MONTAGUE INSTITUTE REVIEW

ISSN 1554-303X

February 2000

Your Taxonomy is Your Future

The way your company organizes information (i.e. its taxonomy) is critical to its future. A taxonomy not only frames the way people make decisions, but also helps them find the information to weigh all the alternatives. A decision that seems right from one perspective (e.g. the supplier) may look different when viewed through another lens (e.g. that of the customer).

For example, would you classify a carrot or a grain of rice as a food or a drug? Would you classify a cotton bole as a fiber or a pesticide? According to two recent articles, they can be classified as both (see "Is Monsanto's Biotech Worth Less Than a Hill of Beans?" and "The voice of reason in the global food fight," (Fortune, February 21, 2000). Moreover, the interests of farmers, consumers, developing countries, environmentalists, and health care professionals are all intertwined in complex ways. A good taxonomy helps decision makers see all the perspectives, "drill down" to get details from each, and explore lateral relationships among them.

How classification adds value within an organization

Outmoded classification schemes can be dangerous to corporate health. Remember the business school case about the railroad industry's failure to think of itself in the "transportation" business. Now forward-thinking shippers are calling themselves "logistics" firms.

Because of the Internet, classification has become a business issue for everyone, not just consultants, computer scientists, and librarians. Examples are everywhere:

- human resources managers want to classify employees according to their real skills and capabilities, not just those listed on their resume;

- customer service managers want to organize their databases by problem and solution as well as products and market segments;

- sales and marketing managers want to know how much profit a customer is likely to generate over several years, not just how much he bought last quarter;

- R&D managers want to be able to find solutions to engineering problems, including those "outside of the box" (in other disciplines or applications);

- Financial managers want better measuring instruments that reflect the growing importance of intellectual capital and that are better able to predict future value.

How classification adds value among several organizations

Not only has classification become a key issue within a company, it has also become a top priority among groups of companies. Here's why:

Electronic marketplaces such as Chemdex and Automotive Network Exchange (ANX) require members to agree on how to describe the elements of a sale. For example, a purchase order may mean a standing order for a certain volume of products per year or a single purchase of one product on a specific date. One way to standardize electronic commerce transactions is to create a "dictionary" using XML (Extensible Mark-up Language, a kind of "Rosetta Stone" that allows computers to translate from one descriptive system to another). Many industries are developing specialized XML dictionaries (NOTE: in this article, "GM" refers to General Motors, not "genetically modified").

Electronic repositories, such as an intranet or consumer Web portal (e.g. Yahoo), need to provide pathways into huge collections of documents and Web sites. A "search engine" (e.g. Google) is not enough because it often returns too many

"hits." A classification scheme helps users zero in on those items of interest and browse related areas.

An example is the Washington State Find-It Web site. The main page contains a search facility as well as a high-level taxonomy. Each of the 25 main categories (e.g. "Health, Medical") is divided into subcategories (e.g. "Drugs, pharmacology" and "Nutrition"), which can be searched individually. Each document is tagged with descriptive information that provides yet another way of finding it (e.g. department, contact name, keywords, jurisdiction, etc.).

Another example is the Society Knowledge Base, which includes three levels of classification -- a top level with 6 categories, a second level with 25 - 30 categories, and a third level (a "controlled vocabulary" of approximately 2000 words).

Shoppers, increasingly sophisticated about computer searching capabilities, are demanding more convenient ways to research purchasing decisions and compare products. The Internet is making it easier to serve the needs of all parties in the purchasing process. For example, the WebMD site serves not only buyers and sellers, but also related professionals, such as teachers, nurses, and administrators. Strategic partners include News Corporation, DuPont, UnitedHealth Group, Merck Medco, HEALTHSOUTH, Microsoft, Intel, CNN and Reader's Digest.

The consumer portion of the WebMD site uses a classification scheme appropriate for patients or customers (e.g. Sports and Fitness, Find a Physician, Health-E-Meters), while the physician portion uses categories appropriate for doctors (e.g. For Your Patients, Practice Enhancements, Clinical Services, Administrative Services). Both parts of the WebMD site share content. The stories are not only classified differently, but they are written for different audiences. For example, the site recently published two articles describing the work of Carl Keen on the health benefits of chocolate. The articles were authored by the same person but had different titles and a different vocabulary (see "Death by chocolate? Maybe Not" for physicians vs. "Chocolate as a health food" for consumers).

Joint venture partners need standardized ways of describing things to facilitate collaboration. There are many examples in the academic and nonprofit worlds. Among them are the Arctic Research Consortium, the EPA (see its Environmental Data Registry), and the Getty Research Institute (see its Index and "Introduction to Metadata" white paper).

Find me a taxonomy

Your company already has multiple classification schemes:

- organization chart listing divisions, departments, and employees;
- chart of accounts with categories for income, expenses, assets, and liabilities;
- bills of materials listing product modules and their component parts;
- product catalogs provide sales territories and competitors.

Each of these schemes serves a necessary purpose but none are useful as a corporate taxonomy. Not only do they provide too much detail, but they also represent only one point of view. Even more important, they fail to describe the company in terms that can be directly related to revenue generation. They don't, for example:

- help customers make purchasing decisions;
- help customers use products;
- enable employees to solve customer problems, especially those that cross organizational lines;
- make it easier to identify and exploit business opportunities.

External taxonomies are also available from government agencies, electronic publishers, professional associations, and standards bodies. These have the same kinds of limitations as internal classification schemes, but sometimes they can be used as a starting point for a corporate taxonomy.

The economics of classification

The leaders in developing and using classification schemes - government, libraries, museums, and researchers, electronic publishers, and electronic commerce participants -- have three things in common:

1. large repositories of information in electronic form (e.g. National Technical Information Service);
2. a culture that is predisposed to collaborate and share information (e.g. librarians, researchers);
3. a demonstrable connection between a classification scheme and economic value (e.g. electronic marketplaces).

While corporations with well-established intranets or document repositories meet the first criteria, they often do not meet the second or third. But it may not matter as long as they create gateways or bridges to internal and external taxonomies. Increasingly, these bridges must be readable by both computers and human beings.

Taxonomy bridges

Some kinds of bridges are:

- *Human experts*, such as knowledge stewards, knowledge brokers, or corporate scholars (see "Capturing know how through interviewing", "Leveraging experts", and "Corporate scholarship").
- *Metadata Registries*, computer files that allow different taxonomies to inter-operate.
- *Thesauri*, lists of subject headings or descriptors with references to broader terms, narrower

terms, and synonyms (see the Cook's Thesaurus, which suggests substitutions for thousands of cooking ingredients, including low-calorie and low-fat alternatives for dieters, inexpensive substitutes for gourmets on a budget, and innovative replacements for hard-to-find ethnic ingredients).

• *Indices*, lists of terms arranged usually in alphabetical order by author, subject, or keyword (for examples, see the Society of Indexers "Indexing the Web").

Gateways and bridges facilitate rather than mandate information sharing. The expense of creating and maintaining a classification scheme is borne by multiple groups, each of which can justify its share of the cost. Gateways can be used by the layman and the specialist, the manager and the employee, and other groups -- customers, investors, suppliers, government agencies, and joint venture partners.

Definitions

Classification -- A systematic arrangement of words or phrases in groups or categories according to established criteria.

Index -- A list (as of bibliographical information or citations to a body of literature) arranged usually in alphabetical order of some specified datum (as author, subject, or keyword). A list of items (as topics or names) treated in a printed work that gives for each item the page number where it may be found.

Lexicon - 1 A book containing an alphabetical arrangement of the words in a language and their definitions; a ictionary. 2 The vocabulary of a language, an individual speaker or group of speakers, or a subject. 3 The total stock of morphemes in a language.

Metadata -- Data that describes information -- a text file, a book, a magazine article, a photograph. A library card catalog consists of metadata about a book. The <Title> tag in an HTML document is a metadata element that describes a Web page.

Ontology - 1. A model that a particular speaker has about the world. This model is populated by concepts, organized in a particular hierarchy. The concepts in the ontology cover things (such as airplanes, ideas, or giraffes), events (e.g., buying, eating), as well as relations. To support retrieval and economy of representation, the ontology is organized as a taxonomy. 2. The hierarchical structuring of knowledge about things by sub-categorizing them according to their essential (or at least relevant and/or cognitive) qualities.

Taxonomy - Classification, especially the orderly classification of plants and animals according to their presumed natural relationships.

Thesaurus - 1. A book or list of words or of information about a particular field or set of concepts; especially a book or list of words and their synonyms. 2. A list of subject headings or descriptors usually with a cross-reference system for use in the organization of a collection of documents for reference and retrieval.

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