

An Overview of Taxonomies and AI

Heather Hedden
Senior Consultant, Taxonomy and Ontology
Enterprise Knowledge, LLC

CILIP K&IM Group
Artificial Intelligence Webinar Series: The promise and the perils
30 January 2024

Outline



Introduction
to Taxonomies

AI for Tagging
and Classification
with Taxonomies

AI for
Taxonomy
Development

AI for
Taxonomy
Analysis and
Improvement

Taxonomies for
Supporting AI
Applications

ENTERPRISE KNOWLEDGE

Introduction: What is a Taxonomy?

A knowledge organization system that is...

1. **Controlled:**

A kind of controlled vocabulary, based on unambiguous concepts, not just words (*things*, not *strings*).

2. **Organized:**

Concepts are organized in a structure of hierarchies, categories, or facets to make them easier to find and understand.

Controlled

Preferred Label
 Organizational skills en

Alternative Labels
 Self-organization en
 Working efficiently
+

Hidden Labels
 Organisational skills en
+

Scope Notes
 Organizing one's own tasks, not organizing others. en
+

Definitions
 The ability to use time and resources efficiently and effectively, including time, energy, and physical workspace so as to accomplish tasks successfully. en
+

Organized

Training Content

- ▶ Content Type (3)
- ▶ Level (3)
- ▶ Role (7)
- ▼ Skill (4)
 - ▶ Business skills (9)
 - ▶ Languages (11)
 - ▼ Soft skills (4)
 - ▶ Communication skills (25)
 - ▶ Observation skills (4)
 - ▼ Organizational skills (7)
 - ▶ Coordination (0)
 - ▶ Fast context switching (0)
 - ▶ Personal time management (0)
 - ▶ Prioritization (0)
 - ▶ Punctuality (0)
 - ▶ Structured thinking (0)
 - ▶ Working independently (0)
 - ▶ Social skills (5)
 - ▶ Technical skills (7)
- ▶ Topic (6)
- ▶ Training Program (6)

Introduction: What is a Taxonomy For?

- ◆ Concepts are used to tag/categorize content to make finding and retrieving specific content easier.
 - Supporting better findability than search alone.

A screenshot of a search interface. The search bar contains the text 'data'. Below the search bar, there is a list of search results. Each result consists of a bolded title and a subtitle indicating the context. The results are:

- Metadata**
Context: Topics
- Linked Open Data**
Context: Topics
- Linked Data enrichment**
Context: PoolParty Product
- Output data unit**
Context: PoolParty Product
- Data processing**
Context: PoolParty Product
- Data engineering**
Context: Topics

Taxonomy implementation:
focus on
controlled

A screenshot of a taxonomy tree structure. The root node is 'Topics (8)'. The tree is expanded to show the following hierarchy:

- Standards - 1029
- Activities and methods - 1028
- Data - 1000
 - Metadata - 345
 - Unstructured data - 213
 - Structured data - 117
 - Spreadsheet - 99
 - Relational data - 19
 - Open data - 72
 - Linked Open Data - 52
 - Linked Open Data
 - Master data
 - Data structure
 - Sensor data
 - Streaming data - 7
- Knowledge organization systems - 898
 - Taxonomies - 598
 - Knowledge graphs - 405
 - Ontologies - 318

Taxonomy implementation:
focus on
organized

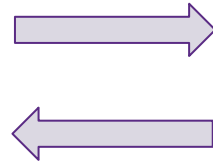
- ◆ The taxonomy is an intermediary that links users to the desired content.

Introduction: How are Taxonomies Created?

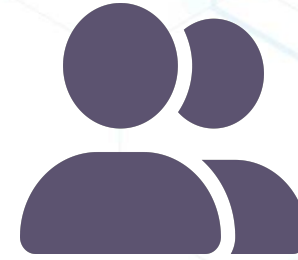
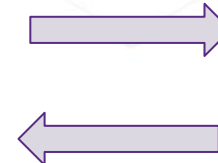
- ◆ The taxonomy needs to be designed to suit both the **users** and the **content**.



Content



Taxonomy



Users



From Bottom Up:

- ◆ Analyze the content
- ◆ Conduct content audit
- ◆ Extract terms from content



From Top Down:

- ◆ Interview stakeholders
- ◆ Brainstorm in workshops and focus groups
- ◆ Get users' terms from search logs



AI for Tagging and Classification

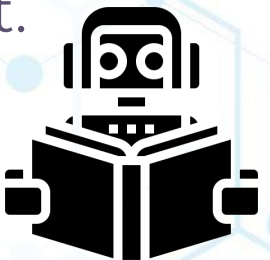
- ◆ Manual tagging and classification has limitations. Not scalable.
- ◆ AI methods, existing since the 1990s, have become more common.
- ◆ Not all “auto-tagging” uses AI. Rules can also be written to auto-tag.
- ◆ Human review remains a feature.



AI-based auto-tagging and auto-classification makes matches based on a linguistic, logical, or mathematical profile it expects and recognizes.

Technologies Include:

- ◆ **Machine learning (ML)** and **deep neural networks**: Use mathematical analysis to find patterns that match known properties (text or images).
- ◆ **Named entity recognition**: Matches proper nouns mentioned in text.
- ◆ **Semantic analysis**: Locates concepts referenced within the content.
- ◆ **Natural language processing (NLP)**: Analyzes sentences.



AI for Taxonomy Development

- ◆ Technologies for auto-tagging can also be used for **extracting terms** as candidate concepts for a taxonomy.
- ◆ **Machine learning** and **NLP** enable term extraction from a “corpus” of documents.
- ◆ Extracted terms are manually reviewed to be added to the taxonomy.
- ◆ Software exists as stand-alone tools or features of taxonomy management systems.

Extracted Terms					Select All	Deselect All	Add Candidate Concepts	Add to Blacklist	Export Documents
Term	Relevance	CTS	MIS	Frequency					
data	139.8	15	0	3178					
U.S.A	82.22	10.563	0	48					
ERIC	72.24	9.603	0	42					
OKS/OKC	60.24	8.342	0	36					


Creating hierarchical “is a” relationships can also be automated. However...


- ◆ Lacking any input from users (interviews, surveys, etc.), results are not user friendly, but may be used in non-displayed taxonomies, such as for search support/SEO.

Generative AI for Taxonomy Development


Using technologies such as ChatGPT with large language models (LLMs):


- ◆ Request to put a list of terms (such as from term extraction) into categories.
- ◆ Request suggested narrower concepts.
- ◆ Request alternative labels (synonyms), including for certain audiences.

 **You**
Provide several narrower concepts for the concept of health care.

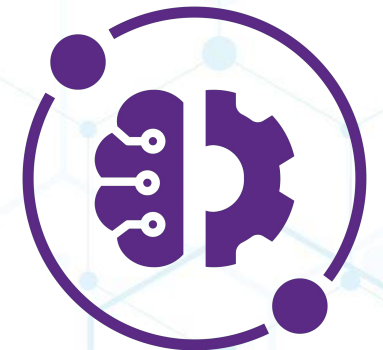
 **ChatGPT**
Certainly! Here are several narrower concepts related to the broader concept of "health care":

- 1. Primary Care:**
 - Family Medicine
 - Internal Medicine
 - Pediatrics
 - General Practice
- 2. Specialized Medical Services:**
 - Cardiology
 - Dermatology
 - Oncology
 - Neurology
- 3. Preventive Health Services:**
 - Vaccinations
 - Screenings (e.g., cancer screenings, blood pressure checks)

 **You**
Provide a list of synonyms for health care.

 **ChatGPT**
Sure, here is a list of synonyms for "health care":

1. Healthcare
2. Medical care
3. Medical treatment
4. Health services
5. Wellness services
6. Health maintenance
7. Healthcare delivery



Do not ask ChatGPT to “create” a taxonomy. It might violate copyrights.

AI for Taxonomy Analysis and Improvement

ML, NLP, and Entity Recognition: Corpus Analysis

- ◆ Rather than extract new, candidate terms, identify taxonomy concepts mentioned in the content.
- ◆ Similar technology to auto-tagging, but may use other, similar content.
- ◆ Identify high and low frequency occurring concepts.

The screenshot displays a web interface for document analysis. At the top, there are filter tabs: 'Concept Schemes' (highlighted with a red box), 'Applications and technologies', 'Knowledge and data assets', 'Organizations', and 'Topics'. Below these are 'Highlights' with checkboxes for 'Concepts' (checked and highlighted with a red box), 'Shadow Concepts', and 'Terms'. The main content area shows a document titled 'A Brief History of Knowledge Graphs.txt' with an 'Export Document' button. The document text discusses the history of knowledge graphs, mentioning key figures like John F. Sowa, C. Hoede, and F.N. Stokman, and concepts like graph theory, RDF, and the Semantic Web. Below the text is a list of extracted tags, each with a count in parentheses, such as 'artificial intelligence', 'cloud (19)', 'context (21)', 'data integration', 'data management', 'DBpedia', 'description (6)', 'enterprise knowledge graph', 'F.N. (6)', 'Gartner', 'Google', 'GQL', 'graph database', 'graph querying', 'graph theory (14)', 'history (14)', 'hoede (6)', 'john (6)', 'Knowledge graph', 'lassila (6)', 'logic (6)', 'machine learning', 'mathematical (100)', 'natural language processing', 'ontology', 'open (22)', 'open-world assumption', 'ora (6)', 'paramount (6)', 'Persona', 'project (15)', 'publish (25)', 'query language (6)', 'RDF', 'reasoning', 'search engine', 'semantic AI', 'Semantic Web', 'SPARQL', 'Standard and directive', 'stokman (6)', 'theory (12)', and 'university (62)'.

AI for Taxonomy Analysis and Improvement

Generative AI: Generating SPARQL Queries

- For various analysis of a taxonomy built on SKOS standard.
- For customized reports, that the taxonomy management software does not provide as a preset option.



You

Create a SPARQL query to identify all the concepts that have more than one broader concept relationships in a SKOS taxonomy.

< 2 / 2 >



ChatGPT

Certainly! To identify all concepts that have more than one broader concept relationship in a SKOS (Simple Knowledge Organization System) taxonomy using SPARQL, you can use the following query:

sparql

Copy code

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT ?concept (COUNT(?broader) AS ?count)
WHERE {
  ?concept skos:broader ?broader.
}
GROUP BY ?concept
HAVING (COUNT(?broader) > 1)
```

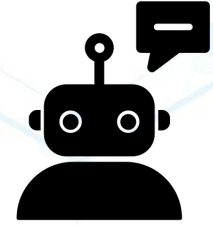
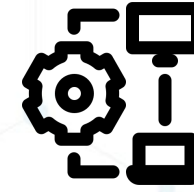
This SPARQL query retrieves concepts and counts the number of broader relationships each concept has. The **HAVING** clause filters out concepts that have only one broader relationship, leaving those with more than one.

Make sure to replace `<http://www.w3.org/2004/02/skos/core#>` with the actual namespace URI if your SKOS vocabulary uses a different one.

Taxonomies Supporting AI Applications

Taxonomies support applications, including:

- Enhanced search and insights
- Recommendation systems
- Natural language question-answering
- Chatbot support
- Generative AI & LLM improved results (with Retrieval Augmented Generation/RAG)



By providing:

- Synonyms, homonyms, antonyms, etc. - *for normalization and disambiguation*
- Hierarchies - *for context*
- Components of an ontology - *for domain knowledge*

Which enables:

- Matches based on relationships, profile and behavior
- Query disambiguation
- Query expansion / refinement



Resources

- ◆ [“The Role of Ontologies with LLMs,”](#) (January 9, 2024) by James Midkiff, *Enterprise Knowledge Blog*.
- ◆ [“How a Knowledge Graph Supports AI: Technical Considerations,”](#) (September 26, 2023), by Urmi Majumder, *Enterprise Knowledge Blog*.
- ◆ [“Expert Analysis: When should my organization use auto-tagging?,”](#) [Part 1 \(July 19, 2022\)](#) and [Part 2 \(December 20, 2022\)](#) by James Midkiff and Sara Duane, *Enterprise Knowledge Blog*.
- ◆ [“ChatGPT and Generative AI for Taxonomy Development”](#) (PPT), presented by Xia Lin, Taxonomy Boot Camp Conference, November 7, 2023.
- ◆ [“ChatGPT, Taxonomist: Opportunities & Challenges in AI-Assisted Taxonomy Development”](#) (PPT), presented by Margie Hlava and Heather Kotula, Taxonomy Boot Camp Conference, November 7, 2023.
- ◆ [“Taxonomies and ChatGPT,”](#) (May 29, 2023), by Heather Hedden, *The Accidental Taxonomist Blog*.

Q&A

Thank you for listening.
Questions?

Heather Hedden
Senior Consultant
Enterprise Knowledge, LLC
www.enterprise-knowledge.com
hhedden@enterprise-knowledge.com
www.linkedin.com/in/hedden

