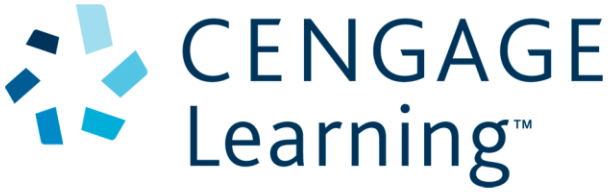
Taxonomy, Thesaurus, or Something in Between

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

Taxonomy Boot Camp
Washington, DC
November 5, 2014



Taxonomy, Thesaurus, or Something in Between

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

CENGAGE Learning™

Classic Comparison of Taxonomies and Thesauri

Pick List	Synonym Ring	Authority File	**Taxonomy**	**Thesaurus**	Ontology
Ambiguity control	Synonym control	Ambiguity control Synonym control	Ambiguity control (Synonym control) Hierarchical relationships	Ambiguity control Synonym control Hierarchical relationship Associative relationships	Ambiguity control (Synonym control) Semantic relationships Classes

Definitions – ANSI/NISO Z39.19-2005

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.

Display Examples

Taxonomy excerpt example

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

Thesaurus entry example

Economic policy

SN	Actions that governments take in the economics field.
UF	Finance policy
	Public policy (Economic)
BT	Domestic policy
NT	Fiscal policy
	Monetary policy
RT	Economic conditions
	Economic reform
	Energy policy
	Macroeconomics

Comparisons: Suitability of a Taxonomy

Taxonomies for:	Thesauri for:
➢ Content/terms that naturally can be categorized	➢ Content/terms that are *not* easily categorized
➢ A subject area with defined scope and limits	➢ Multiple, overlapping subject areas or domains with diverse content.
➢ Browse navigations	➢ Highly specific terms for detailed indexing
➢ Non-expert users, who benefit from guidance of hierarchies	➢ Users who are subject-matter experts and will likely look for specific terms
➢ Relatively small collections of terms	➢ Large and/or constantly changing vocabulary

Different Definitions – In use

"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.

Cengage Learning Case

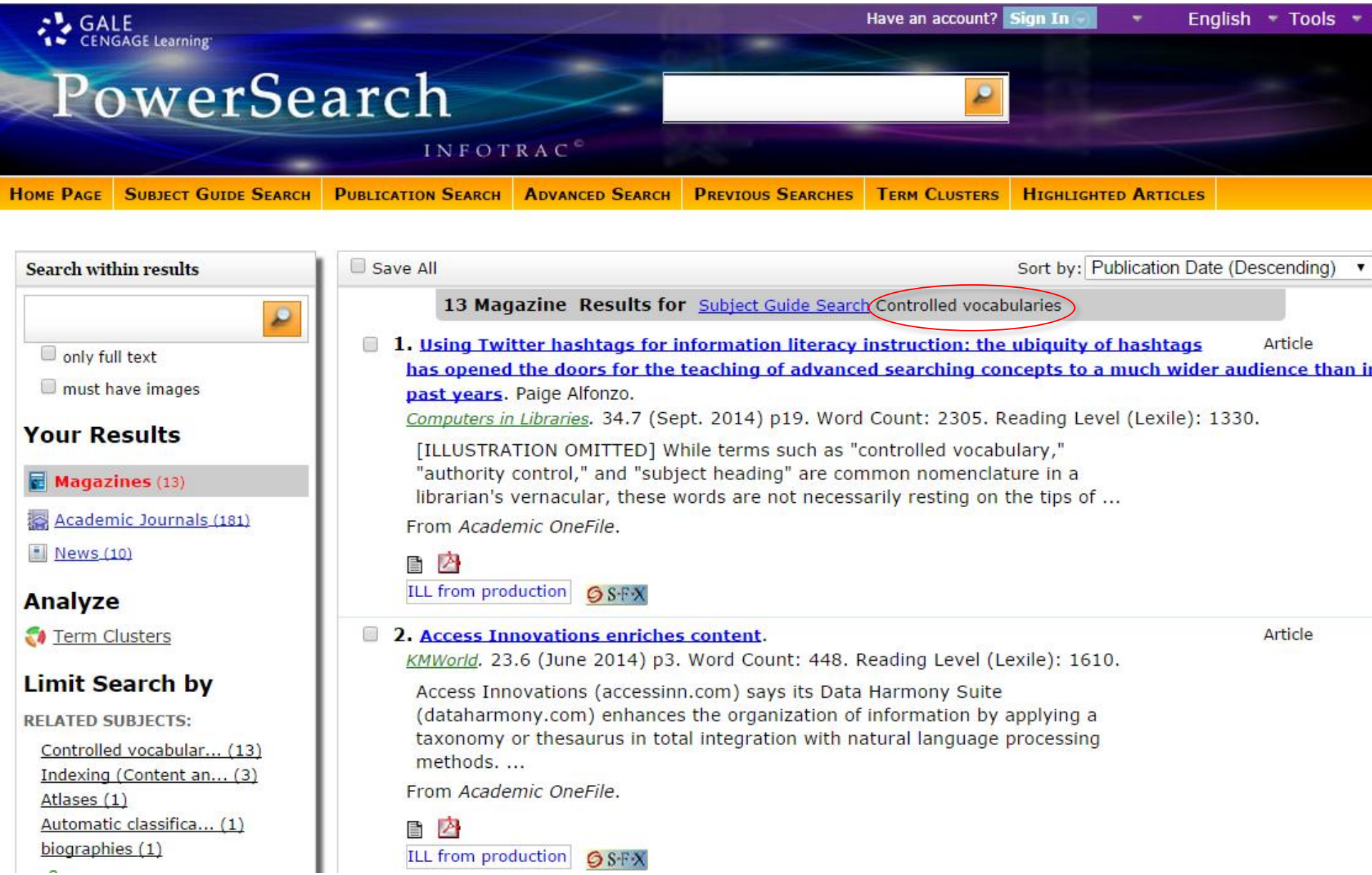
1. "Controlled vocabularies" for research/library products (Gale), comprising both:
   - Multiple name authority files
   - A Subject thesaurus, complies with ANSI/NISO standards.
2. "Taxonomies" for academic learning products

Why the difference?

What's the difference?

CENGAGE Learning

Cengage Learning Case: Research Products



Cengage Learning Case: Learning Products

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 fla-

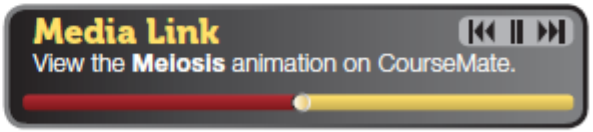
from the carbohydrates, proteins, and fats we eat. Calories released from the decomposition of food are used to synthesize ATP. ATP is then available to be used for maintenance of cellular structure and function.

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.



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.

CENGAGE Learning

Cengage Learning Case: Learning Products

© Piotr Marcinski / Shutterstock.com

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

CENGAGE Learning

Cengage Learning Case

1. One controlled vocabulary or two?
2. Taxonomy with thesaurus features
3. Benefits of discipline taxonomies

Cengage Learning Case: Choices

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.

Cengage Learning Case: Why Build New

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.

CENGAGE Learning™

Cengage Learning Case: Hybrid model

- **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.

CENGAGE Learning

Cengage Learning Case: What's the Difference

Health Insurance term in the Subject thesaurus

NT Capitated payment systems (Medical care) (Subjects)
NT Group health insurance (Subjects)
NT Hospitalization insurance (Subjects)
NT Long term care insurance (Subjects)
NT Managed care plans (Medical care) (Subjects)
NT Medicare (Subjects)
NT Medicare Part D (Subjects)
NT Medigap (Subjects)
NT National health insurance (Subjects)
NT Point of service plans (Subjects)
NT Preferred provider organizations (Medical care) (Subje
NT Prescription drug plans (Subjects)
NT Medicare Part D (Subjects)
NT Prospective payment systems (Medical care) (Subjects
NT Diagnosis related groups (Subjects)
NT Relative value scale payment systems (Medical care) (
NT Single payer system (Health care) (Subjects)
RT Accident insurance (Subjects)
RT Altmark Trans GmbH v. Nahverkehrsgesellschaft Altma
(Court Cases)
RT American Ass'n of Retired Persons v. EEOC 489 F.3d 558 (3d Cir. 2007) (Court Cases)
RT Coinsurance (Subjects)
RT Commission of the European Communities v. Ireland 2012 C.E.C. 229 (C.J.E.U.) (Court Cases)
RT Cooperative insurance (Subjects)
RT Cost shifting (Medical care) (Subjects)
RT Dental insurance (Subjects)
RT Diagnosis related groups (Subjects)
RT Disability insurance (Subjects)
RT Flexible spending accounts (Subjects)
RT Fraternal organizations (Subjects)
RT Germany. Sickness Insurance Act 1883 (Statutes)
RT Health care costs (Subjects)
RT Health care industry (Subjects)
RT Health care reform (Subjects)
RT Health insurance claims processing software (Subjects)
RT Health insurance exchanges (Subjects)
RT Health insurance industry (Subjects)
RT Health insurance tax credits (Subjects)
RT Health maintenance organizations (Subjects)
RT Health savings accounts (Subjects)
RT Individual health insurance mandates (Subjects)
RT Maternity benefits (Subjects)
RT Medicaid (Subjects)
RT Medical care (Subjects)
RT Medical law (Subjects)
RT Medical savings accounts (Subjects)
RT Medically uninsured persons (Subjects)
RT Mental health benefits (Subjects)
RT Multiple line insurance (Subjects)
RT Property and casualty insurance (Subjects)
RT Third party administrators (Insurance) (Subjects)
RT Value-based insurance design (Subjects)
RT Virginia v. Sebelius 728 F. Supp. 2d 768 (E.D. Va. 2010) (Court Cases)
RT Workers' compensation (Subjects)

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

Health Insurance term in the Health discipline taxonomy

NT Health care reimbursement systems (Taxon Health)
NT CMS reimbursement (Taxon Health)
NT Health insurance coverage (Taxon Health)
NT Health insurance history (Taxon Health)
NT Health insurance laws (Taxon Health)
NT ENTITY Health Care and Education Affordability Reconciliation Act o
NT ENTITY Patient Protection and Affordable Care Act (Statutes)
NT Health insurance terminology (Taxon Health)
NT Insurance claims process (Taxon Health)
NT Accept assignment (Taxon Health)
NT Assignment of benefits (Taxon Health)
NT Claims adjudication (Taxon Health)
NT Claims tracking (Taxon Health)
NT Electronic data interchange (EDI) (Taxon Health)
NT Filing insurance claims (Taxon Health)
NT CMS-1500 claim form completion (Taxon Health)
NT National provider identifier (NPI) (Taxon Health)
NT Payer responsibility determination (Taxon Health)
NT Insurance claim files maintaining (Taxon Health)
NT Insurance claim forms (Taxon Health)
NT CMS-1500 claim forms (Taxon Health)
... More levels exist.
NT UB-04 claim forms (Taxon Health)
NT Insurance forms (Taxon Health)
NT First report of injury form (Taxon Health)
NT Insurance claim forms (Taxon Health)
NT CMS-1500 claim forms (Taxon Health)
... More levels exist.
NT UB-04 claim forms (Taxon Health)
NT Insurance payments (Taxon Health)
NT Insurance plan types (Taxon Health)
NT Commercial insurance plans (Taxon Health)
NT Automobile, disability, and liability insurance (Taxon Health)
NT Medigap (Taxon Health)
NT Consumer-driven health plans (Taxon Health)
NT Fee-for-service insurance (Taxon Health)
NT Government health plans (Taxon Health)
NT Medicaid (Taxon Health)
NT Medicare (Taxon Health)
... More levels exist.
NT TRICARE and CHAMPVA (Taxon Health)
NT Managed care plans (Taxon Health)
NT Exclusive provider organizations (Taxon Health)
NT Health maintenance organizations (Taxon Health)
NT Integrated delivery systems (Taxon Health)
NT Point-of-service plans (Taxon Health)
NT Preferred provider organizations (Taxon Health)
NT Managed health care (Taxon Health)
NT Managed care organizations (Taxon Health)

NT Exclusive provider organizations (Taxon Health)
NT Health maintenance organizations (Taxon Health)
NT Integrated delivery systems (Taxon Health)
NT Point-of-service plans (Taxon Health)
NT Preferred provider organizations (Taxon Health)
NT Medicare administrative contractors (Taxon Health)
NT Workers' compensation (Taxon Health)
NT ENTITY Blue Cross and Blue Shield Association (Company Hut
NT Verifying insurance coverage (Taxon Health)
RT Billing and collections (Taxon Health)
RT Fee schedules (Taxon Health)
RT Health insurance specialists (Taxon Health)
RT Health service administration (Taxon Health)

3 UFs
13 direct NTs
(58 NTs total)
up to 4 levels deep
5 RTs (below line)

CENGAGE Learning™

Cengage Learning Case: Additional Benefits

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.

CENGAGE Learning™

Taxonomy, Thesaurus, or Something in Between

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

CENGAGE Learning