

Taxonomies and Classification

Shannon Lucas, INF 385E – Information Architecture

Introduction

Information architects, like building architects, are in the business of building structures. If building architects design the structures that house our bodies, it could be said that information architects design the structures that house our minds. Like a physical structure, these information structures must provide windows, doors, and dividing walls to ensure access and organization. Classification systems and taxonomies provide these capabilities to information architects.

Classification Concepts

The desire to classify and organize things seems to be a fundamental feature of the human condition. An almost universal tendency with how humans organize information is to use the genus as the primary grouping (Taylor, 1999); genus-level items include things such as tree, car and lizard as opposed to elm, Porsche and bearded dragon. These are the groupings we most often use in normal conversation. It is not enough, however, just to say that we organize at the genus level. Numerous organizational schemes exist throughout information systems, particularly the World Wide Web. Analysis of various systems exposes exact and ambiguous as the primary groupings of organizational systems on the Web (Rosenfeld, 2002).

Exact Schemes

Exact schemes create divisional lines along sections that are well-defined or mutually exclusive. Alphabetical, chronological and geographical organization fall into this area. Systems based on exact schemes require known-item searching on the part of the user and may require ancillary knowledge of the item being searched for. Alphabetical schemes require the user to know the spelling of the search item. Geographical systems may require a user to read a map and possess and understanding of relative locations. Take as example that a person who knows to look for an item in Germany may not know where Germany is located on a map of the world.

Alphabetical

The prototypical example of alphabetical organization is the residential listings in a telephone directory. Knowing a person's name, a corresponding phone number can be retrieved by finding the proper place of the name in the alphabet. Web sites that follow this scheme, such as employee directories, are typically performing a function similar to that of the phone book's residential listing.

Chronological

The use of chronology for Web sites almost exactly parallels the use of the scheme in the physical world. Calendars, event schedules and diaries (often in the form of weblogs, or blogs) all mimic their physical counterparts. News headline sites, however, do not mirror their physical equivalents. Chronologically organized news headlines are difficult to find outside the Web. One advantage to chronologies is that the concept of ordering events by time of occurrence is understood universally. Like alphabetical systems, it doesn't require much effort to create or maintain.

Geographical

Even people who are adept at map reading often overlook the information density of the typical geographic map. As an organizational scheme, maps take advantage of a user's geographic or spatial knowledge. Travel-oriented sites often use geographic schemes to provide access to information about vacation destinations. A geographic scheme, however, does not necessarily require a map; an effective geographic system may provide a list of countries or regions instead of a graphical map.

An important consideration with geographical organization is to make sure the users are familiar with the geography. Some bus systems organize route and schedule information with a map of the city they service. This causes problems for newcomers and visitors who have no prior knowledge of the city's geography.

Ambiguous Schemes

Ambiguous schemes base their organization on subjective areas that users may be more familiar with. Most familiar is organization by subject, the system used in virtually all modern libraries. Following this are division by task and division by audience. Additionally, metaphorical systems utilize an idea with which the user may be familiar.

Subject

Subject, or topical, classification has its roots in Classical Greece with Aristotle's idea that everything in the world could be fit into a perfect hierarchy (Taylor, 1999). Objects were typically categorized by their external characteristics or their subjectively determined "aboutness." The utility of subject categorization is that it allows users to browse a selection of items when they do not know exactly what they are looking for. Web sites such as yahoo.com and about.com provide broad categories that the user can then browse to reveal more finely grained categories until a specific item is found. These systems are made even more powerful by cross-listing and providing hyperlinks to other related categories.

Twentieth century research has found holes in classical categorization theory, however. Wittgenstein demonstrated that items in a category such as games did not fit the classical system; there aren't any properties that would lead one to conclude an activity such as chess is in the same category as hide-and-seek. Austin furthered this work with the study of words. Under classical theory, items such as lists, beers, and people should be in the same category since they each possess the property "head" (Taylor, 1999). For the most part, however, subject schemes still prove easy to understand if the architect makes well thought out choices on subjects.

In the 1930s, Ranganathan developed a concept of faceted classification. This system works under the premise that objects and documents can be classified in multiple dimensions, or facets. Ranganathan identified five facets in his system to which he attached the acronym PMEST: personality (most specific subject), material, energy (activity or process), space (location), and time (Taylor, 1999).

Task

Task-oriented schemes focus on the activities a user is expected to perform when they visit a site. This is the scheme around which usability for desktop applications is based; Norman even writes that the user tasks should dominate the user interface (1990). E-commerce sites and corporate intranets are where task-oriented organizations are found most often (Rosenfeld, 2002). Often task-oriented schemes are located beneath a site's primary scheme if the tasks are not the sole purpose of the site. Sites of this type may also implement a process funnel pattern (van Duyne, 2003) that guides the user through specific tasks.

Audience

Audience based systems arrange a site's content based on the types of users that are expected to visit the site. Typically, a site would provide a means of separating user content at the top level and then organize the content for each user type using another scheme. Multiple levels of audience schemes exist starting with very broad user categories down to individual customization of sites such as the My Yahoo! portal system created by Yahoo!. Another example of user level audience organization is the targeted user level recommendations page on amazon.com; users are presented with products their past purchases indicate they might want.

Multinational companies often create audience division along geographic or political boundaries. Initially these sites may appear to be geographically organized since they often present a map as the initial interface, but inspection reveals that the intention is separating the audience. Division along these lines deals with a number of issues that differ from one country to the next, not the least of which are language and local laws. This scheme also allows cultural idioms to be addressed as some wording and imagery can carry vastly different implications from one culture to another.

Though some desktop applications attempt to present different information based on a user's skill level, usually "beginner" and "expert," it can be a disaster for a Web site. The architect of the site must make a judgment call on what a user is capable of doing at each level, and this seldom involves proper usability testing. Raskin generalizes that the beginner-expert dichotomy is mythological (2000).

Metaphor

The goal of metaphor schemes is to relate a system to concepts that the user is already familiar with. Currently, the most pervasive metaphor is the desktop, folder, and file system used by the majority of graphical user interfaces. It is unusual to find this metaphor on Web sites, however; spatial metaphors tend to be the most common on the Web. Lipton's liptonfavorites.com presents the user with a picture of a kitchen, Oscar Mayer's website uses various buildings, and Apple's defunct online service eWorld used a town metaphor. The

inherent danger of using metaphors is that they may not be obvious to the user. The Lipton site's kitchen metaphor offers no indication of how it should be used. Eventually the user may discover that they need to place the mouse over various items in the kitchen to discover what is a link and what it links to. Oscar Mayer's buildings also provide little information. Nothing about a "town hall" indicates what its function within the site is. Apple's eWorld interface provided descriptive labels of the buildings making it clear what the function of each was. Napier University computer science professor David Benyon points out that spatial metaphors generally fail to appreciate how people conceptualize the spaces they use and focus instead on how those spaces are perceived (1998).

Taxonomies

There is no agreed upon definition of exactly what a taxonomy is. In the biological sciences, a taxonomy is generally a hierarchical arrangement beginning with the kingdom level and expanding to individual species. In a system such as the World Wide Web where any item may be linked to any number of other items, the notion of a hierarchy may not seem to fit, but hierarchical parent-child relationships and mutual exclusions are ideas the most people are familiar and comfortable with. The benefit of a taxonomy for a user is that it provides a foundation for developing a mental model of the material being offered and how to access it. Conway identifies two basic types of taxonomies: descriptive and navigational with a special case of the descriptive known as a data management vocabulary.

Descriptive Taxonomies

Descriptive taxonomies use a controlled vocabulary to facilitate retrieval by searching. A controlled vocabulary is a list of terms agreed upon by a system's architects as the definitive names that will be used for organizing and identifying content. Sites such as Mayo Clinic's offer an alphabetical list of terms that are used for locating information. The utility and user-friendliness of descriptive taxonomies can be increased by adding a list of common synonyms or misspellings that a user might search for if the controlled term is not known. Descriptive taxonomies also offer growth potential. By adding parent-child relationships, a classification system emerges. By adding relationship information, a thesaurus is created (Rosenfeld, 2002). The construction of a controlled vocabulary requires careful planning from the architect. Terms should provide a balance between the general and specific. The focus when building the vocabulary should be on the actual vocabulary of a site's users and not so much that of the site's producers.

The data management vocabulary allows the architect to approach the system from the bottom up, starting at the document. Controlled vocabulary meta-data can be used to describe documents in a database. What differentiates this scheme is the power of the underlying relational databases. Relationships between pieces of meta-data can be created arbitrarily as they are needed. This facility proves useful for structures such as catalogs and directories where data type is largely homogenous (Rosenfeld, 2002).

Navigational Taxonomies

The focus of navigational taxonomies is user behavior and not content (Conway, 2002). They are generally “top down” approaches involving an ambiguous classification scheme that starts with a small set of general categories, or “parents” that then expand in a tree like structure. The root, or “super parent” of this tree being the main page of the site itself. The language used in a navigational taxonomy also differs from that of a descriptive taxonomy. A personal health site using a navigational taxonomy may use the label “Loose 10 Pounds” to identify a section on weight loss. Such a label would be inappropriate for a site using a descriptive taxonomy.

One of the decisions that faces the information architect when building navigational taxonomies is the choice of depth and breadth. If a hierarchy is narrow, but deep, a user may lose interest while navigating or get lost. If it is too wide, the cognitive load on the user to decide what to do can be too great.

Future Directions

Santa Barbara company Cooperating Systems is currently testing an online social computing platform called HelloWorld. Renowned interface designer Kai Gradert has developed an interface for the application based on geographic metaphor. The company’s rationale for the application is that the Web is a largely solitary experience. The goal then is to provide a geographic context (via an enlargeable world map) for e-mails, file transfers, and instant messaging. The company believes that this idea can be extended to office floor plans, municipal maps, or logical (non-spatial) coordinate systems. The application is currently only available for Windows, but the company promises future support for Macintosh and Linux platforms.

The advent of the artificial neural network (ANN) has made possible technologies such as speech and handwriting recognition as well as automated decision support systems. Research is also being conducted towards applying ANNs for categorization browsing using self-organizing maps (SOMs) (Taylor, 1999). WEBSOM, a project of the Helsinki University of Technology, is a project to automatically categorize web documents using a two dimensional density map. Related documents exist in close proximity as do related categories.

The initial view of the map resembles a color graded topographic map. Brighter areas on the map represent greater document density. Labels indicate the categorization choices the system made. Clicking on any area of the map will zoom in for a closer view displaying sub-categories. A further click will retrieve specific documents. To date, the largest document set processed by WEBSOM was a collection of 7 million patent abstracts.

Conclusion

In order for information to be used, it must be arranged in a way that allows and encourages its discovery. Through the use of classification systems and taxonomies, and information architect can accomplish this. A number of classification systems have been discovered to date ranging from the familiar subject ordering to more abstract ideas such as metaphors. More systems will be discovered as more research into linguistic anthropology and cognitive science is

done. Taxonomies provide a means of allowing the user to develop a mental model of a collection's content and structure. All taxonomies have some form of classification system as their foundation.

References

- Benyon, D. R., (1998). *Beyond Navigation as Metaphor*. Retrieved October 25, 2003 from <http://www.dcs.napier.ac.uk/~dbenyon/EuroDL.html>
- Conway, S., & Sligar, C., (2002). *Unlocking Knowledge Assets*. Redmond, Washington: Microsoft Press.
- Cooperating Systems, Inc. (2003). *HelloWorld Overview*. Retrieved October 26, 2003, from: <http://www.cooperatingsystems.com/helloworld/overview/index.html>
- Neural Networks Research Centre (NNRC) at Helsinki University of Technology (HUT). (1999). *WEBSOM – A novel SOM-based approach to free-text mining*. Retrieved October 25, 2003 from <http://websom.hut.fi/>
- Norman, D. A., (1990). Why Interfaces Don't Work. In B. Laurel (Ed.), *The Art of Human-Computer Interface Design* (pp. 209-219). Reading, Massachusetts: Addison-Wesley.
- Raskin, J., (2000). *The Humane Interface*. Boston: Addison-Wesley.
- Rosenfeld, L., & Morville, P., (2002). *Information Architecture for the World Wide Web*. Cambridge: O'Reilly.
- Taylor, Arlene G. (1999). *The Organization of Information*. Westport, Connecticut: Libraries Unlimited.
- van Duyne, D., Landay, J.A., & Hong, J. I. (2003). *The Design of Sites*. Boston: Addison-Wesley.