WEB BASED INFORMATION SYSTEMS AND COMPATIBILITIES OF Z 39.50

Devendra. S. Gobbur

Abstract

To enable Interoperability in web based Information Systems there is a need of unique communication protocol. In this respect, Z39.50 playing an important role in Interoperability of the Information Systems. It does not determine how systems will be built, how they will present information to the user and so on. Z39.50 systems have been put together in many ways to suit different needs. It has been used for the compatibility of exchange of data and can be incorporated into all sorts of products and systems only a few of which are currently being exploited. An attempt has been made to provide complete overview of Z39.50 compatibilities for smooth functioning of web based information systems. Further it also provides web standards and search interface support for combination of search terms from different search attribute sets.

Keywords: Information Systems, Web Compatibility, Internet standards, Z39.50

1.   Introduction

Information Systems and Networks play a major role in disseminating and communication of information all over the world. The Information Systems collect, stores and communicate the information in digital form through the major Networks. Among the Networks, Internet is most powerful tool to disseminate and communicate the information to any part of the world. But, in practice there was problem of different methods and structure of databases in the Information Systems. For example, a few Information systems adopted MARC-I, a few adopted MARC-II and a few adopted CDS/ISIS, to store information. Under such circumstances, it was difficult to access the information from different sources (systems). Hence there was a need to develop a Standard Communication Protocol, which must be compatible to all kinds of web environments. In this respect, Z39.50 was developed as Standard web Communication Protocol.
2 Z39.50: Standard Communication Protocol

Z39.50 is an American national standard for information retrieval. It specified a set of rules and procedures for the behavior of two systems communicating for the purposes of database searching and information retrieval.[1] As a network application standard, Z39.50 is an open standard that enables communication between systems that run on different hardware and use different software.

The Z39.50 standard was developed to overcome the problems associated with multiple database searching such as having to know the unique menus, command language, and search procedures of each system accessed. Z39.50 simplifies the search process by making it possible for a searcher to use the familiar user interface of the local system to search both the local library catalogue as well as any remote database system that support the standard.

The latest edition of Z39.50 was approved in 1995 by the National Information Standards Organization (NISO), the only organization accredited by the American National Standards Institute (ANSI) to approve and maintain standards for information services, libraries and publishers.

Z39.50-International: Next Generation covers a number of initiatives by Z39.50 implementers to make the intellectual/semantic content of Z39.50 more broadly available and to make Z39.50 more attractive to information providers, developers, vendors, and users, by lowering the barriers to implementation while preserving the existing intellectual contributions of Z39.50 that have accumulated over nearly 20 years.[2]

Z39.50 is the formal designation of an international standard protocol that facilitates communication between a local software client and a remote information retrieval system. During the 1960s and early 1970s, there emerged a vision among library leaders in the United States to create a national bibliographic network, in order to share resources electronically, and especially cataloging records standardized in the MARC format.

William Moen (1995)[3] offers the following description of Z39.50: Z39.50 is a computer-to-computer communications protocol designed to support searching and retrieval of information — full-text documents, bibliographic data, images, multimedia — in a distributed network environment. Based on client/server architecture and operating over the Internet, the Z39.50 protocol is supporting an increasing number of applications. And like the dynamic network environment in which it is used, the standard is evolving to meet the changing needs of information creators, providers, and users.

3.   Use of Z39.50 in interoperable Web based Information Systems:

It is important to note that Z39.50 provides for an articulated sequence of steps in a search session between two web based Information Systems, known, respectively as the origin and the target, or what we commonly think of as the client and the server. The basic sequence of these interactions is as follows:

- Init: In this phase of the session, the two Information Systems establish baseline parameters for intercommunication, such as determining the maximum file size that will be returned.

- Search: Here the query is formulated and sent to the target system.

- Present: Results are returned to the client and displayed to the user.

There are many services that can be facilitated by Z39.50. For example, complex Boolean searching, keyword searching, proximity searching, truncation or focusing on a particular field or subfield of a formatted record (e.g., MARC21 or GILS). It is vital to maintain a separation in one's thinking between Z39.50 proper and any particular Information System. Z39.50 by itself is a strictly a communications protocol. It can enable intercommunication between, say, a stand-alone software client and a remote database (or databases) returning the results in any number of formats. The target could be a library catalog (i.e., an OPAC), a database containing cultural heritage information in a museum setting, a collection of CD-ROM bibliographic citation databases, or even a collection of Web pages incorporating XML tags. What Z39.50 most essentially provides is a layer of abstraction between any two Information Systems and a common language for interoperability. A Z39.50 query does not directly address the target database or databases. It talks to a Z39.50 server, which, in turn, translates the query, results and so forth, into the language of the Z39.50 protocol.

In addition, to what has already been mentioned, Z39.50 can provide for:

- Authentication: Searches can be restricted to authorized users only.

- Explain: This step provides the user a way of asking the target system to transmit human readable information about its particular capabilities.

- Definition of Record Formats: Records can be returned in MARC, GILS, SUTRS, etc.

- Index Browsing: Facilitates browsing within a controlled index of subject terms.

A full implementation of the Z39.50 protocol also permits the user to save a search session so that it can be run again at a later time, or to save a result set for later use. Result sets may be sorted. Databases may be updated, charges can be assessed and sessions can be closed. However, not every Z39.50 server will support all of these services.

A key issue in relation to Z39.50 is interoperability. There are challenges that must be overcome whenever one attempts to search a remote Web based Information System using something other than its native search interface. For example, when the user launches a search against a remote database using the term author, he or she may not be aware of how variable the result can be. Precision depends upon the manner in which the Z39.50 server has been mapped to different parts of the underlying record structure.

This document identifies a subset of specifications from the Z39.50 Information Retrieval Protocol (ANSI/ NISO Z39.50/ISO 23950) for use in Z39.50 client and server software. Conformance to this profile's specifications will improve international and national search and retrieval among library catalogues, union catalogues, and other electronic resource discovery services worldwide. The profile will evolve as the environment and the standard change, and is intended to facilitate global resource sharing. However, a primary goal of the Bath Group was to publish the first release of the profile as quickly as possible rather than waiting for the complete definition of all specifications before the initial release.[4]

This profile builds upon the experience of other profiling efforts addressing interoperability for library catalogue searching and the development of virtual union catalogues.

National libraries are investing a lot of money to authority control. The costs can be reduced by copying of authority data between countries, or even from the national bibliography to other library databases. Since national libraries have an urgent need for being able to exchanging authority records via Z39.50, it is our wish to extend the Bath profile so that MARC authority records are added among the supported record formats into Level 1.

3.1 Support for combination of search terms from different search attribute sets

Z39.50 standard contains so called attribute sets, which define search terms that can be used. It is possible to define basically unlimited number of attribute sets, and it is easy to add more search terms into an attribute set.

For basic bibliographic searching search terms are defined in attribute set Bib-1. There are 15 other attribute sets at the moment (http://lcweb.loc.gov/z3950/agency/defns/oids.html#3), of which two are national (Dan-1 for Denmark and Fin-1 for Finland).

The reason for defining national attribute sets is that each country has a number of purely domestic search terms (e.g. Finnish subject headings list, Finnish public libraries classification). Adding all of these into Bib-1 would make Bib-1 very complex and difficult to understand.

Terminology issues are ever-present in a document such as this. An example of such an issue is with terms such as "access points," "indexes," "fields," and "data elements." Often, the library community uses the phrase "search a particular field or fields," when at the system level, the search may be executed by matching the search term with entries in a system-generated index. Access points can be considered searchable fields of a record as represented by the index created from data from those fields. For Cross-Domain searching, the concept of "field" may be completely absent. In defining searches for library catalogs, the description references fields and indexes. In defining cross-domain searches, the description references data elements and indexes. [5]

4. Functional Requirements

This section identifies the functional requirements informing the Z39.50 specifications in this release. These requirements focus on search and retrieval between Information Systems Index, the search and retrieval of bibliographic holdings information, and cross-domain search and retrieval for resource discovery. The requirements detailed in the sections below comprise three Functional Areas:

- Functional Area A for Basic Bibliographic Search & Retrieval, with Primary Focus on Library Catalogues

- Functional Area B for Bibliographic Holdings Search & Retrieval

- Functional Area C for Cross-Domain Search & Retrieval.

4.1   Basic Bibliographic Search and Retrieval

Users of Information Systems conduct a variety of search and retrieval transactions. The functional requirements for basic bibliographic search and retrieval delineate a limited number of core searches, the browsing of indexes, and the appropriate retrieval mechanisms needed by library users when interacting with library catalogues and other electronic resources discovery services.

The users of Information System engage in a wide range of searching behaviors. Agreements on a core set of bibliographic searches have evolved through various Z39.50 profiling efforts:

- Author searches that include searching for an established name heading, searching for names not under any authority control, and searching where only part of a name is used as a search term

- Title searches that include searching for the entire title, the first part of a title, and searching using one or more words from a title

- Subject searches that include searching for a complete subject heading, the first part of a subject heading, and searching using one or more subject words as search terms

- Keyword searches that include high recall searches using one or more words from author, title, subject, and other common access points

- Boolean searches that include combining search terms with the Boolean operators of AND, OR, NOT

- Truncation searches where the user wants the system to truncate within the final word of an expression.

4.2   Browsing Indexes

Information users often use a browse function on a local system to identify appropriate search terms to use in a query. Browse-based searching can assist users in improving their selection of search terms for the query. This is a useful feature, often used in conjunction with known-item searching. Such browse-based searching can be seen as an alternative searching strategy for some search requirements listed in Browsing indexes is a requirement for Z39.50 implementations and can be achieved through the Z39.50 Scan service, and this requirement is addressed in Level 1.

4.3   Bibliographic Holdings Search and Retrieval

Identifying which collections contain certain information resources is a key factor in creating an infrastructure for resource sharing among libraries, Information Centres and Information Systems. Librarians and Information Scientists require holdings information for resource sharing, and library users need holdings information for knowing where to go to use or borrow a resource. The holdings information should provide sufficient identification and description of an item to give the user adequate information to make a decision about requesting or retrieving the item. Search and retrieval of bibliographic holdings information require the use of the Holdings Attribute Set, the Generic Record Syntax (GRS-1), and the recently completed Holdings Schema.

4.4   Cross-Domain Search and Retrieval

The Information users all over the world desire integrated access to distributed resources, often in conjunction with resource discovery where searches are across many types of information resources. There is a requirement for effective cross-domain searching of diverse resources including library catalogues, government information, museum systems, and archives. A user may wish to send a single search to one or more these resources[6].

5.   Conclusion

It is easy to become enthusiastic about the capabilities of Z39.50 and the range of benefits it can bring in terms of library cooperation, Information Sharing, reader services and systems integration. When building a Z39.50 Z-client, the designer has to decide which version and which features to implement. The standard defines many facilities and it is not necessary to implement all of them. However the exact nature of the "discussion" is not usually relayed to the user interface and assumptions may be made about what can be achieved and what cannot. When applied to a remote Z-server, an author search may be personal authors only. Combinations of such disparities and differences both in the version of the standard and in extended services supported are common.

Z39.50 software allows computer-to-computer information retrieval. Z39.50 makes it possible for a user in one information system to search and retrieve information from other information systems that have also implemented Z39.50. It is not necessary to know the search syntax that is used by those other systems. In this way, it will provide Interoperability between different Web based Information Systems.

6.   Reference

1.   National Information Standards Organization. (1995). ANSI/NISO Z3950-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Bethesda, MD: NISO Press.

2.   Turner, Fay.(1998). " Selecting a Z39.50 client or web gateway." Library Hi-Tech, Issue 62, Vol. 16, No. 2: 7-18.

3.   Moen, William E. (1999). Handling Bibliographic Holdings Information in Z39.50: Implications for the Z Texas Profile..

4.   Husby, Ole. (1997, January 9). BIB-1 Profile for ONE www.bibsys.no/one-wg/bib-1.profile.html

5.   Conference of European National Librarians. (1997, October 15). Z39.50 bib-1 Attribute Set Profile for CENL, Version 1.1. www.lib.helsinki.fi/z3950/cenl_profile.html accessed on 01st Dec 2005.

6.   Lunau, Carrol D., "Z39.50: a critical component of the Canadian resource sharing infrastructure; implementation activities and results achieved." New Review of Information Networking, 3, 1997, pp.77-92

About Author

Shri.Devendra.S.Gobbur, as Assistant Librarian Gulbarga University, Gulbarga. He has over 7 years of professional experience of Library and Information Science. He has contributed more than 4 papers in Seminars, Conference, and Journals. His area of interest are Internet, Citation Analysis and Bibliometrics etc.