ORGANIZING AND FINDING INFORMATION ON THE WEB

This paper discusses and compares techniques and technologies for making information findable in a web application or on the internet. It was written to help you make better websites, better web directories, and better tools for finding web content. .

Finding things in the physical and cyber worlds

We spend a lot of time trying to find things. At work, we search for information to progress our business objectives. We go home and have to remember where we put whatever we need to pursue our leisure activities. We would be more productive if we could do this finding more quickly.

We can also get pleasure out of searching. Much of the shopping experience revolves around creating a pleasant way to find things, and buy them.

The tools for finding information on the internet have improved significantly in the last decade. However, the quantity of information on the web seems to have grown even faster--even with improved tools, it takes us at least as long to find what we need. This creates a business opportunity. If you can organize web information in a better or more pleasant way, you can attract more visitors to your site, application, or product and have them stay longer. The rest of this paper describes technologies to accomplish this.

Alphabetical lists

These are the equivalent of the "white pages" phone directory. They are the second most primitive way to organize web information. (Unordered lists are the most primitive). Many web directories only provide an alpha list. An alphabetized list has the advantages that (a) it is quick to implement and (b) every item has a pre-defined place in the structure. In most cases, the list is automatically alphabetized. Alpha lists have the disadvantage that you have to know the name of what you're looking for to find it. An example is the UN Islands Directory's alphabetical listing at http://islands.unep.ch/Iindex.htm. Fortunately, this alphabetical listing is not the only one provided, but this is not the case for many sites.

Hierarchical lists

The better way to organize web information is to create a hierarchical topic-oriented directory, analogous to the "yellow pages" phone directory. This has the advantage that you can "browse a category" to find what you're looking for. There are two disadvantages to this approach: (a) it takes effort to categorize items and (2) different people would categorize the same thing into different topics.

In the early days hierarchical directories were the primary way web information was organized, so you could "browse" topics using your "browser." This was continued with (http://www.dmoz.org) the Open Directory (also at http://www.asperon.net/directory) the Yahoo directory (http://www.yahoo.com) and Craigs List (http://www.craigslist.org) . The Open Directory has cataloged over **4 million** websites, but it has taken over 63,000 volunteer editors to do this. In the mean time the web has grown to over **4.2 billion** web pages. The volunteers can't keep up. The Yahoo directory requires a $299 per year payment to be listed or you must be popular enough for them to put the effort in adding you.

If you want to add a hierachical directory to your website, Dataroo™ (http://www.dataroo.com) provides a tool called Surftree™ that makes this easy.

Database search

A technique that was brought over to the web from client/server architecture was to do a database search. This means typing in categories that match database fields. http://www.northerncoloradorentals.com is an example that lets you find a rental property in this way. Database search is easy to implement if you're using a database. The user must know the search field and options in order to get the right result. This can be done with choice boxes or pull-down menus, but these limit data entry flexibility. Database search also works poorly for items that could be in more than one category. If you have a software product that converts Java code to C, would this be listed as Java or C software? There are three improvements that can make database search much better. These can also work in conjunction with other searches methodologies such as keyword search and hierarchical lists:

a. **Sifting the data** by allowing the user to sequentially narrow the total amount of data that he would scan. For example, the website at http://www.reodispo.com lets you first pick a state. It then gives you choices for cities in this state, then type of property. Only valid next level choices are provided.

b. **Sorting or ordering the data**. For example, Northern Colorado Rentals allows you to "Order by" Date listed, Price, City or Area.

c. **Hierarchically organizing the data**. In some cases, you can have the data presented first by one field and then by another. If this had been

implemented on Northern Colorado Rentals, you might then be able to specify that you wanted the results organized first by City and then by Price, for example.

Key word search

Key word search was a big advancement over database search. It worked for any text-based site and didn't require hand organizing into categories. This is why it quickly became the #1 technique for finding information on the web. It does have limitations in that, if you don't know what something is called or you don't know what you're looking for, you won't get meaningful results. This is why other methods continue to also be used.

You can use a product like (http://www.freefind.com) Freefind to add key word search to your web site.

Visual sort

It's more difficult to use the web to find item that are not text. One solution is (http://images.google.com), Google Images which allows you to search for graphics by presenting thumbnails that can quickly be visually scanned, just like scanning a store aisle.

Geospatial organization

Another class of items can be organized geospatially. For example, the job hunting map used by clicking on "FindJobs" at Flipdog (http://www.flipdog.com) allows you to select your state by clicking on it. This is easier than having to scan through an alphabetized list of states.

Comparison of methods

The table below compares the methods discussed. In general, the best approach is to use as many of the above methods as possible. This will make it easiest for the web user to get to the desired information. For example, the plastics industry directory at http://www.selectedtechnologies.com can be searched both by category and by key word.

Additional criteria

In addition to the above, please keep in mind that:

- Whatever approach you use should be implemented in a way that makes it easy to maintain and grow. A website with stale information is not useful.

- It is important to ensure that whatever approach you use is compatible with the web browsers used by your intended audience. Sites like Yahoo and Google do a good job in this regard. The Flipdog site is completely unusable by people who use Netscape.

Conclusion

This paper is an overview. Feel free to give me comments to help improve this. Here is an additional reference to let you further explore this topic:

- Rosenfeld, Louis and Peter Morville. *Information Architecture for the World Wide Web*. © 1998 O'Reilly and Associates.

Please contact me if you would like more info or help on making your business or website more successful.

About the Author

Bert Vermeulen owns Corp21, a company that supports, incubates and advises businesses, entrepreneurs, and inventors around the world. For more information, see http://www.corp21.com. He also hosts Dataroo (http://www.dataroo.com), a site that provides tools and information for making it easy to organize information on a web page.

Technique	Example	Advantages	Disadvantages
Alphabetical list	Islands directory	Easy and fast to implement Everything fits in order	Difficult to use for lists of more than 20.
Hierarchical list	Open directory	Can be used for large lists	It takes labor to categorize.
Database search	Rental property	Easy to implement if you use a database.	Must provide list of valid choices. Difficult for multi-category items
a. Sifting data	Reodispo	Makes site easier to browse	Takes extra programming
b. Sorting data	Rental property	Makes site easier to browse	Takes extra programming
c. Hierarchical sort	None	Makes site easier to browse	Takes extra programming
Key word search	Google	Flexible and easy to use	Must know what you are looking for.
Visual sort	Google images	Can scan visuals quickly. More enjoyable than text	Difficult to use if you have more than 100 results.
Geospatial sort	Flipdog	Easy to spot More enjoyable than text	More difficult to implement. Requires something where location is relevant.