

Distributed Digital Library Architecture: The Key to Success for Distance Learning

William J. Adams

Department of EE & CS
United States Military Academy
West Point, NY 10996
Voice: (914) 938-5575
FAX: (914) 938-5956
adams@exmail.usma.edu

Bernard J. Jansen

Department of EE & CS
United States Military Academy
West Point, NY 10996
Voice: (914) 938-3233
FAX: (914) 938-5956
jjansen@acm.org

Abstract

Cost cutting and personnel restructuring are forcing organizations to make difficult decisions on where to spend money. One of the areas hit hard by budget cuts is education and training. Academic, industrial, and governmental institutions are all seeking means to leverage technology to improve the timeliness, efficiency, and standardization of their required training. One way to extend budgets while continuing to deliver training is by constructing distributed digital libraries. A distributed digital library consists of material on separate machines connected via a network. The challenge of managing this information is deciding how to store the information and how users will connect, search, and retrieve the material. One method, the monolithic library, forces all user interactions through a single, controlling node of the library network. Another is called the distributed library, which hides the actual server architecture by allowing the user to interact with whichever library node is nearest to him. Using the model of the U.S. Army's Army Training Digital Library as an example, this paper will discuss challenges and solutions to indexing, searching, and retrieving material from globally distributed digital libraries. In particular, this paper will compare the costs and benefits of using a monolithic library structure with that of a distributed digital library.

Keywords Digital Libraries, Distance Learning, Information Retrieval

Please Cite: Adams, W. J., Howard, R. and Jansen, B. J. 1998. Distributed digital libraries: The key to success for distance learning. Computers and Technology in Education. Cancun, Mexico.

[See Other Publications](#)

Background

Schools, businesses, and governmental organizations are turning to Distance Learning to bolster enrollments, share expertise, extend the geographical extent of training programs, and broaden their customer base [1] [2]. Within an educational environment, Distance Learning (DL) provides a means to keep faculty employed and low enrollment courses viable through video-conferencing and digital libraries [6] [3]. Delivering an instructor's video taped lecture or printed material is fairly well documented and routine. The challenge is in formulating a method to distribute interactive, multimedia educational resources on large scale, in a timely and cost-effective manner.

Research is continuing on the most effective utilization of networks for DL [7]. Current DL programs are using a mixture of 3.5" floppy disks and CD-ROMs to distribute multimedia files to students [4] [5]. The 1.44 MB size restrictions of a floppy disk make it impractical for anything but text files or application files to be distributed. Likewise, CDs present two challenges to instructional developers. First is the 650 MB capacity limit of the compact disk. With a typical AVI file averaging 12 MB per minute of video, this storage limitation is quickly reached. Second, changes to any file on the disk, no matter how slight, requires

a new master and all of its accompanying charges, effectively erasing most of the savings realized through the use of CDs in the first place.

To provide learning materials quickly and efficiently to students, many schools have turned to digital libraries. Digital libraries are storage, access, and retrieval repositories of multimedia information usually accessed via some type of network connection. These server-based edifices allow any student with network access, either through a local area network or dial-up access, to access or browse learning material of any type. Because of the immediacy of server access by instructors and developers, the material is the most current it can be, without the lag time of disk mastering or disk distribution. As much, digital libraries become a natural extension of any distance learning plan.

Problem

There are some issues with digital libraries, however. Digital libraries must be accessible if they are to be successful. Therefore, where the information is stored and how these storage sites are interconnected is of concern. Of equal importance is that material in the digital library must be arranged and organized so that users do not spend an inordinate amount of time searching for the material they need. To be successful, digital libraries must provide accurate information, of sufficient quantity, and in a reasonable amount of time.

Solution

The best manner to accomplish these tasks is still a subject for research in the networking, information retrieval, and digital library fields. However, some trends are emerging. There are two ways to ensure accessibility. The first method is to consolidate all the material at a single site. This method is suitable for small schools or single colleges, institutions with a narrow range of topics or a small staff of developers. Larger, more diverse organizations need a more distributed solution. The organization and retrieval of networked information is also an on-going research area. The exponential growth of the World-Wide Web (W3) can provide much information about the characteristics one can expect from a digital library's typical users. Although this is valuable information, the W3 may not be the best model for organizing information in a digital library. The closed nature of digital libraries provides organizations with indexing and storage opportunities not available to the W3. Another way to view this is to think of DL as the W3 with some forced structure and chosen proponents for a given subject area.

While the investment in technology has been extensive, one of the remaining challenges is deciding on the best design to facilitate access and retrieval of information from the schools' digital libraries. The remainder of this paper focuses on the discussion and the decision-making rationale so far for a major distance learning project for the US Army. The goal is to share our experiences, successes, and lessons learned in order to aid other institutions in the development of their DL programs and digital libraries.

Background

The United States Army has made a significant investment into distance learning capabilities. The goal is for soldiers to be able to access training material from any location. In 1995, the U.S. Army's Training and Doctrine Command (TRADOC) embarked upon a plan that is intended on "leveraging technology to improve training" at the 21 Army schools. TRADOC is similar to a University and the 21 TRADOC schools are similar to a Department within the University. The 21 TRADOC schools are in various locations across the United States. TRADOC is responsible for developing the training doctrine, the training material, training standards for both military tasks and doctrine. Previously, TRADOC distributed this material via hard copy. The organization then moved to disk distribution; however, this soon proved unworkable on a global, Army-wide scale. TRADOC began exploring other options for distribution.

The long-term intent of the Army Distance Learning Plan (ADPL) is to provide high-quality, standardized training material to soldiers around the world. By providing access to this material, soldiers can receive training or review material in any combination of locations. First, attendees of resident courses will access training material in the classroom through local area networks. Second, upon returning to their workplace the soldiers can review the material from their local education centers, armories, Reserve Centers, or their offices by way of the Internet. In the near future, soldiers will use the Internet from their home, using their personal computer.

There are many ways to quantify the efficiency and cost-effectiveness of this vision. The most immediate means are strictly monetary. The cost of connecting users is undeniably less than the amount currently spent on: mailing and updating course books for correspondence and nonresident courses; travel and per diem costs for resident training; and the lost productivity of students that have to travel. The greater, but harder to quantify, benefit is the guarantee of standardized, on-demand training anywhere in the world. This benefit is especially important for the military, with a large, mobile, and geographically diverse population.

The key component to this architecture is the backbone connectivity that will be used to transfer the training material between school and student. Backbone connectivity is provided via the Internet. The Internet does not provide instant access to the DL; but it does provide a global path. Each DL on the network would contain a cached copy of material from other DLs. If the cached copy were too old or too big to store, then the local DL would request an update from the source DL.

Within the ADLP, the training material that is transported over the network is a combination of multimedia files. Video, audio, and text are the three largest components of learning resources. Once these materials are retrieved from their source, they are temporarily stored at the network access point's Digital Training Access Center (DTAC.) From the DTAC, users can view and replay the material at their convenience, for their use or for a class.

The planned training environment redefines the concept of the training site. The training site could be a school classroom, a learning lab at a military post anywhere in the world, or a National Guard Armory. Regardless of its location, the site is equipped with a set of hardware that enables a specified set of functions. These functions include: Internet access, World Wide Web browsing, multimedia capability, and local area network connectivity.

The distance learning retrieval process works as follows:

1. A soldier arrives for training. He accesses the local DTAC and requests training material.
2. The local DTAC searches its index of material that is currently stored there. If it finds what the user needs, the DTAC notifies the user that requested material is available and returns a list of Uniform Resource Locators (URLs) of the appropriate files. The process now jumps to step 5.
3. If the requested material is not on the local DTAC, the request is forwarded to the ATDL to find the course material he needs. This material supports correspondence courses, individual professional development, on-the-job training, or review of material learned at a previous resident course.
4. The ATDL searches its database to find course material related to the soldier's request. It returns a list of material in the form of URLs of the appropriate files.
5. The DTAC displays the list of material with estimated download times, calculated from the current network load. The user can select any or all of the material, which is then downloaded to the local DTAC.
6. Once this download is complete, the DTAC notifies the user and prepares for delivery.

Delivery is dependent on both media and the user. Text, either in the form of HTML or Adobe Acrobat files, is downloaded to the user via a web browser. Audio and video files represent more challenges however, although streaming products have simplified their delivery.

Discussion of Access

We explored two options concerning storage and access to information: a monolithic and a distributed solution. It is the management of the DTAC and its material that is at the heart of the challenge of a distributed digital library. The challenge is how to manage the storage, access, and retrieval of material from the DTACs.

The Monolithic Solution

Consolidating all an institution's academic material on a single server has several seductive benefits. Consolidated digital libraries can be easier to index, and users only have a single location to access. Unfortunately, after analyzing this option, the problems quickly outweighed the benefits. First, the challenge of providing users with a satisfactory response time is very involved. This is comprised of getting enough bandwidth to the site, spreading the traffic load among several servers, and providing security against the eventual hardware failure or network interruption.

Other, more subtle problems are also present. The first is the fact that a centralized library often has the unintentional effect of absolving the authors of ownership of their material. Each school is the proponent for a specific area of doctrine and training. For example, the Signal School at Ft. Gordon, Georgia is responsible for writing and maintaining all the Army's communications doctrine. This problem exhibits itself in difficulties in getting material updated, defeating the primary purpose of making material available electronically in the first place. Second, access and download times are proportional to the user's distance from the library. This could have the effect of undermining user confidence in the library's services. This is especially important concerning the TRADOC mission of providing training materials and courses to soldiers world-wide.

The schools (which own the DTAC machines) plan to use the DTAC as a place to store developmental materials. This would complicate indexing and retrieval because of the additional layer of marking to prevent users from accessing material that is not ready for distribution. It would also complicate the updating and transferring of information from the schools generating the training material.

The Distributed Solution

The distributed solution has some apparent problems. Namely, the information is not centrally located; therefore, it can be more difficult to manage. This is true in terms of hardware, networking, and information retrieval.

The distributed model has several advantages, however. For example, the distributed model provides a location and space for schools to develop and distribute material. Approved material is placed in a specific directory structure where a master index would know to look for it. A distributed model also has the advantage of eliminating a single point of failure. A mirror plan can be implemented where if one server is down, another server with a copy of the information can still provide access to users. With information stored locally, the authors retain ownership of the documents and multimedia stored on their server. The distributed model also can reduce access time for certain users, especially if highly utilized information is mirrored to different geographical locations. The distributed model would provide space for development and distribution needs. Approved material would be placed in a specific directory structure where the master index would search, catalog, and retrieve it.

Conclusion

Digital libraries, whether distributed across the globe or across a small campus, provide a quick and cost effective means to distribute learning resources to students, employees, or soldiers. The challenge for administrators is to facilitate the indexing, searching, and retrieval of material to prevent users from becoming overwhelmed by the task of finding the proper data. Two methods to manage a digital library were discussed: monolithic and distributed. The monolithic system used a single point to store and deliver all the course material for the organization. While this simplifies the maintenance and indexing of the digital library, user access and information maintenance can present a problem. The monolithic method may be best suited for organizations with small, highly specialized information bases.

The distributed system requires material that is approved for distribution to be placed in a specific directory structure on the dispersed machines for a single master index mechanism to find. Users contact the system through the closest machine, which uses a distributed index to find the material requested by the user. This method is suited for large or widely dispersed organizations, since it optimizes user access and dispersed storage. As such, the distributed method is best suited for the military and other dispersed organizations. The distributed system does, however, increase the maintenance complexity since index updates and material transfers can significantly impact system performance. Choosing the best system requires an analysis of the organization, its location, its user population, and the type of material being provided to the users.

Even with the distributed system, one must have a clear understanding of and a working model of a system's users. It appears from W3 usage, that most users rely on naïve searching techniques to access and retrieve information. However, digital libraries have an advantage over the W3 with more control of content and information organization. As such, digital libraries can take advantage of automated indexing, agreed upon set of keywords, and meta-data to describe documents. These techniques can improve retrieval of relevant information and improve the services that a digital library offers its users.

References

- [1] Dance, Muriel. *The Promise of Distance Learning*. <http://weber.u.washington.edu/~jamesher/mdance.htm>. Accessed November 1997.
- [2] Etter, D.M.; Orsak, G.C.; Johnson, D.H. *A distance learning laboratory design experiment in undergraduate digital signal processing*. 1995 International Conference on Acoustics, Speech, and Signal Processing. Conference Proceedings. p. 2885-7 vol. 5.
- [3] Fox, Edward. *Digital Libraries, WWW, and Educational Technology: Lessons Learned*. Proceedings of the World Conference on Educational Multimedia and Hypermedia. P. 246 - 251.
- [4] Harris, J.A.; Murden, C.; Webster, L.L. *The potential of interactive multimedia on CD-ROM to enhance laboratory work in physical science and engineering*. 1994 IEEE First International Conference on Multi-Media Engineering Education Proceedings. p. 296-301.
- [5] Lollar, R.B. *Distance learning for non-traditional students to study, near home, toward a UNC Charlotte BSET degree*. Proceedings IEEE Southeastcon '95 Visualize the Future. p. 366-7.
- [6] Palounek, Andrea P. T., et.al. *Distributed Computing Network for Science and Math Education in Rural New Mexico*. Proceedings of the World Conference on Educational Multimedia and Hypermedia. P 557-562
- [7] Stanford. *An On-Line Distance Learning System Using Digital Video and Multimedia Networking Technologies*. <http://minas.stanford.edu/project/project.html>. Accessed November 1997.