CHAPTER 1

Defining Information Architecture

We shape our buildings: thereafter they shape us.
—Winston Churchill

What we'll cover:

- What is (and isn't) information architecture
- Why information architecture is important
- The value of explaining and illustrating IA concepts

What is it about buildings that stirs us? Whether we're architectural connoisseurs or just plain folks, we are all emotionally engaged by the physical structures we experience throughout our lives.

Each building serves a different purpose. A bustling café with hardwood floors and large windows facing Main Street provides the ideal place for a quick breakfast meeting. A steel-and-glass high-rise with its mix of cubes and offices envelops inhabitants in a collaborative, high-energy work environment. A dark, smoky bar with tin ceilings and exposed brick walls becomes a sanctuary from the whirl of modern life. And a medieval Gothic cathedral adorned with granite sculptures, stained-glass windows, and towers that reach for the heavens provides an experience both humbling and inspirational.

Each building serves its purpose uniquely. Architecture, design, construction, furnishings, inhabitants, and location all play major roles in shaping the overall experience. All elements must work together. In successful buildings, the whole is greater than the sum of its parts.

Why begin a book about web sites by writing about buildings? Because the architectural analogy is a powerful tool for introducing the complex, multidimensional nature of information spaces. Like buildings, web sites have architectures that cause us to react.

Some web sites provide logical structures that help us find answers and complete tasks. Others lack any intelligible organization and frustrate our attempts to navigate through them. We can't find the product we need; we can't locate the report we found last week; we're lost inside an online shopping cart. These web sites may

remind us of buildings that fail: houses with flat roofs that leak, kitchens with no counter space, office towers with windows you can't open, and mazelike airports with misleading signs.

Bad buildings and bad web sites share similar architectural roots. First, many architects don't inhabit the structures they design. They don't fully understand the needs of their customers, and they're not around to suffer the long-term consequences of poor decisions. Second, creating structures to stand the test of time is really difficult. Needs change. Surprises are the rule. The desire for stability must be balanced against the value of flexibility and scalability. Architects are often faced with complex requirements, competing goals, and high levels of ambiguity. Transforming this chaos into order is extremely hard work that requires rare vision and perspective.

However, as designers of web sites, we should not become trapped by the metaphor of building architecture. Throughout this book, we'll also talk about information ecologies, knowledge economies, digital libraries, and virtual communities. We learn what we can from each analogy, and we leave the baggage behind.

A Definition

If you're new to the field, you may still be wondering: what exactly is information architecture? This section is for you.

> **in•for•ma•tion ar•chi•tec•ture** n.
>
> 1. The structural design of shared information environments.
>
> 2. The combination of organization, labeling, search, and navigation systems within web sites and intranets.
>
> 3. The art and science of shaping information products and experiences to support usability and findability.
>
> 4. An emerging discipline and community of practice focused on bringing principles of design and architecture to the digital landscape.

Were you expecting a single definition? Something short and sweet? A few words that succinctly capture the essence and expanse of the field of information architecture? Keep dreaming!

The reason we can't serve up a single, all-powerful, all-purpose definition is a clue to understanding why it's so hard to design good web sites. We're talking about the challenges inherent in language and representation. No document fully and accurately represents the intended meaning of its author. No label or definition totally captures the meaning of a document. And no two readers experience or understand a particular document or definition or label in quite the same way. The relationship between words and meaning is tricky at best.*

* For a humorous perspective on the trickiness of the English language, see Bill Bryson's *The Mother Tongue: English & How It Got That Way* (William Morrow).

We'll now descend from our philosophical soapbox and get down to basics. Let's expand on our definitions to explore some basic concepts of information architecture.

Information

We use the term *information* to distinguish information architecture from data and knowledge management. Data is facts and figures. Relational databases are highly structured and produce specific answers to specific questions. Knowledge is the stuff in people's heads. Knowledge managers develop tools, processes, and incentives to encourage people to share that stuff. Information exists in the messy middle. With information systems, there's often no single "right" answer to a given question. We're concerned with information of all shapes and sizes: web sites, documents, software applications, images, and more. We're also concerned with *metadata*: terms used to describe and represent content objects such as documents, people, processes, and organizations.

Structuring, organizing, and labeling

It's what information architects do best. Structuring involves determining the appropriate levels of granularity* for the information "atoms" in your site, and deciding how to relate them to one another. Organizing involves grouping those components into meaningful and distinctive categories. Labeling means figuring out what to call those categories and the series of navigation links that lead to them.

Finding and managing

Findability is a critical success factor for overall usability. If users can't find what they need through some combination of browsing, searching, and asking, then the site fails. But user-centered design isn't enough. The organizations and people who manage information are important, too. An information architecture must balance the needs of users with the goals of the business. Efficient content management and clear policies and procedures are essential.

Art and science

Disciplines such as usability engineering and ethnography are helping to bring the rigor of the scientific method to the analysis of users' needs and information-seeking behaviors. We're increasingly able to study patterns of usage and subsequently make improvements to our web sites. But the practice of information architecture will never be reduced to numbers; there's too much ambiguity and complexity. Information architects must rely on experience, intuition, and creativity. We must be willing to take risks and trust our intuition. This is the "art" of information architecture.

* *Granularity* refers to the relative size or coarseness of information chunks. Varying levels of granularity might include: journal issue, article, paragraph, and sentence.

Tablets, Scrolls, Books, and Libraries

Humans have been structuring, organizing, and labeling information for centuries. Back in 660 B.C., an Assyrian king had his clay tablets organized by subject. In 330 B.C., the Alexandria Library housed a 120-scroll bibliography. In 1873, Melvil Dewey conceived the Dewey Decimal System as a tool to organize and provide access to the growing number of books.

In modern times, most of us become familiar with the basics of information organization through our experiences with books and libraries. Table 1-1 shows how the concepts of information architecture (IA) apply to the world of print and the World Wide Web.

Table 1-1. Differences between books and web sites

IA concept	Books	Web sites
Components	Cover, title, author, chapters, sections, pages, page numbers, table of contents, index	Main page, navigation bar, links, content pages, sitemap, site index, search
Dimensions	Two-dimensional pages presented in a linear, sequential order	Multidimensional information space with hypertextual navigation
Boundaries	Tangible and finite with a clear beginning and ending	Fairly intangible with fuzzy borders that "bleed" information into other sites

As we go beyond books to collections of books, the comparisons become even more interesting. Imagine a bookstore with no organization scheme. Thousands of books are simply tossed into huge piles on table tops. Such a bookstore does, in fact, exist: Gould's Book Arcade in Newtown, Australia. It's shown in Figure 1-1.

Figure 1-1. Gould's Book Arcade (image courtesy of Seth Gordon)

From a philosophical perspective, you might feel that this casual jumble of books represents a refreshing break from the rigid structures of everyday life. And this bookstore really can provide a wonderful browsing experience filled with adventure and serendipity. But if you arrive seeking a specific book or if you have a particular author or topic in mind, you're almost guaranteed to have a long and painful needle-in-the-haystack experience.

Compare the chaos of this bookstore to the order of a library (see Figure 1-2). Even on the surface, the contrast is like night and day. But look deeper and you'll see that a library is more than a warehouse for books, magazines, and music. There are complex systems and well-trained professionals operating behind the scenes to select, evaluate, label, describe, structure, and organize the collection so that users of the library can find what they need. And though the library's information environment is highly structured, the subject-oriented approaches of the Dewey Decimal and Library of Congress classification schemes also support exploratory browsing and serendipity.

Figure 1-2. Browsing in a library (image courtesy of http://intergate.sdmesa.sdccd.cc.ca.us/lrc/stacks.jpg)

In short, a major way that libraries and librarians add value to printed materials is by placing them within the framework of an information architecture that facilitates access to those materials. Information architects perform a similar role, but we typically do it within the context of web sites and digital content. Of course, there are major differences between libraries and web sites. Table 1-2 shows just a few.

Table 1-2. Differences between libraries and web sites

IA Concepts	Libraries	Web sites
Purpose	Provide access to a well-defined collection of formally published content.	Provide access to content, sell products, enable transactions, facilitate collaboration, and on and on…
Heterogeneity	Diverse collections with books, magazines, music, software, databases, and files.	Huge diversity of media types, document types, and file formats.
Centralization	Highly centralized operations, often within one or a few physical library buildings.	Often very decentralized operations, with subsites maintained independently.

Developing an information architecture for a library presents many challenges, but a library is a relatively well-defined environment, and there is much collective experience and wisdom to draw upon. Web sites, on the other hand, present an array of new challenges. Virtual spaces are more flexible than physical spaces and can therefore be more complex. And at this point, we have precious few guidelines for creating information architectures for digital spaces.

Obviously, we've made some gross generalizations in these comparisons, and have oversimplified to illustrate key points. As you try to communicate information architecture concepts to others, you'll probably have to do the same.

Explaining IA to Others

One of the most frustrating things about being an information architect is the fact that most of your family members and neighbors will never have a clue what you do. The more you try to explain it, the more confused or bored they become. Their eyes glaze over. They nod politely. Then comes the desperate attempt to change the subject. "Hey, speaking of information architecture, did you hear tomorrow's weather report?"

Friends and relatives aren't the only tough audience. Sometimes you have to sell the concept to colleagues, clients, or managers. Each audience presents its own set of challenges. There's no magic bullet, but it's helpful to be prepared with an "elevator pitch" and an analogy suited to your particular audience.

The elevator pitch explains what you do in a sentence or two of plain language. If you can combine an analogy that resonates with your audience, even better!

Here are a few approaches to try out:

- "I'm an information architect. I organize huge amounts of information on big web sites and intranets so that people can actually find what they're looking for. Think of me as an Internet librarian."
- "I'm an information architect. I help my company by making it easy for customers to find our products on our web site. I'm a kind of online merchandiser. I apply one-to-one marketing concepts on the Internet."
- "I'm an information architect. I'm the one who takes on that information overload problem that everyone's been complaining about lately."

Sometimes we're too close to what we do. That's when it's a good idea to call for help. Ask someone who's familiar with you and your job to describe what you do in one or two sentences. Often you'll be amazed by how well they nail it, and grateful for their clarity and brevity.

What Isn't Information Architecture?

One of the most effective ways to define something is to identify its boundaries. We do this all the time. This is my property. That's your property. This is England. That's Scotland. She's a brain surgeon. He's an ophthalmologist.

Sometimes it's very easy to explain the differences. Mammals breathe with their lungs and give birth to live young. Dogs, cats, dolphins, and humans are all clearly mammals. Fish live in water, breathe with their gills, and lay eggs. Salmon, bass, and guppies are all clearly fish.

But as with many classifications, you quickly run into problems. What about fish with lungs? What about fish that don't look like fish? Are sharks, skates, eels, and sea horses really fish? (Yes, they are.) And where do we put that darned platypus?* Biological taxonomists have argued about these classification issues for centuries.

Mapping the boundaries of information architecture is even more slippery. Some things are clearly not information architecture:

- Graphic design is NOT information architecture.
- Software development is NOT information architecture.
- Usability engineering is NOT information architecture.

Makes sense, right? But as soon as you start working within the messy reality of web site design and construction, you find yourself in the gray areas between disciplines. For example, consider the ubiquitous global navigation bars in Figure 1-3.

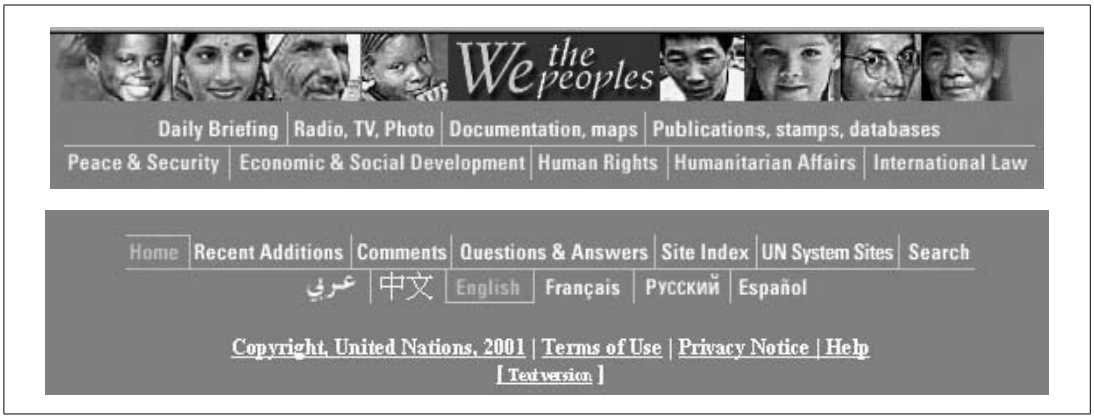


Figure 1-3. Top and bottom navigation bars on the United Nations web site

The navigation bars feature labels and links that lead to other sections and pages within the web site. These labels are dependent upon the underlying structure and categorization of the site. The creation of categories and choice of labels fall clearly inside the domain of information architecture.

* To find out, read *The Platypus and the Mermaid: And Other Figments of the Classifying Imagination*, by Harriet Ritvo (Harvard University Press).

But wait a second. What about the look and feel of the navigation bar? What about the choice of colors, images, font styles, and sizes? Now we enter the realms of graphic design, interaction design, and information design. And what if a designer challenges the labels proposed by an information architect? Perhaps those labels are too long to fit on the navigation bar. What happens then?

What if the information architect wants a search link on the navigation bar, but the software developer says that adding a search capability to the web site is too expensive and time-consuming? And what if the usability engineer says that user tests indicated there are too many options on the navigation bar? What happens then?

These are the questions and challenges that live in the gray areas between disciplines. These gray areas drive some people crazy. Lots of heated arguments have resulted from attempts to draw clear lines. We believe the gray areas are necessary and valuable. They force interdisciplinary collaboration, which ultimately results in a better product.

Gray areas and caveats aside, here is our attempt to draw some boundaries between information architecture and a number of closely related disciplines.

Graphic design
: Traditionally, a graphic designer was responsible for all aspects of visual communication, from the design of corporate logos and identities to the layout of individual pages. On the Web, we're seeing increasing specialization due to the complexity of the environment. Even so, many graphic designers do a great deal of information architecture as part of their work.

Interaction design
: Interaction designers are concerned with the behavior of tasks and processes that users encounter in software and information systems at the interface level. They often have a background in human–computer interaction, and are focused on helping users successfully achieve goals and complete tasks.

Usability engineering
: Usability engineers understand how to apply the rigors of the scientific method to user research, testing, and analysis. Their background in human–computer interaction and their experience observing users provide them with useful insights into design. They are often concerned with testing all aspects of the user experience, inclusive of information architecture and graphic design.

Experience design
: Experience design is an umbrella term that encompasses information architecture, usability engineering, graphic design, and interaction design as components of the holistic user experience. You'll find relatively few "experience designers," as there aren't many people with skills in all these areas. The term is useful insofar as it encourages cross-disciplinary awareness and collaboration.

Software development
: People rarely confuse software development and information architecture, but the two fields are highly interdependent. Information architects rely on developers to

bring our ideas to fruition. Developers help us understand what is and isn't possible. And as the Web continues to blur the distinction between software applications and information systems, these collaborations will become even more important.

Enterprise architecture
: In the 80s and 90s, a movement calling itself *enterprise architecture* arose in the information systems discipline. While the early stages of this movement were focused on data and system integration, later definitions have encompassed business, process, information, and technology architecture.

Content management
: Content management and information architecture are really two sides of the same coin. IA portrays a "snapshot" or spatial view of an information system, while CM describes a temporal view by showing how information should flow into, around, and out of that same system over time. Content managers deal with issues of content ownership and the integration of policies, processes, and technologies to support a dynamic publishing environment.

Knowledge management
: Knowledge managers develop tools, policies, and incentives to encourage people to share what they know. Creating a collaborative knowledge environment means tackling tough issues surrounding corporate culture such as "information hoarding" and "not-invented-here syndrome." Information architects focus more on making accessible what has already been captured.

Why Information Architecture Matters

You now understand what information architecture is and what it isn't. So, why is it important? Why should you care? Why should your company or your clients invest time and money in the design of their information architectures? What is the return on investment (ROI)?

We'll tackle these tough questions in detail later in the book, but for now, let's hit the highlights without getting bogged down in subtleties. When you calculate the importance of information architecture to your organization, you should consider the following costs and value propositions:

The cost of finding information
: What does it cost if every employee in your company spends an extra five minutes per day struggling to find answers on your intranet?* What is the cost of frustrating your customers with a poorly organized web site?

* Jakob Nielsen deserves credit for publicizing the fact that the costs of poor navigation-system design in a large enterprise can add up to millions of dollars of lost employee productivity.

The cost of not finding information
: How many bad decisions are made every day in your organization because employees didn't find the information they needed? How much duplication of effort results from this disconnect? How many customers do you lose because they can't find the product they want on your web site? How much do you spend every day providing telephone support to existing customers because they hate navigating your online technical-support database?

The value of education
: What is the value of educating your customers about new products and services related to the ones they're actively seeking on your web site?

The cost of construction
: What does it cost to design and build a web site? How much does it cost to redo it six months later because it doesn't support findability or doesn't scale?

The cost of maintenance
: Similarly, what does it cost to ensure that good designs don't crumble over time? Will the people who maintain your site know where to put new content and when to remove outdated content?

The cost of training
: For internal, mission-critical information systems that support call centers, for example, what does it cost to train employees to use that system? How much could you save if it wasn't so complicated to use?

The value of brand
: No matter how beautiful your web site is, if customers can't find what they need, your brand loses value in their eyes. How much did you spend on those brand-building TV commercials?

And the list goes on. In your particular situation, there are sure to be a whole slew of opportunities to make money, save money, improve employee or customer satisfaction, or just plain make the world a better place. Figure out what they are and communicate them as clearly and directly as possible.

We're not saying this is easy. In fact, it's very difficult to calculate an exact return on an information architecture investment—there are simply too many variables. This is really no different from most other areas of activity within the business world. It's just that people in more traditional areas like sales, marketing, engineering, human resources, and administration have had more time to get their stories straight.

Bringing Our Work to Life

Information architecture lives beneath the surface. Users rarely look at a web site and exclaim, "Wow, check out this brilliant classification scheme!" In fact, much of our work is intangible; many people who are directly involved in web design have only a superficial understanding of information architecture. They may recognize the need

for clear labels in a navigation bar, but have no clue how a controlled vocabulary could improve the search experience. If you can't see it, touch it, taste it, or smell it, it doesn't exist.

This invisibility is fine with respect to users. We don't want to force users to see our hard work; we want them to complete tasks and find information in blissful ignorance of our labors. But invisibility is a major problem when it comes to justifying our existence to colleagues and making the case for investments to decision makers. We must constantly work to help people see the complexity of the challenges we face and the long-term value of our solutions.

We must find ways to articulate the key concepts of our craft, helping people to understand the sophisticated nature of user needs and behavior. We must show the interconnections between people and content that underpin knowledge networks, and explain how these concepts can be applied to transform static web sites into complex adaptive systems (Figure 1-4*).

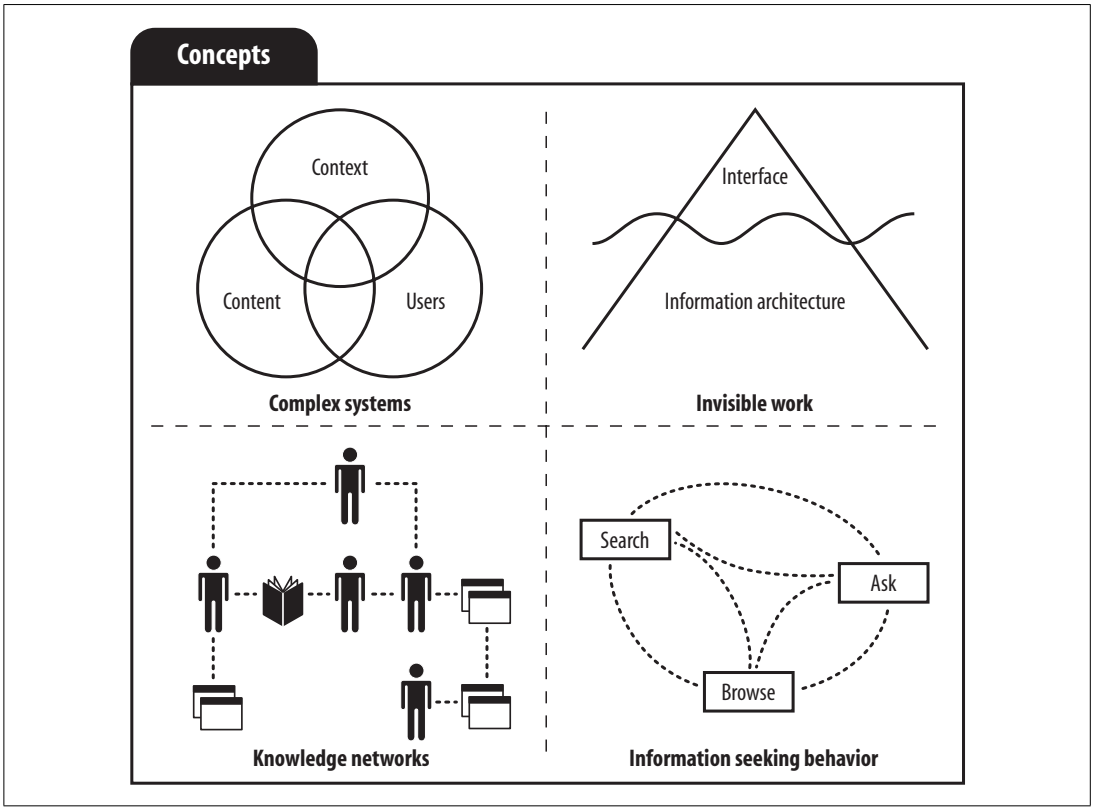


Figure 1-4. Information architecture concepts

* This series of images was designed by Myra Messing Klarman of Studio Mobius (*http://studiomobius.com*).

We must be prepared to dive into detail, identifying and defining the component systems that support our sites (Figure 1-5). We must show how semantic networks can provide a foundation for fluid navigation. And we must convince our clients and colleagues that an effective searching experience requires not just a good engine or a nice interface, but a carefully integrated system of interdependent parts.

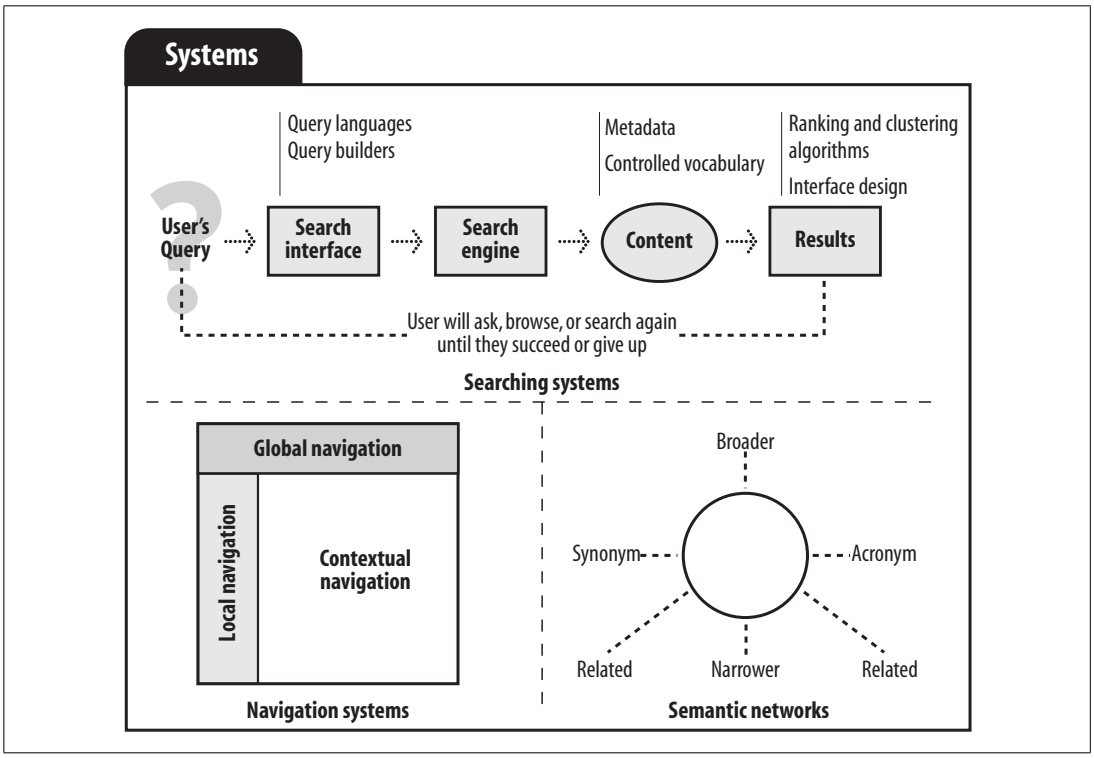


Figure 1-5. Information architecture systems

Finally, we must be ready to produce concrete deliverables (Figure 1-6). We must learn to render our constructs of semantics and structure in clear and compelling ways. In short, we must help people to see the invisible.

In this book, we explain the concepts, systems, and deliverables of information architecture. By drawing upon words, stories, metaphors, and images, we've done our best to bring our work to life. However, no single collection of words and images can serve all purposes. A key to the craft of information architecture is understanding how to shape your message for your audience. This requires some sense of what your managers, clients, and colleagues want to hear and how they want to hear it.

Did we mention that information architecture involves a little magic? How else would you read minds and make the invisible visible? So put on a black hat, bring along your sense of humor, and prepare to enter the secret society of information architects.

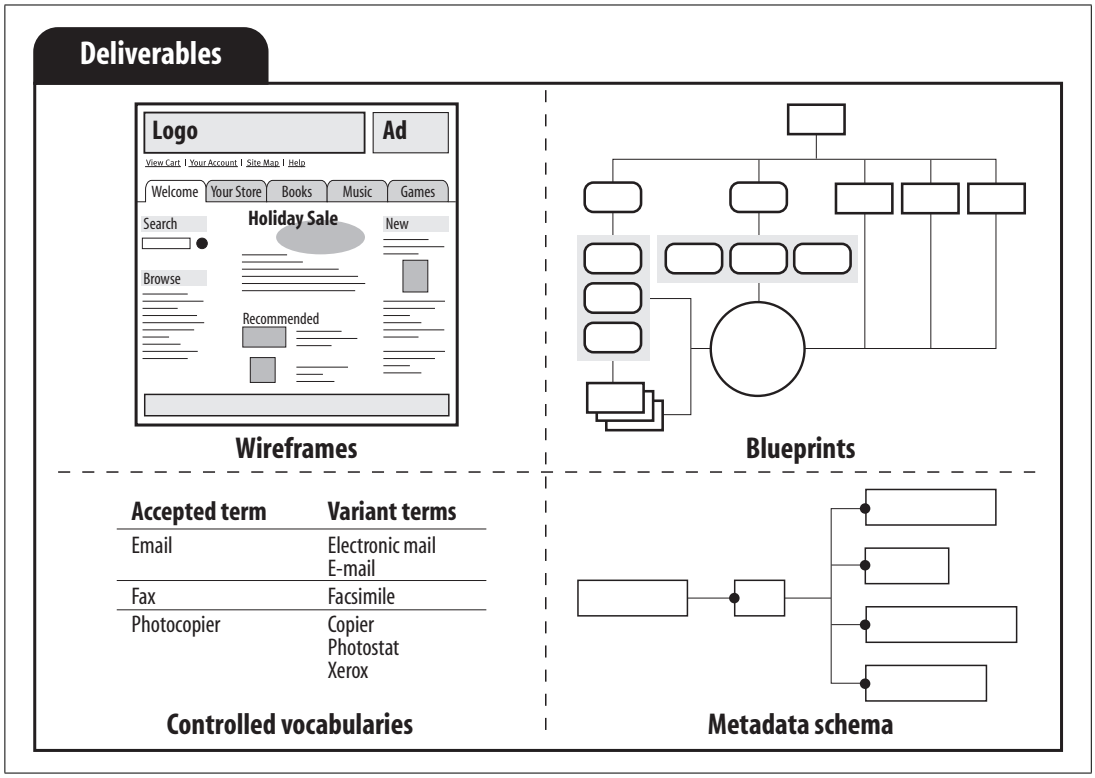


Figure 1-6. Information architecture deliverables

CHAPTER 2

Practicing Information Architecture

What we'll cover:

- Information architecture is everywhere
- Whether the world needs information architects
- Qualifications and source disciplines for information architects
- Information ecologies and their impact on the practice of information architecture

What is information architecture? Is it an art, a science, or a craft? Who should do this work? What qualifications are required? These are the questions we grapple with as a community of information architects. We write articles and publish books. We debate on discussion lists and argue passionately at conferences. We pull out our hair. We lose sleep. This is serious stuff.

And yet, independent of our intellectual theories and existential agonies, something very powerful is taking place. We are being surrounded, quite literally, by information architecture.

Have you ever walked through Times Square in New York City at night? It's quite a spectacle. You're on the corner of 42nd and Broadway. The glassy facades of buildings are pulsing with real-time information, courtesy of the latest in flat-panel display and projection technologies. Business news, financial data, corporate logos, and URLs are lit up in neon. Taxicabs sport billboards on their roofs as they honk their way through traffic. Pedestrians (or shall we say "users") hustle past one another, chattering into their cell phones or stopping on the corner to check email or get directions on their wireless PDAs. This is William Gibson's cyberspace turned inside out, physical architecture meets information architecture, a world of content, labels, and metadata all competing for your attention.*

* See the Flickr photo pool "Everyday Information Architecture": *http://www.flickr.com/groups/everyday-information-architecture/pool.*

And that's nothing compared to the real cyberspace, a new reality where we spend increasing amounts of time. How many hours do you spend staring at a computer monitor each day? How often do you check email or pop open your web browser? When your Internet connection is broken, how do you feel?

The World Wide Web has lived up to its name. It has connected and transformed the world. Want to know what's going on? Check out nytimes.com, bbc.co.uk, or your favorite blogs. Planning a trip? orbitz.com and kayak.com will meet your every need. Having trouble with your green iguana? No need to leave the house. You'll find the answer at iguana.com.

Billions of web pages have sprung up since the Web began. And guess what? Information architects played no role in designing most of them. This has been an emergent, bottom-up, grass-roots phenomenon. But every single web site that exists does have an information architecture. They're riddled with labels and taxonomies, vocabularies and metadata, sitemaps and indexes. There are portals linking to portals linking to search engines. Pure navigation. Some is good. Much of it isn't. We can critique it and we can make fun of it, but we can't stop it. Information architecture happens!

Do We Need Information Architects?

Since information architecture happens anyway, does the world really need information architects? If you've attended any of the IA Summits* in recent years, you know this has been a hot topic. A few speakers in particular have stirred the pot. Andrew Dillon is fond of saying, "I know we need information architecture. I'm not so sure we need information architects." And Peter Merholz suggests that "we need to teach everyone to do information architecture, rather than isolating the practice to a handful of professionals."

We have to give credit to the information architecture community for having the guts to ask these questions in public. But we'd like to respond with a firm assertion that *we absolutely do need information architects*. We're not too particular about the specific job title; if you prefer to call them user-experience designers, knowledge managers, or findability engineers, that's fine with us. What we're focused on is the need for professionals with specialized skills and experience, who know how to create useful, usable information systems within massively complex environments.

Programmers and graphic designers are great at what they do. They're not great at what we do. And information architecture design is not a skill you can pick up by taking a half-day seminar. There's real depth to the discipline. Information architecture resembles the games of *Othello* and *Go*. A minute to learn, a lifetime to master.

* Sponsored by the American Society for Information Science & Technology, the Information Architecture Summit is held in February or March each year. Learn more on the IA Summit web site: *http://iasummit.org*.

Does this mean that all web developers will need a licensed information architect on board before they write their first line of code? Of course not. Information architecture happens, with or without information architects, and that's just fine with us. That's why Peter Merholz is right to emphasize the vital role information architects must play in education. We can have a major positive impact on the world by sharing what we know with all those people who do information architecture in the course of doing something else.

But the most important and complex information environments already rely on professional information architects. Large organizations like IBM, Microsoft, and Vanguard already have teams of information architects dedicated to the long-term strategy and design of their web sites and intranets. Smaller organizations tend to involve information architects in a consulting capacity during a site redesign. This allows the information architect to make a major contribution without breaking the bank.

This selective use of expertise is not isolated to the field of information architecture; in fact, it is quite common. Consider, for example, the practice of law. A huge percentage of legal decisions are made every day by business managers rather than by their lawyers.

> Manager #1: "Should we approve this nondisclosure agreement?"
>
> Manager #2: "Yes, that's fine. It's no big deal. Let's move on."

Most companies don't have lawyers on staff. They get lawyers involved when the situation is particularly messy, complex, or important. The same happens and will continue to happen with information architects.

In fact, as web sites and intranets become more sophisticated and mission-critical, the demand for information architects will only rise. This demand will be partly offset as other professionals learn the basics of information architecture. Our responsibility as information architects will be to continue to push the envelope, to learn how to do what we do faster and better, and then to share our knowledge and experience with those around us. We all have so much to learn and so much to do. We fully expect information architects to be very busy for at least the next few hundred years.

Who's Qualified to Practice Information Architecture?

Unlike medicine and law, the field of information architecture has no official certification process. There are no university consortia, boards, or exams that can prevent you from practicing information architecture. As we explain in Chapter 13, a number of academic programs are emerging to serve the needs of prospective information architects, but for now very few people have a degree in information architecture.

Disciplinary Backgrounds

As you look over this list, you might not find your home discipline listed. Don't be daunted: any field focused on information and its use is a good source of information architects. And the field is still young enough that just about anyone will have to rely on experience from the School of Hard Knocks to practice IA effectively and confidently.

If you're looking for IA talent, keep in mind that, because the field is relatively new and because demand for information architects continues to explode, you can't just post a job description and expect a flock of competent and experienced candidates to show up on your doorstep. Instead, you'll need to actively recruit, outsource, or perhaps *become* the information architect for your organization.

Of course, if you are looking for someone else to fill this role, you might consider the following disciplines as sources for information architects. If you're on your own, it might not be a bad thing to learn a little bit about each of these disciplines yourself. In either case, remember that no single discipline is the obvious source for information architects. Each presents its own strengths and weaknesses.

OK, on to the list:

Graphic design and information design
: Many of the people who have written about and practice information architecture are graphic designers by training. This is not surprising, as both graphic design and information design involve much more than creating pretty pictures. These professions are geared more toward creating relationships between visual elements and determining how those elements can be integrated as a whole to communicate more effectively.

Information and library science
: Our backgrounds in information science and librarianship have proven very useful in dealing with the relationships between pages and other elements that make up a whole site. Librarians have a long history of organizing and providing access to information and are trained to work with searching, browsing, and indexing technologies. Forward-looking librarians understand that their expertise applies in new arenas far beyond the library walls.

Journalism
: Journalists, like librarians, are trained at organizing information, but in a setting that places special emphasis on timeliness. If your web site is geared toward delivering dynamic information, such as a news service or online magazine, someone with a background in journalism might have a great sense of how this information could be best organized and delivered. Because of their writing experience, journalists are also good candidates for architecting sites that will have high levels of edited content.

Usability engineering

Usability engineers are experts at testing and evaluating how people work with systems. These human–computer interaction professionals measure such criteria as how long it takes users to learn how to use a system, how long it takes them to complete tasks and find answers, and how many errors they make along the way. Of all the disciplines we list, usability engineering is probably the most scientific in its view of users and the quality of their experiences.

Marketing

Marketing specialists are expert at understanding audiences and communicating messages effectively. They are particularly valuable in the design of customer-facing web sites, where product sales and brand are critical to success. Marketing expertise can ensure that the message is presented in the language of the target audience. We've run into a number of "online merchandisers" who have become expert information architects.

Computer science

Programmers and software developers bring important skills and sensitivities to information architecture, especially to "bottom-up" processes. For example, developers are often excellent at modeling content and metadata for inclusion in a database or content management system. They're also great at figuring out how all of the component systems and technologies of an information architecture fit together.

Technical writing

Professionals who have spent time writing technical documentation or developing online help systems are often well-sensitized to both the needs of users and the potential for structuring, labeling, and describing textual content.

Architecture

While the transition from bricks and mortar to bits and bytes is obviously a big move, we actually know quite a few building architects turned information architects. These folks tend to have a great deal of experience studying people's needs and seeking behaviors, and an excellent foundation in the concepts and challenges surrounding strategy and design.

Product management

Many information architects play the role of "orchestra conductor." They understand how to tap the motivations and talents of a diverse group of professionals, creating a whole that's greater than the sum of its parts. People with a background in product, program, or project management can become very effective information architects, particularly in the areas of strategy formation and interdisciplinary team management.

…And many more

This list is far from comprehensive. There are dozens of established fields from which we can learn (see Figure 2-1). No list or picture will ever capture the true diversity of practicing information architects.

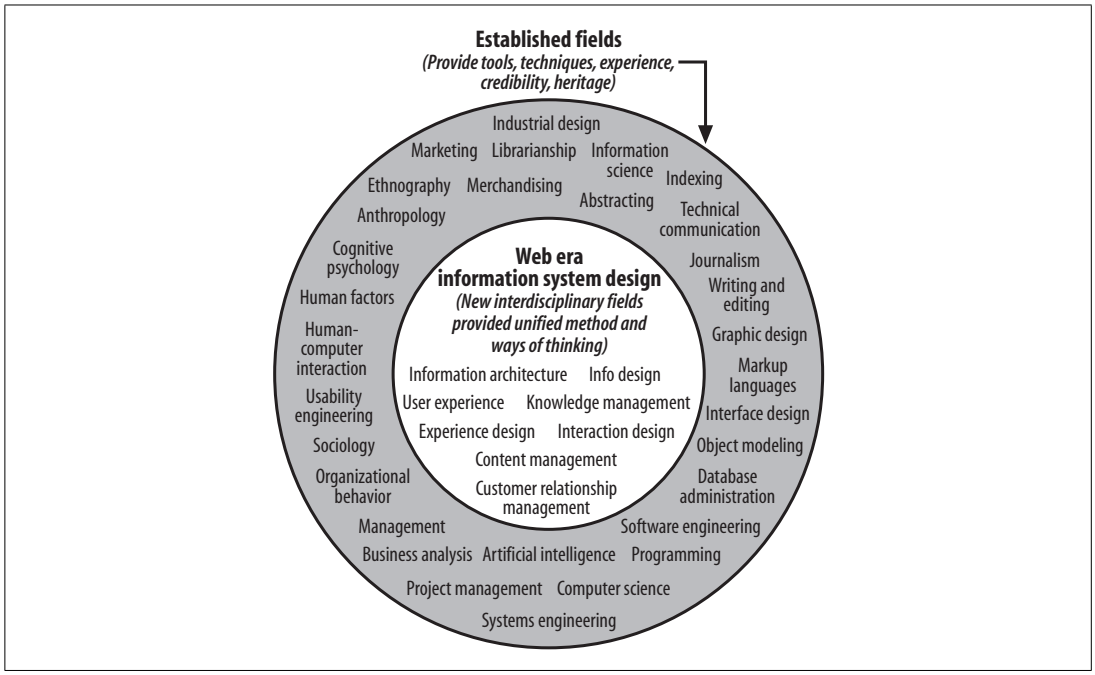


Figure 2-1. Information system design in the Web era (designed with help from Jess McMullin)

Innies and Outies

When staffing an information architecture project, it's also worth considering tradeoffs between insider and outsider perspectives. On one hand, there's value in having an information architect who can think as an "outsider," take a fresh look at the site, and be sensitive to the needs of users without being weighed down by internal political baggage. On the other hand, an "insider" can really understand the organization's goals, content, and audiences, and will also be around for the long haul, helping to design, implement, and manage the solution.

Because it's difficult to choose between these two perspectives, many intelligent organizations put together a balanced team of consultants and employees. The consultants often help with major strategy and design initiatives, and provide highly specialized varieties of IA consulting, while the employees provide continuity as projects transition into programs. Even if you're the lone in-house information architect, you should seek to work with outies—whether by convincing management to hire consultants or specialists for you to collaborate with, or simply by hanging out with and learning from other IAs at local meetups and conferences.

Really, the fact that both innies and outies are flourishing is a sign of the field's maturation: in IA's early years (coinciding with this book's first edition), most practitioners were outies, working at agencies and consultancies. After the bubble burst (see the second edition), many of us ran for cover in the security of working in-house, often assisting with the implementation and customization of large applications like CMS and search engines. And now, as our third edition comes out, the field is in balance—there

is room for both innies and outies, and a symbiotic relationship exists between them. It's truly indicative of a healthy profession, and good insurance against the vagaries of the next sudden economic downturn. We're not going away.

Gap Fillers and Trench Warriors

IA's early practitioners got their jobs by taking on work that no one else wanted or realized existed. Structuring information? Indexing it? Making it findable? Even if these tasks sounded appealing, few had the vocabulary, much less the skills, to address them. So stone-age information architects were, by definition, natural gap fillers who often tackled these tasks out of opportunism or simply because *someone* had to.

Over the past five or seven years, the field has matured and the practice of IA has solidified. What an information architect does is now much better understood and documented; you'll even detect a whiff of standardization among job descriptions. In effect, IA has moved from the exotic to the everyday, and more and more the people filling those roles are heads-down crack experts in the nuts-and-bolts of IA practices. These are the information architects that you'll want and need down in the trenches, grinding out an information architecture amid the guts and gore of your organization's users, content, and context. These trench warriors aren't pioneers, but providers of an important commodity service.

Of course, as trench warriors began to take over, gap fillers didn't disappear. They saw other opportunities that needed filling—only this time, the gaps popped up in the field of IA itself, rather than within specific teams or organizations. Information architects are now making livings as independent consultants, often working in such specialized areas such as taxonomy development, or as user experience team leaders, or as teachers and trainers for in-house IAs. Increasingly, many of us have become independent entrepreneurs who are developing own IA-infused products and services; there are always new gaps to be filled.

As the field continues its healthy evolution, gap fillers and trench warriors will continue to fill changing roles. Whether you're looking to staff your team, hire a consultant, or determine if IA is in your future, it's important to know that the field is now large and healthy enough to accommodate many personality types.

Putting It All Together

Whether you're looking to become an information architect or hire one, keep this in mind: everyone (including the authors) is biased by their disciplinary perspective. If at all possible, try to ensure that various disciplines are represented on your web site development team to guarantee a balanced architecture.

Additionally, no matter what your perspective, the information architect ideally should be responsible solely for the site's architecture, *not* for its other aspects. It can be overly distracting to have to deal with other, more tangible aspects of the site, such as its graphic identity or programming. In that case, the site's architecture can

easily, if unintentionally, get relegated to second-class status because you'll be concentrating, naturally, on the more visible and tangible stuff.

However, in the case of smaller organizations, limited resources mean that all or most aspects of the site's development—design, editorial, technical, architecture, and production—are likely to be the responsibility of one person. Our best advice for someone in this position is obvious but still worth considering. First, find a group of friends and colleagues who are willing to be a sounding board for your ideas. Second, practice a sort of controlled schizophrenia in which you make a point to look at your site from different perspectives: first from the architect's, then from the designer's, and so on. And look for company among others who are suffering similar psychoses; consider joining the Information Architecture Institute* and attending the annual ASIS&T Information Architecture Summit.

Information Architecture Specialists

These general discussions about the role, value, and qualifications of information architects are worthwhile but incomplete. The community of information architects is experiencing what evolutionary biologists call a period of "punctuated equilibrium," marked by rapid change and specialization.

Particularly in large organizations, people who began as all-purpose information architects are gravitating towards specialized niches that match their strengths to their organization's needs. Here are just a few of the titles that already exist:

- Thesaurus Designer
- Search Schema Content Editor
- Metadata Specialist
- Content Manager
- Information Architecture Strategist
- Manager, Information Architecture
- Director, User Experience

There are so many possible variations and so many different facets. For example, information architects can specialize by:

- Industry lines (e.g., financial services, automotive)
- Functional department (e.g., human resources, engineering, marketing)
- Type of system (e.g., intranets, web sites, extranets, online magazines, digital libraries, software, online communities)
- Audience (e.g., small business owners, elementary school teachers, rocket scientists, teenagers, grandparents)

* Information Architecture Institute: *http://www.iainstitute.org.*

Finally, much IA work is centered on making large-scale applications work as advertised. So many information architects find their specializations centered on a variety of tools, most commonly:

- Content management systems
- Search engines
- Portals

As our use of networked information environments grows, the possibilities for specialization are unlimited and unpredictable. We're watching evolution in fast-forward. This is part of what makes it so much fun to be part of the information architecture community.

Practicing Information Architecture in the Real World

Users. Content. Context. You'll hear these three words again and again throughout this book. They form the basis of our model for practicing effective information architecture design. Underlying this model is a recognition that you can't design useful information architectures in a vacuum. An architect can't huddle in a dark room with a bunch of content, organize it, and emerge with a grand solution. It simply won't hold up against the light of day.

Web sites and intranets are not lifeless, static constructs. Rather, there is a dynamic, organic nature to both the information systems and the broader environments in which they exist. This is not the old world of yellowing cards in a library card catalog. We're talking complex, adaptive systems with emergent qualities. We're talking rich streams of information flowing within and beyond the borders of departments, business units, institutions, and countries. We're talking messiness and mistakes, trial and error, survival of the fittest.

We use the concept of an "information ecology"* composed of users, content, and context to address the complex dependencies that exist. And we draw upon our trusty Venn diagram (see Figure 2-2) to help people visualize and understand these relationships. The three circles illustrate the interdependent nature of users, content, and context within a complex, adaptive information ecology.

In short, we need to understand the business goals behind the web site and the resources available for design and implementation. We need to be aware of the nature and volume of content that exists today and how that might change a year from now. And we must learn about the needs and information-seeking behaviors of our major audiences. Good information architecture design is informed by all three areas.

* For more about information ecologies, read *Information Ecology* by Thomas Davenport and Lawrence Prusak (Oxford University Press, USA) and *Information Ecologies* by Bonnie Nardi and Vicki O'Day (MIT Press). Nardi and O'Day define an information ecology as "a system of people, practices, values, and technologies in a particular local environment."

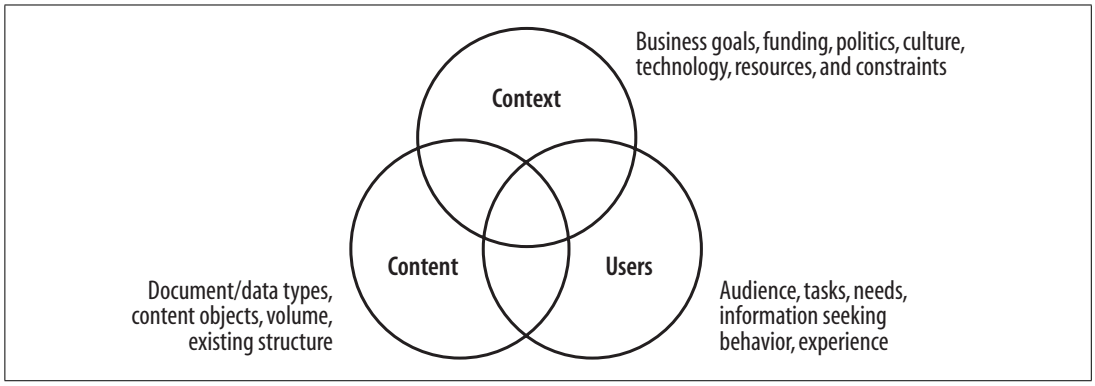


Figure 2-2. The infamous three circles of information architecture

Is this an oversimplified view of reality? Yes. Is it still useful? Absolutely. We've been using this model for over 10 years. It's held up well in all sorts of environments, from global web sites of Fortune 100 corporations to standalone intranet applications within small nonprofits. More importantly, we find these three circles incredibly helpful whenever we're confronted by a difficult question. After mouthing the trusty phrase "It depends"—as all smart information architects do—we develop our answer by deconstructing the question into three parts that coincide with our three circles. For example, when asked what are the most important qualities that an information architect should have, the answer becomes quite simple: some knowledge of users and their needs (which might come from exposure to human–computer interaction and a variety of other fields), content (think technical communication and journalism), and context (read a book on organizational psychology).

The three circles help with other tough questions, too, such as:

- What research and evaluation methods should information architects be familiar with?
- What's the ideal education for an information architect?
- What kinds of people should be part of an information architecture team?
- What kinds of books and blogs should I read to keep up with the field and its practice?
- What should go into the IA strategy that I propose to my new prospect?

The answer to each starts with a balance among the three areas: users, content, and context.

Should technology have its own circle? Maybe. But we find that technology usually gets too much attention—and it would look silly to add a fourth circle.

Incidentally, we think it's important for information architects to have a good sense of humor. Perhaps you've already figured this out. The work we do involves high levels of abstraction, ambiguity, and occasionally absurdity, and to some degree we're all still making it up as we go along. A good information architect knows how to get the work done while having some fun along the way.

If there's one thing that many years of information architecture consulting has taught us, it's that every situation is unique. We don't just mean that web sites are different from intranets or that extranets should vary by industry. We mean that, like fingerprints and snowflakes, every information ecology is unique.

The DaimlerChrysler intranet is vastly different from that of Ford or GM. Fidelity, Vanguard, Schwab, and Etrade have each created unique online financial-service experiences. Despite all the copycatting, benchmarking, and definitions of industry best practices that have surged throughout the business world in recent years, each of these information systems has emerged as quite distinctive.

That's where our model comes in handy. It's an excellent tool for learning about the specific needs and opportunities presented by a particular web site or intranet. Let's take a look at how each of our three circles contributes to the emergence of a totally unique information ecology.

Context

All web sites and intranets exist within a particular business or organizational context. Whether explicit or implicit, each organization has a mission, goals, strategy, staff, processes and procedures, physical and technology infrastructure, budget, and culture. This collective mix of capabilities, aspirations, and resources is unique to each organization.

Does it then follow that the information architecture of each organization must be unique? After all, companies buy generic office furniture. They invest in standard technology platforms. They even outsource important activities to vendors that service their competitors.

Still, the answer is a resounding yes. Information architectures must be uniquely matched to their contexts. The vocabulary and structure of your web site and your intranet is a major component of the evolving conversation between your business and your customers and employees. It influences how they think about your products and services. It tells them what to expect from you in the future. It invites or limits interaction between customers and employees. Your information architecture provides perhaps the most tangible snapshot of your organization's mission, vision, values, strategy, and culture. Do you really want that snapshot to look like that of your competitor?

As we'll explain later in more detail, the key to success is understanding and alignment. First, you need to understand the business context. What makes it unique? Where is the business today and where does it want to be tomorrow? In many cases, you're dealing with tacit knowledge. It's not written down anywhere; it's in people's heads and has never been put into words. We'll discuss a variety of methods for extracting and organizing this understanding of context. Then, you need to find ways to align the information architecture with the goals, strategy, and culture of the business. We'll discuss the approaches and tools that enable this custom configuration.

Content

We define "content" very broadly to include the documents, applications, services, schema, and metadata that people need to use or find on your site. To employ a technical term, it's the *stuff* that makes up your site. Our library backgrounds will be evident here in our bias toward textual information, and that's not such a bad thing, given the heavily textual nature of many web sites and intranets. Among other things, the Web is a wonderful communication tool, and communication is built upon words and sentences trying to convey meaning. Of course, we also recognize the Web as a tool for tasks and transactions, a flexible technology platform that supports buying and selling, calculating and configuring, sorting and simulating. But even the most task-oriented e-commerce web site has "content" that customers must be able to find.

As you survey content across a variety of sites, the following facets bubble to the surface as distinguishing factors of each information ecology.

Ownership
: Who creates and owns the content? Is ownership centralized within a content authoring group or distributed among functional departments? How much content is licensed from external information vendors? The answers to these questions play a huge role in influencing the level of control you have over all the other dimensions.

Format
: Web sites and intranets are becoming the unifying means of access to all digital formats within the organization. Oracle databases, product catalogs, Lotus Notes discussion archives, technical reports in MS Word, annual reports in PDF, office-supply purchasing applications, and video clips of the CEO are just a few of the types of documents, databases, and applications you'll find on a given site.

Structure
: All documents are not created equal. An important memo may be fewer than 100 words. A technical manual may be more than 1,000 pages. Some information systems are built around the document paradigm, with the fully integrated document as the smallest discrete unit. Other systems take a content component or digital asset approach, leveraging some form of structural markup (XML or SGML, for example) to allow management and access at a finer level of granularity.

Metadata
: To what extent has metadata that describes the content and objects within your site already been created? Have documents been tagged manually or automatically? What's the level of quality and consistency? Is there a controlled vocabulary in place? Or have users been allowed to supply their own "folksonomic" tags to the content? These factors determine how much you're starting from scratch with respect to both information retrieval and content management.

Volume
: How much content are we talking about? A hundred applications? A thousand pages? A million documents? How big is your web site?

Dynamism
: What is the rate of growth or turnover? How much new content will be added next year? And how quickly will it go stale?

All of these dimensions make for a unique mix of content and applications, which in turn suggests the need for a customized information architecture.

Users

When we worked on the first corporate web site for Borders Books & Music, back in the mid-90s before Amazon became a household name, we learned a lot about how customer research and analysis was applied towards the design and architecture of physical bookstores.

Borders had a clear understanding of how the demographics, aesthetic preferences, and purchasing behaviors of their customers differed from those of Barnes & Noble. It is no mistake that the physical layout and the selection of books differ significantly between these two stores, even within the same town. They are different by design. And that difference is built upon an understanding of their unique customer or market segments.

Differences in customer preferences and behaviors within the physical world translate into different information needs and information-seeking behaviors in the context of web sites and intranets. For example, senior executives may need to find a few good documents on a particular topic very quickly. Research analysts may need to find all the relevant documents and may be willing to spend several hours on the hunt. Managers may have a high level of industry knowledge but low navigation and searching proficiency. Teenagers may be new to the subject area but really know how to handle a search engine.

Do you know who's using your web site? Do you know how they're using it? And perhaps most importantly, do you know what information they want from your site? These are not questions you can answer in brainstorming meetings or focus groups. As our friend and fellow information architect Chris Farnum likes to say, you need to get out there in the real world and study your "users in the mist."

What Lies Ahead

So, information architecture happens. Information architectures are being created every day by generalists and specialists, by innies and outies, risk takers and people who get things done, and by people who've never heard the term "information architecture." They're being created inside all manner of information ecologies with unique combinations of users, content, and context.

Herein lies the dual challenge to the information architecture discipline. As professionals, we must advance our own understanding and our ability to perform this very difficult work inside massively complex environments. We still have so much to learn! And as a community, we must strive to advance the practice of information architecture by educating those around us who create or influence information architectures while they're focused on doing something else. We still have so much to teach!

In any case, we hope we've done a good job of setting the stage. Now it's time to delve into the guts of information architecture, so roll up your sleeves and dig in.

CHAPTER 3

User Needs and Behaviors

What we'll cover:

- The dangers of an oversimplified view of how we find information
- How our information needs vary
- How our information-seeking behaviors vary
- How and why to learn more about determining users' information needs and information-seeking behaviors

In the last two chapters, we've defined information architecture and placed it within the broader context of where, when, and by whom it's practiced. But before we jump into the actual "stuff" of information architecture—the components that make up an architecture, the methodologies that drive its design, and so on—let's first take a look at users. Information architecture is not restricted to taxonomies, search engines, and the other things that help users find information on a site. Information architecture starts with users and the reason they come to a site in the first place: they have an information need.

This is a truism, but there's more to it than meets the eye. Information needs can vary widely, and each type of information need causes users to exhibit specific information-seeking behaviors. Information architects need to understand those needs and behaviors, and their designs should correspond accordingly. There is no goal more important to designing information architecture than to satisfy users' needs.

For example, if your site is a staff directory, looking up a staff member's phone number is probably a very common information need among your site's users; in fact, this type of need may describe most of your users' finding sessions. When confronted by such a need, users will likely perform a search, and you'd be wise to make sure your information architecture supports searching by name. On the other hand, if your site helps non-savvy investors learn about and select mutual funds for investment, your users may satisfy this need through some other means. They might really benefit from a site wizard that leads them through a tutorial, or they may wish to wander by browsing through categories.

Seeking something you know is there, like your colleague's phone number, is quite a different information need than learning about a topic, like small-cap mutual funds, and your site's information architecture should be designed with those differences in mind. These needs are examples of information-seeking behaviors and, not surprisingly, searching for something you know is a very different behavior than browsing for the unknown. Distinguishing between these needs and behaviors and determining which are your users' highest priorities is an extremely valuable pursuit—it helps you determine where to invest your efforts, resources, time, and money as you design your architecture.

The "Too-Simple" Information Model

There are different models of what happens when users look for information. Modeling users' needs and behaviors forces us to ask useful questions about what kind of information the user wants, how much information is enough, and how the user actually interacts with the architecture.

Unfortunately, "too-simple" is the most common information model, and it's also the most problematic. It looks something like Figure 3-1.

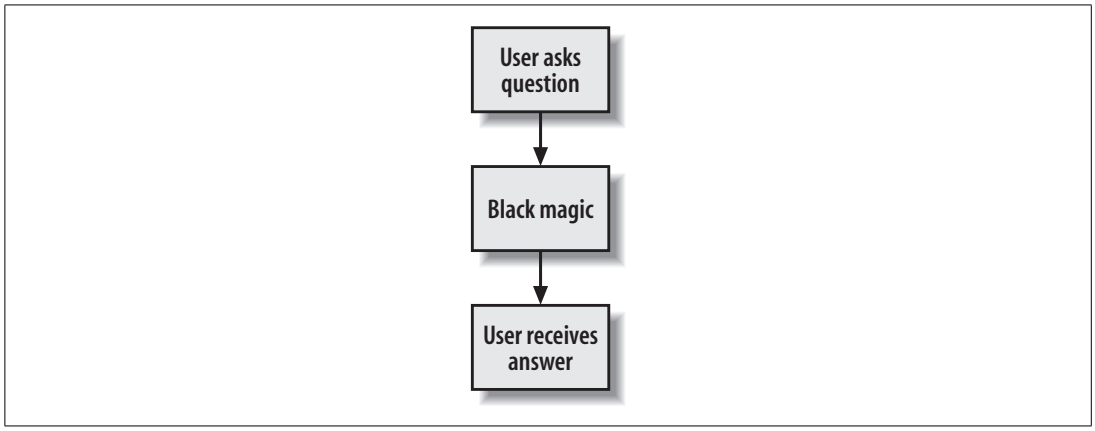


Figure 3-1. The "too-simple" model of information needs

Or, expressed as a simple algorithm:

1. User asks a question.
2. Something happens (i.e., searching or browsing).
3. User receives the answer.
4. Fin.

Input, output, end of story. This is a very mechanistic and ultimately dehumanizing model for how users find and use information on web sites. In fact, in this model, the user, like the site itself, is just another system—predictable in behavior, rational in motivation.

Why do we have a problem with this "too-simple" model? Because it rarely happens this way. There are exceptions—for example, when users know what they're looking for, as in the staff directory scenario. Here, users have a question for which there is a right answer, they know where to find the answer, they know how to state the question, and they know how to use the site to do so.

But users don't always know exactly what they want. Have you ever visited a site just to poke around? By exploring the site, you're trying to find information of a sort; you just don't exactly know what you're looking for. Even when you do, you may not have the language to express it: is it "PDA," "Palm Pilot," or "handheld computer"?

Users often complete their efforts at finding information in a state of partial satisfaction or outright frustration. Example: "I was able to find information on synchronizing my Palm Pilot, but nothing specific on syncing to a Macintosh." Or, during the process of finding, they may learn new information that changes what they're looking for altogether. Example: "I realized that a Keough retirement plan is ideal for me, even though when I started I was trying to learn about IRAs."

We also dislike the "too-simple" model because it narrowly focuses on what happens while the user is interacting with the information architecture. The information need's context—all the related stuff that happens before and after the user ever touches the keyboard—gets left out. It also assumes an ignorant user who brings little, if any, prior knowledge to the table. So the model essentially ignores any context for this scenario.

Finally, by oversimplifying, this model cedes so many great opportunities to understand what goes on in users' heads and observe the richness of what happens during their interactions with an information architecture.

This model is dangerous because it's built upon a misconception: that finding information is a straightforward problem that can be addressed by a simple, algorithmic approach. After all, we've solved the challenge of retrieving data—which, of course, is facts and figures—with database technologies such as SQL. So, the thinking goes, let's treat the abstract ideas and concepts embedded in our semi-structured textual documents the same way.

This attitude has led to the wasting of many millions of dollars on search engine software and other technological panaceas that would indeed work if this assumption were true. Many user-centered design techniques carry this misconception forward, assuming that the process of finding is simple enough to be easily measured in a quantifiable way. So we think we can measure the experience of finding by how long it takes, or how many mouse clicks it takes, or how many viewed pages it takes to find the "right" answer, when often there is no right answer.

OK, enough complaining about this model. Let's take a closer look at information needs and seeking behaviors so that we can build better models.

Information Needs

When a user comes to a web site to find something, what does she really want? In the "too-simple" model, she wants the "right answer" to her question. Indeed, right answers do come from searching databases, which store facts and figures and answer questions that really do have right answers, such as "What is the population of San Marino?" To many of us, database searching is the most familiar model of searching.

But web sites store much more than highly structured data. Not surprisingly, text is the most common type of data stored, and text itself is made up of ambiguous, messy ideas and concepts. When we go to a web site for advice on retirement investing, to learn about restaurants in Mendocino County, or to find out what's happening with the Manchester United football team, we are essentially looking for ideas and concepts that inform us and help us make decisions. The answer, if there is one, is an ambiguous moving target.

So back to the question: What do users want? Let's use the metaphor of fishing to get at the answer.

The perfect catch
: Sometimes users really are looking for the right answer. Let's think of that as fishing with a pole, hoping to hook that ideal fish. What is the population of San Marino? You go to the CIA Fact Book or some other useful site that's jam-packed with data, and you hook in that number (it's 29,251, by the way). And you're done, just as the too-simple model would have it.

Lobster trapping
: What about the times you're looking for more than just a single answer? Let's say you're hoping to find out about good bed-and-breakfast inns in Stratford, Ontario. Or you want to learn something about Lewis and Clark's journey of exploration. Or you need to get a sense of what sort of financial plans can help you save for retirement. You don't really know much about what you're looking for, and aren't ready to commit to retrieving anything more than just a few useful items, or suggestions of where to learn more. You're not hoping to hook the perfect fish, because you wouldn't know it if you caught it. Instead, you're setting out the equivalent of a lobster trap—you hope that whatever ambles in will be useful, and if it is, that's good enough. Perhaps it's a few candidate restaurants that you'll investigate further by calling and checking their availability and features. Or maybe it's a motley assemblage of Lewis and Clark stuff, ranging from book reviews to a digital version of Clark's diary to information about Lewis & Clark College in Oregon. You might be happy with a few of these items, and toss out the rest.

Indiscriminate driftnetting
: Then there are times when you want to leave no stone unturned in your search for information on a topic. You may be doing research for a doctoral thesis, or performing competitive intelligence analysis, or learning about the medical condition

affecting a close friend, or, heck, ego surfing. In these cases, you want to catch every fish in the sea, so you cast your driftnets and drag up everything you can.

I've seen you before, Moby Dick…
: There's some information that you'd prefer to never lose track of, so you'll tag it so you can find it again. Thanks to social bookmarking services like del.icio.us—which were primarily intended to support refindability—it's now possible to toss a fish back in the sea with the expectation of finding it again.

This fishing metaphor is helpful because it illustrates four common information needs. When you're hoping to make the perfect catch, you usually know what you're looking for, what to call it, and where you'll find it—this is called *known-item seeking*. An example is when you search the staff directory to find a colleague's phone number.

When you're hoping to find a few useful items in your traps, you're doing something called *exploratory seeking*. In this case, the user is not exactly sure what he's looking for. In fact, whether he realizes it or not, he is looking to learn something from the process of searching and browsing. For example, the user may go to his employer's human resources site to learn something about retirement plans that the company offers. In the process, he may encounter some basic information on IRA plans, and then change his search to learning more about such plans. As he learns more about the IRA, he shifts his search again to learning whether the simple or Roth IRA plan is best for him. Exploratory seeking is typically open-ended; there is no clear expectation of a "right" answer, nor does the user necessarily know how to articulate what exactly he is looking for. He is happy to retrieve a few good results, and use them as a springboard for the next iteration of the search. It's not always possible to definitively determine when exploratory searching is finished.

When you want everything, you're performing *exhaustive research*. The user is looking for everything on a particular topic, hoping to leave no stone unturned. In this case, the user often has many ways to express what she's looking for, and may have the patience to construct her search using all those varied terms. For example, someone who is trying to learn more about a friend's medical condition might execute multiple searches for "AIDS," "HIV," "acquired immuno-deficiency syndrome," and so forth. Again, there isn't necessarily a "right" answer. And in this case, the user must be patient enough to wade through more results than is typical with other information needs.

Finally, our failing memories and busy schedules continually force us to engage in refinding a piece of useful information that we've happened upon before. For example, while you're at work, you might surf for a few minutes and stumble on a great but long explanation of Django Reinhardt's guitar technique. Naturally, you won't read it now and risk losing your job. You'll refind it later instead. It's no surprise that del.icio.us users often assign such tags as "readme," "toread," or "readlater" to their bookmarks.

Figure 3-2 illustrates these four different types of information needs.

These four information needs are by no means the only ones, but many of your users' needs will fall into these categories.

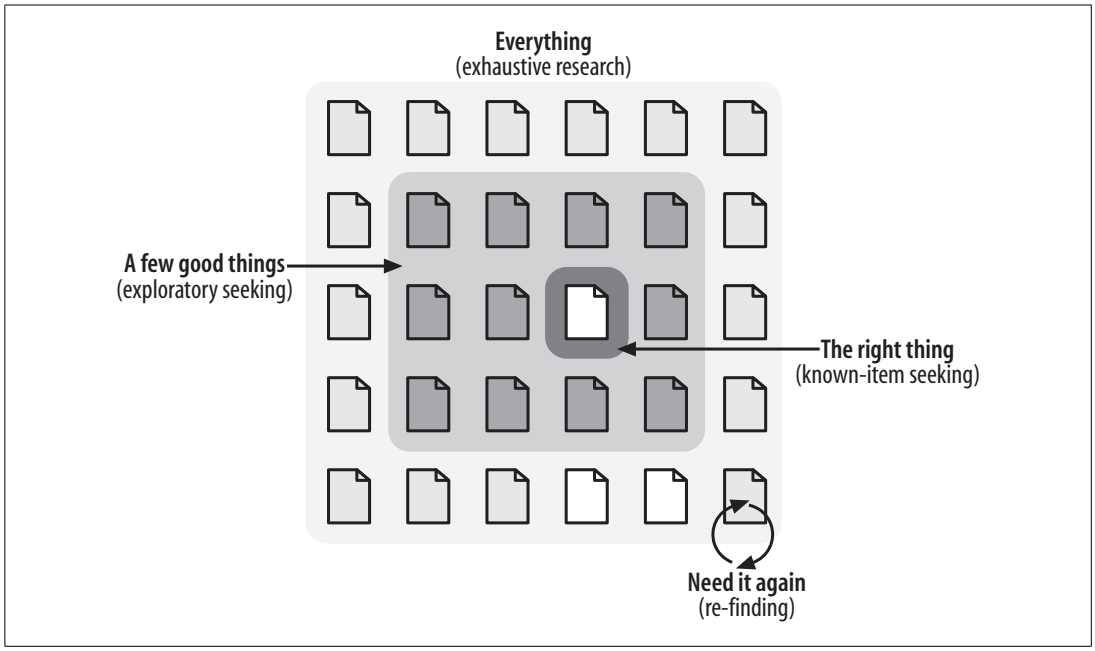


Figure 3-2. Four common information needs

Information-Seeking Behaviors

How do web site users find information? They enter queries in search systems, browse from link to link, and ask humans for help (through email, chat interfaces, and so forth). *Searching*, *browsing*, and *asking* are all methods for finding, and are the basic building blocks of information-seeking behavior.

There are two other major aspects to seeking behaviors: integration and iteration. We often integrate searching, browsing, and asking in the same finding session. Figure 3-3 shows how you might search your corporate intranet for guidelines on traveling abroad. You might first browse your way through the intranet portal to the HR site, browse the policies area, and then search for the policy that includes the string "international travel." If you still didn't get your question answered, you might send an email to Biff, the person responsible for that policy, to ask exactly what your per diem will be while spending the week in Timbuktu. Let's hope your intranet's information architecture was designed to support such integration!

Figure 3-3 also illustrates the iteration you may go through during one finding session. After all, we don't always get things right the first time. And our information needs may change along the way, causing us to try new approaches with each new iteration. So, while you may have begun with a broad quest for "guidelines on traveling abroad," you might be satisfied to find something as specific as "recommended per diem in Timbuktu" by the time you're done. Each iteration of searching, browsing, asking, and interacting with content can greatly impact what it is we're seeking.

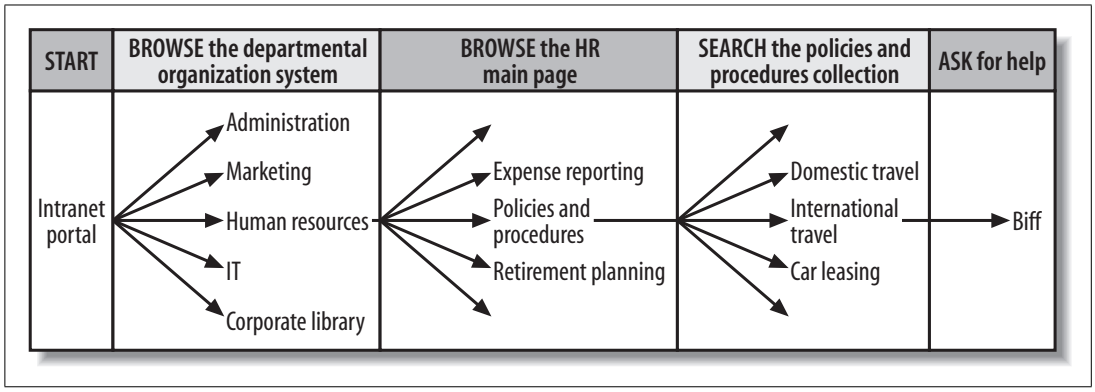


Figure 3-3. Integrated browsing, searching, and asking over many iterations

These different components of information-seeking behaviors come together in complex models, such as the "berry-picking" model* developed by Dr. Marcia Bates of the University of Southern California. In this model (shown in Figure 3-4), users start with an information need, formulate an information request (a *query*), and then move iteratively through an information system along potentially complex paths, picking bits of information ("berries") along the way. In the process, they modify their information requests as they learn more about what they need and what information is available from the system.

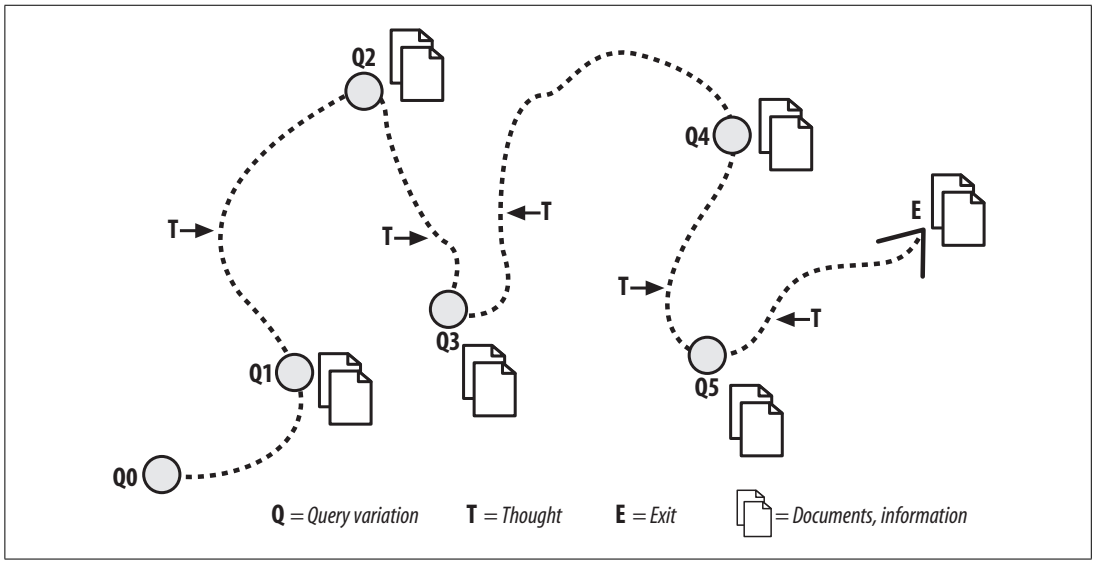


Figure 3-4. The "berry-picking" model of how users move through an information system

The berry-picking diagram looks messy, much more so than the "too-simple" model. It should; that's the way our minds often work. After all, we're not automatons.

* Bates's seminal paper, "The design of browsing and berrypicking techniques for the online search interface" (in *Online Review*, vol.13, no.5, 1989), is required reading for every information architect. See *http://www.gseis.ucla.edu/faculty/bates/berrypicking.html*.

If the berry-picking model is common to your site's users, you'll want to look for ways to support moving easily from search to browse and back again. Yahoo! provides one such integrated approach to consider: you can search within the subcategories you find through browsing, as shown in Figure 3-5. And you can browse through categories that you find by searching, as shown in Figure 3-6.

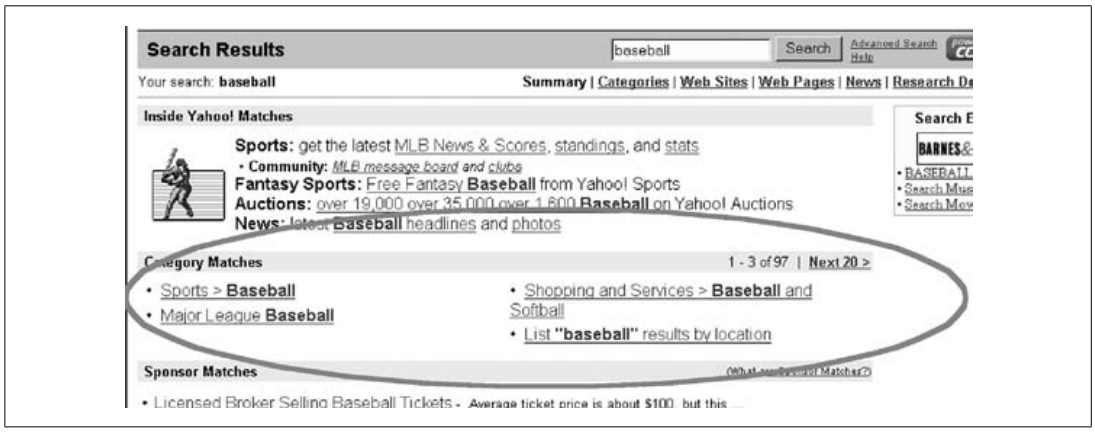


Figure 3-5. First search, then browse: searching Yahoo! for "baseball" retrieves categories that can be browsed

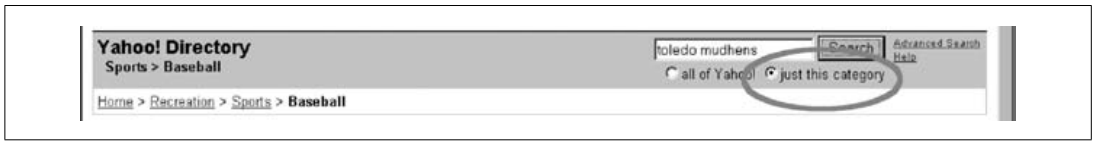


Figure 3-6. First browse, then search: Yahoo!'s categories are themselves searchable

Another useful model is the "pearl-growing" approach. Users start with one or a few good documents that are exactly what they need. They want to get "more like this one." To meet this need, Google and many other search engines allow users to do just that: Google provides a command called "Similar pages" next to each search result. A similar approach is to allow users to link from a "good" document to documents indexed with the same keywords. In sites that contain scientific papers and other documents that are heavy with citations, you can find other papers that share many of the same citations as yours or that have been co-cited with the one you like. Del.icio.us and Flickr are recent examples of sites that allow users to navigate to items that share something in common; in this case, the same user-supplied tag. All of these architectural approaches help us find "more like this one."

Corporate portals and intranets often utilize a "two-step" model. Confronted with a site consisting of links to perhaps hundreds of departmental subsites, users first need to know where to look for the information they need. They might search or browse through a directory until they find a good candidate or two, and then perform the second step: looking for information within those subsites. Their seeking behaviors may be radically different for each of these two steps; certainly, the information architectures typical of portals are usually nothing like those of departmental subsites.

Learning About Information Needs and Information-Seeking Behaviors

How does one learn about their users' information needs and seeking behaviors? There are a variety of user research methods to consider—too many to cover in detail here—so we'll recommend a pair of our favorites: search analytics and contextual inquiry. Search analytics* involves reviewing the most common search queries on your site (usually stored in your search engine's logfiles) as a way to diagnose problems with search performance, metadata, navigation, and content. Search analytics provides a sense of what users commonly seek, and can help inform your understanding of their information needs and seeking behaviors (and is handy in other ways, too, such as developing task-analysis exercises).

While search analytics is based on a high volume of real user data, it doesn't provide an opportunity to interact with users and learn more about their needs directly. Contextual inquiry,† a user research method with roots in ethnography, is a great complement to search analytics because it allows you to observe how users interact with information in their "natural" settings and, in that context, ask them why they're doing what they're doing.

Other user research methods you might look to are task analysis, surveys, and, with great care, focus groups. Ultimately, you should consider any method that might expose you to users' direct statements of their own needs, and when you can, use a combination of methods to cover as many bases as possible.

Finally, remember that, as an information architect, your goal is to do your best to learn about your users' major information needs and likely information-seeking behaviors. A better understanding of what users actually want from your site will, naturally, help you determine and prioritize which architectural components to build, which makes your job much simpler, especially considering how many ways a particular information architecture could be designed. You'll also have great user data to help counterbalance the other drivers that too often influence design, such as budget, time, politics, entrenched technologies, and designers' personal preferences.

* For more on search analytics, read the forthcoming book by Rosenfeld and Wiggins, *Search Analytics for Your Site: Conversations with Your Customers*, which will be published in 2007 (Rosenfeld Media).

† For more on contextual inquiry, read Beyer and Holtzblatt's *Contextual Design: Defining Customer-Centered Systems* (Morgan Kaufmann).

PART II

Basic Principles of Information Architecture

CHAPTER 4

The Anatomy of an Information Architecture

What we'll cover:

- Why it's important (and difficult) to make an information architecture as tangible as possible
- Examples that help you visualize an information architecture from both the top down and the bottom up
- Ways of categorizing the components of an information architecture so you can better understand and explain IA

In the preceding chapters, we discussed information architecture from a conceptual perspective. This chapter presents a more concrete view of what information architecture actually is to help you recognize it when you see it. We also introduce the components of an architecture; these are important to understand because they make up the information architect's palette. We'll cover them in greater detail in Chapters 5–9.

Visualizing Information Architecture

Why is it important to be able to visualize information architecture? There are several answers. One is that the field is new, and many people don't believe that things exist until they can see them. Another is that the field is abstract, and many who might conceptually understand the basic premise of information architecture won't really "get it" until they see it and experience it. Finally, a well-designed information architecture is invisible to users (which, paradoxically, is quite an unfair reward for IA success).

IA's lack of tangible qualities forces all information architects to be salespeople to some degree. Because it's highly probable that you'll need to explain information architecture to several important people, including colleagues, managers, prospects, clients, and perhaps your significant other, it's in your interest to be able to help them visualize what an information architecture actually is.

Let's start by looking at a site's main page. Figure 4-1 shows the main page for Gustavus Adolphus College in Saint Peter, Minnesota, USA.

Figure 4-1. Gustavus Adolphus College's main page

What's obvious here? Most immediately, you see that the site's visual design stands out. You can't help but notice the site's colors (you'll have to take our word for it), typeface choices, and images. You also notice aspects of the site's information design; for example, the number of columns—and their widths—changes throughout the page.

What else? With a careful eye, you can detect aspects of the site's interaction design, such as the use of mouseovers (over main menu choices) and pull-down menus for "Go Quickly To" and search options. Although the college's logo and logotype are prominent, the site relies on textual content (e.g., "Excellence," "Community," and so forth) to convey its message and brand. And although this particular site functions well, you'd learn something about its supporting technology (and related expertise) just from the main page—for example, if it didn't load properly in a common browser, you might guess that the designers weren't aware of or concerned with standards-compliant design.

Thus far, we've noticed all sorts of things that aren't information architecture. So what *is* recognizable as information architecture? You might be surprised by how much information architecture you can see if you know how to look. For example, the information has been structured in some basic ways, which we'll explain in later chapters:

Organization systems
: Present the site's information to us in a variety of ways, such as content categories that pertain to the entire campus (e.g., the top bar and its "Calendar" and "Academics" choices), or to specific audiences (the "I am a…" area, with such choices as "Prospective Students" and "Staff Member").

Navigation systems
: Help users move through the content, such as the "A–Z Directory" and the "Go Quickly To…" menu of popular destinations.

Search systems
: Allow users to search the content. Here, the default is set to search the Gustavus site, but one could also search the Gustavus calendar, its directory, or the whole web from the site's search interface.

Labeling systems
: Describe categories, options, and links in language that (hopefully) is meaningful to users; you'll see examples throughout the page, some (e.g., "Admission") more understandable than others ("Nobel Conference").

Figure 4-2 provides a visualization of these architectural components.

As we can see from this figure and from Figure 4-3, these areas are just the tip of the iceberg. Categories group pages and applications throughout the site; labels systematically represent the site's content; navigation systems and a search system can be used to move through the site. That's quite a lot of information architecture to cram into one screenshot!

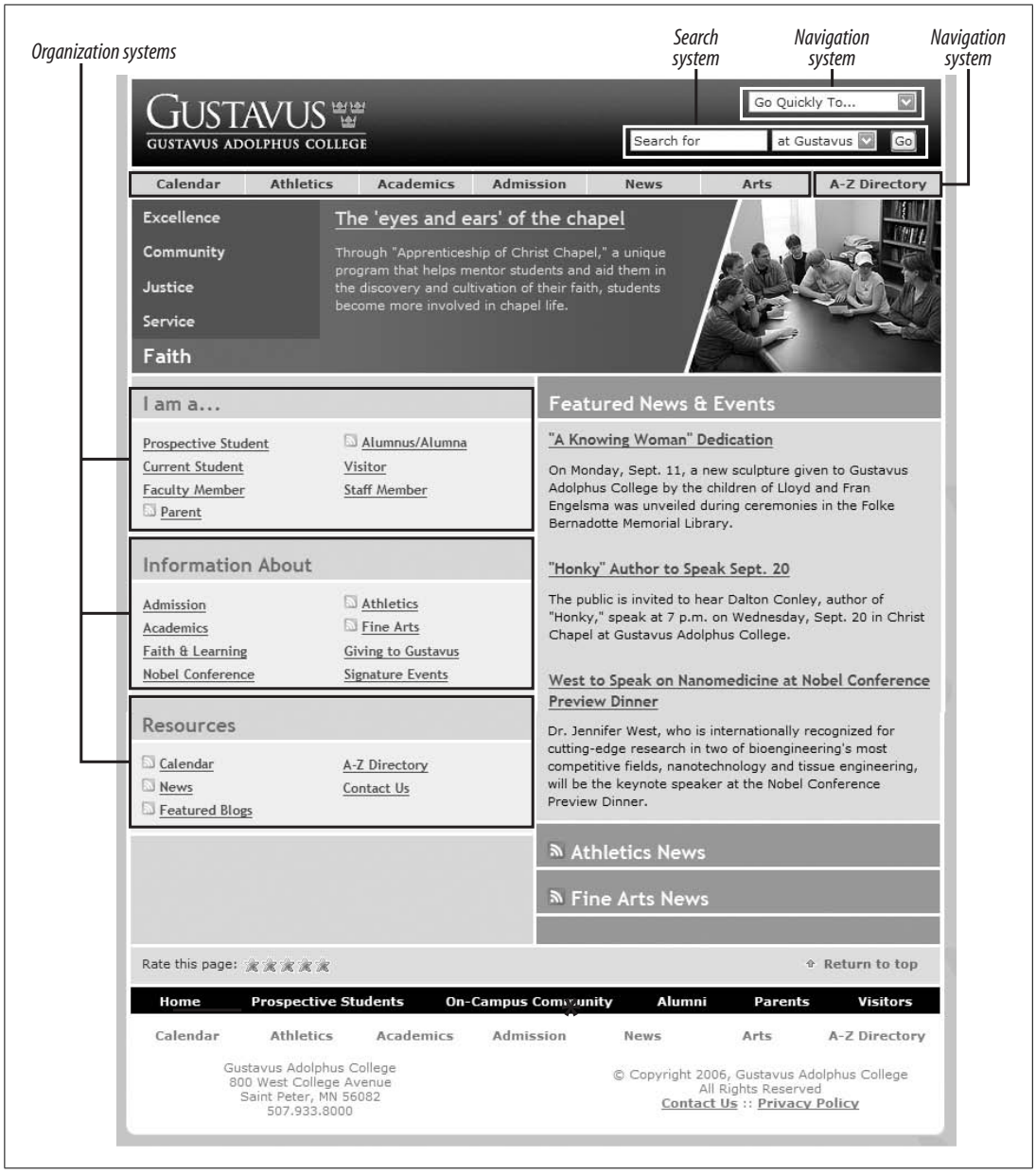


Figure 4-2. This page is crammed with architectural components

In effect, the Gustavus main page tries to anticipate users' major information needs, such as "How do I find out about admissions?" or "What's going on this week on campus?" The site's information architects have worked hard to determine the most common questions, and have designed the site to meet those needs. We refer to this as *top-down information architecture*, and the Gustavus main page addresses many common "top-down" questions that users have when they land on a site, including:

1. Where am I?
2. I know what I'm looking for; how do I search for it?

3. How do I get around this site?
4. What's important and unique about this organization?
5. What's available on this site?
6. What's happening there?
7. Do they want my opinion about their site?
8. How can I contact a human?
9. What's their address?

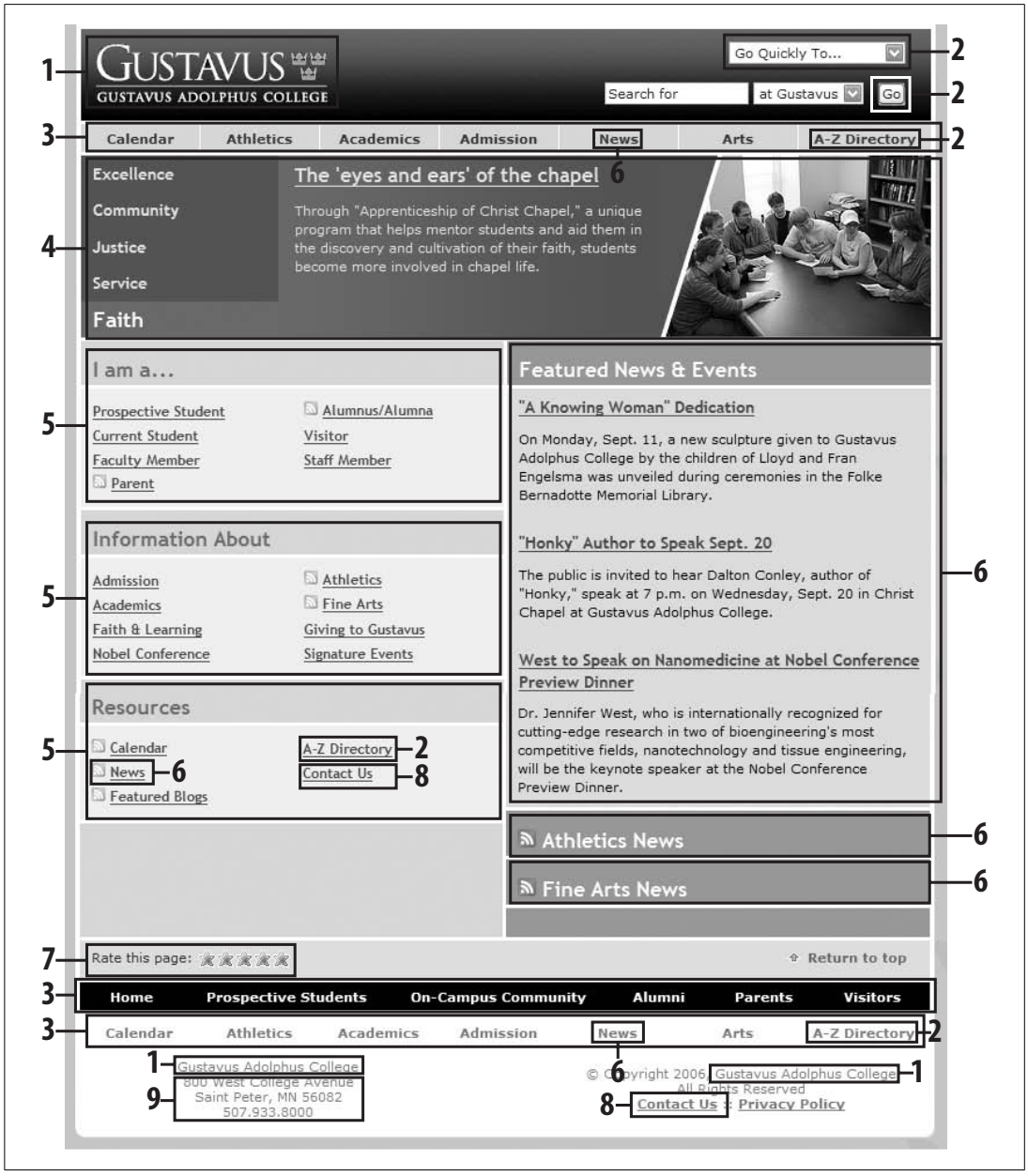


Figure 4-3. A site's main page is crammed with answers to users' questions

Figure 4-4 shows a slightly different example—pages tagged by one of the authors as relevant to enterprise user experience in del.icio.us, a social bookmarking service.

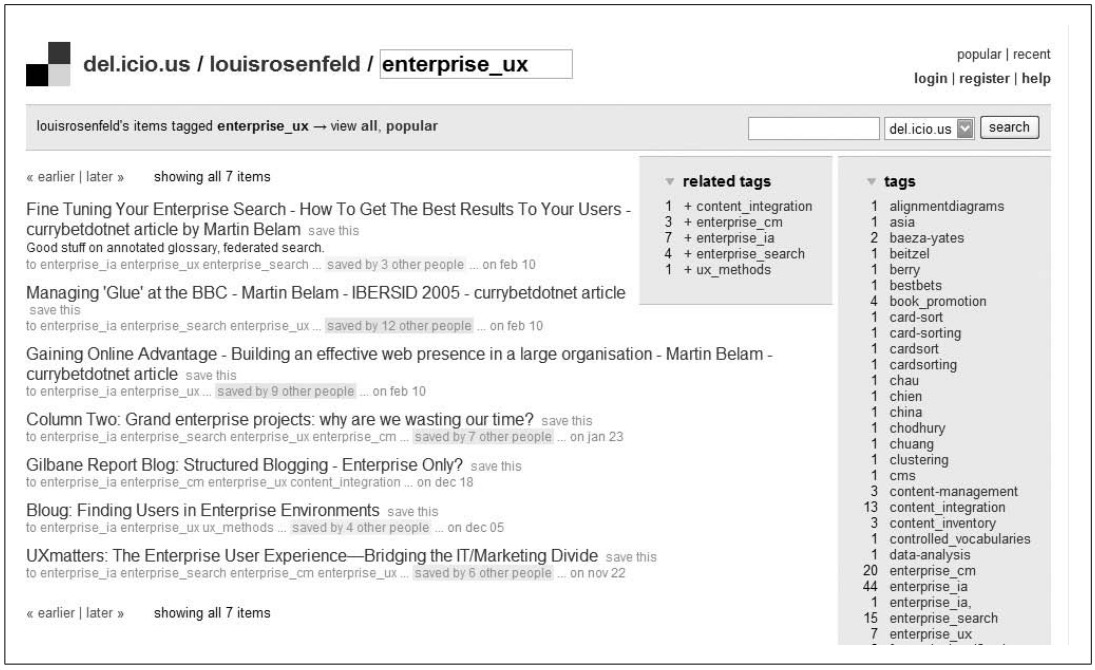


Figure 4-4. Bookmarks tagged as about "enterprise_UX" in del.icio.us, a social bookmarking service.

There is little to see here besides the information architecture and the content itself. In fact, as the content is just a collection of links to bookmarked pages from other web sites, the information architecture is the bulk of the page. It provides context for the content, and tells us what we can do while we're here.

- The information architecture tells us where we are (in del.icio.us, on a page maintained by user "louisrosenfeld" that contains bookmarks he tagged as "enterprise_ux").
- It helps us move to other, closely related pages (by, for example, scrolling through results ("<< earlier | later >>") and to pages we've bookmarked using different tags (under "tags" and "related tags").
- It helps us move through the site hierarchically (for example, we can navigate to the del.icio.us main page, or to recent or popular bookmarks) and contextually (for example, by clicking on "saved by 4 other people" or by seeing who else bookmarked pages using the same tag).
- It allows us to manipulate the content for better browsing (we can display tags in alphabetical order, as is shown, or as a "tag cloud"; a variety of other configuration choices are displayed in the "options").
- It tells us where we can go for basic services, such as logging into our account or getting help ("contact us" and "help").

In many respects, this page from del.icio.us is nothing but information architecture.

Content itself can have information architecture embedded within it. The recipe in Figure 4-5 shows a nutritious drink from the Epicurious site.

Figure 4-5. A recipe for the thirsty from Epicurious.com

Beyond the navigational options at the top of the page, there's not much information architecture here. Or is there?

The recipe itself has a clear, strong structure: a title at the top, a list of ingredients, then preparation directions and serving information. This information is "chunked" so you know what's what, even without subtitles for "ingredients" or "directions." The recipe's native chunking could also support searching and browsing; for example, users might be able to search on the chunks known as "recipe titles" for "salty dog" and retrieve this one. And these chunks are sequenced in a logical manner; after all, you'll want to know the ingredients ("Do I have four ounces of grapefruit juice?") before you start mixing the drink. The definition and sequential placement of chunks help you to recognize that this content is a recipe before you even read it. And once you know what it is, you have a better idea what this content is about and how to use it, move around it, and go somewhere else from it.

So, if you look closely enough, you can see information architecture even when it's embedded in the guts of your content. In fact, by supporting searching and browsing, the structure inherent in content enables the answers to users' questions to "rise" to the surface. This is *bottom-up information architecture*; content structure, sequencing, and tagging help you answer such questions as:

- Where am I?
- What's here?
- Where can I go from here?

Bottom-up information architecture is important because users are increasingly likely to bypass your site's top-down information architecture. Instead, they're using web-wide search tools like Google, clicking through ads, and clicking links while reading your content via their aggregators to find themselves deep in your site. Once there, they'll want to jump to other relevant content in your site without learning how to use its top-down structure. A good information architecture is designed to anticipate this type of use; Keith Instone's simple and practical Navigation Stress Test is a great way to evaluate a site's bottom-up information architecture (*http://user-experience.org/uefiles/navstress/*).

You now know that information architecture is something that can be seen, if you know what to look for. But it's important to understand that information architecture is often invisible. For example, Figure 4-6 shows some search results from the BBC's web site.

Figure 4-6. BBC's search results include three "Best Links"

What's going on here? We've searched for "chechnya," and the site has presented us with a couple of different things, most interestingly three results labeled as a "BBC Best Link." As you'd imagine, all search results were retrieved by a piece of software—a search engine—that the user never sees. The search engine has been configured to index and search certain parts of the site, to display certain kinds of information in each search result (i.e., page title, extract, and date), and to handle search queries in certain ways, such as removing "stop words" (e.g., "a," "the," and "of"). All of these decisions regarding search system configuration are unknown to users, and are integral aspects of information architecture design.

What's different is that the "Best Link" results are manually created: some people at the BBC decided that "chechnya" is an important term and that some of the BBC's best content is not news stories, which normally come up at the top of most retrieval sets. So they applied some editorial expertise to identify three highly relevant pages and associated them with the term "chechnya," thereby ensuring that these three items are displayed when someone searches for "chechnya." Users might assume these search results are automatically generated, but humans are manually modifying the information architecture in the background; this is another example of invisible information architecture.

Information architecture is much more than just blueprints that portray navigational routes and wireframes that inform visual design. Our field involves more than meets the eye, and both its visible and invisible aspects help define what we do and illustrate how challenging it really is.

Information Architecture Components

It can be difficult to know exactly what components make up an information architecture. Users interact directly with some, while (as we saw above) others are so behind the scenes that users are unaware of their existence.

In the next four chapters, we'll present and discuss information architecture components by breaking them up into the following four categories:

Organization systems
: How we categorize information, e.g., by subject or chronology. See Chapter 5.

Labeling systems
: How we represent information, e.g., scientific terminology ("Acer") or lay terminology ("maple"). See Chapter 6.

Navigation systems
: How we browse or move through information, e.g., clicking through a hierarchy. See Chapter 7.

Searching systems
: How we search information, e.g., executing a search query against an index. See Chapter 8.

Like any categorization scheme, this one has its problems. For example, it can be difficult to distinguish organization systems from labeling systems (hint: you organize content into groups, and then label those groups; each group can be labeled in different ways). In such situations, it can be useful to group objects in new ways. So before we delve into these systems, we'll present an alternative method of categorizing information architecture components. This method is comprised of browsing aids, search aids, content and tasks, and "invisible" components.

Browsing Aids

These components present users with a predetermined set of paths to help them navigate the site. Users don't articulate their queries, but instead find their way through menus and links. Types of browsing aids include:

Organization systems
: The main ways of categorizing or grouping a site's content (e.g., by topic, by task, by audiences, or by chronology). Also known as taxonomies and hierarchies. Tag clouds (based on user-generated tags) are also a form of organization system.

Site-wide navigation systems
: Primary navigation systems that help users understand where they are and where they can go within a site (e.g., breadcrumbs).

Local navigation systems
: Primary navigation systems that help users understand where they are and where they can go within a portion of a site (i.e., a subsite).

Sitemaps/Tables of contents
: Navigation systems that supplement primary navigation systems; provide a condensed overview of and links to major content areas and subsites within the site, usually in outline form.

Site indices
: Supplementary navigation systems that provide an alphabetized list of links to the contents of the site.

Site guides
: Supplementary navigation systems that provide specialized information on a specific topic, as well as links to a related subset of the site's content.

Site wizards
: Supplementary navigation systems that lead users through a sequential set of steps; may also link to a related subset of the site's content.

Contextual navigation systems
: Consistently presented links to related content. Often embedded in text, and generally used to connect highly specialized content within a site.

Search Aids

These components allow the entry of a user-defined query (e.g., a search) and automatically present users with a customized set of results that match their queries. Think of these as dynamic and mostly automated counterparts to browsing aids. Types of search components include:

Search interface
: The means of entering and revising a search query, typically with information on how to improve your query, as well as other ways to configure your search (e.g., selecting from specific search zones).

Query language
: The grammar of a search query; query languages might include Boolean operators (e.g., AND, OR, NOT), proximity operators (e.g., ADJACENT, NEAR), or ways of specifying which field to search (e.g., AUTHOR="Shakespeare").

Query builders
: Ways of enhancing a query's performance; common examples include spell checkers, stemming, concept searching, and drawing in synonyms from a thesaurus.

Retrieval algorithms
: The part of a search engine that determines which content matches a user's query; Google's PageRank is perhaps the best-known example.

Search zones
: Subsets of site content that have been separately indexed to support narrower searching (e.g., searching the tech support area within a software vendor's site).

Search results
: Presentation of content that matches the user's search query; involves decisions of what types of content should make up each individual result, how many results to display, and how sets of results should be ranked, sorted, and clustered.

Content and Tasks

These are the users' ultimate destinations, as opposed to separate components that get users to their destinations. However, it's difficult to separate content and tasks from an information architecture, as there are components embedded in content and tasks that help us find our way. Examples of information architecture components embedded in content and tasks include:

Headings
: Labels for the content that follows them.

Embedded links
: Links within text; these label (i.e., represent) the content they link to.

Embedded metadata
: Information that can be used as metadata but must first be extracted (e.g., in a recipe, if an ingredient is mentioned, this information can be indexed to support searching by ingredient).

Chunks
: Logical units of content; these can vary in granularity (e.g., sections and chapters are both chunks) and can be nested (e.g., a section is part of a book).

Lists
: Groups of chunks or links to chunks; these are important because they've been grouped together (e.g., they share some trait in common) and have been presented in a particular order (e.g., chronologically).

Sequential aids
: Clues that suggest where the user is in a process or task, and how far he has to go before completing it (e.g., "step 3 of 8").

Identifiers
: Clues that suggest where the user is in an information system (e.g., a logo specifying what site she is using, or a breadcrumb explaining where in the site she is).

"Invisible" Components

Certain key architectural components are manifest completely in the background; users rarely (if ever) interact with them. These components often "feed" other components, such as a thesaurus that's used to enhance a search query. Some examples of invisible information architecture components include:

Controlled vocabularies and thesauri
: Predetermined vocabularies of preferred terms that describe a specific domain (e.g., auto racing or orthopedic surgery); typically include variant terms (e.g., "brewskie" is a variant term for "beer").Thesauri are controlled vocabularies that generally include links to broader and narrower terms, related terms, and descriptions of preferred terms (aka "scope notes"). Search systems can enhance queries by extracting a query's synonyms from a controlled vocabulary.

Retrieval algorithms
: Used to rank search results by relevance; retrieval algorithms reflect their programmers' judgments on how to determine relevance.

Best bets
: Preferred search results that are manually coupled with a search query; editors and subject matter experts determine which queries should retrieve best bets, and which documents merit best bet status.

Whichever method you use for categorizing architectural components, it's useful to drill down beyond the abstract concept of information architecture and become familiar with its more tangible and, when possible, visual aspects. In the following chapters, we'll take an even deeper look at the nuts and bolts of an information architecture.

CHAPTER 5
Organization Systems

The beginning of all understanding is classification.
—Hayden White

What we'll cover:

- Subjectivity, politics, and other reasons why organizing information is so difficult
- Exact and ambiguous organization schemes
- Hierarchy, hypertext, and relational database structures
- Folksonomies, tagging, and social classification

Our understanding of the world is largely determined by our ability to organize information. Where do you live? What do you do? Who are you? Our answers reveal the systems of classification that form the very foundations of our understanding. We live in towns within states within countries. We work in departments in companies in industries. We are parents, children, and siblings, each an integral part of a family tree.

We organize to understand, to explain, and to control. Our classification systems inherently reflect social and political perspectives and objectives. We live in the first world. They live in the third world. She is a freedom fighter. He is a terrorist. The way we organize, label, and relate information influences the way people comprehend that information.

As information architects, we organize information so that people can find the right answers to their questions. We strive to support casual browsing and directed searching. Our aim is to design organization and labeling systems that make sense to users.

The Web provides information architects with a wonderfully flexible environment in which to organize. We can apply multiple organization systems to the same content and escape the physical limitations of the print world. So why are many large web sites so difficult to navigate? Why can't the people who design these sites make it easy to find information? These common questions focus attention on the very real challenge of organizing information.

Challenges of Organizing Information

In recent years, increasing attention has been focused on the challenge of organizing information. Yet this challenge is not new. People have struggled with the difficulties of information organization for centuries. The field of librarianship has been largely devoted to the task of organizing and providing access to information. So why all the fuss now?

Believe it or not, we're all becoming librarians. This quiet yet powerful revolution is driven by the decentralizing force of the global Internet. Not long ago, the responsibility for labeling, organizing, and providing access to information fell squarely in the laps of librarians. These librarians spoke in strange languages about Dewey Decimal Classification and the Anglo-American Cataloging Rules. They classified, cataloged, and helped you find the information you needed.

As it grows, the Internet is forcing the responsibility for organizing information on more of us each day. How many corporate web sites exist today? How many blogs? What about tomorrow? As the Internet provides users with the freedom to publish information, it quietly burdens them with the responsibility to organize that information. New information technologies open the floodgates for exponential content growth, which creates a need for innovation in content organization (see Figure 5-1).

And if you're not convinced that we're facing severe information-overload challenges, take a look at an excellent study* conducted at Berkeley. This study finds that the world produces between one and two exabytes of unique information per year. Given that an exabyte is a billion gigabytes (we're talking 18 zeros), this growing mountain of information should keep us all busy for a while.

As we struggle to meet these challenges, we unknowingly adopt the language of librarians. How should we *label* that content? Is there an existing *classification scheme* we can borrow? Who's going to *catalog* all of that information?

We're moving toward a world in which tremendous numbers of people publish and organize their own information. As we do so, the challenges inherent in organizing that information become more recognized and more important. Let's explore some of the reasons why organizing information in useful ways is so difficult.

Ambiguity

Classification systems are built upon the foundation of language, and language is ambiguous: words are capable of being understood more than one way. Think about

* "How Much Information?" is a study produced by the faculty and students at the School of Information Management and Systems at the University of California at Berkeley. See *http://www2.sims.berkeley.edu/research/projects/how-much-info-2003*.

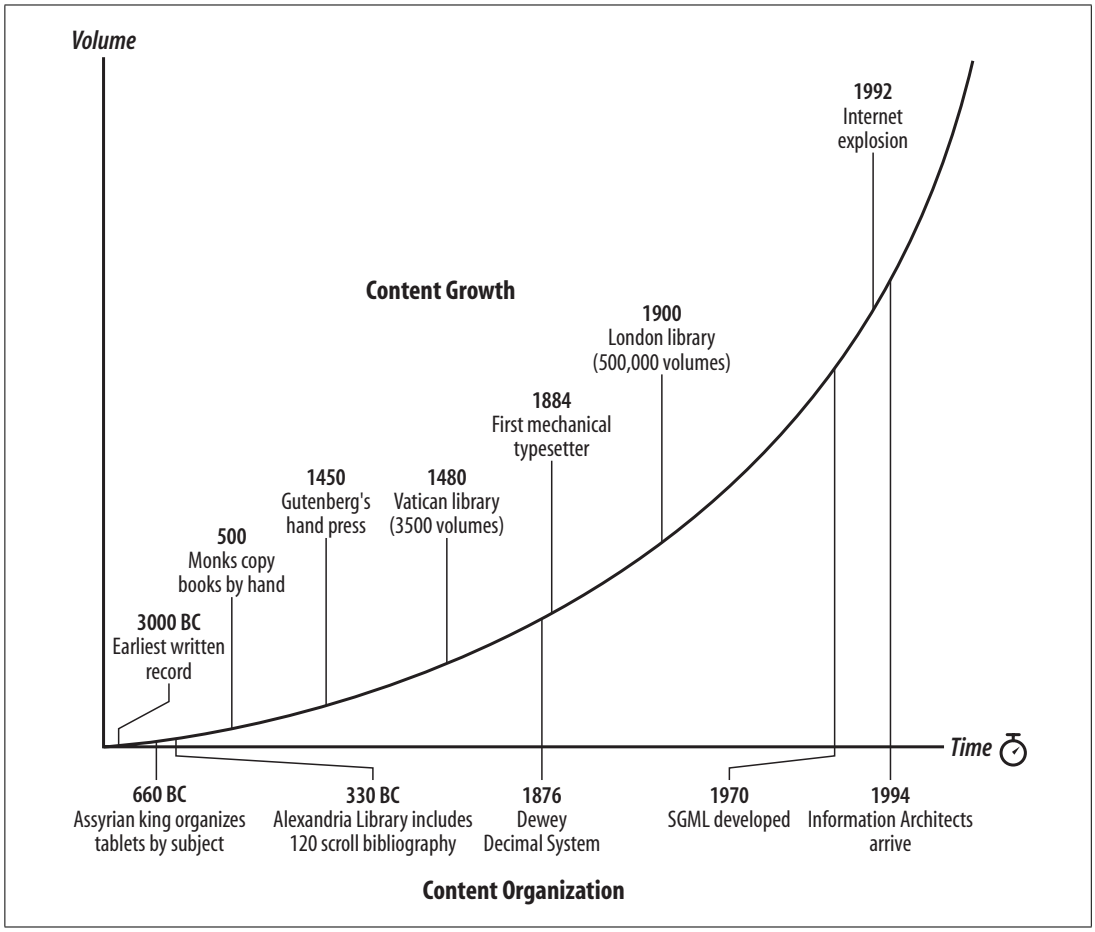


Figure 5-1. Content growth drives innovation

the word *pitch*. When I say pitch, what do you hear? There are more than 15 definitions, including:

- A throw, fling, or toss
- A black, sticky substance used for waterproofing
- The rising and falling of the bow and stern of a ship in a rough sea
- A salesman's persuasive line of talk
- An element of sound determined by the frequency of vibration

This ambiguity results in a shaky foundation for our classification systems. When we use words as labels for our categories, we run the risk that users will miss our meaning. This is a serious problem. (See Chapter 6 to learn more about labeling.)

It gets worse. Not only do we need to agree on the labels and their definitions, we also need to agree on which documents to place in which categories. Consider the common tomato. According to Webster's dictionary, a tomato is "a red or yellowish

fruit with a juicy pulp, used as a vegetable: botanically it is a berry." Now I'm confused. Is it a fruit, a vegetable, or a berry?*

If we have such problems classifying the common tomato, consider the challenges involved in classifying web site content. Classification is particularly difficult when you're organizing abstract concepts such as subjects, topics, or functions. For example, what is meant by "alternative healing," and should it be cataloged under "philosophy" or "religion" or "health and medicine" or all of the above? The organization of words and phrases, taking into account their inherent ambiguity, presents a very real and substantial challenge.

Heterogeneity

Heterogeneity refers to an object or collection of objects composed of unrelated or unlike parts. You might refer to grandma's homemade broth with its assortment of vegetables, meats, and other mysterious leftovers as heterogeneous. At the other end of the scale, homogeneous refers to something composed of similar or identical elements. For example, Ritz crackers are homogeneous. Every cracker looks and tastes the same.

An old-fashioned library card catalog is relatively homogeneous. It organizes and provides access to books. It does not provide access to chapters in books or collections of books. It may not provide access to magazines or videos. This homogeneity allows for a structured classification system. Each book has a record in the catalog. Each record contains the same fields: author, title, and subject. It is a high-level, single-medium system, and works fairly well.

Most web sites, on the other hand, are highly heterogeneous in many respects. For example, web sites often provide access to documents and their components at varying levels of *granularity*. A web site might present articles and journals and journal databases side by side. Links might lead to pages, sections of pages, or other web sites. And, web sites typically provide access to documents in multiple formats. You might find financial news, product descriptions, employee home pages, image archives, and software files. Dynamic news content shares space with static human-resources information. Textual information shares space with video, audio, and interactive applications. The web site is a great multimedia melting pot, where you are challenged to reconcile the cataloging of the broad and the detailed across many mediums.

The heterogeneous nature of web sites makes it difficult to impose any single structured organization system on the content. It usually doesn't make sense to classify

* The tomato is technically a berry and thus a fruit, despite an 1893 U.S. Supreme Court decision that declared it a vegetable. (John Nix, an importer of West Indies tomatoes, had brought suit to lift a 10 percent tariff, mandated by Congress, on imported vegetables. Nix argued that the tomato is a fruit. The Court held that since a tomato was consumed as a vegetable rather than as a dessert like fruit, it was a vegetable.) "Best Bite of Summer," by Denise Grady, July 1997.

documents at varying levels of granularity side by side. An article and a magazine should be treated differently. Similarly, it may not make sense to handle varying formats the same way. Each format will have uniquely important characteristics. For example, we need to know certain things about images, such as file format (GIF, TIFF, etc.) and resolution (640×480, 1024×768, etc.). It is difficult and often misguided to attempt a one-size-fits-all approach to the organization of heterogeneous web site content. This is a fundamental flaw of many enterprise taxonomy initiatives.

Differences in Perspectives

Have you ever tried to find a file on a coworker's desktop computer? Perhaps you had permission. Perhaps you were engaged in low-grade corporate espionage. In either case, you needed that file. In some instances, you may have found the file immediately. In others, you may have searched for hours. The ways people organize and name files and directories on their computers can be maddeningly illogical. When questioned, they will often claim that their organization system makes perfect sense. "But it's obvious! I put current proposals in the folder labeled */office/clients/green* and old proposals in */office/clients/red*. I don't understand why you couldn't find them!"*

The fact is that labeling and organization systems are intensely affected by their creators' perspectives.† We see this at the corporate level with web sites organized according to internal divisions or org charts, with groupings such as *marketing*, *sales*, *customer support*, *human resources*, and *information systems*. How does a customer visiting this web site know where to go for technical information about a product she just purchased? To design usable organization systems, we need to escape from our own mental models of content labeling and organization.

We employ a mix of user research and analysis methods to gain real insight. How do users group the information? What types of labels do they use? How do they navigate? This challenge is complicated by the fact that web sites are designed for multiple users, and all users will have different ways of understanding the information. Their levels of familiarity with your company and your content will vary. For these reasons, even with a massive barrage of user tests, it is impossible to create a perfect organization system. One site does not fit all! However, by recognizing the importance of perspective, by striving to understand the intended audiences through user research and testing, and by providing multiple navigation pathways, you can do a better job of organizing information for public consumption than your coworker does on his desktop computer.

* It actually gets even more complicated because an individual's needs, perspectives, and behaviors change over time. A significant body of research within the field of library and information science explores the complex nature of information models. For an example, see "Anomalous States of Knowledge as a Basis for Information Retrieval" by N.J. Belkin, *Canadian Journal of Information Science,* 5 (1980).

† For a fascinating study on the idiosyncratic methods people use to organize their physical desktops and office spaces, see "How Do People Organize Their Desks? Implications for the Design of Office Information Systems" by T.W. Malone, *ACM Transactions on Office Information Systems 1* (1983).

Internal Politics

Politics exist in every organization. Individuals and departments constantly position for influence or respect. Because of the inherent power of information organization in forming understanding and opinion, the process of designing information architectures for web sites and intranets can involve a strong undercurrent of politics. The choice of organization and labeling systems can have a big impact on how users of the site perceive the company, its departments, and its products. For example, should we include a link to the library site on the main page of the corporate intranet? Should we call it *The Library* or *Information Services* or *Knowledge Management*? Should information resources provided by other departments be included in this area? If the library gets a link on the main page, why not corporate communications? What about daily news?

As an information architect, you must be sensitive to your organization's political environment. In certain cases, you must remind your colleagues to focus on creating an architecture that works for the user. In others, you may need to make compromises to avoid serious political conflict. Politics raise the complexity and difficulty of creating usable information architectures. However, if you are sensitive to the political issues at hand, you can manage their impact upon the architecture.

Organizing Web Sites and Intranets

The organization of information in web sites and intranets is a major factor in determining success, and yet many web development teams lack the understanding necessary to do the job well. Our goal in this chapter is to provide a foundation for tackling even the most challenging information organization projects.

Organization systems are composed of *organization schemes* and *organization structures*. An organization scheme defines the shared characteristics of content items and influences the logical grouping of those items. An organization structure defines the types of relationships between content items and groups.

Before diving in, it's important to understand information organization in the context of web site development. Organization is closely related to navigation, labeling, and indexing. The hierarchical organization structures of web sites often play the part of primary navigation system. The labels of categories play a significant role in defining the contents of those categories. Manual indexing or *metadata tagging* is ultimately a tool for organizing content items into groups at a very detailed level. Despite these closely knit relationships, it is both possible and useful to isolate the design of organization systems, which will form the foundation for navigation and labeling systems. By focusing solely on the logical grouping of information, you avoid the distractions of implementation details and can design a better web site.

Organization Schemes

We navigate through organization schemes every day. Telephone books, supermarkets, and television programming guides all use organization schemes to facilitate access. Some schemes are easy to use. We rarely have difficulty finding a friend's phone number in the alphabetical organization scheme of the white pages. Some schemes are intensely frustrating. Trying to find marshmallows or popcorn in a large and unfamiliar supermarket can drive us crazy. Are marshmallows in the snack aisle, the baking ingredients section, both, or neither?

In fact, the organization schemes of the phone book and the supermarket are fundamentally different. The alphabetical organization scheme of the phone book's white pages is exact. The hybrid topical/task-oriented organization scheme of the supermarket is ambiguous.

Exact Organization Schemes

Let's start with the easy ones. Exact or "objective" organization schemes divide information into well-defined and mutually exclusive sections. The alphabetical organization of the phone book's white pages is a perfect example. If you know the last name of the person you are looking for, navigating the scheme is easy. "Porter" is in the Ps, which are after the Os but before the Qs. This is called *known-item* searching. You know what you're looking for, and it's obvious where to find it. No ambiguity is involved. The problem with exact organization schemes is that they require users to know the specific name of the resource they are looking for. The white pages don't work very well if you're looking for a plumber.

Exact organization schemes are relatively easy to design and maintain because there is little intellectual work involved in assigning items to categories. They are also easy to use. The following sections explore three frequently used exact organization schemes.

Alphabetical

An alphabetical organization scheme is the primary organization scheme for encyclopedias and dictionaries. Almost all nonfiction books, including this one, provide an alphabetical index. Phone books, department-store directories, bookstores, and libraries all make use of our 26-letter alphabet for organizing their contents. Alphabetical organization often serves as an umbrella for other organization schemes. We see information organized alphabetically by last name, by product or service, by department, and by format. Figure 5-2 provides an example of a departmental directory organized alphabetically by last name.

Figure 5-2. A directory of people at Microsoft Research

Chronological

Certain types of information lend themselves to chronological organization. For example, an archive of press releases might be organized by the date of release. Press release archives are obvious candidates for chronological organization schemes (see Figure 5-3). The date of announcement provides important context for the release. However, keep in mind that users may also want to browse the releases by title, product category, or geography, or to search by keyword. A complementary combination of organization schemes is often necessary. History books, magazine archives, diaries, and television guides tend to be organized chronologically. As long as there is agreement on when a particular event occurred, chronological schemes are easy to design and use.

Geographical

Place is often an important characteristic of information. We travel from one place to another. We care about the news and weather that affects us in our location. Political, social, and economic issues are frequently location-dependent. And, in a world where mobile devices such as Blackberries and Treos are becoming location-aware, while companies like Google and Yahoo! are investing heavily in local search and directory services, the map as interface is enjoying a resurgence of interest.

With the exception of border disputes, geographical organization schemes are fairly straightforward to design and use. Figure 5-4 shows an example of a geographical organization scheme. Users can select a location from the map using their mouse.

Figure 5-3. Press releases in reverse chronological order

Ambiguous Organization Schemes

Now for the tough ones. Ambiguous or "subjective" organization schemes divide information into categories that defy exact definition. They are mired in the ambiguity of language and organization, not to mention human subjectivity. They are difficult to design and maintain. They can be difficult to use. Remember the tomato? Do we classify it under fruit, berry, or vegetable?

However, they are often more important and useful than exact organization schemes. Consider the typical library catalog. There are three primary organization schemes: you can search for books by author, by title, or by subject. The author and title organization schemes are exact and thereby easier to create, maintain, and use. However, extensive research shows that library patrons use ambiguous subject-based schemes such as the Dewey Decimal and Library of Congress classification systems much more frequently.

There's a simple reason why people find ambiguous organization schemes so useful: we don't always know what we're looking for. In some cases, you simply don't know the correct label. In others, you may have only a vague information need that you can't quite articulate. For these reasons, information seeking is often iterative and interactive. What you find at the beginning of your search may influence what you

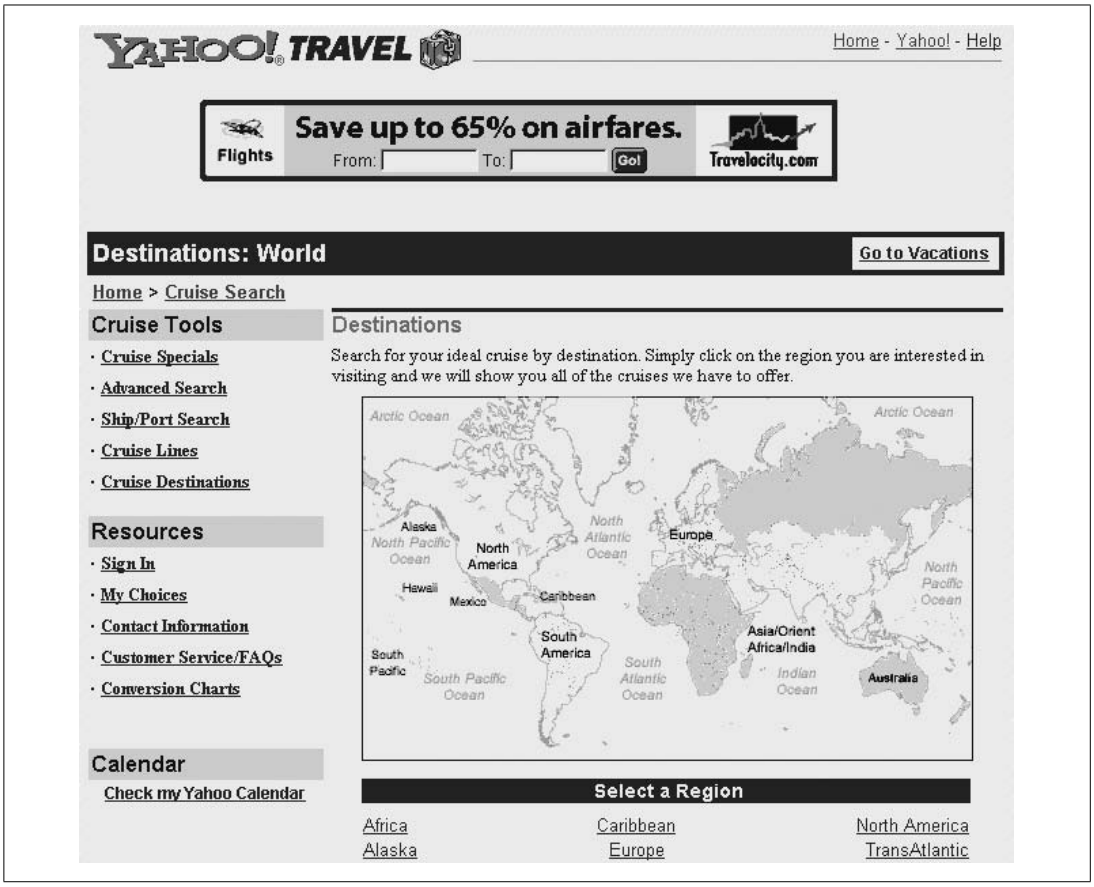


Figure 5-4. A geographical organization scheme

look for and find later in your search. This information-seeking process can involve a wonderful element of associative learning. Seek and ye shall find, but if the system is well designed, you also might learn along the way. This is web surfing at its best.

Ambiguous organization supports this serendipitous mode of information seeking by grouping items in intellectually meaningful ways. In an alphabetical scheme, closely grouped items may have nothing in common beyond the fact that their names begin with the same letter. In an ambiguous organization scheme, someone other than the user has made an intellectual decision to group items together. This grouping of related items supports an associative learning process that may enable the user to make new connections and reach better conclusions. While ambiguous organization schemes require more work and introduce a messy element of subjectivity, they often prove more valuable to the user than exact schemes.

The success of ambiguous organization schemes depends upon the quality of the scheme and the careful placement of individual items within that scheme. Rigorous user testing is essential. In most situations, there is an ongoing need for classifying new items and for modifying the organization scheme to reflect changes in the industry. Maintaining these schemes may require dedicated staff with subject matter expertise. Let's review a few of the most common and valuable ambiguous organization schemes.

Topic

Organizing information by subject or topic is one of the most useful and challenging approaches. Phone book yellow pages are organized topically, so that's the place to look when you need a plumber. Academic courses and departments, newspapers, and the chapters of most nonfiction books are all organized along topical lines.

While few web sites are organized solely by topic, most should provide some sort of topical access to content. In designing a topical organization scheme, it is important to define the breadth of coverage. Some schemes, such as those found in an encyclopedia, cover the entire breadth of human knowledge. Research-oriented web sites such as Consumer Reports (shown in Figure 5-5) rely heavily on their topical organization scheme. Others, such as corporate web sites, are limited in breadth, covering only those topics directly related to that company's products and services. In designing a topical organization scheme, keep in mind that you are defining the universe of content (both present and future) that users will expect to find within that area of the web site.

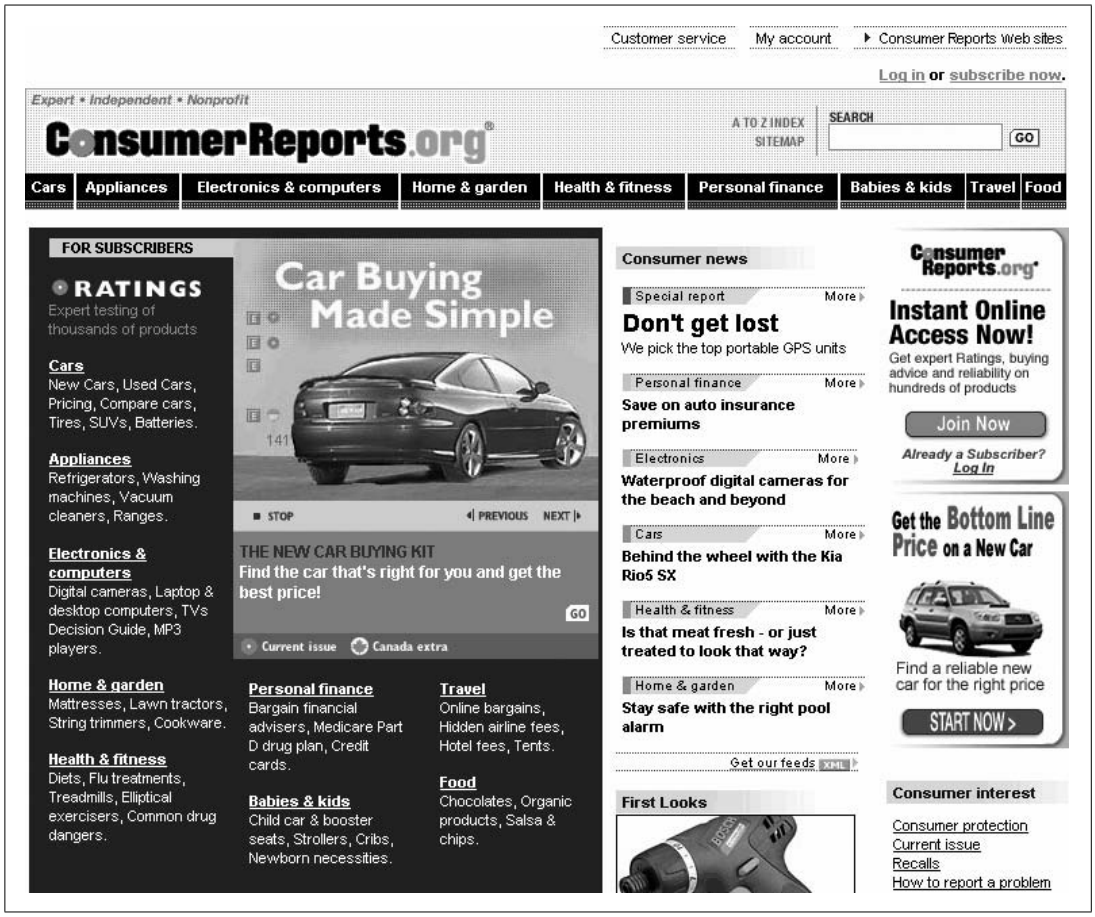


Figure 5-5. This topical taxonomy shows categories and subcategories

Task

Task-oriented schemes organize content and applications into a collection of processes, functions, or tasks. These schemes are appropriate when it's possible to anticipate a limited number of high-priority tasks that users will want to perform. Desktop software applications such as word processors and spreadsheets provide familiar examples. Collections of individual actions are organized under task-oriented menus such as *Edit*, *Insert*, and *Format*.

On the Web, task-oriented organization schemes are most common in the context of e-commerce web sites where customer interaction takes center stage. Intranets and extranets also lend themselves well to a task orientation, since they tend to integrate powerful applications or "e-services" as well as content. You will rarely find a web site organized solely by task. Instead, task-oriented schemes are usually embedded within specific subsites or integrated into hybrid task/topic navigation systems, as we see in Figure 5-6.

Figure 5-6. Task and topic coexist on the eBay home page

Audience

In cases where there are two or more clearly definable audiences for a web site or intranet, an audience-specific organization scheme may make sense. This type of scheme works best when the site is frequented by repeat visitors who can bookmark their particular section of the site. It also works well if there is value in customizing the content for each audience. Audience-oriented schemes break a site into smaller, audience-specific mini-sites, thereby allowing for clutter-free pages that present only the options of interest to that particular audience. The main page of dell.com, shown in Figure 5-7, presents an audience-oriented organization scheme (on the right) that invites customers to self-identify.

Figure 5-7. Dell invites users to self-identify

Organizing by audience brings all the promise and peril associated with any form of personalization. For example, Dell understands a great deal about its audience segments and brings this knowledge to bear on its web site. If I visit the site and identify myself as a member of the "Home & Home Office" audience, Dell will present me with a set of options and sample system configurations designed to meet my needs. In this instance, Dell might make the educated guess that I probably need a modem to connect to the Internet from my home. However, this guess would be wrong,

since I now have affordable broadband access in my community. I need an Ethernet card instead. All ambiguous schemes require the information architect to make these educated guesses and revisit them over time.

Audience-specific schemes can be open or closed. An open scheme will allow members of one audience to access the content intended for other audiences. A closed scheme will prevent members from moving between audience-specific sections. This may be appropriate if subscription fees or security issues are involved.

Metaphor

Metaphors are commonly used to help users understand the new by relating it to the familiar. You need not look further than your *desktop* computer with its *folders*, *files*, and *trash can* or *recycle bin* for an example. Applied to an interface in this way, metaphors can help users understand content and function intuitively. In addition, the process of exploring possible metaphor-driven organization schemes can generate new and exciting ideas about the design, organization, and function of the web site.

While metaphor exploration can be useful while brainstorming, you should use caution when considering a metaphor-driven global organization scheme. First, metaphors, if they are to succeed, must be familiar to users. Organizing the web site of a computer-hardware vendor according to the internal architecture of a computer will not help users who don't understand the layout of a motherboard.

Second, metaphors can introduce unwanted baggage or be limiting. For example, users might expect a digital library to be staffed by a librarian that will answer reference questions. Most digital libraries do not provide this service. Additionally, you may wish to provide services in your digital library that have no clear corollary in the real world. Creating your own customized version of the library is one such example. This will force you to break out of the metaphor, introducing inconsistency into your organization scheme.

In the Teletubbies example in Figure 5-8, the games area is organized according to the metaphor of a physical place, populated by creatures and objects. This colorful approach invites exploration, and children quickly learn that they must go "inside Home Hill" to play with the machine called "Nu Nu." Since most of the target audience can't read, an overarching visual metaphor is a great solution. But unless your web site is aimed at young children, metaphor should probably play only a niche role.

Hybrids

The power of a pure organization scheme derives from its ability to suggest a simple mental model that users can quickly understand. Users easily recognize an audience-specific or topical organization. And fairly small, pure organization schemes can be applied to large amounts of content without sacrificing their integrity or diminishing their usability.

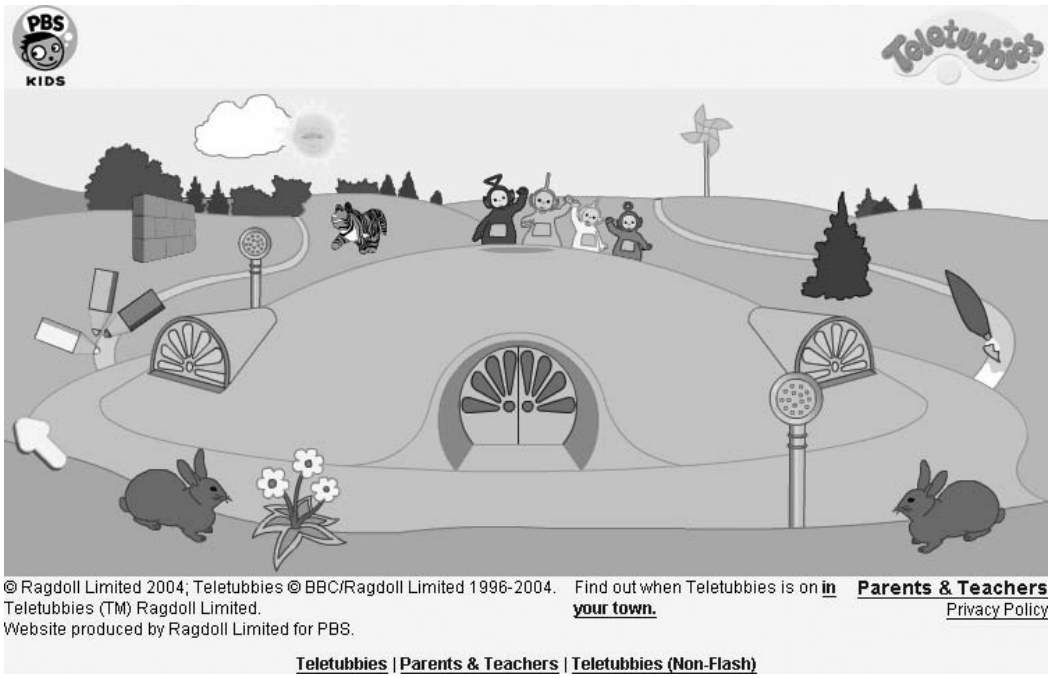


Figure 5-8. The Teletubbies' metaphor-driven games area

However, when you start blending elements of multiple schemes, confusion often follows, and solutions are rarely scalable. Consider the example in Figure 5-9. This hybrid scheme includes elements of audience-specific, topical, metaphor-based, task-oriented, and alphabetical organization schemes. Because they are all mixed together, we can't form a mental model. Instead, we need to skim through each menu item to find the option we're looking for.

The Mixed-Up Library

Adult	*audience-oriented*
Arts and Humanities	*topical*
Community Center	*metaphor-based*
Get a Library Card	*functional*
Learn About Our Library	*functional*
Science	*topical*
Social Science	*topical*
Teen	*audience-oriented*
Youth	*audience-oriented*

Figure 5-9. A hybrid organization scheme

The exception to these cautions against hybrid schemes exists within the surface layer of navigation. As illustrated by eBay (see Figure 5-6), many web sites successfully combine topics and tasks on their main page and within their global navigation. This reflects the reality that both the organization and its users typically identify

finding content and completing key tasks at the top of their priority lists. Because this includes only the highest-priority tasks, the solution does not need to be scalable. It's only when such schemes are used to organize a large volume of content and tasks that the problems arise. In other words, shallow hybrid schemes are fine, but deep hybrid schemes are not.

Unfortunately, deep hybrid schemes are still fairly common. This is because it is often difficult to agree upon any one scheme, so people throw the elements of multiple schemes together in a confusing mix. There is a better alternative. In cases where multiple schemes must be presented on one page, you should communicate to designers the importance of preserving the integrity of each scheme. As long as the schemes are presented separately on the page, they will retain the powerful ability to suggest a mental model for users. For example, a look at the Stanford University home page in Figure 5-10 reveals a topical scheme, an audience-oriented scheme, an alphabetical index, and a search function. By presenting them separately, Stanford provides flexibility without causing confusion.

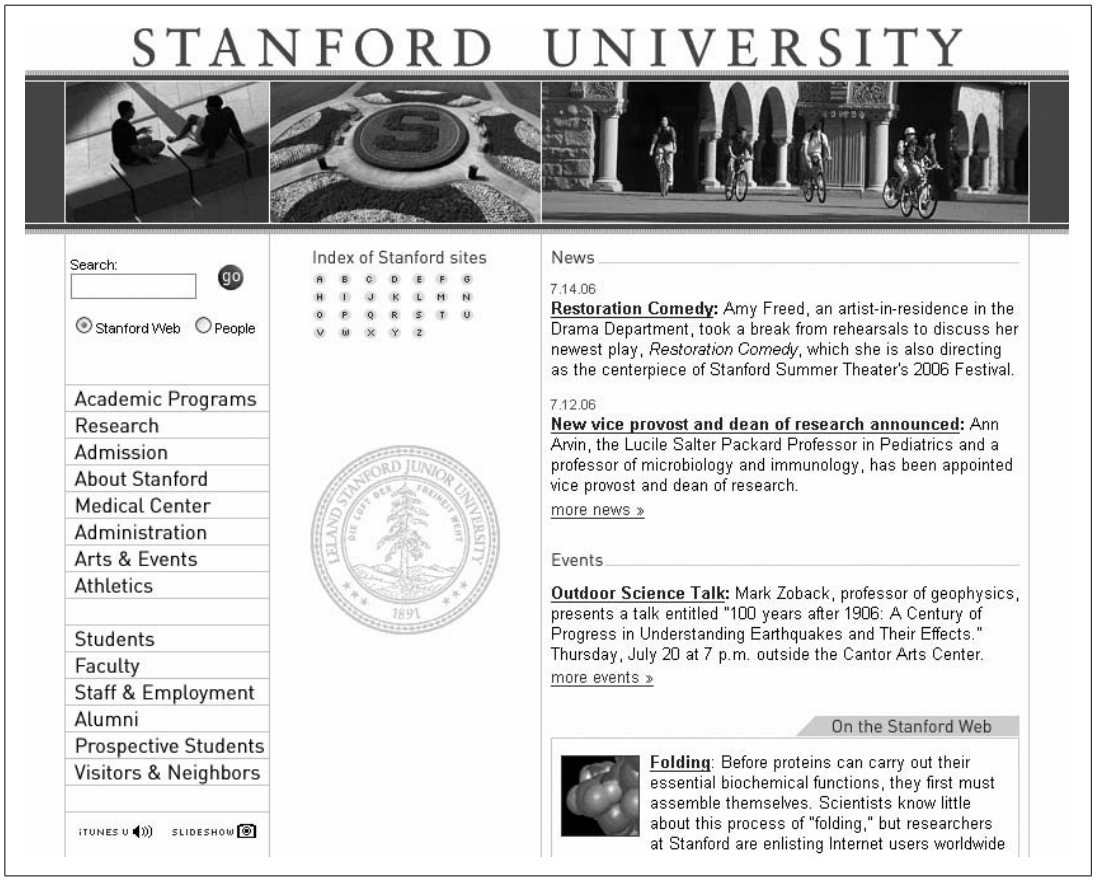


Figure 5-10. Stanford provides multiple organization schemes

Organization Structures

Organization structure plays an intangible yet very important role in the design of web sites. Although we interact with organization structures every day, we rarely think about them. Movies are linear in their physical structure. We experience them frame by frame from beginning to end. However, the plots themselves may be nonlinear, employing flashbacks and parallel subplots. Maps have a spatial structure. Items are placed according to physical proximity, although the most useful maps cheat, sacrificing accuracy for clarity.

The structure of information defines the primary ways in which users can navigate. Major organization structures that apply to web site and intranet architectures include the hierarchy, the database-oriented model, and hypertext. Each organization structure possesses unique strengths and weaknesses. In some cases, it makes sense to use one or the other. In many cases, it makes sense to use all three in a complementary manner.

The Hierarchy: A Top-Down Approach

The foundation of almost all good information architectures is a well-designed hierarchy or *taxonomy*.* In this hypertextual world of nets and webs, such a statement may seem blasphemous, but it's true. The mutually exclusive subdivisions and parent–child relationships of hierarchies are simple and familiar. We have organized information into hierarchies since the beginning of time. Family trees are hierarchical. Our division of life on earth into kingdoms, classes, and species is hierarchical. Organization charts are usually hierarchical. We divide books into chapters into sections into paragraphs into sentences into words into letters. Hierarchy is ubiquitous in our lives and informs our understanding of the world in a profound and meaningful way. Because of this pervasiveness of hierarchy, users can easily and quickly understand web sites that use hierarchical organization models. They are able to develop a mental model of the site's structure and their location within that structure. This provides context that helps users feel comfortable. Figure 5-11 shows an example of a simple hierarchical model.

Because hierarchies provide a simple and familiar way to organize information, they are usually a good place to start the information architecture process. The top-down approach allows you to quickly get a handle on the scope of the web site without going through an extensive content-inventory process. You can begin identifying the major content areas and exploring possible organization schemes that will provide access to that content.

* In recent years, the business world has fallen in love with the term "taxonomies." Many biologists and librarians are frustrated with the exploding abuse of this term. We use it specifically to refer to a hierarchical arrangement of categories within the user interface of a web site or intranet. If you can't beat them, join them.

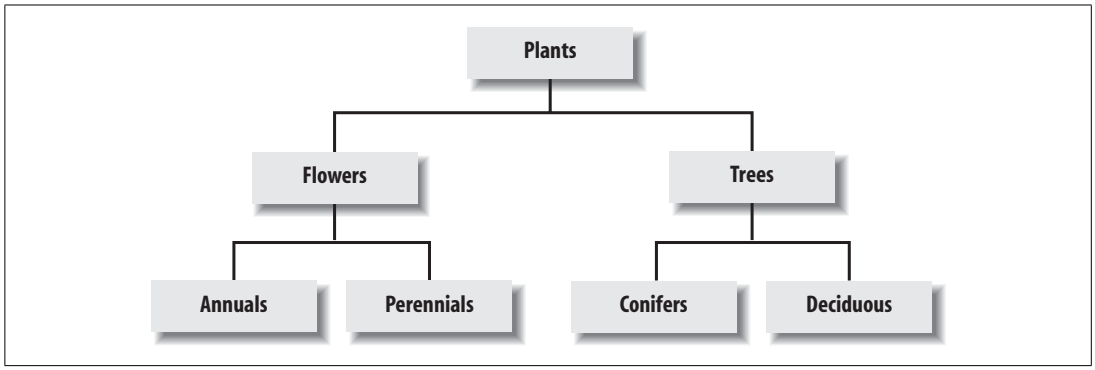


Figure 5-11. A simple hierarchical model

Designing taxonomies

When designing taxonomies on the Web, you should remember a few rules of thumb. First, you should be aware of, but not bound by, the idea that hierarchical categories should be mutually exclusive. Within a single organization scheme, you will need to balance the tension between exclusivity and inclusivity. Taxonomies that allow cross-listing are known as *polyhierarchical*. Ambiguous organization schemes in particular make it challenging to divide content into mutually exclusive categories. Do tomatoes belong in the fruit, vegetable, or berry category? In many cases, you might place the more ambiguous items into two or more categories so that users are sure to find them. However, if too many items are cross-listed, the hierarchy loses its value. This tension between exclusivity and inclusivity does not exist across different organization schemes. You would expect a listing of products organized by format to include the same items as a companion listing of products organized by topic. Topic and format are simply two different ways of looking at the same information. Or to use a technical term, they're two independent *facets*. See Chapter 9 for more about metadata, facets, and polyhierarchy.

Second, it is important to consider the balance between breadth and depth in your taxonomy. Breadth refers to the number of options at each level of the hierarchy. Depth refers to the number of levels in the hierarchy. If a hierarchy is too narrow and deep, users have to click through an inordinate number of levels to find what they are looking for. The top of Figure 5-12 illustrates a narrow-and-deep hierarchy in which users are faced with six clicks to reach the deepest content. In the (relatively) broad-and-shallow hierarchy, users must choose from 10 categories to reach 10 content items. If a hierarchy is too broad and shallow, as shown in the bottom part of Figure 5-12, users are faced with too many options on the main menu and are unpleasantly surprised by the lack of content once they select an option.

When considering breadth, you should be sensitive to people's visual scanning abilities and to the cognitive limits of the human mind. Now, we're not going to tell you

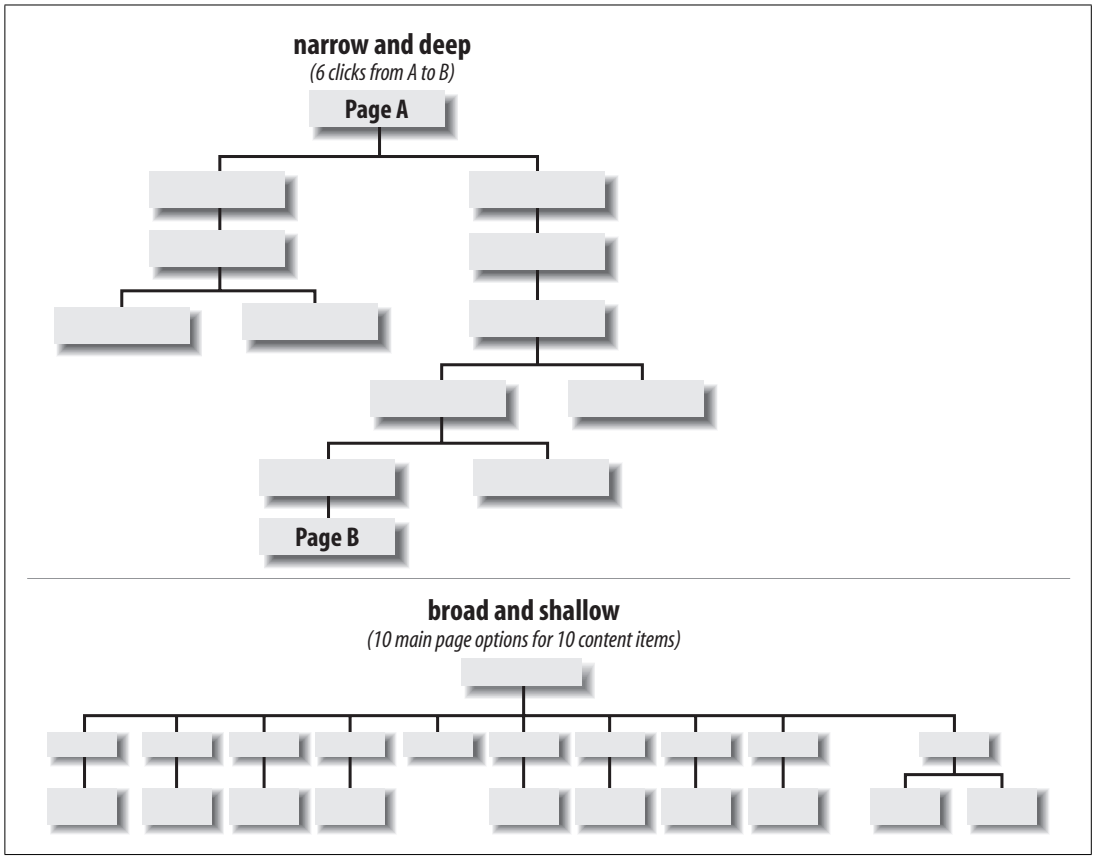


Figure 5-12. Balancing depth and breadth

to follow the infamous seven plus-or-minus two rule.* There is general consensus that the number of links you can safely include is constrained by users' abilities to visually scan the page rather than by their short-term memories.

Instead, we suggest that you:

- Recognize the danger of overloading users with too many options.
- Group and structure information at the page level.
- Subject your designs to rigorous user testing.

Consider the National Cancer Institute's award-winning main page, shown in Figure 5-13. It's one of the government's most visited (and tested) pages on the Web, and the portal into a large information system. Presenting information hierarchically at the page level, as NCI has done, can make a major positive impact on usability.

* G. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review* 63, no. 2 (1956).

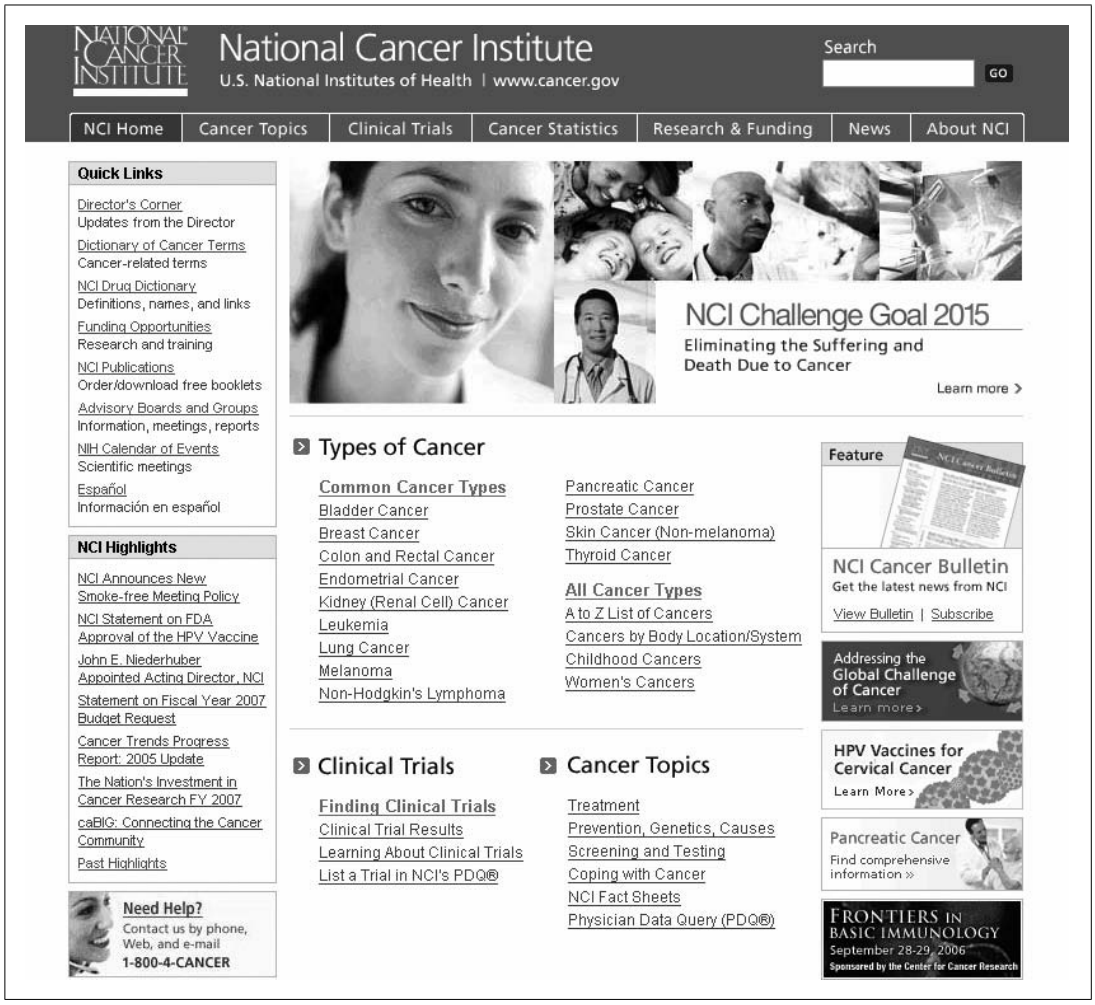


Figure 5-13. The National Cancer Institute groups items within the page

There are roughly 75 links on NCI's main page, and they're organized into several key groupings:

Group	Notes
Global Navigation	Global navigation (e.g., Cancer Topics, Clinical Trials, Cancer Statistics) has 7 links plus Search.
Types of Cancer	Includes 13 Common Cancer Types and 4 alternate ways to explore All Cancer Types.
Clinical Trials	Includes 4 links.
Cancer Topics	Includes 6 links.
Quick Links	Includes 8 links.
NCI Highlights	There are 7 headlines plus a link to the archive.
Features	On the right, there are 5 feature tiles.
Footer Navigation	Includes 11 links.

These 75 links are subdivided into eight discrete categories, with a limited number of links per category.

In contrast to breadth, when considering depth, you should be even more conservative. If users are forced to click through more than two or three levels, they may simply give up and leave your web site. At the very least, they'll become frustrated.

An excellent study conducted by Microsoft Research suggests that a medium balance of breadth and depth may provide the best results.*

For new web sites and intranets that are expected to grow, you should lean toward a broad-and-shallow rather than a narrow-and-deep hierarchy. This allows for the addition of content without major restructuring. It is less problematic to add items to secondary levels of the hierarchy than to the main page for a couple of reasons. First, the main page serves as the most prominent and important navigation interface for users. Changes to this page can really hurt the mental model users have formed of the web site over time. Second, because of the main page's prominence and importance, companies tend to spend lots of care (and money) on its graphic design and layout. Changes to the main page can be more time consuming and expensive than changes to secondary pages.

Finally, when designing organization structures, you should not become trapped by the hierarchical model. Certain content areas will invite a database or hypertext-based approach. The hierarchy is a good place to begin, but it is only one component in a cohesive organization system.

The Database Model: A Bottom-Up Approach

A database is defined as "a collection of data arranged for ease and speed of search and retrieval." A Rolodex provides a simple example of a flat-file database (see Figure 5-14). Each card represents an individual contact and constitutes a *record*. Each record contains several *fields*, such as name, address, and telephone number. Each field may contain data specific to that contact. The collection of records is a database.

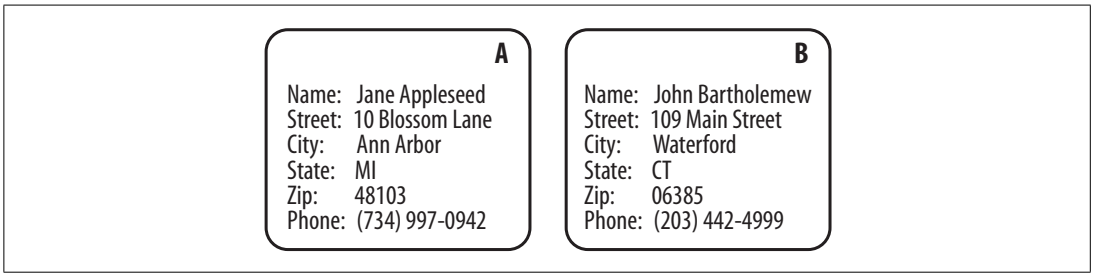


Figure 5-14. The printed card Rolodex is a simple database

* "Web Page Design: Implications of Memory, Structure and Scent for Information Retrieval," by Kevin Larson and Mary Czerwinski, Microsoft Research. See *http://research.microsoft.com/users/marycz/p25-larson.pdf*.

In an old-fashioned Rolodex, users are limited to searching for a particular individual by last name. In a more contemporary, computer-based contact-management system, we can also search and sort using other fields. For example, we can ask for a list of all contacts who live in Connecticut, sorted alphabetically by city.

Most of the heavy-duty databases we use are built upon the relational database model. In relational database structures, data is stored within a set of relations or tables. Rows in the tables represent records, and columns represent fields. Data in different tables may be linked through a series of keys. For example, in Figure 5-15, the au_id and title_id fields within the Author_Title table act as keys linking the data stored separately in the Author and Title tables.

So why are database structures important to information architects? After all, we made a fuss earlier in the book about our focus on information access rather than data retrieval. Where is this discussion heading?

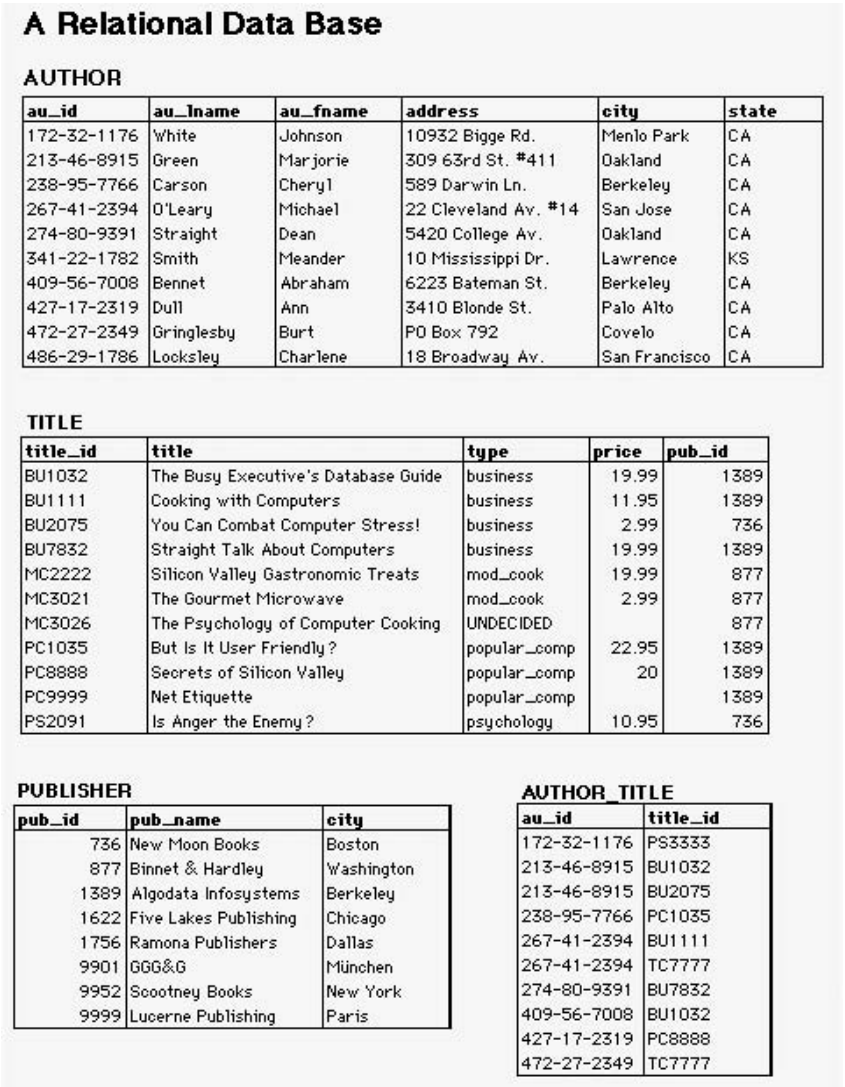
In a word, metadata. Metadata is the primary key that links information architecture to the design of database schema. It allows us to apply the structure and power of relational databases to the heterogeneous, unstructured environments of web sites and intranets. By tagging documents and other information objects with controlled vocabulary metadata, we enable powerful searching, browsing, filtering, and dynamic linking. (We'll discuss metadata and controlled vocabularies in more detail in Chapter 9.)

The relationships between metadata elements can become quite complex. Defining and mapping these formal relationships requires significant skill and technical understanding. For example, the entity relationship diagram (ERD) in Figure 5-16 illustrates a structured approach to defining a metadata schema. Each entity (e.g., Resource) has attributes (e.g., Name, URL). These entities and attributes become records and fields. The ERD is used to visualize and refine the data model before design and population of the database.

We're not suggesting that all information architects must become experts in SQL, XML schema definition, the creation of entity relationship diagrams, and the design of relational databases—though these are all extremely valuable skills. In many cases, you'll be better off working with a professional programmer or database designer who really knows how to do this stuff. And for large web sites, you will hopefully be able to rely on Content Management System (CMS) software to manage your metadata and controlled vocabularies.

Instead, information architects need to understand how metadata, controlled vocabularies, and database structures can be used to enable:

- Automatic generation of alphabetical indexes (e.g., product index)
- Dynamic presentation of associative "see also" links
- Fielded searching
- Advanced filtering and sorting of search results



A Relational Data Base

AUTHOR

au_id	au_lname	au_fname	address	city	state
172-32-1176	White	Johnson	10932 Bigge Rd.	Menlo Park	CA
213-46-8915	Green	Marjorie	309 63rd St. #411	Oakland	CA
238-95-7766	Carson	Cheryl	589 Darwin Ln.	Berkeley	CA
267-41-2394	O'Leary	Michael	22 Cleveland Av. #14	San Jose	CA
274-80-9391	Straight	Dean	5420 College Av.	Oakland	CA
341-22-1782	Smith	Meander	10 Mississippi Dr.	Lawrence	KS
409-56-7008	Bennet	Abraham	6223 Bateman St.	Berkeley	CA
427-17-2319	Dull	Ann	3410 Blonde St.	Palo Alto	CA
472-27-2349	Gringlesby	Burt	PO Box 792	Covelo	CA
486-29-1786	Locksley	Charlene	18 Broadway Av.	San Francisco	CA

TITLE

title_id	title	type	price	pub_id
BU1032	The Busy Executive's Database Guide	business	19.99	1389
BU1111	Cooking with Computers	business	11.95	1389
BU2075	You Can Combat Computer Stress!	business	2.99	736
BU7832	Straight Talk About Computers	business	19.99	1389
MC2222	Silicon Valley Gastronomic Treats	mod_cook	19.99	877
MC3021	The Gourmet Microwave	mod_cook	2.99	877
MC3026	The Psychology of Computer Cooking	UNDECIDED		877
PC1035	But Is It User Friendly?	popular_comp	22.95	1389
PC8888	Secrets of Silicon Valley	popular_comp	20	1389
PC9999	Net Etiquette	popular_comp		1389
PS2091	Is Anger the Enemy?	psychology	10.95	736

PUBLISHER

pub_id	pub_name	city
736	New Moon Books	Boston
877	Binnet & Hardley	Washington
1389	Algodata Infosystems	Berkeley
1622	Five Lakes Publishing	Chicago
1756	Ramona Publishers	Dallas
9901	GGG&G	München
9952	Scootney Books	New York
9999	Lucerne Publishing	Paris

AUTHOR_TITLE

au_id	title_id
172-32-1176	PS3333
213-46-8915	BU1032
213-46-8915	BU2075
238-95-7766	PC1035
267-41-2394	BU1111
267-41-2394	TC7777
274-80-9391	BU7832
409-56-7008	BU1032
427-17-2319	PC8888
472-27-2349	TC7777

Figure 5-15. A relational database schema (this example is drawn from an overview of the relational database model at the University of Texas at Austin)

The database model is particularly useful when applied within relatively homogeneous subsites such as product catalogs and staff directories. However, enterprise controlled vocabularies can often provide a thin horizontal layer of structure across the full breadth of a site. Deeper vertical vocabularies can then be created for particular departments, subjects, or audiences.

Hypertext

Hypertext is a relatively recent and highly nonlinear way of structuring information. A hypertext system involves two primary types of components: the items or chunks of information that will be linked, and the links between those chunks.

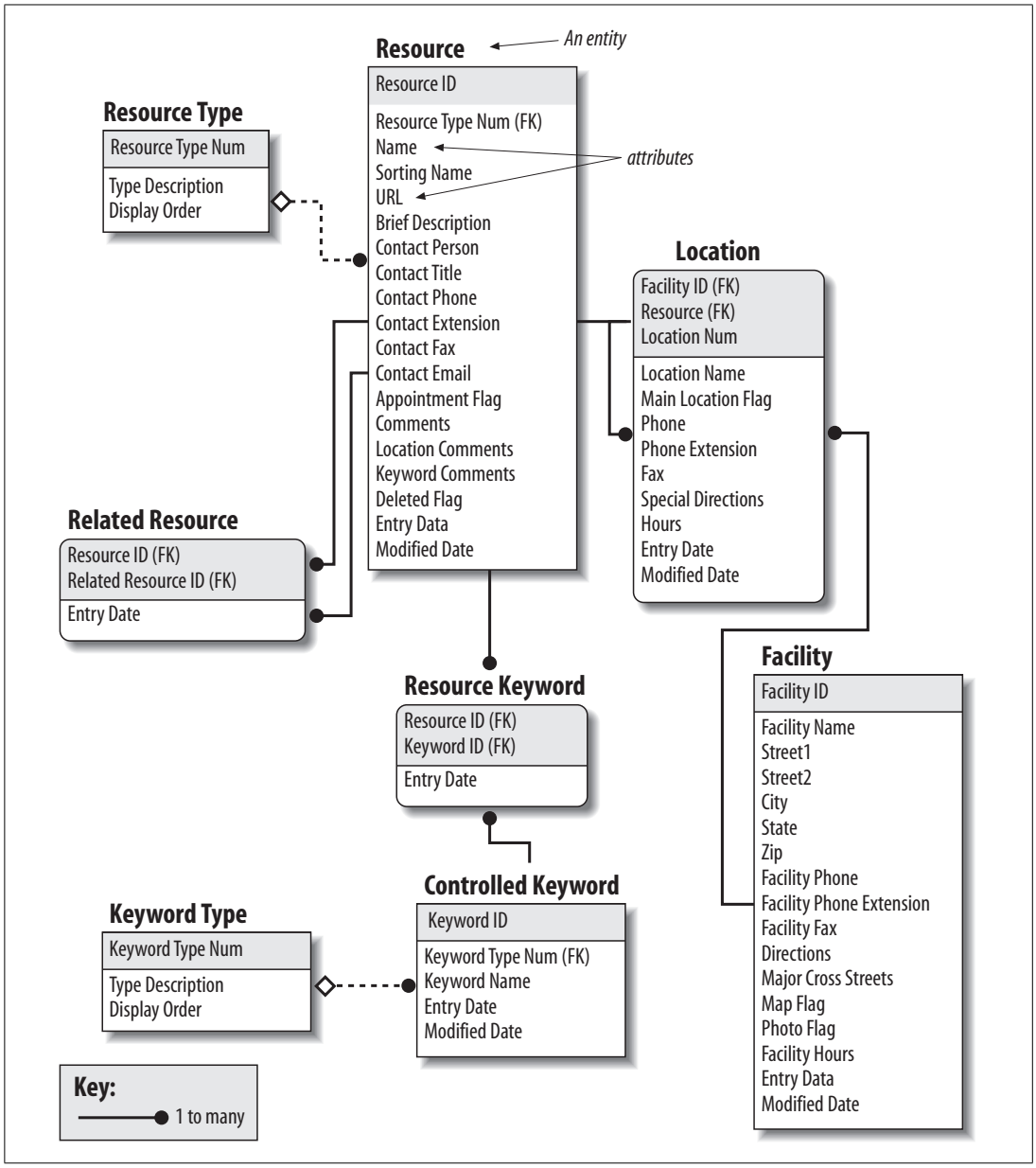


Figure 5-16. An entity relationship diagram showing a structured approach to defining a metadata schema (courtesy of InterConnect of Ann Arbor)

These components can form hypermedia systems that connect text, data, image, video, and audio chunks. Hypertext chunks can be connected hierarchically, non-hierarchically, or both, as shown in Figure 5-17. In hypertext systems, content chunks are connected via links in a loose web of relationships.

Although this organization structure provides you with great flexibility, it presents substantial potential for complexity and user confusion. Why? Because hypertext links reflect highly personal associations. As users navigate through highly hypertextual web

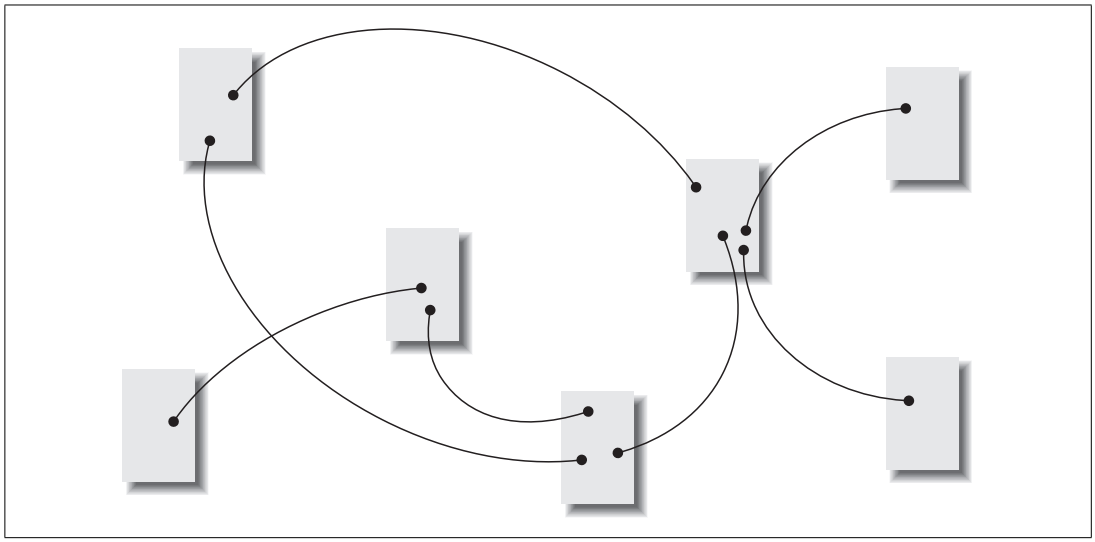


Figure 5-17. A network of hypertextual connections

sites, it is easy for them to get lost. It's as if they are thrown into a forest and are bouncing from tree to tree, trying to understand the lay of the land. They simply can't create a mental model of the site organization. Without context, users can quickly become overwhelmed and frustrated. In addition, hypertextual links are often personal in nature. The relationships that one person sees between content items may not be apparent to others.

For these reasons, hypertext is rarely a good candidate for the primary organization structure. Rather, it can be used to complement structures based upon the hierarchical or database models.

Hypertext allows for useful and creative relationships between items and areas in the hierarchy. It usually makes sense to first design the information hierarchy and then identify ways in which hypertext can complement the hierarchy.

Social Classification

In recent years, user participatory systems have captured the attention and imagination of many in the web design community. High profile successes such as Flickr and del.icio.us have demonstrated the potential to enlist users in content creation and classification, and they've sparked tremendous enthusiasm for tagging as a form of description and organization.

Free tagging, also known as collaborative categorization, mob indexing, and ethnoclassification, is a simple yet powerful tool. Users tag objects with one or more keywords. The tags are public and serve as pivots for social navigation. Users can move fluidly between objects, authors, tags, and indexers. And when large numbers of people get involved, interesting opportunities arise to transform user behavior and tagging patterns into new organization and navigation systems.

For instance, in Figure 5-18, we see that the "IxDG Resource Library" is the most frequently bookmarked site that's been tagged with *interactiondesign*, and we can easily explore related tags such as *design*, *patterns*, *ia*, and *ui*. No single person or centralized team created a taxonomy to define these relationships. Rather, they emerged (and continue to emerge) through the tagging efforts of many individuals.

Figure 5-18. Popular items on del.icio.us

Similarly, Flickr has developed clustering algorithms (Figure 5-19) that group photos with overlapping tag sets, thereby creating emergent, self-describing taxonomies.

From an information architect's perspective, these experiments in the co-creation of structure and organization are fascinating. And, as you might expect, we can't resist labeling the new phenomenon. For instance, on an information architecture mailing list, Gene Smith described the growing use of user-defined tags to organize information, and asked, "Is there a name for this kind of informal social classification?" After a brief discussion, Thomas Vander Wal replied:

> So the user-created bottom-up categorical structure development with an emergent thesaurus would become a Folksonomy?*

* Posted on the members mailing list of the Information Architecture Institute on July 24, 2004.

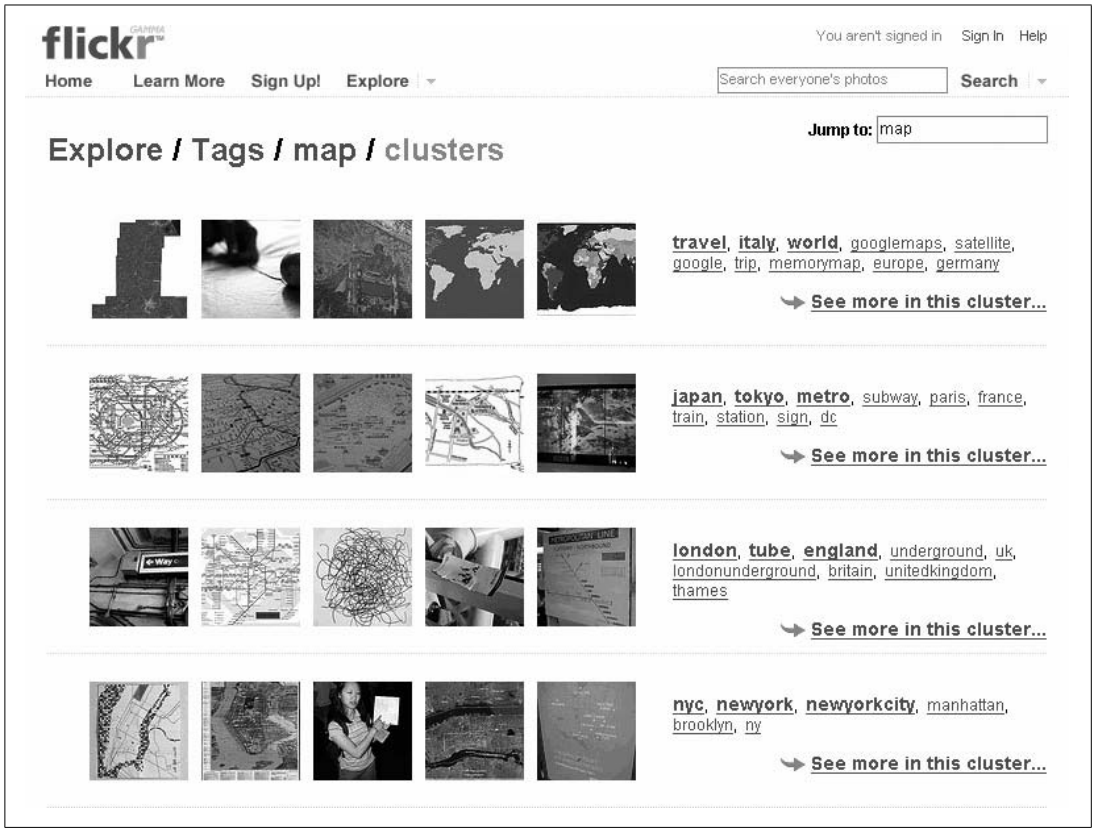


Figure 5-19. Clustering on Flickr

Of course, the tagging revolution hasn't come without a cost. In their enthusiasm for the new, many overzealous pundits have forecast the demise of traditional forms of organization. For example, David Sifry, founder and CEO of Technorati, stated:

> Tags are a simple, yet powerful, social software innovation. Today millions of people are freely and openly assigning metadata to content and conversations. Unlike rigid taxonomy schemes that people dislike, the ease of tagging for personal organization with social incentives leads to a rich and discoverable folksonomy. Intelligence is provided by real people from the bottom-up to aid social discovery. And with the right tag search and navigation, folksonomy outperforms more structured approaches to classification.*

And, in a debate with Lou Rosenfeld, Clay Shirky argued:

> The advantage of folksonomies isn't that they're better than controlled vocabularies, it's that they're better than nothing, because controlled vocabularies are not extensible to the majority of cases where tagging is needed…This is something the 'well-designed metadata' crowd has never understood…the cost of tagging large systems rigorously is crippling, so fantasies of using controlled metadata in environments like Flickr are really fantasies of users suddenly deciding to become disciples of information architecture.†

* This excerpt is from "Technorati Launches Tags," a January 17, 2005 post on the blog of David Sifry, founder and CEO of Technorati, the self-described "authority on what's going on in the world of weblogs."

† From the blog posting "folksonomies + controlled vocabularies" (*http://www.corante.com/many/archives/2005/01/07/folksonomies_controlled_vocabularies.php*).

These colorful statements play well in the blogosphere, but they are neither fair nor accurate. First, there's simply no evidence to suggest that folksonomy outperforms traditional approaches to organization, and anyone who searches Flickr can see firsthand the findability problems that come with the complete absence of vocabulary control. Second, these arguments ignore the critical importance of context. To date, tagging has flourished in a very limited set of environments. This is why the same examples, Flickr and del.icio.us, are used over and over again. It remains to be seen whether social classification can be successfully integrated into a wider range of web sites, intranets, and interactive products.

Hopefully, information architects will embrace this challenge, and play a leadership role in the synthesis of traditional and novel ways of organizing. In many contexts, we will continue to structure and organize information on behalf of our users. In others, we will design environments and tools that enlist our users in folksonomic acts of co-creation. And on some projects, we'll have the opportunity to bridge the gap, using both tags and taxonomies to connect users with the content they seek.

Creating Cohesive Organization Systems

Experience designer Nathan Shedroff suggests that the first step in transforming data into information is exploring its organization.* As you've seen in this chapter, organization systems are fairly complex. You need to consider a variety of exact and ambiguous organization schemes. Should you organize by topic, by task, or by audience? How about a chronological or geographical scheme? What about using multiple organization schemes?

You also need to think about the organization structures that influence how users can navigate through these schemes. Should you use a hierarchy, or would a more structured database model work best? Perhaps a loose hypertextual web would allow the most flexibility? Taken together in the context of a large web site development project, these questions can be overwhelming. That's why it's important to break down the site into its components, so you can tackle one question at a time. Also, keep in mind that all information-retrieval systems work best when applied to narrow domains of homogeneous content. By decomposing the content collection into these narrow domains, you can identify opportunities for highly effective organization systems.

However, it's also important not to lose sight of the big picture. As with cooking, you need to mix the right ingredients in the right way to get the desired results. Just because you like mushrooms and pancakes doesn't mean they will go well together. The recipe for cohesive organization systems varies from site to site. However, there are a few guidelines to keep in mind.

* For an interesting perspective on organizing things, see Nathan Shedroff's Unified Theory of Design at *http://www.nathan.com/thoughts/unified/6.html*.

When considering which organization schemes to use, remember the distinction between exact and ambiguous schemes. Exact schemes are best for known-item searching, when users know precisely what they are looking for. Ambiguous schemes are best for browsing and associative learning, when users have a vaguely defined information need. Whenever possible, use both types of schemes. Also, be aware of the challenges of organizing information on the Web. Language is ambiguous, content is heterogeneous, people have different perspectives, and politics can rear its ugly head. Providing multiple ways to access the same information can help to deal with all of these challenges.

When thinking about which organization structures to use, keep in mind that large web sites and intranets typically require several types of structure. The top-level, umbrella architecture for the site will almost certainly be hierarchical. As you are designing this hierarchy, keep a lookout for collections of structured, homogeneous information. These potential subsites are excellent candidates for the database model. Finally, remember that less structured, more creative relationships between content items can be handled through author-supplied hypertext or user-contributed tagging. In this way, myriad organization structures together can create a cohesive organization system.

CHAPTER 6

Labeling Systems

What we'll cover:

- What labeling is and why it's important
- Common types of labels
- Guidelines for developing labels
- Developing labels: borrowing from existing sources or starting from scratch

Labeling is a form of representation. Just as we use spoken words to represent concepts and thoughts, we use labels to represent larger chunks of information in our web sites. For example, "Contact Us" is a label that represents a chunk of content, often including a contact name, an address, and telephone, fax, and email information. You cannot present all this information quickly and effectively on an already crowded web page without overwhelming impatient users who might not actually need that information. Instead, a label like "Contact Us" works as a shortcut that triggers the right association in the user's mind without presenting all that stuff prominently. The user can then decide whether to click through or read on to get more contact information. So the goal of a label is to communicate information efficiently; that is, to convey meaning without taking up too much of a page's vertical space or a user's cognitive space.

Unlike the weather, hardly anyone ever talks about labeling (aside from a few deranged librarians, linguists, journalists, and, increasingly, information architects), but everyone can do something about it. In fact, we *are* doing something about it, albeit unconsciously: anyone developing content or an architecture for a web site is creating labels without even realizing it. And our label creation goes far beyond our web sites; ever since Adam named the animals, labeling has been one of the things that make us human. Spoken language is essentially a labeling system for concepts and things. Perhaps because we constantly label, we take the act of labeling for granted. That's why the labeling on web sites is often poor, and users suffer the consequences. This chapter provides some advice on how to think through a site's labeling before diving into implementation.

How does labeling fit with the other systems we've discussed? Well, labels are often the most obvious way to clearly show the user your organization and navigation systems. For example, a single web page might contain different groups of labels, with each group representing a different organization or navigation system. Examples include labels that match the site's organization system (e.g., Home/Home Office, Small Business, Medium & Large Business, Government, Health Care), a site-wide navigation system (e.g., Main, Search, Feedback), and a subsite navigation system (e.g., Add to Cart, Enter Billing Information, Confirm Purchase).

Why You Should Care About Labeling

Prerecorded or canned communications, including print, the Web, scripted radio, and TV, are very different from interactive real-time communications. When we talk with another person, we rely on constant user feedback to help us hone the way we get our message across. We subconsciously notice our conversation partner zoning out, getting ready to make her own point, or beginning to clench her fingers into an angry fist, and we react by shifting our own style of communication, perhaps by raising our speaking volume, increasing our use of body language, changing a rhetorical tack, or fleeing.

Unfortunately, when we "converse" with users through the web sites we design, the feedback isn't quite so immediate, if it exists at all. There are certainly exceptions—blogs, for example—but in most cases a site serves as an intermediary that slowly translates messages from the site's owners and authors to users, and back again. This "telephone game" muddies the message. So in such a disintermediated medium with few visual cues, communicating is harder, and labeling is therefore more important.

To minimize this disconnect, information architects must try their best to design labels that speak the same language as a site's users while reflecting its content. And, just as in a dialogue, when there is a question or confusion over a label, there should be clarification and explanation. Labels should educate users about new concepts and help them quickly identify familiar ones.

The conversation between user and site owner generally begins on a site's main page. To get a sense of how successful this conversation might be, look at a site's main page, do your best to ignore the other aspects of its design, and ask yourself a few questions: Do the prominent labels on this page stand out to you? If they do, why? (Often, successful labels are invisible; they don't get in your way.) If a label is new, unanticipated, or confusing, is there an explanation? Or are you required to click through to learn more? Although unscientific, this label testing exercise will help you get a sense of how the conversation might go with actual users.

Let's try it with an average, run-of-the-mill main page from the U-Haul site,* which is shown in Figure 6-1.

* In fairness to the good folks at U-Haul, their site is much improved since we grabbed this screen shot. But as the old design remains a wonderfully useful example of labeling problems, we've decided to keep it.

Figure 6-1. How do you respond to these labels?

The U-Haul main page's labels don't seem terribly out of the ordinary. However, mediocrity isn't an indicator of value or success; in fact, many trouble spots arise from an informal cruise through the page's labels. We've identified them as follows:

Main

"Main" refers to what? In web parlance, "Main" typically has something to do with a main page. Here, it describes a set of useful link labels such as "Get Rates & Reservations" and "Find a U-Haul Location." Why label these important links as "Main"? There are other possible labels, or visual design techniques could have been used to make the links stand out without mixing things up by using a conventional term like "Main." What exactly will be found under "College Connection"? It sounds like a branded program. Although it may represent useful content and functionality, that label sounds like part of U-Haul's corporate-speak, not the language of users.

Products & Services

If I wanted a hand truck, I'd look under "Hand trucks," not "Dollies." This disconnect may be due to regional differences: U-Haul is based in Phoenix, and I'm from New York. But which is the more common usage? Or if both labels are comparably common, should U-Haul list both terms?

SuperGraphics

Have you ever heard this term before? SuperGraphics are not graphics; they're apparently something better ("super"). English is wonderfully flexible, and new words are invented every day. But it's not realistic to expect impatient users to catch up with your linguistic creativity. Are "SuperGraphics" as important as "Products & Services"? What will we find behind the link "Pictorial Tribute to North America"—photos, a travelogue? And just what does such a tribute have to do with leasing trucks anyway?

Corporate

Do users understand what "Corporate" means? The term sounds, well, rather corporate, as if it might be intended for employees, suppliers, and others involved with the corporation. Perhaps the more conventional label "About Us" might be more appropriate. "Company Move" is a service for corporate relocations, not anything about U-Haul moving to new headquarters. Other links don't appear to belong here: like "Corporate Move," "Truck Sales" seems like it should go under "Products & Services." "Real Estate" and "Missing or Abandoned Equipment" are oddities that don't seem to belong anywhere. Is "Corporate" really another way of saying "Miscellaneous"?

Buy Online

Like "SuperGraphics," this label describes a single link, which is wasteful. And that link, "The U-Haul Store," seems to be a place to purchase or lease products and services. Why is "The U-Haul Store" set aside here? Does U-Haul want to accentuate it for some reason? If that reason has little to do with users, perhaps it's got everything to do with internal politics—perhaps one U-Haul VP owns "Products & Services," another owns "The U-Haul Store," and until they battle out their turf issues and one is extinguished, never the twain shall meet.

The results of this quick exercise can be summarized by these categories:

The labels aren't representative and don't differentiate

Too many of U-Haul's labels don't represent the content they link to or precede. Other than clicking through, users have no way to learn what "Corporate Move" means, or what the difference is between "Products & Services" and "The U-Haul Store." Groupings of dissimilar items (e.g., "Truck Sales," "Public Relations," and "Missing or Abandoned Equipment") don't provide any context for what those items' labels really represent. There is too much potential for confusion to consider these labels effective.

The labels are jargony, not user-centric

Labels like "College Connection" and "SuperGraphics" can expose an organization that, despite its best intentions, does not consider the importance of its customers' needs as important as its own goals, politics, and culture. This is often the case when web sites use organizational jargon for their labels. You've probably seen such sites; their labels are crystal clear, obvious, and enlightening, as

long as you're one of the .01 percent of users who actually work for the sponsoring organization. A sure way to lose a sale is to label your site's product-ordering system as an "Order Processing and Fulfillment Facility."

The labels waste money
: There are too many chances for a user to step into one of the many confusing cognitive traps presented by U-Haul's labels. And any time an architecture intrudes on a user's experience and forces him to pause and say "huh?", there is a reasonable chance that he will give up on a site and go somewhere else, especially given the competitive nature of this medium. In other words, confusing labels can negate the investment made to design and build a useful site and to market that site to intended audiences.

The labels don't make a good impression
: The way you say or represent information in your site says a lot about you, your organization, and its brand. If you've ever read an airline magazine, you're familiar with those ads for some educational cassette series that develops your vocabulary. "The words you use can make or break your business deals" or something like that. The same is true with a web site's labeling—poor, unprofessional labeling can destroy a user's confidence in that organization. While it may have spent heavily on traditional branding, U-Haul doesn't seem to have given much thought to the labels on the most important piece of its virtual real estate—its main page. Customers might wonder if U-Haul will be similarly haphazard and thoughtless in the way it services its fleet of vehicles or handles the customer hotline.

Like writing or any other form of professional communication, labels do matter. It's fair to say that they're as integral to an effective web presence as any other aspect of your web site, be it brand, visual design, functionality, content, or navigability.

Varieties of Labels

On the Web, we regularly encounter labels in two formats: textual and iconic. In this chapter, we'll spend most of our time addressing textual labels (as they remain the most common despite the Web's highly visual nature), including:

Contextual links
: Hyperlinks to chunks of information on other pages or to another location on the same page

Headings
: Labels that simply describe the content that follows them, just as print headings do

Navigation system choices
: Labels representing the options in navigation systems

Index terms
: Keywords, tags, and subject headings that represent content for searching or browsing.

These categories are by no means perfect or mutually exclusive. A single label can do double duty; for example, the contextual link "Naked Bungee Jumping" could lead to a page that uses the heading label "Naked Bungee Jumping" and has been indexed as being about (you guessed it) "Naked Bungee Jumping." And some of these labels could be iconic rather than textual, although we'd rather not imagine a visual representation of naked bungee jumping.

In the following section, we'll explore the varieties of labeling in greater detail and provide you with some examples.

Labels As Contextual Links

Labels describe the hypertext links within the body of a document or chunk of information, and naturally occur within the descriptive context of their surrounding text. Contextual links are easy to create and are the basis for the exciting interconnectedness that drives much of the Web's success.

However, just because contextual links are relatively easy to create doesn't mean they necessarily work well. In fact, ease of creation introduces problems. Contextual links are generally not developed systematically; instead, they are developed in an ad hoc manner when the author makes a connection between his text and something else, and encodes that association in his document. These hypertext connections are therefore more heterogeneous and personal than, say, the connections between items in a hierarchy, where links are understood to be connecting parent items and child items. The result is that contextual link labels mean different things to different people. You see the link "Shakespeare" and, upon clicking it, expect to be taken to the Bard's biography. I, on the other hand, expect to be taken to his Wikipedia entry. In fact, the link actually takes us to a page for the village of Shakespeare, New Mexico, USA. Go figure….

To be more representational of the content they connect to, contextual links rely instead upon, naturally, context. If the content's author succeeds at establishing that context in his writing, then the label draws meaning from its surrounding text. If he doesn't, the label loses its representational value, and users are more likely to experience occasionally rude surprises.

Because Fidelity (Figure 6-2) is a site dedicated to providing information to investors, contextual links need to be straightforward and meaningful. Fidelity's contextual link labels, such as "stocks," "mutual funds," and "Learn how to invest," are representational, and draw on surrounding text and headings to make it clear what type of help you'll receive if you click through. These highly representational labels are made even clearer by their context: explanatory text, clear headings, and a site that itself has a few straightforward uses.

On the other hand, contextual links on a personal web log ("blog") aren't necessarily so clear. The author is among friends and can assume that his regular readers possess a certain level of background, or really, contextual knowledge. Or he knows that keeping his

Figure 6-2. The contextual links on this page from Fidelity are straightforward and meaningful

link labels less representational creates some mystery around what they'll lead to. So the author may choose to design contextual link labels that aren't so representational.

In Figure 6-3, the author expects us to know who "Eric Sinclair" is—perhaps he's been mentioned in this blog before. Or the author knows that we'll recognize the label "Eric Sinclair" as a person, and provides some minimal context—the fact that Eric wrote some comments—to entice the user to click through. "They Rule" is equally mysterious; we have no idea what this label represents, but the blog author contextualizes it as "fascinating" and "scary." Nonrepresentational labels have their place; as it's likely that we already trust the author's opinion, we'll probably want to click through and learn more. But without that degree of trust already in place, non-representational links can be damaging.

As we'll see, other varieties of labels derive context, and therefore meaning, from being part of larger sets of labels or labeling systems. But systematic consistency isn't

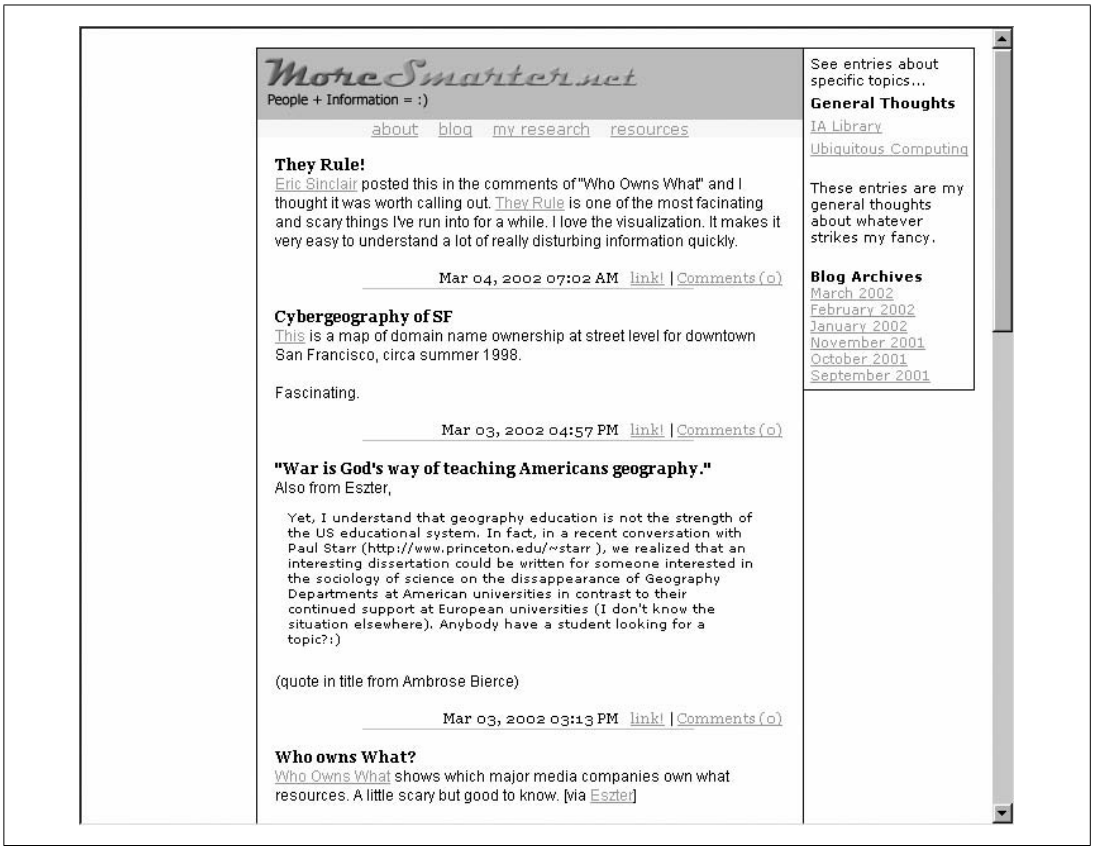


Figure 6-3. These contextual links aren't very representational, but that's acceptable when there is a high degree of trust for the author

quite so possible with link labels. These labels are instead glued together by the copy and context rather than membership in a peer group. However, consistency among these labels and the chunks of information to which they link remains an issue to keep in mind.

An information architect can ensure that contextual link labels are representational by asking herself, "What kind of information will the user expect to be taken to?" before creating and labeling a contextual link. Contextual links are created in such an ad hoc manner that simply asking this question will improve the quality of representation. (An easy way to study users' interpretations of labels is to provide a printout of a page with the labels clearly identified, and have subjects jot down what they'd expect each to link to.)

On the other hand, it's important to acknowledge that contextual links are often not within the information architect's control. Usually, content authors are responsible for contextual links. They are the ones who know the meaning of their content and how to best link it to other content. So while you may want to enforce rules for contextual link labels (such as what an employee's name should always link to), you may be better off suggesting guidelines to content authors (such as suggesting that employees' names link to a corresponding directory listing when possible).

Labels As Headings

Labels are often used as headings that describe the chunk of information that follows. Headings, as shown in Figure 6-4, are often used to establish a hierarchy within a text. Just as in a book, where headings help us distinguish chapters from sections, they also help us determine a site's subsites, or differentiate categories from subcategories.

Figure 6-4. Numbering, bullets, bolding, and vertical whitespace help the reader distinguish heading labels

The hierarchical relationships between headings—whether parent, child, or sibling—are usually established visually through consistent use of numbering, font sizes, colors and styles, whitespace and indentation, or combinations thereof. A visually clear hierarchy, often the work of information or graphic designers, can take some pressure off information architects by reducing the need to create labels that convey that hierarchy. So a set of labels that don't mean much can suddenly take on meaning when presented in a hierarchy. For example, this set of inconsistent headings may be quite confusing:

Our Furniture Selection
Office Chairs
Our buyer's picks

Chairs from Steelcase
Hon products
Herman Miller
Aerons
Lateral Files

However, they are much more meaningful when presented in a hierarchy:

```
Our Furniture Selection
Office Chairs
        Our buyer's picks
                Chairs from Steelcase
                Hon products
                Herman Miller
                        Aerons
Lateral Files
```

It's also important not to be too rigidly bound to showcasing hierarchical relationships. In Figure 6-5, heading labels such as "Background" and "Scouting report" represent the text that follows them. Yet the statistics closer to the top of the page don't merit the same treatment because most readers could visually distinguish these without actually reading them. In other words, inserting the heading "Statistics" before the numbers and applying to it the same typographic style as "Background" and "Scouting report" wouldn't greatly benefit users, who, as baseball fans, would likely recognize them already.

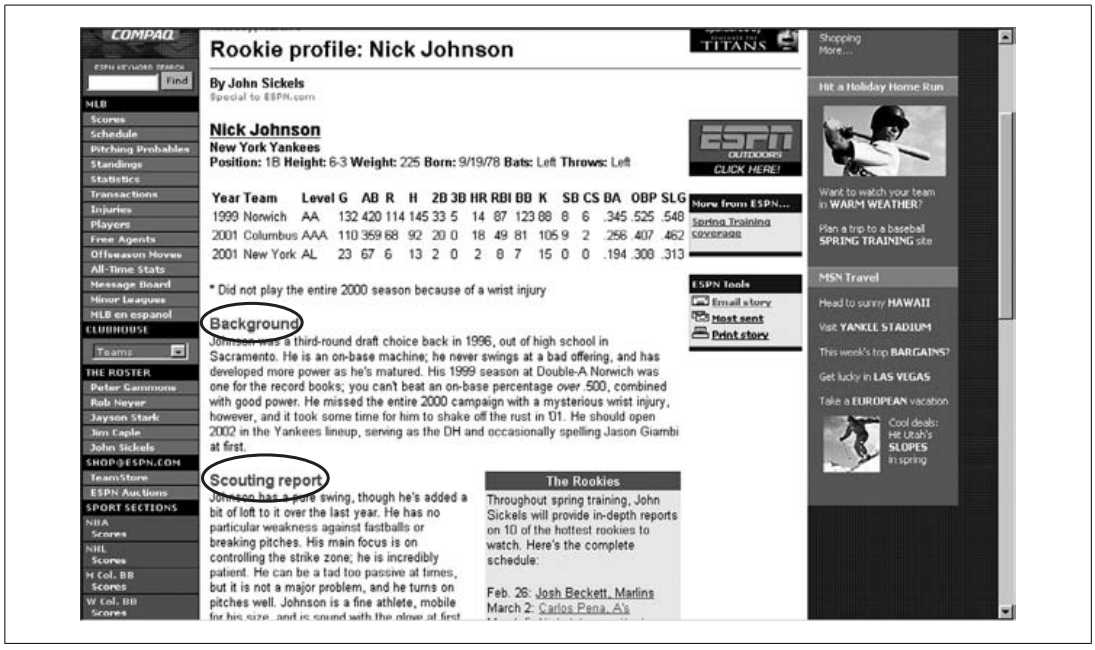


Figure 6-5. This hierarchy of heading labels is inconsistent, but that's OK

It is interesting to note, however, that it's difficult to distinguish one column of statistics from another, so each utilizes its own heading label.

We can be a bit more flexible when designing hierarchical headings, but it's especially important to maintain consistency when labeling steps in a process. To successfully navigate a process, it's typically necessary for users to complete each step along the way, so heading labels have to be obvious and must also convey sequence. Using numbers is an obvious way to communicate progression, and consistently framing the labels as actions—utilizing verbs—also helps tie together the sequence of steps. In effect, the labels should tell users where to start, where to go next, and what action will be involved in each step along the way. Figure 6-6 shows a page from Northwest Airlines in which the heading labels are clearly numbered, are consistently laid out, and utilize a consistent syntax that describes the question addressed in each step of the process.

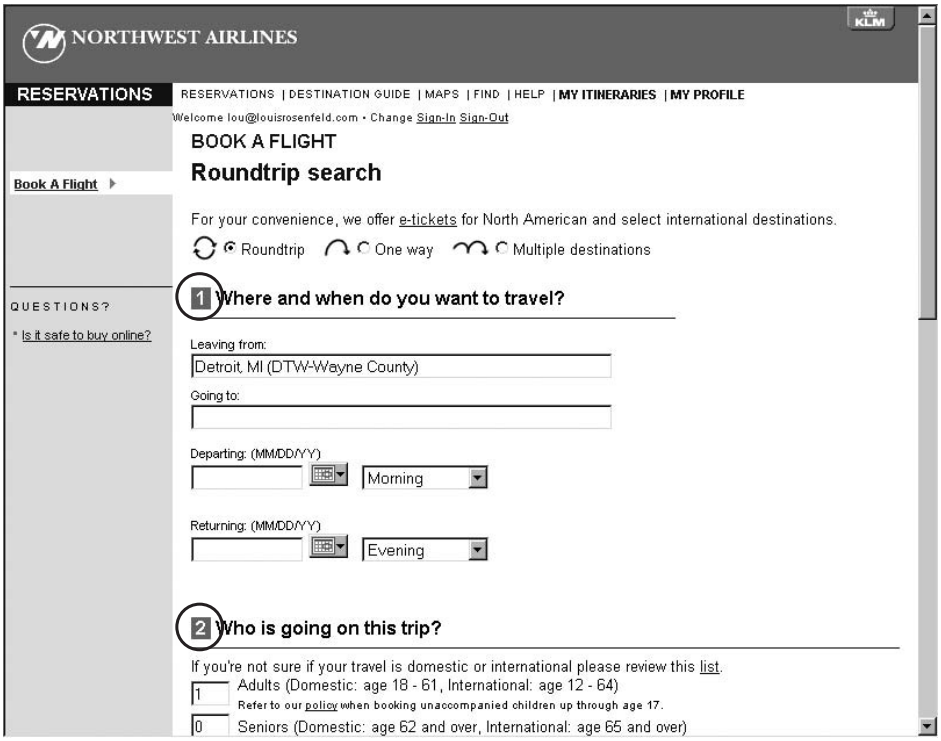


Figure 6-6. Sequential numbering and consistent syntax keep these labels clear

Heading labels, whether hierarchical or sequenced, come in multiples, and should be more systematically designed than contextual link labels.

Labels Within Navigation Systems

Because navigation systems typically have a small number of options, their labels demand consistent application more than any other type of label. A single inconsistent option can introduce an "apples and oranges" effect more quickly to a navigation

system, which usually has fewer than ten choices, than to a set of index terms, which might have thousands. Additionally, a navigation system typically occurs again and again throughout a site, so navigation labeling problems are magnified through repeated exposure.

Users rely on a navigation system to behave "rationally" through consistent page location and look; labels should be no different. Effectively applied labels are integral to building a sense of familiarity, so they'd better not change from page to page. That's why using the label "Main" on one page, "Main Page" on another, and "Home" elsewhere could destroy the familiarity that the user needs when navigating a site. In Figure 6-7, the horizontal navigation system's four labels—"Getting Started," "Our Funds," "Planning," and "My Account"—are applied consistently throughout the site, and would be even more effective if colors and locations were also consistent.

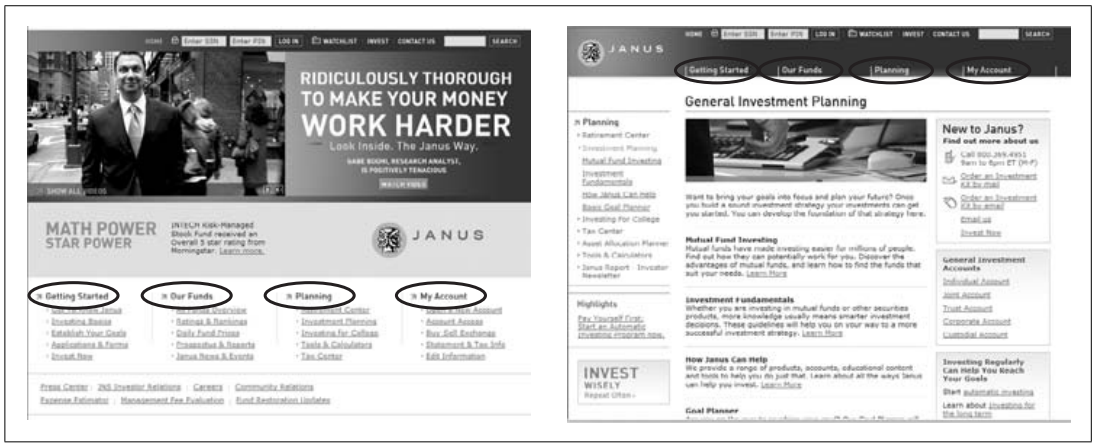


Figure 6-7. Janus' navigation system labels remain consistent throughout the site

There are no standards, but some common variants exist for many navigation system labels. You should consider selecting one from each of these categories and applying it consistently, as these labels are already familiar to most web users. Here is a nonexhaustive list:

- Main, Main Page, Home
- Search, Find, Browse, Search/Browse
- Site Map, Contents, Table of Contents, Index
- Contact, Contact Us
- Help, FAQ, Frequently Asked Questions
- News, News & Events, News & Announcements, Announcements
- About, About Us, About <company name>, Who We Are

Of course, the same label can often represent different kinds of information. For example, in one site, "News" may link to an area that includes announcements of

new additions to the site. In another site, "News" may link to an area of news stories describing national and world events. Obviously, if you use the same labels in different ways within your own site, your users will be very confused.

To address both problems, navigational labels can be augmented by brief descriptions (also known as *scope notes*) when initially introduced on the main page. In Figure 6-8, the navigation system labels appear in brief on the lefthand side, and are described with scope notes in the body of the main page.

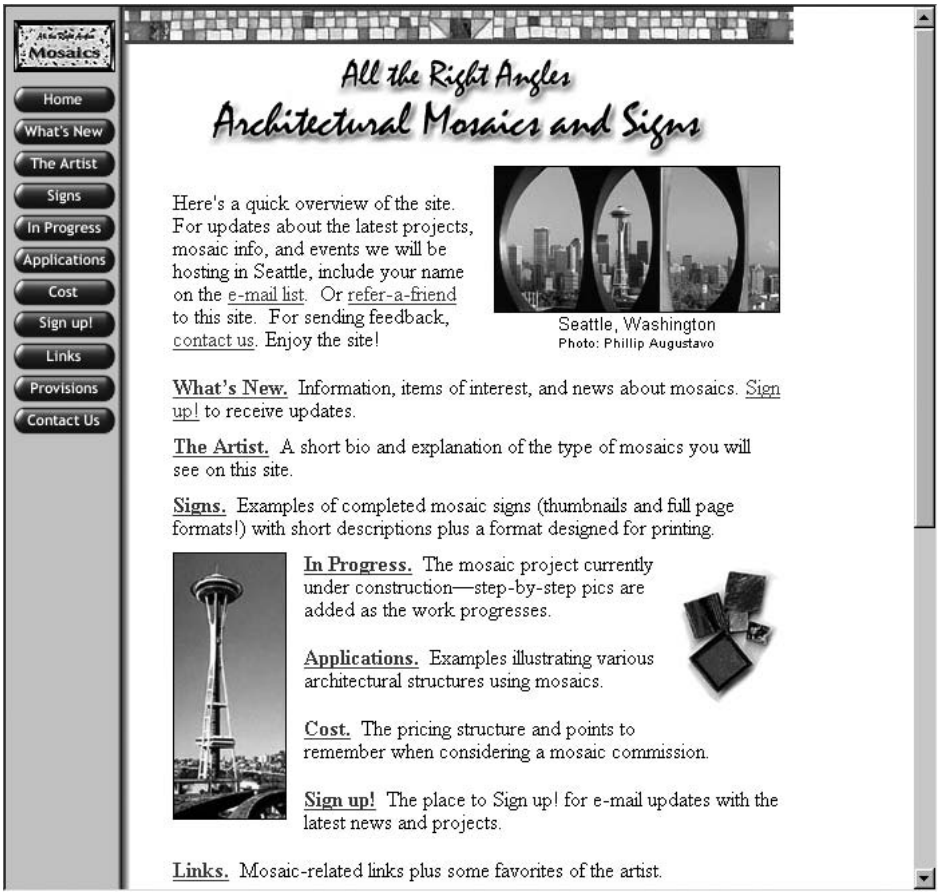


Figure 6-8. Scope notes are provided for each of the navigation system labels

In this case, if more representational navigation system labels had been used in the first place, they may have diminished the need to devote so much valuable main page real estate to scope notes. There are alternatives to scope notes that don't monopolize real estate, such as using JavaScript rollovers and other scripted mouseover effects to display the scope note, but these aren't an established convention. If you feel that your site will be regularly used by a loyal set of users who are willing to learn your site's conventions, then it's worth considering these alternatives; otherwise, we suggest keeping things simple by making your navigation labels representational.

Labels As Index Terms

Often referred to as keywords, tags, descriptive metadata, taxonomies, controlled vocabularies, and thesauri, sets of index term labels can be used to describe any type of content: sites, subsites, pages, content chunks, and so on. By representing the meaning of a piece of content, index terms support more precise searching than simply searching the full text of content—someone has assessed the content's meaning and described it using index terms, and searching those terms ought to be more effective than having a search engine match a query against the content's full text.

Index terms are also used to make browsing easier: the metadata from a collection of documents can serve as the source of browsable lists or menus. This can be highly beneficial to users, as index terms provide an alternative to a site's primary organization system, such as an information architecture organized by business unit. Index terms in the form of site indexes and other lists provide a valuable alternative view by "cutting across the grain" of organizational silos.

In Figure 6-9, this index of the BBC's site is generated from index term labels, which, in turn, are used to identify content from many different Sun business units. Much of the content already accessible through the BBC site's primary organization system is also accessible by browsing these index terms (e.g., keywords).

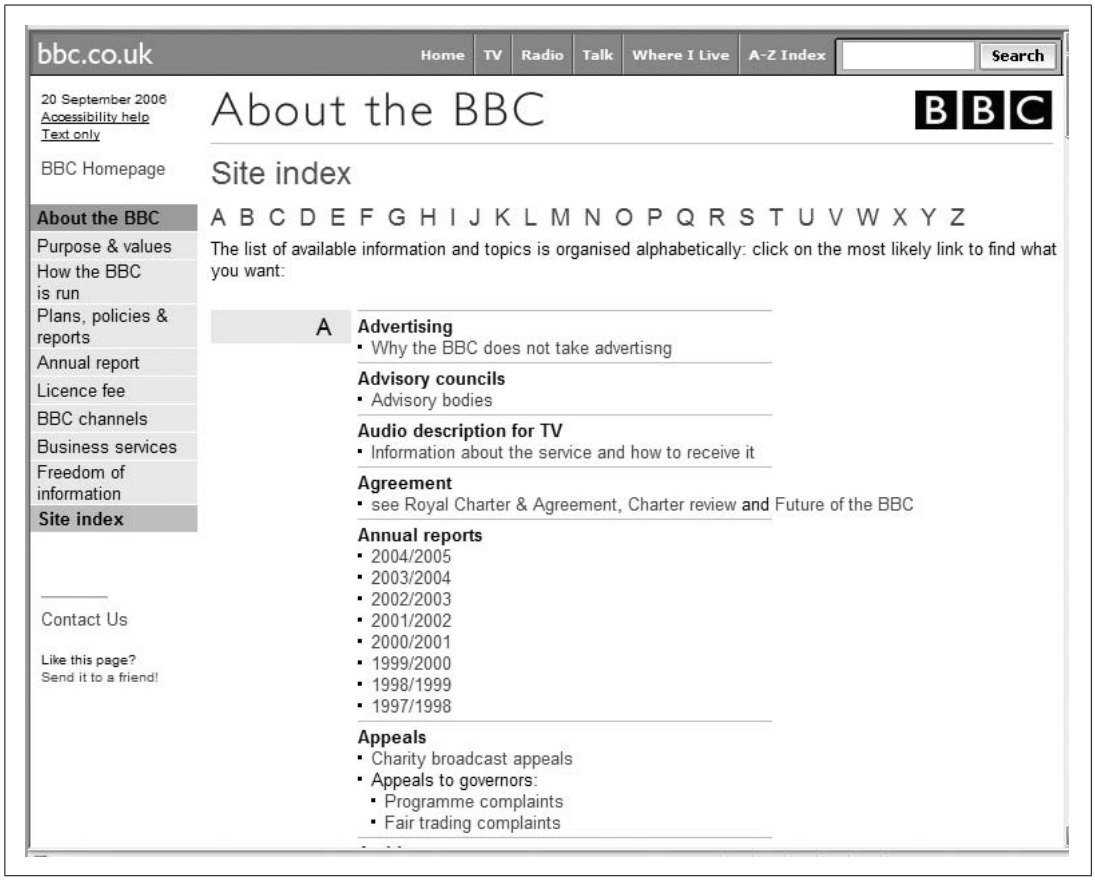


Figure 6-9. The BBC's site index

Frequently, index terms are completely invisible to users. The records we use to represent documents in content management systems and other databases typically include fields for index terms, which are often heard but not seen: they come into play only when you search. Similarly, index terms may be hidden as embedded metadata in an HTML document's <META...> or <TITLE> tags. For example, a furniture manufacturer's site might list the following index terms in the <META...> tags of records for its upholstered items:

```
<META NAME="keywords" CONTENT="upholstery, upholstered, sofa, couch,
loveseat, love seat, sectional, armchair, arm chair, easy chair,
chaise lounge">
```

So a search on "sofa" would retrieve the page with these index terms even if the term "sofa" doesn't appear anywhere in the page's text. Figure 6-10 shows a similar, more delectable example from Epicurious.com. A search for "snack" retrieves this recipe, though there is no mention of the term in the recipe itself. "Snack" is likely stored separately as an index term in a database record for this recipe.

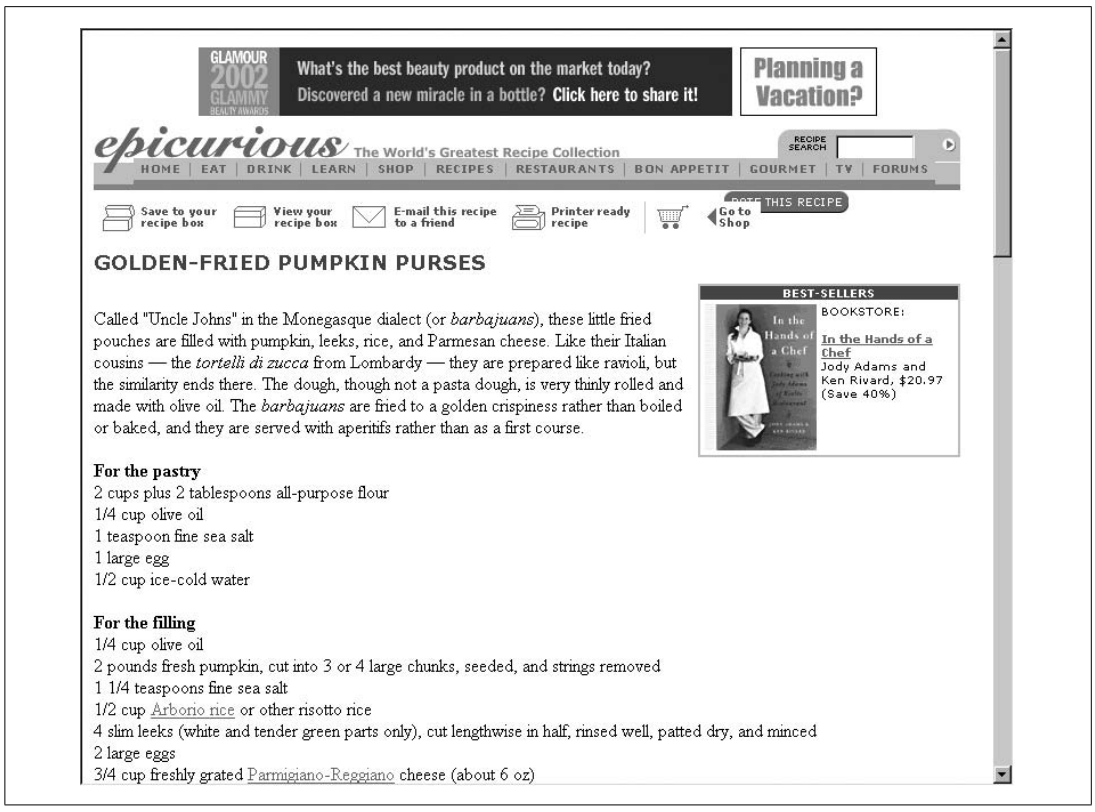


Figure 6-10. A search for "snack" retrieves this recipe, even though the term doesn't appear within the text

It's interesting how many sites' main pages don't feature index terms. Organizations do crazy, expensive things to get their sites noticed, including advertising their URL on banners flown over football stadiums. But using index terms to describe a main

page is a much cheaper way for getting that page, and the site as a whole, indexed and "known" so that users who search the Web are more likely to find it.*

Getting your pages to stand out from one another is a different and much more daunting challenge. That's where a more systematic approach to labeling—using index terms from controlled vocabularies or thesauri—has more value. These sets of labels are designed to describe a delineated domain—such as products and services, or oncology—and to do so in a consistent, predictable manner. We'll describe these vocabularies in great detail in Chapter 9.

Iconic Labels

It's true that a picture is worth a thousand words. But which thousand?

Icons can represent information in much the same way as text can. We see them most frequently used as navigation system labels. Additionally, icons occasionally serve as heading labels and have even been known to show up as link labels, although this is rare.

The problem with iconic labels is that they constitute a much more limited language than text. That's why they're more typically used for navigation system or small organization system labels, where the list of options is small, than for larger sets of labels such as index terms, where iconic "vocabularies" are quickly outstripped. (They also can work well for less text-oriented audiences, like children.)

Even so, iconic labels are still a risky proposition in terms of whether or not they can represent meaning. Figure 6-11 is a navigation aid from jetBlue's web site. But what do the icons mean to you?

Figure 6-11. Icons from jetBlue's navigation system

Even given the fairly specific context of an airline's site, most users probably won't understand this language immediately, although they might correctly guess the meaning of one or two of these labels.

Since the iconic labels are presented with textual labels, our test wasn't really fair. But it is interesting to note that even the site's designers acknowledge that the iconic labels don't stand well on their own and hence need textual explanations.

* Search Engine Watch (*http://www.searchenginewatch.com*) is the most useful resource for learning how web-wide search engines and directories work, and how you can index your site's main and other major pages to "rise to the top" of retrieval results.

Iconic labels like these add aesthetic quality to a site, and as long as they don't compromise the site's usability, there's no reason not to use them. In fact, if your site's users visit regularly, the iconic "language" might get established in their minds through repeated exposure. In such situations, icons are an especially useful shorthand, both representational and easy to visually recognize—a double bonus. But it's interesting to note that jetBlue's subsidiary pages don't use iconic labels alone; they've chosen to maintain the icon/text pairing throughout their site. Unless your site has a patient, loyal audience of users who are willing to learn your visual language, we suggest using iconic labels only for systems with a limited set of options, being careful not to place form ahead of function.

Designing Labels

Designing effective labels is perhaps the most difficult aspect of information architecture. Language is simply too ambiguous for you to ever feel confident that you've perfected a label. There are always synonyms and homonyms to worry about, and different contexts influence our understanding of what a particular term means. But even labeling conventions are questionable: you absolutely cannot assume that the label "main page" will be correctly interpreted by 100 percent of your site's users. Your labels will never be perfect, and you can only hope that your efforts make a difference, as measuring label effectiveness is an extremely difficult undertaking.

If it sounds to you like labeling is an art rather than a science, you're absolutely correct. And, as in all such cases, you can forget about finding incontrovertible rules, and hope for guidelines instead. Following are some guidelines and related issues that will help you as you delve into the mysterious art of label design.

General Guidelines

Remember that *content, users*, and *context* affect all aspects of an information architecture, and this is particularly true with labels. Any of the variables attached to users, content, and context can drag a label into the land of ambiguity.

Let's go back to the term "pitch." From baseball (what's thrown) to football (the field where it's played in the United Kingdom), from sales (what's sometimes made on the golf course) to sailing (the angle of the boat in the water), there are at least 15 different definitions, and it's hard to make sure that your site's users, content, and context will converge upon the same definition. This ambiguity makes it difficult to assign labels to describe content, and difficult for users to rely on their assumptions about what specific labels actually mean.

So what can we do to make sure our labels are less ambiguous and more representational? The following two guidelines may help.

Narrow scope whenever possible

If we focus our sites on a more defined audience, we reduce the number of possible perspectives on what a label means. Sticking to fewer subject domains achieves more obvious and effective representation. A narrower business context means clearer goals for the site, its architecture, and therefore its labels.

To put it another way, labeling is easier if your site's content, users, and context are kept simple and focused. Too many sites have tried to take on too much, achieving broad mediocrity rather than nailing a few choice tasks. Accordingly, labeling systems often cover too much ground to truly be effective. If you are planning any aspect of your site's scope—who will use it, what content it will contain, and how, when, and why it should be used—erring toward simplicity will make your labels more effective.

If your site must be a jack of all trades, avoid using labels that address the entire site's content. (The obvious exception are the labels for site-wide navigation systems, which do cover the entire site.) But in the other areas of labeling, modularizing and simplifying content into subsites that meet the needs of specific audiences will enable you to design more modular, simpler collections of labels to address those specific areas.

This modular approach may result in separate labeling systems for different areas of your site. For example, records in your staff directory might benefit from a specialized labeling system that wouldn't make sense for other parts of the site, while your site-wide navigation system's labels wouldn't really apply to entries in the staff directory.

Develop consistent labeling systems, not labels

It's also important to remember that labels, like organization and navigation systems, are systems in their own right. Some are planned systems, some aren't. A successful system is designed with one or more characteristics that unify its members. In successful labeling systems, one characteristic is typically *consistency*.

Why is consistency important? Because consistency means predictability, and systems that are predictable are simply easier to learn. You see one or two labels, and then you know what to expect from the rest—if the system is consistent. This is especially important for first-time visitors to a site, but consistency benefits all users by making labeling easy to learn, easy to use, and therefore invisible.

Consistency is affected by many issues:

Style
: Haphazard usage of punctuation and case is a common problem within labeling systems, and can be addressed, if not eliminated, by using style guides. Consider hiring a proofreader and purchasing a copy of Strunk & White.

Presentation
: Similarly, consistent application of fonts, font sizes, colors, whitespace, and grouping can help visually reinforce the systematic nature of a group of labels.

Syntax
: It's not uncommon to find verb-based labels (e.g., "Grooming Your Dog"), noun-based labels (e.g., "Diets for Dogs"), and question-based labels (e.g., "How Do You Paper-Train Your Dog?") all mixed together. Within a specific labeling system, consider choosing a single syntactical approach and sticking with it.

Granularity
: Within a labeling system, it can be helpful to present labels that are roughly equal in their specificity. Exceptions (such as site indexes) aside, it's confusing to encounter a set of labels that cover differing levels of granularity. For example: "Chinese restaurants," "Restaurants," "Taquerias," "Fast Food Franchises," "Burger Kings."

Comprehensiveness
: Users can be tripped up by noticeable gaps in a labeling system. For example, if a clothing retailer's site lists "pants," "ties," and "shoes," while somehow omitting "shirts," we may feel like something's wrong. Do they really not carry shirts? Or did they make a mistake? Aside from improving consistency, a comprehensive scope also helps users do a better job of quickly scanning and inferring the content a site will provide.

Audience
: Mixing terms like "lymphoma" and "tummy ache" in a single labeling system can also throw off users, even if only temporarily. Consider the languages of your site's major audiences. If each audience uses a very different terminology, you may have to develop a separate labeling system for each audience, even if these systems are describing exactly the same content.

There are other potential roadblocks to consistency. None is particularly difficult to address, but you can certainly save a lot of labor and heartache if you consider these issues before you dive into creating labeling systems.

Sources of Labeling Systems

Now that you're ready to design labeling systems, where do you start? Believe it or not, this is the easy part. Unless you're dealing with ideas, concepts, and topics that until now were unknown to humanity, you'll probably have something to start with. And already having a few labels generally beats starting from scratch, which can be prohibitively expensive, especially with large vocabularies.

Existing labeling systems might include the labels currently on your site, or comparable or competitors' sites. Ask yourself who might have taken this on before. Study, learn, and "borrow" from what you find on other sites. And keep in mind that a major benefit of examining existing labeling systems is that they're systems—they're more than odd, miscellaneous labels that don't necessarily fit together.

As you look for existing labeling systems, consider what works and what doesn't. Which systems can you learn from, and, perhaps more importantly, which of those labels can you keep? There are a variety of sources for labels that you should examine.

Your site

Your web site probably already has labeling systems by default. At least some reasonable decisions had to have been made during the course of the site's creation, so you probably won't want to throw all those labels out completely. Instead, use them as a starting point for developing a complete labeling system, taking into consideration the decisions made while creating the original system.

A useful approach is to capture the existing labels in a single document. To do so, walk through the entire site, either manually or automatically, and gather the labels. You might consider assembling them in a simple table containing a list or outline of each label and the documents it represents. Creating a labeling table is often a natural extension of the content inventory process. It's a valuable exercise, though we don't recommend it for indexing term vocabularies, which are simply too large to table-ize unless you focus on small, focused segments of those vocabularies.

Following is a table for the navigation system labels on jetBlue's main page.

Label	Destination's heading label	Destination's <TITLE>label
Top-of-page navigation system labels		
Buy tickets	-	Online booking
Hotels/cars	Book hotels and rent cars online	Hotels - jetBlue
Travel info	-	Travel info - JetBlue
Work here	-	Work here - JetBlue
Learn more	Welcome from our CEO	Learn more - JetBlue
Speak up	-	Speak up - JetBlue
ShopBlue	Now you're ready to shopBlue	Welcome to shopBlue!
Body navigation system labels		
Track your flight	Real-Time Flight Tracking	Travel info - JetBlue
Our cities	Route map	Travel info - JetBlue
What to expect at the airport	Important security information	JetBlue Airways
Have fun	-	Have fun - jetBlue
Register with us	-	Member Profile
Bottom-of-page navigation system labels		
Home	jetBlue	JetBlue
Sitemap	Sitemap	siteMap - JetBlue
Faqs	FAQs	Get help - jetBlue
Your privacy	Privacy	Privacy policy - JetBlue
Contact us	Contacts	Learn more - JetBlue
Jobs	-	Learn more - JetBlue
Travel agents	Travel agency login	Agency and Corporate Bookings
Espanol	jetBlue en espanol	jetBlue en espanol

Arranging labels in a table provides a more condensed, complete, and accurate view of a site's navigation labels as a *system*. Inconsistencies are easier to catch; in jetBlue's case, we encounter three variants of the company's name alone: "jetBlue," "JetBlue," and "JetBlue Airways." We find inconsistencies for a single page's labels: the contact page is labeled "Contact us," "Contacts," and "Learn more - JetBlue." Many pages don't have main headings. We encounter various other style inconsistencies that may confuse users. We may decide that, personally, we just don't like certain labels. We may also decide that some of the problems aren't worth changing. In any case, we now have a sense of the site's current labeling system and how it could be improved.

Comparable and competitive sites

If you don't have a site in place or are looking for new ideas, look elsewhere for labeling systems. The open nature of the Web allows us to learn from one another and encourages an atmosphere of benevolent plagiarism. So, just as you might view the source of a wonderfully designed page, you can "borrow" from another site's great labeling system.

Determine beforehand what your audiences' needs are most likely to be, and then surf your competitors' sites, borrowing what works and noting what doesn't (you might consider creating a label table for this specific purpose). If you don't have competitors, visit comparable sites or sites that seem to be best in class.

If you surf multiple competitive or comparative sites, you may find that labeling patterns emerge. These patterns may not yet be industry standards, but they at least can inform your choice of labels. For example, in a recent competitive analysis of eight financial services sites, "personal finance" was found to be more or less a de facto label compared to its synonyms. Such data may discourage you from using a different label.

Figure 6-12 shows labeling systems from Compaq, Gateway, Dell, and IBM, all competing in the PC business. Do you notice a trend here?

Controlled vocabularies and thesauri

Another great source is existing controlled vocabularies and thesauri (a topic we'll cover in depth in Chapter 9). These especially useful resources are created by professionals with library or subject-specific backgrounds, who have already done much of the work of ensuring accurate representation and consistency. These vocabularies are often publicly available and have been designed for broad usage. You'll find these to be most useful for populating labeling systems used for indexing content.

But here's a piece of advice: seek out narrowly focused vocabularies that help specific audiences to access specific types of content. For example, if your site's users are computer scientists, a computer science thesaurus "thinks" and represents concepts in a way your users are likely to understand, more so than a general scheme like the Library of Congress subject headings would.

A good example of a specific controlled vocabulary is the Educational Resources Information Center (ERIC) Thesaurus. This thesaurus was designed, as you'd guess, to

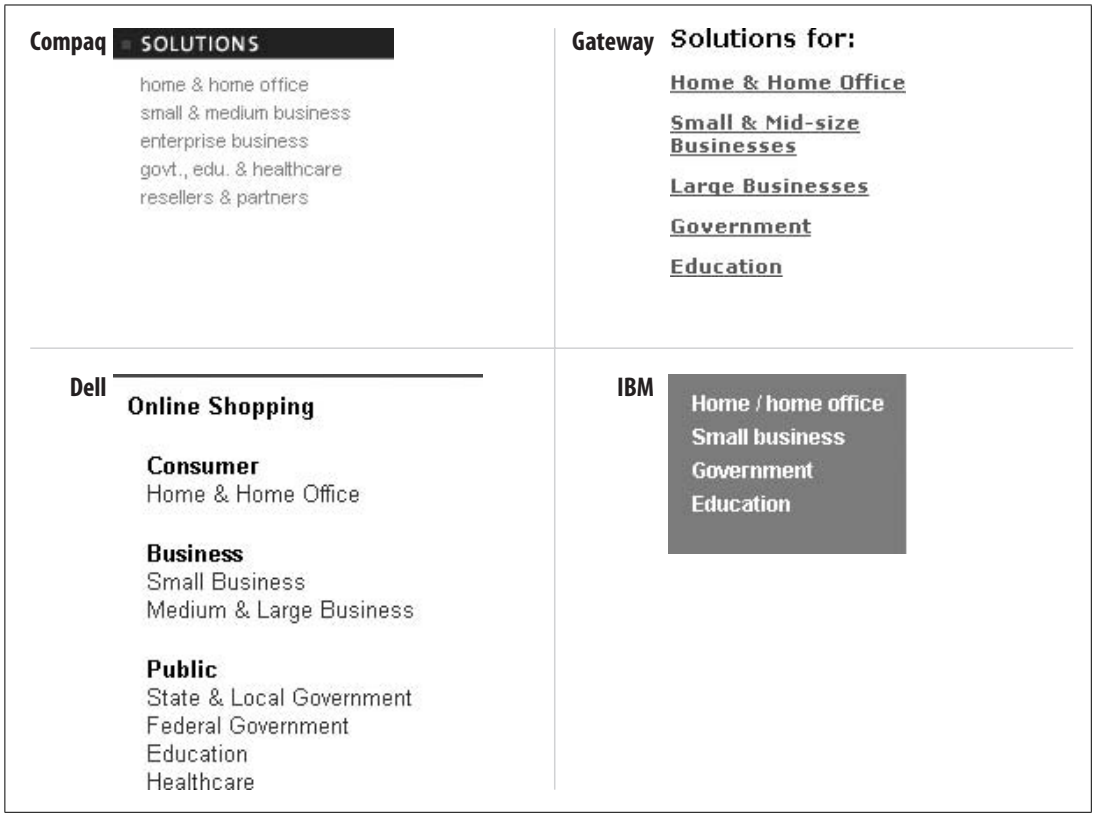


Figure 6-12. Labeling systems from Compaq, Gateway, Dell, and IBM

describe the domain of education. An entry in the ERIC Thesaurus for "scholarships" is shown in Figure 6-13.

If your site has to do with education or if your audience is comprised of educators, you might start with ERIC as the source for your site's labels. You can use a thesaurus like ERIC to help you with specific labeling challenges, like determining a better variant for a particularly knotty label. You might go as far as to license the entire vocabulary and use it as your site's labeling system.

Unfortunately, there aren't controlled vocabularies and thesauri for every domain. Sometimes you may find a matching vocabulary that emphasizes the needs of a different audience. Still, it's always worth seeing if a potentially useful controlled vocabulary or thesaurus exists before creating labeling systems from scratch. Try these four excellent lists as you hunt for sources of labels:

- Taxonomy Warehouse: *http://taxonomywarehouse.com/*
- ThesauriOnline (American Society of Indexers): *http://www.asindexing.org/site/thesonet.shtml*
- Controlled vocabularies (Michael Middleton): *http://sky.fit.qut.edu.au/~middletm/cont_voc.html*
- Web Thesaurus Compendium (Barbara Lutes): *http://www.ipsi.fraunhofer.de/~lutes/thesoecd.html*

Term:	**Scholarships**
Record Type:	Main
Scope Note:	Awards, usually of money or reduced tuition, given to students primarily in recognition of achievement or potential but also for other specific characteristics such as financial need, residence, or academic interest
Category:	620
Broader Terms:	Student Financial Aid;
Narrower Terms:	Merit Scholarships; Tuition Grants;
Related Terms:	Awards; Educational Finance; Educational Vouchers; Eligibility; Fellowships; Grants; Instructional Student Costs; Noninstructional Student Costs; Scholarship Funds; Student Costs;
Used For:	Endowed Scholarships
Use Term:	
Use And:	
Add Date:	07/01/1966

Figure 6-13. Controlled vocabularies and thesauri are rich sources of labels

Creating New Labeling Systems

When there are no existing labeling systems or when you need to do more customizing than you'd expected, you face the tougher challenge of creating labeling systems from scratch. Your most important sources are your content and your site's users.

Content analysis

Labels can come directly from your site's content. You might read a representative sample of your site's content and jot down a few descriptive keywords for each document along the way. It's a slow and painful process, and it obviously won't work with a huge set of documents. If you go this route, look for ways to speed up the process by focusing on any existing content representations like titles, summaries, and abstracts. Analyzing content for candidate labels is certainly another area where art dominates science.

There are software tools now available that can perform *auto-extraction* of meaningful terms from content. These tools can save you quite a bit of time if you face a huge body of content; like many software-based solutions, auto-extraction tools may get you 80 percent of the way to the finish line. You'll be able to take the terms that are output by the software and use them as candidates for a controlled vocabulary, but you'll still need to do a bit of manual labor to make sure the output actually makes

sense. (And it's worth noting that auto-extraction tools—and the training and tuning to make them work well—can be quite expensive.) We provide pointers to some auto-extraction tools in Chapter 16.

Content authors

Another manual approach is to ask content authors to suggest labels for their own content. This might be useful if you have access to authors; for example, you could talk to your company's researchers who create technical reports and white papers, or to the PR people who write press releases.

However, even when authors select terms from a controlled vocabulary to label their content, they don't necessarily do it with the realization that their document is only one of many in a broader collection. So they might not use a sufficiently specific label. And few authors happen to be professional indexers.

So take their labels with a grain of salt, and don't rely upon them for accuracy. As with other sources, labels from authors should be considered useful candidates for labels, not final versions.

User advocates and subject matter experts

Another approach is to find advanced users or user advocates who can speak on the users' behalf. Such people may include librarians, switchboard operators, or subject matter experts (SMEs) who are familiar with the users' information needs in a larger context. Some of these people—reference librarians, for example—keep logs of what users want; all will have a good innate sense of users' needs by dint of constant interaction.

We found that talking to user advocates was quite helpful when working with a major healthcare system. Working with their library's staff and SMEs, we set out to create two labeling systems, one with medical terms to help medical professionals browse the services offered by the healthcare system, the other for the lay audience to access the same content. It wasn't difficult to come up with the medical terms because there are many thesauri and controlled vocabularies geared toward labeling medical content. It was much more difficult to come up with a scheme for the layperson's list of terms. There didn't seem to be an ideal controlled vocabulary, and we couldn't draw labels from the site's content because it hadn't been created yet. So we were truly starting from scratch.

We solved this dilemma by using a top-down approach: we worked with the librarians to determine what they thought users wanted out of the site. We considered their general needs, and came up with a few major ones:

1. They need information about a problem, illness, or condition.
2. The problem is with a particular organ or part of the body.
3. They want to know about the diagnostics or tests that the healthcare professionals will perform to learn more about the problem.

4. They need information on the treatment, drug, or solution that will be provided by the healthcare system.
5. They want to know how they can pay for the service.
6. They want to know how they can maintain their health.

We then came up with basic terms to cover the majority of these six categories, taking care to use terms appropriate to this audience of laypersons. Here are some examples:

Category	Sample labels
Problem/illness/condition	HIV, fracture, arthritis, depression
Organ/body part	Heart, joints, mental health
Diagnostics/tests	Blood pressure, X-ray
Treatment/drug/solution	Hospice, bifocals, joint replacement
Payment	Administrative services, health maintenance organization, medical records
Health maintenance	Exercise, vaccination

By starting with a few groupings, we were able to generate labels to support indexing the site. We knew a bit about the audience (laypersons), and so were able to generate the right kinds of terms to support their needs (e.g., *leg* instead of *femur*). The secret was working with people (in this case, staff librarians) who were knowledgeable about the kind of information the users want.

Directly from users

The users of a site may be telling you, directly or indirectly, what the labels should be. This isn't the easiest information to get your hands on, but if you can, it's the best source of labeling there is.

Card sorting. *Card sort* exercises are one of the best ways to learn how your users would use information. (Card sorting methodologies* are covered more extensively in Chapter 10.) There are two basic varieties of card sorts: open and closed. *Open card sorts* allow subjects to cluster labels for existing content into their own categories and then label those categories (and clearly, card sorting is useful when designing organization systems as well as labeling systems). *Closed card sorts* provide subjects with existing categories and ask them to sort content into those categories. At the start of a closed card sort, you can ask users to explain what they think each category label represents and compare these definitions to your own. Both approaches are useful ways to determine labels, although they're more appropriate for smaller sets of labels such as those used for navigation systems.

* We also anticipate that Donna Maurer's book, *Card Sorting: The Book* will be quite helpful here; it will be published by Rosenfeld Media in early 2007 (*http://www.rosenfeldmedia.com/books/cardsorting*).

In the example below, we asked subjects to categorize cards from the owner's section of a site for a large automotive company (let's call it "Tucker"). After we combined the data from this open card sort, we found that subjects labeled the combined categories in different ways. "Maintenance," "maintain," and "owner's" were often used in labels for the first cluster, indicating that these were good candidates for labels (see Table 6-1).

Table 6-1. Cluster 1

Subject	Categories
Subject 1	Ideas & maintenance
Subject 2	Owner's guide
Subject 3	Items to maintain car
Subject 4	Owner's manual
Subject 5	Personal information from dealer
Subject 6	-
Subject 7	Maintenance upkeep & ideas
Subject 8	Owner's tip AND owner's guide and maintenance

But in other cases, no strong patterns emerged (see Table 6-2).

Table 6-2. Cluster 2

Subject	Categories
Subject 1	Tucker features
Subject 2	-
Subject 3	Shortcut for info on car
Subject 4	Auto info
Subject 5	Associate with dealer
Subject 6	Tucker web site info
Subject 7	Manuals specific to each car
Subject 8	-

In a corresponding closed card sort, we asked subjects to describe each category label before they grouped content under each category. In effect, we were asking subjects to define each of these labels, and we compared their answers to see if they were similar or not. The more similar the answers, the stronger the label.

Some labels, such as "Service & Maintenance," were commonly understood, and were in line with the content that you'd actually find listed under this category (see Table 6-3).

Table 6-3. Service & Maintenance

Subject	Content
Subject 1	When to change the fluids, rotate tires; a place to keep track when I had my vehicle in for service (sic)
Subject 2	How to maintain vehicle: proper maintenance, features of car, where to find fuse box, etc., owner's manual
Subject 3	Find service that might be open on Sunday sometimes
Subject 4	When I will need service and where to go to get it
Subject 5	Reminders on when services is recommended (sic)
Subject 6	Timeline for service and maintenance
Subject 7	Maintenance schedule and tips to get best performance out of car and longevity of car
Subject 8	Maintenance tips, best place to go to fix car problem, estimated price

Other category labels were more problematic. Some subjects understood "Tucker Features & Events" in the way that was intended, representing announcements about automobile shows, discounts, and so on. Others interpreted this label to mean a vehicle's actual features, such as whether or not it had a CD player (see Table 6-4).

Table 6-4. Tucker Features & Events

Subject	Content
Subject 1	New items for my vehicle; upcoming new styles—new makes & models; financial news—like 0% financing
Subject 2	Local & national sponsorship; how to obtain Tucker sponsorship; community involvement
Subject 3	Mileage, CD or cassette, leg room, passengers, heat/AC control dull or not, removable seats, automatic door openers
Subject 4	All information regarding the Tucker automobile I'm looking for and any sale events going on regarding this auto
Subject 5	Looking for special pricing events
Subject 6	Site for outlining vehicles and options available. What automobile shows are available and where
Subject 7	About Tucker, sales, discounts, special events
Subject 8	No interested (sic)

Card sort exercises are very informative, but it's important to recognize that they don't present labels in the context of an actual site. Without this natural context, the labels' ability to represent meaning is diminished. So, as with all other techniques, card sorts have value but shouldn't be seen as the only method of investigating label quality.

Free-listing. While card sorting isn't necessarily an expensive and time-consuming method, free-listing is an even lower-cost way to get users to suggest labels.* Free-listing is quite simple: select an item and have subjects brainstorm terms to

* The best summary of this method is Rashmi Sinha's short but highly useful article in the Februrary 2003 *Boxes & Arrows*, "Beyond cardsorting: Free-listing methods to explore user categorizations" (*http://www.boxesandarrows.com/view/beyond_cardsorting_free_listing_methods_to_explore_user_categorizations*).

describe it. You can do this in person (capturing data with pencil and paper will be fine) or remotely, using a free or low-cost online-survey tool like SurveyMonkey or Zoomerang. That's really all there is to it.

Well, not quite: you'll want to consider your subjects: who (ideally representative of your overall audience) and how many (three to five may not yield scientifically significant results, but it is certainly better than nothing and may yield some interesting results). You might also consider asking subjects to rank the terms they've suggested as a way to determine which are the most appropriate.

You'll also need to choose which items to brainstorm terms for. Obviously you can only do this with a subset of your content. You could choose some representative content, such as a handful of your company's products. But even then, it'll be tricky—do you choose the most popular products or the more esoteric ones? It's important to get the labeling right for your big sellers, but conventions for their labels are already fairly established. The esoteric items? Well, they're more challenging, but fewer people care about them. So you may end up with a balance among the few items you select for a free-listing exercise. This is one of those cases where the art of information architecture is at least as important as the science.

What do you do with the results? Look for patterns and frequency of usage; for example, most of your subjects use the term "cell phone" while surprisingly few prefer "mobile phone." Patterns like these provide you with a sense of how to label an individual item, but may also demonstrate the tone of users' language overall. You might note that they use jargon quite a bit, or the reverse; perhaps you find a surprising amount of acronyms in their labels, or some other pattern emerges from free-listing. The result won't be a full-fledged labeling system, but it will give you a better sense of what tone and style you should take when developing a labeling system.

Indirectly from users

Most organizations—especially those whose sites include search engines—are sitting on top of reams of user data that describe their needs. Analyzing those search queries can be a hugely valuable way to tune labeling systems, not to mention diagnose a variety of other problems with your site. Additionally, the recent advent of folksonomic tagging has also created a valuable, if indirect, source of data on users' needs that can help information architects develop labeling systems.

Search-log analysis. Search-log analysis (also known as search analytics) is one of the least intrusive sources of data on the labels your site's audiences actually use. Analyzing search queries* is a great way to understand the types of labels your site's visitors

* Naturally, we have one more book to recommend that's not yet quite available at press time, but that should be useful nonetheless: *Search Analytics for Your Site: Conversations with Your Customers*, by Louis Rosenfeld and Rich Wiggins. It will be published by Rosenfeld Media and should be available in early 2007 (*http://www.rosenfeldmedia.com/books/searchanalytics*).

typically use (see Table 6-5). After all, these are the labels that users use to describe their own information needs in their own language. You may notice the use (or lack thereof) of acronyms, product names, and other jargon, which could impact your own willingness to use jargony labels. You might notice that users' queries use single or multiple terms, which could affect your own choice of short or long labels. And you might find that users simply aren't using the terms you thought they would for certain concepts. You may decide to change your labels accordingly, or use a thesaurus-style lookup to connect a user-supplied term (e.g., "pooch") to the preferred term (e.g., "dog").

Table 6-5. The top 40 most common queries from Michigan State University's site, February 8–14, 2006; each query tells us something about what the majority of users seek most often and how they label their information needs

Rank	Count	Cumulative	Percent of total	Query
1	1184	1184	1.5330	capa
2	1030	2214	2.8665	lon+capa
3	840	3054	3.9541	study+abroad
4	823	3877	5.0197	angel
5	664	4541	5.8794	lon-capa
6	656	5197	6.7287	library
7	584	5781	7.4849	olin
8	543	6324	8.1879	campus+map
9	530	6854	8.8741	spartantrak
10	506	7360	9.5292	cata
11	477	7837	10.1468	housing
12	467	8304	10.7515	map
13	462	8766	11.3496	im+west
14	409	9175	11.8792	computer+store
15	399	9574	12.3958	state+news
16	395	9969	12.9072	wharton+center
17	382	10351	13.4018	chemistry
18	346	10697	13.8498	payroll
19	340	11037	14.2900	breslin+center
20	339	11376	14.7289	honors+college
21	339	11715	15.1678	calendar
22	334	12049	15.6002	human+resources
23	328	12377	16.0249	registrar
24	327	12704	16.4483	dpps
25	310	13014	16.8497	breslin

Table 6-5. The top 40 most common queries from Michigan State University's site, February 8–14, 2006; each query tells us something about what the majority of users seek most often and how they label their information needs (continued)

Rank	Count	Cumulative	Percent of total	Query
26	307	13321	17.2471	tuition
27	291	13612	17.6239	spartan+trak
28	289	13901	17.9981	menus
29	273	14174	18.3515	uab
30	267	14441	18.6972	academic+calendar
31	265	14706	19.0403	im+east
32	262	14968	19.3796	rha
33	262	15230	19.7188	basketball
34	255	15485	20.0489	spartan+cash
35	246	15731	20.3674	loncapa
36	239	15970	20.6769	sparty+cash
37	239	16209	20.9863	transcripts
38	224	16433	21.2763	psychology
39	214	16647	21.5534	olin+health+center
40	206	16853	21.8201	cse+101

Tag analysis. The recent explosion in sites that employ folksonomic tagging (i.e., tags supplied by end users) means another useful indirect source of labels for you to learn from. In many of these sites, users' tags are publicly viewable. When you display them in aggregate, you'll find a collection of candidate labels that approximates the results of a free-listing exercise. Additionally, the data that comes from tag analysis can be used in much the same way as search-log analysis. Look for common terms, but also look for jargon, acronyms, and tone; even misspellings are useful if you're building a controlled vocabulary.

In the examples shown in Figures 6-14 and 6-15, you might be wondering how to develop labels for a new web-based iPod accessories store. To start, you might look at a popular folksonomic system like del.icio.us and see whether users have tagged a few common iPod accessories, and what terms they used. Let's try a pair of iPod accessories, a radio remote and a leather case. After searching both terms in del.icio.us, we found a variety of results, and chose those that had been bookmarked the most times.

Some of the tags are too broad to be particularly useful (e.g., "iPod" or "shopping"). But some will help you determine labels for categories; in the first example, "hardware" is more common than "media." Knowing that will clarify your category labeling. In the second example, you might choose "case" over the less popular "cases" as a product label.

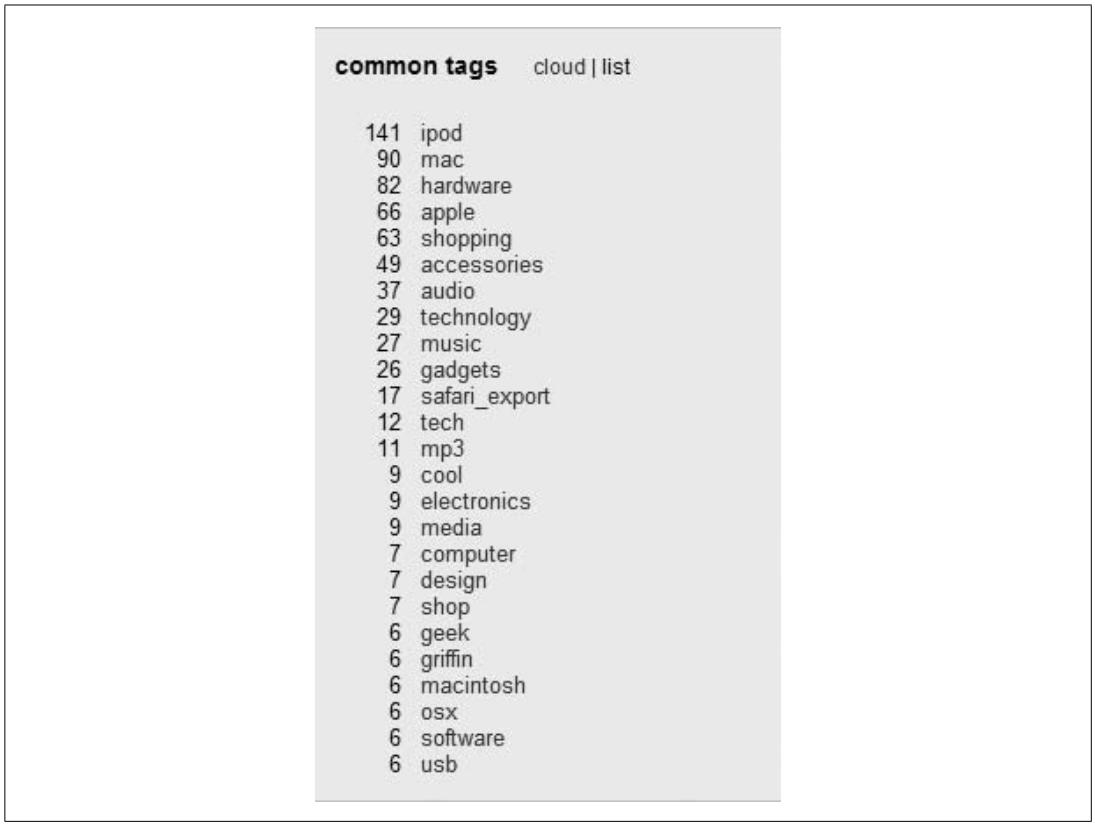


Figure 6-14. Griffin Technology's IPod Radio Remote (as tagged by 298 del.icio.us users)

Tuning and Tweaking

Your list of labels might be raw, coming straight from the content in your site, another site, your site's users, or your own ideas of what should work best. Or, it may come straight from a polished controlled vocabulary. In any case, it'll need some work to become an effective labeling system.

First, sort the list of terms alphabetically. If it's a long list (e.g., from a search log), you'll likely encounter some duplicates; remove these.

Then review the list for consistency of usage, punctuation, letter case, and so forth, considering some of the consistency issues discussed earlier in this chapter. For example, you'll remember that the label table drawn from the jetBlue web site had inconsistencies that were immediately obvious: sometimes there were periods after labels, sometimes there weren't; the usage of link labels versus the heading labels on the corresponding pages was inconsistent; and so on. This is a good time to resolve these inconsistencies and to establish conventions for punctuation and style.

Decisions about which terms to include in a labeling system need to be made in the context of how broad and how large a system is required. First, determine if the labeling system has obvious gaps. Does it encompass all the possibilities that your site may eventually need to include?

Figure 6-15. Vaja's leather products for PDAs (as tagged by 92 del.icio.us users)

If, for example, your e-commerce site currently allows users to search only a portion of your product database, ask yourself if eventually it might provide access to all products. If you're not certain, assume it will, and devise appropriate labels for the additional products.

If the site's labeling system is topical, try to anticipate the topics not yet covered by the site. You might be surprised to see that the addition of these "phantom" labels has a large impact on your labeling system, perhaps even requiring you to change its conventions. If you fail to perform this predictive exercise, you might learn the hard way that future content doesn't fit into your site because you're not sure how to label it, or it ends up in cop-out categories such as "Miscellaneous," "Other Info," and the classic "Stuff." Plan ahead so that labels you might add in the future don't throw off the current labeling system.

Of course, this planning should be balanced with an understanding of what your labeling system is there to accomplish *today*. If you try to create a labeling system that encompasses the whole of human knowledge (instead of the current and anticipated content of your web site), don't plan on doing anything else for the rest of your life. Keep your scope narrow and focused enough so that it can clearly address the requirements of your site's unique content, the special needs of its audiences, and the

business objective at hand, but be comprehensive within that well-defined scope. This is a difficult pursuit, to be sure; all balancing acts are. Consider it justification #64 for information architects—like yourself—to be paid well.

Finally, remember that the labeling system you launch will need to be tweaked and improved shortly thereafter. That's because labels represent a relationship between two things—users and content—that is constantly morphing. Stuck between two moving targets, your labeling system will also have to change. So be prepared to perform user tests, analyze search logs on a regular basis, and adjust your labeling system as necessary.

CHAPTER 7
Navigation Systems

Just wait, Gretel, until the moon rises, and then we shall see the crumbs of bread which I have strewn about; they will show us our way home again.
—Hansel and Gretel

What we'll cover:

- Balancing context and flexibility in web navigation
- Integrating global, local, and contextual navigation
- Supplemental navigation tools such as sitemaps, indexes, guides, wizards, and configurators
- Personalization, visualization, tag clouds, collaborative filtering, and social navigation

As our fairy tales suggest, getting lost is a bad thing. It is associated with confusion, frustration, anger, and fear. In response to this danger, humans have developed navigation tools to prevent us from getting lost and to help us find our way home. From bread crumbs to compass and astrolabe, to maps, street signs, and global positioning systems, people have demonstrated great ingenuity in the design and use of navigation tools and wayfinding strategies.

We use these tools to chart our course, to determine our position, and to find our way back. They provide a sense of context and comfort as we explore new places. Anyone who has driven through an unfamiliar city as darkness falls understands the importance these tools and strategies play in our lives.

On the Web, navigation is rarely a life or death issue. However, getting lost in a large web site can be confusing and frustrating. While a well-designed taxonomy may reduce the chances that users will become lost, complementary navigation tools are often needed to provide context and to allow for greater flexibility. Structure and organization are about building rooms. Navigation design is about adding doors and windows.

In this book, we have split navigation and searching into individual chapters. This chapter focuses on navigation systems that support browsing; the next chapter digs deep into searching systems that are clearly components of navigation. In fact, structure, organization, labeling, browsing, and searching systems all contribute toward effective navigation.

Types of Navigation Systems

Navigation systems are composed of several basic elements, or subsystems. First, we have the global, local, and contextual navigation systems that are integrated within the web pages themselves. These *embedded navigation systems* are typically wrapped around and infused within the content of the site. They provide both context and flexibility, helping users understand where they are and where they can go. These three major systems, shown in Figure 7-1, are generally necessary but not sufficient in themselves.

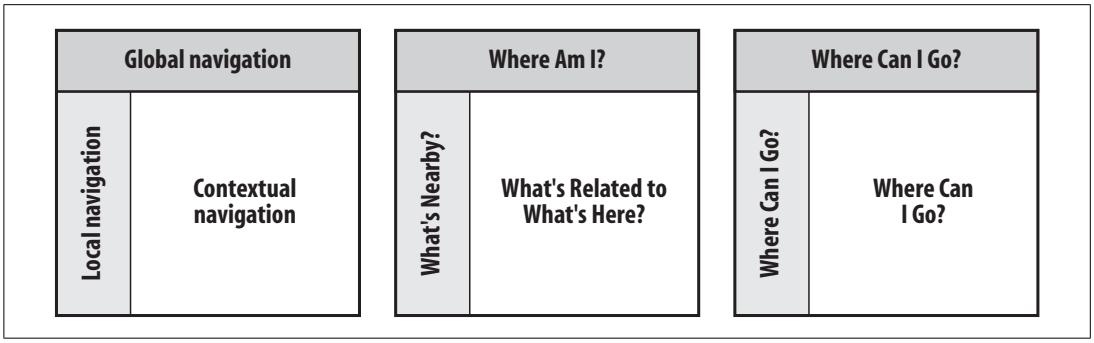


Figure 7-1. Global, local, and contextual embedded navigation systems

Second, we have *supplemental navigation systems* such as sitemaps, indexes, and guides that exist outside the content-bearing pages. These are shown in Figure 7-2.

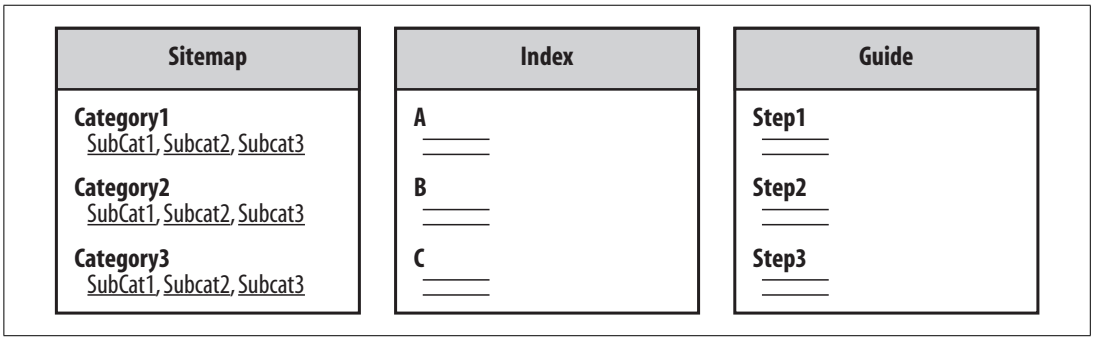


Figure 7-2. Supplemental navigation systems

Similar to search, these supplemental navigation systems provide different ways of accessing the same information. Sitemaps provide a bird's-eye view of the site. A to Z indexes allow direct access to content. And guides often feature linear navigation customized to a specific audience, task, or topic.

As we'll explain, each type of supplemental navigation system serves a unique purpose and is designed to fit within the broader framework of integrated searching and browsing systems.

Gray Matters

The design of navigation systems takes us deep into the gray area between information architecture, interaction design, information design, visual design, and usability engineering, all of which we might classify under the umbrella of user experience design.

As soon as we start talking about global, local, and contextual navigation, we find ourselves on the slippery slope that connects strategy, structure, design, and implementation. Does the local navigation bar work best at the top of the page, or is it better running down the left side? Should we use pull-downs, pop-ups, or cascading menus to reduce the required number of clicks? Will users ever notice gray links? Isn't it better to use the blue/red link color convention?

For better or for worse, information architects are often drawn into these debates, and we are sometimes responsible for making these decisions. We could try to draw a clear line in the sand, and argue that effective navigation is simply the manifestation of a well-organized system. Or we could abdicate responsibility and leave the interface to designers.

But we won't. In the real world, the boundaries are fuzzy and the lines get crossed every day. Information architects do design and designers do information architecture. And the best solutions often result from the biggest debates. While not always possible, interdisciplinary collaboration is the ideal, and collaboration works best when each of the experts understands something about the other areas of expertise.

So in this chapter, we roll up our sleeves, cross lines, step on toes, and get a little messy in the process. We tackle navigation design from the information architect's perspective. But before we drag you deep into this swampy gray matter, let us throw you a lifeline. In the Appendix, we have included references to a few truly excellent books that cover these topics from a variety of perspectives. We encourage you to read them all!

Browser Navigation Features

When designing a navigation system, it is important to consider the environment in which the system will exist. On the Web, people use web browsers such as Mozilla Firefox and Microsoft Internet Explorer to move around and view web sites. These browsers sport many built-in navigation features.

Open URL allows direct access to any page on a web site. Back and Forward provide a bidirectional backtracking capability. The History menu allows random access to pages visited during the current session, and Bookmark or Favorites enables users to save the location of specific pages for future reference. Web browsers also go beyond the Back button to support a "bread crumbs" feature by color-coding hypertext links. By default, unvisited hypertext links are one color and visited hypertext links are another. This feature helps users see where they have and haven't been and can help them to retrace their steps through a web site.

Finally, web browsers allow for a prospective view that can influence how users navigate. As the user passes the cursor over a hypertext link, the destination URL appears at the bottom of the browser window, hinting at the nature of that content. A good example is shown in Figure 7-3, where the cursor is positioned over "Pricing." The prospective view window at the bottom shows the URL of that page. If files and directories have been carefully labeled, prospective view gives the user context within the content hierarchy. If the hypertext link leads to a web site on another server, prospective view provides the user with basic information about this offsite destination.

Much research, analysis, and testing has been invested in the design of these browser-based navigation features. However, it is remarkable how frequently site designers unwittingly override or corrupt these navigation features. The most common design crimes are:

- Cluelessly modifying the visited/unvisited link colors
- Killing the Back button
- Crippling the Bookmark feature

Should you plan to commit any of these grave transgressions, make sure you've got a really good reason or an even better lawyer.

Building Context

With all navigation systems, before we can plot our course, we must locate our position. Whether we're visiting Yellowstone National Park or the Mall of America, the *You Are Here* mark on fixed-location maps is a familiar and valuable tool. Without that landmark, we must struggle to triangulate our current position using less dependable features such as street signs or nearby stores. The *You Are Here* indicator can be the difference between knowing where you stand and feeling completely lost.

When designing complex web sites, it is particularly important to provide context within the greater whole. Many contextual clues in the physical world do not exist on the Web. There are no natural landmarks, no north and south. Unlike physical travel, hypertextual navigation allows users to be transported right into the middle of an unfamiliar web site. Links from remote web pages and search engine results allow users to completely bypass the front door or main page of the web site. To further complicate matters, people often print web pages to read later or to pass along to a

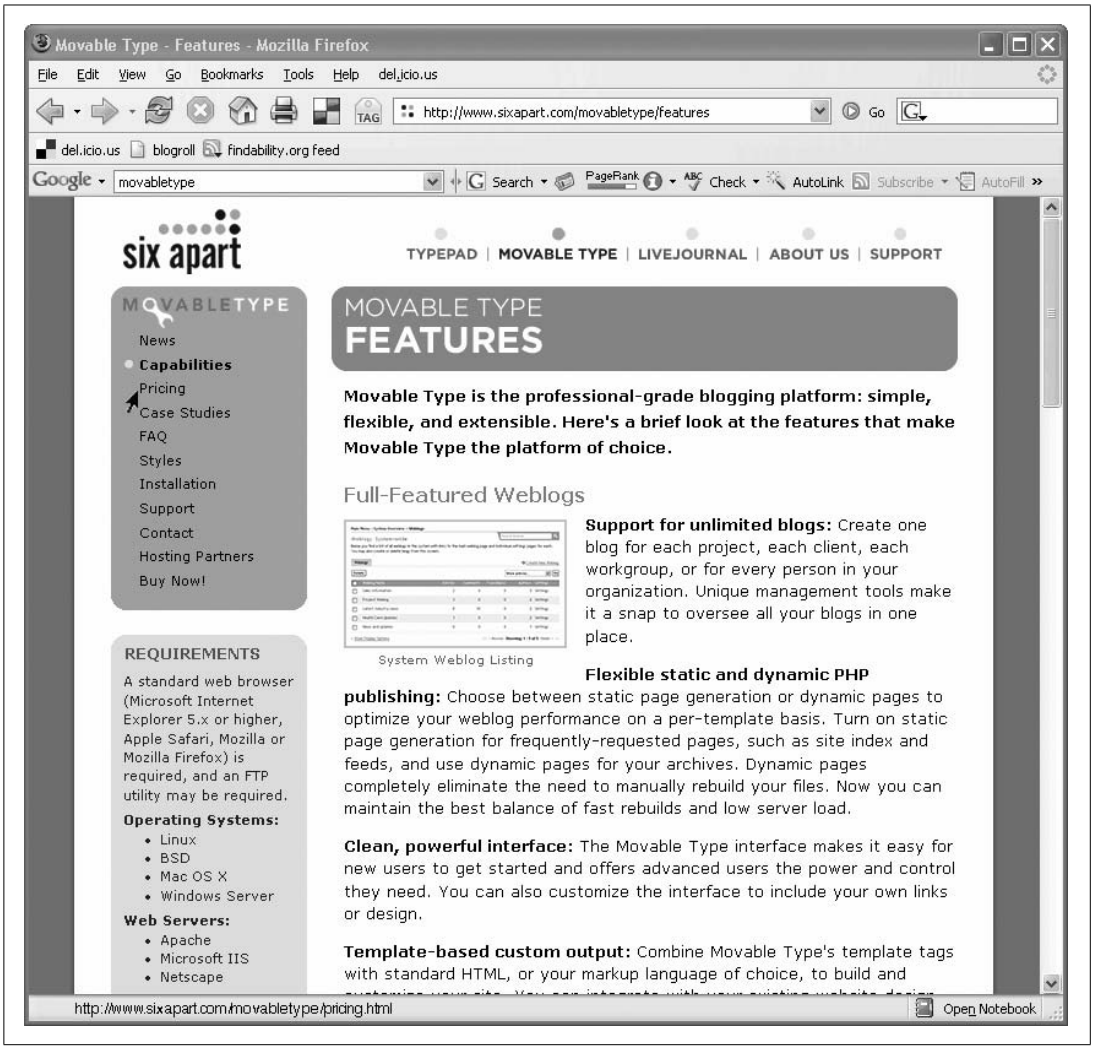


Figure 7-3. Prospective view is built into the browser

colleague, resulting in even more loss of context. For all these reasons, in the design of navigation systems, context is king!

You should always follow a few rules of thumb to ensure that your sites provide contextual clues. For example, users should always know which site they're in, even if they bypass the front door and enter through a search engine or a link to a subsidiary page. Extending the organization's name, logo, and graphic identity through all pages of the site is a fairly obvious way to accomplish this goal.

The navigation system should also present the structure of the information hierarchy in a clear and consistent manner, and indicate the user's current location, as shown in Figure 7-4. Wal-Mart's navigation system shows the user's location within the hierarchy with a variation of the You Are Here sign near the top of the page. This helps the user to build a mental model of the organization scheme, which facilitates navigation and helps her feel comfortable.

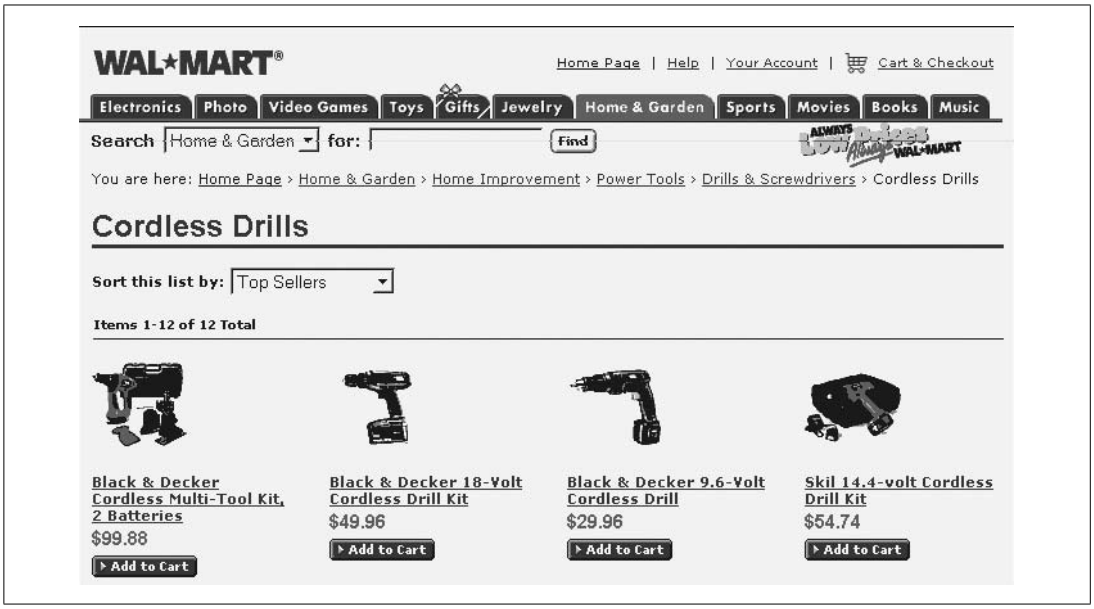


Figure 7-4. Wal-Mart's navigation system shows the user's location within the hierarchy

If you have an existing site, we suggest running a few users through a Navigation Stress Test.* Here are the basic steps as outlined by Keith Instone:

1. Ignore the home page and jump directly into the middle of the site.
2. For each random page, can you figure out where you are in relation to the rest of the site? What major section are you in? What is the parent page?
3. Can you tell where the page will lead you next? Are the links descriptive enough to give you a clue what each is about? Are the links different enough to help you choose one over another, depending on what you want to do?

By parachuting deep into the middle of the site, you will be able to push the limits of the navigation system and identify any opportunities for improvement.

Improving Flexibility

As we explained in Chapter 5, hierarchy is a familiar and powerful way of organizing information. In many cases, it makes sense for a hierarchy to form the foundation for organizing content in a web site. However, hierarchies can be limiting from a navigation perspective. If you have ever used the ancient information-browsing technology and precursor to the World Wide Web known as Gopher, you will understand the limitations of hierarchical navigation.† In Gopherspace, you were forced to move up

* Keith Instone popularized the notion of a Navigation Stress Test in his 1997 article, "Stress Test Your Site." See *http://user-experience.org/uefiles/navstress*.

† If you're too young to remember Gopher, consider the category/subcategory navigation on an iPod instead.

and down the tree structures of content hierarchies (see Figure 7-5). It was impractical to encourage or even allow jumps across branches (lateral navigation) or between multiple levels (vertical navigation) of a hierarchy.

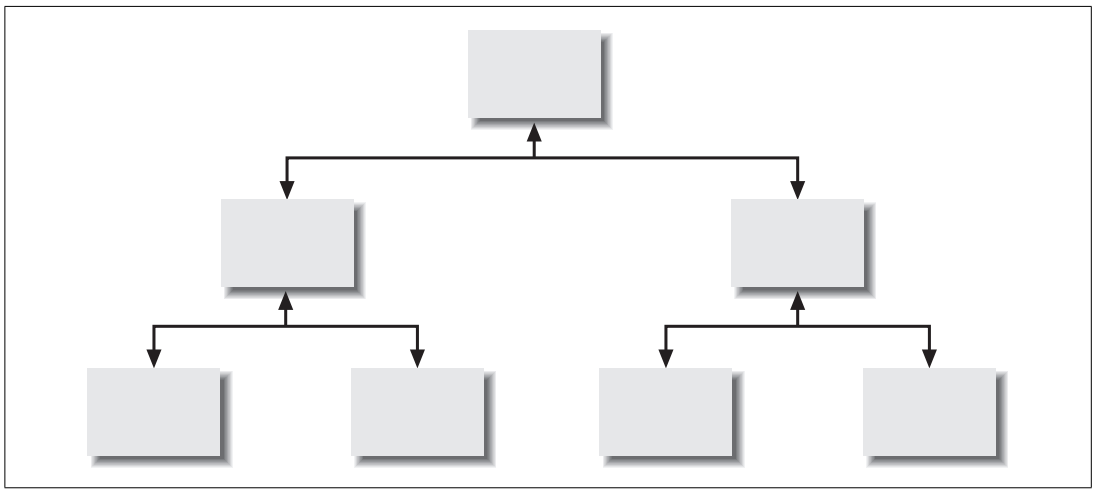


Figure 7-5. The pure hierarchy of Gopherspace

The Web's hypertextual capabilities removed these limitations, allowing tremendous freedom of navigation. Hypertext supports both lateral and vertical navigation. From any branch of the hierarchy, it is possible and often desirable to allow users to move laterally into other branches, to move vertically from one level to a higher level in that same branch, or to move all the way back to the main page of the web site. If the system is so enabled, users can get to anywhere from anywhere. However, as you can see in Figure 7-6, things can get confusing pretty quickly. It begins to look like an architecture designed by M.C. Escher.

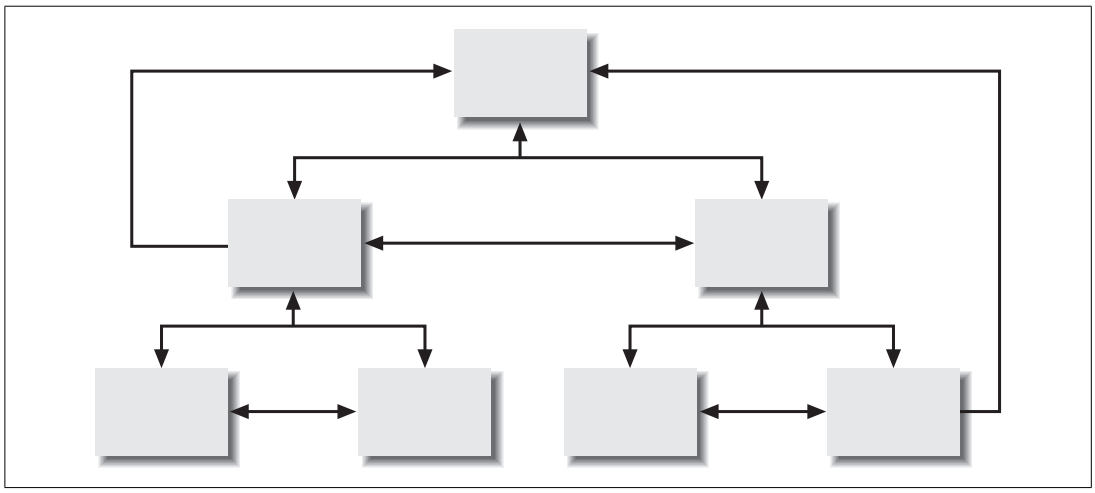


Figure 7-6. A hypertextual web can completely bypass the hierarchy

The trick to designing navigation systems is to balance the advantages of flexibility with the dangers of clutter. In a large, complex web site, a complete lack of lateral and vertical navigation aids can be very limiting. On the other hand, too many navigation aids can bury the hierarchy and overwhelm the user. Navigation systems should be designed with care to complement and reinforce the hierarchy by providing added context and flexibility.

Embedded Navigation Systems

Most large web sites include all three of the major embedded navigation systems we saw back in Figure 7-1. Global, local, and contextual navigation are extremely common on the Web. Each system solves specific problems and presents unique challenges. To design a successful site, it is essential to understand the nature of these systems and how they work together to provide context and flexibility.

Global (Site-Wide) Navigation Systems

By definition, a global navigation system is intended to be present on every page throughout a site. It is often implemented in the form of a navigation bar at the top of each page. These site-wide navigation systems allow direct access to key areas and functions, no matter where the user travels in the site's hierarchy.

Because global navigation bars are often the single consistent navigation element in the site, they have a huge impact on usability. Consequently, they should be subjected to intensive, iterative user-centered design and testing.

Global navigation bars come in all shapes and sizes. Consider the examples shown in Figure 7-7.

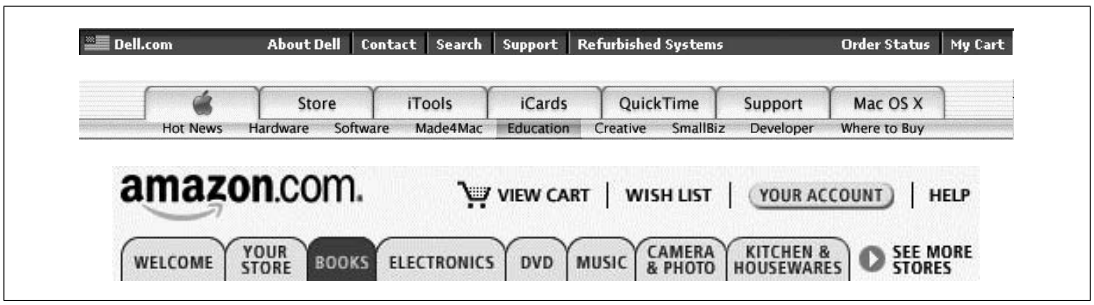


Figure 7-7. Global navigation bars from Dell, Apple, and Amazon

Most global navigation bars provide a link to the home page. Many provide a link to the search function. Some, like Apple's and Amazon's, reinforce the site's structure and provide contextual clues to identify the user's current location within the site. Others, like Dell's, have a simpler implementation and don't do either. This pushes the burden of providing context down to the local level and opens the door for

inconsistency and disorientation. Global navigation system design forces difficult decisions that must be informed by user needs and by the organization's goals, content, technology, and culture. One size does not fit all.

It's often not possible to identify the global navigation system from the main page of a web site. The main page is sometimes the sole exception to the omnipresence of the global navigation bar. In some cases, designers choose to show an expanded view of the global navigation system on the main page. In other cases, the main page presents a variety of navigation options, and it's impossible to tell which ones have been carried throughout the site without exploring further.

This is the case with Microsoft's main page, as shown in Figure 7-8. There are three distinct navigation bars, and it's unclear whether any or all of them constitute a global navigation system. However, check out a few subsidiary pages, and it becomes obvious that only one is truly global. The other two are simply the designer's way of exposing important dimensions of the site's structure on the main page.

Figure 7-8. Microsoft's main page navigation

As Figure 7-9 shows, Microsoft's global navigation bar is very compact, and for good reason. This global navigation bar represents a massive investment in screen real estate, occupying a prominent position on several hundred thousand pages. These pages exist within dozens of subsites that are "owned" by powerful business units and functions within Microsoft.

Figure 7-9. Microsoft's global navigation bar

Despite convincing user-centered design arguments, it is still not easy to drive consistency throughout the subsites of modern, decentralized organizations. Most large enterprises are lucky if they can get the company logo and a simple global navigation bar implemented on 80 percent of their pages.

Local Navigation Systems

In many web sites, the global navigation system is complemented by one or more local navigation systems that enable users to explore the immediate area. Some tightly controlled sites integrate global and local navigation into a consistent, unified system. For example, the *New York Times* web site presents a global navigation bar that expands to provide local navigation options for each category of news. A reader who selects "Business" sees different local navigation options than a reader who selects "Sports," but both sets of options are presented within the same navigational framework (see Figure 7-10).

Figure 7-10. Local navigation at nytimes.com

In contrast, large sites like Microsoft.com (Figure 7-11) often provide multiple local navigation systems that may have little in common with one another or with the global navigation system.

Figure 7-11. Local navigation at Microsoft.com

These local navigation systems and the content to which they provide access are often so different that these local areas are referred to as *subsites*,* or sites within sites. Subsites exist for two primary reasons. First, certain areas of content and functionality really do merit a unique navigation approach. Second, due to the decentralized nature of large organizations, different groups of people are often responsible for different content areas, and each group may decide to handle navigation differently.

In Microsoft's case, it makes sense to provide different ways to navigate the Jobs Area, the Support Database, and the Product Catalog. These local navigation systems are aligned with user needs and the local content. Unfortunately, there are

* The term *subsite* was coined by Jakob Nielsen in his 1996 article "The Rise of the Subsite" to describe a collection of web pages within a larger site that invite a common style and shared navigation mechanism unique to those pages. See *www.useit.com/alertbox/9609.html*.

many bad examples on the Web where the variation between local navigation systems is simply a result of multiple design groups choosing to run in different directions. Many organizations are still struggling with the question of how much central control to exercise over the look and feel of their local navigation systems. Grappling with these local navigation issues can make global navigation systems look easy.

Contextual Navigation

Some relationships don't fit neatly into the structured categories of global and local navigation. This demands the creation of *contextual* navigation links specific to a particular page, document, or object. On an e-commerce site, these "See Also" links can point users to related products and services. On an educational site, they might point to similar articles or related topics.

In this way, contextual navigation supports associative learning. Users learn by exploring the relationships you define between items. They might learn about useful products they didn't know about, or become interested in a subject they'd never considered before. Contextual navigation allows you to create a web of connective tissue that benefits users and the organization.

The actual definition of these links is often more editorial than architectural. Typically an author, editor, or subject matter expert will determine appropriate links once the content is placed into the architectural framework of the web site. In practice, this usually involves representing words or phrases within sentences or paragraphs (i.e., prose) as embedded or "inline" hypertext links. A page from Stanford University's site, shown in Figure 7-12, provides an example of carefully chosen inline contextual navigation links.

This approach can be problematic if these contextual links are critical to the content, since usability testing shows that users often tend to scan pages so quickly they miss or ignore these less conspicuous links. For this reason, you may want to design a system that provides a specific area of the page or a visual convention for contextual links. As you can see in Figure 7-13, REI designs contextual navigation links to related products into the layout of each page. Moderation is the primary rule of thumb for guiding the creation of these links. Used sparingly (as in this example), contextual links can complement the existing navigation systems by adding one more degree of flexibility. Used in excess, they can add clutter and confusion. Content authors have the option to replace or complement the embedded links with external links that are easier for the user to see.

The approach used on each page should be determined by the nature and importance of the contextual links. For noncritical links provided as a point of interest, inline links can be a useful but unobtrusive solution.

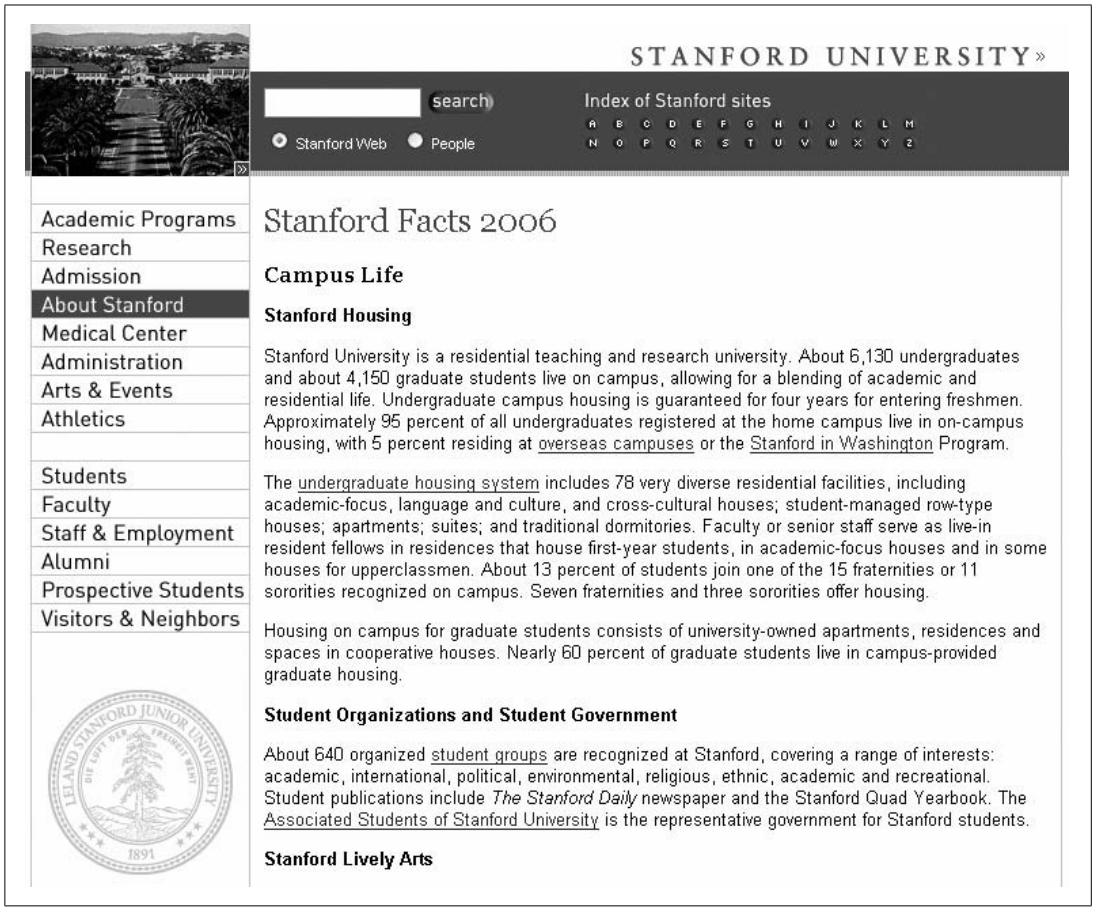


Figure 7-12. Inline contextual navigation links

When designing a contextual navigation system, imagine that every page on the site is a main page or portal in its own right. Once a user has identified a particular product or document, the rest of the site fades into the background. This page is now his interface. Where might he want to go from here? Consider the REI example. What additional information will the customer want before making a buying decision? What other products might he want to buy? Contextual navigation provides a real opportunity to cross-sell, up-sell, build brand, and provide customer value. Because these associative relationships are so important, we'll revisit this topic in Chapter 9.

Implementing Embedded Navigation

The constant challenge in navigation system design is to balance the flexibility of movement with the danger of overwhelming the user with too many options. One key to success is simply recognizing that global, local, and contextual navigation elements exist together on most pages (consider the representation of a web page shown in Figure 7-14). When integrated effectively, they can complement one another.

Figure 7-13. External contextual navigation links

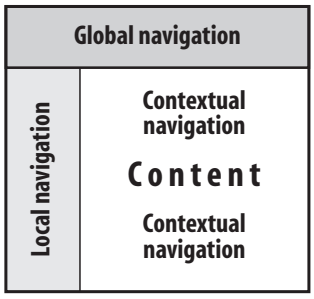


Figure 7-14. Navigation can drown out the content

But when designed independently, the three systems can combine to monopolize a great deal of screen real estate. Alone, they may each be manageable, but together on one page, the variety of options may overwhelm the user and drown out the content. In some cases, you may need to revisit the number of options within each navigation bar. In others, the problem may be minimized through careful design and layout.

In its simplest form, a navigation bar is a distinct collection of hypertext links that connect a series of pages, enabling movement among them. They can support global, local, and contextual navigation. You can implement navigation in all sorts of ways, using text or graphics, pull-downs, pop-ups, rollovers, cascading menus, and so on. Many of these implementation decisions fall primarily within the realms of interaction design and technical performance rather than information architecture, but let's trespass briefly and hit a few highlights.

For example, is it better to create textual or graphical navigation bars? Well, graphic navigation bars tend to look nicer, but can slow down the page-loading speed and are more expensive to design and maintain. If you use graphic navigation bars, you need to be sensitive to the needs of users with low bandwidth connections and text-only browsers. People who are blind and people using wireless mobile devices are two important audiences to consider. Appropriate use of the <ALT> attribute to define replacement text for the image will ensure that your site supports navigation for these users.

And where do the navigation bars belong on the page? It has become convention to place the global navigation bar along the top of the page and the local navigation bar along the lefthand side. However, all sorts of permutations can be successful. Just make sure you do lots of user testing, particularly if you deviate from convention.

What about textual labels versus icons? Textual labels are the easiest to create and most clearly indicate the contents of each option. Icons, on the other hand, are relatively difficult to create and are often ambiguous. It's difficult to represent abstract concepts through images. A picture may say a thousand words, but often they're the wrong words—particularly when you're communicating to a global audience.

Icons can successfully be used to complement the textual labels, however. Since repeat users may become so familiar with the icons that they no longer need to read the textual labels, icons can be useful in facilitating rapid menu selection. In Figure 7-15, Scott McCloud combines text and images to create a global navigation system that balances form and function. But can you guess what lies behind icons *b* through *e*? On this comic creator's web site, the mystery icons provoke curiosity and create a playful experience. On a business web site, they would simply be frustrating.

How about the increasingly common use of DHTML and JavaScript rollovers to show the navigation options behind a category or menu option (as shown in Figure 7-16)? Well, it depends. On one hand, this prospective view on steroids can make valuable

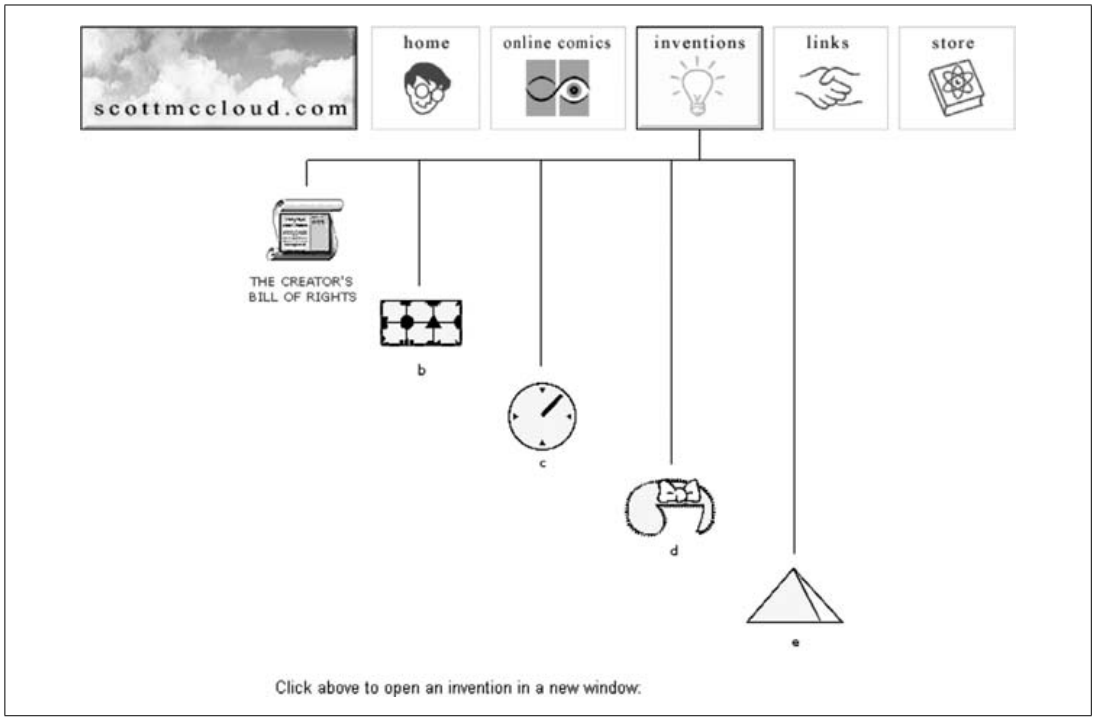


Figure 7-15. Navigation with integrated text and images

use of limited screen real estate, enhancing the scent of information and often reducing the number of pages and clicks, while simultaneously adding a dynamic, interactive feel to the web site. On the other hand, rollover navigation can be difficult to do well. Usability and accessibility often suffer due to poor design and implementation. Also, the use of rollover navigation is no substitute for the careful selection of the omnipresent major categories and labels, which lend themselves to rapid visual scanning. You can't expect the user to "mine sweep" her mouse cursor over every option.

And finally, what about frames? In the 1990s, designers went a little crazy with frames, implementing navigation bars and banner advertisements alike inside non-scrollable panes. We don't see too many frames these days, and that's a very good thing. Even beyond the technical design and performance problems, frames tend to cripple usability. After all, the Web is built upon a model of pages, each of which has a unique address or URL. Users are familiar with the concept of pages. Frames confuse this issue by slicing up pages into independent panes of content. By disrupting the page model, the use of frames frequently disables important browser navigation features such as bookmarking, visited and unvisited link discrimination, and history lists. Frames can also confuse users who are trying to perform simple tasks such as using the Back button, reloading a page, and printing a page. While web browsers have improved in their ability to handle frames, they can't remove the confusion caused by violating the page paradigm.

Figure 7-16. Audi's rollover navigation

Supplemental Navigation Systems

Supplemental navigation systems (shown back in Figure 7-2) include sitemaps, indexes, and guides. These are external to the basic hierarchy of a web site and provide complementary ways of finding content and completing tasks. Search also belongs to the supplemental navigation family but is so important that we've dedicated all of Chapter 8 to it.

Supplemental navigation systems can be critical factors for ensuring usability and findability within large web sites. However, they're often not given the care and feeding they deserve. Many site owners still labor under the misconception that if they could only get the taxonomy right, all users and all user needs would be addressed.

Usability pundits feed this fantasy by preaching the gospel of simplicity: users don't want to make choices, and they resort to sitemaps, indexes, guides, and search only when the taxonomy fails them.

Both statements are theoretically true but miss the point that the taxonomy and the embedded navigation systems will always fail for a significant percentage of users and tasks. You can count on this like death and taxes. Supplemental navigation systems give users an emergency backup. Do you really want to drive without a seatbelt?

Sitemaps

In a book or magazine, the table of contents presents the top few levels of the information hierarchy. It shows the organization structure for the printed work and supports random as well as linear access to the content through the use of chapter and page numbers. In contrast, a print map helps us navigate through physical space, whether we're driving through a network of streets and highways or trying to find our terminal in a busy airport.

In the early days of the Web, the terms "sitemap" and "table of contents" were used interchangeably. Of course, we librarians thought the TOC was a better metaphor, but *sitemap* sounds sexier and less hierarchical, so it has become the de facto standard.

A typical sitemap (Figure 7-17) presents the top few levels of the information hierarchy. It provides a broad view of the content in the web site and facilitates random access to segmented portions of that content. A sitemap can employ graphical or text-based links to provide the user with direct access to pages of the site.

A sitemap is most natural for web sites that lend themselves to hierarchical organization. If the architecture is not strongly hierarchical, an index or alternate visual representation may be better. You should also consider the web site's size when deciding whether to employ a sitemap. For a small site with only two or three hierarchical levels, a sitemap may be unnecessary.

The design of a sitemap significantly affects its usability. When working with a graphic designer, make sure he understands the following rules of thumb:

1. Reinforce the information hierarchy so the user becomes increasingly familiar with how the content is organized.
2. Facilitate fast, direct access to the contents of the site for those users who know what they want.
3. Avoid overwhelming the user with too much information. The goal is to help, not scare, the user.

Finally, it's worth noting that sitemaps are also useful from a search engine optimization perspective, since they point search engine spiders directly to important pages throughout the web site.

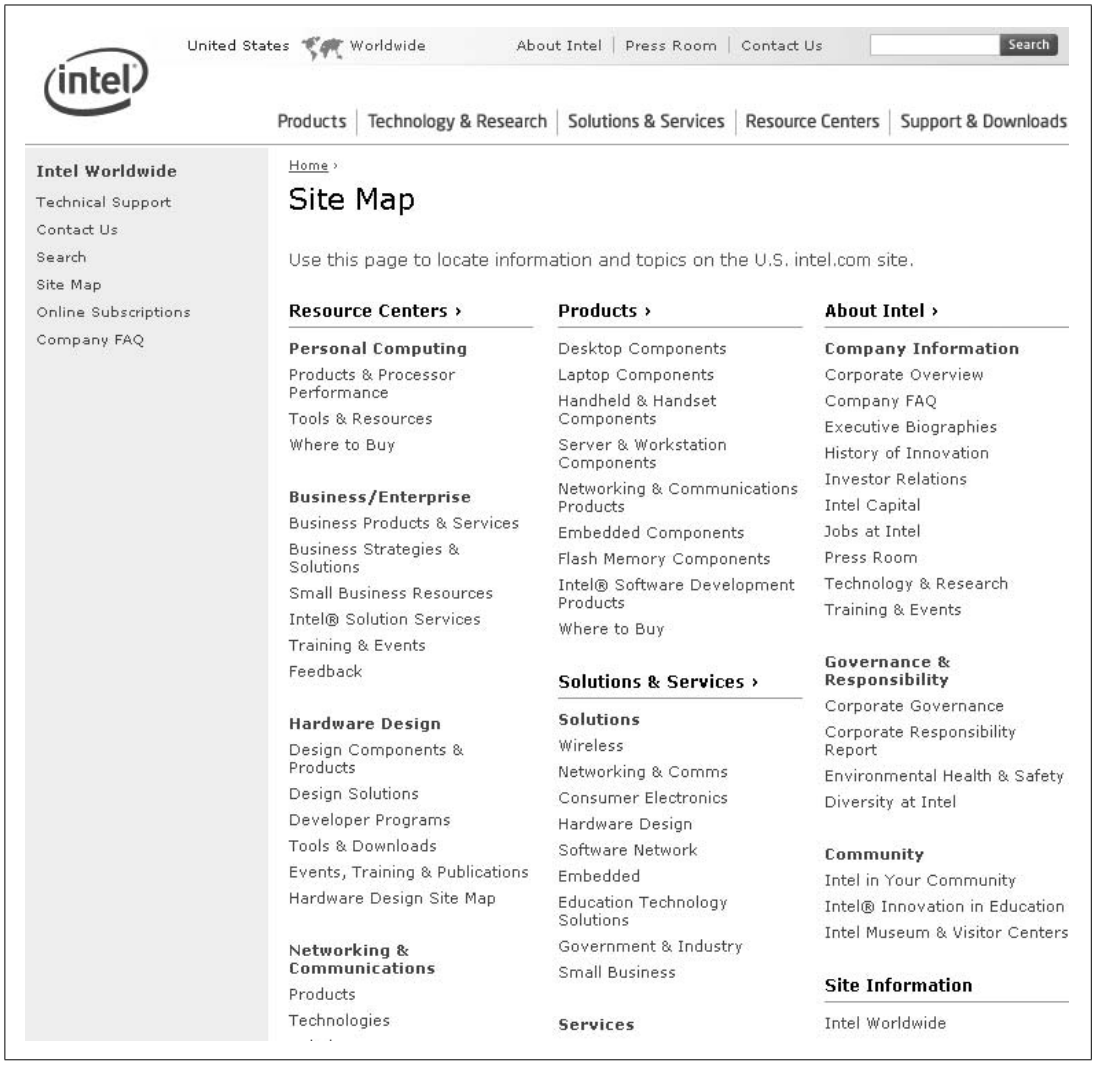


Figure 7-17. Intel's sitemap

Site Indexes

Similar to the back-of-book index found in many print materials, a web-based index presents keywords or phrases alphabetically, without representing the hierarchy. Unlike a table of contents, indexes are relatively flat, presenting only one or two levels of depth. Therefore, indexes work well for users who already know the name of the item they are looking for. A quick scan of the alphabetical listing will get them where they want to go; there's no need for them to understand where you've placed that item within your hierarchy. In Figure 7-18, AOL presents a simple but useful alphabetical site index. Handcrafted links within the index lead directly to destination pages.

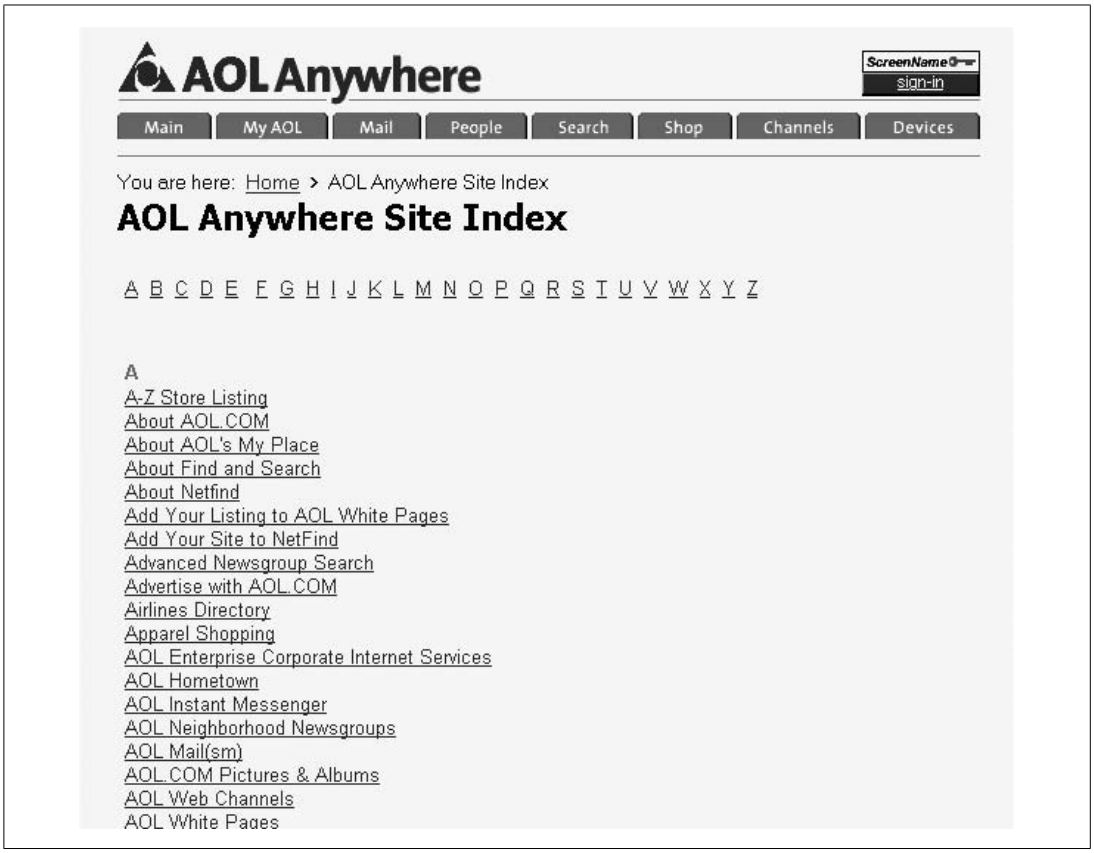


Figure 7-18. AOL's simple but useful alphabetical site index

Large, complex web sites often require both a sitemap and a site index (and a search capability, for that matter). The sitemap reinforces the hierarchy and encourages exploration, while the site index bypasses the hierarchy and facilitates known-item finding. For small web sites, a site index alone may be sufficient. On Usable Web (see Figure 7-19), Keith Instone has made his site index even more useful by indicating the number of items behind each link.

A major challenge in indexing a web site involves the level of granularity. Do you index web pages? Do you index individual paragraphs or concepts that are presented on web pages? Or do you index collections of web pages? In many cases, the answer may be all of the above. Perhaps a more valuable question is: what terms are users going to look for? The answers should guide the index design. To find those answers, you need to know your audience and understand their needs. You can learn more about the terms people will look for by analyzing search logs and conducting user research.

There are two very different ways to create a site index. For small web sites, you can simply create the index manually, using your knowledge of the full collection of content to inform decisions about which links to include. This centralized, manual

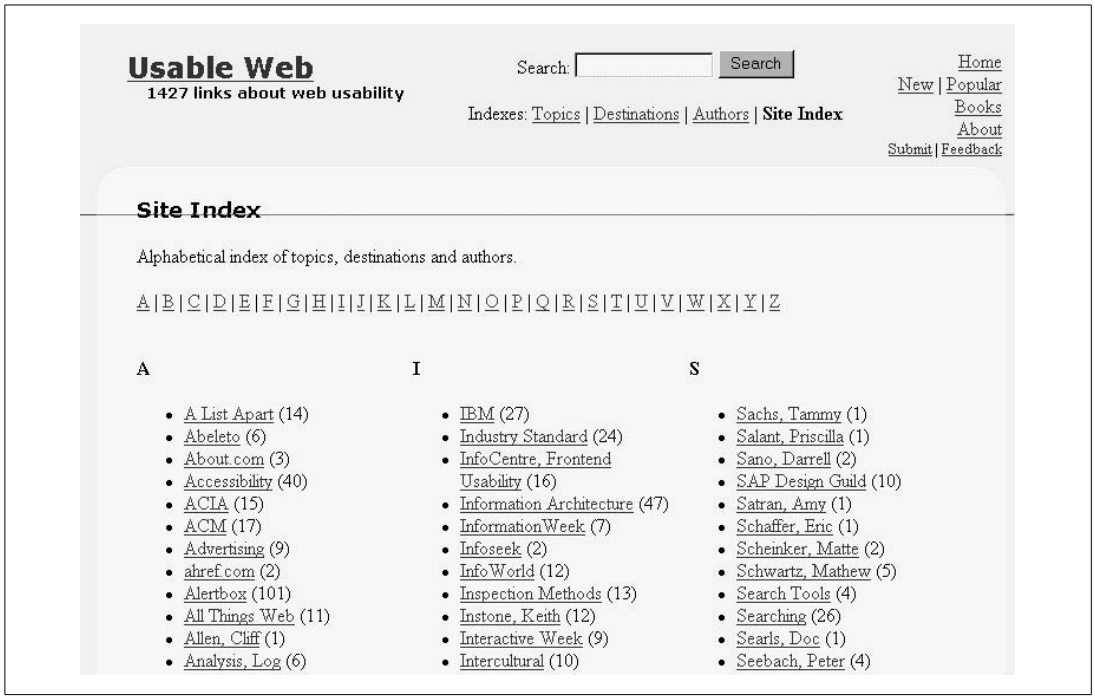


Figure 7-19. Usable Web's highly usable site index

approach results in a one-step index such as the one in Figure 7-18. Another example is shown in Figure 7-20, where Vanguard's dynamically generated two-step site index features term rotation and see/see-also references.

In contrast, on a large site with distributed content management, it may make sense to use controlled vocabulary indexing at the document level to drive automatic generation of the site index. Since many controlled vocabulary terms will be applied to more than one document, this type of index must allow for a two-step process. First the user selects the term from the index, and then selects from a list of documents indexed with that term.

A useful trick in designing an index involves *term rotation*, also known as *permutation*. A permuted index rotates the words in a phrase so that users can find the phrase in two places in the alphabetical sequence. For example, in the Vanguard index, users will find listings for both "refund, IRS" and "IRS refund." This supports the varied ways in which people look for information. Term rotation should be applied selectively. You need to balance the probability of users seeking a particular term with the annoyance of cluttering the index with too many permutations. For example, it would probably not make sense in an event calendar to present Sunday (Schedule) as well as Schedule (Sunday). If you have the time and budget to conduct focus groups or user testing, that's great. If not, you'll have to fall back on common sense.

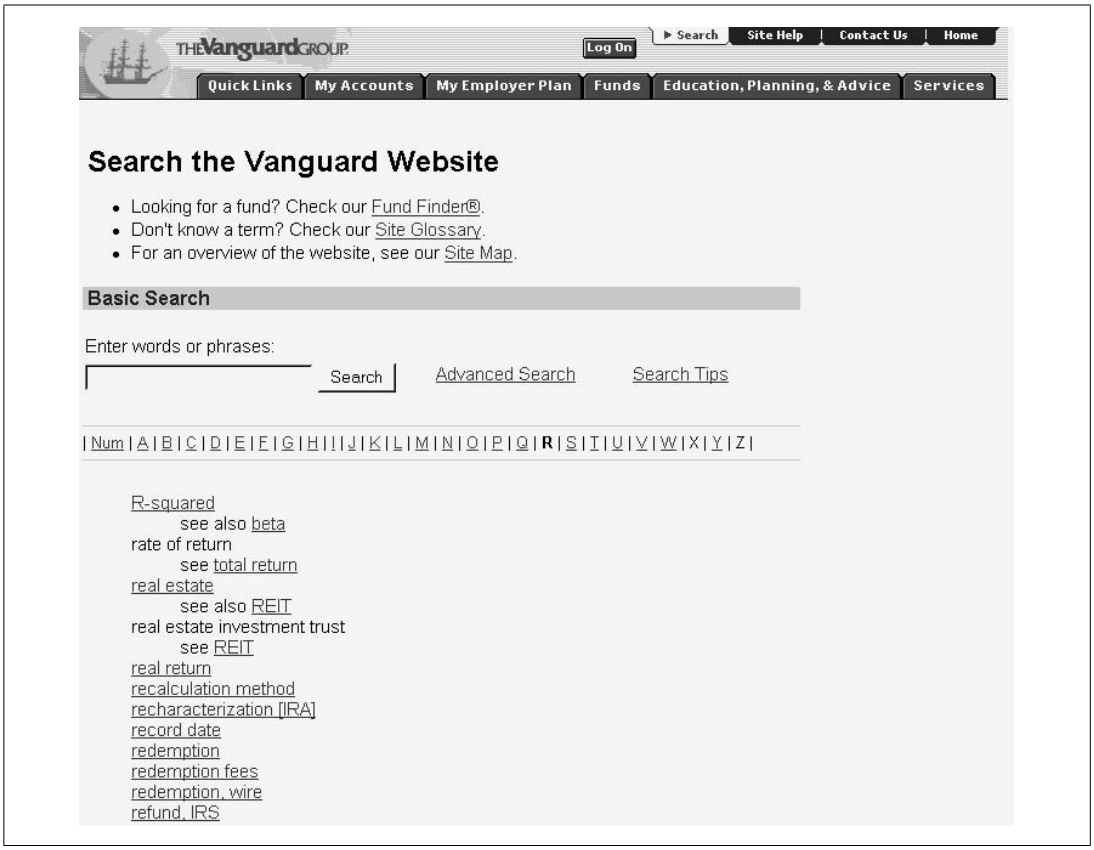


Figure 7-20. Vanguard's dynamically generated site index

Guides

Guides can take several forms, including guided tours, tutorials, and micro-portals focused around a specific audience, topic, or task. In each case, guides supplement the existing means of navigating and understanding site content.

Guides often serve as useful tools for introducing new users to the content and functionality of a web site. They can also be valuable marketing tools for restricted-access web sites (such as online publications that charge subscription fees), enabling you to show potential customers what they will get for their money. And, they can be valuable internally, providing an opportunity to showcase key features of a redesigned site to colleagues, managers, and venture capitalists.

Guides typically feature linear navigation (new users want to be guided, not thrown in), but hypertextual navigation should also be available to provide additional flexibility. Screenshots of major pages should be combined with narrative text that explains what can be found in each area of the web site.

The *Wall Street Journal*, shown in Figure 7-21, uses a guided tour to showcase navigation and editorial features of the web site that, for the most part, are accessible only to subscribers.

Figure 7-21. The Wall Street Journal's guided tour

Rules of thumb for designing guides include:

1. The guide should be short.
2. At any point, the user should be able to exit the guide.
3. Navigation (Previous, Home, Next) should be located in the same spot on every page so that users can easily step back and forth through the guide.
4. The guide should be designed to answer questions.
5. Screenshots should be crisp, clear, and optimized, with enlarged details of key features.
6. If the guide includes more than a few pages, it may need its own table of contents.

Remember that a guide is intended as an introduction for new users and as a marketing opportunity for the web site. Many people may never use it, and few people will use it more than once. You should balance the inevitable big ideas about how to create an exciting, dynamic, interactive guide with the fact that it will not play a central role in the day-to-day use of the web site.

Wizards and Configurators

Though they could be considered a special class of guide, wizards that help users to configure products or navigate complex decision trees deserve separate highlighting. Sophisticated configurators, like the Mini Cooper example shown in Figure 7-22, blur the lines between software application and web site.

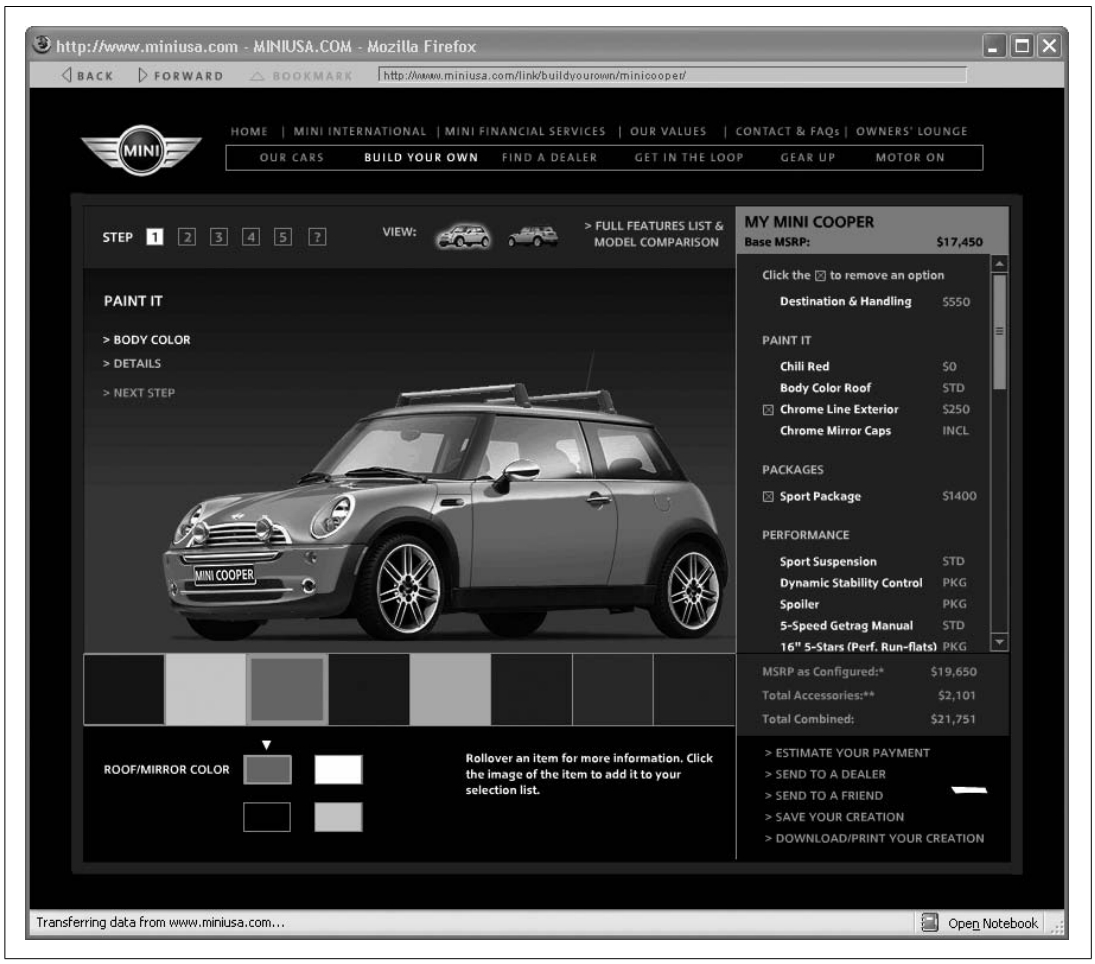


Figure 7-22. The Mini Cooper configurator

Mini successfully combines a rich suite of navigation options without causing confusion. The user can move through a linear process or jump back and forth between steps, and the site's global navigation is always present, providing context and possible next steps.

Search

As we noted earlier, the searching system is a central part of supplemental navigation. Search is a favorite tool of users because it puts them in the driver's seat, allowing them to use their own keyword terms to look for information. Search also

enables a tremendous level of specificity. Users can search the content for a particular phrase (e.g., "socially translucent systems failure") that is unlikely to be represented in a sitemap or site index.

However, the ambiguity of language causes huge problems with most search experiences. Users, authors, and information architects all use different words for the same things. Because the design of effective search systems is so important and so complex, we've devoted all of the following chapter to the topic.

Advanced Navigation Approaches

So far, we've focused attention on the bread-and-butter components of navigation systems, the elements that form the foundation of useful, usable web sites. Good navigation design is really important and really hard. Only after you've mastered the integration of these fundamental building blocks should you dare wander into the minefield of advanced navigation.

Personalization and Customization

Personalization involves serving up tailored pages to the user based upon a model of the behavior, needs, or preferences of that individual. In contrast, *customization* involves giving the user direct control over some combination of presentation, navigation, and content options. In short, with personalization, we guess what the user wants, and with customization, the user tells us what he wants.

Both personalization and customization can be used to refine or supplement existing navigation systems. Unfortunately, however, both have been hyped by consultants and software vendors as the solution to all navigation problems. The reality is that personalization and customization:

- Typically play important but limited roles
- Require a solid foundation of structure and organization
- Are really difficult to do well

Personalization has preoccupied marketing folks in recent years, partly due to the influential book by Don Peppers and Martha Rogers, *The One to One Future* (Doubleday). On a web site, you might use demographic data (e.g., age, sex, income level, zip code) and previous purchasing behavior to make educated guesses about which products to feature in the contextual navigation system during a customer's next visit. On an intranet, you might use role and job function as a basis for filtering views of news and e-service applications; for example, personalization is essential for controlling access to human-resource applications involving compensation and benefits.

Amazon is the most cited example of successful personalization, and some of the things it's done are truly valuable. It's nice that Amazon remembers our names, and it's great that it remembers our address and credit card information. It's when Amazon

starts trying to recommend books based on past purchases that the system breaks down (see Figure 7-23). In this example, Peter already owns two of the top three recommended books, but the system doesn't know this because he didn't purchase them from Amazon. And this ignorance is not the exception but the rule. Because we don't have time to teach our systems, or because we prefer to maintain our privacy, we often don't share enough information to drive effective personalization. In addition, in many cases, it's really hard to guess what people will want to do or learn or buy tomorrow. As they say in the financial world, past performance is no guarantee of future results. In short, personalization works really well in limited contexts, but fails when you try to expand it to drive the entire user experience.

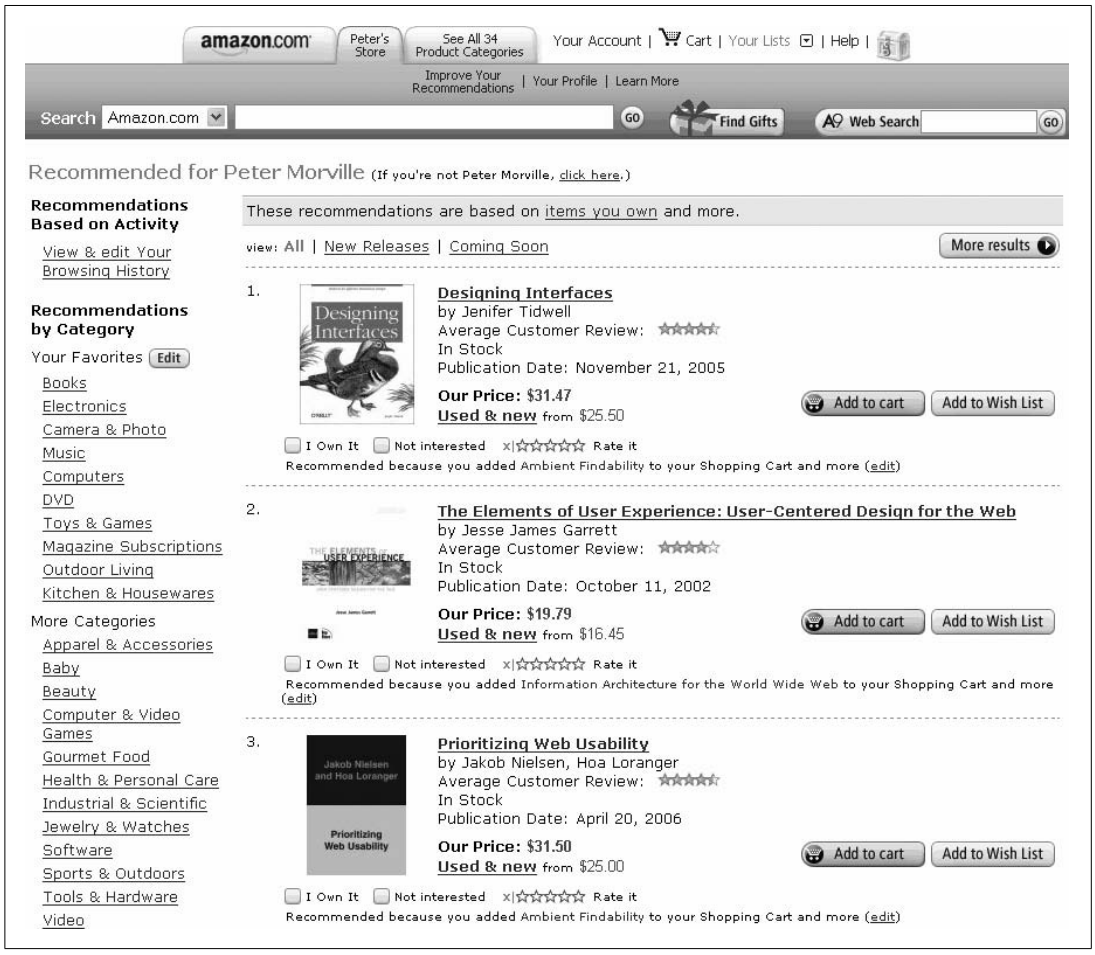


Figure 7-23. Amazon's personalized recommendations

Customization introduces a similar set of promises and perils. The idea of giving users control and thereby alleviating some of the pressures on design is obviously very compelling. And customization can sometimes deliver great value. My Yahoo! (Figure 7-24) and more recently, MySpace, are flagship examples and provide all sorts of customization capabilities, which many users take full advantage of—for better or for worse.

Figure 7-24. Customization at My Yahoo!

The problem with customization is that most people don't want to spend much (if any) time customizing, and will do this work only on a small handful of sites that are most important to them. Since corporate intranets have a captive audience of repeat visitors, customization has a much better chance of being used there than it does on most public web sites.

However, there's another problem. Even users themselves don't always know what they will want to know or do tomorrow. Customization works great for tracking the sports scores of your favorite baseball team or monitoring the value of stocks you own, but not so well when it comes to broader news and research needs. One day you want to know the results of the French elections; the next day you want to know when dogs were first domesticated. Do you really know what you might need next month?

Visualization

Since the advent of the Web, people have struggled to create useful tools that enable users to navigate in a more visual way. First came the metaphor-driven attempts to display online museums, libraries, shopping malls, and other web sites as physical places. Then came the dynamic, fly-through "sitemaps" that tried to show relationships

between pages on a web site. Both looked very cool and stretched our imaginations. But neither proved to be very useful. Even today, high-profile companies such as Groxis continue to explore the potential of visualization for navigation. Grokker, its enterprise search product, allows you to create visual navigation experiences for users (see Figure 7-25). It's worth keeping an eye on these experiments, but we remain skeptical that these approaches will prove useful for mainstream search and navigation.

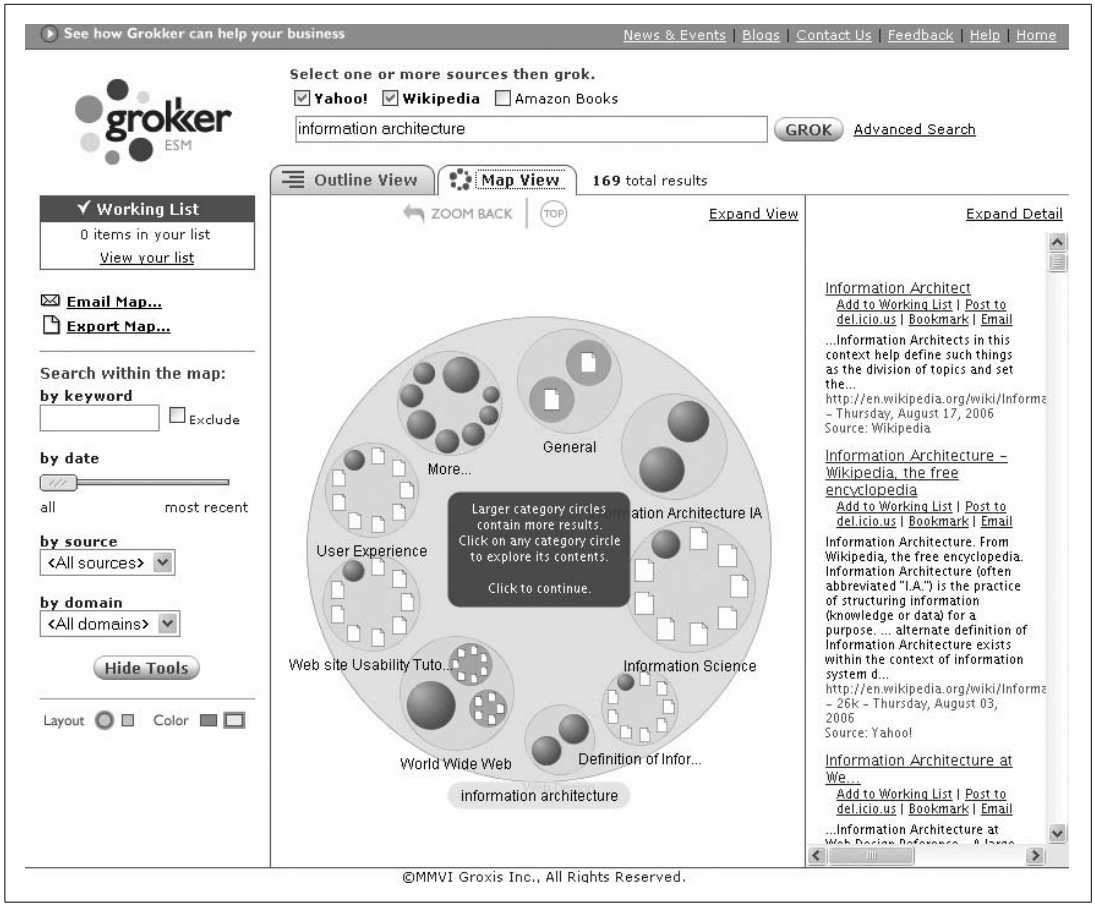


Figure 7-25. Grokker's visual search results

Social Navigation

On a more positive note, *social navigation*, built on the premise that value for the individual user can be derived from observing the actions of other users, continues to hold great promise and is already on the fast track to mainstream adoption. Simple examples include lists of most popular resources, such as the *New York Times*' Most Popular (see Figure 7-26).

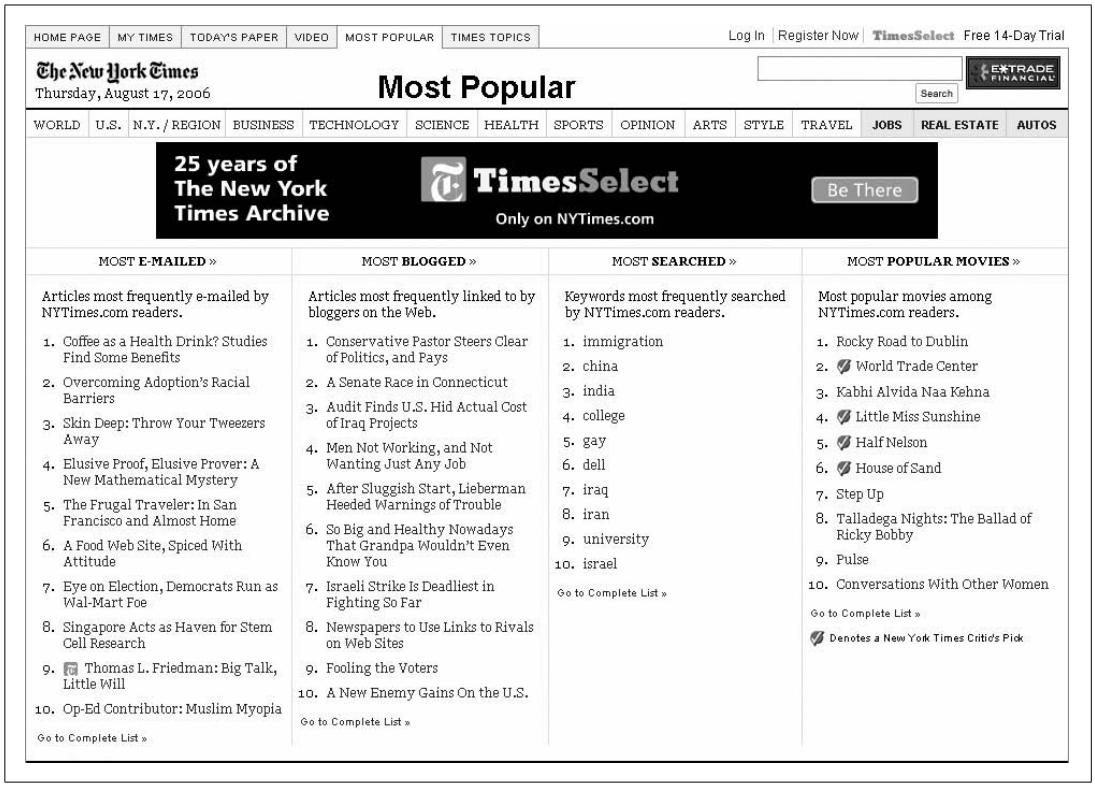


Figure 7-26. Most Popular at the New York Times

More sophisticated examples include Amazon's collaborative filtering (see Figure 7-27), Epinions' recommendation engine (see *http://www.epinions.com*), and Flickr's beloved tag clouds, shown in Figure 7-28, which use font size to show tag popularity. Perhaps there is a future for visualization after all.

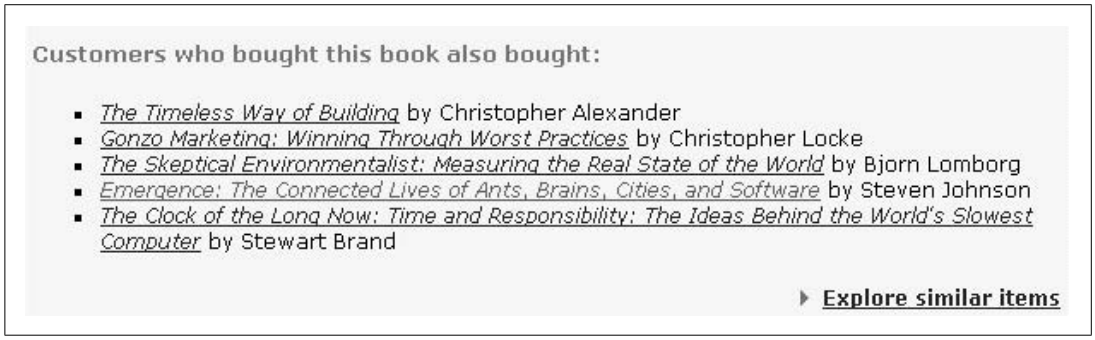


Figure 7-27. Amazon's collaborative filtering

While most companies aren't yet employing social navigation approaches on their web sites and intranets, we expect the practice to become increasingly common in the coming years. At a minimum, companies will find ways to unlock the value currently

trapped in their search logs, usage statistics, and customer databases to drive more effective contextual navigation. We also hope to see more ambitious solutions that tap this feedback loop between design and behavior, creating adaptive navigation systems that significantly advance the usability of our web sites and intranets.

Figure 7-28. Flickr's tag clouds

In the past several years, the design of navigation systems has improved in a rapid and highly visible manner. If you need convincing, just check out a few sites from the mid-90s using the Internet Archive's Wayback Machine (*http://www.archive.org/*). Let's hope we can keep up the pace, because there's still a long way to go.