Connecting Users to Content
An Introduction to
Taxonomy Design & Creation

Workshop

Information Architecture Conference

April 21, 22, 23, 2021

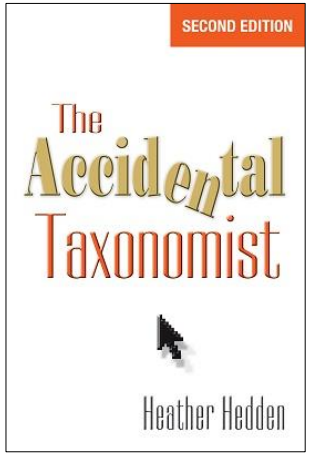
Heather Hedden

About Heather Hedden

- Data and Knowledge Engineer, Semantic Web Company, vendor of PoolParty software
- Instructor of online, virtual, and onsite taxonomy courses and workshops. Hedden Information Management
- Former consultant; former staff taxonomist at Gale/Cengage Learning, Viziant, First Wind
- Author of *The Accidental Taxonomist* (2010, 2016, Information Today, Inc.)

SEMANTIC WEB COMPANY

poolparty



Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

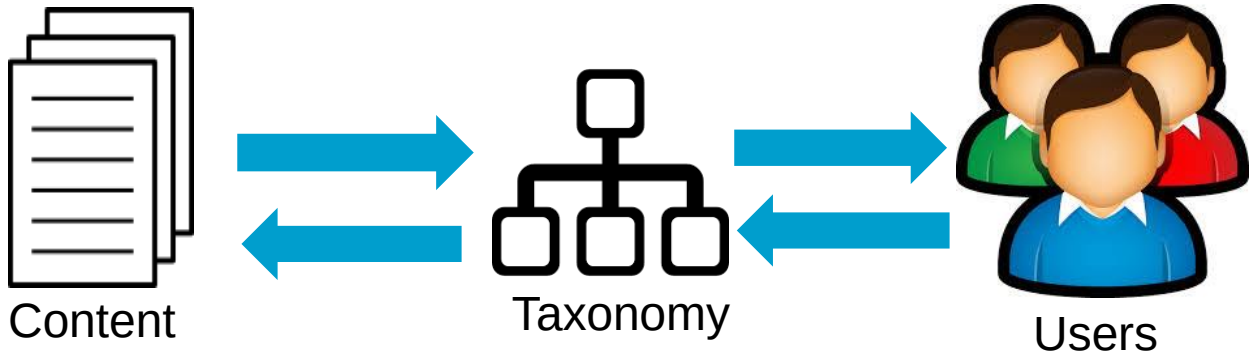
Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Introduction to Taxonomies

What is a taxonomy for?

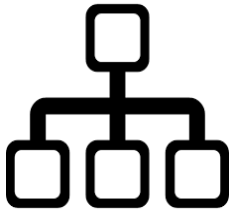
- Concepts/terms are used to tag/index/categorize pages or content to make them easier to be found and retrieved
  - supporting better findability than search alone
- The taxonomy is an intermediary that links the user to the desired content.
- The taxonomy should suit the content and the users.



Introduction to Taxonomies

Two major approaches to taxonomies

1. A hierarchy of terms/topics/categories arranged with narrower topics/subcategories displayed under their broader/parent categories.



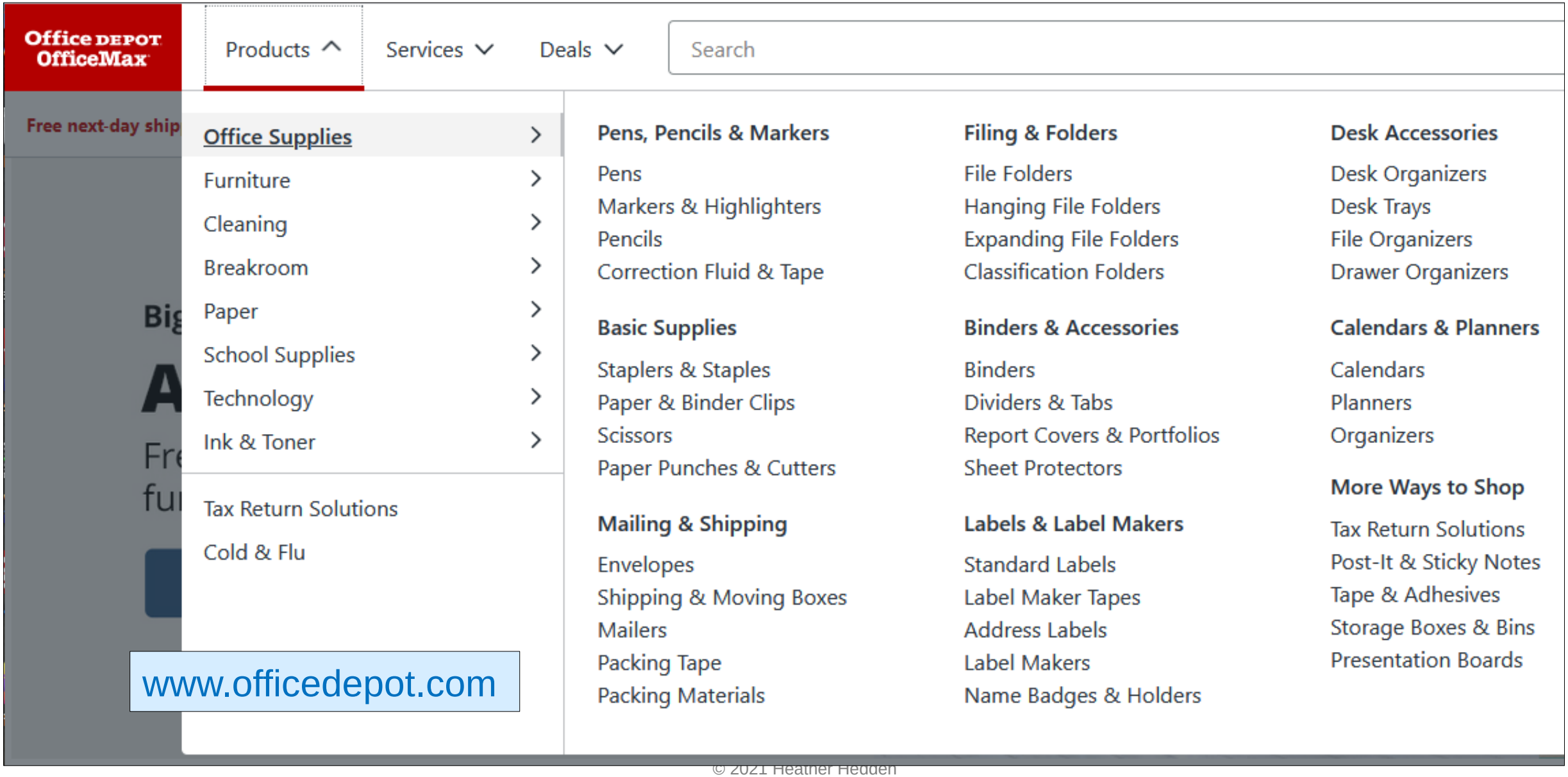
   - To guide users to find the desired topic (and its linked content of pages or documents)
   - Similar to navigation and site maps, but more topical and not just based on page titles

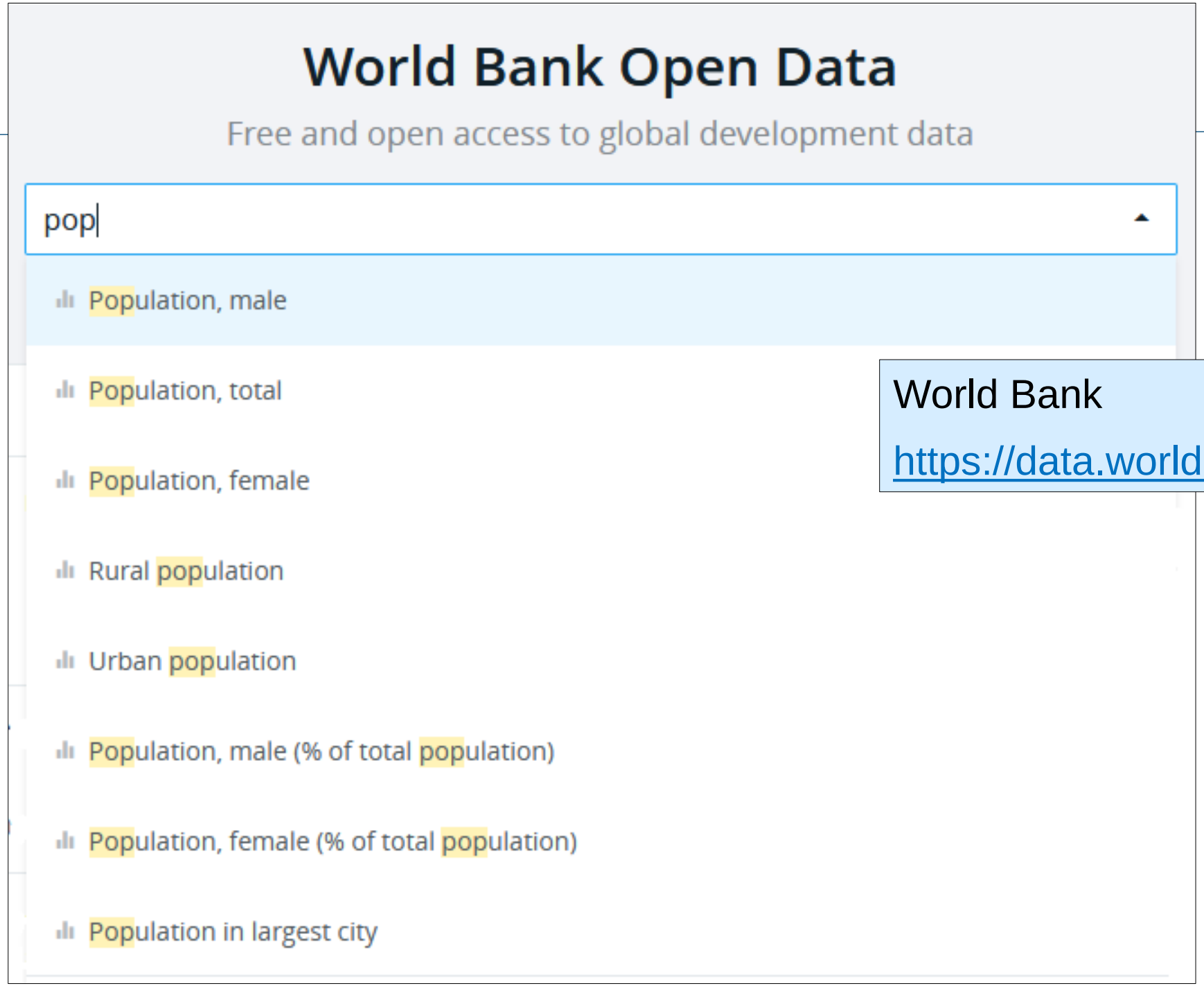
2. A controlled vocabulary of metadata tags/labels to apply to pages, posts, or documents, so that they can be more precisely and comprehensively retrieved (than by search algorithms alone on keywords in text).

   - Implemented as search suggestion terms, search refinement filters, or related topics and searches

Which approach is most relevant to you? Complete the Zoom poll.

Introduction to Taxonomies





World Bank

https://data.worldbank.org

Introduction to Taxonomies

What is a taxonomy?

Controlled and *organized*

- A kind of controlled vocabulary or knowledge organization system, based on unambiguous concepts, not just words
- Terms are arranged in a structure of hierarchies, categories, or facets to organize the terms.

Why called “taxonomy”?

- From ancient Greek "taxis,” meaning arrangement + "nomia,” meaning method.
- Originally meant the science and practice of naming and classifying.
- Aspect of "classification" is still relevant to the definition of taxonomies.

000 Computer science, knowledge & systems
010 Bibliographies
020 Library & information sciences
030 Encyclopedias & books of facts
040 [Unassigned]
050 Magazines, journals & serials
060 Associations, organizations & museums
070 News media, journalism & publishing
080 Quotations
090 Manuscripts & rare books

100 Philosophy
110 Metaphysics
120 Epistemology
130 Parapsychology & occultism
140 Philosophical schools of thought
150 Psychology
160 Logic
170 Ethics
180 Ancient, medieval & eastern philos
190 Modern western philosophy

200 Religion
210 Philosophy & theory of religion
220 The Bible
230 Christianity & Christian theology
240 Christian practice & observance
250 Christian pastoral practice & religious orders
260 Christian organization, social work & worship
270 History of Christianity
280 Christian denominations
290 Other religions

300 Social sciences, sociology & anthropology
310 Statistics
320 Political science
330 Economics
340 Law
350 Public administration & military science 360 Social problems & social services
370 Education
380 Commerce, communications & transportation
390 Customs, etiquette & folklore 400 Language

400 Language
410 Linguistics
420 English & Old English languages
430 German & related languages
440 French & related languages
450 Italian, Romanian & related languages
460 Spanish & Portuguese languages
470 Latin & Italic languages
480 Classical & modern Greek languages
490 Other languages

500 Science
510 Mathematics
520 Astronomy
530 Physics
540 Chemistry
550 Earth sciences & geology
560 Fossils & prehistoric life
570 Life sciences; biology
580 Plants (Botany)
590 Animals (Zoology)

600 Technology
610 Medicine & health
620 Engineering
630 Agriculture
640 Home & family management
650 Management & public relations
660 Chemical engineering
670 Manufacturing
680 Manufacture for specific uses
690 Building & construction

700 Arts
710 Landscaping & area planning
720 Architecture
730 Sculpture, ceramics & metalwork
740 Drawing & decorative arts
750 Painting
760 Graphic arts
770 Photography & computer art
780 Music
790 Sports, games & entertainment

800 Literature, rhetoric & criticism
810 American literature in English
820 English & Old English literatures
830 German & related literatures
840 French & related literatures
850 Italian, Romanian & related literatures
860 Spanish & Portuguese literatures
870 Latin & Italic literatures
880 Classical & modern Greek literatures
890 Other literatures

900 History
910 Geography & travel
920 Biography & genealogy
930 History of ancient world (to ca. 499)
940 History of Europe
950 History of Asia
960 History of Africa
970 History of North America
980 History of South America
990 History of other areas

Dewey Decimal Classification

100s level

Is a classification system a taxonomy?

Introduction to Taxonomies

Taxonomy vs. a Classification System

Classification systems: LC Classification, Dewey Decimal Classification, NAICS, etc.

Classification Systems

- Have codes
- Don’t have synonyms (alternate labels)
- Comprehensive and balanced coverage of a domain
- Designed to be browsed hierarchically, top-down
- Built for a domain but independent of specific content
- Provide for limited expansion
- Has “not elsewhere classified” or “other”

Taxonomies

- Don’t have codes
- Often have synonyms (alternate labels)
- Just the terms needed and may be hierarchically unbalanced
- Designed to be browsed, searched, or may not be fully displayed to end-users
- Custom-built to include the topics in the content to be tagged
- Can grow and adapt without limits
- Does not have terms for the miscellaneous

Introduction to Taxonomies

Taxonomy vs. site/intranet navigation hierarchy

What's the relation between a taxonomy and navigation?

1. The two are names used interchangeably
2. A navigation is one form of a taxonomy
3. A navigation is usually part of a larger taxonomy
4. A taxonomy informs navigation design
5. A taxonomy and navigation are distinct and different

Complete the Zoom poll.

Introduction to Taxonomies

Taxonomy vs. site/intranet navigation hierarchy

Navigation

- Single-site use and implementation
- Reflect the site-map structure
- Often includes task-based labels
- Labels based on page titles
- Designed to be browsed hierarchically, top-down
- 1-3 level hierarchy
- One-to-one label-to-page
- Limited size; does not cover all pages
- Biased to emphasize what is important
- Not so flexible for updating
- Paths and links, not metadata

Taxonomies

- May be re-used in multiple implementations
- Reflects organic relations of the topics
- Based on topics, subjects
- Labels based on concepts/topics
- Designed to be browsed, searched, or may not be fully displayed to end-users
- Options for deeper hierarchy and/or facets
- One-to-many label to multiple pages
- Can be large; can cover all pages/content
- Neutral to topic importance
- Can grow and adapt without limits
- Often is metadata

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Knowledge Organization System Types

Knowledge organization system (KOS)

- The most general, broadest all applications
- Any system of terms, terminology, classification, etc.
  to organize, define, manage, and/or retrieve information.
- Not any method to organize knowledge directly, but a scheme to organize concepts/terms for organizing, classifying, defining, tagging, or retrieving information.
- Broader, includes more than just “controlled vocabularies”

KOS types:

term lists
synonym rings
name authorities
taxonomies
thesauri
glossaries
dictionaries
gazetteers
terminologies
categorization schemes
classification schemes
subject heading schemes
semantic networks
ontologies

Controlled Vocabularies for information retrieval

Controlled Vocabulary Types

Controlled Vocabulary

- An authoritative, restricted list of terms (words or phrases) mainly used for indexing/tagging content to support information retrieval.
- Terms refer to unambiguous concepts.
- Controlled in who and when new terms can be added.
- Usually makes use of variants/synonyms/alternative labels to point to the correct term names.
- May or may not have structured relationships between terms.

Controlled Vocabulary Types

Types of controlled vocabularies

- Simple term list
- Synonym ring (search-support “thesaurus”)
- Authority file (controlled list with variants; no hierarchy)
- Taxonomy
  - Hierarchical taxonomy
  - Faceted taxonomy
- Thesaurus

“Taxonomy” sometimes means any controlled vocabulary.

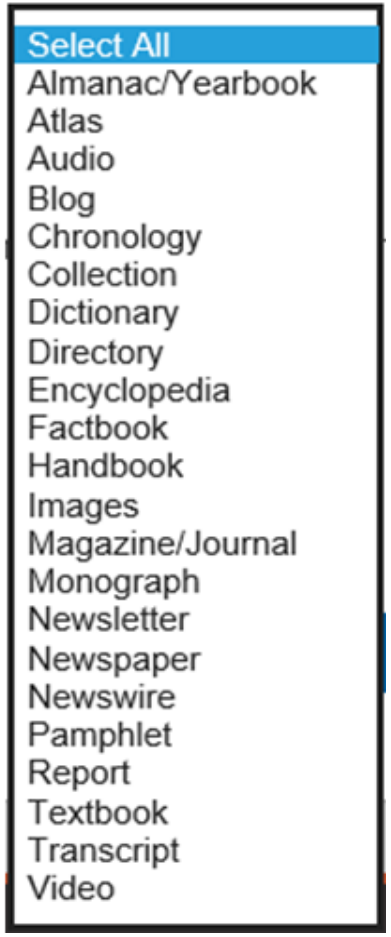
Controlled Vocabulary Types

Term list

- A simple list of terms
- Usually alphabetical, but could be in other logical order
- Lacking synonyms, it is usually short enough for quick browsing
- Can appear in drop-down scroll boxes
- May be used for various metadata values
- Part of a larger set of controlled vocabularies

Country of publication

Language

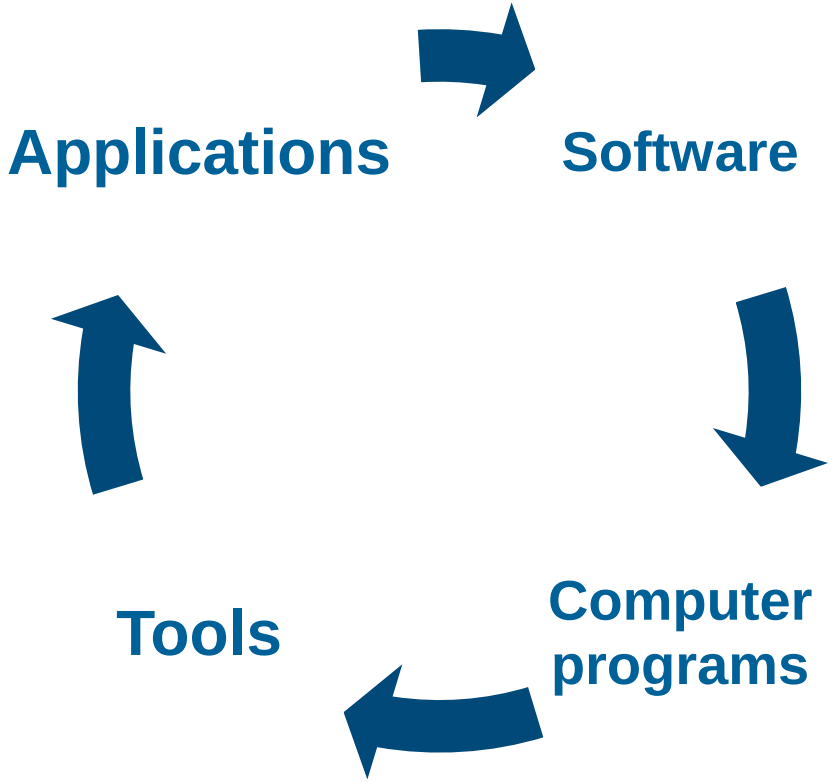


Format

Controlled Vocabulary Types

Synonym ring

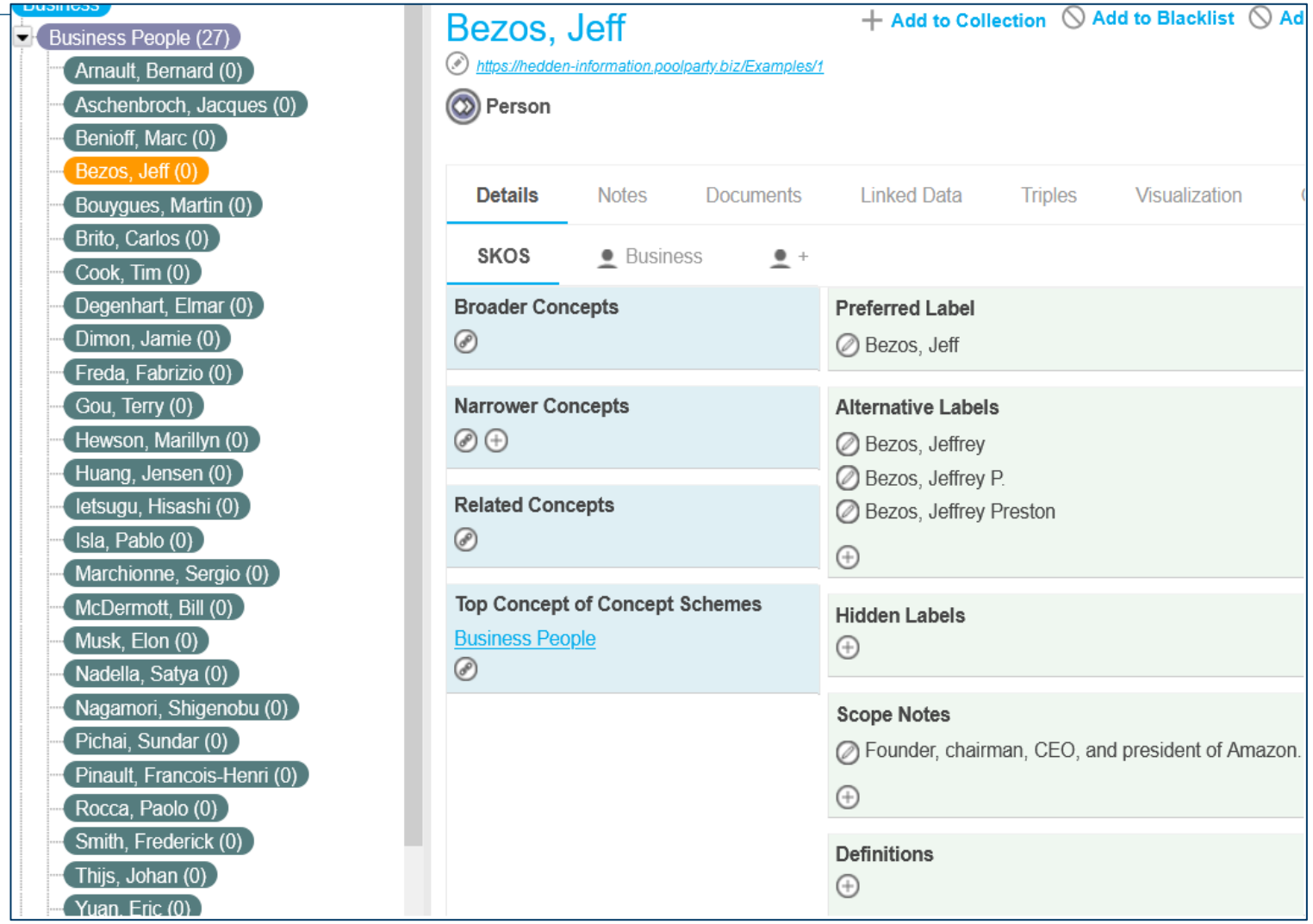
- A controlled vocabulary with synonyms or near-synonyms for each concept
- No designated “preferred” term: All terms are equal and point to each other, as in a ring.
- Synonyms are usually not displayed to the user.
- Usually used to support search.
- Also called a “search thesaurus.”



Controlled Vocabulary Types

Name authority

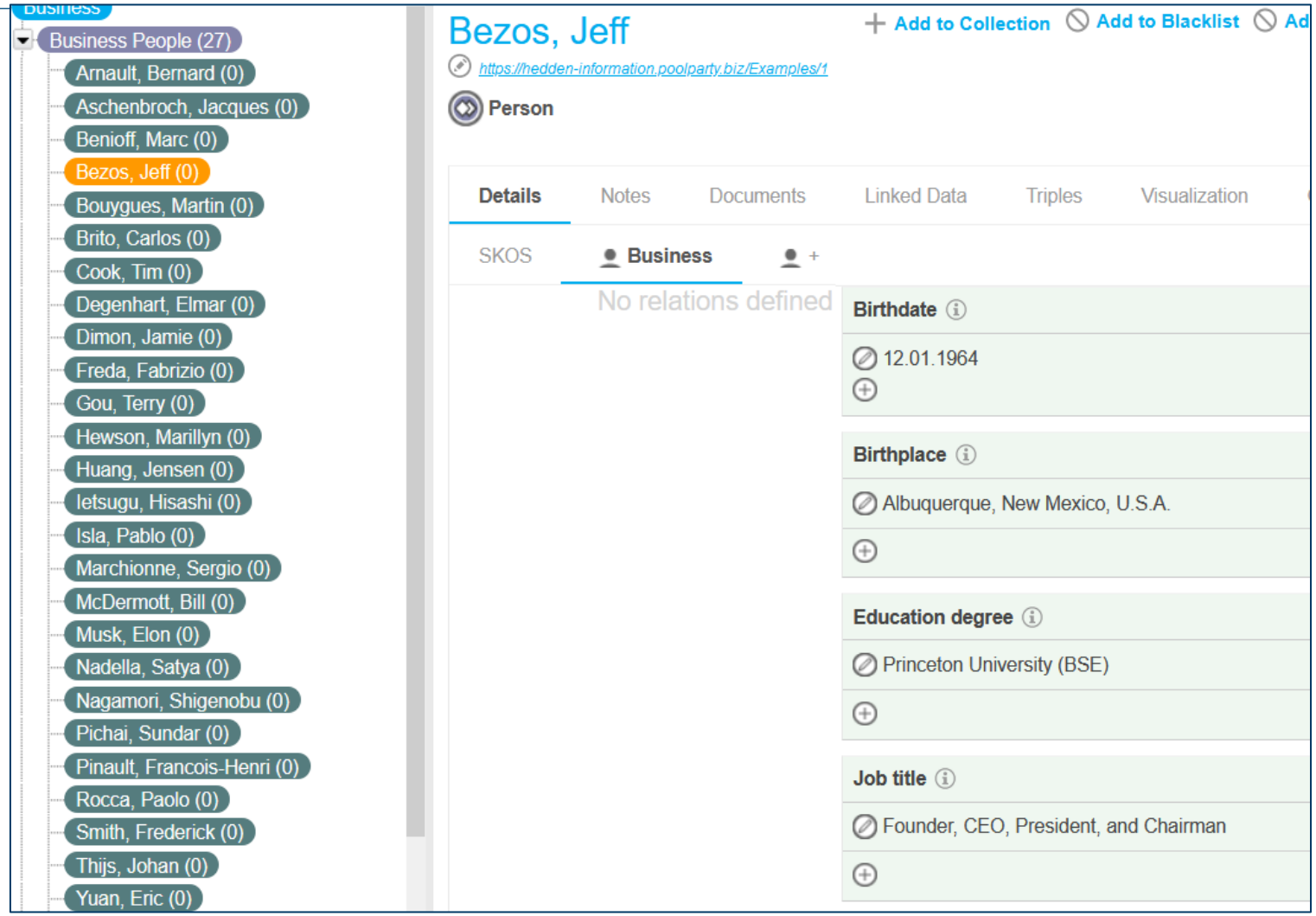
- For named entities, concrete entities,
  proper nouns
- A controlled vocabulary with preferred names and variant/alternative names.
- May or may not have hierarchical relationships between named entities.
- Usually has additional information/attributes (metadata) for each named entity.



Controlled Vocabulary Types

Name authority

- The basic model can be extended with a custom scheme to support extended attributes, often desired for named entities.



Controlled Vocabulary Types

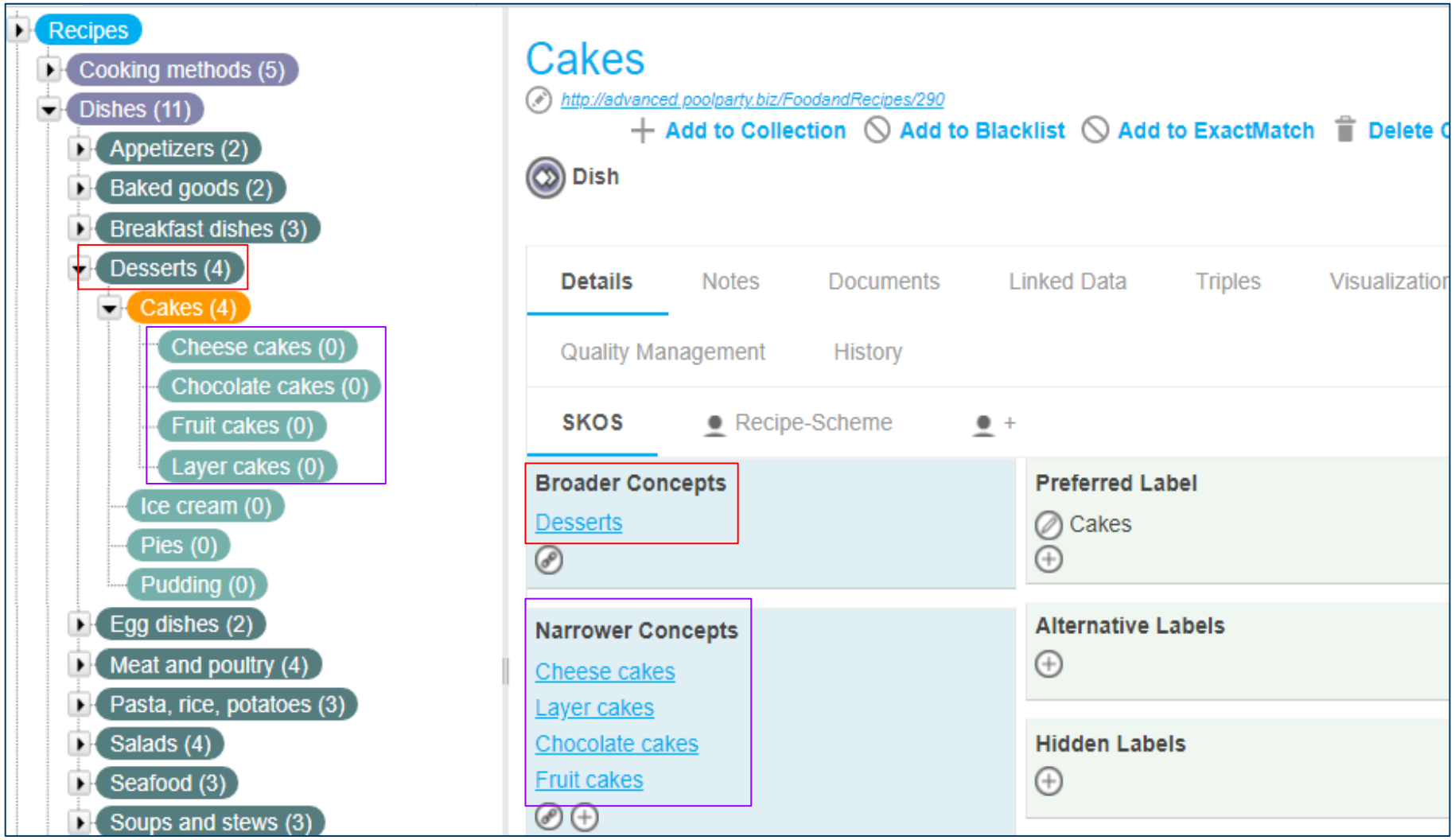
Taxonomy

- A controlled vocabulary with broader/narrower (parent/child) term relationships that include all terms to create a hierarchical structure
- With focus for categorizing and organization concepts
- May or may not have “synonyms” to point to the correct, preferred terms
- May comprise several hierarchies or facets
  (A facet can be considered a hierarchy.)

Controlled Vocabulary Types

Hierarchical taxonomy

Concepts have broader-concept and/or narrower-concept relationships to other concepts.



Controlled Vocal

Taxonomy Examples

Leisure and culture
. Arts and entertainment venues
. . Museums and galleries
. Children's activities
. Culture and creativity
. . Architecture
. . Crafts
. . Heritage
. . Literature
. . Music
. . Performing arts
. . Visual arts
. Entertainment and events
. Gambling and lotteries
. Hobbies and interests
. Parks and gardens
. Sports and recreation
. . Team sports
. . . Cricket
. . . Football
. . . Rugby
. . Water sports
. . Winter sports
. Sports and recreation facilities
. Tourism
. . Passports and visas
. Young people's activities

Hierarchical Taxonomy Example

Career Level
- Student
- Entry Level
- Experienced
- Manager
- Director
- Executive

Function
- Customer Service & Support
- Delivery
- Engineering
- Finance
- General Management
- Legal & Regulatory Affairs
- Marketing & Advertising [more]

Industry
- Agriculture
- Apparel & Fashion
- Automotive
- Aviation & Aerospace
- Banking
- Biotechnology
- Broadcast Media
- Chemicals [more]

Faceted Taxonomy Example

Controlled Vocabulary Types

Thesaurus

- A controlled vocabulary that has standard structured relationships between terms
  - Hierarchical: broader term / narrower term (BT/NT)
  - Associative: related terms (RT)
  - Equivalence: preferred term (“use for”) / non-preferred term (use) (USE/UF)
- Created in accordance with standards:
  - ANSI/NISO Z39.19 (2005, renewed 2010) *Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies*
  - ISO 25964 (2011, 2013) *Thesauri and Interoperability with Other Vocabularies*
- “Thesaurus” is most often the kind of controlled vocabulary used in indexing periodical literature

ASIS&T thesaurus excerpt

library personnel
UF library staff
BT information workers
NT librarians
paraprofessional library personnel
RT media specialists

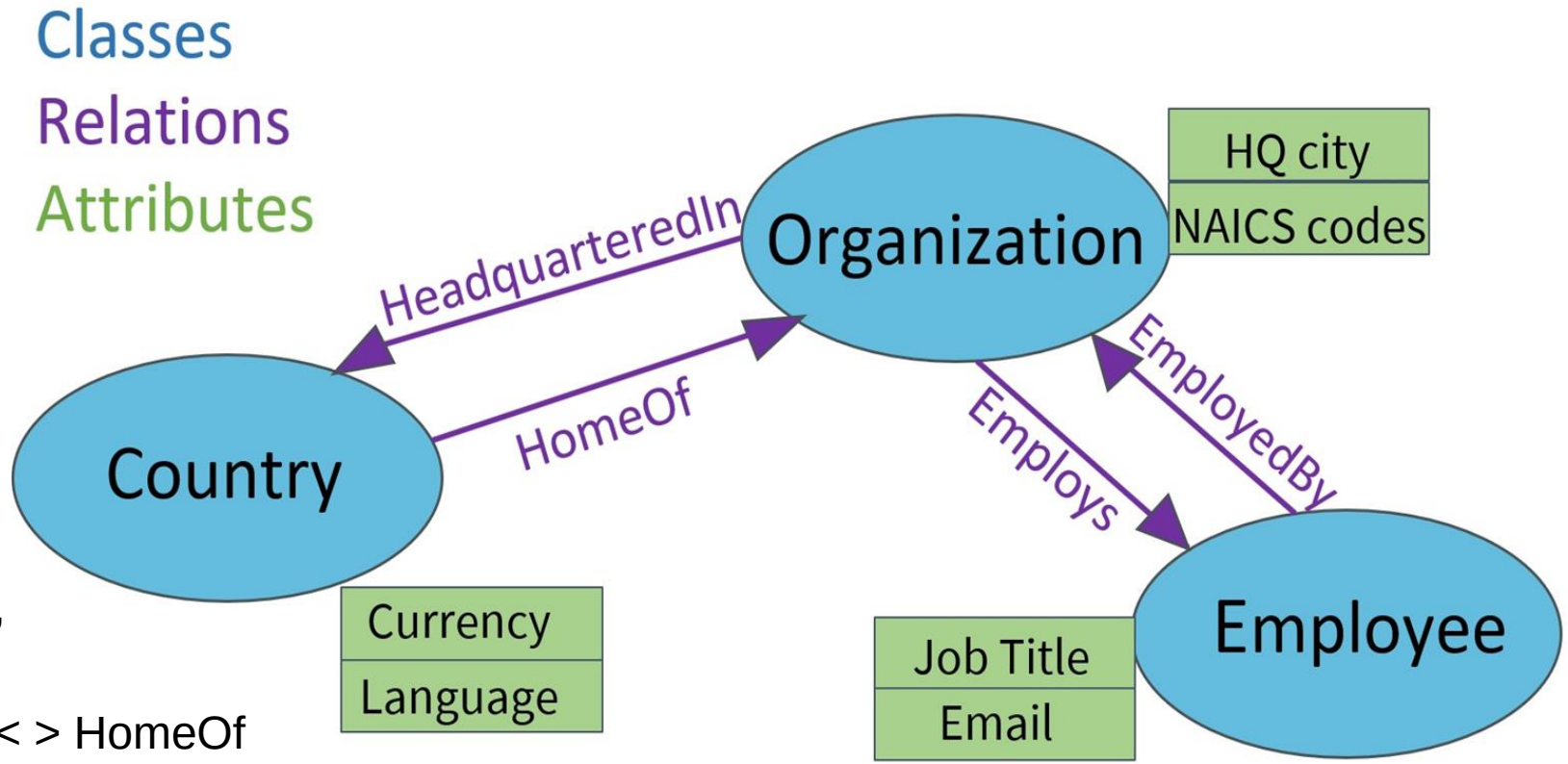
Knowledge Organization System Types

Ontology

- A more abstract layer in describing a knowledge organization system.
- A formal naming and definition of the types, properties and interrelationships of entities in a particular domain.
- Relationships contain meaning, are “semantic.”
- A form of “knowledge representation”
- Created according to W3C guidelines: OWL Web Ontology Language Guide W3C Recommendation (2004)
  http://www.w3.org/TR/2004/REC-owl-guide-20040210/
- Comprises classes, relations, and attributes. Theses are linked in triples.

Knowledge Organization System Types

Ontology excerpt example



Classes

Employee, Country, Organization

Relations: HeadquarteredIn < > HomeOf
EmployedBy < > Employs

Attributes: Email address, Job title, HQ city, NAICS codes, Currency, Language

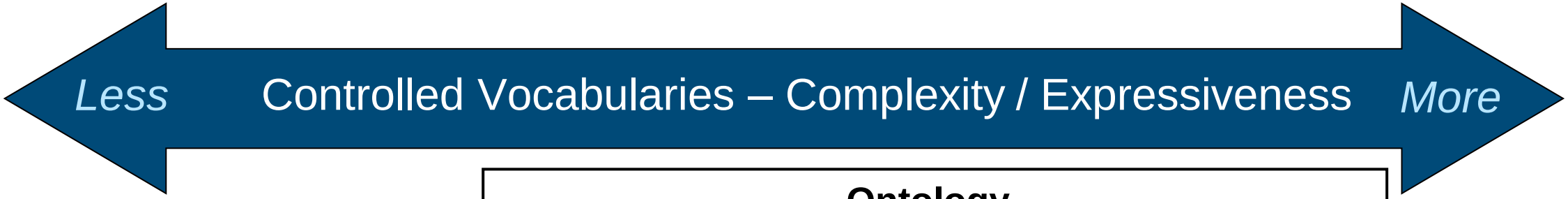
Knowledge Organization System Types

Summary Types

Pick List	Synonym Ring	Name Authority	Taxonomy	Thesaurus	Ontology
Ambiguity control		Ambiguity control	Ambiguity control	Ambiguity control	
	Synonym control	Synonym control	(Synonym control)	Synonym control	
			Hierarchical relationships	Hierarchical relationship	
				Associative relationships	Semantic relationships
					Classes
		(Attributes)			Attributes

Knowledge Organization System Types

Summary of Types



		Ontology		
Pick List	**Synonym Ring**	**Name Authority**	**Taxonomy**	**Thesaurus**
Ambiguity control		Ambiguity control	Ambiguity control	Ambiguity control
	Synonym control	Synonym control	(Synonym control)	Synonym control
			Hierarchical relationships	Hierarchical relationship
				Associative relationships
		(Attributes)		

Knowledge Organization System Types

Taxonomies

- All terms belong to a limited number of major hierarchies (or facets).
- Does not strictly follow standards.
- Supports classification, categorization, concept organization. (Like Linnaean taxonomy)
- Approach is a top-down browse navigation.
- Especially serving end-users when browsing.

Thesauri

- All terms have relationships, but “hierarchies” may be as few as 2 terms.
- Follows ANSI/NISO thesaurus standards.
- Supports concept scoping, disambiguation, and relationships with similar concepts. (Like Roget’s)
- Approach is term-centered and what terms are linked to/from it.
- Especially serving indexers/ indexing.

Knowledge Organization System Types

Quiz-Poll:
What kind of knowledge organization system is best for each?

- A collection of published research articles
- A collection of digital assets (images, etc.)
- An ecommerce website
- Enterprise/intranet search
- A government agency public website
- Pharmaceutical research system

Complete the Zoom polls.

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Benefits and Uses

Benefits of taxonomies/controlled vocabularies

1. Controlled vocabulary

   Brings together different wordings (synonyms) for the same concept

   - Helps people search for information by different names

2. Classification and structure

   Organizes information into a logical structure

   - Helps people browse for information
   - Provides context and meaning for concepts, to support indexing and retrieval

Benefits and Uses

Synonym Control

As a controlled vocabulary…

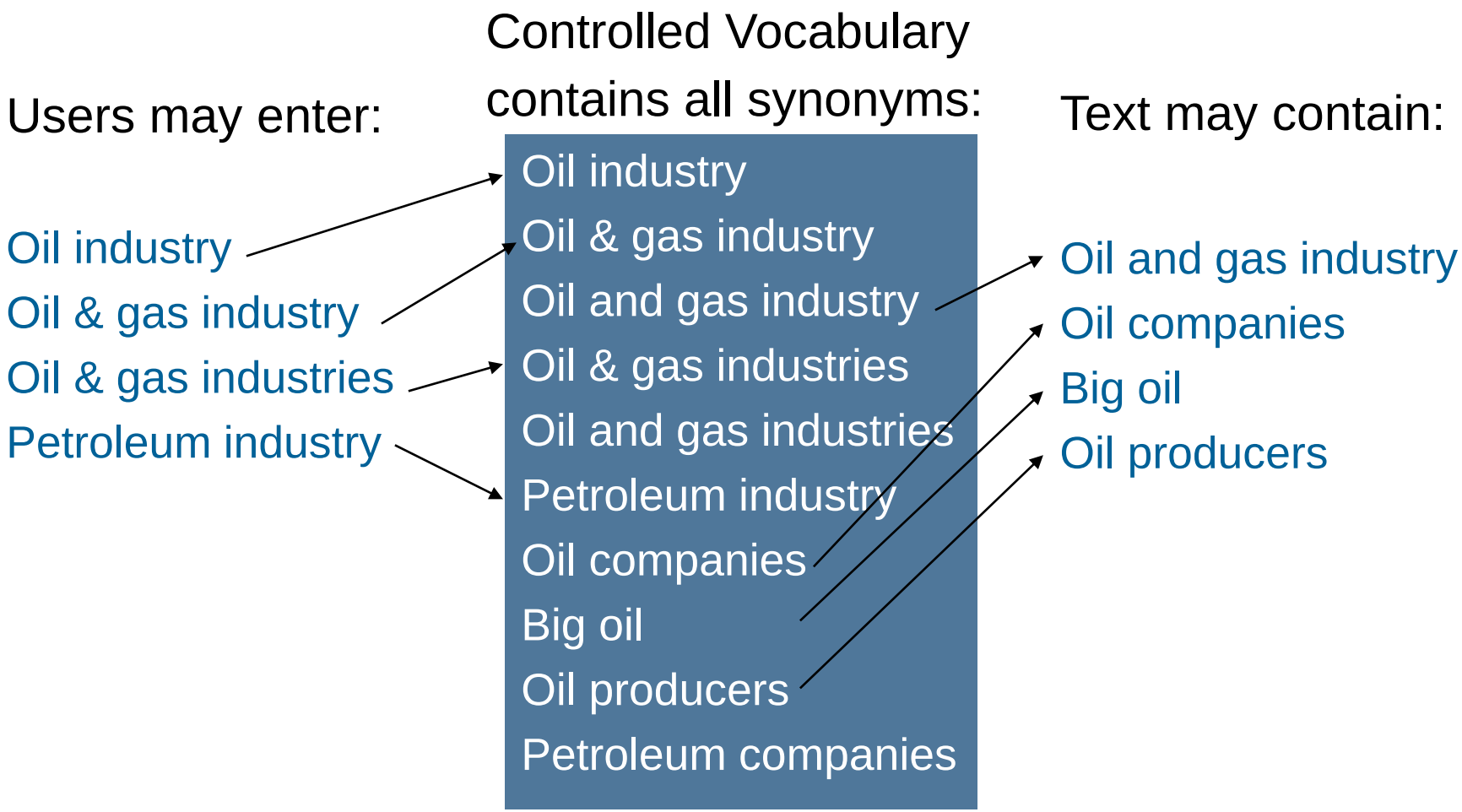
A search restricted on the controlled vocabulary retrieves concepts not just words.

- Documents excluded for mere text-string matches (e.g. **monitors** for computers, not the verb “observes”)
- Improves accuracy and **precision** in retrieving information

Gathers synonyms, acronyms, variant spellings, etc.

- Documents not missed due to use of different words (e.g. **Automobiles**, instead of **Cars**)
- Improves comprehensive **recall** of information

Benefits and Uses



Benefits and Uses

Classification and structure

- Categorization scheme for information organization, classification, guided search
- More often for end-users guidance than for indexers

Example:
Craigslist Boston
https://boston.craigslist.org

community

activities
artists
childcare
classes
events
general
groups
local news
lost+found
missed connections
musicians
pets
politics
rants & raves
rideshare
volunteers

services

automotive
beauty
cell/mobile
computer
creative
cycle
event
farm+garden
financial
household
labor/move
legal
lessons
marine
pet
real estate
skilled trade
sm biz ads
travel/vac
write/ed/tran

discussion forums

android
apple
arts
frugal
gaming
garden
pets
philos
photo

housing

apts / housing
housing swap
housing wanted
office / commercial
parking / storage
real estate for sale
rooms / shared
rooms wanted
sublets / temporary
vacation rentals

for sale

antiques
appliances
arts+crafts
atv/utv/sno
auto parts
aviation
baby+kid
barter
beauty+hlth
bike parts
bikes
boat parts
boats
farm+garden
free
furniture
garage sale
general
heavy equip
household
jewelry
materials
motorcycle parts
motorcycles
music instr
photo+video

jobs

accounting+finance
admin / office
arch / engineering
art / media / design
biotech / science
business / mgmt
customer service
education
etc / misc
food / bev / hosp
general labor
government
human resources
legal / paralegal
manufacturing
marketing / pr / ad
medical / health
nonprofit sector
real estate
retail / wholesale
sales / biz dev
salon / spa / fitness
security
skilled trade / craft
software / qa / dba
systems / network

Benefits and Uses

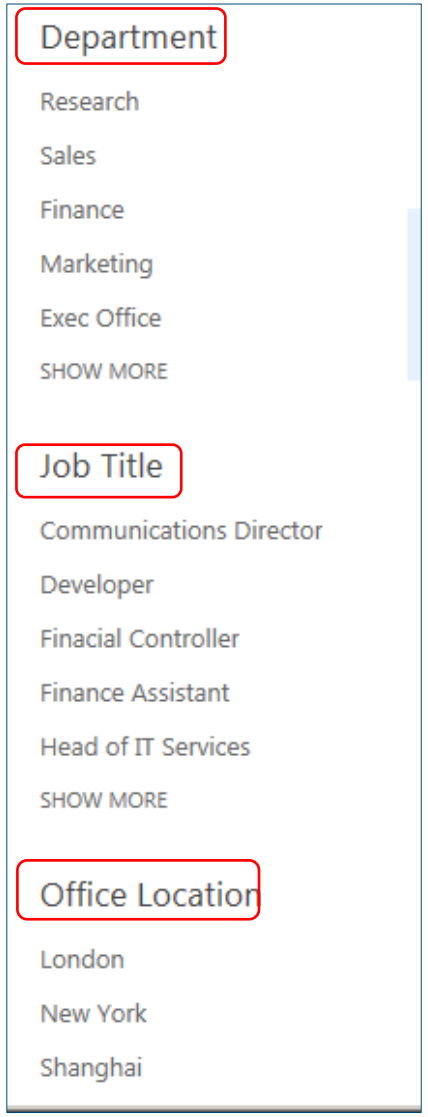
Structure: Hierarchy purposes and benefits

1. Serving users who are browsing, exploring, discovering, not searching, to whom the hierarchy is displayed.
   - Users don't even have to know the first word or few letters, as in search or alphabetical browsing.
2. Instructing users on appropriate classification
3. Providing context to terms for manual indexers/taggers so that they apply the correct term.
4. Providing the context of a broader concept and thus meaning to aid in auto-classification.
5. Enabling "recursive"/"rolled up" retrieval results (A term retrieves what is indexed to it *and* what is indexed to each on of its narrower terms, all together.)

Benefits and Uses

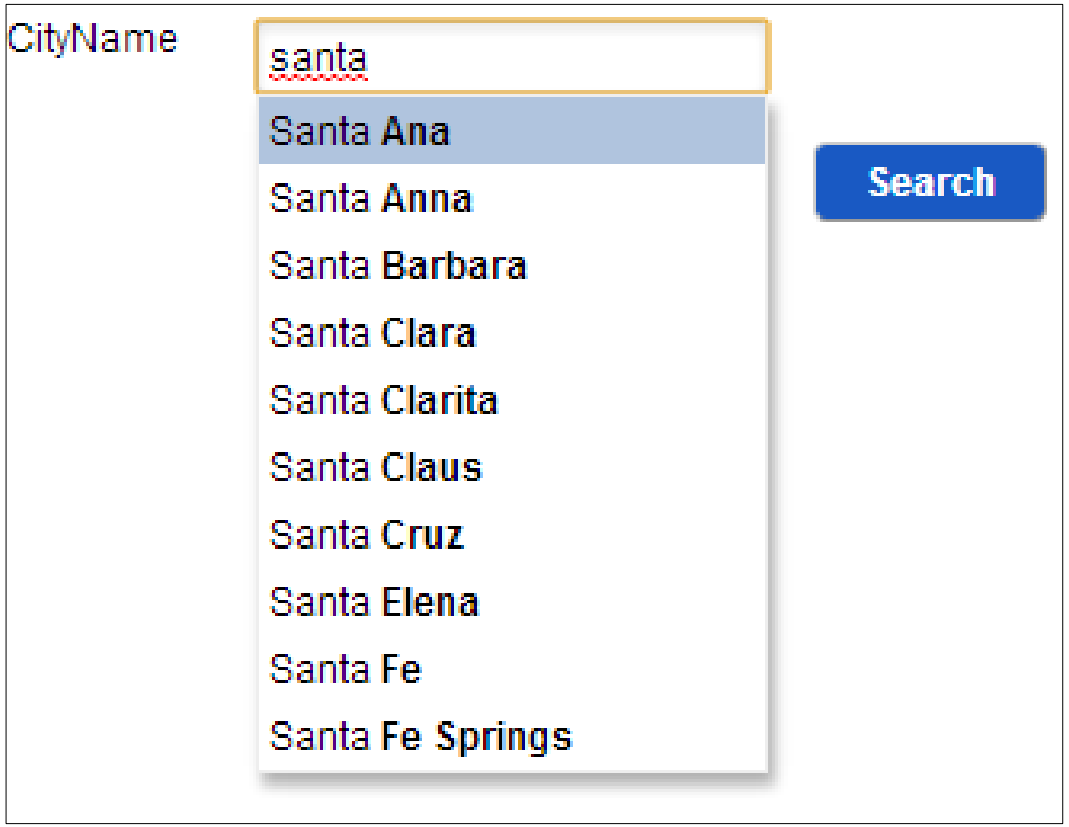
Structure: Facet purposes and benefits

- Ensures comprehensive tagging and comprehensive search/retrieval by multiple different aspects/vocabulary types.
- Supports filtering search results by different aspects/vocabulary types.
- Supports more complex search queries by users.
- Allows users to control the search refinement, narrowing or broadening in any manner or order.
- Provides guided Boolean “AND” searching upon a combination of terms in different facets.

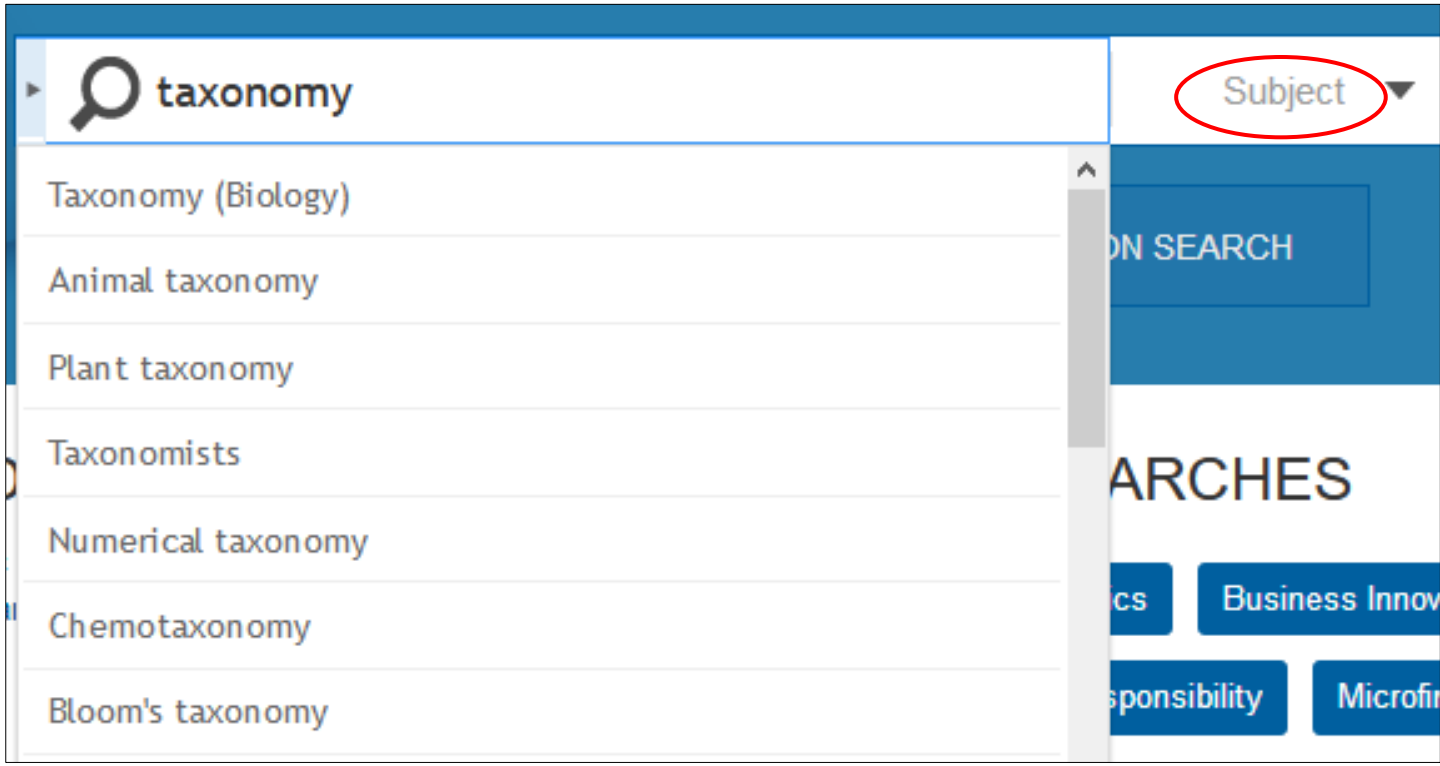


Benefits and Uses

Use of controlled vocabularies in search



Type-ahead



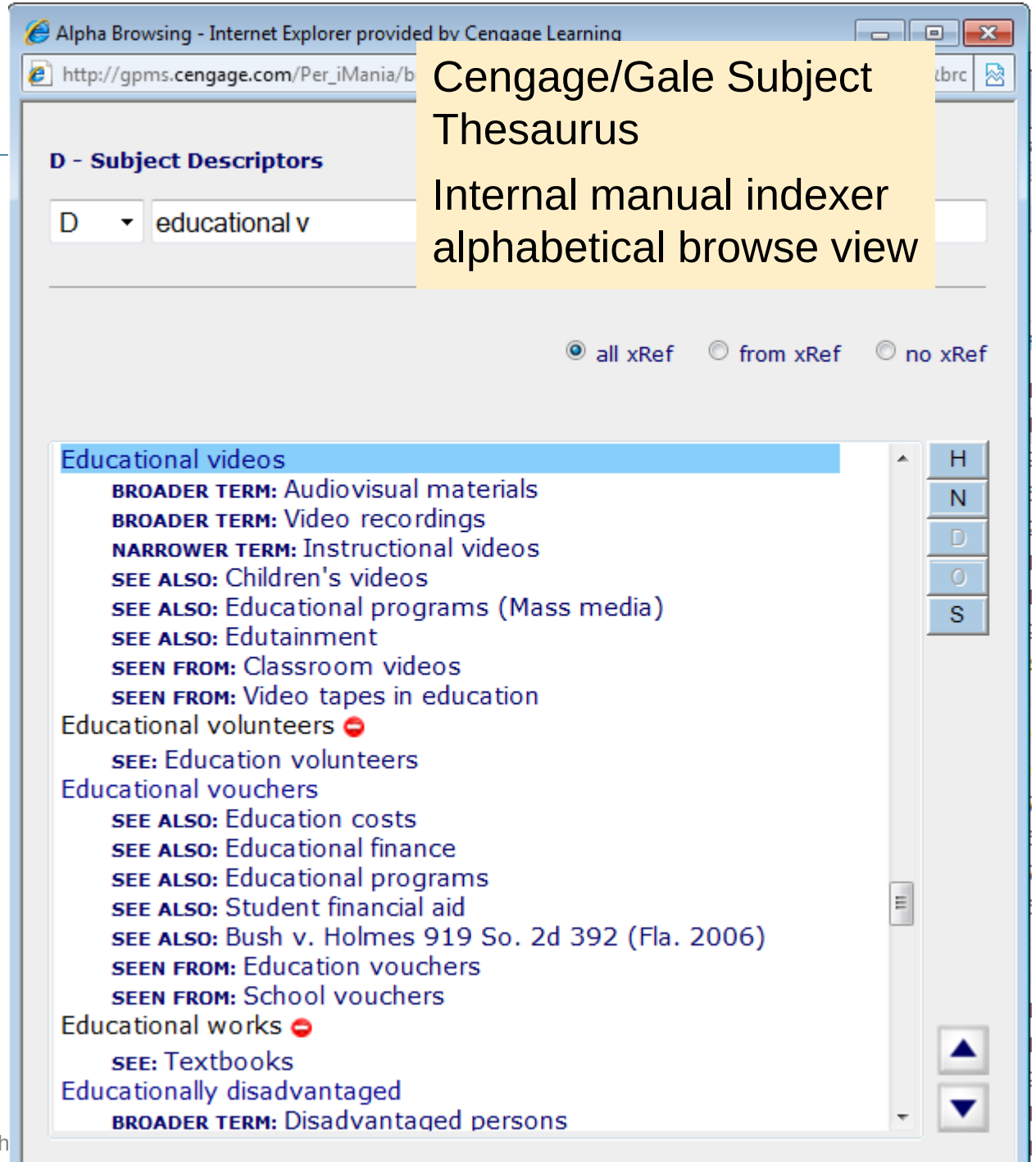
Search-suggest

Benefits and Uses

Indexing Support

As a structured list of agreed-upon terms to ensure consistent indexing

- Across multiple documents or content items, where different synonyms describe the same concepts
- By multiple indexers working on the same collection
- By machine-aided indexing / autoclassification, where taxonomy terms have rules, clues, or sample tuned documents



Benefits and Uses

Purposes and uses of controlled vocabularies

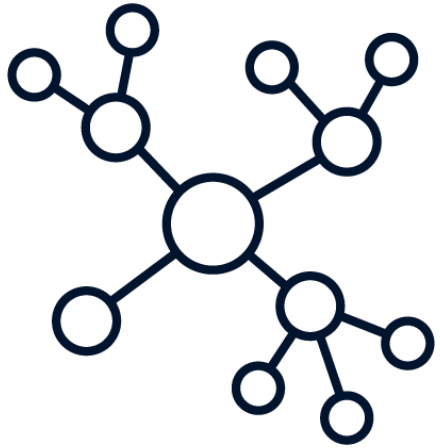
- Consistent tagging/indexing
- Topic/category browsing
- Search (matching search strings to concepts)
- Discovery (related concept links or contents sharing the same concepts)
- Filtering results
- Sorting results
- Content management workflow (rights, audience, retention, etc.)
- Consistent metadata for identification, comparison, analysis
- Visualization of topics (importance and/or relations)
- Curated content in feeds or info boxes
- Automatic linking of relevant topics for personalization or recommendation
- Enhancing knowledge graphs

Benefits and Uses

Knowledge Graphs

- Combination of:
  - An ontology as a semantic model
  - A taxonomy (or multiple taxonomies name authorities) of specific concepts and named entities
  - Instance data, which may be attributes of named entities or other data not managed in a taxonomy or name authority (data in a database, not necessarily indexed content)
- The instance data is typically managed in a graph data base (a triple store or a property graph)

What knowledge graphs can do



- Integrate knowledge
- Serve data governance
- Provide semantic enrichment
- Bring structured and unstructured data together
- Provide unified view of different kinds of unconnected data sources
- Provide a semantic layers on top of the metadata layer
- Improve search results beyond machine learning and algorithms
- Answer complex user questions instead of merely returning documents on a topic

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Standards and Models

Standards are of two basic types:

1. *Standards for design* – supports an expected experience and results by varied users without requiring training
2. *Standards for specifications* (measurements, protocols, coding, etc.) – supports exchange and interoperability.

Standards for knowledge organization systems of each type:

1. *For design*: ISO 25964 and ANSI/NISO Z39.19-2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* www.niso.org/publications/ansiniso-z3919-2005-r2010
2. *For specifications and interoperability*: Dublin Core, MARC, ZThes, DD 8723-5, SKOS, RDF, RDFS, and OWL (Some may be used in combination.)

Standards and Models

ANSI/NISO Z39.19-2005 (R2010)

Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

- Published by the National Information Standards Organization in 2005 with minor revisions in 2010
- Despite its title, focus is on thesauri, but much is applicable to taxonomies, in general
- Comparable international standard ISO 25964
  (which also includes multilingual thesauri and more on interoperability)
- Provides both standards and guidelines/guidance
- Free, 167 pages + index PDF

Standards and Models

ANSI/NISO Z39.19-2005 (R2010)

Guidelines/Standards on:

- Definitions, types, structure of various controlled vocabularies
- Term choice, scope, form, and style
- Forming relationships: hierarchical, associative, and equivalence (nonpreferred terms/variants/alternative labels)
- Display formats
- Interoperability and merging vocabularies
- Vocabulary construction approaches
- Vocabulary management system considerations

➢ Select from it what you need.

Standards and Models

ANSI/NISO Z39.19-2005 (R2010)

Examples of quality standards:

- Concepts/terms are nouns or noun phrases.
- Use plural for countable nouns.
- No duplicates: Concept labels must be unique.
- No relationship clashes: A pair of concepts can be either hierarchically or associatively related to each other, but not both.
- No circular relationships: hierarchical relationship logic extend:
    - Concept A is narrower to Concept B, and
    - Concept B is narrower to Concept C,
    - Concept C cannot be narrower to Concept A

Standards and Models

SKOS (Simple Knowledge Organization System)

- A data model to representation knowledge organization systems
- A World Wide Web (W3C) recommendation (2009)
- “A common data model for sharing and linking knowledge organization systems via the Web” https://www.w3.org/TR/skos-reference/
- A KOS built on SKOS is machine-readable and interchangeable.
- Encoded using XML and RDF (Resource Description Framework).
- To enable easy publication and use of such vocabularies as linked data.

➢ Different KOS types (name authority, thesaurus, taxonomy, ontology) can all be built on the SKOS standard (although ontologies are usually based on the OWL standard instead).

Standards and Models

SKOS principles

- A KOS is a group of Concepts identified with URIs and grouped into a Concept scheme.
- Concepts can be labeled with any number of lexical strings (labels) in any natural language, such as prefLabel and altLabel.
- Concepts can be documented with notes of various types: scope notes, definitions, editorial notes, etc.
- Concepts can be linked to each other using hierarchical and associative semantic relations.
- Concepts can be grouped into Collections, which can be labeled and/or ordered.
- Concepts of different concept schemes can be mapped using four basic types of mapping links.

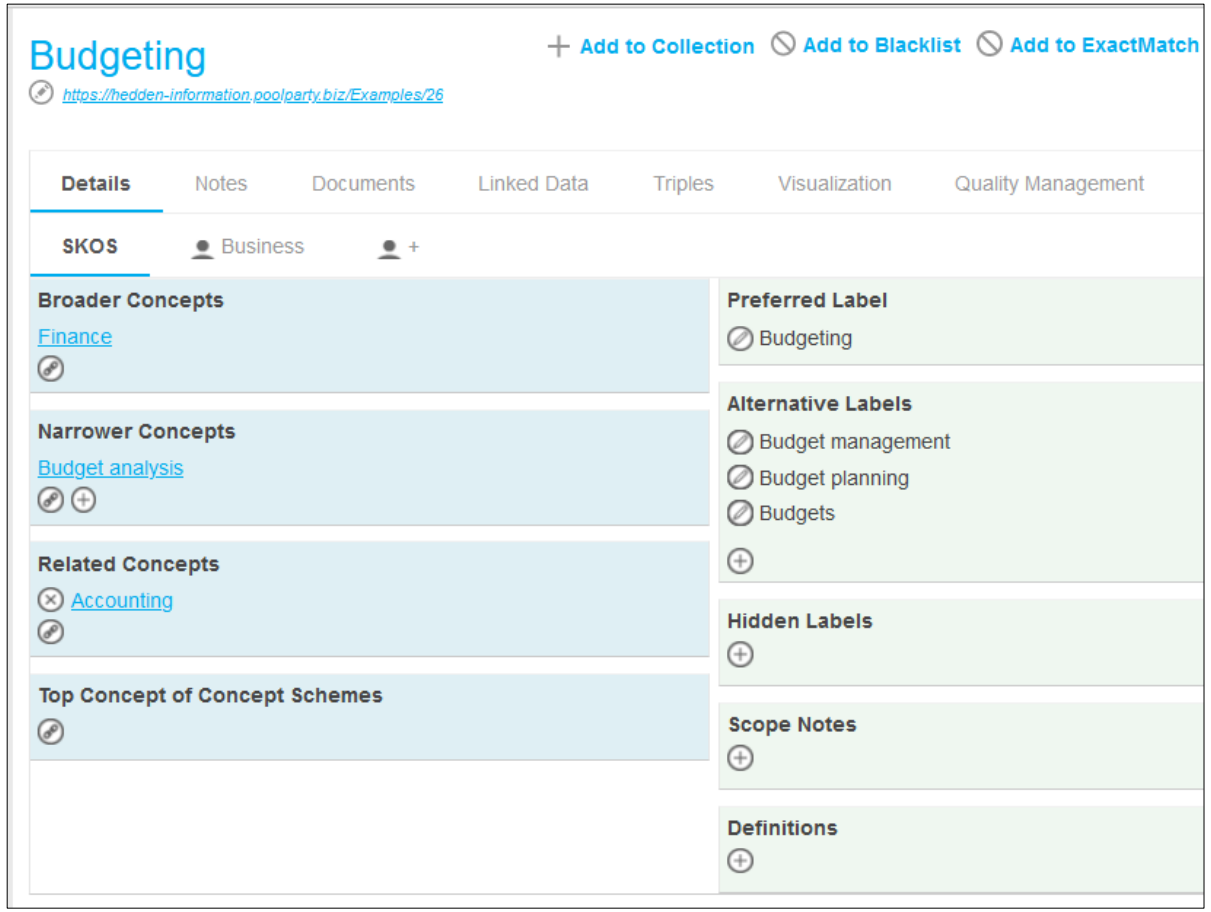
Standards and Models

Full SKOS Elements

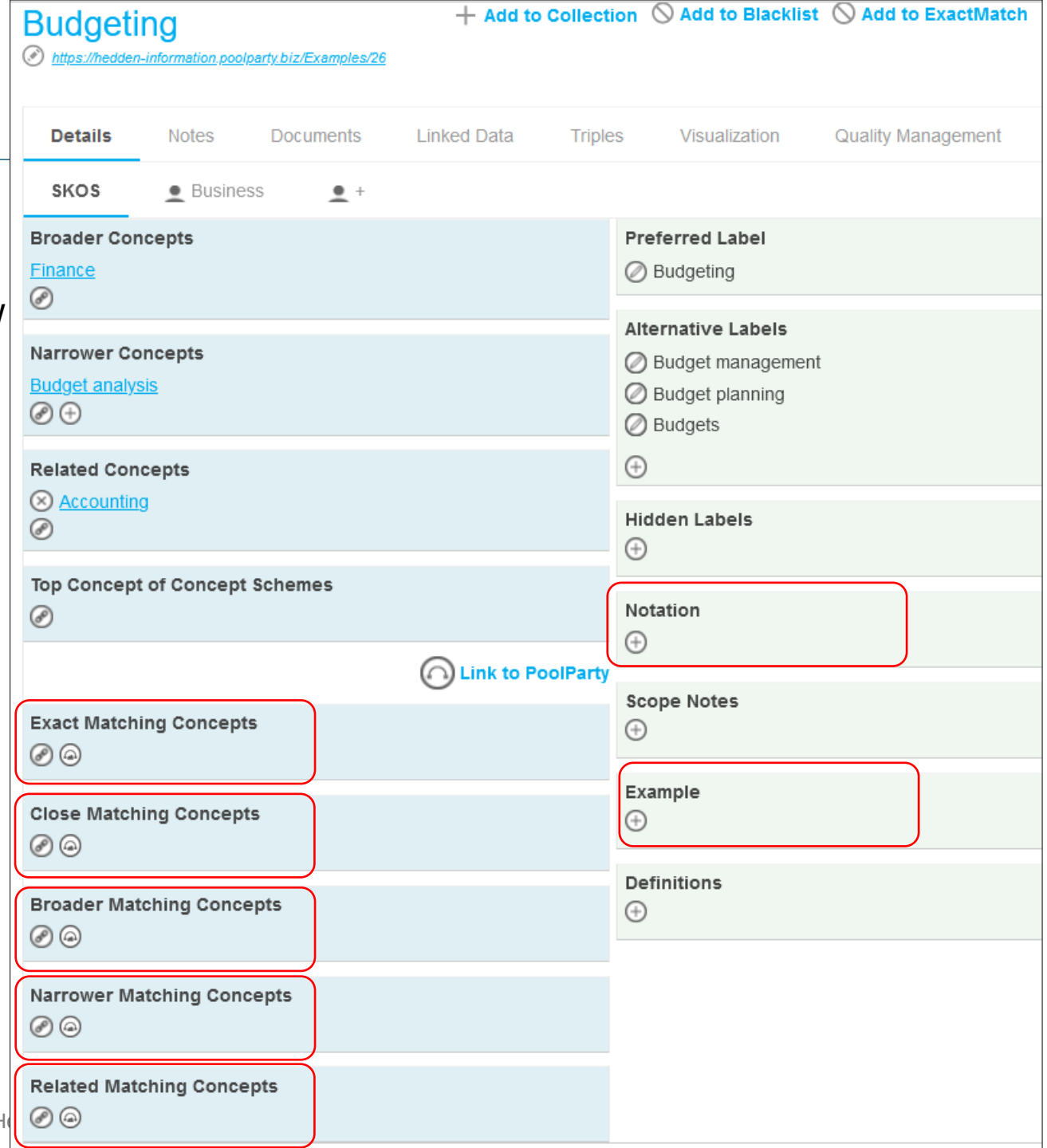
Concepts	Labels & Notation	Documentation	Semantic Relations	Collections	Mapping Properties
Concept	prefLabel	note	broader	Collection	broadMatch
ConceptScheme	altLabel	changeNote	narrower	orderedCollection	narrowMatch
inScheme	hiddenLabel	definition	related	member	relatedMatch
hasTopConcept	notation	editorialNote	broaderTransitive	memberList	closeMatch
topConceptOf		example	narrowerTransitive		exactMatch
		historyNote	semanticRelation		mappingRelation
		scopeNote			

Standards and Models

Basic SKOS view for taxonomy editing



Fuller SKOS view



Screenshots from PoolParty

Thesaurus Model vs. SKOS Model

Thesaurus model:

Terms

Of two *types*:

1. Preferred terms
2. Nonpreferred terms

- Content is indexed/tagged to preferred terms only.
- Hierarchical and associative relationships are between preferred terms.
- Preferred terms and nonpreferred terms are linked by the equivalence relationships.

SKOS vocabulary model:

Concepts

Which each *have*:

- Preferred labels (one in each language)
- Alternative labels

- Content is indexed/tagged to concepts as described by various labels.
- Hierarchical and associative relationships are between concepts.
- Preferred and alternative labels are attributes of a single concept.

Standards and Models

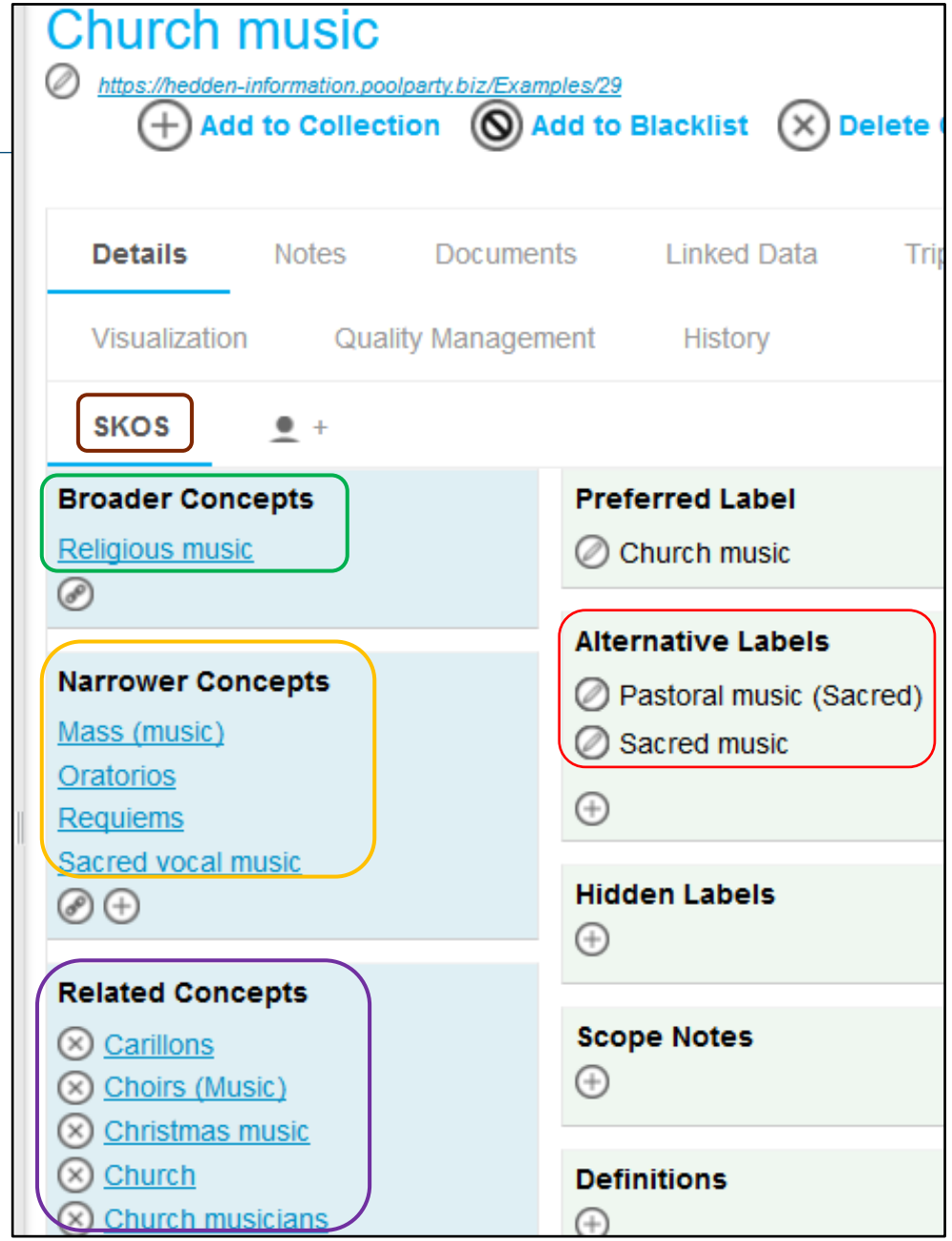
Thesaurus entry example

ANSI/NISO model vs. SKOS model

Church music

UF Pastoral music (Sacred) (Subjects)
UF Sacred music (Subjects)
BT Religious music (Subjects)
NT Mass (Music) (Subjects)
NT Oratorios (Subjects)
NT Requiems (Subjects)
NT Sacred vocal music (Subjects)
RT Carillons (Subjects)
RT Choirs (Music) (Subjects)
RT Christmas music (Subjects)
RT Church (Subjects)
RT Church musicians (Subjects)
RT Classical music (Subjects)
RT Contemporary Christian music (Subj
RT Devotional exercises (Subjects)
RT Easter music (Subjects)
RT Liturgics (Subjects)
RT Organ music (Subjects)

Gale Subject Thesaurus



Thesaurus in a SKOS-based tool, PoolParty

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Sources for Concepts/Terms

Concepts vs. Terms

Terms

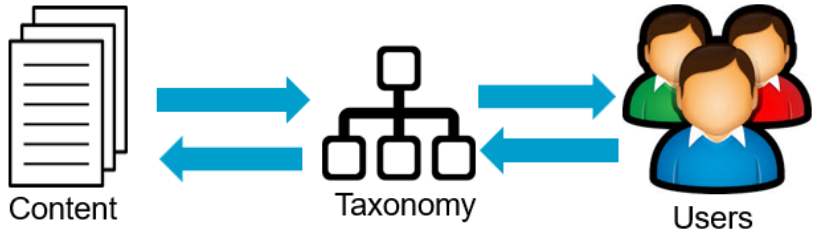
- One or more words designating a concept
- What is described in thesauri (ANSI/NSO Z.39.19 standard).
- What exists in term lists and taxonomies that have no synonyms.
- In thesauri, two kinds: preferred terms and nonpreferred terms.
- Relationships are between preferred terms.

Concepts

- An idea of something that may have different names/labels.
- What is described in the SKOS (Simple Knowledge Organization System) standard.
- For taxonomies, thesauri, or other controlled vocabularies.
- Concepts have a preferred label and any number of alternative labels.
- Relationships are between concepts.

Sources for Concepts/Terms

1. Content to be tagged
   - Manual survey/audit of sample content
   - Automatic term extraction from corpus of content
2. Users and other stakeholders
   - Brainstorming workshops and sessions
   - Interviews of sample users
   - Submitted lists of concepts/terms
   - Search query logs and other internal usage data



3. External sources: websites, Wikipedia, other taxonomies and controlled vocabularies, book tables of contents, etc.

➢ Whether creating a new or enhancing an existing taxonomy.

Sources for Concepts/Terms: Content

Content as a source for concept: manual survey of sample content

- From a representative sample of the content to be tagged with the taxonomy
- Of different types, sources, file formats, etc.
- Look for concepts especially within titles, section headings, lead paragraphs.
- Look for main idea concepts, as if you were indexing or categorizing the content.
- Keep it general for the content item/document/page as a whole
- Consider desired search strings to retrieve the content item.
- Consider different aspects: activity, location, event, person type,

➢ Manual survey can get the main idea more reliably than automated methods.

➢ Use manual methods to start a taxonomy; automatic methods to supplement it.

Sources for Concepts/Terms: Content

Content as a source for concept: manual survey of sample content

Exercise 1: Skim the text and identify candidate concepts for a taxonomy

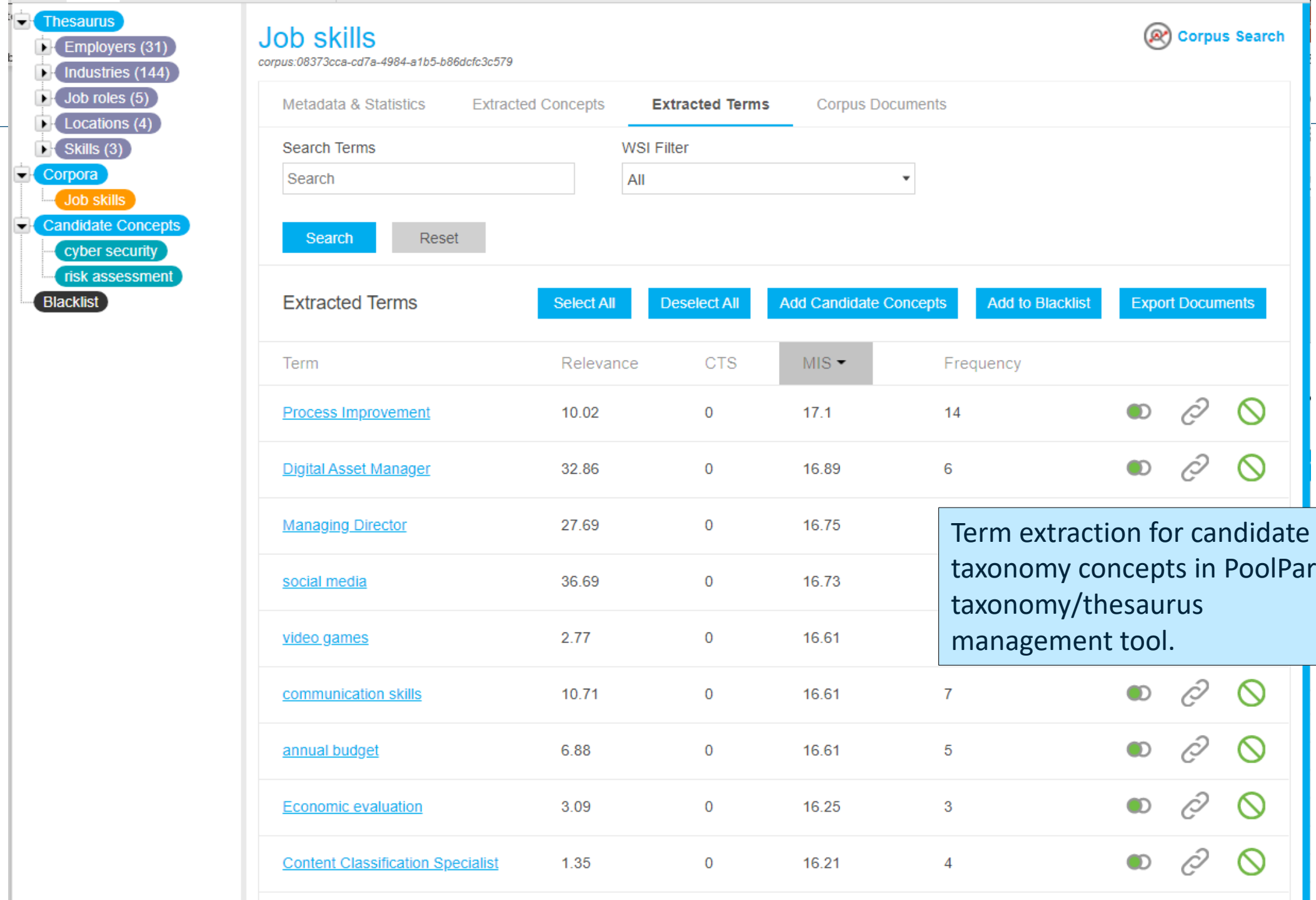
https://www.cnbc.com/2021/03/20/say-goodbye-to-30-plane-tickets-the-era-of-dirt-cheap-flights-is-ending.html

Insert into Zoom chat.

Sources for Concepts/Terms: Content

Content as a source for concept: automatic term extraction

- Using text analytics technologies, such as natural language processing (NLP)
- All extracted terms are suggestions only and should be reviewed for inclusion.
- Considers combinations of general frequency, frequency within a document, relevancy, multiple word co-occurrences.
- Term extraction, intended for auto-tagging, is usually too specific/granular for most taxonomy development.
- Candidate term extraction is most practical when the tool is integrated into taxonomy management software.
- Useful for *enriching* a taxonomy already started from manual identification of terms.



Term extraction for candidate taxonomy concepts in PoolParty taxonomy/thesaurus management tool.

Sources for Concepts/Terms: Users

Users and other stakeholders

- Owner/manager of the controlled vocabulary
  - for some of the top-level concepts
  - more for concepts, rather than for term labels
- Subject matter experts
  - for some or many specific terms and their relationships
  - for concepts and preferred wording
- Various users

  Both the those who tag and those who retrieve content
  - for some or many specific terms
  - for concepts and preferred wording

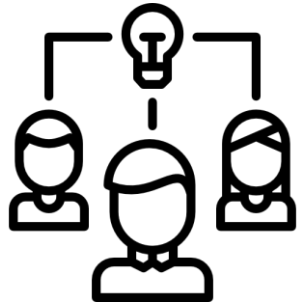
Sources for Concepts/Terms: Users

Users' input for concepts, from:

- Brainstorming sessions
  - For designating vocabularies and facets; gathering top and sample concepts
- Interviews of sample users and stakeholders
  - For scoping, developing user cases, designating vocabularies and facets, and obtaining some sample terms.
- Requested lists of suggested terms
  - For gathering detailed concepts/terms
  - Especially from subject matter experts in technical subjects
- Search log reports
  - For indirect input from users, of how the word what they are looking for

Sources for Concepts/Terms: Users

Brainstorming for user input



1. Identify and name things
2. Consolidate duplicates
3. Remove outliers
4. Group them into categories

Using whiteboards/sticky notes or virtual collaborate tools (e.g. Miro or Mural)

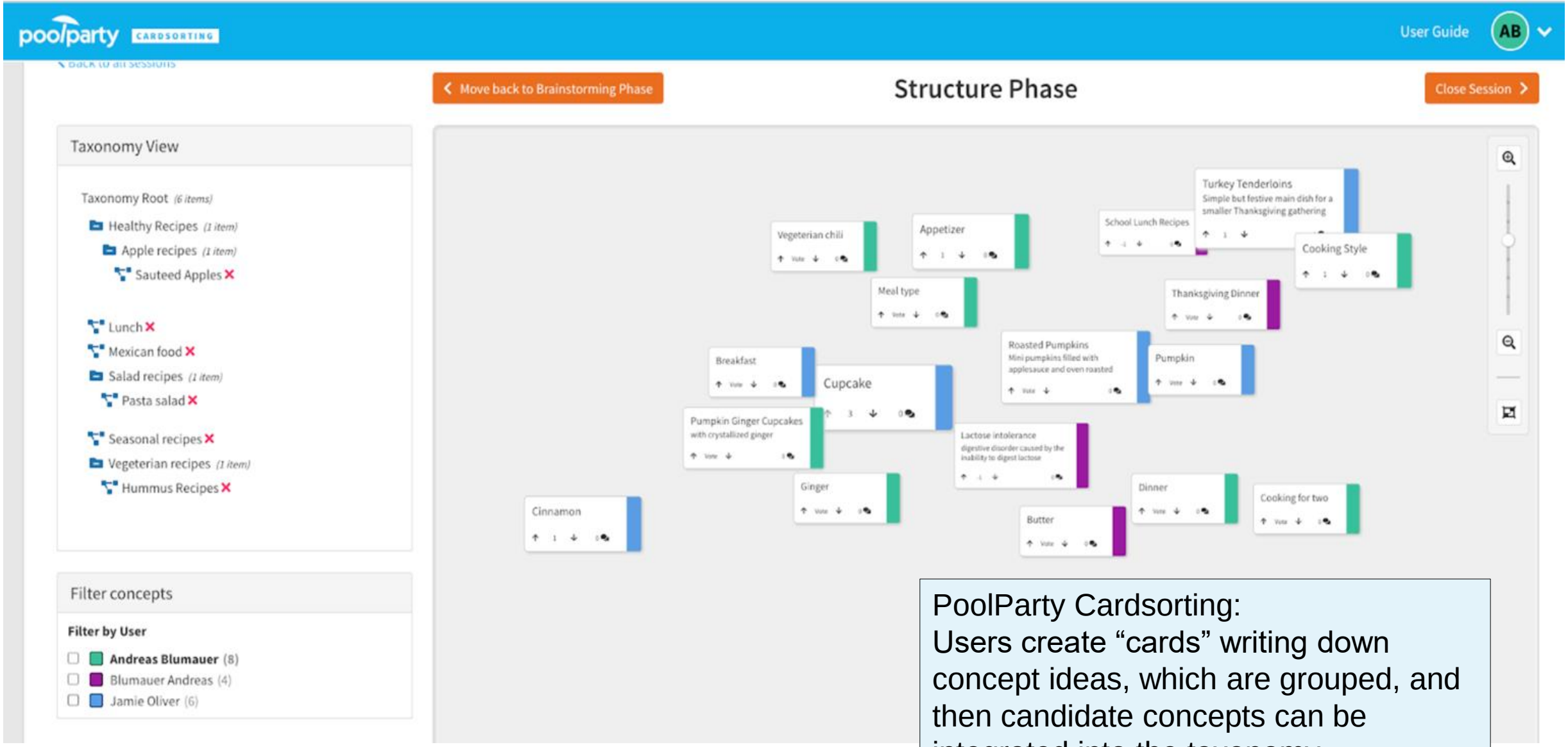
Workshops

- Larger number (5-15), usually in a room together with a facilitator
- Good for intranet/enterprise taxonomies with representatives from different teams
- Often a component of a longer, taxonomy training/introduction workshops

Brainstorming sessions

- Smaller number (3-8) in-person or remote
- Suitable for a taxonomy for any purpose

Sources for Concepts/Terms: Users



PoolParty Cardsorting:
Users create “cards” writing down concept ideas, which are grouped, and then candidate concepts can be integrated into the taxonomy.

Sources for Concepts/Terms: Users

Breakout Session Activity (10 minutes)

Brainstorm taxonomy concepts for the field of **information architecture** for information resources on a new professional association website.

- Join the Zoom breakout room of your group, IAC 1, IAC 2, or IAC 3
- Log into https://cardsorting-demo.poolparty.biz/ and enter your session.
- Start in Brainstorming Phase
  - Create concept cards: enter the name in the Title field (Description is optional.)
  - Vote and comment on other's concept cards
  - Drag to group similar cards together
- Move to Structure Phase
  - Drag and drop concepts into the Taxonomy View tree on the left

Sources for Concepts/Terms: Users

Interviews for user input

- Interviews of sample users are for multiple purposes:
  - To obtain use cases to better design the taxonomy and its UX
  - To obtain use cases to later test the taxonomy
  - To identify taxonomy facets and scope
  - To collect some concepts for the taxonomy
- From different functions that deal with the content
- 1-2 people at once (if from the same function), in person, phone, or video call
- Have prepared sets of questions sent to stakeholders in advance
- Different sets of questions for information users and for information curators (uploading/tagging)
- For information users, different question about how they:
  - find/discover information
  - find desired content items
- For information curators, questions about decisions and issues for tagging/categorizing content

Sources for Concepts/Terms: Users

Questions asked in Berkeley Music & Health taxonomy project of sample potential users

1. If you were to search a music and health database collection of articles, what kind of research would you be doing?
2. What kind of topics might you look for?
3. What would be any specific keywords you might search?
4. Given the different aspects of research studies in music & health, which (any number) of the following would you most likely *initially* want to search (enter into a search box) on (with the remaining aspects getting answered after you retrieve the articles and read them)? [facets listed]
5. For the above aspects that you initially search on, how specific would you want to search? How specific would your search terms be?
6. How would you look up "patient population/type" for your purposes (if at all)? By medical condition, setting (inpatient/outpatient), demographic, etc.? brain lesion sites.
   Acute care, chronic care, country: e.g. Korea

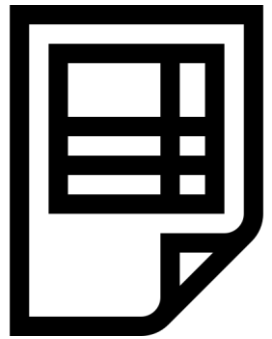
1. When searching a database of articles, are you more likely to…
   1) start with basic search
   2) go straight to advanced search

	A	B	C	D	E	F	G	H	I
1	Proff	Name	**Topics**	**Keywords**	**Initial search aspects**	**Specificity**	**Population**	**Search type**	**Comments**
2	MT	CH	Clinical trials, music therapy methods, Neuro-rehab standardized protocols, Day training, physical therapy, occupational therapy		music therapy method	Start broad, and then get specific. Example of specific: Walking for balance	By condition, such as stroke, traumatic brain injury, cerebral palsy; Not usually by age, but possibly aging/geriatric, not pediatric Not as inpatient outpatient, but rather as acute or chronic	Basic search	Might also look up by country, such as Korea
3	MT	JK	Trauma, mind-body health, adverse child experiences ACEs music and neurobiology, crisis music and disaster, community trauma, collective trauma	GIM – guided imagery, Bonny method of GIM, analytical music therapy, PTSD, anxiety, depression, mindfulness, psychoeducation, resilience	Health issues, including psychological issues Qualitative is important than quantitative studies, so not looking at test type. Whether involving a music therapist or not. Client selected music is of interest.		By diagnosis first. Demographic might be a factor; in person, setting is third	Could go either way. Less experienced so will probably go with defaults	Consider the population as "clients," not "patients"
4	MD	IY	Palliative care, pain management, quality of life, geriatric population, pediatric population, behavior, communication, stress, symptoms, music therapy impact,	Analgesics, PTSD, coping strategies, preparation for death,	1) Patient population, 2) condition (Alzheimer's, cancer, etc.) 3) Location - Hospice, nursing home, hospital; not initially but 4) impact: pain reduction, quality of life, etc.	Would start broad and stay broad and review the articles returned	chronically ill, hospitalized, patients with disabilities, developmental disabilities	Basic Search.	I would want review articles, rather than individual studies. Needs to be evidence-based. What's the impact?
5	PhD	PR	Music and wellness; music and biomedical markers; specific aging-related issues (by diagnosis), community music, chorus/choir, artistic endeavor and aging, cultural engagement	Aging related searches; cultural engagement; social connectedness; pain; enrichment. I currently search for "cultural engagement and aging", "music and wellness", "music and stress", etc.	Who is providing the musical intervention (physician, nurse, music therapist, etc.) ; Health issue, disease, condition; Medical procedure, intervention, or care type; Objective, outcome, or target (pain reduction, anxiety reduction, quality of life, physical rehabilitation, shortened hospital stay, etc.)	Articles indexed with music THERAPY, rather than general music interventions. I'd prefer the terms to be more specific than less; I also mine citations of the specific articles.	By demographic (Older adults) and setting (community vs. hospital vs adult day health care)	Advanced search	
6	MD	DB	How music can support staff sustainability, pain management, stress, anxiety, depression. Distinguishing music therapy and music medicine.	pain, burnout (professional fatigue, compassion fatigue), resilience	1) Type of patients (cancer patients, hospice patience, palliative care patients) and 2) age demographic. Music genre is less important. Genre itself could be a topic.		By condition, site of care (facility, home) and age.	Basic search	Search should be easy and accessible to many people, not just researchers.

Notes from sample users for taxonomy terms from interviews: Berkeley Music & Health taxonomy

Sources for Concepts/Terms: Users

Requesting term lists for user input



- Usually from subject matter experts (SMEs) or domain experts
- In spreadsheets, allowing columns for narrower concepts, synonyms, notes, and comments
- Define the scope to a very specific subject domain.
- Provide example terms to SMEs.
- Provide instructions, guidance, and clarify specificity:
  - need concepts for tagging and search: common topics the content is about
  - do not need a classification scheme, do not need terminology/glossary
  - number range of terms, such as 10-50
- Meet to review, discuss, and clarify suggestions
- Taxonomist edits suggestions and sends back to SME for approval
- Possible multiple iterations

	A	B	C	D	E	F	G
1	**L1**	**L2**	**L3**	**L4**		**Comments**	**Synonyms**
23		Transport and Logistics Services				RT customs and border management	
24			Freight Services				
25			Passenger Services				
26			Transport Integrators				freight forwarders, travel agencies,
27			Transport Logistics Providers				Port Services, Transport Terminals
28			Customs and Border Management			Moved from under Transport Policy and Regulation	
29		Transport and Sustainable Development				Changed from Transport and Deve	Sustainable Transport, Transport and Sustainability
30			Transport and Poverty				
31			~~Transport Impact on Trade~~			move under Transport Economice	
32			~~Transport and Economic Geography~~			move under Transport Economice	
33			Transport and Urban Development				
34			Transport and Millennium Development Goals (MDGs)			[will probably remove and make a synonym to Transport and Sustainable Development--HH]	
35		~~Transport and Social Responsibility~~					
36			Gender and Transport				
37			Transport and HIV-AIDS				
38			Universal Access in Transport			RT Disability	
39			Transport Safety			Move from under Transport Policy and Regulation	
40				Traffic and Road Safety		changed by Heather Hedden	Traffic Safety, Road Safety, Highway Safety
41			Transport Impact on the Environment			Change from Transport and the Environment; move down under Transport and Sustain	
42				Carbon Emissions and Transport			Low-Emissions Transport
43				Transport Infrastructure and Environment			
44		~~Transport Information Systems~~				Move under Transport Infrastructure	
45		Transport Economics				Transport Finance	
46			Congestion				
47			Fuel Taxes				
48			Road Funds				
49			Road Tolls			Toll Roads, Highway Tolls, Road Pricing, Road User Fees	
50			Transport and Economic Geography				
51			Transport Efficiency			Transport Productivity	
52			Transport Employment				
53			Transport Impact on Trade			moved from under Transport and Sustainabile Development	

Key:
Red - Added
Blue - Moved
Purple - Changed
~~Cross-out~~ - to Delete

Sample taxonomy section suggestions from a SME with taxonomist review

Sources for Concepts/Terms: Users

Search log search terms

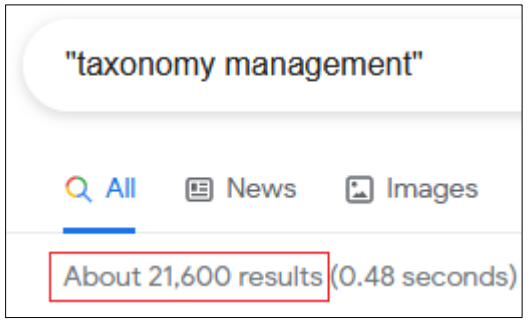
	A	B	C	D	E	F
1	**Search Term**	**Total Unique Searches**	**Include?**	**Type/facet**	**Term equivalencies**	**Other notes**
2	shif	451	y	department/service		
3	dsa	449	n			on Tools menu
4	travel	414	y	topic		
5	staff regulations	380	y	document type		
6	igds	223	n			on Tools menu
7	pardev	220	y	department		
8	home leave	193	y	topic		
9	implementation agreement	191	y	document type		
10	procurement	178	y	department		
11	normes	157	y	department?		
12	exchange rate	151	y	topic		
13	infotec	147	y	department		
14	normlex	136	y	document database		add to Tools menu?
15	teleworking	135	y	topic		
16	dcomm	132	y	department		
17	actrav	126	y	department		
18	iris	126	y	document database	same as iuc	ERP system
19	cafeteria menu	123	n			Put it on the menu or home page
20	437	121	n		Travel policies	IGDS number
21	hrd	118	y	department		
22	iuc	114	y	document database	same as iris	
23	prodoc	112	y	department		
24	education grant	110	n			on Tools menu

Sources for Concepts/Terms

External reference sources

Websites of:

- Wikipedia
- Trade, industry, professional organizations
- Government agencies
- If external/public web taxonomy, competitor/similar service websites



Web search engines (such as Google)

- to compare term usage counts to determine the preferred wording

Other published controlled vocabularies on a term-by-term basis

- Basic Register of Thesauri, Ontologies, and Classifications (BARTOC) https://bartoc.org

➢ Mostly for label names, not for determining concepts to include.

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Term and Label Creation

Whether a concept should be included as a term

1. Is it within the subject-area scope of the controlled vocabulary?
2. Is there enough information on the subject?
   - ➢ sufficient number of current and *anticipated* documents, pages, assets
3. Is it important, likely to be looked up?
4. Do users want and expect it?
   - ➢ Look at search logs

Term and Label Creation

Choosing the preferred term/label wording (the displayed form)

Choosing between two “synonyms”:

Doctors vs. **Physicians**

Movies vs. **Motion pictures**

Cars vs. **Automobiles**

Consider:

1. Wording of terms most likely looked up by the intended users/audience, especially in browsed controlled vocabulary
2. Enforcing organizational/enterprise controlled vocabulary
3. Conforming to academic or professional standards
4. Consistency in style throughout the controlled vocabulary
5. Wording with in the documents/content indexed

Term and Label Creation

Choosing the preferred term/label wording

The other becomes a nonpreferred term/alternative label.

Differentiate closely related terms, or use one as preferred:

Foreign policy vs. **International relations**

Differentiate topics from actions, or use one as preferred:

Contracts vs. **Contracting**

Differentiate broader and more specific concept, or use one:

Electric power plants vs**. Hydroelectric power plants**

Consider likely occurrences of the more specific topic in the content.

Term and Label Creation

Term format and style

- Unambiguous; understood even out of context of the hierarchy.
  Example: **Nursing Certification,** rather than **Certification** as narrower to Nurses
- Consistent capitalization: initial capitalization is recommended.
  Example: **Corporate finance**, rather than **corporate finance** or **Corporate Finance**
- Single words or multi-word phrases; Nouns or noun phrases
  Example: **Employment**; **Part-time employment**
- Countable nouns are usually plural
  Example: **Occupational accidents** (countable); **Occupational health** (not countable)
- Adjectives alone may exist within term lists of characteristics/properties (metadata or facets), but not within hierarchical taxonomies or thesauri. For example, colors, sizes.
- Parenthetical qualifiers may be used for disambiguation, not modification.
  Example: **Walnut (wood)**
- Avoid term inversions (e.g. noun, adjective) because labels are searchable
  Example: **Racial discrimination**, not **Discrimination, racial**

Concept and Label Creation

Term format and style: choices to make and document

- Capitalization style
- Use of abbreviations and acronyms
- Use of special characters
- Use of parenthetical qualifiers
- Use of singular or plural for Concept Scheme names
- Label length limit
- Approved authoritative sources for preferred label names of technical concepts
- Special style for named entities (person names, organization names, place names, etc.)
- Any different style rules for Alternative labels
- When to use Hidden labels
- When to add scope notes and preferred style for notes
- When to add definitions and preferred style for definitions

Variants, Nonpreferred Terms, or Alternative Labels

- Defined: Approximately synonymous words or phrases to refer to an equivalent concept, for the context of the controlled vocabulary and the set of content.
- Purpose: To capture different wordings of how different people might describe or look up the same concept or idea and used as alternative entries.
  - Differences between that of the author and the user/reader
  - Differences between that of the indexers and the end-users
  - Differences among different users/readers
- Serving as “multiple entry points” to look up and retrieve the desired content, as do double posts or *See* references in an index.
- Enabling consistent indexing

Variants, Nonpreferred Terms, or Alternative Labels

Examples

from
Gale Subject
Thesaurus

Conflict management
Conflict resolution
Managing conflict

Wills
Codicils
Last will and testament
Testaments (Wills)

Influenza
Flu
Grippe

Movies
Cinema
Films (Movies)
Motion pictures
Movie genres

Telecommunications industry
Communications industry
Digital transmission industry
Interexchange carriers
Telecommunications services industry
Telephone holding companies
Telephone industry
Telephone services industry

Environmental management
Adaptive management (Environmental management)
Environmental control
Environmental stewardship
Natural resource management
Stewardship (Environmental management)

Piano music [no variants]

Variants, Nonpreferred Terms, or Alternative Labels

When to use

Not needed:

- A very small, browsable taxonomy, where all can be seen or easily scrolled to (such as in facets) *and* tagging is manual

Needed:

- If a controlled vocabulary is too large to be all seen in one view with minimal scrolling.
- If a taxonomy will be searched upon and not just browsed.
- If automated indexing/auto-classification/auto-categorization is implemented.
- Whether it’s called a taxonomy or thesaurus does not matter.

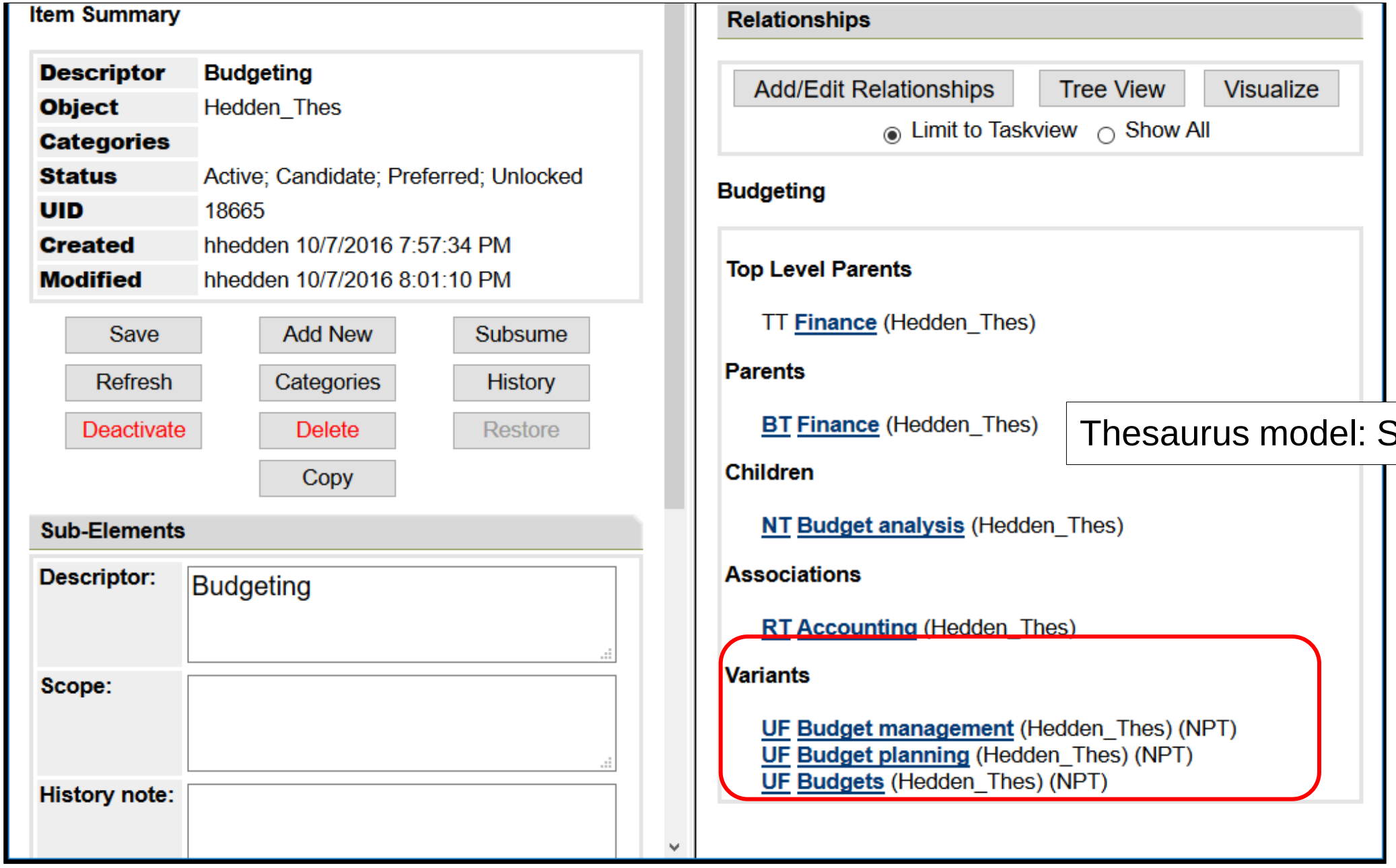
Variants, Nonpreferred Terms, or Alternative Labels

Nonpreferred Term

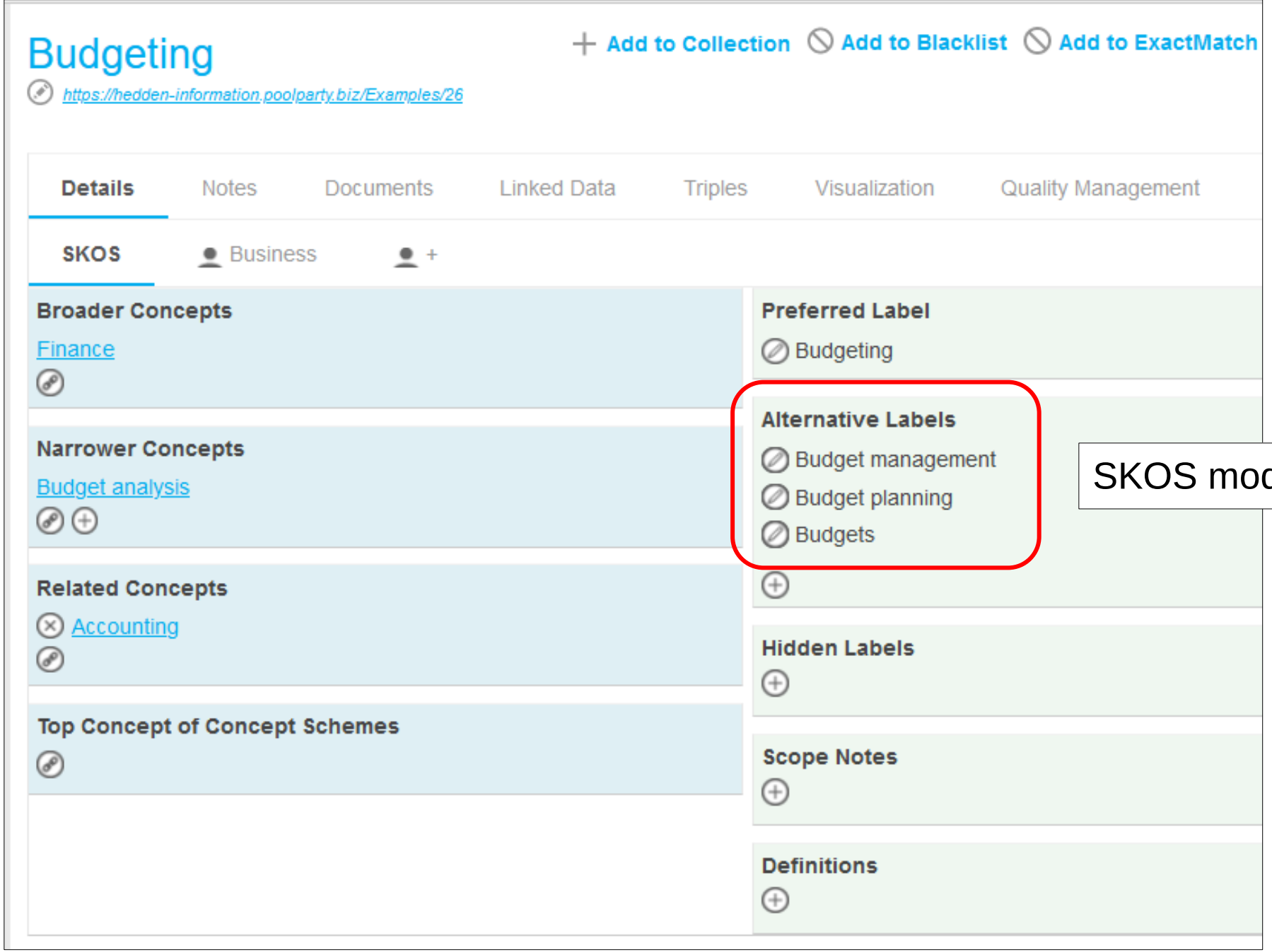
- Formal designation in thesauri, in accordance with ANSI/NISO Z.39-19 and ISO 25964 and standards.
- Shortened as *NPT.*
- Points to a *Preferred term.*
- Considered a kind of **relationship** of the Equivalency type.
- Reciprocity of relationship, pointing in both directions: USE and UF (use and used for/use for).
- Both preferred terms and non-preferred terms are “terms.”

Alternative Label

- Formal designation for SKOS (Simple Knowledge Organization System) vocabularies, a World Wide Web (W3C) recommendation.
- Shortened as *altLabel.*
- Associated with a *Preferred label.*
- Alternative labels are part of a concept’s **attributes**, not equivalent terms and not connected by relationships.
- Instead of terms, there are concepts, each with any number of labels
- Concepts have one preferred label (for each language) and any number of alternative labels and hidden labels (for each language).



Thesaurus model: Synaptica



SKOS model: PoolParty

Variants, Nonpreferred Terms, or Alternative Labels

Guidelines for using variants

- A concept may have any number of (multiple) variants, or it may have only a single preferred name (no variants).
- A variant points to only a single preferred term.
  Exceptions:
  - Print thesauri may permit a “multiple-use” reference (Use A and B), but for simplification, most vocabulary management software does not permit it.
  - Auto-categorization synonym ring variants that have weighting or scoring
- Variants may be displayed to the end-user or they may not be.

Variants, Nonpreferred Terms, or Alternative Labels

Sources for variant terms

- Same sources as for concepts and preferred terms
  - Survey/audit of the content and terms used
  - Search query logs and other internal usage data
  - External sources: websites, Wikipedia, other taxonomies and controlled vocabularies, book tables of contents, etc.
- Creative changes of terms (after verification of variant term usage in search)
- Not to be used as a source:
  Dictionary-type thesaurus, such as *Roget's Thesaurus of English Words and Phrases* or thesaurus-dictionary websites

Variants, Nonpreferred Terms, or Alternative Labels

Types of variants include:

- synonyms
- quasi-synonyms
- variant spellings
- lexical variants
- foreign language names
- acronyms/spelled out
- scientific/popular names
- antonyms (for characteristics)
- older/current names
- phrase variations (in print)
- narrower terms that are not preferred terms

Variants, Nonpreferred Terms, or Alternative Labels

Types of variants include:

- synonyms: Cars / Automobiles
- quasi-synonyms: Politics / Government
- variant spellings: Taoism / Daosim; Email / E-mail
- lexical variants: Selling / Sales; Hair loss / Baldness
- foreign language names: Ivory Coast / Côte d'Ivoire
- acronyms/spelled out: GDP / Gross domestic product
- scientific/popular names: Neoplasms / Cancer
- antonyms (for characteristics): Flexibility / Rigidity
- older/current names: Near East USE Middle East
- phrase variations (in print): Unions, labor USE Labor unions
- narrower terms that are not preferred: Genetic engineering USE Biotechnology

Variants, Nonpreferred Terms, or Alternative Labels

Exercise 1: Create multiple variant terms/labels for each in the context of an Intranet

Ethics

Compensation

Insert into Zoom chat.

Concept and Label Creation

Terms in multiple languages

- Options:
  1. Link separate language taxonomies at equivalent concepts
     - Allows for differing concepts and differing hierarchies, relationships
  2. Develop a single multilingual taxonomy with concept labels in multiple languages.
     - Assumes concept exists in both/all languages (although occasional missing language for a concept is OK) the same hierarchy and relationships

- Links or preferred labels in each language must be exact translations for the vocabulary to function bi-directionally.
- Most vocabulary management software supports multilingual vocabularies.

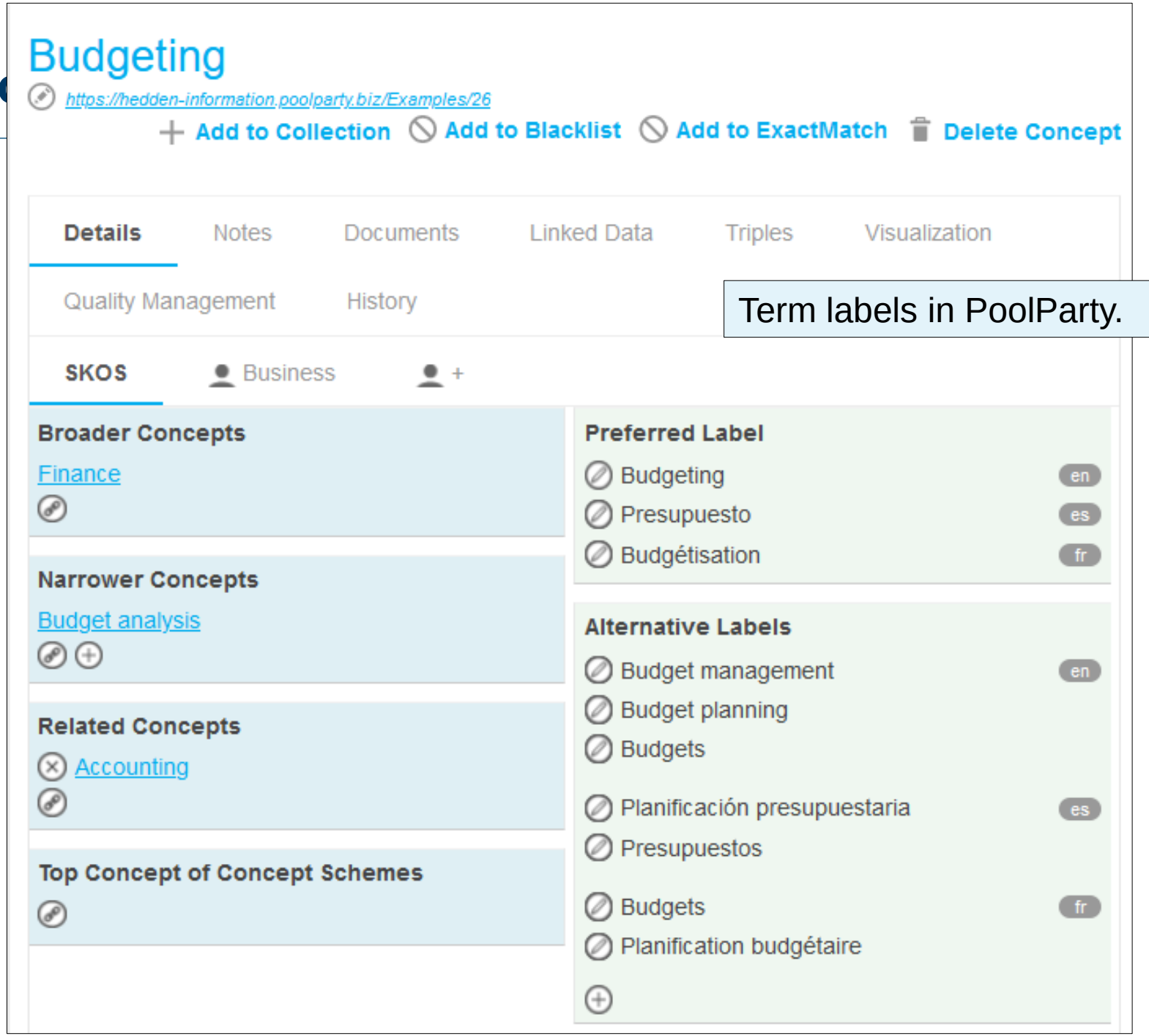
Concept and Label Creati

The SKOS model is especially suited for multilingual vocabularies.

A single concept has a preferred label, one in each language, and any number of alternative labels in each language.

Alternative labels do not link to translations of each other, so may vary for each language.

Broader and narrower concepts are presented in the in default selected language, but default language can be changed.



Term labels in PoolParty.

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Relationships

Types of relationships between concepts or terms

Between concepts or preferred terms:

1. Hierarchical
   Broader term (BT) / Narrower term (NT)
   Broader concept / Narrower concept
2. Associative
   Related term (RT)
   Related concept
3. Customized, semantic relationships (more specific)

Relationships are reciprocal (bi-directional) between terms/concepts.

Between preferred and nonpreferred terms in the traditional thesaurus model

4. Equivalence
   Use (USE) / Used for (UF)

Relationships: Hierarchical

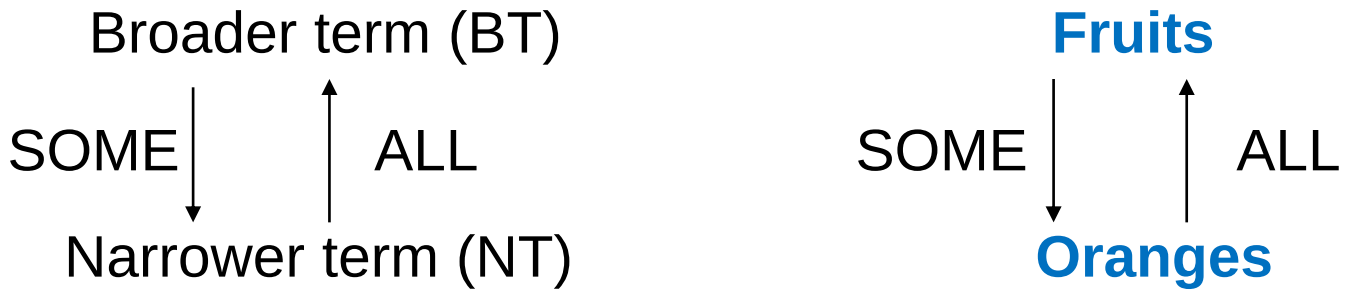
Hierarchical relationships

- Broader-narrower / Topic-subtopic / Parent-child / Superordinate-Subordinate
- Required feature of both thesauri and taxonomies
- Thesaurus designation of BT / NT (broader term / narrower term)
- SKOS designation: Broader concept / Narrower concept
- Terms usually have more than one narrower term (NT), unless they are the most specific terms in the vocabulary.
- On occasion, a term may have more than one broader term (BT), referred to as polyhierarchy.

Relationships: Hierarchical

Hierarchical relationships

Reciprocal (bi-directional) relationships, but asymmetrical



Fruits NT **Oranges** **Oranges** BT **Fruits**

Three types:

1. Generic – Specific
2. Generic – Named entity instance: Common noun – Proper noun
3. Whole – Part

Relationships: Hierarchical

Generic – Specific

Category or class

– members

– more specific types

Narrower term “is a” or
“are a kind of” broader term

Dogs

NT **Puppies**

Financial services

NT **Investment services**

Romance languages

NT **Italian language**

Relationships: Hierarchical

Generic – Instance

Common noun

– Proper noun

Narrower term “is a” broader term

Smartphones

NT **Samsung Galaxy**

Golfers

NT **Woods, Tiger**

Holidays

NT **Thanksgiving**

Relationships: Hierarchical

Whole – Part

Concept or entity

– part

– subentity

Narrower term “is in” broader term.

Must be an integral part that cannot be or is not intended to be taken out.

New England

NT **Massachusetts**

Gastrointestinal system

NT **Stomach**

United Nations

NT **United Nations General Assembly**

Computers

NT **Motherboards**

Relationships: Hierarchical

Exercise 2: Suggest a broader term and a narrower term for each of the following:

Hospitals

Teachers

Webinars

Insert into Zoom chat.

Relationships: Associative

Associative relationships

- Suggestions to the user of possible related terms of interest
- Like *See also* in an index
- Required feature of thesauri; Optional feature of taxonomies
- RT (Related term) in thesaurus model; Related concept in SKOS
- Symmetrical bi-directional relationship
- Between concepts within the same hierarchy or in different hierarchies (within the same concept scheme or across different concept schemes)

Related concept relationship purposes

- Guidance for end-user browsing and navigation, if supported in front-end applications: suggestions of related concepts of potential interest
- Support for those doing manual tagging, to find potential additional concepts to tag.
- Leveraged and can be weighted in auto-tagging, providing context for NLP.

Relationships: Associative

Associative relationships between terms in different hierarchies:

- Process and agent
- Process and instrument
- Process and counter-agent
- Action and property
- Action and product
- Action and target
- Cause and effect
- Object and property
- Object and origins
- Raw material and product
- Discipline and practitioner
- Discipline and object

Relationships: Associative

Associative relationships between terms in different hierarchies:

- Process and agent: Skiing RT Skiers
- Process and instrument: Ventilation RT Fans (Equipment)
- Process and counter-agent: Bacterial infections RT Antibiotics
- Action and property: Environmental protection RT Pollution
- Action and product: Glassblowing RT Glass containers
- Action and target: Appliance repair RT Appliances
- Cause and effect: Hurricanes RT Storm surges
- Object and property: Plastics RT Elasticity
- Object and origins: Petroleum RT Oil wells
- Raw material and product: Timber RT Wood products
- Discipline and practitioner: Chemistry RT Chemists
- Discipline and object: Literature RT Books

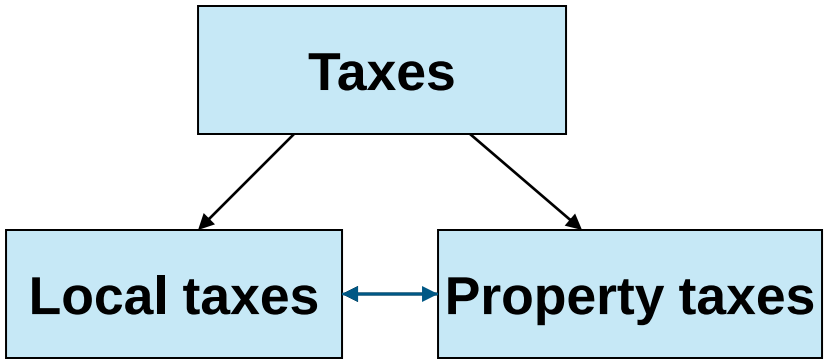
Relationships: Associative

Between terms/concepts within the same hierarchy

1. Having a shared broader term (siblings) *and* overlapping meaning

   Required associative relationship, according to thesaurus standards

Example:



Local taxes RT **Property taxes**

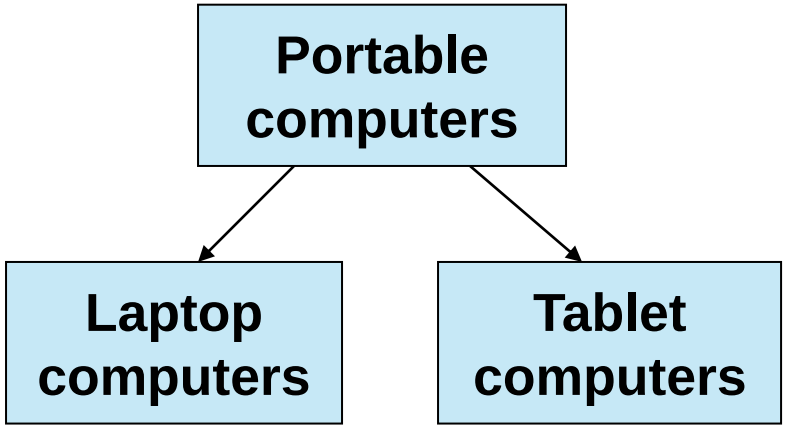
Relationships: Associative

Between terms/concepts within the same hierarchy

2. Having a shared broader term (siblings) *without* overlapping meaning

   While not incorrect, better style is to avoid adding such "valueless" relationships.

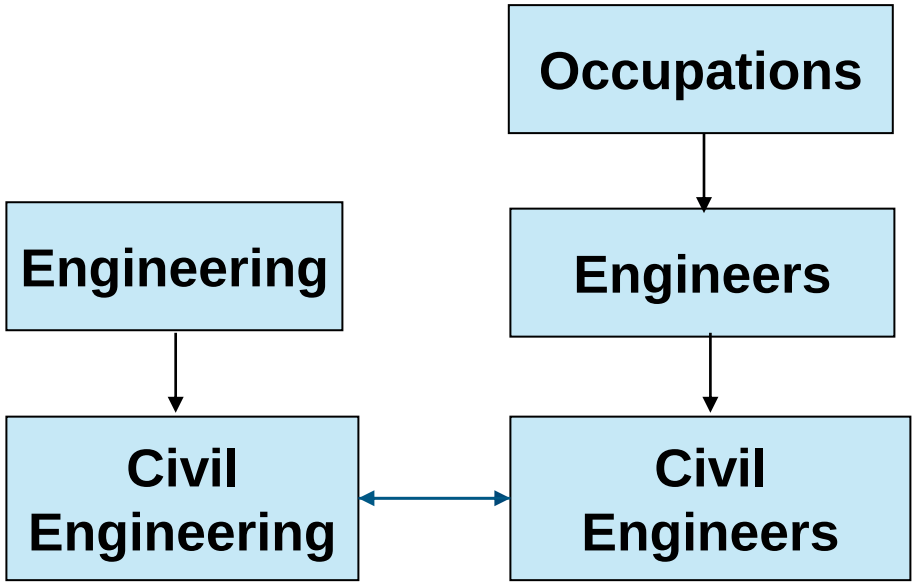
Example:



Not related terms

Relationships: Associative

Between terms/concepts in different hierarchies



Acceptable, by standards.

However, if they are in separate concept schemes, the concepts could be related by semantic relations instead.

Civil Engineering RT **Civil Engineers**

Relationships: Customized, Semantic

Specific/customized relationships

- Relationships containing meaning: “semantic”
- May be variations on hierarchical or associative relationships, but usually associative (more specific types of Related concept).
- Reciprocal, but asymmetrical, or directional.
- Specific enough to convey the necessary meaning, but not uniquely specific.
- Relationships are between terms of different types, across different designated *categories* or *classes*.
- Taxonomist defines the relationships (and their codes) and the categories/classes.
- Trending toward ontology. May be called “ontology lite.”

Relationships: Customized, Semantic

Specific/customized relationships

Sample variations on the associative relationship (RT):

- Has produced the work (WRK) / Created by (CRE)
  - **Twain, Mark** WRK **The Adventures of Tom Sawyer**
  - **The Adventures of Tom Sawyer** CRE **Twain, Mark**
- Produces the product (PRD) / Is manufactured by (MAN)
  - **Apple Inc.** PRD **iPod**
  - MAN **Apple Inc.**
- Has member affiliation with (AFF) / Has members (MEM)
  - **Saudi Arabia** AFF **OPEC**
  - MEM **Saudi Arabia**
- For treating (TRE) / Can be treated with the drug (DRUG)
  - **ACE inhibitors** TRE **Hypertension**
  - DRUG **ACE inhibitors**

Relationships: Customized, Semantic

Specific/customized relationships

Sample semantic relationships between the following terms (on next slide):

Term:	Category or class:
Toyota Corolla	Product name
Automobiles	Product type
Motor vehicle industry	Industry
Toyota Motor Corporation	Company name
Akio Toyoda	Executive name

Relationships: Customized, Semantic

Sample relationships

Product type – Product name
Automobiles – Toyota Corolla
Hierarchical relationship (instance)

Type of (TYP) / Includes the brand (BRD)
Toyota Corolla TYP Automobiles
Automobiles BRD Toyota Corolla

Industry – Company name
Motor vehicle industry – Toyota Motor Corp.
Associative relationship

Has members (MEM) / Is affiliation with (AFF)
Motor vehicle industry MEM Toyota Motor Corporation
Toyota Motor Corporation AFF motor vehicle industry

Company name – Product name
Toyota Motor Corp. – Toyota Corolla

Produces (PRO) / Manufactured By (MAN)
Toyota Motor Corporation PRO Toyota Corolla
Toyota Corolla MAN Toyota Motor Corporation

Industry – Product type
Motor vehicle industry – Automobiles

Motor vehicle industry PRO Automobiles
Automobiles MAN Motor Vehicle Industry

Company name – Executive name
Toyota Motor Corp. – Akio Toyoda

Is managed by person (MPN)/Manages company (MC)
Toyota Motor Corp. MPN Akio Toyoda
Akio Toyoda MC Toyota Motor Corporation

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Software

Types of software used in vocabulary management

- Spreadsheet software (Excel)
- Dedicated thesaurus/ontology (taxonomy) management software
- Taxonomy creation & editing module of a content management, document management, digital asset management, collaborative software (SharePoint)
- Taxonomy creation & editing module of auto-classification (automated indexing) software
- Vertical market software for creating classification structures
- Proprietary programs developed in-house in organizations with large/core taxonomy management needs

Software

Using spreadsheet software for a simple hierarchy

- For hierarchical relationships only, not for associative, and limited use for alternative labels
- Not more than 4-5 levels deep
- Subject to errors, such as duplications
- For small taxonomies or for starter top levels of larger taxonomies
- Individual worksheets can be saved as CSV files and imported into other systems, including thesaurus management software

Software

Excel use example

	A	B	C	D	E	F	G	H
1	**Level 1**	**Level 2**	**Level 3**	**Level 4**	**Level 5**			
2	Diseases and disorders							
3		Age related diseases						
4			Alzheimer's disease					
5			Cataracts					
6			Gout					
7			Osteoarthritis					
8			Osteoporosis					
9			Sarcopenia					
10			Senile dementia					
11			Dehydration					
12			Hyponatremia					
13			Metabolic syndrome X					
14			Water intoxication					
15		Cardiovascular diseases and disorders						
16			Aneurysm					
17			Arteriosclerosis					
18				Atherosclerosis				
19					Coronary artery disease			
20			Heart diseases and disorders					
21				Angina pectoris				
22				Arrhythmias				
23					Atrial arrhythmias			
24					Heart block			
25					Ventricular arrhythmias			
26				Athletic heart syndrome				
27				Cardiomyopathy				
28				Carditis				
29				Coronary artery disease				

Software

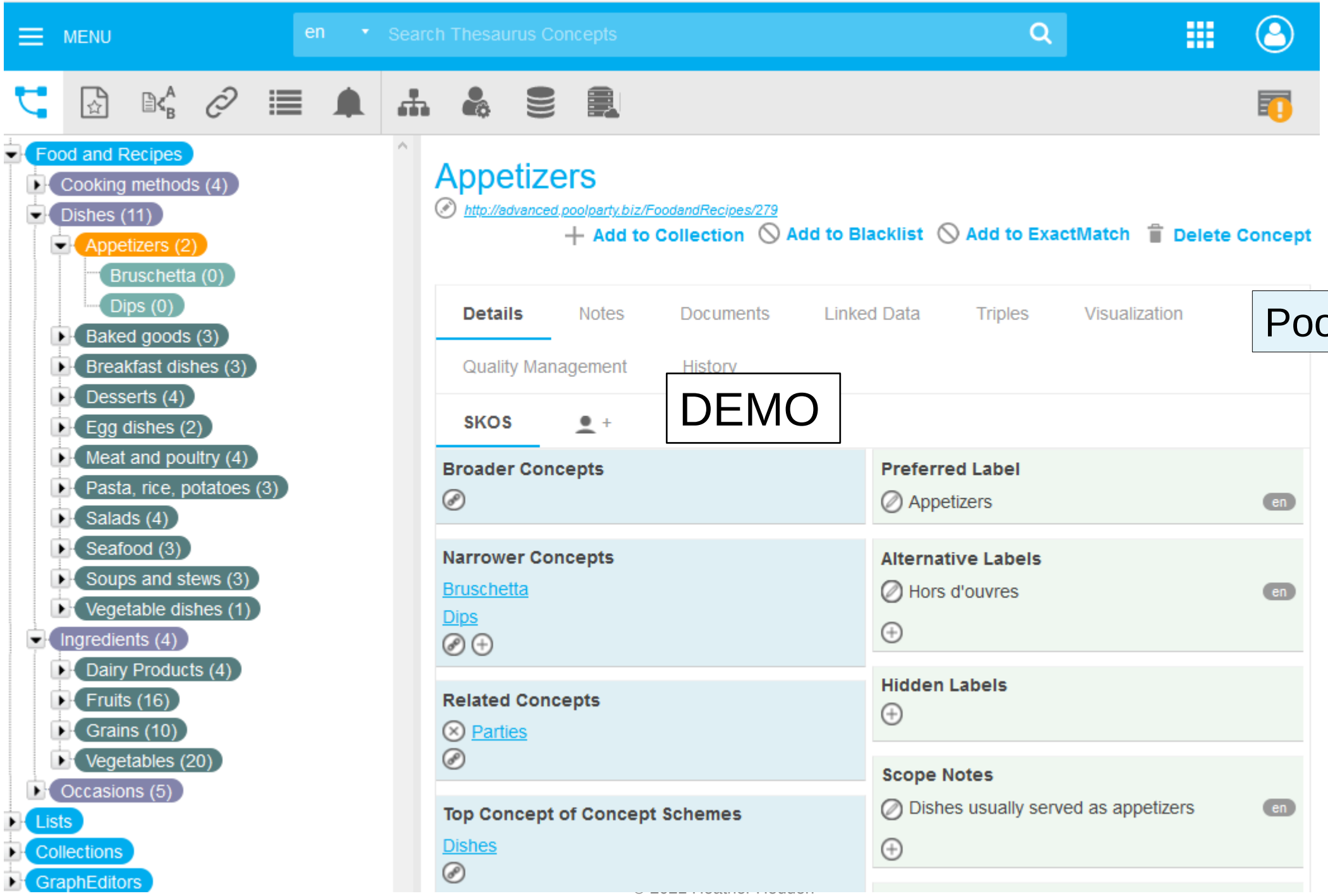
Why dedicated taxonomy/thesaurus management software

- Quality control
  - Support of ANSI/NISO Z39.19 quality standards
  - Maintain terms/concepts and all their relationships and labels
    - As reciprocals
    - When renaming, merging, subsuming, or deleting concepts or branches
- Centrally manage taxonomies used in multiple systems (website, intranet, CMS, DAM, custom applications, etc.)
- Vocabularies in SKOS and/or other interoperable standards
- Scales from taxonomies to thesauri to ontologies
- Ease of use
- Additional features

Software

Taxonomy/thesaurus management software additional features

- scope notes, definitions, and attributes on concepts
- term searching
- batch additions of concepts
- single-step new term & relationship creation
- single-step branch (term and narrower terms) moving
- drag & drop relationship adding
- user-defined relationships, attributes, classes (or ontology support)
- multilingual taxonomy support
- taxonomy project linking or mapping
- support for linked data
- APIs or connectors to CMSs or search systems
- auto-indexing or auto-categorization add-on module


MENU
en
Search Thesaurus Concepts
Food and Recipes
Cooking methods (4)
Dishes (11)
Appetizers (2)
Bruschetta (0)
Dips (0)
Baked goods (3)
Breakfast dishes (3)
Desserts (4)
Egg dishes (2)
Meat and poultry (4)
Pasta, rice, potatoes (3)
Salads (4)
Seafood (3)
Soups and stews (3)
Vegetable dishes (1)
Ingredients (4)
Dairy Products (4)
Fruits (16)
Grains (10)
Vegetables (20)
Occasions (5)
Lists
Collections
GraphEditors
Appetizers
http://advanced.poolparty.biz/FoodandRecipes/279
Add to Collection
Add to Blacklist
Add to ExactMatch
Delete Concept
Details
Notes
Documents
Linked Data
Triples
Visualization
Quality Management
History
SKOS
DEMO
Broader Concepts
Narrower Concepts
Bruschetta
Dips
Related Concepts
Parties
Top Concept of Concept Schemes
Dishes
Preferred Label
Appetizers
en
Alternative Labels
Hors d'ouvres
en
Hidden Labels
Scope Notes
Dishes usually served as appetizers
en

PoolParty

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Humanities & Social Sciences

- Anthropology
- Art & Humanities
- College Success and Study Skills
- Communication Studies
- Counseling
- Criminal Justice
- Developmental English
- Early Childhood Education
- Education
- English
- History
- Human Services
- Mass Communication
- Music
- Philosophy
- Political Science
- Psychology
- Radio Television & Film
- Religion
- Social Work
- Sociology
- Theatre
- World Languages

Science, Technology & Mathematics

- Agriculture
- Astronomy
- Biology
- Chemistry
- Computing & Information Technology
- Developmental Math
- Earth Sciences
- Engineering
- Environmental Science
- Forensic Science
- Geography
- Health
- Mathematics
- Nutrition
- Physics
- Statistics
- Veterinary Technology

Business & Economics

- Accounting
- Business Communication
- Business Law
- Career Success
- Decision Sciences
- Economics
- Finance
- General Business
- Introduction to Business
- Keyboarding
- Keyboard/Computer Education
- Management
- Marketing
- Office Technology
- Taxation

Professional & Career

- Automotive & Trucking
- Beauty & Wellness
- Career Education
- Culinary, Hospitality, Travel & Tourism
- Emergency Services
- General Interests & Hobbies
- Health Care
- Media Arts & Design
- Paralegal
- Trades

Higher education disciplines of Cengage Learning

www.cengage.com/all-disciplines

AGROVOC thesaurus

http://aims.fao.org/standards/agrovoc/functionalities/search

- processes
- products
  - agricultural products
    - bioproducts
    - out of season products
    - perishable products
  - animal products
    - carcasses
    - cocoons
    - eggs
      - egg white
      - egg yolk
      - shell-less eggs
      - soft shelled eggs
    - fish
    - hive products
      - beeswax
      - honey
      - honeycomb extracts
      - propolis
      - royal jelly
    - ivory
    - livestock products
    - meat
    - milk
    - wildlife products
  - artificial products
  - biodegradable products
  - byproducts
  - commodities

Structural Design: Hierarchies

Hierarchies:

The extension of hierarchical relationships to include all concepts in a single or limited number of tree structures. (A single taxonomy may have more than one top-term hierarchy.)

Hierarchical taxonomy purposes

1. Serving users who are browsing, exploring, discovering, not searching, to whom the hierarchy is displayed. Users don't even have to know the first word or few letters, as in search or alphabetical browsing.
2. Instructing users on appropriate classification.
3. Providing context for concepts for manual indexers/taggers so that they tag with the correct concept.
4. Providing the context of a broader concept and thus meaning to aid in auto-classification.
5. Enabling "recursive"/"rolled up" retrieval results (A concept retrieves what is indexed to it *and* what is indexed to each on of its narrower concepts, all together.)
6. Better organizing and management of a taxonomy.

Structural Design: Hierarchies

Hierarchical taxonomy suitability:

- Content that is naturally categorizable:
  products, industries, government agencies, academic disciplines, scientific things, technologies
- Small and medium-sized taxonomies. (100s of concepts)
  Very large controlled vocabularies are not so suitable.

Otherwise, a thesaurus structure might be better.

Structural Design: Hierarchies

Issues in creating hierarchies

- Depth and breadth of hierarchy levels
- Top concept usage – As indexing concept or just labels
- Recursive retrieval – Broader term additionally retrieves content indexed with all narrower terms or not
- Sorting display options for concepts at the same level.
- Polyhierarchies – Terms with more than one broader term
- Accommodating user interface designs

Structural Design: Hierarchies

Depth vs. breadth of hierarchy levels

Deep: Many levels (industrial classification style with 10-20 upper level terms)

Industries
- Transportation services
-- Air transportation
--- Schedule air transportation services
---- Scheduled air freight transportation services

Broad: Many terms per level (job search sites, 50 - 80 terms per level)

Industries Second levels at select terms only: **Healthcare, Sales**
- Accounting/Auditing
- Administrative Support Services
- Advertising/Marketing/Public Relations
- Aerospace/Aviation/Defense
- Agriculture, Forestry, & Fishing
- Airlines

Structural Design: Hierarchies

Depth vs. breadth of hierarchy levels: decision factors

- Display interface horizontal and vertical space
- Multiple clicks to deeper levels
- More levels lead to less consistency across levels.
- User needs, and expectations
  Industry experts, internal employees, general public, students, etc.

Structural Design: Hierarchies

Top concept usage for indexing or just as an organizational label

Fictional work
- NT Folktale
- NT Novel
- NT Novella
- NT Play
- NT Poem
- NT Short story

Information technology occupations
- NT Computer consultant
- NT Computer engineer
- NT Computer imaging specialist
- NT Computer scientist
- NT Information technology manager
- NT Information technology specialist
- NT Software developer
- NT Systems analyst

Health care
- NT Anatomy and physiology
- NT Clinical procedures
- NT Diagnosis
- NT Diseases and disorders
- NT Drugs
- NT Health care profession
- NT Health care services
- NT Medical equipment and supplies
- NT Medical specialties
- NT Patients
- NT Personal health
- NT Treatments

Examples of **Broad Level Access Terms** not used for indexing, from Cengage Learning vocabularies

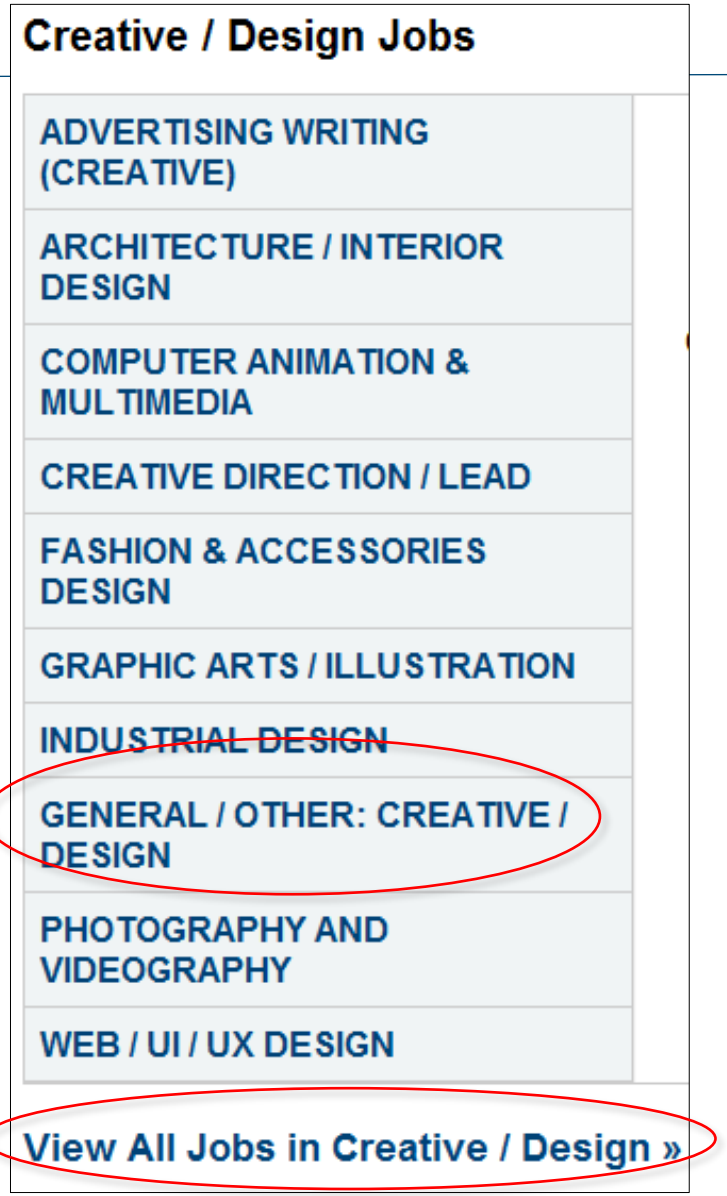
Structural Design: Hierarchies

Recursive or “rolled up” retrieval

Does a term for which there are narrower terms retrieve the sum or all of what is indexed with all of the narrower terms?

- If not, then its meaning is the broader topic “in general.”
- If yes, then an additional term for “in general” might need to be created.

http://jobmarket.nytimes.com/jobs

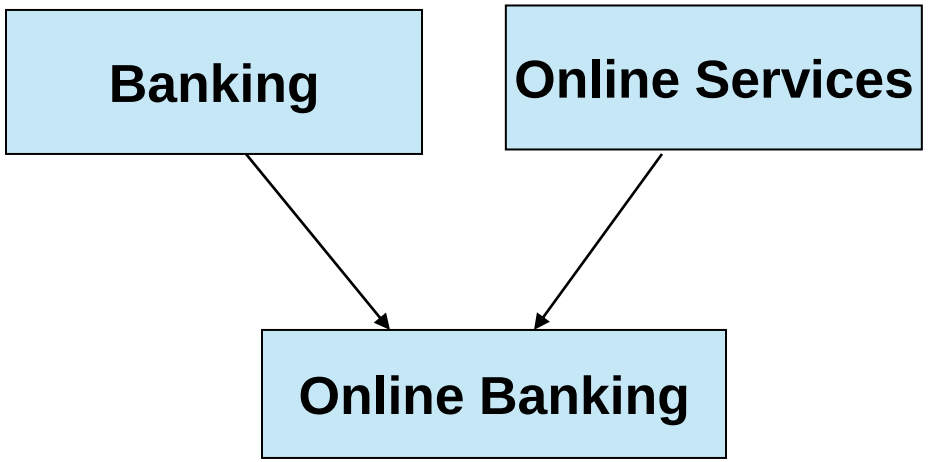


Structural Design: Hierarchies

Polyhierarchies

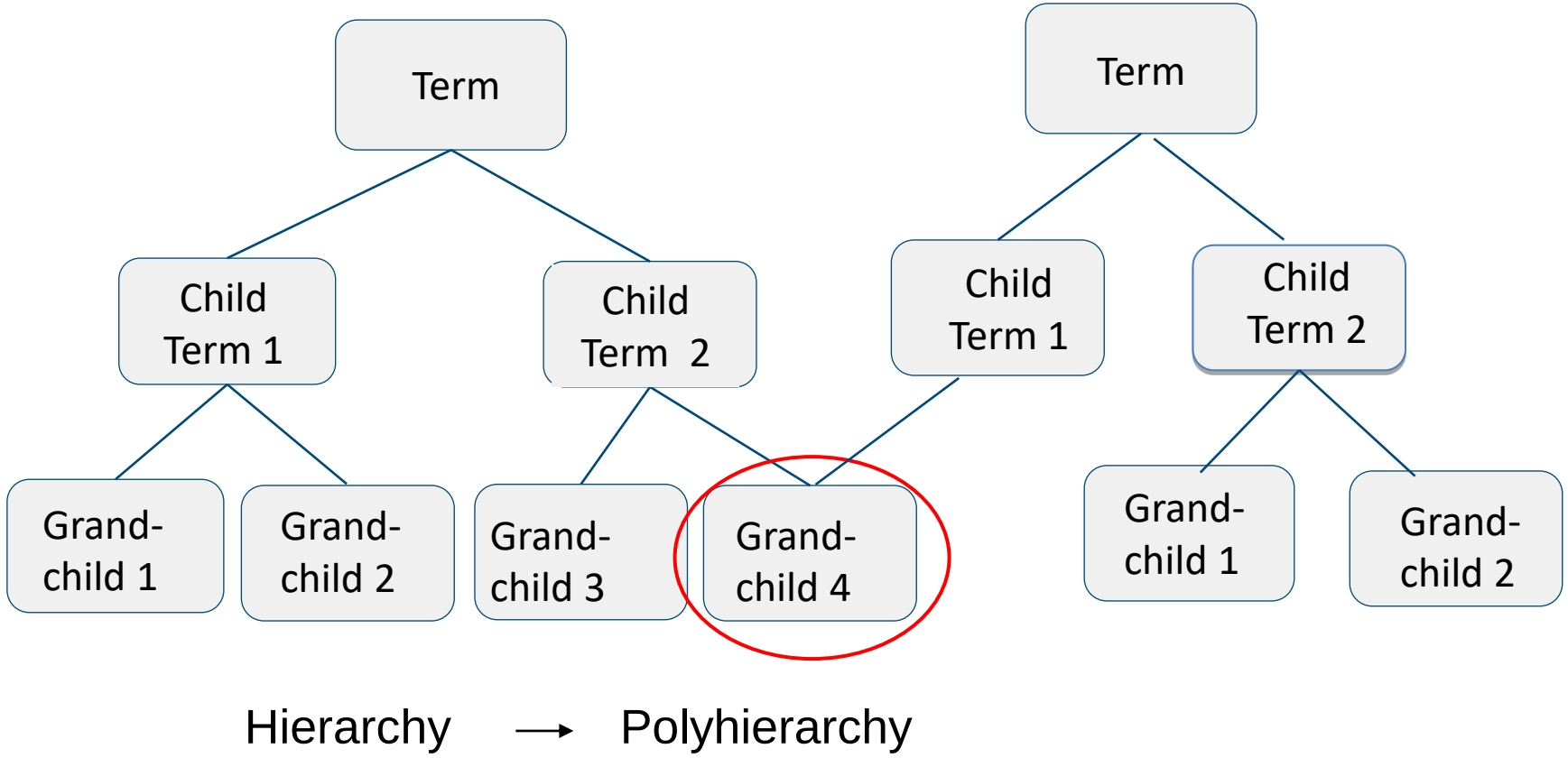
Sometimes a term can have two or more broader terms.

- Polyhierarchy is permitted if the hierarchical relationship is valid in both/all cases
- Remember “All-and-Some” test for each generic hierarchical relationship
- Systems may or may not support it.



Structural Design: Hierarchies

Polyhierarchies

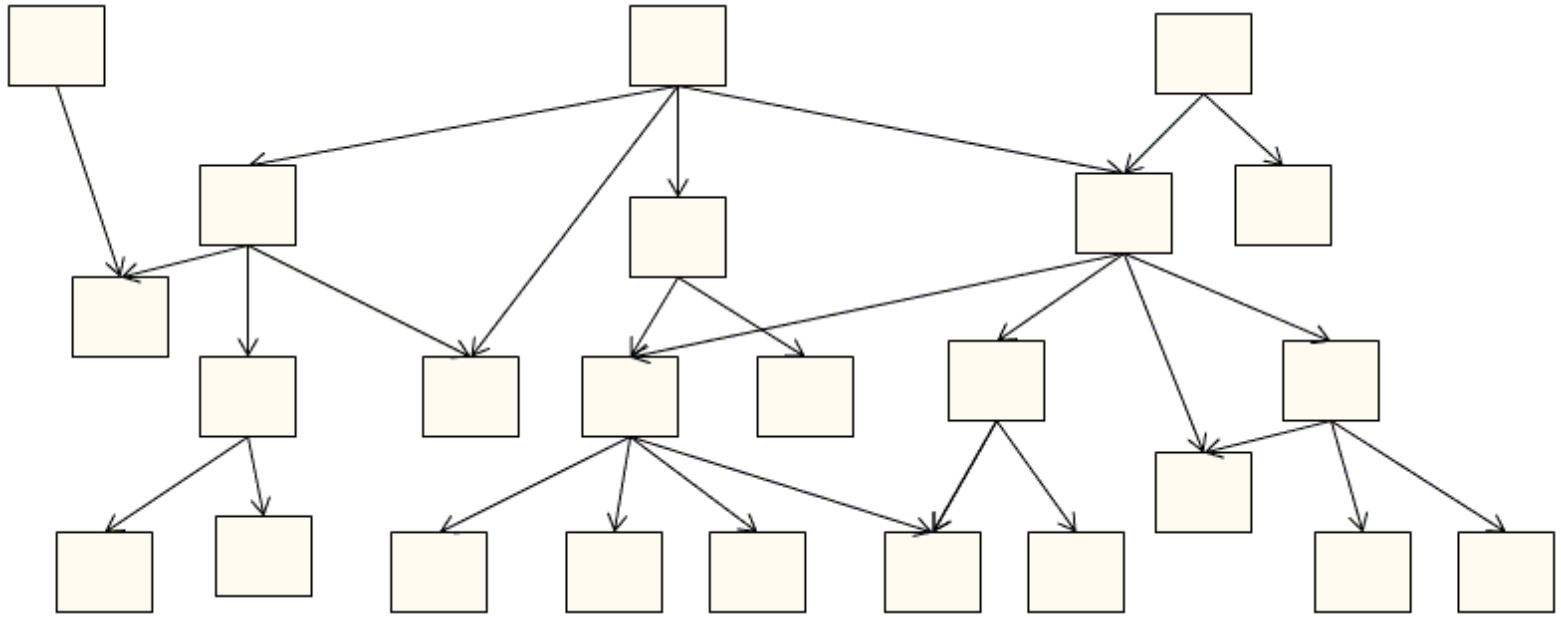


Structural Design: Hierarchies

Polyhierarchies – Disadvantages

Problems with excessive polyhierarchies:

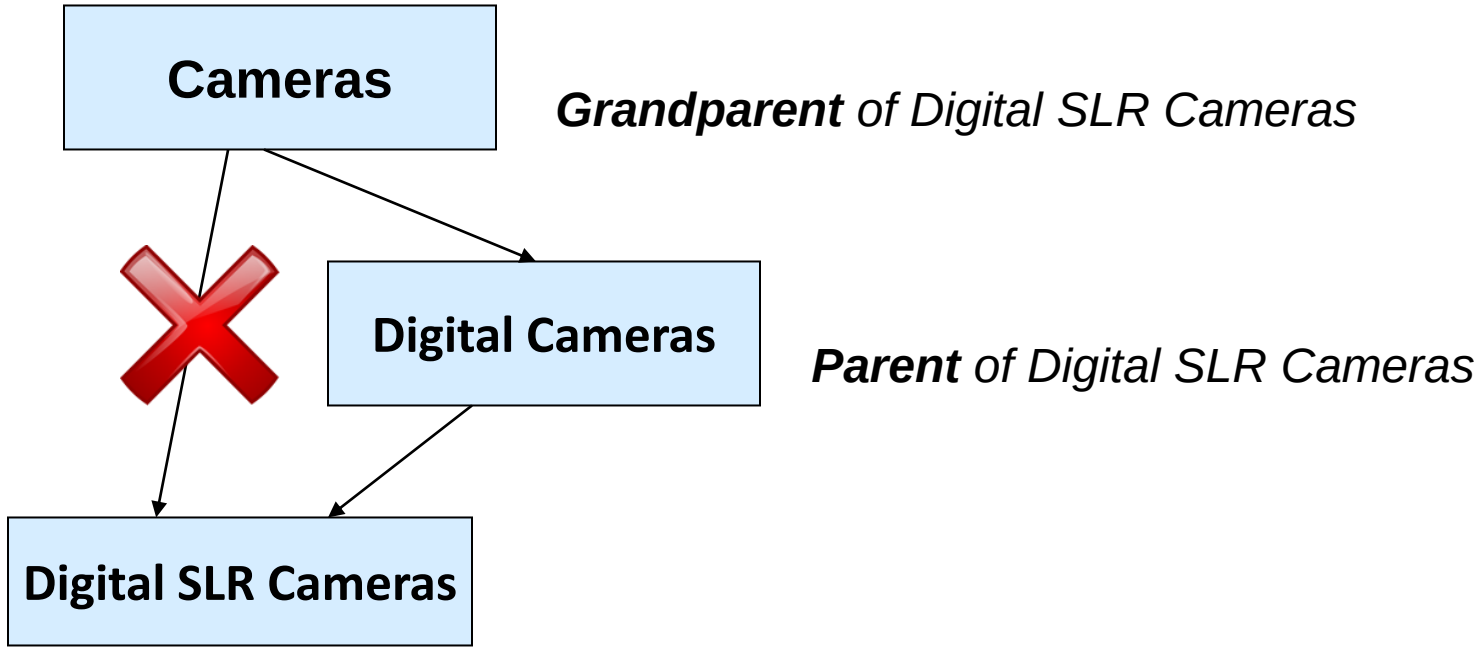
- Familiar tree structure is lost.
- Users cannot see the logical hierarchy.
- Users spend too much time clicking through categories.



Structural Design: Hierarchies

Polyhierarchies – Disadvantages

Do not create a polyhierarchy to both a “parent” and a grandparent” of a term.



Structural Design: Hierarchies

Exercise 3: Suggest two broader terms for each term to create a polyhierarchy:

Printers

Financial reports

Hotel managers

Insert into Zoom chat.

Structural Design: Hierarchical UI Displays

End-user browse display options

Hierarchical taxonomy displays tend to be on public websites with a collection of content, such as ecommerce.

Display of multiple levels of hierarchy may be implemented in different ways:

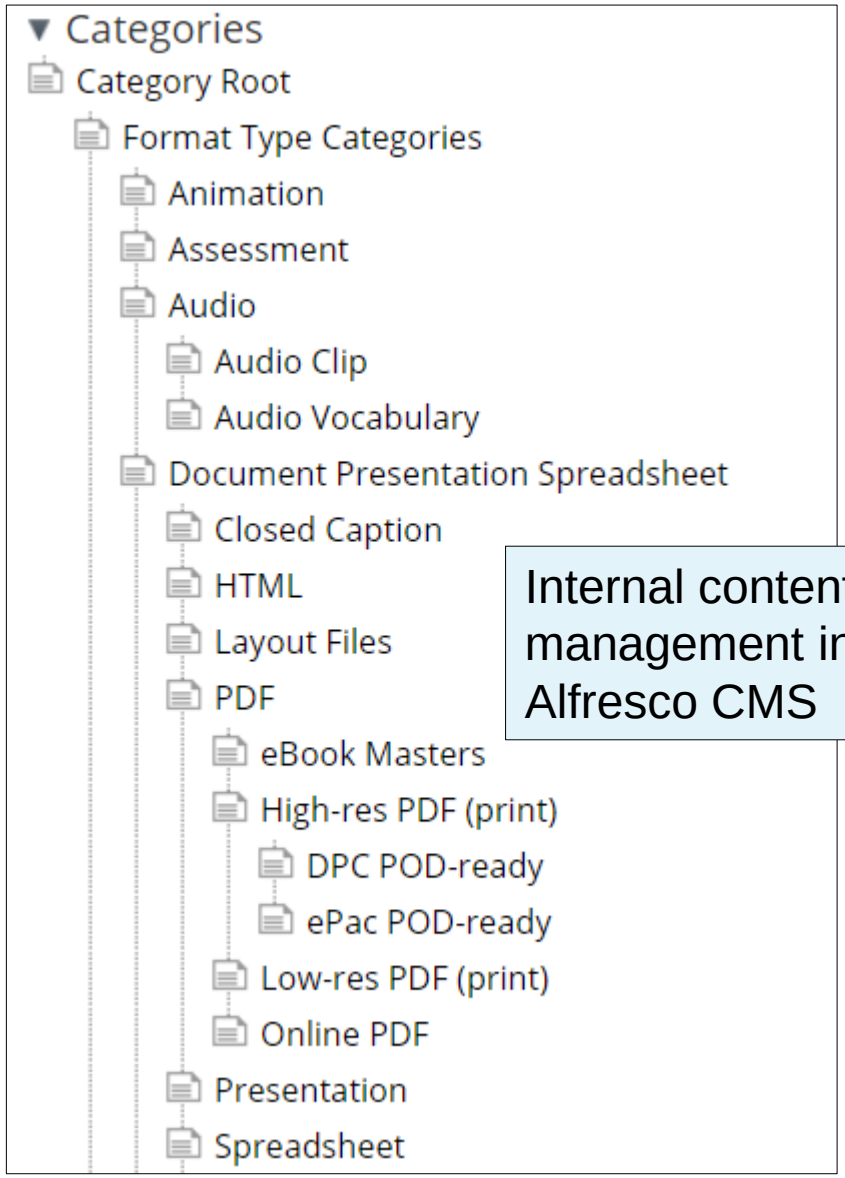
- Expandable tree in place
- One level per new web page, screen, or window
- Fly-out subcategories (expanding to the right)

Or a combination (becoming more common)

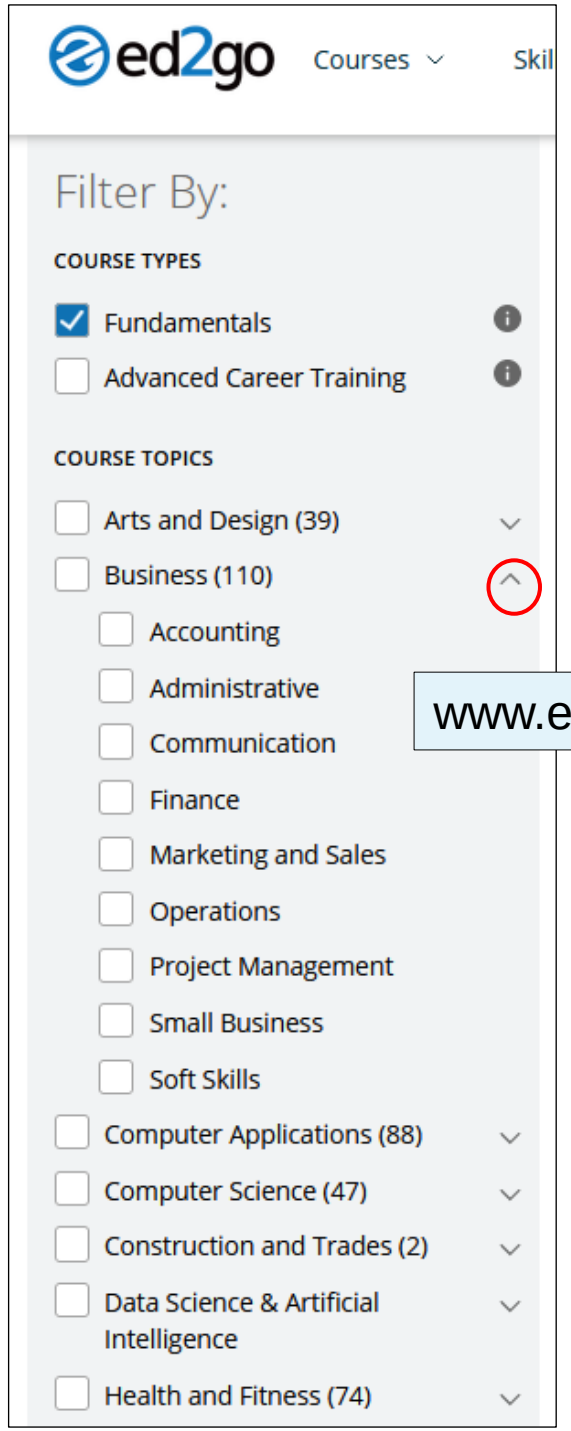
➢ The taxonomy needs to be designed to accommodate the display method, with respect to size, depth, and breadth.

Structural Design: Hierarchical UI Displays

Expandable hierarchies



Internal content management in Alfresco CMS



www.ed2go.com

Economic Policy

Fiscal Policy

Fiscal sustainability
Public Expenditure Policy
Debt Policy
Tax policy
Subnational Fiscal Policies

Macro-financial policies

External Finance
Monetary and Credit Policies
Macroeconomic Resilience

Economic Growth and Planning

Inclusive Growth
Structural Transformation and Economic Diversification
Green Growth
Spatial Growth
Migration, Remittances and Diaspora Engagement
Macroeconomic & Structural Policy Modelling

Trade

Trade Facilitation
Trade Logistics
Trade Policy

Environment and Natural Resource Management

Finance

Human Development and Gender

Private Sector Development

Public Sector Management

Expands from first level to combined 2nd + 3rd levels. Wide and narrow screen variations.

Economic Policy

Fiscal Policy

Fiscal sustainability
Public Expenditure Policy
Debt Policy
Tax policy
Subnational Fiscal Policies

Macro-financial policies

External Finance
Monetary and Credit Policies
Macroeconomic Resilience

Economic Growth and Planning

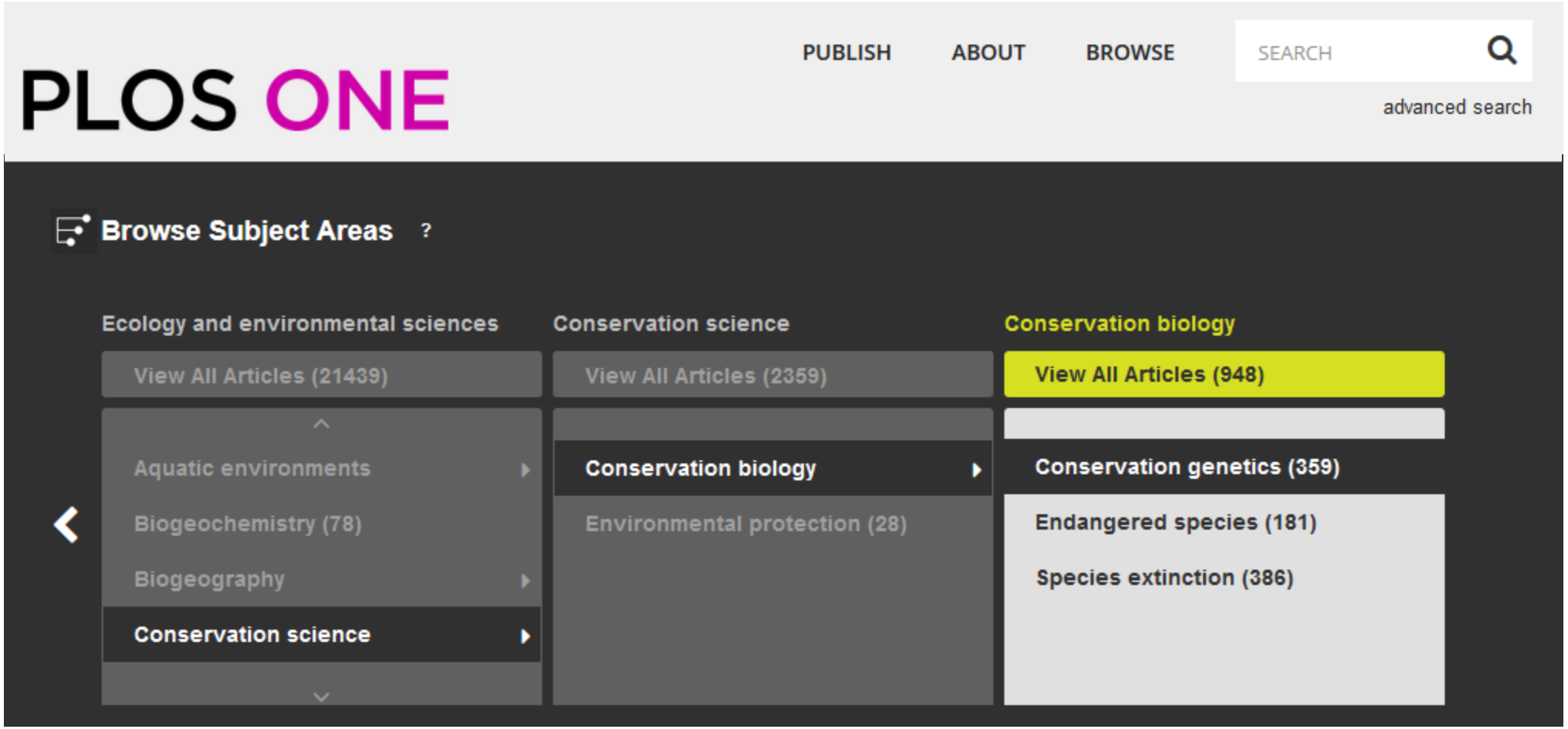
Inclusive Growth
Structural Transformation and Economic Diversification
Green Growth
Spatial Growth
Migration, Remittances and Diaspora Engagement
Macroeconomic & Structural Policy Modelling

Trade

Trade Facilitation
Trade Logistics
Trade Policy

Environment and Natural Resource Management

Structural Design: Hierarchical UI Displays



Fly-out subcategories: PLOS ONE scholarly articles https://journals.plos.org/plosone/

All Categories

- eBay Motors
- Fashion
- **Books, Movies & Music** ✓
- Electronics
- Collectibles & Art
- Home & Garden
- Sporting Goods
- Toys & Hobbies
- Business & Industrial
- Health & Beauty
- Others

Books, Movies & Music

Musical Instruments & Gear

- Guitars & Basses
- Pro Audio Equipment
- Percussion Instruments
- Wind & Woodwind Instruments
- String Instruments
- Pianos, Keyboards & Organs

View all in Musical Instruments & Gear →

DVDs & Movies

- DVDs & Blu-ray Discs
- VHS Tapes
- Movie Film Stock
- Other Movie Formats
- LaserDisc Movies
- DVD & Movie Wholesale Lots

View all in DVDs & Movies →

Books

- Nonfiction Books
- Fiction & Literature Books
- Antiquarian & Collectible Books
- Magazines
- Textbooks, Educational & Referenc...
- Children & Young Adult Books

View all in Books →

Music

- Music CDs
- Vinyl Records

Top Brands

- Fender
- Gibson
- Ibanez
- EMG
- Roland

Popular Topics

- DVDs
- Nonfiction Books
- Gibson Les Paul Electric Guitars
- Antiquarian & Collectible Books
- Fiction & Literature Books

https://www.ebay.com/n/all-categories

Musical Instruments & Gear

Shop by Category

Brass
- See all in Brass
- Alto Horns
- Baritones
- Bugles
- Cornets
- Euphoniums
- Flugelhorns
- French Horns
- Mellophones
- Other Brass Instruments
- Parts & Accessories
- Sousaphones
- Trombones
- Trumpets
- Tubas

DJ Equipment

Equipment

Guitars & Basses

Shop by Category


Electric Guitars


Acoustic Guitars

Guitar Amplifiers

Guitar Parts & Accessories


Percussion Instruments


Pro Audio Equipment


DJ Equipment


Brass Instruments


Pianos, Keyboards & Organs

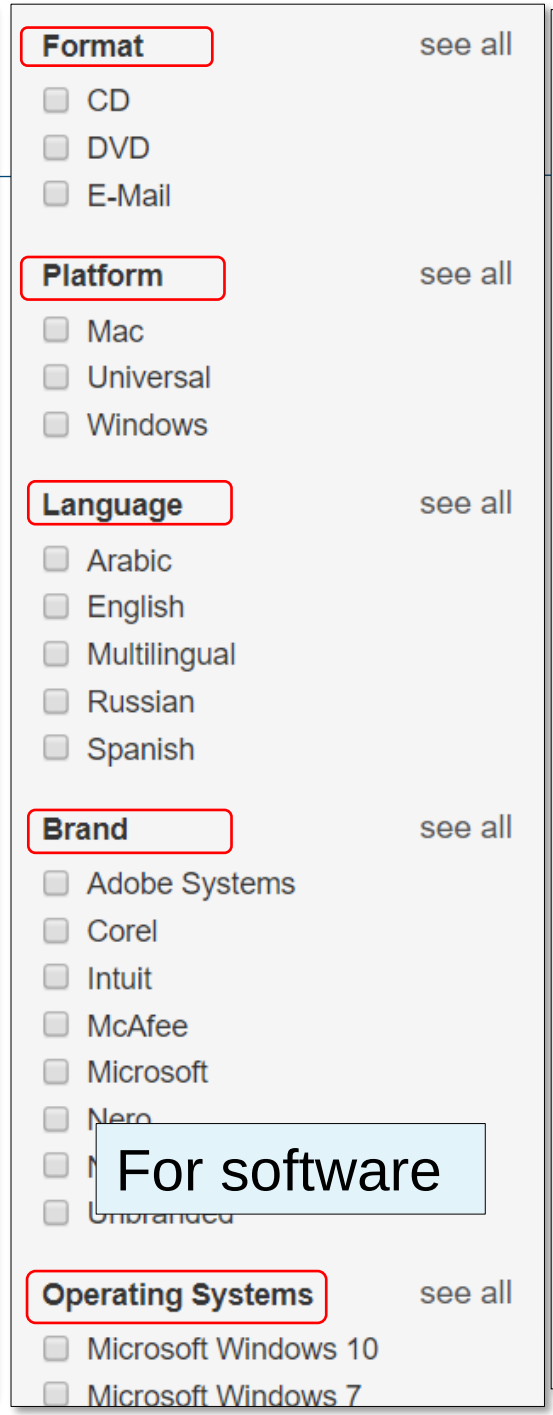

Vintage Musical Instruments

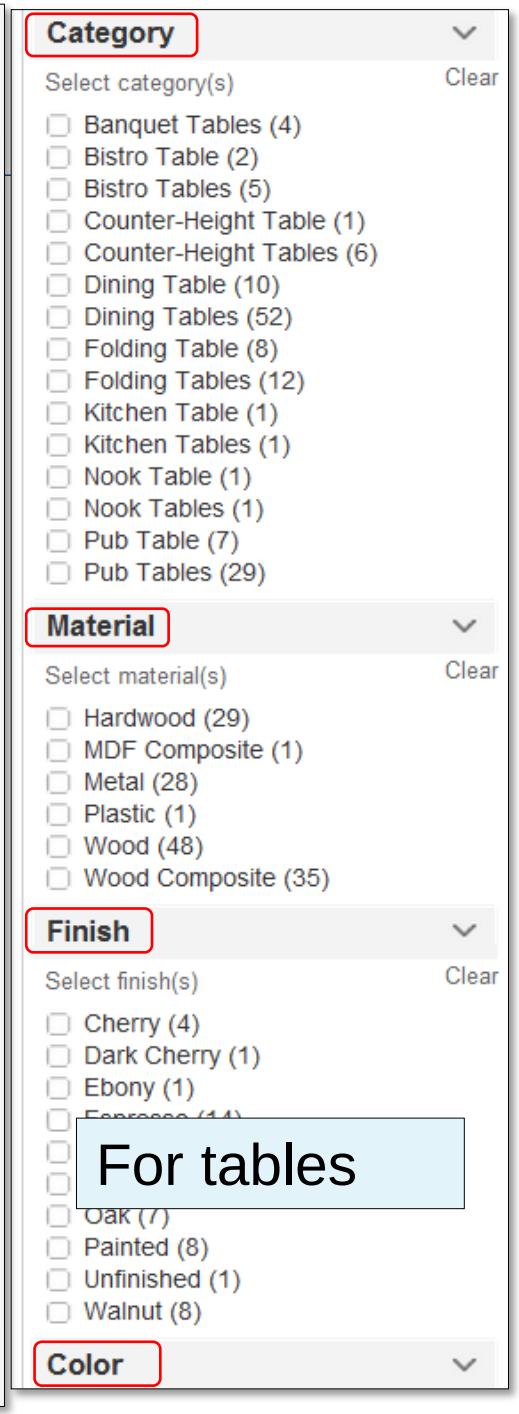
Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Structural Design: Facets

Examples of ecommerce facets





Structural Design: Facets

Facets

- For serving faceted classification, which allows the assignment of multiple classifications to an object
- A “dimension” of a query; a type of concept
- Intended for searching with multiple terms in combination (post-coordination), one from each facet
- A refinement, filter, limit by, narrow by
- Can be for topics or for named entities, but generally not both
- Reflect the domain of content

Start over | Format: **Thesis** ✖

Limit your search

Cornell University Libraries
https://newcatalog.library.cornell.edu/

Access

At the Library	194,967
Online	79,941

Format

Thesis ✖	204,007
Book	178,404
Microform	17,548
Musical Score	152
Miscellaneous	23
Database	6
Video	6
Computer File	2
Journal/Periodical	1
Map	1

more »

Author, etc.

Publication Year

Language

Subject

Political science	3,254
Economic history	1,996
Diplomatic relations	1,505
Dogs > Diseases	1,299
Social history	1,137

more »

Subject: Region

Subject: Era

Genre

Fiction/Non-Fiction

Library Location

Call Number

Date Acquired

ther

Project status

- Active (197)
- Completed (231)

County

- Adams (8)
- Benton (6)
- Chelan (10)
- Clallam (14)
- Clark (29)
- Cowlitz (20)
- Franklin (2)
- Grant (3)

Show more

Division

- Facilities (1)
- Rail (1)
- Washington State Ferries (7)

Funding

- ARRA (2)
- CWA (108)
- Nickel (14)
- PEF (168)
- Tiger Grant (2)
- TPA (29)

Highway

- I-5 (62)
- I-82 (3)
- SR 3 (4)
- SR 4 (8)
- SR 6 (6)
- SR 7 (2)
- SR 8 (1)

Show more

Location

Washington State Department of Transportation
https://wsdot.wa.gov/projects/index

FILTER BY CATEGORY

- Evaluations
- Event resources
- Multimedia
- News and stories
- Reports and publications

TOPICS

- Air
- Biosafety
- Chemicals & waste
- Cities and lifestyles
- Climate change
- Disasters & conflicts
- Ecosystems and biodiversity
- Education & environment
- Energy
- Environmental rights and governance
- Environment under review
- Extractives
- Forests
- Gender
- Green economy
- Oceans & seas
- Resource efficiency
- Sustainable development goals
- Technology
- Transport
- Water

REGION

Choose

COUNTRY

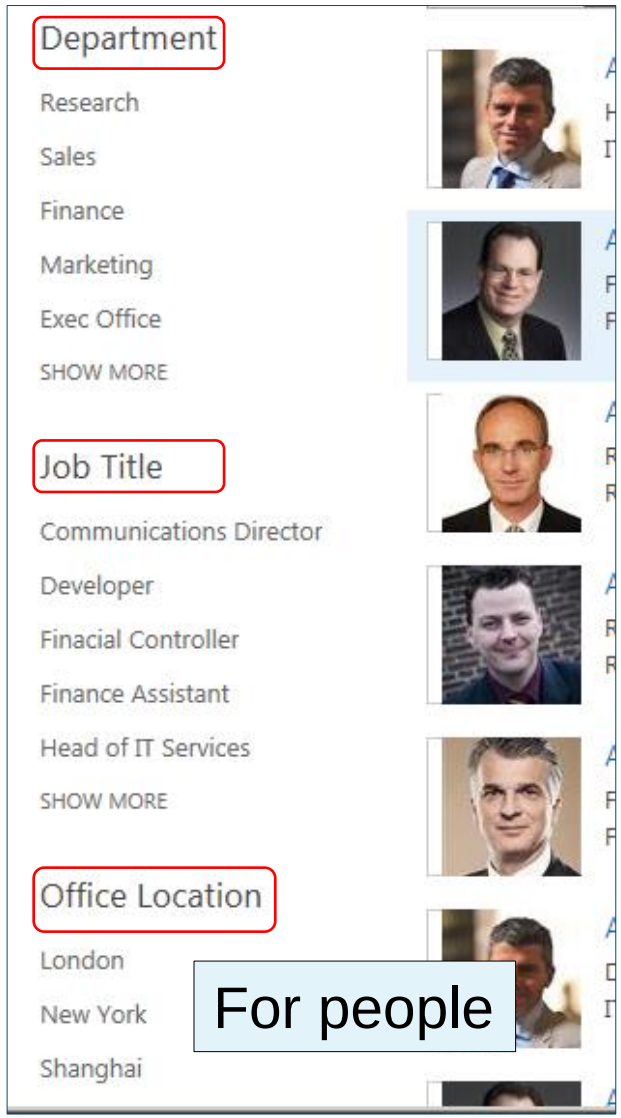
Select country

Select country
Afghanistan
Albania

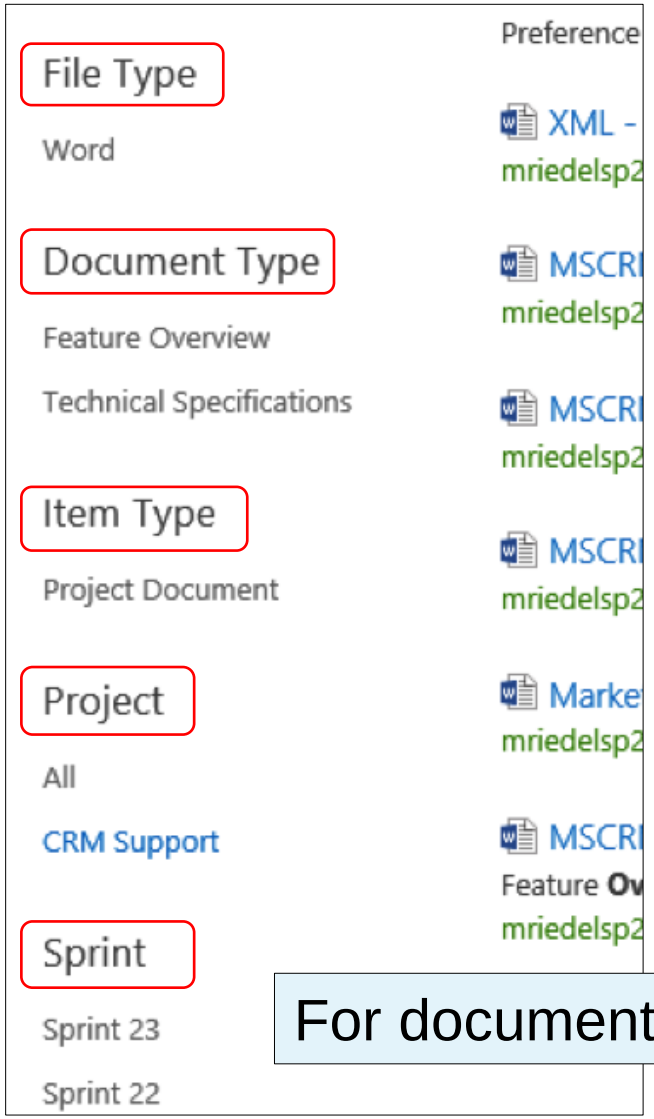
www.unep.org/resources

Structural Design: Facets

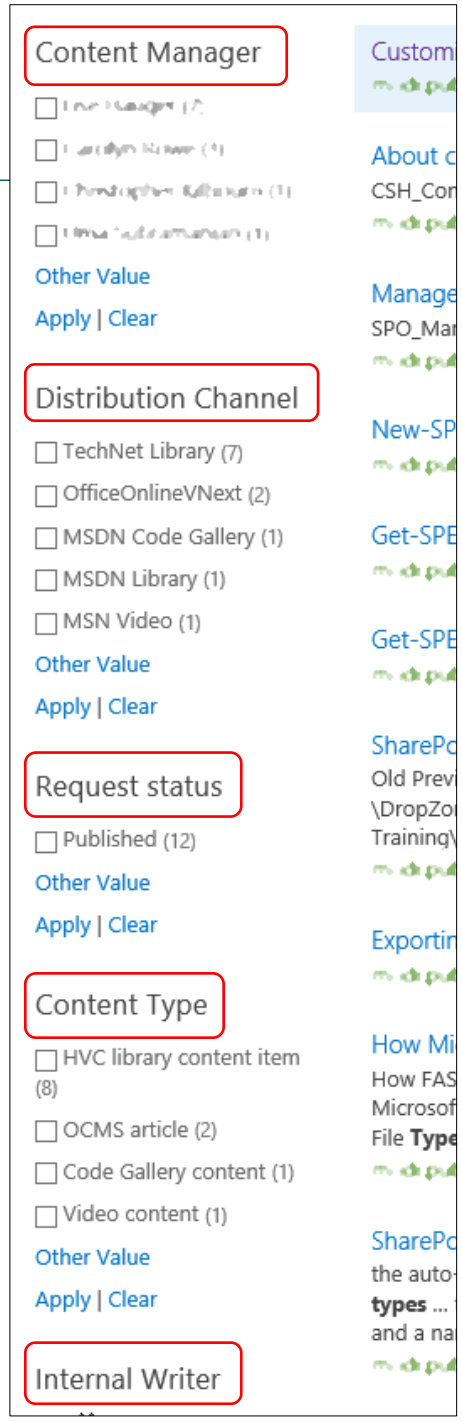
Examples of internal content facets



For people



For documents



Structural Design: Facets

Facet advantages

- Supports more complex search queries by users
- Allows users to control the search refinement, narrowing or broadening in any manner or order
- Familiar to novice users; suitable for expert users

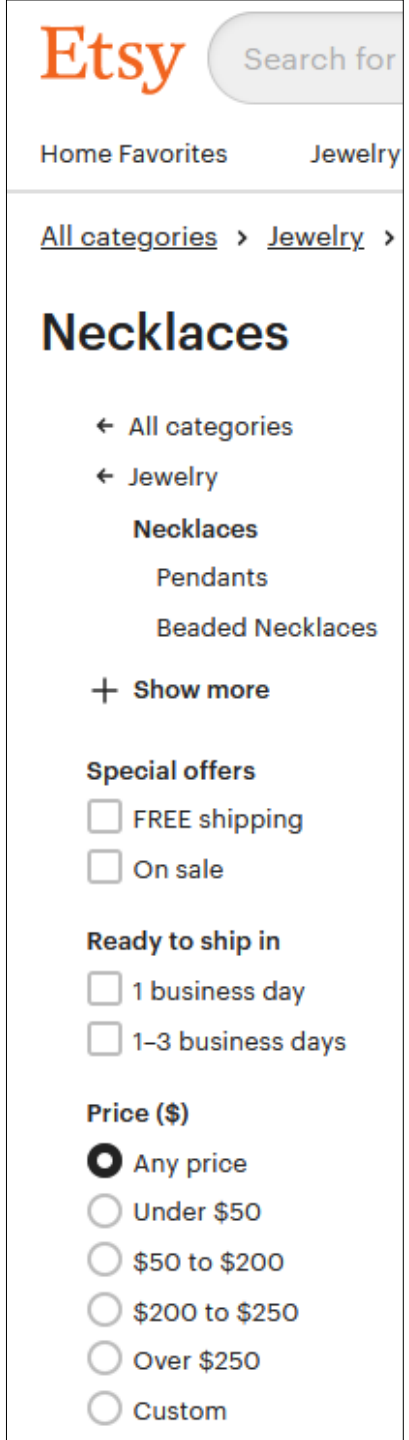
Facet disadvantages

- Only suitable for somewhat structured, unified type of content that all share the same multiple facets
- Not practical for extremely large topical controlled vocabularies
- May not support “advanced search” of multiple terms selected at once (“or”) from the same facet
- Requires investment of thorough indexing/tagging

Structural Design: Facets

Facet design options and issues

- Facets as post-search filters, as initial browse, or both
- Display of partial/start of list of values within a facet
- Ability to select multiple values from within the same facet at once (with tick boxes)
- Inclusion of other metadata (not “taxonomy”) in the same set of facets (date, author, price, etc.)
- Have all generic facets or also category-specific facets
- Designating kinds of topics into different facets


Etsy
Search for
Home Favorites
Jewelry
All categories › Jewelry ›
Necklaces
← All categories
← Jewelry
Necklaces
Pendants
Beaded Necklaces
+ Show more
Special offers
FREE shipping
On sale
Ready to ship in
1 business day
1–3 business days
Price ($)
Any price
Under $50
$50 to $200
$200 to $250
Over $250
Custom
Recipient
Boys
Girls
Unisex kids
Men
Women
+ Show more
Material
Aluminum
Bone & horn
Brass
Bronze
Ceramic
+ Show more
Gemstone
Amethyst
Ametrine
Aquamarine
Bloodstone
Citrine
+ Show more
Cut type
Asscher
Baguette
Brilliant
Crescent
Cushion
+ Show more
Number of strands
1
2
3
4
5 or more
Style
Art deco
Art nouveau
Avant garde
Boho & hippie
Brutalist
+ Show more
Theme
Animals
Beach & tropical
Evil eye
Fantasy & Sci Fi
Floral
+ Show more
Chain style
Ball
Bead
Box
Byzantine
Cable
+ Show more
Closure
Alligator clip
Ball & joint
Barrel & torpedo
Box
Buckle
+ Show more
Setting
Bar
Bezel
Channel
Cluster
Double halo
+ Show more
Can be personalized
Yes
No
Occasion
Anniversary
Baby shower
Bachelor party
Bachelorette party
Back to school
+ Show more
Holiday
Christmas
Cinco de Mayo
Easter
Halloween
Hanukkah
+ Show more
Color
Beige
Black
Blue
Bronze
Brown
+ Show more
Shop location
Anywhere
United States
Custom
Enter location
Item type
All items
Handmade
Vintage
Ordering options
Accepts Etsy gift cards
Can be gift-wrapped
Customizable

Structural Design: Facets

Facet design tips

- Number of facets: 3-8, with 5-6 as ideal
- Facets listed in logical, not alphabetical order
- Number of terms per facet: 2-25
  - Ideally not much more than can be viewed in a scroll box
  - If the list is obvious (US states), then more is OK.
  - Exception can be made for hierarchical “Topics” facet
- If <12 terms, then a logical display order
  If >12 terms, then alphabetical order
- A two-level hierarchy (indented) within a facet is possible

Structural Design: Facets

Breakout Session Activity (10 minutes)

Create a set of facets and several example terms within each facet.

1. Digital assets
2. Intranet content
3. Books or movies
4. Health/medical information
5. Ecommerce for computers
6. Events
7. Training content
8. Employee or member profiles

- Facets: multiple methods to filter/limit by
- Consider: content items, scope, and users.
- List several terms within each facet.
- Determine if there will be a small hierarchy within a facet.

Use the Google sheet:
https://docs.google.com/spreadsheets/d/1_FLIwP6iXjgheQ0U4u1Rq7Gqc1Mx8OZF-rIyz5-D08I/edit#gid=869315193

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Taxonomy Project Planning

Justification for the need of a taxonomy

- Quantitative or qualitative data for ROI
  - Quantitative data: click-through and search long reports, precision and recall statistical data
  - Qualitative data: success or failure stories in users being able to find content
- What problems will the taxonomy solve?
  - User frustration, pain points, in difficulty finding something
  - Need to save time
  - Need to get better results

Taxonomy Project Planning

Initial project considerations

- What the taxonomy will be used for?
- Who/what department is the taxonomy project sponsor and taxonomy owner?
- Who will build the taxonomy?
  - Staff taxonomist, external consultant, contractor(s), content owner?
  - Individual or a team? Group or cross-functional team?
  - Separate tasks of taxonomy designing/planning and detailed building?
- Who are the stakeholders to also be involved?
- Who will be the users of the taxonomy (tagging and end-users)?
- What are the resources and constraints?: time, money, staffing, stakeholder involvement, skills, technology

Taxonomy Project Planning

Further taxonomy considerations

- Topic area and scope
- Build a new taxonomy or reuse, adapt and expand and existing taxonomy
- Method of indexing/tagging content: manual (by whom), automated, or both
- Content management workflow
- Systems for
  - Taxonomy management
  - Tagging with the taxonomy
  - End-user access via browsing and/or searching
- How the taxonomy is displayed to the users

Taxonomy Project Planning

Determining the taxonomy scope

- Scope of the content to be tagged with the taxonomy
- Types of content to be tagged
- Subject area scope
- Term types to include
- Taxonomy general breadth and depth
- Types/communities of users

Determining the taxonomy type

- Hierarchical, faceted, a combination, etc.
- Separate taxonomies for different purposes, or one master taxonomy for the enterprise?

Taxonomy Project Steps

Taxonomy building steps

1. Information gathering
   - Sample content, of different types, to be tagged
   - Input from user and other stakeholders
   - Existing vocabularies, term lists, search logs
2. Taxonomy design and review
3. Taxonomy building
4. Taxonomy testing and/or validation
5. Taxonomy maintenance and policy drafting

Taxonomy Project Steps

Building taxonomies

1. Gather term concepts from the content audit, interviews, search logs, etc.
2. Start initial term lists in Excel, then import into thesaurus software
3. Consult subject matter experts and reference sources (glossaries, other thesauri) as needed
4. Build taxonomy of concepts, alternative labels, relationships, and scope notes
5. Test with sample indexing and revise
6. Add apparently “missing” terms and relationships

Approach is a combination of top-down and bottom-up.

Taxonomy Project Steps

Taxonomy review, revision and testing

Review, test, use, and maintain. And edit at each stage.

Use difference views/run different reports (alphabetical, hierarchical).

- Check for balance of coverage; analyze gaps.
- Run quality control checks available in the software.
- Proofread, verify spelling.
- Remove unnecessary duplication.

Taxonomy Project Steps: Testing

Different tests for different stages of taxonomy development

- Development phase: to test ideas
  - Card sorting (with web-based software)
  - A/B testing (with mockup/wireframe user interfaces)
- Draft completion phase: to test usability/functionality (may also be considered “validation”)
  - User/use case content retrieval testing
  - Content indexing testing
- Implemented taxonomy: to periodically test quality
  - QA testing

Taxonomy Project Steps: Testing

Test tagging/categorizing – using sample actual content items that would need tagging and categorizing

- Test users browse the UI of metadata tagging and category options.
- Can they find the metadata they think is needed for tagging the item?
- Can they easily find the desired category into which to put the content?
- Do they categorize and apply the metadata that the policies (the taxonomist) intended?

Test search and retrieval – based on use cases described during stakeholder interviews

- Test users walk through use cases and browse the UI for search/browse retrieval and look for the categories and terms they think would be relevant.
- Can they find the metadata and categories they think is relevant?
- (The content item does not need to be in the system yet to execute the search.)

Taxonomy Project Steps: Testing

Summary of taxonomy testing methods

Test	Testers	Purpose
Card sorting	Stakeholders or sample users/searchers	To test taxonomy structure ideas - but often not practical for taxonomies (in contrast to testing website menu navigation)
A/B Testing	Sample users/searchers	
Use case Testing	Sample users/searchers	To test suitability to retrieve desired content
Indexing testing	Sample indexers/taggers or taxonomist	To test suitability to tag content and appropriateness for content scope
QA testing	Taxonomist	To test continued precision and recall in retrieval results

Taxonomy Project Steps

Implementation of taxonomies

- The taxonomy/thesaurus creation tool is often separate from the indexing program.
- Might have a combined thesaurus creation/indexing system, but separate from the end-user search system.
- The controlled vocabulary is likely to need to be ported (for indexing and/or end-use searching).
  Consider:
  - Export format, schedule, and method for updates
  - Integrating with APIs or custom integrations

Taxonomy Governance and Maintenance

Taxonomy governance comprises:

- Taxonomy descriptive documentation (purpose, type, scope, users, indexing method, history/sources)
- Descriptive metadata element rules and governance
- Taxonomy editorial policy/guidelines for maintenance
- Indexing or tagging policy/guidelines
- Instructional/how-to documents (system-specific)
- Maintenance (updating): responsibility, roles, processes, procedures

➢ Taxonomy governance process starts with the start of creating the taxonomy. As issues come and get resolved, they get documented as policy.

➢ Taxonomy governance is often part of a larger metadata specification.

Taxonomy Governance and Maintenance

Taxonomy editorial policy

- Concept label editorial style (case, abbreviations, acronyms, special characters, diacritics, punctuation, plurals, spelling)
- Extent of compound/pre-coordinated concepts and specificity of concepts
- Types of concepts included and types of terms not
- Authoritative sources for concepts and labels
- Guidelines for including alternative labels
- Relationship policy (hierarchy depth, polyhierarchy, associative relationships, relationships between topics and named entities, orphans)
- Use of categories or classes for concept
- Scope notes, definitions and other notes policy
- Extended attributes for terms

Taxonomy Governance and Maintenance

Maintenance – kinds of changes impacting vocabularies

- Terminology changes
- New content, bringing up new concepts
- Content that gets dropped
- New requirements, users, needs, trends, markets, etc.
- User feedback suggesting improvements

Taxonomy Governance and Maintenance

Maintenance

For maintenance policy and procedures, need to determine:

- What kinds of changes are routine, which are not
- What information needed to determine the changes
- What group should maintain the taxonomy
- What role indexers play in suggesting changes
- The processes for proposing and resolving changes
  - Comment-handling, appeals, issue logs, announcements, update schedules, etc.

For maintenance, review:

- Newly added content sources
- Search logs, click-trail reports
- Sections of the controlled vocabulary covering high-change topics (technology, current events)
- Indexing statistics (human or auto) to find:
  - high-use terms needing further differentiation
  - low/no-use terms that should be merged
  - identifiable indexing errors

Outline

1. Introduction to taxonomies
2. Controlled vocabulary types
3. Benefits and uses
4. Standards and models
5. Sources for concepts/terms
6. Term and label creation
7. Relationships
8. Taxonomy management software
9. Structural design: hierarchies
10. Structural design: facets
11. Taxonomy project planning, steps, and governance
12. Applying taxonomies: search, metadata, categories, tags

Applying Taxonomies

- Integrating taxonomies into search
- Taxonomies and metadata
- Categories and tags in content management systems
- Tagging with a taxonomy

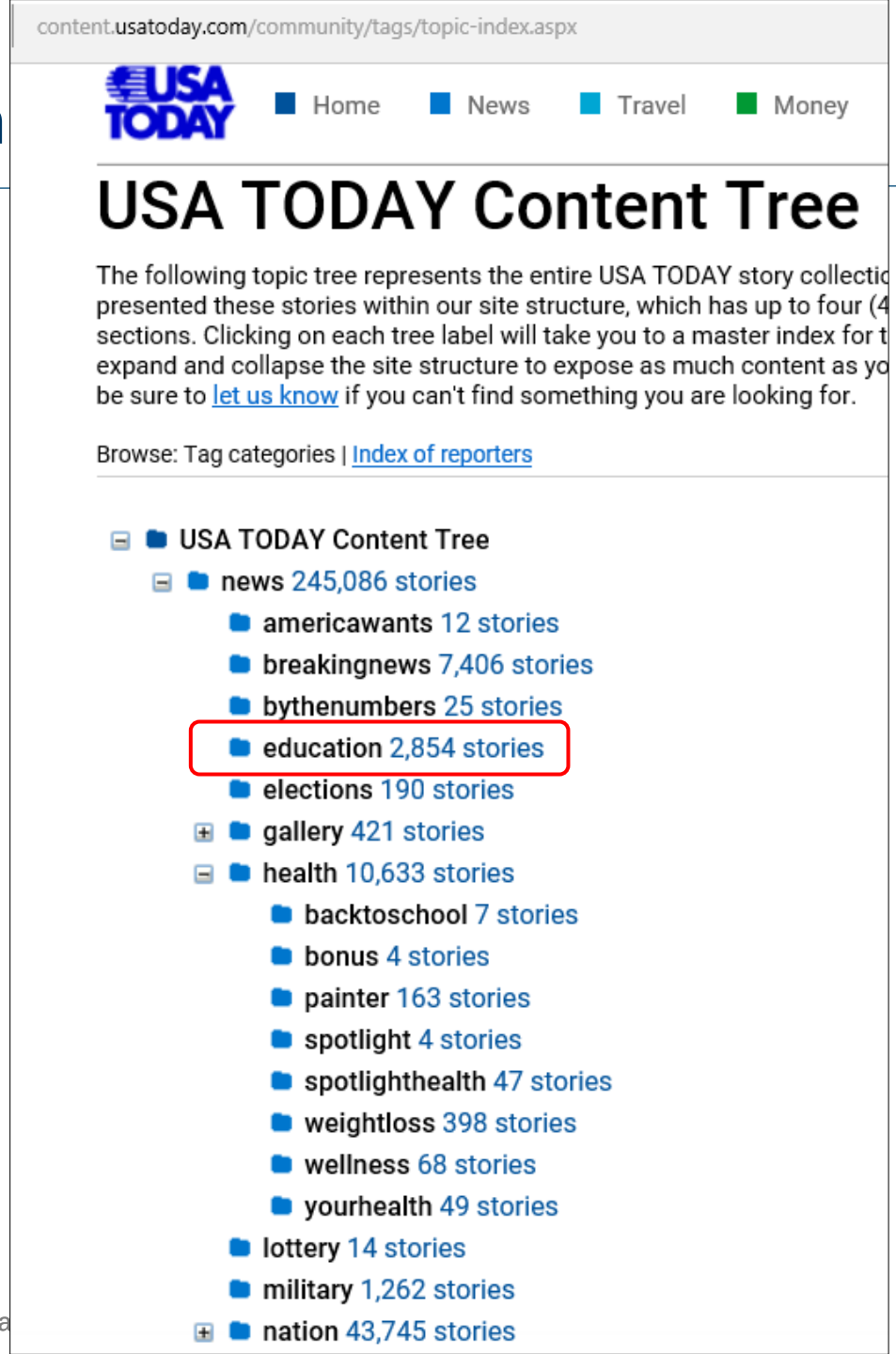
Taxonomy Integration with Search

Historic trends in taxonomy implementation, search integration and display

1. Originally, full taxonomy hierarchical browse or thesaurus alphabetical browse. Search on content was separate from browsing taxonomies.
2. Full large taxonomies came to be displayed less. Search on taxonomy terms.
3. Search on more than just the taxonomy terms.
   Search on a combination of taxonomy terms and words in titles, texts, etc.
4. Faceted taxonomies – fully displayed, combining search and browsing
5. Post-search filters – facets of terms and metadata used, for limiting search results
6. Knowledge graphs – collected data results from a knowledge base, often with a graph database and an ontology, based on queries

Taxonomy Integration with Search

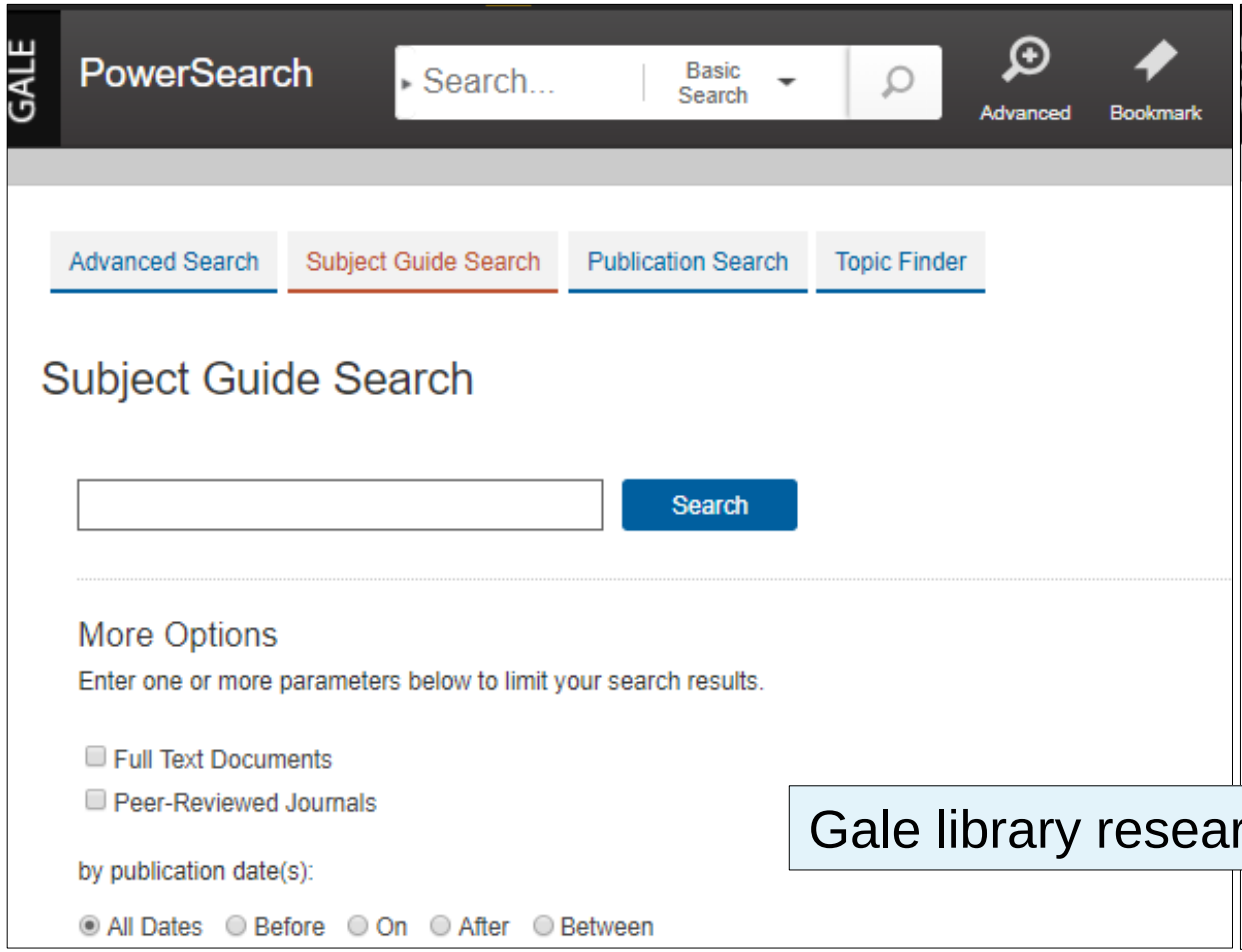
Originally
full taxonomy browse



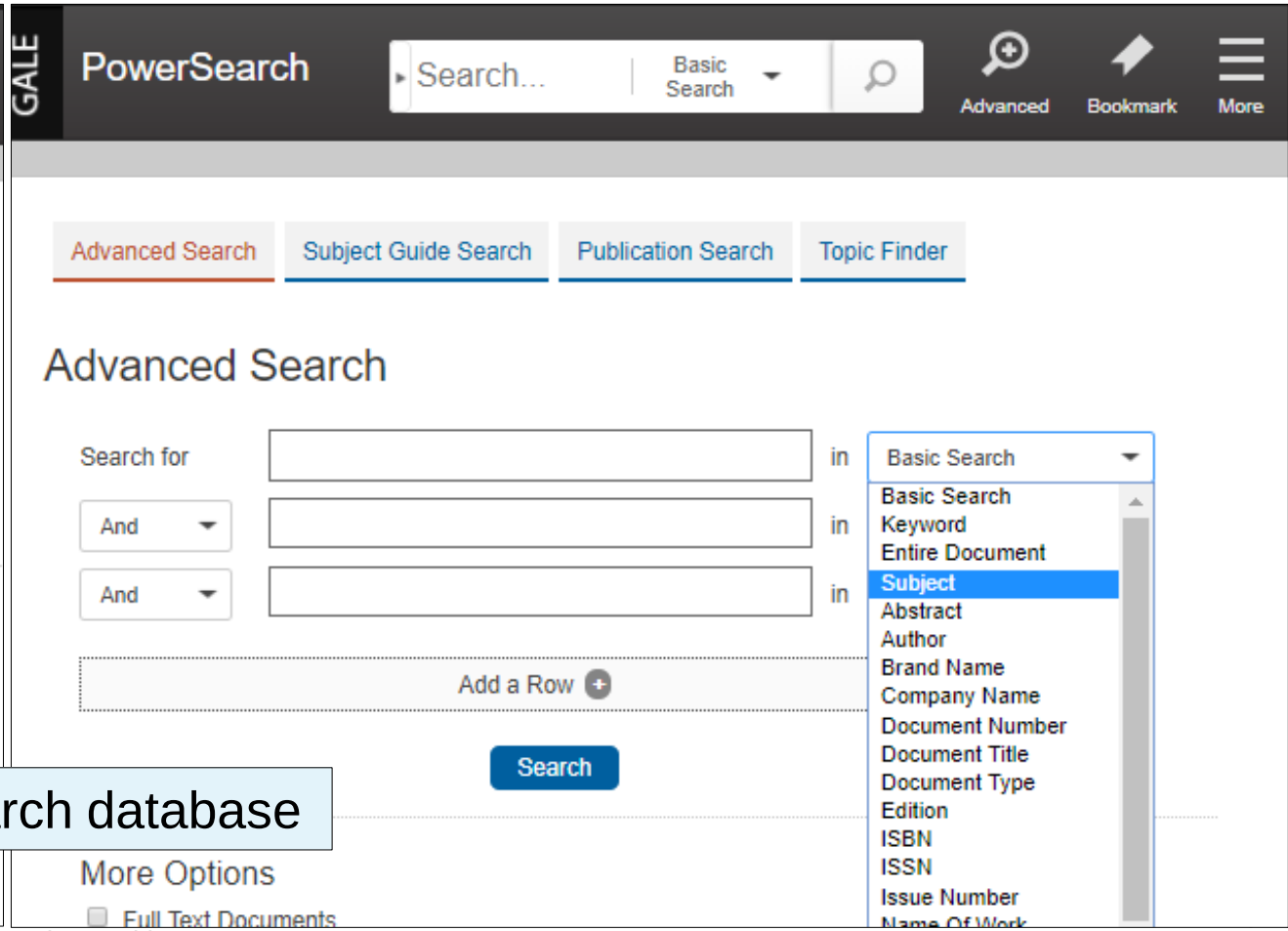
Taxonomy Integration with Search

Search on Taxonomy Terms

Search on terms in a taxonomy



Search on words/phrases in the Subject field



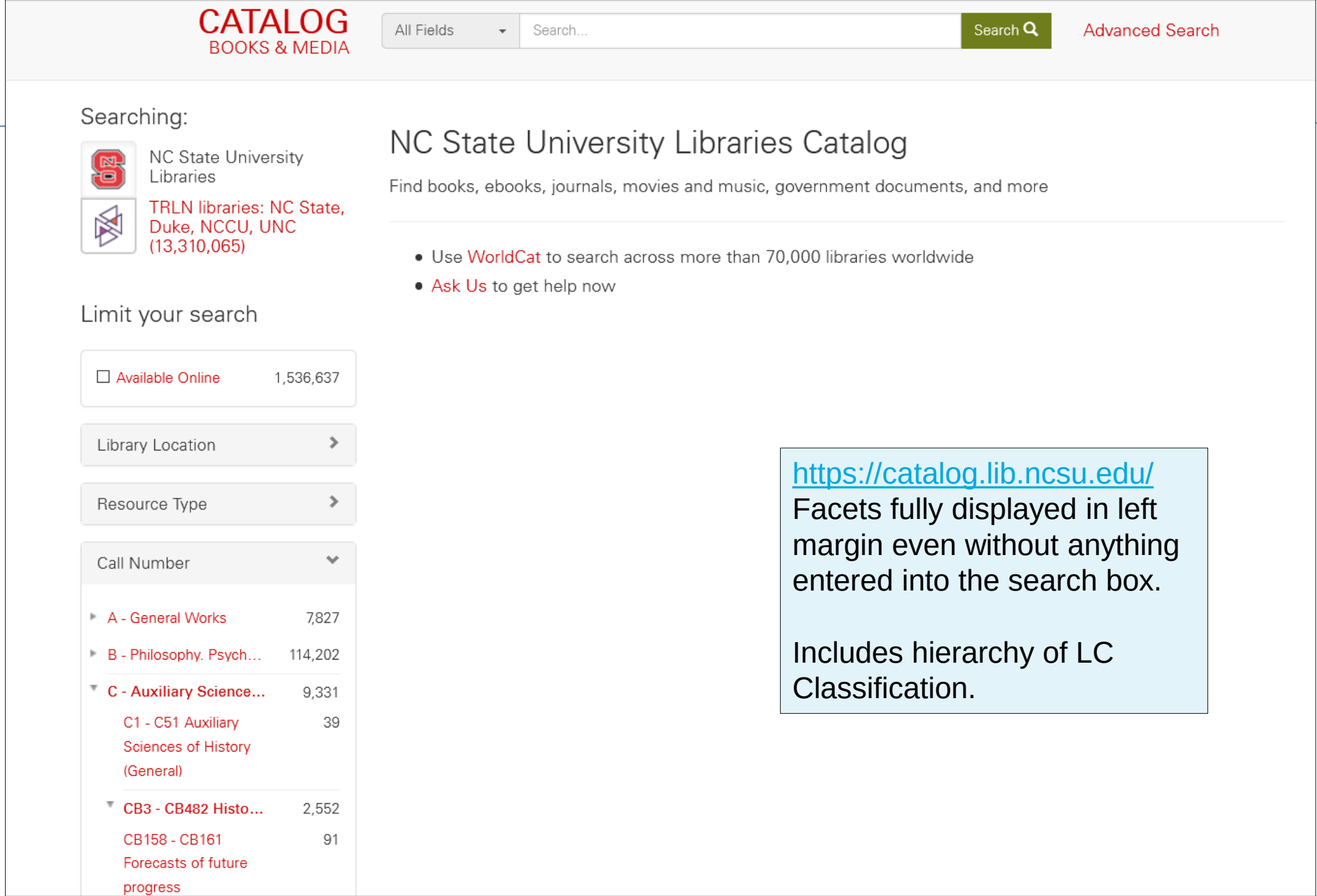
Gale library research database

Taxonomy Integration with Search

Faceted taxonomy for initial search

- The full taxonomy is displayed (or can be scrolled).
- Any hierarchy within a facet would also be displayed
- The user selects terms each from multiple facets to execute the search, rather than entering text into the search box.
- Further refinements can be made after search results are retrieved.

Best when a taxonomy is not too large


CATALOG
BOOKS & MEDIA
All Fields
Search...
Search
Advanced Search
Searching:
NC State University Libraries
TRLN libraries: NC State, Duke, NCCU, UNC (13,310,065)
Limit your search
Available Online 1,536,637
Library Location
Resource Type
Call Number
A - General Works 7,827
B - Philosophy. Psych... 114,202
C - Auxiliary Science... 9,331
C1 - C51 Auxiliary Sciences of History (General) 39
CB3 - CB482 Histo... 2,552
CB158 - CB161 Forecasts of future progress 91
NC State University Libraries Catalog
Find books, ebooks, journals, movies and music, government documents, and more
Use WorldCat to search across more than 70,000 libraries worldwide
Ask Us to get help now
https://catalog.lib.ncsu.edu/
Facets fully displayed in left margin even without anything entered into the search box.
Includes hierarchy of LC Classification.

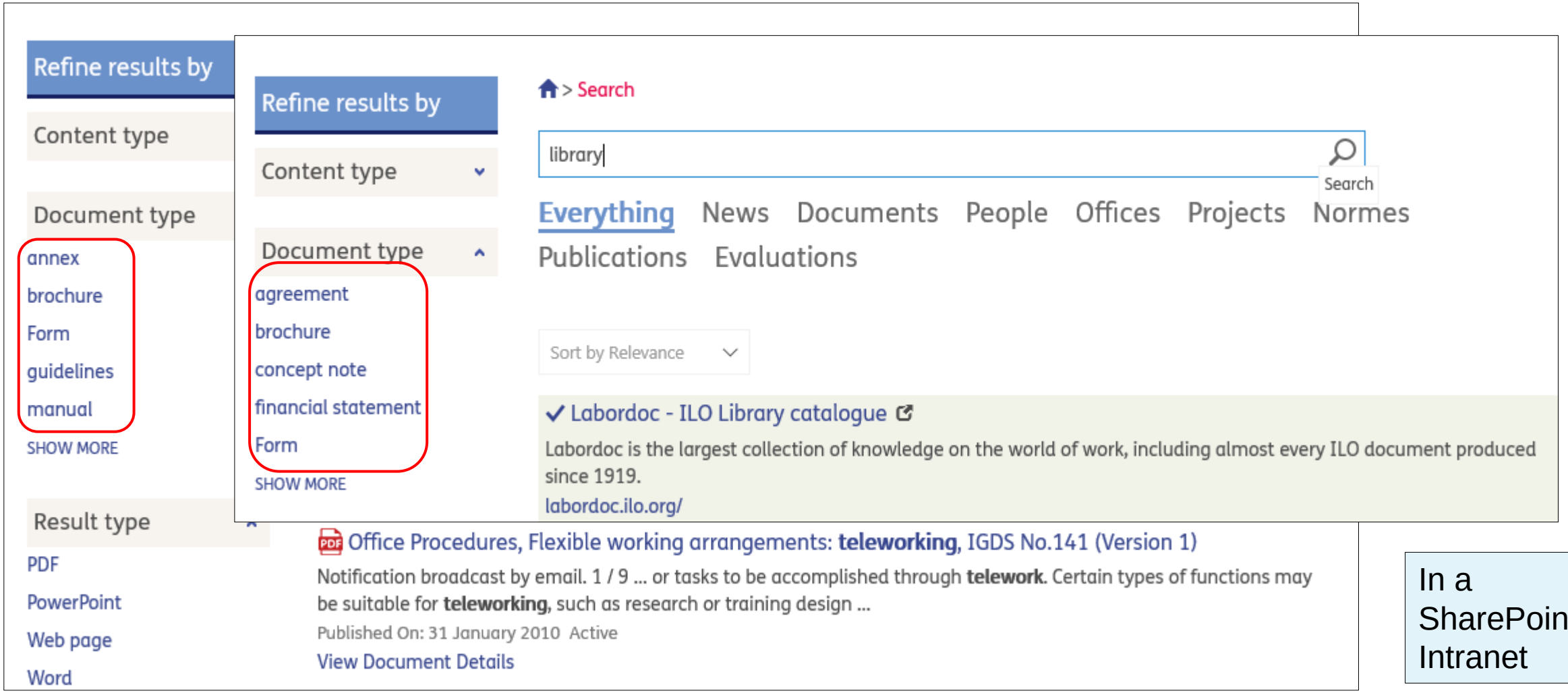
Taxonomy Integration with Search

Taxonomy terms in facets to refine post-search results

- A display of taxonomy terms that have been used to index the content items in the search result set (not all the taxonomy terms).
- A display of selected terms from the taxonomy, not the full taxonomy itself.
- Displayed terms thus vary depending on the search.
- Any relationships between terms are not indicated.
- Displayed alphabetically or in order of usage frequency of the search result.

Suitable when a large taxonomy or thesaurus does not fit into a facet.

Taxonomy Integration with Search



In a SharePoint Intranet

Taxonomy Integration with Search

Magazines (3,776)

Search Terms: *Basic Search* ("Internet of things")

Sort by: Relevance

Citation: *ARTICLE*

NO BATTERIES INCLUDED: New forms of energy harvesting and low-powered sensors promise to eliminate batteries and enable the Internet of Things

Amit Katwala.

Professional Engineering Magazine.

30.9 *(Oct. 2017)* p24+. *Word Count: 1990.* Reading Level (Lexile): 1360.

Link to CDL Materials Catalog ILL My Library OpenURL Link

Save

Citation: *BRIEF ARTICLE*

INTERNET OF THINGS IS 'COMING OF AGE'

Amit Katwala.

Professional Engineering Magazine.

30.9 *(Oct. 2017)* p6. *Word Count: 217.* Reading Level (Lexile): 1110.

Link to CDL Materials Catalog ILL My Library OpenURL Link

Save

Full-text: *ARTICLE*

The internet of things that can be hacked

Modern Healthcare.

48.4 *(Jan. 22, 2018)* p0022. *Word Count: 1493.* Reading Level (Lexile): 1290.

Byline: Rachel Z. Arndt When doctors replaced then-Vice President Dick Cheney's pacemaker in 2007, they asked the

Content Types

Academic Journals (6,397)

Magazines (3,776)

Books (24)

News (38,942)

Images (2)

Videos (4)

Search within results

Limit Search by:

- Full Text
- Contains Images

Publication Dates:

Past Week

Past Month

Past Year

Custom Date Range

Subjects

Internet of things (724)

Software industry (197)

Semiconductor indust... (120)

In a library periodical research database

Taxonomy Integration with Search

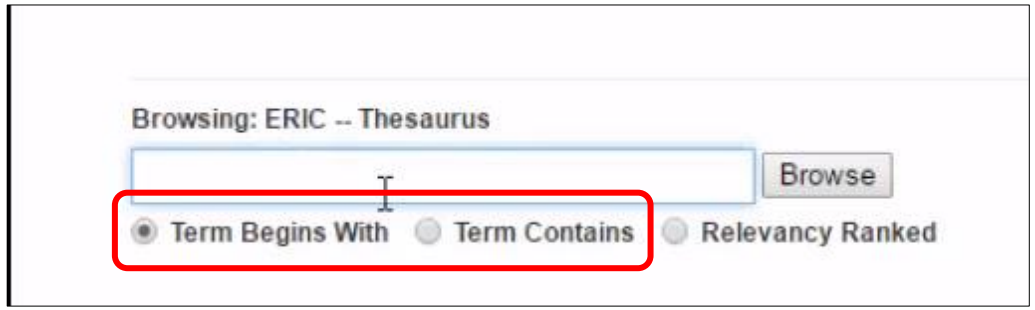
Different types of search on terms

- Exact – exact match
- Contains – exact match phrase with additional words before or after
- Begins – alphabetical from start, but allows end truncation
- Smart – words within the term in any order and also internal word stemming (singular/plural)

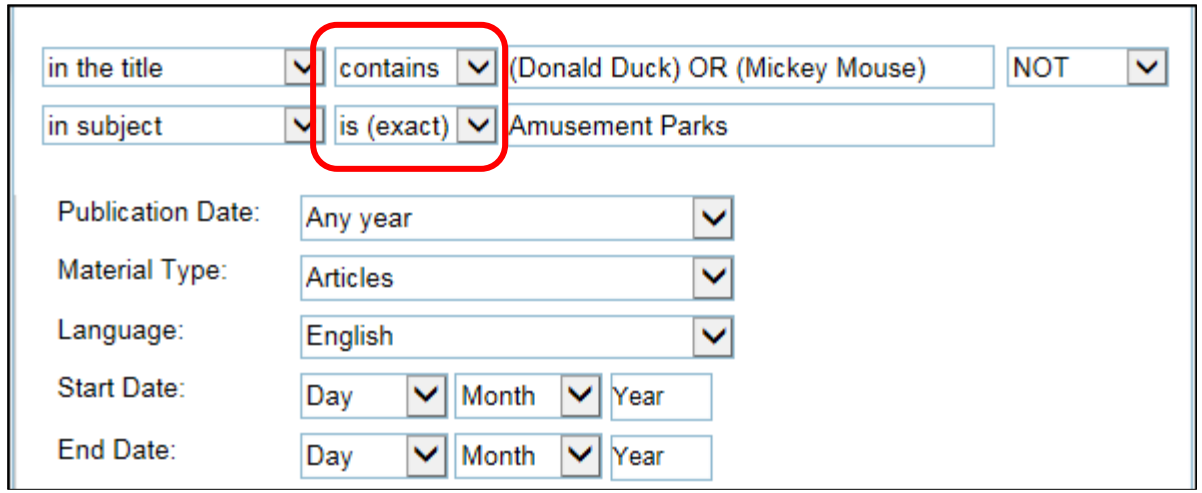
Implementations

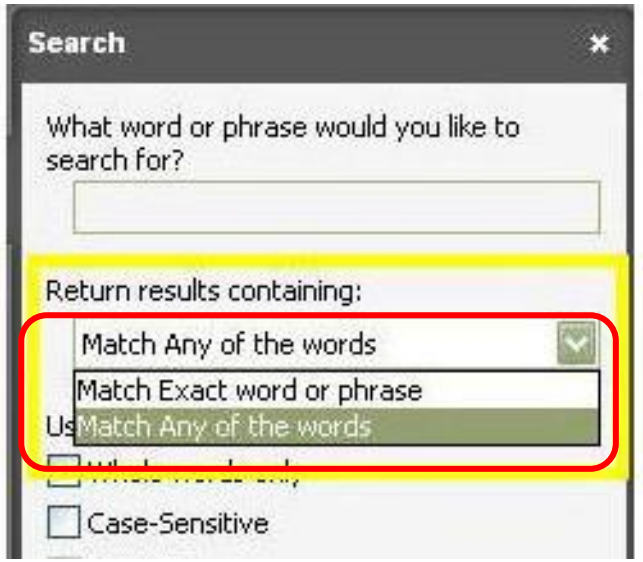
- Exact – option for experienced or repeat users
- Contains – also called phrase search. Can be done with quotation marks.
- Begins – option if there is a type-ahead display of terms
- Smart – sometimes the default, if no other options, and terms are not displayed

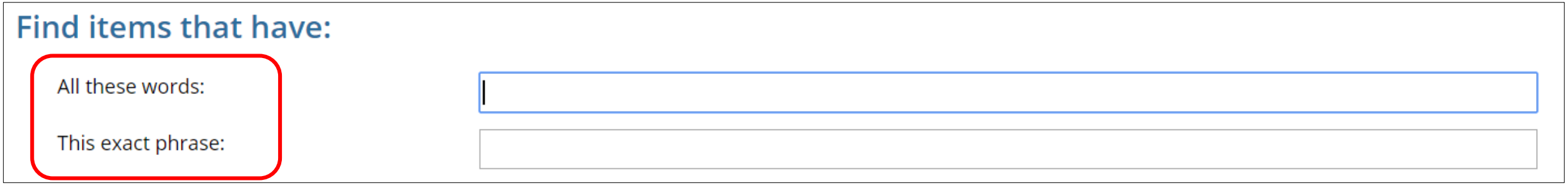
Taxonomy Integration with Search



Different types of search on terms







Taxonomies and Metadata

Metadata

> *"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information."*
> -- NISO (National Information Standards Organization) "*Understanding Metadata*"

- Data about data
- Standardized set of data for labelling multiple resources/content items
- Defined set of standard elements/fields/properties for a content set, with varying values within the fields, depending on the specific content item

Taxonomies and Metadata

Metadata – Example Structure

Metadata Element/Property/Field	Metadata Value/Term
Author	Hedden, Heather
Company	Semantic Web Company
Title	Connecting Users to Content
Document Type	Workshop training
Date	4/21/21
Format	PowerPoint
Audience	Conference attendees

Taxonomies and Metadata

Metadata Types and Uses

- Descriptive
  - for content discovery/retrieval (via searching or browsing)
  - for content identification (to cite sources, contact authors, compare content, resolve content issues, etc.)
- Administrative
  - for content management (for building collections, information products, web pages, content maintenance)
  - for workflow/process management (assigning content, archiving, preserving)
- Structural
  - for content navigation (within a large content resource)
  - for content presentation (as markup/style metadata)

Taxonomies and Metadata

Metadata & Controlled Vocabulary

- Some, but not all, metadata elements are filled/populated with values/terms in a controlled vocabulary.
- When a controlled vocabulary term is assigned/tagged to content in a designated metadata element, then it becomes metadata.
- Most implementations of controlled vocabularies are as metadata.

Exceptions (non-metadata controlled vocabularies):

- A controlled vocabulary that serves only search (such as a synonym ring)
- A taxonomy that is a model for navigation
- A controlled vocabulary used off-line to support indexing of print materials

Taxonomies and Metadata

Metadata schema/rules

Set of metadata elements, element definitions, and rules

Rules for an element:

1. Values from controlled vocabulary or not
2. Free text according to style rules
3. Source, such automatically or human-created (by whom)
4. Cardinality
   - Required/not required
   - One only/multiple permitted (with a limit/with no limit)
5. Use of subfields
6. Dependency on other values in other elements

Taxonomies and Metadata

Metadata schema/rules: values from controlled vocabulary or not

Uncontrolled/Free Text:

Title
Abstract/Description
Key Words

Form of Entry Rules:

Date
Author
Page number
Record number

Controlled Vocabulary Use:

Subject
Person - about
Place name
Language
Journal Name
Document Type
Publication Country
Lexile Reading Level

Taxonomies and Metadata

Metadata, controlled vocabularies/taxonomy are planned and designed together.

1. Identify possible metadata elements/properties and, at the same time, sample controlled vocabulary metadata values.
2. Determine if any of the controlled vocabularies should be hierarchical taxonomies.
3. Determine metadata element rules (required, multiple, dependent, etc.)
4. Start defining and documenting all metadata and tagging policies (governance).

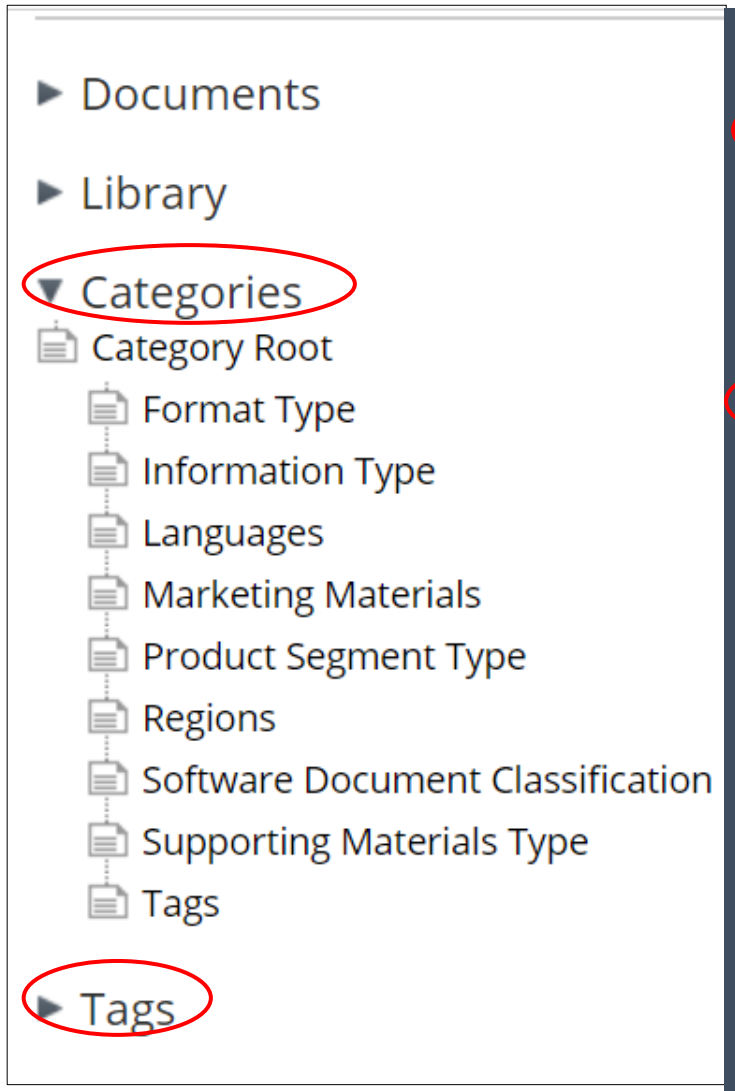
Categories and Tags

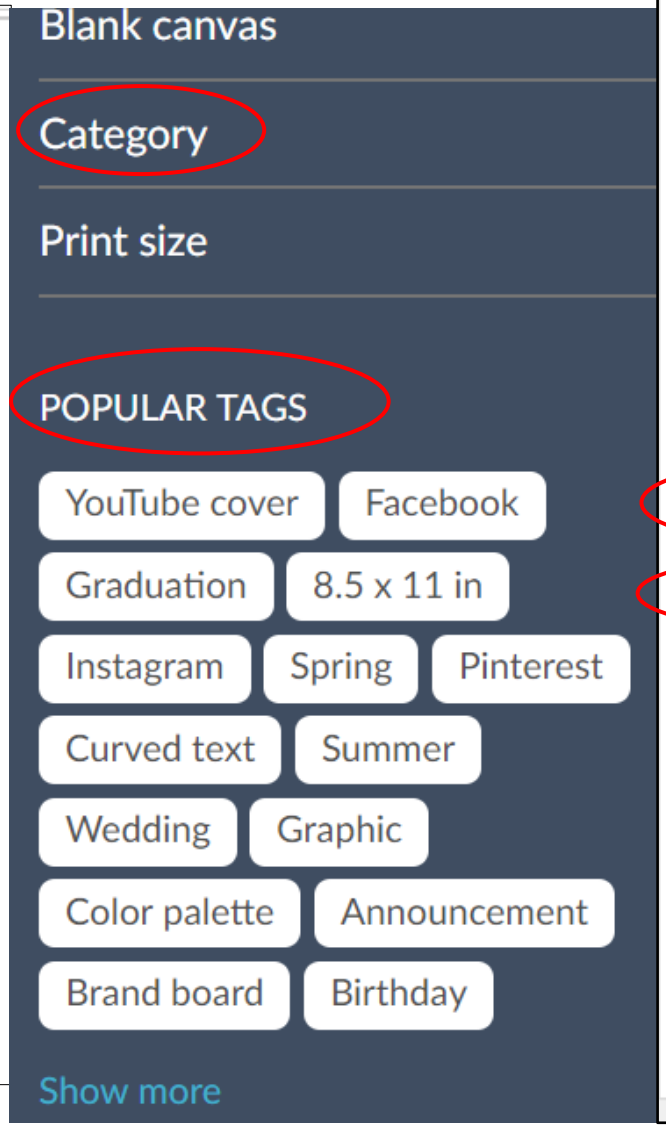
- A single taxonomy may include both approaches/features of (1) organized guiding classification structure (2) controlled vocabulary search terms.
- Sometime two different taxonomies exist at the same time, as supported by many content management systems, one for supporting hierarchical classification (as categories) one for supporting search (as tags).

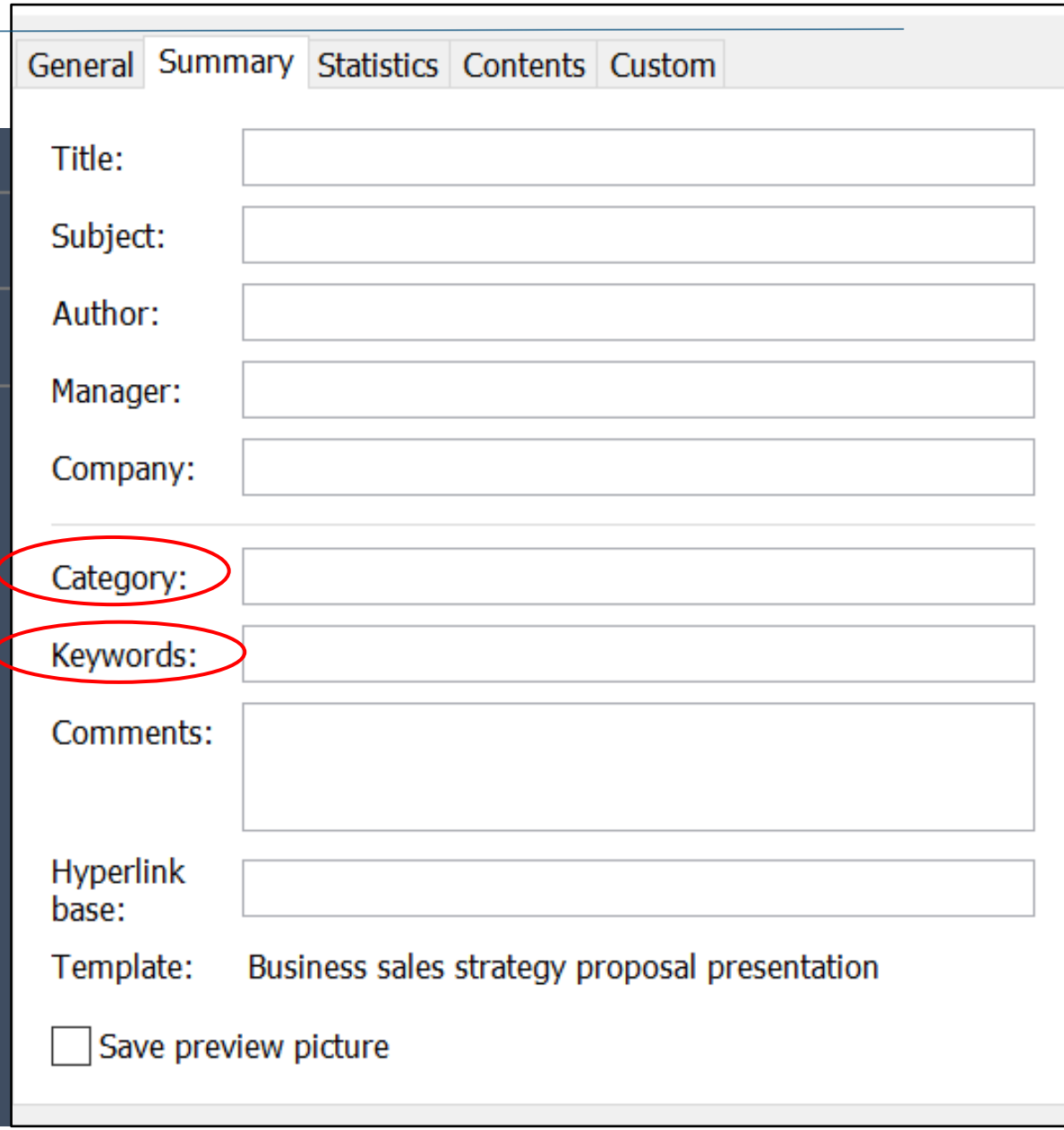
- Categories or tags both are familiar to users
- Categories are often virtual folders to “contain” or group content
- Tags are searchable metadata associated with content
- The same content may be managed by both categories and tags.

vs.

Categories and Tags

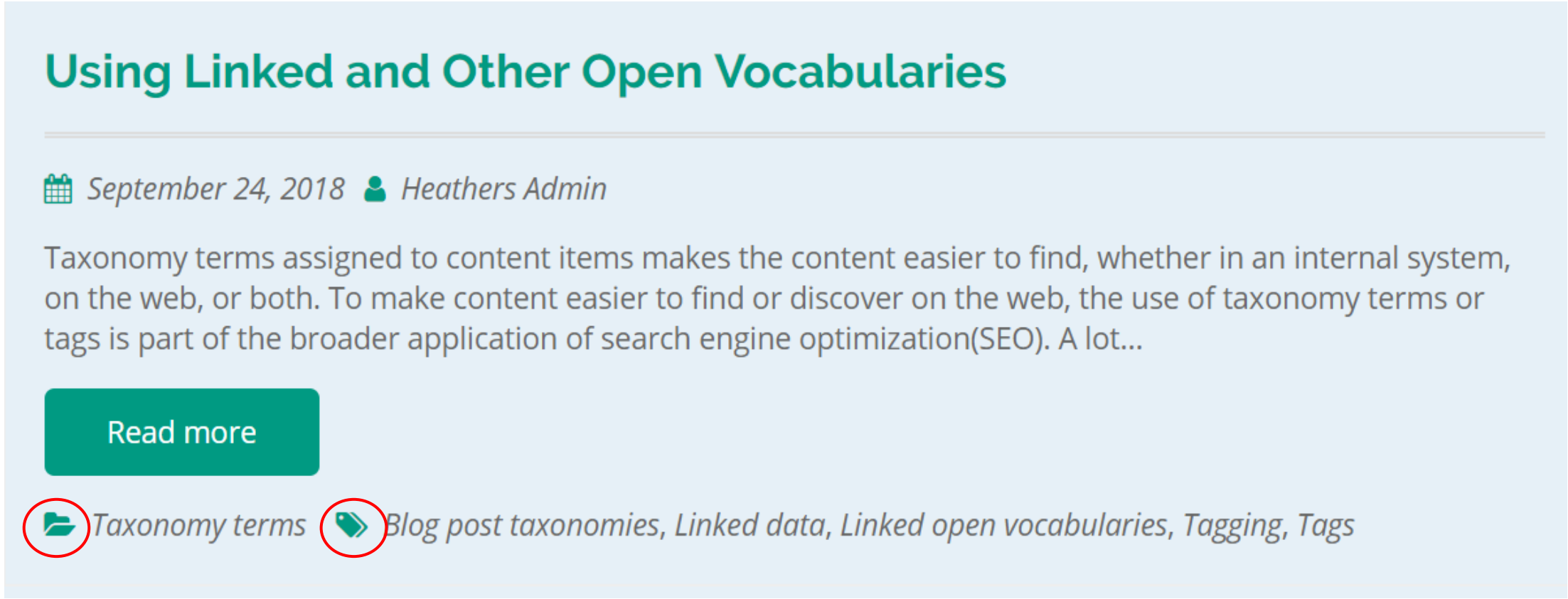






Categories and Tags

Categories vs. Tags



Categories and Tags

When to create categories instead of topical tags

- If a dominant means of categorizing (e.g. Content Type, Discipline, Brand) is identified through conversations with users.
- If the same set of users usually work in the same category, accessing their “go-to” folder.
- If the files always stay in this repository rather than “travel” downstream to other applications (so don’t need taxonomy as attached metadata).
- If the taxonomy is relatively small (and there is no need for synonyms).
- If there is the desirability for a hierarchical taxonomy but the CMS metadata fields don’t support it.
- If there are problems with user compliance in tagging.
- If users clearly prefer category folders (based on use cases).

Categories and Tags

Categories and Tags comparison

Categories

- What “buckets” the content goes into
- Like a table of contents
- Relatively broad
- Limited in number
- Mutually exclusive
- Sometimes hierarchical
- More controlled
- Pre-planned
- Supports browsing

Tags

- What topics the content contains
- Like an index
- More specific
- More numerous
- Overlapping
- Often unstructured
- Less controlled
- Ad hoc
- Supports searching

Categories and Tags

Issues in Taxonomy Approach

- **Inconsistent user experience / view of the taxonomies**

  If creating completely separate taxonomies (e.g. categories and tags)…

  - User sees certain terms in a browsed hierarchy, but differently worded terms for same concept in search refinement topics

- **Risk of wrong design**

  If using the design approach of a traditional hierarchical taxonomy for the purpose of tagging and search support…

  - Creating a classification taxonomy won't be effective when what is needed are simply topics to limit results by.

- **Compromised taxonomy**

  If creating a single taxonomy for both/multiple purposes…

  - Might be too detailed for hierarchical categories and not detailed enough for search terms/tags

Tagging with a Taxonomy

Indexing – assigning index terms (Subject metadata and related elements)

- By trained indexers, often as their primary job responsibility
- By selecting terms from a large controlled vocabulary/thesaurus/taxonomy
- May or may not be metadata on content; could be links instead
- Library science term

Tagging – assigning metadata labels (“tags”)

- By content creators or editors (minimally trained in tagging or not at all), not as their primary job responsibility
- For metadata both with and without controlled vocabularies (could be just user-created keywords)
- Can also be semi-automated
- Assigning “tags” or “categories”
- Information technology term

Tagging with a Taxonomy

Indexing or tagging with a controlled vocabulary or not

Controlled vocabulary

- Using only pre-approved terms
- Used by indexers and content managers
- Ensures consistent indexing
- Slower to change and updates

Keywords

- Assigning any terms desired
- Used by authors and editors
- Tends to inconsistent terms and indexing
- Responsive to trends and dynamic
- May supplement a controlled vocabulary

Folksonomy

- Assigning any terms and reusing terms
- Used by authors, editors, content managers, users
- Tends to inconsistent terms and indexing
- Responsive to trends and dynamic
- May supplement a controlled vocabulary
- More collaborative as “social tagging”

Tagging with a Taxonomy

Tagging can be:

1. Automatic
   - automatically derived from physical file (file format, file size, etc.)
   - automatically applied based on business rules (document type, title, author, topic, etc.)
   - using auto-categorization/text analytics software
2. Automatically suggested and manually reviewed and edited
3. Manually applied

- Choice of method depends on the nature of the content and on the metadata.
- Often, it's a combination.

Tagging with a Taxonomy

Tagging Recommendations

- Automate what can tagged/captured through auto-detection

  File format, Language, Date, often Title, etc.

- Implement system-enforced rules for required/not required elements, and for the number of values permitted and required per element (“cardinality”)

- Tagging can be a burden on people:
  - Do not have available more metadata fields than necessary.
  - Do not require values for more metadata fields than necessary.

Tagging with a Taxonomy

Choosing human vs. automated indexing/classification/tagging

Human methods

- Manageable number of docs
- Higher accuracy in indexing
- May include non-text files
- Investing in people
- Low-tech: can build your own indexing tool/user interface
- Internal control

Automated methods

- Very large number of docs
- Greater speed in indexing
- Generally text files only
- Investing in technology
- High-tech: must purchase auto-indexing/classification software
- Software vendor relationship

Taxonomy Governance and Maintenance

Indexing/tagging policy

For each taxonomy, concept scheme, term list, facet, or metadata element
Rules for indexing/tagging:

- Required or not required for each content item
- Required for certain kinds of content
- Single only or multiple terms from the same taxonomy, facet, or metadata field permitted for each content item
- Any dependency of one metadata field on another metadata field (such as a subfield)

General indexing guidelines/policy:

- Depth of indexing: approximate range of number of terms per content item
- Indexing with broader term in addition to specific terms
- Can terms be assigned “weight” in indexing

Tagging with a Taxonomy

Feedback from indexing to improve the taxonomy, add new terms

Often based on statistics on term usage in indexing

- Underused terms may need added alternative labels or relationships.
- Overused terms may need to be split into more specific terms.
- Misused terms may need rewording, scope note, and/or alternative labels.
- Correctly used low-use terms can be dropped.

Also based on indexers' individual requests and queries

Resources

Books

Hedden, Heather. (2016) *The Accidental Taxonomist, 2nd edition.* Medford, NJ: Information Today Inc. http://www.hedden-information.com/accidental-taxonomist/

Lambe, Patrick. (2007). *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Oxford, England: Chandos Publishing.

Abbas, June. (2010) *Structures for Organizing Knowledge*. New York: Neal-Schuman Publishers.

Hlava, Marjorie M.K. (2015) *The Taxobook*. Morgan & Claypool Publishers.

Resources

Standards and Guidelines

ANSI/NISO Z39.19-2005 (2010) *Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies.* Bethesda, MD: NISO Press. www.niso.org/publications/ansiniso-z3919-2005-r2010

ISO 25964-1 Part 1, Thesauri and interoperability with other vocabularies www.iso.org/standard/53657.html

NISO TR-06-2017 Issues in Vocabulary Management www.niso.org/publications/tr-06-2017-issues-vocabulary-management

Resources

Websites

Accidental Taxonomist book websites
http://www.hedden-information.com/accidental-taxonomist/websites/

Construction of Controlled Vocabularies: A Primer
http://marciazeng.slis.kent.edu/Z3919/index.htm

Thesaurus Construction tutorial by Tim Craven
http://publish.uwo.ca/~craven/677/thesaur/main00.htm

The Accidental Taxonomist Blog
http://accidental-taxonomist.blogspot.com

Hedden Information Management past presentations
http://www.hedden-information.com/presentations/

Resources

Courses, Conferences, Webinars

Taxonomies and Controlled Vocabularies" self-paced online course from Hedden Information Management (15% off with code TAXON15)
http://www.hedden-information.com/courses-workshops/taxonomy-course/

Taxonomy Boot Camp
November 15-16, 2021, Washington DC
https://www.taxonomybootcamp.com

Taxonomy Boot Camp London
Virtual half-day conference
http://www.taxonomybootcamp.com/london

SLA Taxonomy Division webinars
http://taxonomy.sla.org

Questions/Contact

Heather Hedden
Data & Knowledge Engineer
Semantic Web Company, Inc.
One Boston Place, Suite 2600
Boston, MA 02108
www.semantic-web.com
www.poolparty.biz
heather.hedden@semantic-web.com

Hedden Information Management (taxonomy training)
www.hedden-information.com
accidental-taxonomist.blogspot.com
heather@hedden.net