

---

# Tools for Taxonomies

---

Heather Hedden

Taxonomy Consultant

Hedden Information Management

---

# About Heather Hedden

- Taxonomy consultant, Earley & Associates
- Indexer, Hedden Information Management
- Instructor, Continuing Education, Simmons College Graduate School of Library & Information Science
- Formerly taxonomist at Viziant Corporation and Thomson Learning (Gale)
- Author, *The Accidental Taxonomist* (Information Today, Inc., May 2010)

---

# What are Taxonomy Tools?

- No authoritative industry list of taxonomy software
- “Taxonomy software” can mean different things
  - Auto-categorization vs. taxonomy management
- Existing Web lists are miscellaneous taxonomy-related tools or out-of-date
  - <http://taxocop.wikispaces.com/TaxoTools>
  - [www.taxotips.com/resources/tools](http://www.taxotips.com/resources/tools)
  - [www.searchtools.com/info/classifiers-tools.html](http://www.searchtools.com/info/classifiers-tools.html)
  - [www.willpowerinfo.co.uk/thessoft.htm](http://www.willpowerinfo.co.uk/thessoft.htm)

---

# Taxonomy Tool Types

1. Thesaurus/ontology management software
2. Other software with thesaurus/taxonomy modules
3. Auto-categorization/text mining software
4. Other software supporting creating taxonomies
  - ❑ mindmapping or concept modeling tools
  - ❑ Cardsorting tools
  - ❑ Web analytics

---

# Taxonomy Types & Software Types

1. Controlled vocabularies, authority files, or synonym rings
2. Hierarchical or faceted taxonomies
3. Thesauri
4. Ontologies

---

# Taxonomy Types & Software Types

1. Controlled vocabularies, authority files, or synonym rings
  - ❑ Support correspondence between term and synonyms (equivalencies)
  - ❑ No other relationships between terms
  - ❑ Optional additional data for terms
- *Spreadsheet software suffices*

---

# Taxonomy Types & Software Types

## 2. Hierarchical or faceted taxonomies

- ❑ Broader/narrower (parent/child) relationships between terms
- ❑ Optional synonyms (equivalencies) or additional data for terms
  
- *If very small and simple, spreadsheet software*
- *If including synonyms and other data, then use thesaurus management software*

---

# Taxonomy Types & Software Types

## 3. Thesauri

- ❑ Broader/narrower (parent/child) relationships between terms
- ❑ Associative (symmetrical reciprocal related/“See also”) relationships between terms
- ❑ Synonyms (equivalencies) for terms
- ❑ Other data for terms
  
- *Thesaurus management software*



---

# Taxonomy Types & Software Types

## 4. Ontologies

- ❑ Classes for terms
- ❑ Customizable semantic relationships between terms (hierarchical and associative), dependent on class
- ❑ Optional synonyms (equivalencies) for terms
- ❑ WC3 or RDF standard outputs
  
- *High-end thesaurus management software (with ontology support features)*
- *Ontology software*

## Tools Used by “Taxonomists”

Commercial dedicated thesaurus/taxonomy management software	38.1%	24
An internally developed thesaurus/taxonomy management system	23.8%	15
Commercial software, of which taxonomy management is a feature, module or component	15.9%	10
Other commercial software that is not intended for taxonomy creation	19.0%	12
Open-source ontology/taxonomy management software	3.2%	2

---

# Thesaurus Management Software Basics

- Maintains terms and their relationships (equivalencies, hierarchical, and associative)
  - As reciprocals
  - When renaming, merging, subsuming, or deleting terms
- Disallows invalid relationships (according to standards)
- Supports term notes and other attributes for terms
- Supports candidate/approved terms; includes term creation and update dates
- Generates reports in various thesaurus display formats (hierarchical, alphabetical)
- Exports data in interoperable formats for importing into a content management, indexing, search, retrieval system
- Supports thesaurus standards: ANSI/NISO Z39.19 or ISO 2788

---

# Thesaurus Management Software Feature Comparisons

- interface design and ease of use
  - multiple taxonomy display options
  - term searching
  - spell-checking
  - speed (limited mouse clicks) for repeated term and relationship additions
  - single-step new term & relationship creation
  - single-step branch (term and narrower terms) moving
  - drag & drop relationship adding
  - user-defined (customizable) relationships
  - user-defined term notes and term attributes
  - bilingual or multilingual taxonomy support
  - importing and exporting formats
  - connectors to enterprise search systems
-

---

# Thesaurus Management Software

Lower-end, single-user desktop software (may also have multi-user versions):

- MultiTes Pro (Windows)
- Cognatrix (Mac)
- One-2-One (Windows)
- TheW32 (Multi-platform)

Other tools have come and gone:

<http://www.willpowerinfo.co.uk/thessoft.htm>

---

# Thesaurus Management Software

## MultiTes Pro

Multisystems (Miami, FL)

[www.multites.com](http://www.multites.com)

- Since 1983. Hector Echeverria, president.
- \$295 single user; \$1295 for 5 users  
\$2495 for 10 users; \$3950 enterprise deployment
- Add-on products: web development kit, enterprise development kit
- Imports text; exports text, HTML (as a web page), XML, CSV
- Free limited-time downloadable trial and online tutorial
- Online discussion group for tips

Term	Status	Flag	Relationships	Notes	Categories	Translation
Account Management	A		USE(1)	-	-	ENG
Account Relationship	A		USE(1)	-	-	ENG
Accountability	A		BT(1), UF(2)	SN	Theme	ENG
Acquisition	A		USE(1)			ENG
Acquisitions	A		USE(1)			ENG
Actions	A		USE(1)			ENG
Adaptability	A		USE(1)			ENG
Adaptable	A		USE(1)			ENG
Adapting	A		USE(1)			ENG
Advancement	A		USE(1)			ENG
Advertise	A		USE(1)			ENG
Advertising	A		USE(1)			ENG
Agility	A		BT(1), U			ENG
Alliances	A		USE(1)			ENG
Anticipating Change	A		BT(1), U			ENG
Art of the Turnaround	A		USE(1)			ENG
Articulating Your Message	A		USE(1)			ENG
Assessing Performance	A		USE(1)			ENG
Automation	A		USE(1)			ENG
Behavior of Leaders	A		USE(1)			ENG
Being Accountable	A		USE(1)			ENG
Borrowing	A		USE(1)			ENG
Brand	A		USE(1)			ENG
Brand Management	A		BT(1), U			ENG
Brand strategy	A		USE(1)			ENG
Branding	A		USE(1)			ENG
Brands	A		USE(1)			ENG
Budget	A		USE(1)			ENG
Budget Management	A		USE(1)			ENG
Budget Planning	A		USE(1)			ENG
Budget Strategy	A		USE(1)			ENG
Budgeting	A		USE(1)	-	-	ENG
Budgeting & Costs	A		BT(1), UF(9)	SN	Theme	ENG
Budgets	A		USE(1)	-	-	ENG

**Budgeting & Costs**

Clipboard Edit Print Window

Record Details Multilevel Hierarchy 2-way Hierarchy Edit

**Budgeting & Costs**

- General
- Classification
  - FAC: Theme Theme
- Notes
  - SN: Managing expenses, costs, and budgets
- Relationships
  - UF: Budget
  - UF: Budget Management
  - UF: Budget Planning
  - UF: Budget Strategy
  - UF: Budgeting
  - UF: Budgets
  - UF: Cost Cutting
  - UF: Costs
  - UF: Expenses
  - BT: Finance

Buttons: Add Rels..., +Rel Wiz..., Delete..., Personal Note, Web search

Stay on top

**Add relationships**

**Budgeting\_Costs**

SC  Browse... Save

SC

SN

UF

USE

BT

NT

RT

Close

ENG

---

# Thesaurus Management Software

## Cognatrix

LGOSystems Pty. Ltd. (Australia)

[www.cognatrix.com](http://www.cognatrix.com)

- For Mac OS X 10.4.5 and later
- US \$499, or \$199 for an “Education” version limited to 500 terms.
- Imports from plain text with tab separations.  
With CognatrixImporter add-on, imports from various XML schemas: Cognatrix, MultiTes, Term Tree, and Zthes.
- Exports to XML and HTML.
- Free limited-time downloadable trial and manual



Untitled

Term Attribute Relate Delete

Scope Note User-defined Field

list view filter

Tree View List View

Type/Name	Preferred	Published	Marked	Duplicate	Relationships	Changed
Thesaurus		<input checked="" type="checkbox"/>	<input type="checkbox"/>			2009-04-17 13:42
Top Terms		<input checked="" type="checkbox"/>	<input type="checkbox"/>			2009-04-17 13:42
clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:36
children's clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 14:38
toddler clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 14:37
infant clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		2	2009-04-17 14:35
baby clothes	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:35
baby clothing	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:37
boys' clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 14:18
boys' pants	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 14:13
girls' clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:43
girls' dresses	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 14:39
girls' shirts	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 14:39
girls' pants	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 14:39
clothes	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:36
junior/miss clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		3	2009-04-17 14:38
misses clothing	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:29
juniors clothing	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:29
men's clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 13:46
shoes	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:36
footwear	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		1	2009-04-17 14:36
women's clothing	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		0	2009-04-17 13:47

Scope Note

Girls clothing sizes 4 - 14

Type	Relationship with
RT	junior/miss clothing
(NT)	girls' shirts
(NT)	girls' pants
(NT)	girls' dresses
(BT)	children's clothing

---

# Thesaurus Management Software

## One-2-One

This to That Pty. Ltd. t/a A.C.S.

Active Classification Solutions (Australia)

[www.acs121.com](http://www.acs121.com)

- Price: \$800 Australian (approximately \$700-750)
- For thesauri and classification systems, but also has features and connectors for records management
- Replaces a previous thesaurus-only product, Term Tree
- Drag-and-drop hierarchy feature
- Free limited-use and limited-size trial

One-2-One - connected to Access DB

File Search Insert Delete View Disposal Import Export Reports Options Help

Hierarchy Builder Thesaural Change recording - active

Hierarchy COMMERCIAL ACTIVITIES > Joint Ventures

[6] Function  
 + - A Top down hierarchy view

- 1 COMMERCIAL ACTIVITIES
  - 2 Fees and Charges
  - 2 Joint Ventures
    - 3 *Contracts*
    - 3 *Co-research*
    - 3 *Interdepartmental*
    - 3 *Service Agreements*
  - 2 Marketing
  - 2 Projects
  - 2 Proposals
  - 2 Recruitment
    - 3 *Applications*
  - 2 Reporting
- 1 COMMUNITY RELATIONS
  - 2 Addresses
    - 3 *Seminars*
  - 2 Arrangements
    - 3 *Applications*
  - 2 Awards
  - 2 Celebrations, Ceremonies and
  - 2 Community Consultation
  - 2 Contacts
  - 2 Corporate Image
  - 2 Enquiries
  - 2 Greetings
  - 2 Issues Management
  - 2 Marketing
  - 2 Media Liasion
  - 2 Memorials
  - 2 Signs
  - 2 Sponsorships
  - 2 Visits and Tours
- 1 COMMUNITY SERVICES
- 1 CORPORATE MANAGEMENT
- 1 FLEET MANAGEMENT

Source: Keyword AAA

[1] Broader

1 COMMERCIAL ACTIVITIES

[3] Narrower

3 Co-research

3 Interdepartmental

3 Service Agreements

Drag and Drop Hierarchy Builder

Loaded DB External DB

Copy of terms in currently loaded database. Drag terms from here to main term list to copy selected hierarchy.

Term	C Notes	L Notes
1 COMMERCIAL ACTIVITIES	Replace	Write
2 Fees and Charges	Replace	Write
2 Joint Ventures	Replace	Write
2 Marketing	Replace	Write
2 Projects	Replace	Write
2 Proposals	Replace	Write
2 Recruitment	Replace	Write
2 Reporting	Replace	Write
1 COMMUNITY RELATIONS	Replace	Write
1 COMMUNITY SERVICES	Replace	Write
1 CORPORATE MANAGEMENT	Replace	Write
1 FLEET MANAGEMENT	Replace	Write

DEMO

Close

Database: lgexamphes.mdb in C:\Program Files\One2One\

---

# Thesaurus Management Software

## TheW32

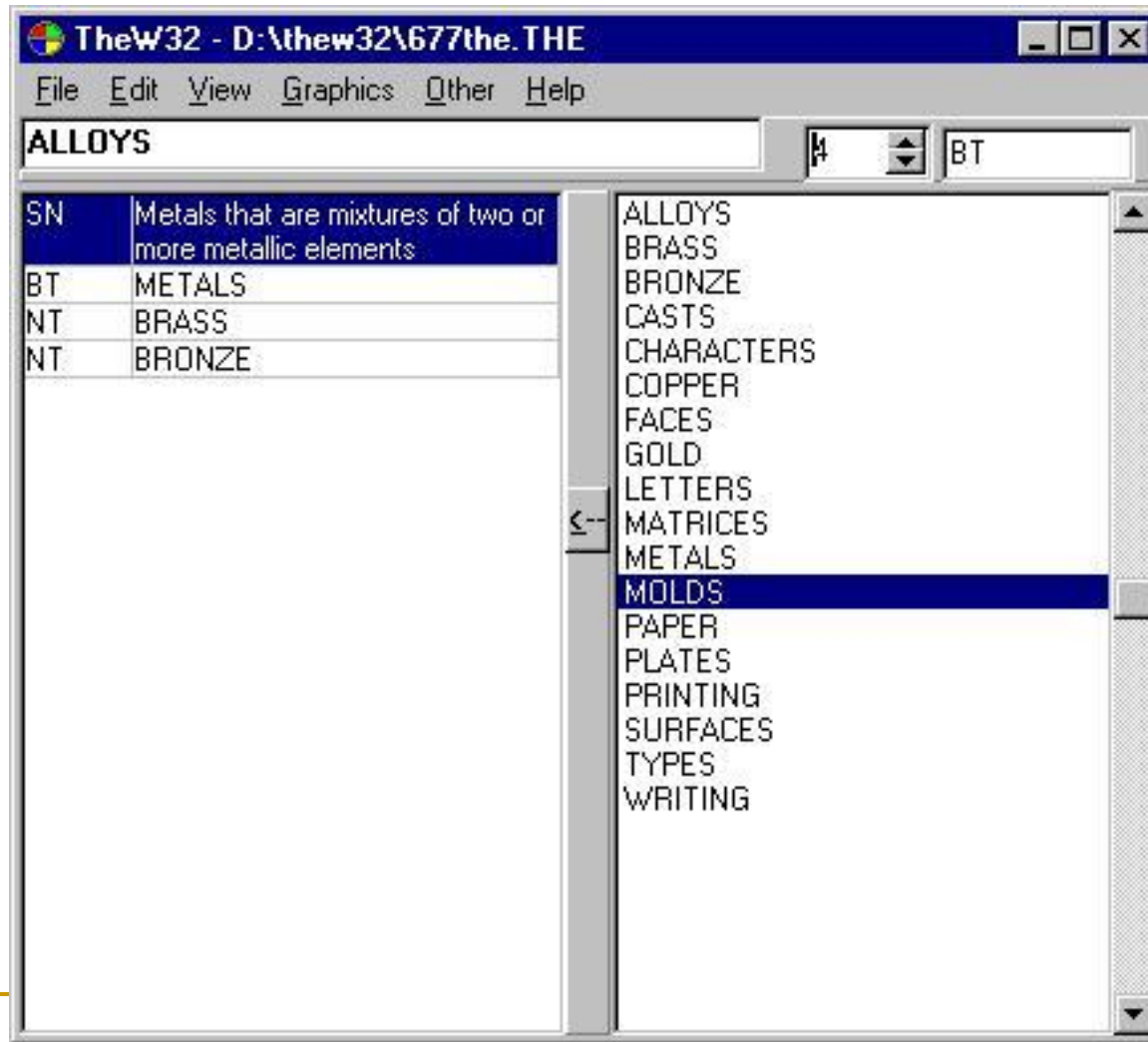
Tim Craven Freeware (Ontario, Canada)

<http://publish.uwo.ca/~craven/freeware.htm>

- Started as thesaurus feature in his TEXNET auto-abstracting tool in early 1990s.
- Free
- Also provides web site indexing software XRefHT and machine-aided indexing and abstracting software

# Thesaurus Management Software

## TheW32 Interface



---

# Thesaurus Management Software

High-end, multi-user client-server, large scale systems  
(\$3000/single user - \$75,000+; or annual hosted options)

- Data Harmony
- Synaptica
- Smartlogic
- Wordmap
- SchemaLogic
- STAR/Thesaurus
- SoutronTHESAURUS
- Mondeca ITM T3
- PoolParty
- TheMa
- a.k.a.

---

# Thesaurus Management Software

## Data Harmony Thesaurus Master

Access Innovations (Albuquerque, NM)

[www.dataharmony.com](http://www.dataharmony.com)

- Indexing services since 1978, commercial software (originally used in-house) offered since 1998.
- Multi-platform java-based (used on Windows, Mac, Solaris, Linux). Client software allows remote access.
- All standard thesaurus displays types as view options
- User defined associative and equivalence, but no user-defined hierarchical relationships
- Sold separately or combined with M.A.I. (Machine Aided Indexer) as MAIstro. Other software extensions available.

MAIstro -- newsindexerthes

Thesaurus Master MAI Rule Builder Test MAI

File Edit View Help

- Business and finance
- Community life
- Government and politics
- Health and wellness
  - Health sciences
    - Allied health sciences
    - Alternative medicine
    - Food engineering
    - Health care
    - Health care facilities**
    - Health concerns
    - Health supplies
    - Health support services
    - Medical conditions and diseases
    - Medical patients
    - Medical prostheses
    - Medical research
    - Medicine
    - Nursing
    - Patient care
    - Physical rehabilitation
    - Public health
    - Veterinary sciences
  - Wellness
- Locations
- News
- Science and technology
- Sports

Term: Health care facilities

Broader Term + - V

Health sciences

Narrower Term + - V

Hospices  
Hospitals  
Mental health facilities  
Nursing and retirement homes  
Rehabilitation centers

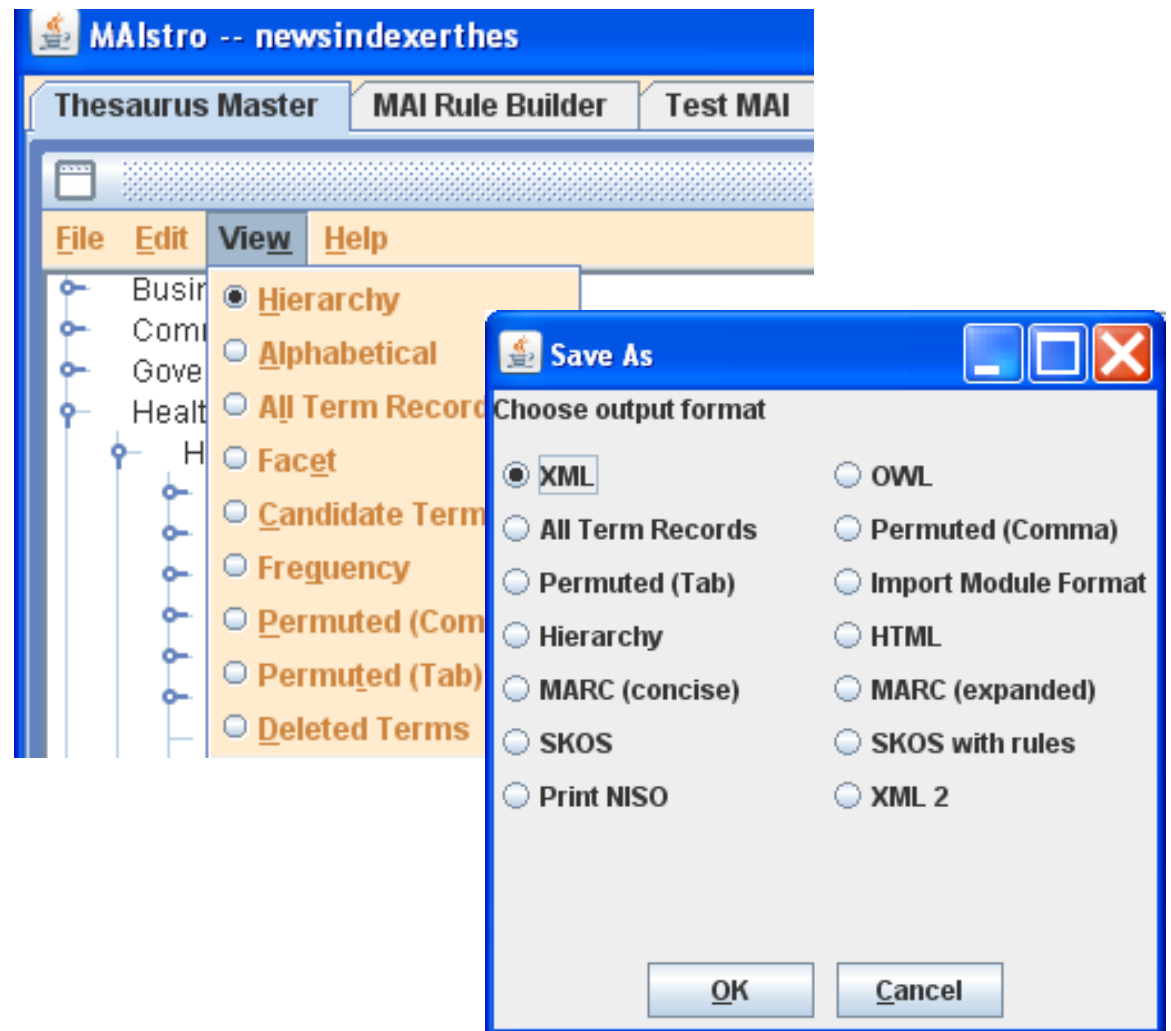
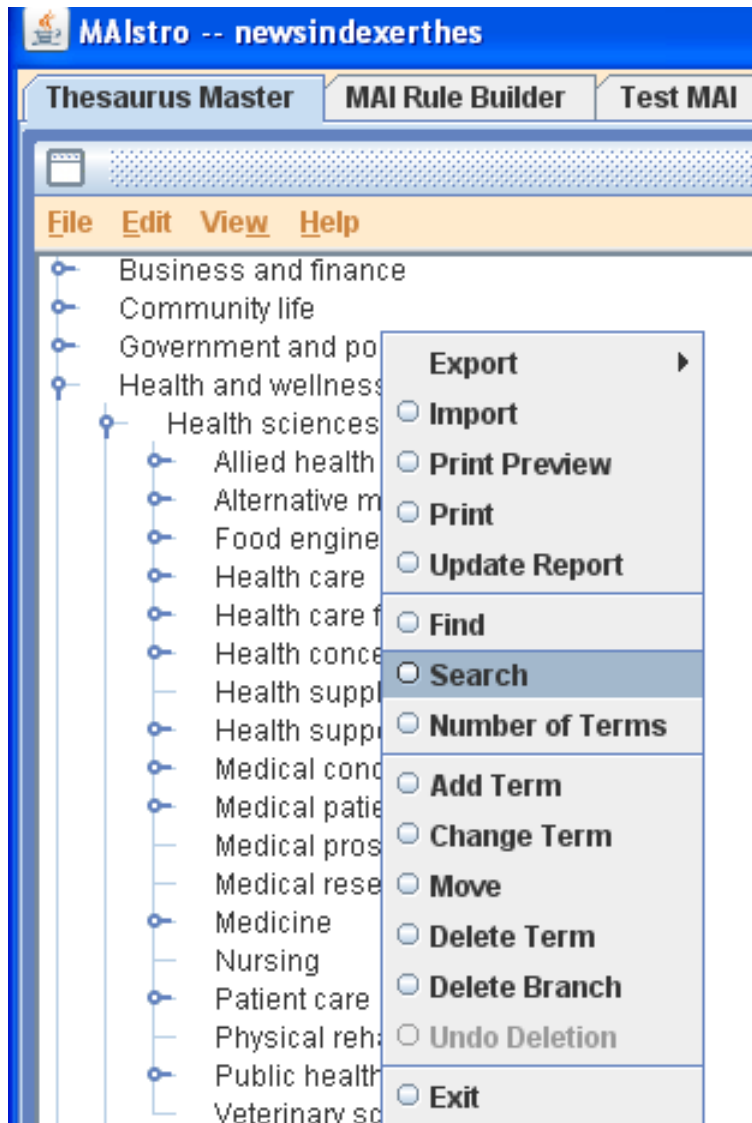
Status  Candidate  Accepted

Related Term + - V

Non-Preferred Term + - V

Scope Note	Editorial Note
Facet+	NNID
History+	





---

# Thesaurus Management Software

## Synaptica

Synaptica Software LLC (Franktown, CO)

<http://synapticasoftware.com>

- Since 1995. Owned by Dow Jones 2005-2009.
- Web browser-based, priced per user, per year, per vocabulary
- 12 graduations of permission levels
- Can assign relationship weights
- Global term and relationships editor, creating a list of terms to edit
- Side-by-side editor with drag-and-drop
- Imports: CSV, text, MS Excel, XML (including schemas of ZThes, RDF, SKOS, and OWL)
- Exports: CSV, HTML, MS Word, MS Excel, XML (including schemas of ZThes, RDF, SKOS, and OWL)

Synaptica® from Dow Jones

Taskviews Search **Tree Browse** A-Z Browse Add Reports

**Browse**

[Refine by Category](#)

[-] demo\_training\_thes

[-] Botany

[-] Applied botany

Agriculture

[-] Forestry

Afforestation

Reforestation

Research Botany

+ Exobiology

+ Medicine

+ term one

Synaptica® from Dow Jones

Taskviews Search Tree Browse **A-Z Browse** Add Reports

**A-Z Browse**

Select Object Class(es)

demo\_training\_thes  
Global  
Local\_1  
Local\_2  
Training Thesaurus

Select Letter / Number

[A](#)[B](#)[C](#)[D](#)[E](#)[F](#)[G](#)[H](#)[I](#)[J](#)[K](#)[L](#)[M](#)  
[N](#)[O](#)[P](#)[Q](#)[R](#)[S](#)[T](#)[U](#)[V](#)[W](#)[X](#)[Y](#)[Z](#)  
[\\*](#)[0](#)[1](#)[2](#)[3](#)[4](#)[5](#)[6](#)[7](#)[8](#)[9](#)

\* returns characters less than 0

Search Results

- [Afforestation](#) (Training Thesaurus)
- [Agriculture](#) (Training Thesaurus)
- [Applied Botany](#) (Training Thesaurus)
- [Applied Zoology](#) (Training Thesaurus)

4 Items Found

There are also browse options including a Tree Browse and Alpha-numeric browse option to review terms. © 2010 Heather Hedden

### Item Summary

**Descriptor** Applied botany  
**Object** Training\_Thes  
**Categories**  
**Status** Active; Approved; Preferred; Unlocked  
**UID** 11570  
**Created** dclarke 3/31/2008 12:09:12 PM  
**Modified** dclarke 3/31/2008 12:09:12 PM

<input type="button" value="Save"/>	<input type="button" value="Add New"/>	<input type="button" value="Subsume"/>
<input type="button" value="Refresh"/>	<input type="button" value="Categories"/>	<input type="button" value="History"/>
<input type="button" value="Deactivate"/>	<input type="button" value="Delete"/>	<input type="button" value="Restore"/>
	<input type="button" value="Copy"/>	

### Sub-Elements

**Descriptor:**

**Scope:**

### Administrative Attributes:

**Approval:**

**Workflow:**

**Language:**

**Locked:**

**BLAT:**  Yes  No

**BTN:**  Yes  No

### Relationships

Limit to Taskview  Show All

### Applied botany

#### Top Level Parents

TT [Botany](#) (Training\_Thes)

#### Parents

[BT Botany](#) (Training\_Thes)

#### Children

[NT Agriculture](#) (Training\_Thes)

[NT Forestry](#) (Training\_Thes)

#### Associations

#### Variants

---

# Thesaurus Management Software

## Semaphore Ontology Manager

Smartlogic Semaphore Ltd. (London, UK) + US office

[www.smartlogic.com](http://www.smartlogic.com)

- Supports creating thesauri according to ISO 2788 standard
- Supports creating ontologies, through customizable relationships and user-created classes
- User-defined term attributes and metadata
- Multiple user access/privilege levels
- Imports/export in CSV, XML, Zthes, SQL databases, and MultiTes files
- Related products:
  - Classification Server for automated classification
  - Ontology Service for a navigation system

IPSV - Semaphore Ontology Manager

File Terms Tools Publish Configure Windows Help

Information and communication technology

**Hierarchical**

Type	Term
BT	Information and communication
BT	Science, technology and innovation
NT	Communication channels
NT	Computer applications
NT	e-Learning
NT	Electronic media
NT	Hardware
NT	ICT services
NT	Networking
NT	Radio
NT	Smartcards
NT	Software
NT	Systems development
NT	Telecommunications
NT	Television

**Associative**

Type	Term
RT	Communications industries
RT	Data security
RT	e-Commerce
RT	Information management

**Equivalence**

Type	Term
UF	<i>JCT (information and communication technology)</i>
UF	<i>Information Technology</i>
UF	<i>IT (information technology)</i>

Hierarchical    Alphabetical

---

# Thesaurus Management Software

## Wordmap Designer

Wordmap Inc. (Concord, MA)

[www.wordmap.com](http://www.wordmap.com)

- In UK since 1998. Acquired by Earley & Associates in 2007.
- Multi-platform java-based
- One of a suite of products: Wordmap Navigator, Wordmap Content Author (for manual indexing), Wordmap Intelligent Text Classifier, Wordmap Taxonomy + Search.
- User-defined relationships; can also turn on or off relationship name display.
- Can display two taxonomies side by side and drag and drop.
- User access/privileges can be set at the individual node level.
- Imports: text, CSV, XML; Exports: CSV, XML

PRAKASHGOVIND.IPSV (1.00)

PSV

- [is a] Business and industry
  - [is a] Business people
  - [is a] Business practice and regulation
  - [is a] Business sectors
  - [is a] Companies
  - [is a] Consumer affairs
  - [is a] Energy and fuel
    - [is a] Energy conservation
    - [is a] Energy efficiency
    - [is a] Fossil fuels
    - [is a] Nuclear energy
    - [is a] Renewable energy
      - [related] Combustion processes licences
      - [related] Combustion processes licences
      - [related] Electricity supply
      - [related] Electricity supply
      - [related] Gas supply
      - [related] Gas supply
      - [related] Utilities
      - [related] Utilities
- [is a] International trade
- [is a] Economics and finance
- [is a] Education and skills
- [is a] Employment, jobs and careers
- [is a] Environment
- [is a] Government, politics and public administration
- [is a] Health, well-being and care
- [is a] Housing
- [is a] Information and communication
- [is a] International affairs and defence
- [is a] Leisure and culture
- [is a] Life in the community
- [is a] People and organisations
- [is a] Public order, justice and rights
- [is a] Science, technology and innovation
- [is a] Transport and infrastructure

### Members

Pos	Text	Language
1	Consumer affairs	en
2	Consumer law	en
3	Shopping	en

### Features

Feature	Value	Language
A to Z	true	
Added in version	1.00	
Category	true	
Concept Id	500	
Id	500	
Last updated in version	1.00	
Obsolete	false	
Preferred	true	
Scope notes		
WM ITL	LGCL/409	

### Children

- [is a] Energy conservation
- [is a] Energy efficiency
- [is a] Fossil fuels
- [is a] Nuclear energy
- [is a] Renewable energy
- [related] Combustion processes licences
- [related] Combustion processes licences
- [related] Electricity supply
- [related] Electricity supply
- [related] Gas supply
- [related] Gas supply

### Other Locations

IPSV > Business and industry > Energy and fuel

IPSV > Environment > Energy and fuel



- + [is a] Downloadable content
- + [is a] Home monitoring & control systems
- [is a] Phones
  - + [is a] Cordless phones
  - [is a] Mobile phones
    - [is a] MOTOPEBL
    - [is a] MOTORAZR V3
    - + [is a] MOTORAZR V3i
    - + [operate on a] Networks
    - + [have a] Data transmission technologies
    - + [have a] Form factors
    - + [can have service with a] Carriers
    - + [can have feature] Phonebook
    - + [can have feature] Speakerphone
- + [is a] Two-way radios
- + [have a] Availability
- + [have a] Product states

Parent Child Relationship

Define custom associative relationships through simple drag & drop

- + [is a] Downloadable content
- + [is a] Home monitoring & control systems
- [is a] Phones
  - + [is a] Cordless phones
  - [is a] Mobile phones
  - [is a] MOTOPEBL
  - [is a] MOTORAZR V3
  - + [is a] MOTORAZR V3i
  - + [operate on a] Networks

Wordset Attributes		
Satellites Archived		
Members		
Pos	Text	Language
1	Mobile phones	en
2	Cell phones	en
3	Cellular phones	en
4	Téléphones mobiles	fr

Synonyms  
Translation

---

# Thesaurus Management Software

## SchemaLogic Enterprise Suite

SchemaLogic Inc. (Kirkland, WA)

[www.schemalogic.com](http://www.schemalogic.com)

- Provides thesaurus management according to ANSI/NISO standards, plus broader structural metadata support
- Can create customizable relationships
- Can assign various permission levels to vocabularies or terms
- Classification module supports 3<sup>rd</sup> party auto-indexing
- Connectors to SharePoint, EMC Documentum, and FAST ESP
- Can import CSV or XML files

---

# Thesaurus Management Software

## STAR/Thesaurus

Cuadra Associates, Inc. (Los Angeles)

[www.cuadra.com](http://www.cuadra.com)

- Stand-alone or integrates with STAR family of products for records mngmt, collections mngmt, archives mngmt, DAM
- Supports standard thesaurus relationship but not customizable relationships
- Supports unlimited user-defined notes and categories
- Various output report display formats
- Import/export ASCII text and CSV, but not XML

---

# Thesaurus Management Software

## SoutronTHESAURUS

Soutron Ltd. (United Kingdom)

[www.soutron.com](http://www.soutron.com)

- Markets in the U.S. through partnership with InMagic
- Stand-alone or integrates with SoutronGLOBAL or SoutronSOLO library management systems, or with InMagic Presto social knowledge management software
- Supports standard thesaurus and user-defined relationships
- Supports term merging
- Imports from XML; exports to XML or CSV

---

# Thesaurus Management Software

Mondeca ITM T3 (Intelligent Topic Manager:  
Thesaurus, Taxonomies, Terminologies)

Mondeca S.A. (Paris, France)

[www.mondeca.com](http://www.mondeca.com)

- Since 2008, addition to Intelligent Topic Manager set of products for knowledge management, semantic portals, and e-catalogs
- web-based collaborative application
- conforms to both SKOS vocabularies and OWL-standard ontologies
- connectors to text mining, classification, and search tools
- Imports/exports XML, RDF, and SKOS

---

# Thesaurus Management Software

## PoolParty

punkt. netServices GmbH (Vienna, Austria)

<http://poolparty.punkt.at>

- A new thesaurus tool 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
- Wordpress plugin to build glossaries for blogs
- Imports ZThes XML, CSV
- Integrated text extraction and semi-automatic tagging to enable semantic search

---

# Ontology Software

## Tools for ontologies, not thesauri

- TopBraid Composer, [www.topquadrant.com](http://www.topquadrant.com)
- Altova SemanticWorks, [www.altova.com/semanticworks.html](http://www.altova.com/semanticworks.html)
- Protégé, <http://protege.stanford.edu>
- SMORE, [www.mindswap.org/2005/SMORE/](http://www.mindswap.org/2005/SMORE/)
- SWOOP, <http://code.google.com/p/swoop>
- CMAP Tools Ontology Editor, <http://cmap.ihmc.us>



---

# Other Software with Taxonomy/thesaurus Creation & Editing Components

- Metadata or cataloging software, especially for archives and libraries
  - Adlib Information Systems, [www.adlibsoft.com](http://www.adlibsoft.com)
- Content management and document management systems
  - Open Text Collections Server Webtop Thesaurus Manager, [www.opentext.com](http://www.opentext.com)
- Records management systems
  - a.k.a. from Synercon Management Consulting, [www.a-k-a.com.au](http://www.a-k-a.com.au)
- Auto-categorization and enterprise search software

---

# Auto-categorization and Text Mining

## Auto-categorization

- Algorithms, statistics, and training documents – utilize a large set a sample documents per taxonomy term to “train” the system to learn to index
- Rules base – generate and edit or write “rules” for each term based on co-existing words, proximity, Boolean logic, etc.

## Text mining

- Extracts relevant terms from texts to generate a candidate taxonomy or supplement an existing taxonomy

---

# Auto-categorization/Search Software

Auto-categorization, text mining, and search systems that utilize taxonomies handle these taxonomies in different ways:

- With pre-installed taxonomies that cannot be edited
- With pre-installed taxonomies that the user may edit and extend through the user interface
- Automatically generate a taxonomy that can be edited
- Support the import of taxonomies but do not support the editing of those taxonomies
- Support the import of taxonomies and then the editing of those taxonomies
- Various combinations of above

---

# Auto-categorization/Search Software

Software that can import and use taxonomies but lacking user-interface features to edit those taxonomies includes:

- Microsoft SharePoint
- IBM Classification Module
- Fast
- Endeca
- Temis
- Vivisimo
- Mindbreeze
- Exalead
- PerfectSearch

They collaborate with other vendors that develop taxonomies and/or have taxonomy editing capabilities.

---

# Auto-categorization/Search Software

Examples of tools with *some* taxonomy management capabilities:

- Inxight SmartDiscovery Analysis Server  
[www.inxightfedsys.com](http://www.inxightfedsys.com)
- Autonomy Collaborative Classifier  
[www.autonomy.com](http://www.autonomy.com)
- Autonomy Interwoven MetaTagger  
[www.interwoven.com](http://www.interwoven.com)
- Lexalytics Classifier  
[www.lexalytics.com](http://www.lexalytics.com)
- Conceptsearching  
[www.conceptsearching.com](http://www.conceptsearching.com)

---

# Auto-categorization/Search Software

Examples of tools with *full* thesaurus management capabilities:

- Data Harmony MAIstro  
[www.dataharmony.com/products/maistro.html](http://www.dataharmony.com/products/maistro.html)
- Smartlogic Semaphore Classification Server  
[www.smartlogic.com](http://www.smartlogic.com)
- Wordmap Intelligent Text Classifier  
[www.wordmap.com](http://www.wordmap.com)
- SAS Enterprise Content Categorization  
(formerly Teragram TK240)  
[www.sas.com/text-analytics](http://www.sas.com/text-analytics)
- Nstein Text Mining Engine (part of Open Text)  
[www.nstein.com](http://www.nstein.com)

---

# Auto-categorization Software combined with Thesaurus Management

## Data Harmony MAIstro (combines Data Harmony Thesaurus Master and Machine-Aided Indexer)

- Automatically creates a basic rule for every term and its variants in the Thesaurus Master's thesaurus
- Rules may be edited and additional rules can be manually written
- statistics module tracks the editor's term choices and compares them with M.A.I. term suggestions, sorting them as hits, misses, and noise to guide and prioritize the editor's fine-tuning of rules
- Can be used for machine-aided indexing or fully automated indexing
- Connectors to Sharepoint and search engines

---

# Auto-categorization Software combined with Thesaurus Management

## Smartlogic Semaphore Classification Server

(connects with Semaphore Ontology Manager)

- Automatically creates classification rulebases directly from a taxonomy/thesaurus/ontology, and applies these rules to content as it is received to automatically classify content.
  - Rules are based on the term, its equivalencies, and broader/narrower/related terms
  - Employs 20 different kinds of rules
  - Rules have weights and scores
  - Variants based on spelling, plurals, and stemming may also be considered.
  - Manual rules can take precedence over generated rules.
-



---

# Auto-categorization Software combined with Thesaurus Management

## Wordmap Intelligent Text Classifier

(connects with Wordmap Designer for thesaurus creation)

- Auto-classification based on statistical method based of Support Vector Machine algorithms and machine learning with training documents
- Pre-packaged with statistical algorithms based on a generic taxonomy, the U.K.'s Integrated Public Services Vocabulary (IPSV), for which each of hundreds of terms have already been “trained” with representative documents
- Wordmap also offers Taxonomy Connectors for taxonomy-driven tagging and search within SharePoint and Endeca

---

# Auto-categorization Software combined with Thesaurus Management

## SAS Enterprise Content Categorization (ECC)

(Formerly Teragram TK240 Taxonomy Manager)

Supports taxonomy building or connects with SAS Ontology Manager

- ECC supports equivalent, hierarchical & related relationships
- Ontology Manager supports customized relationships and attributes
- Utilizes both auto-categorization and entity/concept extraction
- Auto-categorization bases on rules
- Rules-writing supported with a graphical tree view of Boolean operators and commands
- User can define weighting of terms

---

# Auto-categorization Software combined with Thesaurus Management

## Nstein Text Mining Engine (TME)

- Modules include concept extractor, entity extractor, auto-categorizer, automated abstract creation, sentiment analysis
- Taxonomy Manager module (not sold separately) supports creating & editing hierarchical, associative and equivalence relationships according to ANSI/NISO standard
- Auto-categorization technology based on use of training sets for taxonomy terms, combined with concept extraction technology
- Ships with pre-installed taxonomies already “trained” for auto-categorization

---

# Concluding Remarks

- Some “taxonomy tools” are stronger in taxonomy/thesaurus/ontology management
- Some “taxonomy tools” are stronger in auto-categorization
- A few tools combine both, but vendor partnerships and connectors and also achieve high results

---

# Questions/Contact

Heather Hedden

Carlisle, Massachusetts

978-371-0822

978-467-5195 (mobile)

heather@hedden.net

[www.hedden-information.com](http://www.hedden-information.com)

[www.accidental-taxonomist.com](http://www.accidental-taxonomist.com)