VISUALIZING WEB USAGE:
USING DATA VISUALIZATION TO IMPROVE WEB SITE PERFORMANCE

A Thesis
Presented for the
Master of Science Degree
The University of Tennessee, Knoxville

Murray Browne
August, 1998

DEDICATION:
To non-traditional students

Acknowledgments

I would like to express my appreciation to the people who are responsible for helping me complete this thesis. Besides the usual thesis mechanics of proper form and methodology, "Visualizing Web Usage" required assistance from many different areas of expertise -- scientific communication, information sciences, graphics, Internet technology, and mathematics to name just a few. Fortunately for me, these subject experts willingly shared their knowledge and provided encouragement throughout the course of this research.

First of all, I would like to express my gratitude to the committee chairperson Dr. David Penniman, who understood the "big picture" of this project from the beginning and to Dr. William Robinson who concentrated on the equally important finer details of this work. I also appreciate the insights of Dr. Carol Tenopir and the graphic expertise of Dr. Julie Little of the Innovative Technologies Center.

Besides the committee members, I must also thank Shawn Collins of the School of Information Sciences for his technical help and Dr. Gonghua Liu of the Innovative Technologies Center for his assistance in transferring this work back and forth from 3-D to "flatland." I also appreciate the support of Dr. Michael Berry of the Computer Science Department who introduced me to the wonders of matrices.

But this is only half of the thesis story. This research involved studying how the web is being developed and used by a well-respected, scientific communication organization, the Human Genome Management Information System (HGMIS) led by

Betty Mansfield. Every student likes his or her project to be relevant and this one certainly was, thanks to her trust in my abilities. I would like to thank Mansfield, HGMIS web master Sheryl Martin, and the rest of the HGMIS staff for their help, support, and friendship throughout my three-year tenure there.

Abstract

As the World Wide Web continues to grow in popularity, better tools of understanding, web usage are needed. One popular system of measurement is tracking the number of "hits" (or files transferred) to each page of a particular web site. In addition to the web industry's misgivings about using "hits" as units of measurements, most reportage of web statistics is done as a series of bar and pie graphs or large printouts of tabular data, which fail to account for a site's structure or the link-relationship between the pages. Conversely, academic researchers have concentrated on the structure of sites, but have not added in the usage factor.

In order to improve web statistics reportage, a three-dimensional web visualization graphic was developed using an off-the-shelf, PC based, software package. The Human Genome Management Information System at the Oak Ridge National Laboratory in Oak Ridge, Tennessee provided the usage data from their web site for the month of January, 1998. Included in the graphic were cubes representing pages (URLs) and the relationship of the links, which was built based on a link matrix designed by Botafogo, Rivlin and Shneiderman. The graphic followed the principles of good data visualization by taking into account the aspects of correct size and proportionality, effective use of color and accurate portrayal of relationships.

The graphic was presented to the HGMIS group for evaluation based on information and communication principles set forth by Shannon and Weaver.

In general, the staff found the web visualization graphic representing 26 pages and 276 links (representing approximately 36% of the activity on the site) appealing and functional. Some minor changes at the web site were made as the result of the graphic.

With its ability to compress usage data accurately and reveal semantic relationships, the web visualization usage graphic (even though it is still evolving), is one of the tools that can help clear up some of the current ambiguities surrounding web usage.

Table of Contents

1. Introduction

Last year, *Wired* magazine's media critic Jon Katz (1997) wrote "to browse the Web effectively, you almost have to be a college student or unemployed; no one else has the time." By accepting the slightest portion of Katz's sarcasm as the truth, one can only wonder if no one has time to browse the World Wide Web, what does that imply about those who dare to measure its use? Moreover, the Web still enjoys the flash and excitement of being a relatively new creative medium, further dwarfing the "dull and dreary" business of identifying who is visiting individual web sites and what they do once they get there.

For better or worse, there is a driving force behind the web usage measurement cause ---- the marketing and advertising community, which reminds one of Marshall McLuhan's time-honored adage that "new media typically borrow from old media." In other words, today's attempts at web ratings are not unlike the early days of television, which began developing rating methods back in the 1950s (Comstock, 1989). Furthermore, it seems logical that no new medium could be introduced without those who wish to generate revenue via advertising. In this respect, the Web is no different, and there are a myriad of commercial enterprises (including TV rating stalwarts such as Nielsen and Arbitron) who have a strong interest in determining how people use web sites.[1]

At this stage in the Web's brief history, pinpointing usage has proven to be a difficult task as web researchers and marketers alike struggle to understand and measure

[1]Of course, not for profit sites such as those of government agencies, which are interested in web usage as well. They are interested in "customer satisfaction by making their web sites relevant to the users' need." (Mansfield, 1998)

how people use a web site. According to Ariel Poler (1997), founder of I/PRO, a Web measurement vendor, measuring web usage suffers from the same state of constant flux as does the Web. Browser technology changes rapidly, forcing designers to rebuild sites and creating a need for improved measuring technologies to test the effectiveness of those designs.

There are three basic problems with measuring web usage: (a) defining and measuring the users in terms of demographics, (b) the nature and extent of how a site is being used and (c) accurately summarizing the massive information generated from *a* and *b* into a form analysts, marketers and Web designers can use to more easily comprehend how sites are being used. This thesis addresses these problems by building a graphic representation that includes: how many users visit which pages, the relationship between the site's link structure and these pages, and the navigational tendencies of those who visit a particular web site.

Defining and Measuring Users

To date, current Web usage statistics have concentrated on the size and demographics of the user population --- often with inconclusive results. CyberAtlas (1997) <http://www.cyberatlas.com>, a Web-based collection of marketing studies, referred to the disparate studies on user population "as one of the most contentious of Internet issues." Novak and Hoffman (1996) in their study of Web measurements gleaned from various Web measurement agencies, reported nine definitions of the term "user" and “visitor” ranging from “a uniquely identified person” to “anyone who visits the site at least once.”

Distinguishing between the types of the users is just the beginning when studying web usage. There are also discrepancies surrounding “hits” or “files transferred,” the basic unit of measurement in Web land. These terms are tossed about rather loosely, implying that when a site receives a “hit” some individual has purposely visited a site, whereas that is not necessarily the case. According to Novak and Hoffman, hits are a weak measurement in determining Web traffic:

> While the definitions of hits are quite consistent, the weakness of hits as a valid measure of traffic to a Web site is quite evident. Since hits include all units of content (images, text, sound files, java applets) sent by a Web server when a particular URL is accessed, hits are inherently noncomparable across Web sites...Other than ignorance of the meaningless of hits, the only reason we feel a Web site would report numbers of hits is that this is typically a large and very impressive sounding number. (p.17)

Evidence of Novak and Hoffman's contention that hits are tossed around with little regard to their meaning is not difficult to find, especially in the popular media. In an article about hyped-up sports web sites, *Sports Illustrated* used "hits" to illustrate the popularity of Hall of Fame pitcher and notorious spitballer Gaylord Perry (13,227 hits) compared to NASCAR driver Jeff Gordon (32 million hits). *The Knoxville News Sentinel* in its story about National Football League consultant Gil Brandt wrote (without any explanation) that "his (referring to Brandt) NFL draft analysis on the Internet went from 1.2 million hits in two days three years ago to 93 million hits in three days earlier this year." (Hyams, 1997). Even *The New York Times* flippantly refers to the growth of web by reporting that the "various Mars Web sites were visited more than 200 million times." (Tierney, 1997)

"Hits are the 'weeds' of Internet measurement," writes Poler, "they won't die." However, despite his railing about the continued use of hits, Poler believes that the technology of tracking web activity is rapidly improving to the point that an individual's activity on a web site will not only be routinely collected and tabulated, but also a better indication of the number of individuals who are visiting.[2]

Problems in Creating Graphs of Web Statistics

Data Overload

One major problem concerning web usage is the magnitude of the data being generated by the server's log analyzers. A log analyzer is a program that runs on the server capturing data representing activity on the server. In general, log analyzers provide such information as which files were accessed how many times, Internet protocol addresses, and dates and times of accesses. Some logs analyzers provide "referrer" information on the URL of the Web site that linked to your site. Some log analyzers will identify the brand and version of browser software used by visitors. These log files can be quite large and often have to be archived to free up disk space. (Allen and Kania, 1997).

For example, the Human Genome Management Information System (HGMIS) at the Oak Ridge National Laboratory (ORNL) estimates that in the last six months of 1997 between 9,000 and 17,000 users visited the HGMIS site each month at <http://www.ornl.gov/hgmis>, accessing approximately 1700 unique text files that make up the Web site. In this same period, between 100,000 and 180,000 files are transferred per month and the HGMIS home page average approximately 13,000 "hits."

HGMIS uses wwwstat HTTPd Logfile Analysis Software (via the Oak Ridge National Lab) to track web use of the files (URLs) those users visit, but it also lists the unique host machines where the information request originated. This creates a large data file of nearly 1 million characters. It would require approximately 350 8 1/2 by 11 pages to print out the entire HGMIS log for a single month.[3] Fortunately, the data can be viewed using HTML format, but a web analyst must still wade through a mass of data in order to generate a simple usage chart.

This example illustrates the staggering amount of data generated, an amount that can be expected to grow with more sophisticated log analyzers. Such sophistication includes collection of “click through” data which is data on how a web surfer navigates from one page to the next (how long between clicks, from which link to which link, etc.).

The Dynamic Nature of Web Sites

Another problem with interpreting web statistics is the ever-changing nature of a web site. Files are constantly added and deleted, links are changed, or sites are

---

[2]Moreover, despite the complaints about the accuracy of "hits," there still isn't any widely accepted replacement for the "hits" unit of measurement.

[3]Approximately 70 pages of such a printout would record the number of hits by URL. Most of the data generated is a list by machine address (reversed subdomain) of who visited the site.

reorganized, which sometimes causes a dramatic change in user habits. Again using the HGMIS site as an example, a major restructuring of pages went into affect in July, 1997. While doing customary web site data collection for that month, HGMIS staff noticed that one of the URLs in the HGMIS family of pages (<.../project/history>) experienced a dramatic drop from 317 hits in June, 1997 to only 17 hits in July, 1997. (In the peak month of April the <.../project/history> URL had over 1400 hits.) Investigation by the HGMIS web master revealed that the recent reorganizing had cut the main link to the <...project/history> page, which led to the precipitous 90 per cent drop in visits. (The page has since been removed permanently.)

This may be an overly dramatic situation, but it does demonstrate that if sites are reorganized, there is the potential to generate a different set of usage numbers even if the number of visitors remain constant. This can make comparisons and trend analysis misleading.

Graphing Web Usage

Adding to the confusion about Web usage is the sparse collection of user statistics being generated and displayed from individual Web sites. These user statistics, which sometimes appear courtesy of webmasters of the individual Web sites, take the form of simple counters, bar graphs, line graphs, or a simple "data dump" of hosts, lists of time of files transferred and lists of total files transferred . (For examples: See Appendix A).

At first glance, such Web-use statistics are impressive, but closer examination usually reveals that the numbers show little about how people use a Web site. According to Novak and Hoffman, such numbers fail to:

> ...provide even the most basic descriptive statistics about their sites, including how many unique visitors are coming, how often users visit, where they tend to come from, how long they stay, the average number of pages per visit, the four or five most popular navigation patterns through the site, the most popular pages, the least popular pages... (p.17)

Nowhere is this more evident than on Novak and Hoffman's own set of links at the New Metrics for New Media web site at <http://www/2000.ogsm.vanderbilt.edu /novak/web.standards/webstand.html> where the authors list 17 Industry Players in their Visitor and Measurement Analysis section.[4] (For a summary table listing their product lines, usage statistics and preference of informational graphic, see Appendix B.)

Browsing through the web sites of the 17 companies several conclusions emerge:[5]

w The "iffy" nature of the Web measurement business. Of the 17 web measurement companies listed by Novak and Hoffman in their 1996 survey, four have gone out of business (Net Genesis, Web Watch, Open Market, and Group Cortex, Inc.), while two others (Interse and Net Count) have been purchased by

---

[4]Allen Marketing Group, Andromeda, Bien Logic,Clickshare, Everyware, Group Cortex Inc. (out of business) Inc. Interse, (now owned my Microsoft) I/PRO, Net Count, net.Genesis (out of business), The NPD Group, Inc., Open Market (out of business), Software, Inc.,Webthreads,WebWatch (out of business), W3.Com.

[5] One caveat concerning this summary is that not all the companies chose to display their graphic representations on the Web. Therefore there is the possibility that more elaborate representations of web usage data do exist.

competitors, Microsoft and I/Pro respectively. Also, Arbitron and Nielsen standard bearers in the television ratings industry, have formed alliances with Interse and I/PRO.

w Usually web usage software is part of a web site management product mix such as transaction software. For example, Guest Track, a product from Allen Interactive, Inc. is just one product in a mix that includes interactive marketing and web site development.

w Most of the packages offer interfaces to standard log analyzer software.

w Most of the packages offer the standard fare of usage statistics:

- usage by hour, day, (i.e. hourly, daily server load)
- accesses by domain (the IP address of the machine which visits the site)
- breakdowns by country, geographic location, or domain (.com, .org)

But most do not include:

- navigational patterns or
- relationships between the files transferred

w Of the companies that presented examples of their work, most of the graphic representations of statistics were bar graphs, pie graphs and tables. This use of informational graphics was how companies elected to show the "analytical" nature of their services. One exception was Everyware's product Bolero, which exported its Web server data into an SQL relational-based database or Oracle database.

The most troublesome aspect of this brief survey of web analyzing reports is the wide use of what Cleveland (1994) derisively calls "pop charts," a collective term for pie charts, divided bar charts, and area charts, which he believes are better suited for business and the media than for technical or scientific information. Edward Tufte (1983), an expert in the area of graphic representation, also has serious reservations about the use of bar charts and pie charts, which he says lack "data density" (too little information for the space they take). According to Tufte, "a table is nearly always better than a dumb pie chart; the only worse design than a pie chart is several of them."

Thus an over-reliance on "pop charts" certainly opens a window of opportunity for the information designer. Anne Senechal (1997), in her cover article for *Adobe* Magazine on the importance of designing information systems that are both appealing and functional, summed up the current state of chaos and opportunity:

> The shapeless morass of information engulfing the Internet has only complicated matters further. On the one hand, the Internet leads us to believe we can find anything, anytime. On the other hand, it sucks us into a vortex each time we try. Yet, by bringing the world to our doorstep, the Internet has made a convincing case for information design. (p. 40)

2. Literature Review

Selecting a display to communicate information or data visualization effectively has always been a relevant topic in information sciences because data visualization involves aspects of information communication, information content analysis and information content representation.

Data visualization as information communication. In Claude Shannon and Warren Weaver's generic theory of the levels of communication, the properties of data visualization and communication are strikingly similar. Shannon and Weaver (1949) approached communication problems from three perspectives: (a) The *technical* problem of how accurate symbols are communicated, (b) the *semantics* of how precisely the symbols convey the meaning (c) and how *effective* the communication is in changing behavior. The same three perspectives are relevant to data visualization. For example, in building a web usage visualization graphic based on the Shannon and Weaver perspectives, the developer should have the three aforementioned criteria in mind. The symbols representing the pages and links must be consistent and mathematically correct while offering possible interpretations of web usage. Finally, is the graphic convincing enough to cause the web master to make changes in the web site's design?

Data visualization as information representation. The use of maps, the most common method of visually displaying information, dates back centuries, but the increase of interest in data visualization comes as a result of the computer's ability to generate large amounts of data as well as software that can display the data (Hall, 1991). Another contributor to this increased interest is Edward Tufte. Tufte's influential books, *The Visual Display of Quantitative Information* published in 1983, *Envisioning Information* published in 1990, and the most recently published *Visual Explanations* (1997), are dedicated to explaining effective ways to display information in charts and graphs. His books cover topics, ranging from choosing the graphic display (text tables, semigraphics), to the density of the data, to the multi-functioning of graphic elements (i.e. not only are the numbers are important, but how we arrange those numbers), all of which fall under information sciences' representation umbrella. Just as the library index card represents a book in the stacks, a nutrition label represents the contents of a food package, the web usage visualization graphic represents the web site.

Data visualization as information content analysis. One of the changing roles of information science involves the shift from librarian-curator to information analyst and manipulator. Some information professionals are asked to evaluate and/or synthesize information. Information science professionals can use graphics to explain how such information is classified or synthesized. Most notably, graphics are continually being developed in the information retrieval field of information sciences.

Gerald Kowalski, (1997) in his book *Information Retrieval Systems* dedicates an entire chapter to visualization of retrieval systems. According to Kowalski, there are two aspects of information retrieval that must be addressed -- organizing the documents that come back from the query and analyzing the search statement. Kowalski describes the basics of several such systems ranging from the three-dimensional Information Visualizer developed Xerox PARC, which uses a mixture of formats to show the hierarchies of a database, to Mitre corporation's product, which "enables a user to see a structure visual representation of the information space they have navigated through." (Kowalski, p.200)

In related research, Lin, Marchionini, and Soergel (1995) in their study of map displays of retrieved documents asked human subjects to take 133 documents, whose titles had been placed on note cards and arrange them on a two-dimensional grid by subject. Later, Lin (1997) developed a similar map from information retrieval data and placed it on a computer screen. Lin also developed maps to represent the contents of a researcher's personal document collection. Lin’s work shows how information such as documents from a search or a collection can be mapped in a meaningful way to display relationships in the information, while further demonstrating the increasing role of visualization plays in information sciences.

This changing role of using graphics in information sciences can also be expanded to demonstrate how people use and navigate through individual web sites. Although the literature does not yet address work in this area, groundwork has been laid in developing metrics to analyze the structure of web sites.

Using Matrices to Develop Structures of Hypertext Documents

In the beginning of his 1983 book *Semiology of Graphics*, Jacques Bertin wrote:, "A graphic should not show only the leaves; it should show the branches as well as the entire tree." Those words seem to be encapsulate the first goal of visualizing web usage --- to develop an information graphic that reveals general structure of site use. Moreover, this structure can serve as a starting point to compute the activity within the site. Bertin's quote also indirectly describes the work of Rodrigo Botafogo, Ehud Rivlin and Ben Shneiderman (1992) whose research on identifying hierarchies, and developing metrics for hypertext, is one of the few articles that deals specifically with hypertext structures:

> User interface techniques and textual analysis do not make use of an important part of hypertexts: their structure. A hypertext is usually defined as a set of nodes that contain information, and as a set of links that connect interrelated nodes. To be understandable, a hypertext must be carefully crafted; each superfluous or missing link contributes to user disorientation. A link between two nodes is not only glue that keeps the hypertext together, but each link represents a semantic relationship between the nodes.
>
> Authoring hypertexts is not an easy task. Although authors are improving their skills, as hypertexts grow, management becomes extremely difficult even for the most qualified professionals. So software tools that would help hypertext authors see global and local properties would be beneficial. Similarly, metrics that would indicate the hypertext complexity of special properties of nodes would be welcome (p.143).

In the lengthy article, "Structural Analysis of Hypertexts: Identifying Hierarchies, and Useful Metrics," Botafogo, et. al. studied the structures of three hypertext systems by computing matrices and then graphing them: the University of Maryland Computer Science department which had 106 nodes and 402 links; a volunteer archeology web site, which had 222 nodes and 1609 links; and Hypertext Hands On!, an electronic book of 243 nodes and 803 links.

To develop these matrices, they counted the number of nodes, links and the "distances" from one node to another. If Node A links to Node B, and B to C, (and node A doesn't directly link to C) then the "distance" value is two.[6]

Using the converted distance matrix, Botafogo et. al. were able to compute which nodes were the central nodes, "ones whose distance to all other nodes in the hypertext is small" (p. 147) and the root, defined as a node "that has to reach every or almost every node in the hypertext" (p. 148). By identifying such nodes, Botafogo et. al. were able to represent the hierarchy of a site and determine which links were hierarchical in nature and which were more cross-referential.

But Botafogo et. al. were not content with just showing the hierarchies and cross references, they wanted to be able to compare the structures of different hypertext documents. One way was to develop a formula for determining a site's compactness and strata -- again using data from the matrices. In short, compactness in the site is the aggregate distance between nodes. In a high compactness site, nodes can easily reach another in hypertext. Conversely, low compactness would indicate that many links between nodes are missing. (Compact values range anywhere from 0 to 1, with 0 being no links, and a value of 1 would indicate that all the nodes are inter connected.)

Stratum measures the presence of some predetermined order where there is a specific sequence in the node linkage. In other words, the site must be negotiated in a specific manner. Stratum values range from 1, which indicates that the nodes must be followed in a certain order (which the authors say defeats the advantage of hypertext) to a

---

[6] Although this seems simple at first, problems arise when one node doesn't link to another node, creating an infinite value. However, the authors alleviate the problem by implementing a conversion constant and

value of zero, where the user can enter any part of the hypertext document at any node and it makes no difference in the understanding of the document. An example of the latter would be a site that is circular in nature --- node A links to node B, which in turn links to node C, C to D, and the D links only back to A.

The authors also refer to a site's COD value --- "converted out distance" value, which mathematically determines which nodes that are the most centrally located in the site.

Botafogo et. al. designed their metrics to characterize the structure of sites as a new technique for users (authors and readers alike) to identify the key nodes in hypertext document. They maintained that such metrics would lead to better design of web sites and help users to later avoid the "lost in hyperspace" problem.

Developing Structures of Hypertext Using Different Visual Platforms

In March, 1996, IEEE's *Computer Graphics and Applications* published a special series of short reports on using computer graphics and visualization to meet the rising challenge of handling the information expected with the Global Information Infrastructure. Not only did the series include a promotional blurb on the necessity of graphics and visualization by Vice President of the United States Albert Gore, but it contained a warning that graphics and visualization were critical elements in the web's future. "Otherwise, the information universe will become a cemetery, a massive burial site for an infinitely large amount of data and information." (Gershon and Brown, 1996, p. 61)

---

hence, infinite distance values become converted to finite values.

In his report, Stuart K. Card (1996) of the previously mentioned Xerox PARC project team listed some of the approaches researchers use in developing and visually mapping individual web sites or hypertext documents.

- The 3-D Conetree has been used to map over 20,000 nodes on the Xerox web site. Each node is so full of links that it appears to be nothing more than a stack of construction cones. All the cone figures are piled on top of each other, representing the hierarchy of the site. "When the users touches some node in the tree, the parts of the tree rotate, elements around the touched element grow, and less relevant parts of the tree shrink." (Card, p. 65)
- In Web Forager, the retrieved documents in this graphic look like playing cards sitting on a 3-D plane service. (This format resembles the Z-diagram format/MAPA navigational product designed by Dynamic Diagrams <http://www.dynamicDiagrams.com>.)
- In Web Book, the retrieved documents are attached like a book and displayed on a 3-D plane and so the user literally can "thumb through" them to search for a specific document.
- Sticks and Spheres. In the same IEEE report, Gershon (1996) describes a 3-D information graphic, which uses spheres to represent pages (nodes) or groups of pages and sticks to show the links in the hypertext document.

Several designs could be used to develop a visual representation of a web site's structure and how it is being navigated. One can even argue that the best strategy may be to develop several presentations and allow the user to choose his or her own representation. A similar strategy was suggested by Catarei (1996), another one of the authors in the special IEEE issue, who mentioned offering different visual interfaces in a database retrieval system.

No single information graphic representing a web site has dominated the web scape, leaving an open invitation for new visual concepts. However, the literature indicates that over the past twenty years a broad set of guidelines has developed that require certain elements to be used in a particular manner to create an accurate information graphic. These elements are:

- *Size and proportionality.* The size of an object must be in proper proportion in its relations to other objects in the same graphic while maintaining integrity with the original data set. Tufte (1983) takes designers to task for failing in such a seemingly simple task. Both Tufte and Cleveland (1994) stress the importance of maintaining the proper proportions. Tufte writes, "The representation of numbers, as physically measured on the surface of the graphic itself, should be proportional to the numerical quantities represented" (p.56). Bertin (1981) gives designers guidelines and formulas for determining the proper sizes, although he admits that the human eye can distinguish only a limited amount of difference in sizes. Wurman (1996) uses examples of graphic designer Nigel Holmes "The World's Longest River and Highest Mountain " graphic to demonstrate the effectiveness of properly displaying size and proportionality.

w *Links and relationships.* Bertin says that network or topographic drawings are an excellent way to suggest relationships (such as the web.) Not only can varying the thickness of such links indicate size and proportionality, but according to Tufte it is a far more dynamic to the user. Eick (1996), another author in the IEEE special issue, writes that, "the color and thickness represents the strength of the relationship with thicker and brighter lines tied to stronger relationships" (p. 69).

w *Color*. Eick, Cleveland, Bertin and specifically cite color's ability to provide visual coding on a data display. Bertin includes texture, orientation, and shape along with color as methods of increasing associative meaning in a graphic. (Tufte is less convinced that color is an asset; he prefers shades of gray.)

w *Interactive capabilities*. When a user uses a filter, or rotates or zooms in on a graphic, they are interacting with the graphic. Gershon and Brown (1996) more specifically refer to real time updating, using filters, and activiating/de-activating portions as techniques that will improve the effectiveness of a graphic to communicate. Cleveland also sees the advantages of real-time updating of a graphic.

Summary

The literature surrounding the visualizing of web usage can best be described as multi-disciplinary (i.e. information sciences, human computer interaction, mathematics, marketing, and graphic design) and gleaned from a host of formats (technical books, journal articles, newspaper and magazine articles, and web pages). Perhaps this is a reflection of the changing methods of scholarship as well as rapid changes in communication brought on by the development of the worldwide web.

Designing and building an information graphic to represent the structure and usage of the Human Genome Management Information System's web site -- the objective of this thesis -- should help those who are creating, developing and managing the HGMIS web site on a daily basis. Monitoring and evaluating this computerized information system based on this web visualization graphic should lead to better analysis of future information systems of this type and "could improve user productivity by optimizing system performance or data base structuring." (Dominick and Penniman, 1980).

3. Methodology

Designing and building a graphic representation of a web site's usage statistics, required two main steps. The first was to accurately determine the structure of the web site. The second step was to "overlay" how viewers used and navigated such a site. As mentioned in the literature review, other researchers and vendors had worked in these areas but no one had sought to combine structure and usage to create a succinct, accurate and useful graphic representation of how a web site was being used.

In order to compute the structure of a web site, formulas introduced in Botafogo et. al. were adapted to create a matrix of nodes and links, from which a structure was developed. The structure was built in a 3-D computer program called Virtual Home Space Builder (VHSB) 2.0. This program from Paragraph <http://www.paragraph.com> had limited capabilities in terms of its graphic offerings, but it was able to create boxes, (for pages or nodes), thin walls (for links), while adding color, text and URLs for further identification.

To test the ability create a structure from the matrix, a smaller pilot project of 22 nodes and 49 links from the Book Shopper web site <http://funnelweb.utcc.utk.edu /~mubrowne> was built first. It was available for viewing; however, it required a VRML (Virtual Reality Machine Language Browser) browser or VHSB's viewer. This browser is minor part of the Silicon Graphics product mix. It is free and available for downloading at <http://www. paragraph.com>.

There are several advantages of having the graphic run on a VRML browser:

- A VRML browser not only allows a user to see a graphic, but navigate through it.
- VRML is a free downloadable browser, which only requires the user to have Netscape 3.01 or better (such as Netscape Communicator). The graphic is loaded into a standard Web browser window, but there is an extra set of controls at the bottom of the screen that allows the user to navigate by zooming in, zooming out and or rotating the graphic. (The plan to use VRML was eventually dropped. See footnote 12 in the section Visualization Graphic Building Problems.)
- More importantly, the information being collected in these usage studies was about the Web and thus, those most interested in this information are undoubtedly Web users with skills necessary to use the browser. Users of this tool could be web designers, information managers, information science students, advertisers and marketers.

After completion of the pilot project, a larger and more complicated web structure was built and studied for the Human Genome Management Information Systems (HGMIS) located at the Oak Ridge National Laboratory <http://www.ornl.gov/hgmis>. The HGMIS site has an estimated 2000 text files on its web site with an unknown number of links. For this study, a link-node matrix was built for any page which receives over 1% of the HGMIS' site "files transferred." In January, 1998 over 175,000 files were transferred at the HGMIS web site, therefore any page that received over 1750 hits would be included in the graphic. In addition, key pages designated by the HGMIS staff were included to make the graphic more comprehensive.

Adding Usage Statistics

As the representation was built, in 3-D space, usage values based on January, 1998 log data were added to the model. In order to address some of the current problems with web statistics, the graphic model was developed within the parameters of good information design. The following list contains some of the features the graphic representation of web usage had to have in order to be an effective representation of what was happening at a specific Web site:

- *Size of boxes.* Depending on number of hits, nodes were scaled accordingly. The home page cube (node) was considerably larger than, for example, a page (node) from the on-line version of the *Primer on Molecular Genetics.*
- *Color identification.* Color was used to identify different pages.
- *Page-node identification.* The user was able to click somewhere on the graphic to identify specific files (pages or URLs). Graphics built in VHSB did allow URLs to be imbedded in the graphic, so the user could click on the URL and the specific page was downloaded.
- *Most frequently used and least frequently used navigational paths*. This was the most difficult information to include because the log analyzer data usage statistics did not include what is known as "click through" data. However, there

was another acceptable approach. By the development of a matrix of nodes and links, it was possible to "decompose" the matrix using stochastic probabilistic modeling. In other words, given the structure of a certain web site, there was a high probability that the site was navigated in a certain way.

After the graphic was completed, it was tested and evaluated by the HGMIS staff who were the experts/clients interested in this project. In two informal presentations, the HGMIS staff were asked to compare it to their current methods of reportage and determine whether the graphic might cause them to "see their site" differently or change their web site in terms of organization or content.

Limitations

Limitations apparent at this stage included:

- The data could not mark where visitors entered the site despite the knowledge that surfers do not always enter through the home page.
- Data used from HGMIS' January 1998 web site usage logs had to be translated and plotted manually along with number of links between these pages. In a perfect world, top log analyzers from any server that uses Common Log File Format (CLF) which includes ORNL's HTTPd Logfile Analysis Software, Netsite (Commercial and Regular) from Netscape Communications Corporation, CERN httpd, and NCSA httpd would be sorted and dumped into a computerized matrix program.

4. Data Collection

On February 5, 1998 the HGMIS webmaster ran ORNL's HTTPd Logfile Analysis Software for the HGMIS web site at http://www.ornl.gov/hgmis. The program produced two large data files covering web activity from January 1, 1998 to January 31, 1998. These files included the number of files transferred by each URL at this site. One data file originates from the former domain name of the HGMIS web site, <http://www.ornl.gov/TechResources/ Human_Genome/home.html> which was shortened to its current domain name (http://www.ornl.gov/hgmis) in January, 1996. Historically, these data files and the individual files within the site are totaled when activity on the web site is recorded.[7] Currently, reports are drawn from the data files and made on a monthly basis to HGMIS' group leader Betty Mansfield and the webmaster Sheryl Martin. The January 1998 report, which consists of a table of URLs and the number of hits on each file, can be found in Appendix C.

---

[7]The breakdown of hits generated by the TechResources domain name and the hgmis domain name merits some discussion. In general, the "hits" generated off identical pages (i.e....TechResources/ Human_Genome/CAL.HTML and ...hgmis/CAL.HTML) favor the TechResources name approximately 3 to 1. This ratio may reflect the fact that HGMIS in the early days of the web established itself as a popular science site and many other sites linked to HGMIS using the TechResources domain name. It also brings into the limelight the difficulties of those who attempt to establish new sites or a web presence.

Based on personal observation only, one wonders if recently constructed sites have a tendency not to include links that take surfers off their pages. After all, why go through the time and expense of building a site only to provide a convenient way to exit? (The trend is to build what I call "Hotel California" sites---you can come, but you can never leave.) This makes it difficult for newer web sites to find sites willing to link to them. Conversely, more established sites have an advantage in generating traffic.

Also, according the HGMIS webmaster, once a visitor enters a site all the hits recorded during that session are recorded as TechResource hits, which would also explain why the statistics favor the older URL domain name.

Other statistical evidence supports this explanation. On HGMIS web pages that were developed and uploaded after January, 1996 with hgmis domain names exclusively (i.e...hgmis/genechoice.html) the ratio of hits between domains was closer to 1:1 and in some instances favored the new hgmis domain name. Keep in mind that if a user came to the web site through the TechResource URL all subsequent

Twenty-six URLs from the estimated 2000 HGMIS web site URLS were selected to be included in the web visualization graphic. Any URL which received over 1750 hits, reflecting at least 1% of the site's usage was automatically selected. Also several key URLs (with less than 1 % of the site hits) reflecting some of the HGMIS sites' specific content elements were selected. Part of HGMIS' mission is to communicate Human Genome Project information to researchers as well as the interested general public and the web site reflects this mission. The research section was developed to meet the needs of the HGP researchers, but numerically there are fewer hits because there are simply fewer researchers. Other important pages included in the graphic despite their lack of numerical strength, was the Table of Contents navigational page, the page for those interested in locating support groups for those afflicted by genetic diseases, and the acronym and glossary pages. Again each page was selected because it received over 1 % of the hits or it filled a specific information niche in the web site.

Closer study of the log files reports (in Appendix C) revealed several inconsistencies in deciding which URLs were being evaluated.[8]

w Recorded hits for <.../publicat/publications#hgn> and <.../publicat /publications#publications.html> were actually hits for the same page. (In HTML, the pound sign (#) is indication of an internal anchor of the same page.)

---

hits on that visit would log as TechResource hits. (This explains why a URL with the hgmis domain name can log TechResource hits.)

[8]If blame for previous HGMIS web reportage inconstancies must be assigned, the blame should go to the author of this thesis, who was the HGMIS person responsible for the reports. If this thesis accomplishes nothing else, at least it will improve the data collection aspects of web statistics within the HGMIS group.

w HGMIS statistics for the <...genetics.html> were not included in this study or the graphic. Although this URL does shows up in the log stat report, (See Appendix C) this site is actually a separate web entity, which HGMIS is responsible for, although it operates outside the framework of the main site.

Previously, the HGMIS web statistics report provided information on 18 web site pages representing 33% of the files transferred on the HGMIS site. The web visualization graphic provides information on 26 web site pages representing 36% of the files transferred on the HGMIS site. In general, usage on the 2000-file HGMIS site is heavily concentrated on a relatively few files with approximately 1/3 of the activity occurring on approximately 1 % of the pages.[9]

Building the Link Matrix

After the URLs were selected, a link matrix similar to the one used by Botafogo et. al was constructed. Data from the link matrixes was used in the visual representation graphic to determine the structure of the site and the thickness of the links between the selected URLs. The 26 by 26 matrix (676 cells) contained 26 URLs and several extra columns and rows for additional computations. Each designated URL was listed across and down in the same order. Then a worksheet listing each URL (in the same order) was created to gather data on how these URLs were interconnected.

[9]One reason this is such a startling statistic is that approximately 1127 files are files of the *HGN* newsletter and an additional 600 files are abstracts from past meetings of Department of Energy contractor-grantees. These 1800 files are archival in nature.

To determine the structure of a web site and its interconnectedness, links between the 26 URLs were counted. In order to do this accurately, each URL was brought to the screen using the Netscape browser, and the source code was then saved to a PC's hard drive as an HTML (ASCII) file. The file was then pulled into a word processing program and then using the program's search and replace feature, the links from the selected file to the other 26 were counted and the value placed onto the link matrix.

A small section of the link matrix is shown in Table 1 with the complete matrix shown in Appendix D.

By reading across the matrix, one knows that there three links from <...home.html> to <...research.html>; one link to <...toc_expand.html> and two links to <...archive/news.html>. By reading *down* the matrix, one knows that <...home.html> has two links from <...research.html>, two from <...toc_expand.html>, and two from <...archive/news.html>. The link matrix is very useful in organizing the relationship of links between the pages of a web site.

After the link matrix was built, the next step was to compute each node's (page or URL) converted out distance (COD). Botafogo et. al. used the COD formula in order to locate the more centrally located nodes of a particular site. The formula is:

$$COD_i = \text{Sum}_{j} C_{ij}$$

To compute COD, one adds all the converted distance (CD) of links between Page A across the row of selected URLs. If the Page A is linked to Page B, the converted distance is 1. If it is not directly linked, one must count the number of links (or clicks of the mouse) it takes to get the from Page A to Page B. If it takes two clicks to navigate from one page to another, the converted distance is 2. For example, the HGMIS home page is directly linked to the <..research.html> page (CD = 1), the <...publications.html> page (CD =1), but indirectly to the <...funding/fund.html page>

Table 1. Sample of HGMIS Link Matrix. Value indicates number of links between pages.

	home.html	research.html	toc_expand.html	archive/news.html
home.html	----	3	1	2
research.html	2	----	0	1
toc_expand.html	2	6	----	3
arch./news.html	2	2	1	----

One must go to the <...research.html> and from there to <...funding.html> page (CD = 2). The more links one must traverse, the higher the converted distance. To compute the COD, all the converted distances must be added together for a particular URL. According to Botafogo, the nodes with the lowest CODs are the more centrally located pages in the site.

In the HGMIS web site, COD values range from 25 to 71. As one would expect, the home page should be centrally located, and it was with a COD value of 26, but other centrally located nodes were the Table of Contents page (<...toc_expand.html>; 25 COD) and the publications page (<.../publicat/publications.html; 25 COD>. Conversely, the less centralized glossary and the training calendar web pages had high CODs of 71 and 63 respectively.

Building the Graphic

Using HTTPd analyzer log data for the month of January, 1998, the data from the link matrix (including COD values) and Paragraph's Virtual Home Space Builder (VHSB) 2.0 software, a web usage visualization graphic was built. Pages with lower COD values were placed in the center of the graphic and those with higher CODs were placed towards the perimeter. Other characteristics of the graphic included:

- The size of boxes representing pages was dependent on the number of hits that page received.
- The size of the links between pages was dependent on the number of links between pages.
- Different colors represent different pages of the web site (i.e. red represents the home page).
- Each page-node was identified on the graphic.

Size of Boxes

VHSB allows the designer the flexibility to take numerical values such as 1000 hits and convert it to units. By assigning a single unit the value of 100 hits and rounding "hit" totals from the HGMIS log data to the nearest 100 hits, each of the selected 26 URLs was converted to a numerically accurate cube size, which would allow for comparison of hit totals between pages on the web site. For example, the HGMIS FAQ page (<.../faq/faqs1.html>) received 5707 hits in the month of January. Rounding the figure off left 5700 hits, so a graphic the size of 9 by 7 (63 units multiplied by 100 = 6000) was used. Fortunately to improve accuracy, VHSB has a "cutout" tool to remove extraneous units. In this case, for the FAQ URL, six units were removed from the 9 by 7 graphic. (See Figure 1.)

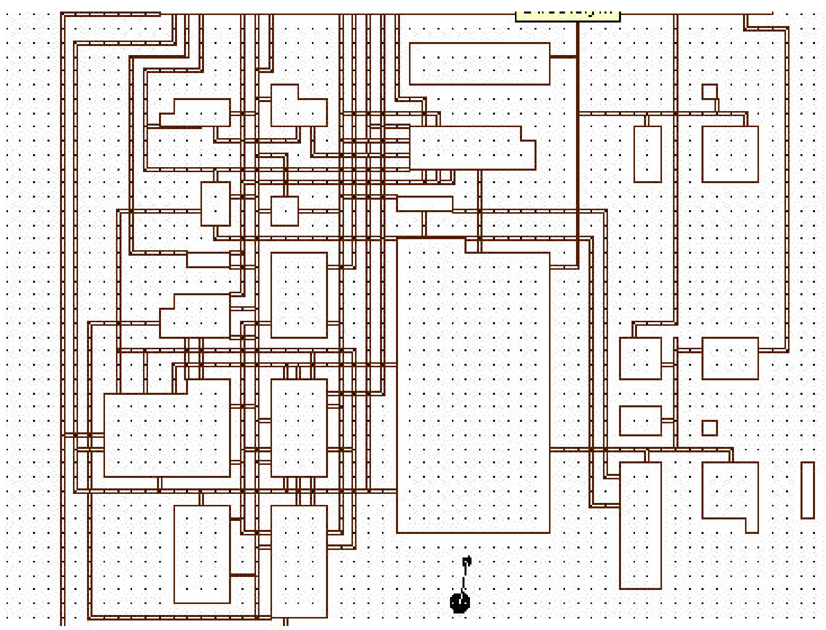


Figure 1: VHSB Design Grid: Note how the irregular FAQ box on the left of the home page representation has the six squares removed.

Although it would have been ideal to have certain types of rectangles or squares represent certain characteristics, the shape selected depended more on how it could fit cleanly into the graphic. This proved to be a good idea because: (a) the hit totals on two of the pages were very small, limiting shape choice (i.e.<...funding.fund.html> and <.../publicat/TKO/index.html> had only 128 hits and 135 hits respectively) and (b) later it became important to keep the graphic from spreading out "too far," making it more difficult to print a paper version. Besides, the VHSB 2.0 software package only comes equipped with cube rectangles and cube squares.

To improve accuracy, VHSB allows the designer to place graphics on an X and Y coordinated grid and cube value density is displayed during the building of the cube graphic.These features were utilized during the graphic construction.

Constructing Links

To build the links between the URLs, the "wall" feature of VHSB was used. VHSB allows the user to control the height of each wall mathematically. Making the wall very short, creates a "square" pipe. Walls also can be made to look "suspended in air." This allows links to criss cross or occupy the same space (but different planes), which is one advantage three dimensional figures have over two dimensional representations.

The height of the walls can range in value from 0.00 to 3.50 units.. (For uniformity in the graphic, the cube height was also set at 3.50) A single link has the value .25 which allows for a possible 14 links between Page A and Page B (3.50/.25). If there were three links between Page A and Page B, a wall or link of .75 height would be constructed.

Color

VHSB's airbrush color feature allows the user to select 256 different colors to color a link (wall) or page (cube). Each side of a link or cube, must be colored individually. Rather than trying to keep track of the colors for 26 different entities and anticipating that the colors could significantly change between different browsers, monitors and eventually color printers, the decision was made to select families of colors to correspond with families or groups of web pages within the HGMIS site.

Pages were grouped and colored by function as shown in Table 2:

Links colors were assigned depending on the original destination of the link. For example all links from the home page are red. A link from <...links.html> to <...resource/elsi.html> would be yellow but a link from <...resource/elsi.html> to <...links.html> would be assigned a tan or light brown color. Using this color scheme was helpful in identifying links between different groups of pages.

Page Node Identification

To enhance page node identification, the upper left corner of each page representation (cube), a small box (known in VHSB vernacular as a picture frame) was hung and a page title was typed in. During the navigation of the 3-D graphic, the user can place the arrow cursor on the frame and a title with the name of the page (home,

Table 2: HGMIS Web Pages Grouped by Function

Page Group	# of Pages in Group	Function	URL examples	Designated Color
Home	1	Main page of web site	home.html	Red
Research	7	Scientific pages for researchers	...research.html, ...TRCAL.html ...informatics.html	Blue
Public Resources				

research, sequencing) along with the number of January hits will appear in the lower left window.

In the upper right hand corner another picture frame was hung and the URL of the graphic representation was typed in. Again, while navigating the 3-D graphic, the user can place the cursor on the URL frame, click, and the corresponding URL will appear as a web page. Even if the user decides not to view the URL, the URL address will still appear in the lower left window to enhance identification.

One unforeseen advantage of the page node identifiers was that the picture frames helped orientate the user to identify the "front" of the 3-D graphic.

Visualization Graphic Building Problems

The main difficulty in building the web visualization graphic came from trying to enter all the links between the 26 pages. In contrast to the Book Shopper prototype, the HGMIS pages are heavily linked (646 links). Even though some of these link drawings can be combined (because a wall can represent more than one link), the graphic quickly became congested and difficult to draw and thus difficult for potential users to navigate.

Another problem was that as box and link elements were added to the visualization graphic VHSB 2.0 slowed down dramatically. Midway through the construction, it took over ten minutes to add a link.

To reduce the graphic clutter, a threshold value of three was set on the link matrix. In other words, connections between pages with three or more links were the ones added to the visualization graphic and connections with only one or two connections were left off. This reduced the number of links from 646 to 253. Because it was by far the most popular page on the site, 11 single (11 links) and 6 double links (12 links) from the home page were also included in this figure. This raised the total to 276 total links.[10]

While reducing the number of links reduced some congestion, it did not reduce all the clutter. In the second attempt to draw the graphic, a more uniform "street like" pattern was adopted to help organize the pages and links better. This type of pattern made it easier to draw links in between groups of pages.

After the graphic was completed, a print version of the graphic was made (See Appendix E [11]) along with an electronic version, which can be displayed on screen and navigated using a VHSB viewer.[12]

Initial Evaluations of the Graphic

Both a print version of the web visualization graphic and 3-D screen representation of the same graphic and a print copy of the link matrix was presented to the HGMIS webmaster on April 7, 1998. The print version was introduced first to show the general overview of the graphic which included the main categories of: the home page, the

---

[10] Considering that, for the most part, the minimum link value was three, it only took 88 walls to represent the 276 links. Unfortunately, VHSB 2.0 continued to perform slowly after the first 20 or 30 link drawings. At the end it would take nearly 15 minutes to draw a single link.

[11] Because VHSB does not have a print (to paper) command, printing the graphic was a challenge. Eventually, using the "Print Screen" command, the graphic was saved as a bitmap file into a Windows 95 clipboard and then pulled into a Photoshop where the titles were added to the major groups of pages. One limitation was that the graphic had to be small enough to fit into screen and secondly, as predicted some of the colors varied. For example, VHSB screen red looks more like the color of dried blood in the printed version.

[12] Not only did VRML navigation prove to be difficult, but it also altered the colors of the graphic and removed all the click-on identifiers, which eliminated the titles and URL identifying features of each web page. The VHSB "user manual" (a readme file) warned users that certain VRML browsers would not be entirely compatible with VHSB-built graphics. Therefore attempts to use VRML were discontinued.

links page, the research pages, the public resource pages and the publications pages. Next the webmaster was shown the electronic representation of the web visualization graphic from a desktop computer. Using the mouse and the walk through capabilities of the software, the entire 3-D graphic was traversed. The webmaster listened to a short informal presentation describing the significance of the color scheme, the box and link sizes and how they were positioned in the graphic. The link matrix served as a reference sheet, when there were questions on how many specific links there are between different pages.

The webmaster's comments about the HGMIS web visualization graphic can be divided into three categories: General reaction, specifics and other uses.

General reaction. The webmaster's initial commentary was that having volume of the boxes dependent on the number of hits gave her “a better of feel” of site activity, more than a list of numbers --- especially in terms of how the numbers related to each other. The webmaster noticed (and was pleased) that a page she had designed (<...resources/ information.html>) was being used by visitors.

Specific comments. After looking at the graphic, the webmaster voiced concerned that there were not enough links from the links page <.../links.html> to other pages on the site. The graphic indicated that if a person came to the web site and visited the links page first, it would be difficult to know about or navigate to other pages on the site. Although one usually thinks of a links page as the place you visit before you leave, the webmaster believed it also might be an entry point for others and an opportunity to draw people into the site to stay longer.

Moreover, there were few links from publication pages such as *Your Genes, Your Choice* (<.../genechoice/index.html>), *To Know Ourselves*, (<.../tko/index.html>) and the *Primer on Molecular Genetics* <.../primer/intro.html> to other locations on the HGMIS web site. These are also pages where visitors may enter the site as a result from a direct link of a search engine or from another web site. The graphic also underscored how comparatively few hits went to *To Know Ourselves* as compared to the other HGMIS publications.

The HGMIS webmaster also voiced concern about the lack of hits on the well-connected and centrally located navigation page <.../toc_expand.html>.

Other Uses. Another unforeseen advantage of the graphic is that it could be modified to be a navigational tool for the web site. One HGMIS concern has always been that various audiences (educators, scientists, high school students...) are not always being routed to the information they were looking for. Secondly, if they did find the place they were “supposed to be,” were they sufficiently aware of other types of material on the web site? The information graphic might be adapted to be used as a navigational tool which can show the "Big Picture" of the site, but then allow the user to go to specific areas of interest easier.

Because of other responsibilities on April 7th, the group leader was unable to see the walk-through of the 3-D visualization graphic and only briefly looked at the print overview. Jokingly, her initial comment was “Looks like Y-12," a reference to Y-12 building complex, which is near the Oak Ridge National Lab.[13] (A second presentation was scheduled for later on that month.)

Another topic of discussion was that even though the graphic was an accurate representation of the HGMIS web site, the data density was so "thick" that analysis was limited reinforcing the idea that the graphic needed to be compressed. Furthermore, even though the first representation indicated the heavy linkage in the public resources group of pages it did not reveal specific navigational paths.

Compressing the Web Usage Visualization Graphic

In order to accurately compress the web usage visualization graphic, the information in the graphic had to combined into groups. Fortunately, the pages in the web site divide well by function as shown previously in Table 2. Pages from the same group were combined into one page. For example, hits from the five publications pages, (<.../primer/intro.html>, <.../publications.html>, <...pub/genechoice/index.html>, <.../tko/index.html>, and <.../97santafe.html>) were added together (2566 hits + 3044 hits + 1590 hits + 135 hits + 741 hits respectively) for a total 8076 hits. This was done for each of the seven groups and then each group page was placed into a data visualization graphic, using the same value as the previous graphic (1 cube = 100 hits). For consistency, the cube depth also remained the same between the first graphic and this revision.

---

[13]Normally such a comment might be easily passed over but there is significance in the group leader's joke, because it indicated that she quickly interpreted the graphic into something she could understand -- a map of a building complex. In *Information Anxiety*, Wurman (1989) talks about one of the keys to understanding any new material is relating "the information to ideas that you already understand" (p.53).

A new link matrix was also constructed --- again showing the links between "groups of pages" as opposed to just links between pages. There were two major differences in the construction of the graphic.

In the original graphic, the height of the walls ranged in value from 0.00 to 3.50. A single link had the value .25 which allowed for the possibility of 14 links between Page A and Page B (3.50/.25). In the compressed graphic the cube height was also set at 3.50 but the value .25 represented five links.[14] Also the number of links was rounded to nearest 5 or zero. There was an exception: If the number of links between groups was only one or two, it was rounded up to the five value. The rounding was implemented to show even minor linkage between groups, *if* any existed at all.

The second difference in the compressed graphic is in computing the links to oneself. Because the home page and the links page groups contained only one page in their group, their value was not computed. But in the other four groups there was a tendency for group pages to link heavily within themselves. Therefore, a "link back" was drawn in the graphic to show these important relationships.

Table 3 shows the entire link matrix for the group pages. (A print version of compressed graphic appears in Appendix F, Figure 1 with a corresponding design grid similar to Figure 1 located in Appendix F, Figure 2.)

This link matrix was then converted into a table that could be used to compute transition diagrams. These transition diagrams can be used to "determine the probability of next user action given previous action " (Penniman and Dominick, 1980, p.32). In the

[14]One drawback was that one wall (or link graphic) was limited to 70 links or (14 X 5 links). To overcome this, a second wall running parallel to first one was constructed.

Table 3. Number of Links Between Groups of Pages

Page Group	**Home**	**Res.**	**Publ. Res,**	**Publ.**	**Links**	**Utilities**	**Total**
Home 21,400 hits	--	7	20	7	2	5	41
Research 7,522	9	31	14	11	9	0	74
Public Res 20042	26	46	132	88	32	45	369
Public. 8076	5	12	20	4	3	7	51
Links 4401	2	9	12	15	--	5	43
Utilities 2154	4	12	33	12	4	3	68
Total Links	46	117	231	137	50	65	646

case of the HGMIS web site, transition diagrams were used to determine the such as most and least traveled navigational paths.

To do this, the link counts were converted to percentages. If home had 41 total links to other pages in the web site and seven of those were to the research pages. Then the value was changed to 17%. This was done throughout the table as shown in Table 4. Using the percentages, a set of transition graphs was built. The information from these graphs were used to compute the most probable navigational path (See Appendix G, Figures 1 & 2) and the least probable navigational path (See Appendix H, Figures 1 & 2).

The most probable navigational path showed that half of the visitors to the home page would first go from the home page to the one of public resources pages, from there approximately one third of those visitors would go to another public resources page. Later, they'd migrate to a publications page.[15]

The least traveled path from the home page would be to go the links page and then from the links page back to the home page.[16]

Second Presentation to HGMIS

During a second presentation on April 20, 1998, both the print version and 3-D screen version of the web visualization graphic and print versions of the two navigational paths were shown to HGMIS staffers. This informal demonstration was attended by the

[15]There is also the possibility that the user would stay in the public resources pages and not leave.
[16]Obviously this is a scenario that isn't typical for users nor is it one that should be. The significance of identifying the least traveled path is to insure that the site isn't set up in a way that a navigational path you are encouraging users to follow, *isn't*

Table 4. Number of Links Between Groups of Pages as Percentages

Page Group	**Home**	**Resear.**	**Public Res.**	**Publ.**	**Links**	**Utilities**	**Total**
Home	--	17%	49%	17%	5%	13%	100%[1]
Resear.	12%	42%	19%	15%	12%	0	100%
Public Res.	7%	12%	36%	24%	9%	12%	100%
Public.	10%	23%	39%	8%	6%	14%	100%
Links	5%	21%	28%	35%	--	12%	100%
Utilities	6%	18%	49%	18%	6%	4%	100%

1 Because of rounding, percentages do not necessarily add up to 100 %.

site's experts including the group leader, webmaster, a web production assistant, ascience writer and an interested web and publications manager from another group. In contrast to the first presentation, the print version of the link matrix was not shown, because of its technical nature. Because of the larger number of attendees, the second presentation was more focused on the overview of the graphic and the site rather than the details of how the graphic was built.

Just as in the first presentation, the print version was introduced first to show the general overview of the graphic/site. Next, while using the mouse and the walk through capabilities of the software, the attendees were shown the significance of the graphic's color scheme, the box and link sizes, and the positioning of the boxes and links. In the second half of the presentation, the group saw the compressed web visualization graphic, and the print versions of the most and least traveled navigational paths. The group's commentary about the presentation can be categorized as: General reaction, specifics and other uses.

General reaction. In general, the reaction was favorable to the graphic and the HGMIS staffers seemed to have little trouble understanding it after introductory remarks. However, the group leader, although not negative in her remarks, was more lukewarm in her reaction. According to her, she is a "text person" preferring to have a table of numbers.[17]

final graphic. (She had seen the original link matrix, which was wall-sized and hung in one of the HGMIS cubicles throughout the months of March and April).

Others found the graphic appealing and functional. For instance, after the presentation the webmaster spent several minutes playing with graphic, zooming around and even distorting it.[18]

Specific comments. Some of the minor problems on the site,[19] which the graphic revealed poignantly , were discussed. The biggest problem was considered to be the lack of links from the links page and the publication family of pages to the rest of the site. This problem had already been corrected (more links have been built in the HGMIS publication pages) from the time it was presented to the webmaster on April 7th. Also there was some surprise at the navigation page's weak hit totals (for such a centrally located page -- this page was not even listed in the tabular web usage report). There was also concern about the lack of hits to the electronic version of the *To Know Ourselves* publication.

Other uses. The graphic also reinforced one of basic tenets of HGMIS' mission that it does have two distinct clienteles -- researchers and the general public. Also it is important for HGMIS to continue work on communicating the breadth of the site because web visitors may want to cross over from the research portion of the site to the public resources pages or vice versa.

---

[18] This latter activity should serve as a reminder that there is a caveat with this 3-D graphic, because depending on where the VHSB viewer "eye" is placed, pages and links alike can look much larger than they actually are.

[19] Personally, I feel this is a fairly well organized site, especially considering the number of files and wide range of topics that must be organized. In my opinion, the problems at the HGMIS web site are relatively minor. In a way, the graphic helps one appreciate the abilities of the developers and maintainers of the web site.

This led to discussion about whether the graphic could be adapted as a navigational tool (especially since the current navigational page is heavily linked, but rarely used). As a navigational tool, a visitor could use the graphic as the "Big Picture" and then zoom into specific pages by clicking on the research or publication family of pages. The user then could click again to go into more depth, but always have the option to zoom out quickly to maintain a sense of orientation.[20]

[20] Remember earlier in the Botafogo et. al. link matrix article, the authors made the comment that "software tools that would help hypertext authors see global and local properties would be beneficial."

5. Results

As Jacques Bertin wrote in the opening pages of *Graphics and Graphic Information Processing*, "A graphic is not only a drawing; it is a responsibility, sometimes a weighty one, in decision-making. A graphic is not 'drawn' once and for all: it is 'constructed' and reconstructed until it reveals all the relationships constituted by the interplay of the data." p.16)

With Bertin's comment in mind, one is compelled to keep the book open on the use of the HGMIS web usage visualization graphic a little longer. This project must be evaluated from several perspectives -- not only from this specific Web usage graphic, but the idea of developing usage graphics in general.

Evaluating the Web Visualization Graphic

In order to completely understand web usage visualization graphics, they should be evaluated at several levels:

A representation level. Referring back to the need of web industry's need for better measuring technologies, the evaluator must ask, "Did the web visualization graphic accurately summarize the mass of information into a form that an analyst could use?" Furthermore with respect to Shannon and Weaver, were the symbols (walls for links, cubes for pages) accurate and did they precisely convey the meaning of the site? Also, those who have studied and written about data visualization believe the correct proportionality, link relationships, and color are critical to any data representation.

A communication level. With respect to Shannon and Weaver again, the effect of the web visualization graphic as a communication tool should be judged as to whether its creation changed behavior. In other words, did the decision makers make changes in their web site based on what they saw in the graphic?

An evaluation level. The final question concerning the web visualization graphic is whether it was an effective tool and whether it has potential to be more widely used by web analysts in general.

Web Usage Visualization Graphic: Representation Level -- Present and Future

Even by using an off the shelf software package with a limited product features, the HGMIS web usage visualization graphic fulfilled its obligation to accurately represent the HGMIS web site. The cube size did accurately indicate the number of hits on each URL and with the help of the link matrix, the graphic was able to display the relationships between the pages. Moreover, the web usage visualization graphic compressed the pages upon pages of data dump that often comes with web site log analyzer statistics into a series of brief, more analytical reports (the graphic, the compressed graphic, the link matrix and the transition stage diagrams).

In the future, a web graphic usage designer may want to look for other ways (cones, spheres) to represent pages and nodes and evaluate other "less-congested" looking ways to represent links. A more upscale version of VHSB has such features. Also other VRML browsers should be evaluated because viewing the 3-D graphic was somewhat limited and slow. Having the graphic easier to navigate in VRML (while maintaining the graphic's integrity) would allow people from other sites on the Internet to view and evaluate the web usage graphic.

Web Usage Visualization Graphic: Communication Level -- Present and Future

The HGMIS Web Usage Visualization Graphic did immediately impact HGMIS' web site, thus fulfilling Shannon and Weaver's criteria for successful communication. In the two weeks between the first presentation and the second presentation, the webmaster added links from the publication pages to other parts of the site. The graphic clearly showed that these types of links were missing in the web site.

More graphic visualization work could be done on the HGMIS web site, in regards to the newsletter and abstract URLs. As mentioned earlier, (see footnote 9) almost 1800 files on the HGMIS web site are either articles from the HGMIS newsletter and abstracts from contractor-grantee meetings. They don't have as many hits per page, but collectively they could be a factor in the site and should be investigated.

Web Usage Visualization Graphic: Evaluation Level -- Present and Future

For HGMIS, in terms of being an effective evaluation tool, the web usage visualization graphic had only a slight short term impact (adding links to publication pages). It's difficult to say what the long term impact might be. At this time, there were no specific plans to add the graphic as a navigational tool, add more links or re-allocate web sources based on the two presentations. Also there are no plans to redo or update the graphic using statistics from other time periods.

The Future of Web Usage Visualization Graphics

Although the impact at the HGMIS site was minimal, web usage visualization graphics merit further use and consideration. However, before final judgment on whether web usage visualization graphics have a role in web measurement, more work needs to be done in certain areas:

- Build additional web usage graphics.This should be done for different types of web sites and the findings should be compared.
- Automate the process. Considering that the data are generated from the Unix based HTTPd log analysis software, and that the links are built in (or programmed in HTML language), this usage data and links data could be fed into a template and a representation could be automatically built and updated on a regular basis. Doing computations manually is time consuming and not cost effective.
- Gather better data. "Hits" have the reputation as a weak statistic and "click through" data are not readily available. Improved data integrity in these areas would improve the web usage graphic's accuracy.
- Better evaluative techniques. The presentations and evaluation of the HGMIS web usage graphic were informal and only included a small group of experts. A larger or different group may respond differently. A more methodical quantitative survey-type evaluation should be developed.

6. Final Discussion

Even if improvements are made in the design and construction of web usage visualization graphics, there will always be a degree of uncertainty in their effectiveness as an exclusive web measurement instrument. The uncertainty lies in the fact that each person views a graphic differently. For example, some people referred to the HGMIS graphic as "a city," "a building complex," "a mall," or as a "a printed circuit board." Such comments indicate, as Wurman says, that the user is attempting to understand new material by relating it to material they already understand.

On the other hand, others who learn differently (and its no fault of their own) are unable "to see" the web site via the usage graphic. They find the information more meaningful in a tabular or text form (such as the link matrix).

With these types of limitations in mind, the best strategy may be to develop several presentations and allow the user to choose the representation that promotes better understanding. This thesis has explored one such presentation, the web usage visualization graphic in its current early evolutionary stages. With its ability to compress usage data accurately and reveal semantic relationships, the web visualization usage graphic deserves to be one of the tools that can help clear up some of the current ambiguities surrounding web usage.

References

Allen, C. and D. Kania. 1997. Web log analysis tools. Internet: http://www.allen.com/amg /a-netrev2.html. Accessed: July 8, 1997.

Botafogo, R.A., E. Rivlin, and B. Shneiderman. 1992. “Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics.” *ACM Transactions and Information Systems,* Vol 10, No. 2, (April) 142-180.

Bretin, J. 1981 *Graphics and Graphics Processing*. Translated by William J. Berg and Paul Scott. Walter de Gruyter & Co. Berlin.

Bretin, J. 1983. *Semiology of Graphics: Diagrams, Networks and Maps.* Translated by William J. Berg). University of Wisconsin Press: Madison.

Card, S.K. 1996. "Visualizing Retrieved Information: A Survey," *Computer Graphics and Applications.* Vol. 16, No. 2, (March) 61-63.

Catarei. T. 1996. "Interaction with Databases." *Computer Graphics and Applications.* Vol. 16, No. 2, (March) 67-69.

Cleveland, W.S. 1994. *The Elements of Graphing Data.* Murray Hill, New Jersey, Bell Laboratories.

Comstock, G. 1989. *The Evolution of American Television*. Newbury Park, California: Sage Publications.

CyberAtlas. 1997. Market Size. Internet: http://www.cyberatlas.com/market.html. Accessed: March 27, 1997.

Eick, S.1996. "Aspects of Network Visualization," *Computer Graphics and Applications.* Vol. 16, No. 2, (March) 61-63.

Gershon N. 1996. "Moving Happily through the World Wide Web," *Computer Graphics and Applications.* Vol. 16, No. 2, (March) 72-74.

Gershon N. and J.R. Brown 1996. "Computer Graphics and Visualization in the Global Information Structure," *Computer Graphics and Applications.* Vol. 16, No. 2, (March) 72-75.

Gershon N. and J.R. Brown 1996. "The Role of Computer Graphics and Visualization in the GII," *Computer Graphics and Applications.* Vol. 16, No. 2, (March) 61-63.

Hall, S. S. 1992. *Mapping the Next Millennium.* New York: Vintage Books.

Hyams, J. 1997. "Winning may mean big loss for Saints," *The Knoxville News Sentinel*. 7 October. Sec. D, p 4.

Katz, J. 1997 “Old Media, New Media and a Middle Way,” *The New York Times*. 19 January, Sec. 2 p.1+

Kowalski, G. 1997. "Information Visualization" (Chapter 8) in *Information Retrieval Systems: Theory and Implementation*. Kluwer Academic Publishers: Boston. 180-201.

Lin, X. 1997. “Map Displays for Information Retrieval,” *Journal of the American Society for Information Science.* 48: 40-54.

Lin, X., G. Marchionini and D. Soergel. 1995. “Category-Based and Association-Based Map Displays by Human Subjects,” Advances in Classification Research, Vol IV.

Mansfield, B. 1998. Written comments to author about thesis draft. 12 February.

Novak T.P. & Hoffman D.L. 1996. New metrics of new media: Toward the development of web measurement standards. Internet .http://www2000.ogsm.vanderbilt.edu /novak/web.standards/webstand.html. Accessed: July 4, 1997.

Penniman W.D. and W.D.Dominick 1980 “Monitoring and Evaluation of On-Line Information System Usage,” *Information and Processing Management*. 36: 17-35.

Poler, A. 1997. Internet measurement: past & future (Part I). Internet: http://www.cyberatlas.com/column1.html. Accessed: November 13, 1997.

Senechal A. 1997. It's all in the process. *Adobe*. Vol. 8, No. 4 (Spring, 1997--North American Edition) 34-40.

Tierney, J. 1997. "Out Oldest Computer Updated," *The New York Times Magazine*. 28 September. p. 46+.

Tufte, E. R. 1983. *The Visual Display of Quantitative Information*. Cheshire CT: Graphics Press.

———. 1990. *Envisioning Information.* Cheshire CT: Graphics Press.

———. 1997. *Visual Explanations.* Cheshire CT: Graphics Press

Weaver, Warren. 1949. “Recent Contributions to the Mathematical Theory of Communication.” In *The Mathematical Theory of Communication*, Claude Shannon and Warren Weaver, ed. Champaign, IL: University of Illinois Press. 93-117.

Wurman, R.S. 1989. *Information Anxiety*. New York: Doubleday.

Wurman, R.S. 1996. *Information Architects*. Graphis Press Corp. Zurich.

Appendices

Appendix A: Web site examples

The following are two examples of how a site handles its own web statistics:

The first example (Figure 1) is from the University of Tennessee home page at <http://www.utk.edu/usage> as of May 18, 1998. The second, (Figure 2, next page) also dated May, 18, 1998, is from AccessWatch <http://www.accesswatch.com> a company specializing in analyzing WWW servers.

Figure 1

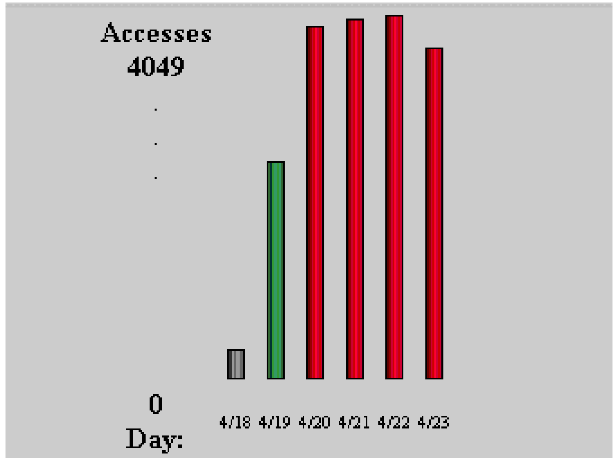


Figure 2

Appendix B: Industry players in visitor and measurement analysis (Part 1)

Company & Web Address	**Name of Web Stat Product & relation. to co. product mix**	**Type of Usage Statistics**	**Type of Information Graphics Used**
Allen Interactive,Inc., Raliegh, NC www.allen.com	GuestTrack Product mix that includes interactive marketing, & website development	Assigns a unique ID to each registered user, builds a profile, enters ID into access log, but requires visitor register Personalizes web visit "on the fly"	Information inaccessible
Andromedia San Francisco www.andromedia.com	Aria 2.0 Recorder Reporter Main product of company	Captures data off the server in realtime and encrypts it for protection Claims to be better than conventional log file reading or packet sniffing	Tables and bar graphs of navigation, visitor, content, and traffic reports
Bien Logic LaJolla, CA www.bienlogic.com	SurfReport 2.1 Similar prod. called Proxy Report. Another product Clubhouse builds "on-line" communities"	Surf Report, does web traffic analysis, lists pages, hits,	Straightforward tables of information, but instead of listing URLs it will identify what the page is about.
Clickshare Williamstown, MA www.clickshare.com	Clickshare Service "transaction processing and audience measurement service"	More focused on billing for site use than audience measurement.Collects info. from NSCA Httpd common log format then does value added with respect to billing	Unavailable. Limited information about Clickshare
Accesswatch www.accesswatch.com	Used to measure Project 2000 website, but not on Novak & Hoffman list.	Standard fare of daily access, accesses by domain, accesses by host	Standard information display with bar chart and line chart graphics.
Media Metrix, Inc. New York, NY www.npd.com	Offers a syndicated WWW Audience Measurement Service	Information inaccessible	Information inaccessible.
Logical Design Solutions www.lds.com	LDS WebTrac	Information inaccessible	Information inaccessible
Software, Inc. www.webtrends.com	Web Trends	Standard fare of Usage Statistics	3-D bar graphs, can be viewed on browser

Appendix B: Industry players in visitor and measurement analysis (Part 2)

Company & Web Address	**Name of Web Stat Product w/ relation. to co. product mix**	**Type of Usage Statistics**	**Type of Information Graphics Used**
Everyware Mississauga, Ontario www.everyware.com	Bolero Used with a Everyware's program/product called Tango	Complete standard fare of usage by hour day, month, top client domains. Also has top referrers, a listing of URLs that refers to a site.	Some bar graphs, rest is tabular material Dumps log info into a database like SQL or Oracle for manipulation
Webthreads Vienna, VA www.webthreads.com	Threadtrack Fits into their major line of interactive direct marketing products	Standard fare of usage statistics and some "clickstream" data. "Clickstream" reporting (tracks how users navigate through a site)	No graphics. All information is displayed in tabular form.
Microsoft Redmond, WA www.backoffice.microsoft.com	Usage Analyst (formerly known as a product from Interse) Part of Site Server Enterprise Edition	Standard usage fare. Also analyzes referring URLs, tracks length of visit, compares user behavior on various levels of the site. Handles data from various servers, user can customize database. Highly detailed reports	No graphic examples. Bar graphs and tabular material *Note:May determine link usage through "inference algorithms" instead of clickstream*
I/PRO Redwood City, CA www.ipro.com (Analyzes big sites like USA Today, NBC.)	Recently purchased another web analysis company called Net Count. Host of products for traffic verification, analysis and research	All kinds of variations on standard usage statistics such as Weekly Traffic Summary, Weekly Traffic Trends, Top 50 pages, Top 100 files, pages accessed per visit	Bar graphs and tables
W3.Com http://www.w3.com Mountain View, CA	Personal Web Site System Part of a product package interested in personalized web sites	Only tracks enough for commercial confirmation on how web site is used	Standard fare of charts, graphs and tables

Companies Discontinued since the Novak & Hoffman report in 1996: Net Genesis, WebWatch, Open Market, Group Cortex Inc.

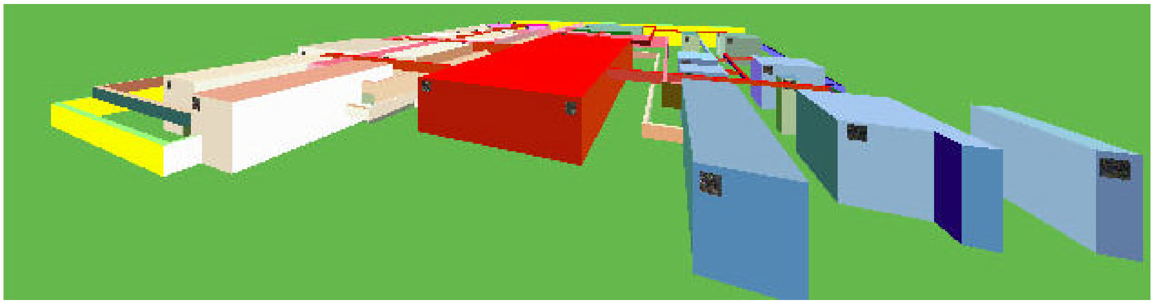
Appendix C: HGMIS report on web activity -- January 1998

1998 HGMIS Web Statistics (by URL)					
URL					
	Jan	Feb	March	April	May
home	21,431				
genetics	7,003				
CAL TRCA	2,021				
centers	421				
research	2,655				
links (tot.)	4,401				
mission	157				
arch/news	2,399				
fund'g.fund	326				
proj/hgp	3,189				
HGP faq/faq	5,707				
proj/5yr/5yr	671				
proj/proj	533				
publ/publ	2,566				
res/informa	2,775				
resour/ELSI	3,259				
res./educ	1,431				
res./map	1,166				
res./seq.	917				
news/toc	436				
Subtotal	63,464				
% Rep.	35%				
Tot. files	**183,867**				
Host mach.	15,054				

Appendix D: HGMIS Link Matrix

	H	R	T	A	C	P	F	H	I	E	P	T	G	A	G	L	T	F	E	C	S	A	S	M	I	S
Home		3	1	2	1	3	4	3	3	4	1	1	1	2	2	2	0	0	2	1	1	2	1	1	1	1
Res	2		0	1	1	1	0	0	1	1	0	0	0	0	0	1	1	1	0	0	0	0	1	1	1	1
Toc	2	6		3	2	4	5	4	3	5	2	2	2	1	2	4	1	1	3	2	2	3	2	2	2	2
Arc	2	2	1		1	3	3	3	3	3	1	1	1	1	2	3	0	0	2	1	1	2	1	1	1	1
Cal	1	1	0	0		0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Pub	2	2	1	1	2		3	3	2	4	1	2	1	4	2	3	1	1	3	1	1	2	2	2	2	1
FAQ	2	2	0	2	1	3		4	2	4	0	2	0	1	1	3	0	0	2	1	1	2	1	1	1	0
HGP	2	2	1	2	1	6	3		2	3	2	2	1	2	2	3	0	0	2	1	1	2	1	1	1	1
Info	2	2	1	2	0	5	3	3		3	2	1	2	2	2	3	0	0	2	1	1	2	1	1	1	1
Elsi	1	2	1	2	1	4	4	3	0		2	2	2	2	2	5	0	0	0	0	0	0	0	0	0	1
Pri	1	1	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tko	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gene	1	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0
Acro	1	1	0	0	0	1	0	0	0	0	0	0	0		0	0	1	0	0	0	0	0	0	0	0	0
Glos	0	0	0	0	0	0	0	0	0	0	1	0	0	0		0	1	0	0	0	0	0	0	0	0	0
Link	2	2	1	2	1	3	3	3	2	2	1	1	1	2	2		0	0	2	1	1	2	2	2	2	1
Tcal	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0
Fund	1	1	0	1	1	1	0	0	0	1	0	0	0	0	0	1	0		0	0	0	0	1	1	1	0
Educ	2	2	1	2	1	4	4	3	2	3	2	2	2	3	3	4	0	0		0	1	1	2	1	1	1
Car	2	2	1	1	1	4	4	4	2	3	1	1	1	0	2	3	0	0	2		1	2	1	1	1	1
Stud	2	3	1	2	1	5	4	3	2	3	2	2	2	2	2	3	0	0	2	1		2	1	1	1	1
Asst	2	1	1	2	1	4	4	3	2	3	2	1	2	2	1	5	0	0	2	1	1		1	1	1	1
Seq	1	1	0	1	1	1	0	1	0	1	1	0	0	0	0	3	0	1	0	0	0	0		1	1	0
Map	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	3	0	0	0	0	0	0	1		1	0
Info	1	1	0	0	1	1	0	1	0	1	1	0	0	0	0	1	0	1	0	0	0	0	1	1		1
SF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix E: Full graphic of HGMIS web site



Appendix F: Compressed graphics of HGMIS web site

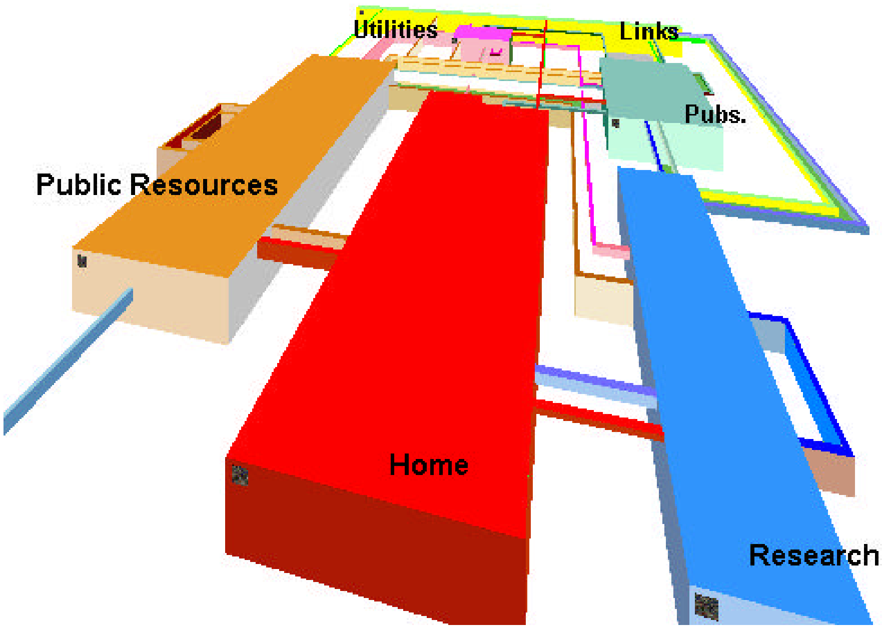


Figure 1: Three-dimensional version of compressed web graphic.

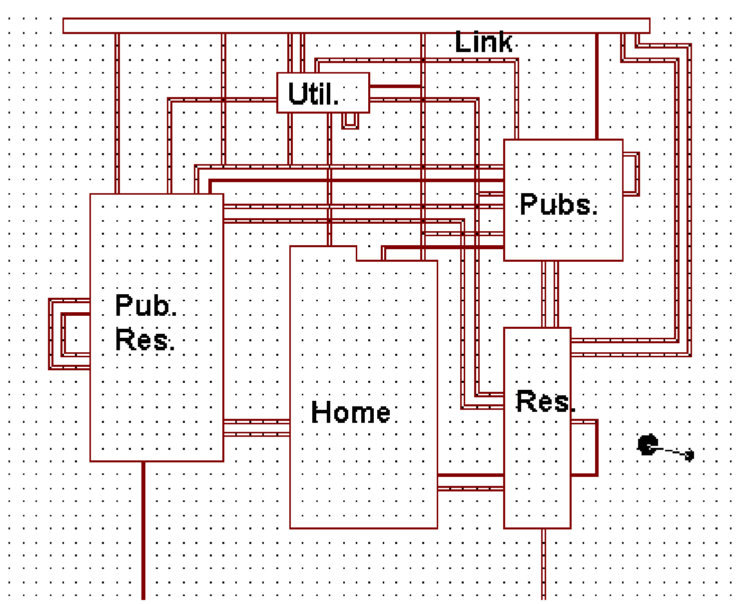


Figure 2: Grid version of HGMIS Compressed Graphic

Appendix G: Transition stages -- most traveled navigational path.

Figure 1 shows the most traveled navigational path on the two dimensional grid. Users arrive first at the home page (1) , travel to a public resources page (2), stay in the public resources pages (3) and finally migrate to a publications page (4). Figure 2 (next page) shows the same information in stages using the data from Table 4.

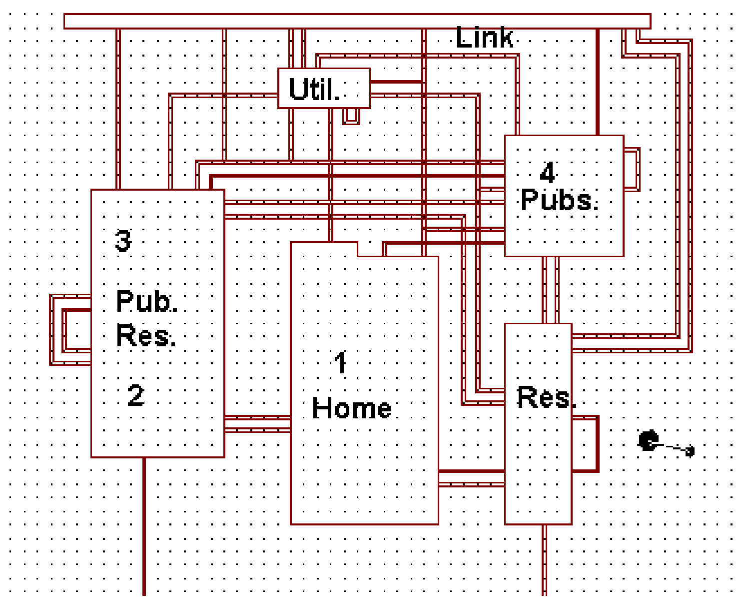
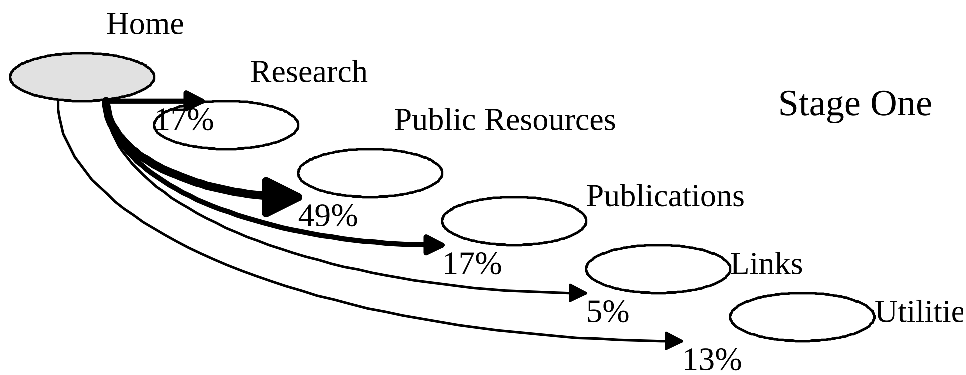
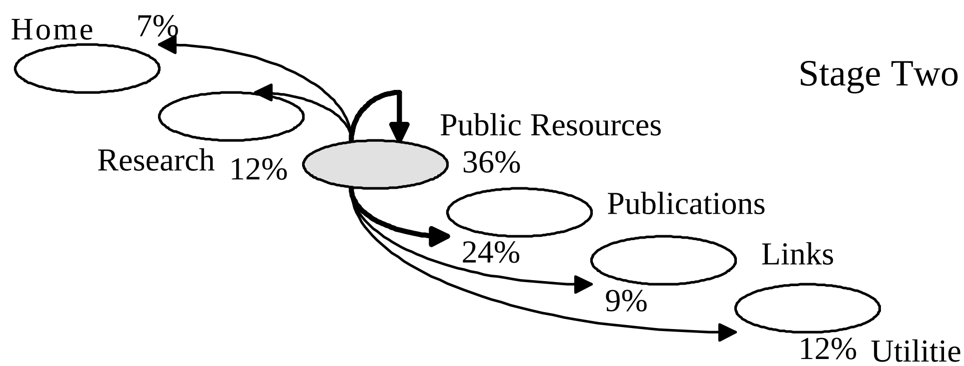
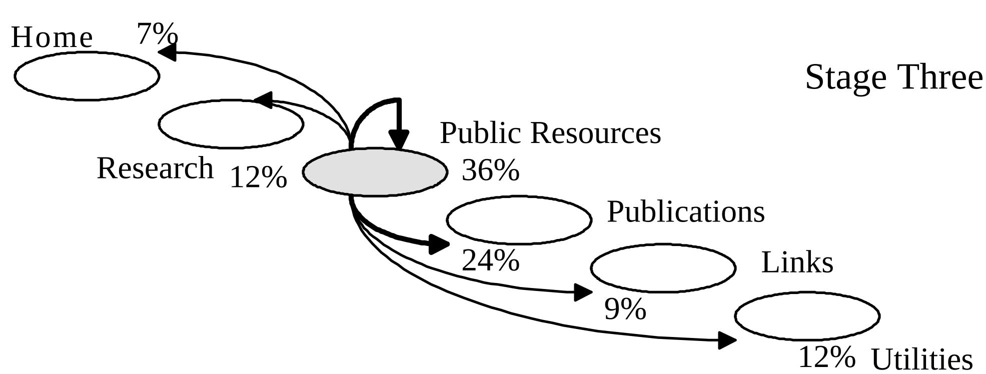


Figure 1


Home
Research
17%
Public Resources
Stage One
49%
Publications
17%
Links
5%
Utilitie
13%


Home
7%
Stage Two
Public Resources
Research
12%
36%
Publications
24%
Links
9%
12%
Utilitie


Home
7%
Stage Three
Public Resources
Research
12%
36%
Publications
24%
Links
9%
12%
Utilities

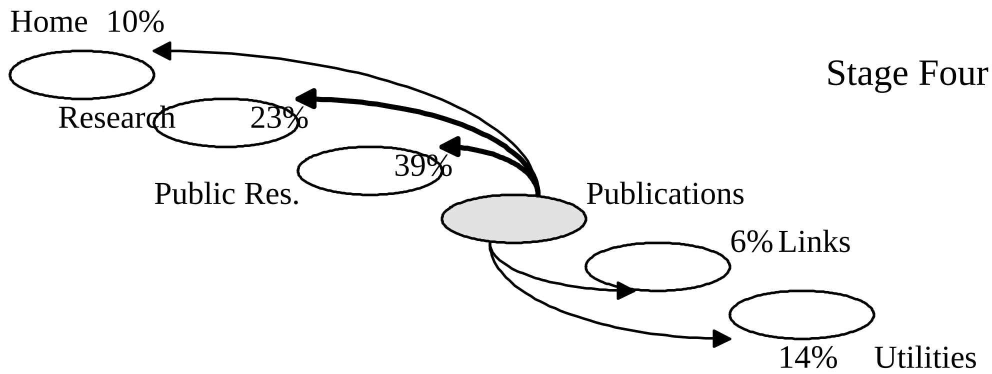

Home 10%
Stage Four
Research
23%
39%
Public Res.
Publications
6%
Links
14%
Utilities

Figure 2

Appendix H: Transition stages -- least traveled navigational path.

Figure 1 shows the least traveled navigational path on the two dimensional grid. Users arrive first at the home page (1) , travel to the links page (2), and back to the home page (3). Figure 2 (next page) shows the same information in stages using the data from Table 4.

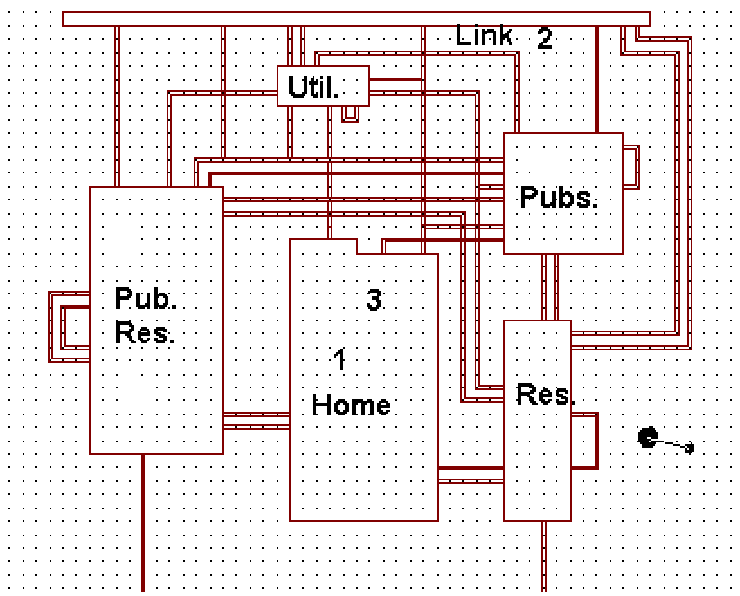


Figure 1

Vita

Murray Browne is a Master of Science degree candidate in the University of Tennessee's School of Information Sciences. He graduated with a Bachelor of Arts in English and Radio/Television from Indiana University in 1976 and a Master of Arts in Gerontology from Wichita State University in 1985.

While attending the University of Tennessee for the past three years, Browne worked at the Human Genome Management Information System at the Oak Ridge National Laboratory in Oak Ridge Tennessee, specializing in studying web usage activity, scientific communication and assisting in the production of print and electronic publications. Currently he is working on a book with University of Tennessee Computer Science professor, Dr. Michael Berry on how to build better search engines.

As a writer, Browne has also published numerous essays, book reviews and feature stories. He is married with two daughters.