

Taxonomy Creation with Context

Mini Workshop

Heather Hedden

World IA Day Boston

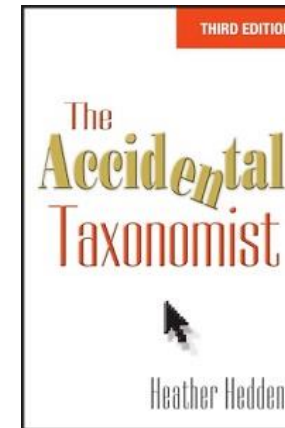
WORLD IA DAY 2024
LOCAL CONNECTIONS. GLOBAL IMPACT.

02 March 2024

#Context

About Heather Hedden

- Senior Consultant, Enterprise Knowledge, LLC
- Instructor of online and corporate taxonomy courses and workshops, Hedden Information Management
- Former taxonomy-related roles at Semantic Web Company, Gale/Cengage, Viziant, First Wind
- Author of *The Accidental Taxonomist*, 3rd ed. (2022, Information Today, Inc.)



Taxonomy Creation with Context: Outline

- Introduction to taxonomies
- Taxonomy creation sources with context
 - Content as a source
 - Users and stakeholders as sources
- Structural design: taxonomy hierarchies and facets
 - Faceted taxonomy exercise
- Concept and label creation

Introduction to Taxonomies

The screenshot shows the eBay website's navigation menu. The 'Shop by category' dropdown is open, displaying several categories. The 'Motors' category is expanded, showing sub-categories like 'Parts & accessories', 'Cars & trucks', 'Motorcycles', and 'Other vehicles'. The 'Clothing & Accessories' category is also expanded, showing sub-categories like 'Women', 'Men', 'Handbags', and 'Collectible Sneakers'. The 'Sporting goods' category is partially visible at the bottom.

The screenshot shows a search results page for the term 'data'. The search bar contains the text 'data'. Below the search bar, there are several filter categories with their respective counts and contexts:

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

Topic

- Advanced Content (73)
- Agile, Design Thinking, & Facilitation (60)
- Artificial Intelligence (25)
- Change Management & Communications (23)
- Company (19)
- Content & Brand Strategy (5)
- Enterprise Learning (23)
- Enterprise Search (49)
- Knowledge Graphs & Data Modeling (108)
- Knowledge Management Strategy & Design (243)
- Taxonomy & Ontology Design (113)
- Technology Solutions (98)

Article Type

- Blog (324)
- Podcast (71)
- Presentation (56)
- Case Study (40)
- White Paper (31)

– Computer and information sciences

Artificial intelligence

- Artificial neural networks
- Expert systems
- Genetic programming
- Machine learning

– Artificial intelligence

Machine learning

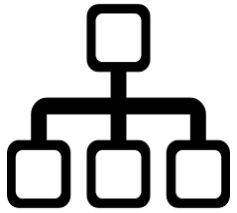
- Decision tree learning
- Deep learning
- Ensemble methods
- Machine learning algorithms
- Relevance vector machines
- Supervised machine learning
- Support vector machines
- Unsupervised machine learning

Taxonomies help people find information

Introduction to Taxonomies

Two major approaches to (or features of) 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 do you think of primarily as a taxonomy?

Introduction to Taxonomies

What is a taxonomy?

A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy.

– *ANSI/NISO Z39.19-2005 r2010 section 4.1 Definitions*

What is a taxonomy not?

- Not just any metadata or tags ➤ Needs to be controlled, structured, related
- Not business glossary ➤ For finding, not defining
- Not a classification scheme ➤ For tagging, not classifying
- Not a navigation scheme ➤ For searching, not just browsing

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 philosophy
- 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

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?

Also, does not reflect the context of a particular content set.

Introduction to Taxonomies

Site navigation hierarchy vs. a taxonomy

Navigation

- Often reflects 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

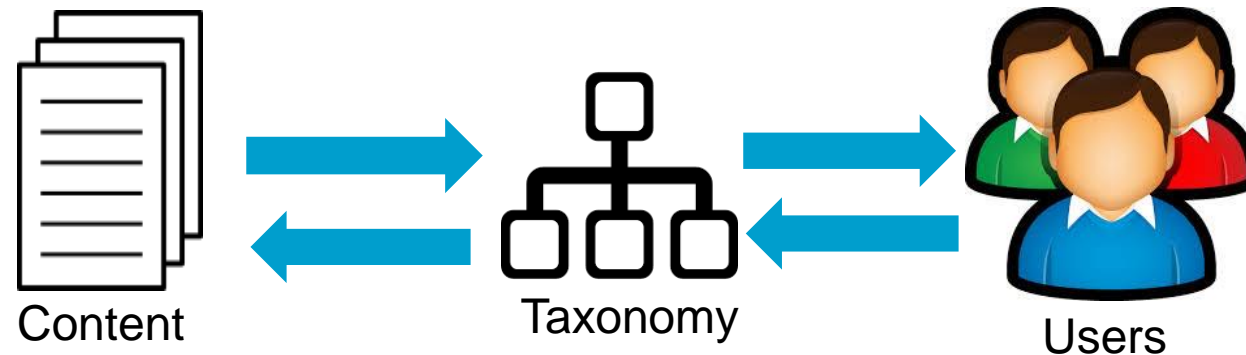
Taxonomies

- Reflect organic relations of the topics
- Based on topics, subjects
- Labels based on terms/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; covers all pages/content
- Neutral to topic importance
- Can grow and adapt without limits
- Often used as metadata

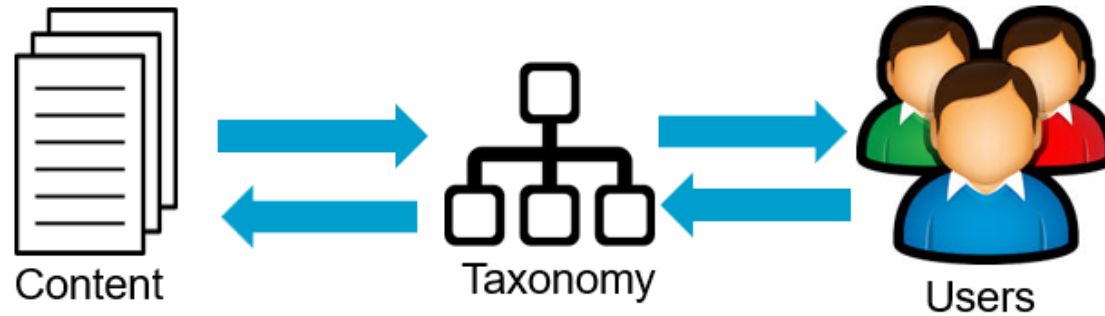
Introduction to Taxonomies

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 **context** of the content and the users.



Taxonomy Creation Sources



1. Content to be tagged: source of concepts
2. Users and other stakeholders: source of concepts and structure

External sources (websites, Wikipedia, other taxonomies and controlled vocabularies, book tables of contents, etc.) *may supplement* the taxonomy.

➤ Whether creating a new or enhancing an existing taxonomy.

Source for Taxonomy Terms: Content

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 terms especially within titles, section headings, lead paragraphs.
 - Look for main idea terms, 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 often get the main idea more reliably than automated methods.



Google, Apple facing anti-competitive complaint in Mexico

PUBLISHED SEP 9 2022 Reuters

MEXICO CITY, Sept 9 (Reuters) - Apple (AAPL.O) and Alphabet's (GOOGL.O) Google are facing a probe over anti-competitive practices in Mexico after the country's former telecommunications chief filed a complaint, he said in a statement on Twitter on Friday.

The complaint was brought to Mexico's telecommunications regulator IFT yesterday by Mony de Swaan Addati, who once headed the regulator before being replaced by IFT.

His complaint accuses the companies of abusing their monopoly in app stores.

In his online statement, he accused the companies of a commission, forcing prices up.

A wave of lawsuits and regulatory actions over the years has led Google to look for more changes are needed.

De Swaan Addati added that he has requested an investigation into the companies, prompting him to take his case to the regulator.

"I have full confidence that (the IFT) will investigate and exercise its powers - in line with international best practices - so that these companies stop abusing their market power to the detriment of developers and consumers," he said.

According to data firm Statista, over 30 million smartphones were forecast to be sold in Mexico last year.

Google's Android has the largest market share in Mexico with 77% penetration, according to the latest data from Statcounter.

Main terms:

[Mexico](#)

[Google](#)

[Apple](#)

[Anti-competitive practices](#)

[App stores](#)

[Instituto Federal de Telecomunicaciones \(IFT\)](#)

[De Swaan Addati, Mony](#)

"taking advantage of in-app purchases."

are charge 15%-20%

s countries over recent years and other critics say

rest to open an

L

N

O

P

Q

R

S

T

U

V

1

Abstract

Population (Patients)

Medical procedure

Music interaction (Recommend changing this to cover other

Health Aspect or Condition

Music type or genre

Study method

Outcome Measures

Setting

Other

10

Does live harp playing have an effect on patient

Post-operative patients (Post-surgical patients)

Vascular and thoracic surgery

Post-operative inpatient music listening

surgical pain, anxiety

Live harp playing

Quasiexperimental, longitudinal (repeated)

visual analog scale, patient satisfaction questionnaire, heart rate, systolic and diastolic blood pressure, respiratory rate, and oxygen saturation

Hospital Vascular Thoracic Unit

Patient satisfaction

11

postoperative cardiovascular surgery patients were

Post-surgical patients (post-operative patients)

Cardiac surgery

Post-surgery music listening

Surgical pain, anxiety, relaxation

Special music including nature sounds

Randomized control trial

Visual analog scales

Hospital Cardiovascular surgical unit

Nature sounds, Patient satisfaction

12

Most research has examined the efficacy of hypnosis,

Pediatric patients (children)

Medical procedures

acute medical procedure pain

Identified terms then grouped into types for developing the Berklee Music & Health taxonomy

literature review

Complementary and alternative medical therapies (Complementary and alternative medicine)

13

systematic review of the efficacy of music therapy (MT) on

Medical procedure patients

Bone marrow transplant

Music listening sessions

bone marrow transplant pain and nausea

Case controlled study

Visual analog scale

relaxaton imagery

systematic review of the efficacy of music therapy (MT) on pain and anxiety in children undergoing clinical procedures. METHODS: We searched 16 electronic databases of published and unpublished studies, subject bibliographies, reference lists of relevant articles, and trials registries. Two reviewers independently screened

Pediatric medical procedure patients,

Music listening sessions; Music therapist involved;

Clinical procedure pain; clincial

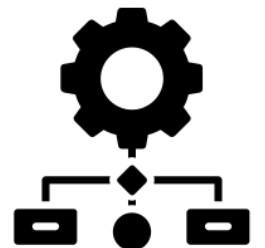
Review of multiple trials (Literature

Active music therapy, Passive

Source for Taxonomy Terms: Content

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.



- Thesaurus
- Employers (31)
- Industries (144)
- Job roles (5)
- Locations (4)
- Skills (3)
- Corpora
 - Job skills
- Candidate Concepts
 - cyber security
 - risk assessment
- Blacklist

Job skills

corpus:08373cca-cd7a-4984-a1b5-b86dcfc3c579

Metadata & Statistics Extracted Concepts **Extracted Terms** Corpus Documents

Search Terms WSI Filter

Search All

Extracted Terms

Term	Relevance	CTS	MIS <input type="button" value="v"/>	Frequency			
Process Improvement	10.02	0	17.1	14	<input checked="" type="checkbox"/>	link	<input type="checkbox"/>
Digital Asset Manager	32.86	0	16.89	6	<input checked="" type="checkbox"/>	link	<input type="checkbox"/>
Managing Director	27.69	0	16.75				
social media	36.69	0	16.73				
video games	2.77	0	16.61				
communication skills	10.71	0	16.61	7	<input checked="" type="checkbox"/>	link	<input type="checkbox"/>
annual budget	6.88	0	16.61	5	<input checked="" type="checkbox"/>	link	<input type="checkbox"/>
Economic evaluation	3.09	0	16.25	3	<input checked="" type="checkbox"/>	link	<input type="checkbox"/>
Content Classification Specialist	1.35	0	16.21	4	<input checked="" type="checkbox"/>	link	<input type="checkbox"/>

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

Sources for Terms & Structure: Users and Other Stakeholders

Gather input from:

- Brainstorming sessions of stakeholders
 - For designating vocabularies and facets; gathering top and sample terms
- 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 from subject experts
 - For gathering detailed terms
- Search log reports
 - For indirect input from end users, of how the word what they are looking for

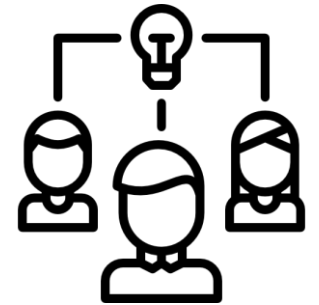


Sources for Terms & Structure: Users and Other Stakeholders

Brainstorming for user input

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

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 Terms & Structure: Users and Other Stakeholders

The screenshot displays the PoolParty Cardsorting interface. At the top left is the 'poolparty CARDSORTING' logo. The top right shows 'User Guide' and a user profile 'AB'. Below the header, there are navigation buttons: '< Move back to Brainstorming Phase' and 'Close Session >'. The main area is titled 'Structure Phase' and contains a collection of cards. Each card represents a term or concept, such as 'Vegeterian chili', 'Appetizer', 'Meal type', 'Breakfast', 'Cupcake', 'Pumpkin Ginger Cupcakes with crystallized ginger', 'Ginger', 'Cinnamon', 'Butter', 'Dinner', 'Cooking for two', 'Lactose intolerance', 'Roasted Pumpkins', 'Pumpkin', 'Thanksgiving Dinner', 'School Lunch Recipes', 'Turkey Tenderloins', and 'Cooking Style'. The cards are arranged in a hierarchical structure, with some cards having colored bars (green, blue, purple) indicating their parent categories. On the left side, there is a 'Taxonomy View' panel showing a tree structure of categories: 'Healthy Recipes (1 item)' containing 'Apple recipes (1 item)' and 'Sauteed Apples X'; 'Lunch X'; 'Mexican food X'; 'Salad recipes (1 item)' containing 'Pasta salad X'; 'Seasonal recipes X'; 'Vegeterian recipes (1 item)' containing 'Hummus Recipes X'. Below the taxonomy view is a 'Filter concepts' section with 'Filter by User' options: 'Andreas Blumauer (8)', 'Blumauer Andreas (4)', and 'Jamie Oliver (6)'. A text box in the bottom right corner explains the process: 'PoolParty Cardsorting: Users create “cards” writing down term ideas, which are grouped, and then candidate terms can be integrated into the taxonomy.'

Sources for Terms & Structure: 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 use later test the taxonomy
 - To identify taxonomy facets and scope
 - To collect *some* terms 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 Terms & Structure: Users

Questions asked in Berklee 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 do?
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 of the following would you most likely *initially* want to search (enter into a search box) on?
 - Specific music therapy method, musical technique, or medical music intervention
 - Who is providing the musical intervention (physician, nurse, music therapist, etc.)
 - Who is choosing the music (patient, medical provider, music therapist)
 - Music type
 - Patient population/type
 - Health issue, disease, condition
 - Medical procedure, intervention, or care type
 - Objective, outcome, or target (pain reduction, anxiety reduction, quality of life, etc.)
 - Test type (Visual analog scale, Hospital Anxiety and Depression Scale, etc.)
5. For the aspects above that you initially search on, how specific would you want to search?
6. How would you look up “patient population/type” for your purposes (if at all)? By medical condition, setting (inpatient/outpatient), demographic, etc.?
7. When searching a database of articles, are you more likely to start with basic search or got 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.	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 patients, 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

Workshop Exercise 1: User Interviews

Taxonomy/IA consultant interviews an information-seeking person.

The interviewee chooses *one* of the seven scenarios:

An information seeking employee at – searching for –

1. Professional services firm – knowledge base information
2. Large employer – employee intranet content
3. Software vendor – technical support information
4. Engineering/manufacturing company – policies and procedures

A consumer seeking information from – to obtain –

5. Edtech/course publisher – courses or learning content
6. Health information service – medical/health information
7. Retailer – products of a certain category

Interviewer asks questions to inform the taxonomy creation.

Then switch roles.



Sources for Terms & Structure: Users and Other Stakeholders

Requesting term lists for user input

- Usually from subject matter experts (SMEs) or domain experts
- In spreadsheets, allowing columns for narrower terms, 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 terms 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			<p>Key: Red - Added Blue - Moved Purple - Changed Cross-out - to Delete</p>	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 Economic		
32		Transport and Economic Geography				move under Transport Economic		
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 Sustainable Development		

Sample taxonomy section suggestions from a SME with taxonomist review

Sources for Terms & Structure: Users

Search log search terms – software documentation help pages

	A	B	C	D	E	F	G
1	Search Term	Total Unique Searches	Results Page View	% Search Exits	% Search Refine	Time After Search	Avg. Search Depth
2	import	104	1.25	0.96%	23.08%	00:08:57	5.95
3	extractor	82	1.43	4.88%	26.50%	00:06:50	6.35
4	corpus	75	1.16	8.00%	19.54%	00:08:00	5.13
5	sparql	73	1.48	12.33%	17.59%	00:04:11	5.26
6	api	65	1.12	6.15%	21.92%	00:06:39	7.68
7	ontology	64	1.19	6.25%	30.26%	00:09:00	5.77
8	excel	56	1.21	7.14%	19.12%	00:10:00	3.8
9	export	43	1.09	11.63%	25.53%	00:03:50	3.21
10	excel import	40	1.1	2.50%	0.00%	00:03:26	3.58
11	workflow	39	1.31	2.56%	27.45%	00:08:10	5.18
12	API	38	1.08	5.26%	26.83%	00:08:09	8.21
13	graphsearch	38	1.18	13.16%	26.67%	00:05:22	5.05
14	snapshot	38	1.16	2.63%	20.45%	00:04:29	3.66
15	language	37	1.54	18.92%	24.56%	00:06:05	4.11
16	collection	35	1.26	0.00%	36.36%	00:11:51	6.26
17	blacklist	34	1.85	14.71%	19.05%	00:07:04	3.76
18	suggest	34	1.41	0.00%	35.42%	00:12:18	4.71
19	linked data	31	1.23	9.68%	13.16%	00:08:17	5.23

Sources for Terms & Structure: Users

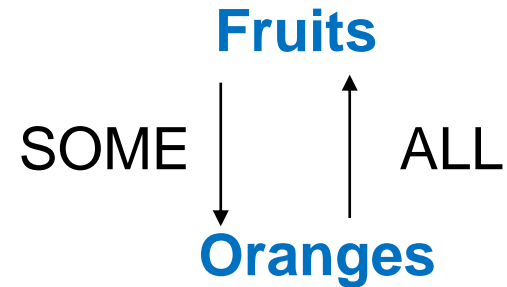
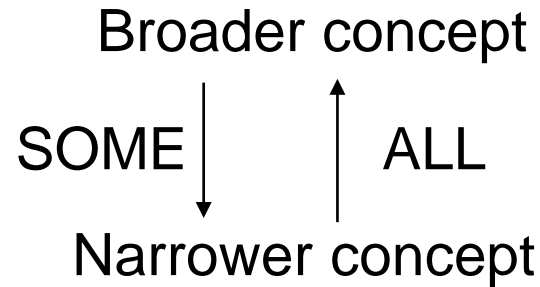
Search log search terms – international organization employee intranet

	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

Structural Taxonomy Design: Hierarchical Relationship

Broader / Narrower

Reciprocal (bi-directional) relationships, but asymmetrical



Fruits (has) narrower **Oranges** **Oranges** (has) broader **Fruits**

Three subtypes (according to ANIS/NISO and ISO thesaurus standards):

1. **Generic – Specific**: “is/are a kind of”
2. **Whole – Part**: “is/are within”
3. **Generic – Instance**: “is an instance of”

- 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

Structural Taxonomy Design: Facets

Facets

- For serving faceted classification, which allows the assignment of multiple concepts of different aspects/types to a content item or record
- “Dimension” of a query, type of term, refinement, filter, limit
- Intended for searching with multiple terms, from different facets, in combination
- Together, reflect the domain of content (content context)
- Reflect the various ways that users would initiate search for content (user context)

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 [\[more\]](#)

Taxonomy Facets

Examples

Department

- Research
- Sales
- Finance
- Marketing
- Exec Office
- SHOW MORE

Job Title

- Communications Director
- Developer
- Financial Controller
- Finance Assistant
- Head of IT Services
- SHOW MORE

Office Location

- London
- New York
- Shanghai

For people

poolparty

Ontologies

Refine search

REMOVE FILTERS | START NEW SEARCH

Document Sources

- User Guide for Knowledge Engineers (12)
- Developer Guide (7)
- PoolParty FAQs (2)
- Administrator Guide (1)
- PoolParty Release Notes (1)

PoolParty Product

- Custom schemes (7)
- Attributes (6)
- Custom ontologies (6)
- Core ontologies (5)
- Details view (5)
- Classes (3)

LOAD MORE

Topics

- Ontologies (24)
- URI (7)
- RDF
- Semantic W
- OWL
- SKOS
- Taxonomies (2)

Search results

Custom scheme a

Custom Scheme and Ontol services can be used to ma custom schemes and ontol Scheme Web Service Meth

Web service meth

Web Service Method: Requ Information Description Ge Content-Type application/x Description Type Required application/json JsonOntol

User groups onto

User Groups - Ontologies Ta the Ontologies tabs in the D available here (1). It displ

Web service meth

Web Service Method: List C Schemes and Ontologies D This service is the equivaler Ontologies URL: /PoolParty

What is a thesaur an ontology

What is a thesaurus, what is thesaurus, what is the diffe to improve most enterprise

Available core on

ologies A logies in AME pro network

Create a custom c

For help documentation

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
- T
- T
- Water

REGION

Choose

COUNTRY

Select country

- Select country
- Afghanistan
- Albania
- Algeria
- Annona

For projects

Content Manager

- Content Manager (7)
- Content Manager (9)
- Content Manager (1)
- Content Manager (1)
- Content Manager (1)

Other Value
Apply | Clear

Distribution Channel

- TechNet Library (7)
- OfficeOnlineVNext (2)
- MSDN Code Gallery (1)
- MSDN Library (1)
- MSN Video (1)

Other Value
Apply | Clear

Request status

- Published (12)

Other Value
Apply | Clear

Content Type

- HVC library content item (8)
- OCMS article (2)
- Code Gallery content (1)
- Video content (1)

Other Value
Apply | Clear

Internal Writer

For documents

Workshop Exercise 2: Design a Faceted Taxonomy

Choose one of the following taxonomy uses:

Internal:

1. Professional services firm – knowledge base
2. Large employer – employee intranet content
3. Software vendor – technical support information
4. Engineering/manufacturing company – policies/procedures

External:

5. Edtech/course publisher – course or learning content
6. Health information service – medical/health information
7. Retailer – ecommerce products of a certain category

- Multiple methods to filter/limit by
- List several terms within each facet.
- Determine if there will be a small hierarchy within a facet.

- Discuss and draft facets.
- Brainstorm concepts.
- Revise facets based on concepts.

Also consider the possible users:

- All employees
- Potential customers
- General public
- Employees in certain knowledge worker roles
- Current customers/clients/partners
- External experts/subject specialists



Concept and Label Creation

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

Format and style: choices to make and to 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

Concept and Label Creation: Synonyms 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 end-user
 - Differences between that of the tagger and the end-user
 - Differences among different end-users
- 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 tagging and comprehensive retrieval

Concept and Label Creation: Synonyms 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]

Conclusions

Building a taxonomy based on...

- Analysis of the actual content
- User interviews and other user input
- User search logs

Results in a more suitable, customized taxonomy based on context,

Results in greater engagement, commitment, and support for the ongoing use, maintenance, and expansion of the taxonomy.

Context is needed for selecting concepts, preferred and alternative labels, determining the degree of hierarchy, and the choice of facets.

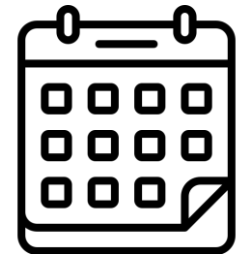
Taxonomy Resources

- ANSI/NISO Z39.19-2005 (2010) Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies
www.niso.org/publications/ansiniso-z3919-2005-r2010
- The Accidental Taxonomist Blog
<http://accidental-taxonomist.blogspot.com>
- Accidental Taxonomist book websites
www.hedden-information.com/accidental-taxonomist/websites
- Hedden Information Management past presentations
www.hedden-information.com/presentations
- Hedden Information Management taxonomy training
www.hedden-information.com/courses-workshops
- Taxonomy Talk, taxonomists community on Discord
<https://discord.com/invite/3qyMVYCAsw>



Upcoming Taxonomy Events

- “Controlled Vocabularies & Taxonomies” HS Events live online course
Wednesdays, February 28, March 6, 13, and 27, 11:00 am – 12:15 pm
www.henrystewartconferences.com/events/controlled-vocabularies-and-taxonomies
- “Taxonomy Creation for Content Tagging” STC live online course
Tuesdays, June 11, 18, and 25, 4:00 – 5:30 pm
www.stc.org/course/taxonomy-creation-for-content-tagging
- Bite-Sized Taxonomy Boot Camp London virtual half-day events. March 20, May 1 (Taxonomy Basics), June 19, and October 9, 9:00 – 11:45 am EDT
www.taxonomybootcamp.com/London
- Taxonomy Boot Camp Conference, November 18 – 19, Washington, DC
www.taxonomybootcamp.com



Questions/Contact

Heather Hedden

Taxonomy Instructor/Author
Hedden Information Management
Carlisle, MA

www.hedden-information.com

accidental-taxonomist.blogspot.com

www.linkedin.com/in/hedden

heather@hedden.net

Senior Consultant (Taxonomy/Ontology)
Enterprise Knowledge, LLC
Arlington, VA

enterprise-knowledge.com

hhedden@enterprise-knowledge.com



ENTERPRISE KNOWLEDGE

Hedden, Heather. (2022) *The Accidental Taxonomist*, 3rd edition.
Medford, NJ: Information Today Inc.

www.hedden-information.com/accidental-taxonomist/

