

User-Driven Taxonomies

Christine Connors

iKMS, Singapore, 13 March 2008

The problem with...

- **Formal taxonomies**
 - High cost
 - Taxonomy creation experts
 - Subject Matter Experts (SMEs)
 - Software & Hardware
 - Purchase & modify
 - Consultants
 - Scope and timeline
 - Implementation
 - Maintenance
 - Hard to sell an ROI

The problem with...

- **Informal taxonomies**
 - Consistency, clarity, context
 - Scope and timeline
 - Implementation
 - Maintenance
 - Hard to sell an ROI

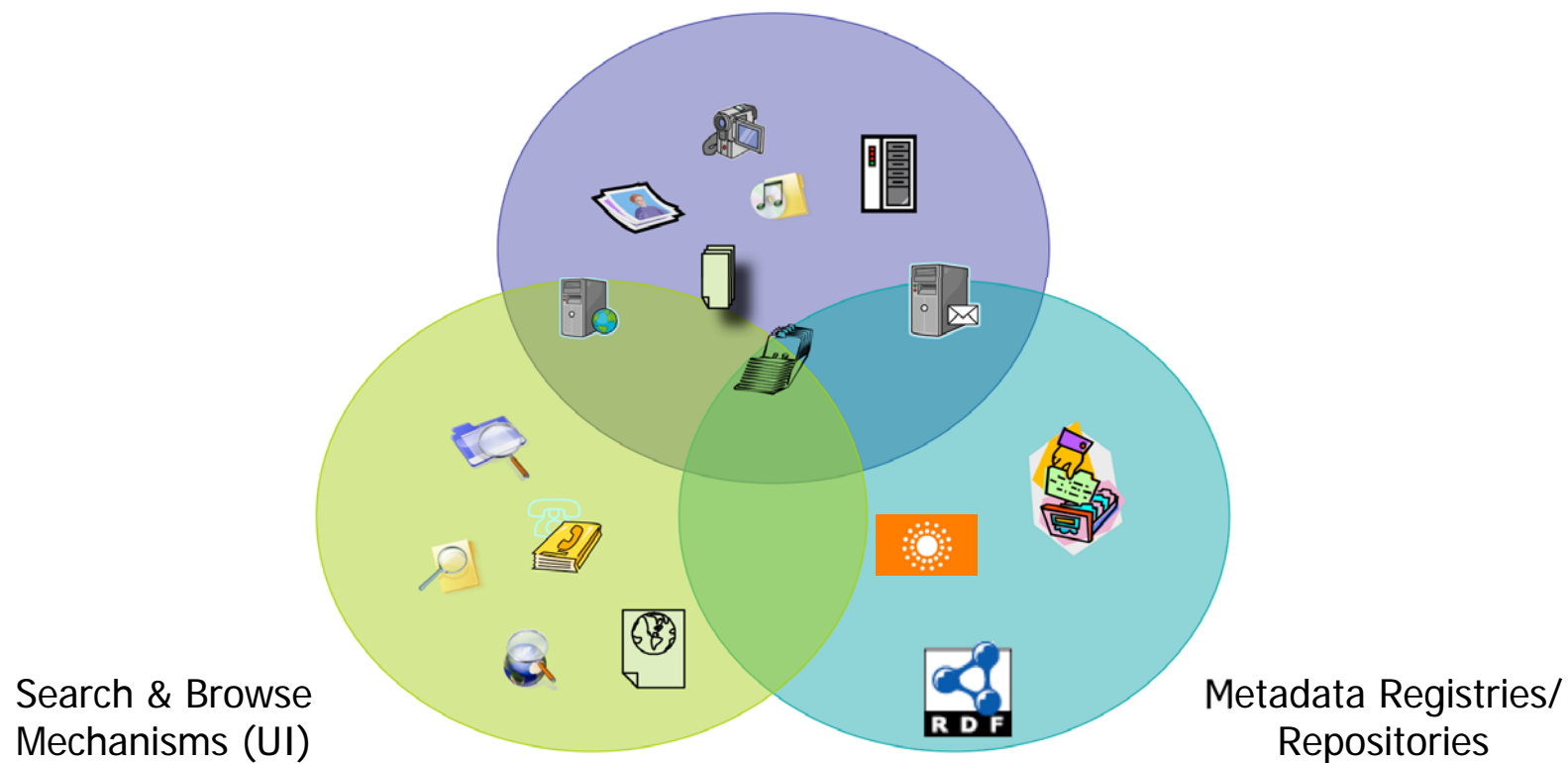
The benefits of a hybrid approach

- **Expertise in taxonomy design**
- **User-centered language**
- **Contextual variety**
- **User-driven prioritization of knowledge modeling**
- **Grow the model faster**
 - Guided by taxonomists to avoid chaos
- **Distributed costs**

- **Does require**
 - A champion
 - Change Control Board / Taxonomy Advisory Board

Literary and user warrant in the enterprise

Object Repositories



What is a folksonomy?

- **“People’s classification management”**
- **Wisdom of the crowd**
- **User-generated tags applied to digital objects**
- **Informal, uncontrolled vocabularies**
- **Usually Subject or Task based**
- **Provide little to no context on their own**

Examples

- Primary examples are del.icio.us and flickr

del.icio.us / CJMConnors / taxocop
 your bookmarks | your network | subscriptions | links for you | post

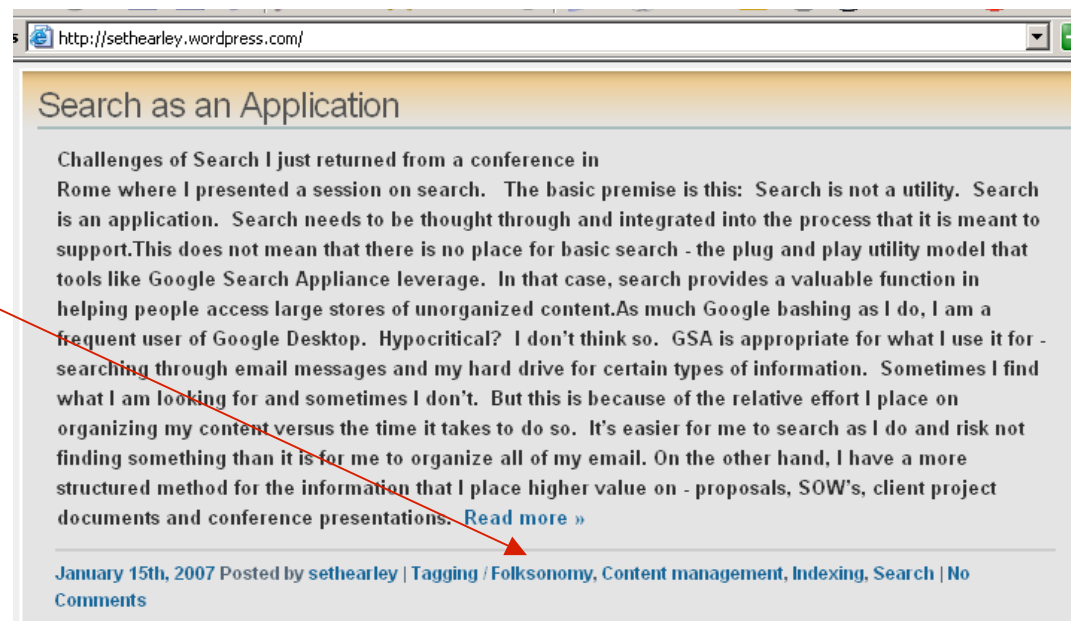
Your items tagged **taxocop** ([create tag description](#)) → view **all**, **popular**

« earlier | later » showing both items

[Resources for Taxonomy Creation and Content Management](#) [edit](#) / [delete](#)
 to taxocop taxonomy ontology standards books weblinks ... [saved by 6 other people](#) ... on april 07

[taxocop » Discussion summaries](#) [edit](#) / [delete](#)
 to taxocop CJMConnors ... [saved by 1 other person](#) ... on april 04

- Blogs are another good place to look



http://sethearley.wordpress.com/

Search as an Application

Challenges of Search I just returned from a conference in Rome where I presented a session on search. The basic premise is this: Search is not a utility. Search is an application. Search needs to be thought through and integrated into the process that it is meant to support. This does not mean that there is no place for basic search - the plug and play utility model that tools like Google Search Appliance leverage. In that case, search provides a valuable function in helping people access large stores of unorganized content. As much Google bashing as I do, I am a frequent user of Google Desktop. Hypocritical? I don't think so. GSA is appropriate for what I use it for - searching through email messages and my hard drive for certain types of information. Sometimes I find what I am looking for and sometimes I don't. But this is because of the relative effort I place on organizing my content versus the time it takes to do so. It's easier for me to search as I do and risk not finding something than it is for me to organize all of my email. On the other hand, I have a more structured method for the information that I place higher value on - proposals, SOW's, client project documents and conference presentations. [Read more »](#)

January 15th, 2007 Posted by [sethearley](#) | [Tagging](#) / [Folksonomy](#), [Content management](#), [Indexing](#), [Search](#) | [No Comments](#)

Lessons learned: Hybrid methods and Social tagging pilots

Evolution

- **In the beginning...**
 - Best Bets were created by the search administrator
 - Search terms parsed out of the query sent from the browser to the search engine
 - Terms compared to manually created list of Best Bet sites
 - Matches were programmatically inserted into the SERP before the #1 hit, with special formatting to highlight their existence
 - Intranet site owners called the search administrator to beg inclusion

An early pilot

Library Services
Library Catalogs
Worldwide News Resources
Web Links
Special Collections
Directory of Research Libraries

Research Libraries Home

Browse Bookmarks
Find Bookmark
Add Bookmark

Administration

Legend	
	Rate
	Review
	Inspect

Public Bookmarks

- Select a category within which to base your search
- Enter keyword(s)

Category

Keyword

- Each resource can only be placed in one bucket, need to duplicate entries for full coverage
- Not integrated with any other system - ILMS, DMS, CMS, FS
- Administered by Research Librarians
- Rarely used!
- How do we integrate Enterprise Search, Suggested Sites, Public Bookmarks and Social Tagging?

Updates to enterprise search

- In search, search terms are tagged to bring back certain websites
 - Users call, email or submit via web-form sites they would like to see added
 - Taxonomy team reviewed the submission for appropriateness, accuracy of tags, uniqueness of tags
 - Sites and associated terms are manually entered into a flat file
 - During the regular index refresh cycle the flat file is programmatically converted to XML and ingested into search

Make a site suggestion for your query

Please make a suggestion for a site to be added to the "Suggested Sites" list. Please tell us what you are looking for, and what site you expected to find. If you know the URL for the specific site you were looking for and have time to get it, please provide it. If not, simply tell us the name and we will find it. Provide any description you think will help us.

Suggestions for websites that may be of general interest/use to everyone are welcome, as are very specific suggestions. Your input will enable us to make Search better for all users. We evaluate suggestions on a weekly basis, and provide new releases once a month. Your last search query will be automatically sent to us as part of your suggestion to help us.

Thanks for your help.

Your Employee Number:

Your Suggestion:

Suggested Sites

- **Research Libraries**
Electronics Resource Page
- **Research Libraries**
Research Libraries
- **Make a suggestion**

2006 Social bookmarking pilot

- We wanted to see **what** would happen if we “opened” up the tagging
- Goal was to help our users find commonly requested information and most useful information by
 - Tagging favorite internal websites
 - Maintain security by NOT posting intranet URLs to public sites like del.icio.us
 - Linking directly to a resource, be it internal or external
 - Sharing and searching other user’s bookmarks
 - Removing a bottleneck and relieving resource constraints in a moderated hybrid system
 - Reviewed available systems
 - Public sites not an option due to security considerations
 - Connotea, Scuttle, del.icio.us, Freetag

How can folksonomies improve discovery?

As Inputs

- **To taxonomies, thesauri, ontologies?**
 - What folksonomy terms are popular?
 - What synonyms can you derive?
 - What relationships can you identify?
 - What entity types are you discovering?
- **To search**
 - Identify Best Bets
 - As inputs to a recommendation engine
- **To the content management strategy**
 - What do they tell you about how your content is perceived?
 - What do they tell you about how your content is used?
 - Do they tell you when your users go elsewhere for their content needs?

User driven

- **Enables user warrant**
 - Useful for understanding users – how do they think about the objects you are providing to them?
 - Allows the users to find things their own way, rather than forcing them to do it the site's way

- **Improves user experience**
 - Combine with search and web logs to
 - Improve navigation
 - Improve browse mechanisms
 - Improve search
 - Identify content gaps
 - Prioritize content and UI related tasks

How can you implement folksonomy tools?

Sample Pure methods

- **Install a tagging tool**
 - Tools similar to del.icio.us
 - Connotea
 - Scuttle
 - ConnectBeam
 - Semantic applications such as Annotea (W3C) or semantic blogging tools
 - Modules for blogs/CMSs, examples:
 - Taxonomy modules for Drupal
 - Tagging system in Wordpress or Typepad
 - Extensions for MediaWiki
- **Make sure you review the reports available in the tools you consider**
 - Can you get actionable data?

Sample implementations of social/hybrid methods

- **Best bets**
 - Allow users to submit sites, along with keywords, to improve search results

Search Support Site

Example: Create a report [Search Tips](#)

You are searching **QuickBooks**

Search Results 1 - 25 of 45 < prev | next 20 > [Search With](#)

Closest Matches:

- [Windows Vista Resource Center](#)
 Microsoft has made significant changes to how software runs on its new operating system, Windows Vista. Planning to upgrade to the Windows Vista operating system, no action is required. You can continue to use QuickBooks products.

Suggested Sites

- **Research Libraries**
Electronics Resource Page
- **Research Libraries**
Research Libraries
- **Make a suggestion**

- **File properties / repository check-in form**
 - Encourage (or require!) that users fill out the properties of the files they create, using any terms they deem appropriate
 - Automate whenever possible

Why not just get Google Properties

General | Security | Custom | Summary

Title:

Subject:

Author:

Category:

Keywords:

Comments:

Commercial Example

- **Buzzillions.com**
 - Combines formal taxonomy with folksonomy terms to guide users to the products right for them

The screenshot shows the Buzzillions.com website interface. At the top, the logo 'Buzzillions BETA' is displayed with the tagline 'Find the right product for you ... guided by real users!'. Navigation tabs include Home, Digital Cameras, Sporting Goods, Footwear, Entertainment, Home & Garden, and See All. A search bar is present with 'Find' and 'go' buttons. Below the search bar, the breadcrumb trail reads 'Home > Consumer Electronics > Cameras & Photo > Digital Cameras'. The main heading is 'Digital Cameras' with 'Results matched: 321 of 321 Products'. A 'Narrow Results' sidebar on the left lists categories and brands. The 'Consumers Speak' section contains four steps: 'Who am I?', 'How will I use it?', 'What do I like?', and 'What do I dislike?'. A tooltip explains the 'Affinity Recommendation Wizard'. The product listing for the 'Nikon D200 Digital SLR Camera- (Camera Body)' shows a price of \$1,339.95 (crossed out) and \$1,499.99, a 4.8-star rating, and a quote from a user: 'This is an excellent camera. It incorporates a of of the D2X features but at less than half the price. The performance... - dopey from Chicago, IL'. A 'RITZ CAMERA' logo with 'FREE SHIPPING' and 'NO SALES TAX' is also visible.

Thank you!

Christine Connors
Global Director, Semantic Technology Solutions
Dow Jones & Company
Christine.Connors@dowjones.com

Announcing Synaptica 7.0!

- **Synaptica 7.0 provides standardized, Semantic Web-enabled tools to manage your global business vocabulary in order to add structure and value to existing information assets, improve the online user experience and connect professionals in your organization with the information they need, where and when they need it.**

- **Customer Benefits**
 - **Easy configuration**
 - **Scalable for the enterprise with multi-user permissions**
 - **Customizable and flexible with audience-centric views**
 - **Supports collaboration and workgroups**
 - **Standards based, semantic Web enabled**
 - **Multiple data formats (HTML,XML,etc.)**
 - **API level access for simple integration**

- **<http://solutions.dowjones.com/djcs/index.asp>**

Easily find and edit a key term and multiple related terms

Synaptica's new side by side relationship editor makes the creation and editing of terms a one step process.

Synaptica® from Dow Jones Taskview: Eurovoc DTS; User: Jim S

Taskviews Search Tree Browse A-Z Browse Add Reports Tools Admin

Add Relationships

Key Term Search (Select One):

Search phrase:

Smart Begins Contains Exact

Object Class filter: Eurovoc_DTS_en

- 56 AGRICULTURE, FORESTRY AND FISHERIES (Eurovoc_DTS_en)
- 329 tropical agriculture (Eurovoc_DTS_en)**
- 559 part-time agriculture (Eurovoc_DTS_en) (NPT)
- 645 Mediterranean agriculture (Eurovoc_DTS_en)
- 968 United Nations Food and Agriculture Organisation (Eurovoc_DTS_en)
- 1534 water management in agriculture (Eurovoc_DTS_en)
- 1849 mechanisation of agriculture (Eurovoc_DTS_en)

Related Term Search (Select One or More):

Search phrase:

Smart Begins Contains Exact

Object Class filter: Eurovoc_DTS_en

- 3340 foreign tourism (Eurovoc_DTS_en)
- 3341 green tourism (Eurovoc_DTS_en) (NPT)
- 3341 rural tourism (Eurovoc_DTS_en)
- 3705 tourism policy (Eurovoc_DTS_en)
- 4470 tourism (Eurovoc_DTS_en)**
- 4471 mass tourism (Eurovoc_DTS_en)
- 5909 tourism statistics (Eurovoc_DTS_en)

Relationship:

show weic BT

Add Relatio NR

NT

RT

UF

Use

Current Relationships: 329 tropical agriculture (Eurovoc_DTS_en)

<ul style="list-style-type: none"> DEL 329 tropical agriculture (Eurovoc_DTS_en) DEL 329 tropical agriculture (Eurovoc_DTS_en) DEL 329 tropical agriculture (Eurovoc_DTS_en) DEL 329 tropical agriculture (Eurovoc_DTS_en) DEL 329 tropical agriculture (Eurovoc_DTS_en) DEL 329 tropical agriculture (Eurovoc_DTS_en) 	<ul style="list-style-type: none"> BT 5631 agricultural activity (Eurovoc_DTS_en) BT 2723 crop production (Eurovoc_DTS_en) EN>FR 329 culture tropicale (Eurovoc_DTS_fr) EN>HU 329 trópusi mezőgazdaság (Eurovoc_DTS_hu) RT 2418 tropical plant (Eurovoc_DTS_en) UF 329 tropical crops (Eurovoc_DTS_en)
--	--

Synaptica © Software Copyright © 1995-2008 Dow Jones & Company, Inc. All Rights Reserved.

Easily Manage and edit vocabulary hierarchies

Synaptica drag and drop hierarchical relationship editing provides a simple, convenient way to manage vocabulary hierarchies.

Synaptica® from Dow Jones Taskview: Eurovoc DTS; User: Jim Sweeney; Permission: SA

Taskviews Search Tree Browse A-Z Browse Add Reports Tools Admin

Drag and Drop Hierarchical Relationship Editor

Select drag mode: Move Copy

Object Class filter Refresh Left Tree Object Class filter Refresh Right Tree

Drag and drop terms between terms or trees to create relationships

- U8 INTERNATIONAL RELATIONS
- 10 EUROPEAN COMMUNITIES
- 12 LAW
- 16 ECONOMICS
 - 1606 economic policy
 - 1611 economic growth
 - 1616 regions and regional policy
 - 431 regional development
 - 462 regional disparity
 - 1468 regional integration
 - 1542 rural development
 - 1900 Mez. 57 tax authorities
 - 2515 regional policy
 - 2516 Community regional policy
 - 3034 agricultural region
 - 2042 coastal region

Delete Selected Relationship Edit Selected Term

Drag and drop terms between terms or trees to create relationships

- 2411 monetary economics
- 2416 credit and financial institutions
- 2421 free movement of capital
- 2426 financing and investment
- 2431 insurance
- 2436 public finance and budget policy
- 2441 budget
- 2446 taxation
 - 57 tax authorities
 - 170 taxpayer
 - 365 tax relief
 - 504 double taxation
 - 530 registration tax
 - 561 tax law
 - 562 international tax law
 - 634 tax avoidance

Delete Selected Relationship Edit Selected Term

Synaptica © Software Copyright © 1995-2008 Dow Jones & Company, Inc. All Rights Reserved.

Gain quick views of term information without leaving current interface

Term Information Summary Window provides quick views of term details

The screenshot shows a hierarchical tree view on the left and a detailed summary window on the right. The tree view is expanded to show '614 trading operation' selected. The summary window displays the following information:

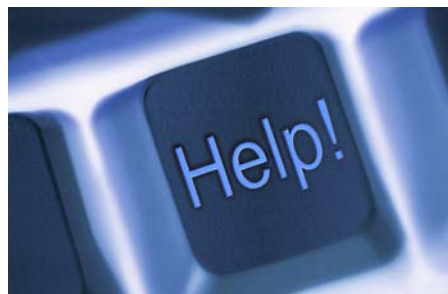
Summary for: 614 trading operation	
Property	Value
Item UID:	1223450
Preferred Term:	Yes
Object Class:	Eurovoc_DTS_en
Subelements	
Label	Value
descripteur_id (df)	614
descripteur (df)	trading operation
Relationship Metrics	
Broader Concepts:	1
Narrower Concepts:	8
Associated Concepts:	7
Preferred/Non-Preferred Concepts:	0
Inter-Object Class Relationships	
Eurovoc_DTS_fr	1

Expanded Reporting Functionality for Easier, More Flexible Information Sharing

In addition to CSV, HTML and XML formats, reports may be created in Microsoft Word and Excel.

	A	B	C	D	E
1	TERM_LEVEL	TERM_UID	Descriptor	PARENT_TERM_UID	
2	1	1217469	POLITICS		
3	2	1217588	political framework	1217469	
4	3	1220349	Communism	1217588	
5	3	1218304	conservatism	1217588	
6	3	1221014	democracy	1217588	
7	3	1221024	people's democracy	1217588	
8	3	1221614	dictatorship	1217588	
9	3	1223606	ecologism	1217588	
10	3	1224431	State	1217588	
11	4	1224435	Welfare State	1224431	
12	4	1222907	non-secular State	1224431	
13	5	1222910	Islamic State	1222907	
14	4	1222908	rule of law	1224431	
15	4	1222909	federal State	1224431	
16	5	1218558	Federation State	1222909	
17	4	1222911	secular State	1224431	
18	4	1222912	unitarian State	1224431	

Quickly and easily access Help right from the application



Synaptica User and Administrative Guides are now available online directly from the application to browse and search

The screenshot shows a Mozilla Firefox browser window with the title bar '- Getting Started Guide - Mozilla Firefox'. The page content includes a search bar with 'Search' and 'Close' buttons, and a 'Getting Started Guide' section. The main content area displays a search result for 'botany' with three items found: 'Botany (Training Thesaurus)', 'Applied Botany (Training Thesaurus)', and 'Research Botany (Training Thesaurus)'. Below the search results, there are instructions for navigating through the documentation. A 'Relationships' section is partially visible at the bottom of the page.

Online Documentation > Search Close

Getting Started Guide

<< [Previous Topic](#) | [Up a Level](#) | [Table of Contents](#) | [Next Topic](#) >>

4. [Adding Terms and Relationships](#)

4.2. Review Terms Added so far

1. On the main browser window in the keyword search field enter the search phrase "botany" and press the Search button. The search should yield three results:

Search Results

 - [Botany](#) (Training Thesaurus)
 - [Applied Botany](#) (Training Thesaurus)
 - [Research Botany](#) (Training Thesaurus)

3 Items Found
2. Click on the second item in the results, "Botany", to view the Item Summary popup screen for that term. This was the first term you created before you added a tree of narrower terms beneath it.
3. In the right-hand frame select the Tree View option press the "[+]" plus button next to the Tree Expansion label to expand the tree of terms under "Botany".

Relationships

Done

Dow Jones Client Solutions Offers Comprehensive, Business Taxonomy Solutions for Fast, Relevant Information Retrieval

**License & Integrate
Industry-focused
Taxonomies**

**Build & Customize
To Suit Your
Information Needs**



**Industry-focused
integrated solutions**

**Optimize and Manage
With Synaptica**

**Stay Informed
with
Taxonomywarehouse.com**