Getting Off to a Strong Start with Your Search Taxonomy

Heather Hedden

Principal, Hedden Information Management

Gilbane Conference, Boston
December 4, 2008

Outline

1. Why have a taxonomy
2. Deciding on the taxonomy user interface
3. Deciding on the taxonomy scope
4. Resources

Outline

1. Why have a taxonomy
2. Deciding on the taxonomy user interface
3. Deciding on the taxonomy scope
4. Resources

Why a Taxonomy

Taxonomy Definitions

1. Controlled vocabulary: An authoritative, restricted list of terms (words or phrases)

   ➢ Mainly used for indexing/tagging content to support retrieval

2. A hierarchical system of classification, with broader/parent and narrower/child concepts

   ➢ For categorizing and organizing things

3. Both: a controlled vocabulary that has hierarchical structure among the terms

   ➢ To support both information retrieval and information organization

Why a Taxonomy

1. **Controlled vocabulary (CV):**
   Brings together different wordings (synonyms) for the same concept

   ➢ Helps people search for information by different names

2. **Hierarchical classification:**
   Organizes information into a logical structure

   ➢ Helps people browse or navigate for information

Why a Taxonomy

- A controlled vocabulary gathers synonyms, acronyms, variant spellings, etc.
  - Documents not missed due to use of different words (e.g. *Automobiles*, instead of *Cars*)
  - User does not need to guess the spelling of unusual or foreign names (e.g. *Condoleezza Rice*)

- 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")

Why a Taxonomy

Users may enter:

Oil industry

Oil & gas industry

Oil & gas industries

Petroleum industry

CV contains all synonyms:

Oil industry

Oil & gas industry

Oil and gas industry

Oil & gas industries

Oil and gas industries

Petroleum industry

Oil companies

Big oil

Oil producers

Petroleum companies

Text may contain:
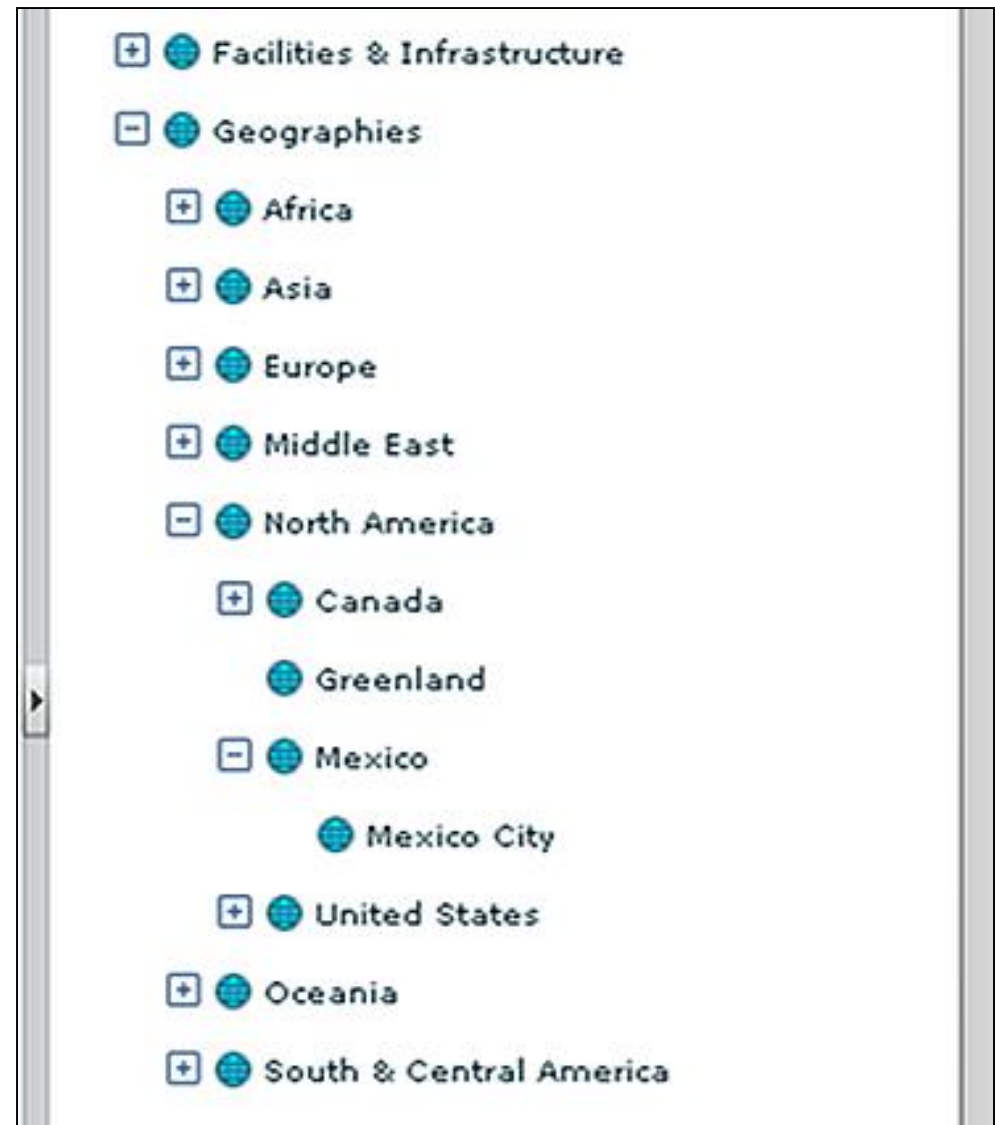
Oil and gas industry

Oil companies

Big oil

Oil producers

Why a Taxonomy

A hierarchical taxonomy provides guided search.

- Users can browse and locate narrower (more specific) subjects of interest.
- Users find out what is included and what is not.
-Users may find related subjects of interest.

Taxonomies reflect natural categories.

Outline

1. Why have a taxonomy
2. Deciding on the taxonomy user interface
3. Deciding on the taxonomy scope
4. Resources

Taxonomy User Interface

Deciding on the Taxonomy User Interface:

- Alphabetical browse
- Hierarchical tree
- Facets
- Search box

- Combination of one or more

Taxonomy User Interface

Alphabetical browse

- Simple list
- Back-of-the-book style index with indented subentries (site A-Z index)

O
off-campus work, 78
office hours responsibility, 80
Office of Research Administration (ORA), 96
office space
    faculty, 109
    librarian, 93
    MSP/Lowell, 10
Official Designee of the Chancellor, 3
Official Personnel File, 65–66
orientation work of faculty, 81

P
paid leaves of absence. *see* leaves of absence, paid
parental/adoptive leave, 134
parking lots and spaces, 108–109
part-time faculty, 59
    collective bargaining representative, 5
    consecutive contracts, 60
    full-time faculty retrenchment impact, 63
    hiring process, 62
    temporarily less than full-time, 122
    temporary part-time (visiting), 60–61
    union membership exclusion, 61–62
patent policy, 95–96
payroll deduction
    agency service fee, 148
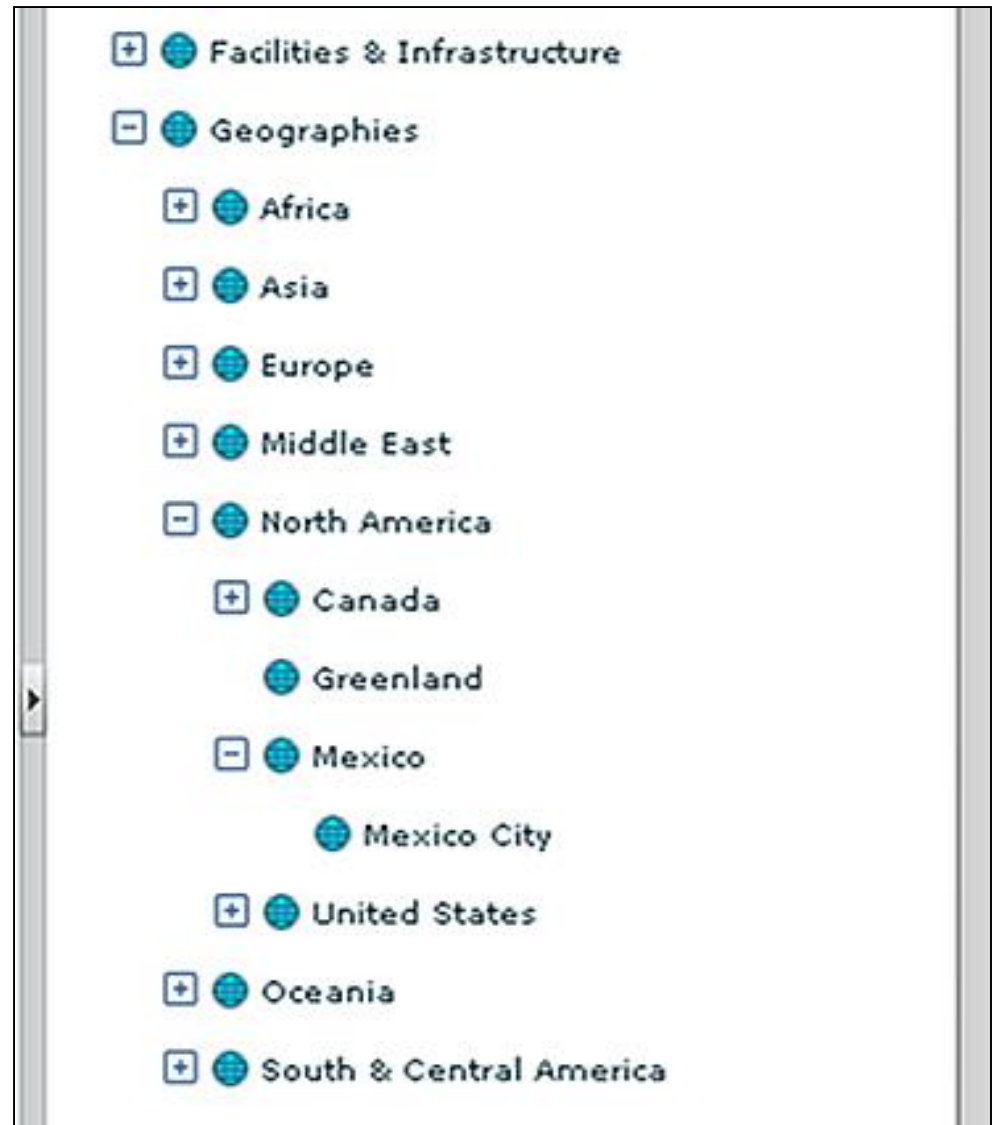
Taxonomy User Interface

Alphabetical browse is suitable for:

- Names (proper nouns), separated out or as a significant component
- Content within a book and book-like narrative texts, including instructional texts
- Unlimited-scope subject area content
- Content vast in volume and depth necessitating a large controlled vocabulary
- Resources for either index updating, or for frequent updates/additions of content
- Content not easily categorized
- Taxonomy browse for human indexers/taggers

Taxonomy User Interface

Hierarchical tree

1. Expandable,
    on same page

Taxonomy User Interface

Hierarchical tree

2. Separate page for each level of the hierarchy

Arts & Entertainment >
Automotive >
Business & Professional Services >
Clothing & Accessories >
Community & Government >
Computers & Electronics >

Find business listings for Arts & Entertainment.

Arcades & Amusements
Amusement Equipment & Supplies, Video Arcades & Internet Gaming Centers, ...

Bars

Carnivals, Fairs, & Festivals
Fairgrounds, Music Festivals, ...

Children's & Family Entertainment
Miniature Golf Courses, Theme Parks, Zoos, ...

Concert Tickets

Cultural Attractions, Events, & Facilities
Historical Places & Services, Libraries, Museums, ...

DVD Rentals

Entertainment Clubs & Nightlife
Comedy Clubs, ...

Entertainment Industry
Entertainment Agencies & Bureaus, Motion Picture Producers & Studios, Music & Recording Industry, ...

Event Planning
Convention & Meeting Planning Services, Party Planners, ...

Movie Theaters

Movies, Videos, & DVDs
Video Production Services, ...

Music
Music Instruction, Music Stores, Musical Instruments Retail, ...

Night Clubs

Performing Arts
Entertainers, Live Theater, Orchestras, Symphonies, & Bands, ...

Sports Tickets
Concert Tickets, ...

Tours & Charters
Boat & Yacht Charters, Rental, & Leasing, Fishing Guides & Charters, ...

Video Game Rentals

Visual Arts
Art Supplies & Art Supply Stores, Artists & Art Studios, Professional Photographers, ...

Taxonomy User Interface

Hierarchical categories are suitable for:

- Content naturally categorizable:
  products, industries, government agencies, academic disciplines, scientific things, technologies

- Limited scope resulting in a relatively small controlled vocabulary

- Large volume and changing content (e.g. news feeds), resulting in multiple records per term (category) and the frequent need for new subcategories

- Automatic indexing/ auto-categorization, which puts documents into approximate categories

Taxonomy User Interface

Facets

- For serving faceted classification, which allows the assignment of multiple classifications to an object

- A "dimension" of a query; a type of concept

- Intended for searching with multiple terms in combination (post-coordination), one from each facet

- Can be for topics or for named entities, but generally not both

- Reflect the domain of content

Taxonomy User Interface

Facets suitable for:

- Structured data with discernable metadata fields or database records

- Homogeneous data with similar types of characteristics

Examples:

- Products (name, series number, category, size, color, price, etc.)

- People (name, job title, gender, birth year, location, department)

- Reports (title, author, subject, audience, document type)

Taxonomy User Interface

Search box

Ideally should:

- Support Boolean, truncation, wildcards, phrase search, and other "advanced" searching

- Search on the taxonomy, not merely on the full-text

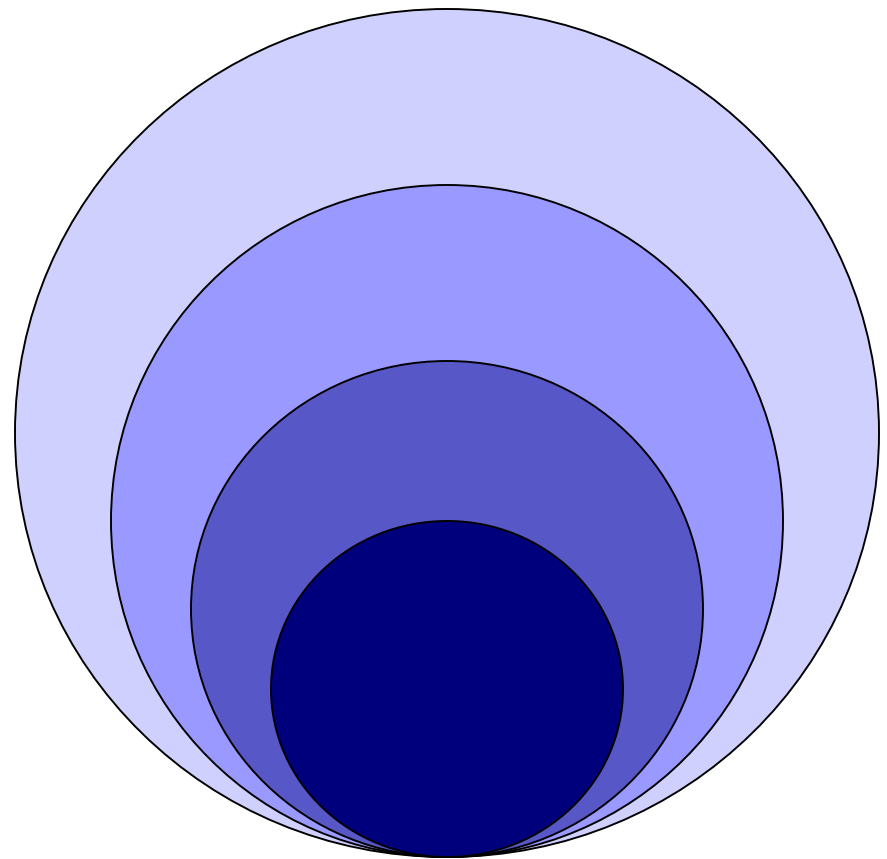- Be in addition, not instead of a taxonomy browse

Outline

1. Why have a taxonomy

2. Deciding on the taxonomy user interface

3. Deciding on the taxonomy scope

4. Resources

Taxonomy Scope

Deciding on the Taxonomy Scope

- Content scope
- Subject area scope
- Term types to include
- Usage
- Breadth and depth
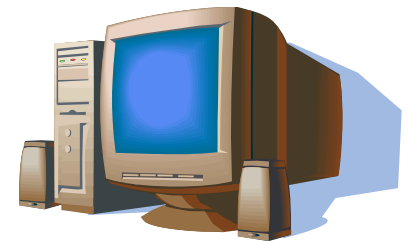
Taxonomy Scope

Content Scope

Content to be covered by the taxonomy can be limited to some or all of:

Content types:
- Employee handbooks
- Manuals
- Policies
- Standards
- Product data
- Transaction records
- Reports & white papers
- Marketing literature
- External publications

Document/file formats:
- Word documents
- PDF documents
- Presentations
- Spreadsheets
- Web pages
- Image files
- Videos
- Database records
- Other published formats

Taxonomy Scope

Subject Area Scope

Example: Area to be covered by an enterprise taxonomy can be limited to some or all of:

- Products and services (names, descriptions, features, user issues)

- Technology and technical terminology (science, research & development oriented)

- People and their areas of expertise

- Human resources

- All information in an enterprise

Taxonomy Scope

Scope of Term Types
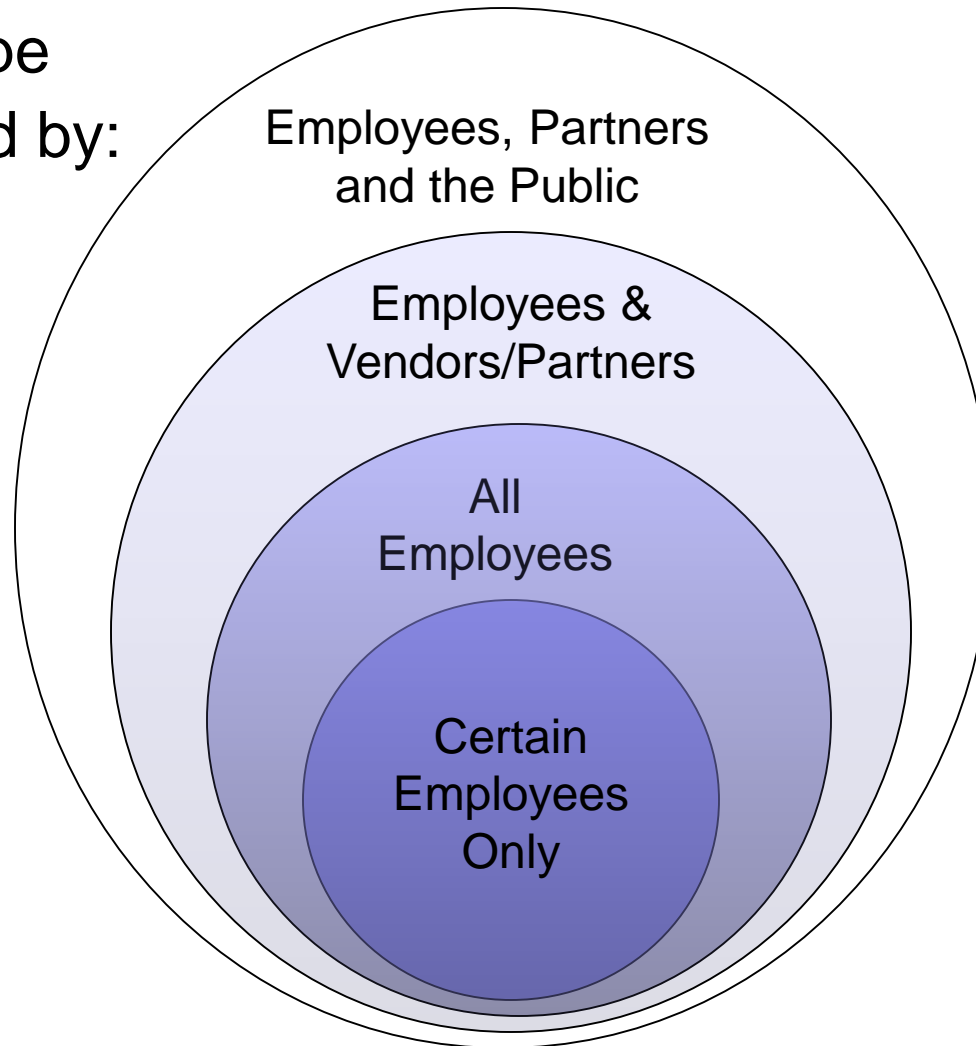 (could also be considered as metadata or "facets")

Term types to be included can be limited to some or all of:

- Subjects/Topics
- Person names
- Organization/company names
- Product names
- Locations
- Activities/Actions
- Job titles
- Price ranges

  etc.

Taxonomy Scope

Usage Scope
– to be used by:

Employees, Partners
and the Public

Employees &
Vendors/Partners

All
Employees

Certain
Employees
Only

•Public Web Site

•Extranet

•Intranet

•Intranet
restricted
sections

Outline

1. Why have a taxonomy
2. Deciding on the taxonomy user interface
3. Deciding on the taxonomy scope
4. Resources

Resources

Books

- Aitchison, J., Gilchrist, A. & Bawden, D. (2000). *Thesaurus Construction and Use: A Practical Manual* (4th ed.). Chicago, IL: Fitzroy Dearborn.

- ANSI/NISO Z39.19 (2005). *Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies.* Bethesda, MD: NISO Press.

- Lambe, Patrick. (2007). *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness.* Oxford, England: Chandos Publishing.

- Stewart, Darin L. (2008). *Building Enterprise Taxonomies.* Portland, OR: Mokita Press.

Resources

Organizations

- American Society for Indexing - Taxonomies and Controlled Vocabularies Special Interest Group
  http://www.taxonomies-sig.org

- Information Architecture Institute
  http:// iainstitute.org

- Special Libraries Association (SLA)
  http:// www.sla.org

- American Society of Information Science & Technology
  http:// www.asis.org

Resources

Discussion Groups

- Taxonomy Community of Practice
    http://finance.groups.yahoo.com/group/TaxoCoP

- Taxonomies & Controlled Vocabularies SIG, ASI
    http://finance.groups.yahoo.com/group/taxonomies

- Metadatalibrarians
    http://metadatalibrarians.monarchos.com

Resources

Workshops and Seminars

- Taxonomy Community of Practice Webinar phone calls
    - $50 each. Occasionally free vendor-sponsored calls.
    - Usually first Wednesday of the month, 1:00-2:00 pm EST
    - www.earley.com/TaxoCoP.asp

- "Taxonomies and Controlled Vocabularies"
  Simmons College Graduate School of Library and Information Science Continuing Education Program
    - Online: 5 weeks, $250, next session February 2009
    - Onsite: Boston, full day, $220, next session March 28, 2009
    - www.simmons.edu/gslis/continuinged/workshops

- Taxonomy Boot Camp conference
    - Information Today Inc.
    - Next: November 19-20, 2009, San Jose, CA
    - www.taxonomybootcamp.com/2009

Resources

Web Sites

- **Taxonomy Community of Practice Wikispace**, http://taxocop.wikispaces.com

- **Taxonomy Guide, Faculty of Information Studies, University of Toronto**
  http://plc.fis.utoronto.ca/tgdemo/default.asp

- **Construction of Controlled Vocabularies: A Primer**
  http://www.slis.kent.edu/%7Emzeng/Z3919/index.htm

- **Thesaurus Construction tutorial by Tim Craven**
  http://publish.uwo.ca/~craven/677/thesaur/main00.htm

- **Willpower Information: Publications on thesaurus construction and use**
  http://www.willpowerinfo.co.uk/thesbibl.htm

- **Taxonomy Watch Blog by Linda Farmer**, http://taxonomy2watch.blogspot.com

- **Earley & Associates** http://www.earley.com
- **Taxonomy Strategies** http:// www.taxonomystrategies.com

Questions

Heather Hedden

Hedden Information Management

98 East Riding Drive

Carlisle, MA 01741

www.hedden-information.com

heather@hedden.net

978-371-0822

978-467-5195 (mobile)