

Libraries for Digital Documents

Ray R. Larson

Associate Professor

School of Information Management and Systems

University of California, Berkeley

ray@sims.berkeley.edu

A digital document -- whether it is a simple text file, an HTML-based WWW page, a multimedia encyclopedia, or a collection of simple page images -- participates in larger aggregations of other digital documents. These aggregations of documents may be organized based on a variety of criteria, ranging from simple "file name order" to complex multimedia databases and Web sites. With increasing frequency, these collections of digital documents are being described as "Digital Libraries." But what is a library for digital documents? What are the features and characteristics of such libraries? There have been libraries of text documents for thousands of years, in media ranging from clay tablets to vellum to paper, with a rich tradition of organization and behavior for these libraries, but the character of digital documents demands a reevaluation of the practices and principles of organization of the "Digital Library".

The notion of the "Digital Library" as it has been developing over the past several years, is beginning to have a profound effect on the development of the National (and Global) Information Infrastructure. This notion assumes that Digital Libraries will not be single, stand-alone, repositories of digital data. Instead they will be a heterogeneous collection of network-based repositories using a variety of protocols for user (and repository) interaction, data encoding and transmission. The repositories may range size from small personal collections of information housed on a PC, to multi-terabyte repositories of remote sensing data downloaded from satellites and made available to researchers.

The vision of Digital Libraries is the vision of global, or "Universal Libraries", where a vast population of users scattered around the globe are able to access, easily and conveniently, the complete contents of thousands of large and small repositories containing texts, images, sound recordings, videos, maps, scientific and business data, as well as hypermedia combinations of these elements. The Digital Library

must, therefore, be a network-based distributed system with disparate local servers responsible for maintaining individual collections of digital documents ranging from sets of electronic texts to video-on-demand services. In effect, this Universal Library will actually consist of a set of "publishers" of electronic information, a set of "consumers" of that information, and a set of "service providers" offering added-value services (such as indexes, search services, and document delivery) distributed across the global network. In many ways, the "service providers" who attempt to provide some sort of coherent organization and integrated access to the information contained in the diverse resources provided by the "publishers" will be the true Digital Libraries of the future.

In this minitrack we will be concerned with a variety of issues relating to libraries for digital documents: ranging from historical perspectives on digital library issues, to the ever-present issue of how a user or consumer with a particular information need gets access to the digital documents containing the information that satisfies that need. Joao Ferreira, Joaquim Jorge and Jose Delgado discuss the digital library as a metaphor and how that metaphor has been materialized in the ArquITex project. Chaomei Chen describes the development of a virtual reality interface for information retrieval in digital libraries. Alan R. Heminger and Steven B. Robertson examine the critical issue of how to maintain long-term access to digital documents and formats and machine architectures change. David-Michael Lincke, Beat Schmid, Petra Schubert, and Dorian Selz explore the application of digital library methods to create a new medium to aid in creating, integrating, reviewing, and disseminating domain-specific knowledge. Raymond R. Panko examines the lessons learned in the development of the NLS/Augment Journal System, the first online hypertext library. Martha Waters and James McDonald discuss the application of digital library methods to managing the large scientific collections of Los Alamos National Laboratory.