Teaching Taxonomy Methodology
Who I am: Joseph Busch

- Over 25 years in the business of organized information.
  - Founder, Taxonomy Strategies LLC
  - Director, Solutions Architecture, Interwoven
  - VP, Infoware, Metacode Technologies
    - (acquired by Interwoven, November 2000)
  - Program Manager, Getty Foundation
  - Manager, Pricewaterhouse

- Metadata and taxonomies community leadership.
  - President, American Society for Information Science & Technology
  - Director, Dublin Core Metadata Initiative
  - Adviser, National Research Council Computer Science and Telecommunications Board
  - Reviewer, National Science Foundation Division of Information and Intelligent Systems
  - Founder, Networked Knowledge Organization Systems/Services
What we do
What we do

Organize Stuff
Recent & current projects:
http://www.taxonomystrategies.com/html/clients.htm
What is a Taxonomy?

- A categorization framework agreed upon by business and content owners (with the help of subject matter experts) that will be used to tag content.
  - 6 broad, discrete divisions (called facets)
  - 2-3 levels deep.
  - Up to 15 terms at each level.
  - 1200 terms total.
  - With some logic—hierarchical, equivalent and associative relationships between terms.
What uses must a Taxonomy support?

- **Primary categorization**
  - Navigation
  - Content Management

- **Secondary categorization**
  - Search
  - Tagging

“When we talk about a taxonomy, we are not only talking about a website navigation scheme. Websites change frequently, we are looking at a more durable way to deal with content so that different navigation schemes can be used over time.”

– R. Daniel “Taxonomy FAQs”
Effectiveness of taxonomies

- Categorize in multiple, independent, categories.
- Allow combinations of categories to narrow the choice of items.
- 4 independent categories of 10 nodes each have the same discriminatory power as one hierarchy of 10,000 nodes ($10^4$):
  - Easier to maintain.
  - Easier to reuse existing material.
  - Can be easier to navigate, if software supports it.

<table>
<thead>
<tr>
<th>Main Ingredients</th>
<th>Meal Type</th>
<th>Cuisines</th>
<th>Cooking Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chocolate, Dairy, Fruits, Grains, Meat &amp; Seafood, Nuts, Olives, Pasta, Spices &amp; Seasonings, Vegetables</td>
<td>Breakfast, Brunch, Lunch, Supper, Dinner, Snack</td>
<td>African, American, Asian, Caribbean, Continental, Eclectic/Fusion/International, Jewish, Latin American, Mediterranean, Middle Eastern, Vegetarian</td>
<td>Advanced, Bake, Broil, Fry, Grill, Marinade, Microwave, No Cooking, Poach, Quick, Roast, Sauté, Slow Cooking, Steam, Stir-fry</td>
</tr>
</tbody>
</table>

42 values to maintain ($10+6+11+15$)
9900 combinations ($10x6x11x15$)
Why build and apply a Taxonomy? Taxonomy enables usability and re-usability

- Tagging content enables content re-use and dynamic web publishing.
- Tagged content exponentially increases the ability to aggregate related content, making it easier to present users with relevant content.
- The presentation of relevant related content provides users with a “scent” or context.
- Readily offering content-related web services—RSS feeds, bookmarking, user tagging—provide a more rewarding experience.
- Googlers are oriented—even when they land on a page fifteen layers deep.
Taxonomy trends on the Web

- Lesson from shopping websites – a few small taxonomies, rather than one large one.
  - Easier to tag
  - Easier to build services
  - Easier to maintain
Hi all,

I am forwarding a question from my department head, in hopes that you may have leads on this topic.

Does anyone have experience with material that is being produced on campuses (teaching materials, lab experiments, etc.) and uploaded to youtube. are any of you aware of staff that is cataloging this material? or any discussion around this in the larger community? seems valuable to capture and expose.....

Thanks,

Louise Ratliff

UCLA Library Cataloging & Metadata Center

Metadatalibrarians mailing list
Metadatalibrarians@lists.monarchos.com
http://lists.monarchos.com/listinfo.cgi/metadatalibrarians-monarchos.com
YouTube metadata
You Tube taxonomies

Channel
- All
- Comedians
- Directors
- Gurus
- Musicians
- Non-Profit
- Partners
- Sponsors
- YouChoose 08

Audience
- Autos & Vehicles
- Comedy
- Education
- Entertainment
- Film & Animation
- Howto & Style
- Music
- News & Politics
- Nonprofits & Activism
- People & Blogs
- Pets & Animals
- Science & Technology
- Sports
- Travel & Events

Category
- Autos & Vehicles
- Comedy
- Education
- Entertainment
- Film & Animation
- Howto & Style
- Music
- News & Politics
- Nonprofits & Activism
- People & Blogs
- Pets & Animals
- Science & Technology
- Sports
- Travel & Events

Rating
- Featured
- Most Subscribed
- Most Viewed
- Most Discussions
- Most Members
- Most Videos
- Most Active
- Most Discussed
- Most Recent
- Most Members
- Most Responded
- Most Viewed
- Previously Popular
- Top Favorites
- Top Rated

Audience & Category use the same controlled vocabulary
You Tube services:
From just a little bit of tagging

- Populate all Video, Channel & Community pages.
- Automatic RSS subscriber based on tags for any page.
- Identify more videos from the same creator.
- Identify related videos.
- Create, publish & monetize (place ads) your playlists & favorites.
Taxonomy development overview

- Any taxonomy development process needs to engage key stakeholders such as:
  - End users
  - Content managers
- Engaging busy professionals and bureaucrats so that they have the maximum impact with the minimum effort is a tricky business.
Singapore Government-wide Metadata Standard (SGMS)

Case Study
SGMS project objectives

- Enable effective and efficient searches that meet the 30% improvement in search performance target.
- Provide taxonomy to ease searching and retrieving content.
- Establish government-wide taxonomy and metadata standard.
  - Singapore Government-wide Metadata Standard (SGMS)
- Facilitate automated meta-tagging of content for agencies.
## Domain and Subject Owner participation

<table>
<thead>
<tr>
<th>Stage</th>
<th>Domain Owners</th>
<th>Subject Owners</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Content Analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Content analysis</td>
<td>✓</td>
<td>✓</td>
<td>Jun-Jul 04</td>
</tr>
<tr>
<td>b) Mass briefing</td>
<td>✓</td>
<td>✓</td>
<td>05 Aug 04</td>
</tr>
<tr>
<td>c) Pre-Alpha Taxonomy workshops</td>
<td>✓</td>
<td>✓</td>
<td>06 Aug 04 – 17 Aug 04</td>
</tr>
<tr>
<td><strong>2. Taxonomy Build</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Alpha Taxonomy workshops</td>
<td>✓</td>
<td>✓</td>
<td>28 Sep 04 – 08 Oct 04</td>
</tr>
<tr>
<td>b) Beta Taxonomy workshops</td>
<td>✓</td>
<td>✓</td>
<td>17 Nov 04 – 26 Nov 04</td>
</tr>
<tr>
<td>c) Taxonomy Release 1.0 (Final)</td>
<td>✓</td>
<td>✓</td>
<td>20 Dec 04 – 10 Jan 05</td>
</tr>
<tr>
<td><strong>3. Taxonomy &amp; Methodology Use</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Review Standard &amp; Methodology</td>
<td>✓</td>
<td></td>
<td>17 Nov 04 – 10 Jan 05</td>
</tr>
<tr>
<td>b) Domain Owners Training</td>
<td>✓</td>
<td></td>
<td>11 Jan 05 – 24 Jan 05</td>
</tr>
<tr>
<td>c) Subject Owners Training</td>
<td></td>
<td>✓</td>
<td>07 Jan 05 – 17 Jan 05</td>
</tr>
</tbody>
</table>
Pre-Alpha SGMS Taxonomy

- Taxonomy development requires multiple stages
- The Pre-Alpha is a first cut taxonomy used as a discussion basis with the Domain teams
- The inputs to the pre-alpha are:
  - Crawled content from existing Singapore Government websites.
  - Review of US Govt & Commonwealth countries’ taxonomies (UK, Canada, NZ, Australia)
  - Best practices, background & experiences of the taxonomists.
Proposed high-level faceted SGMS Taxonomy

- **Recommendation:** Large faceted taxonomy
  - Attributes shared by all content areas.
  - Some universal facets, some optional facets.
  - One facet, Topics, is a subject taxonomy with sub-facets for various Subject Areas (Domains)
Pre-Alpha SGMS Taxonomy workshop objectives

- **Goal:** To identify broad and narrow Subject Areas in each Domain.
  - List of Subject Areas will become more detailed later in project. At this stage we want to identify the broadest and most important Subject Areas.
- To discuss Subject Area and e-Services definitions, identify core content and information services.
- To review and discuss the Pre-alpha version of the Taxonomy, and each domain’s Subject Area hierarchy.
- To gather feedback on how things should be changed for the Alpha version
  - Identify additional Subject Areas and synonyms.
  - Determine additional Subject Owners, if applicable.
  - Identify additional sources for existing vocabularies.
  - Discuss Subject Areas that may overlap across Domains.
## Taxonomy Workshops Schedule

<table>
<thead>
<tr>
<th>Date and Time</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>06-Aug-2004, 9:30am - 12:00pm</td>
<td>Transport &amp; Travel</td>
</tr>
<tr>
<td>11-Aug-2004, 9:30am - 12:00pm</td>
<td>Family &amp; Community Development</td>
</tr>
<tr>
<td>11-Aug-2004, 9:30am - 12:00pm</td>
<td>Defence &amp; Security</td>
</tr>
<tr>
<td>11-Aug-2004, 2:30pm - 5:00pm</td>
<td>Health &amp; Environment</td>
</tr>
<tr>
<td>12-Aug-2004, 9:30am - 12:00pm</td>
<td>Business (Workshop #1)</td>
</tr>
<tr>
<td>12-Aug-2004, 2:30pm - 5:00pm</td>
<td>Education, Learning &amp; Employment</td>
</tr>
<tr>
<td>13-Aug-2004, 2:30am - 5:00pm</td>
<td>Housing &amp; Property</td>
</tr>
<tr>
<td>16-Aug-2004, 9:30am - 12:00pm</td>
<td>Infocomm Technology</td>
</tr>
<tr>
<td>16-Aug-2004, 2:30pm - 5:00pm</td>
<td>Culture, Recreation &amp; Sports</td>
</tr>
<tr>
<td>17-Aug-2004, 9:30am - 12:00pm</td>
<td>Finance &amp; Human Resource</td>
</tr>
<tr>
<td>17-Aug-2004  2:30pm - 5:00pm</td>
<td>Business (Workshop #2)</td>
</tr>
<tr>
<td>18-Aug-2004, 9:30am - 12:00pm</td>
<td>Public Administration</td>
</tr>
</tbody>
</table>
### Sample Worksheet: Transport and Travel Subject Area

<table>
<thead>
<tr>
<th>Key Subject areas</th>
<th>Broad Subject Area (e.g. “Transport”)</th>
<th>Narrow Subject Area (e.g. “ticket”)</th>
<th>Location (e.g. “CHY”)</th>
<th>Organization (e.g. “CAA”)</th>
<th>Content Type (e.g. “Map”)</th>
<th>Audience (e.g. “Passenger”)</th>
<th>Industry (e.g. “Shipbuilding”)</th>
<th>e-Service (e.g. “Check Scheduler”)</th>
<th>Synonyms, Related subject areas, and Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Transportation</td>
</tr>
<tr>
<td>Aviation</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Aeronautics, Airlines</td>
</tr>
<tr>
<td>Airports</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Aerodromes, “Changi Airport” is specific location, “CAAS” is related organization.</td>
</tr>
<tr>
<td>Flight schedules</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>“Check schedules online” is a particular service</td>
</tr>
<tr>
<td>Car Parks</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Parking Lots</td>
</tr>
<tr>
<td>Dockyards</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cargo, Transport of Goods</td>
</tr>
<tr>
<td>Freight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Transport</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Trains, Light Rail</td>
</tr>
<tr>
<td>Buses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Railways</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxis</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Next 3 Months Activities...

- Additional feedback by email after the workshops.
- Follow-up meetings to gather more detailed comments.
- Taxonomy Strategies continues refining the taxonomy from the bottom-up and the top-down, based on feedback from workshops and follow-up meetings.
- Alpha Taxonomy will:
  - Finalize top-level of facets and top-level of Subject Areas.
  - Provide draft of next levels of facets and Subject Areas.
- Alpha Taxonomy circulated.
- Another round of workshops to review the Alpha version and work on next levels of facets and Subject Areas.
Singapore Government Online Search

Technology: Autonomy

http://mysearch.internet.gov.sg/

Facetted taxonomy navigation

Typical type & go full text relevance ranked search
Singapore Government Online Search

Your search contains:
- Industry: Manufacturing
- Topics: Science & Engineering
- Topics: Electrical & Electronic Engineering
- Topics: Robotics

Search within results

Browse Results by
- Topics
- Organisation
- Audience
- Industry
- Location
- Content Type

Filter & sort results. Click to see options.

Results 1-9 of about 9

A*STAR Corporate Site - About A*STAR - News - Press Release Details
These include four new awards, such as the A*STAR Best New Comer Award, which aims to recognise the fighting spirit of new participants, the A*STAR Most Creative Journal Award and the A*STAR Best Cheer Team Award. In view of the overwhelming response to the A*STAR Robotics Games 2002 - Dancing Robot and Bmobile Side-show.

More from this site | Similar Pages

Nanyang Polytechnic - School Of Engineering

More from this site | Similar Pages

Engineering Week 2004
During the workshop, participants will experiment and find the best way for the robot to hit the golf ball cleanly off the tee and finish with a good follow-through. 2. Fun with Design and Modeling Participants will be exposed to simple 3D software modeling techniques used in engineering design.

More from this site | Similar Pages

Team Champion - BOON LAY PRIMARY SCHOOL
ROBOCHAMPION credits Thanks to: We would like to thank the following people for their guidance that leads to our success. We have little knowledge about robots but we are ready to face the challenges in the future for further studies about robot. The Principal of Boon Lay Primary School, Mr www boonlaypri.moe.edu.sg/blps/ADD/robotics/robochampioncredits.html

More from this site | Similar Pages
EPA Web Taxonomy

Case Study
EPA Taxonomy objectives

- Provide a single methodology for categorizing information across EPA offices, programs, and regions.
  - Provide incentives for quality user-contributed metadata (search results, more links)
  - Provide almost instantaneous feedback using that metadata.
- Reduce the time it takes to successfully target and find cross-Program/Region information
  - Enable and enforce content linking across the agency
- Build common agency-wide terminology resources
  - Eliminate multiple, ambiguous taxonomies
  - Eliminate multiple glossaries, abbreviations and acronyms
- Group things differently depending on the context
  - e.g., ground water with drinking water, or ground water with water quality
- Make it easy to change website organization without having to re-index or re-code pages.
Case Study: EPA web taxonomy and metadata project

1. Define the value of taxonomy
   - Interview stakeholders.
   - “As is” and “to be” review—website taxonomy, tagging procedures, editorial guidelines, business objectives, etc.
   - Enterprise terminology value proposition.
   - Cost benefit analysis model.

2. Metadata and vocabulary spec
   - Inventory & normalize keywords, vocabularies & classification systems.
   - Factor out discrete facets (people, places, organizations, and things) from existing vocabularies.
   - Metadata & taxonomy specification.
   - Initial testing & agreement.

3. Governance and training
   - Training materials & staff training.
   - Agency content tagging.
   - Editorial & maintenance process guides.
How to define taxonomy value

- Interview staff involved in managing content.
- Review interview notes to identify themes & issues.
- Review internal & external resources mentioned in interviews.
- Review existing agency websites.
- Prepare cost benefit analysis.
- Facilitate business case working session.
EPA website search
Technology: Northern Light
http://nlquery.epa.gov/epasearch/searchutils/folders.html

Website taxonomy navigation trees with the term “water”

Typical full text relevance ranked search results for term “water”

Full text search results for term “water” grouped by associated Topics and Types
Oracle.com Taxonomy Project

Case Study
Oracle.com taxonomy project

- Clean-up existing taxonomy
  - Validate for completeness, arrangement of categories & relationships.
  - Compile synonyms.
- Expand taxonomy
  - Include products, product lines, industries, services, solutions, technologies, locations, languages, content types, people & audiences.
- Test taxonomy
  - Build automated tagging application, re-tag Pressroom & mount in hosted Seamark environment.
  - Validate Pressroom with old & new taxonomies side-by-side.
- Develop taxonomy editorial & governance guidelines.
- Identify and prioritize collections to tag with taxonomy.
- Replicate & extend taxonomy tagging & navigation to new collections.
Oracle.com top-level taxonomy architecture
Oracle press release finder

Technology: Siderean
http://pressroom.oracle.com/

- Traditional search
- Time line search widget
- More filters based on this result
- Subscribe to RSS feed based on the criteria set on this page
- View details, or email info
- Add keyword tags
- Bookmark & share
Oracle event finder

Technology: Siderean
http://events.oracle.com/

- Traditional search
- Filter on Location “Germany” and Date “This Month”
- Time line search widget
- More filters based on this result
- Results shown on Google maps UI
- Subscribe to RSS feed based on the criteria set on this page
- Register, view details, or email info
What is the semantic web

- Making content web-accessible in a format that can be read and used by automated tools, so that people and machines can find, share and integrate information more easily.

- Some current examples, especially if they use semantics as the basis for that integration.
  - **Dynamic Web Pages** – Content changes in response to different contexts or conditions.
  - **Personalization** – Tailoring to a user based on personal details or characteristics they provide.
  - **Mashups** – Combining data from more than one source into an integrated application.
Questions?

Joseph A. Busch, 415-377-7912,

jbusch@taxonomystrategies.com