Smartlogic webinar

Semaphore Knowledge Modeling with The Accidental Taxonomist

Thursday, March 28th, 2019
12 – 1pm Eastern

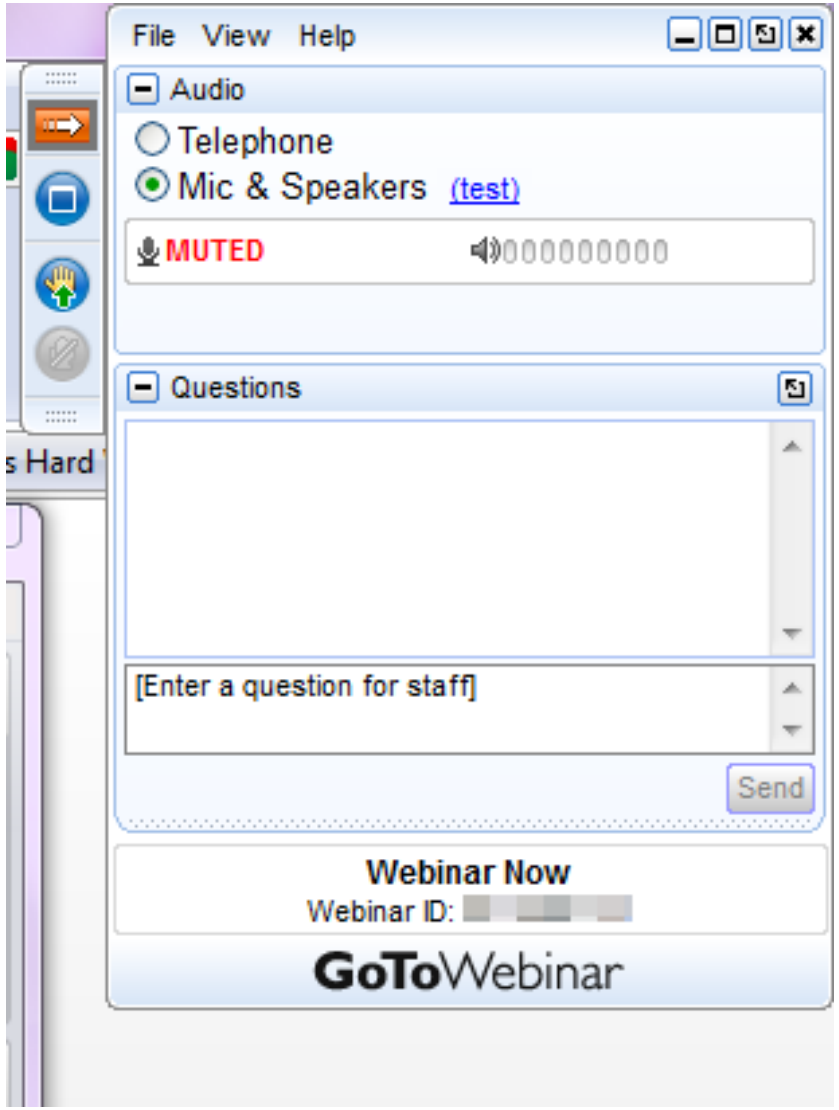
Speakers

Heather Hedden
Senior Vocabulary Editor
Gale, A Cengage Company

James Morris
Senior IS Consultant
Smartlogic

A few housekeeping items

- This webinar is in broadcast mode all participants are muted.
- Please put your questions in the GoToWebinar panel and we'll answer as many as we can in the Q & A session.
- This broadcast is being recorded – replay information will be sent to all registrants following the broadcast.



About Heather Hedden

- Senior Vocabulary Editor at Gale, a Cengage Company
- Previous taxonomy consultant
- Author of *The Accidental Taxonomist* (2010, 1st ed.; 2016, 2nd ed.), published by Information Today Inc.
- Instructor of taxonomy-creation online courses

Agenda

- What are Knowledge Models?
- Concepts and Labels
- Relationships
- Concept Classes
- Q & A

Knowledge Models

What is a Knowledge Model?

- Usually more than just a single controlled vocabulary / simple term list.
- It could be a thesaurus or ontology.
- It could be a taxonomy depending on how it is defined: a set of top-level hierarchies, not just one.
- It could be another kind of knowledge organization system: terminology, classification scheme, etc.

→ A knowledge model comprises the concepts, their labels, metadata,
and their relationships, and the rules for usage,
to tag/classify/index a content repository,
and to find and discover information in that content repository.

Knowledge Models

What is a Knowledge Model?

Same as a knowledge organization system (KOS). Not quite the same as a controlled vocabulary.

Term 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

Knowledge Models

“Standards” for Knowledge Models

Standard for *specifications*: SKOS model recommendation

- A World Wide Web (W3C) recommendation
- “A common data model for sharing and linking knowledge organization systems via the Web”
- https://www.w3.org/TR/skos-reference/

Standard for *best practice design*: Thesaurus standards/guidelines

- ISO or BS ISO 25964-1 Information and documentation - Thesauri and interoperability with other vocabularies
  Part 1: Thesauri for information retrieval [2011]
- ANSI/NISO Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies
  http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf

Knowledge Models

Knowledge models comprise

- Concepts
- Relationships between concepts

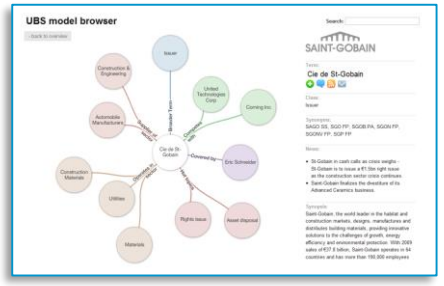
Concepts may

- Be organized into groupings, called *concept schemes*
- Be assigned categories, also called *concept classes*
- Have various names or *labels*
- Have additional descriptive values or *metadata*

Relationships may

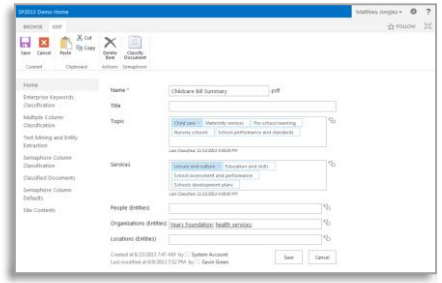
- Be of standard hierarchical or associative types, or of customized types

Semaphore in a Nutshell



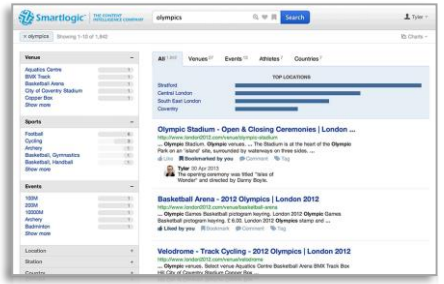
Build and manage semantic knowledge models

→ Simplify ingestion, development, customization and governance



Enrich, extract and harmonize

→ Enrich information assets with complete, consistent and precise metadata
→ Extract critical facts, entities and relationships for further processing
→ Harmonize different information sources for unified access



Apply semantics to your business problem

→ Leverage knowledge models to support investigative analytics
→ Enable knowledge discovery - Contextual metadata-driven search and navigation for a rich user experience
→ Automate manual processes for higher precision

Semaphore delivers these capabilities at an enterprise scale

Concepts and Labels

Knowledge models comprise:

Terms, Values, Nodes, Topics, Descriptors, Identifiers, Individuals, Instances, etc.

Best to treat them as “concepts.”

- A concept is a unique, unambiguous entity in a knowledge model, with its own definition and usage.
- The same concept may have multiple names, and the same name/word may refer to multiple concepts, so the focus should be on concepts, not names/words/terms.
- Concepts are tagged/indexed/assigned to content items.
- It should be clear to both those tagging/indexing and those browsing and searching for content what the concept means.

Concepts and Labels

Labels for concepts

- A concept has a single preferred label (in each language).
- The preferred label is what displays in hierarchical views or other visualizations of the knowledge model with relationships.
- A concept may have any number of alternative labels, which can match searches and aid users in finding the concept, but do not display.

Concepts and Labels

Choosing the preferred label wording

Choosing between two “synonyms”:

- **Doctors** vs. **Physicians**
- **Movies** vs. **Motion pictures**
- **Cars** vs. **Automobiles**

Consider:

- Wording of labels most likely looked up by the intended users/audience, especially in a browsed display
- Enforcing organizational/enterprise controlled vocabulary
- Conforming to academic or professional standards
- Consistency in style throughout the knowledge model
- Wording with in the documents/content indexed
- A meaning that is sufficiently broad

Concepts and Labels

Label format and style

- Consistent capitalization: lower case or initial capitalization; not title caps
  Corporate finance; **corporate finance**; *not* **Corporate Finance**
- Single words or multi-word phrases
- Nouns or noun phrases
- Adjectives alone can be concepts/labels only in special circumstances (e.g. facets for attributes).
- Countable nouns are usually plural.
- Parenthetical qualifiers may be used for disambiguation, not modification.
- Avoid inversions in labels (e.g. noun, adjective).

- Both preferred and alternative labels should follow the same style.
- Document specific style policies.

Concepts and Labels

Concepts in multiple languages

- The concept model is especially suited for multilingual vocabularies.
- A concept has a single preferred label in each language and alternative labels in each language.
- ISO 2-letter language codes may indicate the language.
- Preferred labels should be exact translations of each other.
- Alternative labels do not link to translations of each other, so may vary for each language.

Concepts and Labels: Alternative Labels

Alternative labels

- Defined: Approximately synonymous words or phrases to refer to an equivalent concept, for the context of the knowledge model and the set of content.
- Purpose: To capture different wordings of how different people might describe or look up the same concept or idea.
  - Differences between that of the author and the user/reader
  - Differences between that of the indexers and the end-users
  - Differences among different users/readers
- Serving as “multiple entry points” to look up and retrieve the desired content
- Enabling consistent indexing/tagging

Concepts and Labels: Alternative Labels

Examples from Gale Subject Thesaurus

Conflict management
- *Conflict resolution*
- *Managing conflict*

Wills
- *Codicils*
- *Last will and testament*
- *Testaments (Wills)*

Influenza
- *Flu*
- *Grippe*

Movies
- *Cinema*
- *Films (Movies)*
- *Motion pictures*
- *Movie genres*

Telecommunications industry
- *Communications industry*
- *Digital transmission industry*
- *Interexchange carriers*
- *Telecommunications services industry*
- *Telephone holding companies*
- *Telephone industry*
- *Telephone services industry*

Environmental management
- *Adaptive management (Environmental management)*
- *Environmental control*
- *Environmental stewardship*
- *Natural resource management*
- *Stewardship (Environmental management)*

Piano music [no alternative labels]

Concepts and Labels: Alternative Labels

When to use alternative labels

Not needed:

- If a very small, browsable taxonomy, where all can be seen or easily scrolled to (such as in facets) *and* tagging is manual.

Needed:

- If knowledge model is too large to be all seen in one view with minimal scrolling.
- If knowledge model will be searched upon and not just browsed.
- If automated indexing/auto-classification/auto-categorization is implemented.
- Whether it's called a taxonomy or thesaurus does not matter.

Concepts and Labels: Alternative Labels

Guidelines for creating alternative labels

- A concept may have any number of (multiple) alternative labels, or it may have no alternative labels.
- An alternative label is associated with a single concept, points to only a single preferred label.
- In implementations, alternative labels may be displayed to the end-user or they may not be.
- In implementations, alternative labels may point (re-direct) to the preferred label, or they can point directly to the content.

Concepts and Labels: Alternative Labels

Sources for alternative labels

- Same sources as for concepts and preferred labels
  - Survey/audit of the content and terms used
  - Search query logs and other internal usage data
  - External sources: websites, Wikipedia, other taxonomies and controlled vocabularies, book tables of contents, etc.
- Creative changes of labels - only after verification of alternative label usage in search on the content repository

Concepts and Labels: Alternative Labels

Types of alternative labels

- synonyms
- quasi-synonyms
- variant spellings
- lexical variants
- foreign language names
- acronyms/spelled out
- scientific/popular names
- antonyms (for characteristics)
- older/current names
- phrase variations (in print)
- narrower concepts that are not preferred

Concepts and Labels: Alternative Labels

Types of alternative labels

- synonyms: Cars / Automobiles
- quasi-synonyms: Politics / Government
- variant spellings: Taoism / Daosim; Email / E-mail
- lexical variants: Selling / Sales; Hair loss / Baldness
- foreign language names: Ivory Coast / Côte d'Ivoire
- acronyms/spelled out: GDP / Gross domestic product
- scientific/popular names: Neoplasms / Cancer
- antonyms (for characteristics): Flexibility / Rigidity
- older/current names: Near East USE Middle East
- phrase variations (in print): Unions, labor USE Labor unions
- narrower concepts that are not preferred: Genetic engineering USE Biotechnology

Concepts and Labels: Metadata

Notes for concepts

- Concepts may have notes.
- If utilized, not all concepts need notes.
- May have multiple types/purposes of notes.
- For searcher, indexer, or both.
- Basic standard note is: **Scope Note** (SN)
- Others: **History Note**
  Indexer Note
  Usage Note
- Definitions are possible, but not as practical.

Concepts and Labels: Metadata

Additional attributes for concepts

- Additional data may be free text or controlled.
- Additional data may be a field to sort upon.
- Typically used for name entity concepts, not subjects

Examples:

- *For Companies*: address, industry code, private/public status
- *For Person names*: title/occupation, birth date, nationality
- *For Products*: part number, price, market, intro date
- *For Places*: latitude and longitude

Semaphore Demo

Relationships

Types of relationships between concepts

1. Hierarchical: Broader concept / Narrower concept
2. Associative: Related concept
3. Customized relationships: More specific types of hierarchical or associative

Relationships are reciprocal between concepts.

- Broader concept, Narrower concept, Related concept are *relationship* types, not *concept* types.
- More accurately: “has broader concept,” “has narrower concept,” and “has related concept”

Relationships

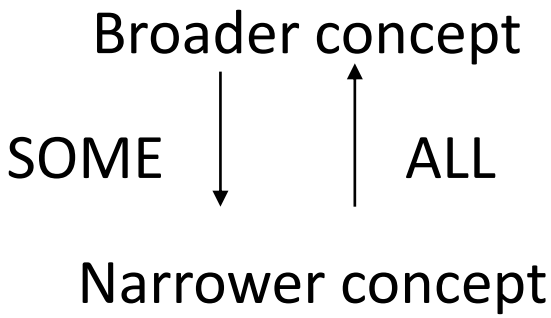
Hierarchical relationships

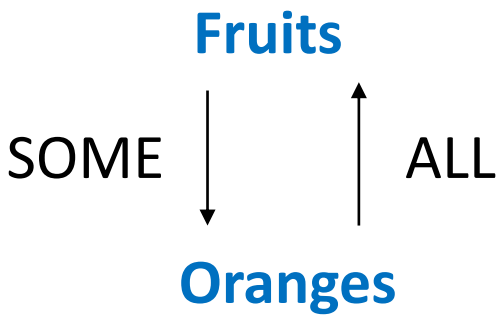
- Broader-narrower / generic-specific / topic-subtopic / parent-child
- Required feature of most knowledge models, including both thesauri and taxonomies
- Concepts usually have more than one narrower concept, unless they are the most specific concepts in the vocabulary.
- On occasion, a concept may have more than one broader concept, referred to as polyhierarchy.

Relationships

Hierarchical relationships

Reciprocal (bi-directional) relationships, but asymmetrical





Fruits *has narrower concept* **Oranges**

Oranges *has broader concept* **Fruits**

Three types:

1. Generic – Specific
2. Generic – Named entity instance: Common noun – Proper noun
3. Whole – Part

Relationships

Hierarchical relationships: Generic – Specific

Category or class
– members
– more specific types

Narrower concept “is a”
or “are a kind of”
broader concept.

“has broader” = “is/are a kind of”

Dogs *has narrower* **Puppies**
Puppies *has broader* **Dogs**

Financial services *has narrower* **Investment services**
Investment services *has broader* **Financial services**

Romance languages *has narrower* **Italian language**
Italian language *has broader* **Romance languages**

Relationships

Hierarchical relationships: Generic – Instance

Common noun – Proper noun

Narrower concept “is a” broader concept.

“has broader” = “is a”

Not used as often, since named entities are often in separate concept schemes.

Smartphones *has narrower* **Samsung Galaxy**
Samsung Galaxy *has broader* **Smartphones**

Rivers *has narrower* **Nile**
Nile *has broader* **Rivers**

Festivals *has narrower* **Oktoberfest**
Oktoberfest *has broader* **Festivals**

Relationships

Hierarchical relationships: Whole – Part

Concept or entity
– part
– sub-entity

France *has narrower* **French Alps**
French Alps *has broader* **France**

Narrower concept “is in” broader concept.

Gastrointestinal system *has narrower* **Stomach**
Stomach *has broader* **Gastrointestinal system**

“has broader” = “is in”

U.S. Congress *has narrower* **U.S. Senate**
U.S. Senate *has broader* **U.S. Congress**

Must be an integral part that cannot be taken out.

Relationships

Associative relationships

- Suggestions to the user of possible related concepts of interest
- Like *See also* in an index
- Common feature of knowledge models, required of thesauri, optional in taxonomies
- Symmetrical bi-directional relationship
- Between concepts within the same hierarchy or in different hierarchies

Relationships

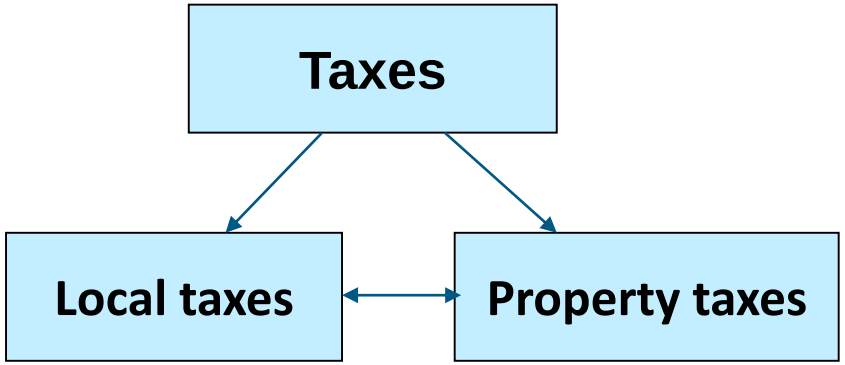
Associative relationships

Between concepts within the same hierarchy

1. Having a shared broader concept (as siblings) *and* overlapping meaning

   Required associative relationships, according to thesaurus standards

Example:



Local taxes *has related* **Property taxes**
Property taxes *has related* **Local taxes**

Relationships

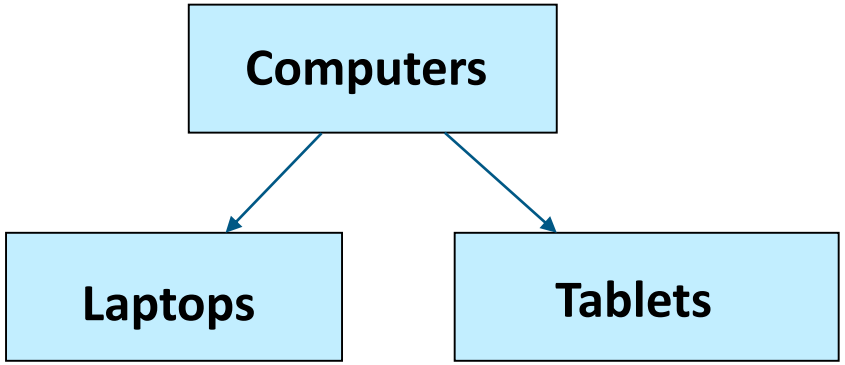
Associative relationships

Between concepts within the same hierarchy

2. Having a shared broader concept (as siblings) *without* overlapping meaning

   While not incorrect, better style is to avoid adding such associative relationships.

Example:

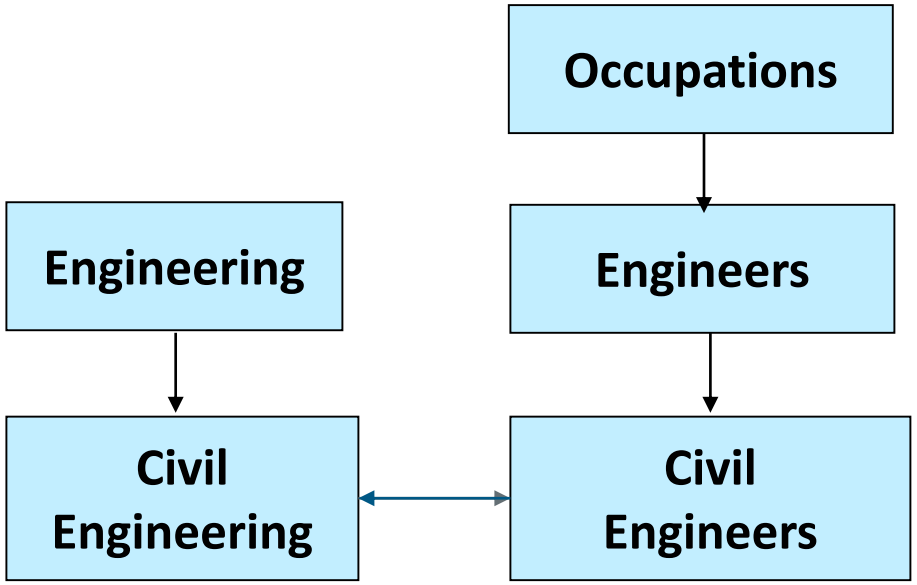


Not related concepts

Relationships

Associative relationships

Between concepts in different hierarchies



Civil Engineering *has related* **Civil Engineers**

Civil Engineers *has related* **Civil Engineering**

Relationships

Specific/customized relationships

- Relationships containing meaning: “semantic.”
- Variations on associative or hierarchical relationships, but usually associative.
- Reciprocal, but asymmetrical, or directional, not plain *has related*.
- Specific enough to convey the necessary meaning, but not uniquely specific.
- Relationships are between concepts of different types, across different designated categories or *concept classes*.
- Taxonomist/knowledge modeler defines the relationships, their names and abbreviations, and the classes.
- A defining characteristic of ontologies.

Relationships

Specific/customized relationships

Sample variations on the associative relationship (*related concept*):

Has produced the work (WRK) / Created by (CRE)

Twain, Mark WRK **The Adventures of Tom Sawyer**

The Adventures of Tom Sawyer CRE **Twain, Mark**

Produces the product (PRD) / Is manufactured by (MAN)

Apple Inc. PRD **iPod**

iPod MAN **Apple Inc.**

Has member affiliation with (AFF) / Has members (MEM)

Saudi Arabia AFF **OPEC**

OPEC MEM **Saudi Arabia**

For treating (TRE) / Can be treated with the drug (DRUG)

ACE inhibitors TRE **Hypertension**

Hypertension DRUG **ACE inhibitors**

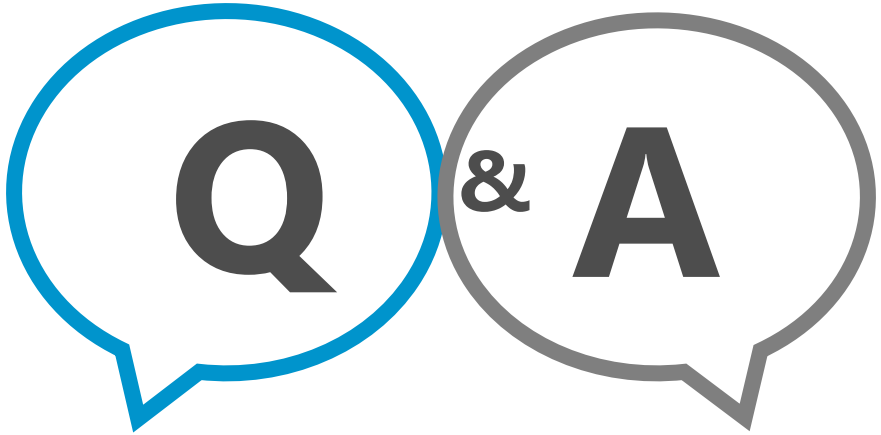
Concept Classes

A knowledge model may have concept classes

- Concept classes may match a concept scheme, but do not have to.
- Concept classes are needed when creating customized semantic relationships, as certain relationships are defined between certain concept classes.
- Examples: Organizations, Product types, Person types, Locations
- Concept classes may be used for other purposes in traditional taxonomies or thesauri, such as designating concepts for different:
  - audiences, internal or external
  - market segments
  - user interface displays
  - concept metadata field requirements

Semaphore Demo

Q&A



Thanks for joining the webinar

Thank You

SMARTLOGIC – AMERICAS
111 N MARKET ST.
SAN JOSE, CALIFORNIA, 95113
TEL: +1 408 213 9500

SMARTLOGIC – EUROPE, MIDDLE-EAST AND AFRICA
200 ALDERSGATE
LONDON, EC1A 4HD
TEL: +44 203 176 4500

WWW.SMARTLOGIC.COM
INFO@SMARTLOGIC.COM

Heather Hedden
Senior Vocabulary Editor
Gale, A Cengage Company
Boston, MA
heather.hedden@cengage.com

www.cengage.com, www.gale.com

www.hedden-information.com
https://accidental-taxonomist.blogspot.com
heather@hedden.net