

Final Assignment

Stuart Maxwell

IMT 530B - Organization of Information Resources
Winter Quarter 2010
Mike Doane, Instructor

Part I: Authority Control Document

Summary

This is the authority control document for a taxonomy of IT documents on Galera, an intranet site. The full list of information objects for this version of the taxonomy can be viewed at the following URL: <http://www.evernote.com/pub/stumax/msimimt530classproject>. For the complete list of controlled vocabularies, refer to spreadsheet "IMT530B Final - Maxwell.xls".

Intranet address: <http://galera.herrera.local/>

Client: Herrera Environmental Consultants, Inc., Seattle, WA

Date: March 14, 2010

Taxonomy version: 1.0

Taxonomist: Stuart Maxwell

Governance

The taxonomy will be controlled by Herrera's Technical Communications Manager.

Refer to ANSI/NISO Z39.19-2005 for editorial guidance.

Maintenance

The taxonomy will be reviewed quarterly, but may be updated more frequently, if necessary. Quarterly maintenance will include a review of site search logs to identify search trends, and especially to review most popular searches and failed searches. The Technical Communications Manager will also take into consideration suggestions from employees, emails about the site, inquiries to the IT help desk, and keywords entered by users in the Tags field. In addition, user testing should be employed to validate the efficacy of the taxonomy.

Scope

The subject taxonomy provides a metadata scheme that describes discrete documents on Galera. ("Documents" here includes journal posts and discussions.) The current version of the taxonomy describes only documents found under the **Employee Info > IT Help** space of the site.

The referenced taxonomy spreadsheet lists a subset of 40 documents available on IT Help, out of 137 total documents available. It is intended that describing this subset will provide sufficient precedent for describing the remaining documents. Some decisions have also been made with an eye to expanding the

taxonomy to describe the documents in other spaces on Galera, with the eventual goal of bringing all content under one common space.

Content Scheme Objectives and Rationale

The goal of this taxonomy is to support a task-oriented approach to finding information on Galera. The strategy is to organize content primarily by Category and Action. The Category facet is comprised of nouns on which some Action can be taken. The search and browse schemes should map to a variety of users' mental frameworks for accomplishing a task. So, if a user thinks "I need to set up my email account," he could approach the site in any of the following ways to find the same result:

User thinks...	...intranet search/browse affords
I need to set up my email	User chooses "set up" from the Action facet, then sees all things that can be set up, including email.
My email needs to be set up	User chooses "email" from Category facet, then sees all available actions that can be taken with email, including "set up".
How to I go about setting up my email?	User chooses "setting up" from Activity facet and sees an associated list of items (Categories and Products) to be acted upon.

The metadata schema is borrowed from two main sources: the Australian Government Locator Service (AGLS) metadata standard, and the schema for MSWeb, Microsoft's intranet.

The facets and controlled vocabularies were developed through a task analysis of Galera users, and terms were drawn first from vocabulary used within the subject documents, and then supplemented with terms and product names used by vendors of Herrera's IT products (literary warrant). User warrant was given priority in determining preferred terms.

The list of verbs in the Action facet was drawn from Blooms Taxonomy and supplemented by verbs suggested by users. Search logs should supply suggestions for future additions to the Action facet. It is expected that some verbs (e.g., learn about, use) will be less helpful in narrowing down a search. But some (e.g., sync, map, conference in) will be very useful because of the narrower context in which they are likely to occur.

The Activity facet is a list of the gerund forms of terms in the Action facet. The Activity facet exists to provide flexibility to the user in constructing a search. So,

for example, a query built either as “file restore” or “restoring files” will return the same result set.

User Stories

Jane is a wetland scientist at Herrera. She frequently works on reports, but she's not terribly tech savvy, so she needs help using email and finding her way around Word and Excel. She also suspects she has a virus on her computer, and wants to know who to turn to for help.

Bill is a new engineer at Herrera. He needs to figure out how to set up his email and voice mail, how to use devices like copiers and phones, and how to find and save files on the network.

Erika is a hazardous materials scientist getting ready to do a site assessment at a remote location. Since she's going to be on the road, she needs to figure out how she's going to communicate with her client and colleagues from her hotel. She'll also need to check out some equipment, and set her out of office messages for email and phone.

Facet definitions and usage guidelines

Galera runs on Jive Software's Clearspace. Several of the facets referenced below will draw from metadata that is automatically captured by Clearspace.

General

GUID: Globally Unique Identifier. Each information object has a unique ID.

URL: The web address of the resource. This is automatically generated by Clearspace when a document is published.

TITLE: The name of the resource. This is entered by the user prior to publishing the document.

Descriptive Metadata

CATEGORY: Terms from the vocabulary of category labels; these are generic nouns, not specific products, and are derived from terms contained within documents on Galera.

ACTION: A list of verbs; things you can do to or with a category label (e.g., change, delete, fix, scan). Verbs are drawn from Bloom's Taxonomy and supplemented with verbs suggested by users.

ACTIVITY: The gerund form of terms found in the Action list. Selection of compound terms has been adjusted to suit natural language preferences (see ANSI/NISO Z39.19-2005, Section 7.8). The activity facet is intended to add flexibility in constructing a query.

DESCRIPTION: A brief description of the resource; suitable for display in a search result. The description will be entered by the author prior to publishing the document.

TAGS: This facet allows free text entry of keywords by users. The vocabulary is not controlled, but Clearspace will programmatically capture these terms and present them as suggestions for other users. It is expected that tags will be reviewed periodically for inclusion in the controlled vocabularies in this taxonomy.

Administrative Metadata

PRODUCT: Terms that describe the specific subject matter of the resource; can include product, standards, and technology names. Terms are drawn from the vocabulary used in documents on Galera, and supplemented with proper industry-standard or vendor-specific terms as appropriate (by literary warrant).

CREATOR: Resource creator (UF author). The document creator is the person logged on to Clearspace who publishes the document. Generated automatically by Clearspace.

CONTRIBUTOR: Resource contributor (e.g., editor; co-author). A document contributor has saved a version of the document on Clearspace. Generated automatically by Clearspace.

VENDOR: If applicable, the name of the company that provides or created the software, hardware, or service. The provider is preferred; so, for example, enter the phone vendor (Firstline) instead of the manufacturer (Polycom).

CONTACT: The name of the person or group responsible for the resource. In this taxonomy, IT is always the contact, but as the taxonomy moves out to other sections of Galera, the contact will change as appropriate.

SUBJECT: A short list of high-level concepts that describe the content of the document. Used mainly behind the scenes to group related items in search results.

LOCATION: If applicable, the Herrera office where the resource is used or located. If the document applies to all office locations, leave this field blank.

STATUS: The record's status; e.g., "active" (the default), "deleted," "inactive," and "suggestion";

COMMENT: Administrative information that helps manage a record (not seen by the end user)

Structural Metadata

ID: Unique ID; shorter than the GUID, for use in URLs. Generated automatically by Clearspace.

CREATION DATE: The date the resource was created. Should be displayed in full text, as, for example: May 28, 2009. Generated automatically by Clearspace.

REVIEW DATE: The date that the resource should be next reviewed. Entered by the user; default setting is six months from when the record was created or last updated. Should be displayed in full text, as, for example: May 28, 2009.

DATE MODIFIED: The date the resource was last modified. Should be displayed in full text, as, for example: May 28, 2009. Generated automatically by Clearspace.

TYPE: The medium through which the resource is presented.

Editorial Guidance Notes

Unless otherwise noted, refer to ANSI/NISO Z39.19-2005 for editorial guidance.

STEMMING: Stem all terms except for proper nouns, or if an exception is noted in the scope notes.

PRECOORDINATED AND COMPOUND TERMS: Precoordinated and compound terms are permitted when the resulting term has a specific meaning that is not implied by the separate, uncoordinated terms. The rationale for specific precoordinated and compound terms are referred to in the scope notes.

Part II: Project Review

Introduction

Herrera Environmental Consultants maintains a knowledge sharing intranet site built on Jive Software's Clearspace collaboration software. Known as Galera internally, the site's siloed organization reflects the company's previous approach to capturing knowledge as a series of Word and Excel files kept in a network folder structure. Content on Galera is organized by department, as it was in the network folder, yet employees often complain that finding the information they need is difficult. With recent reductions in IT and Admin support staff available, better information retrieval is a critical part of supporting business needs.

Approach

To develop a new approach to organizing information on Galera, I started with a task analysis to determine what intranet users at Herrera might search for. To keep the first phase of development manageable, I narrowed down the field of potential information objects to documents in the IT Help section of the site. I reviewed current content on the site as well as drawing on my own deep experience with the site. I validated my observations through informal interviews with users, especially with subject matter experts in the IT group. (Stewart, 2008)

This review of content generated a list of terms which I captured in FreeMind, a mind mapping program. Sorting this list revealed general subject terms as well as specific product and technology terms. At the Boxes and Arrows website I found an article that described the schema for MSWeb, Microsoft's intranet site, and their use of Category and Product facets seemed a logical starting point for my taxonomy.

In order to support a task-based approach to finding documents, I tried to design a scheme that would support users forming action-based queries. Using Blooms Taxonomy as well as drawing terms from the existing documents, I developed a list of verbs that might be associated with Category and Product terms. (Anderson, Sosniak, Bloom, & National Society for the Study of Education, 1994)

This approach resulted in a relatively flat set of term lists, one nearly devoid of hierarchy. While this flies in the face of much of what we've learned this quarter, it also seemed to make some sense within the context of this intranet site. With a relatively small set of terms to categorize and a low context environment to work within, findability should be enhanced by keeping the facets shallow. And, to some extent, hierarchy is represented within the structure through the Vendor and Subject facets, each of which are short lists that can contain a large number of related terms.

Furthermore, using verbs (what am I trying to do?) and nouns (what am I trying to affect?) should support “probing” the system through a contextual approach to information retrieval. The addition of the Tags facet allows a user to further personalize context within the system.

The schema as designed is intended to afford the users the following views at minimum:

- an alphabetical list of all products by their preferred term
- a list of all products filtered by category
- a list of all categories filtered by action or activity
- a list of all products filtered by action or activity
- a list of actions or activity filtered by category or product
- a list of categories, products, actions, or activities filtered by subject

This flexibility should help users quickly find the content they’re looking for within a few clicks.

Implementation

As I moved from the midterm to the final taxonomy, I started by clarifying the Subject and Category facets. Subject became relegated to the Administrative metadata, and serves the supporting role of facilitating return of related items into search/browse results. The Category facet now contains generic nouns that describe higher concepts (as opposed to terms in the Product facet, which identify specific products or technologies).

Nailing down the Category facet was by far the most challenging part of building out the final taxonomy. It wasn’t until I went through each information object and tried to use the controlled vocabularies from the midterm that I discovered what worked and what didn’t... and there was a lot that didn’t. I did some significant modifications to the Category facet in order for the pick list to make more sense. I’m happy with how it’s turned out now, but it was a frustrating process to get everything to fall into place.

I also discovered that I needed a Vendor facet. I did this so that I could include Konica Minolta and other vendors without adding them as top levels in the hierarchy. I’m really not sure if this was the right approach, but my rationale was that, for instance, the full proper name of a copier is "Konica Minolta BizHub 350"; however, most of our users would search by the copier name (Brown) or model (BizHub 350). So I created Konica Minolta as a Vendor term and included BizHub 350 and Brown as related terms. In the end, I think that having the Vendor facet will help connect related items. And, it provides a useful search facet.

Extending the controlled vocabularies was difficult at times. The term “copier” was a particular challenge. Within the current selection of IOs, all printers are also copiers. However, that may not remain the case, and Herrera does have

printers that are not also copiers. When building this taxonomy out for the real site, I would probably add printer as a term under the Category facet.

In general, as I settled on specific definitions for each facet and understood better the role each would play in the final taxonomy, the vocabulary for each facet became more obvious. In the end, I was satisfied that my facets were properly chosen and my controlled vocabularies were sufficient to describe the information objects.

My concept of what an Authority Control should look like is still hazy, but discovering the documents at the IPSV site was tremendously helpful. I liked their approach of developing documents such as Editorial Policy, Maintenance Guide, and Guidance Notes. These documents provided some structure for my own authority control, as well as informing my approach to the content of my AC.

Impact

Our employees are like employees at most other businesses: they've got too much to do and not enough time to do it. According to a survey by Accenture, "Managers spend up to two hours a day searching for information, and more than 50 percent of the information they obtain has no value to them." (Business Wire, 2007) Employees at Herrera are certainly no different in this regard, either, so efficient information design is critical to helping our staff do more billable work, and work of higher quality. Furthermore, recent reductions in the IT and administrative support staff have made it more critical that employees are able to easily and quickly find information on Galera.

My observation is that people mostly use the intranet when they have a task to accomplish (as opposed to just browsing for general information). If so, then a task-based approach to sorting and displaying information should dramatically improve our employees' experience with Galera. The addition of a Review Date – something we don't currently track – will help ensure that the information employees do find will be relevant and up to date.

Problems

I'm not sure I fully captured all the relationships implied in my facets, nor am I sure that the approach of avoiding a deep hierarchy in the controlled vocabularies was the right one. When I expand this taxonomy, it will probably turn polyhierarchical. For example, "spam" exists only in the Category facet, but it should probably also appear in the "Product" facet.

As I expand the taxonomy to include other documents outside of the IT Help space, I'm likely to encounter unexpected issues. Other spaces contain more complex documents (documents that cover multiple topics). For the one complex document that I dealt with in this project, I was able to apply multiple terms in a

way that made sense within the spreadsheet, but I'm not clear that it will be the right approach in the real world.

The Subject facet will probably also need some clarification. The basic idea is for the facet to serve the administrative function of relating content using a few broad terms, but this may be redundant or unwieldy; it's a weak spot in the current taxonomy that will need to be addressed.

I would also put more effort into finding a prebuilt taxonomy that could be used for Galera. Selecting terms internally was the right approach for this project, but may become an overwhelming task as the taxonomy evolves.

There might be a significant amount of work to be done were this taxonomy to be moved into an ontology. The documents in the IT Help space on Galera have a limited relationship to the majority of documents that Herrera generates. The greatest benefit to Herrera would be if there were a way to connect the internal document set to some external repositories; in other words, to show users what documents are located on external web sites that relate to the one they're currently viewing.

Certainly, Herrera could make use of an ontology for mapping the hundreds of thousands of documents contained on the network. As an interdisciplinary environmental firm, the company depends on the synergies created among the various scientists, engineers, and planners working at the company. Yet, much of the company's work product is more or less invisible to employees, as details of the location and subject matter of the documents are mostly held in the heads of the authors, project managers and principals. An ontology that allowed easy discoverability of documents based on concepts as well as titles, authors, and project location would be of great benefit to the company. But that will have to be a project for another day.

Resources Used

Australia, N. A. o. (2002). *AGLS Metadata Element Set, Part 1: Reference Description, Version 1.3.*

Rosenfeld, P. M. a. L. (2002). *MSWeb: An Enterprise Intranet. Boxes and Arrows.* Retrieved from http://www.boxesandarrows.com/view/msweb_an_enterprise_intranet_1

These two resources formed the framework of my schema. The MSWeb schema has proven itself in a high-volume, high-profile intranet, and the AGLS schema was useful for its focus on categorizing documents.

(IPSV), I. P. S. V. IPSV-Integrated Public Sector Vocabulary. 2.0. Retrieved March 14, 2010, from <http://www.esd.org.uk/standards/ipsv/>

I used the documents at the IPSV site to provide structure and inspiration for my Authority Control document, especially for an approach to maintenance and editorial control.

Anderson, L. W., Sosniak, L. A., Bloom, B. S., & National Society for the Study of Education. (1994). *Bloom's taxonomy : a forty-year retrospective*. Chicago, Ill.: NSSE : Distributed by the University of Chicago Press.

Bloom's Taxonomy verbs, referenced in this work, formed the basis of my Action and Activity facets.

NISO. (2005). ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. Bethesda, MD: National Information Standards Organization.

The Z39 standard provided invaluable guidance for the selection of terms and for an overall editorial approach to building the taxonomy.

Stewart, D. L. (2008). *Building Enterprise Taxonomies*: Mokita Press.

Stewart was particularly useful in providing guidance on collecting terms internally, and for structuring the Authority Control document.

Managers Say the Majority of Information Obtained for Their Work Is Useless, Accenture Survey Finds. (2007). *Business Wire*. Retrieved from http://www.businesswire.com/portal/site/google/index.jsp?ndmViewId=news_view&newsId=20070104005159&newsLang=en

This article provided support for the impact of creating a useful and efficient taxonomy.