Maintaining Quality of Library Web Sites Using Cluster and Path Analysis

By John Matylonek, Oregon State University, Valley Library

Introduction

Library web services are changing and expanding rapidly. The increasing complexity and quantity of subject electronic information resources, novel ways of providing information and services, cross functionality and complexity of the online catalog and general web sites, and the increasing expectations of users have made web design decision making all the more difficult. These factors must be incorporated in an evolving design so that library users are assisted by the changes, not stymied. Furthermore, the skills of a library web site user cover a spectrum between beginner to expert while user needs run the gamut from highly comprehensive research to vague information requests based on expediency and convenience. A library web site must cater to all these audiences by gauging the effectiveness of the services.

It is not always apparent how users are actually being served by the plethora of service links on a library web site. Measuring quality and effectiveness of a library web site is not as simple as counting books, library seating, and circulation statistics. Many librarians have resorted to increased instruction to deal with this complexity and subtlety. Measuring how library web-site visitors use and perform on a library web site can, however, provide a baseline for future enhancements. Special design techniques and usage statistics can help library web developers in adapting their web sites to the needs of library web site users as services and technology change.

Development of quality criteria based on performance measures

A library web site's quality can be increased across many criteria. Part of developing criteria for quality involves agreeing on what measures to use in determining success. Fortunately, the web environment provides many performance factors by which effectiveness can be gauged. How users associate services, what they feel is important and relevant, where and how they use the library web site, and how successful they are at finding the information they are seeking are all performance factors that can be determined by studying user data and server log statistics.

Purpose of this paper

This paper will provide methods for the systematic redesign of a library web site through gathering baseline data and comparing that data to changes in use brought about by design changes. I will show how complementary information, based on web users' log statistics and direct observation of users, can enhance a library web site. This paper will relate this information to practical web-site decisions that library web developers can make to help them design, modify, and evolve their web sites so that quality is maintained. Identifying common problems with web library services and finding solutions for these problems are the ultimate goals of these studies. The development of general guidelines that arise from different library perspectives may benefit the profession as a whole and may help to identify standards and best practices.

Measuring the success of library web users

Gathering statistics on library web-user behavior does no good unless a method of determining success is formulated. Success can only be measured by agreeing on objectives, establishment of specific criteria, measurement of those criteria, and observation of the results. If a baseline for a particular performance measure is observed, then the effect of a change in layout, logical flow, prominence, etc. should be reflected in a web performance factor. The following method can help library web-site developers design a library web site in a systematic way by guiding a web-design team to identify a purpose, track changes, and document the evolution of its library web site.

1. **Establish Objectives.** Reach an agreement between all parties involved about what you are trying to achieve (for example, the promotion, placement, and prominence of services on the web site can all be reflected in web-usage statistics).

2. **Determine Criteria for Success.** Define success specifically (for instance, the desired number of user sessions, increase in usage, decrease in usage, the paths users take, time they spend, target population for the site).

3. **Develop a Baseline**. This step is essential because benchmarking is a comparative process. As library web sites are increasingly relied upon for basic services, knowledge of the trend of usage in these services will be invaluable.

4. **Select a Benchmark.** Baselines provide an essential benchmark if you are evaluating the web site's progress. There may, however, be other benchmarks that are more helpful. For instance, library home pages sometimes have a promotional function. Comparing web graphic designs with those of comparable institutions is appropriate if extensive prototyping is impractical.

5. **Compare the results to objectives.** Once the design change has been implemented, examine the resulting differences in web statistics. Decide if the change succeeded based on Step 2.

6. **Act on the results.** If the change succeeded, decide if further action is necessary. If the change failed, try something else. Record what succeeded in the general design guidelines for the web site.

Useful web measurement methods and the information garnered

Usability studies measure how patrons actually use library web sites. These studies have resulted in many web style guidelines that can be incorporated into any web-site design plan. The reader is strongly encouraged to use these guidelines in developing basic web-site designs. Although not a comprehensive list of usability study methods, the following outline of methods and web statistical procedures is especially efficient and complementary in determining library users' web usage. Each section will tie the nature and results of the method to library web-design procedure. Each result eventually leads to web-site design questions, changes, or decisions.

Evaluating the design and prototyping stage

Refining the purpose, content, and intended audience of the library web site is the major task of the first stage in the design. Test the purpose of the site by determining whether the information or services presented fulfill a real need. Content usability is heightened by matching the terminology, hierarchy, and layout so that hyperlinks make sense, providing proper associations between hyperlinks and providing efficient routes with sufficient navigation cues. Confirm the audience to assess whether users you wish to serve are being served. Surveys, cluster analyses of service categories, and aggregate statistics

can aid in evaluating the prototyping stage before the web site is created or re-designed. This will increase the chances that the new web site will serve the intended users and their needs.

Surveys

General surveys published and distributed in electronic and print format can determine users' feelings about how an actual or potential library web site should be designed. Major services can be identified as priority services. These surveys will also provide confirmation on what was done right and possible gaps in the service or design structure. In this way, needless backtracking can be avoided and initial prototypes will be more successful. Following is the online survey Oregon State University Valley Library uses. (For a detailed description see: http://osu.orst.edu/dept/library/libweb/survey.htm).

The Valley Library is currently reviewing the design and organization of its web pages. To help insure that our web pages meet the needs of our users, we would appreciate a few minutes of your time to answer the following survey questions.

In completing this survey, you are participating in a research project to evaluate user responses to the OSU Libraries Web. The ultimate purpose of the project is to improve the usability of the library web. As an anonymous participant, your confidentiality will be maintained. If you have further questions about this project, please contact Loretta Rielly, Head of Reference & Instruction, (541)737-2642. For questions about your rights as a research subject, please contact the IRB Coordinator, OSU Research Office, (541)737-8008.

Question 1: What is your status?

freshman/sophomore   junior/senior   graduate student   faculty
staff   community user

Question 2: How would you characterize your level of familiarity with the World Wide Web?

   beginner   intermediate   advanced

Question 3: In the past year, how often have you used the Library's Web page?

   this is my first visit   daily   weekly   monthly   occasionally

Question 4: What kinds of services or tools do you use on the OSU Libraries Web Site? Please be as specific as possible.

Question 5: How successful are in you in finding the information that you want on the OSU Libraries Web Site?

   very successful   successful   somewhat successful   not successful

Question 6: How would you change the presentation of information on the OSU Libraries Web Site? What would make it better?

Question 7: What services or resources would you like to see on the OSU Libraries Web Site?

Question 8: Overall, how satisfied are you with the Library's Web Site?

very satisfied
somewhat satisfied
not satisfied

Question 9: If you responded to the question above with "not satisfied," please comment on the source of dissatisfaction.

Cluster analysis of web service categories

Determining how to arrange categories, the terminology, hierarchy, and layout so that hyperlinks make sense is one of the goals of cluster analysis. This design technique provides the associations between hyperlinks and possible navigation cues that users select to enhance web navigation. Cluster analysis attempts to map the users' mindset and classification of services to the web-page creators. Card sorting, in which library users freely associate terminology on three by five cards, performs the same function as cluster analysis. The statistical procedures used in the cluster analysis method, however, make analyis much easier.

Cluster analysis asks users to categorize a list of possible service content of a web site. This content (service categories) is then grouped by users and statistically added to create the optimum organization. Sub-populations of the sample can be pulled from the study to determine differences in web terminology preferences among different user groups, e.g., faculty, staff, and students. (For a detailed description see: http://osulibrary.orst.edu /staff/matylonj/cluster/sld001.htm).

The ultimate goal of cluster analysis is to

· Gather the demographics of the audience.

· Find which service descriptions the user prefers to lump together.

· Find which service descriptions the user considers important, lacking, or irrelevant.

· Reveal the preferred terminology of the user.
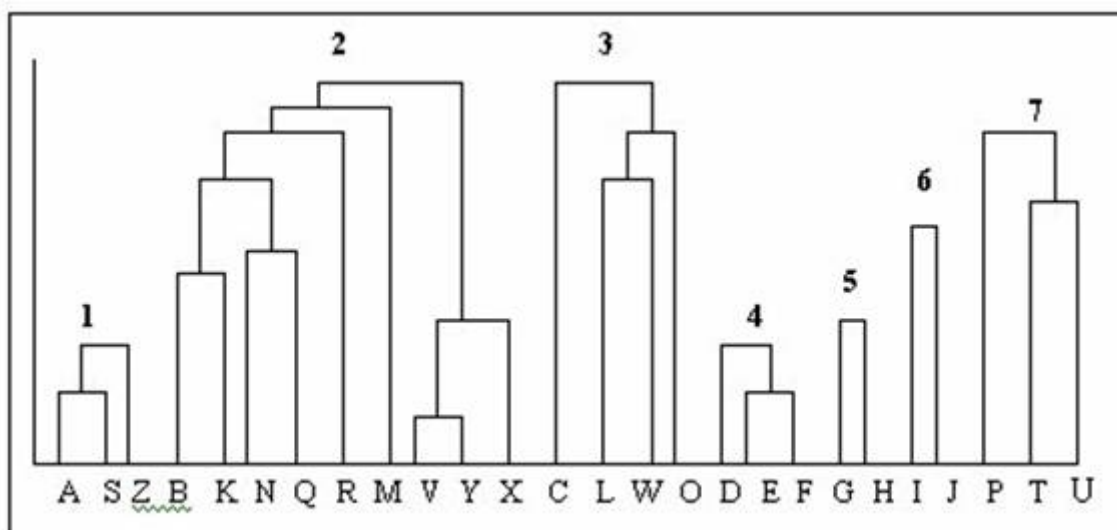
Methodology

An alphabetized list of service categories, representing all the content of the site, is shown to users. The user is asked to freely associate these categories in broader groupings, and these groupings are then statistically added and graphed to show the natural groupings of a population of users.

Library and Computer Service Categories

A. Library collections and what they contain

B. Problem resolution procedures for the network
C. Telephones and pagers
D. Newsletter of Information Services
E. Contact list of staff working for Information Services
F. Organization chart of IS; outline of responsibilities
G. Description of technology resource fees
H. Student computer lab resources
I. Resource for faculty developing multimedia for instruction
J. Databases to support university administration
K. Assistance for operating systems, software, mainframe, dial-in problems
L. Signage, printed materials, illustrations, logos, and animations
M. Acquiring site-licensed software
N. Dorm computer connectivity
O. Adaptive technology for visual, speech, hearing, mobility, or learning special needs
P. Departmental (not run by IS) computing labs
Q. Network wiring for computer and phones
R. Central campus network description
S. Articles, books, newspapers, government information
T. Network support for departments
U. Computer classroom scheduling
V. Modem services (dialing-in by computer)
W. Event and documentary photography for special occasions
X. Workshops on instructional design using media, graphics, presentation software
Y. Documents for assistance on installing and configuring software
Z. Library hours

The resulting graph, called a dendrogram, displays all those groups of services (here represented by a letter) that were naturally associated with one another during the survey. The seven groupings shown below may be considered to serve as major headings on the web site. The vertical axis represents the decreasing frequency of association: the smaller the peak, the more strongly it was associated by the target population.

Testing the website

Evaluating prototypes by cognitive walkthroughs

Testing the prototype is a prerequisite to presenting the design to the public. The library web site can be put through its paces by submitting the design to an information-retrieval challenge based on what the library web designers believe is important. A cognitive walk-through is best performed as a one-person focus group. It is best to be encouraging and helpful while not giving the answer away. By encouraging free association and exploration, obstacles to navigation become conspicuous.

Users are physically observed operating the web site. They are asked to find information derived from the priorities identified from the online surveys, content analyses, and the educational agenda of the site. The subjects are encouraged to verbalize their search strategies and to explore the hyperlink options. These strategies are audio-recorded. The actual performance of the subject is recorded on a work sheet. These are analyzed later to determine any web-design elements that act as obstacles to navigation and retrieval. The most valuable information is the time, number of alternate paths, and major search strategy used in finding the information.

A sample cognitive walkthrough worksheet follows.

Challenge Question: Find a journal article in your major.

Topic: Education

Strategies and path used:

Attempt 1—Subject Resource Guide …Education…Educational Abstracts

Attempt 2—Electronic Reference Center…InfoTrac Databases …???

Attempt 3—Electronic Reference Center…Database list

Comments by Subject:

The subject guide has too many links to get to the databases.

Obstacles observed :

‑

‑
She didn't know what oasis is.

Changing the web site: Decisions, actions, or questions

The major aims and functions of the library web site are validated from the survey results. Significant omissions in services and information content are also uncovered from the survey. At a more detailed level, the cluster study uncovers user preferences for web categories and associations. From these decisions, layout, terminology, and explanatory phrasing of the web site can be modified to reflect user preferences. The above example shows that the computer services available to students (Peak 2 -B, K, N R, M, V, Y, X) are very loosely associated. In this case, these services could be made more apparent by explanatory phrasing or better definition and promotion.

Often, paths that lure users to dead-ends can be identified. Sometimes explanatory material meant to help can be shown to be an obstacle.   Sometimes not enough explanation is offered. For instance, during our last study, students simply did not recognize narrower subject headings linked under broader headings.

Preferences for certain placement of links can become apparent in the design. One user was biased toward links on the left side of the page. Another did not find the search engine window buried beyond the observable window. These and other phenomena will become apparent with more studies.

Evaluating an ongoing web site

Once the web site is made available to the public, ongoing aggregate statistics can indicate whether the site is achieving basic goals. Web statistical software can aggregate web performance measures over time, which is especially useful if baselines are observed at the onset of record-keeping. Some useful web performance measures include:

Top visitors and domain activity

Identification of library web users and where they are accessing the library web site is important for several reasons. In the past, library gate counts, circulation statistics, etc. have provided a real gauge to library usage. Now, the library user community must include remote network users. The relative impact of remote access versus in-house service will become more important in library resource budgeting. In addition, these users have special technical needs that often depend on the local internet service provider (ISP) and quirks of their local networks. Beyond the technical aspects of access, distance education initiatives have guidelines for service and content that may be enhanced by knowledge of where users are accessing the library web site.

The following table shows the breakdown of campus ISP users of the library web site versus those users having private ISPs.
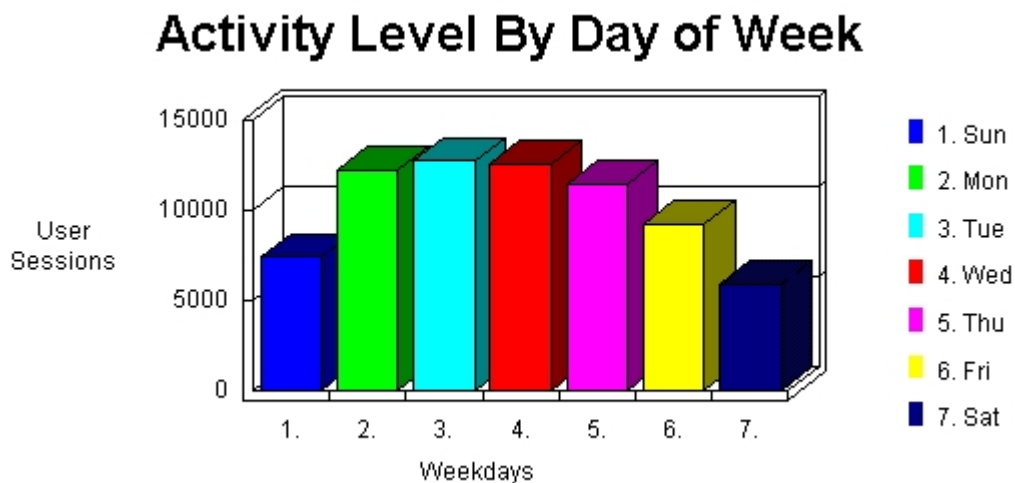
	Most Active Organizations			
	Organizations	**Hits**	**Percentage of Total Hits**	**User Sessions**
1	**Oregon State University** Orst.edu	19,068	50.55%	4515
2	**Rubis Network** Rubis.net	1,128	2.99%	45
3	**America Online** Aol.com	852	2.25%	551
4	**Peak.org** Peak.org	439	1.16%	116
	Subtotal for domains	21487	56.95%	**5227**
	Total for Log File	38369	100%	**9178**

In the table above, 712 user sessions (defined as activity by a unique IP address within any thirty-minute time span) or 7.75% originated from remote internet service providers. This number ought to be watched closely and considered in developing web services.

Improvement in remote network services often depends on how and where users are connecting to the site. For instance, library web sites are increasingly turning to full-text sources and databases that have licensing and access restrictions that may require proxy-server translation. Developing agreements with these companies or considering other access agreements may be necessary.

Peak times

Statistics regarding peak times of usage can be useful to network administrators who need to know the optimal time to perform maintenance and/or upgrades and will help minimize adverse effects on web-site users who regularly depend on the services. The following graph shows activity level by day of week. Such graphs can also be generated by hour of day.

Activity Level By Day of Week

Popular and most requested pages

The number of user sessions experienced by a library web site can show the most/least successful parts of the site. This can determine whether the placement, prominence, and promotion of services on the web site are working, especially if current statistics are compared with a baseline. Whenever there is a change in placement, prominence, or promotion, developers and maintainers of the site can determine effect.

Most Requested Pages					
	Pages	**Views**	**Percentage of Total Views**	**User Sessions**	**Average Time**
1	Oregon State University Valley Library http://osu.orst.edu /dept/library	48,642	17.78%	31,670	00:00:42
2	Oregon State University Electronic Research Databases http://osu.orst.edu	16,931	6.19%	12,279	00:00:43

	/dept/library/database.htm				
3	**Oregon State University FirstSearch Access Page** http://osu.orst.edu /dept/library/fsaccess.htm	13,095	4.78%	9,100	00:04:51
4	**Oregon State University InfoTrac AccessPage** http://osu.orst.edu /dept/library/iac.htm	9,804	3.58%	7,037	00:03:33
5	**302 Found** http://osu.orst.edu/cgi-bin /library-oasis/	8,827	3.22%	7,431	00:04:53
	Sub Total For the Page Views Above	**145,600**	**53.24%**	**N/A**	**N/A**
	Total for Log File	**273,450**	**100%**	**N/A**	**N/A**

Path analysis

A path (sometimes called a thread) shows us how groups of users navigated through a web site. Cluster analysis tries to determine these likely paths before the web site is created. The relative activity of library web users may be determined by finding the most likely paths users take. If you have linked your library web site from obvious gateway pages or have configured home-page defaults on library workstations, you may believe users are naturally disposed to using your site. In libraries that have unrestricted web access, however, the plethora of hyperlink choices may lead to many alternative sources of information, some of which may be considered more convenient than the library web site resources.

Designing a library web site for users outside the library is even more demanding. These users cannot be helped by walking up to the reference desk and will many times choose the most expedient and convenient network information sources, often not your library web site. Constructing an environment where users are intuitively led to the resources would be optimal. Competing goals for library web sites, however, almost always hamper this goal.

Library web designers can also study various web performance factors that may provide a picture of how a site is used.

1. Entrance—Studying what percentage of users enter a site on the homepage in comparison to the percentage who go directly to a page within a site can help determine where to put crucial services and information.

2. Intermediate pages—If the path to crucial services requires specific paths, then the most-frequently-used-paths statistics can determine if the web site is leading users to the correct services.

3. Exit—Determine what page tends to be the last page users are on before they leave the site. What may be found is that a large number of users leave a site on a specific page, and typically this is not on purpose. If one of the purposes of the site is to lead users on a particular path to crucial services, then this information can gauge how well the site is performing this goal.

4. Time—The amount of time users spend on particular pages may indicate how effectively they are using that section.

Other uses of log analysis

Browser analysis

Lynx is still used by some people who access the Web with older computers that cannot support the graphical browsers as well as by users who are accessing the Web over slow modem connections. Today's Web is obviously a lot more graphical then the Web of three years ago; however, we should design our site so that users with text-based browsers can navigate through and so that we meet ADA standards.

Search engines

Many referrals that a site receives are from search engines. Analyzing the web-site logs can tell us exactly what words users entered within the search engine request. Knowing this information can help us determine what terms should be used in our meta tags, which may help others find the site more easily from within a search engine.

Keeping track of error 404 messages and bad referrer links

Whenever changes occur on a web-site, the bookmarks and links people have made are not changed. If we haven't created "re-directs" (pages or files that refer people to the proper URL) for the changes this will show up as a "404 file not found" to the user. Log reports note the offending URL and referrer. Internally, maintenance of bad links can be performed. Externally, owners of web-sites with obsolete or bad links can be informed and your links modified or deleted based on the response.

Conclusion

Libraries have always tried to remove obstacles to information access. A poorly designed web site is certainly a barrier to the library user. A systematic method of designing, testing, and evaluating a web site can insure quality and enhance usability of the site. There is no need to guess or intuit what your library users are doing when visiting your web site. By creating benchmarks for use, carefully experimenting with design changes, and then testing the results of those changes, library web developers can improve the experience of the user and provide better service.

References

Chisman, Janet K., Diller, Karen R., and Walbridge, Sharon L. "Usability Testing: A Case Study at Washington State University." *College & Research Libraries* 60 (November 1999): 552-69.

Dickstein, Ruth and Mills, Victoria. "Usability Testing at the University of Arizona Library: How to Let the Users in on Design." *Information Technology and Libraries* 19 (September 2000): 144-51.

McMullen, Susan. "Usability Testing in a Library Web Site Redesign Project." *Reference Services Review* 29 (1): 7-22.

Levi, Michael D. and Conrad, Frederick G. "Usability Testing of World Wide Web Sites." Bureau of Labor Statistics Research Papers. Available Online [August 2001] URL: http://stats.bls.gov/ore/htm_papers/st960150.htm.

For More Information

Links to library web site usability studies. Available online [August 2001] URL: http://www-hsc.usc.edu/~benjes/usability.html.

Cluster Analysis in Web Design. Available online [August 2001] URL: http://osulibrary.orst.edu/staff/matylonj/cluster/sld001.htm.

Webtrends User Manual. Available online [2001] URL: http://www.webtrends.com /products/log/default.htm.