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.)

© 2019 Hedden Information Management

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  - Web lists are miscellaneous taxonomy-related tools or out-of-date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)

© 2019 Hedden Information Management

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

© 2019 Hedden Information Management

Background to Software

Types of software used by taxonomists

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

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

© 2019 Hedden Information Management

Background to Software

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 Scope notes	Ambiguity control (Synonym control) Semantic relationships Classes Attributes (Linked data)

Standards Supported by Software

Standards for thesaurus design and best practices

- ISO 25964-1
- ANSI/NISO Z39.19-2005 (R2010)

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

Standards Supported by Software

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.

© 2019 Hedden Information Management

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

© 2019 Hedden Information Management

Software Features

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

© 2019 Hedden Information Management

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

© 2019 Hedden Information Management

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

© 2019 Hedden Information Management

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

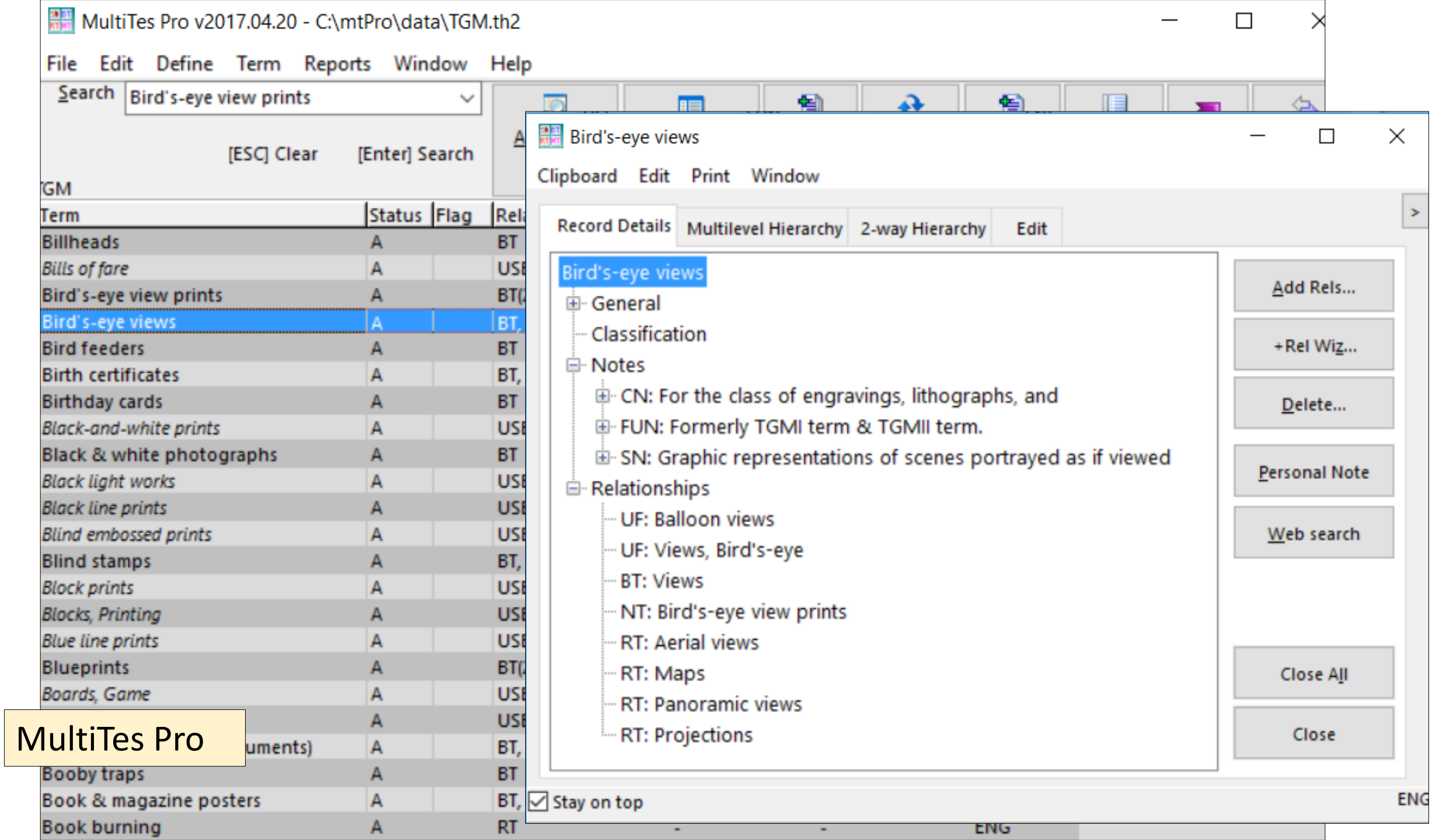
© 2019 Hedden Information Management

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

© 2019 Hedden Information Management



MultiTes Pro

Software Products: Synaptica

Synaptica Software LLC (Franktown, CO)

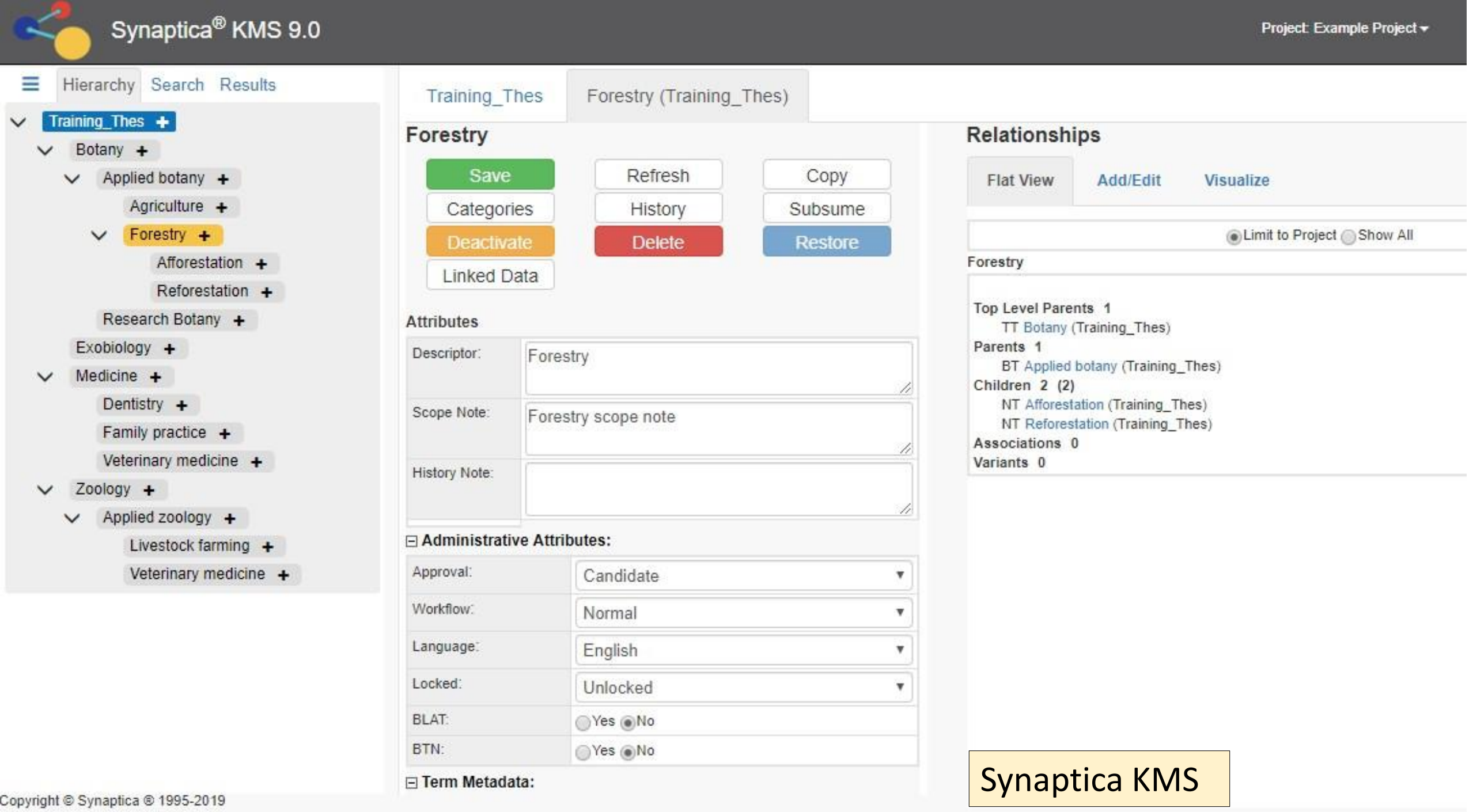
www.synaptica.com

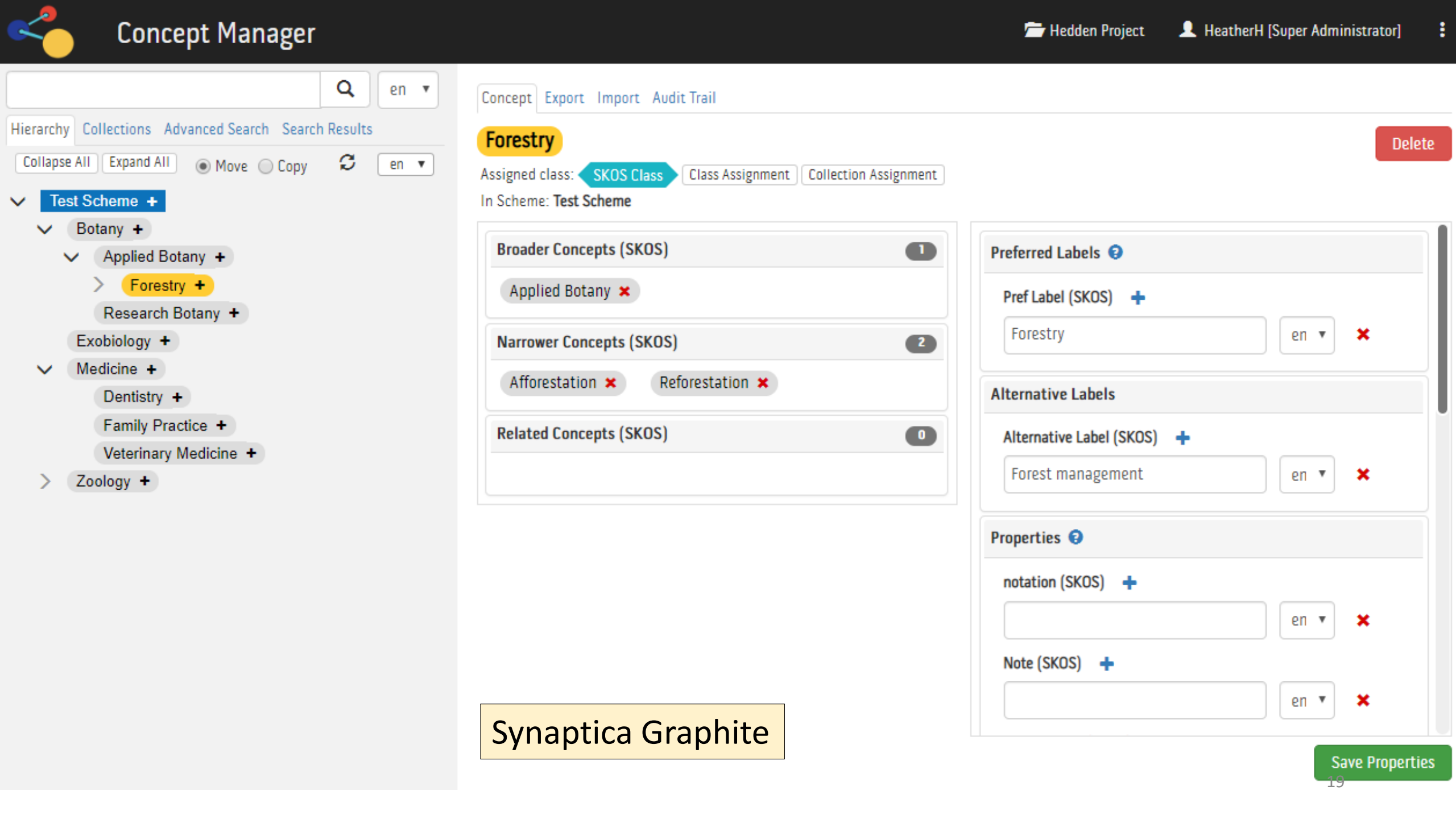
Synaptica KMS (Knowledge Management System) – thesaurus model (since 1995)

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

© 2019 Hedden Information Management


Synaptica® KMS 9.0
Project: Example Project
Hierarchy
Search
Results
Training_Thes
Botany
Applied botany
Agriculture
Forestry
Afforestation
Reforestation
Research Botany
Exobiology
Medicine
Dentistry
Family practice
Veterinary medicine
Zoology
Applied zoology
Livestock farming
Veterinary medicine
Training_Thes
Forestry (Training_Thes)
Forestry
Save
Refresh
Copy
Categories
History
Subsume
Deactivate
Delete
Restore
Linked Data
Attributes
Descriptor:
Forestry
Scope Note:
Forestry scope note
History Note:
Administrative Attributes:
Approval:
Candidate
Workflow:
Normal
Language:
English
Locked:
Unlocked
BLAT:
Yes
No
BTN:
Yes
No
Term Metadata:
Relationships
Flat View
Add/Edit
Visualize
Limit to Project
Show All
Forestry
Top Level Parents 1
TT Botany (Training_Thes)
Parents 1
BT Applied botany (Training_Thes)
Children 2 (2)
NT Afforestation (Training_Thes)
NT Reforestation (Training_Thes)
Associations 0
Variants 0
Synaptica KMS
Copyright © Synaptica ® 1995-2019



Synaptica Graphite

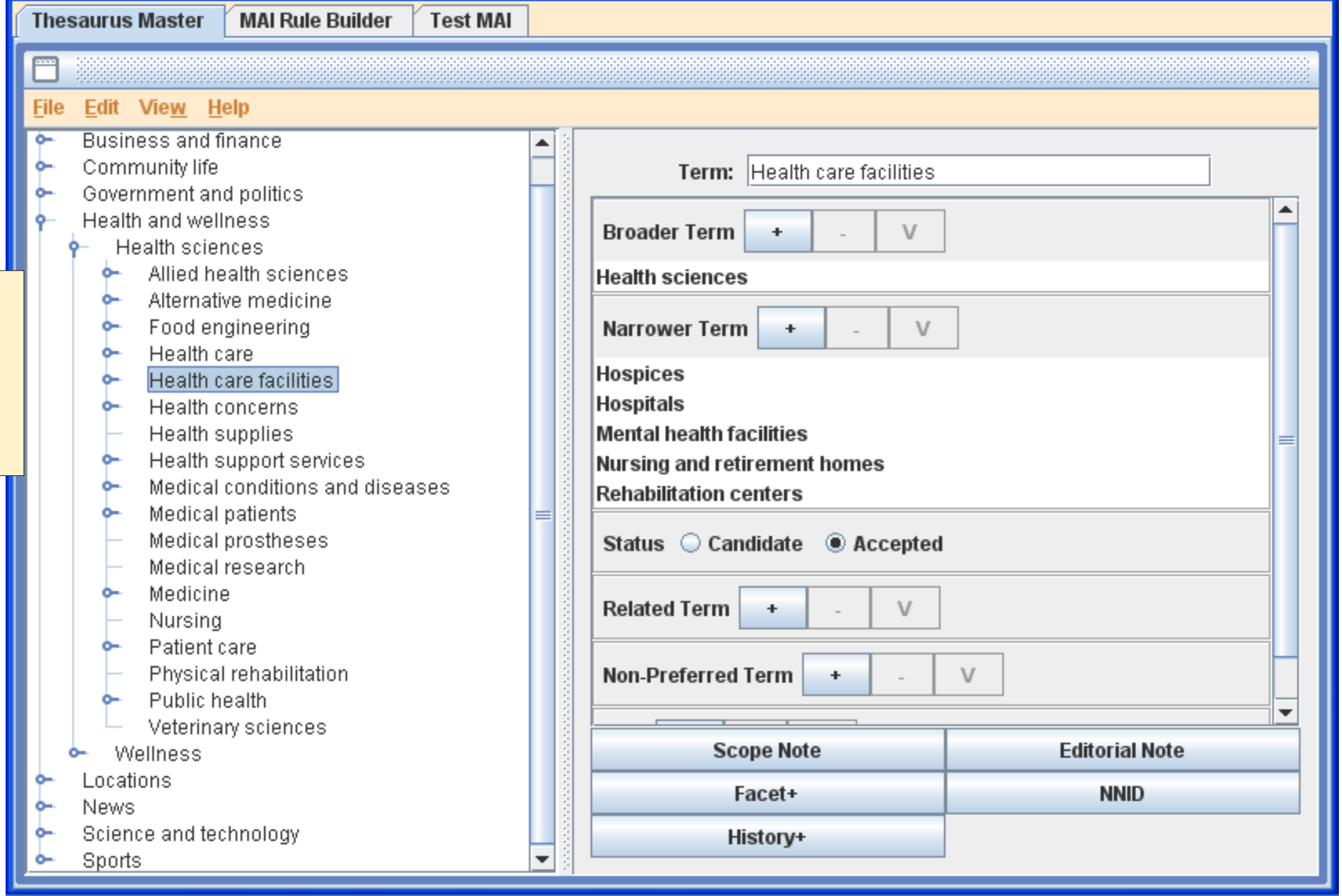
Software Products: Data Harmony Thesaurus Master

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.

© 2019 Hedden Information Management

Data Harmony Thesaurus Master



Software Products: Semaphore

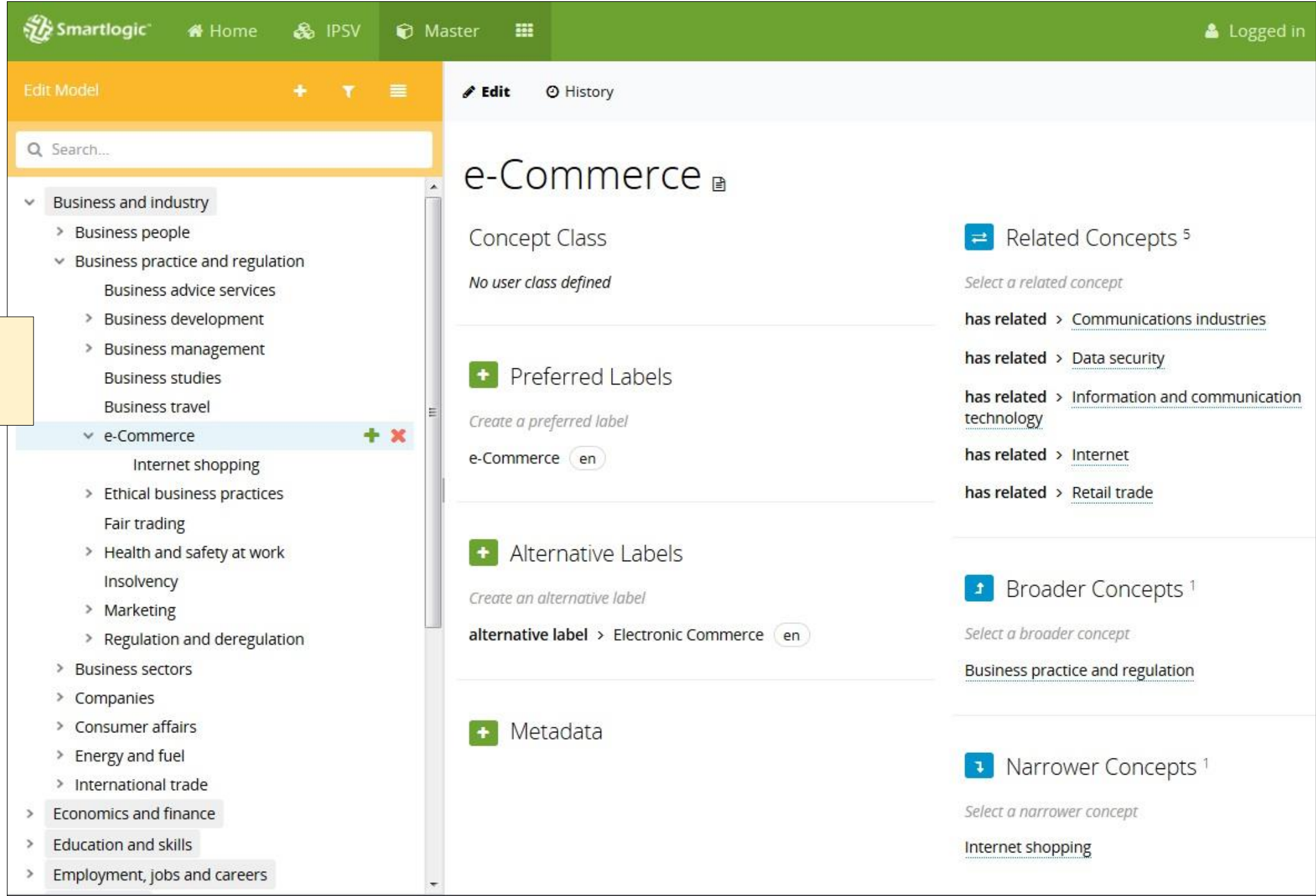
Smartlogic Semaphore Ltd. (London, UK)

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.)

© 2019 Hedden Information Management

Smartlogic Semaphore



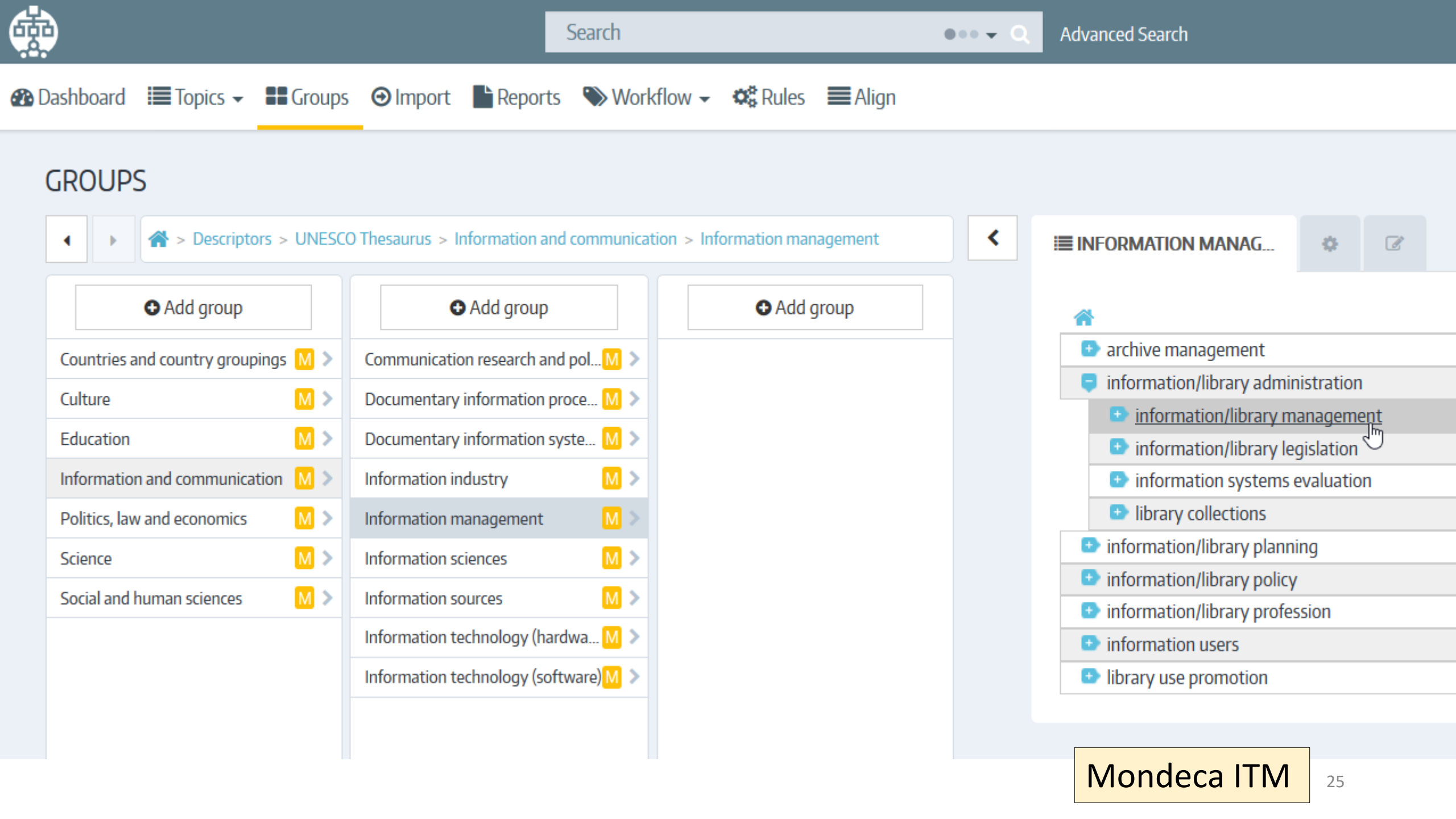
Software Products: Mondeca Intelligent Topic Manager (ITM)

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

© 2019 Hedden Information Management


Search
Advanced Search
Dashboard
Topics
Groups
Import
Reports
Workflow
Rules
Align
GROUPS
Descriptors > UNESCO Thesaurus > Information and communication > Information management
Add group
Add group
Add group
Countries and country groupings
Culture
Education
Information and communication
Politics, law and economics
Science
Social and human sciences
Communication research and pol...
Documentary information proce...
Documentary information syste...
Information industry
Information management
Information sciences
Information sources
Information technology (hardwa...
Information technology (software)
INFORMATION MANAG...
archive management
information/library administration
information/library management
information/library legislation
information systems evaluation
library collections
information/library planning
information/library policy
information/library profession
information users
library use promotion

Mondeca ITM

UNESCO CONCEPT

INFORMATION/LIBRARY MANAGEMENT

Edit | Delete | Actions

ATTRIBUTES

Name	es información/biblioteca, gestión
	fr information/bibliothèque, gestion
	en information/library management
URI	http://skos.um.es/unescothes/C02010
in group	▸ UNESCO Thesaurus ▸ Information management ▸ Information and communication
used for	fr gestion de la bibliothèque
	en library management
	es gestión de las bibliotecas

ASSOCIATIONS

Filter by type

Broader-Narrower

information/library administration

Related concepts

related	▸ information/library management ▸ library use promotion

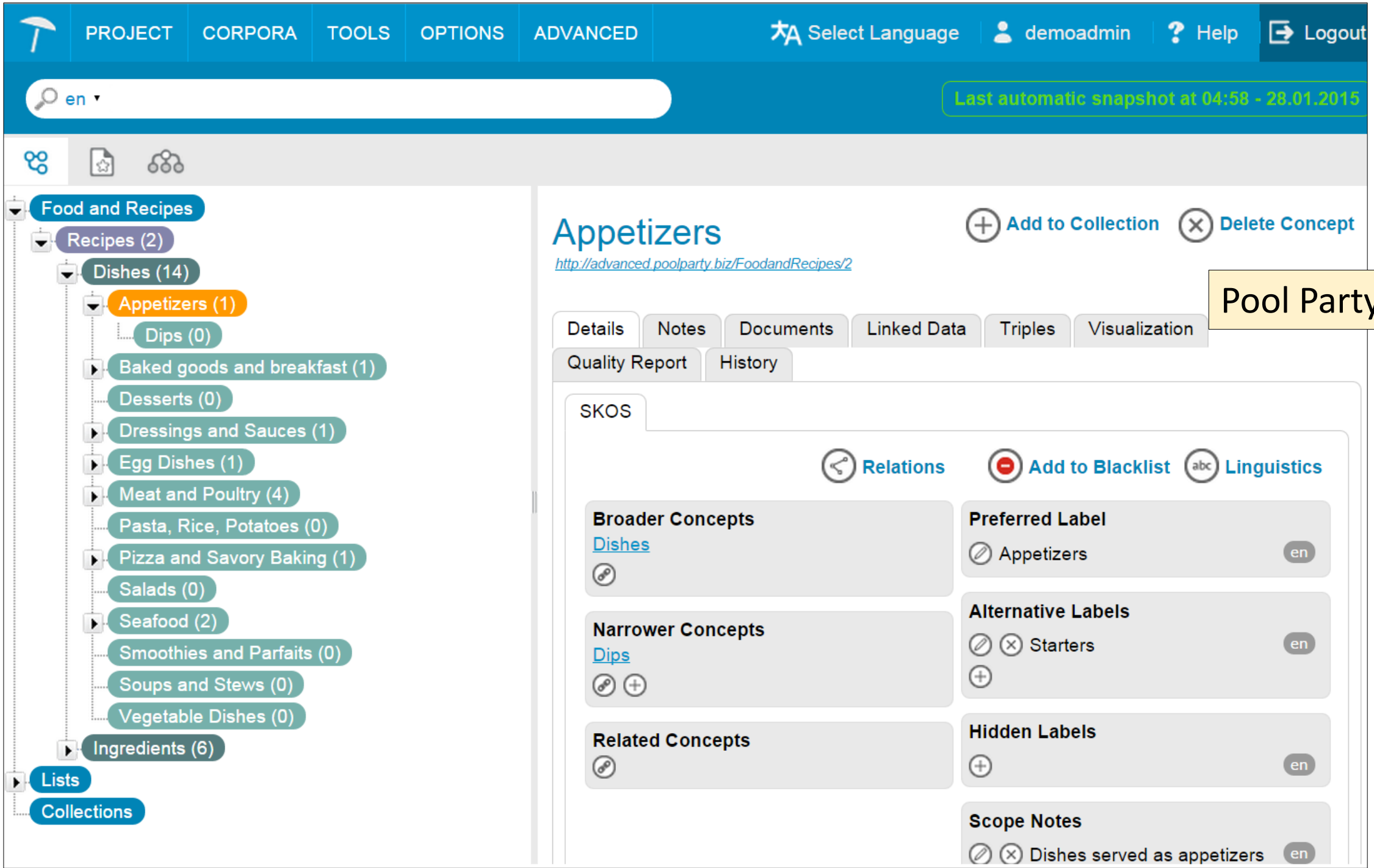
Mondeca ITM

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

© 2019 Hedden Information Management


PROJECT
CORPORA
TOOLS
OPTIONS
ADVANCED
Select Language
demoadmin
Help
Logout
en
Last automatic snapshot at 04:58 - 28.01.2015
Food and Recipes
Recipes (2)
Dishes (14)
Appetizers (1)
Dips (0)
Baked goods and breakfast (1)
Desserts (0)
Dressings and Sauces (1)
Egg Dishes (1)
Meat and Poultry (4)
Pasta, Rice, Potatoes (0)
Pizza and Savory Baking (1)
Salads (0)
Seafood (2)
Smoothies and Parfaits (0)
Soups and Stews (0)
Vegetable Dishes (0)
Ingredients (6)
Lists
Collections
Appetizers
http://advanced.poolparty.biz/FoodandRecipes/2
Add to Collection
Delete Concept
Pool Party
Details
Notes
Documents
Linked Data
Triples
Visualization
Quality Report
History
SKOS
Relations
Add to Blacklist
Linguistics
Broader Concepts
Dishes
Narrower Concepts
Dips
Related Concepts
Preferred Label
Appetizers
en
Alternative Labels
Starters
en
Hidden Labels
en
Scope Notes
Dishes served as appetizers
en

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

© 2019 Hedden Information Management

Concept Hierarchy

Peru

- Geography
  - Africa
  - Antarctica
  - Asia
  - Atlantic Ocean
  - Europe
  - Global
  - Indian Ocean
  - Latin America
    - Central America
    - South America
      - Argentina
      - Bolivia
      - Brazil
      - Chile
      - Colombia
      - Ecuador
      - French Guiana
      - Guyana
      - Paraguay
      - Peru
      - Suriname
      - Uruguay
      - Venezuela
  - North America
  - Pacific Ocean
  - West Indies

Peru (Country)

Edit | Delete | Print | Start Workflow | (0) Tasks | (0) Comments | Country

Hide Properties | (0) | Show History

http://topquadrant.com/ns/examples/geography#Peru

Labels and Description | **Label Objects (SKOS-XL)**

Geopolitical Characteristics | **Location** | **Standard Relationships** | **Notes**

Matching Relationships | **Other Properties**

preferred label:	Peru (en) Perú (es) Pérou (fr)
alternative label:	
hidden label:	
notation:	
type:	Country
definition:	
top concept of:	

TopBraid EDG
Vocabulary Management

Search for Concept

Type Country

Free Text

Filters | Columns

Return local results only

Details | Export

Show 25 entries **Refine:**

Country	broader concept	area (KM)
Afghanistan	Asia	647500
Albania	Europe	28748
Algeria	Africa	2381741
American Samoa	Pacific Ocean	
Andorra	Europe	468
Angola	Africa	1246700
Anguilla	West Indies	
Antarctica		
Antigua and Barbuda	West Indies	442
Argentina	South America	
Armenia	Asia	
Aruba	West Indies	193

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

- Dedicated ontology software; not so suitable for taxonomies/thesauri

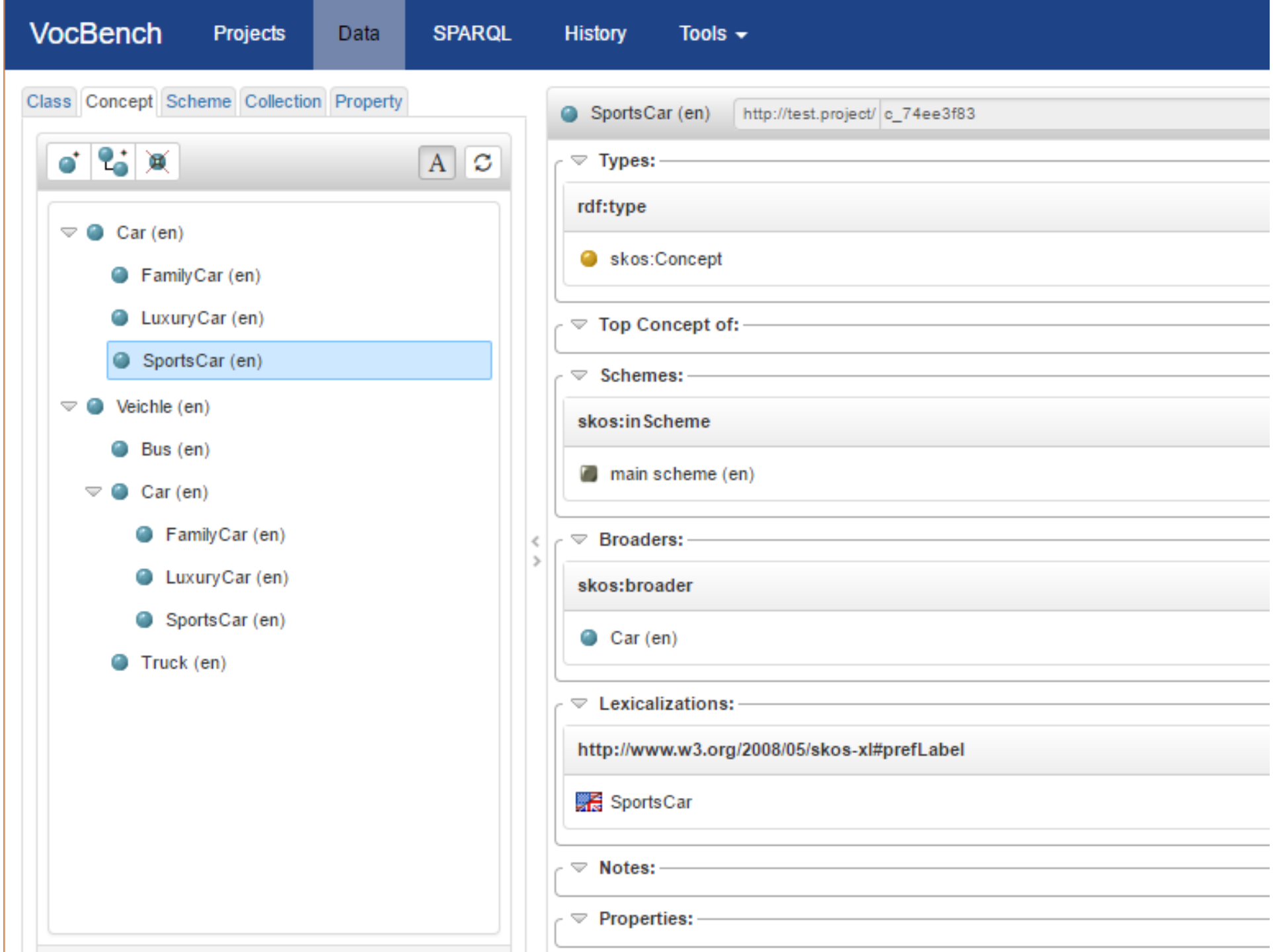
VocBench – Developed by the Artificial Intelligence Research group at University of Tor Vergata, Rome http://vocbench.uniroma2.it

- 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 ISA^2 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

- 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

© 2019 Hedden Information Management


VocBench
Projects
Data
SPARQL
History
Tools
Class
Concept
Scheme
Collection
Property
A
Car (en)
FamilyCar (en)
LuxuryCar (en)
SportsCar (en)
Veichle (en)
Bus (en)
Car (en)
FamilyCar (en)
LuxuryCar (en)
SportsCar (en)
Truck (en)
SportsCar (en)
http://test.project/
c_74ee3f83
Types:
rdf:type
skos:Concept
Top Concept of:
Schemes:
skos:inScheme
main scheme (en)
Broaders:
skos:broader
Car (en)
Lexicalizations:
http://www.w3.org/2008/05/skos-xl#prefLabel
SportsCar
Notes:
Properties:

VocBench

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

© 2019 Hedden Information Management

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.

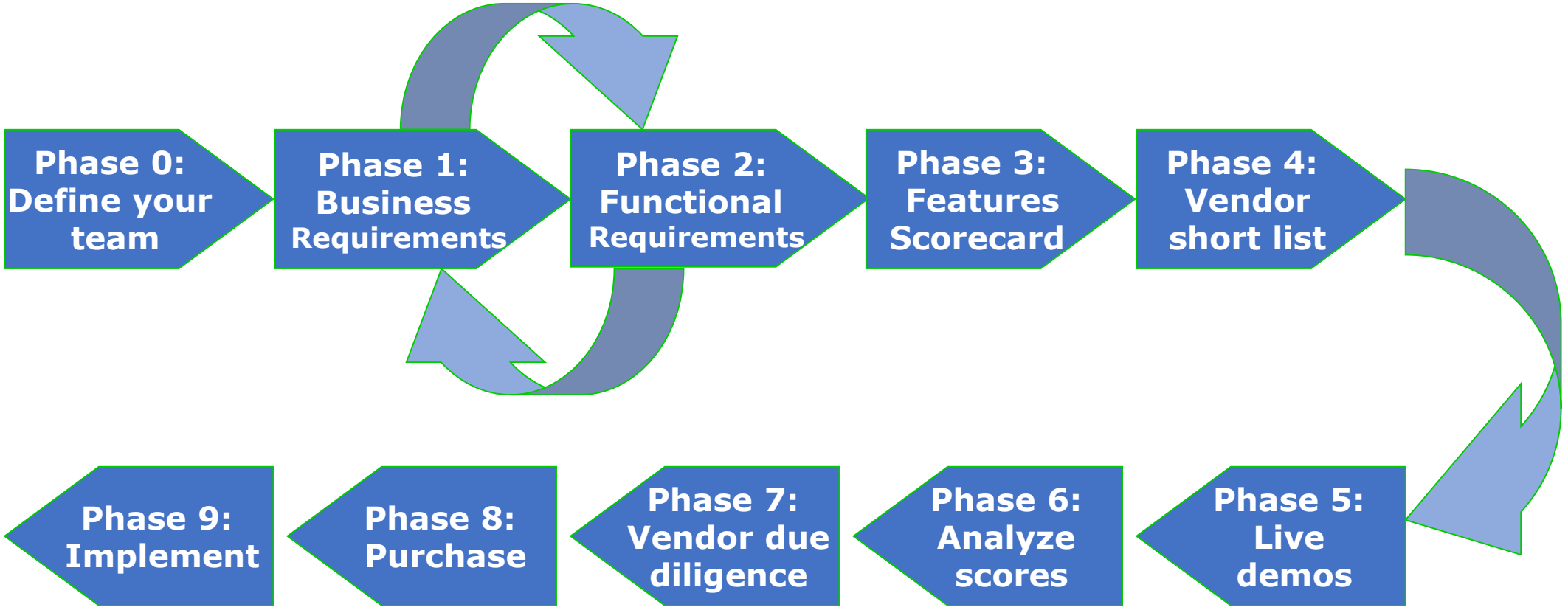
Not evaluating software risks…

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

© 2019 Hedden Information Management

Software Evaluation

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



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

© 2019 Hedden Information Management

Taxonomy Features Review Sheet

Priority 1=Must have 2=Important but don't need in first release 3=Nice to have

Relevant User Group	Feature	Priority	Vendor 1	Comments	Vendor 2	Comments
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - Supports multiple Languages	1	0		1	
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - ANSI/NISO Z39.19 compliant	1	1		1	
All (Content Mgrs, Taxonomy Mgrs, End Users)	General - Windows operating system compliant	1	1		1	
Taxonomy Managers	Editing - Ability to select multiple terms.	1	0		1	
Taxonomy Managers	Editing - Built in Spell checker	1	0		0	
Taxonomy Managers	Editing - Prevent duplicate term entry within a category. Warn when duplicate is entered.	1	1		1	
Taxonomy Managers	Administration - terms should not be deleted by taxonomy managers, just changed to an inactive status.	1	0.5	business process	1	
System	Relationship Mgmt - Poly-hierarchies (i.e. Ability for a single term to be in more than one category).	1	1		1	
Taxonomy Managers	Workflow - There should be an audit trail showing changes.	1	0		1	
Taxonomy Managers	Workflow - roles should be used when defining taxonomy managers responsible for each workflow step. Ability to change the roles without developer assistance.	2	1		1	
All (Content Mgrs, Taxonomy Mgrs, End Users)	Display - Ability to see full record of a term on the screen (relationships, attributes, comments, etc.)	1				
All (Content Mgrs, Taxonomy Mgrs, End Users)	Reporting - Ability to export reports in a common format (CSV, XML, etc.)	1	1		1	
Taxonomy Managers	Import/Export - Import/Export terms, categories, taxonomies in a standard format (CSV, XML, etc.)	1	0.5	not intuitive	1	
Taxonomy Managers	Search - Search across all terms and all categories (taxonomies) and attributes.	1	1		0.5	no wild cards?
Taxonomy Managers	Security - Ability to set security at a category (taxonomy) level.	2	0		1	
All (Content Mgrs, Taxonomy Mgrs, End Users)	Help - On line help, quick reference guide	1	0.5	pdf guide	0	print manual
			9.3	**TOTAL**	13.1	**TOTAL**

Sample Completed Score Card

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

© 2019 Hedden Information Management

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

© 2019 Hedden Information Management