Taxonomy, Thesaurus, or Something in Between

Heather Hedden & Patricia Parsons
Indexing and Controlled Vocabulary Services, Cengage Learning
heather.hedden@cengage.com, patricia.parsons@cengage.com
Agenda

1. Comparisons of Taxonomies and Thesauri
   - Definitions, examples, and comparisons
   - Comparisons of suitability
   - When it’s a compromise / in-between

2. Cengage Learning case of thesaurus and taxonomies
   - Why we have both a thesaurus and taxonomies
   - How our taxonomy has thesaurus features (is in-between)
   - Benefits of discipline taxonomies
## Classic Comparison of Taxonomies and Thesauri

<table>
<thead>
<tr>
<th>Pick List</th>
<th>Synonym Ring</th>
<th>Authority File</th>
<th><strong>Taxonomy</strong></th>
<th><strong>Thesaurus</strong></th>
<th>Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguity control</td>
<td>Synonym control</td>
<td>Ambiguity control</td>
<td>Ambiguity control (Synonym control)</td>
<td>Ambiguity control</td>
<td>Ambiguity control (Synonym control)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hierarchical relationships</td>
<td>Hierarchical relationship</td>
<td>Semantic relationships</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Associative relationships</td>
<td>Classes</td>
</tr>
</tbody>
</table>
**Taxonomy**

A taxonomy is a controlled vocabulary consisting of preferred terms, all of which are connected in a hierarchy or polyhierarchy.

**Thesaurus**

A thesaurus is a controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators.

Its purpose is to promote consistency in the indexing of content objects…

*Sections 4.1 Definitions, 5.4.3 Taxonomy, and 5.4.4 Thesaurus*
## Further Comparisons

### Taxonomies
- All terms belong to a limited number of major hierarchies (or facets)
- May bend ANSI/NISO hierarchical rules.
- Supports classification, categorization, and concept organization. (Like Linnaean taxonomy.)
- Approach is a top-down navigation.
- Especially serving end-users when browsing.

### Thesauri
- All terms have relationships, but “hierarchies” can comprise as few as 2 terms.
- ANSI/NISO rules are strictly followed.
- Supports concept scoping, disambiguation, and relationships with similar concepts. (Like looking up in Roget’s.)
- Approach is term-centered and what terms are linked to/from it.
- Especially serving indexers/indexing.
Politics & Government
  . Domestic policy
  . . Agricultural policy
  . . Economic policy
  . . . Fiscal policy
  . . . Monetary policy
  . . Energy policy
  . . Health policy
  . . Social policy
  . Foreign policy
  . . Appeasement
  . . Bilateralism
  . . Foreign assistance
  . . . Foreign military assistance
  . . Foreign intervention

Economic policy
  SN Actions that governments take in the economics field.
  UF Finance policy
  BT Domestic policy
  NT Fiscal policy
  RT Economic conditions
  . Economic reform
  . Energy policy
  . Macroeconomics
## Comparisons: Suitability of a Taxonomy

<table>
<thead>
<tr>
<th><strong>Taxonomies for:</strong></th>
<th><strong>Thesauri for:</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>➢ Content/terms that naturally can be categorized</td>
<td>➢ Content/terms that are <em>not</em> easily categorized</td>
</tr>
<tr>
<td>➢ A subject area with defined scope and limits</td>
<td>➢ Multiple, overlapping subject areas or domains with diverse content.</td>
</tr>
<tr>
<td>➢ Browse navigations</td>
<td>➢ Highly specific terms for detailed indexing</td>
</tr>
<tr>
<td>➢ Non-expert users, who benefit from guidance of hierarchies</td>
<td>➢ Users who are subject-matter experts and will likely look for specific terms</td>
</tr>
<tr>
<td>➢ Relatively small collections of terms</td>
<td>➢ Large and/or constantly changing vocabulary</td>
</tr>
</tbody>
</table>
"Taxonomy"
Any kind of controlled vocabulary in a/an...
- enterprise, corporate setting
- content management system
- website navigation (e.g. ecommerce site)

"Thesaurus"
Any kind of controlled vocabulary...
- for indexing articles / literature retrieval databases
- used by librarians, indexers, or other information professionals
- that includes nonpreferred terms (synonym rings)
Compromises: Something in Between / Hybrid

1. You want features of both taxonomies and thesauri.
   Ex. Structured as hierarchies or facets and also related term (RT) relationships.

2. Issues of suitability are a mix.

3. Stakeholders or users prefer to call it one thing, even though the other is needed.
   Ex. Users referring to a “taxonomy” for research content that needs a thesaurus: varied and broad in scope and manually indexed.

And “thesaurus management” software is for both/either.
1. "Controlled vocabularies" for research/library products (Gale), comprising both:
   • Multiple name authority files
   • A Subject thesaurus, complies with ANSI/NISO standards.

2. "Taxonomies" for academic learning products

Why the difference?
What’s the difference?
1. **Using Twitter hashtags for information literacy instruction: the ubiquity of hashtags has opened the doors for the teaching of advanced searching concepts to a much wider audience than in past years.** Paige Alfonzo.


[ILLUSTRATION OMITTED] While terms such as "controlled vocabulary," "authority control," and "subject heading" are common nomenclature in a librarian's vernacular, these words are not necessarily resting on the tips of ...

From *Academic OneFile*.

2. **Access Innovations enriches content.**


Access Innovations (accessinn.com) says its Data Harmony Suite (dataharmony.com) enhances the organization of information by applying a taxonomy or thesaurus in total integration with natural language processing methods. ...

From *Academic OneFile*. 
Peroxisomes
Membranous sacs that contain oxidase enzymes are called peroxisomes. These enzymes help to digest fats and detoxify harmful substances.

Cytoskeleton
The cytoskeleton is the internal framework of a cell. It consists of microtubules, intermediate filaments, and microfilaments. The filaments provide support for the cells; the microtubules are thought to aid in movement of substances through cytoplasm.

Pinocytic Vesicles
Large molecules such as protein and lipids, which cannot pass through the cell membrane, will enter a cell by way of the pinocytic vesicles. The pinocytic vesicles form when the cell membrane folds inward to create a pocket. The edges of the pocket then close and pinch away from the cell membrane, forming a bubble or vacuole in the cytoplasm. This process, by which a cell forms pinocytic vesicles to take in large molecules, is called pinocytosis (pye-noh-sigh-TOH-sis) or "cell drinking."

Cilia and Flagella
Cilia and flagella are protrusions from the cell membrane. Cilia have short hair-like protrusions, whereas flagella have a singular tail-like protrusion. They are composed of fibrils that protrude from the cell and beat or vibrate. Cilia move materials across the surface of a cell. An example is the respiratory tract cells, which move the mucous-dust package from the respiratory tree to the throat. The sperm cell of the male has a flag.

from the carbohydrates, proteins, and fats we eat. Calories released from the decomposition of food are used to synthesize ATP. ATP is then available to be used for maintenance of cellular structure and function.

Cell Division
Cells divide for two purposes: growth or maintenance of cells in the human body (mitosis) and for reproduction (meiosis). In mitosis each cell carries a complete set of chromosomes, 46; however, in meiosis each cell carries only half of the chromosomes, 23.

Meiosis
Meiosis is the process of cell division of the sex cell or gamete. During meiosis, the ovum from the female and the spermatozoa from the male reduce their respective chromosomes by half, from 46 to 23. When fertilization (the union of the ovum and the spermatozoa) occurs, the two sex cells combine to form a simple cell called the zygote, with the full set of 46 chromosomes, 23 from each parent, Figure 3-3.

Media Link
View the Meiosis animation on CourseMate.

Mitosis
Cell division is divided into two distinct processes; the first stage is the division of the nucleus and the second stage is the division of the cytoplasm.
Objectives

After completing this chapter, you will be able to:

1. Define system development and list the system development phases
2. Identify the guidelines for system development
3. Discuss the importance of project management, feasibility assessment, documentation, and data and information gathering techniques
4. Discuss the purpose of and tasks conducted in each system development phase
5. Differentiate between low-level languages and procedural languages
6. Identify the benefits of object-oriented programming languages and application development tools
7. List other programming languages and application development tools
8. Describe various ways to develop webpages

System Development

Recall from the previous chapter that an information system is a collection of hardware, software, data, people, and procedures that work together to produce quality information. An information system supports daily, short-term, and long-range activities of users. The type of information that users need often changes. When this occurs, the information system must meet the new requirements. In some cases, members of the system development team modify the current information system. In other cases, they develop an entirely new information system.
1. One controlled vocabulary or two?
2. Taxonomy with thesaurus features
3. Benefits of discipline taxonomies
Our subject thesaurus for research products:

- Multiple, overlapping subject areas.
- Hierarchical (NT, BT) and Related Term (RT) relationships important for communicating meaning and disambiguation.
- Terms can be broad because "sub-headings" are applied to focus meaning.
- Developed and tuned for use by trained indexers, with plenty of RTs and cross references.
- Vocabulary is very large and growing.
Discipline taxonomies for learning content:

- Separate taxonomies for each discipline (subject area).
- Term choice/form reflects the discipline.
- Each relatively small (1,000 - 2,000 preferred terms).
- A single hierarchy for each discipline.
- Once built and tested, vocabularies will be slow-growing.
- Easier use for subject-matter experts & product developers.
- Could be implemented in an end-user browse interface.
- **Thesaurus-like:**
  - Have all standard thesaurus relationships
    - Broader Term (BT)/Narrower Term (NT)
    - Related Terms (RT) (but not as extensive)
    - Nonpreferred Terms (UF)
    - Scope Notes (SN)

- **Taxonomy-like:**
  - Each has a single top term and a limited number of second-level terms.
  - Sometimes bending ANSI/NISO hierarchical relationship rules at the top levels.
Health Insurance term in the Subject thesaurus

4 UF
14 direct NTs
(16 NTs total)
up to 2 levels deep
36 RTs (below line)

Health Insurance term in the Health discipline taxonomy

3 UF
13 direct NTs
(58 NTs total)
up to 4 levels deep
5 RTs (below line)
Hierarchical structure of the taxonomies for learning disciplines:

- Single hierarchy encourages focus on topics within the discipline.
- Easier to understand and review by non-taxonomist subject-matter experts.
- Gaps more easily identified.
- Enables inclusion of the very granular terms that can be typical in instructional material.
Questions?

**Heather Hedden**  
Senior Vocabulary Editor  
Cengage Learning  
20 Channel Center St., Boston, MA 02210  
heather.hedden@cengage.com  
www.accidental-taxonomist.com

**Patricia Parsons**  
Vocabulary Editor  
Cengage Learning  
500 Terry Francois Blvd, San Francisco, CA 94158  
patricia.parsons@cengage.com

www.cengage.com