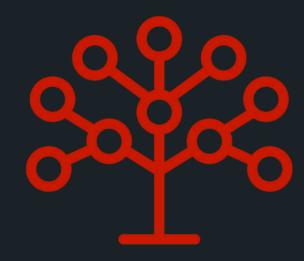


Heather Hedden

JULY 15, 2020 – 10AM PT/1PM ET



Introduction to Taxonomies For Technical Writers









Heather Hedden



Introduction to Taxonomies for Technical Writers

Heather Hedden
Taxonomy Consultant
Hedden Information Management

BrightTALK: The Content Wrangler Webinar

July 15, 2020

About Heather Hedden

- Taxonomy consultant
 - For intranets, enterprise content management, SharePoint, research databases, websites, ecommerce, etc.
 - Independent, through Hedden Information Management
 - Previously as an employed and contract consultant
- Instructor of online and onsite taxonomy courses
 - Independently through Hedden Information Management
 - Previously at Simmons University Library & Information Science School
- Former staff taxonomist
 - At various companies: Gale/Cengage Learning, Viziant, First Wind
- Former indexer
 - Back of the book freelance, periodical article indexer at Gale
- Author of The Accidental Taxonomist (2010, 2016, Information Today, Inc.)

Outline

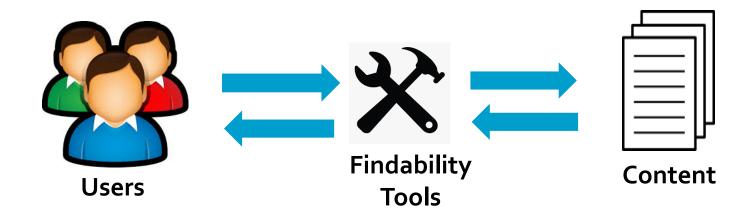
- Methods for Content Findability
- What is a Taxonomy
- Categories and Tags
- Taxonomy Features
- Taxonomy Creation

Outline

- Methods for Content Findability
- What is a Taxonomy
- Categories and Tags
- Taxonomy Features
- Taxonomy Creation

Tools to connect users to specifically sought content

- 1. Indexes browsable, alphabetically displayed back-of-the-book style
- 2. Search engines
- 3. Content/document management systems with (taxonomy) tagged content



Indexes

"A systematic guide designed to indicate topics or features of documents in order to facilitate retrieval of documents or parts of documents"

--NISO TR02-1997

Print

- Book/manual indexes
- Periodical article indexes (formerly)

Electronic

- Ebook hyperlinked indexes
- Periodical article/reference/database indexes
- Web site A-Z indexes (have become rare)
- Online help A-Z indexes (have become rare)

Α

access of web pages. See web page access algorithms. See computational approaches anchor points. See entries for landmarks artificial intelligence, 14–16

E

bookmarks, web, 22 usage, 48-49, 74-75, 98-99 buildings as landmarks, 24-26, 118

Chi, E. H., 2, 113-114
cognitive landmarks. See semantic landmark characteristics
cognitive maps, 1, 9, 26. See also spatial cognition
landmarks and, 28-29, 110
computational approaches, 32-33, 43, 47, 118-119
content, text, in web pages. See semantic landmark
characteristics

Γ

data. See statistical data design, urban, 8—9 disorientation, hypertext, 20—21

E

environments, types of, 7

\mathbf{F}

favorite sites. See bookmarks, web

ϵ

Golledge, R. G., 1, 26, 28 graph theory, 12–13 graphical overview maps, 22 graphics, web. See visual landmark characteristics guided tours, web, 22, 112

\mathbf{H}

hierarchies, 22

A-Z Displayed Indexes

Advantages

- Best method for print
- Familiar and intuitive to users
- Standardized features

Disadvantages

- Does not scale well to large-volume electronic content
- Is not dynamic for changing, growing content
- Does not support discovery or multi-faceted queries well



Search engines (enterprise search, site search)

Searching within a large document (ebook) or across multiple documents.

Advantages



- Scales well for large volumes of content
- Is dynamic for growing, changing content

Disadvantages

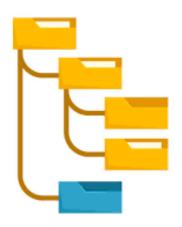
- Misses relevant documents if identical words are not used in the search.
 Problem of synonyms: User searches cars, but text has automobiles.
- Retrieves incorrect documents with search words having more than one meaning (homographs) or are negated.
- Quality of results depend on search expertise of the user.

You get what you pay for.

Content/document management systems and tagged content

- Allow tagging of content for customizable attributes (topics, other)
- May support both categories and tags
- Support controlled vocabularies/taxonomies and uncontrolled keywords

- For multiple small documents
- For larger documents broken into smaller content units



Content/document management systems and tagged content

Advantages

- Scales well and adapts to growing, changing content
- Has better (more comprehensive and accurate) results than search alone
- Supports discovery and multi-faceted queries
- Supports novice and experienced users

Disadvantages (Caveats)

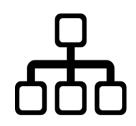
- Dependent on flexible, well-designed user interface
- Quality of results depend on quality of controlled vocabularies/ taxonomies and quality of tagging/indexing.

Outline

- Methods for Content Findability
- What is a Taxonomy
- Categories and Tags
- Taxonomy Features
- Taxonomy Creation

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 or documents, so that they can be more precisely and comprehensively retrieved (than by search algorithms alone on words in text).



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

community

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

services

volunteers

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

discussion forums

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

jobs

accounting+finance admin / office arch / engineering art / media / design biotech / science business / mgmt customer service

education

etc / misc

general labor

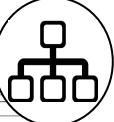
government



food / bev / h Craigslist Boston boston.craigslist.org

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



Free next-day ship	Office Supplies	>	Pens, Pencils & Markers	Filing & Folders	Desk Accessories
	Furniture	>	Pens	File Folders	Desk Organizers
	Cleaning	>	Markers & Highlighters Pencils	Hanging File Folders Expanding File Folders	Desk Trays File Organizers
	Breakroom	>	Correction Fluid & Tape	Classification Folders	Drawer Organizers
Big	Paper	>	Basic Supplies	Binders & Accessories	Calendars & Planners
^	School Supplies	>	Staplers & Staples	Binders	Calendars
A	Technology	> Pa	Paper & Binder Clips	Dividers & Tabs	Planners
F	Ink & Toner	>	Scissors	Report Covers & Portfolios	Organizers
Fre fu	Tax Return Solutions		Paper Punches & Cutters	Sheet Protectors	More Ways to Shop
101			Mailing & Shipping	Labels & Label Makers	Tax Return Solutions
Cold & Flu			Envelopes	Standard Labels	Post-It & Sticky Note
			Shipping & Moving Boxes	Label Maker Tapes	Tape & Adhesives
			Mailers	Address Labels	Storage Boxes & Bins
	vw.officedepot.com		Packing Tape	Label Makers	Presentation Boards
WW			Packing Materials	Name Badges & Holders	

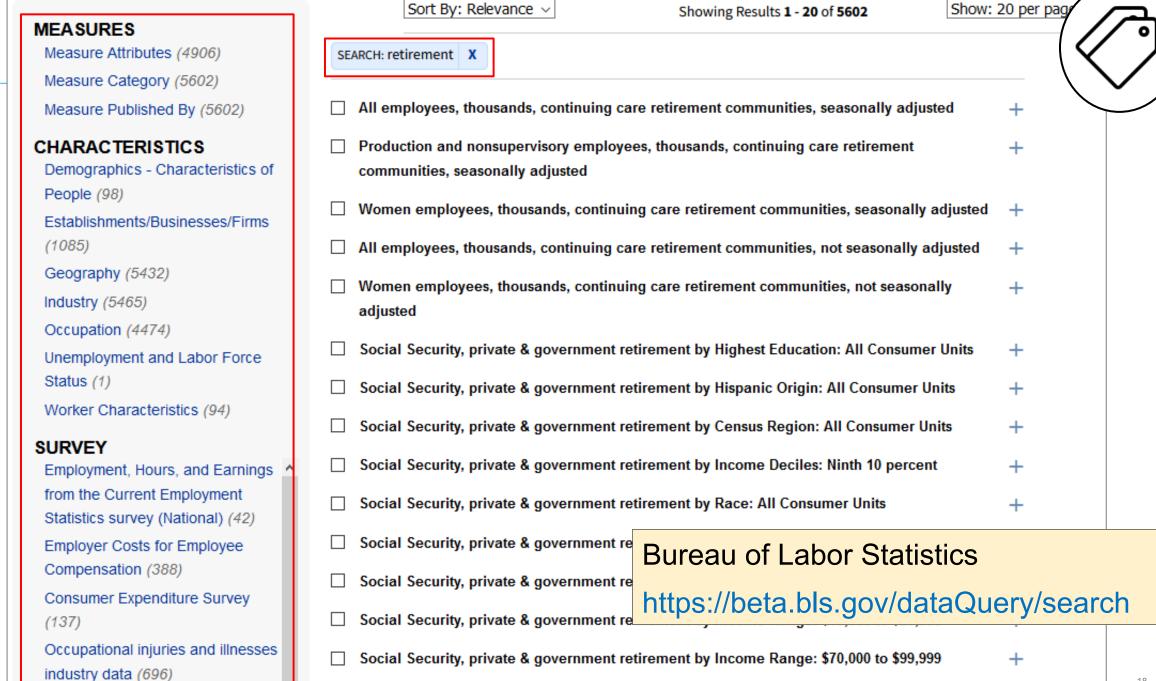
© 2020 Hedden information Management

World Bank Open Data

Free and open access to global development data



pop In Population, male World Bank II Population, total https://data.worldbank.org Il Population, female ili Rural population II Urban population Population, male (% of total population) Population, female (% of total population) Population in largest city



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 organized, 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.

19

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

Faceted

Taxonomy

Example

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]

Taxonomy Examples

Taxonomy

- A controlled vocabulary with broader/narrower (parent/child) term relationships that include all terms to create a hierarchical structure
- Has a 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 as a hierarchy.)

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" or "used for")/non-preferred term (use) (USE/UF)
- Created in accordance with ANSI/NISO Z39.19 standards
- The kind of controlled vocabulary most used in indexing periodical and other research literature.
- Can form the basis of an index of articles/multiple documents.

animal nutrition

Subject Category

L Animal Science and Animal Products S Biological Sciences

RDF/XML Format:

http://lod.nal.usda.gov/nalt/9178.rdf

Persistent URI:

http://lod.nal.usda.gov/nalt/9178

Broader Term

animal science nutrition

Narrower Term

feed conversion feed intake fish nutrition insect nutrition livestock nutrition maternal nutrition poultry nutrition ruminant nutrition

Related Term

animal feeding
animal production
cannulas
diet-related diseases
digestive physiology
feedlots
finishing
forage and feed science
nutrition research

USDA National Agricultural Library Thesaurus

https://agclass.nal. usda.gov

Thesaurus term example

Controlled Vocabulary Types

Less Complexity More

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

Standards for Taxonomies

ANSI/NISO Z39.19 Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies (first published 1974; last update 2010)

- Developed by the National Information Standards Organization (NISO)
- A standard for consistent design for user experience
- Specifies best practices for when and how to created hierarchical, associative, and equivalence relationships, what wording to use to use in terms, and display options.
- Intended especially for thesauri, but applicable (as relevant) to taxonomies

SKOS Simple Knowledge Organization System (first published 2004; last update 2009)

- Developed by the World Wide Web Consortium (W3C)
- A standard for interoperability, both machine and human readable
- Specifies types and notations for relationships, labels, concepts, and schemes
- Intended for all kinds of controlled vocabularies, including taxonomies and thesauri

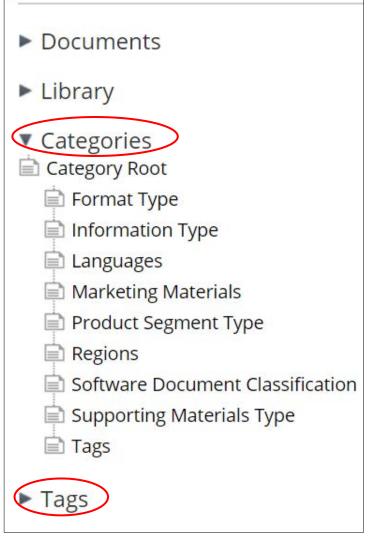
Outline

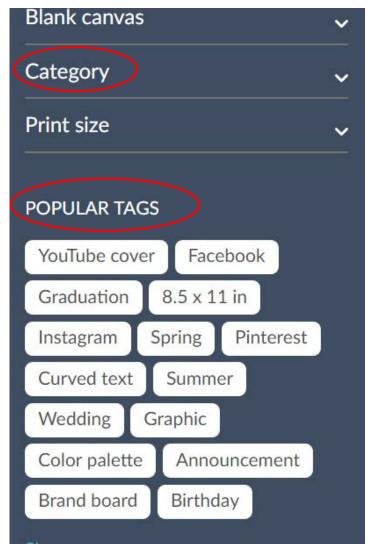
- Methods for Content Findability
- What is a Taxonomy
- Categories and Tags
- Taxonomy Features
- Taxonomy Creation

- A single taxonomy may include both approaches/features of (1) organized guiding classification structure, and (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.



Categories and tags within the same application





General	Summary	Statistics	Contents	Custom	
Title:					
Subject	t:				
Author	:				
Manage	er:				
Compa	nny:				
Catego	ry:				
Keywoi	rds:				
Comm	ents:				
Hyperli base:	ink				
Templa	ate: Busi	ness sales	strategy pı	roposal presentation	
Save preview picture					

Categories and tags within the same application

Using Linked and Other Open Vocabularies

🛗 September 24, 2018 💄 Heathers Admin

Taxonomy terms assigned to content items makes the content easier to find, whether in an internal system, on the web, or both. To make content easier to find or discover on the web, the use of taxonomy terms or tags is part of the broader application of search engine optimization(SEO). A lot...

Read more



Categories and Tags comparison





- ➤ What "buckets" the content goes into
- > Like a table of contents
- Relatively broad
- ➤ Limited in number
- ➤ Mutually exclusive
- Often hierarchical
- > More controlled
- > Pre-planned
- > Supports browsing

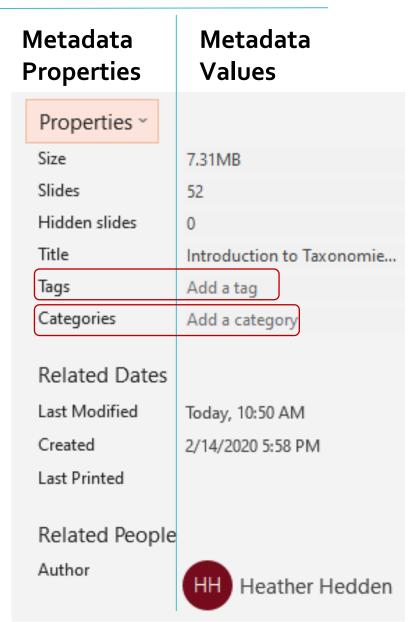
- > What topics the content contains
- ➤ Like an index
- ➤ More specific
- ➤ More numerous
- ➤ Overlapping
- Often unstructured
- > Less controlled
- ➤ May be Ad hoc
- Supports searching

Metadata

Both categories and tags can be metadata.

Metadata and controlled vocabulary

- Some, but not all, metadata elements/properties are populated with 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.
- A hierarchical topical taxonomy could serve a single metadata property, or each facet of a faceted taxonomy could serve a separate metadata property_{© 2020 Hedden Information Management}



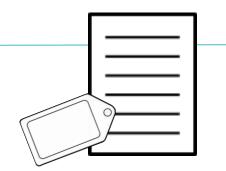
Metadata and Taxonomy Example

Properties	Version	Platform	Language	User	Content Type	Function
Metadata Values	4	Windows	English	Administrator	Key concept	Customizing
	5	Mac	French	Editor	Window/menu description	Messaging
	6	Web	German	Publisher	Trouble-shooting	Navigating
2			Spanish	End-user	Step-by-step	Records updating

Tagging/Indexing

Whether as categories or as tags...

- For structured content:
 Taxonomy terms can become metadata values associated with documents.
- For unstructured content:
 Taxonomy terms can have links to IDs of unstructured content items.
- Content is divided into units; tagging is done on the unit level.
- Tagging functionality is usually a component of content management system.
- Tagging guidelines/policies are needed for each taxonomy hierarchy or facet.
- Those doing tagging should receive instruction/training.
 Tagging should be reviewed.



Outline

- Methods for Content Findability
- What is a Taxonomy
- Categories and Tags
- Taxonomy Features
- Taxonomy Creation

Taxonomy/Thesaurus Features

Synonym Control

As a controlled vocabulary...

Gathers synonyms, acronyms, variant spellings, etc.

Documents not missed due to use of different words (e.g. Automobiles, instead of Cars)

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")

Taxonomy/Thesaurus Features

Synonym Control

Taxonomy or thesaurus contains all variants Text may contain: Users may enter: Oil industry Oil & gas industry Oil industry Oil and gas industry Oil and gas industry Oil & gas industry Oil companies Oil & gas industries Oil & gas industries Oil and gas industries Petroleum industry Oil producers Petroleum industry Oil companies Big oil Oil producers Petroleum companies

Taxonomy/Thesaurus Features

Synonym Control Variants, Synonyms, Alternative Labels, Nonpreferred Terms

 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, thus 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
- To serve as "multiple entry points" to look up and retrieve the desired content, as do double posts or See references in an index.
- To enable consistent indexing

Taxonomy/Thesaurus Features

Synonym Control

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

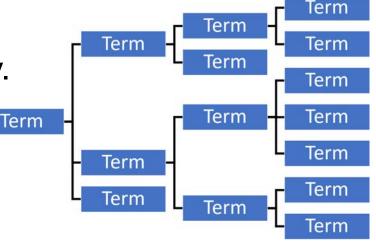
Stewardship (Environmental management)

Piano music [no variants]

Taxonomy/Thesaurus Features

Structure: 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.



Taxonomy/Thesaurus Features

Structure: Hierarchical Relationships

Permitted scenarios:

- Generic specific
 Category and more specific types
- Generic instance
 Common noun proper noun
- Whole Part
 Concept or entity and an integral part

System software

- > Device Drivers
- > Operating systems

Operating systems

- > Linux
- > Windows

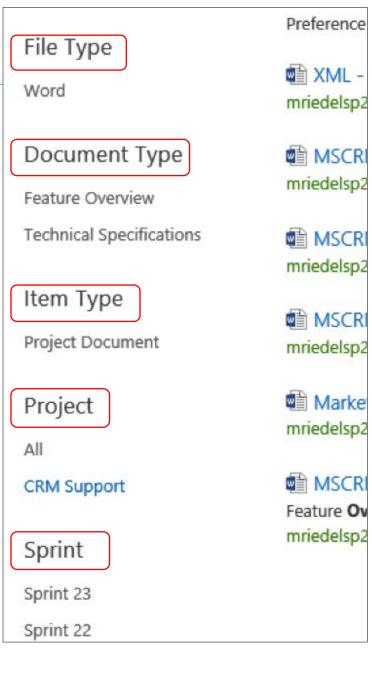
Operating systems

- > Kernel
- > Operating system user interface

Taxonomy Features

Structure: Facets

- For serving faceted classification, which allows the assignment of multiple classifications to a content item.
- A "dimension" of a query; a type of concept; an attribute of a thing; an aspect
- Intended for searching with multiple terms in combination, selected from multiple facets
- A refinement, filter, limit by, narrow by
- A single facet may contain a taxonomy of topics.
- A single taxonomy may be divided into multiple facets.



Taxonomy Features

Structure: 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 unified type of content that share the same multiple facets
- Not practical for extremely large topical taxonomies
- Requires investment of thorough indexing/tagging

Outline

- Methods for Content Findability
- What is a Taxonomy
- Categories and Tags
- Taxonomy Features
- Taxonomy Creation

Taxonomy planning

Determine:

- Who are the stakeholders that need to be involved.
- The scope of subject coverage and content coverage
- Who the users are and what their purposes are (use cases)
- The taxonomy type: hierarchical, faceted, thesaurus-like
- The content/document management system's features and limitations
- How the taxonomy will be maintained



Sources for concepts, labels, and variants

Consider

- Your content (table of contents, headings, descriptive metadata)
- Your users (through interviews, questionnaires, feedback, search logs)
- Existing in-house vocabularies, glossaries, file folder names, navigation menu labels – from different departments (Marketing, Sales, Support, IT)

Be cautious and only consider selected terms from:

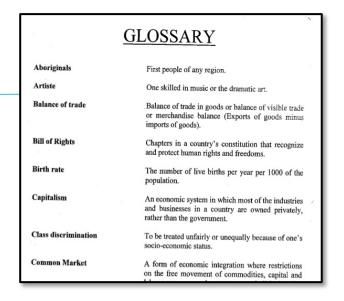
- External taxonomies, classifications, thesauri, etc. for license or free
- Suggestions from engineers, SMEs, or stakeholders who do not know how the taxonomy will be used



Sources for concepts, labels, and variants

Glossaries vs. Taxonomies

- Only some glossary terms would belong in a taxonomy; most would not.
- Glossaries include all terms needing definitions; Taxonomies include concepts only for which there is a significant amount of content.
- Glossaries may include words/terms with multiple meanings; Taxonomy concepts must each have their own unique meaning.
- Glossaries may include terms that a nouns or adjectives; Taxonomies include only nouns or noun phrases.
- Taxonomy concepts usually do not require definitions, rather they may have "scope notes" to explain the scope of the concept in the context of the content.



Sources for concepts, labels, and variants

Tables of contents vs. Taxonomies

All items in a ToC should *covered* by the taxonomy, but ToC chapter or heading names should not become taxonomy terms.

 Chapter 1-Basic of InDesign
 1

 History of InDesign
 1

 InDesign in Ancient Greek
 1

 The Birth of InDesign
 2

 The Death of InDesign
 2

 The Best Publishing Software
 2

 What About Quark
 3

 Why Use inDesign?
 3

 I Love InDesign?
 4

 I Also Love Photoshop
 4

 Who Create InDesign?
 4

 Chapter 2-Typography
 5

 Definition of Typography
 5

 Typography is Text In Visual
 5

 Typhography is Text In Visual
 5

 Typhography Himself
 6

 Mr. Typhography Himself
 6

- Chapters are based on roughly uniform length.
- Chapter and heading names are usually too long, sometimes vague, might be questions or commands and not noun phrases, might be for "introduction" to a topic and repeated later in detail.
- Hierarchies of headings are balanced and don't reflect true taxonomy hierarchy of concepts.
- Topics are presented an instructional or chronological order.

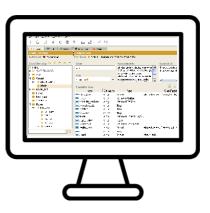
Taxonomy project steps

- 1. Content analysis/audit (look at ToCs, headings, etc.)
- 2. Stakeholder interviews (and use-case creation)
- 3. Review of any existing vocabularies, search logs
- 4. Taxonomy build
- 5. Stakeholder review of taxonomy and feedback
- 6. User testing
- 7. Taxonomy governance documentation
 - Taxonomy maintenance and editorial style policies/guidelines
 - Taxonomy use (tagging) policies/guidelines
- May be part of a larger information or content strategy initiative.



Taxonomy building & maintenance tools

- Spreadsheets (Excel, Google Sheets, etc.) for hierarchies or facets
- Taxonomy/thesaurus management software
 - MultiTes
 - Data Harmony (from Access Innovations)
 - Synaptica
 - PoolParty (from the Semantic Web Company)
 - Semaphore (from Smartlogic)
- Taxonomy editing feature of a content/document management system
- Additional options: term extraction tools and auto-classification tools



Build yourself or contract a consultant

 Depends on: time, taxonomy size and complexity, interest and background in information management

In addition to taxonomy creation, a consultant can determine:

- What kind of controlled vocabulary/taxonomy
- For multiple purposes (indexing, search, and browse);
 a single or multiple separate (or linked) taxonomies
- Taxonomy size, breadth, and depth
- How many facets
- Indexing rules for each facet (required, single or multiple term from a facet)
- Editorial style/format for terms
- What taxonomy management software to consider

Resources

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

Resources

Workshops and Conferences

"Taxonomies & Controlled Vocabularies" self-paced online course through Hedden Information Management

www.hedden-information.com/taxonomy-course-online.htm

Taxonomy creation corporate onsite training through Hedden Information Management www.hedden-information.com/courses-workshops

"Connecting Users to Content: An Introduction to Taxonomy Design & Creation" Virtual conference workshop, September 21-25, 2020 https://euroia.eu

Taxonomy Boot Camp conference, November 16-17, 2020, Washington, DC (or online, TBD) www.taxonomybootcamp.com

Questions/Contact

Heather Hedden Hedden Information Management

heather@hedden.net

98 East Riding Dr. Carlisle, MA 01741 978-467-5195

www.hedden-information.com
accidental-taxonomist.blogspot.com
twitter.com/hhedden
linkedin.com/in/hedden





Heather Hedden





Brought to you by



www.widen.com



Brought to you by

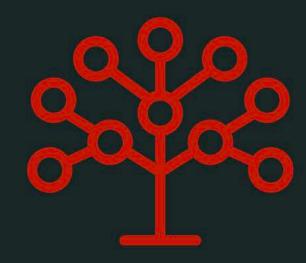






Val Swisher Philipp Baur

JULY 23, 2020 — 10AM PT/1PM ET



Intelligent Authoring Assistance: Reusing Content with Congree

The Electronic Style Guide with Linguistic Intelligence



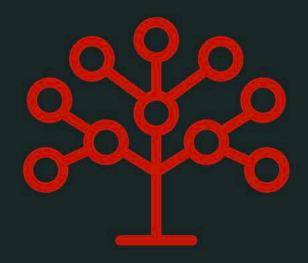






Ann Rockley Val Swisher

AUGUST 12, 2020 - 10AM PT/1PM ET



Using Structured Content to Manage Complex Global Medical Device Content







Val Swisher

AUGUST 19, 2020 – 10AM PT/1PM ET

Best Practices for Creating Reusable Content

Kickstart your content reuse strategy and discover what works and what doesn't

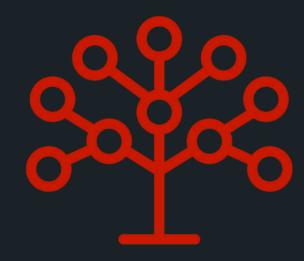






Heather Hedden

JULY 15, 2020 — 10AM PT/1PM ET



Introduction to Taxonomies For Technical Writers



