SPE 90006

Taxonomy: A Knowledge Sharing Enabler

David L. Smith, Jr., SPE, Halliburton Energy Services, Joseph Busch, Taxonomy Strategies, Ron Daniel, Jr. Taxonomy Strategies

Abstract

As companies have begun to share information using intranets and internets, more information is available than the average person can absorb. A simple internet search using any search engine can return hundreds to thousands of documents that may contain only bits of information that one is required to integrate into useful knowledge. Successful implementation of internet and intranet portals, knowledge management, and content management requires a systematic approach to classifying information. The classification framework, or taxonomy, allows content owners to categorize their content so that it becomes more useful to a broader audience. Taxonomy allows knowledge seekers, whether they are researchers or the general public, to easily find and re-use information rather than recreating it.

Taxonomies are not new; in fact most work groups organize their information around a hierarchy that developed over several years. In most corporate environments, each product line or functional group organizes its information from its internal viewpoint rather than from a broader viewpoint. Use of a broader viewpoint would however, encourage knowledge sharing beyond the product line or group. To be successful in the digital world must eventually merge these independent hierarchies toward a common taxonomy.

This paper details the development of an enterprise-wide taxonomy for a diverse, global oil and gas services company. The paper describes how the taxonomy has been implemented, including the application of taxonomy in a comprehensive content management system, knowledge management initiatives, and effective portal design. Challenges and lessons learned are also presented, the most significant of which are issues surrounding the means to get people to think differently and do things in different ways.

Background

Our company's previous content management efforts had evolved from published hard copy content to placing anything that was important to anyone on the intranet. Although document control procedures had been used for hard copies, controls were typically not properly applied to electronic information. This problem was exacerbated because of mergers and acquisitions, bringing in more collections of content which followed different policies regarding creation, organization and retention. The gains made in sharing were lost in information quality, with many authors forgetting about their content altogether. Thus, text searches retrieved too much unwanted content that had questionable information.

In many websites, content was broadly organized according to the content owner's view of their functional organization. Once written, content was typically sent to a web administrator for actual placement in the system, a step that added to delays in document deployment. Also, commonly no one bothered to see if there were any similar documents already in the system, because it was too difficult to find them.

The seemingly logical placement of content varied with each functional unit. For example, some product lines organized by their business strategy, some by their functional sub-organizations and one actually organized all their content by the geographical area that produced it. One had to actually know where content may have been produced or they would not have a chance of finding it!

Search results returned too much content and there was not a simple way to focus results toward specific topics. Adding to the search problem, terminology varied among the functional groups.

A content management process was needed to ensure that valid content was displayed and an information management solution was needed to deliver content about specific subjects to people seeking the information.

A content management software system was deployed to manage the content. This system allows content to be tagged with attributes defining a content owner, author, review date and expiration date and tagged with taxonomy terms defining the information contained in the document.

A Corporate Taxonomy
The taxonomy described in this paper is a multi-dimensional classification scheme. Each dimension or, taxonomy facet is a hierarchical list of detail taxonomy terms. Information entered in a repository is categorized with many taxonomy terms which can later be used to access it. The taxonomy provides a common way to search the information by people with different viewpoints. Each facet in table 1 has many terms which allow for granular categorization of information.

Table 1. Taxonomy Facets for the Subject Oil and Gas Service Company

- *Content Type*
- *Geographical Location.*
- *Organizations*
- *E & P Lifecycle*
- *Business Process*
- *HSE*
- *Customer Challenges*
- *Product Groups*
- *Product Names*
- *Tools & Component.*
- *Oil, Gas, Chemicals & Lubricants*
- *Other Materials & Equipment*
- *General Subjects*

Detailed taxonomy terms for the Content Type Facet can be seen in table 2 and detailed terms for the Customer Challenges Facet can be seen in table 3.

Table 2. Useful Terms for Content Type Facet
- *Technical Documents*
  - *Research Notes*
  - *Data Sheets*
- *Sales and Marketing Documents*
  - *Advertisements*
  - *Case Histories*
  - *Price Lists*
- *Business Requirements*
  - *Proposals*

Table 3. Terms for Customer Challenges Facet
- *Borehole Challenges*
- *Drilling Challenges*
- *Environmental Challenges*
- *Extending Economic Life*
- *Fluids Challenges*
- *Formation Challenges*
- *Maintenance Challenges*
- *Production Challenges*
- *Real Time Decisions*
- *Reducing Costs*
- *Reserve-Resource Calculation*
- *Reservoir Challenges*
- *Well Control Challenges*

Additional attributes as in table 4, were defined for content management allowing the system to notify content owners for review and allowing content to expire if reviews were not done. Only information that met the review requirements would be published to the end user.

Table 4: Additional Document Management Attributes
- *Title*
- *Description*
- *Author*
- *Content Owner*
- *Launch Date*
- *Review Date*
- *Expiration date*
- *Security Access Control*

Developing the Taxonomy
Each product line and functional organization had already developed a hierarchy with which to present their information as in table 5. There were documents about equipment, processes, and chemicals. Some of these documents related to particular organizational functions or locations. Studying how people from different job roles would look for specific documents led to the facets displayed in Table 1.

Table 5: Intranet Product Line Hierarchies

PL 1
  Chemical Process
  Services
PL 2
  Organization
   Sub-Organization
    Tools
     Content type
PL 3
  Technology Center Location
   Tool Type
    Tool 1
    Tool 2
SAP
  Tools
  Organization
Sales
  Product Groups
   Part of PL 1 + Part of PL 3
  Service area or location

When the individual product line hierarchies were merged into the facets listed in Table 1, redundancies were removed. Since the goal was to create a single language for sharing information, synonyms derived from redundant product line categories were added to the taxonomy terms. This process was used to develop the terms in each facet, for example, in Table 2.

Any previous enterprise wide hierarchy agreements were

considered a priority. SAP was deployed within the subject company in 2000 and agreements about equipment purchased and equipment sold had created an enterprise hierarchy that led to the Tools & Components facet and the Oil, Gas, Chemicals, & Lubricants facet. The SAP hierarchy for our organization also provided the terms for the Organization facet.

Re-use of current enterprise hierarchy terms was considered very important for three reasons:
- Faster initial development of the taxonomy
- Reduced confusion by using existing hierarchy agreements
- Lower maintenance cost for the taxonomy since other groups would maintain the authoritative data.

Look Outside of the Company
There are several examples of taxonomies which can help one develop and validate the internal corporate taxonomy. Any industry standards or general standards that exist can shorten the taxonomy development time and allow for easier sharing of information beyond your company.

Petrotechnical Open Standards Consortium, Inc (POSC) is an international not-for-profit membership corporation which facilitates E&P information sharing and provides open specifications for information modeling and information management (Reference 1). They have an E&P cataloguing initiative to develop and deploy standards for the identification, classification, and cataloguing of both document and data model based E&P information which can be a useful starting point in developing taxonomy. Their current version of the catalogue standards is based on material submitted to POSC beginning in March 2002 from work done by Shell Expro and continuing through refinement submitted through the end of 2002 by Flare Consultants.

The National Aeronautics and Space Administration (NASA) also has an interesting taxonomy web site that can be seen at http://nasataxonomy.jpl.nasa.gov/ (Reference 2). Their taxonomy is very detailed and a good reference.

Purpose of the Taxonomy
An enterprise wide taxonomy applied to documents and other objects provides a means to classify information in common terms that everyone understands. The use of terms in common, allows the classified information to be pushed to portal pages and directly to end users.

Searching on specific taxonomy terms lets an information seeker quickly focus on a set of agreed-to terms and move toward a specific area of interest.

Sharing Information
Content categorized with taxonomy terms can now be deployed automatically to various published sites. The site defines a rule of taxonomy terms and actually searches for content that satisfies the rule. Figure 1 is an example of a

portal tool using taxonomy terms to search for the latest Health Safety and Environment (HSE) information. The system searches for documents using logical taxonomy rules. Clicking on Flash Alerts retrieves 12 documents that are less than 30 days old categorized with the Alerts and HSE organization taxonomy terms.

Figure 1. Display information using taxonomy. The following rules are used for Flash Alerts:
        Content Type = Alerts
        And Organization = HSE
        And Launch Date = < 30 days ago

Additionally one can perform text searches of the content management system and the resulting documents retrieved are displayed along with the taxonomy terms of the entire set of retrieved documents. Figure 2 shows the taxonomy list of CPS documents retrieved in a search. Note that each taxonomy term also indicates the number of documents that are categorized with the term. Clicking on any taxonomy term filters the search result to documents that contain the term. This allows one to quickly reduce the number of retrieved documents to a smaller set that more closely meets their need.

Figure 2: Taxonomy terms related to a set of documents retrieved in a text search.
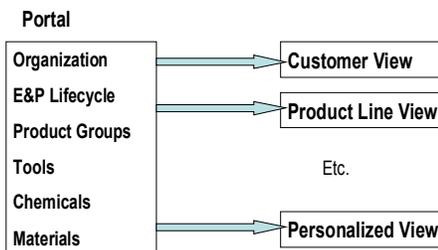
Personalization
The killer application that comes from the taxonomy is the ability to prioritize and deliver information to people focused on personalized requirements. That is, if the system knows

that an information seeker is located in North America and works for a particular product line then searches should prioritize retrieved information toward North America and the seeker's product line.

As an example, the SAP organization hierarchy was an input for the Organization facet. When one looks for information, the system can pull the seekers organization from an identity management database fed by SAP and establish a priority for displaying information tagged with the particular organization's taxonomy term.

The identity management data base drives the personalization functionality. It contains the latest basic information concerning the user's location, job title, product line and other work related information. The presentation in the portal can then be filtered to provide specific views. See Figure 3.

Figure 3: Personalization of displayed information

Information providers can use taxonomy to focus their information toward key concepts (i.e. taxonomy terms) understood by specific information seekers. New and revised information of interest for specific information seekers is automatically displayed in the information seekers portal.
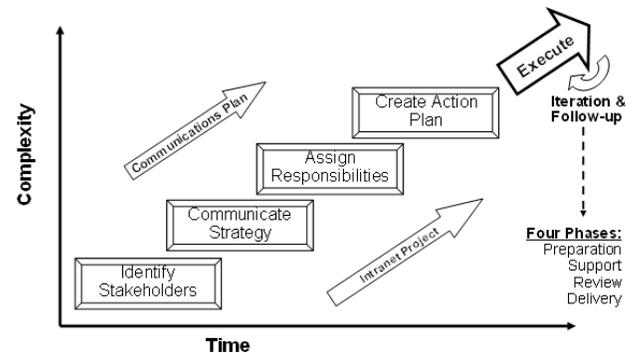
Implementing Taxonomy and Content Management

Since application of taxonomy to documents begins as a manual process, we chose to implement content management on a project by project basis rather than attempting an enterprise wide solution. One of the first projects for the subject company was for a customer portal, with about 6000 documents. Values from taxonomy (Table 1) were used to select documents for display in specific locations in the portal. To minimize the work and ensure proper classification, rules were used to automatically classify documents with common terms most often used by a single content owner. That is, a content owner for a particular product line would have their documents automatically tagged with the product line

taxonomy term and other often used terms. Thus the content owner spends their time classifying the document with the detail terms that differentiated it.

The software systems used in Knowledge Management do not automatically provide value without a process and a culture to share. Enterprise-wide change is called for and requires a clear plan. See Figure 4.

Figure 4: Creating Change Requires a Plan

First identify the stakeholders for the content within a project. Persons, who own, manage and use the content daily will have a stake in any change that will occur. They will need to buy-in to the change and the work required and will ultimately provide the resources to categorize and manage their content.

Second, communicate constantly with the stake holders. They need to know the overall process and how it will affect them and their associates. They will be concerned about any change to how they access and publish their information before and after beneficial content management. A particular issue may be that they previously owned their particular hierarchy yet they now will share taxonomy with the enterprise. Classifying a document with a taxonomy term will allow that document to be shared with everyone interested in that term.

Third, define the stakeholder's responsibilities. It must be clear that the process belongs to the stakeholders and is not a corporate process that will be done for them. The results of their work will not only be used by them but also will be used by others outside of their current influence.

Fourth, create an implementation action plan. Assign tasks and deadlines to move, categorize and publish content.

Finally, the content owners actually do the work of moving information to the content management system, assigning taxonomy terms, reviewing the validity of their content, and publishing it to the enterprise portals.

Lessons Learned

Categorizing content with agreed taxonomy terms creates work that must show some immediate benefit to both content providers and content users. Automation should be used to simplify the categorization process by automatically classifying documents with taxoanomy terms for each content owner. People will not want to categorize information that is not needed for immediate display in a particular website. However, content owners will have an incentive to categorize if the search engine has the ability to use the taxonomy terms to help them find content.

Additionally, content owners will question the taxonomy terms in an effort to rebuild their individual view of content organization. The taxonomy owner must control the taxonomy, but be flexible with proposed alternate terms which help the content owners enhance their view.

In summary:

- Automate/simplify document categorization
- Use a search engine with the taxonomy terms
- Manage the taxonomy from an enterprise viewpoint.

Conclusions

For a large enterprise to share information across diverse product lines and functions, a common language or taxonomy is required to classify the information. The best way to develop the common taxonomy is to look at the hierarchies currently in use. A simple procedure includes:

1. Combine current hierarchies
2. Remove any redundancy
3. Associate alternate terms to the common taxonomy terms
4. Test the proposed taxonomy

Developing the taxonomy is the easy part of a major change management process that is beginning to take place. Storing and managing fresh, quality information that can be retrieved and re-used will make the work place more efficient, but that efficiency is not gained without individuals actually doing the work of managing and classifying information.

References:
1. Petrotechnical Open Standards Consortium, POSC E&P Cataloguing initiative, http://www.posc.org/, February 3, 2004.

2. Dutra, Jayne, The National Aeronautics and Space Administration (NASA), Jet Propulsion Laboratory, http://nasataxonomy.jpl.nasa.gov/ , March 11, 2004.