Building Your Library's Portal

This text proposes a definition for Web portal, describes how information architecture plays a critical role in the development of a library's website, and briefly describes one database-driven website application designed for libraries, MyLibrary.

What is a portal?

The defining characteristic of a library portal is the user-driven customizability of a website's content. It is a website who's output is tailored for an individual and is retained by the underlying system so the user's customizations are echoed on subsequent visits to the site.

There are as many types of websites as there are types of human activity. As the size of a website grows so does the need for better searchability and browsability. As the size of a website grows even more and as the intended audience of the website's content becomes more diversified so does the need for user-driven customization and personalization.

A portal is only one possible component to a library's Web presence. A library website can be made up of three types of content:

1. *Information about the library* - staff directories, departmental descriptions, maps of the building, hours, etc.
2. *Electronic versions of traditional library services* - online tutorials, book renewals, interlibrary loan requests and status reports, requests for purchase, online chat/reference, virtual tours of the building(s), etc.
3. *Access to library content* - catalogs, indexes, full-text magazines and journals, digitized special collections, free and commercial ebooks, government documents, freely accessible Internet resources, electronic encyclopedias and dictionaries, licensed content from vendors, etc.

Our goal as librarians who maintain library websites is to implement these services and collections in a functional, scalable, usable, and an aesthetically pleasing manner. In order to achieve these goals is necessary to first practice a bit of information architecture.

Information architecture

Information architecture begins by answering questions about your institution's purpose, your user's needs and desires, and the types of content you have to communicate – research. Keeping the answers to these questions in mind, information architecture is then about organizing your content, labeling it effectively, providing the means for browsing and searching the content, and maintaining metadata used to describe it – strategy and implementation.

Research

The goal of research is to attempt to answer a myriad of questions. Your answers will often not be definitive, and they will change over time.

The answers to these questions will help formulate a set of milestones for the website. The questions fall into three categories: questions about your organization, questions about your intended audience, and questions about your content. Some of these questions include:

- What is the mission of your organization, and in turn, what is the purpose of the website?
- Who is the primary audience of your website?
- What content does the organization have to communicate via the website?
- Who in your organization will do the work to create and maintain the website?
- What task does your audience expect your website to facilitate, and what technical resources do they have at their disposal?

To answer these questions talk your users, read your organization's mission statement, and take a serious inventory of your existing content. Do not rely solely on your professional judgment. The only way you are going to get a accurate picture of user expectations is to ask them.

Strategy

Articulate a plan for putting the research answers into practice.

User's want to accomplish specific tasks to do something else. "I need a list of articles on... Do you have the book whose title is... What is the status of my interlibrary loan request? I need a synopsis of the American Revolutionary War." Go back to the user's and ask them if it makes sense. Again, focus group interviews are indispensable here. Ask the user.

On pieces of paper, draw rough outlines and tree structures of your website. In general it is better to create shallow and wide websites as opposed to deep and narrow ones.

As the website becomes larger, say more than twenty five pages, it becomes more important to index the website's content and provide a search engine against the index. The key to success is accurately describing each item of the indexed content thoroughly with user-centered metadata.

As your website becomes even larger, say more than seventy-five pages, consider using a relational database to maintain and create your website's content. By putting your content into a database you will be effectively separating your content from presentation. In turn, this

will allow you repurpose your content for different venues. Thus, you will be able to create things like:

- a comprehensive list of all your content by subject, audience, format, etc.
- pathfinders describing very specific subject areas or class assignments
- generalized home pages fulfilling the needs of most users
- portal applications whose content is tailored to specific individuals
- content intended to by syndicated and integrated into your host institution's website

Organizing your content in a relational database application is the key to implementing a portal. The database must contain fields describing users and those same fields must be used to bring together information resources pertinent to their interest. For example, your database might classify users in terms of their status in the organization such as grade or education level. Similarly, you will have to classify your information resources with these same grades or education levels. You might classify users in terms of their primary areas of subject interest or expertise. That's easy, librarians classify information resources with subjects all the time, but remember to keep these classifications user-centric. Do not use things like Library of Congress Subject Headings. Instead use subject terms that are directly identifiable to your users. Seriously consider creating a sort of used bookstore model for your subject classification.

Implementation

The final phase of the redesign process is the implementation phase. It is in this phase where the strategy is put into practice. For example, ROT (redundant, outdated, and trivial content) will be removed. Tools and processes for creating and maintaining website content will be refined, documented, and taught. Massive amounts of HTML will be "retrospectively converted" into the new design. Indexing will be more systematically applied. During this part of the process usability testing will come to a head and more rigorously applied. The website will also be vigorously marketed and promoted in an effort to make the user population more aware of the changes.

Ironically, this part of the process should be easy. All the thinking was done before hand, and now all you have to do is the work.

MyLibrary

MyLibrary is a database-driven website application designed for libraries. It provides the means for librarians to describe sets of Internet resources in terms of user characteristics. It then provides the means for repurposing this content in the form of home pages, lists of resources organized in various fashions, pathfinders, a portal, and syndicated content in the form of XML streams. MyLibrary also provides the means to report on what resources are being used and by whom. It includes a search engine indexing its content. It facilitates librarians sending targeted email messages to their constituents. It also includes a "virtual new bookshelf" service.

MyLibrary is distributed as open source software. Being freely available, anybody can download the software, examine it, pick it apart, and try it out before making a commitment to using it. Technically speaking, MyLibrary is a set of CGI Perl modules and scripts running on top of an HTTP server and against a relational database. The relational database can be either MySQL or PostgreSQL. It can run on just about any ol' computer, Windows or Unix.

MyLibrary has been available since 1998 and it is in production in about two dozen libraries across the world. It has also been the inspiration for many other database-driven applications and portals including implementations at the Los Alamos National Laboratory and the recently developed MyLibrary service at the University of Rochester. The term MyLibrary is slowly becoming a part of the library vernacular in the same way name brands such as Kleenex and Xerox have become part of our language.

The University Libraries of Notre Dame is in the very beginning stages of redesigning its Web presence. The Libraries' Digital Access and Information Architecture Department is leading this effort and plan to use the processes outlined above to ensure the redesign goes smoothly. Consequently, we are spending the time to learn user needs, explicitly articulating the purpose of the website, and taking a long hard look at the content we have to offer. We will then use some form of a database-driven website to repurpose our content for many venues. One of those venues will most likely be a portal application.

Futher reading

For more information about information architecture read Louis Rosenfeld and Peter Morville's book, *Information Architecture for the World Wide Web*, O'Reilly Publishers, 2002.

For more information about MyLibrary, see it's home page at http://dewey.library.nd.edu/mylibrary/.

Eric Lease Morgan
University Libraries of Notre Dame
March 25, 2003

http://dewey.library.nd.edu/morgan/musings/portals/