Internet Information Discovery and Retrieval Tools - Cost Effective Building Blocks for Asset Libraries

Timothy Stockwell

The MITRE Corporation
MS-Z381
7525 Colshire Drive
McLean, VA 22102-3481
Tel: (703) 883-6695
Email: tim@mitre.org

Mark Krause

The MITRE Corporation
MS-W273
7525 Colshire Drive
McLean, VA 22102-3481
Tel: (703) 883-7642
Email: mkrause@mitre.org

Abstract

The MITRE Corporation's Center for Advanced Aviation System Development (CAASD) is addressing the issue of institutionalizing reuse within the corporation. Implementation of a distributed, heterogeneous asset library has been recognized as a critical step in achieving the vision of institutionalized information reuse. Recent advances in the field of network information discovery and retrieval technology have led to the CAASD development of a reuse library mechanism, at minimal cost, using widely available Internet resources. This position paper supports the use of Internet resource discovery tools as cost-effective building blocks for implementing distributed, heterogeneous asset libraries.

Keywords: library, architecture, interoperability, standards, full-text indexing, network information discovery and retrieval, Internet Gopher, WAIS

Workshop Goals: learning; sharing CAASD reuse initiative results; advocating the importance of standards for implementing interoperable libraries.

Working Groups: tools and environments, education, reuse management, organization and economics.

1 Background

In November 1992 CAASD started an initiative to support and promote software and information reuse within the center. The CAASD Resource Discovery System is a network information discovery and retrieval system built from Internet resource discovery tools and implemented as a client-server computing model. The system supports an asset library that is implemented across distributed, heterogeneous computer platforms. Client software is available on a wide variety of computer platforms, including VT220s, IBM PCs, Macintoshes, NeXT workstations, UNIX workstations, and X Window terminals. Server software is available for UNIX platforms. Client-server communication occurs over a network using TCP/IP. The major Internet resource discovery tools used in implementing the CAASD Resource Discovery System are the Internet Gopher and Wide Area Information Servers (WAIS). Descriptions of these tools and how they are used within the CAASD Resource Discovery System are presented below. The CAASD Resource Discovery System has been available for on-line access since January 1993, and has since received wide support at all levels in the corporation.

2 Position

Implementation of an on-line asset library was viewed as a crucial step in institutionalizing reuse within CAASD. The library mechanism had to be low cost, easy to set up and maintain, accessible from the wide variety of computer platforms used by CAASD staff, and it had to meet the requirements for an asset library mechanism. Specifically, the asset library had to allow for the automated search and retrieval of on-line information in the form of text documents, images, software, and pointers to sources of more information (meta-assets) with minimal barriers between CAASD staff and the desired information.

The CAASD Reuse Team informally reviewed a number of asset library mechanisms. None fit the set of candidate requirements. The search was expanded. The two main tools that were evaluated and eventually used to build the CAASD asset library were the Internet Gopher and Wide Area Information Servers (WAIS). Given the recent advances in network information discovery and retrieval technology, this class of tools seemed to lend themselves to the solution of the problem.

The Internet Gopher client/server provides a distributed information delivery system around which a campus-wide information system (CWIS) can be readily constructed. While providing a delivery vehicle for local information, Gopher also facilitates access to other Gopher and information servers throughout the world. [1]

WAIS is a networked information retrieval system that uses TCP/IP to connect client applications to information servers. Client applications request documents using user-supplied keywords and are able to retrieve text or multimedia documents stored on the servers. Servers search the full text index for the documents and return a list of documents containing the keywords. The client may then request the server to send a copy of any of the documents found. [2]

Gopher allows a user to browse for assets using a menu structure. If a user wants to access an item, Gopher will "go fer" it and, if possible, "show" it to the user. Different Gopher clients have varying levels of capability with respect to "showing" assets to the user. All clients can display ASCII text documents. Other document types which the Gopher protocol defines, and which may or may not be supported by the specific client software for a computer platform, are sounds, bit-mapped images, and multimedia documents. If desired, the user may save the document on the local disk drive, print the document on a local printer, or mail the document to an email address. Gopher can connect to other asset library systems via the Gopher protocol, TELNET protocol (VT100 or TN3270 emulation), or anonymous ftp. Gopher also has interfaces to several information servers, including Archie servers, CSO name servers (a class of electronic phone books), and WAIS servers. Archie is a system which allows searching indexes of files available on anonymous ftp servers on the Internet [3]. CSO name servers allow users to access certain electronic phone books. The Gopher-WAIS interface allows the user to specify keywords to be used for a WAIS search. The results of the WAIS search are presented to the Gopher user as a custom directory. Items in the directory are assets which match the search criteria. In addition, WAIS uses a complex scoring algorithm for the assets and presents the assets in sorted order, based upon the score of each asset. This feature saves the user effort by automating the process of determining which assets are most relevant, based on the user's search criteria.

The implementation of the CAASD Resource Discovery System consists of an Internet Gopher server and a WAIS server running on a UNIX workstation. Gopher client software is available for UNIX workstations (X Window and character based), Macintoshes, IBM PCs (DOS, Windows, and OS/2), and NeXT workstations. Dumb terminals (for example, VT220s) may access the system by logging into a special account and running a "secure mode", character-based Gopher client. The CAASD Resource Discovery System has links to other Internet Gopher servers maintained within MITRE, telnet links to the MITRE on-line library catalogs, and links to anonymous ftp sites within MITRE.

The CAASD Resource Discovery System acts as a library mechanism for many different types of assets available for reuse by CAASD staff. Asset types include technical reports, letters, memos, briefings, meeting minutes, numerous databases, and software. All of these asset types are searched, browsed, and retrieved using the common mechanisms provided by the CAASD Resource Discovery System.

A number of tasks have been accomplished to assist CAASD staff to utilize the CAASD Resource Discovery System for the purpose of software reuse. Software description templates have been designed and distributed to CAASD staff. These templates are used to capture information about software developed within MITRE as well as information about software that was developed externally from MITRE, but used within MITRE. When the templates are filled in, these software descriptions are collected, indexed, and made available through the CAASD Resource Discovery System. In addition, catalogs and information about external software repositories, such as ASSET, AdaNET, and the NETLIB server at research.att.com, have been reformatted, indexed, and made available through the CAASD Resource Discovery System. Both of these efforts provide a means for searching through collections of descriptions of software assets available for reuse to CAASD staff. The software descriptions also provide information on how to retrieve the software assets. Finally, using the anonymous ftp facility available within the CAASD Resource Discovery System, a large number of software assets have been made available for direct retrieval by CAASD staff.

The CAASD Reuse Team is currently evaluating the use of enhancements to both Gopher and WAIS as well as the use of World-Wide Web (WWW) technology. Gopher enhancements include serving documents in multiple types (for example, Microsoft Word, PostScript, and ASCII text). WAIS enhancements include the use of boolean operators, partial keywords, and literal phrases in WAIS index searches. WWW technology merges hypertext and network information discovery and retrieval technologies in a very powerful paradigm.

All of the technology and software used to build the CAASD Resource Discovery System is freely available from the Internet community. Many information resources related to these technologies are also available in the form of network news groups and mailing lists. For more information, please consult the following sources.

The Internet Gopher. Software is available via anonymous ftp to *boombox.micro.umn.edu* in directory */pub/gopher*.

Network news is discussed in *comp.infosystems.gopher*.

WAIS. Software is available via anonymous ftp to *ftp.cnidr.org* in directory */pub/NIDR.tools*.

Network news is discussed in *comp.infosystems.wais*.

WWW. Software is available via anonymous ftp to *ftp.ncsa.uiuc.edu* in directory */Mosaic*.

Network news is discussed in *comp.infosystems.www*.

3 Comparison

An asset library mechanism must make available a set of services for both the library staff and the users of the library. Both the North Atlantic Treaty Organization (NATO) and the United States (U.S.) Department of Defense (DoD) have put together documents specifying the recommended services to be provided by asset library mechanisms. This section presents the recommended services and discusses how the CAASD Resource Discovery System performs these services.

3.1 NATO

NATO has issued a three volume set of standards to address the issue of software reuse.

Volume 1 - The Standard for the Development of Reusable Software Components is designed to provide guidance in the creation of software products with maximum potential for software reuse.

Volume 2 - The Standard for Management of a Reusable Software Component (RSC) Library provides guidance in the establishment and operation of a NATO-controlled resource to support the reuse of software life-cycle products in NATO contracts. This volume also includes a section on library tool management that discusses tool requirements of both the library staff and the library user. [4]

Volume 3 - The Standard for Software Reuse Procedures is designed to provide guidance for software project teams who wish to practice reuse by making significant use of reusable software components available in the NATO Reuse Library.

3.1.1 Library Staff Tool Requirements

Support the cataloguing of RSCs. WAIS provides a mechanism to perform full-text indexing on asset descriptions and keyword searching on these asset descriptions. Each asset description exists as an individual file, and these files may be bulk-loaded into the library system. Gopher provides a mechanism to allow a hierarchy of libraries to be created so that domain-specific assets may be classified as such.

Support the maintenance of the classification mechanism. A classification system is not formally used. Instead, the full-text searching mechanism of WAIS is utilized.

Generate transaction and status reports. Both Gopher and WAIS maintain logs of accesses and queries. These logs may be processed to determine asset and user statistics.

Support configuration management of the RSCs and of the classification vocabulary. Gopher and WAIS do not directly address the issue of configuration management. The Gopher server administrator is responsible for configuration management of the RSCs.

Support problem-report tracking. Gopher and WAIS do not directly address the issue of problem-report tracking. Problem reports are tracked by library staff.

Support user and project tracking. Gopher and WAIS do not directly address the issue of user and project tracking, except by the logging mechanism. Tracking of users and projects is performed by library staff.

Include adequate tool documentation. Gopher allows for the inclusion of help files that a user may browse in the same directory as the search items and other tools.

3.1.2 Library User Tool Requirements

Provide reusers an effective RSC search and retrieval system. Gopher and WAIS provide an excellent search and retrieval system accessible from a wide variety of computer platforms. The response to a keyword search will be a custom directory showing asset titles which match the search parameters. If a title is selected, the full text of the asset description will be made available to the user.

Furnish documentation explaining the features and use of the tools. As mentioned above, Gopher allows documentation to be provided on-line in help files, where desired. The library staff also provide paper copies of documentation to users.

3.2 U.S. DoD

Another document which identifies and categorizes a set of asset library services is the *Asset Library Open Architecture Framework (ALOAF) - Version 1.2*, developed under the Software Technology for Adaptable, Reliable Systems (STARS) program [5]. Sponsored by the U.S. Defense Advanced Research Projects Agency (DARPA), the STARS program is supported by the military services, SEI, and MITRE, with the U.S. Air Force as the executive contracting agent. The STARS Asset Library Framework (ALF) Service Model defines the categories of services that individual library frameworks make available.

The ALF service categories and how each is addressed by the tools used in the CAASD Resource Discovery System are discussed below.

Session Services. The Gopher protocol manages connections to the asset library, provides other connection services through FTP and TELNET protocols, and logs accesses to the asset library.

Library Management Services. Gopher provides services for managing and manipulating an asset library. The asset library maps into a UNIX file hierarchy, with provisions for providing symbolic names to the files.

Data Model Services. Gopher and WAIS do not directly address data model services. Data models are managed and maintained by library staff.

Asset description services. Asset descriptions are managed and manipulated outside of the Gopher and WAIS tools.

Query Services. Gopher and WAIS provide a keyword query service on the full text of asset descriptions. Results are returned and are sorted by asset relevance as determined by the WAIS scoring algorithms.

Asset Processing Services. Asset processing varies on a case-by-case basis. Some assets may be available through the anonymous ftp mechanism available through Gopher. Other assets may require contacting the internal point of contact mentioned in the asset description.

Metrics Services. Gopher and WAIS provide logging mechanisms to allow asset library administrators to collect statistics on library usage and asset usage, as well as unsuccessful library queries.

Access Control Services. Gopher and WAIS provide access control on a machine-by-machine basis. User level authentication has been prototyped but is not a part of the standard Gopher software distribution. Currently, sensitive information is not made available through the CAASD Resource Discovery System.

References

[1] UGT - UofMN Gopher Team, *Gopher (comp.infosystems.gopher) Frequently Asked Questions (FAQ).* University of Minnesota, May 1993.

[2] A. Edguer, *comp.infosystems.wais Frequently asked Questions [FAQ] (with answers).* Case Western Reserve University, May 1993.

[3] E. Krol, *The Whole Internet - User's Guide and Catalog.* O'Reilly and Associates, Inc., 1992.

[4] NATO, *NATO Standard for Management of a Reusable Software Component Library.* NATO Communications and Information Systems Agency.

[5] ALOAF, *Asset Library Open Architecture Framework - Version 1.2.* No. STARS-TC-04041/001/02, Software Technology for Adaptable, Reliable Systems (STARS), August 1992.

4 Biographies

Timothy Stockwell is a Member of the Technical Staff at The MITRE Corporation, McLean, Virginia. He is a lead researcher on information asset reuse within the Center for Advanced Aviation System Development (CAASD) and is also working with MITRE Corporate Informations Systems in exploring the use of network information discovery and retrieval tools for searching and accessing corporate databases. Mr. Stockwell was previously involved with the design and implementation of oceanic air traffic control prototypes and simulation models. He received a B.S. in Computer and Systems Engineering from Rensselaer Polytechnic Institute in 1988.

Mark Krause is a Member of the Technical Staff at The MITRE Corporation, McLean, Virginia. Mr. Krause is a lead researcher specializing in network information discovery and retrieval technologies and software and information reuse in support of the CAASD Reuse Initiative and is also providing input at the corporate level on information infrastructure re-engineering. He was previously involved in the design and implementation of en-route air traffic control systems. Mr. Krause received a B.S. in Computer Science from George Mason University in 1990.