NCSA Mosaic Technical Summary

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

Revision 2.1 — May 8, 1993

Abstract

We describe a networked information system for wide-area distributed asynchronous collaboration and hypermedia-based information discovery and retrieval currently being designed and built at NCSA. The system, called NCSA MOSAIC, integrates cleanly into existing network protocols, formats, data sources, and environments, and provides powerful new capabilities for using and sharing information.

1 Introduction

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

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

2 NCSA Mosaic Clients

The user interface into the NCSA MOSAIC system is provided by clients currently being developed at NCSA. These clients, NCSA MOSAIC for the X Window System and NCSA MOSAIC for the Macintosh, are designed for Unix workstations running the X Window System and for the Apple Macintosh respectively. The first official version of NCSA MOSAIC for X has been released and NCSA MOSAIC for the Mac is currently nearing beta. (Development of NCSA MOSAIC for Microsoft Windows is slated to begin in early summer, 1993.)

The initial versions of the NCSA MOSAIC clients have the following functionality:

- Graphical display of plain text, rich (formatted) text, and hypertext, as well as inlined access to graphs, images, audio clips, video sequences, and scientific data in multimedia and hypermedia documents.
- An intelligent graphical user interface featuring on-the-fly font and style selection, cut-and-paste support for formatted text, and extensive customization and user support options.
- Information space navigation and history-tracking facilities, including persistent global history mechanisms.
- Support for "hotlists" of interesting documents that can be monitored for changes over time.
- In-document and network-based search capabilities, including those capabilities required to access WAIS databases.

- Preliminary asynchronous collaboration capabilities, including text and voice annotations for documents located anywhere on the network.
- Full TCP/IP-based communications support, including native HTTP, Gopher, FTP, and NNTP protocol support, and gatewayed access to WAIS, Techinfo, TeXinfo, Archie, finger, Hyper-G, X.500, Whois, and other data sources.[1]
- NCSA Data Transfer Mechanism communications support, for integration with NCSA Collage and other network-based DTM clients and information servers.
- NCSA Data Management Facility communications support, for exploring and retrieving data stored in DMF databases and directory servers.

Projected functionality for future releases of the NCSA Mosaic clients includes:

- Full support for the asynchronous collaboration mechanisms described in Section 4.
- Graphical history and information space representation and control, including fully configurable graphical construction of information subspaces.
- WYSIWYG hypermedia document editing.
- Full hypermedia-based mail- and news-handling capabilities.
- Support for collaborative filtering of information spaces — convenient construction of shared overviews and threads through large and dynamic masses of information.
- "Intelligent agent"-based information filtering and retrieval capabilities.
- Support for interacting with "knowledge daemons" — network-level sedentary intelligent mechanisms for tracking, organizing, caching, and sharing information.

A screen snapshot of NCSA Mosaic for X (viewing the Mosaic home page – the document that is retrieved and displayed when Mosaic is launched) is in Figure 1. Each underlined word or phrase is a *hyperlink* to another document anywhere on the network.

[1] NCSA Mosaic's communications support is courtesy of the CERN World Wide Web project's common client library.

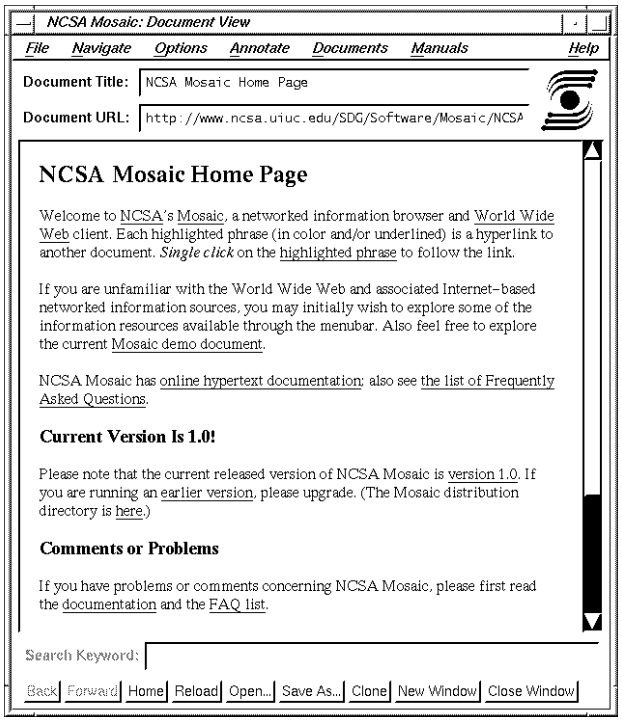

Figure 1: NCSA Mosaic for X — Home Page

3 Global Hypermedia

NCSA Mosaic provides extensive distributed hypermedia capabilities that take advantage of the information base on the global Internet. Any piece of information — any document — on the Internet in any of a dozen common and flexible information systems can be uniquely named and thereby retrieved, displayed, annotated, and remembered by Mosaic.

The scheme that NCSA Mosaic uses to name information resources on the global network is the *Uniform Resource Locator* mechanism, a developing IETF-guided standard specifically designed to envelop the existing universe of digital information within a single nomenclature. Uniform Resource Locators can point to documents residing on FTP or HTTP servers, news articles on NNTP servers, queries to WAIS databases, documents within WAIS databases, documents or search engines residing on Gopher servers, a specific sub-part of a document residing on any server, a set of metadata served from a relational database in NCSA DMF, a piece of data served from a mass storage system in NCSA DMF, or almost anything else that is available somewhere on the network through some network-based access mechanism.

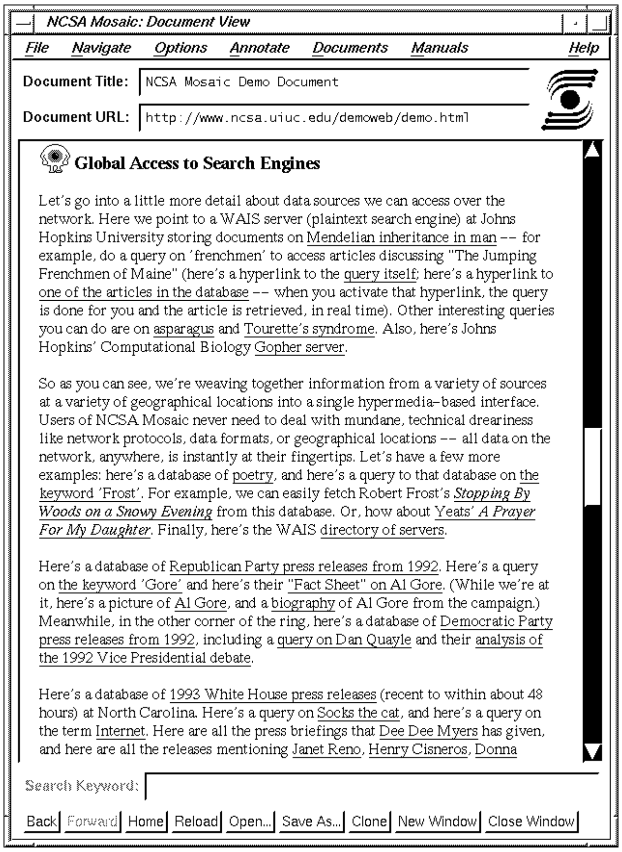

Figure 2: NCSA MOSAIC for X — Query Hyperlinks

The URL mechanism is *inclusive* and *open-ended*: it can encapsulate virtually any information system that accepts remote network queries, and it can evolve to match the capabilities of new systems that will no doubt be appearing on the Internet throughout the next several years. It is also extremely *flexible*: an example can be seen in Figure 2, which illustrates a hypermedia document within Mosaic pointing directly to queries on remote databases. Activating a hyperlink resolves the query and presents the user with the query results — for example, activating the hyperlink called "the keyword 'Gore"' in the document viewed in Figure 2 resolves the query and the results are shown in Figure 3 as a list of hyperlinks to the matching documents. Hyperlinks can also point straight documents within a database — activating the hyperlink called "analysis of the 1992 Vice Presidential debate" causes the referenced document to be pulled out of the database and displayed, as seen in Figure 4, without the user even having to be aware that the document was stored in a query-based database (in this case, a WAIS server).

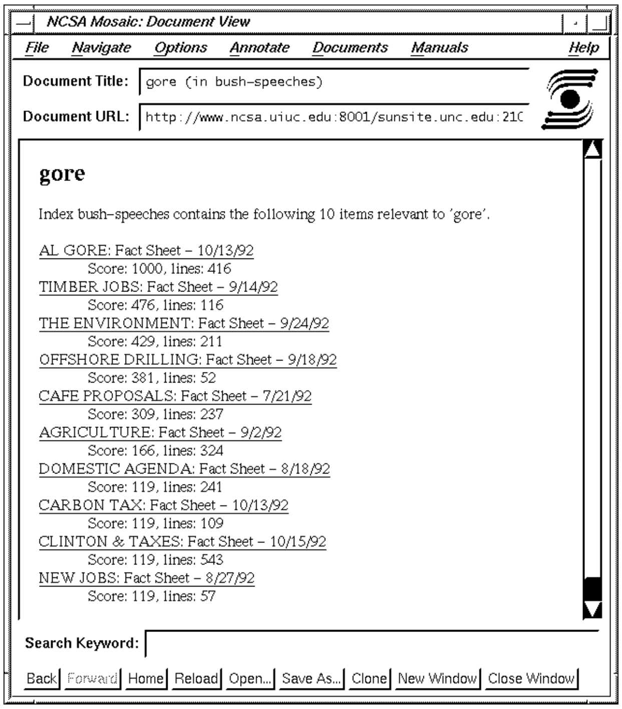

Figure 3: NCSA MOSAIC for X — Query Results

The result is complete transparency of data location and retrieval process: NCSA MOSAIC does everything it can to avoid burdening the user with the technical details of information navigation and retrieval over a wide-area network to allow the user to focus all of his or her attention on the task at hand — interacting with the information itself.

4 Asynchronous Collaboration

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

One of the primary mechanisms for asynchronous collaboration within NCSA MOSAIC is 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 NCSA MOSAIC:

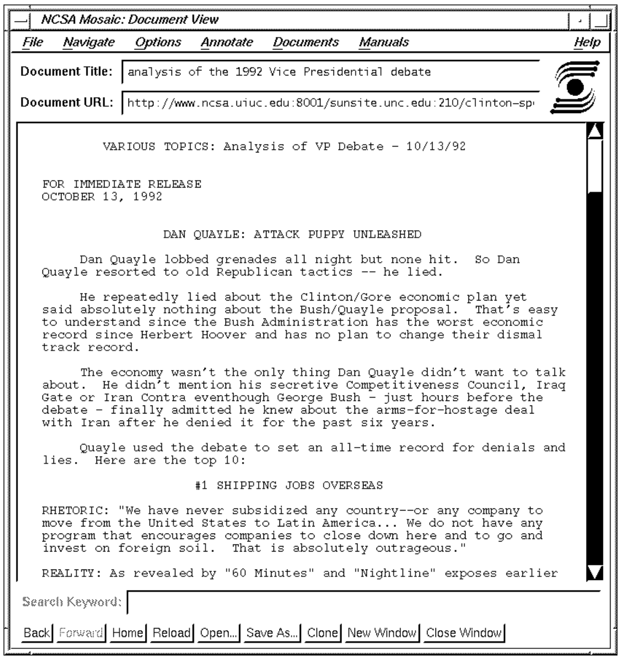

Figure 4: NCSA Mosaic for X — Retrieved Document

- Document annotation — the ability to attach personal, workgroup, or public annotations to documents arbitrarily located on the network. Specifically, the NCSA Mosaic group annotation server mechanism will allow multiple, possibly overlapping groups of people to share annotations and to easily inform one another of new or modified annotations.
- Document crosslinking — the ability to attach hyperlinks to part or all of a document, and to share those hyperlinks with collaborators as with annotations.
- 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 interface.

Permission mechanisms for collaboration within the NCSA 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 NCSA 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 NCSA 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 collaboration capabilities we want to provide.

5 Scientific Data

Historically, hypermedia systems — and networked information systems in general — have not supported scientific data as a basic information type. Since NCSA Mosaic is specifically oriented towards scientific and academic research, it 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 sequences. The scientific data format of choice for NCSA 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 NCSA Mosaic-compliant hypermedia documents, with display and processing of received data dependent on local system capabilities.)
- NCSA 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.) NCSA Mosaic will also be able to similarly communicate with future scientific data analysis tools developed by NCSA's Software Development Group.
- NCSA Mosaic clients can communicate directly with NCSA's Data Management Facility, a scientific data management and archival system.

NCSA 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 NCSA 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 or operate on any subset or portion of the data.

6 Technical Underpinnings

The core of the NCSA Mosaic system is based on the World Wide Web, a distributed hypertext-based information system developed at CERN. As such, NCSA 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. NCSA 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.

NCSA Mosaic uses the World Wide Web projects' HTML format for internal and document-based hypermedia capabilities and will use the Bellcore MIME format to support documents containing multiple heterogeneous data elements. NCSA Mosaic initially supports the X bitmap (XBM) and GIF image formats directly (an example can be seen in Figure 5), and provides interfaces to external viewers to handle other multimedia data formats (e.g., JPEG, XWD, TIFF, RGB, MPEG, DVI, PostScript, and several types of audio).

NCSA 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.

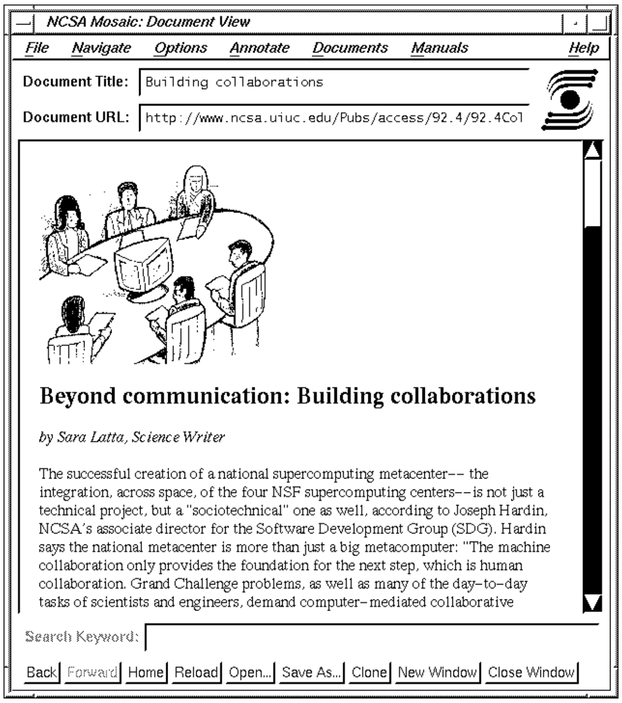

Figure 5: NCSA Mosaic for X — Inlined Image

7 Availability

NCSA intends to make the complete NCSA Mosaic system freely and publicly available and distributable for all academic and research organizations and purposes. Currently, version 1.0 of NCSA Mosaic for X is 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 `/Mosaic`.

NCSA Mosaic is free for internal use by commercial organizations, and is also available for licensing by commercial organizations for modification and/or distribution. For information, please contact the author.

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

8 Acknowledgements

NCSA Mosaic for X is being implemented by Eric Bina and Marc Andreessen; NCSA Mosaic for the Mac is being implemented by Aleks Totic. The project lead for NCSA Mosaic is Joseph Hardin. We are indebted to Tim Berners-Lee and the World Wide Web project at CERN for their vision, ideas, and common client library code, as well as to the many people on the Internet in two dozen countries who are actively providing feedback on NCSA Mosaic's capabilities.