

Introduction to World Wide Web Servers

A presentation by Eric Lease Morgan to MUGLNC, March 31, 1995
<http://www.lib.ncsu.edu/staff/morgan/eric-talks-to-muglnc.html>

It is a pleasure and an honor to be able speak with you this morning. I hope to provide you with a presentation that is informative, inspirational, intellectually stimulating, and slightly "intertaining". I also hope you go away from this presentation learning three things:

- Some background about the World Wide Web (WWW)
- 3 Qualities of well constructed information systems
- Possible uses of WWW servers for libraries

WWW is synonymous with HTTP

The World Wide Web (WWW) is a popular term used to denote the hypertext transfer protocol (HTTP). Based on the concept of "hypertext" put forth by Vannevar Bush in 1945 and coined by Ted Nelson, the purpose of HTTP is to share and distribute information. Other applications have tried to implement the concept of hypertext (like HyperCard of Apple Computer), but HTTP seems to be the best implementation to date.

HTTP is older than gopher

While the popularity of HTTP servers is relatively recent, it is interesting to note that it is older than the gopher protocol. It has its beginnings in 1990 with Tim Berners-Lee, then of the CERN particle physics laboratory in Switzerland. Gopher was developed at the University of Minnesota in 1991.

The first HTTP server application was developed on a NeXT computer. While more than functional, these computers were not very popular and consequently, not very many people owned the necessary hardware to make HTTP services available.

Similarly, the client applications ran on NeXT computers or was implemented using vt100 emulation. Again, few people owned the necessary hardware to use the client applications and/or the vt100-based

client application was so ugly that no one wanted to use it.

On the other hand, the gopher protocol had been adapted to a "host" of hardware platforms. Additionally, graphical user-interface gopher client applications were developed for a wide range of computers. Consequently, the gopher protocol, at least initially, was more popular than HTTP.

It wasn't until 1994 when Marc Andreessen and Rob McCool then of the National Center for Supercomputing Applications (NCSA) developed HTTP client and server applications (Mosaic and httpd, respectively) for more general "flavors" of Unix that HTTP really became popular.

HTTP is more efficient than gopher

Besides the ability to distribute formatted text in the form of hypertext markup language (HTML) documents, gopher servers and HTTP server are very similar. They are both based on the client/server model of computing. But in terms of computing resources, HTTP is much more efficient than gopher. The main reason behind this is because HTTP distributes much of the computing load to the client applications, whereas gopher server applications do all of the work for gopher clients.

For example, both gopher and HTTP servers allow end-users to retrieve files from file transfer protocol (FTP) servers. Using the gopher protocol, client applications make requests for files. The server then FTPs the files and delivers them to the client. Doing searches of Wide Area Information Servers (WAIS) work in the same manner. The client makes the request and the server does the work.

By contrast, using a HTTP client like Mosaic, Netscape, or Lynx, when an end-user requests a file from an FTP server, the end-user's HTTP client application retrieves the file directly without going through the

server. Since HTTP servers are only handling HTTP requests, the computer hosting an HTTP server can handle many more requests more efficiently.

3 Qualities of information systems

At its very heart, HTTP servers are information systems. Information systems, especially in today's world, abound. For our purposes, an information system is broadly defined as a collection of information. Consequently, a book can be described as an information system, as well as a library, an advertisement, or even the dashboard of your car. Similarly, things like gopher, FTP, and HTTP server too can be defined as information systems.

With everybody creating information systems these days, there is a need for some guidelines describing qualities of effective information systems. In my opinion, these guidelines can be distilled into three qualities:

1. Readability
2. Browsability
3. Searchability

All of these qualities (readability, browsability, and searchability) do not have to be equally represented in every information system. As a collection of information increases, different aspects of these qualities take on greater significance. Thus, the amount of readability, browsability, and searchability an information system exhibits depends on the type and quality of the collected data, as well as the information needs of the clientele.

Readability means good visual design

All information systems, no matter how small must incorporate principles of good graphic design. You and your information system are competing with a myriad of other information systems. If your data is not presented in a visually appealing, easy-

to-read manner, then your chances of retaining the attention of your intended audience are significantly reduced. Try to follow these guidelines:

- Use a consistent layout
- White space is good
- Visually organize the page; employ horizontal rules
- Keep pages short
- Include elements of contrast
- Use all stylistic elements in moderation

Browsability connotes logical organization

As the size of an information system grows, so does the need to logically organize its data. This implies grouping conceptual sets of data with similar conceptual sets of data. Browsability becomes apparent when it is coupled with hypertext and logical groupings of information.

Despite the dynamic nature of logical groupings of information, the organization of information and knowledge seem to be a necessary part of human existence. Since the primary purpose of information servers is to disseminate knowledge, facts, and ideas, it then follows the information they disseminate must be organized in some reasonable fashion. Thus:

- Know your audience
- Provide "about" texts
- Use the vocabulary of your intended audience
- Create a hierarchal system of ideas
- Create a system that is both flexible and exhaustive
- Classify by format last

Searchability addresses specific needs

The largest of information systems must include search features. These features help overcome the disadvantages of the purely browsable system.

First, your conception of the information universe is not necessarily the same as your reader's. While you try to group things in the most logical manner, your

reader's "logic" will be different than yours. Searchability can help overcome this discrepancy by allowing the reader to create their own set of logically similar items. Searchability readily lends itself to locating known items rather than making the reader browser down a number of menus to get what they want. Searchability works independently of your collection's size.

On the other hand, in order to effectively search an information system, the reader must know the query language of the search engine. The ability to search an information system assumes your readership has a preconceived idea describing what they need. Totally searchable systems require the searcher to know the data structure of the indexed collection.

All hope is not lost. Try to follow these guidelines:

- Include help texts
- Map located items to similar items
- Provide simple as well as "power user" search mechanisms

HTTP and libraries

Given the existence of these new technologies, how can they be put to use in libraries? There are a number of ways including but not limited to:

- "about" texts describing services and staff
- collections of Internet resources
- interfaces to databases
- electronic "librarians"

First of all, libraries can use HTTP server to supplement much of their print materials describing services and staff. This sort of information can include things like resumes, presentation, essays, or other activities demonstrating the qualifications of the staff to the library's clientele. It can include any publications a library creates. It can include instructions on how to use various library tools as well as library guides.

Using a database program, collections of Internet resources can be collected. Once in a database program, reports can be generated (HTML files) which in turn can be used to make up the body of an HTTP server. This is how the "Study Carrels" of

the NCSU Libraries are maintained.

Coupled with the use of special HTML constructs called FORMS and the use of common gateway interface (CGI) scripts, HTTP servers are extensible. CGI scripts allow programmers to create applications using HTTP servers as front-ends. These applications take input from the FORM, process it, and return results back to the end-user. Many times these CGI scripts are used to search databases. This is how the Alcuin database is being utilized. Sometimes they are used to generate specialized email.

There is no reason this same sort of programming can't be applied to create an electronic "librarian." Such a program could ask the end-user questions. Based on the answers, it could ask other questions. At the end of the question/answer process, the program could generate a list of search statements (in the form of URLs). These URLs would then form an outline of how the end-user could find the sort of information they seek in the electronic library.

Conclusion

In conclusion, HTTP (or more commonly known as World Wide Web) servers are Internet-based applications for disseminating information. As information systems, they can embody many of the qualities of libraries and need to be readable, browsable, and searchable in order to be most effective. HTTP servers represent the fundamental change society is looking upon information. If we as librarians can learn to incorporate these new technologies with our traditional ideals, then we can look forward to enjoying providing needed information services for a long time to come.