Z39.50 and the World Wide Web

Contributed by:

Sebastian Hammer
Index Data
Copenhagen, Denmark
quinn@index.ping.dk

John Favaro
Intecs Sistemi
Intecs Sistemi
Pisa, Italy
favaro@pisa.intecs.it

The tremendous success of the World Wide Web (WWW) and the increasing use of WWW front-ends to library catalogs and other information systems have caused decision-makers to question whether the investments required to establish additional Z39.50 services are still warranted. Meanwhile, the increased versatility of the 1995-version of the Z39.50 protocol -- which enables it to provide powerful services outside of the strictly bibliographic application domain -- leads information specialists to wonder where the WWW and Z39.50 fit together in the evolving information infrastructure.

The Web as a Simple Networked Access System

Using the forms-based interface of the World Wide Web (WWW) in conjunction with graphical clients or browsers such as Mosaic or Netscape has become an inexpensive and popular method of providing user-friendly access to on-line catalog systems. The tools required to publish information on the WWW in this fashion are inexpensive or even free, and are generally straightforward to use. The results are rewarding: It is a remarkably simple task to produce attractive, graphical interfaces which have similar appearances across many different desktop platforms. No specialized software beyond a normal WWW "browser" is required on the client side, and facilities for File Transfer (FTP) and simple searching (WAIS, etc.) are well-integrated into the WWW suite of protocols.

However, there are serious tradeoffs involved when using this approach. The individual WWW client has no knowledge of the application domain in which it operates. It receives a stream of graphical user interface primitives (such as buttons, text-entry fields, and formatted response data) from the server, and naively displays these to the user. The WWW inherits a problem that has haunted users since the first information systems went on-line using simple character terminals: no two information systems share the same interface characteristics. Each new system requires the user to master a new interface structure, and, with the advent of graphical interfaces such as the WWW, a new set of custom-designed icons and symbols.

Information systems often support the notion of a search "session", in which the results of previous queries can be re-used or refined. The HTTP (HyperText Transfer Protocol), which is at the core of the WWW, is inherently stateless: Numerous problems arise when the interface is adapted to host systems that have a notion of a continuous session between the client (user) and the server. There are currently efforts underway to add state-managing mechanisms to the underlying protocols, but the basic paradigm remains essentially a stateless one, which fits poorly with the session-oriented interfaces to most on-line information systems.

Searching on the Internet - the Role of Z39.50

The Web is an ideal vehicle for organizations that are "vertically integrated," that is, which are owners of content that they can present to the user in a structure of their own choosing. That is why many media and entertainment companies are showing a great interest in the Web today. But when users must actively search the Web for information across organizations, they encounter a sea of largely unstructured data.

The library community has much to offer in the way of providing structure to information resources on the Internet. The Z39.50 standard is a concrete representation of this fact. Currently, the search engines and indices of Web resources suffer from the same weaknesses as the interfaces to library systems. No two are alike, and there is no way to make structured use of the data that they return. With the current growth of the Web, the search engines are becoming increasingly important - a significant portion of the Web community now spends more time looking at search engine output than on any other type of Web page. However, it may eventually become impossible for any one organization to index it all in a useful way. We will need more well-structured access methods to allow searching across multiple indices. Here the power of Z39.50 as a true, mature information retrieval protocol becomes evident.

The Z39.50 standard specifies an abstract information system with a rich set of facilities for searching, retrieving records, browsing term lists, etc. At the server side, this abstract system is mapped onto the interface of whatever specific database management system is being used. The communication taking place between the server and the client application is precisely defined. The client application is unaware of the implementation details of the software hiding behind the network interface, and it can access any type of database through the same, well-defined network protocol. On the client side, the abstract information system is mapped back onto an interface which can be tailored to the unique requirements of each user: a high-school student may require a simple, graphical interface with limited functionality, while an information specialist may need a complex, highly configurable information retrieval engine. Finally, casual users may prefer an interface which blends in smoothly with their word processor, database software, or, indeed, WWW browser.

In summary, the essential power of Z39.50 is that it allows diverse information resources to look and act the same to the individual user. At the same time, it allows each information system to assume a different interface for every user, perfectly suited to his or her particular needs.

Navigation Between Resources - the Strength of the Web

Z39.50 was born as a point-to-point, client/server mechanism. It provides very powerful

means of locating records within one or more databases on a single server. The problem that remains is that of navigation between servers or information resources:.

- How do we find the server and the database that has the information we are looking for?
- How do we learn about the contents of a server?

For learning about new servers or information providers, the Explain facility of Z39.50 is an important resource. Explain provides a structured mechanism for the information provider to publish information about the capabilities of the server software, and about the characteristics of the information stored in each database on the server. The rich set of information elements defined by the Explain facility includes contact information for the host instutution, as well as specifications of the available access points (indices) for searching. The rigid structuring of the information allows the client software to automatically configure itself and adapt to each server system, while the uniform interface to the descriptive information about the database helps the user quickly orient himself to the contents of a new information resource.

The truly difficult issue, however, is establishing an infrastructure between servers that allow users to locate the right servers for their purpose in an easy way. The Z39.50 URLs are useful in this respect, because they make Z39.50 servers appear to be "just another kind of document" in the Internet space. People can collect and categorize collections of servers the same way they do other kinds of documents or information resources. WWW search engines can even be used to discover new Z39.50 servers.

Our preferred approach would be to use Z39.50 itself to find Z39.50 servers. That is what locator services can do. GILS defines an application profile for Z39.50 that is useful for locating information resources (although admittedly, GILS is optimized for US government documents, and as such it is probably less than ideal for some other purposes). These documents can be anything - from books to reports to archives of photographs to on-line databases to WWW-documents (and since a WWW-document can be a Z39.50 server, the locator service can be used for exactly the purpose we have in mind).

With a slightly simpler and more general profile than GILS, Z39.50 could become a very powerful tool for accessing indices of information resources. In effect, we are postulating that we replace or supplement all of the existing WWW-crawlers with Z39.50 servers. In that way, we would be able to access all of the different indices with a uniform interface, and because the access structure is fully standardized, it would be simple to gateway or replicate information between servers - we would potentially only need a single starting point to search for any kind of information anywhere in the world. Indeed, this is an important part of the vision behind the Global Information Locator Service currently being investigated by the G-7 Group of industrial nations.

Again, static documents containing Z39.50 URLs will provide an increasingly important means of discovering and accessing information resources, as WWW- browsers with Z39.50 client-capabilities become commonplace. When these documents are, themselves, served or located by Z39.50-aware systems, the circle is complete.

In summary, we believe that there is a strong potential for a profitable and synergetic relationship between the WWW and Z39.50. We see the two worlds merging together, with each one growing stronger by using the best elements of the other: Hyperlinks between systems and document types from the WWW - structured searching and

document discovery from Z39.50.

Comments and discussion is invited - please send email to Sebastian Hammer at quinn@index.ping.dk.

Copyright © 1996 Sebastian Hammer, John Favaro

---

hdl://cnri.dlib/march96-hammer