Selecting Software for Taxonomy, Thesaurus and Ontology Management

Taxonomy Boot Camp London
15 October 2019

Presented by
Heather Hedden
Hedden Information Management
About Heather Hedden

- Taxonomy consultant
  - Independent, through Hedden Information Management
  - Previously as an employed and contract consultant

- Former staff taxonomist
  - At various companies: Gale/Cengage Learning, Viziant, First Wind

- Instructor of online and onsite taxonomy courses
  - Independently through Hedden Information Management
  - Previously at Simmons University - Library & Information Science School

- Author of *The Accidental Taxonomist* (2010, 2016, Information Today, Inc.)
Background to Software

- Taxonomy, thesaurus, or ontologies - The distinctions are blurred. Most software enables the creation of a combination: taxonomy/thesaurus or taxonomy/thesaurus/ontology
  - “Ontology” may be variously defined.

- Resources to research software are lacking.
  - Market is too small and specialized to be followed by industry analysts
  - Web lists are miscellaneous taxonomy-related tools or out-of-date

- Excel suffices for flat term lists (such as for facets), and small hierarchical taxonomies, but not for the complexities of large taxonomies, thesauri, ontologies, taxonomy-terminologies, or multilingual vocabularies.

- Software tools enforce/support standards, but not all the same standards: thesaurus (ANSI/NISO or ISO) and/or ontology (SKOS/RDF, OWL)
Background to Software

Types of software used for vocabulary management

- Spreadsheet software (Excel)
- Dedicated thesaurus/ontology management software
- Taxonomy creation & editing module of a content management, document management, digital asset management, collaborative software (SharePoint)
- Taxonomy creation & editing module of auto-classification (automated indexing) software
- Vertical market software for creating classification structures
- Proprietary programs developed in-house in organizations with large or core taxonomy management needs
## Background to Software

### Types of software used by taxonomists

<table>
<thead>
<tr>
<th>Type of Software</th>
<th>Percentage</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other commercial software that is not intended for taxonomies (such as a word processor, spreadsheet, or database management)</td>
<td>30.5%</td>
<td>43</td>
</tr>
<tr>
<td>An internally developed taxonomy/thesaurus management system</td>
<td>25.5%</td>
<td>36</td>
</tr>
<tr>
<td>Commercial dedicated thesaurus/taxonomy/ontology management software</td>
<td>22.7%</td>
<td>32</td>
</tr>
<tr>
<td>Commercial software, of which taxonomy management is a feature, module or component</td>
<td>12.1%</td>
<td>17</td>
</tr>
<tr>
<td>Open-source ontology/taxonomy management software</td>
<td>9.2%</td>
<td>13</td>
</tr>
</tbody>
</table>

Results of author survey for *The Accidental Taxonomist*, 2nd ed., conducted May 2015
# Background to Software

![Knowledge Organization System - Complexity Diagram](image)

## Knowledge Organization System - Complexity

<table>
<thead>
<tr>
<th>Term List</th>
<th>Synonym Ring</th>
<th>Authority File</th>
<th>Taxonomy</th>
<th>Thesaurus</th>
<th>Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguity control</td>
<td>Synonym control</td>
<td>Ambiguity control</td>
<td>Ambiguity control</td>
<td>Ambiguity control</td>
<td>Ambiguity control</td>
</tr>
<tr>
<td>Synonym control</td>
<td>Synonym control</td>
<td>Synonym control</td>
<td>(Synonym control) Hierarchical relationships</td>
<td>Synonym control</td>
<td>(Synonym control)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></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>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Scope notes</td>
<td>Attributes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Linked data)</td>
</tr>
</tbody>
</table>
Standards Supported by Software

Standards for thesaurus design and best practices
- ISO 25964-1

Standard for taxonomy or thesaurus modeling and interoperability
- SKOS (Simple Knowledge Organization System), RDF

Standards for ontology modeling and interoperability
- OWL, RDF, RDF Schema (RDFS)

Standards for interoperability (import/export formats)
- RDF serialization formats (RDF/XML, Turtle, N-Triples, N-Quads, JSON-LD, RDF/JSON, N3)
- CSV (comma separate values)
- ZThes

© 2019 Hedden Information Management
Feature of enforcing thesaurus design standards (ISO or ANSI/NISO)

- Preferred terms (preferred labels) must be unique; no duplicates
- A nonpreferred term (alternative label) can point to only one preferred term (concept)
- A pair of terms (concepts) can be either hierarchically (broader/narrower) or associatively (related) linked to each other, but not both.
- Hierarchical relationship logic extends:
  - A is narrower to B, and
  - B is narrower to C,
  - C cannot be narrower to Term A.
Standards Supported by Software

Following thesaurus standards does not require the thesaurus model.

**Thesaurus model:**

**Terms**

Of two *types*:

1. Preferred terms
2. Nonpreferred terms

- Separate terms linked to each other with the Equivalence relationship.
- Hierarchical and associative relationships are between preferred terms.
- Content is indexed/tagged to preferred terms only.

**SKOS model:**

**Concepts**

Which each *have*:

- Preferred labels
- Alternative labels

- Labels are attributes of a single concept.
- Hierarchical and associative relationships are between concepts.
- Content is indexed/tagged to concepts as described by preferred labels.
Software Features

Thesaurus software basic features (in addition to supporting thesaurus standards)
Software can:

- Maintain terms/concepts and their relationships
  - as reciprocals
  - when renaming, merging, subsuming, or deleting terms
- Support controlled variants/NPTs/synonyms/alternative labels
- Support notes/definitions and other attributes for terms
- Automatically assign unique identifier numbers
- Manage categories or classes for terms/concepts
- Manage candidate and approved terms; term creation and update dates
- Import lists of terms and their relationships
- Export data in recognizable format for importing into content management system, indexing, or search system
Additional features to consider and compare:

- **Editing features**
  - single-step new term & relationship creation
  - single-step branch (term and narrower terms) moving
  - drag & drop relationship adding

- **User-defined (customizable) relationships (if not an ontology)**

- **User-defined note fields and term attributes**

- **Multilingual taxonomy support**

- **Automated mapping between vocabularies**

- **Connectors to SharePoint, content management systems, and enterprise search systems**

- **Candidate term extraction from corpus documents**

- **Auto-categorization or machine-aided indexing add-on module**

- **Support for URIs and linked data**
Software Features

Additional points of comparison:

- Interface design (default view) and ease of use
- Term searching
- Multiple taxonomy display options
- Generating customized reports
- Importing and exporting formats
- Quality control methods and customization - flags, alerts, or reports
- Multi-user management, including permission levels

Generic points of comparison:

- Price/pricing model, number of users, platform/hosting, administration requirements, technical support, training, community, additional products and services
Software Products

Taxonomy/thesaurus software – thesaurus model (but can export as SKOS):
- Data Harmony Thesaurus Master
- MultiTes
- Synaptica KMS

Taxonomy/thesaurus/ontology software – SKOS/RDF model:
- Mondeca ITM
- Poolparty
- Smartlogic Semaphore
- Synaptica Graphite
- TemaTres
- TopBraid EDG
- VocBench

Dedicated ontology software
- Protégé
- TopBraid Composer
Software Products

Product summaries and screenshots

Single-user Windows desktop software
- MultiTes (also multi-user version)

Multi-user client-server or web-based commercial software
- Data Harmony Thesaurus Master
- Mondeca ITM
- Poolparty
- Semaphore
- Synaptica KMS, Synaptica Graphite
- TopBraid EDG

Free, open-source software
- VocBench
Software Products: MultiTes Pro

Multisystems (Miami, FL)
www.multites.com

- Single product independent vendor since 1983.
- Windows single user $295 (multi-user and enterprise packages also available)
- Web/cloud-based option: $4950/year per thesaurus for 20 accounts
- Thesaurus model ANSI/NISO Z39.19 based
- Supports user-defined relationships, classes, and notes; multilingual thesauri
- Imports delimited text. Outputs text, HTML, XML, SKOS/RDF, and CSV
- Add-on products: web development kit, enterprise development kit
- Free 1-month downloadable trial and online video tutorials
### Bird's-eye views

**General**
- CN: For the class of engravings, lithographs, and
- FUN: Formerly TGM term & TGMII term.
- SN: Graphic representations of scenes portrayed as if viewed

**Notes**

**Relationships**
- UF: Balloon views
- UF: Views, Bird's-eye
- BT: Views
- NT: Bird's-eye view prints
- RT: Aerial views
- RT: Maps
- RT: Panoramic views
- RT: Projections

**Term** | **Status** | **Flag** | **Rel** |
--- | --- | --- | --- |
Billheads | A | BT | |
Bills of fare | A | US | |
Bird's-eye view prints | A | BT | |
Software Products: Synaptica

Synaptica Software LLC (Franktown, CO)
www.synaptica.com

Synaptica KMS (Knowledge Management System) – thesaurus model (since 1995)

Synaptica Graphite – SKOS ontology model on a linked data graph database (2018)

- Web browser-access, inside the firewall or hosted.
- Supports user-defined relationships, classes, and notes; multilingual vocabularies
- Features drag-and-drop editing, automatic term mapping
- Imports: CSV, Excel, XML (Zthes, RDF SKOS, RDF OWL). Exports also HTML, Word.
- Related add-on products: Indexing Management System (IMS), Text Analytics Platform (TAP), Image Annotation & Indexing, Linked Data Manager, SharePoint connector
- Online video tutorial for editing terms in Synaptica KMS
Synaptica Graphite
Software Products: Data Harmony Thesaurus Master

Access Innovations (Albuquerque, NM)
www.dataharmony.com

- Commercial software (originally used for indexing in-house) offered since 1998
- Multi-platform java-based (used on Windows, Mac, Solaris, Linux). Client software allows remote access. Also a web-hosted version.
- Thesaurus model ANSI/NISO Z39.19 based
- Separately or combined with M.A.I. (Machine Aided Indexer) as MAIstro.
- Related products: XIS (XML Intranet System), Inline Tagging, Search Harmony
- API connectors for SharePoint, MarkLogic, OpenText, Oracle, SAP
- Access Innovations also offers taxonomy creation services.
Software Products: Semaphore

Smartlogic Semaphore Ltd. (London, UK)

www.smartlogic.com

- Introduced in 2006.
- Supports SKOS, RDF ontology standard, and ISO 25964 thesaurus standard
- Imports/export CSV, XML (RDF SKOS, Turtle, N Triple), SQL databases, and MultiTes files
- Related products: Classification Server for automated classification; Ontology Service for a navigation system
- Download free 30-day trial: https://trial.smartlogic.com/S4Trials/
  (Sign in with a LinkedIn account.)
e-Commerce

Concept Class
No user class defined

Preferred Labels
Create a preferred label
e-Commerce

Alternative Labels
Create an alternative label
alternative label > Electronic Commerce

Metadata
Software Products: Mondeca Intelligent Topic Manager (ITM)

Mondeca S.A. (Paris, France)

https://mondeca.com/itm

- Introduced in 2008
- Supports SKOS vocabularies and OWL-standard ontologies
- Linked data feature
- SharePoint term store connector
- Visualization of hierarchies and relationships
- Exports to Excel, XML, RDF, SKOS, and Topic Maps
## Attributes

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>es</td>
<td>información/biblioteca, gestión</td>
</tr>
<tr>
<td>fr</td>
<td>information/bibliothèque, gestion</td>
</tr>
<tr>
<td>en</td>
<td>information/library management</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://skos.um.es/unescothes/C02010">http://skos.um.es/unescothes/C02010</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>in group</th>
</tr>
</thead>
<tbody>
<tr>
<td>→ UNESCO Thesaurus</td>
</tr>
<tr>
<td>→ Information management</td>
</tr>
<tr>
<td>→ Information and communication</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>fr</td>
</tr>
<tr>
<td>en</td>
</tr>
<tr>
<td>es</td>
</tr>
</tbody>
</table>

## Associations

**Broader-Narrower**
- information/library administration

**Related concepts**
- related: information/library management, library use promotion
Software Products: Poolparty

Semantic Web Company (Vienna, Austria)
www.poolparty.biz

- Introduced in 2009.
- Built on W3C Semantic Web standards: SKOS, RDF, OWL, SPARQL
- Installed server or web-hosted options
- Can link domain-specific thesauri to Linked Open Data
- Import/export formats: Excel, N3, N-Quads, Trix, Binary-RDF, MultiTes, RDF/XML, Turtle, N-Triples, RDF/JSON, Trig, JSON-LD, and Zthes
- Add on modules: Concept Tagging, Linked Data Management, Semantic Search, Text Mining & Entity Extraction, Classification, Data Analytics & Visualization
- Connectors for SharePoint, Drupal, WordPress, Confluence, Alfresco, FontoXML
- Download free 30-day trial: http://www.poolparty.biz/test-demo/thesaurus-server-entity-extractor
Software Products: TopBraid Enterprise Data Governance (EDG)

TopQuadrant Inc. (Raleigh, NC)
www.topquadrant.com/products/topbraid-edg-vocabulary-management/

Originally Enterprise Vocabulary Net (2010), now as a module of TopBraid EDG, introduced in 2016, or as stand-alone Vocabulary Management tool

- Web-based access to a Linux server installation, with browser access
- Based on a graph database
- Taxonomies in SKOS or SKOS-XL; ontologies based on SHACL or OWL
- Import/export formats Excel/CSV, XML, RDF/OWL
- Automatic creation of crosswalks between two vocabularies
- Video demos at: www.topquadrant.com/knowledge-assets/videos
Software Products: Free and Open Source Software

**Protégé** – Developed by the Center for Biomedical Informatics Research at Stanford University School of Medicine. [https://protege.stanford.edu](https://protege.stanford.edu)
- Dedicated ontology software; not so suitable for taxonomies/thesauri

**VocBench** – Developed by the Artificial Intelligence Research group at University of Tor Vergata, Rome [http://vocbench.uniroma2.it](http://vocbench.uniroma2.it)
- For OWL ontologies, SKOS(/XL) thesauri
- Introduced in 2010 for UN Food & Agriculture Organization’s AGROVOC thesaurus.
- Now funded by the EU European Commission's ISA² program. Current version 3.
- Can be installed on a web server or on a single desktop

**TemaTres** – *Originally* developed by the Library and Information Science program of the University of Buenos Aires [https://www.vocabularyserver.com](https://www.vocabularyserver.com)
- Available On-Premise on a web server; Software as a Service, or On-Demand
- Version 3.0, November 2017
- Uses SKOS model and supports ISO thesaurus standards
Software Products: Other Software

- **a.k.a. by Synercon** – information management software with taxonomy/thesaurus/ontology component; Australian company
- **Coreon** – taxonomy/thesaurus + terminology management; German company
- **Lucidia’s STAR/Thesaurus** – part of the CuadraSTAR (2008 acquisition) suite, software marketed at libraries, archives, and museums
- **Soutron Global** – library management system with thesaurus component
- **Unilexicon** – web hosted open source, but all vocabularies are open, too.
- **Wordmap** – offered by a consulting company (Earley Information Science), not their main focus
Software Evaluation

Investing time and effort in evaluating software is worth it, if…

- A taxonomy program underway
- You have sufficient corporate support and funding
- You have sufficient business support that you have subject matter experts involved in the process
- You’ve reached the point where you sense taxonomy management software is needed.

Not evaluating software risks…

- Buying more than you need
- Not buying enough features that you need
- Conflicting with IT infrastructure or other systems
Software Evaluation

Spec. & Select Process Blueprint
(courtesy of Marti Heyman)

Phase 0: Define your team
Phase 1: Business Requirements
Phase 2: Functional Requirements
Phase 3: Features Scorecard
Phase 4: Vendor short list
Phase 5: Live demos
Phase 6: Analyze scores
Phase 7: Vendor due diligence
Phase 8: Purchase
Phase 9: Implement
Software Evaluation

Sample business requirements

- Make the enterprise vocabulary accessible and available to a geographically dispersed set of taxonomy managers
- Make the enterprise vocabulary accessible to all applications that depend on it
- Ensure scalability of the enterprise vocabulary
- Low operating costs
- Unix environment compliant

Functional requirements

- Are based on specific used cases
- Need to consider technical restraints
- Need to consider what is truly necessary
<table>
<thead>
<tr>
<th>Relevant User Group</th>
<th>Feature</th>
<th>Priority</th>
<th>Vendor 1</th>
<th>Comments</th>
<th>Vendor 2</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (Content Mgrs, Taxonomy Mgrs, End Users)</td>
<td>General - Supports multiple Languages</td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>All (Content Mgrs, Taxonomy Mgrs, End Users)</td>
<td>General - ANSI/ISO Z39.19 compliant</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>All (Content Mgrs, Taxonomy Mgrs, End Users)</td>
<td>General - Windows operating system compliant</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Editing - Ability to select multiple terms.</td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Editing - Built in Spell checker</td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Editing - Prevent duplicate term entry within a category. Warn when duplicate is entered.</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Administration - terms should not be deleted by taxonomy managers, just changed to an inactive status.</td>
<td>1</td>
<td>0.5 business process</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>System</td>
<td>Relationship Mgmt - Poly-hierarchies (i.e. Ability for a single term to be in more than one category)</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Workflow - There should be an audit trail showing changes.</td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Workflow - roles should be used when defining taxonomy managers responsible for each workflow step. Ability to change the roles without developer assistance.</td>
<td>2</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>All (Content Mgrs, Taxonomy Mgrs, End Users)</td>
<td>Display - Ability to see full record of a term on the screen (relationships, attributes, comments, etc.)</td>
<td>1</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>All (Content Mgrs, Taxonomy Mgrs, End Users)</td>
<td>Reporting - Ability to export reports in a common format (CSV, XML, etc.)</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Import/Export - Import/Export terms, categories, taxonomies in a standard format (CSV, XML, etc.)</td>
<td>1</td>
<td>0.5 not intuitive</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Search - Search across all terms and all categories (taxonomy) and attributes.</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0.6 no wild cards?</td>
<td></td>
</tr>
<tr>
<td>Taxonomy Managers</td>
<td>Security - Ability to set security at a category (taxonomy) level.</td>
<td>2</td>
<td>0</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>All (Content Mgrs, Taxonomy Mgrs, End Users)</td>
<td>Help - On line help, quick reference guide</td>
<td>1</td>
<td>0.5 pdf guide</td>
<td></td>
<td>0 print manual</td>
<td></td>
</tr>
</tbody>
</table>

9.3 TOTAL  13.1 TOTAL
Software Evaluation

Spec. & Select Key Points

- **Requirements:**
  to purchase the software that meets your needs and expectations, you need a strong, clear, unambiguous definition!

- **Process Blueprint:**
  follow a tried and true path to success and ensure you have the data to explain your decisions

- Ensure stewardship of corporate funds
Questions/Contact

Heather Hedden
Taxonomy Consultant
Hedden Information Management
Carlisle, MA USA
+1 978-467-5195
www.hedden-information.com
accidental-taxonomist.blogspot.com
www.linkedin.com/in/hedden
Twitter: @hhedden