

# Indexes, Search, and Taxonomies: Paths to Findability

by  
Heather Hedden

STC India

Bengaluru Learning Session: 13 January 2025

Noida Learning Session: 16 January 2025



# About Heather Hedden

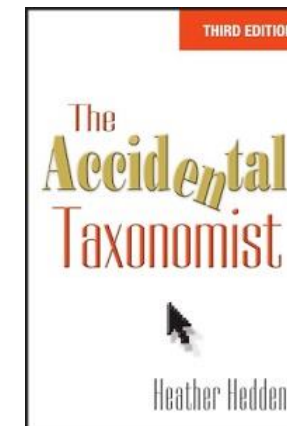
---

- Independent taxonomy consultant, Hedden Information Management
- Instructor of online and corporate taxonomy courses and workshops
- Previously a taxonomy consultant in consulting firms, Enterprise Knowledge and PPC. Also, a contract consultant for others.
- Former taxonomy-related roles at Semantic Web Company, Gale/Cengage, Viziant, and First Wind.
- Author of *The Accidental Taxonomist*, 3<sup>rd</sup> ed. (2022, Information Today, Inc.)



**Hedden Information  
Management**

Making information findable



# Outline

---

- Background and definitions
- Indexes
- Controlled vocabularies and thesauri
- Taxonomies
- Search in combination with thesauri or taxonomies
- Taxonomy uses and applications
- Information findability and AI

# Outline

---

- Background and definitions
- Indexes
- Controlled vocabularies and thesauri
- Taxonomies
- Search in combination with thesauri or taxonomies
- Taxonomy uses and applications
- Information findability and AI

# Introduction to Information Findability

---

- Challenge: Make desired targeted information findable to users
- Users of documentation don't read it from start to finish.
- They look for content units or sections it to get answers to their questions, from among:
  - sections within narrative descriptions
  - step-by-step how-to explanations
  - lists or tables of information
  - images/diagrams/screenshots/video clips
  - glossary definitions
- So, they look up what they consider their topic to be called, in:
  - alphabetical indexes or lists
  - hierarchical topic trees
  - search boxes



# Introduction to Information Findability

---

## Definitions

- **Index** – browsable, alphabetical list of terms, with cross-references, which are linked to or point to content
- **Controlled vocabulary** – an authoritative list of terms approved for indexing or tagging to support consistency and comprehensive retrieval
- **Thesaurus** – a kind of controlled vocabulary with standard relationships between terms, for improved use of a large controlled vocabulary
- **Taxonomy** – a kind of controlled vocabulary arranged into a hierarchy or categories to support guided browsing for topics
  - As a topical tree
  - As facets for filters, based on types/aspects

# Introduction to Information Findability

---

## Definitions (continued)

- **Indexing** – associating terms with the content they describe and gathering the terms in an index for the user to browse or search.  
Can be for books or articles.
- **Tagging** – associating terms (tags) with content they describe, but the tags don't have to be arranged in an index, and they don't have to be from a controlled vocabulary.  
For articles, digital assets, content components (e.g. DITA Topics)

# Outline

---

- Background and definitions
- **Indexes**
- Controlled vocabularies and thesauri
- Taxonomies
- Search in combination with thesauri or taxonomies
- Taxonomy uses and applications
- Information findability and AI



# Indexes

---

## Index definitions:

- “A systematic arrangement of entries designed to enable users to locate information in a document.”  
--*British indexing standard (BS3700:1988)*

- “A systematic guide to facilitate retrieval of content.

Indexes include the following major components:

- (a) headings representing topics or features;
- (b) cross-references or other linking devices among synonymous, equivalent, broader, narrower, and other related terms;
- (c) locators.

An index serves as an information retrieval aid to the text or documents it accompanies.”

-- *Criteria for Indexes (ANSI/NISO Z39.4-2021) standard*

- Locator: The part of an index entry that indicates where the information is found.
- A system for “pointing” to information
  - The “index finger” is for pointing.
- Traditionally is displayed alphabetically, but this is not a requirement.

A→Z



## A

Absolute extrema *See also* Optimization  
critical-point theorem, 705  
finding with graphing calculator, 704  
on other intervals, 705–710  
Absolute maximum on the interval, 703  
Absolute minimum on the interval, 703  
Absolute value, 8  
Absolute-value equations, 46–47  
Absolute-value function, 125, 613  
Absolute-value inequalities, 98–100  
Acceleration, 694  
Addition  
of matrices, 299–301  
in order of operations, 4  
of polynomials, 13–14  
of rational expressions, 27–29  
Addition rule  
for counting, 418  
for disjoint events, 434  
for probability, 432  
for rational expressions, 27  
Additive identity, 3  
Additive inverse, 3, 303  
Adjacency matrix, 334  
Amortization  
defined, 256  
loans and, 256–258  
payments, 256  
remaining balance, 258  
schedules, 258–259  
Annual percentage yield (APY), 238–239  
Annuities  
computations, 244–245  
defined, 244  
future value of, 244, 245  
ordinary, 244–251  
payment period, 244

Architectural arches case study, 181–182

## Area

approximations, 765  
continuous functions and, 779  
between curves, 786–790  
finding, 778  
formula, 765  
fundamental theorem of calculus and, 776–779  
between graph and  $x$ -axis, 776–777  
under normal curves, 547, 550–551  
rectangles, 764–766  
region, 765  
total under a curve, 766  
Arithmetic mean, 528  
Associative properties, 2  
Asymptotes  
horizontal, 169–172, 587  
slant, 736  
vertical, 169, 171  
Augmented matrix, 279  
Average cost, 635–637  
Average rate of change, 594  
Average speed, 594, 596  
Average velocity, 597  
Axis of the parabola, 146

## B

Back substitution, 278  
Base  $a$  logarithms, 201–202  
Base  $e$  logarithms, 201  
Bases  
defined, 11  
exponential equations, 212  
exponential functions, 183  
negative, 36  
Basic feasible solutions  
defined, 367, 389

## Bonds

corporate, 226  
coupon rate, 255  
interest rates, 255  
maturity, 226  
as two-part investments, 256  
zero-coupon, 236–237  
Boundary lines, 338, 339  
Bounded feasibility region, 347  
Bounded population growth case study, 812–813  
Boxplots  
defined, 553  
first quartile, 553  
five-number summary, 553  
graphing calculators and, 554  
second quartile, 553  
third quartile, 553  
uses of, 554  
Break-even analysis, 138–139  
Break-even point, 138–139

## C

Calculators *See* Graphing calculators  
Cancellation property, 25  
Carbon dating, 214  
Cartesian coordinate system, 65–66  
Case studies  
architectural arches, 181–182  
bounded population growth, 812–813  
consumers and math, 63–64  
continuous compounding, 267–268  
cooking with linear programming, 403–404  
extrapolation and interpolation for prediction, 112–113  
gapminder.org, 221–223  
global warming, 856–857  
medical diagnosis, 463  
price elasticity of demand, 675–676

Index in print at  
the back of a book

## Index

A-B testing, [349–350](#)  
abbreviations in printed thesauri, [127–128](#)  
ABI/INFORM, [39](#)  
abilities and skills of taxonomists, [66–71](#)  
academic institutions, [77](#)  
Access Innovations  
    Data Harmony software, [42](#), [179–180](#), [240](#)  
    TaxoBank, [177](#), [455](#)  
accidental taxonomists, [59–61](#)  
accuracy of manual tagging, [201](#)  
acronyms, [110](#), [114](#)  
adaptability of taxonomists, [69–70](#)  
administrative attributes of concepts, [174–175](#)  
administrative metadata, [19](#)  
advanced (fielded) search, [270](#), [309](#), [313–316](#)  
AGROVOC Thesaurus, [187](#), [387](#)  
AIIM, [452](#), [457](#)  
Airtable, [157](#)  
a.k.a. software, [166](#)  
all/some rule, [129–130](#), [137](#)  
alphabetical displays, [289–294](#), [308](#)  
alternative labels  
    for auto-categorization, [252–253](#)  
    in controlled vocabularies, [5](#)  
    customized, [149–150](#)  
    display of, [318–320](#)  
    for manual tagging, [205–207](#)  
    in merging concepts, [367](#), [375](#)  
    number of, [117–120](#)  
    purpose and format, [110–112](#)  
    vs. synonyms, [111](#)  
    in thesauri, [10–11](#), [14](#)  
    types of, [112–116](#)  
    updating taxonomies and, [367](#), [369](#)

## A – Z Index

A–Z Index (\* denotes members-area pages)

[A](#) | [B](#) | [C](#) | [D](#) | [E](#) | [F](#) | [G](#) | [H](#) | [I](#) | [J](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [P](#) | [Q](#) | [R](#) | [S](#) | [T](#) | [U](#) | [V](#) | [W](#) | [X](#) | [Y](#) | [Z](#)

### A

AAE (Association of Art Editors)  
AAP (Association of American Publishers)  
aboutness, [On Aboutness: Determining What It's All About \(webinar\)](#)  
abstracting, professional organization for  
Academic Info Subject Guides  
*Accidental Indexer, The* (Badgett)  
Accounting, Business Studies, and Economics Dictionary  
Acquisitions Editor  
Acronym Directory  
address, ASI  
ADE (Association for Documentary Editing)  
[administrative office, ASI](#)  
Adobe Acrobat Reader, PDFs in the Indexing Process (webinar)  
advertising rates, *Key Words*  
AEJMC (Association for Education in Journalism and Mass Communication)  
AESE (Association of Earth Science Editors)  
African names, indexing  
a.k.a. classification software  
Alcohol Industry and Policy Database  
aliaINDEXERS (discussion group)  
AllPages.com (business directory)  
AllWords dictionaries  
amazon.com  
Ament, Kurt  
American Association of University Presses  
American Indian terminology resources  
[American Library Association](#)  
    annual meeting  
    contact information  
American Medical Writers Association (AMWA)  
*American National Standard Guidelines for the Construction, Format, and Management of*

Book-style  
indexes online

Contents [Index](#) Search x

inde

merge import  
Help Explorer  
Help Express  
Help Wizard Dialog  
HH Project Editor  
HTML Help 1.x  
    Batch Mode  
    Creating HTML Help  
    Examine Chm Dialog  
    Help Express  
    Help Wizard Dialog  
    HH Project Editor  
    TOC/Index Editor  
HTML Help 2.x  
    Batch Mode  
    Creating HTML Help  
    Examine Chm Dialog  
    Filters Dialog  
    H2 Primer  
    H2 Project Editor  
    H2 Registration Dialog  
    H2 Viewer Dialog  
    H2Reg Import  
    Help Express  
    Help Wizard Dialog  
    New File Wizard  
    TOC/Index Editor  
    Utilities  
HxF Editor  
HxM Project Editor  
Index Editor Notes



# Indexes for Articles

Expert • Independent • Nonprofit

ConsumerReports.org®

Home Customer service My account

Autos Appliances Electronics & computers Home & garden Health & fitness Personal finance Babies

A to Z Index Search Consumer protection Donate Recalls Web site e-Ratings

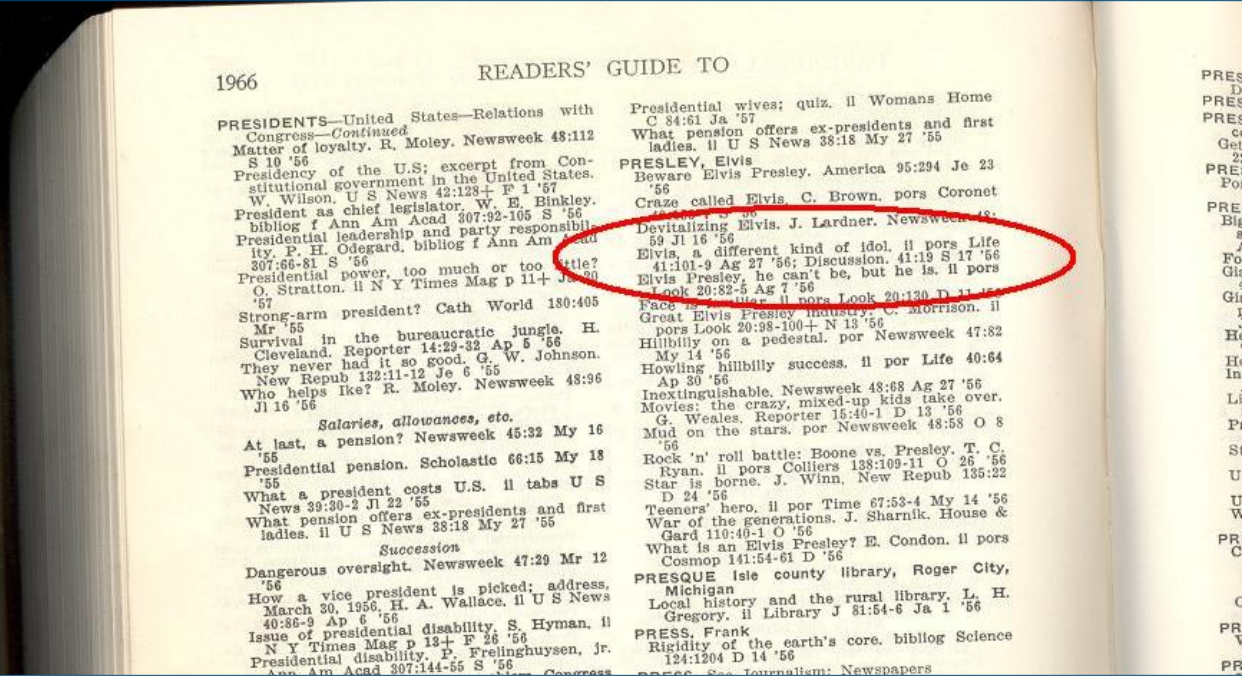
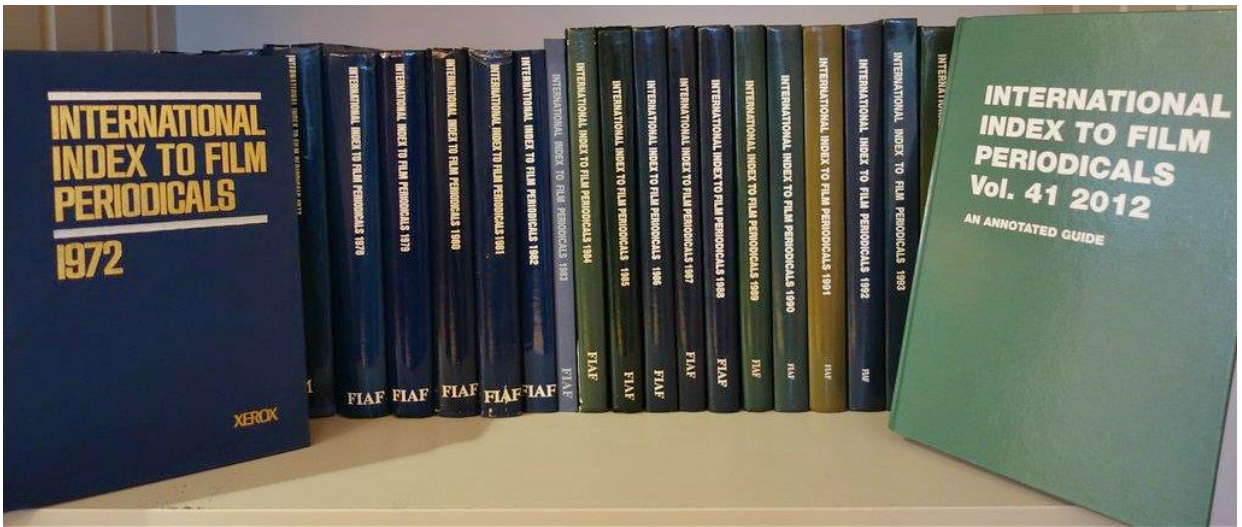


The following index includes all reports available on **ConsumerReports.org**. Reports listed with a month and year indicate either the *Consumer Reports* magazine publication date or when the content was posted and/or last updated on our site. A \* indicates a product that is part of our continual-testing program. Use "Ctrl F" to search for a specific word in the index. Don't see a particular product report listed here? Our [Index of previous reports](#) lists articles that are no longer available on our site. You'll also find information on how to obtain copies of these articles via fax or mail.

**A B C D E F G H I J K L M N O P Q R S T U V W X Y Z**

- Abdominal exercises** 3/02
- Abdominal fat and disease risk** 8/03
- Abortion** (FREE) 2/05
- AC power supplies** 1/05
- Accounting, automobile** (FREE) 2/03
- Accounting reform laws** (FREE) 3/03
- Adhesive bandages** (Ratings) 8/02
- Adjustable rate mortgages** (FREE) 10/03
- Advertising**
  - [drug, federal regulation of](#) (FREE) 2/03
  - [misleading](#) (FREE) 2/02
- Agents**
  - [insurance, fraudulent deals from](#) (FREE) 8/04
  - [travel](#) 6/03
  - [airfare and](#) (FREE) 7/01
- Air cleaners** (Ratings) 10/03
- Air conditioners**
  - [central](#) 7/03
  - [room](#) (buying advice and Ratings) \*
- Air mattresses** (Ratings) 11/04
- Air pollution, indoor** (FREE) 2/02
- Air travel**

- Automobiles continued**
  - [Federal rollover ratings](#) (FREE) 4/04
  - [fifty-year auto retrospective](#) (FREE) 4/03
  - [First Look: Ford Escape Hybrid](#) 2/05
  - [Ford Escape Hybrid: First Look](#) 2/05
  - [fuel economy](#) (FREE) 8/04
    - [cargo capacity and](#) 10/04
    - [corporate average fuel economy \(CAFÉ\) standards](#) 5/04
    - [vs. performance](#) 8/04
  - [fuel efficiency](#) (FREE) 8/04
  - [fuel-efficient](#) 7/03
  - [fuel-saving tips](#) 7/04
  - fuels**
    - [alternative](#) 10/04
    - [emergency](#) 3/04
  - [gas-saving tips](#) 7/04
  - [gasoline, premium](#) 9/04
  - [gasoline substitutes, emergency](#) 3/04
  - [headlights, HID](#) (FREE) 4/03
  - [HID headlights](#) (FREE) 4/03
  - [hoses and belts, inspecting](#) (FREE) 3/04
  - [IIHS crash testing](#) (FREE) 4/03
  - [incentives & dealer discounts](#) 4/04



# Indexes

---

## Back-of-the-book indexing

- Continues to be done manually, mostly offline, for print first
- Indexes are converted to hyperlinks in ebooks by only a minority of publishers
- Embedded indexing, whereby index terms are linked to anchors in the text, accounts for a small but growing number of book indexes.
  - Useful for books that are updated often for new editions.

## Article indexing

- Is done both manually and automatically, but now digitally/online only
- User display is not an alphabetical index anymore.
  - Now search suggest, topic tree browse, facets for filtering
- No longer creating an “index,” it is less commonly called “indexing,” but rather “tagging”
- Same method and principles for tagging content/pages in a CMS

# Index Types

---

## Book indexes

- “closed” indexing – indexed and completed
- For books, manuals, long documents (100+ pages)
- Index terms point to page numbers or ranges, and the user must scan the page to find the topic of interest.
- Typically or traditionally done in print but can be created for ebooks, too with same presentation format, but hyperlinked.
- Index terms are derived from the text by the indexer.
- Index created by a single indexer
- Manually indexed

## Article indexes

- “open” indexing – continually indexing new content (articles) that get added
- For large numbers of short documents (0.25 - 10 pages)
- Index terms link to the entire document, file, web page, CMS page, blog post, PDF files, Word/Office file
- No longer created for print, but online only
- Index terms come from an approved controlled vocabulary to provide consistency in indexing documents from different authors and indexed by different indexers.
- Often indexed by multiple indexers.
- Manually or automatically indexed

# Index Types

---

## For technical documentation

- What about online help documentation?
  - Is that closed like a book/manual, or open like a series of articles?
  - Originally indexed like books/closed indexing; now more like articles/open indexing
- What about intermediate-length documents 10-100 pages long?
  - Best if indexed by units/sections
- What about sections of documents?
  - Components of documents as managed in [component content management systems](#)
  - Then it's a form of "open indexing" or tagging
- Open indexing/tagging requires a [controlled vocabulary](#) for *consistency* and better search results.

# Outline

---

- Background and definitions
- Indexes
- **Controlled vocabularies and thesauri**
- Taxonomies
- Search in combination with thesauri or taxonomies
- Taxonomy uses and applications
- Information findability and AI



# Controlled Vocabularies & Thesauri

---

## Controlled vocabulary

- An authoritative, restricted list of terms (words or phrases) mainly used for indexing/tagging content to support retrieval.
- Controlled in who and when new terms can be added.
- Usually makes use of synonym terms/alternative labels to point to the correct, preferred terms.
- “A list of terms that have been enumerated explicitly. This list is controlled by and is available from a controlled vocabulary registration authority. All terms in a controlled vocabulary must have an unambiguous, non-redundant definition.”  
-- *Guidelines for the Construction, Format, and Managed of Monolingual Controlled Vocabularies (ANSI/NISO Z39.19-2005) standard*

## Thesaurus

- A kind of structured controlled vocabulary that has synonyms/nonpreferred terms/alternative labels, hierarchical relationships, and associative (related term) relationships
- “A controlled vocabulary arranged in a known order and structured so that the various **relationships** among terms are displayed and clearly identified by standardized relationship indicators.”  
-- *ANSI/NISO Z39.19-2005 standard*

# Controlled Vocabularies & Thesauri

- Select All
- Argentina
- Australia
- Austria
- Bangladesh
- Barbados
- Belgium
- Bosnia and Herzegovina
- Brazil
- Canada
- Chile
- China
- Colombia
- Costa Rica
- Croatia
- Cuba
- Czech Republic
- Denmark
- Ecuador
- Egypt
- Estonia
- Ethiopia
- Finland
- France
- Germany
- Ghana
- Greece
- Hong Kong
- Hungary
- India

Country of publication

- Select All
- Danish
- English
- French
- German
- Italian
- Portuguese
- Spanish
- Turkish
- Ukrainian

Language

- Select All
- Almanac/Yearbook
- Atlas
- Audio
- Blog
- Chronology
- Collection
- Dictionary
- Directory
- Encyclopedia
- Factbook
- Handbook
- Images
- Magazine/Journal
- Monograph
- Newsletter
- Newspaper
- Newswire
- Pamphlet
- Report
- Textbook
- Transcript
- Video

Format

Simple controlled vocabulary flat lists.  
Often used as values for metadata properties.

Each term has been tagged to content.

# Thesauri

## Thesaurus features

- A controlled vocabulary that has standard structured relationships between terms
  - Hierarchical: broader term/narrower term (BT/NT)
  - Associative: related terms (RT)
  - Equivalence: preferred term (“use for” or “used for”)/ non-preferred term (use) (USE/UF)
- Created in accordance with standards:
  - [ISO 25964](#) (2011, 2013) *Thesauri and Interoperability with Other Vocabularies*
  - [ANSI/NISO Z39.19](#) (2005, renewed 2010) *Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies*  
[www.niso.org/publications/ansiniso-z3919-2005-r2010](http://www.niso.org/publications/ansiniso-z3919-2005-r2010)
  - “Thesaurus” is most often the kind of controlled vocabulary used in indexing periodical literature

### materials acquisitions

- UF acquisitions (of materials)  
library acquisitions
- BT collection development
- NT accessions  
approval plans  
gifts and exchanges  
materials claims  
materials orders  
subscriptions
- RT book vendors  
jobbers  
subscription agencies  
subscription cancellations

ASIS&T thesaurus

# Thesauri

## Article indexing with a thesaurus

- Across multiple documents or content items, where different words/terms describe the same concepts
- By multiple indexers working on the same collection
- By machine-aided indexing / autotclassification, where terms have rules, clues, or sample tuned documents.
- The thesaurus terms on the back end become the index on the front end.

Alpha Browsing - Internet Explorer provided by Cengage Learning

http://gpms.cengage.com/Per\_iMania/brAbrowse.jsp?termType=D&termValue=survey&browse=A&brc

**D - Subject Descriptors**

D educational v

all xRef  from xRef  no xRef

**Educational videos**

- BROADER TERM:** Audiovisual materials
- BROADER TERM:** Video recordings
- NARROWER TERM:** Instructional videos
- SEE ALSO:** Children's videos
- SEE ALSO:** Educational programs (Mass media)
- SEE ALSO:** Edutainment
- SEEN FROM:** Classroom videos
- SEEN FROM:** Video tapes in education

**Educational volunteers** -

- SEE:** Education volunteers

**Educational vouchers**

- SEE ALSO:** Education costs
- SEE ALSO:** Educational finance
- SEE ALSO:** Educational programs
- SEE ALSO:** Student financial aid
- SEE ALSO:** Bush v. Holmes 919 So. 2d 392 (Fla. 2006)
- SEEN FROM:** Education vouchers
- SEEN FROM:** School vouchers

**Educational works** -

- SEE:** Textbooks

**Educationally disadvantaged**

- BROADER TERM:** Disadvantaged persons

Navigation buttons: H, N, D, O, S, and up/down arrows.

# Thesaurus Benefits

---

## Synonym Control

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

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

Gathers synonyms, acronyms, variant spellings, etc.

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

# Thesauri

## History of thesauri for information retrieval

- Insufficiencies of library classification systems for technical and scientific research/articles, led to the creation of thesauri based on indexing terms
  - “Thesaurus” first used for a controlled vocabulary at IBM (1957)
  - E. I. Dupont de Nemours Company’s thesaurus (1959)
  - *Thesaurus of ASTIA Descriptors* (1960)
  - *Chemical Engineering Thesaurus* of the American Institute of Chemical Engineers (1961)
  - Professional organization publishers of scientific journals created their own thesauri in the 1960s
- Standard guidelines for thesauri were published:
  - *UNESCO Guidelines for the establishment and development of monolingual thesauri* (1970)
  - *DIN 1463 Guidelines for the establishment and development of monolingual thesauri* (1972)
  - *ISO 2788 Guidelines for the establishment and development of monolingual thesauri* (1974)
  - *ANSI American National Standard for Thesaurus Structure, Construction, and Use* (1974) (superseded since by ANSI/NISO Z39.19-1993)

pesticides	UF: fumigants BT: agrochemicals NT: fungicides herbicides insecticides pests RT: pests	plant products	NT: cereals fruits spices vegetables plants RT: plants
pests	NT: pest insects plant pests pesticides RT: pesticides	plants	RT: plant products
pigs	UF: hogs porkers sows BT: livestock	porkers	USE: pigs
plant pests	BT: pests	poultry	BT: livestock NT: chickens ducks geese turkeys RT: eggs
		sheep	BT: livestock RT: wool

# Thesauri

## Thesauri in user interfaces

- The same thesaurus may be displayed alphabetically or hierarchically
- Example: ERIC (Education Resources Information Center) Thesaurus  
<https://eric.ed.gov>
- Full display is suitable for scholars/researchers, not product users searching help docs

**ERIC** Collection Thesaurus

Search thesaurus descriptors  [Browse Thesaurus](#)

Include Synonyms  Include Dead terms

[Notes](#) [FAQ](#) [Contact Us](#)

### Purpose and Scope

The ERIC Thesaurus is a list of terms representing research topics in the field of education. Descriptors from the ERIC Thesaurus are assigned to every document in the ERIC digital library to describe its subject content.

Terms in the ERIC Thesaurus represent the vocabulary used in the documents that comprise the ERIC digital library collection. Learn more about how the Thesaurus is updated in this [archived webinar](#).

The ERIC Thesaurus contains a total of 11,875 terms. There are 4,578 descriptors and 7,165 synonyms. There are also 132 dead terms, which are no longer used as descriptors but remain in the Thesaurus to aid in searching older records. The ERIC Thesaurus was last updated in 2023. The updated Thesaurus includes 25 new descriptors, 32 new synonyms, and changes to 37 existing terms. For more information, view the [full list of updates](#) and [download the updated file](#).

### Browse Alphabetically

# A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

### Browse by Category

<a href="#">Agriculture and Natural Resources</a>	<a href="#">Facilities</a>	<a href="#">Mental Health</a>
<a href="#">Arts</a>	<a href="#">Government and Politics</a>	<a href="#">Occupations</a>
<a href="#">Bias and Equity</a>	<a href="#">Health and Safety</a>	<a href="#">Peoples and Cultures</a>
<a href="#">Business, Commerce, and Industry</a>	<a href="#">Human Geography</a>	<a href="#">Physical Education and Recreation</a>
<a href="#">Communications Media</a>	<a href="#">Humanities</a>	<a href="#">Publication/Document Types</a>
<a href="#">Counseling</a>	<a href="#">Individual Development and Characteristics</a>	<a href="#">Reading</a>
<a href="#">Curriculum Organization</a>	<a href="#">Individual in Social Context</a>	<a href="#">Research and Theory</a>
<a href="#">Disabilities</a>	<a href="#">Information/Communications Systems</a>	<a href="#">Science and Technology</a>
<a href="#">Economics and Finance</a>	<a href="#">Labor and Employment</a>	<a href="#">Social Problems</a>
<a href="#">Educational Levels, Degrees, and Organizations</a>	<a href="#">Language and Speech</a>	<a href="#">Social Processes and Structures</a>
<a href="#">Educational Process: Classroom Perspectives</a>	<a href="#">Languages</a>	<a href="#">Students, Teachers, School Personnel</a>
<a href="#">Educational Process: School Perspectives</a>	<a href="#">Learning and Perception</a>	<a href="#">Subjects of Instruction</a>
<a href="#">Educational Process: Societal Perspectives</a>	<a href="#">Mathematics</a>	<a href="#">Tests and Scales</a>
<a href="#">Equipment</a>	<a href="#">Measurement</a>	

# Outline

---

- Background and definitions
- Indexes
- Controlled vocabularies and thesauri
- **Taxonomies**
- Search in combination with thesauri or taxonomies
- Taxonomy uses and applications
- Information findability and AI



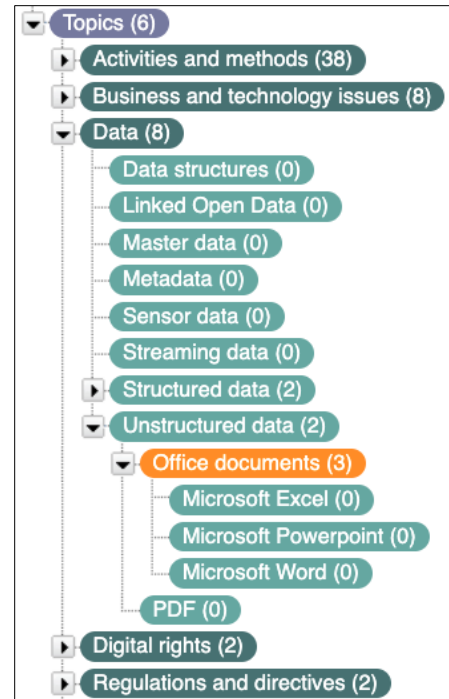
# Taxonomies: History

## Origins of modern information taxonomies

1. Classification systems
2. Subject heading schemes and thesauri

Dewey Decimal Classification	
<b>000 GENERAL WORKS</b> 020 Library and Information Science 030 General Encyclopedias 050 General Periodicals 060 General Organizations	<b>600 TECHNOLOGY</b> 610 Medical Sciences 620 Engineering 630 Agriculture 640 Domestic Sciences 650 Business and Management 660 Chemical Technology 670 Manufacturers 690 Building Construction
<b>100 PHILOSOPHY</b> 110 Metaphysics 130 Psychology, occultism 140 Philosophy 150 Psychology 160 Logic	<b>700 THE ARTS</b> 710 Landscape and Civic Art 720 Architecture 730 Sculpture, Plastics 740 Drawing, Decorative Arts 750 Painting 760 Prints and Print Making 770 Photography 780 Music 790 Recreation, Performing Arts
<b>200 RELIGION</b> 220 The Bible 230 Christian Doctrine 290 Comparative and Other Religions	<b>800 LITERATURE</b> 810 American Literature 820 English Literature 830 German Literature 840 French Literature 850 Italian, Rumanian Literature 860 Spanish, Portuguese Literature 870 Latin and Other Italic Literatures 880 Classical and Modern Greek Literature 890 Other Literature
<b>300 SOCIAL SCIENCES</b> 310 Statistics 320 Political Science 330 Economics 340 Law 350 Public Administration 360 Social Welfare 370 Education 380 Public Service 390 Customs and Folklore	<b>900 HISTORY, GEOGRAPHY</b> 910 Geography, Travel 920 Biography, Genealogy 930 Ancient History 940 Europe 950 Asia 960 Africa 970 North America 980 South America 990 Pacific Ocean Islands 991 Indonesia 993 New Zealand and Melanesia 994 Australia 995 New Guinea (Papua) 996 Polynesia 997 Atlantic Ocean Islands 998 Arctic Region 999 Antarctic Regions
<b>400 LANGUAGE</b> 410 Comparative Linguistics 420 English and Anglo Saxon 430 German Language 440 French 450 Italian, Rumanian 460 Spanish, Portuguese 470 Latin and Other Italic Languages 480 Classical and Modern Greek 490 Other Languages	
<b>500 SCIENCE</b> 510 Mathematics 520 Astronomy 530 Physics 540 Chemistry 550 Earth Sciences 560 Paleontology 570 Life Sciences 580 Botanical Sciences 590 Zoological Sciences	

*classification system*



*taxonomy*

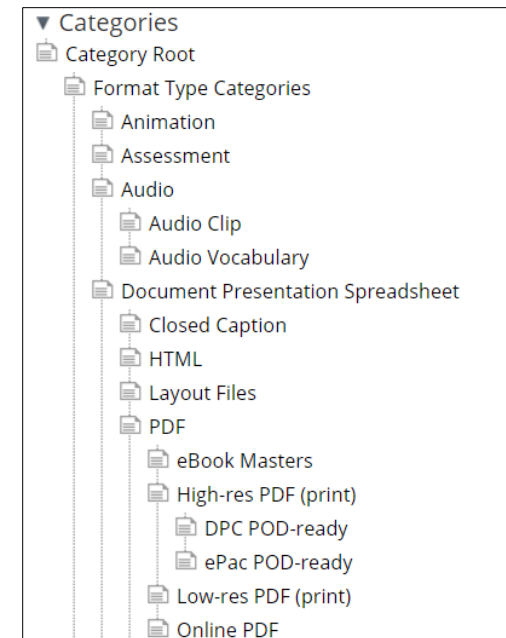
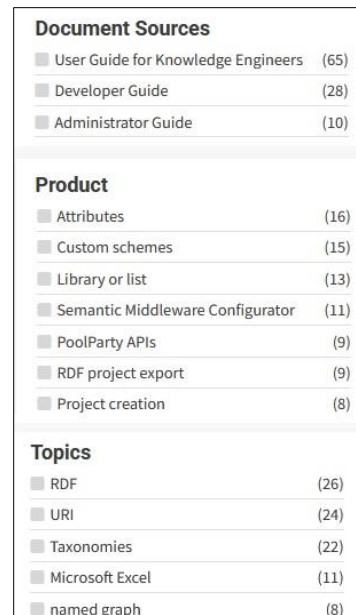
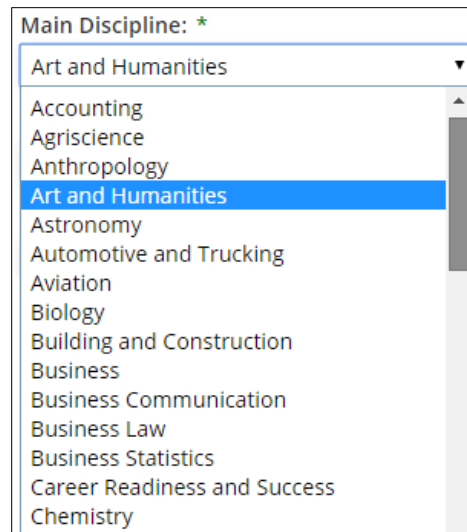
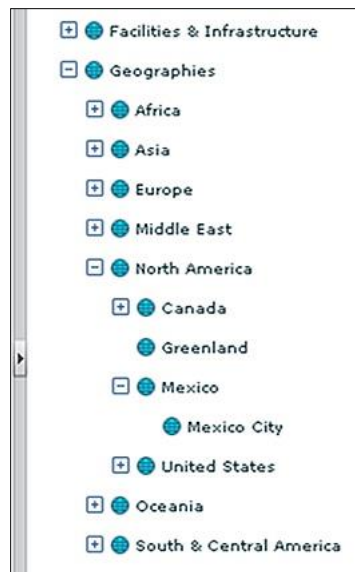
pesticides UF: fumigants BT: agrochemicals NT: fungicides herbicides insecticides RT: pests	plant products NT: cereals fruits spices vegetables RT: plants
pests NT: pest insects plant pests RT: pesticides	plants RT: plant products
pigs UF: hogs porkers BT: livestock	porkers USE: pigs
plant pests BT: pests	poultry BT: livestock NT: chickens ducks geese turkeys RT: eggs
	sheep BT: livestock RT: wool

*thesaurus*

Modern information taxonomies arose in 1990s with the user interface capabilities of the Web and the rise of information architecture.

# Taxonomies: Features

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



# Taxonomies vs. Thesauri

---

## Thesauri

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

## Taxonomies

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

Thesauri may have taxonomy features, and taxonomies may have thesaurus features. There is somewhat of a continuum between thesauri and taxonomies.

# Taxonomy Features

## **Controlled** and **organized**

### 1. Controlled:

A kind of controlled vocabulary or knowledge organization system, based on unambiguous concepts, not just words:  
**things, not strings**

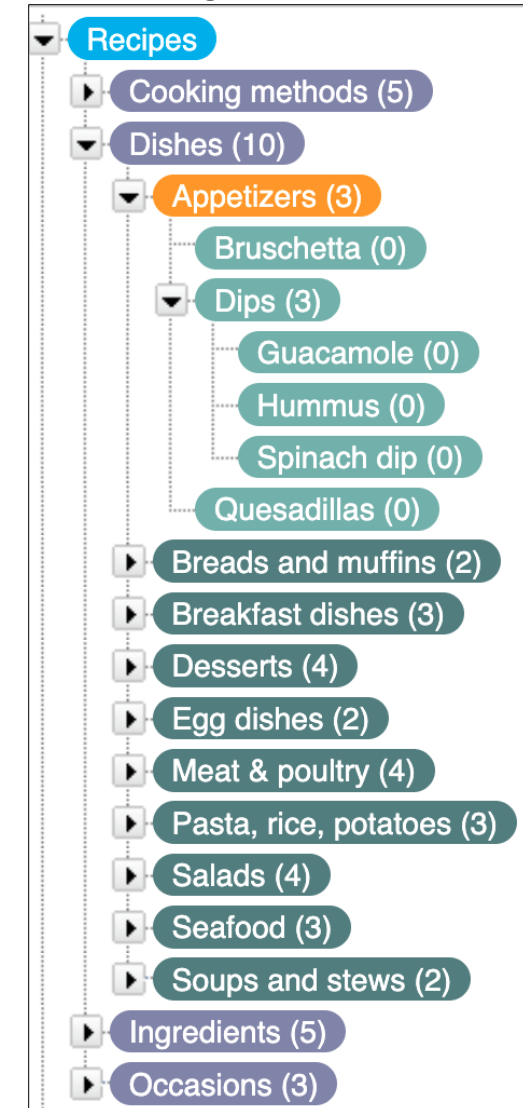
### 2. Organized:

Concepts are arranged in a structure of hierarchies, categories, or facets to organize them.

## Controlled

<b>Preferred Label</b>	
<input type="radio"/> Vorspeisen	de
<input type="radio"/> Appetizers	en
<b>Alternative Labels</b>	
<input type="radio"/> Hors d'oeuvres	en
<input type="radio"/> Starters	
<input data-bbox="1149 654 1187 675" type="button" value="+"/>	
<b>Hidden Labels</b>	
<input type="radio"/> Appetisers	en
<input type="radio"/> Horderves	
<input data-bbox="1149 903 1187 925" type="button" value="+"/>	
<b>Scope Notes</b>	
<input type="radio"/> Dishes usually served as appetizers	en
<input data-bbox="1149 1103 1187 1125" type="button" value="+"/>	
<b>Definitions</b>	
<input type="radio"/> A small dish of food or a drink taken before a meal or the main course of a meal to stimulate one's appetite	en

## Organized



# Taxonomy Features in the Front End

data

- 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

**Controlled:**  
PoolParty product Help  
<https://help.poolparty.biz>

## Topics (8)

- > Standards - 1029
- > Activities and methods - 1028
- ▼ Data - 1000
  - Metadata - 345
    - > Unstructured data - 213
    - ▼ Structured data - 117
      - > Spreadsheet - 99
        - Relational data - 19
    - ▼ Open data - 72
      - Linked Open Data - 52
    - Linked Open Data - 52
    - Master data - 42
    - Data structures - 21
    - Sensor data - 7
    - Streaming data - 7
  - ▼ Knowledge organization systems - 898
    - > Taxonomies - 598
    - > Knowledge graphs - 405
    - Ontologies - 318

**Organized:**  
Semantic Web Company  
Knowledge Hub Topics

# Taxonomy Benefits

---

## 1. Controlled vocabulary aspect

Brings together different wordings (synonyms) for the same concept

- Helps people search for information by different names

## 2. Classification and structure aspect

Organizes information into a logical structure

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

# Taxonomy Types

- 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



# Taxonomy Structure Benefits

---

## Structure: Hierarchy purposes and benefits

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



# Taxonomy Structure Benefits

## Structure: Facet purposes and benefits

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

### 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

# Outline

---

- Background and definitions
- Indexes
- Controlled vocabularies and thesauri
- Taxonomies
- Search in combination with thesauri or taxonomies
- Taxonomy uses and applications
- Information findability and AI

# Search in Combination with Thesauri or Taxonomies

**Terms**

- Marketing
- Marketing industry
- Marketing executives
- Marketing consulting services
- Marketing management
- Marketing communications
- Marketing consultants
- Marketing management software
- Marketing personnel
- Marketing cooperatives
- Marketing models

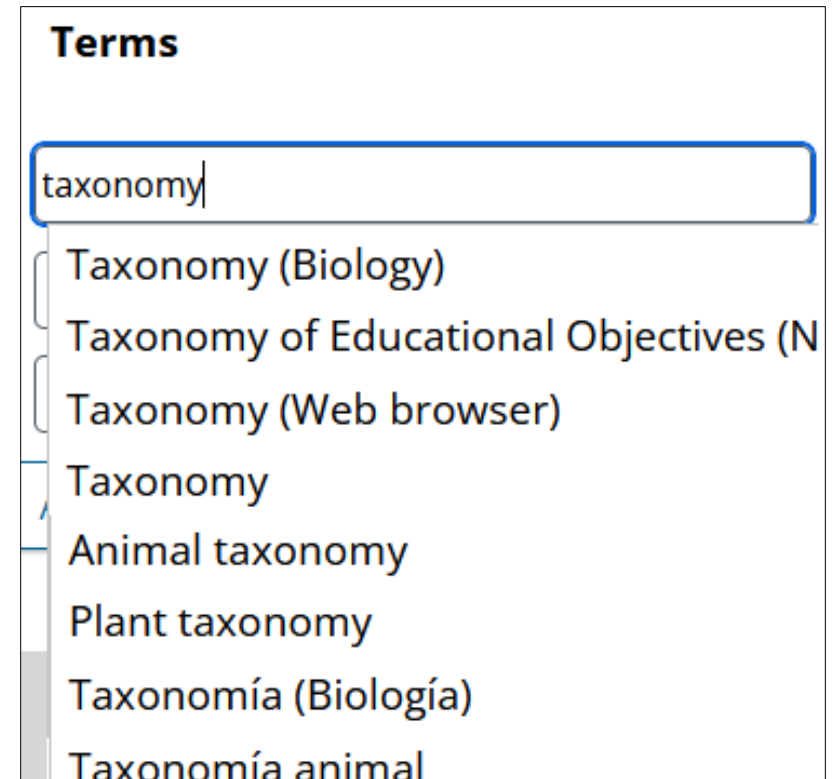
- Users often prefer searching over browsing to save time.
- Thesauri or taxonomies might not always be fully displayed to end-users.
- There could be a partial display of a lookup of terms.

# Search in Combination with Thesauri or Taxonomies

---

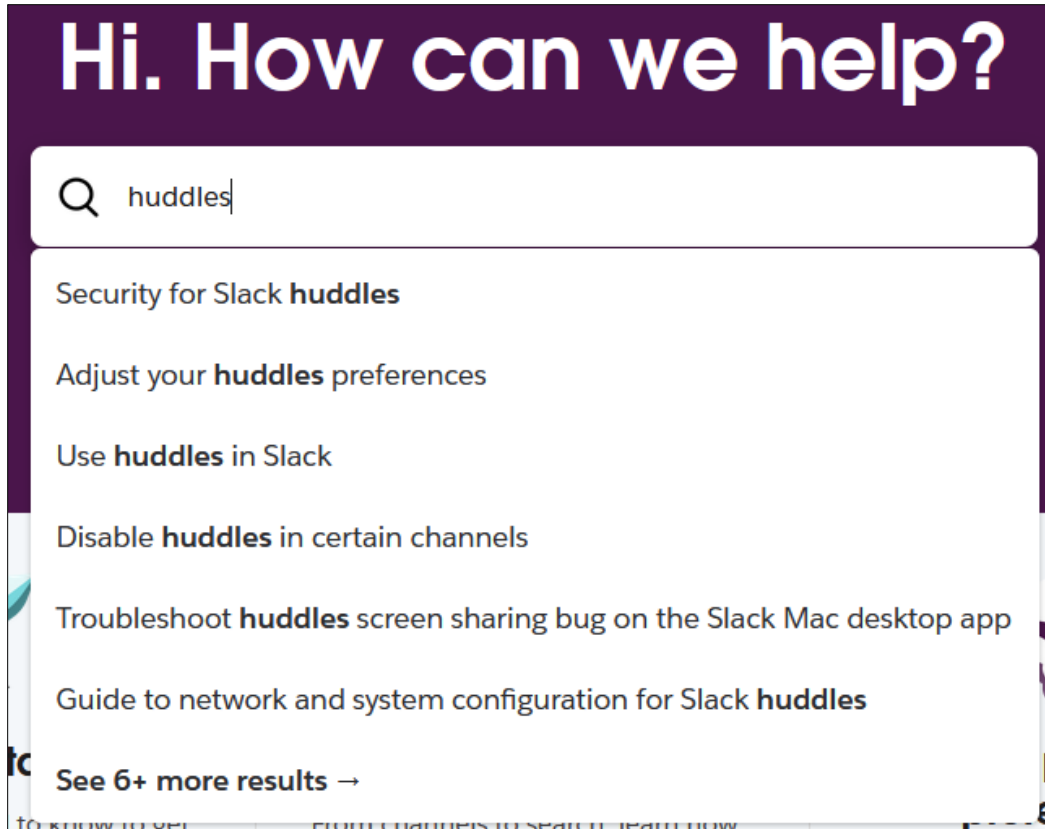
## Search Types

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

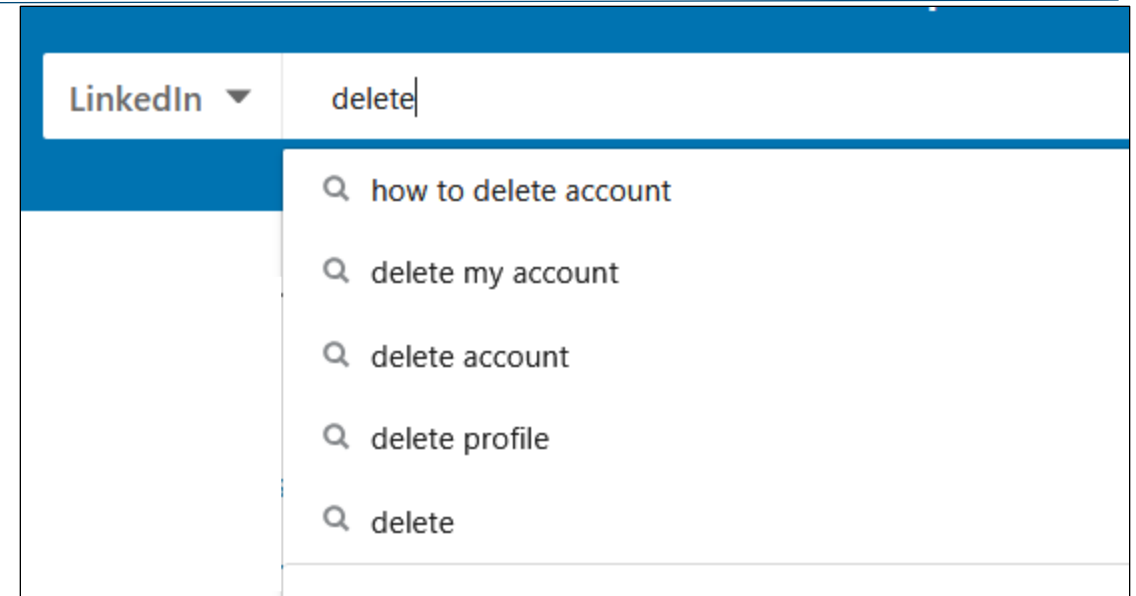


Smart search on  
taxonomy terms

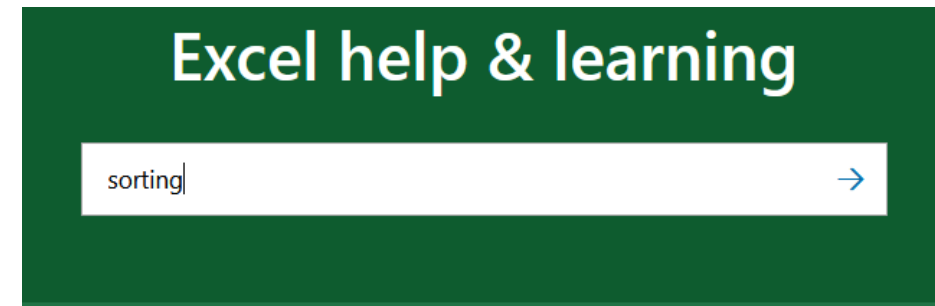
# Search *Not* in Combination with Thesauri or Taxonomies



Search terms match to Help article titles



Search terms match past searches



Search terms have no drop-down matches

# Outline

---

- Background and definitions
- Indexes
- Controlled vocabularies and thesauri
- Taxonomies
- Search in combination with thesauri or taxonomies
- **Taxonomy uses and applications**
- Information findability and AI

# Taxonomy Uses and Applications

---

## Purposes and uses of taxonomies

- Consistent tagging/indexing
- Topic/category browsing
- Search (matching search strings to concepts)
- Discovery (related concept links or contents sharing the same concepts)
- Automatic linking of relevant topics for personalization or recommendation

## Purposes and uses of taxonomies, along with other controlled vocabularies

- Filtering results
- Sorting results
- Content management workflow (rights, audience, retention, etc.)
- Consistent metadata for identification, comparison, analysis
- Visualization of topics (importance and/or relations)
- Curated content in feeds or info boxes

# Taxonomy Uses and Applications

---

## Broad applications of taxonomies

- Information tagging and external retrieval:
  - Technical (Help) documentation for users
  - Website marketing content (news, blog posts, white papers, articles)
  - Ecommerce product search, discovery, and recommendation
  - Marketplace and job opportunity matching online services
  - Media or publishers' articles for subscribers
- Information tagging and internal retrieval:
  - Managing content, including content reuse (proposals, presentations, reports, etc.)
  - Images, video, audio and other digital asset findability and reuse
  - Intranet (such as SharePoint) content search
  - Research information for research and development
  - Knowledge base searchability for customer support representatives
  - Training course search and recommendation
  - Document management for regulatory compliance



# Taxonomy Uses and Applications

- Taxonomies, unlike thesauri, tend to be structured into high-level “concept schemes.”
- Concept schemes align with facets.
- Faceted taxonomies are easily implemented in content management systems.
- Faceted taxonomies are popular and easy for users to interact with.

[enterprise-knowledge.com/knowledge-base](https://enterprise-knowledge.com/knowledge-base)

## Topic

- Agile, Design Thinking, & Facilitation (60)
- Artificial Intelligence (50)
- Change Management & Communications (26)
- Company (21)
- Content Strategy and Operations (87)
- Data Management Services (4)
- Enterprise Learning (23)
- Enterprise Search (54)
- Knowledge Graphs & Data Modeling (130)
- Knowledge Management Strategy & Design (267)
- Taxonomy & Ontology Design (122)
- Technology Solutions (103)

## Article Type

- Blog (343)
- Podcast (83)
- Presentation (74)
- Case Study (50)

## Document Sources

- PoolParty FAQs (5)
- Developer Guide (4)
- PoolParty Release Notes (2)
- User Guide for Knowledge Engineers (1)
- PoolParty Overview (1)

## PoolParty Product

- Inference tagging (8)
- Workbench (5)
- Concept labels (5)
- PoolParty for SharePoint (3)
- Authorization (2)
- Attributes (2)
- PowerTagging (2)
- Term store synchronization (2)
- LLM (2)
- Reporting (1)

[help.poolparty.biz](https://help.poolparty.biz)

## Topics

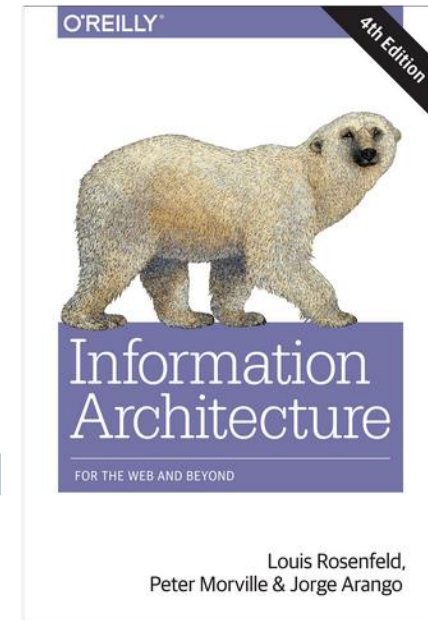
- Semantic tagging (7)
- Taxonomies (6)
- URI (5)

# Taxonomy Uses and Applications - Information Architecture

## Information architecture definitions

1. The structural design of shared information environments.
2. The synthesis of organization, labeling, search, and navigation systems within digital, physical, and cross-channel ecosystems.
3. The art and science of shaping information products and experiences to support usability, findability, and understanding.
4. Organizing and labeling web sites, intranets, online communities, and software to support findability and usability.
5. An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape.

-- Louis Rosenfeld, Peter Morville, and Jorge Arango, *Information Architecture*, 4th ed., O'Reilly, 2015, p. 24

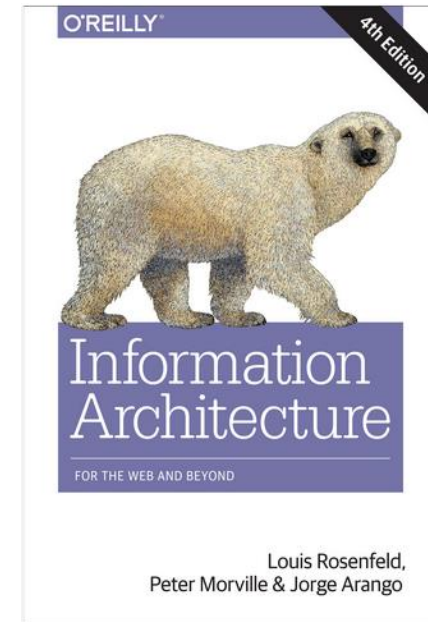


# Taxonomy Uses and Applications - Information Architecture

## Information architecture definitions

1. The structural design of shared information environments.
2. The synthesis of organization, labeling, search, and navigation systems within digital, physical, and cross-channel ecosystems.
3. The art and science of shaping information products and experiences to support usability, findability, and understanding.
4. Organizing and labeling web sites, intranets, online communities, and software to support findability and usability.
5. An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape.

-- Louis Rosenfeld, Peter Morville, and Jorge Arango, *Information Architecture*, 4th ed., O'Reilly, 2015, p. 24



# Taxonomy Uses and Applications - Information Architecture

---

## Information architecture components

1. Organization systems
  - Taxonomies are knowledge organization systems
2. Labeling systems
  - Taxonomies provide labels for concepts
3. Navigation systems
  - Many taxonomies can be navigated by browsing
4. Searching systems
  - Taxonomies support search

Indexing and thesauri are more closely related to documentation, library, and information science.

Taxonomies are more closely related to information architecture, and are related to information access user experience.



# Outline

---

- Background and definitions
- Indexes
- Controlled vocabularies and thesauri
- Taxonomies
- Search in combination with thesauri or taxonomies
- Taxonomy uses and applications
- **Information findability and AI**

# Information Findability and AI

---

What do user prefer to do?

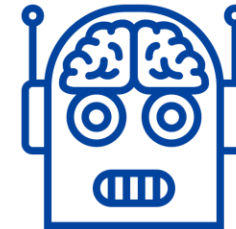
1. to explore topics and interact to find information
  - provided by interactive taxonomies and information architecture

Or

2. to have answers and information served to them
  - provided by AI



vs.



It depends on the user, the information, and the situation (use case).

Taxonomies can also support AI by providing context.

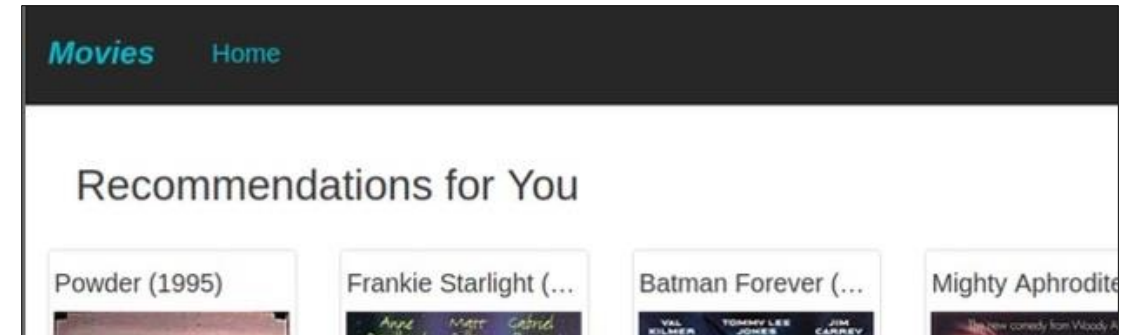
Taxonomies improve AI results.

# Information Findability and AI

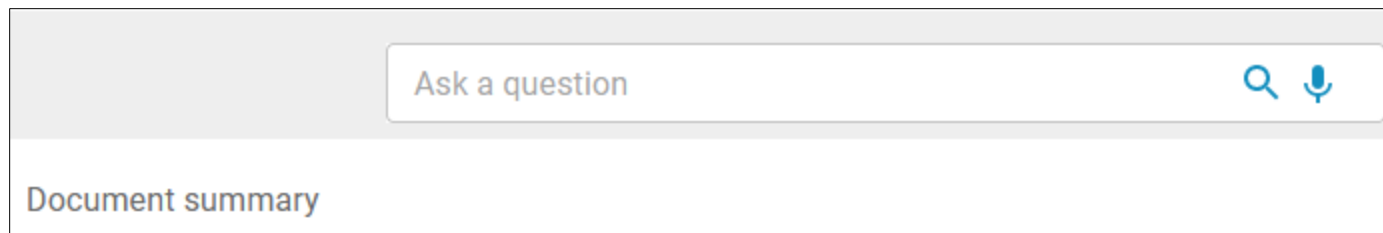
## AI in information management

- Summarize a document
- Translate text into another language
- Predict the completion of text
- ▶- ▶- ▶- Label a piece of content
- Answer a question
- Recommend similar content

Taxonomies can further help with these tasks.



AI-based recommendation systems



Question answering

# AI in Taxonomy Creation and Tagging

---

## Taxonomy development with AI:

- Extraction of candidate terms from a body of text
  - Using text analytics technologies, such as natural language processing (NLP)
- Suggestions of narrower terms and alternative labels
  - Using generative AI and LLMs in taxonomy management software

## Taxonomy concepts need to be tagged to all content and constantly tagged to new content.

- Manual tagging/classification is not scalable for large backlog or large workflow.
- AI methods, existing since the 1990s, have become more common.
- Not all “auto-tagging” uses AI. Rules can also be written to auto-tag.
- Human review remains an option.



# Summary

---

- **Indexes** – best for a single publication (book, manual) or limited set (e.g. multi-volume encyclopedia) that is indexed and done (“closed”)
- **Controlled vocabularies** – for descriptive metadata values
- **Thesaurus** – for open/article/content indexing, especially manual, and for supporting search (matching search strings to thesaurus terms)
- **Taxonomy** – for open/article/content tagging (manual or automated) and a displayed form (hierarchies, categories, facets) for end-user browsing and filtering in addition to supporting search; and for CMS integration

# 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](http://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](http://www.hedden-information.com/accidental-taxonomist/websites)
- Hedden Information Management past presentations  
[www.hedden-information.com/presentations](http://www.hedden-information.com/presentations)
- Hedden Information Management taxonomy training  
[www.hedden-information.com/courses-workshops](http://www.hedden-information.com/courses-workshops)
- Taxonomy Talk, taxonomists community on Discord  
<https://discord.com/invite/3qyMVYCAsw>

# Upcoming Taxonomy Presentations and Workshops

---

- “Connecting Users to Content Through Taxonomies” full-day pre-conference workshop  
Information Architecture Conference, Philadelphia, April 29, 2025  
[www.theiaconference.com/sessions/connecting-users-to-content-through-taxonomies/](http://www.theiaconference.com/sessions/connecting-users-to-content-through-taxonomies/)
- Virtual Bite-Sized Taxonomy Boot Camp London (3 sessions/hours each time)  
March 12, June 18, and October 8, 2025  
[www.taxonomybootcamp.com/London/](http://www.taxonomybootcamp.com/London/)

Any time:

- “Controlled Vocabularies and Taxonomies” training course of four 75-minute video recordings,  
HS Events  
[www.henrystewartconferences.com/events/events-demand-content/Controlled\\_Vocabularies\\_and\\_Taxonomies](http://www.henrystewartconferences.com/events/events-demand-content/Controlled_Vocabularies_and_Taxonomies)
- “Taxonomies and Controlled Vocabularies” training course of 5 web-based lessons  
and exercises, Hedden Information Management  
[www.hedden-information.com/taxonomy-courses-workshops/taxonomy-course/](http://www.hedden-information.com/taxonomy-courses-workshops/taxonomy-course/)

# Questions/Contact

---

## Heather Hedden

Taxonomy Consultant

Hedden Information Management

Carlisle, MA USA

[heather@hedden.net](mailto:heather@hedden.net)

[www.hedden-information.com](http://www.hedden-information.com)

[accidental-taxonomist.blogspot.com](http://accidental-taxonomist.blogspot.com)

[www.linkedin.com/in/hedden](http://www.linkedin.com/in/hedden)



**Hedden Information  
Management**

Making information findable