PART I

Principles of Portal Design

CHAPTER ONE

Creating Frameworks for Organizing Information

Portal designers can learn much from architects and builders. Well-designed buildings are easy to use and structurally sound. We can find what we want, components like doors and windows appear in logical places, and, most importantly, the building stands up over time. We cannot go into a building and find the structural integrity the same way we can find the heating and ventilation units or the corner office. Structural integrity is a property of the way the building was designed and constructed; it is not a single feature added at some point in the construction process. The structural integrity of a portal is similar to that of a building. It is a fundamental property of the portal design, reflected in turn in visible characteristics, such as ease of use, functionality, and reliability.

In this chapter, we look into structural integrity from a user's perspective. The core question we address is "How will portal users find what they need?" Actually, we break this question into a number of more specific questions to which we can provide general but concrete answers.

First, we discuss how to organize information on a page. This may sound insignificant compared to other challenges that await us in portal development, but poorly designed pages hamper the portal's adoption. Next, we look at design patterns for logically grouping related content and applications to provide a sense of context for our users. We can all appreciate the sense of being in a particular section of a department store and knowing in general

how to find other sections. We should provide something analogous for portal users. Without contexts users can easily become lost in an apparent jumble of hyperlinked pages. Finally, we look at specialized techniques (such as taxonomies, faceted content models, and visualization) that can aid navigation, especially in large and diverse portals. A case study shows how visualization and logical restructuring techniques improved customer care services for one organization.

Much has been written about usability and Web design techniques, and this book does not try to add to these well-discussed areas. The main concern here tends more toward architectural issues, which sometimes abut or even overlap with usability issues. For questions about usability and design layout, I defer to any of the well-written books on the subject such as *Don't Make Me Think* by Steve Krug and Roger Black [2000] and *Designing Web Usability* by Jakob Nielsen [1999]. I will address the types of structural elements required in well-designed portals but won't try to describe the finer details of their layout, formatting, and other visual elements.

The Need for Structure in Portal Interface Design

When considering portal interface structure, it is useful to distinguish between the visible structures and the underlying structures. The visible structures provide the organization reflected in the designs of pages, groups of pages (known as subsites), and the entire portal itself. These structures are readily apparent to users.

The underlying structures are core services, such as authentication, access controls, and metadata management, as well as the policies and procedures that govern the evolution of the portal. These structures are not necessarily visible when they work well, but their absence is all too apparent. When users cannot work with essential applications because of access control problems or when navigation tools direct searchers to inappropriate content because of miscategorized metadata, users become all too aware of these underlying services.

Page-Level Structures

Page-level structures include the distribution of content, applications, and navigation tools. Many pages use the basic three-panel structure shown in Figure 1.1. The top area contains global information about the site, the left side area contains navigation controls and links to commonly used objects, and the large central panel is home to the substantive content of the portal.

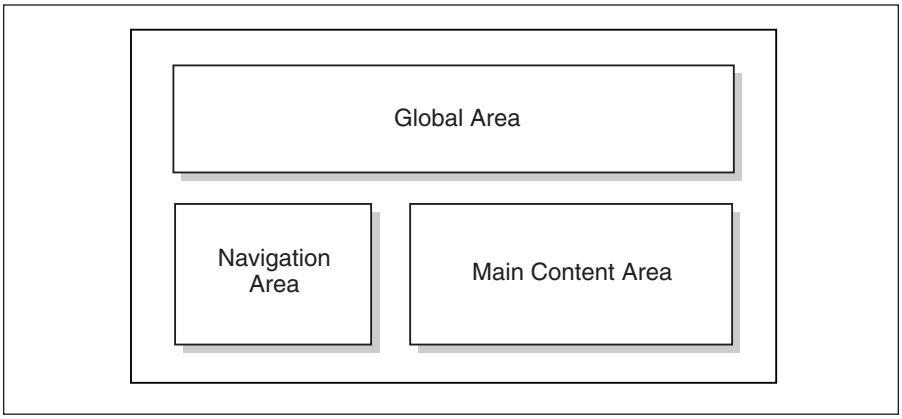

Figure 1.1 Many portal pages use a basic three-part layout.

The global area is consistent across the portal and often provides links to a home page, contact information, accessories, or other frequently used applications.

The navigation area provides a localized context for users. If you went to the human resources area of a portal you would expect to find navigation links to training, policies and procedures, benefits information, and related material; in a health and safety area of the portal you'd expect to find information on material safety, accident prevention, and reporting procedures. The role of the site navigation area is to provide an immediately visible and easily accessible path to related components in the portal while keeping the user from being overwhelmed by the full breadth of the portal.

There are several common approaches to organizing the navigation area. First, the area can be organized by subsite or neighborhood. The CNN Web site (http://www.cnn.com), for example, uses this approach by consistently listing subsites (such as Weather, Politics, Business, and Sports) in the navigation area. A variation on this model is to display subtopics when a topic is selected. A third approach focuses on tasks rather than content and is more appropriate for portals or subsites oriented toward content management. Yet another approach is a hybrid that combines content-oriented with task-oriented links. Care should be taken to clearly distinguish the two types of links, remembering that the purpose of the navigation section is to provide a sense of context. Intermixing content and task links could make it more difficult for users to perceive their location within the portal.

The main content area delivers the core information and application access that the users seek. By framing this information and the applications in navigational frameworks, you provide users immediate access to locally related topics as well as global landmarks, such as the portal home page.

Grouping Pages: More Than One "Right" Way to Do It

An organizational model describes how entities are related. In the case of a portal, it describes how content, applications, and other resources are made accessible to users. The simplest model offers hyperlinking without restrictions. In this case, any page or resource can provide links to any others. This is the general model of the Web and the de facto organizational scheme for ungoverned intranets as well as the Internet. The advantage of this model is that decision making is completely decentralized so anyone can add content at any time to any part of the intranet. The disadvantage, so clear from the World Wide Web, is that this organizational scheme provides no point of reference for users. For example, if you find yourself at a page in such an intranet, there is no absolute reference point such as a home page or directory. All pages are equally important with regard to navigation, as depicted in Figure 1.2.

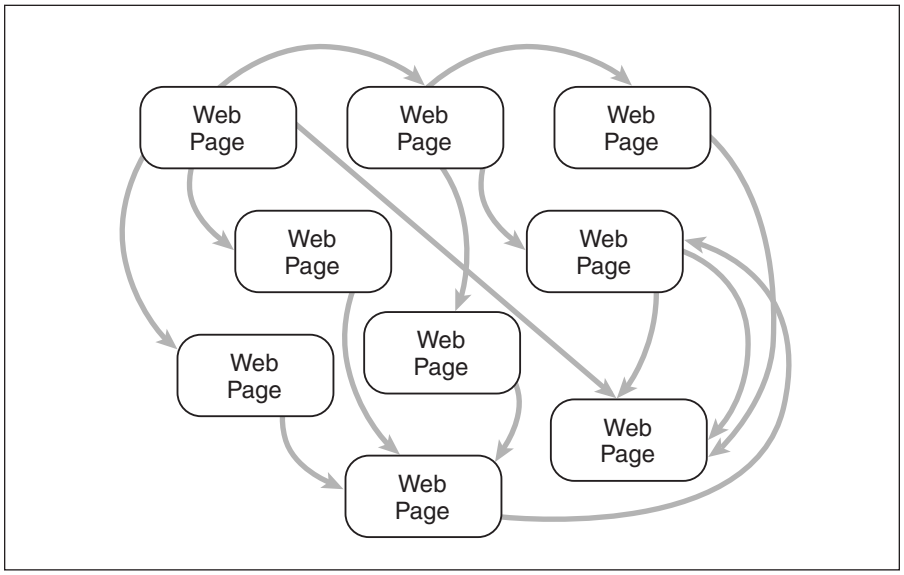

Figure 1.2 Simple hyperlinking schemes provide no point of reference.

Fortunately, most sites are no longer so freewheeling that we are left to navigate without some fixed references. Hierarchical organizational schemes provide an organizational structure with a top-level starting point and one or more levels of content. The simplest form of hierarchical organization is a tree with a root and links to lower-level pages. For practical purposes, most portals and Web sites also link across the hierarchy, as shown in Figure 1.3.

This type of nonhierarchical linking is required because simple hierarchies do not adequately model the way we think about information and how to find it. Consider a general topic such as "wireless phones." Where should this fit into a hierarchy? Some possibilities include:

1. Business > Services > Telecommunications > Mobile Services
2. Business > Office Products > Phone Systems > Wireless Phones
3. Consumer > Telecommunications > Wireless Phones

Each hierarchical categorization is reasonable; the most intuitive one depends on the context of the search. If the user is looking for a mobile phone service, the first categorization above appears logical; if he or she is interested in purchasing a wireless phone, the second makes the most sense; and if the user is looking for a wireless phone for home, the third categorization is the most likely

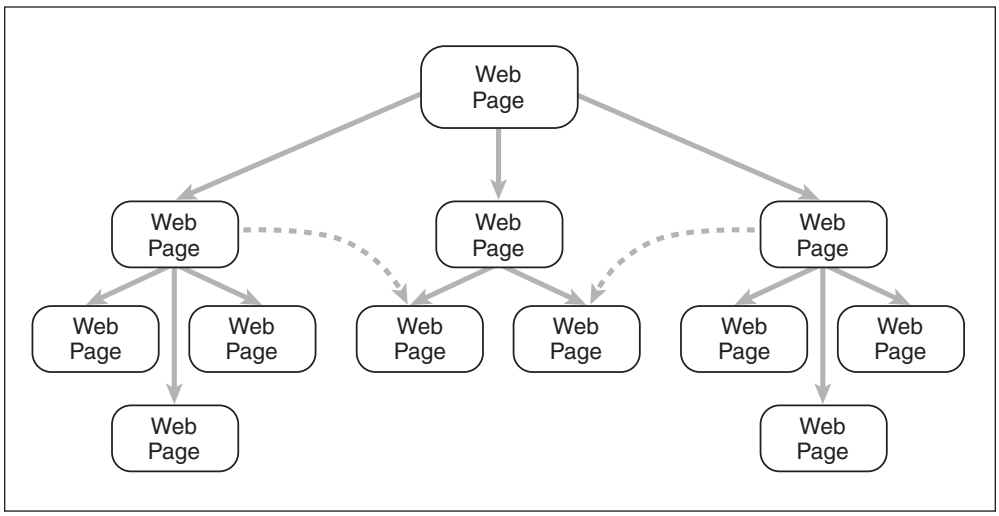

Figure 1.3 Even hierarchical patterns need to cross-reference nodes to support the way users navigate a site.

to be followed. Clearly, a single hierarchical structure is not sufficient for even moderately complex portals. More importantly, there is no single correct answer about where to place a page in a hierarchy, so don't bother trying to find one.

Organizing Multiple Ways with Facets

Instead, use a more flexible approach to navigating between pages. Multifaceted organizational schemes avoid the problem of hierarchies by accounting for the fact that an entity such as a product, application, or Web page can be classified along multiple dimensions. For example, let's assume a user is searching for a wireless phone, for less than $150, in silver or black with multiline support. Ideally, the user could navigate to information about phones, find wireless phones as a subcategory, shift to navigating by price, select the $125–$175 category, and finally narrow the search by color and functionality. In this example we have four facets: product category, cost, color, and feature set. When designing an information model, it is best to consider several facets or dimensions along which content is organized.

Metadata about categorization and content classification constitute facets or dimensions for organizing content. For example, a document published by the Health, Environment, and Safety Department on the proper disposal of chemical waste may be an official policy, published on a particular date, constituting compliance with a government regulation and broadly categorized as a safety document. The document type (policy), publication date, category (safety), and regulation status are all facets or attributes useful for organization and retrieval. One way to think about facets is as dimensions in a multidimensional space. Figure 1.4, for example, depicts the location of documents in a multidimensional space.

Facet-based information retrieval can help users target specific content more quickly than simple keyword searching or navigation through a directory. Facet-based searches should allow a combination of keyword searches and attribute searches, such as searching for all "policy" type documents that contain the phrase "toxic disposal" and were published between May 1, 2003, and July 1, 2003. This technique is especially powerful when working with product catalogs that list items described by several dimensions (e.g., cost, size, color, feature set).

Dimensions can use a list of values, a range of values, or hierarchical values and in this way is similar to many online analytic processing (OLAP) tools. Relational taxonomies can model hierarchical values, but continuous value attributes (e.g., cost, time) are best modeled with scalar variables such as strings, numbers, and Boolean values.

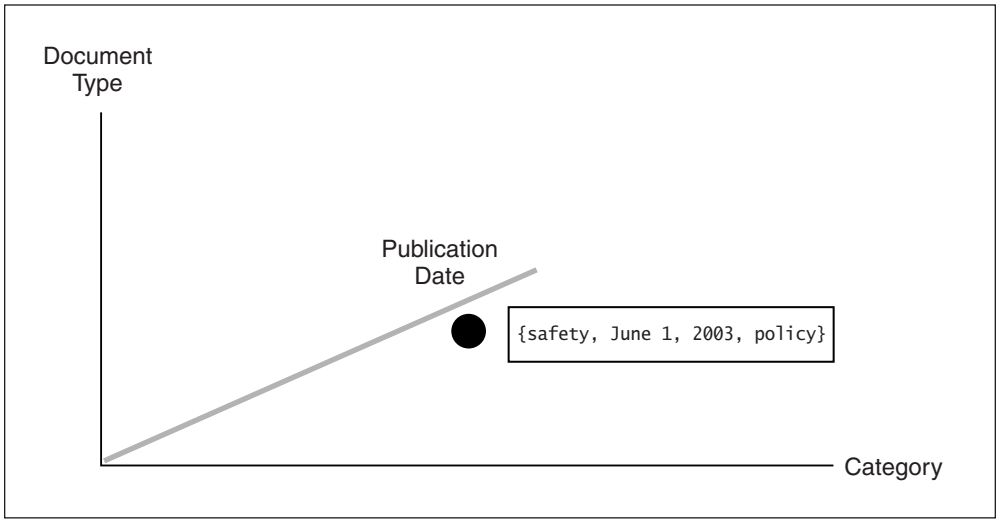

Figure 1.4 When organizing documents by facets, documents can be considered points in multidimensional space where each dimension is a facet.

The dimensions should reflect the ways users describe or understand content since organizational structure, like a search tool, is a key method for facilitating information retrieval. A particularly large site, FirstGov (http://www.firstgov.gov), the official site of the U.S. government, uses multiple facets, including:

- Audience
- Topic
- Service
- Organization

The main page provides links to citizen, business, and government audience channels. The target pages of those links then list options by service. The service page in turn organizes content by topic. FirstGov's organizational model uses multiple facets and mixes those facets within paths from the home page to the content pages. This approach works well when you have information about usage patterns and frequently accessed pages. Analyzing log files from Web servers can provide key information about the most frequently accessed pages. By analyzing and grouping those pages, you can develop a rough categorization scheme based on facets.

Understanding how users think about the content and other resources in a portal is essential to developing a logical organizational model. Neither free-form

links across a site nor rigid adherence to a hierarchical structure will serve the user community. Multifaceted organizational models provide the organizing structure of hierarchical systems and some of the flexibility of free-form linking within a controlled framework. Later in this chapter, we examine how complex facets can be organized using taxonomies.

Flexible Organization with Navigation Sets and Other Design Patterns

Another approach to organizing links is to use a hybrid of the hierarchical and free-form hyperlinking approaches. With this technique, we make a decision that some dimension is more important than others, such as the organizational structure of a company or the categories of products. The hierarchy is based on this dimension. Within each branch of the hierarchy pages can be linked as needed to other pages in the same branch. The advantage of this approach is that it allows users to quickly find high-level topic areas (such as the human resources section of a company portal or the camping equipment offerings of an online store) while still allowing the site designer to customize links between related pages. These relatively closed-off areas of related pages are called **navigation sets**. Figure 1.5 shows this common navigation pattern.

A number of other patterns have evolved along with the development of the Web. These patterns provide a sense of well-defined location within a portal and

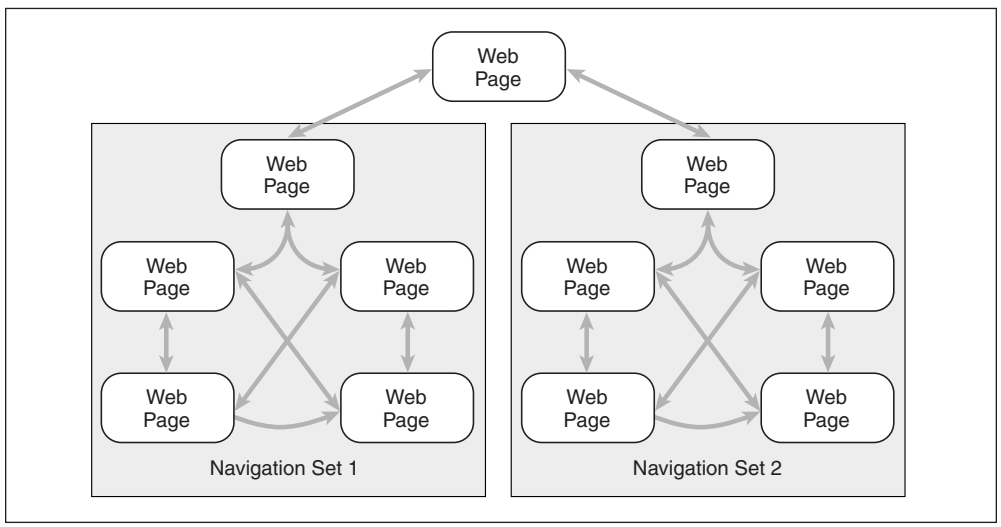

Figure 1.5 Navigation sets are groups of highly linked pages with relatively few links outside the neighborhood. Exceptions include links to home pages, contact information, and other global content.

provide rapid access to other well-defined places. Some of the most useful patterns (in addition to navigation sets) are listed below [Rossi et al.]. These patterns can be used independently but are often found together.

- *Landmarks:* Landmarks are links to entry points, large subsections of portals, or frequently used applications. CNN.com, for example, displays links to weather, politics, business, sports, and so on in the left side menu.
- *Nodes in context:* A variation on navigation sets are nodes in context. With this navigation pattern, the same content is repurposed for multiple uses. Depending on the use, the links associated with content vary. For example, an online retailer could list a new tent in both the camping equipment area and the new products area. In the former case, links from the tent display would lead to other camping equipment while the latter category would include links to other new products. Again, the point is to provide navigation links to logically related content to create a sense of context and intuitive navigation paths out of one navigation set and into another. Both navigation sets and nodes in context provide a fine-grained sense of context. Users also need a sense of context relative to the portal or site as a whole. That is the job of active references.
- *Active references:* Active references are indicators of one's position relative to the site as a whole or relative to a landmark. Directories such as Yahoo! and Google Directory use a list of nodes traversed in the directory. For example, see http://directory.google.com/Top/Computers/Internet/Web_Design_and_Development/, which uses an active link: Computers > Internet > Web Design and Development.

  Visual active references are excellent methods for depicting location within a larger context. This is especially applicable when content corresponds to a physical location, such as a room in a building or a street address. As the New York Metropolitan Museum of Art timeline of art history shows (http://www.metmuseum.org/toah/splash.htm), two or more contexts, in this case time and location, can be depicted simultaneously.
- *News:* Another common navigation pattern is the news section, which is used to prominently display important corporate information or new portal features. Many public news feeds are also available in XML formats, particularly the Rich Site Summary (RSS) scheme.

Together, these and other navigation patterns constitute an essential part of the overall information architecture of a portal. They provide a sense of context to the user and offer easily accessible links to significant or related sections of the

portal. To ensure that the links are named logically and the user's experience is consistent across the portal, we must define labeling standards to identify the links that constitute these navigation patterns.

Labeling: Pointing Users in the Right Direction

Labeling standards dictate how content, links, and other objects are named within a portal. At first blush, this may seem like a trivial consideration compared to others you have to deal with in a portal implementation. However, users constantly see and use the labeling system in a portal. Well-designed systems aid navigation and should be almost unnoticed by end users. Typically, when end users notice the labeling scheme it is because of a problem. The scheme may be inconsistent or ambiguous, or it may have some other aspect that puts an additional burden on the user to determine an appropriate action or understand the meaning of a link. The issues you must contend with include the following:

- Multiple terms that mean the same thing
- Terms with multiple meanings
- Controlled vocabularies

Not long after you begin work on labeling you realize there are many ways to describe an object. Choosing one term often pleases some users and leaves others disagreeing. Some objects, such as products, departments, and projects, have official names and so have an obvious label. Even in these cases, product names change over time and departments are reorganized; as a result, outdated terms can be found in older content. Nonetheless, standard terms should be used consistently throughout the portal. In many industries, controlled vocabularies or standard lists of terms have been developed by corporate librarians, information scientists, and others who have had to deal with information retrieval problems long before the advent of the Web. These industry standards can provide the basis for a labeling standard and minimize the time and effort required to develop your own.

When controlled vocabularies are not available, search log analysis can provide a starting point for labeling conventions. Search logs identify the terms used to query an intranet or portal and provide information about results as well. The frequency with which terms appear in the log can guide the selection of terms for the labeling standard.

Successful labeling schemes are built on two factors. First, the choice of labels should be based on either their use among the portal audience (as measured by search log analysis) or the terms used by other external sources (e.g., industry

vocabularies). Second, the labels should be applied consistently through the portal. This also entails maintenance because labeling schemes change to reflect changes in the organization and general business environment.

Organizing Content around Taxonomies

The organizational techniques just described provide the skeletal structure of well-designed portals. They provide a logical organization that spans individual pages up to the entire portal. It is now time to turn our attention to the fine-grained structures required to organize the content that surrounds the coarser-grained organizational structure. In this section we discuss taxonomies—a commonly used organizing framework—that organize information reflected in facets.

Classifying Content with Taxonomies

Taxonomies are quickly gaining prominence as navigational tools in portals and with good reason. Taxonomies, or classification schemes, provide a high-level view of the content and other resources available in a portal. Search tools are useful when we are looking for a targeted piece of information, but taxonomies provide an easy-to-use browsing method. Users do not have to know what terms to search for or even whether specific information exists. Taxonomies allow us to move quickly from high-level groupings (e.g., Business, Weather, Politics, and Sports) to narrow subjects (e.g., Marketing, Finance, Investment, and so on).

While it is often easy to start constructing taxonomies, the process becomes more difficult as you move to more specific categories and realize there may be multiple ways to classify the same topic. This leads to the first rule of taxonomy development: There is not a single correct taxonomy. There are many. For example, suppose you want to find a speech by the president of the Federal Reserve Bank of Chicago using the Yahoo! directory (http://www.yahoo.com). You could find the bank's Web site in at least two different ways using the directory. One way is to start at Home and then follow the taxonomy based on organizational structure :

Home > Government > U.S. Government > Agencies > Independent > Federal Reserve System > Federal Reserve Banks

Alternatively, you could follow the taxonomy based on geographical organization:

Home > Regional > U.S. States > Illinois > Cities > Chicago > Community > Government > Federal Reserve Bank of Chicago

What seems like a logical organization is a function of how we think about a topic, not how the topic is organized according to a predefined scheme. We are not under the same constraints as librarians who have to manage physical assets. A book can be in only one location at a time, so librarians need to adopt a single arbitrary scheme (such as the Library of Congress Subject Headings or the Dewey Decimal System) to effectively manage these assets. Digital assets are easily categorized with multiple schemes. However, with this flexibility comes a new problem: integrating these multiple schemes.

For those developers convinced that a taxonomy is needed for their portals, the next question is where to begin.

Building Taxonomies

Portal developers have a number of options for building taxonomies.

- Start with an existing third-party taxonomy.
- Use enterprise structures (e.g., directory structures).
- Use automated clustering.

Each has its benefits and drawbacks, but using a combination of these techniques can often meet most needs.

The quickest way to develop a taxonomy is to simply use an existing one. Publicly available classification schemes, such as the Library of Congress Subject Headings, cover a wide range of topics but may not be suited to commercial organizations because of their focus on comprehensive coverage of top-level topics. Industry- and discipline-specific taxonomies are widely available and often provide a good starting point. Remember to match the coverage of the taxonomy to your specific needs. For example, a taxonomy from an electrical engineering organization will work well for electrical engineers but may not work as well for teams that combine electrical engineering, computer science, and chemistry experts. Multidisciplinary teams tend to focus on particular problems (e.g., low power consumption circuit design), and taxonomies organized around those problems are better suited than discipline-centric ones. General business taxonomies are available from news aggregation services that have often developed the classification schemes for their own use. Even if a third-party taxonomy is not a perfect fit "out of the box," it can be combined with schemes developed in-house.

For as long as we have had subdirectories in file systems, we have been categorically organizing content. Many organizations have large shared directories

organized around business processes, organizational structures, and ad hoc practices. These directory structures are useful starting points for building taxonomies because they tend to reflect the way users, at least some users, organize their work. When using network directory structures as a guide we need to remember that some subdirectories are created for ad hoc tasks, some are used simply to share files much like an ftp site, and some are no longer used but continue to exist because of poor directory management practices. Nonetheless, within the sometimes sprawling directory structures we can find elements of organizational structures that reflect existing business processes.

Automatic clustering of documents can also provide insight into the logical grouping of content. Basically, the process involves analyzing patterns within documents and grouping documents with similar patterns. This technique is useful when the logical grouping of documents is not clear, for example, when doing research in an unfamiliar domain. Clustering can definitely help discern the groups of documents, but it cannot be the sole technique used to define taxonomies. Not all groups identified by the clustering tool will make sense. Clustering tools can name the groups using terms that frequently occur in the member documents, but these are generally insufficient labels for end users.

When developing taxonomies, our primary focus must be on the way users think about their domain, not what third-party experts have decreed and not on the output of automatic categorization algorithms. There is no "right" answer. The best taxonomies are the ones that match the users' model of the organization and its processes.

Taxonomies are typically, although not exclusively, hierarchical. Taxonomies allow us to think about topics in relation to broader and narrower topics. When we make these distinctions between broader and narrower topics we are doing it based on some overriding concept. For example, when we navigate from a point labeled "United States" to "Illinois" to "Chicago," the overriding concept is geography. When we navigate from "United States" to "Federal Government" to "Supreme Court," the overriding concept is government structure. Clearly we can categorize topics in different ways depending on our particular interest at the time.[1]

[1] For readers especially interested in keeping up-to-date on developments in taxonomies and other information organization schemes, I suggest subscribing to Ramana Roa's *Information Flow* newsletter (http://www.ramanarao.com/informationflow/).

Similar structures exist in OLAP applications. We can think about sales figures by product, by sales region, and by time. Traditionally, these organizing principles are called **dimensions**. In the world of taxonomies and content management, we refer to these as **facets**.

Visualization Tools for Portal Organization

Structural organization techniques like taxonomies, facets, and metadata can aid information retrieval, but even with well-designed search queries that target particular subsets of content, a user can still be overwhelmed with the number of items returned by conventional search engines. Visualization tools effectively reduce information overload by mapping content to visual representations that aggregate content while highlighting significant relationships. For example, visualization tools can rapidly show a high-level structure, such as a site map, while allowing users to easily navigate to a particular area for more detail. In addition, these schemes depict content areas that users may not know exist and shows a broad context for content.

A number of different techniques are used to visualize content repositories such as portals. One technique, called **focus+context**, highlights one area (the focus) while showing the relationship of that area to other areas (the context). InXight's VizServer uses focus+context to allow users to navigate hyperlinked documents. A study at Xerox Palo Alto Research Center (PARC), which spun off InXight to commercialize the visualization tools, found that users could browse 62% faster with a focus+context tool than with a traditional two-dimensional tree layout such as Microsoft Explorer [Pirolli et al.].

Relational navigation is a visualization technique that depicts database relationships rather than hyperlinks and is useful for browsing database-driven sites. ThinkMap from Plumb Design software is just one example of this type of tool.

Portals frequently integrate content from hyperlink sources, databases, file systems, and other repositories. In those cases, visualization constructs based on business processes and structures tend to work best. The case study on Center Partners describes one such application.

CASE STUDY: Visualization and Logical Restructuring Improve Customer Care Services at Center Partners

The outsourced customer care industry is demanding on all parties involved. Customer care center providers have to understand the details of their clients' business policies, products, and services and convey that understanding to callers with widely varying needs. Client companies entrust customer care centers to attract new customers, build customer loyalty, provide technical support, and provide other critical points of contact with their customers. Maintaining quality standards is essential. The agents who operate the centers are under the sometimes conflicting demands of maintaining quality service while minimizing the average call time. Training agents and providing them with rapid information retrieval tools is essential to the success of these operations. One customer care center, Center Partners (http://*www.centerpartners.com*), used enterprise portal and visualization-based information retrieval software to improve quality measures while reducing call time to the point of generating an additional $500,000 per year in revenue.

The Problem: Poorly Organized Distributed Content

Center Partners operates seven offices with 2,500 customer care agents throughout the United States to serve Fortune 500 clients. This customer care firm generates annual revenue of $100 million and serves customers in industries ranging from insurance and financial services to pharmaceuticals and high tech. This service provider, like many in this area, measures operational success by the time required to handle customer calls and the level of service quality provided. Reducing the average call-handling time increases the efficiency of the center; quality measures are required by contractual agreements with customers. Both objective measures are served by improving agents' access to information.

The breadth and depth alone of the information customer care agents must tap in the course of their work demands a structured content management mechanism. Unfortunately, Center Partners does not control its information sources; the clients do. It is not unusual for the firm to depend on corporate extranets designed for multiple purposes that lack the features needed for fast-paced call center operations. David Geiger, Chief Information Officer of Center Partners, reported one client changed 2,800 extranet

pages in a single month. "The dynamic nature of the content made it nearly impossible for a busy agent to be up-to-date on the latest information. Moreover, the site, as is common with most Web sites, was difficult to navigate and did not offer any useful search capabilities" [TheBrain Technologies 2002]. Even well-designed customer sites are more likely suited for online shopping than the kind of troubleshooting tasks faced by agents.

Since Center Partners could not redesign its customers' sites, the company instead deployed a middle-tier content organization application in its portal that allowed them to organize content in a manner that better fits the way the agents work. Figure 1.6 shows how a middle tier hides the underlying complexity of the application.

The Solution: Visualization and Logical Restructuring

Center Partners chose BrainEKP from TheBrain Technologies Corporation (http://www.thebrain.com), an enterprise knowledge portal tool with strong visualization and search features, to provide a middle layer for the new navigation scheme without redesigning customer content. BrainEKP

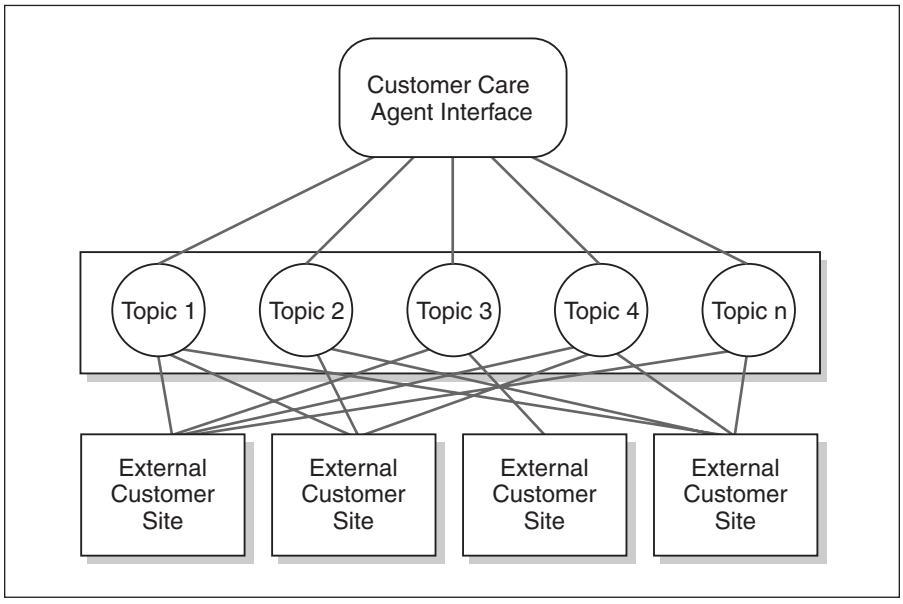

Figure 1.6 A middle tier can provide a logical organization that reflects users' understanding of a domain without requiring changes to the underlying implementation.

completely resides on Center Partners servers; it is external to customer sites and requires no changes on the customers' part. Domain experts at Center Partners first organized content into topics that support particular tasks, such as describing a refund policy or troubleshooting a service problem. Each topic (called a *thought* in TheBrain's terminology) provides links to content distributed across the Web, intranets, file systems, and databases.

Unlike typical search engines, selecting a topic in BrainEKP immediately displays related topics as well as content specific to the topic of interest. For example, navigating to the Repair topic displays the Tier 1 Repair Website while depicting links to vendor-specific repair information, as shown in Figure 1.7. BrainEKP also provides full-featured search using an embedded version of Convera RetrievalWare. To further improve the effectiveness of BrainEKP, Center Partners allows agents to customize their information sources and add their own notes and links.

The return on investment was clear for Center Partners. Quality increased enough to generate an expected additional $500,000 in revenue

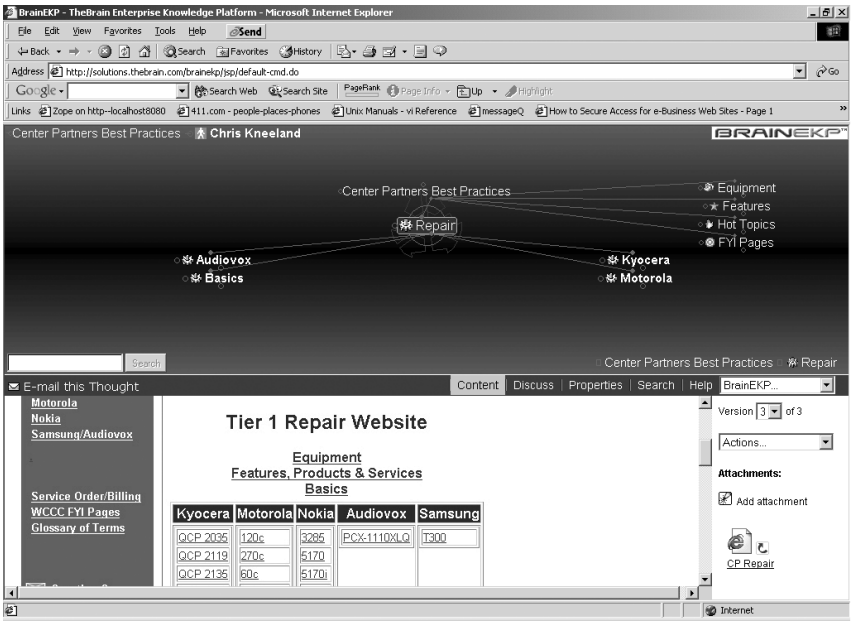

Figure 1.7 BrainEKP depicts the logical relationship between topics while displaying detailed content on the topic of focus.

over the course of a year and dropped the average call-handling time by 43 seconds.

Best Practices

Center Partners realized these gains by recognizing a number of best practices, including the following:

- **Use visualization to reduce the time required to retrieve information.**
- **Customize the logical organization of content through a middle tier rather than implementing a single organizational scheme that serves all usage models.**
- **Allow users to customize the portal to meet their particular needs and style.**

Content used with permission of Harlan Hugh, chief technology officer and co-founder of TheBrain; David Gerger, CIO Call Center Partners, Chris Kneeland, chief learning officer, Call Center Partners, TheBrain Technologies Corporation.

Emerging Trends and Technologies

As the scope of enterprise portals grows, the associated information retrieval problems will worsen if new technologies and techniques are not enlisted. One of the primary values of portals is that they allow us rapid access to a broad range of content and applications; however, that benefit is undermined by the very success and growth of portals. To ensure that we can keep ahead of the information overload problem today we must design portals with effective search, directory, and information architectures. In the near future we will require additional tools to manage content and to allow our applications to assume more responsibility for weeding out irrelevant information.

Researchers have worked on problems in natural language processing and knowledge representation for over 40 years, and the practical, commercial benefit of those undertakings will be realized in enterprise information portals (among other applications). The most pressing problem for advanced portal users and designers is to create applications customized to particular users' needs. Several systems and technologies are especially important to this effort.

- *Cyc:* Cyc is a knowledge base of over 100,000 terms, over 1,000,000 facts or assertions, and a reasoning engine for drawing conclusions about those facts. Cyc was developed by Cycorp (http://www.cyc.com) and has been used in

organizations ranging from Lycos, which uses it to improve search engine results, to the U.S. Department of Defense, which has invested heavily in Cyc development for military applications. The use of large-scale, general knowledge bases may help improve search and navigation in portals by improving the modeling of user behavior.

- *MESH:* Medical Subject Headings (MESH) is a controlled vocabulary thesaurus developed by the U.S. National Library of Medicine with over 21,000 descriptors, over 132,000 supplementary descriptors in a separate chemical thesaurus, and thousands of cross-references between terms. MESH is used to index articles from over 4,000 biomedical journals as well as the MEDLINE database.
- *WordNet:* WordNet is an online lexical reference developed by researchers at Princeton University based on psycholinguistic theories about human memory. WordNet contains over 146,000 words and over 195,000 word senses. WordNet contains both word senses (e.g., ten meanings of the word *book* and five meanings of *search*) and synonym sets of related terms. This lexical resource is currently used in some search engines to improve search results.

Cyc, MESH, and WordNet are currently used or have the potential for use in enterprise portals. These are just three examples of the general and specialized knowledge representation tools that are of growing importance to portals. Much of the work now under way in knowledge representation centers around three approaches.

1. Ontologies are organized representations of concepts, often for specific domains, such as pharmaceuticals, health care, and electronics. The Cyc knowledge base supports multiple ontologies. (See http://ksl-web.stanford.edu/kst/ontology-sources.html.)
2. Topic maps are groups of addressable information objects (e.g., documents) around topics and the relationship between those objects. The TopicsMaps.org consortium is developing standards to use XML to develop topic maps for the Web. (See http://www.topicmaps.org/.)
3. Semantic Web is an effort to embody semantic information in Web resources so that both humans and automated agents can more effectively manage those resources. (See http://www.w3.org/2001/sw/.)

With these technologies emerging in the portal arena, what can users realistically expect in the next few years? First, anticipate improved search capabilities in highly specialized domains like pharmaceuticals and medicine. Second, expect

incremental improvements in general search and categorization. Third, do not expect radical breakthroughs. Technologies like the semantic Web hold great promise, but much of the work is still in the research phase. Finally, we will continue to have significant amounts of manual work, from developing and tuning ontologies to defining topic maps. These technologies, however, will make it easier to share knowledge bases across applications.

Conclusion

Sound structural design is as important to portals as it is to buildings. The interface is a user's introduction to the portal. It is also the key to accessing the content and services provided by the portal. As we design and deploy portals we should consider how to organize information and applications in a way that makes sense to users. This of course is a problem for several reasons. First, there are many users and their needs vary. Second, how even a single user uses the portal depends on the task he or she is trying to accomplish. If performing a routine task, for example, entering a time card, the user will want rapid access to the application. This is no place for needless clicks through a hierarchical set of applications. On the other hand, if the user is researching a new product line, he or she will want to browse through related content, follow promising paths of related information, and quickly narrow the search in response to hunches about new angles on the problem. The only way to meet these needs is to provide multiple ways to navigate and to keep the overall organization consistent.

Content within pages should follow a pattern. The three-panel model balances formal structure with flexible organization of content. Landmarks, active links, and other navigation techniques will help users quickly move around within the portal.

Organizing content around navigation sets provides a sense of context for users. They will be able to move easily between related pages while the page-level navigation patterns, like landmarks back to a home page, will help the user move to other areas quickly.

As the amount of information in the portal grows, users will need additional support to find what they need. Taxonomies and faceted models provide users with tools to see the forest and quickly focus on a specific tree. For especially large portals, visualization techniques can further improve navigation.

With a grounding in techniques for designing interfaces, in the next chapter we will turn our attention to the underlying frameworks that support the portal's core functions and services.

References

Krug, Steve, and Roger Black. 2000. *Don't Make Me Think: A Common Sense Approach to Web Usability.* Indianapolis, IN: Que.

Nielsen, Jakob. 1999. *Designing Web Usability: The Practice of Simplicity.* Indianapolis, IN: New Riders.

Pirolli, Peter, Stuart K. Card, and Mija M. Van der Wege. (Year not stated.) "The Effect of Information Scent on Searching Information Visualization of Large Tree Structures." Accessed in June 2003 at *http://www.inxight.com/pdfs/PARCstudy.pdf.*

Rossi, Gustov, Daniel Schwabe, and Fernando Lyardet. (Year not stated.) "Improving Web Information Systems with Navigational Patterns." Accessed in June 2003 at *http://www8.org/w8-papers/5b-hypertext-media/improving/improving.html.*

TheBrain Technologies. 2002. "Center Partners: Improving Call Handle Times and Agent Best Practices." Received via e-mail from TheBrain.