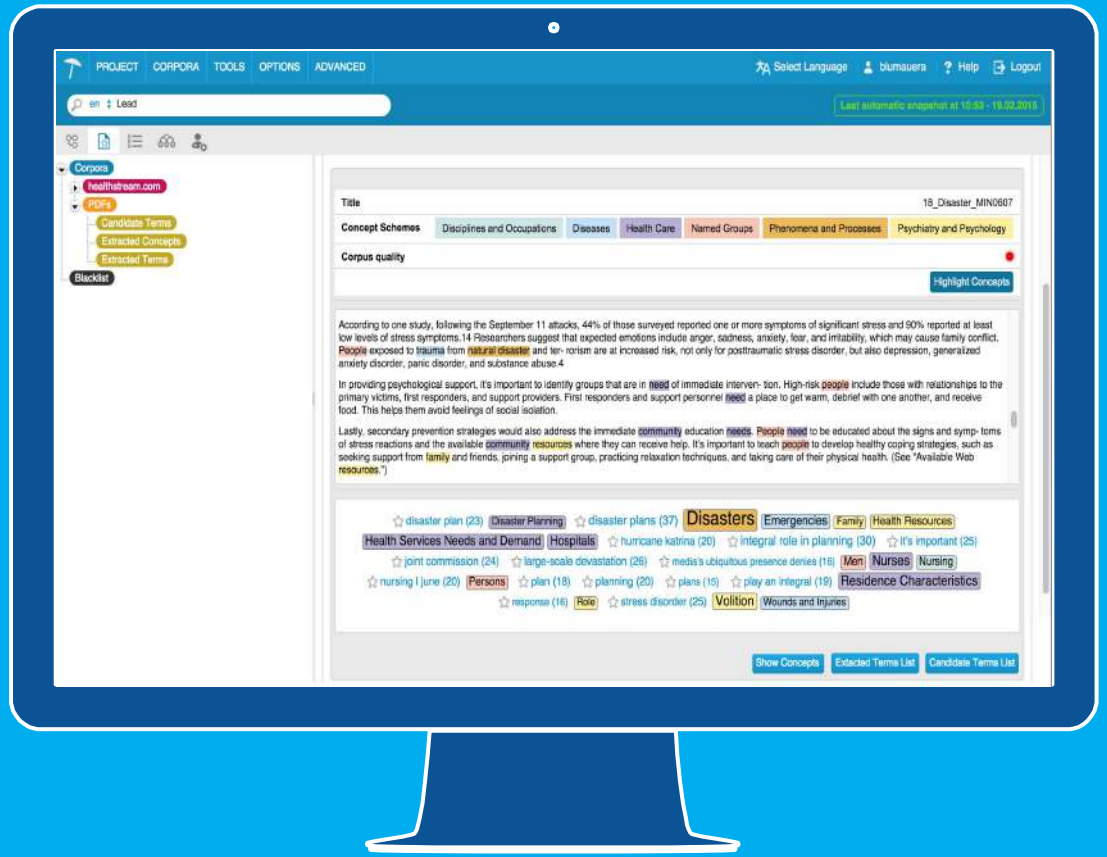
61

Corpus Analysis

PoolParty can automatically analyze reference text corpora.

The calculation of a statistical model of a ‘typical vocabulary’ of a specific domain helps to suggest candidate concepts for the expansion of a taxonomy.

By this means, the quality of term extraction improves over time and potential relations between concepts and terms can be suggested by the system.



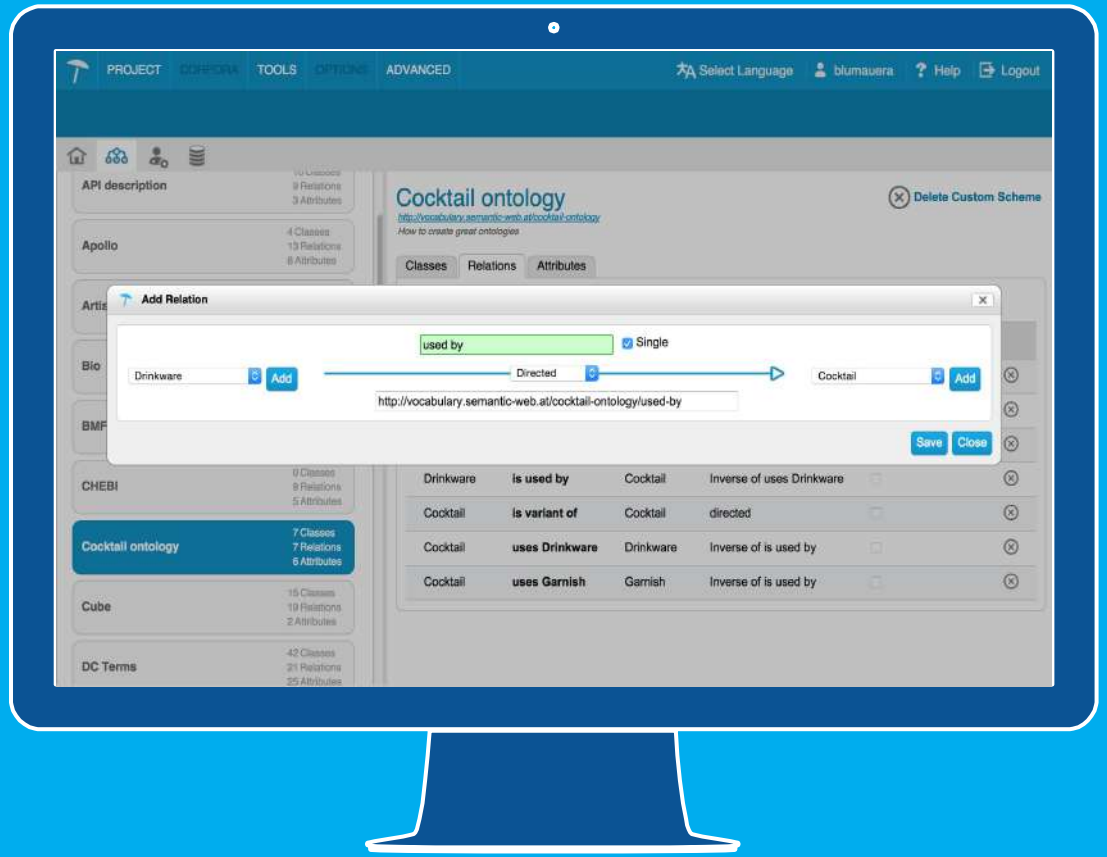
62

Custom Schemes & Ontologies

SKOS is based on a simple schema. This can be expanded by additional custom schemes.

Custom schemes can be created with help of PoolParty's ontology & schema editor.

For an increased interoperability, PoolParty provides a rich set of preconfigured ontologies like schema.org or FOAF.



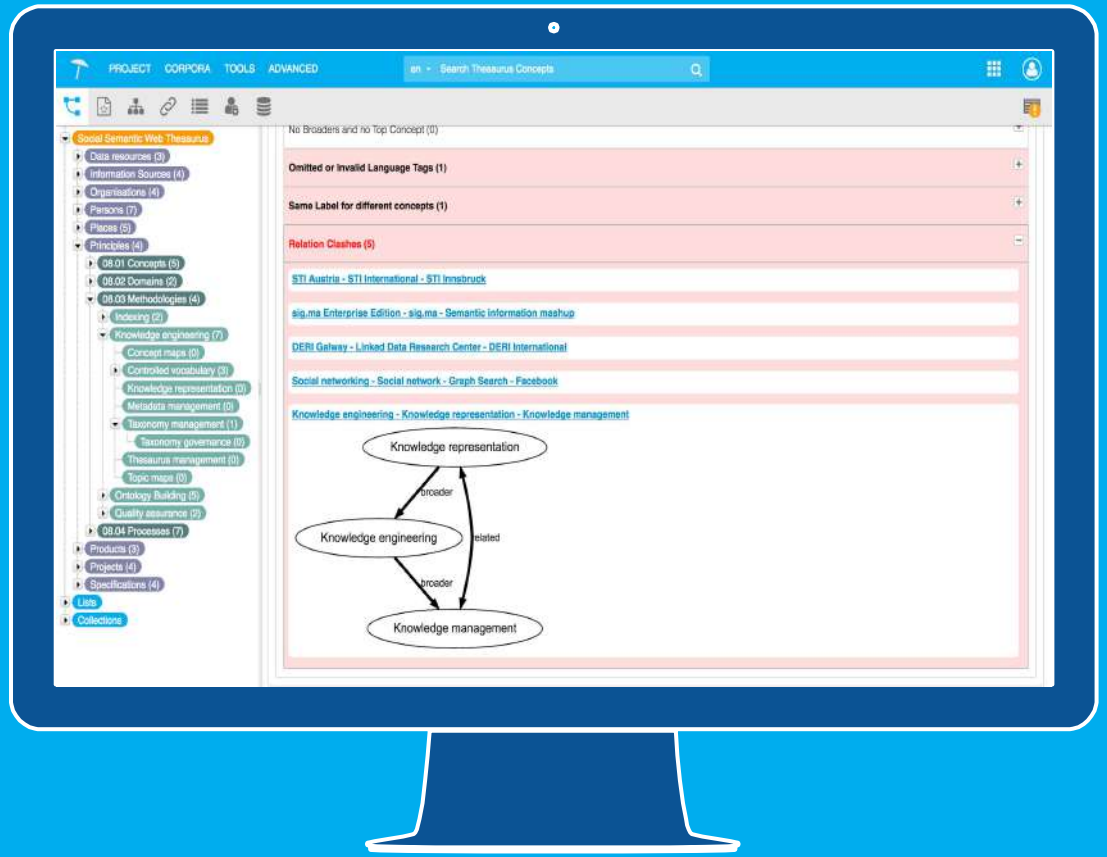
63

Quality Management & Import Validation

Data quality and especially the quality of metadata is key to a more efficient information management.

PoolParty Server provides several built-in quality checks (e.g. to avoid circularities).

Checks can be executed when imports are made, at run-time or at any time to generate a quality report.



64

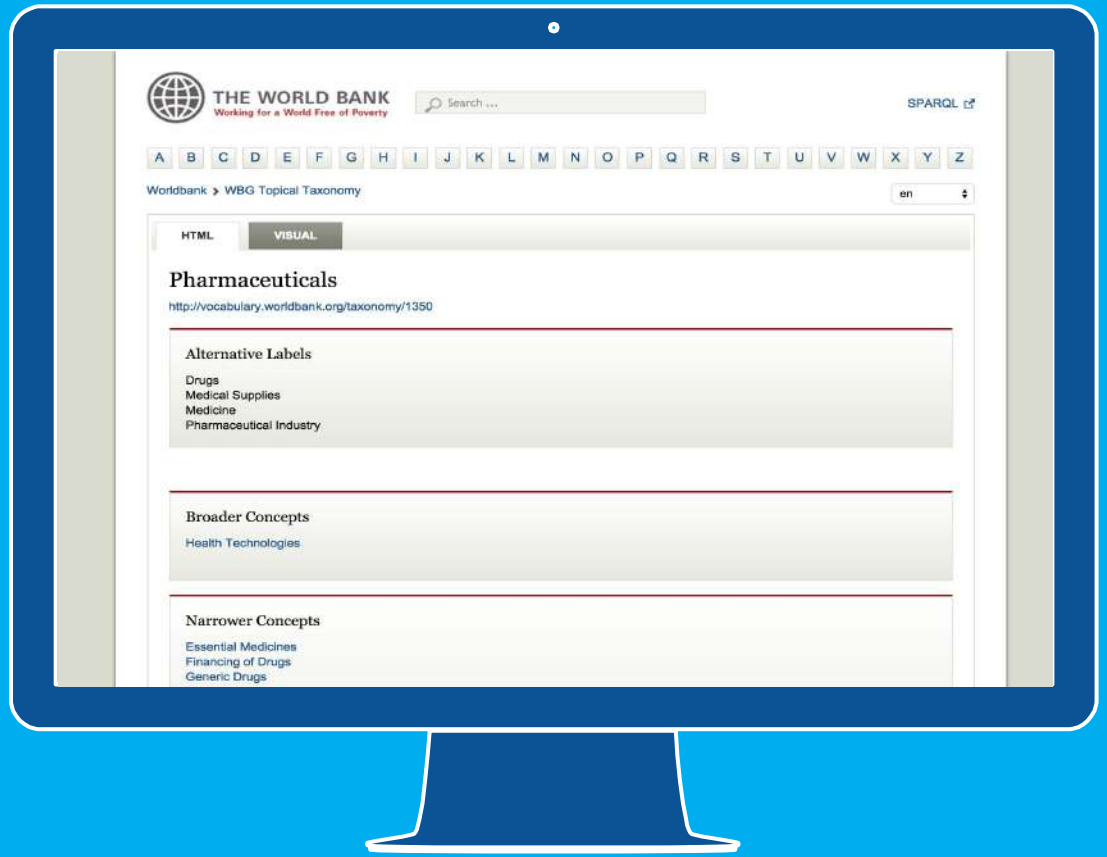
Linked Data

The use of Linked Data standards increases interoperability of your knowledge graphs & metadata.

With PoolParty, each thesaurus and ontology can be provided as a Linked Data graph.

In return, every linked data source can potentially be used to enrich a thesaurus.

PoolParty supports scenarios like ‘Enterprise Linked Data’ as well as ‘Linked Open Data’.



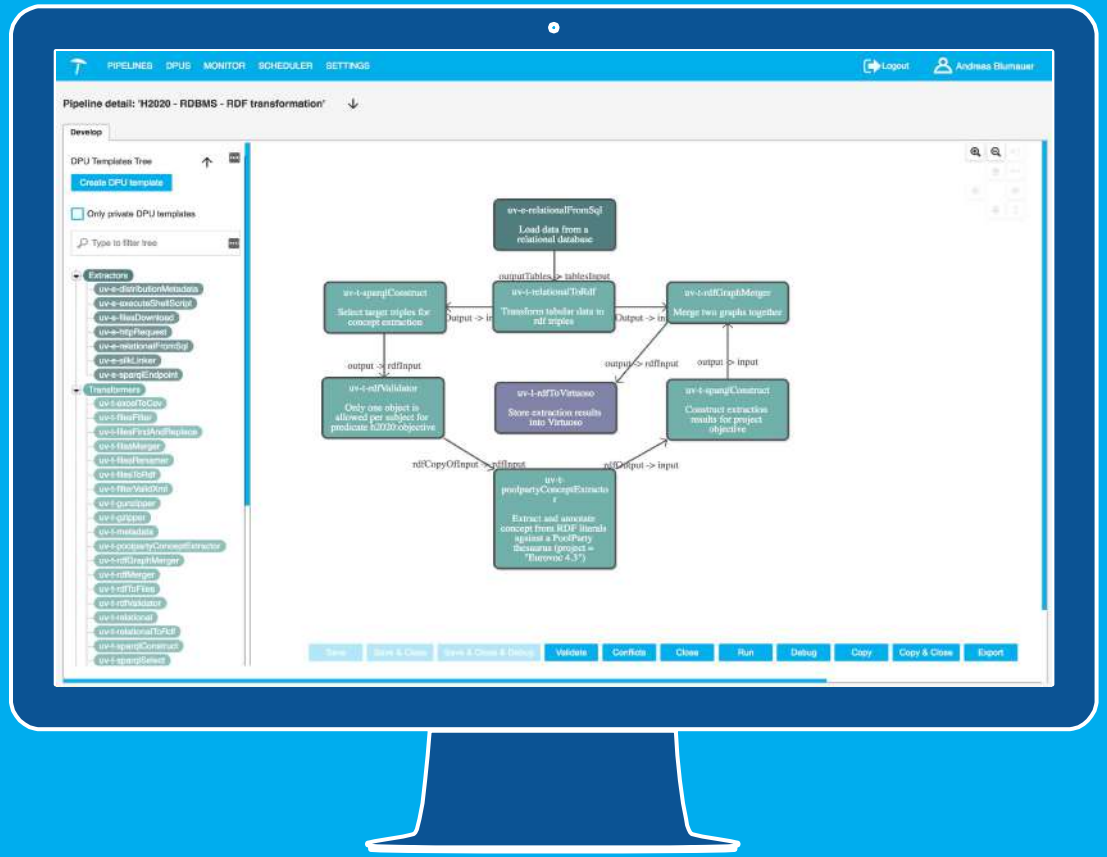
65

Linked Data Orchestration

With UnifiedViews, data processing tasks can be modelled as pipelines: Make use of the intuitively usable graphical interface.

Versatile data integration platform: Link data from internal and external data sources in a central NoSQL linked data warehouse.

Custom plugins: Your pipelines are highly customizable by creating your own data processing units (DPUs).



66

GraphSearch

Semantic search at the highest level: PoolParty Graph Search Server combines the power of graph databases and SPARQL engines with features of ‘traditional’ search engines.

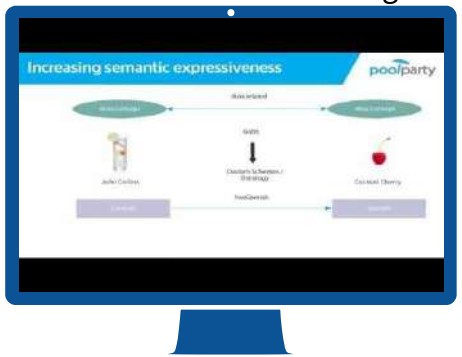
Document search and visual analytics: Benefit from additional insights through interactive visualizations of reports and search results derived from your data lake by executing sophisticated SPARQL queries.

67

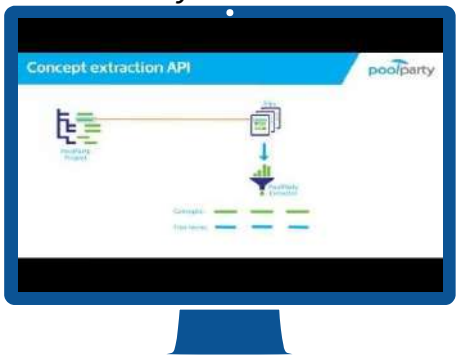
SELECTED VIDEOS

> [PoolParty on YouTube]

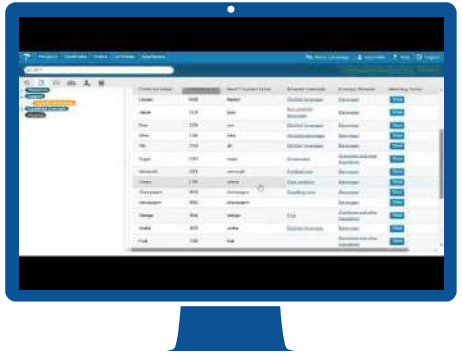
Custom Schemes & Ontologies



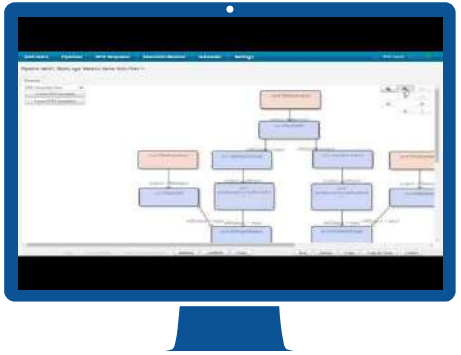
Entity Extraction

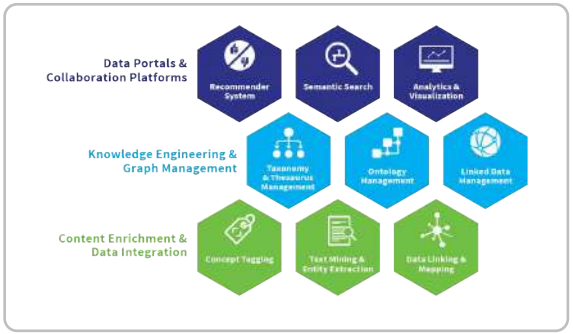


Corpus Analysis



UnifiedViews





W3C Semantic Web

SharePoint

INTEGRATION WITH A CMS

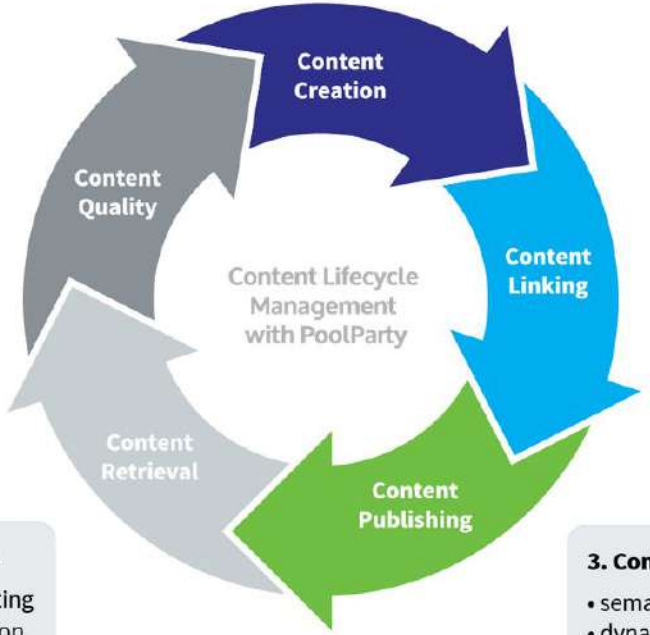
Benefiting from a Semantic Layer

69

INTEGRATING POOLPARTY ALONGSIDE THE CONTENT LIFE CYCLE

poolparty

1. Content Creation
- semi-automatic enrichment with metadata
- workflow management for different user roles
- content analytics



2. Content Linking
- semi-automatic content linking
- Integration of external sources
- Multimedia content recommendation

3. Content Publishing
- semantic annotation for SEO
- dynamic widgets / mash-ups
- visualization of content correlations

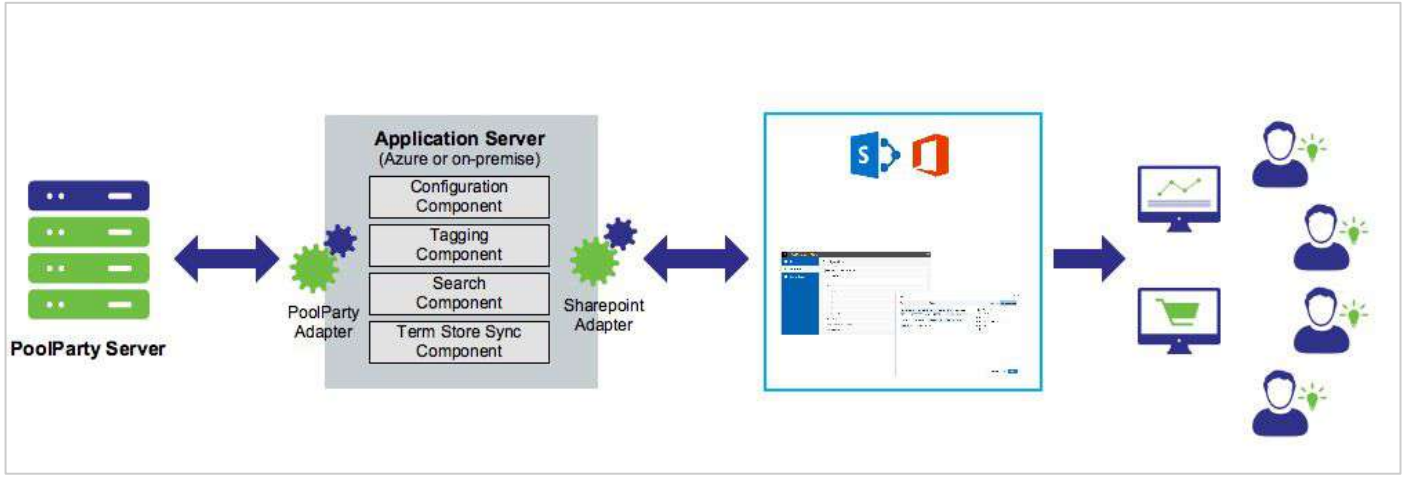
4. Content Retrieval
- classification & faceting
- search personalization
- push-services & alerts

5. Content Quality
- content reusage
- knowledge modelling
- sentiment analysis

70

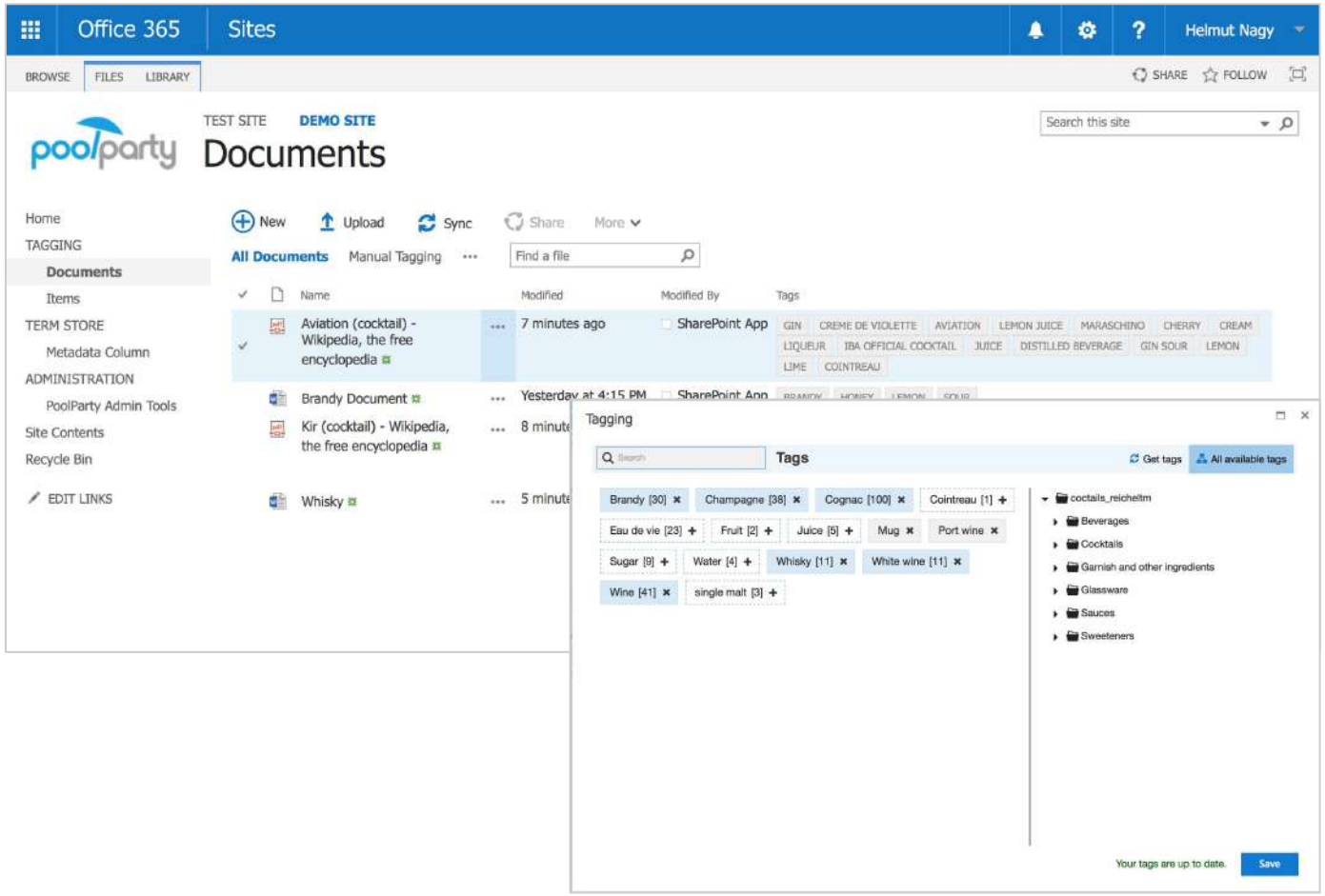
SharePoint and PoolParty at a Glance

> Learn more



71

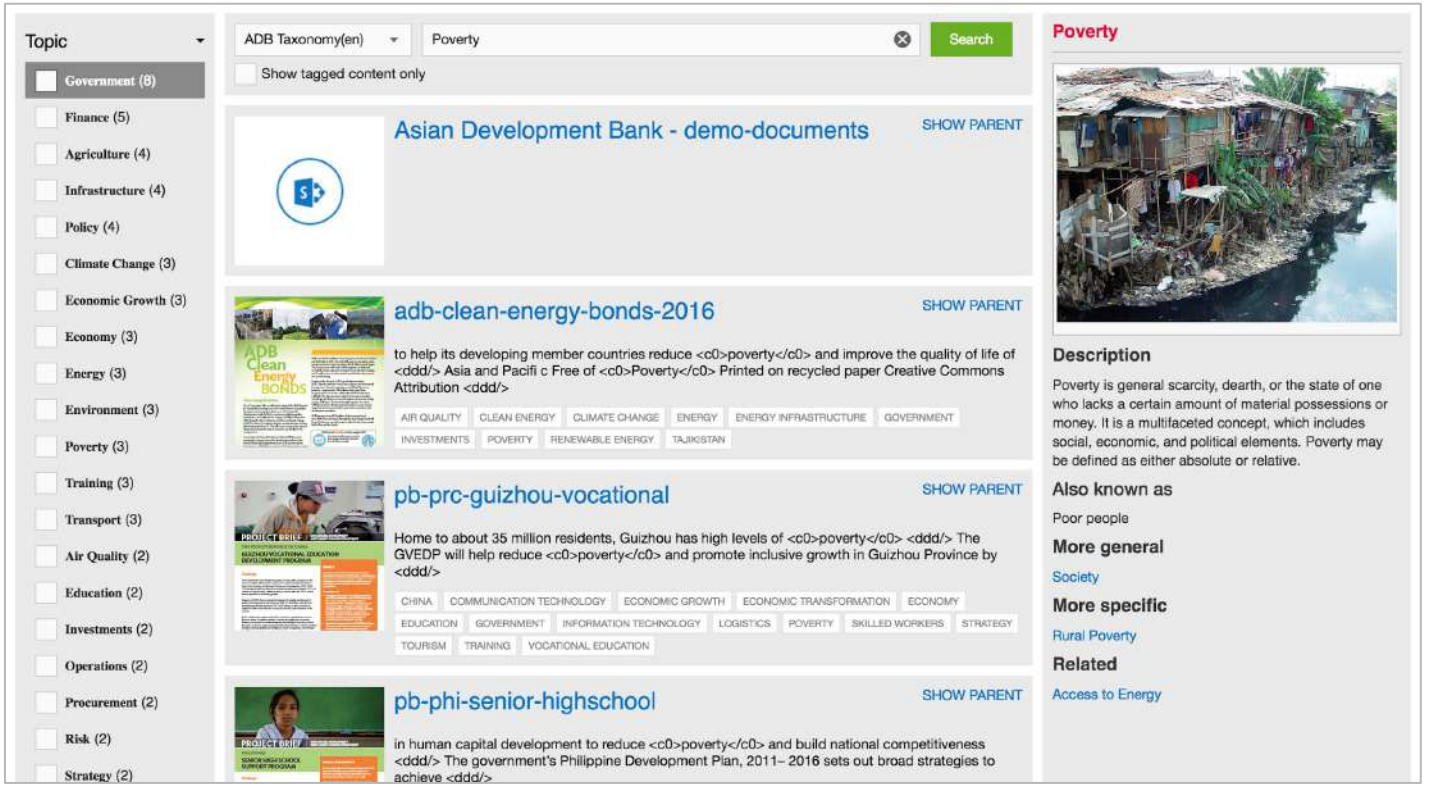
Autotagging & Consistent Tagging based on controlled vocabularies

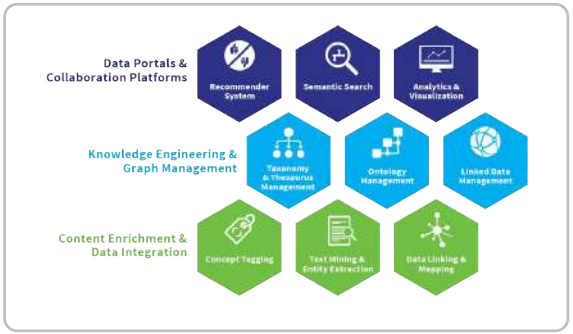


72

Semantic Search for SharePoint and Office 365

poolparty





INTEGRATION WITH MARKLOGIC

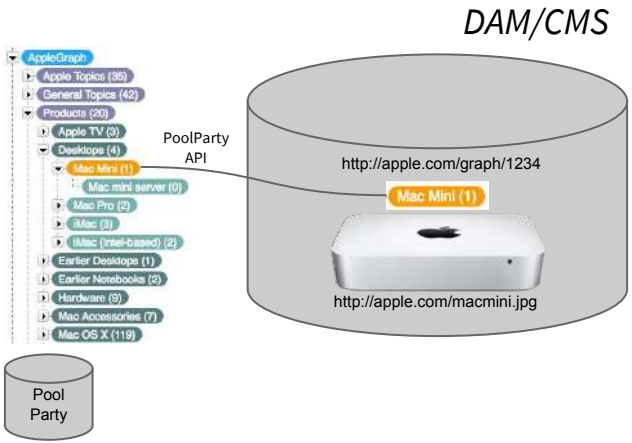
Benefiting from a Full Semantics Stack

74

TWO INTEGRATION SCENARIOS

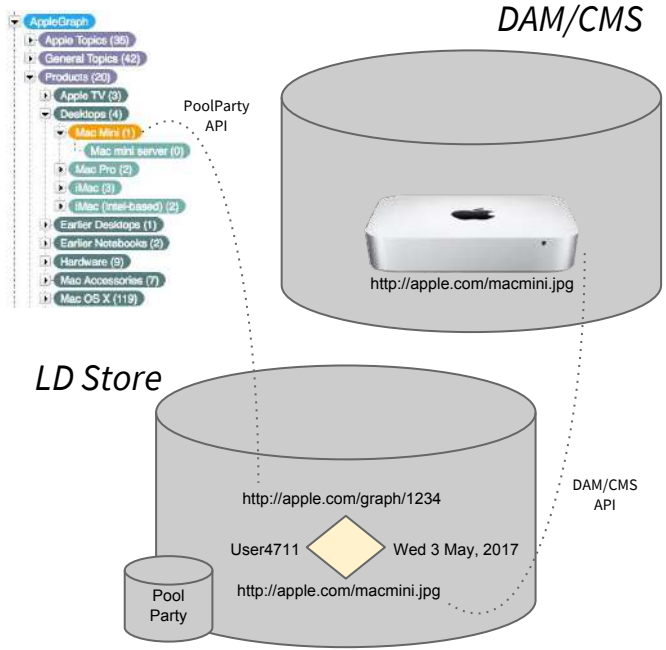
Option 1:

Concepts are derived from taxonomy and tagging is stored together with the asset in the DAM/CMS



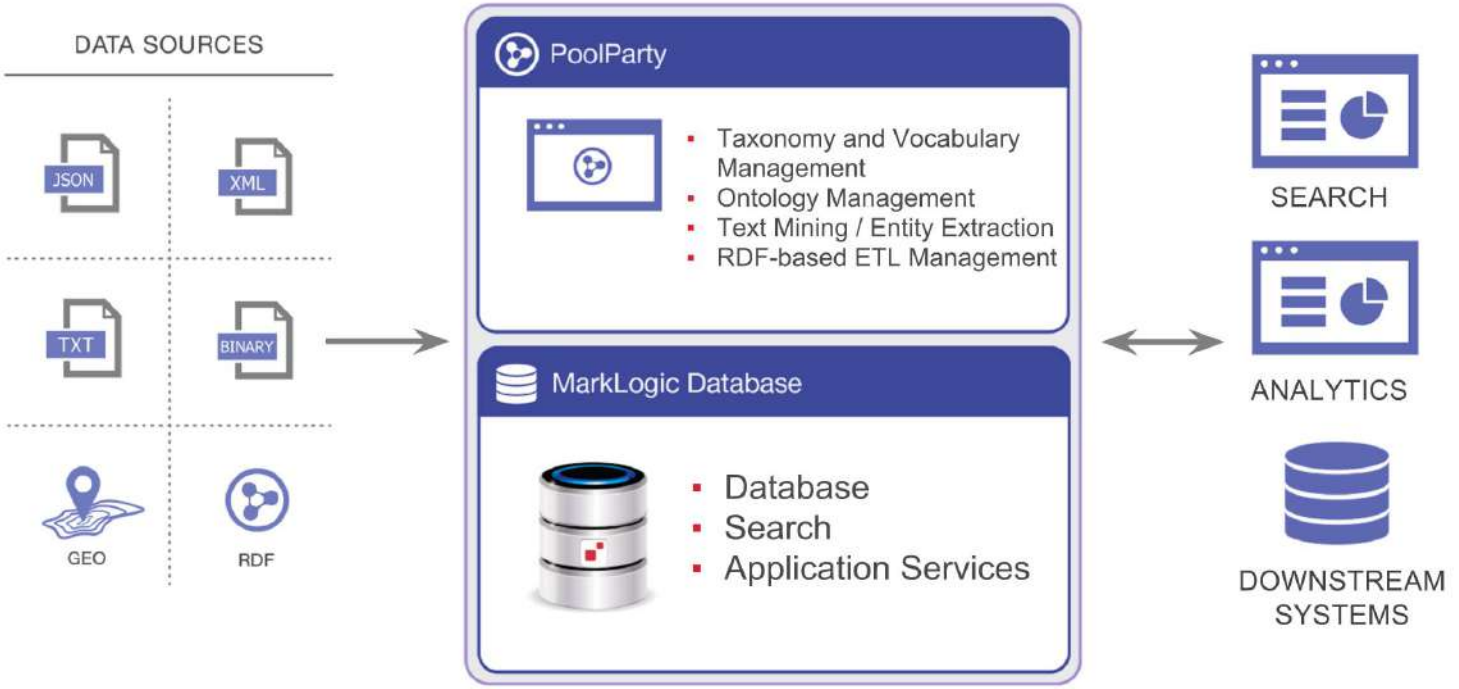
Option 2:

Concepts are derived from taxonomy, and tagging event is stored in a Linked Data Store by tying together assets with concepts from graph.



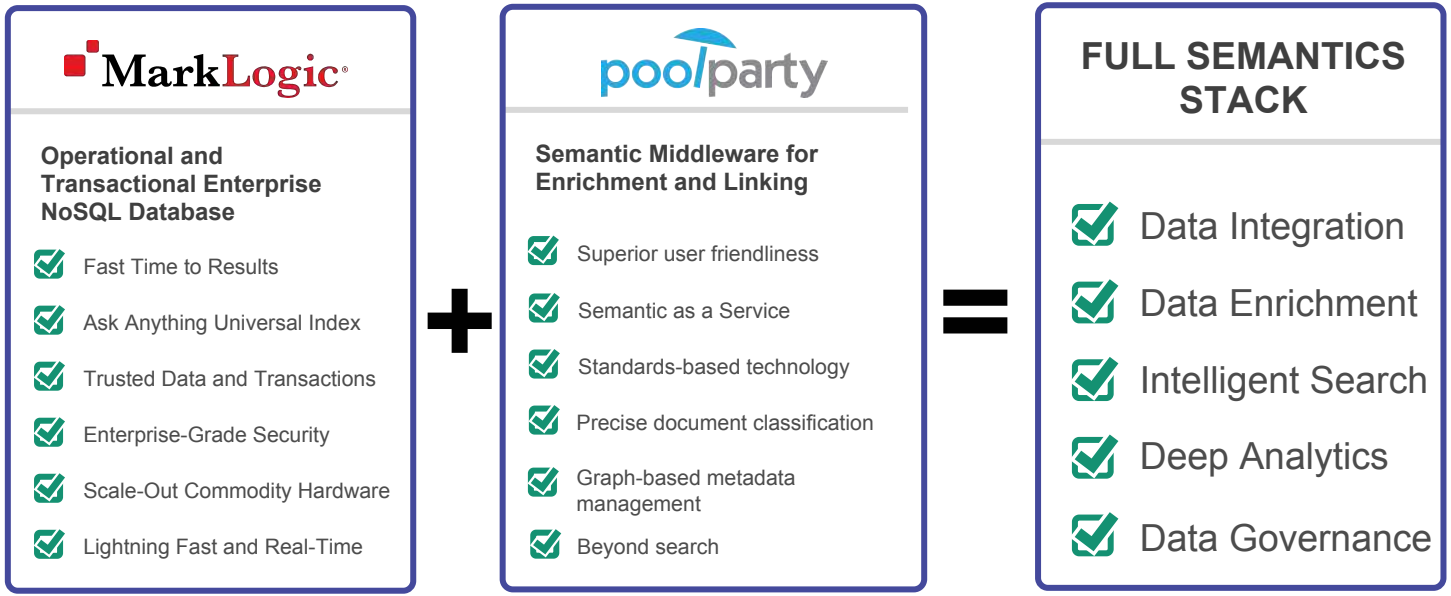
75

MarkLogic and PoolParty at a Glance



76

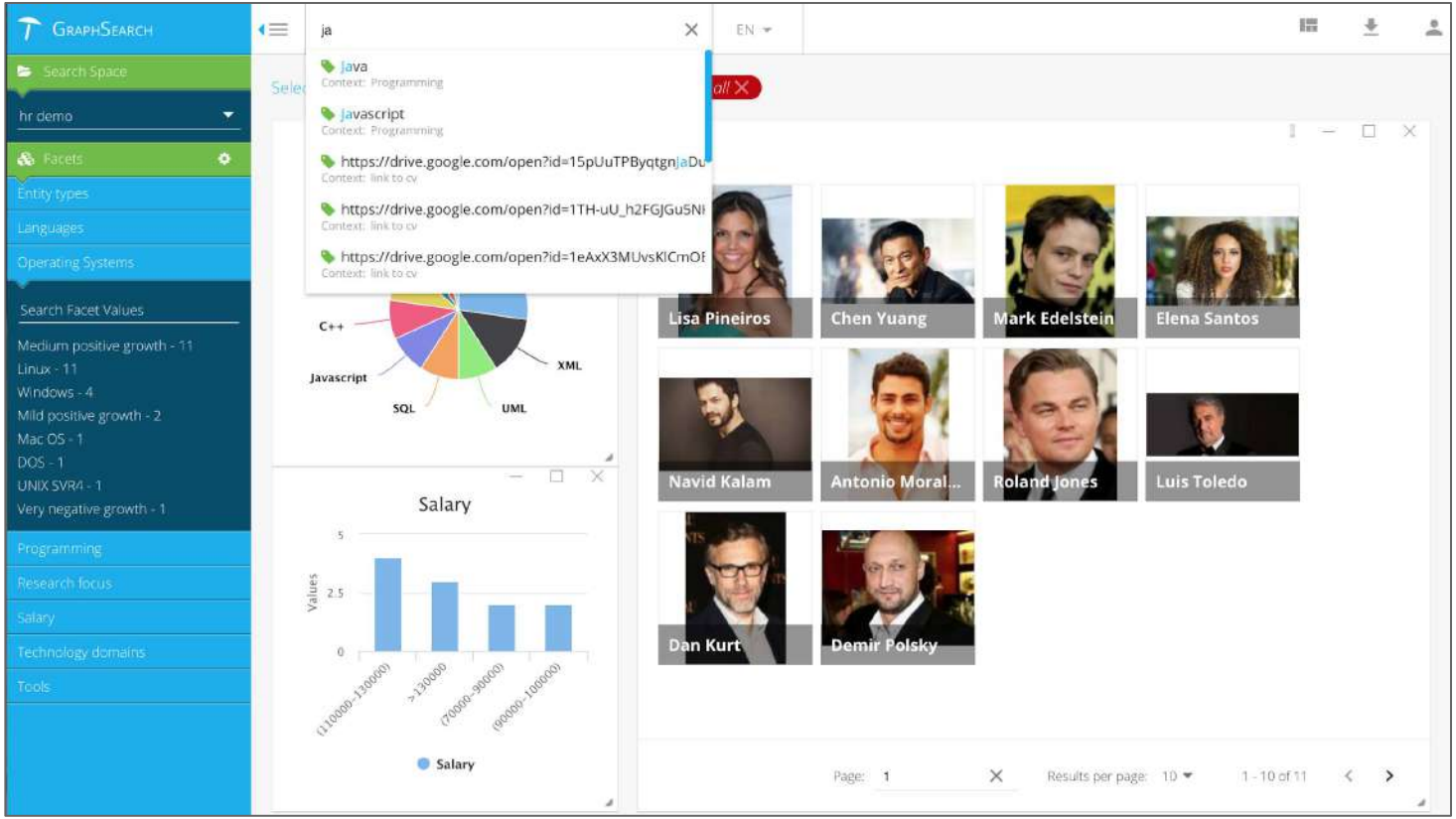
YOUR BENEFIT



77

MarkLogic / PoolParty Demo Application

> [Try it out!]



Learn more about [MarkLogic and PoolParty as a bundle]

Some Use Cases that make use of PoolParty

The potential of AI spans across industries and function. The digital workplace is becoming enhanced by smart applications. Business leaders have to decide where to start. They need to keep pace with the technological advancement and how their competitors embrace it.

Operations
Robotic Process Automation
Office Automation

Sales
Predictive Sales Forecasting
Sales Data Input Automation

Marketing
Ad Targeting
Adaptive Content

Service
Customer Intent Detection
Call Classification

HR
Skill Matching
New Hire Onboarding

Healthcare
Treatment Recommendation
Diagnostic Error Prevention

Media & Publishing
Auto-Tagging
Automated Journalism

Financial Services
Fraud Detection
Price Optimization

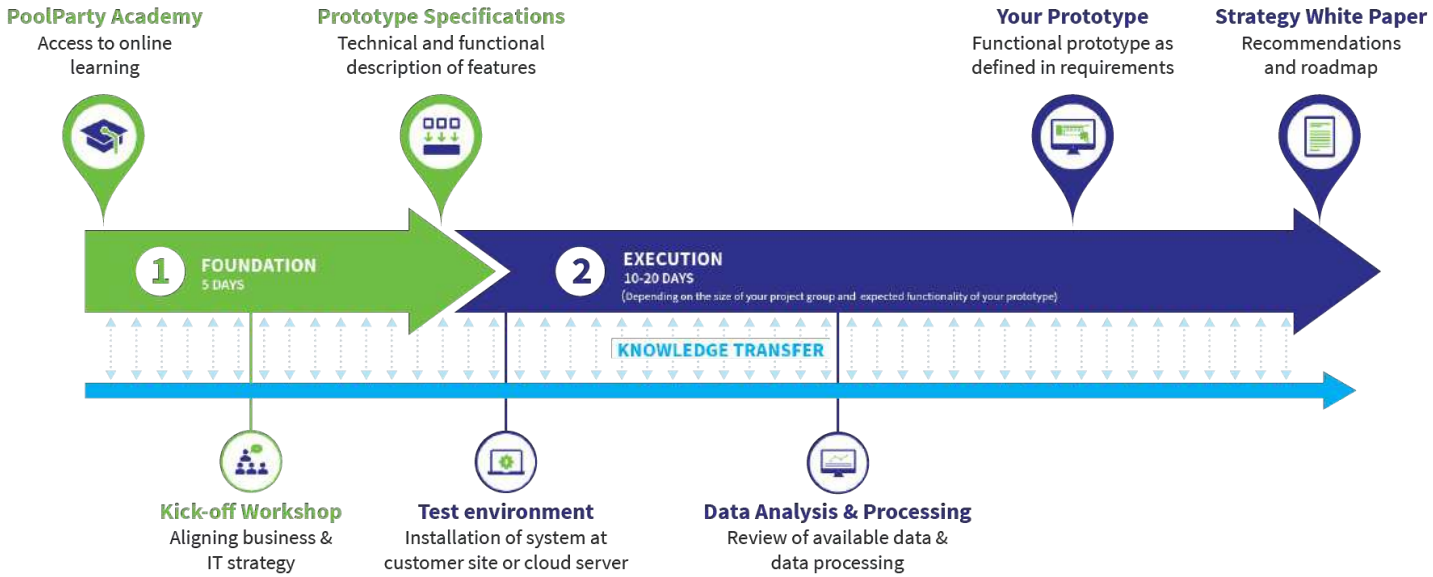
E-Commerce & Retail
Shopping Recommendation
Inventory Management

Education
Personalized Learning
Automated Grading

Government
Intelligent Citizen Portal
Automated Case Management

79

Semantic Web Starter Kit



80

CONNECT

Andreas Blumauer

CEO, Semantic Web Company

- andreas.blumauer@semantic-web.com
- https://www.linkedin.com/in/andreasblumauer
- https://twitter.com/semwebcompany
- https://ablvienna.wordpress.com/