OncoLink: A Multimedia Oncology Information Resource on the Internet

E. Loren Buhle, Jr. Ph.D., Joel W. Goldwein, M.D., Ivor Benjamin, M.D.
University of Pennsylvania School of Medicine, Philadelphia, PA

ABSTRACT

This paper describes OncoLink, the first multimedia World-Wide-Web (WWW) and gopher server focusing on cancer information for both the health care professional and the patient. OncoLink provides an internetworked hypertext and multimedia resource linking people, computers and information together in an easy to use fashion. Our objective in developing OncoLink is to provide comprehensive and timely information about many aspects of oncology for both patients and healthcare providers.

Specifically, OncoLink's purposes are: (1) the rapid dissemination of information relevant to treatment of cancer and concomitant problems; (2) education of health care personnel (at all levels) in the field; (3) education of patients and families of patients who have cancer; (4) posting of clinical trials and eligibility criteria; (5) the rapid collection and dissemination of quality, peer-reviewed information pertinent to oncology in general and specific subspecialities; (6) provide a well-organized, frequently updated hypertext system to access other quality cancer information resources on the Internet. OncoLink attempts to provide one-stop shopping for the patient, healthcare provider, researcher or Internet browser searching for cancer-related information.

Since its inception on March 7, 1994, OncoLink has averaged more than 36,000 accesses per month from around the world. While also accessible by text-based gopher servers, preliminary observations infer increased use of multimedia and hypertext documents over traditional text-only resources. From the large following of users, it is clear that electronic dissemination of high quality, peer-reviewed cancer information is very popular.

We conclude OncoLink is both useful and has wide interest in the international community. We believe, in the future, such systems will become common media for the international dissemination of cancer and other medical information.

INTRODUCTION

The Internet is a global computer network that facilitates educational, corporate, government and private communications throughout the world. Currently, there are an estimated 20 million users, with many more with potential access. One of the components on the Internet is World-Wide-Web
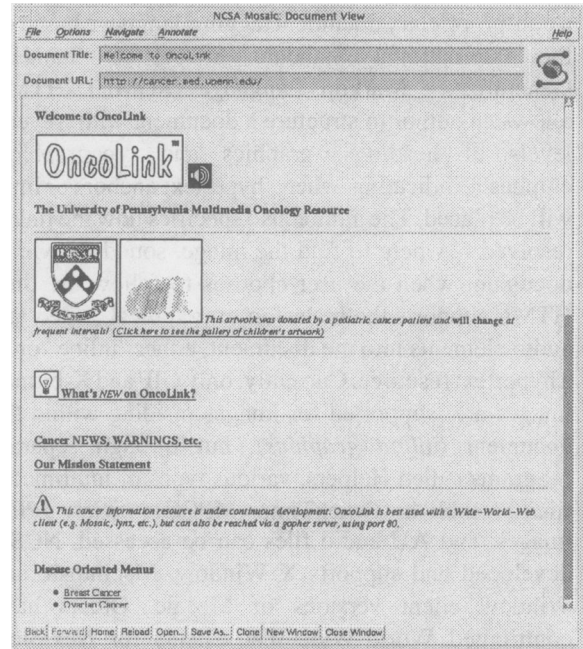

Figure 1 - Top of OncoLink's home page.

(WWW) servers[1,6], providing a mechanism for the distribution and navigation through multimedia information.

World-Wide-Web clients, such as the public domain browser Mosaic, allow users to easily navigate information available on the Internet. Mosaic is an ongoing software project of the National Center for Supercomputing Applications (NCSA) at the University of Illinois. One of their goals is to implement the emerging standards that unify the various Internet data browsing and retrieval services (Wide Area Information Server (WAIS), gopher, Archie, WWW, etc.) as they are developed by the Internet Engineering Task Force (IETF). The NCSA calls Mosaic a "networked information system for wide area distributed asynchronous collaboration and hypermedia-based information discovery and retrieval." After installing the WWW client, such as Mosaic, it is fairly trivial to browse audio, graphic images, and video, as well as make use of hypertext links between information resources. The hypertext elements are activated by a simple point-and-click on underlined "key" portions of the displayed text or images, causing Mosaic to retrieve the related document and display it.

WWW software is based on three primary

components: Hypertext Markup Language (HTML), Hypertext Transport Protocol (HTTP) and Uniform Resource Locator (URL). Both HTML and HTTP are drafts, not yet officially published as part of the Internet Request for Comments (RFC) process.

The Hypertext Markup Language (HTML) is a non-proprietary document tagging language based on the International Standard 8879:1986 Standard Generalized Markup Language (SGML). HTML allows an author to structure a document with several levels of headings, graphics and typographical emphasis, indicating where hypertext anchors or links will be placed. The author also specifies how the link is resolved -- where to find the image, sound, movie or document when the user chooses to follow the link. HTML supports the inclusion of graphical, video and audio elements into the document, either "inline" or as a hypertext resource. Currently, only GIF and X-bitmap images are supported for images residing within the document (*inline graphics*), but through separate programs called Helpers, various types of multimedia files -- such as QuickTime, MPEG movies, JPEG images, and AU audio files can be accessed. NCSA developed and supports X Window, Macintosh, and Window client versions of Mosaic. Others have contributed World-Wide-Web clients for NextStep, VMS, and DESQview/X, and non graphical terminal screen windows (e.g.Lynx, www).

Hypertext Transport Protocol (HTTP) is a relatively simple data communications protocol designed for speedy transport of text files, graphics and the like over wide-area networks. The heart of the World-Wide-Web is the server, which commonly is the HTTPD, or Hypertext Transport Protocol Daemon. Other WWW servers include GN [3] and the Perl implementation, Plexus. Microsoft Windows NT WWW servers are under development in several locations [2]. HTTP runs on top of TCP and maps each request to a TCP connection (default TCP port of 80).

The third component of WWW is the concept of the Uniform Resource Locator (URL), sometimes called the Universal Resource Locator or Unified Resource Locator. The URL provides a standardized specification for objects or resources on a network. The URL allows both document retrieval and full-text search operations as an HTTP object. HTTP objects are identified by the HTTP protocol type, the corresponding server's name, and the path name to the file where the objects' contents reside. Parts or subsections of a document can also be specified. If a search operation is requested, the URL or HTTP object identifier carries the set of specified keywords instead of the path name.

An example of a URL for obtaining a document containing an inline image is the following: "http://www.mit.edu:8001/usa.html" will connect to the WWW at MIT via port 8001 and obtain the file usa.html, a weather map of the United States. The first part of this URL is the protocol specification, in this cast is "http". The second part of the URL is the name of the host where the resource resides. If there is a colon after the resource name, this is the particular port to be used to access the file following the final "/". The port number may be optional. Other URL protocols include telnet, File Transfer Protocol (FTP), WAIS, gopher[4], and Network News Transport Protocol (NNTP). Depending on the protocol other pieces of a URL may be optional.

MATERIALS AND METHODS

The WWW server GN software [3] was implemented on an existing DEC 3000-800 computer running Digital Equipment Corporation's implementation of the OSF/1 operating system. The GN server allows rapid dissemination of both hypertext and plain-text information to both WWW and gopher clients, respectively. The GN server software was selected on the basis of its simplicity of design, efficiency of serving documents to both the WWW and gopher clients, and the security it afforded the server.

OncoLink is accessed at no charge by WWW clients via the URL: http://cancer.med.upenn.edu/. Gopher clients can access OncoLink by pointing to cancer.med.upenn.edu. The security offered by the GN server software is due to tight control of server behavior by allowing only the behavior specified in a menu file. While this menu file is cached to provide speedy service, no externally suggested documents or actions are supported if they are not explicitly stated in this menu. This arrangement tightly controls the behavior of the WWW server and appears to preserve the integrity of the computer and its resources.

OncoLink's top level menus are divided into four general categories: (1) disease sites (e.g. Breast Cancer, Ovarian Cancer), (2) medical specialty (e.g. Pediatric Oncology, Gynecologic Oncology), (3) news items (e.g. What's New, Public Announcements, etc.), and (4) links to other cancer-related resources on the Internet. Extensive hypertext links between the documents contained within OncoLink and documents contained elsewhere on the Internet enhance the navigational value of WWW in seeking information. The user may start in a promising document and follow hypertext links to other documents by merely clicking on the hypertext keys, focusing their attention on seeking specific information and not concerning themselves with logistics of accessing documents.

In order to track usage patterns, a detailed log

of all accesses to OncoLink is recorded. The log contains the Internet name, resource requested, and the time of the request for each OncoLink access. This log is a very important and powerful aspect of information publishing via the Internet, providing a passive feedback to OncoLink's authors and maintainers. In addition to furnishing an insight to OncoLink's most popular resources, the time between hypertext key requests and the path between the selection of one key and the next hypertext key gives an insight to the information sought, what portions the documents were viewed and how long the requester spent on each section of a document. By examining the interchange between the WWW client and OncoLink, as well as the gopher client and OncoLink, this log also revealed how many requests were from (a) graphical WWW clients, (b) text-based hypertext capable WWW clients and (c) gopher clients. This detailed tracking information is invaluable in assessing OncoLink's use.

We define an access as a specific request for a document or resource within OncoLink. A gopher client requests a minimal number of OncoLink resources (ie. one access per document), while a multimedia WWW client such as Mosaic, may request five or six items to download a document containing four or five inline images (i.e. one for the document, and additional accesses for each inline image or icon). Since these accesses are logged only with the Internet name of the WWW or gopher client (and not the specific user), we cannot further quantitate the actual number of individuals using OncoLink. Like a scientific journal subscription, a single copy of journal to a library may be read by many people while a single copy of a journal to an individual is read by only a few people. In the case of commercial services like America OnLine, the usage is represented by 2-3 addresses accessing extensive portions of OncoLink with great frequency.

RESULTS

OncoLink was released on the Internet as a WWW and gopher server in the evening of March 7, 1994. Between March and July 31, 1994, a total of 233,784 accesses were recorded from more than 62 identifiable countries around the world. Approximately 80% of these access were from WWW clients. On a month by month analysis, March yielded 34,972 accesses, increasing steadily to 63,956 access in July of 1994. The average weekday access rate is generally three times heavier than weekend use. Within a given day, OncoLink is accessed heaviest between 8am to 11pm EST, though use between 1-6am EST typically is typically heaviest from Australia, New Zealand, and Japan, reflecting their working hours.

From the Internet names, we have determined the scope of OncoLink users to be international, with accesses from countries originate from essentially all settled continents., even if English is not the primary language of the country (e.g. Japan, Germany, Poland, Israel, Brazil, etc.). OncoLink has only been advertised in the electronic environment, with minimal conventional print-media announcements.

Gopher access could be discerned by the selection of the non-HTML files. In the early days of OncoLink, the cancer resources served by gopher and WWW were essentially identical, the gopher resources lacking only the multimedia and the hypertext. As OncoLink developed, the gopher menus were extended to allow access to contents formerly only available within the hypertext keys of HTML documents. Documents such as the "Pediatric Oncology Case of the Month" progressed in a more non-linear direction, requiring access to hypertext keys to access the treatment given, etc. While still represented in the classical linear or sequential fashion present in the printed literature (i.e. the entire document is laid out in a chronological sequence), the hypertext documents such as the "Case of the Month" remains the most popularly requested document.

When the HTML files are selected, inline images were sent to the multimedia capable WWW clients. WWW clients operating in a text-mode do not request the inline images, allowing our discrimination of multimedia and text-only WWW clients. Approximately one third of the WWW clients accessing OncoLink appear to be text-only clients. This is reflected by their predominate selection of non-image based hypertext keys.

The choice of image based hypertext keys, existing as small icons of what was to follow, appeared to be more popular than a text-based keys. In the March "Case of the Month", the diagnostic pathology and radiology used pictorial icons of the resource contained in the hypertext keys. These icons were selected much more frequently than when the hypertext key was the text: "a CT reveals...". This observation was made on only two case studies and does not take into account the newness of OncoLink, the novelty of icons in an image, the interest of the case to the reader, etc. Further study is necessary to evaluate the correlation of pictorial hypertext keys versus text-based hypertext keys and use.

Large documents were subdivided to allow the user to read the document in a non-linear or arbitrary order. This subdivision was performed for both gopher and WWW services. While this was originally done to allow certain subdivisions of a document, such as "Coping with Survival" to be the hypertext destination

of other HTML documents, we observed most users read the documents in a non-linear fashion. For larger documents (e.g. >30 printed pages), the user frequently viewed approximately 10% of the document before selecting another OncoLink hypertext key (sending the user to a different part of the same document or another document). Tracing the hypertext pathway of a user's session revealed a focus on acquiring particular information. Users proceeded through the disease oriented documents, navigating through the chemotherapy agents used for the particular disease, searching the support lists for groups relevant to the particular disease, etc. Qualitatively, we believe that increasing the number of hypertext links to documents contained elsewhere on OncoLink increased their frequency of being viewed. Documents accessed only by descending through several menus appeared to be accessed less frequently than documents at a similar menu hierarchy containing many hypertext links. This observation is very preliminary and must be taken only in a qualitative sense. This observation appears to reflect the greater ease of navigation using hypertext links, versus the requirement of choosing relevant documents from a menu.

In late May, OncoLink won the international award for "Best of the Web'94" in the Professional Services division, clearly setting the pace for future World-Wide-Web development.

DISCUSSION

OncoLink was originally constructed around the medical specialties of pediatric oncology, radiation oncology, medical oncology, surgical oncology, medical physics, psycho social support and other resources. These subjects reflected the expertise of two co-Founders at the University of Pennsylvania Cancer Center (ELB and JWG). Gynecologic Oncology was added as the result of a strong contributing interest from another member of the Cancer Center (IB). Within the Pediatric Oncology menu, several submenus appeared, including a "Case of the Month", medical manuscripts, surveys (e.g. Pediatric Radiation Oncology Sedation Survey, Cancer Survivor Survey), and teaching files for instruction in pediatric oncology. Hypertext links to other documents, such as specific keys to documents discussing side-effects of chemotherapy, were liberally placed throughout these documents to enhance the navigational value of WWW in seeking information. This forum also provides the opportunity for education and wide distribution of "rare" entities, such as the short movie clip of opsoclonus, a fairly rare condition seen only in large medical centers. Nonetheless, physicians are still expected to recognize this condition.

While this model using medical specialties appeared to suit the academic community well, patients and other users of OncoLink informally requested (via email) a disease oriented approach. The contents of the disease oriented menus, such as breast cancer, were rapidly assembled from existing resources present in medical, surgical and radiation oncology sections.

The issue of presentation and navigation through the immense quantity of information regarding cancer make resources like OncoLink valuable. Despite large amounts of journals, pamphlets, etc. on cancer, patients and physicians alike comment on the difficulty in accessing and understanding this information. Printed information may be outdated and thus no longer reflect current practice. In addition to providing original material, OncoLink functions as a lens to focus the bewildering array of cancer-related Internet resources for the information browser. The benefits for patients and healthcare providers of being able to easily access timely and detailed cancer-related information via a multimedia point-and-click interface is formidable.

The selection and placement of information on OncoLink was largely a function of the interests and expertise of its founders. Currently, OncoLink material is solicited for original publication on OncoLink either from within the University of Pennsylvania or from outside authorities. Other sources of OncoLink material include information reprinted (with written permission) from other sources, reports from cancer and medical agencies (e.g. FDA, CDC, etc.), and unsolicited material of interest to the OncoLink audience. All material is peer reviewed prior to release on OncoLink.

While electronic dissemination of cancer information has advantages in both speed and the ability to navigate through immense quantities of information, OncoLink is only as good as the quality and timeliness of information it contains. Peer review has worked well in the areas for the limited number of specialties covered to date. Rapid peer review of the contents of OncoLink is critical. In cases where there is controversy, we believe the controversy should be noted and the cogent aspects of each side presented in a fair fashion. We are in the process of establishing an editorial board of experts in a wide-variety of cancer-related disciplines. This will broaden the appeal and usefulness of OncoLink to patients and healthcare providers.

To our knowledge, there are only a handful of resources devoted to cancer on the Internet: the National Cancer Institute's PDQ gopher (helix.nih.gov) and the gopher at the M.D. Anderson Cancer Center (utmdacc.utgh.tmc.edu) (see [8] for a review of online resouces). While there will certainly be more gopher servers providing detailed cancer information in the future, we believe gopher severs are hampered by their

inability to present multimedia information and provide hypertext navigation through the potentially vast quantities of information available on cancer. WWW servers devoted to a particular topic currently seem to be developing along lines following the discipline of Diagnostic Radiology[7]. However, most WWW servers located at major medical centers devote most of their resources to information pertaining to the facility, academic programs of the facility and the research pursuits of their faculty. Very few WWW servers are devoted to particular topics of broad interest to the international population, a feature that makes OncoLink stand out among the rapidly increasing number of WWW participants.

Electronic manuscripts must be recognized modes of publication, akin to publication in conventional peer reviewed journals. While issues such as copyright clearly must be addressed [5], the ability to rapidly communicate cancer information to a wide population in a easily navigated fashion addresses many of the problems of cost, storage and timeliness in conventional medical libraries.

CONCLUSION

OncoLink is a electronic library of cancer information potentially available to millions of people throughout the world. OncoLink addresses physicians, health care personnel, patients and their supporters on both medical and non-medical issues related to cancer. Using a simple point-and-click interface, Oncolink allows the user to navigate the rich multimedia resources maintained locally within OncoLink as well other oncology related Internet resources. We believe OncoLink is the first WWW and gopher server devoted to detailed information regarding cancer.

Based on our results, we believe: (1) There is a significant interest in cancer-related resources in the Internet community; (2) we believe users tend to browse cancer related information in a non-linear fashion; (3) we believe the graphical hypertext links stimulate users to browse documents more frequently and extensively; (4) we feel the content of OncoLink should be carefully peer-reviewed by experts spanning the various cancer-related disciplines. Therefore, we intend to expand the editorial board of OncoLink.

We conclude that OncoLink as a medical information resource is both useful and has wide interest in the international community. We believe, in the future, such systems will become common media for the international dissemination of cancer and other medical information.

References

[1] T. Berners-Lee et al. "World-Wide-Web: The Information Universe", Electronic Networking: Research, Applications, and Policy, Vol 2, No. 1, Spring 1992, pp. 52-58.

[2] HTTPS is a multithreaded Windows NT "service" accessible by anonymous ftp from enwac.ed.ac.uk and is available for both Intel and DEC Alpha architectures. HTTPS was produced as part of the European Microsoft Windows NT Academic Centre (EMWAC) project. HTTPS supports HyperText Transport Protocol (WWW) and gopher services under Microsoft Windows NT.

[3] The GN software is available by anonymous FTP or by setting a WWW URL to ftp: // ftp.acns.nwu.edu/ pub/gn/ gn-2.12.tar.gz. A later version may be present on this server, so the "2.12" may be different. GN was written by: John Franks from: Dept of Math. Northwestern University.

[4] Gopher originated at the University of Minnesota as a campus-wide online information system. Gopher is based on using a hierarchy of menus to access information contained either locally or elsewhere on the Internet. The information retrieved by a gopher server is usually text based, though there have been some discussions of extending the support to other types of information.

[5] The Journal of Medical Imaging has written a whitepaper on electronic publications in the medical imaging community. This document is accessible on the WWW by: http://jmi.gdb.org/JMI/ejourn.html.

[6] T. Berners-Lee, R. Cailliau, A. Luotonen, H.F. Nielsen, A. Secret, "The World-Wide Web", Comm. ACM 37:76-82(1994).

[7] J.R. Galvin, M.P. D'Alessandro, W.E. Erkonen, T.A. Knutson, D.L. Lacy, "The Virtual Hospital: A New Paradigm for Lifelong Learning in Radiology" Radiographics 14:875-879(1994).

[8] K. Nagy, "Cancer Information Seekers Cruise the 'Highdays'" J. NCI 86:1115-1118(1994).