MONTAGUE INSTITUTE REVIEW

ISSN 1554-303X

December 2007

Content vs. Navigation Taxonomies:
Why are they different?

What do you say when someone asks why the terms in the corporate taxonomy don't match those used on the Public Web site? For example, why use "learning activity" internally and "training" externally? Do organizations need both a "content" taxonomy and a "navigation" taxonomy?

The simple answer is that because a taxonomy is a working tool designed to support a specific business activity, it's inevitable that an organization will need more than one. But the question exposes fault lines in the way many organizations manage taxonomy development and deployment. Why should the question need to be asked in the first place? Why is there such a wide range of opinion about how taxonomies should be implemented in applications, especially search? Why is there still so much interest in taxonomy return on investment?

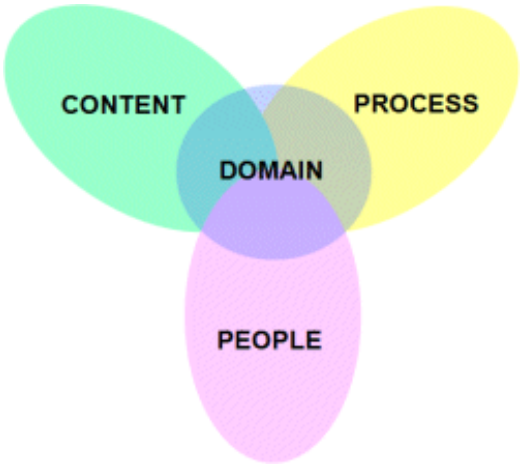
In this article we look once again at why organizations need multiple taxonomies, what are the major kinds of taxonomies needed, what tools should be used to manage taxonomies, and how they should be deployed in applications. The topic is worth revisiting now that many organizations are implementing new search engines and coming to grips with the "bottom up" search and navigation features of collaboration systems like Microsoft Office SharePoint Server (MOSS 2007).

Why multiple taxonomies?

Taxonomies express a certain world view or domain and are used to help people do their work. By "work" we mean buying something, writing a business report, processing an insurance claim, or learning how to do something new.

> A domain combines content, a business process or activity, and a specific kind of user.

As an example, the content (e.g. information about consumer drugs) might be the same



while different processes and/or people might represent three different domains:

• A person with a medical condition wants to buy a prescribed drug at the lowest price.

• A doctor wants to prescribe the most effective drug for his patient.

• A retailer wants to maximize sales by offering other products to people who visit a physical store to fill a prescription.

Each of these three scenarios requires a different taxonomy, although they may overlap. See the two screen shots on the next page.

Common business taxonomies

While there might be dozens of different domains in an organization, in practice taxonomies fall into a limited number of categories based on the activities they're designed to support:

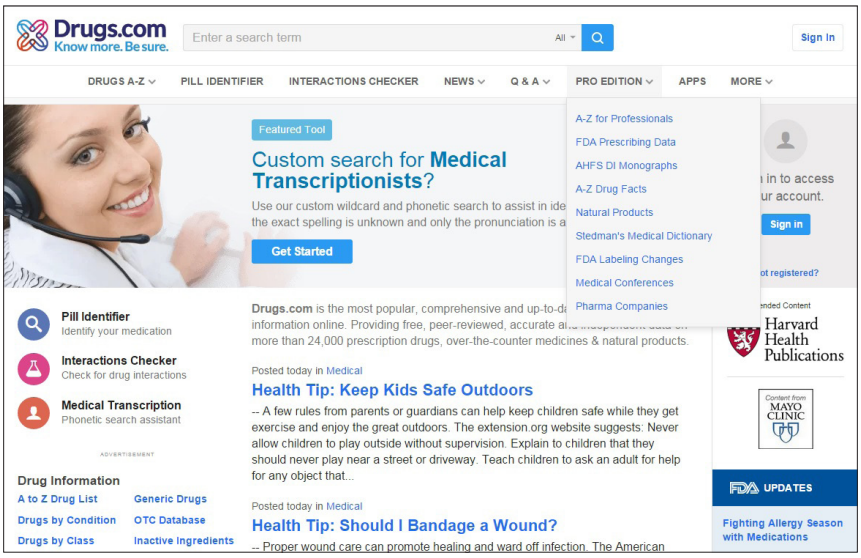
• *Corporate organization chart.* A taxonomy where the categories represent organization units is of most use to managers who need to know who's responsible for what and where their unit fits into the "chain of command." Categories often are closely related to income and expense categories in the accounting system. Department

names, managers, and relationships among departments are constantly shifting. Even if the basic structure is stable, the labels can change. For example, the "Personnel" department becomes the "Human Resources" division. Organizational taxonomies are of little interest to outsiders unless they are looking for a job or doing competitive intelligence.

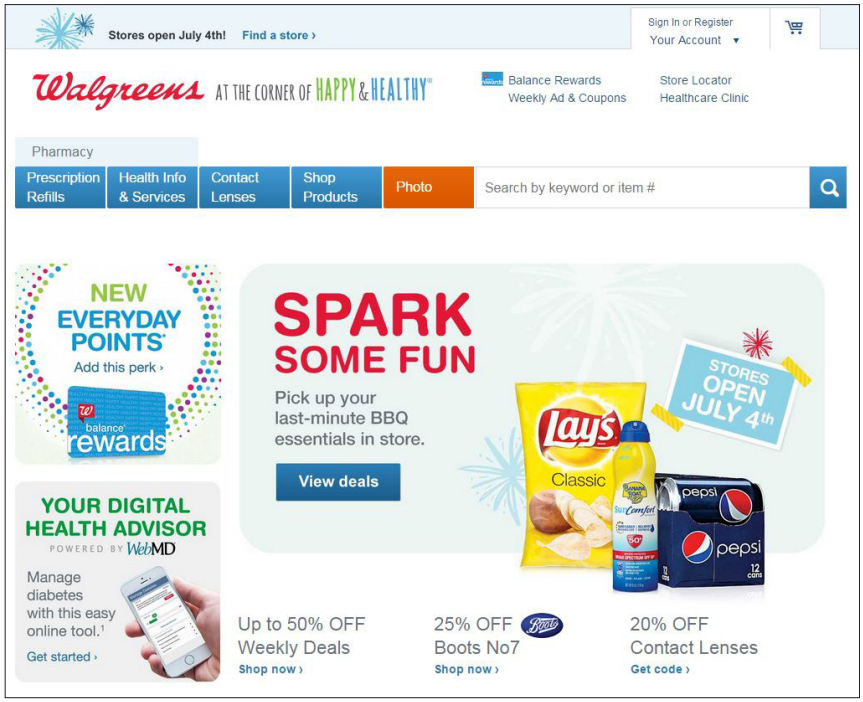
• *Inter-agency collaboration*. Organizations that want to exchange data (e.g. securities brokers and mutual fund companies) or collaborate (e.g. government agencies working with geo-spacial data) need to agree on the terminology used, the meaning of terms, and the relationships among them. Since many of these taxonomies are designed to be used primarily by computers, they must be very precise and detailed. They are slow to evolve because all parties affected must reach a consensus on changes.

• *Navigation*. Users prefer to browse rather than search when they're not sure what to type in the search box, when they want to get an overview of what a work or content collection is "about," or when they're unfamiliar with the lingo used in the collection. In the print and physical world, navigational taxonomies take the form of tables of contents, back-of-the-book indexes, and aisles in a retail store. On the Web, the word "taxonomy" is often used to mean a hierarchical list of topics, similar to a table of contents.

• *Search enhancement*. Full text search programs (search engines) are designed to find documents in which a word or phrase occurs and then rank them according to "relevance." As the technology has matured, these products have become capable of detecting and correcting spelling mistakes (e.g. "restarant" instead of "restaurant") and recognizing words with the same "stem" (e.g. run, runs, ran, running). Some can even group the results into categories or allow the user to narrow the search by selecting a specific "facet" (e.g. geographic region). To get faceted search to work well in a specific organizational setting usually requires that the search



Drugs.com *accommodates the first two domains on the same Web page — the consumer who wants to buy and the doctor who wants to prescribe — in a unique display. The subcategory called "Pro Edition" includes information aimed at doctors and other medical professionals. The A - Z drug index can be displayed for consumers or professionals.*



Walgreens.com *is designed for the third domain — the retailer that wants to maximize sales. Not surprisingly, its display is not as user friendly as drugs.com. Most of the screen (which requires a lot of scrolling) is taken up by ads.*

The objective here is sales, and the focus is on products and services, not so much on the specialized needs of different users. Maximizing sales is a legitimate business objective, but maybe it would be better to break that down into discrete domains, such as products and services for people with specific conditions (e.g. heart disease) who need not only drugs, but also dietary supplements and mobility equipment and who may find it more convenient to buy online rather than visit a store. This might not only boost sales but would also make the taxonomist's job easier!

engine be able to access taxonomy data created or tweaked by humans.

• *Management reporting*. In large companies, management information is typically stored in multiple computer systems, each with its own technology and organization scheme (taxonomy). This is especially true when companies are merged or acquired. To get consistent aggregated reports that allow apples-to-apples comparisons, one taxonomy must become dominant or a new taxonomy must be created. Then all the other taxonomies in use must be mapped to it. Unlike navigational and search engine taxonomies, which should include common terms used by real people, a management reporting taxonomy must above all be consistent and complete.

Looking at business taxonomies in this way answers two questions:

1. Should one taxonomy take precedence over another?
2. How do you show the return on taxonomy investments?

In the general scheme of things, it's clear that no single taxonomy is more important than the others. Which taxonomy you use depends on what you're trying to accomplish — in other words, the specific combination of content, users, and process that constitutes a domain. If your goal is to reduce the time it takes for users to find things on the intranet through browse or search, then you need a taxonomy with common, everyday terms. Look in search logs to find out what people are typing in the search box. Be prepared to have multiple navigation taxonomies for different groups of users (e.g. marketing vs. R&D) or common business activities (e.g. career and retirement planning or qualifying sales leads).

It's not possible to show the value of taxonomies as stand-alone entities, but it is possible to show how a good taxonomy that is well implemented can affect performance by making employees more productive or reducing the cost and time delays of management reporting and inter-agency collaboration. In other words, to show the value of taxonomy, focus on the domain.

Managing multiple taxonomies

Most business taxonomies are stored within the applications that use them. Navigation taxonomies live in content management systems. Synonym lists and topic hierarchies live in search engines. Management reporting taxonomies live in financial applications. The upside of this arrangement is that each application has quick, efficient access to the metadata it needs. The downside is that:

• Indirect users (those that need to aggregate or associate data) have difficulty accessing and massaging the metadata stored in the primary system.

• Alternate terms that may proliferate among multiple systems may cause inconsistencies in reporting and make it difficult for users to find all information relevant to a topic.

• IT staff must invest in special programming to allow multiple applications to access the same metadata.

Recognizing the need to use taxonomies in multiple systems and to map terms across domains, some organizations have implemented a centralized taxonomy function that serves multiple applications. At the low end of the technology scale, centralized taxonomy staff use Excel or an inexpensive thesaurus management system like MultiTes. At the high end, they use systems like Factiva Synaptica and Schemalogic SchemaServer. At the Montague Institute, we have created our own taxonomy management system which supports site search, our topic index, and our internal search and discovery needs. A version of the system is used as a lab for people who take our courses.

The centralized taxonomy management approach has its own problems. No longer directly tied to a single business application, the taxonomy manager finds it harder to justify the software and personnel costs. Divorced from the business objectives, staff must spend time learning the needs of multiple domains and creating interfaces to other applications. Moreover, the centralized taxonomy management function is often staffed with people whose training and experience tilts them toward a content rather than a user focus. When the function is managed within IT, there's often an additional bias toward using full text search as the only finding tool. This is the scenario that gives rise to the original question, "Why don't the terms in the corporate taxonomy match those used on the Public Web site?" The obvious answer — that the different terminologies serve different (and legitimate) purposes — is not so obvious unless you remember that taxonomies play a supporting role in increasing productivity in a specific domain.

To head off potential problems inherent in the centralized taxonomy management approach, you can:

• Base funding for the unit on reducing the cost of maintaining local taxonomies as well as global applications like reporting, innovation, and collaboration;

• Develop and promote demonstrations that show how metadata can reduce costs using real data from key business processes (see "Get ready for end-user development";

• Select an experienced "boundary spanner" to manage the unit. This is a person who can synthesize concepts from journalism, library science, and computer science and is capable of forging strong relationships with business units and subject matter experts.

• Develop a training program for business unit staff that includes exposure to enterprise metadata standards, how-to information and productivity tips, and representation on the enterprise taxonomy policy committee.

Deploying taxonomies

Increasingly, taxonomists and search administrators are asking the question, "How much effort should we put into customizing our search engine with things like custom dictionaries and

synonym lists, tweaks to the relevancy ranking algorithm, "Best Bets," and category search?" At one end of the spectrum are those who say that if you buy the right search product, customization should be minimal. At the other are those who want to have extensive control over search behavior and believe in giving users as many finding options as possible.

One of the benefits of using a relational database to manage taxonomies (as opposed to Excel spreadsheets) is that you can have the best of both worlds. Let the search engine do what it does best — find documents in a relevant collection that match a specific query term. If you want search to display "Keymatches," "Quicklinks," or "Best Bets" (human-selected links that appear above the standard results list), you can periodically export terms and URLs from the taxonomy management in the format required by the search engine or another application.

As an alternative, you can leave the search engine alone and display terms and URLs directly from the taxonomy management system in a format similar to the Montague Institute index, complete with cross references (called "aliases" in IT lingo or "see" references in library lingo). The downside of this strategy is that users must look in two places (search and index). For some applications, this may be an obstacle, but for many others, it's a natural extension of the print world, where tables of contents and back-of-the-book indexes are frequently used tools.

Characteristics of a taxonomy management system

To implement a system where search and browse are complimentary, the taxonomy management system should have the following characteristics:

• Basic thesaurus capability with broader terms, narrower terms, related terms, and USE/Use for relationships;

• The ability to store "card catalog" metadata for documents and other content objects, and the ability to link documents and terms;

• The ability to store metadata about people and link them to content objects in an "authored by" relationship;

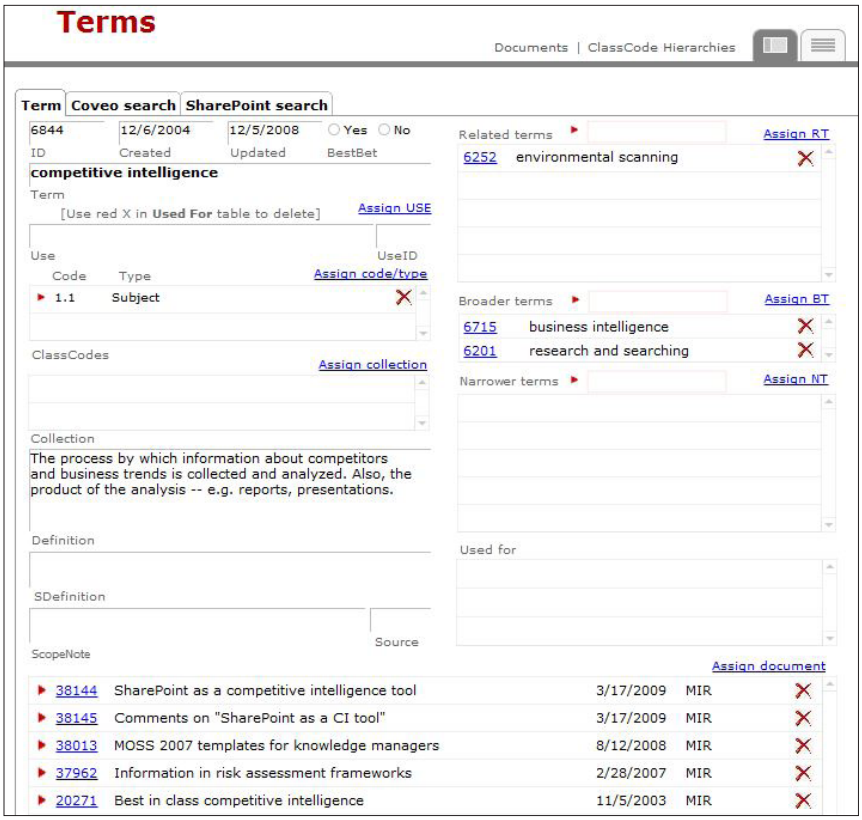
• A user-friendly format that allows business users who are not librarians, indexers, or ontologists to add data and relationships for localized indexes and search scopes;

• The ability to add new fields and non-standard relationships, such as "acquired by" or "member of."

As far we know, a commercially available off-the-shelf system that has all these characteristics does not exist. The high-end taxonomy management and ontology development systems we've seen have thesaurus capability and the ability to create non-standard thesaurus relationships, but they are designed for professional taxonomists or IT specialists, not departmental knowledge stewards. For this reason, we began developing our own taxonomy management system in 1998 (see below).

In the early stages of development, we used data stored in the system to customize our Ultraseek search engine with categories, "Best Bets," "see also" references, and custom search summaries. Over time, we decided to let the search engine stand on its own two feet without any tweaks except the custom search summaries. Categories, "Best Bets," and "see also" references could be provided most effectively in the topic index, which is driven directly by data from the taxonomy database. Our Web and database logs tell us that half of our public Web site visitors use the index while the other half use the search engine. That's consistent with how we use search and index browse internally. They are complimentary tools.

The Montague Institute Review *is published by the Montague Institute and edited by Jean Graef.*



The Montague Institute taxonomy management system has a controlled vocabulary with definition and scope note, a thesaurus of related terms, and links to documents and other content objects. Other components of the system allow links between documents and people.