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

- 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

1. 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.

3. 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

Yahoo! Directory

<table>
<thead>
<tr>
<th>Arts &amp; Humanities</th>
<th>News &amp; Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photography, History, Literature...</td>
<td>Newspapers, Radio, Weather, Blogs...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Business &amp; Economy</th>
<th>Recreation &amp; Sports</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2B, Finance, Shopping, Jobs...</td>
<td>Sports, Travel, Autos, Outdoors...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Computer &amp; Internet</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware, Software, Web, Games...</td>
<td>Phone Numbers, Dictionaries, Quotes...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education</th>
<th>Regional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colleges, K-12, Distance Learning...</td>
<td>Countries, Regions, U.S. States...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entertainment</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movies, TV Shows, Music, Humor...</td>
<td>Animals, Astronomy, Earth Science...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Government</th>
<th>Social Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elections, Military, Law, Taxes...</td>
<td>Languages, Archaeology, Psychology...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Health</th>
<th>Society &amp; Culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease, Drugs, Fitness, Nutrition...</td>
<td>Sexuality, Religion, Food &amp; Drink...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>New Additions</th>
<th>Subscribe via RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:13, 4:12, 4:11, 4:10, 4:9...</td>
<td>Arts, Music, Sports, TV, more...</td>
</tr>
</tbody>
</table>

13 April 2013
Trends:
Full taxonomy browse

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: Search on taxonomy terms

In Advanced Search

Gale current displays
Trends: Search on more than the taxonomy

Search boxes in library databases
Trends: Search on more than the taxonomy

Search on subjects and titles in the same field
Examples of ecommerce facets

- Clothes
- Books
- Software
- Furniture
Examples of internal content facets

For people

For documents

Result type
- Word
- Author
- System Account
- AATISH AGARWAL
- SHOW MORE

Modified date
- One Year Ago
- Today

Community
- Technical
- Recreational
- All
- CRM Support
- Exporting
- Request status
- Published
- Apply | Clear

Content Type
- Office Online Video (1)
- OCMS article (2)
- Code Gallery content (1)
- Video content (1)
- Apply | Clear

Language
- English
- German

Internal Writer
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.
Subjects for post-searching filtering in Gale search 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 industry (120)
Subjects for post-searching filtering in EBSCO search results
Subjects for post-searching filtering in ProQuest search results

- Hydraulic fracturing (129)
- Natural gas (76)
- Groundwater (75)
- Groundwater pollution (74)
- Environmental protection (60)
NO BATTERIES INCLUDED: New forms of energy harvesting and low-powered sensors promise to eliminate batteries and enable the Internet of Things

Amit Katwala


Source Citation (MLA 8th Edition)

Gale Document Number: GALEJ514405340

Related Subjects

- Electric power generation
- Internet of things
- Renewable energy
- Sensors
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

- Santa
- Santa Ana
- Santa Anna
- Santa Barbara
- Santa Clara
- Santa Clarita
- Santa Claus
- Santa Cruz
- Santa Elena
- Santa Fe
- Santa Fe Springs

Search-suggest

- Taxonomy (Biology)
- Taxonomy of Educational Objectives (Nonfiction work)
- Taxonomy (Web browser)
- Taxonomy of My Fossil Megafaunal Heart (Poem)
- Animal taxonomy
- Plant taxonomy
- Taxonomia vegetal
- Taxonomia animal
- Taxonomists
“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

AGROVOC Multilingual Thesaurus

Alphabetical

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>Ç</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>M</td>
<td>N</td>
<td>O</td>
<td>P</td>
<td>Q</td>
<td>R</td>
<td>S</td>
<td>Š</td>
<td>T</td>
<td>U</td>
<td>V</td>
</tr>
<tr>
<td>W</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td>0-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C (symbol) → carbon
C horizons
C3 plants
C4 plants
Ca (symbol) → calcium
CA storage (controlled atmosphere) → controlled atmosphere storage
caapi → Banisteriopsis caapi
cabbage (plant) → Brassica oleracea capitata
cabbage lettuce → head lettuce
cabbage palm → Sabal palmetto
cabbage palmetto → Sabal palmetto cabbages
Sabal palmetto cabbages
Synonyms are rarely displayed for end-users.

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
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.co.uk, www.gale.cengage.co.uk
heather.hedden@cengage.com

blog.gale.com/subject-searching