Indexing & Tagging:
Taxonomies & Folksonomies

National Federation of Advanced Information Services (NFAIS)
Improving the User Search Experience
October 13, 2010, Philadelphia, Pennsylvania

Heather Hedden, Taxonomy Consultant

Heather Hedden's Background

- Periodical database indexer for Information Access Company (IAC): Trade & Industry and PROMT
- Controlled vocabulary editor, IAC/Gale
- Freelance indexer and taxonomy consultant
- Taxonomy manager, Viziant and First Wind
- Continuing education instructor, Simmons GSLIS
- Author, *The Accidental Taxonomist*

What's the difference?

- **Cataloging** ➢ with a CV
- **Indexing** ➢ with a CV
- **Keywording**
- **Tagging**

Definitions

- **Controlled Vocabulary** – A controlled list of terms for concepts, usually with nonpreferred terms ("synonyms"). May or may not have structure and relationships between terms.
  - ➢ Broader and includes taxonomies.
- **Taxonomy** – A hierarchical structure of terms, which may or many not include nonpreferred terms.
  - ➢ Has popularly replaced "controlled vocabulary" as a broader concept, which may or may not be hierarchical.

Indexing

- Closed does not use CV; open does
- Trained, dedicated indexers
- Indexing guidelines and policies used
- Controlled vocabularies designed to support
  - Indexer-focused scope notes
  - Nonpreferred terms, displayed
  - Extensive hierarchical & associative relationships
  - Browsable, alphabetical display, type-ahead, etc.

Indexers and Index Terms

- Indexers are closest to the content
- They see new concepts and terminology
- They need methods of:
  - ➢ suggesting additional nonpreferred terms and relationships between terms
  - ➢ suggesting new terms to add to the controlled vocabulary
  - ➢ adding terms that supplement the controlled vocabulary

Levels of Vocabulary Management

1. Only controlled vocabulary used
   - Indexer suggests terms for approval prior to use.

2. Controlled vocabulary plus unapproved terms
   - Indexer may use terms prior to approval.

3. Controlled vocabulary plus keywords
   - Indexer supplements with terms never reviewed.

4. Controlled vocabulary plus shared keywords
   - Indexer supplements with terms never reviewed, but re-usable.

Levels of Vocabulary Management

1. Only controlled vocabulary used
   - Simple to manage and implement
     Indexers email taxonomist with suggestions
   - For small controlled vocabularies
   - For limited scope, relatively static content
   - For small indexing operations

Levels of Vocabulary Management

2. Controlled vocabulary plus unapproved terms

   ➢ Indexer may create "candidate", "unapproved," or "override terms" for immediate use, but also entered into the system for taxonomist review

   ➢ Unapproved indexed terms could convert to nonpreferred terms.

   ➢ Can be restricted to only certain types of terms (usually named entities) or all terms

   ➢ For large indexing operations, large controlled vocabularies, extensive content

   ➢ More technological complex to implement

Levels of Vocabulary Management

3. Controlled vocabulary plus keywords

   ➢ Indexer supplements controlled vocabulary indexing with any keywords entered into a separate field

   ➢ Keywords may be for new concepts, but are often for more specific concepts and names

   ➢ Keywords do not become part of the controlled vocabulary. Taxonomist may or may not look at them.

   ➢ Varies: (a) more indexing is with controlled vocabulary or (b) more with keywords and CV is only broad categories

   ➢ Less indexing technology, but more complex end-user retrieval options (3 types: taxonomy, keywords, freetext)

   ➢ Variants/synonyms are not controlled, redundant

Levels of Vocabulary Management

4. Controlled vocabulary and shared keywords

    ➢ Indexer supplements controlled vocabulary indexing with any keywords entered into a separate field

    ➢ Keywords do not become part of the controlled vocabulary, but are stored in another database.

    ➢ All keywords become immediately available for all indexers to use and reuse

    ➢ Indexers can browse the list of previously used keywords; redundancies are reduced, not eliminated

Folksonomy

- Shared keywords, created and shared by those "indexing"
- Indexing is *not* by indexers, but by end-users, consumers of the content.
  - ➢ Common people, the "folk."
  - ➢ "Tagging"
- Users might also contribute structure, creating (hierarchical) relationships
- Originally defined as:
  "user-created bottom-up categorical structure development with an emergent thesaurus"
  --*Thomas Vander Wal, July 2004*

Folksonomy: Who Contributes to it?

- Indexers
- Editors
- Employees
- End-user subscribers
- Any end-users

Folksonomy: Where used?

- **Public web sites of high volume content and users:**
    - ☐ Delicious (http://delicious.com)
    - ☐ Connotea (www.connotea.org)
    - ☐ Diigo (www.diigo.com)
    - ☐ Flickr (www.flickr.com)
    - ☐ LibraryThing (www.librarything.com)
- **Large enterprises that want to foster collaboration and innovation**
- **Subscription content providers?**

Social Tagging

- Also called collaborative tagging, social classification, social indexing

- Folksonomies plus Web 2.0/social networking features

- Social communities can be built around shared sets of popular content or popular tags

- Tags for popularity ratings

- Now moving into enterprises

Folksonomies/Social tagging

- **Advantages**
  - ☐ Reflects trends, up-to-date, can monitor change and popularity. Dynamic.
  - ☐ Cheaper and quicker than building and maintaining a taxonomy
  - ☐ Facilitates workplace democracy and the distribution of management tasks
  - ☐ Responsive to user needs

Folksonomies/Social tagging

- **Disadvantages**
  - ☐ Inconsistent – precision & recall deficiencies
  - ☐ Biased
  - ☐ Requires critical mass of involvement to be useful
  - ☐ Does not scale well to a large volume of content

Folksonomies/Social tagging

- **Solutions/trends:**
  - ☐ Some degree of vocabulary control
  - ☐ Applicable to certain areas of content, not all

- **Partial vocabulary control:**
  - ☐ Dual taxonomy & folksonomy system
    - ➢ Requires policy of taxonomy first
  - ☐ Single vocabulary system with some terms managed/edited, and some not (yet)
    - ➢ like a wiki with editors

Conclusions

- Different people often apply taxonomies or folksonomies.

- Taxonomies and Folksonomies may supplement each other.

- Technology facilitates the application of both.

- Users understand the distinction and can use both.

Questions/Contact

Heather Hedden

978-467-5195

heather@hedden.net

www.hedden-information.com

www.accidental-taxonomist.com