## Taxonomies in support of search

Taxonomy Boot Camp London

17 October 2018

Presented by
Heather Hedden
Senior Vocabulary Editor
Gale, A Cengage Company

## **About Heather Hedden**

- Controlled vocabulary editor at a library database vendor, Gale, a Cengage company, 1996 – 2004, 2014 – present
- Taxonomy consultant
- Author, The Accidental Taxonomist (Information Today, Inc.)
- Instructor, 5-week online course in taxonomy creation
- ISO and NISO index/controlling vocabulary working group member
- North American representative for BARTOC (Basel Register of Thesauri, Ontologies & Classifications)



#### **Outline**

- Trends in taxonomy implementation and end-user display
  - Taxonomy for browsing
  - Taxonomy to be searched on
  - Facets and post-search refinements
- Options in searching on taxonomies
  - Different types of search on taxonomy
  - Displayed and nondisplayed taxonomy terms
  - Alternative labels / synonyms / nonpreferred terms
- Benefits of taxonomies over search alone



## Trends in taxonomy implementation and display

- Originally, full taxonomy hierarchical browse or thesaurus alphabetical browse.
  - Search was separate from browse and from taxonomies.
- 2. Full large taxonomies came to be displayed less. Search on taxonomy terms.
- Search on more than just the taxonomy terms.
   Search on a combination of taxonomy terms and words in titles, texts, etc. (default keyword/basic search).
- 4. Faceted taxonomies combining search and limited browsing. Especially with smaller taxonomies.
- 5. Large taxonomies/thesauri also integrated into faceted displays. Taxonomy terms as post-search filters.



## **Trends: Full taxonomy browse**



## **Trends: Full taxonomy browse**



Web Images Groups News Shopping Maps Scholar more >

Google Search

Directory Help

#### The web organized by topic into categories.

	•	٠	-	в.	-
•	•		_		

Music, Movies, Performing Arts, ...

#### Business

Industrial Goods and Services, Finance, ...

#### Computers

Software, Internet, Programming, ...

#### Games

Video Games, Roleplaying, Board Games, ...

#### Home

Cooking, Family, Gardening, ...

#### Kids and Teens

International, School Time, Games, ...

#### News

Newspapers, Media, Colleges and Universities, ...

#### Recreation

Reference

Pets, Outdoors, Food, ...

#### Health

Conditions and Diseases, Medicine, Animal, ... Education, Biography, Museums, ...

#### World

Deutsch, Español, Français, Italiano, Japanese, Korean, Nederlands, Polski, Svenska, ...

#### Regional

North America, Europe, Oceania, ...

#### Science

Biology, Social Sciences, Technology, ...

#### Shopping

Home and Garden, Crafts, Sports, ...

#### Society

Religion and Spirituality, Law, Issues, ...

#### Sports

Soccer, Equestrian, Football, ...

19 July 2011

## Trends: Full taxonomy browse



Home

News

s Travel

Money

Sports

Lif

Те

Weather

### **USA TODAY Content Tree**

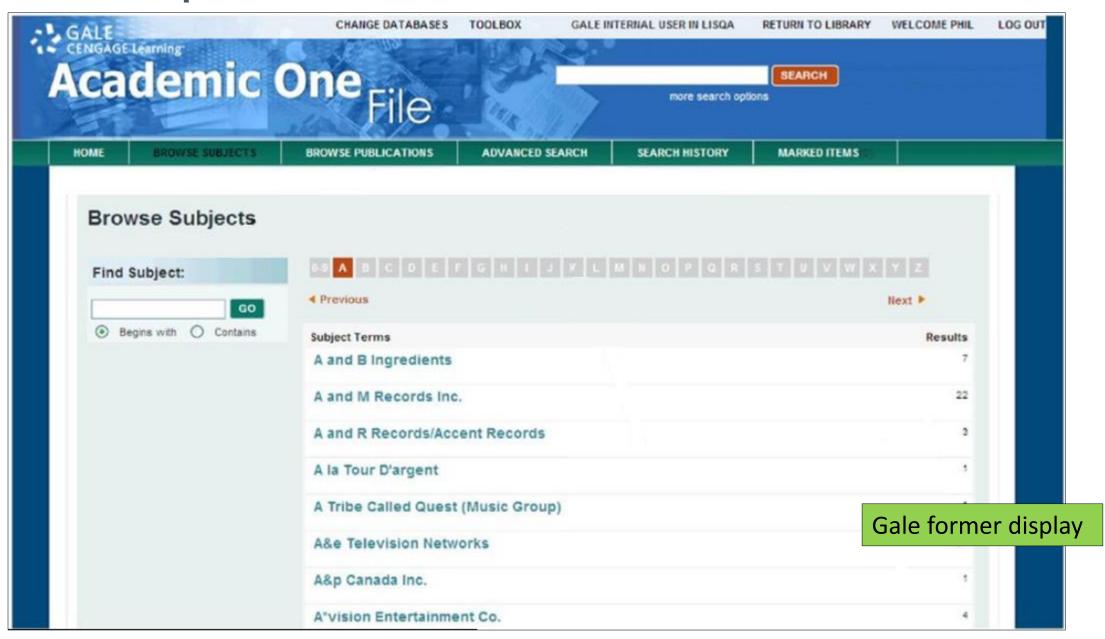
The following topic tree represents the entire USA TODAY story collection going back as far as late 2003. We've presented these stories within our site structure, which has up to four (4) levels in each of our seven (7) site sections. Clicking on each tree label will take you to a master index for that part of the site structure. You can expand and collapse the site structure to expose as much content as you would like to see. Happy browsing and be sure to let us know if you can't find something you are looking for.

Browse: Tag categories | Index of reporters

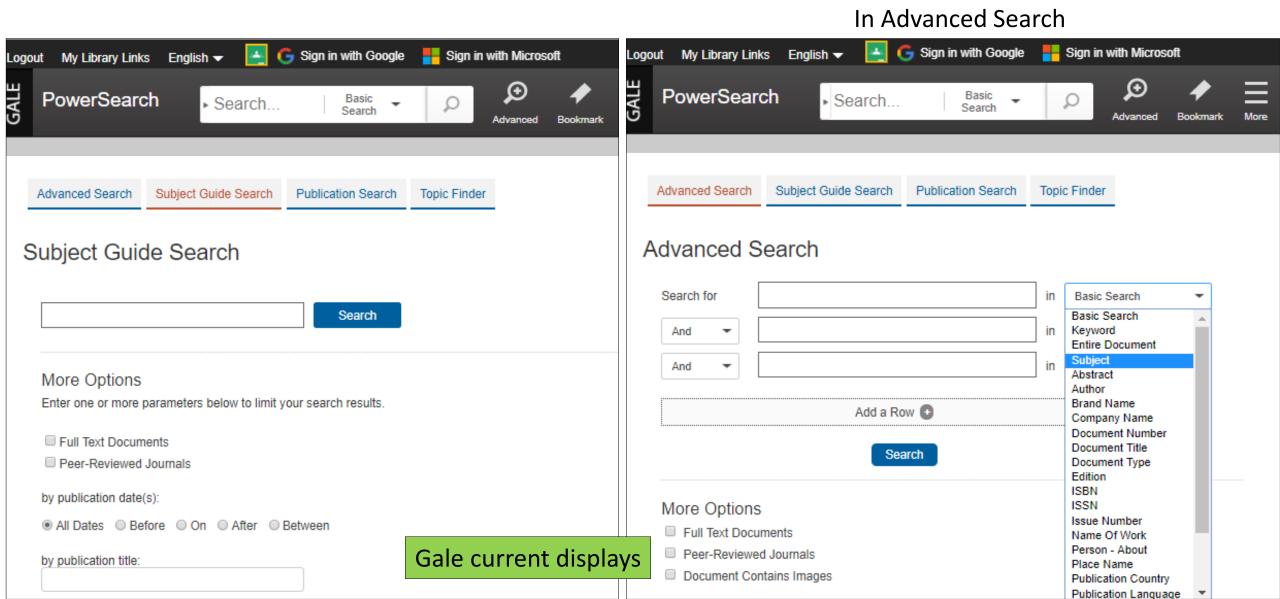
- USA TODAY Content Tree
  - news 245,086 stories
    - americawants 12 stories
    - breakingnews 7,406 stories
    - bythenumbers 25 stories
    - education 2,854 stories
    - elections 190 stories
    - gallery 421 stories
    - health 10,633 stories
      - backtoschool 7 stories
      - bonus 4 stories
      - painter 163 stories
      - spotlight 4 stories
      - spotlighthealth 47 stories
      - weightloss 398 stories
      - wellness 68 stories
      - yourhealth 49 stories
      - lottery 14 stories
      - military 1,262 stories
    - nation 43,745 stories



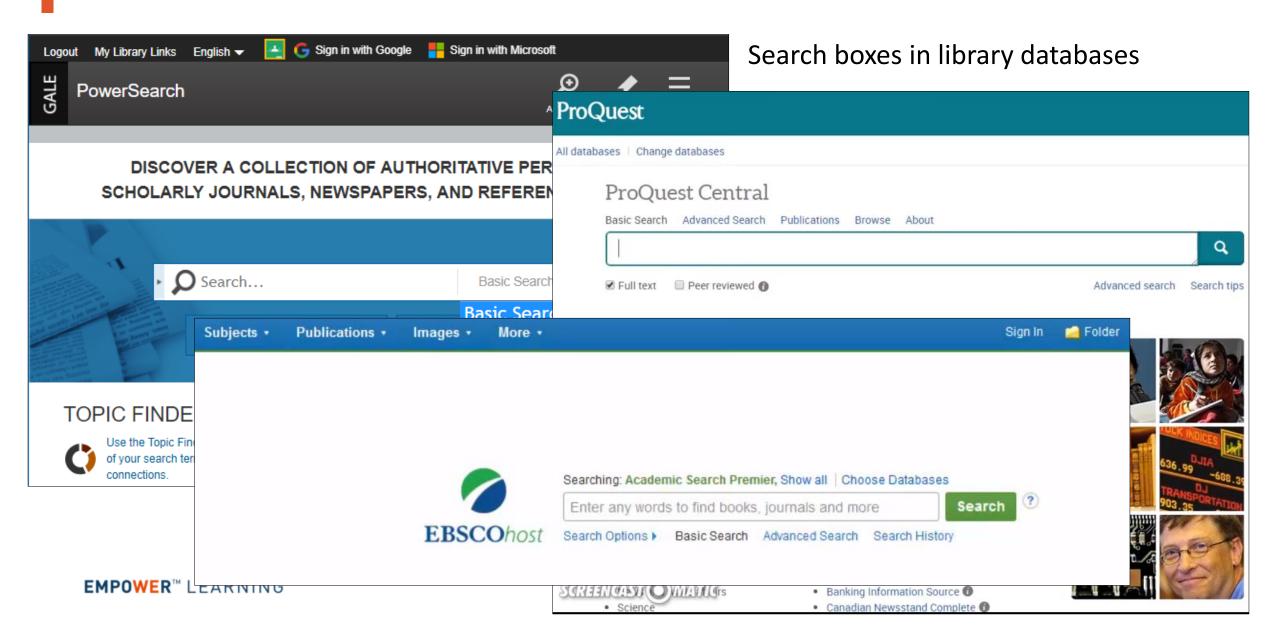
## **Trends: Alphabetical browse**



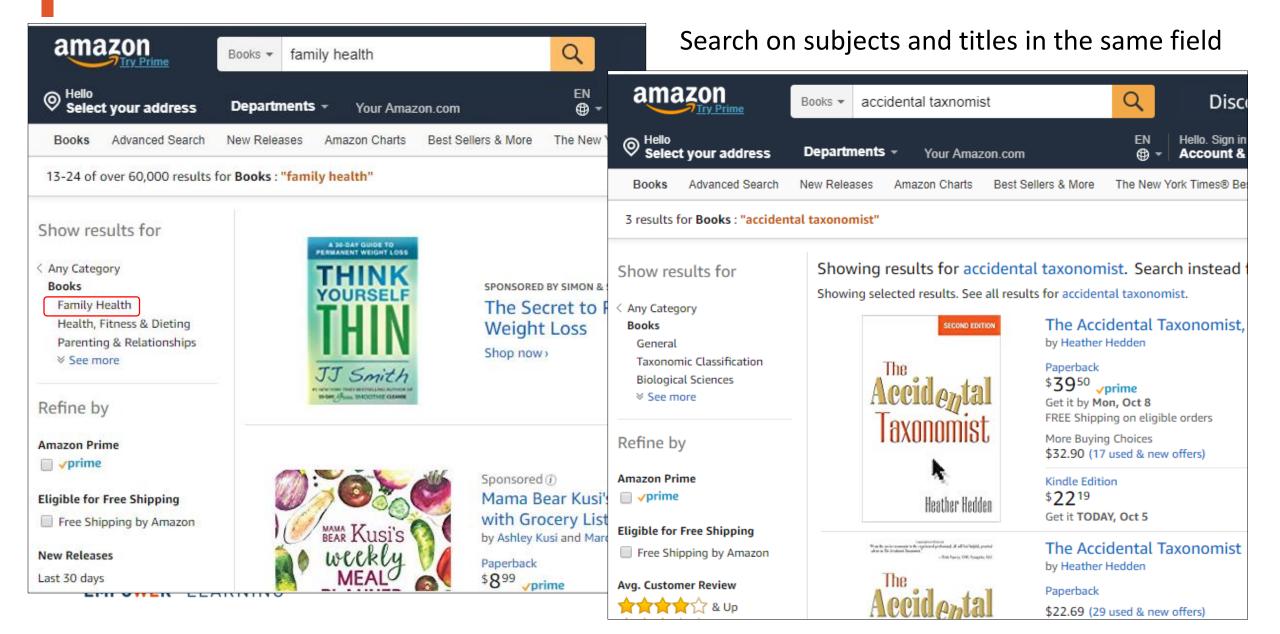
## **Trends: Search on taxonomy terms**



## Trends: Search on more than the taxonomy

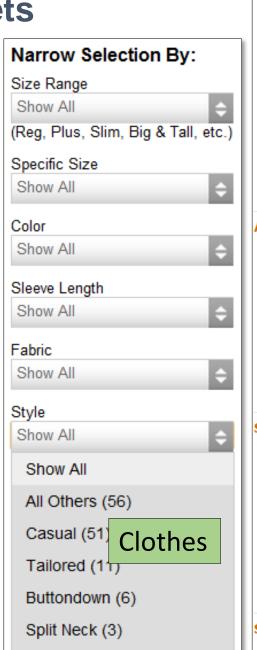


## Trends: Search on more than the taxonomy

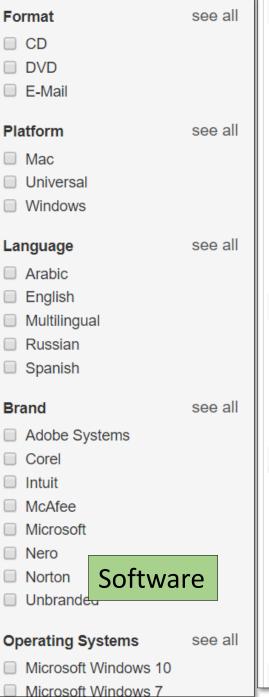


### **Trends: Facets**

Examples of ecommerce facets



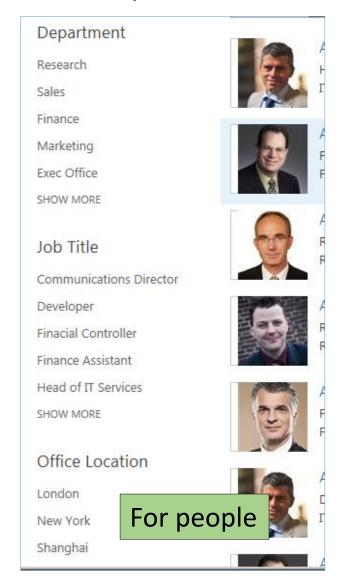


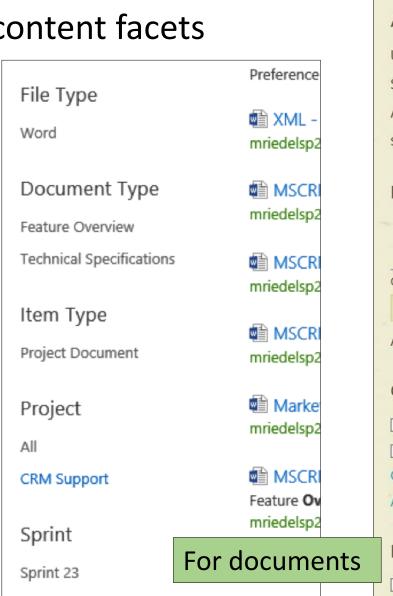


Select category(s)  Banquet Tables (4)	Clear
Bistro Table (2) Bistro Tables (5) Counter-Height Table (1) Counter-Height Tables (6) Dining Table (10) Dining Tables (52) Folding Tables (8) Folding Tables (12) Kitchen Tables (1) Kitchen Tables (1) Nook Tables (1) Pub Tables (7) Pub Tables (29)	
Material	~
Select material(s)	Clear
Hardwood (29) MDF Composite (1) Metal (28) Plastic (1) Wood (48) Wood Composite (35)	
Finish	~
Select finish(s)	Clear
Cherry (4) Dark Cherry (1) Ebony (1) Espresso (14) Mahogar Natural (Oak (7) Painted (8) Unfinished (1) Walnut (8)	
Color	~

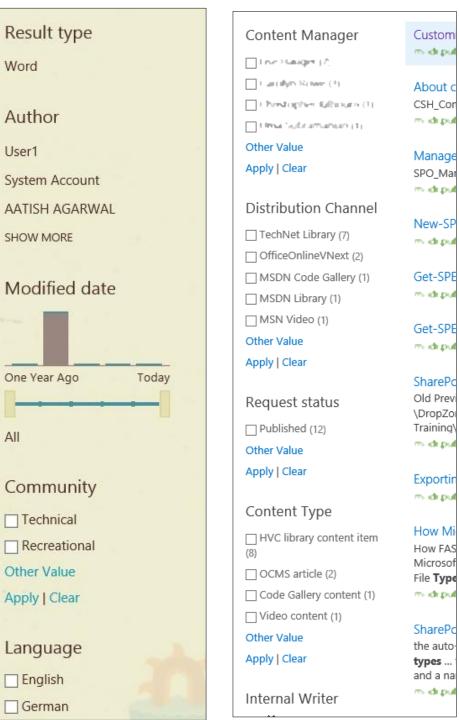
### **Trends: Facets**

## Examples of internal content facets





Sprint 22



## **Trends: Facets**

## Issues with faceted taxonomy

- Content needs to be of a uniform type sharing the same set of multiple facets
  - Examples: employees, colleges, films, music, clothing products, museum artifacts
- Controlled vocabularies for each facet should be small to display well.
  - A large subject thesaurus or multi-level taxonomy cannot easily be browsed.
  - A large topic taxonomy could be the first step, with facets for further refinement.



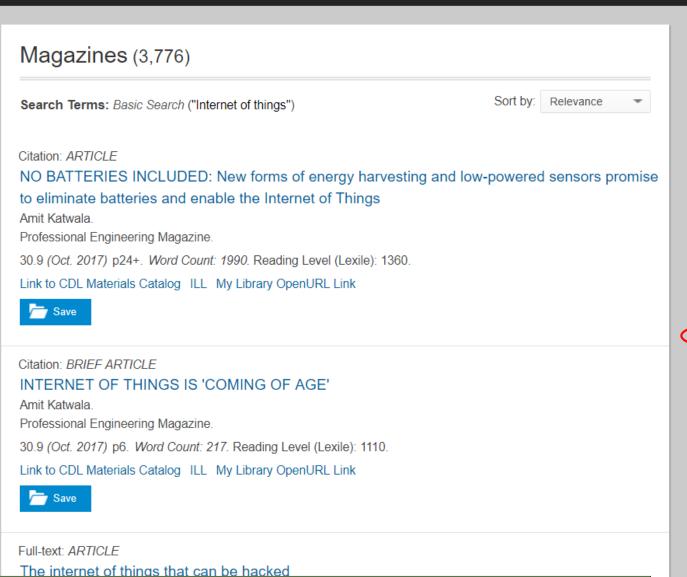
## Trends: Taxonomy terms to refine post-search results

- A display of taxonomy terms that have been used to index the content items in the search result set (not all taxonomy terms).
- A display of selected terms from the taxonomy, not the taxonomy itself.
- Any relationships between terms are not indicated.
- Displayed in order of usage frequency of the search result set.

Suitable when a large taxonomy or thesaurus does not fit into a facet.



Browse by Discipline



**Content Types** Academic Journals (6,397) Magazines (3,776) Books (24) News (38,942) Images (2) Videos (4) Search within results Limit Search by: Full Text Contains Images **Publication Dates:** Past Week Past Month Past Year Custom Date Range + Subjects Internet of things (724) Software industry (197)

Semiconductor indust... (120)

Byline: Rachel 7, Arndt When doctors replaced then-Vice President Dick Cheney's pacemaker in 2007, they asked the

Subjects for post-searching filtering in Gale search results



#### 2. Pitfalls in the Use of Time Series Methods to Study **Deterrence** and **Capital** Punishment.







Journal

By: Charles, Kerwin; Durlauf, Steven. Journal of Quantitative Criminology. Mar2013, Vol. 29 Issue 1, p45-66. 22p. DOI: 10.1007/s10940-012-9169-7. , Database: Criminal Justice Abstracts with Full Text

Subjects: CAPITAL punishment; CRIMINAL law; PUNISHMENT in crime deterrence; CRUELTY; CRIMINAL investigation; DECISION theory



#### 3. **Deterrence** and the death penalty: Why the statistics should be ignored.







Journal

By: Nagin, Daniel. Significance. Apr2014, Vol. 11 Issue 2, p9-13. 5p. DOI: 10.1111/j.1740-9713.2014.00733.x., Database: Academic Search Complete

Subjects: Punishment in crime deterrence; Capital punishment; Murder - Law & legislation; Statistics; National Research Council (U.S.); Nagin, Daniel





#### 4. Deterrence and the Death Penalty: Partial Identification Analysis Using Repeated Cross Sections.







Journal

By: Manski, Charles; Pepper, John. Journal of Quantitative Criminology. Mar2013, Vol. 29 Issue 1, p123-141. 19p. DOI: 10.1007/s10940-012-9172-z., Database: Criminal Justice Abstracts with Full Text

Subjects: CAPITAL punishment; CRIMINAL law; PUNISHMENT; EXECUTIONS & executioners; HOMICIDE; CRIMINOLOGY

#### Show all 6 images

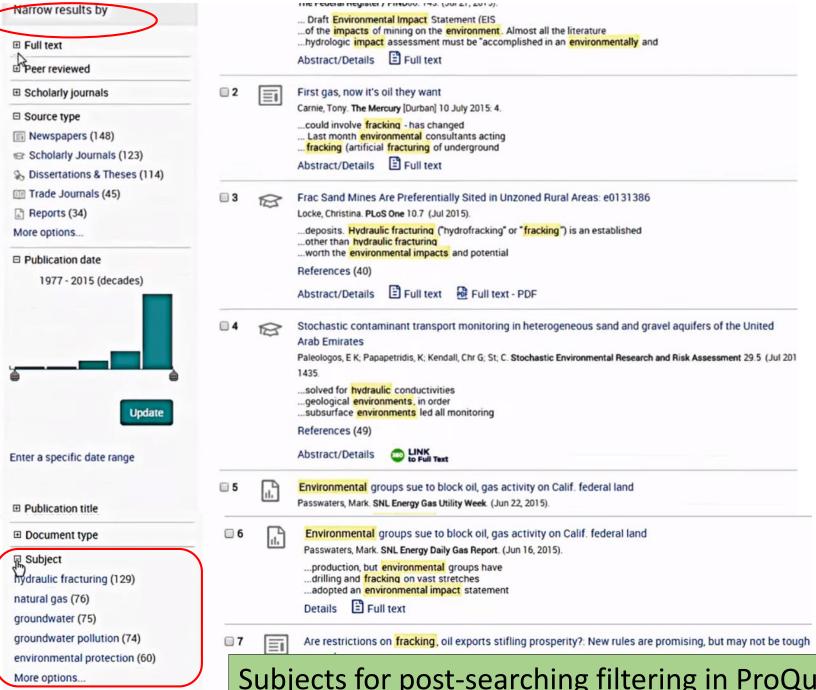
Spike 1. Howevall reproper 100,807 wealther for your sec-	Tier	One		. Total
scame ann a 167		Desail	See	
	609	64	90.E	116
	187	40	92	10
		100 %	44.2	

		100
brigation of a fact assists away (4.5,)		4.0
Notes on a loss of the first pool, firstly for all	Otto:	911
Name and Address of the Address of t	the contract of	40

1 349	Selle A. Common office (10) for the common office (10)
1.52	Sept.
1	Profes manual III
1	Character recent cities

and the same of th			
HART.		76	
	#1 Place II Place II	1000	-

Subjects for post-searching filtering in EBSCO search results



Subjects for post-searching filtering in ProQuest search results

minimize groundwater pollution

Related Resources

# NO BATTERIES INCLUDED: New forms of energy harvesting and low-powered sensors promise to eliminate batteries and enable the Internet of Things

Amit Katwala

Professional Engineering Magazine. 30.9 (Oct. 2017): p24+.

#### Source Citation (MLA 8th Edition)

Katwala, Amit. "NO BATTERIES INCLUDED: New forms of energy harvesting and low-powered sensors promise to eliminate batteries and enable the Internet of Things." *Professional Engineering Magazine*, Oct. 2017, p. 24+. *Academic OneFile*,

http://link.galegroup.com/apps/doc/A514405340/AONE?u=gale&sid=AONE&xid=61541bac.

Accessed 15 Mar. 2018.

**Gale Document Number:** GALE|A514405340

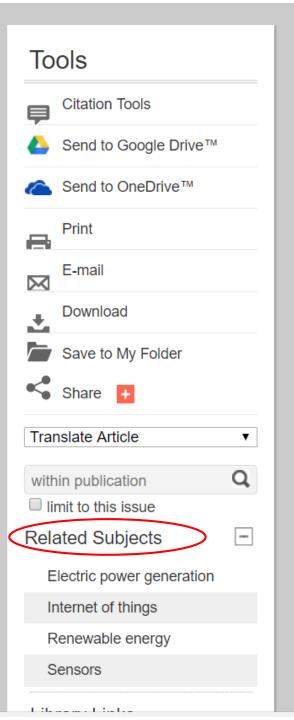
#### **RELATED RESOURCES**

#### Studies Conducted at Zhongy...

Energy Weekly News. (Mar. 16, 2018) p155. Word Count: 391.

#### Global Al Market in Agricultur...

PR Newswire. (Mar. 9, 2018) Word Count: 924.



## **Taxonomy search options**

- Different types of search on taxonomy
- Displayed taxonomy terms in type-ahead or search-suggest
- Alternative labels / nonpreferred terms / synonyms



## Taxonomy search options: Different types of search on terms

## Types

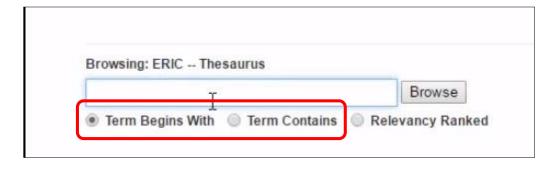
- Exact exact match
- Contains exact match phrase with additional words before or after
- Begins alphabetical from start, but allows end truncation
- Smart words within the term in any order and also internal word stemming (singular/plural)

## **Implementations**

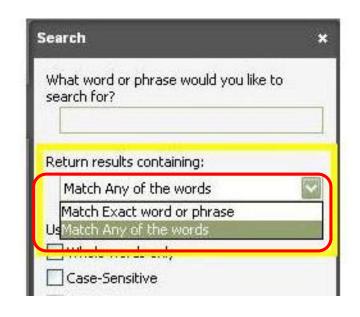
- Exact option if there is a type-ahead display of terms
- Contains also called phrase search. Can be done with quotation marks.
- Begins option if there is a type-ahead display of terms
- Smart sometimes the default, if no other options, and terms are not displayed



## Taxonomy search options: Different types of search on terms







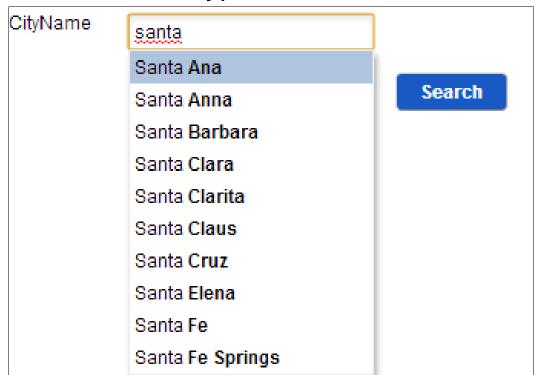
#### Find items that have:

All these words:

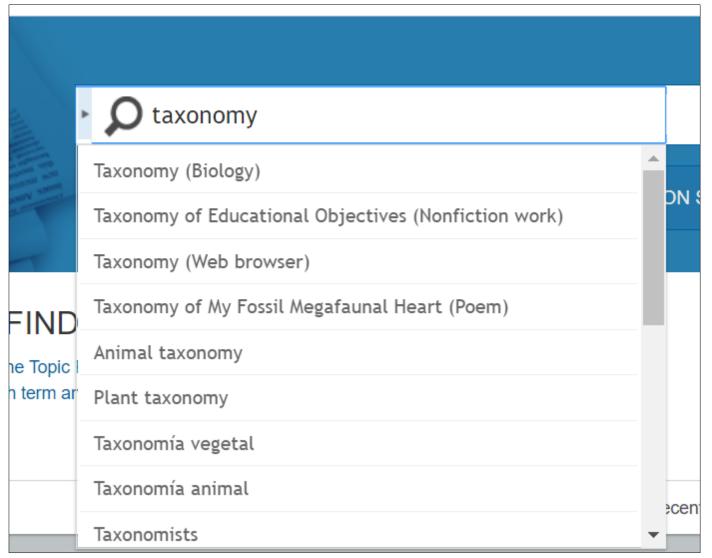
This exact phrase:

## Taxonomy search options: Type-ahead or search-suggest

Type-ahead



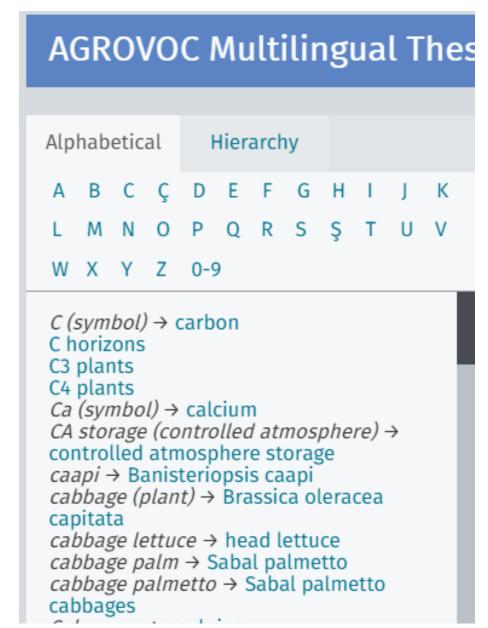
Search-suggest





## Taxonomy search options: Alternative labels/nonpreferred terms/synonyms

- "Synonyms" are very important in helping users find desired content.
- Synonyms are especially important if the taxonomy is not fully displayed.
- Synonyms are displayed in back-end view of the taxonomy for taxonomists and indexers, but rarely displayed for end-users
- Type-ahead/search suggest could make use of synonyms without displaying them.



## Taxonomy search options: Alternative labels/nonpreferred terms/synonyms

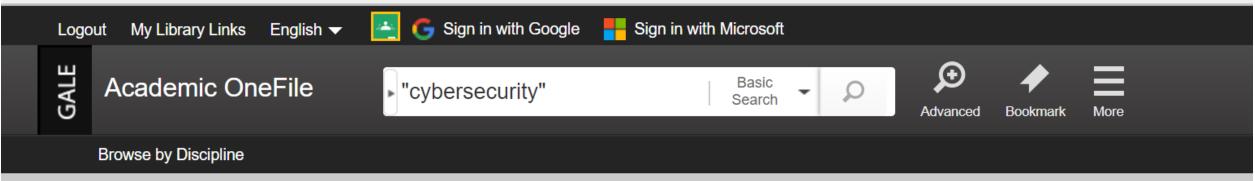


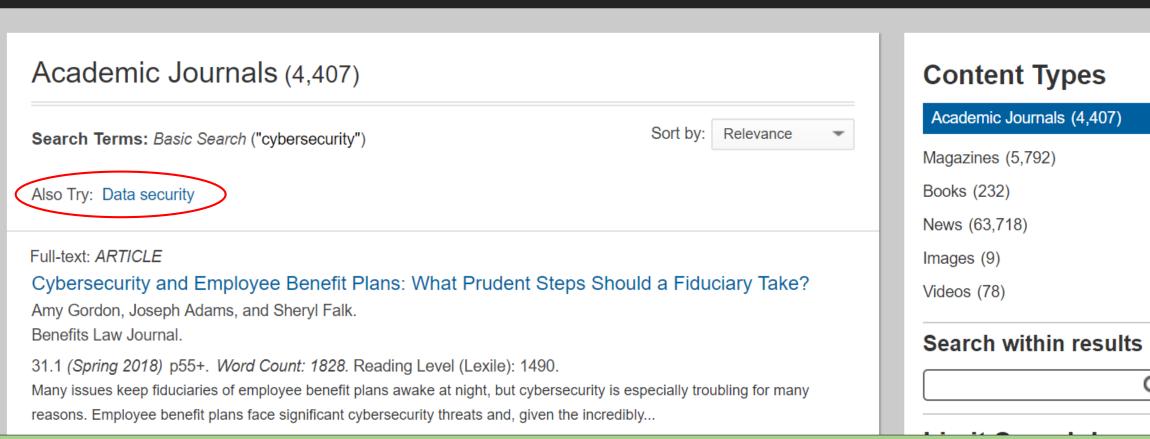
https://designnotes.blog.gov.uk/wpcontent/uploads/sites/53/2017/04/location-pickergeneral-lq.gif

- Synonyms are rarely displayed for endusers.
- Users who enter synonyms are often taken directly to the indexed content.
- Preferred term matches to synonyms typed in by users may be proposed/shown to the users.
- If the taxonomy is not displayed at all, not even in type-ahead, then no preferred label is needed. All labels for a concept are synonyms.
  - Called "synonyms ring" or "search thesaurus"

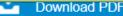








Preferred term (Data security) matched to synonym entered by user (cybersecurity) proposed/shown to the user





Full Text

## Benefits of taxonomies over search alone

- Indexing and retrieval based on concepts, not just words/phrases improves search results in both precision (accuracy) and recall (comprehensiveness)
- Taxonomy terms allow limiting/filtering search by topic.
- Broader categories/terms allow users to choose a broad subject first and then limit by other metadata.
- Relationships between terms allow users to explore related topics.
- Subjects displayed on search results (post-search) allow users to refine and focus their searches by precise topic or explore related topics.
- Taxonomies support the indexing of nontext content (images, video, audio).
- Multilingual taxonomies support accurate search and retrieval across multilingual content.



## **Conclusions**

- The growing size of taxonomies and thesauri may make full displays impractical.
- Searching on the taxonomy is an important end-user feature.
- There are different methods for searching on and displaying taxonomy terms.
- Faceted taxonomies combine search and browse functions, and are useful for many but not all content and taxonomy implementations.
- Taxonomy terms as post-search filters are especially suitable when taxonomies are too large for full display browsing.
- Taxonomies have many benefits over search alone.



## **Questions/Contact**

#### **Heather Hedden**

Senior Vocabulary Editor

Gale, A Cengage Company

Boston, MA U.S.A.

www.cengage.com, www.gale.com

www.cengage.co.uk, www.gale.cengage.co.uk

heather.hedden@cengage.com

blog.gale.com/subject-searching

