

Custom Taxonomy Creation

Workshop

World Information Architecture Day - New Hampshire

Heather Hedden



About Heather Hedden

- Data & Knowledge Engineer, Semantic Web Company, vendor of PoolParty software
- Instructor of online and corporate taxonomy courses and workshops and taxonomy review consulting, Hedden Information Management
- Former taxonomy consultant staff taxonomist at Gale/Cengage Learning, Viziant, First Wind
- Author of *The Accidental Taxonomist*, 3rd ed. (2022, Information Today, Inc.)



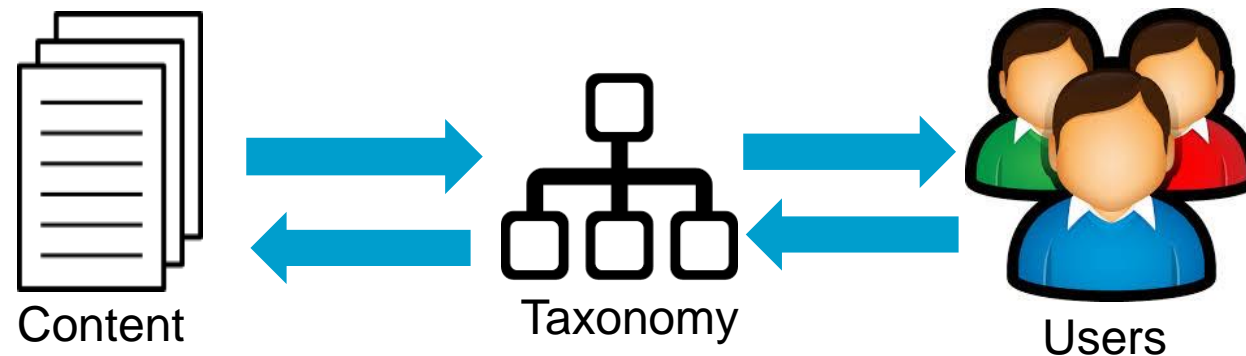
Custom Taxonomy Creation: Outline

- Introduction to taxonomies
- Sources for terms
 - Content as a source
 - Users and stakeholders as sources
- Term and label creation
- Structural design: taxonomy facets

Introduction to Taxonomies

What is a taxonomy for?

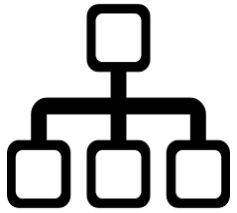
- Concepts/terms are used to tag/index/categorize pages or content to make them easier to be found and retrieved
 - supporting better findability than search alone
- The taxonomy is an intermediary that links the user to the desired content.
- The taxonomy should suit the content and the users.



Introduction to Taxonomies

Two major approaches to (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

Introduction to Taxonomies

The screenshot displays the Office Depot website's navigation menu. The top navigation bar includes the Office DEPOT OfficeMax logo, a search bar, and main categories: Products (expanded), Services, and Deals. The expanded Products menu is organized into a grid of sub-categories and their respective items.

Product Category	Sub-Category	Items
Office Supplies	Pens, Pencils & Markers	Pens Markers & Highlighters Pencils Correction Fluid & Tape
	Basic Supplies	Staplers & Staples Paper & Binder Clips Scissors Paper Punches & Cutters
	Mailing & Shipping	Envelopes Shipping & Moving Boxes Mailers Packing Tape Packing Materials
	Filing & Folders	File Folders Hanging File Folders Expanding File Folders Classification Folders
	Binders & Accessories	Binders Dividers & Tabs Report Covers & Portfolios Sheet Protectors
	Labels & Label Makers	Standard Labels Label Maker Tapes Address Labels Label Makers Name Badges & Holders
	Desk Accessories	Desk Organizers Desk Trays File Organizers Drawer Organizers
	Calendars & Planners	Calendars Planners Organizers
	More Ways to Shop	Tax Return Solutions Post-It & Sticky Notes Tape & Adhesives Storage Boxes & Bins Presentation Boards
	Other Products	Furniture Cleaning Breakroom Paper School Supplies Technology Ink & Toner Tax Return Solutions Cold & Flu

www.officedepot.com

World Bank Open Data

Free and open access to global development data

- Population, male
- Population, total
- Population, female
- Rural population
- Urban population
- Population, male (% of total population)
- Population, female (% of total population)
- Population in largest city

World Bank

<https://data.worldbank.org>

Introduction to T

Structural types of taxonomies

- 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

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

Dewey Decimal Classification

100s level

Is a classification system a taxonomy?

- 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

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

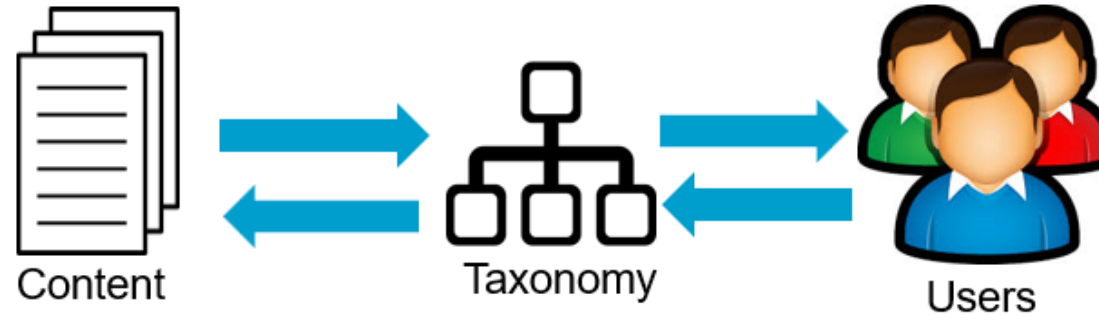
Navigation

- Single-site use and implementation
- 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

- May be re-used in multiple implementations
- 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; can cover all pages/content
- Neutral to topic importance
- Can grow and adapt without limits
- Often are metadata

Sources for Terms



1. Content to be tagged
 2. Users and other stakeholders
 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 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%

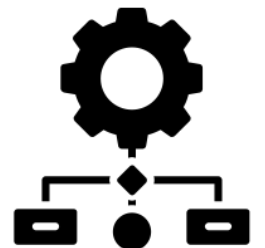
s countries over recent years and other critics say

rest to open an

Sources for 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-b86dffc3c579

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

Search Terms WSI Filter

Extracted Terms

Term	Relevance	CTS	MIS	Frequency	
Process Improvement	10.02	0	17.1	14	<input checked="" type="checkbox"/>
Digital Asset Manager	32.86	0	16.89	6	<input checked="" 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"/>
annual budget	6.88	0	16.61	5	<input checked="" type="checkbox"/>
Economic evaluation	3.09	0	16.25	3	<input checked="" type="checkbox"/>
Content Classification Specialist	1.35	0	16.21	4	<input checked="" type="checkbox"/>

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

Sources for Terms: Users and Other Stakeholders

Gather input from:

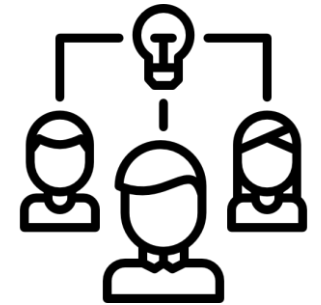
- Brainstorming sessions of stakeholders
 - For designating vocabularies and facets; gathering top and sample terms
- Interviews of sample users and stakeholders
 - For scoping, developing user cases, designating vocabularies and facets, and obtaining some sample terms.
- Requested lists of suggested 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: 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: Users and Other Stakeholders

The screenshot displays the PoolParty Cardsorting interface. At the top, the 'poolparty CARDSORTING' logo is on the left, and 'User Guide' with a user profile icon 'AB' is on the right. 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 and includes a title, a brief description, and a color-coded bar. The cards are arranged in a hierarchical structure. On the left side, there is a 'Taxonomy View' panel showing a tree structure of terms, and a 'Filter concepts' panel with a 'Filter by User' section. The 'Filter by User' section includes three checkboxes: 'Andreas Blumauer (8)', 'Blumauer Andreas (4)', and 'Jamie Oliver (6)'. The cards in the 'Structure Phase' include: 'Vegeterian chili', 'Appetizer', 'School Lunch Recipes', 'Cooking Style', 'Meal type', 'Thanksgiving Dinner', 'Breakfast', 'Cupcake', 'Roasted Pumpkins', 'Pumpkin', 'Pumpkin Ginger Cupcakes with crystallized ginger', 'Lactose intolerance', 'Dinner', 'Cinnamon', 'Ginger', 'Butter', and 'Cooking for two'. The 'Taxonomy View' shows a 'Taxonomy Root (6 items)' with sub-items like 'Healthy Recipes (1 item)', 'Apple recipes (1 item)', 'Sauteed Apples X', 'Lunch X', 'Mexican food X', 'Salad recipes (1 item)', 'Pasta salad X', 'Seasonal recipes X', 'Vegeterian recipes (1 item)', and 'Hummus Recipes X'.

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: 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 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: 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 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
7. 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.	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

Sources for Terms: 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: Users

Search log search terms

Search Terms_PP_Help_2018_2020 ☆ 🗑️ 📄

File Edit View Insert Format Data Tools Extensions Help Last edit was made on 4 May 2020 by Ali Marhubi

🖨️ ▾ 125% ▾ View only ▾

A39 | draft concept

	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: Users

	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 Terms

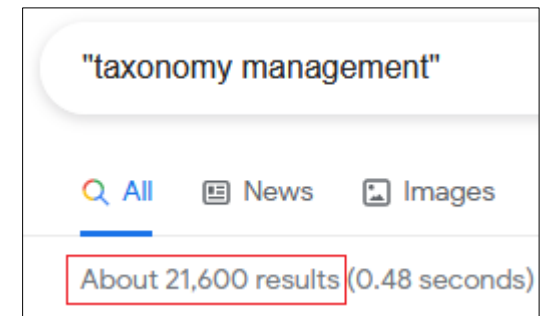
External reference sources

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 terms to include.

Term and Label Creation

Whether a concept should be included as a term

1. Is it within the defined subject-area scope and use of the taxonomy?
2. Is there enough content 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 taxonomies
2. Enforcing organizational/enterprise controlled vocabulary
3. Conforming to academic or professional standards
4. Consistency in style throughout the controlled vocabulary
5. Wording within the documents/content tagged

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***

Term and Label Creation

Term 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

Term 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 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

Term 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]

Structural Taxonomy Design: Facets

Facets

- For serving faceted classification, which allows the assignment of multiple classifications to a content item or record
- “Dimensions” of a query; a type of term
- Intended for searching with multiple terms in combination (post-coordination), one from each facet
- Refinements, filters, limit by, narrow by
- Together, reflect the domain of content

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

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 share the same multiple facets
- May not support “advanced search” of multiple terms selected at once (“or”) from the same facet
- Requires investment of thorough indexing/tagging

Taxonomy Facets

Examples of internal content facets

Department

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

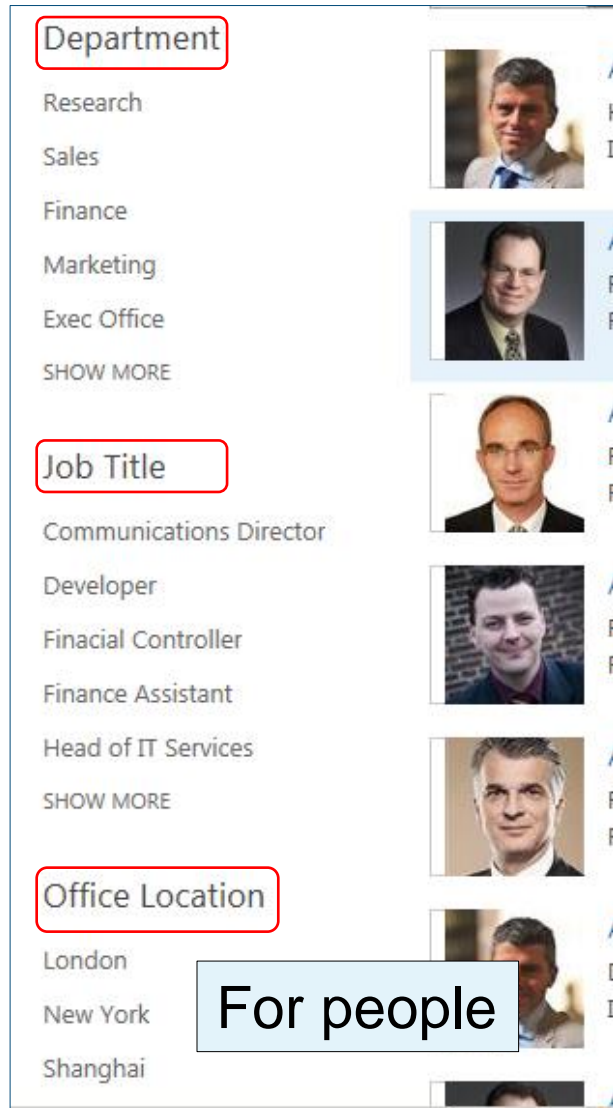
Job Title

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

Office Location

- London
- New York
- Shanghai

For people



File Type

- Word

Document Type

- Feature Overview
- Technical Specifications

Item Type

- Project Document

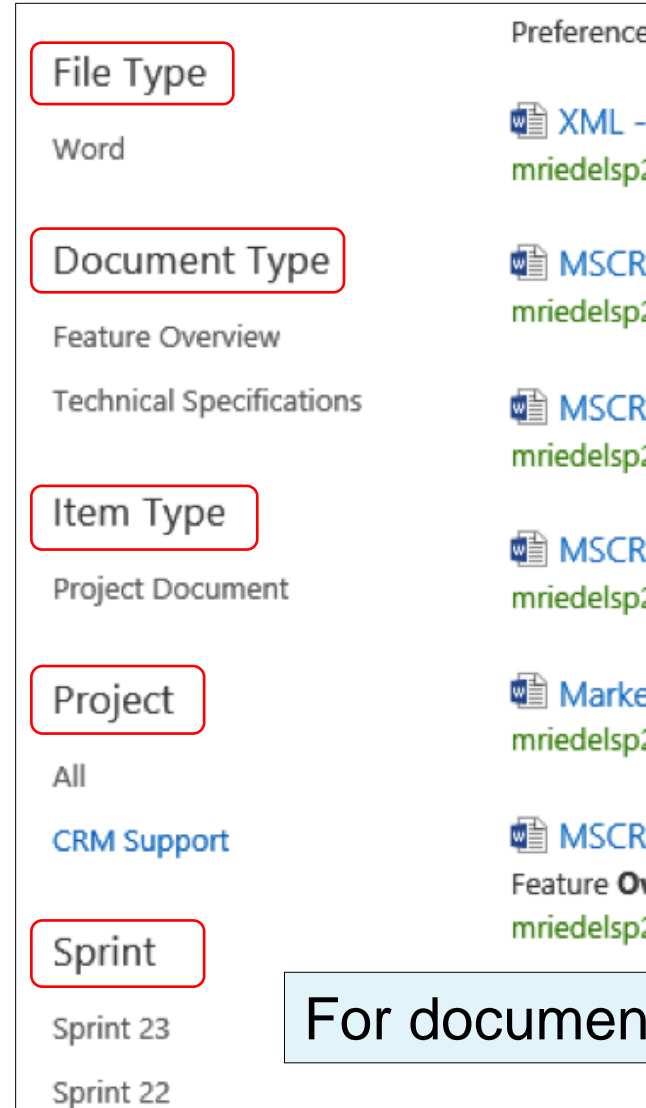
Project

- All
- CRM Support

Sprint

- Sprint 23
- Sprint 22

For documents



Result type

- Word

Author

- User1
- System Account
- AATISH AGARWAL
- SHOW MORE

Modified date



All

Community

- Technical
- Recreational
- Other Value
- Apply | Clear

Language

- English
- German



Content Manager

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

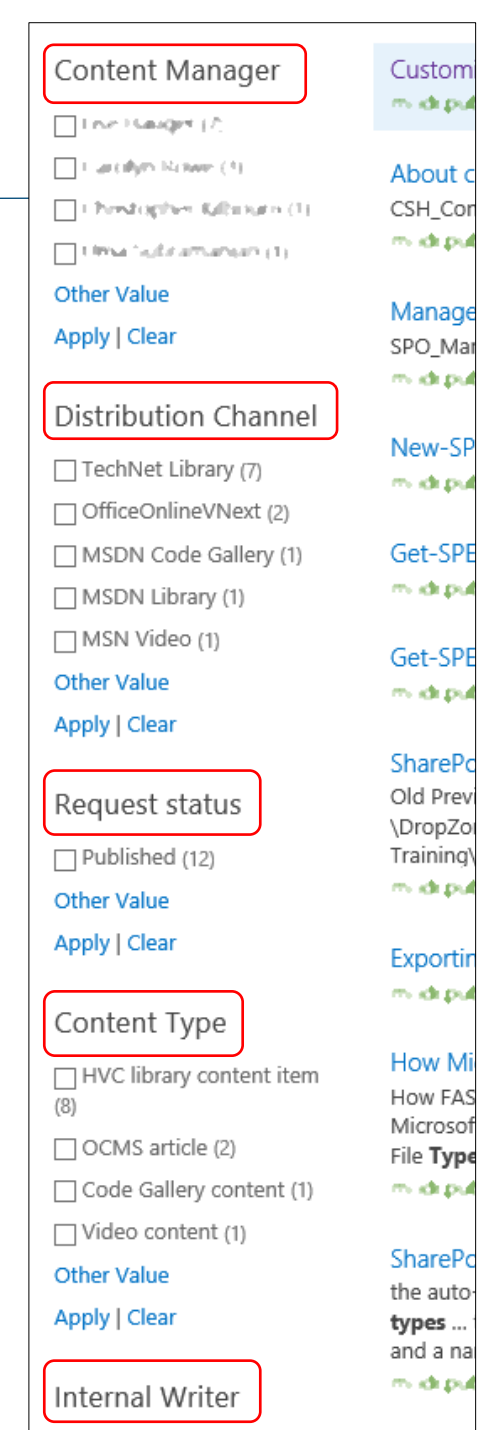
Request status

- 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





Search Space

Semantic Web Research Pap...

Facets

Data resources (2)

Information Sources (3)

Organisations (4)

Persons (2)

Places (4)

Principles (4)

Products (3)

Projects (1)

Specifications (4)

Selected facet values: empty

10 results

[JWS/Mimir-An-open-source-semantic-search-framework-for-interactive-information-seeking-and-discovery 2015 Jou...](#)

http://vocabulary.semantic-web.at/papers/JWS/Mimir-An-open-source-semantic-search-framework-for-interactive-information-seeking-and-discovery_2015_Journal-of-Web-Semantics.pdf

May 28, 2017 8:00:00 PM

JWS/Mimir-An-open-source-semantic-search-framework-for-interactive-information-seeking-and-discovery_2015_Journal-of-Web-Semantics.pdf

Show similar

[JWS/Querying-and-integrating-spatial-temporal-information-on-the-Web-of-Data-via-time-geography 2015 Journal-o...](#)

http://vocabulary.semantic-web.at/papers/JWS/Querying-and-integrating-spatial-temporal-information-on-the-Web-of-Data-via-time-geography_2015_Journal-of-Web-Semantics.pdf

May 28, 2017 8:00:00 PM

JWS/Querying-and-integrating-spatial-temporal-information-on-the-Web-of-Data-via-time-geography_2015_Journal-of-Web-Semantics.pdf

Show similar

[Semantics/2010 a10-celino.pdf](#)

Facets

Taxonomy terms within a facet can be displayed as:

- a list, ranked by order of documents linked to, or
- with the hierarchy of the taxonomy terms

GRAPHSEARCH

Search Space

Semantic Web Research Pa...

Facets

Data resources

Information Sources

Search Facet Values

- Conferences - 8
- SEMANTICS Conference - 8
- Journals - 7
- International Semantic Web Conferen
- Extended Semantic Web Conference
- Magazines - 3
- Twitter - 3
- Wikipedia - 3
- Blogs - 3
- Books - 3
- Social networking - 3
- Flickr - 2
- World Wide Web Conference - 2
- Web sites - 2
- Wikis - 2
- Web of Data - 2
- Facebook - 2

List view

GRAPHSEARCH

Search Space

Semantic Web Research Pa...

Facets

Data resources (2)

Information Sources (3)

- Print media - 8
 - Journals - 7
 - Magazines - 3
 - Books - 3
- Face to face media - 8
 - Conferences - 8
 - Semantic web conferences - 8
 - Web conferences - 3
- Web media - 6
 - Blogs - 5
 - Personal blogs - 3
 - Social networking - 5
 - Wikis - 3
 - Web sites - 2

Organisations (4)

Persons (2)

Tree view

Facets

The number of facets can be limited for the users.

The screenshot shows the PoolParty web application interface. At the top, there is a blue header with the PoolParty logo on the left and 'PoolParty Support Desk' on the right. A search bar in the center contains the text 'Ontologies' and a magnifying glass icon. Below the header, the page is divided into two main sections: 'Refine search' on the left and 'Search results' on the right.

Refine search

At the top of the 'Refine search' section, there are two buttons: 'REMOVE FILTERS' and 'START NEW SEARCH'. Below these are three facet categories:

- Document Sources**: A list of document sources with checkboxes and counts: 'User Guide for Knowledge Engineers' (12), 'Developer Guide' (7), 'PoolParty FAQs' (2), 'Administrator Guide' (1), and 'PoolParty Release Notes' (1).
- PoolParty Product**: A list of product categories with checkboxes and counts: 'Custom schemes' (7), 'Attributes' (6), 'Custom ontologies' (6), 'Core ontologies' (5), 'Details view' (5), and 'Classes' (3). Below this list is a 'LOAD MORE' button.
- Topics**: A list of topics with checkboxes and counts: 'Ontologies' (checked, 24), 'URI' (7), 'RDF' (4), 'Semantic Web' (3), 'OWL' (2), 'SKOS' (2), and 'Taxonomies' (2).

Search results

The 'Search results' section displays three search results, each with a title and a brief description:

- Custom scheme and ontology services**: Custom Scheme and Ontology Services Abstract Custom Scheme and Ontology Services The following services can be used to manage custom classes, relations and attributes. Note To learn more about defining custom schemes and ontologies in PoolParty please see: Create a Custom Ontology and Create a Custom Scheme Web Service Method: ...
- Web service method request ontology information**: Web Service Method: Request Ontology Information Abstract Web Service Method: Request Ontology Information Description Get Ontology information in json format. URL: /PoolParty/api/schema/ontology Content-Type application/x-www-form-urlencoded Supported Methods GET HTTP Parameters Parameter Description Type Required uri URI of the requested ontology. String true Response Content-Type application/json JsonOntologyCollection Attribute Type Comment attributes Array ...
- User groups ontologies tab**: User Groups - Ontologies Tab Abstract User Groups - Ontologies Tab This section contains a short guide on the Ontologies tabs in the Details View of a group. The Ontologies tab is the fourth tab in the row of tabs available here (1). It displays a list of all ontologies ...

Below the search results, there are two more sections:

- Web service method list custom schemes and ontologies**: Web Service Method: List Custom Schemes and Ontologies Abstract Web Service Method: List Custom Schemes and Ontologies Description Retrieve a list of all available ontologies and custom schemes. Note This service is the equivalent to the PoolParty frontend functionality Import, Export & Publish Schemes and Ontologies URL: /PoolParty/api/schema/list Request Supported ...
- What is a thesaurus, what is the difference between a taxonomy and an ontology**: What is a thesaurus, what is the difference between a taxonomy and an ontology? Abstract What is a thesaurus, what is the difference between a taxonomy and an ontology? A thesaurus is expressive enough to improve most enterprise applications significantly, but it is not too complex to create and maintain ...

At the bottom of the 'Search results' section, there are two more sections:

- Available core ontologies**: Available Core Ontologies Abstract Available Core Ontologies This section provides an overview of the available core ontologies in PoolParty. Ontology Description BIBFRAME vocabulary Initiated by the Library of Congress, BIBFRAME provides a foundation for the future of bibliographic description, both on the web, and in the broader networked world that ...
- Create a custom ontology**: A link to create a custom ontology.

Taxonomy 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
- Hierarchy (indented) within a facet is possible

Conclusions

Building a taxonomy based on...

- Analysis of the actual content
- User interviews
- User/stakeholder workshops
- User search logs

Builds a more suitable, customized taxonomy.

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

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

- PoolParty Summit, March 7-8, 8:00 am - 2:00 pm EST, free virtual conference featuring use cases of PoolParty taxonomy/ontology/knowledge graph software www.poolparty.biz/events/poolparty-summit-2023
- "Taxonomy For Delivering Targeted Technical Content" free webinar, hosted by The Content Wrangler, Wednesday, April 12, 1:00 - 2:00 EDT www.brighttalk.com/webcast/9273/557969
- Information Architecture Conference, March 30 - April 1 (March 28-29 pre-conference workshops), New Orleans www.theiaconference.com
- "Taxonomy Basics," online hosted by Bite-sized Taxonomy Boot Camp London, Wednesday, April 26, 9:00 -11:30 am EDT.
Other half-day morning events March 22, June 21, October 11 www.taxonomybootcamp.com/London
- Taxonomy Boot Camp Conference, November 6-7, Washington, DC www.taxonomybootcamp.com

Questions/Contact

Heather Hedden

Data and Knowledge Engineer
Semantic Web Company Inc.
One Boston Place, Suite 2600
Boston, MA 02108
857-400-0183

heather.hedden@semantic-web.com

www.semantic-web.com

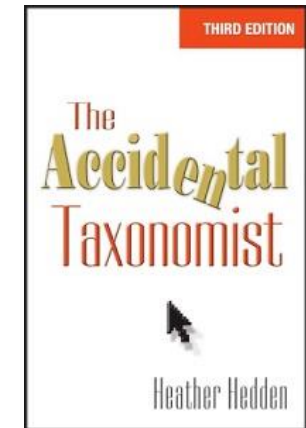
www.poolparty.biz

www.linkedin.com/in/hedden

twitter.com/hhedden

<http://accidental-taxonomist.blogspot.com>

heather@hedden.net



Hedden, Heather. (2022) *The Accidental Taxonomist, 3rd edition*. Medford, NJ: Information Today Inc. www.hedden-information.com/accidental-taxonomist/