

Web-based content organization and the transformation of traditional classification systems

Joseph Busch

Locating objects

- ❖ Traditional classification schemes physically locate objects in one dimension.



Shelf Location

https://upload.wikimedia.org/wikipedia/commons/f/fd/Shelves_of_Language_Books_in_Library.JPG

Table of Contents

Core Concepts	6	Weapons and Armour.....	114
The Basics.....	6	Hunting Traps.....	115
How It Plays.....	7	Sword and Armour	117
Responsibilities.....	9	Armour.....	117
Peoples	10	Assembling Armour.....	118
Bardur.....	11	Melee Weapons.....	123
Dwarves.....	16	Ranged Weapons.....	129
Elnar.....	19	Damage To Arms.....	130
Elves.....	24	Weapon Materials.....	131
Heartlanders.....	29	Rolls and Rules	133
Maktiti.....	31	Task Resolution.....	133
Orcs.....	33	Fatigue.....	139
Senshou.....	36	Triggers.....	141
Toth.....	39	Fear.....	144
Vren.....	42	Social Conflict.....	146
Character Creation	46	Crafting and Smithing.....	149
Group Concept.....	46	Foraging.....	151
Creating a Character.....	47	Survival.....	153
People.....	47	Hunting.....	154
Attributes.....	49	NPC & Beast Skills.....	156
Skills.....	52	Character Development	157
Experience Points.....	54	Experience Points (XP).....	157
Path.....	56	NPC Development.....	159
Aspirations.....	57	Combat	161
Characteristics.....	59	Rounds.....	161
Trigger Ammo.....	61	Melee Attack.....	163
Relationships.....	62	Ranged Attack.....	165
Gear.....	63	Defending.....	167
Finishing Touches.....	65	Tactics.....	169
Paths	73	Movement.....	170
Hero Points.....	73	Reach Advantage.....	172
Way of the Warrior.....	73	Secondary Action.....	173
Way of the Barbarian.....	75	Multiple Attackers.....	173
Way of the Swordsman.....	76	Improvising.....	173
Way of the Archer.....	78	Wrestling.....	176
Way of the Fist.....	79	Throw.....	178
Way of Command.....	80	Cover.....	178
Way of the Rogue.....	82	Combat Scene.....	179
Way of the Beast.....	84	Trauma	180
Way of the Witch.....	85	Wounds.....	180
Way of the Sorcerer.....	88	Inflicting Wounds.....	184
Way of the Demon.....	89	Post Trauma.....	185
Skills	91	Trauma Companion.....	187
Skills by Attribute.....	91	Healing.....	188
Racial Skill Restrictions.....	92	Burns.....	189
Skill Dependencies.....	92	Suffocation.....	190
Skill Descriptions.....	92	Explosions.....	190
Barter	104	Falling.....	191
Wealth.....	104	The Ambush	192
Values.....	105	Game Mastery	195
Coin.....	106	Scene Framing.....	195
Barter.....	106	Pace and Plot.....	197
Regional Values.....	106	NPCs.....	199
Equipment.....	107	Large Scale Conflict.....	202



Stores with a large assortment use well known departments and sub-categories

Supermarket Categories

- ❖ Bakery
- ❖ Beverages & Snacks
- ❖ Dairy
- ❖ Deli & Prepared Foods
- ❖ Grocery
- ❖ Household
- ❖ Meat & Seafood
- ❖ Produce

Department Store Categories

- ❖ Men's Clothing
- ❖ Women's Clothing
- ❖ Baby & Children's Clothing
- ❖ Home Furnishings
- ❖ Electronics
- ❖ Toys & Sports
- ❖ Food

Hierarchical classification problem

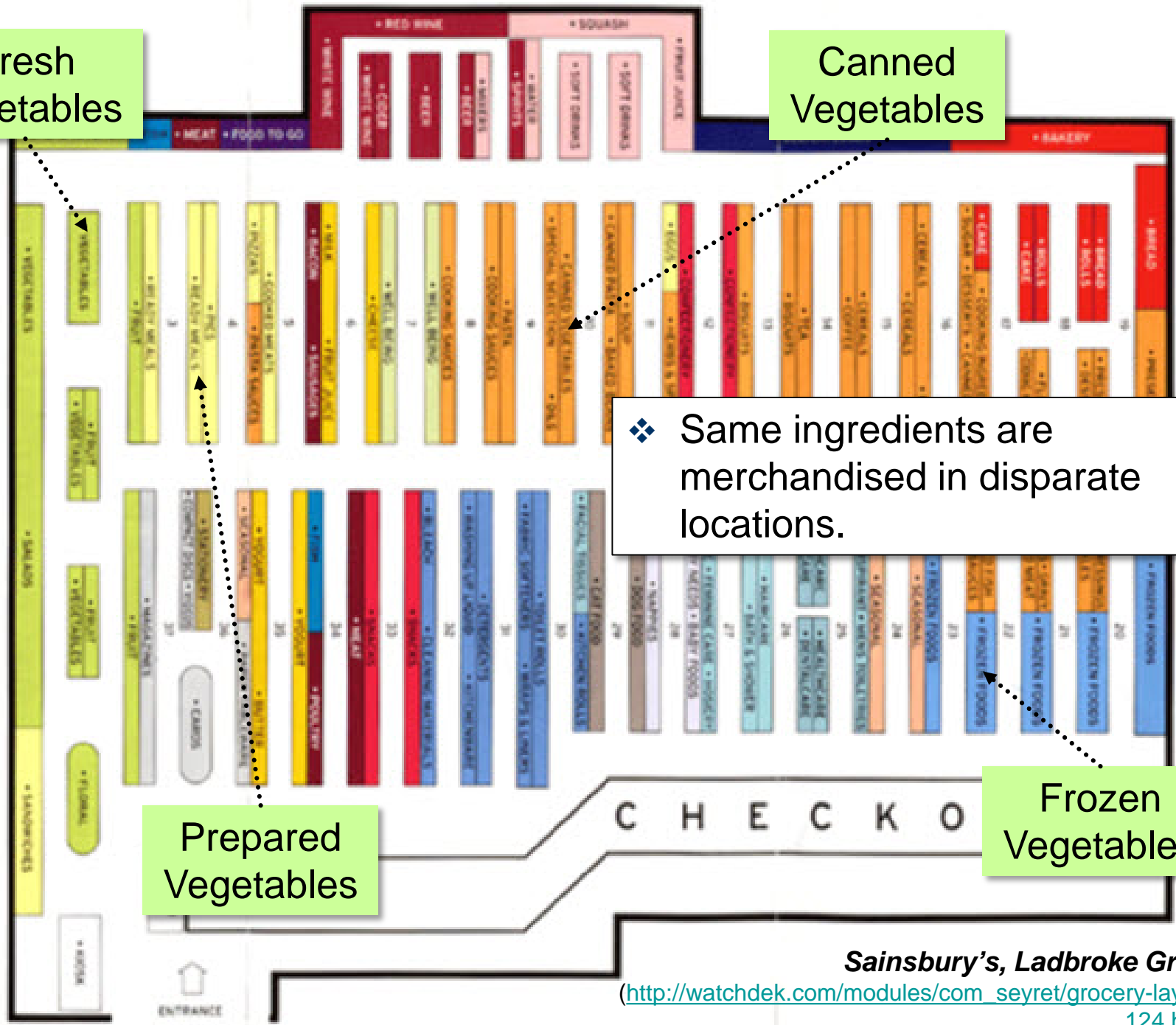
Fresh Vegetables

Canned Vegetables

Prepared Vegetables

Frozen Vegetables

❖ Same ingredients are merchandised in disparate locations.



Physics & Astronomy Classification Scheme (PACS 2010)

- ❖ Alphanumeric codes
- ❖ Strict (single-parent) hierarchy
- ❖ Incomplete and complex labels

- ❖ Some concepts are repeated many times.

The screenshot displays the PACS2010 web interface. On the left, a taxonomy tree shows the hierarchy of physics concepts. On the right, a search interface is shown with the following details:

Search Interface:

- Search input: `semiconductor`
- Filter By: Custom Class: `Concept`, Preferred label: `contains`
- Results: 62 matching concepts!

Search Results Table:

Concept	Alternative Labels	Broader Concepts	# Narrower Concepts	Delete
Amorphous semiconductors		Specific materials: fabrication, treatment, testing, and analysis	0	⊗
Amorphous semiconductors, glasses		Structure of clean surfaces	0	⊗
Amorphous semiconductors, metallic glasses, glasses		Electronic structure of disordered solids	0	⊗
Amorphous semiconductors, metals		Disordered solids	0	⊗

Left Panel (Taxonomy Tree):

- PACS2010 (10)
 - ATOMIC AND MOLECULAR PHYSICS (6)
 - CONDENSED MATTER: ELECTRONIC STRUCTURE, ELECTRICAL, MAGNETIC, AND OPTICAL PROPERTIES (9)
 - CONDENSED MATTER: STRUCTURAL, MECHANICAL, AND THERMAL PROPERTIES (8)
 - ELECTROMAGNETISM, OPTICS, ACOUSTICS, HEAT TRANSFER, CLASSICAL MECHANICS, AND FLUID DYNAMICS (7)
 - GENERAL (7)
 - GEOPHYSICS, ASTRONOMY, AND ASTROPHYSICS (8)
 - INTERDISCIPLINARY PHYSICS AND RELATED AREAS OF SCIENCE AND TECHNOLOGY (8)
 - Biological and medical physics (20)
 - Electronic and magnetic devices; microelectronics (13)
 - Electronics; radiowave and microwave technology; direct energy conversion and storage (11)
 - Materials science (11)
 - Methods of crystal growth; physics and chemistry of crystal growth, crystal morphology, and orientation (8)
 - Methods of deposition of films and coatings: film growth and epitaxy (12)
 - Methods of materials synthesis and materials processing (8)
 - Methods of materials testing and analysis (8)
 - Methods of micro- and nanofabrication and processing (9)
 - Nanoscale materials and structures: fabrication and characterization (11)
 - Other topics in materials science (0)
 - Phase diagrams and microstructures developed by solidification and solid-solid phase transformations (6)
 - Specific materials: fabrication, treatment, testing, and analysis (18)
 - Amorphous semiconductors (0)
 - Carbon/carbon-based materials (5)
 - Ceramics and refractories (0)
 - Cermets, ceramic and refractory composites (0)
 - Dispersion, fiber-, and platelet-reinforced metal-based composites (0)
 - Elemental semiconductors (0)

PACS 2010 semiconductor-related categories are in disparate locations

00—General

- 01. Communication, education, history, and philosophy
- 02. Mathematical methods in physics
- 03. Quantum mechanics, field theories, and special relativity
- 04. General relativity and gravitation
- 05. Statistical physics, thermodynamics, and statistical mechanics
- 06. Metrology, measurements, and standards
- 07. Instruments, apparatus, and computers

60—Condensed Matter: Structural, Mechanical and Thermal Properties

- 61. Structure of solids and liquids; crystallography
- 62. Mechanical and acoustical properties of condensed matter
- 63. Lattice dynamics
- 64. Equations of state, phase equilibria, and phase transitions
- 65. Thermal properties of condensed matter
- 66. Nonelectronic transport properties of condensed matter
- 67. Quantum fluids and solids
- 68. Surfaces and interfaces; thin films and nanosystems (structure and nonelectronic properties)

70—Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties

- 71. Electronic structure of bulk materials
- 72. Electronic transport in condensed matter
- 73. Electronic structure and electrical properties of surfaces, interfaces, thin films, and low-dimensional systems
- 74. Superconductivity
- 75. Magnetic properties and materials
- 76. Magnetic resonances and relaxations in condensed matter, Mössbauer effect
- 77. Dielectrics, piezoelectrics, and ferroelectrics and their properties
- 78. Optical properties, condensed-matter spectroscopy and other interactions of radiation and particles
- 79. Electron and ion emission by liquids and solids; impact phenomena

The *real* world of concepts is multidimensional

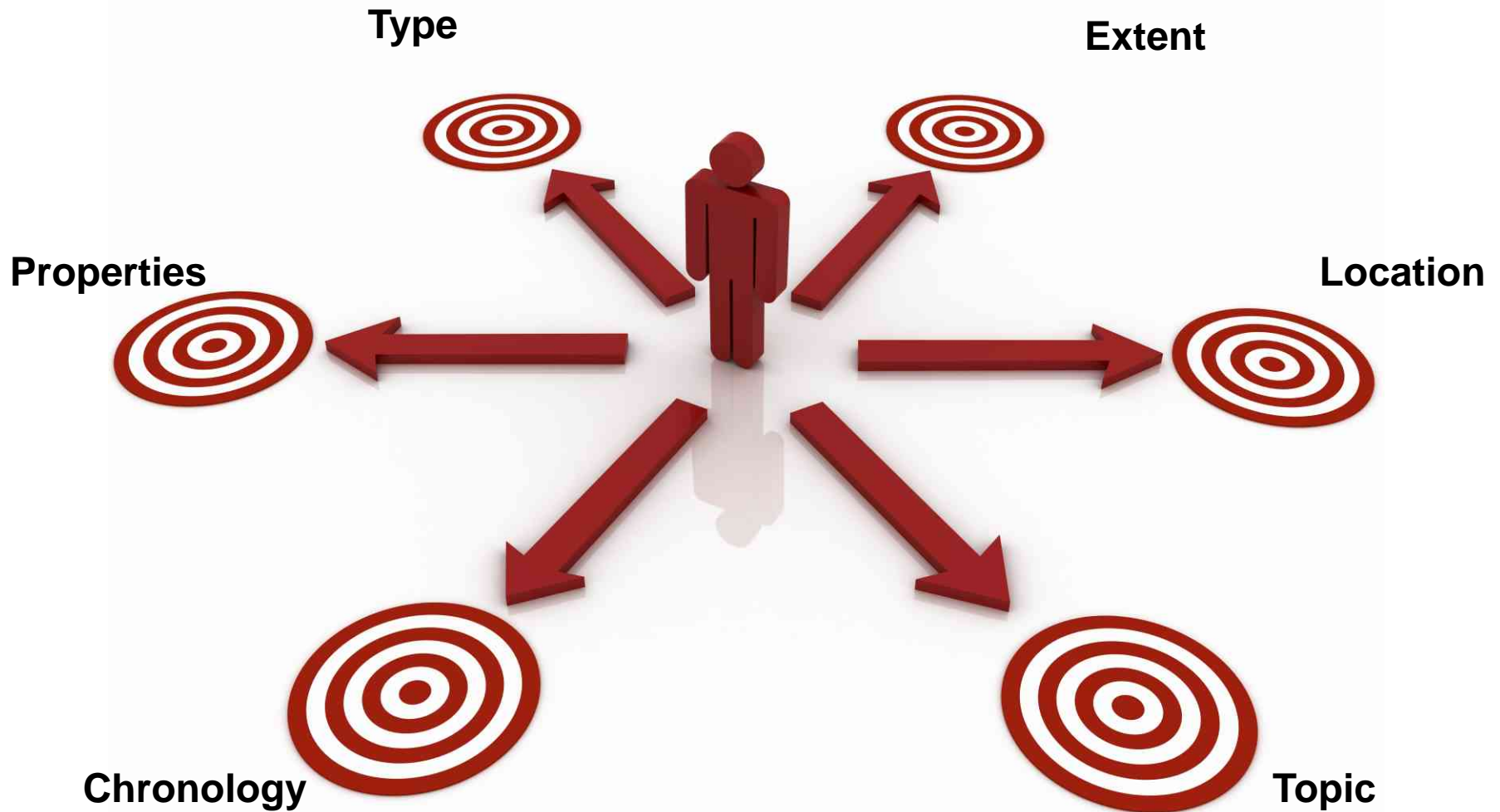


Image source: http://stratconcept.com/yahoo_site_admin/assets/images/Comp_ist2_6702437-many-targets-concept.12743816.jpg

Shopping for shoes on Zappos.com



- ❖ Size
- ❖ Width
- ❖ Styles
- ❖ Occasion
- ❖ Color
- ❖ Brand
- ❖ Price
- ❖ Heel Height
- ❖ Materials
- ❖ Pattern
- ❖ Accents

Filtering Search Results on rwjf.org



Robert Wood Johnson Foundation

The screenshot shows a search results page on rwjf.org. At the top, it indicates 'You are now viewing 1 - 10 of 1660 results'. Below this, there are several article snippets. On the right side, there is a 'Refine Your Results' sidebar. The sidebar has a 'Clear all' button at the top right. Below that, it shows the current filter: 'Topic: Nurses and Nursing'. There is a search box for keywords and a 'GO' button. The sidebar is divided into three sections: 'BY TOPIC', 'BY CONTENT', and 'BY DEMOGRAPHICS'. Each section has a list of categories with counts. 'BY TOPIC' includes: Health Leadership, Education, and Training (353); Health Care Quality (251); Disease Prevention and Health Promotion (48); Public and Community Health (38); and Social Determinants of Health (31). 'BY CONTENT' includes: Content Type. 'BY DEMOGRAPHICS' includes: Age; Gender; Race/Ethnicity; Location; and States and Territories. There are plus signs next to the 'BY CONTENT' and 'BY DEMOGRAPHICS' headers, indicating they can be expanded. There are also 'VIEW MORE' links at the end of the 'BY TOPIC' and 'BY DEMOGRAPHICS' lists.

- ◆ By Topic
- ◆ By Content Type
- ◆ By Age
- ◆ By Gender
- ◆ By Race/Ethnicity
- ◆ By Location
- ◆ By States and Territories



Filtering search results (a.k.a. faceted navigation)

- ❖ Invites end users to **refine** their search results without typing in a new query, e.g.,
 - Expose contextually relevant metadata.
 - Show how many matching hits by attribute.
 - Allow removal and addition of filters.
- ❖ Allows user to **explore** collection of search results
 - Drill down or up, by applying or removing filters.



Online shopping vs. content collections

- ❖ Using faceted navigation in online shopping has become a commonly understood metaphor.
- ❖ But applying multi-dimensional classification and faceted navigation to content collections is not as common.





What are use cases?

- ❖ Understanding how an organization or individual might actually use or wish to use organized information.
- ❖ A “use case” explores various scenarios with multiple stakeholders.
 - Formal and informal interviews.
 - Analysis of quantitative data on content use and search behavior.
 - Learning about organizational goals and expectations.
 - Identifying potential activities or likely uses for a set of organized information.
- ❖ Use cases facilitate the development of a specialized taxonomy
 - Describes activities and uses, or contexts important for particular applications in particular settings.

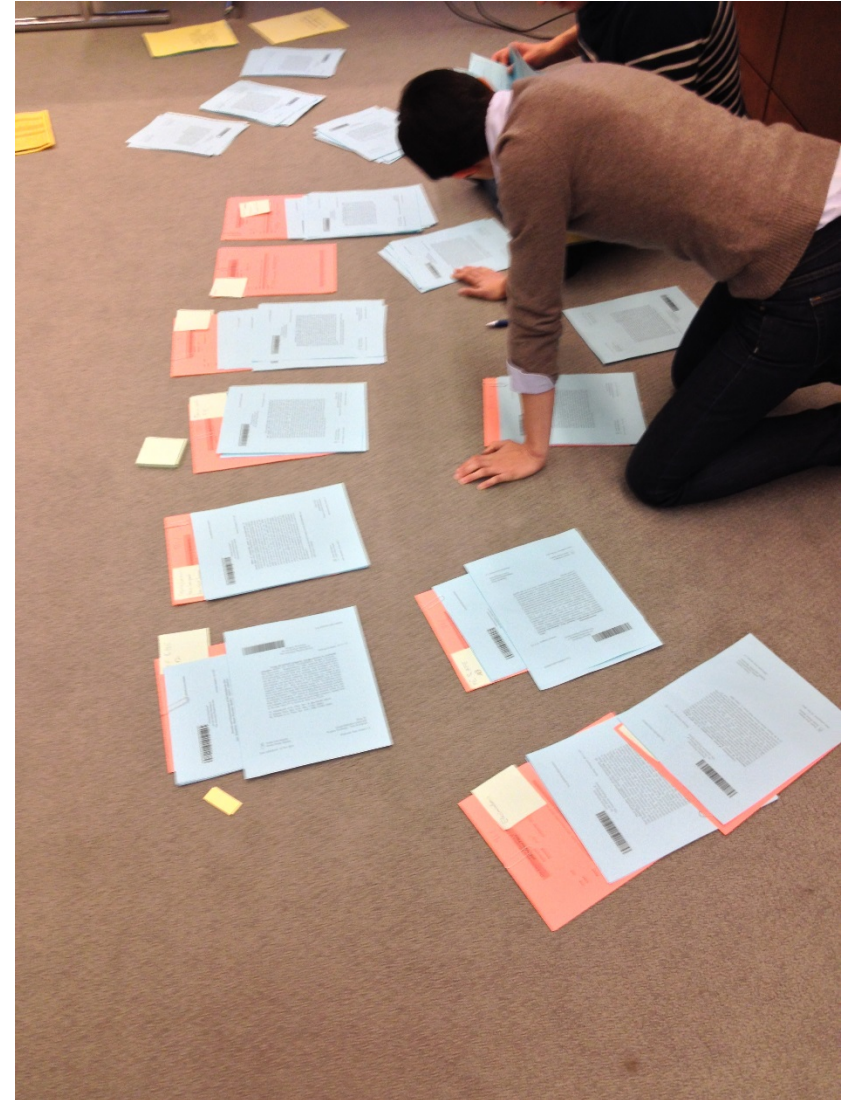
The most common classification use case is ordering a set of related content items in a search results.



APS Taxonomy use cases are not focused on search

- ❖ Organizing & facilitating editorial & publishing process.
 - Selection of taxonomy terms (indexing) for articles.
 - Authors assigning topics to their submissions.
 - Defining areas of responsibility and interest for editors.
 - Assigning articles to APS editors.
 - Referees describing their areas of expertise.
 - Selecting referees to review articles.
 - Assigning articles to journal sections.
 - Generating statistical reports and lists of articles by various subject criteria.

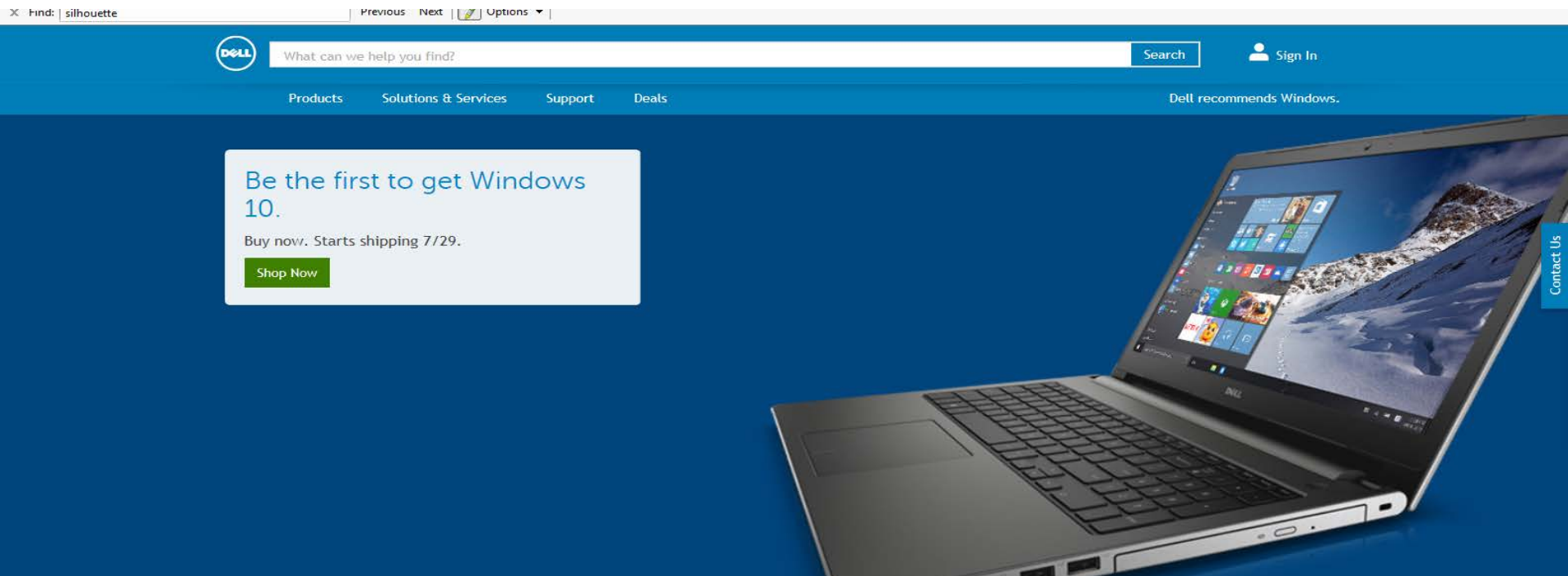
Grouping papers by topic and so that attendees can get from one session is another use case



APS sorting meeting (Dec 5, 2013)

The “big use case”: Overall strategy for locating products so consumers can buy them

- ❖ How to facilitate the identification and linking of a large and changing collection of content items with a large and changing assortment of related products.



Break down of specific tactics by IA method

Use Case	Contextual Navigation	Site Architecture	Synonyms	Import Files
Improve Google search.			X	
Consistent experience across sites.		X		
Consistent terminology.	X		X	
Use technology to pivot between service & product.	X			
Associate learn content w/ specific products.	X			
Move from learn to product content.	X			
Provide context w/in industry solutions.	X			
Consistent solutions & best practices experiences.			X	X
Consolidate community content.	X	X		X
Unify support & community content.	X	X		
Integrate product support w/ product details.	X			
Surface software & peripherals info.				X
Surface parts & accessories w/ products.	X			X
De-segmentation.		X		
Integrate external content.	X			X



Importance of facets and relationships

- ❖ Facetted classifications deconstruct complex concepts into a grammar

- Statements of named entities modified by types and topics, e.g.

Named Entity	Type or Topic
PowerVault NX400	Network Attached Storage
Barak Obama	U.S. President
Fender Stratocaster	Electric Guitar

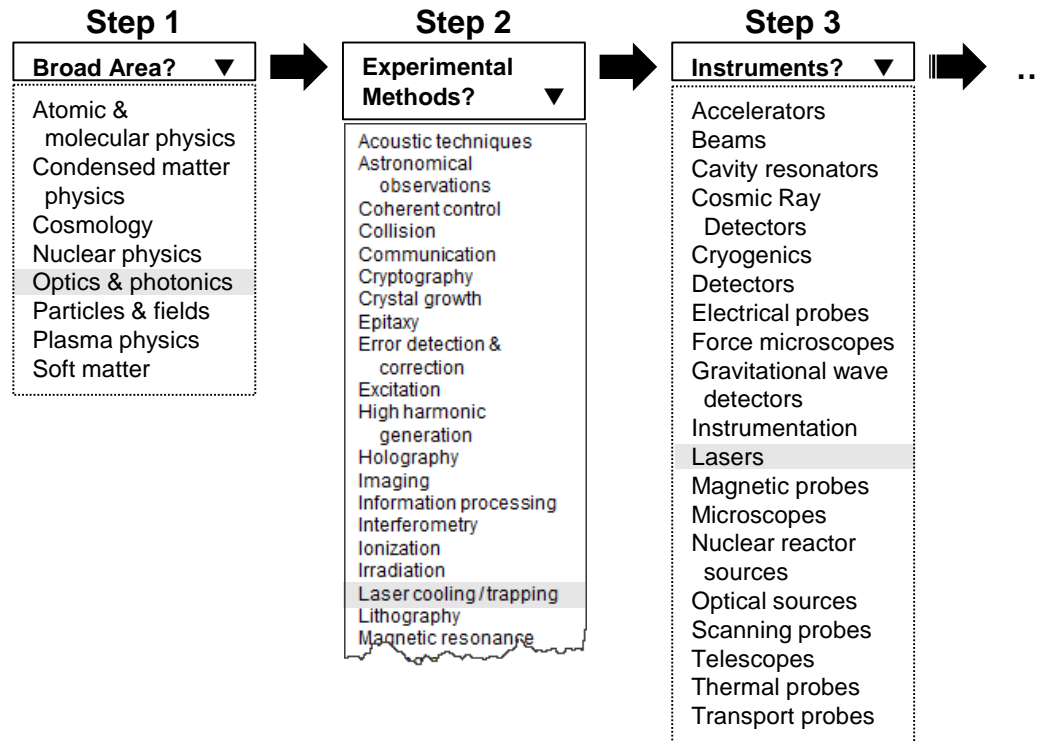
- ❖ Key semantic relationships commonly manifested in web classifications

- Equivalent (synonyms)
- Hierarchical (broader/narrower)
- Associative (related)

Facetted taxonomy example

❖ APS Taxonomy breaks up a complex categorization task into smaller chunks

- What you're studying → Broad area, materials & systems
- Why you're studying it → Phenomena & properties
- How you go about studying it → Apparatus, theory & techniques





9 Common taxonomy facets

Facet	Definition	Example Source
Content Type	Types of content created, managed and used to record or communicate information.	AGLS Document Type, AAT Information Forms , Records management policy, etc.
Audience	Subset of constituents to whom a content item is directed or intended to be used.	Market segments, Educational stages/grade levels, etc.
People	Names of important people such as authors, politicians, leaders, actors, etc.	LC Name Authority File, NY Times Topics-People, etc.
Organization	Names of organizations, their aliases and the relationships between them.	LC Name Authority File, NY Times Topics-Organizations, etc.
Industry	Broad market categories such as industry sector codes.	North American Industry Classification System, ISIC, etc.
Location	Names of places of operations, activities, constituencies, etc.	ISO 3166, Geonames, NY Times Topics-Places, etc.
Function	Activities and processes performed to accomplish goals.	Federal Enterprise Architecture Business Reference Model, AAT Functions, etc.
Product	Names of products and services that are produced by an organization or people.	Household Products Database, UNSPSC, etc.
Topic	Topical subjects and themes that are not included in other facets.	LC Subject Headings, NY Times Topics-Subjects, etc.



Facetted taxonomy example for business content

- ❖ Describe the various attributes of a content item to scope its context:
 - What type is it?
 - Who was it created for?
 - What business activity is it related to?
 - What people, organizations and/or products is it about?
 - Is it related to particular location?
 - Is it related to a particular industry sector or market?
 - etc.

Facetted classification is more like filling in the attributes of a product, than choosing the single most important aspect of a content item.

Describing named entities: Schema.org

- ❖ Identifies named entities in web published content.
- ❖ Rich description that enables Google Knowledge Graph.

People
Organizations

The screenshot shows a Google Knowledge Graph entry for the University of London. At the top left is the university's crest. To its right is a map of London with a red pin marking the university's location. Below the map is the title "University of London" with a star icon. Underneath the title is the text "University in London, England" and two buttons: "Directions" and "Write a review". A paragraph of text describes the university as a collegiate research university with 18 constituent colleges, 10 research institutes, and central bodies, with a link to Wikipedia. Below this is the address: "Senate House, Malet St, London WC1E 7HU, United Kingdom". Other details include "Founder: Jeremy Bentham", "Founded: 1836", "Enrollment: 174,000 (2014)", "Phone: +44 20 7862 8000", and "Colors: White, Blue, Red". A section titled "Notable alumni" features five circular portraits: Mahatma Gandhi, Lucian Freud, Chris Martin, Alexander Graham Bell, and John Stuart Mill. Below this is a section titled "People also search for" with five icons and labels for Imperial College London, City University London, University of Westmins... London, London Metropolitan... University, and University of Cambridge. Red dashed lines connect the "People" label to the founder and alumni sections. Green dashed lines connect the "Organizations" label to the university title, the "People also search for" section, and the map area.



Semantic relationships that can be used to improve information retrieval in search engines

Relationship	Description
Equivalent Label	An alternative label for the Concept such as a synonym or quasi-synonym.
Hierarchical Concept	Hierarchical link between two Concepts where one Concept is more general or more specific than the other.
Related Concept	Link between two Concepts where the two are inherently "related", but that one is not in any way more general than the other.



The equivalence problem

- ❖ An “exact match” is a high bar, yet some applications such as medical diagnostics require an exact match.
- ❖ Matching codes, acronyms and abbreviations to descriptive labels can be very helpful in improving search results.
- ❖ Sometimes it can be effective to roll-up narrower concepts into a broader category
 - E.g., when there is not a critical mass of content for narrower concepts.



The value of name authority files

Union List of Artist Names - Netscape

File Edit View Go Communicator Help

New Search


GETTY VOCABULARY PROGRAM

calendar of events
return home
Research Institute for the Humanities
Getty Research Institute

ULAN Name Record

[16023]

Rembrandt Harmensz van Rijn [PR,VP]	
(Dutch artist, 1606-1669)	[RV,WC,WI]
(Dutch painter and printmaker, 1606-1669)	[GC]
(Dutch painter, 1606-1669)	[PR]
(Dutch painter, draftsman and etcher, 1606-1669)	[VP]
(Dutch painter, printmaker, 1606-1669)	[BA]
(Dutch, 1606-1669)	[JG]
After Rembrandt	[PR]
from Rembrandt	[PR]
Paul Rembran	[PR]
Paul Rembrandt	[PR]
Rambrandt	[PR]
Rebranch	[PR]
Reimbrant	[PR]
Rem.	[PR]
Rembrach'	[PR]
Rembradt	[PR]
Rembrand	[PR]
Rembrande	[PR]
Rembrands	[PR]
Rembrandt	[GC,PR]
Rembrandt Harmenszoon van Rijn	[RV,PR]
Rembrandt Harmensz Van Rijn	[WI]
Rembrandt Harmensz. van Rijn	[BA,GC,PR]
Rembrandt Harmensz. van Rijn or Rhijn	[WC]
Rembrandt Hermanszoon van Rijn	[PR]
Rembrandt Hermansz van Rijn	[PR]
Rembrandt Olandese	[PR]
Rembrandt (school of)	[PR]
Rembrandt Van Rhyn	[PR]
Rembrandt van Rijn	[GC,JG,PR]




Document: Done

Common types of hierarchy

Whole-part

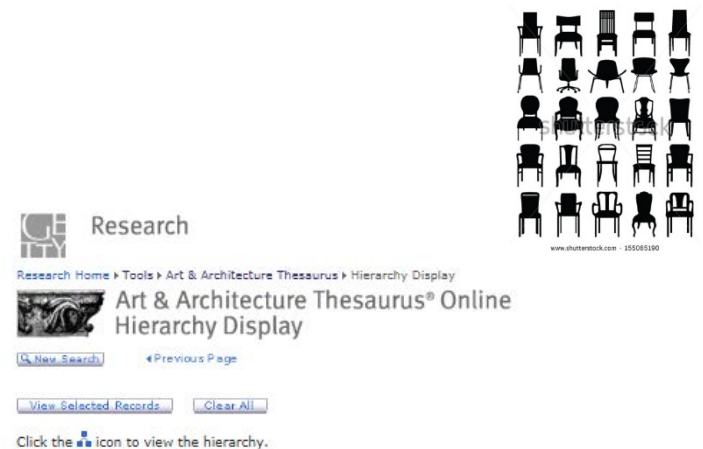


Research
 Research Home ▶ Tools ▶ Thesaurus of Geographic Names ▶ Hierarchy Display
 Getty Thesaurus of Geographic Names® Or Hierarchy Display
 [New Search] [Previous Page]
 [View Selected Records] [Clear All]


Click the  icon to view the hierarchy.
 Check the boxes to view multiple records at once.

-  Top of the TGN hierarchy (hierarchy root)
-  World (facet)
-  North and Central America (continent)
-  Canada (nation)
-  [[view physical features](#)]
-  Alberta (province)
-  British Columbia (province)
-  <lost & found/Canada> (miscellaneous)
-  Manitoba (province)
-  New Brunswick (province)
-  Newfoundland and Labrador (province)
-  Northwest Territories (territory)
-  Nova Scotia (province)
-  Nunavut (territory)
-  Ontario (province)
-  Prairies (general region)
-  Prairies, Lake of the (reservoir)
-  Prince Edward Island (province)
-  Quebec (province)
-  Saskatchewan (province)
-  Trans-Canada Highway (highway)
-  Ungava district (general region)
-  Yukon Territory (territory)

























Classified Lists



Research
 Research Home ▶ Tools ▶ Art & Architecture Thesaurus ▶ Hierarchy Display
 Art & Architecture Thesaurus® Online Hierarchy Display
 [New Search] [Previous Page]
 [View Selected Records] [Clear All]

Click the  icon to view the hierarchy.

Check the boxes to view multiple records at once.

-  Top of the AAT hierarchies
-  Objects Facet
-  Furnishings and Equipment (Hierarchy Name)
-  Furnishings (Hierarchy Name)
-  furnishings (artifacts)
-  Costume (Hierarchy Name)
-  costume (mode of fashion)
-  Tools and Equipment (Hierarchy Name)
-  apparatuses
-  equipment
-  tools [N]
-  Weapons and Ammunition (Hierarchy Name)
-  weapons
-  ammunition
-  Measuring Devices (Hierarchy Name)
-  measuring devices (instruments)
-  Containers (Hierarchy Name)
-  containers (receptacles)
-  Sound Devices (Hierarchy Name)
-  sound devices (equipment)
-  Recreational Artifacts (Hierarchy Name)
-  recreational artifacts (equipment)
-  Transportation Vehicles (Hierarchy Name)
-  vehicles (transportation)

<http://www.getty.edu/research/tools/vocabularies/>



Related term guidelines are complicated

In Same Hierarchy

- ❖ Overlapping sibling terms.
- ❖ Mutually exclusive sibling terms.
- ❖ Derivational relationships.

In Different Hierarchies


- ❖ Disciplines or fields of study and objects or phenomenon studied or practitioners.
- ❖ Operations or processes and their agents or instruments.

- ❖ Objects or processes and their counteragents.
- ❖ Actions and their products.
- ❖ Actions and their targets.
- ❖ Objects or substances and their unique properties.
- ❖ Concepts linked by causal dependence.
- ❖ Concepts and their units or mechanisms of measurement.
- ❖ Phrases in which the noun is not a true broader term.

NISO Guidelines for the Construction, Format, and Management of Monolingual Thesauri. ANSI/NISO Z39.19 – 2003. “Section 5. The Associative Relationship”, pp. 19-20

Curated hyperlinks: Google knowledge graph

- ❖ Images
- ❖ Wikipedia snippet
- ❖ Birth & death dates & locations
- ❖ Full name
- ❖ Nationality
- ❖ Periods
- ❖ Works
- ❖ Related searches



More images

Rembrandt

Painter

Rembrandt Harmenszoon van Rijn was a Dutch painter and etcher. He is generally considered one of the greatest painters and printmakers in European art and the most important in Dutch history. [Wikipedia](#)

Born: July 15, 1606, [Leiden, Netherlands](#)

Died: October 4, 1669, [Amsterdam, Netherlands](#)






Full name: Rembrant Harmenszoon van Rijn

Nationality: Dutch

Periods: Baroque, Dutch Golden Age painting, Dutch Golden Age






Artwork

[View 15+ more](#)

 The Night Watch 1642	 The Anatomy Lesson of... 1632	 The Storm on the Sea of Galilee 1633	 The Return of the Prodigal...	 The Jewish Bride 1669
--	---	--	--	---

People also search for

[View 15+ more](#)

 Vincent van Gogh	 Peter Paul Rubens	 Johannes Vermeer	 Caravagg...	 Claude Monet
---	--	---	--	---

Information retrieval has been the key enabling technology for the Web

- ❖ We are surrounded by named entities and clues about the relationships between and among them.
- ❖ Underpinning web search is the analysis and annotation of the relationships between entities.
- ❖ One goal of schema.org is to uniquely identify named entities URI's (Universal Resource Identifiers such as URL's).
- ❖ URI's are similar to traditional classifications whose codes aim to provide unique locations for physical objects.
- ❖ Perhaps we have come full circle from universal classifications systems to systems of universal unique named identifiers.



Crowd-sourced classification



Joseph Busch, jbusch@taxonomystrategies.com, @joebusch

QUESTIONS?