NCSA Mosaic Technical Summary

Marc Andreessen
Software Development Group
National Center for Supercomputing Applications
605 E. Springfield, Champaign IL 61820
marca@ncsa.uiuc.edu

February 20, 1993

Abstract

We describe a networked information system targeted at wide-area distributed asynchronous collaboration and hypermedia-based information discovery and retrieval currently being designed and built at NCSA. The system, called Mosaic, provides powerful capabilities for dealing with scientific data in an asynchronous collaboration context and integrates cleanly into existing data sources and environments.

1 Introduction

The National Center for Supercomputing Applications at the University of Illinois is developing an environment to facilitate wide-area network-based information discovery and retrieval and asynchronous collaborative work. The environment, called Mosaic, provides a portal into a broad universe of network-based and standalone information, and any information thus accessible to Mosaic becomes a potential part of a user's asynchronous collaboration environment — including information not under Mosaic's direct control or on servers too simple to support complex collaboration-based activities.

Mosaic uses a blend of distributed hypermedia, hierarchical organization, and search functionality for information discovery and retrieval. In addition, Mosaic is fully capable of dealing with scientific data as a standard component type within its information space. Further, Mosaic is designed to be fully distributed across a wide-area network, in part by supporting a wide variety of common information protocols and formats and by providing simple, convenient methods for expanding the domain of information upon which Mosaic can operate.

2 Mosaic Clients

The user interface into the Mosaic system is provided by clients currently being developed at NCSA. These clients, X Mosaic and Mac Mosaic, are designed for the X Window System and the Apple Macintosh respectively. (Currently, the first version of X Mosaic is in beta release and Mac Mosaic is nearing alpha.)

The initial versions of the Mosaic clients will have the following functionality:

- Graphical display of plain text, rich (formatted) text, and hypertext, as well as inlined display of graphs, images, audio clips, video sequences, and scientific data as sections of multimedia and hypermedia documents.
- Information space navigation and history-tracking facilities.
- Support for "hotlists" of interesting documents that can be monitored for changes over time.
- Complex search capabilities including those required to access WAIS databases.
- Support for the asynchronous collaboration mechanisms described below.
- Full TCP/IP-based communications support (including native HTTP, Gopher, FTP, and NNTP protocol support).
- DTM communications support, for integration with NCSA Collage and other network-based DTM clients and information servers.
- DMF communications support, for exploring and retrieving data stored in NCSA's Data Management Facility.

Projected functionality for future releases of the Mosaic clients includes:

- Full hypermedia-based mail- and news-handling capabilities.
- Graphical history and information space representation and control, including fully configurable graphical construction of information subspaces.
- WYSIWYG hypermedia document editing.
- Support for collaborative filtering of information spaces — convenient construction of shared overviews and threads through large and dynamic masses of information.
- "Intelligent assistant"-based information filtering and retrieval capabilities.

A screen snapshot of the current beta version of X Mosaic is in Figure 1.

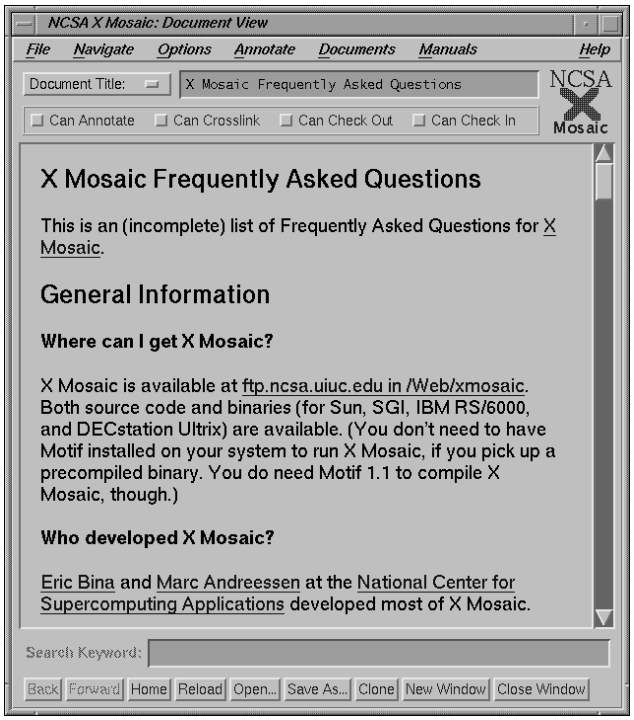

Figure 1: X Mosaic Document View Window

3 Asynchronous Collaboration

Asynchronous collaboration in Mosaic is focused on allowing users distributed across the network to collaboratively construct, modify, and annotate plaintext, multimedia, or hypermedia documents existing within the information space, and to easily form information subspaces — workgroup spaces for intense and sustained collaboration — within the universal information space.

One of the primary mechanisms for asynchronous collaboration within Mosaic is support for distributed document control and modification. (Note that the term "document" denotes, in this context, any network-based information resource, up to and including hypermedia constructs.) The following capabilities will be specifically provided by the first version of Mosaic:

- Document annotation — the ability to attach personal, workgroup, or public annotations to documents arbitrarily located on the network. Specifically, the Mosaic group annotation server mechanism will allow multiple, possibly overlapping groups of people to share annotations and to inform one another of new or modified annotations.
- Document crosslinking — the ability to attach hyperlinks to part or all of a document for yourself and others to see.
- Document revision control — the ability to check out, lock, and check in documents over the network and to control revision logs and version trees with a hypermedia-based front end.

Permission mechanisms for collaboration within the Mosaic system are being designed with an eye towards user convenience — while we want documents and data to be well-protected, we do not want security to unduly interfere with legitimate work. The ultimate security mechanism for Mosaic will most likely be Kerberos, but since Kerberos is not yet running on a wide scale, it is not an immediate solution. In the interim, we will use passwords and machine and domain identification, with all protection occurring on a per-document basis.

We are building both the previously mentioned Mosaic annotation server as well as specialized information servers that can support our server-based collaboration requirements. For example, we are building a HTTP server that uses RCS as a back end for document revision control. Other server technology will be built as it is necessary to support the capabilities we want to provide.

4 Scientific Data

Historically, hypermedia systems — and broad-based networked information systems in general — have not

supported scientific data as a basic information type. As it is specifically oriented towards scientific and academic research, the Mosaic system will support scientific data via the following functionality:

- Scientific data can be included within a hypermedia document in the same manner as images, audio, and video. The scientific data format of choice for Mosaic is NCSA's Hierarchical Data Format (HDF), a machine-independent file format that allows arbitrary grouping and annotation of heterogeneous data elements. (Data elements in formats other than HDF can also be included in Mosaic-compliant hypermedia documents.)

- Mosaic clients can communicate directly with Collage, NCSA's synchronous collaboration tool for scientific data analysis and manipulation, as well as with Polyview, NCSA's collaborative tool for three-dimensional geometric and polygonal data analysis. Scientific data contained within a hypermedia document can be sent across the network to one of these programs for graphical and statistical inspection and analysis. (In fact, the data can be sent to any program that understands DTM, NCSA's TCP-based data protocol — e.g., AVS — or, via Collage, to any program whose communications protocol has been included in Collage, e.g., Mathematica.)

- Mosaic clients can communicate directly with DMF, NCSA's scientific data management and archival facility. Mosaic provides a convenient and flexible hypermedia-based front end to DMF's relational database and mass storage model of information control and access.

- Data in HDF format will be directly browsable by Mosaic clients. Given an HDF file, the clients will be able to examine its structure and contents and generate a hypermedia document containing human-readable information describing the HDF file and its data. Hotlinks into the HDF file will allow the user to directly access any subset or portion of the data.

5 Technical Underpinnings

The core of the Mosaic system is based on the World Wide Web, a distributed hypertext-based information system developed at CERN. As such, Mosaic speaks a variety of common TCP/IP network protocols, including HTTP, Gopher, FTP, and NNTP, and has gateways into other network protocols such as WAIS, Techinfo, and Hytelnet. Mosaic also is fluent in NCSA's Data Transfer Mechanism (DTM) protocol, which allows direct communication with Collage and other DTM sources (e.g., running simulations and remote instruments) and clients. Mosaic uses the MIME format to support documents containing multiple heterogeneous data elements and the World Wide Web's HTML format for internal and document-based hypermedia capabilities. Datatypes directly supported in the Mosaic browsers include NCSA's Hierarchical Data Format (HDF), GIF, JPEG, XWD, MPEG, and several types of audio.

Mosaic is being based on standard and existing formats and protocols to encourage widespread use and applicability, and it will be extended to take new methodology and technology in Internet-based networked information systems and data formats into account as time progresses.

6 Availability

NCSA intends to make the complete Mosaic system publicly available and distributable. Currently, beta versions of X Mosaic are available in both source and binary form for common platforms. NCSA's anonymous FTP server is `ftp.ncsa.uiuc.edu`; project files are in directory `/Web`.

For more information on the Mosaic project, please feel free to contact the author.

7 Acknowledgements

X Mosaic is being implemented by Eric Bina and the author; Mac Mosaic is being implemented by Aleks Totic. The project lead for Mosaic is Joseph Hardin. We are indebted to the World Wide Web project at CERN for their vision, ideas, and code, as well as to the many people who are actively providing feedback on Mosaic's capabilities.