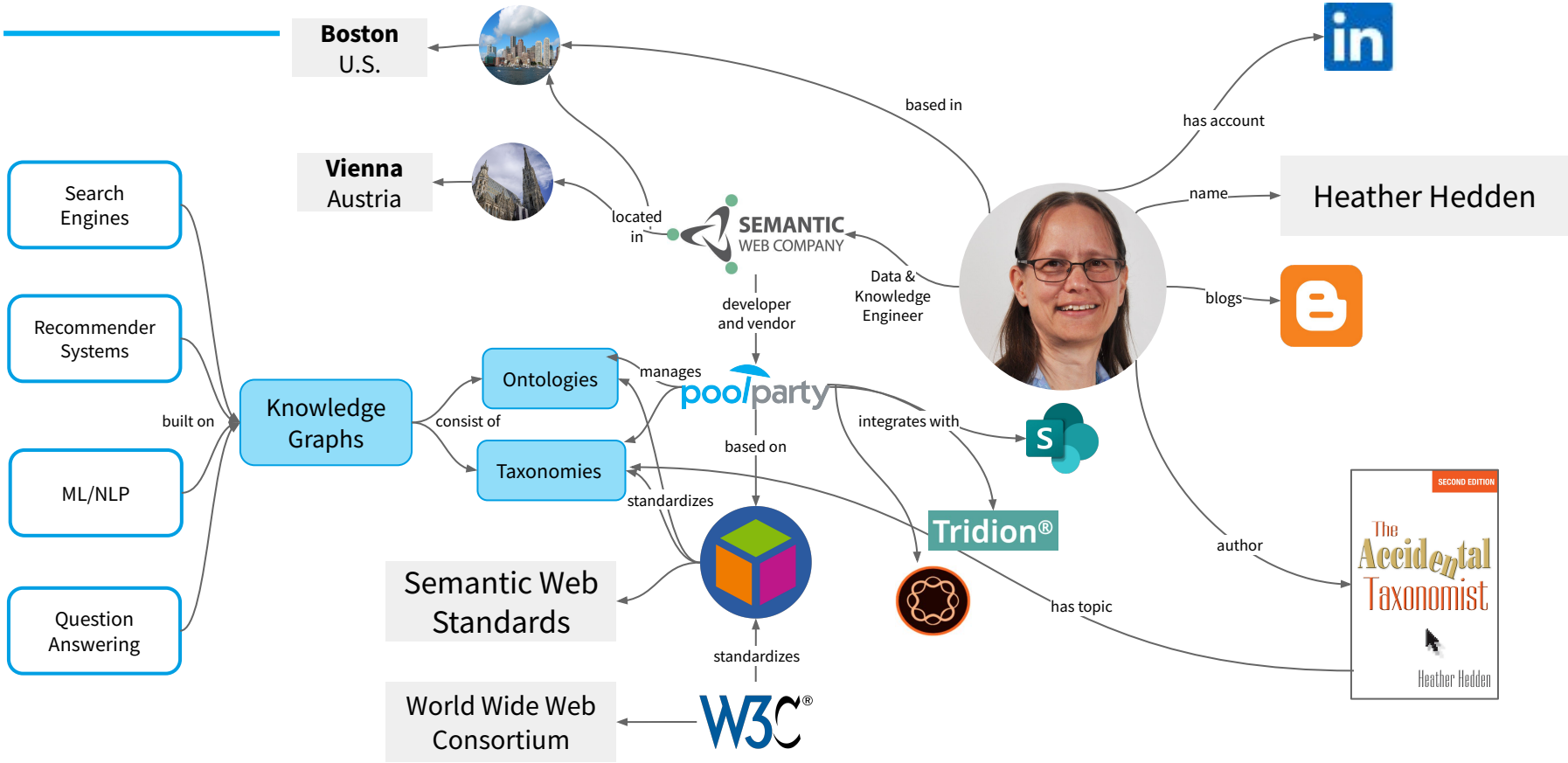
Enhancing Taxonomy and Tagging in SharePoint

North American Collaboration Summit
October 14, 2022

Heather Hedden
Data and Knowledge Engineer,
Semantic Web Company

Click the Graph—get in contact with us!



About the Speaker

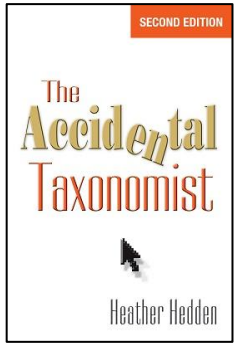
Heather Hedden
Data and Knowledge Engineer
Semantic Web Company

Over 25 years of experience in developing and managing taxonomies, metadata, and other knowledge organization systems for various organizations and applications.

Prior taxonomy consultant and staff taxonomist.

Instructor of self-paced online taxonomy courses.

Author of the book *The Accidental Taxonomist* (3rd edition, 2022)



About Semantic Web Company and PoolParty

SWC is developer / vendor of **PoolParty Semantic Suite**

Most complete and secure **Semantic AI platform** on the global market

W3C standards compliant

ISO 27001:2013 certified (since 2019)

First release in 2009

Current version **8.1**

On-premises or **cloud-based**

Over **170** **customers world-wide**

Semantic AI: Fusion of graphs, NLP, and machine learning

Gartner named SWC a Visionary in their **Magic Quadrant** for Metadata Management Systems 2019 and 2020

KMWorld listed PoolParty as **Trend-Setting Product** 2015 - 2022 and SWC in the **AI 50** list of companies in 2020 and 2022

Forrester listed SWC as sample vendor in their **report** on *The Document-Oriented Text Analytics Platforms Landscape 2022*

Named a Visionary in Gartner's Magic Quadrant for Metadata Management (2019 and 2020)

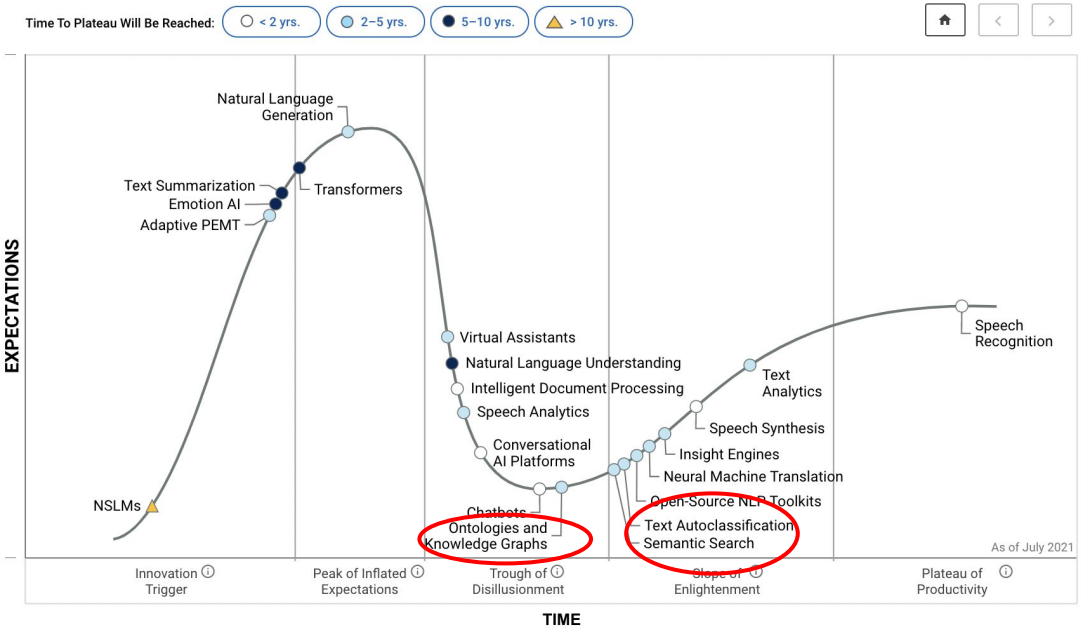
Figure 1. Magic Quadrant for Metadata Management Solutions

Source: Gartner (November 2020)

The market for metadata management solutions continues to demand **more innovation** and better execution in order to address the **needs of automation**, and combined cloud and on-premises deployments. Metadata management technology continues to expand in terms of capabilities and support for **multiple use cases**.
(Gartner, Inc: 'Magic Quadrant for Metadata Management Solutions', Guido De Simoni, Mark Beyer, and Ankush Jain, October 2019)

The Enterprise metadata management market has been valued at USD 2.72 billion in 2017 and is expected to grow at a **CAGR of 22.6%** during the forecast period (2018 - 2023) to reach USD 9.34 billion by 2023.
(Research And Markets, August 2018)

PoolParty Named a Sample Vendor in Gartner's Hype Cycle for Natural Language Technologies, 2021

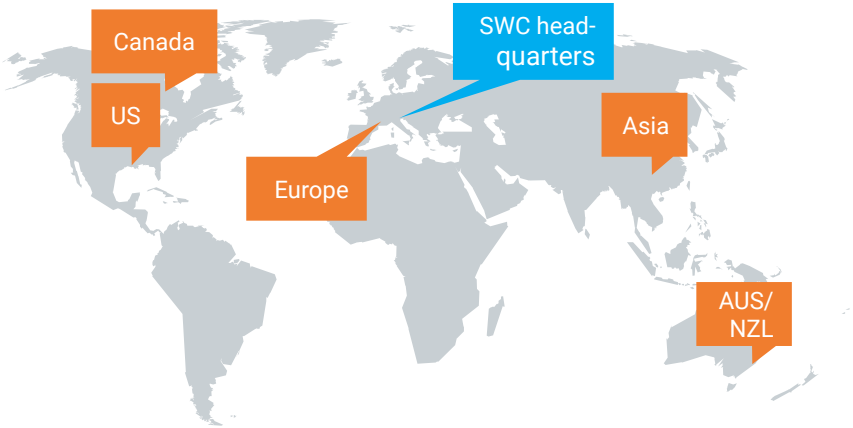


SWC/PoolParty is named a 'Sample Vendor' in three categories:

- **Ontologies and graphs** as components of broader hybrid AI systems used in various NLT markets, such as text mining or conversational systems.
- **Semantic search** as an important part of Insight Engines, as it complements the underlying search technology.
- **Text auto-classification**, which is used for automation purposes (e.g., within platforms for content services, RPA, etc.).

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 Corporation
- Harvard Business School
- Wolters Kluwer
- Philips
- RGP
- Becton Dickinson
- Springer Nature
- Novartis
- Healthdirect Australia
- Healthstream
- Walmart
- Oxford University Press
- International Atomic Energy Agency
- Dutch Parliament
- Microsoft
- Inter-American Development Bank

Selected Partners

- RWS
- Enterprise Knowledge
- Tahzoo
- Ontotext
- Oxford Semantic Technologies
- data.world
- Capco
- Taxonomy Strategies
- Mekon
- Factor
- Ordina
- BAON Enterprises
- Tieto/Findwise
- PANTOPIX
- The QA Company
- MMG Management Consulting
- Kompetenzzentrum Records Management (KRM)
- Cambrica

Outline

- Introduction: Why Taxonomies and Tagging
  - Use Cases to Connect to Content
  - Search Challenges
- Benefits of Managed Metadata/Taxonomies
- Challenges of Managed Metadata/Taxonomies
  - Taxonomy Design Issues and Tips
  - Taxonomy Governance
- Auto-Tagging
- Central Taxonomy Management

Introduction:
Why Taxonomies & Tagging

Why Taxonomies:
Connecting Content to People, Intents, and Other Data

Product Information

PoolParty is a semantic AI platform that can run sophisticated text mining over Office documents while using machine learning.

List of AI platform vendors

Ashok
Customer

User Intent

Show AI systems that can mine PPTX, DOCX etc. using ML to fulfill the requirements of EU's ethical guidelines for AI.

Why Taxonomies:
Connecting Content to People, Intents, and Other Data

Text mining

Semantic AI platform

Product Information

PoolParty is a semantic AI platform that can run sophisticated text mining over Office documents while using machine learning.

Machine learning

Office documents

Market Data

List of AI platform vendors

AI platform vendors

AI systems

User Intent

Show AI systems that can mine PPTX, DOCX etc. using ML to fulfill the requirements of EU's ethical guidelines for AI.

DOCX

PPTX

EU's ethical guidelines for AI

AI software

SharePoint

EU guidelines

Ashok
Customer

Personal Profile

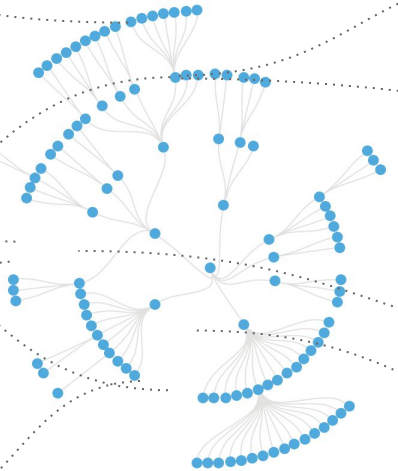
Why Taxonomies:
Connecting Content to People, Intents, and Other Data

Market Data

List of AI platform vendors

Product Information

PoolParty is a semantic AI platform that can run sophisticated text mining over Office documents while using machine learning.



AI software

SharePoint

Text analytics

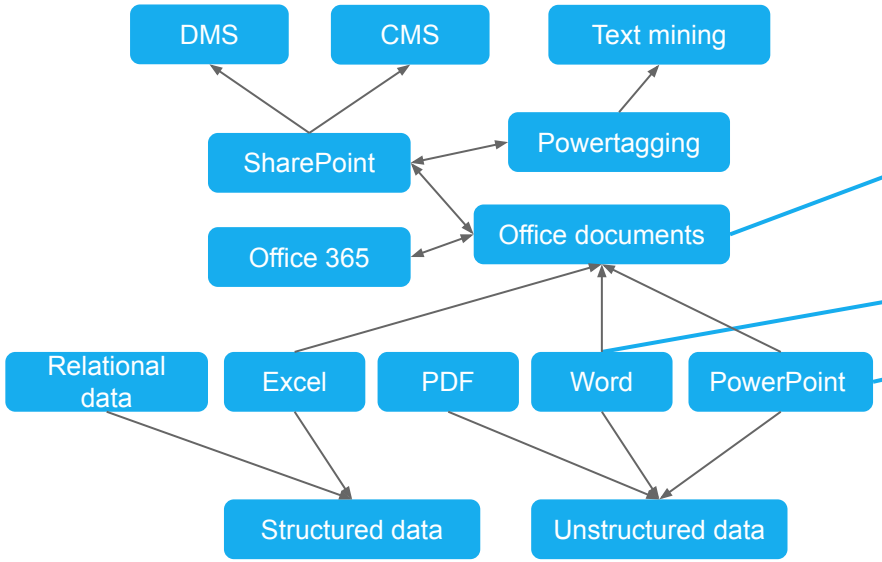
Ashok
Customer

Personal Profile

User Intent

Show AI systems that can mine PPTX, DOCX etc. using ML to fulfill the requirements of EU's ethical guidelines for AI.

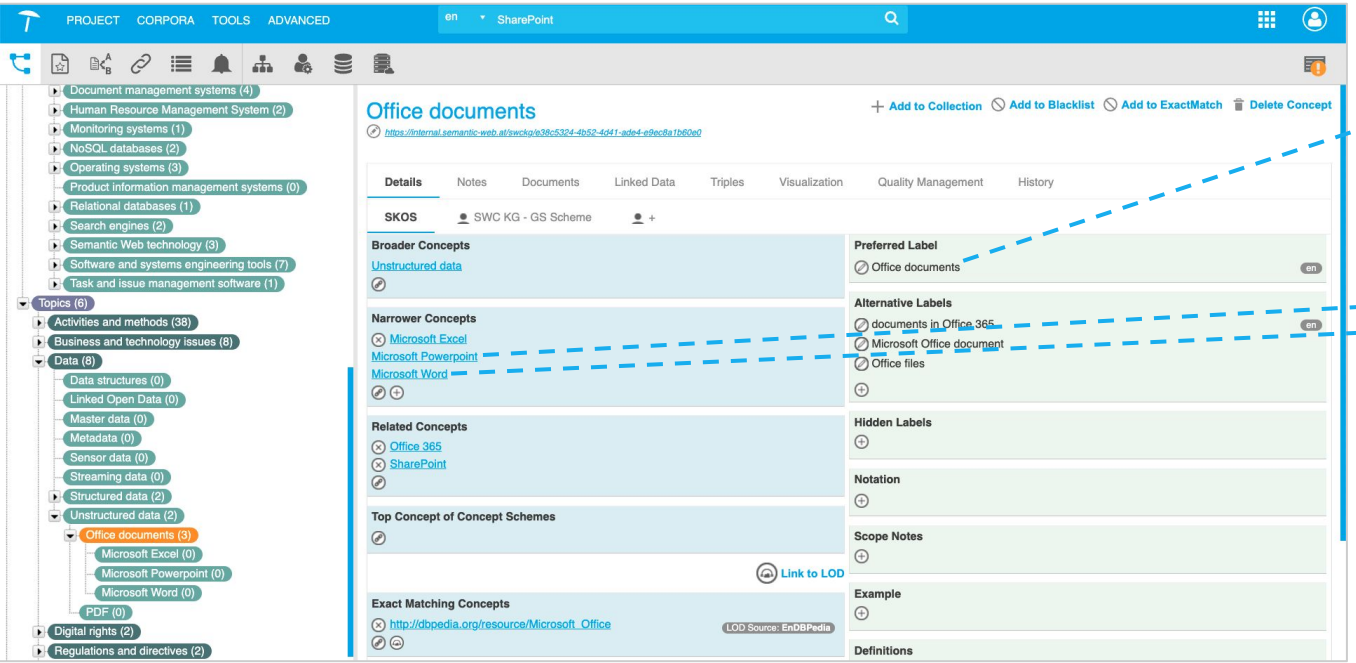
Why Taxonomies:
Connecting Content to People, Intents, and Other Data



PoolParty is a semantic AI platform that can run sophisticated text mining over Office documents while using machine learning.

Show AI systems that can mine PPTX, DOCX etc. ausing ML to fulfill the requirements of EU's ethical guidelines for AI.

Why Taxonomies:
Connecting Content to People, Intents, and Other Data



PoolParty is a semantic AI platform that can run sophisticated text mining over Office documents while using machine learning.

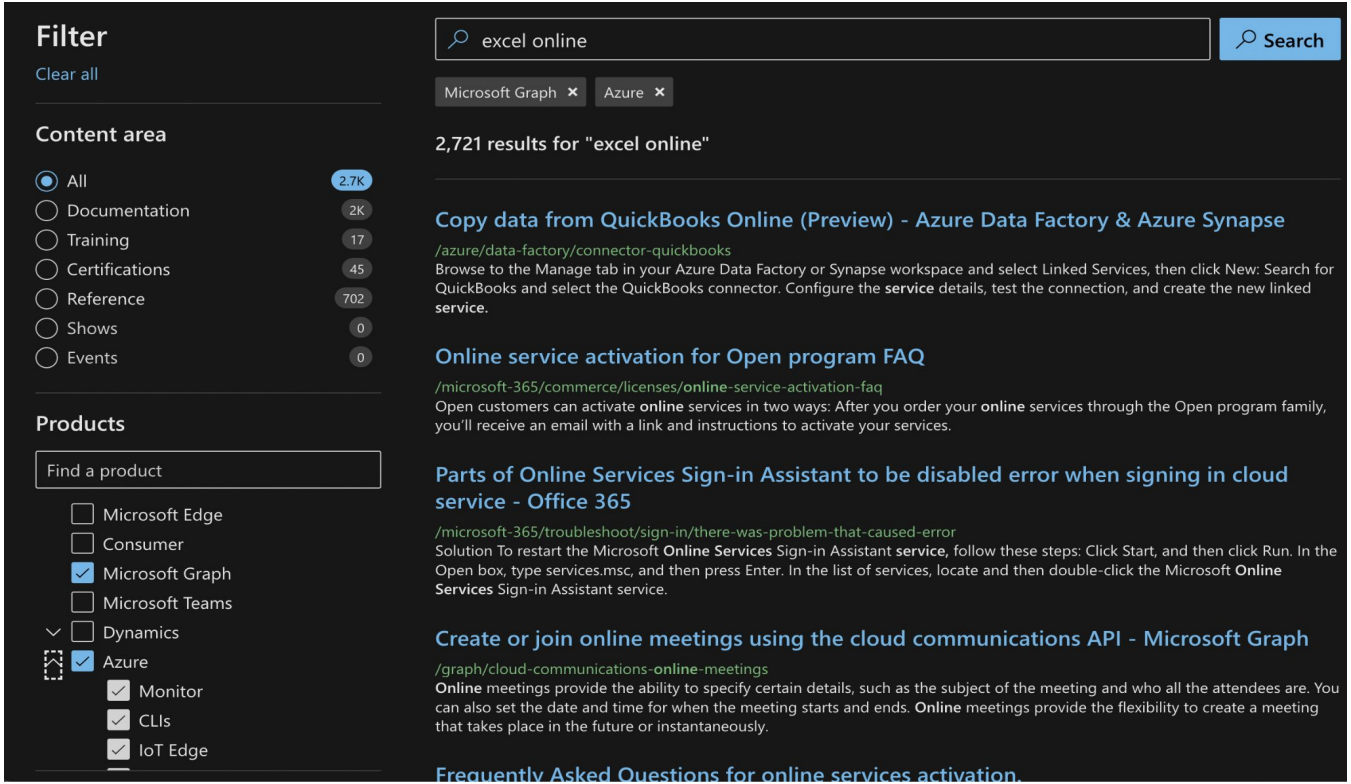
Show AI systems that can mine DOCX, PPTX etc. ausing ML to fulfill the requirements of EU's ethical guidelines for AI.

Why Taxonomies and Tagging

Example Use Case

Organization	Large multinational technology company
Use case	Developer community needs to find online documentation.
Objective	Consistently and efficiently tag the online documentation
Solution	• Continually update and manage a taxonomy in a taxonomy management system (PoolParty). • Autotag content: indexer/crawler is run each day to pick up any content that has changed. • Implement with search in front end.

Why Taxonomies and Tagging



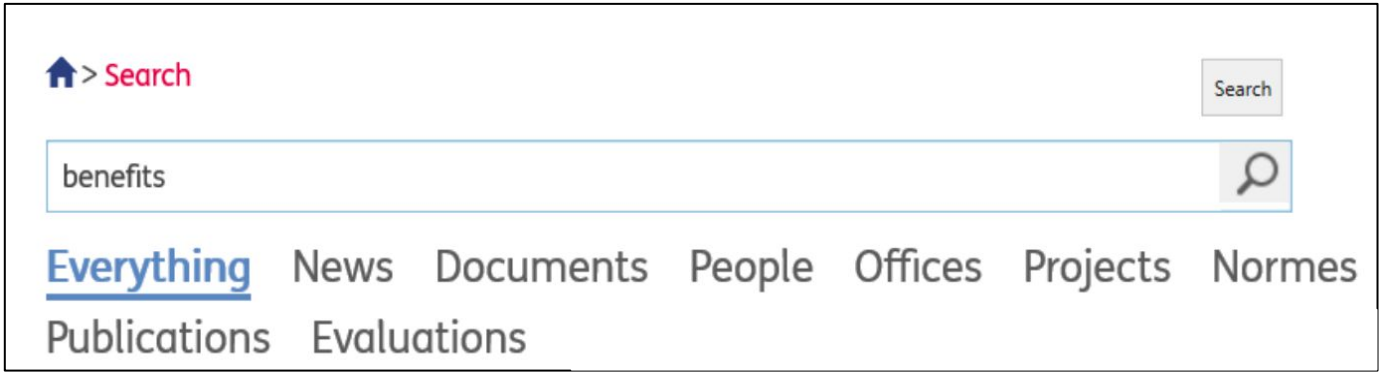
Why Taxonomies and Tagging

Example Use Case with SharePoint

Organization	Green Climate Fund 
Use case	Managing over 100,000 documents in different subject domains in numerous SharePoint libraries to support internal knowledge management
Objective	Enhancement of search experience
Solution	• Manage a taxonomy in a taxonomy management system (PoolParty) • Sync taxonomy with SharePoint Term Store and support NLP-based auto tagging (PoolParty PowerTagging for SharePoint)

Why Taxonomies and Tagging: Search Issues

Search (in SharePoint or in general) is good, but not perfect.



- What are the issues?

Why Taxonomies and Tagging: Search Issues

User Objectives

- Finding the right information quickly
- Finding sufficient information
- Discovering related relevant information not known to search for
- Being confident that the information is complete
- Being able to expand and limit the search
- Having results that can be trusted and are explainable

User Expectations

- Getting search results as comprehensive as web search results

Why Taxonomies and Tagging: Search Issues

Organization Objectives

- Saved time through efficient search
- Knowledge discovery
- Cost savings through content reuse
- Better decisions from complete information
- Improved employee and customer satisfaction
- Improved competitiveness through better information

Why Taxonomies and Tagging: Search Issues

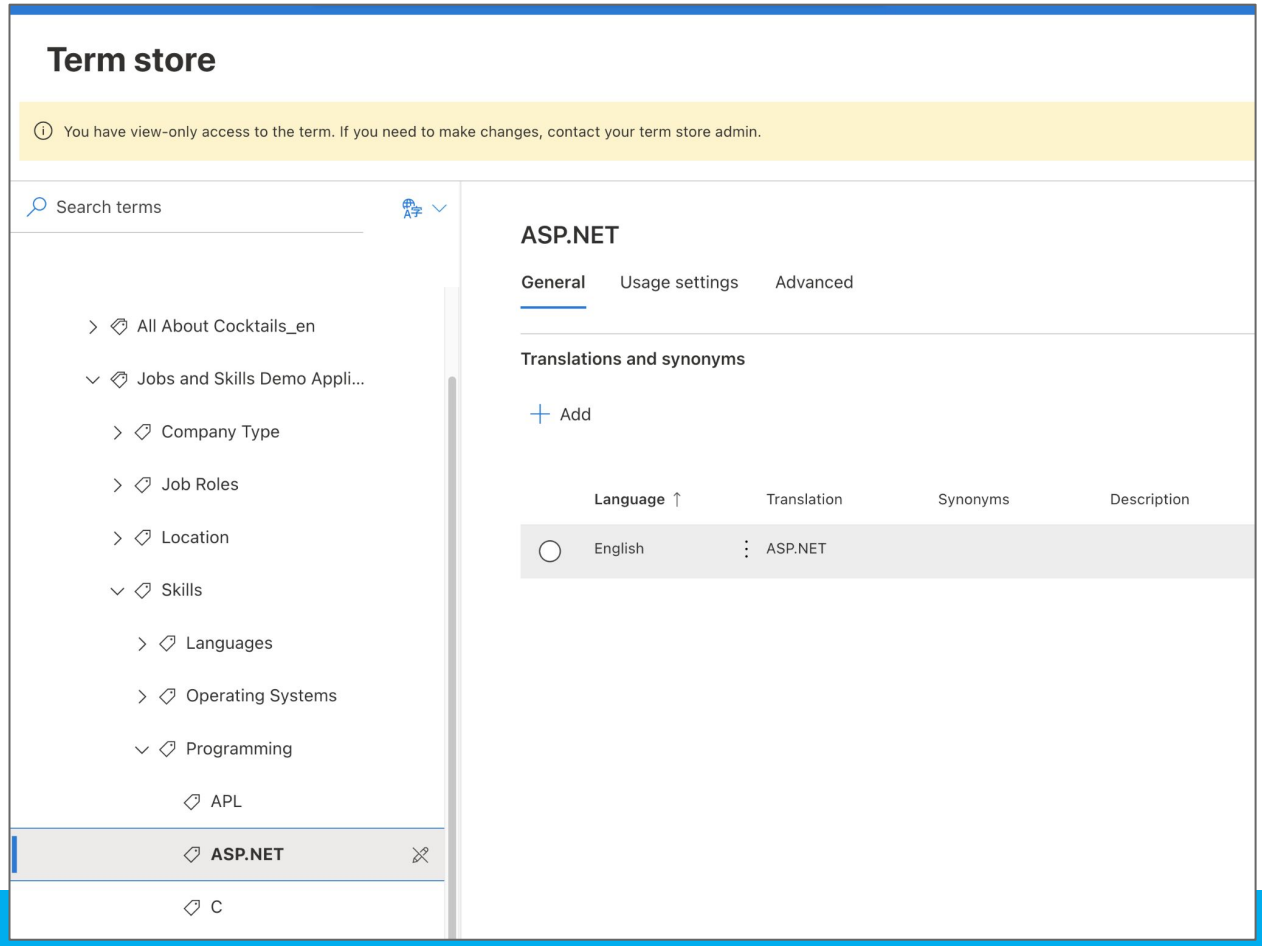
Internal Search User Pain Points

- Retrieving false search results
  - Wasting time reading through irrelevant information
- Not confident that that *all* relevant information was retrieved
  - Not finding content believed to exist
  - Not knowing *if* the desired information exists
- Repeating searches in different, unconnected content repositories

Benefits of Managed Metadata/ Taxonomies

Solution: Metadata & Taxonomies

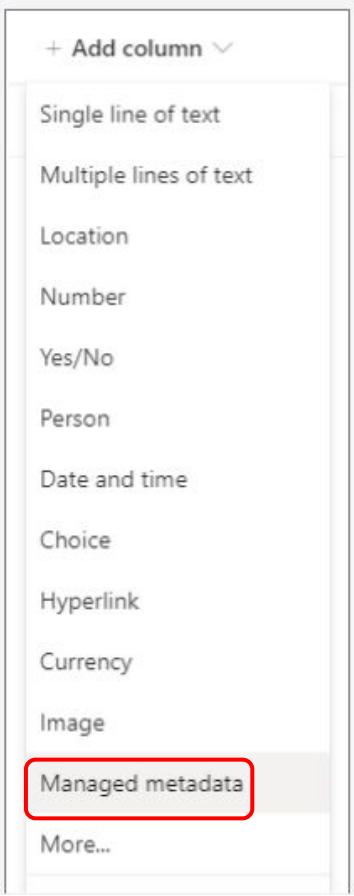
SharePoint Term store for managing metadata/ taxonomies



Solution: Metadata & Taxonomies

In general	In SharePoint	Description
Metadata	Column type	Organized data about data/content, grouped into consistent types/elements/ properties/fields/columns, filled with a specific value for each content item.
Taxonomy	Managed metadata (what is managed in the Term Store)	Metadata type that uses a controlled vocabulary and is structured into subsets and/or hierarchies

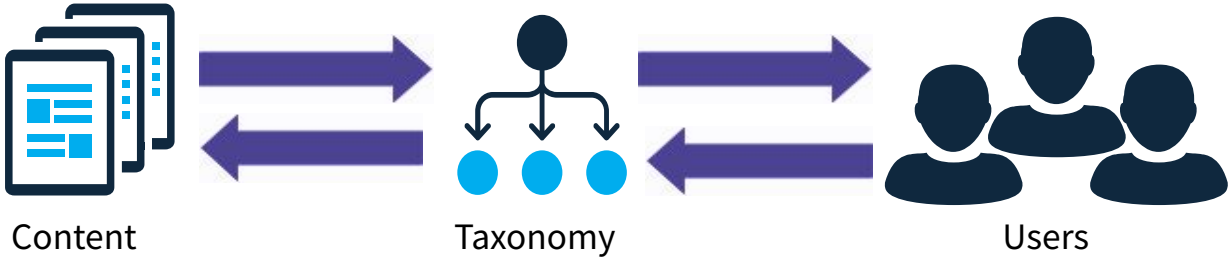
A taxonomy is a source of values for a metadata type (e.g. “managed metadata”).



Solution: Metadata & Taxonomies

What and why taxonomies?

- A taxonomy is a controlled vocabulary organized into a hierarchical structure.
- Concepts/terms are used to tag/index/categorize content to make it easier to be found and retrieved
  - supporting better findability than search alone
- The taxonomy is an intermediary that links the user to the desired content.



- Terms are based on unambiguous concepts, rather than using text strings.
- “Things, not strings.”

Solution: Metadata & Taxonomies

Benefits of taxonomies

- Providing consistent terms for tagging
- Bringing together synonyms for the same concept
- Organizing terms in to refinement filters
- Organizing terms into guiding hierarchies
- Supporting term definitions and
- Supporting multilingual concepts
- Potential to link to other taxonomies and thus other content/data

Benefits of Taxonomies

Providing consistent terms for tagging and retrieval

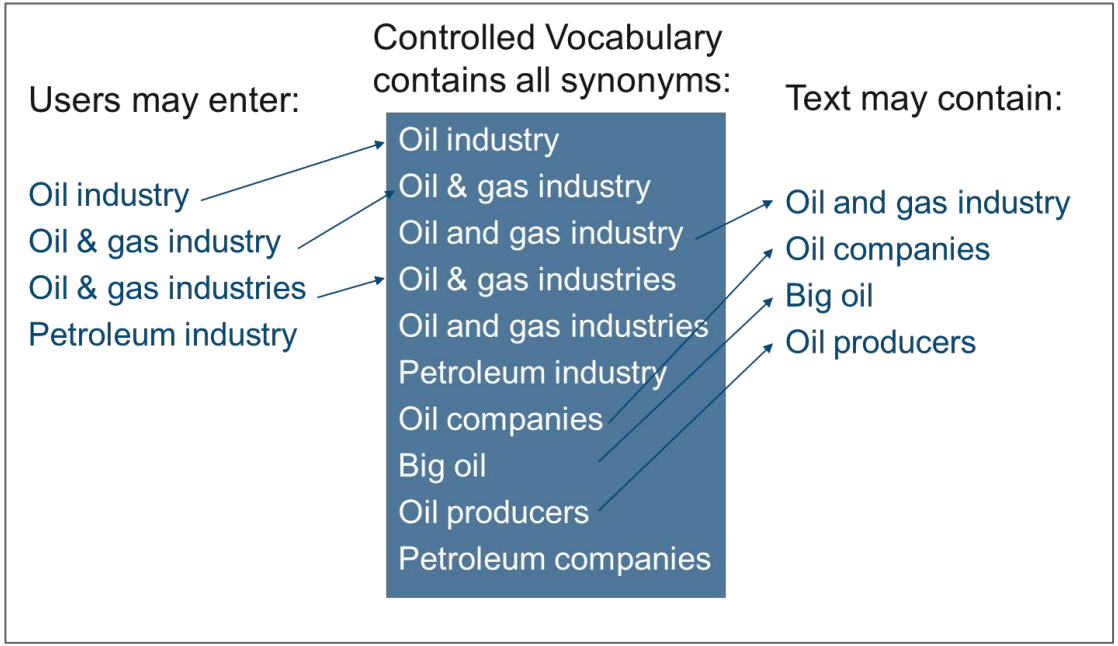
Potential inconsistent tagging for the same document without a controlled vocabulary:

Author	Doc Type	Subject	Source	Region
John Smith	Article	Financial	Finance	U.S.A
J. Smith	Report	Finance	Finance Dept.	United States
John A. Smith	Document	Corporate financials	XYZ Company	New York
Smith, John	Summary	Profit/Loss	(unknown)	NY

Benefits of Taxonomies

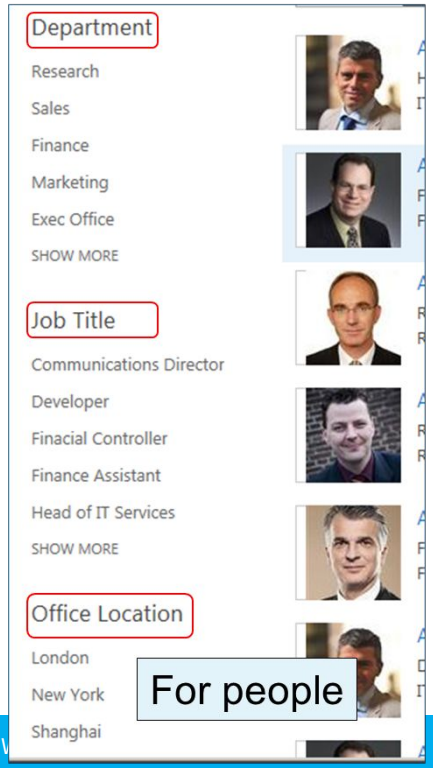
Bringing together synonyms for the same term:

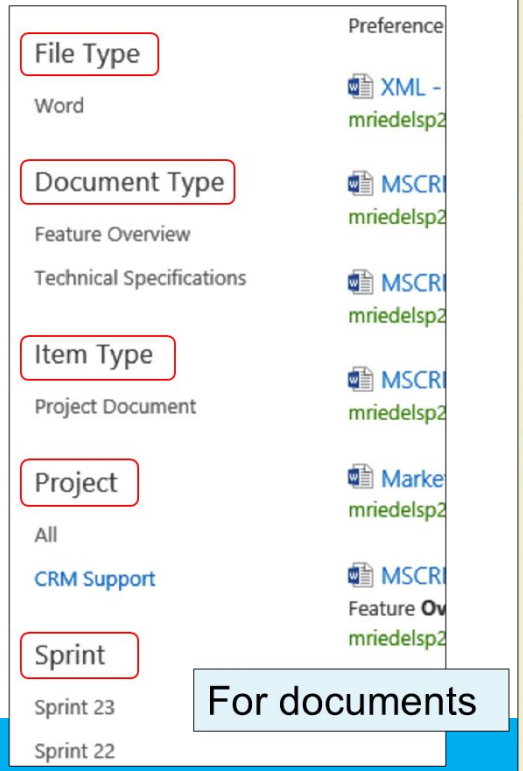
- Retrieves content containing different synonyms for the same thing.
- Supports searchers using different synonymous search strings find the same relevant content.

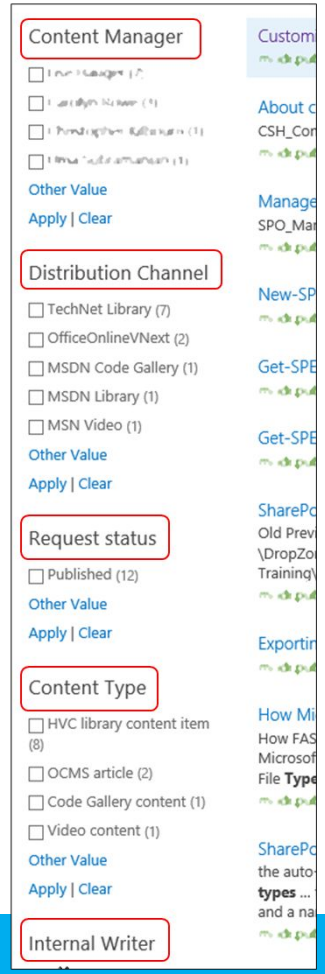


Benefits of Taxonomies

Organizing terms into search refinement filters

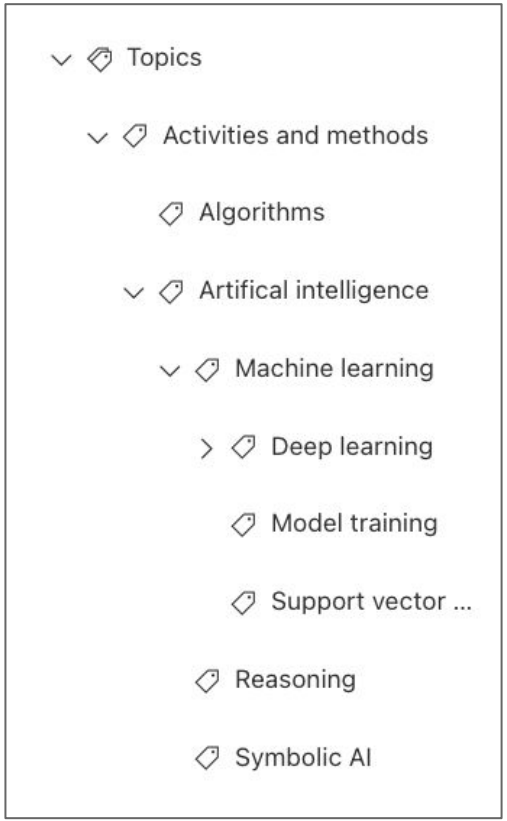


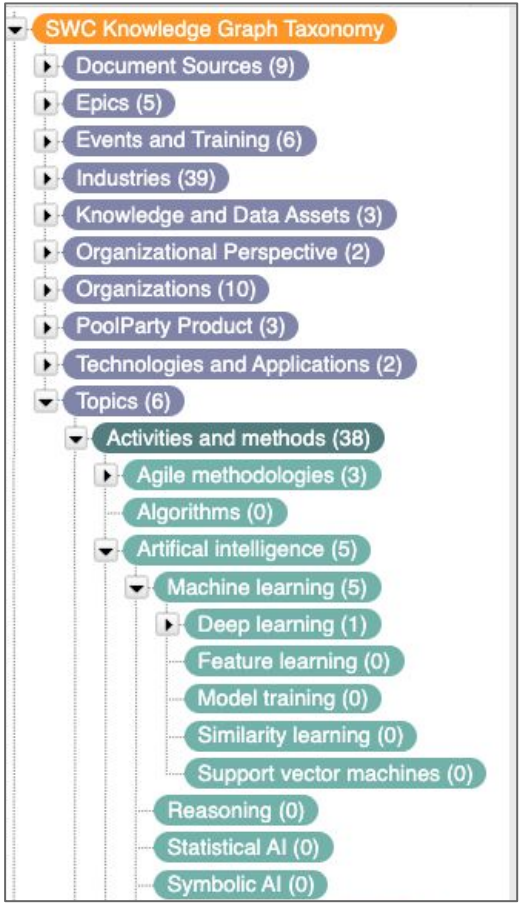




Benefits of Taxonomies

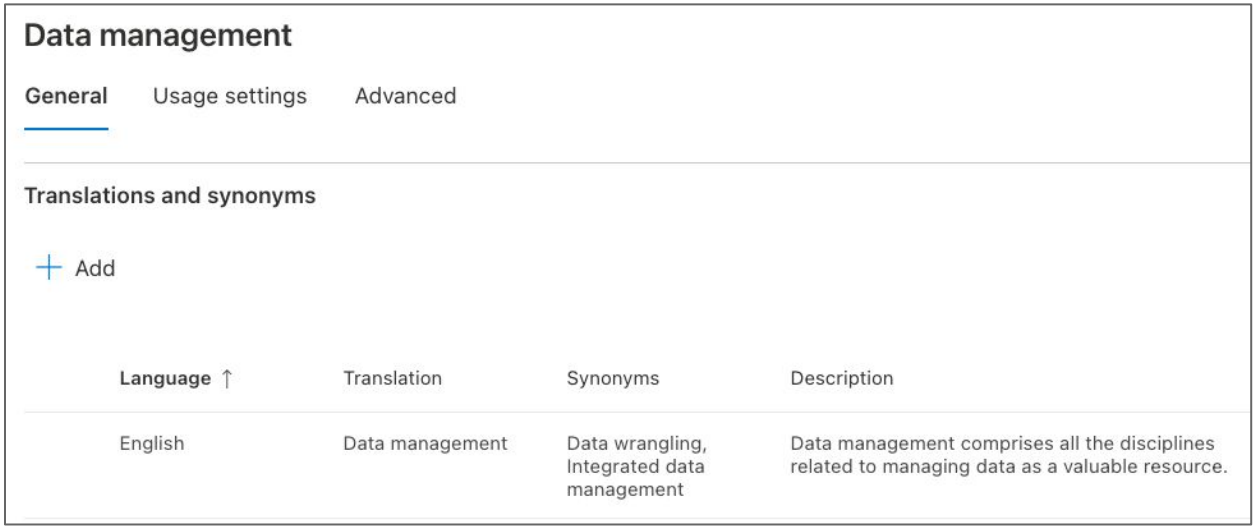
Organizing terms into hierarchies to guide users to to specific topics





Benefits of Taxonomies

Supporting term definitions/notes for clarification



Taxonomy Use in SharePoint

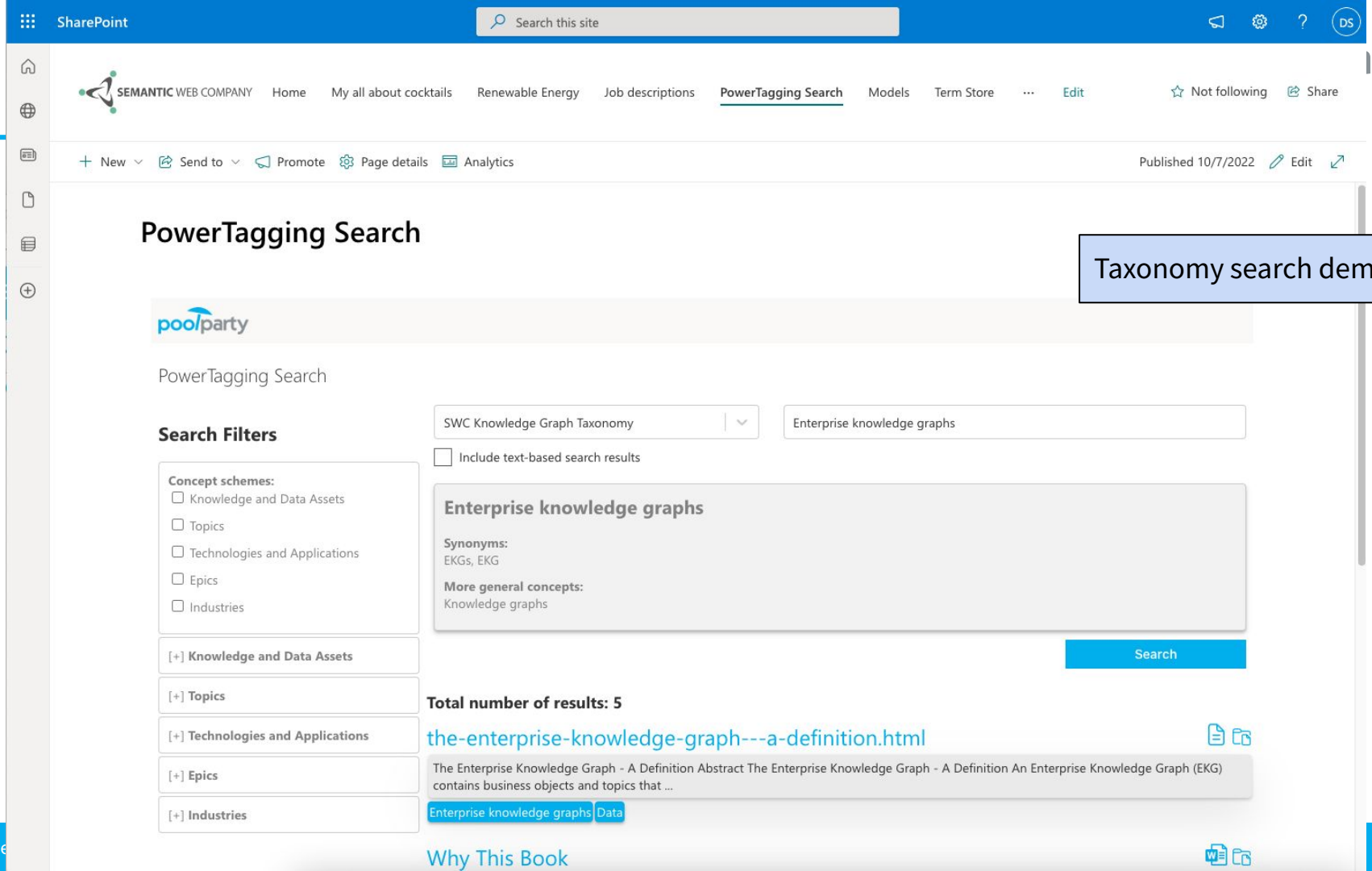
SharePoint can make use of taxonomies various ways

1. As facets or post-search refinement/filters
2. Metadata terms for filtering or sorting columns of content items (documents, images, etc.) in document libraries
3. Implemented in navigation as menu labels, headers, and page URL names
4. Search support (with synonyms), along with added technologies
5. As a hierarchical model for naming and organizing libraries and folders

SharePoint Term Store Capabilities

Taxonomy management in SharePoint Term Store features

- Multiple taxonomies or multiple facets
- Unique identifiers on terms
- Hierarchical taxonomies, many levels deep
- Synonyms, also called “Other labels,” in support of tagging
- Scope notes or definitions in a “Description” field
- Type-ahead search on taxonomy terms and on synonyms for those tagging
- Can create broader categories, which are not terms for tagging, to group terms
- Customizable sort order of terms at the same level


SharePoint
Search this site
DS
SEMANTIC WEB COMPANY
Home
My all about cocktails
Renewable Energy
Job descriptions
PowerTagging Search
Models
Term Store
Edit
Not following
Share
New
Send to
Promote
Page details
Analytics
Published 10/7/2022
Edit
PowerTagging Search
Taxonomy search demo
poolparty
PowerTagging Search
Search Filters
SWC Knowledge Graph Taxonomy
Enterprise knowledge graphs
Include text-based search results
Concept schemes:
Knowledge and Data Assets
Topics
Technologies and Applications
Epics
Industries
Enterprise knowledge graphs
Synonyms:
EKGs, EKG
More general concepts:
Knowledge graphs
Search
[+] Knowledge and Data Assets
[+] Topics
[+] Technologies and Applications
[+] Epics
[+] Industries
Total number of results: 5
the-enterprise-knowledge-graph---a-definition.html
The Enterprise Knowledge Graph - A Definition Abstract The Enterprise Knowledge Graph - A Definition An Enterprise Knowledge Graph (EKG) contains business objects and topics that ...
Enterprise knowledge graphs
Data
Why This Book

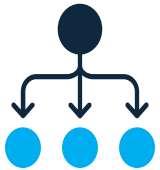
Challenges and Enhancements with Taxonomies

Challenges with Metadata & Taxonomies

1. Poorly designed and managed metadata and taxonomies
2. Poorly applied metadata or taxonomy tagging
3. Taxonomies and metadata residing within separate applications, siloed

Challenges with Metadata & Taxonomies

1. Poorly designed and managed metadata and taxonomies
   - Metadata or taxonomy terms vague or not clearly distinguished
   - Taxonomies too specific or too general
   - Too many, or too few metadata columns or term sets
   - Incorrect hierarchies
   - Too many or too few synonyms for terms



- Addressed by:
  - Elevate taxonomy development to a funded project
  - Obtain taxonomy training and/or contract a taxonomy consultant
  - Manage taxonomies in a full-featured taxonomy management system, connected to your applications(s)

Challenges with Metadata & Taxonomies

2. Poorly applied metadata or taxonomy tagging
   - Insufficient tagging - content cannot be found
   - Over-tagging - taxonomy terms retrieve too many irrelevant results
   - Inconsistent tagging - search results aren't comprehensive;

- Addressed by Auto-tagging
  - For more comprehensive, consistent, and faster tagging
  - Semantic tagging tags terms and not just words
  - Use technologies: text mining, NLP, NER
  - Auto-tagging with human review to identify improvements

Challenges with Metadata & Taxonomies

3. Taxonomies and metadata residing within separate applications, siloed
   - Separate taxonomies for each CMS or other application
   - Taxonomies managed by different groups
   - Lack of central taxonomy governance

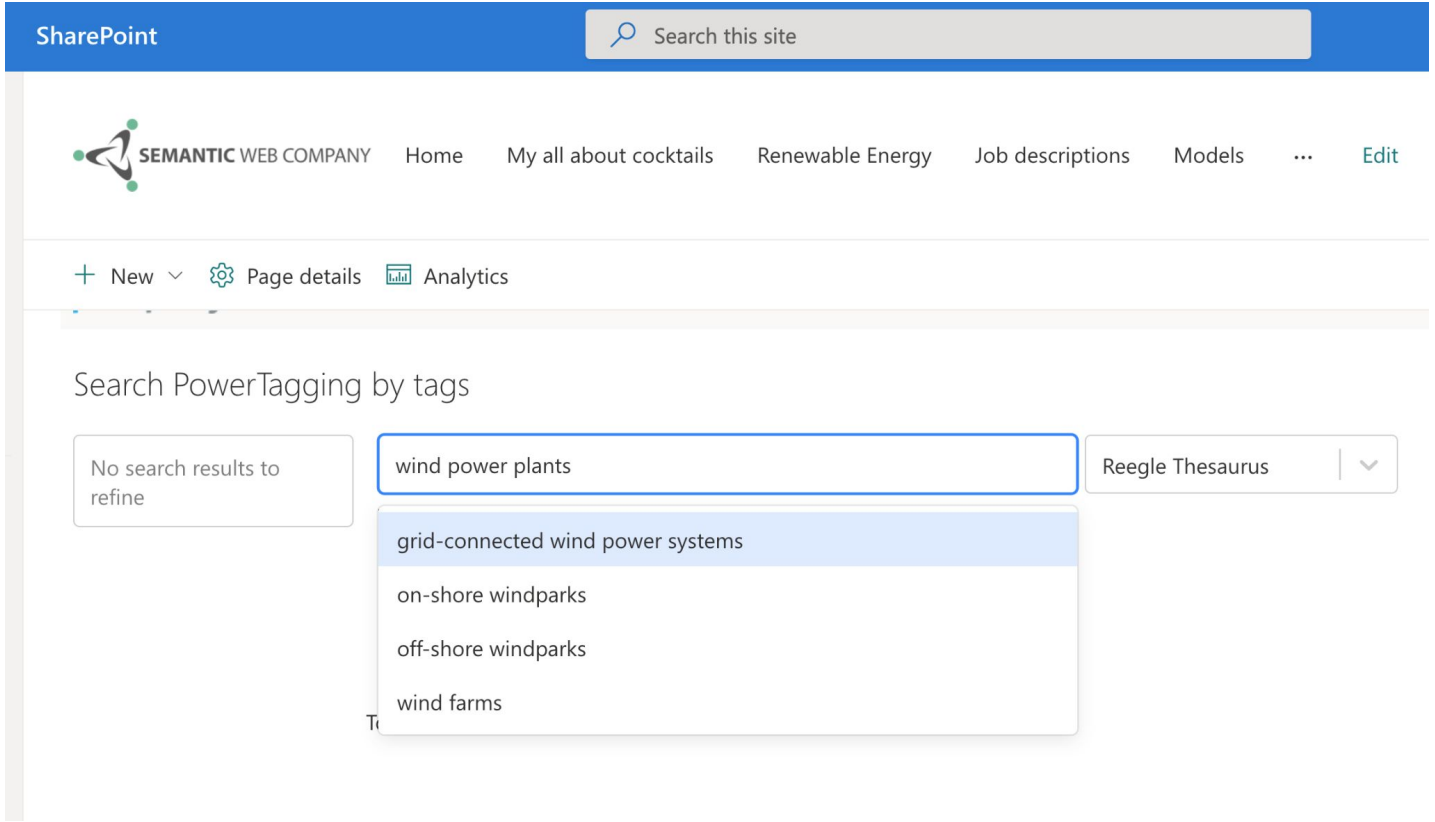
- Addressed by:
  - Centrally managing taxonomy and metadata in dedicated taxonomy management system, as middleware, connected to other systems
  - Documentation and creating and following a taxonomy governance plan to manage taxonomy change

Addressing Taxonomy Design Issues

Selected Taxonomy Design Tips

- Granularity: Create terms that are common enough to tag multiple documents, but specific enough to be useful and not tagged to too many documents.
- Synonyms: Create sufficient but not excessive alternative labels (synonyms), to support comprehensive but not incorrect excessive tagging.
- Structure: Group taxonomy terms into term sets for facets/refinements that are useful and intuitive for refining searches, based on user research.
- Standards: Balance following thesaurus standards with providing a good *customized* user experience, especially with hierarchy.

Addressing Taxonomy Design Issues



Addressing Taxonomy Design Issues

Taxonomy Governance

Taxonomies need to adapt to growth and change of terms.

- Plan: Begin addressing governance issues from the start of creating a taxonomy
- Scope: Identify all the kinds of changes, and policies for each kind
- Roles and Responsibilities: Identify *who* makes and approves changes
- Policies and Procedures: Determine *how* changes are suggested and approved
- Documentation: Document the taxonomy and governance policies
  - Taxonomy description
  - Taxonomy updating style guige
  - Tagging policies

Auto-Tagging:
PoolParty PowerTagging

Tagging with a Taxonomy

- Auto-tagging improves comprehensive, consistent, and faster tagging.
- Use semantic tagging to tag terms and not just words
- Use technologies: text mining, NLP, NER
- Implement auto-tagging with human review to identify improvements

Auto-Tagging

Term/entity extraction - tagging example

Work From Home During the COVID-19 Outbreak

The Impact on Employees' Remote Work Productivity, Engagement, and Stress

Objective:

The COVID-19 pandemic made working from home (WFH) the new way of working. This study investigates the impact that family-work conflict, social isolation, distracting environment, job autonomy, and self-leadership have on employees' productivity, work engagement, and stress experienced when WFH during the pandemic.

Methods:

This cross-sectional study analyzed data collected through an online questionnaire completed by 209 employees WFH during the pandemic. The assumptions were tested using hierarchical linear regression.

Results:

Employees' family-work conflict and social isolation were negatively related, while self-leadership and autonomy were positively related, to WFH productivity and WFH engagement. Family-work conflict and social isolation were negatively related to WFH stress, which was not affected by autonomy and self-leadership.

[Demo: PoolParty Auto-Classification]

Auto-Tagging

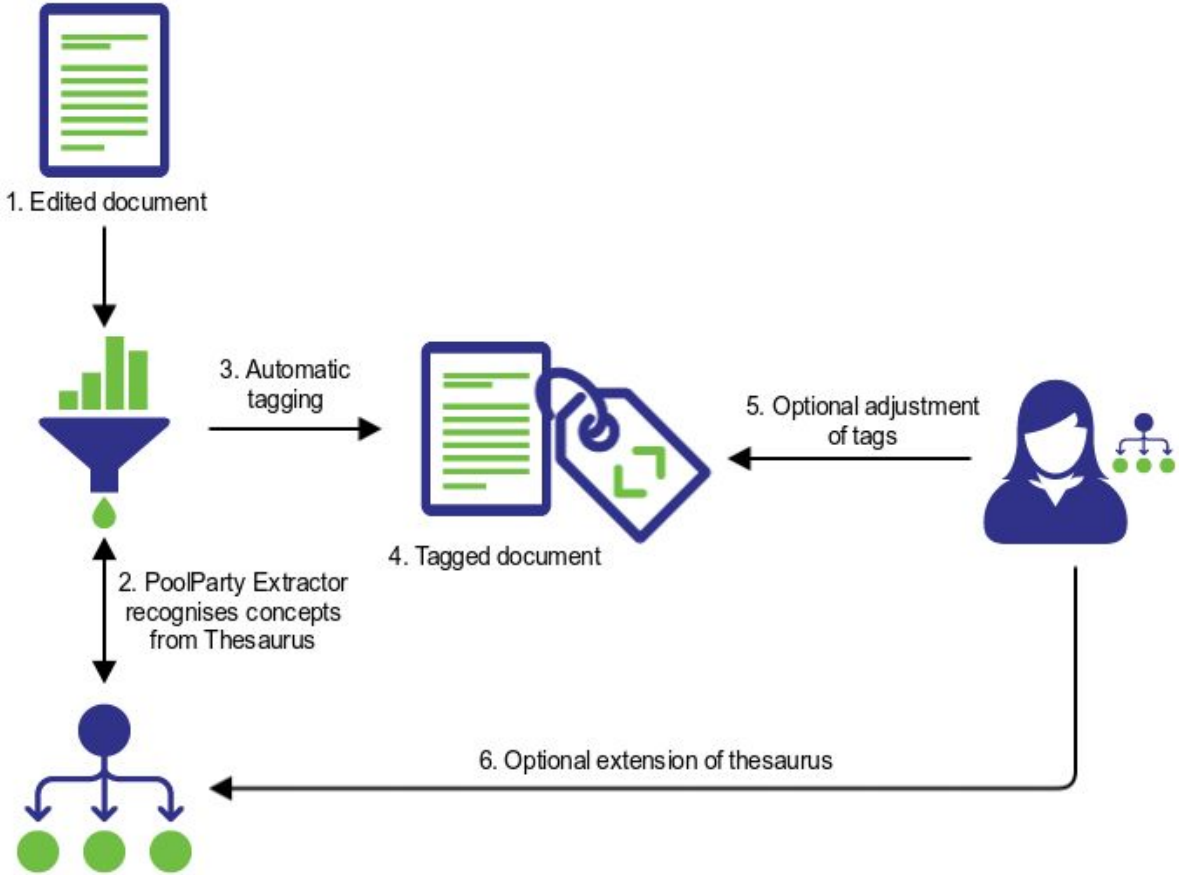
Auto-tagging components in PoolParty

- **Concepts**, extracted through Concept extraction based on PoolParty Extractor and taxonomies.
- **Shadow Concepts**, to extract concepts that are not explicitly mentioned in a document, but are nevertheless implicitly relevant to it.
- **Named Entities**, extracted through NER (Named Entity Recognition) based on machine learning algorithms.
- **Important terms**, identified through statistical text analysis.
- **Document genre**, identified by ML-based classification algorithms.

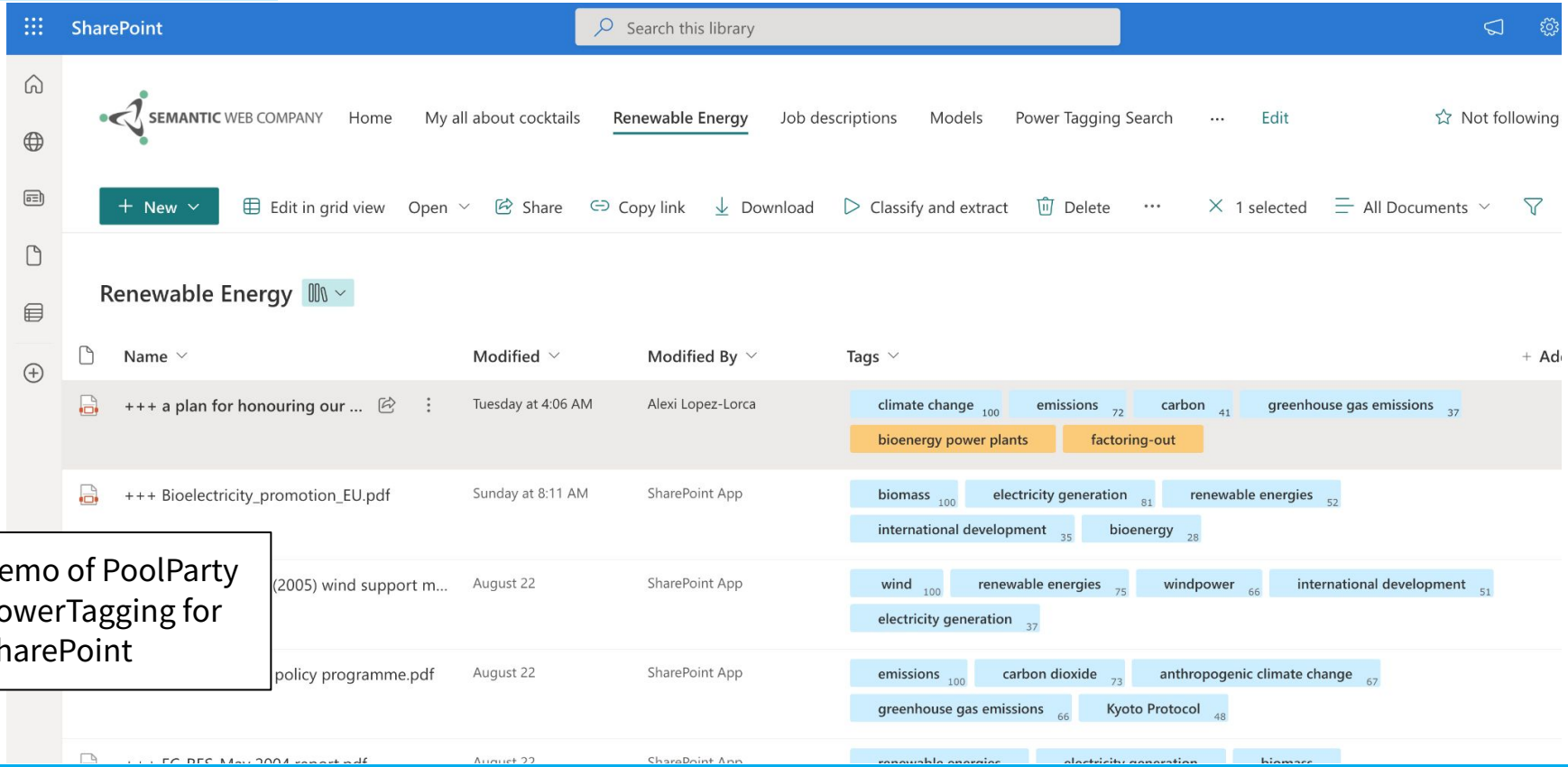
Auto-tagging configuration:

- Concepts can be configured for **Exact label** match only (including capitalization)
- **Blacklist** of concepts and terms *not* to be tagged
- **Disambiguation** based on graph and context
- **Negation rules**

Auto-Tagging with PoolParty



Tagging with a Taxonomy



Demo of PoolParty PowerTagging for SharePoint

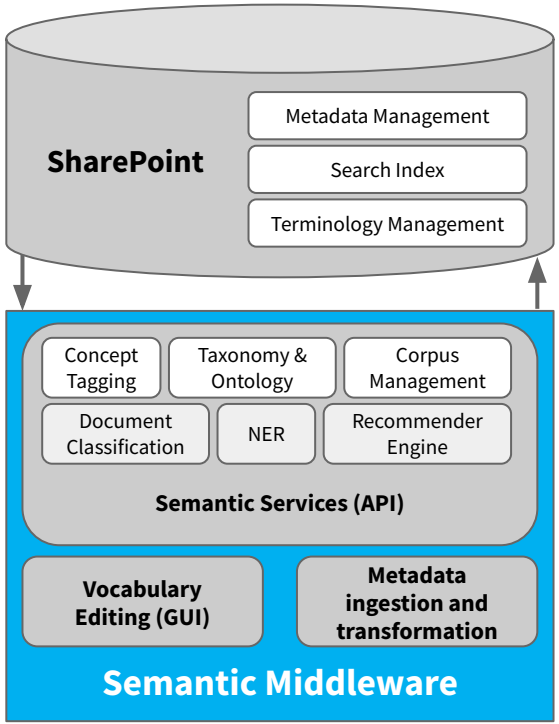
Centrally Managed Taxonomy in Middleware (PoolParty)

Centrally Managed Taxonomies in Semantic Middleware

We have SharePoint, why do we need a Semantic Middleware?

Semantic Middleware (such as PoolParty) offers more services

Enhancing SharePoint and Office 365, with auto-tagging, rich taxonomy and ontology, corpus management, classification, better taxonomy management UI.



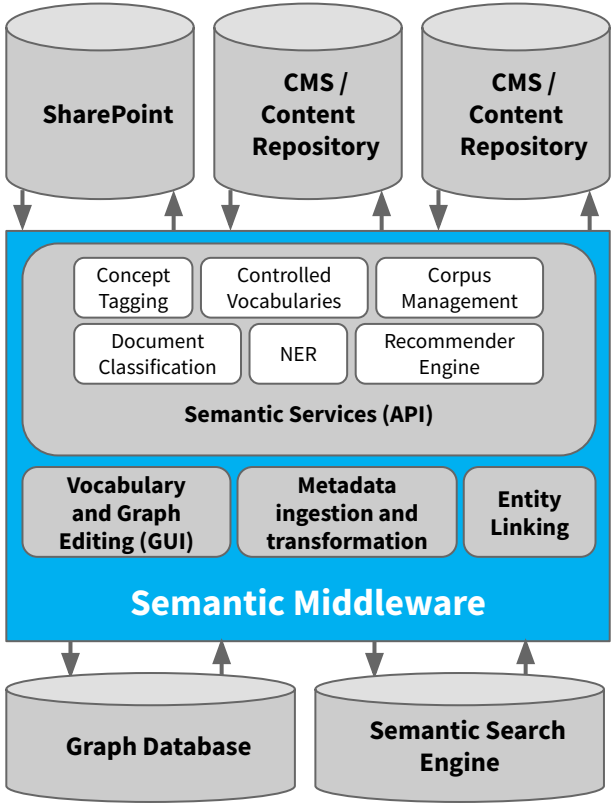
Centrally Managed Taxonomies in Semantic Middleware

We have several CMSs, why do we need a Semantic Middleware?

The multi-source scenario

Semantic Middleware spans an umbrella over existing content management systems, including SharePoint, to link them together.

The meaning of all content becomes interpretable independent of the inherent logic of individual content silos.
360-degree views will become possible.

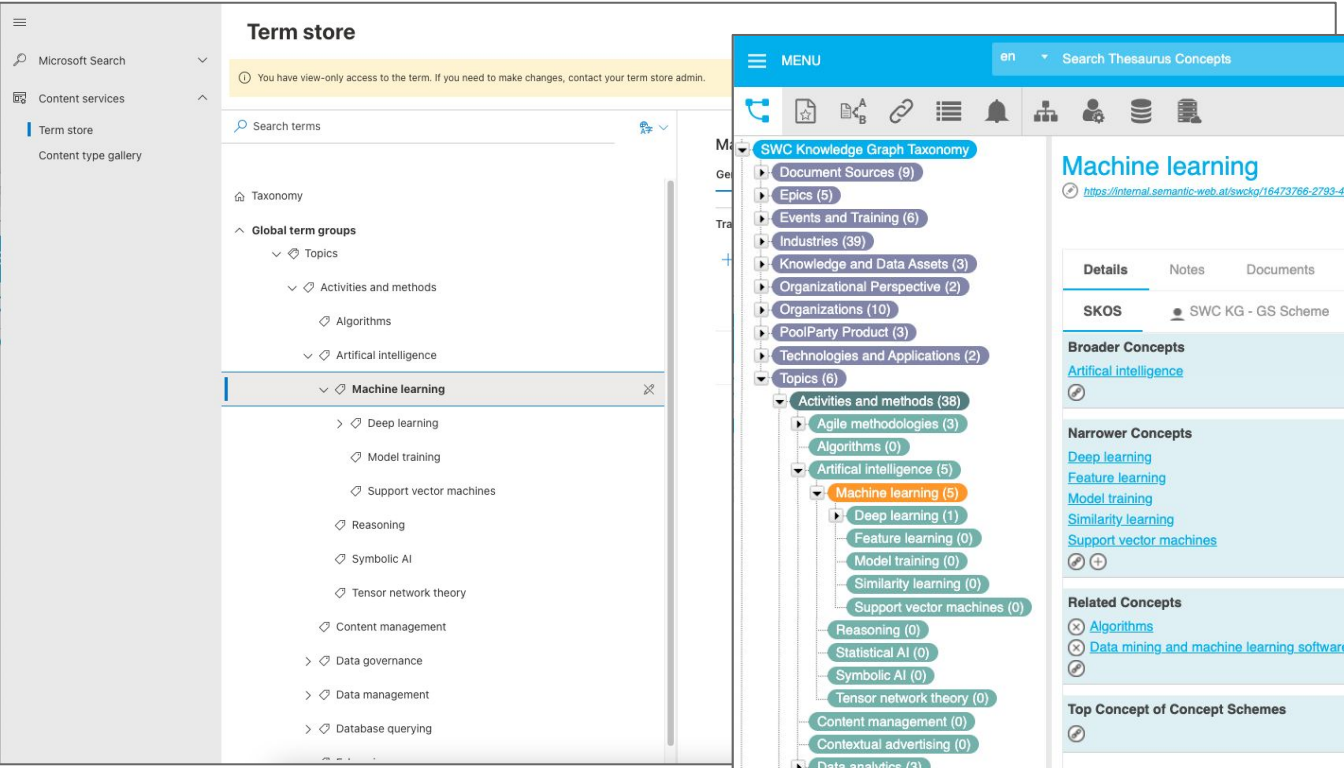


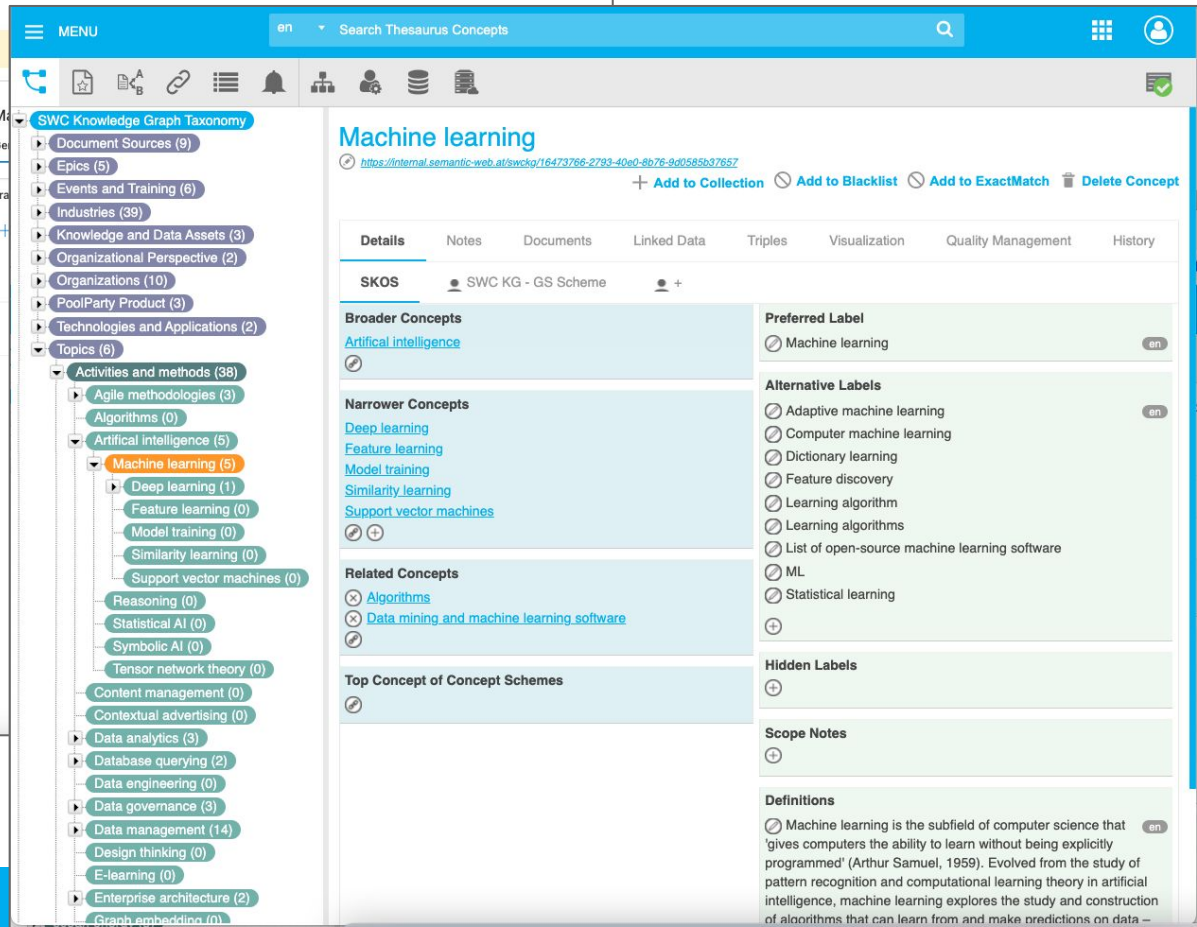
Centrally Managed Taxonomies in Semantic Middleware

What can be built on top of SharePoint to extend Term Store functionality in dedicated taxonomy management middleware:

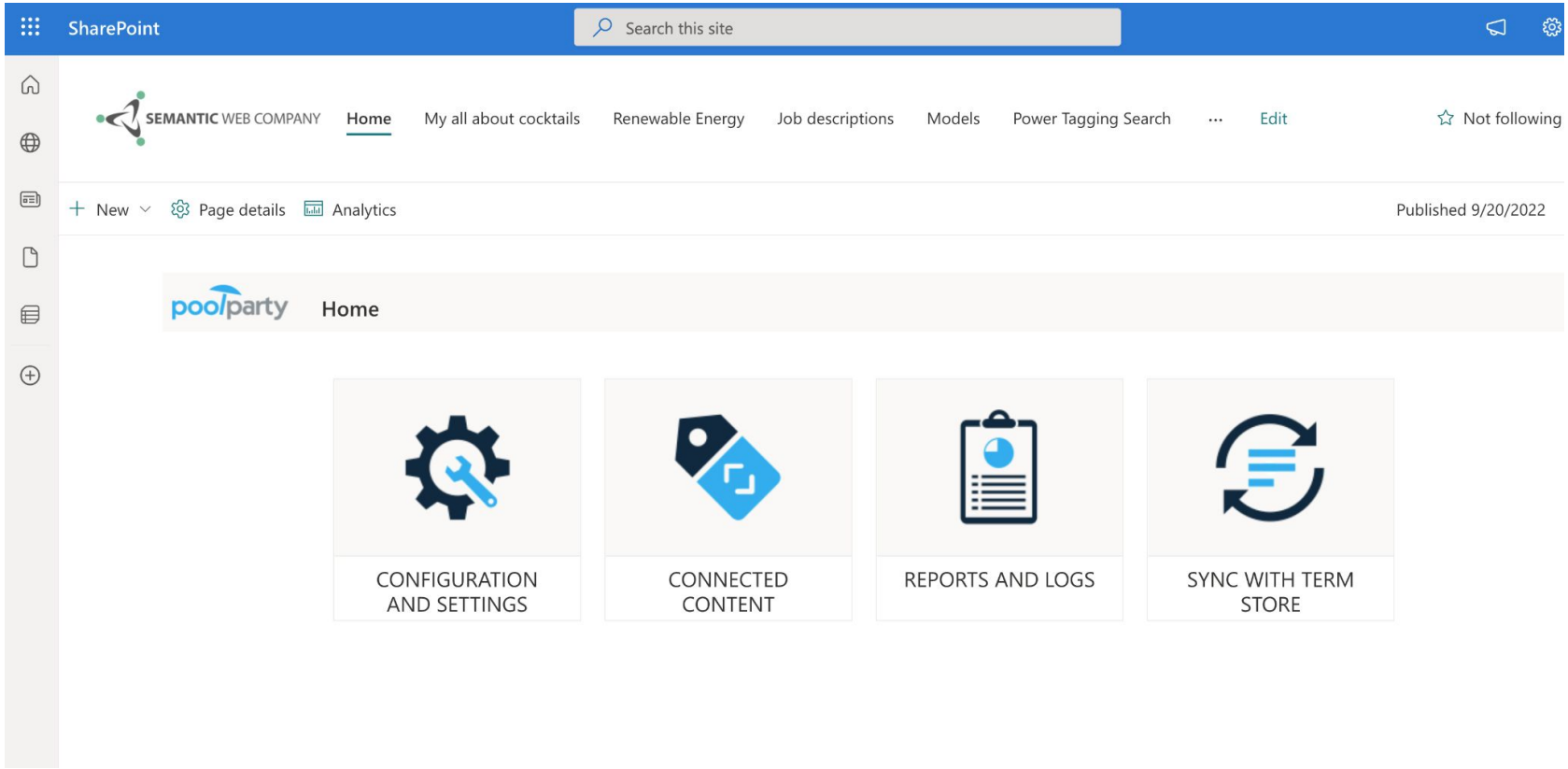
- Synonyms supported in search in addition to tagging
- Supporting thesaurus quality standards (preventing duplicates, circular references, same terms as both preferred and synonyms)
- Managing multiple linked taxonomies
- Supporting associative (non-hierarchical relationships)
- Support for polyhierarchy
- Supporting additional attributes for terms
- Full compliance with W3C standards for vocabularies (SKOS) for import/export interoperability
- Link to content tagged in different systems (other than SharePoint), with a central frontend
- Optimized taxonomy management UI

Centrally Managed Taxonomies in Semantic Middleware





Centrally Managed Taxonomies in Semantic Middleware



Resources

“Optimizing Your Taxonomy in SharePoint Online: Search Filters” Mergan Salerno, Enterprise Knowledge White Paper (September 2022)
https://enterprise-knowledge.com/wp-content/uploads/2022/09/Optimizing_Your_Taxonomy_in_SharePoint_Online_Search_Filters.pdf

“A beginner’s guide to SharePoint metadata,” Matt Wade (November 2019)
https://medium.com/jumpto365/a-beginners-guide-to-sharepoint-metadata-cec1e989272b

“Taxonomies in SharePoint,” Heather Hedden, Accidental Taxonomist Blog (August, 2017)
http://accidental-taxonomist.blogspot.com/2017/08/taxonomies-in-sharepoint.html

The Accidental Taxonomist, 3rd edition, Heather Hedden (fall 2022)
https://accidental-taxonomist.com

Questions/Contact

Heather Hedden
Data and Knowledge Engineer
Semantic Web Company Inc.
One Boston Place, Suite 2600
Boston, MA 02108
857-400-0183
heather.hedden@semantic-web.com
www.linkedin.com/in/hedden

Visit our PoolParty booth at NACS

Semantic Web Company www.semantic-web.com

PoolParty software www.poolparty.biz

THANK YOU TO OUR SPONSORS

DIAMOND LEVEL

tyGraph

Kudzu

Microsoft

PLATINUM LEVEL

AvePoint

WEBCON

backupify
a datto company

Quest

VEEAM

LightningTools

GOLD LEVEL

ScriptRunner®
The #1 for PowerShell Management

poolparty

ENOW