Journal of Science Education and Technology, Vol. 11, No. 2, June 2002 (© 2002)

Broken Links: The Ephemeral Nature of Educational WWW Hyperlinks

John Markwell[1,3] and David W. Brooks[2]

The use of distributed (Internet) resources to enhance both traditional and distance education has caused much excitement in the science education community. However, one of the difficulties with relying on such freely available distributed resources has been the lack of certainty that the resources will be available for students next month, next semester, or next year. We have recently been involved in the development of three graduate-level biochemistry courses designed for high school teachers. Development of these courses relied heavily upon distributed science education resources. As a consequence, they represented a set of authentic science education resources that could be monitored over time to determine their rate of extinction. In total, the three courses contained 515 nonredundant URLs representing either scientific content of science education pedagogy. These have been monitored on a monthly basis during the 14 months since the creation of the courses (August 2000). During this period 85 (16.5%) of the URLs have ceased to function or had their content changed. The most attrition was seen in URLs with the "edu," "com," and "org" domain names, in which 17.5, 16.4, and 11% have already become inaccessible.

KEY WORDS: chemical education; computer assisted instruction; distance learning; Internet.

INTRODUCTION

Many of us teaching chemistry and biochemistry classes use links to WWW materials embedded in our class notes, or point our students to sites with appropriate background information and useful animations. However, use of these distributed resources has also resulted in a new problem that was not previously found with textbooks and library reserve materials—broken links. These are a problem for the student, who knows that the instructor included the link because it pointed to relevant supplementary information. They also are frustrating for the faculty member who must decide how to respond to the sudden, unexpected deficiency in study materials.

Early surveys of Web use cited broken links as the second most serious problem on the Web—following behind slow connections (Rogers, 1998). "Link rot" is another term used to describe this phenomenon (Daniels, 2001). Broken links continue to be cited as a serious problem with Web usage (Kobayashi and Takeda, 2000). Writers often discuss strategies for maintaining links (Davis, 1999). Descriptions of Web site development often contain explicit references to tools for dealing with broken links (Fowler *et al.*, 2000; van Harmelen and van der Meer, 1999). Numerous software tools are available commercially to address this problem, and services are available for hire to survey sites over the web with the intent of testing links and reporting broken links.

While the problem of broken links often is discussed, specific information about the seriousness of the problem is largely nonexistent. We had the impression that "disappearance" of distributed materials was accelerating because of the migration of class materials from open WWW servers to password-protected servers running course software such as

[1]Department of Biochemistry, University of Nebraska, Lincoln, Nebraska 68588.
[2]Center for Curriculum and Instruction, University of Nebraska, Lincoln, Nebraska 68588.
[3]To whom correspondence should be addressed; e-mail: markwell@unl.edu

BlackBoard or *WebCT*. This has certainly been the case for course materials created by faculty at the University of Nebraska.

A STUDY OF BROKEN LINKS

We recently collaborated in the development of three one-credit graduate-level Biochemistry courses intended for distance delivery to high school science teachers (NSF, ESI-9819377). These courses are currently being offered to North American teachers for credit by the University of Nebraska as Biochemistry, Chemistry, or Curriculum and Instruction 869K (Biomolecules), 869P (Energy and Metabolism), and 869N (Molecular Biology). These courses may be viewed at the following URLs:

Biomolecules: http://dwb.unl.edu/Teacher/NSF/C10/C10.html
Metabolism: http://dwb.unl.edu/Teacher/NSF/C11/C11.html
Molecular Biology: http://dwb.unl.edu/Teacher/NSF/C08/C08.html

One of the unusual aspects of these distance education courses was the strong reliance on distributed resources freely available over the Internet. The above three Biochemistry courses contain 515 non-redundant links to WWW sites containing scientific and pedagogic materials. Following completion of these courses in August 2000, we realized that it would be useful to follow the lifetime of these authentic science education links. Starting in November 2000, each of the URLs was checked at midmonth for viability and to ensure that the content was still accessible. These checks were conducted during the work week, since it is common to schedule computer down-time for weekends.

During the past thirteen months, the number of viable links has steadily decreased (Fig. 1). By mid September 2001, 16.5% of the links viable in August 2000 had disappeared or were nonviable. To study whether some types of distributed materials might be more ephemeral than others, we broke down the URLs by domain designation (Table I). The most stable URLs contained the "gov" designation and were 96% viable after 13 months. Three types of URLs that were more transient were the "edu," "com," and the "org" domains. We were not surprised that the "edu" URLs were volatile, since they are usually maintained by individual faculty, for individual courses, and both are subject to frequent changes. Fully 17.5% of these URLs had disappeared during the monitored 11 months. Also surprising were the proportion of "com" (16.4%) and "org" (11.6%) which had disappeared in slightly over a year.

As can be seen from the exponential curve-fit analysis displayed in Fig. 1, the URL extinction is a steady, progressive phenomenon and not correlated with end of the academic year. This type of "extinction equation" is commonly used to follow natural processes such as radioactive decay and the absorption of light. The points fit the equation $N = N_0 * e^{-(0.0124*x)}$ with an r^2 equal to 0.95. This

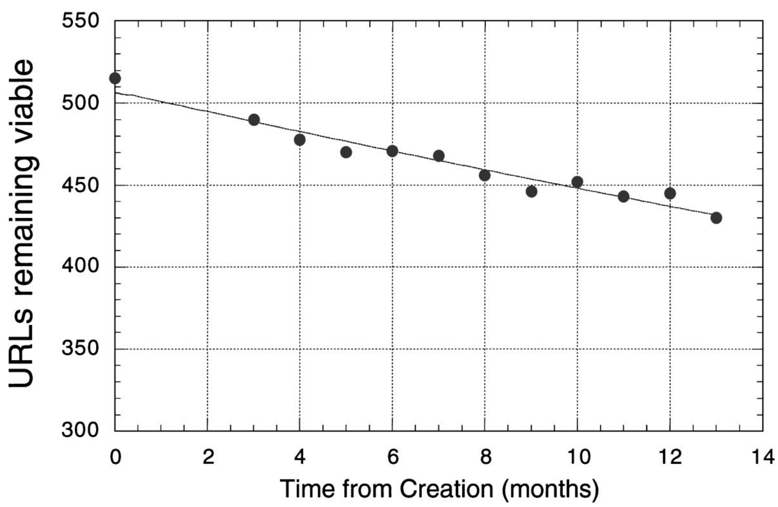

Fig. 1. Rate of loss of all URLs associated with this project from August 2000 through September 2001.

Table I. Distribution, by Domain Designation, of the 515 URLs in the Three Courses and Their Viability During the 13 Months of Study

Domain	Total URLs	Number remaining	Percent remaining
.edu	234	193	83
.com	73	52	84
.org	60	53	88
.gov	53	51	96
.uk	34	33	94
.de	17	16	94
.ca	14	11	93
.net	7	7	[a]
.se	4	4	[a]
.us	4	3	[a]
.jp	3	1	[a]
.kr	3	0	[a]
.za	3	3	[a]
.at	1	1	[a]
.fr	1	0	[a]
.il	1	0	[a]
.nz	1	1	[a]
.mo	1	1	[a]
(numeric)[b]	1	0	[a]
Total	515	430	83

[a]Too few to be significant.
[b]A URL with a numeric identifier rather than alphabetical domain name.

indicates a URL half-life (time for decrease in N to one-half N_0) of 55 months. This value is, of course, an estimate of the initial loss of URLs and the actual extinction process will probably be a more complex function and not so simple to extrapolate. For example, the "gov" domain URLs, which account for over 10% of the total population, have been much more stable than other domains for which we have large enough samples to use statistical analyses.

CONCLUSIONS

We are now less enthusiastic about large-scale utilization of Internet distributed resources for distance education than we were at the start of this project. The progressive disappearance of materials presents a major problem for courses developed to utilize these resources extensively. Distributed resources can evolve and can become extinct. They are not equivalent to an eclectic library instantly accessible from your office or home computer.

Not all distributed science education resources are created equal. Some are clearly intended by their creators to be a distributed resource, are regularly updated, and continue to evolve. Some are created on an *ad hoc* basis for a particular course or even for a single semester; these resources may have a brief lifetime. Another trend that has been noticed is that some "com" URLs which were accessible without cost in August 2000 (e.g. www.britannica.com) have changed their policy and are no longer accessible without cost. It was also noticed that material offered by a news magazine (www.usnews.com) was moved to an archive after a period of time and no longer freely available. An additional problem with use of the "com" URLs has come to light during the past month. Two of these URLs, one from the courses representing the current study, and one from another Biochemistry course taught at University of Nebraska, have apparently been sold and are now presenting advertisements for sexually explicit material.

In the short term, the best alternative for insuring that hyperlinks to distributed material remain accessible is probably a technique known as "ripping." This involves making a local copy of materials available on a different server. Ethics demand that one contact the author of a site for permission prior to making such a copy. Not only is seeking permission an ethical consideration and professional courtesy, but it alerts the source provider that you should be notified when material is updated or substantially modified.

A longer term solution is needed to provide the most valuable resources for a wide science education audience. A professional society interested in education could play a valuable role by actively reviewing and archiving (mirroring) the best and most relevant educational materials developed by its members. Such peer-reviewed hosting could prove useful to educators by providing documentation of authentic scholarship in science education.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grant No. ESI-9819377. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Daniels, J. (2001). *iBoost Journal.* Retrieved September 30, 2001, from http://www.iboost.com/promote/marketing/visitor_retention/20018.htm

Davis, H. C. (1999). Hypertext link integrity. *ACM Computing Surveys* 31(Suppl.): U164–U168.

Fowler, S. L., Novack, A. M. J., and Stillings, M. J. (2000). The evolution of a manufacturing Web site. *Computer Networks—The International Journal of Computer and Telecommunications Networking* 33: 365–376.

Kobayashi, M., and Takeda, K. (2000). Information retrieval on the Web. *ACM Computing Surveys* 32: 144–173.

Rogers, J. D. (1998). GVU's Ninth WWW user survey graphs. Retrieved September 30, 2001, from http://www.gvu.gatech.edu/user_surveys/survey-1998-04/graphs/use/q6.htm

van Harmelen, F., and van der Meer, J. (1999). WebMaster: Knowledge-based verification of Web-pages. In *Multiple Approaches to Intelligent Systems, 12th International Conference on Industrial Engineering and Applications of Artificial Intelligence and Expert Systems.* Vol. 1611, pp. 256–265.