
Combining CORBA and the World-Wide Web in the Stanford Digital Library Project

Andreas Paepcke

Scott Hassan

Stanford University , Gates Computer Science 426A, Stanford, CA. 94305
paepcke@cs.stanford.edu
hassan@cs.stanford.edu

Since the Fall of 1994, the Stanford Digital Library Project has been developing what we call our InfoBus. This is a design and implementation which allows interface builders and the designers of automated on-line service clients to access publication-related services on the Web and on other protocol domains through a convenient, method-based interface.

The basic approach is to construct Library Service Proxies which are CORBA objects representing on-line services. At the back-end, these proxies communicate with the services via the service-native communication channels. At the front, they provide method interfaces.

For example, we have constructed proxies for Web-based repositories, such as Lycos, WebCrawler, and Alta Vista. We also have proxies for Web services that are not repositories, such as ConText, which is Oracle's document summarization tool. Another such service we interface to is InterBib which is our own BibTex/Refer to HTML/Framemaker format converter and document bibliography construction tool.

Other proxies access services that are not on the Web, although users of the proxies cannot tell the difference. Knight-Ridder's Dialog Information Service, and Stanford's on-line catalog are examples of systems available via Telnet. Our proxies conduct the necessary sessions for us.

For communication with proxies to information repositories, we have developed a special, experimental protocol, the Digital Library Interoperability Protocol. It allows our CORBA clients to interact efficiently with the library service proxies. The protocol borrows some concepts from Z39.50, but it is specially designed to take advantage of distributed object technology.

For example, it allows relocation of both data and computational resources while an interaction between client and server is in progress. Clients and servers are able to negotiate whether they want to interact synchronously, or whether they prefer fully asynchronous operation. We use asynchronous communcation when interacting with the Web, because this allows us to smoothly stream information from its source to the client whenever new data becomes available.

Our implementation platform is ILU, a free implementation of a CORBA superset developed at XEROX PARC. Our research partners at the University of Michigan and UC Santa Barbara are using the same approach, and we access each others' Web-based and non-Web services through the interop protocol under CORBA.(see " Towards Interoperability in Digital Libraries---Overview and Selected Highlights of the Stanford Digital Library Project", as well as "A Distributed Interface for the Digital Library" for more details)

Our plans for the coming few months include the use of Java to distribute our proxy client and server objects. Once the code arrives, it begins to use CORBA method calls for further operation (subject to security limitations). This will have the advantage that clients of our InfoBus will not need to pre-install CORBA software and will instead be able to use the Web for code distribution. At the same time, once the code is transferred via the Web, the more easily programmable and maintainable method call semantics and advantages of implementation- and interface inheritance can be exploited.
