safaribooksonline.com/library/view/information-architecture-4th/9781491913529/ch06.html


# Chapter 6. Organization Systems

> *The beginning of all understanding is classification.*
>
> *Hayden White*

In this chapter, 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
- 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.

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

Digital media provide us with wonderfully flexible environments in which to organize. We can apply multiple organization systems to the same content and escape the physical limitations of the analog world. So why are many digital products so difficult to navigate? Why can't the people who design these products make it easy to find information? These common questions focus attention on the very real problem 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 Cataloguing Rules. They classified, cataloged, and helped you find the information you needed.

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 6-1).

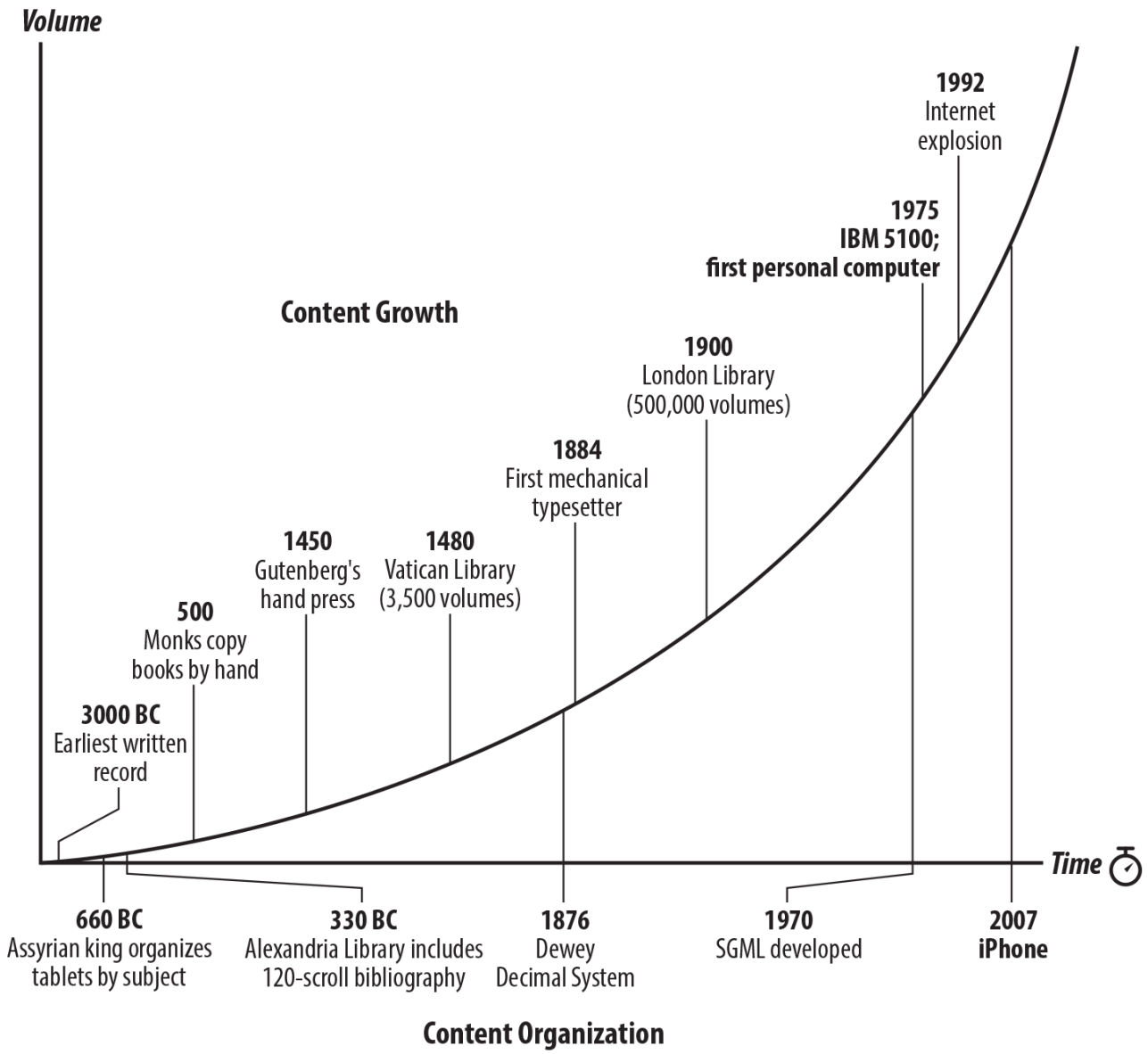



Figure 6-1. Content growth drives innovation

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 living in 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 made of language, and language is ambiguous: words are capable of being understood in more than one way. Think about 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 7 to learn more about labeling.)

It gets worse. Not only do we need to agree on the labels and their definitions, but 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?[1] And of course, this assumes that the user reads English to begin with—an unrealistic assumption in our increasingly multicultural digital media.

If we have such problems classifying the common tomato, consider the challenges involved in classifying website 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," "religion," "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 it works fairly well.

Most digital information environments, on the other hand, are highly heterogeneous in many respects. For example, websites often provide access to documents and their components at varying levels of *granularity*. A site might present articles and journals and journal databases side by side. Links might lead to pages, sections of pages, or other websites. And websites 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 website 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 information environments makes it difficult to impose any single structured organization system on the content. It usually doesn't make sense to classify 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 (JPG, PNG, etc.) and resolution (1024 × 768, 1280 × 800, etc.). It is difficult and often misguided to attempt a one-size-fits-all approach to the organization of heterogeneous 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 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!"[2]

The fact is that labeling and organization systems are intensely affected by their creators' perspectives.[3] We see this at the corporate level with websites 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 website 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 most information environments 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 system 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.

## 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 can involve a strong undercurrent of politics. The choice of organization and labeling systems can have a big impact on how users of the system 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,” “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 a designer, 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 users. 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 Information Environments

The organization of information environments is a major factor in determining their success, and yet many 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. Both organization schemes and structures have an important impact on the ways information is found and understood.

Before diving in, it’s important to understand information organization in the context of system development. Organization is closely related to navigation, labeling, and indexing. The organization structures of information environments often play the part of the 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 grouping of information, you avoid the distractions inherent in implementation details (such as the design of the navigation user interface) and can design a better product.

# Organization Schemes

We navigate through organization schemes every day. Contact directories, supermarkets, and libraries all use organization schemes to facilitate access. Some schemes are easy to use. We rarely have difficulty finding a particular word's definition in the alphabetical organization scheme of a dictionary. 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 dictionary and the supermarket are fundamentally different. The dictionary's alphabetical organization scheme 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. For example, country names are usually listed in alphabetical order. If you know the name of the country you are looking for, navigating the scheme is easy. "Chile" is in the Cs, which are after the Bs but before the Ds. 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 ("What's the name of that country that borders Guyana and French Guiana?").

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 schemes

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. Most address book applications organize contacts alphabetically by last name, as shown in Figure 6-2.

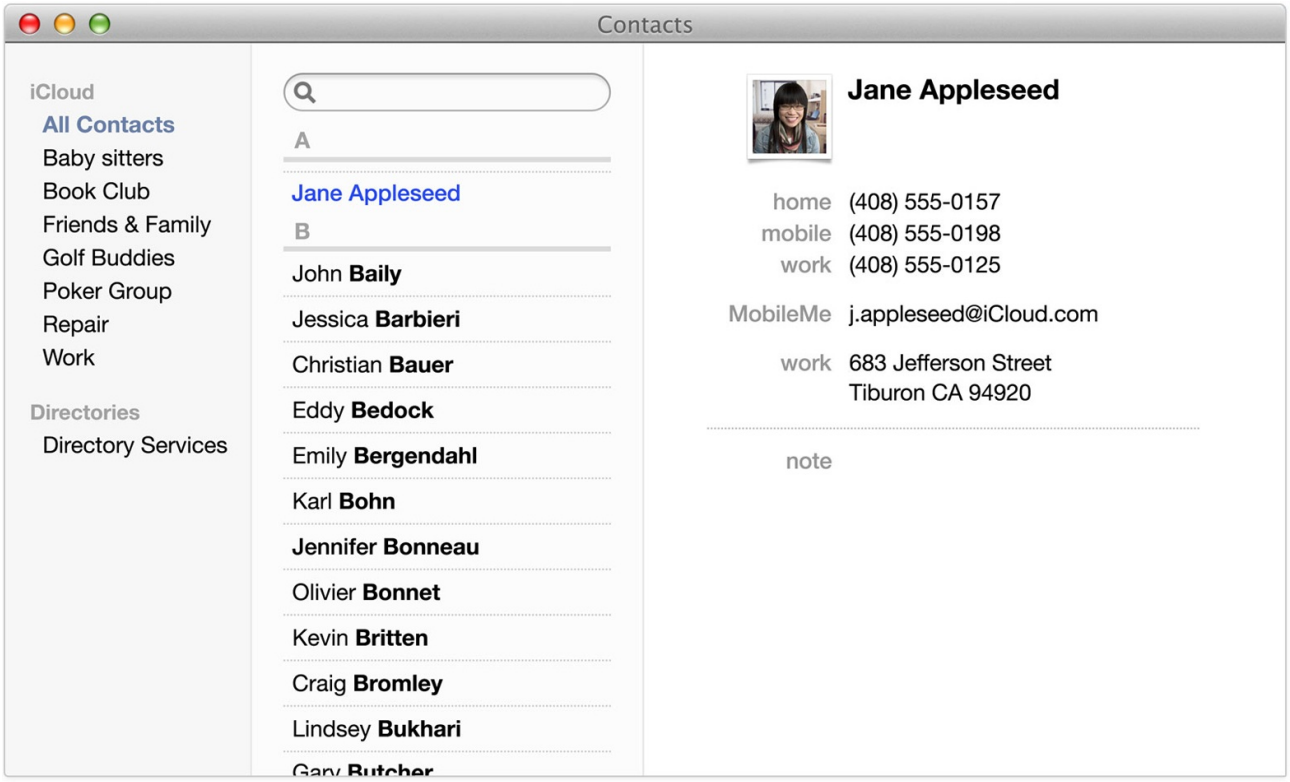
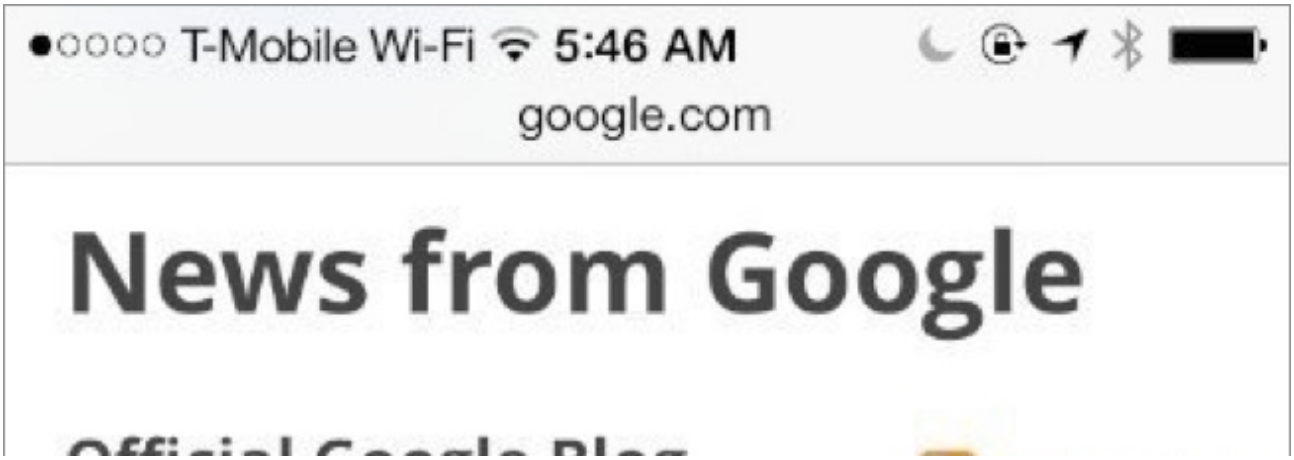



Figure 6-2. The OS X Contacts application (image: *https://www.apple.com/osx/apps/#contacts*)

## Chronological schemes

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 6-3). 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.



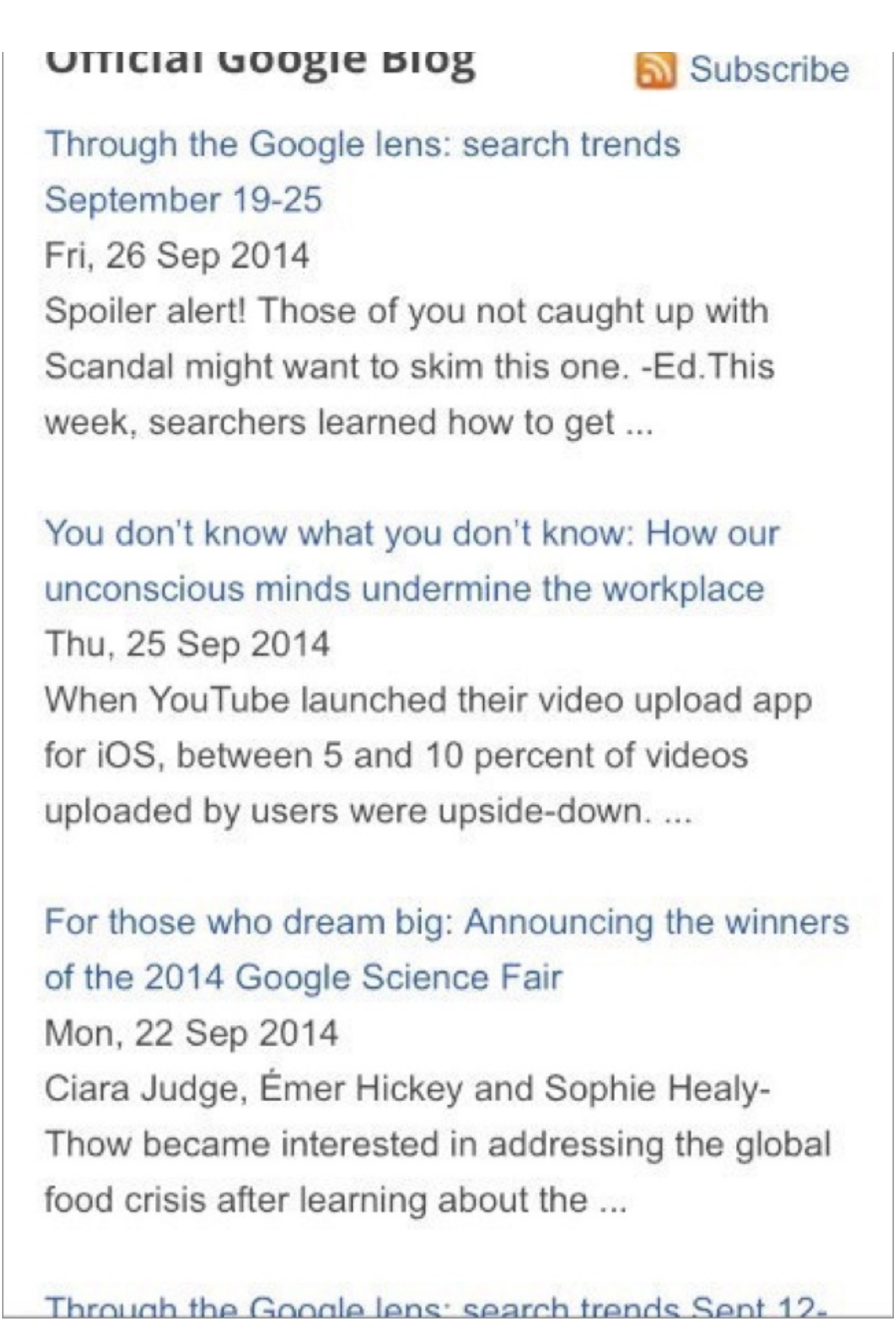



Figure 6-3. Press releases in reverse chronological order

### Geographical schemes

Place is often an important characteristic of information. We travel from one place to another. We care about the news and weather that affect us in our location. Political, social, and economic issues are frequently location dependent. And in a world where location-aware mobile devices have become the main way in which many people interact with information, companies like Google and Apple are investing heavily in local search and directory services, with the map as the main interface to this information.

Border disputes aside, geographical organization schemes are fairly straightforward to design and use. Figure 6-4 shows an example of a geographical organization scheme from Craigslist. The user can select her nearest local directory. If her browser supports geolocation, the site navigates directly to it.


Figure 6-4. A geographical organization scheme with geolocation

## 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, these schemes 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. As we mentioned in Chapter 3, information seeking is often iterative and interactive. What you find at the beginning of your search may influence what you 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.

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 an ambiguous organization scheme 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.

## Topical organization schemes

Organizing information by subject or topic is one of the most useful and challenging approaches. Newspapers are organized topically, so if you want to see the scores from yesterday's game, you know to turn to the sports section. Academic courses and departments, and the chapters of most nonfiction books, are all organized along topical lines. Many people assume that these topical groupings are fixed, when in fact they are cultural constructs that can vary over time.

While few information environments 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 websites such as Consumer Reports (shown in Figure 6-5) rely heavily on their topical organization schemes. Others, such as corporate websites, 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 system.

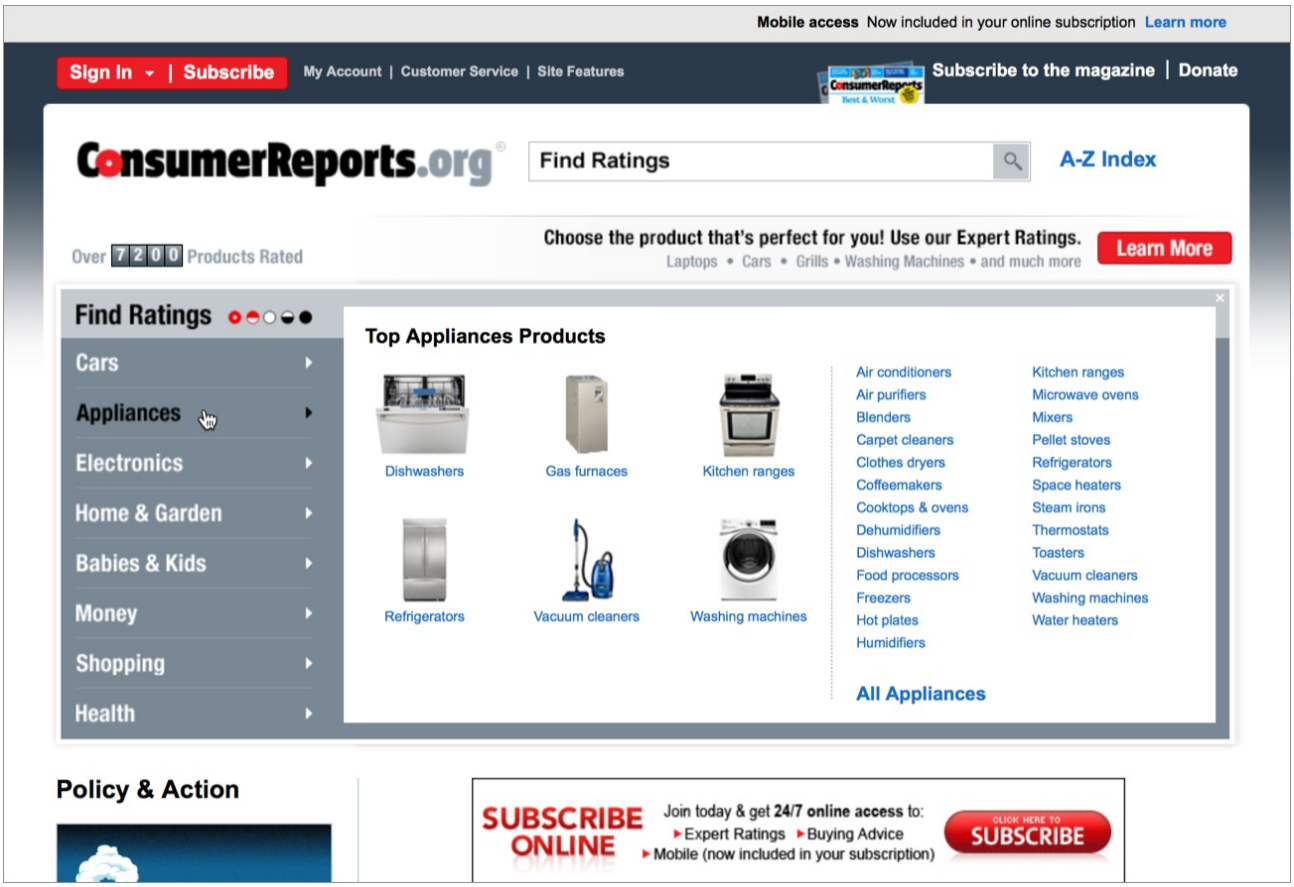



Figure 6-5. A topical taxonomy showing categories and subcategories

## Task-oriented schemes

Task-oriented schemes organize content and applications into collections 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. Task-oriented organization schemes are common in desktop and mobile apps, especially those that support the creation and management of content (such as word processors and spreadsheets; see Figure 6-6).

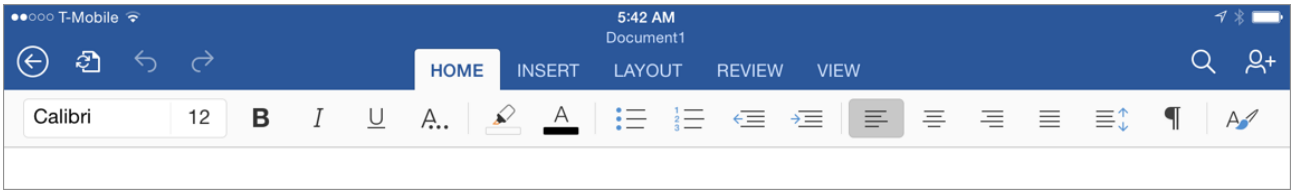



Figure 6-6. Like many apps, Microsoft Word on iOS features a task-oriented organization scheme

On the Web, task-oriented organization schemes are most common in the context of websites

where customer interaction takes center stage. Intranets and extranets also lend themselves well to a task orientation, because they tend to integrate powerful applications as well as content. You will rarely find a website 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 6-7.

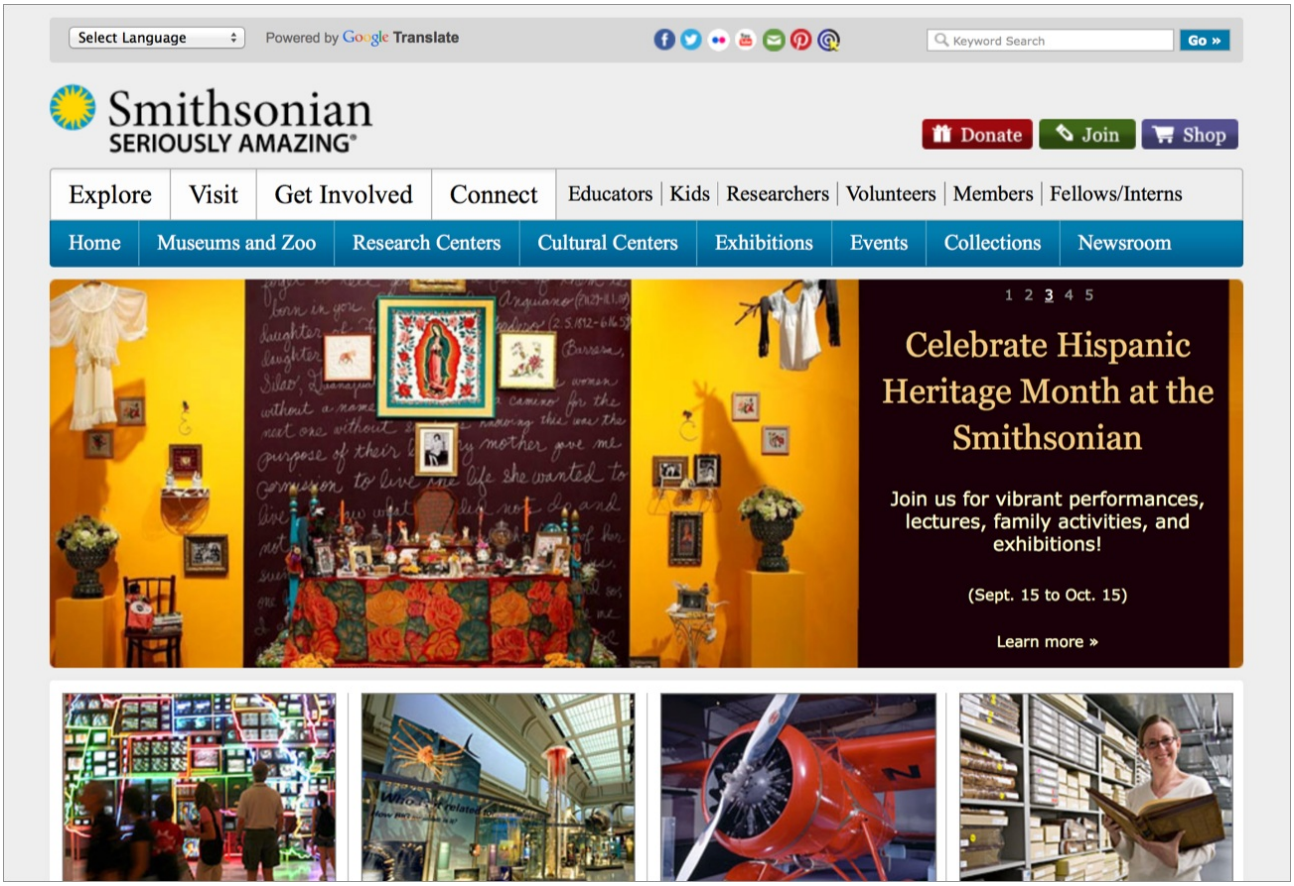



Figure 6-7. Task, topic, and audience coexist on the Smithsonian home page

## Audience-specific schemes

In cases where there are two or more clearly definable audiences for a product or service, an audience-specific organization scheme may make sense. This type of scheme 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. CERN, shown in Figure 6-8, presents an audience-oriented organization scheme that invites users to self-identify.

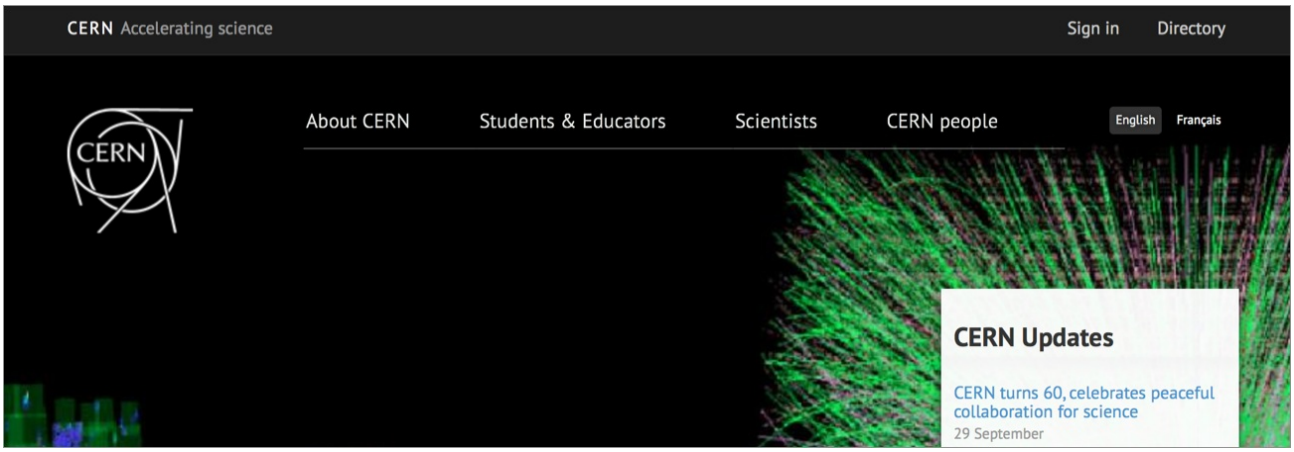



Figure 6-8. CERN invites users to self-identify

Organizing by audience brings all the promise and peril associated with any form of personalization. For example, CERN understands its audience segments and brings this knowledge to bear on its website. If I visit the site and identify myself as a member of the "Scientist" audience, CERN will present me with research results, papers from CERN researchers, and other information of interest to the scientific community. This information is not readily available in the "Students & Educators" section of the site. But what if I'm a science student doing research, and need access to research papers? All ambiguous schemes require us 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-driven schemes

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 a website.

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 website 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.

Another, perhaps less obvious, example: when you first log into Facebook, you are greeted by a "news feed" of content published by your Facebook friends. Initially, the news feed metaphor was apt, because the stream of posts consisted of the latest (chronologically) published friend content. However, as the frequency of posts grew, Facebook eventually introduced a different algorithm for choosing which posts to show first. The result is a news feed that can show posts that are several days old above more recent posts, breaking the chronological order that is expected in a news feed and potentially causing confusion. As shown in Figure 6-9, Facebook allows users to choose between "top stories" and "most recent" to determine which algorithm to use when ordering posts shown in the feed—an awkward solution at best.

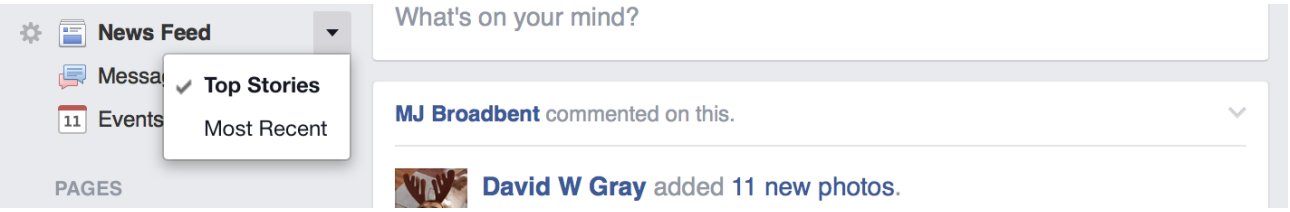



Figure 6-9. Facebook allows users to select which algorithm controls the sequence of posts in their news feed

## Hybrid schemes

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.

However, when you start blending elements of multiple schemes, confusion often follows, and solutions are rarely scalable. Consider the example in Figure 6-10. 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.

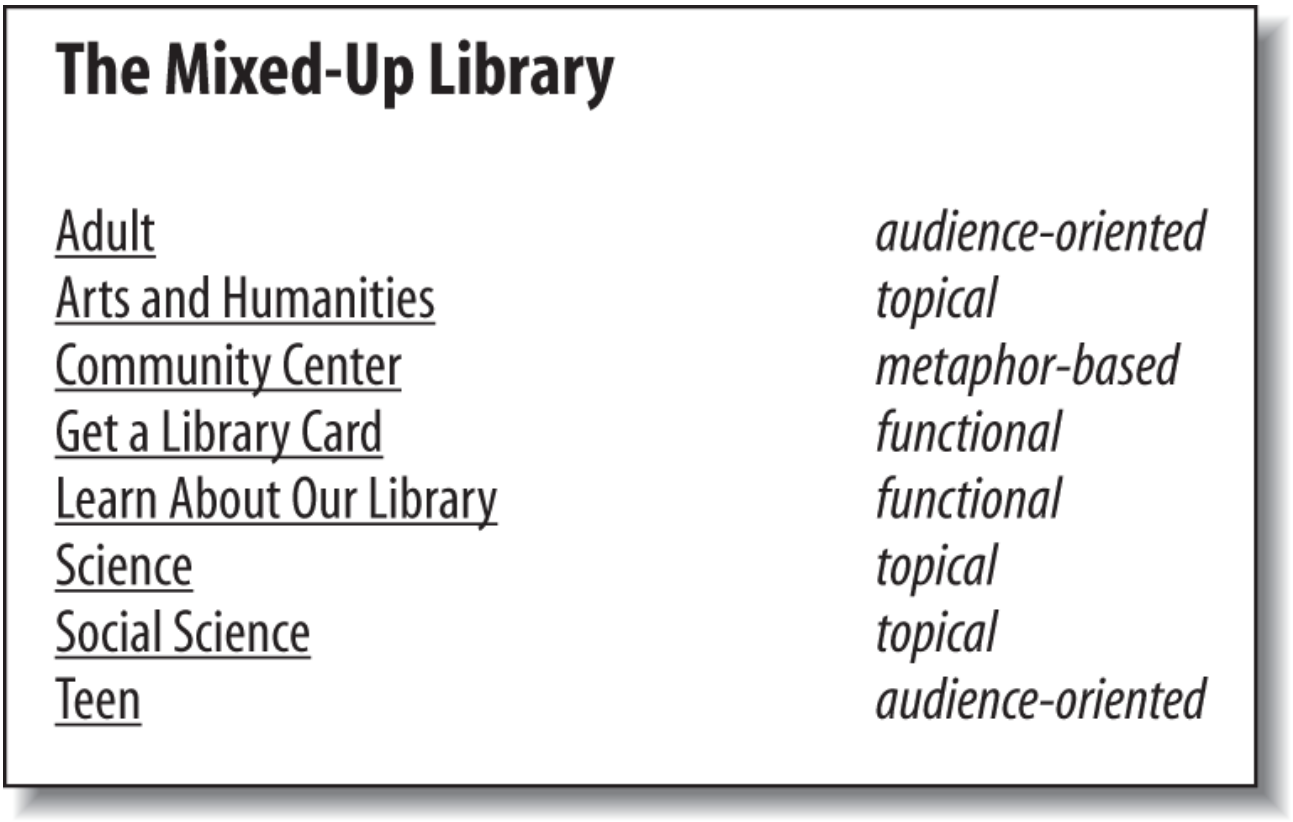
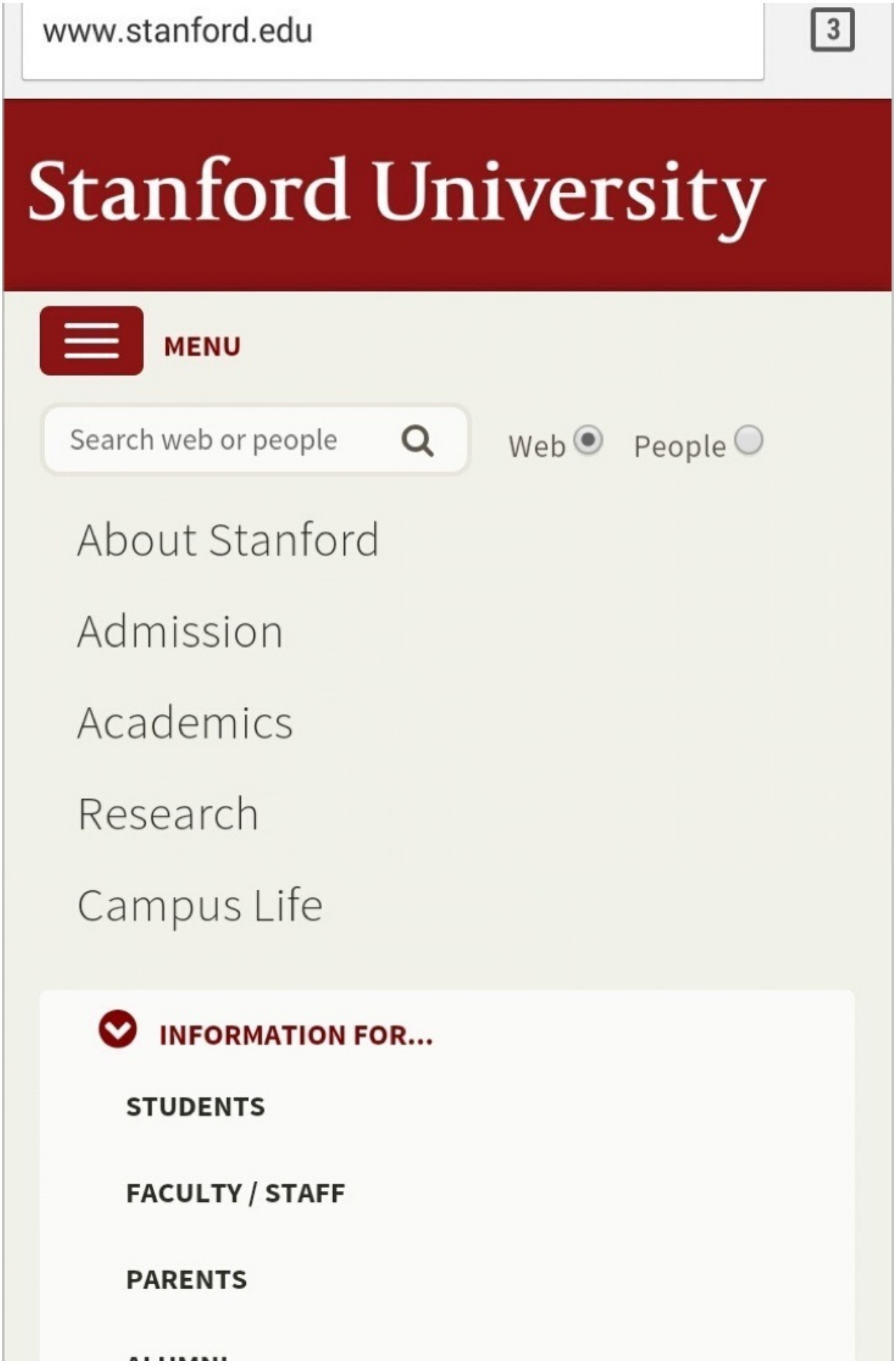



Figure 6-10. A hybrid organization scheme

The exception to these cautions against hybrid schemes exists within the surface layer of navigation. As illustrated by the Smithsonian example (Figure 6-7), many websites 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 only the highest-priority tasks are included, 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 main menu in the Stanford University website in Figure 6-11 reveals a topical scheme, an audience-oriented scheme, and a search function. By presenting them separately, Stanford provides flexibility without causing confusion.


www.stanford.edu
3
Stanford University
MENU
Search web or people
Web
People
About Stanford
Admission
Academics
Research
Campus Life
INFORMATION FOR...
STUDENTS
FACULTY / STAFF
PARENTS

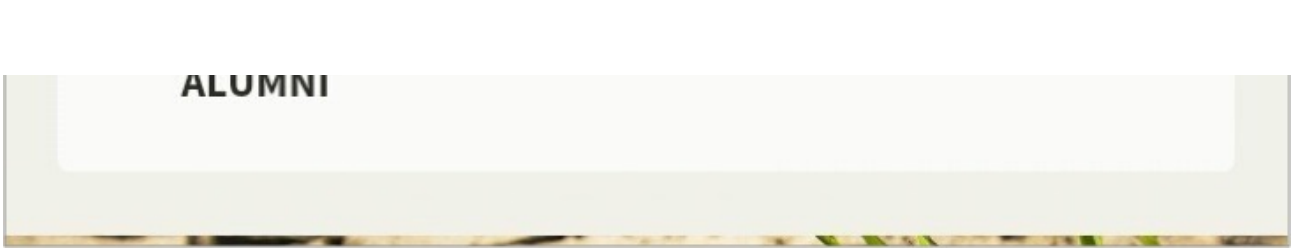



Figure 6-11. Stanford provides multiple organization schemes

## Organization Structures

Organization structure plays an intangible yet very important role in the design of information environments. 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 information 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 many good information architectures is a well-designed hierarchy. In this hypertextual, free-ranging 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 information environments that use hierarchical organization models. They are able to develop a mental model of the environment's structure and their location within that structure. This provides context that helps users feel comfortable. Figure 6-12 shows an example of a simple hierarchical model.

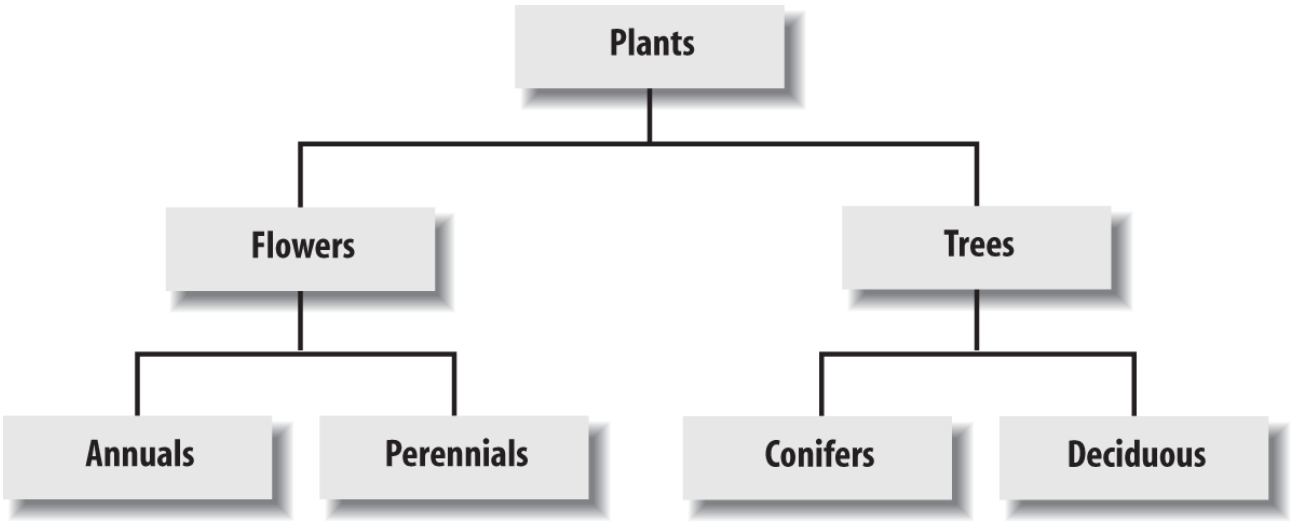



Figure 6-12. A simple hierarchical model

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 information environment 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.

## Designing hierarchies

When designing hierarchies, 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. Hierarchies 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 10 for more about metadata, facets, and polyhierarchy.)

Second, it is important to consider the balance between breadth and depth in your hierarchy. *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 or tap through an inordinate number of levels to find what they are looking for. The top of Figure 6-13 illustrates a narrow-and-deep hierarchy in which users are faced with six clicks to reach the deepest content. The bottom shows a broad-and-shallow hierarchy, where users

must choose from 10 categories to reach 10 content items. If a hierarchy is too broad and shallow, as in this case 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.

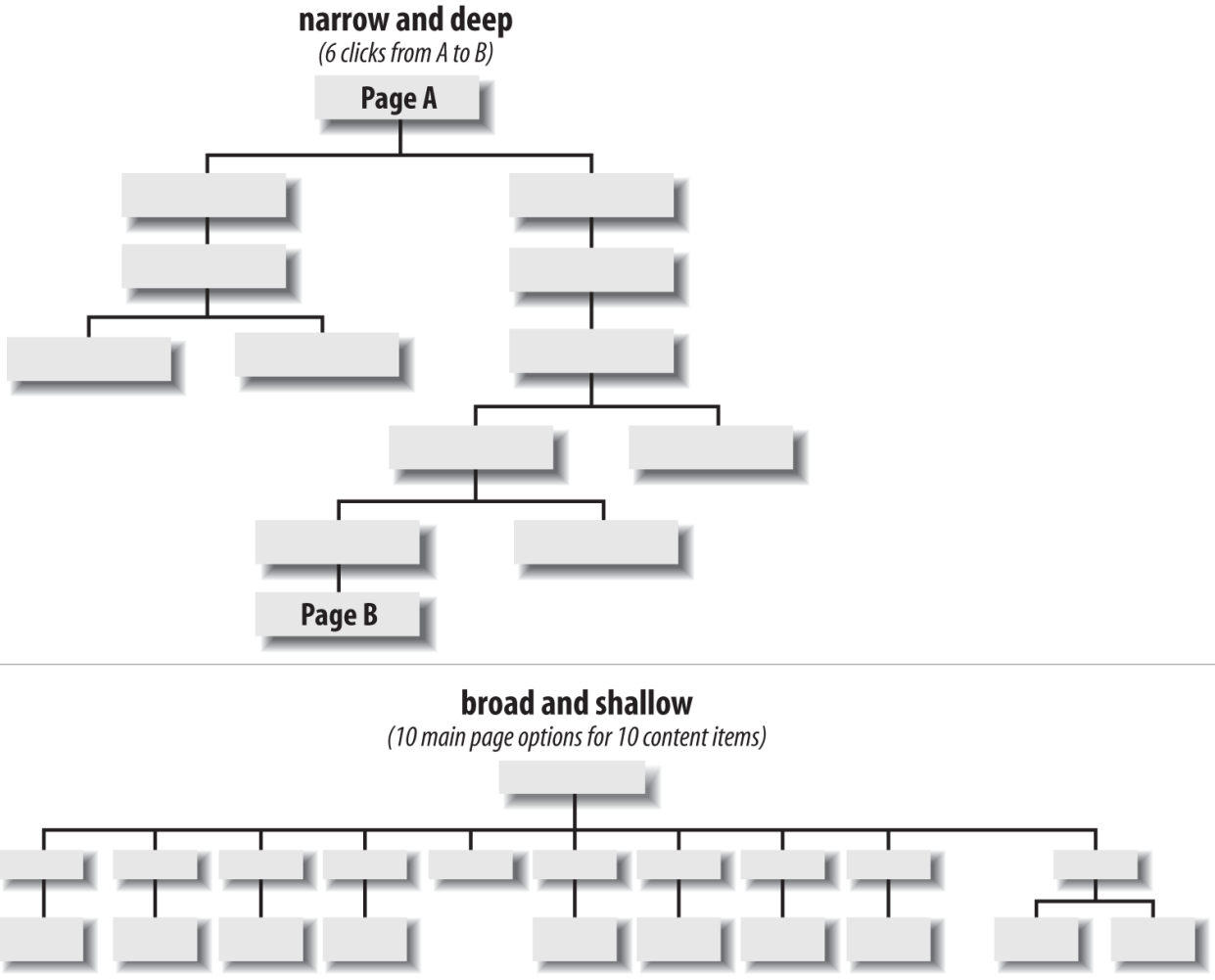



Figure 6-13. Balancing depth and breadth

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 to follow the infamous seven plus or minus two rule.[4] 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, when dealing with issues of breadth versus depth 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 6-14.[5] It's one of the US 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.


Figure 6-14. The National Cancer Institute groups items within the page

There are roughly 85 links on NCI's main page, and they're organized into several key groupings (Table 6-1).

| Group | Notes |
|---|---|
| Global navigation | Global navigation (e.g., Cancer Topics, Clinical Trials, Cancer Statistics) has seven links plus Search. |
| Highlighted stories | Includes 9 links. |
| Types of Cancer | Includes 12 Common Cancer Types and 4 alternate ways to explore All Cancer Types. |
| Clinical Trials | Includes 4 links. |
| Cancer Topics | Includes 9 links. |
| Cancer Statistics | Includes 3 links. |

| Group | Notes |
|---|---|
| Research & Funding | Includes 5 links. |
| NCI Vision & Priorities | Includes 4 links. |
| News | There are 3 headlines plus a link to the archive. |
| Resources | Includes 7 links. |
| Footer navigation | Includes 20 links. |

Table 6-1. Links on NCI's main page

These 80-odd links are subdivided into 10 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 website. At the very least, they'll become frustrated. An excellent study conducted by Microsoft Research suggests that a balance of breadth and depth may provide the best results.[6]

For new information environments 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, in many systems, the main page or screen serves as the most prominent and important navigation interface for users, helping set their expectations of what they can do in the system. Second, because of the main page's prominence and importance, companies tend to put lots of care (and money) into 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 6-15). Before computers became commonplace, Rolodexes were a common tool to store people's contact information. They consisted of rolls of physical cards, with each card representing an

individual contact: a *record* in the system. 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.

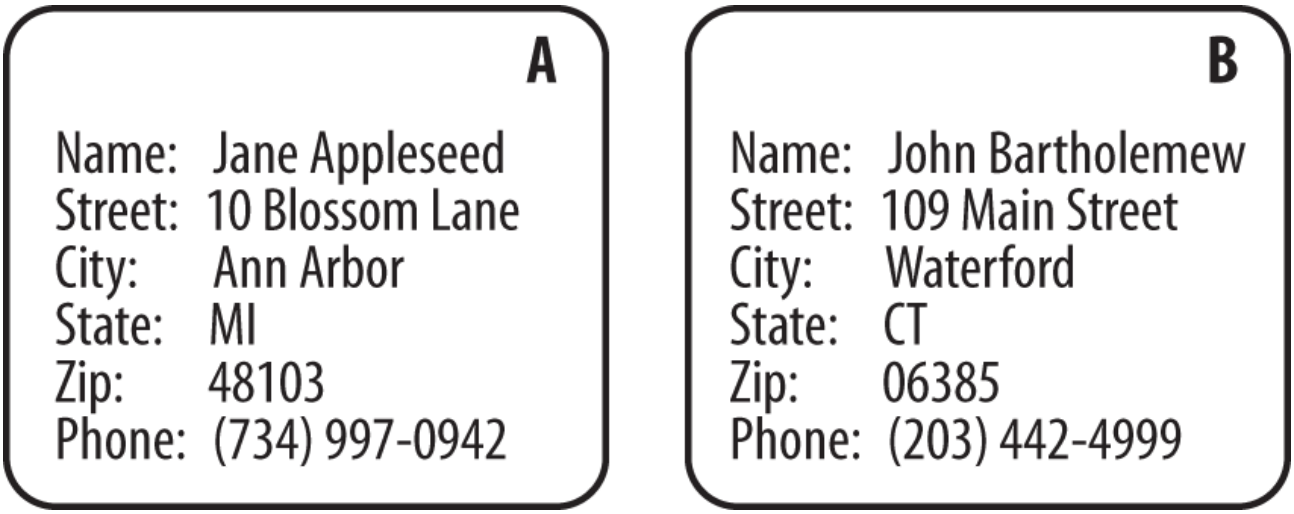



Figure 6-15. The printed card Rolodex is a simple database

In an old-fashioned Rolodex, users are limited to searching for a particular individual by last name. In a digital 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 6-16, 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.

## 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 6-16. A relational database schema (image: http://bit.ly/relational_model).

So why are database structures important to information architects? In a word, *metadata*. Metadata is the primary key that links information architecture to the design of database schemas. It allows us to apply the structure and power of relational databases to the heterogeneous, unstructured environments of websites and intranets. By tagging documents and other information objects with 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 6-17 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 you must become an expert 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 websites, you will hopefully be able to rely on content management system (CMS) software to manage your metadata and controlled vocabularies.

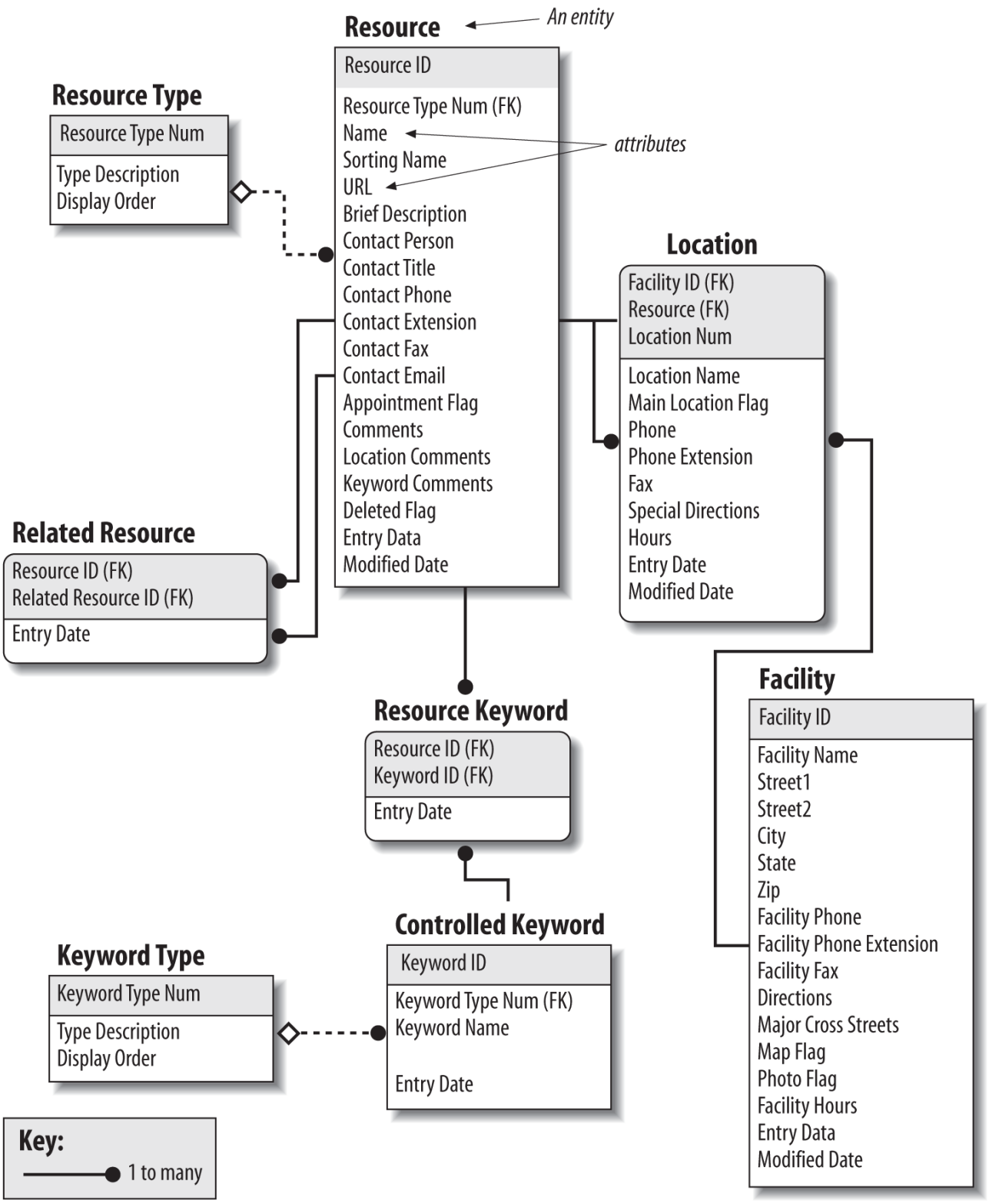



Figure 6-17. An entity relationship diagram showing a structured approach to defining a metadata schema (courtesy of Peter Wyngaard of Interconnect of Ann Arbor)

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

- Automatic generation of alphabetical indexes (e.g., a product index)

- Dynamic presentation of associative “see also” links and content
- Fielded searching
- Advanced filtering and sorting of search results

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 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.

These components can form hypermedia systems that connect text, data, image, video, and audio chunks. Hypertext chunks can be connected hierarchically, nonhierarchically, or both, as shown in Figure 6-18. In hypertext systems, content chunks are connected via links in a loose web of relationships.

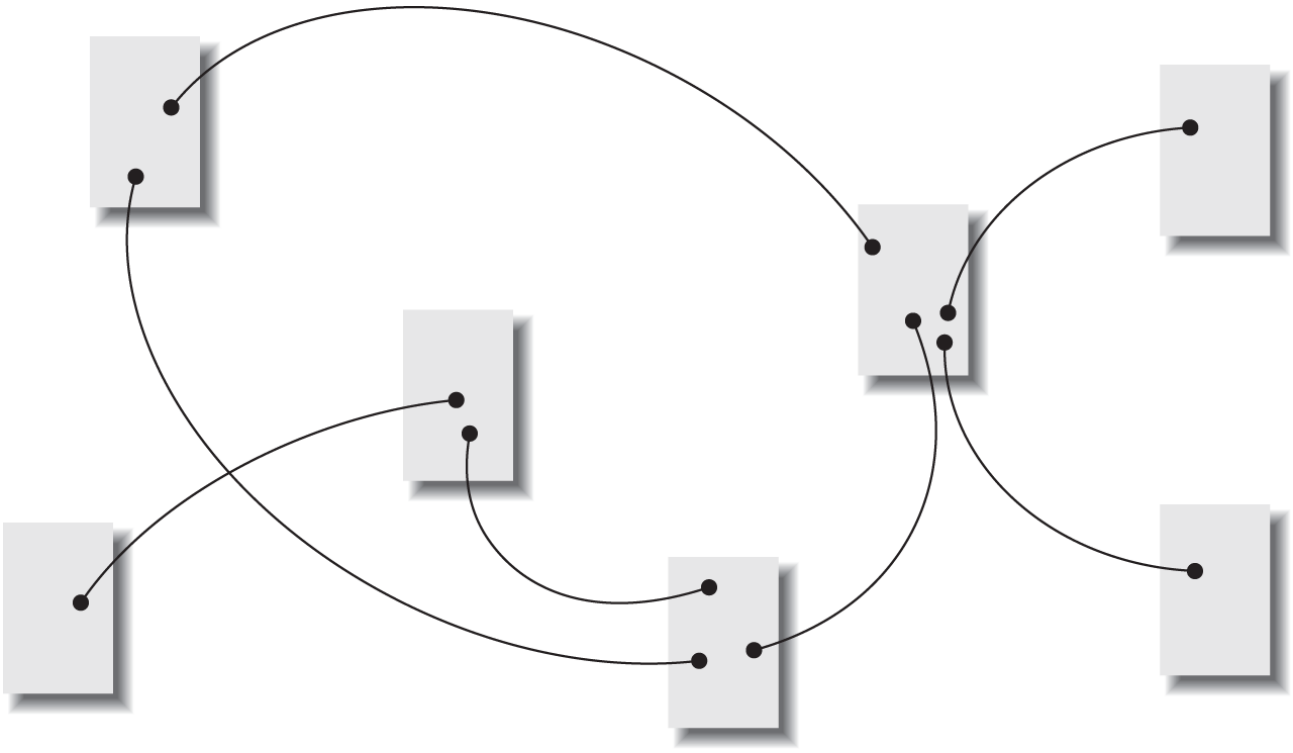


Figure 6-18. A network of hypertextual connections

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. The relationships that one person sees between content items may not be apparent to others. Additionally, as users navigate through highly hypertextual websites, 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 environment's organization. Without context, users can quickly become overwhelmed and frustrated.

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

Social media has become a mainstay of the digital experience. Platforms like Facebook and Twitter have enabled hundreds of millions of people to share their interests, photos, videos, and more with one another and with all of us. As a result, social classification—primarily driven by user-generated content tagging—has emerged as an important tool for organizing information in shared information environments.

*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. These tags can be informally supported in text fields, or they can be provided for with bespoke fields in the formal structure of content objects. 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 example, in Twitter, words with a prepended hash (#) have a special meaning: the system picks them up as tags. When you include one of these tagged words in a tweet, the system marks that post as belonging to a group of posts that has been informally defined by the users of Twitter (Figure 6-19). 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.[7]

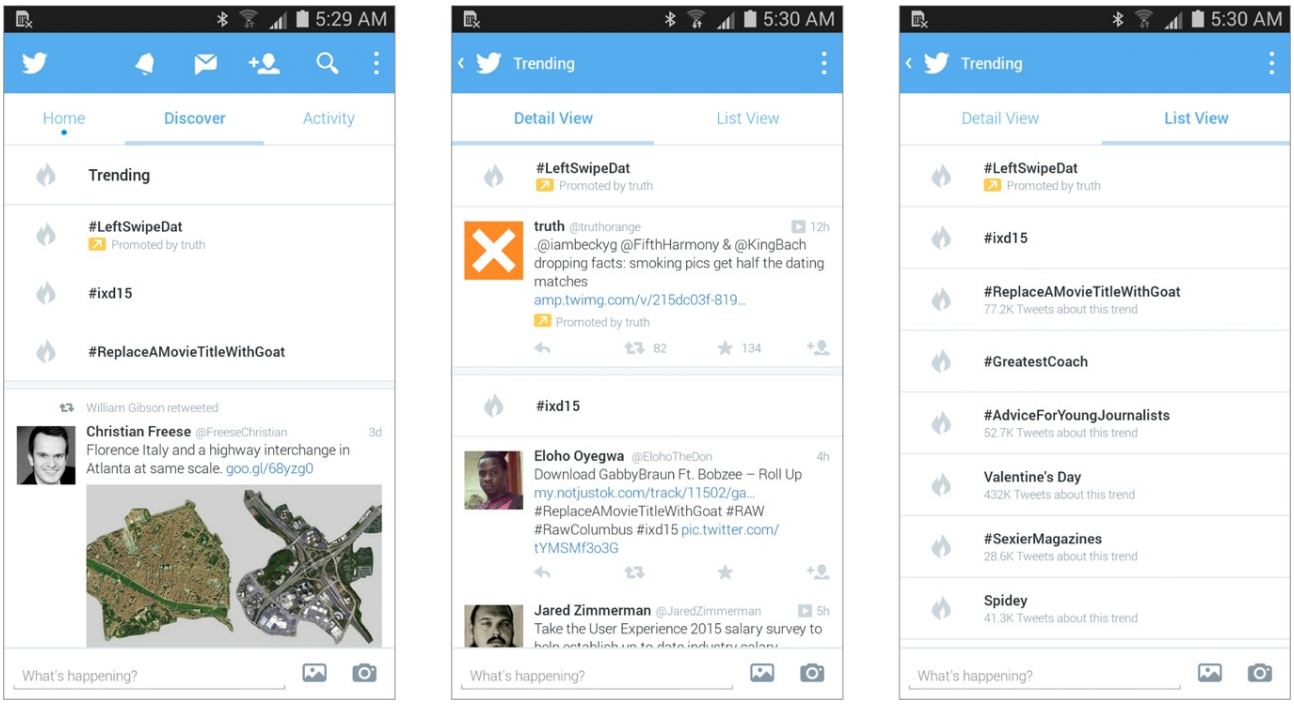



Figure 6-19. The “Discover” and “Trending” features in Twitter, which allow you to discover new and potentially interesting content, are driven by user-generated tags

Similarly, LinkedIn allows users to “endorse” their professional contacts as possessing certain individual professional skills (Figure 6-20). These endorsements are in effect tags: they allow users to describe their business contacts in a granular way that informs how the system groups them with similar people. Though users can suggest new endorsement labels, these are not free-form, unstructured tags like the ones that Twitter employs; they have been built as bespoke, dedicated structures within the architecture of LinkedIn.

In the early days of information architecture, an impassioned debate raged over whether or not free-form tag structures (or “folksonomies,” as information architect Thomas Vander Wal cleverly christened them) would eliminate the need for top-down, centrally defined information structures. The passage of time has proven the value of top-down structures: high-profile experiments in tag-driven systems—such as the bookmarking service Delicious.com—fizzled in the marketplace, and most of these systems employed tags within centrally defined structures anyway. Still, free-form tagging has proven its usefulness in specific situations, and it remains a valuable tool in the information architect’s toolset.

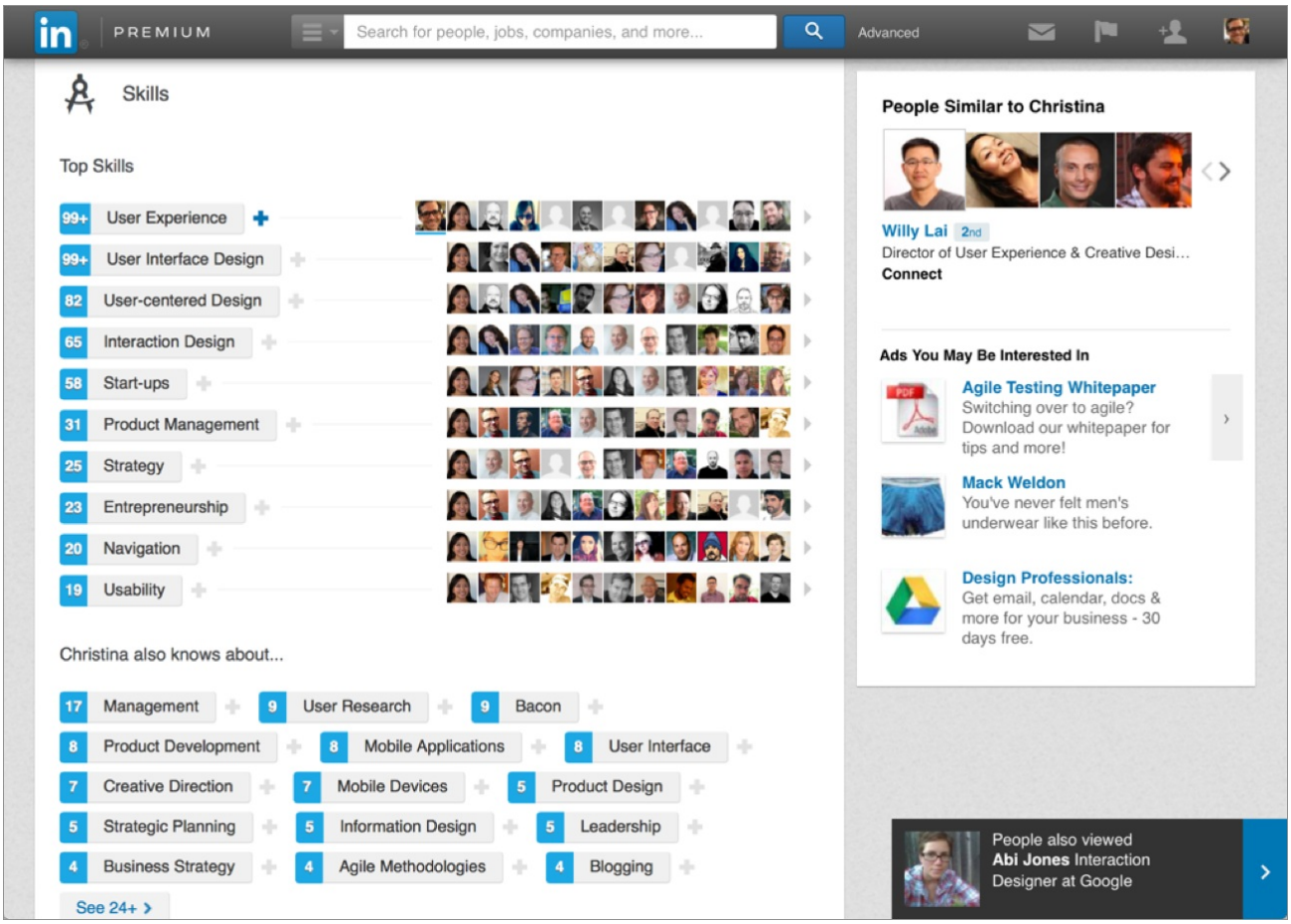



Figure 6-20. LinkedIn allows you to “endorse” your contacts as having certain professional skills, from a set of predefined tags

## Creating Cohesive Organization Systems

User 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 website development project, these questions can be overwhelming. That’s why it’s important to break down the information enviornment 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 one information environment to another. However, there are a few guidelines to keep in mind.

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 their 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 systems typically require several types of structures. The top-level, umbrella architecture for the environment will almost certainly be hierarchical. As you are designing this hierarchy, keep a look out for collections of structured, homogeneous information. These potential subenvironments 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.

## Recap

Let's recap what we've learned in this chapter:

- Our understanding of the world is informed by how we classify things.
- Classifying things is not easy; we have to deal with ambiguity, heterogeneity, differences in perspective, and internal politics, among other challenges.
- We can organize things using exact organization schemes or ambiguous organization schemes.
- Exact organization schemes include alphabetical, chronological, and geographical groupings.
- Ambiguous organization schemes include topical, task-based, audience-based, metaphorical, and hybrid groupings.
- The structure of organization schemes also plays an important role in the design of information environments.
- Social classification has emerged as an important tool for organizing information in shared digital environments.

Now let's move on to cover another critical component of an information architecture: labeling systems.

[1] The tomato is technically a berry and thus a fruit, despite a 1893 US Court decision that declared it a vegetable. (John Nix, an importer of West Indies tomatoes, had brought suit to lift a 10% tariff, mandated by Congress, on imported vegetables. Nix argued that the tomato is a fruit. The Court held that because a tomato was consumed as a vegetable rather than as a dessert-like fruit, it was a vegetable.) Source: Denise Grady, "Best Bite of Summer" (Self 19:7, 1997, 124–125).

[2] 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 N.J. Belkin, "Anomalous States of Knowledge as a Basis for Information Retrieval" (*Canadian Journal of Information Science* 5, 1980, 133–143).

[3] For a fascinating study on the idiosyncratic methods people use to organize their physical desktops and office spaces, see T.W. Malone, "How Do People Organize Their Desks? Implications for the Design of Office Information Systems" (*ACM Transactions on Office Information Systems* 1, 1983, 99–112).

[4] G. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information" (*Psychological Review* 63:2, 1956, 81–97).

[5] Just before this book went to press, the National Cancer Institute launched a new, improved version of this page—which we like quite a bit!

[6] Kevin Larson and Mary Czerwinski, Microsoft Research, "Web Page Design: Implications of Memory, Structure and Scent for Information Retrieval".

[7] Twitter tags weren't originally included in the system: they emerged informally, included by the users of the platform in unstructured text fields.