Why the Bath Profile Makes Z39.50 Work

by PETER GETHIN

INTRODUCTION

The Z39.50 standard has been in productive use for several years now. However, due to the way the standard has been interpreted by some Integrated Library System (ILS) vendors, the results of using the standard have often been somewhat varied at best and, at worst, non functional. The lack of consistency in the use of Z39.50 has lead to the standard being held in rather lower regard than it deserves in some quarters.

It is a matter that can be debated whether the inconsistency in the implementation of the standard is the fault of the implementers ignoring aspects of the standard to make their task simpler or simply a lack of precision in the standard itself. To circumvent the debate, the Z39.50 Implementers Group (ZIG) has defined profiles for the use of the standard that are clearly defined and can be measured.

This paper takes a look back at why the Z39.50 standard was invented in the first place and defines briefly how Z39.50 works. The paper then considers why implementations of the standard appear to fail and how those failures manifest themselves. With this information in hand, it is possible to make sense of the answer to the question, how does the Bath profile make Z39.50 work?

WHY DO WE HAVE Z39.50?

Z39.50 facilitates resource discovery by providing broadcast searching across a variety of information resources, such as catalogues, museums, citation and full-text databases. It supports resource sharing for Inter Library Loans, document delivery and copy cataloguing. It is also the most powerful resource searching option currently available, at least until an Internet query language

which accommodates the rich semantics of Z39.50 is developed and widely used.

A user enters a query in a user interface and the client (user interface) translates the query to the Z39.50 standard and sends it to a server, which in turn translates the query into the server's native query language. The response is then encoded by the server and decoded by the client for presentation to the user.

The Z39.50 standard provides a uniform way to access all types of data. The fact that a particular client or server supports the Z39.50 standard, however, does not automatically mean that you can use it to connect seamlessly to a variety of information sources. Differences in implementation of the standard and differences in local information retrieval systems such as search functionality and indexing policies and decisions are threats to interoperability between client and server. Both of these threats must be addressed with explicit Z39.50 specifications and configuration as well as recommendations for local indexing decisions.

WHAT DID WE HAVE BEFORE Z39.50?

Prior to the definition of Z39.50, every different ILS and Information Retrieval (IR) System had its own user interface. To be able to interrogate different ILS's and IR Systems, the user had to learn and become proficient in the syntax and techniques of each of the different systems being used. Not only did this make end user research impractical, it made automated machine to machine inquiries very difficult indeed. It was possible to programme systems so that they could retrieve information from other, disparate systems, but the algorithms were often so complex and expensive to develop and maintain that very few ILS vendors attempted to take them on.

There was an attempt in the early 1980's at defining a Common Command Language (CCL) for IR Systems. In practical terms, however, this definition was pretty much the lowest common denominator between the available IR systems on the market and in use at the time. The major IR System vendors did not have a vested interest in promoting or supporting CCL because it failed to capitalize on the more advanced features that differentiated the more advanced IR Systems from one another. With the possible exception of some Scandinavian users, CCL never attracted much support, and soon fell into disuse.

HOW Z39.50 WORKS

Z39.50 overcomes the problems of being the lowest common denominator that was experience with CCL because Z39.50 is not a command syntax. A Z39.50 query consists of a search term and six Attribute Types.

The Bib-1 Attribute Set is a standard Z39.50 search definition tool which is widely used to express Z39.50 queries against library catalogues. The six Attribute Types, are each designated by a name and number.

The Use (1) or Index attribute specifies type of search, such as author, title, subject, etc.

The Relation (2) attribute controls the order of numerical expressions, like date ranges.

The Position (3) attribute specifies the location of terms, such as first in field (where field is the indexed term, not the MARC tag) or anywhere in the field.

The Structure (4) attribute specifies whether a single word, a phrase, or a word list is searched.

The Truncation (5) attribute specifies whether or not a partial search word can be entered, ie whether or not to use right hand truncation.

The Completeness (6) attribute specifies whether or not the search must contain the entire, exact field, or part of the field.

Each attribute type can take on values (also designated by name and number value). For example, a Use attribute characterizes the access point that should be searched. One Use attribute value is „Title" or „4" to designate a title access point. Attribute types and values are expressed as integer pairs; the pair (1,4) tells the server to execute a title search. The combination of attribute types and values provides a way to express the semantic intention of the search and prescribe the behavior expected when the server executes the query.

Examples:

A search for the exact title Nature, utilizing all six attributes, is represented in the following way.

(1,4) (2,3) (3,1) (4,1) (5,100) (6,3) Nature

Term = „nature"

Bib-1 Use Attribute (1) = 4 (title)

Relation Attribute (2) = 3 (equal)

Position Attribute (3) = 1 (first in field)

Structure Attribute (4) = 1 (phrase)

Truncation Attribute (5) = 100 (do not truncate)

Completeness Attribute (6) = 3 (complete field)

A keyword author search for Twain, utilizing all six attributes, is represented in the following way.

(1,1003) (2,3) (3,3) (4,2) (5,100) (6,1) Twain

This representation would be interpreted like this.

Term = „twain"

Bib-1 Use Attribute (1) = author (1003)

Relation Attribute (2) = equal (3)

Position Attribute (3) = any position in field (3)

Structure Attribute (4) = word (2)

Truncation Attribute (5) = do not truncate (100)

Completeness Attribute (6) = incomplete sub field (1).

WHY Z39.50 SOMETIMES „FAILS"

Problems have crept into the use of Z39.50 because of the interpretation of the use of default values for some, or all of the attributes. The standard is clear that the client software does not have to specify all of the attributes and if the server receives a Z39.50 request that does not have all of the attributes

specified it may apply default values, at the discretion of the person writing the software. This discretion has been interpreted somewhat liberally by some, if not all, implementers of Z39.50 servers.

It was never the intention that all servers should be able to support all combinations of attributes. If the server receives a combination of attributes that it cannot satisfy, it is supposed to reject the query and return an error diagnostic.

There are some implementations of Z39.50 that are so basic that the combination of attributes that they are capable of supporting is so small, that almost all queries are rejected and return an error diagnostic. With those systems you only get a response if you submit one of the few combinations that are supported. These implementations are at least within the letter of the standard, albeit that they return no information most of the time.

The bigger problem is with systems that have taken a more liberal interpretation of the allowed use of defaults. A depressingly large number of implementations of Z39.50 replace attributes in the client query that they cannot satisfy with ones that they can.

A detailed study of the Z39.50 servers that are commercially available has revealed that the Z39.50 server of one vendor processes all structure attributes as „phrase", regardless of the client query. Another vendor's implementation always processes „incomplete field" as „complete field". A third vendor's implementation processes „do not truncate" as „right truncation". The worst problem encountered, however, is a vendor who ignores the bib-1 use attribute and always uses „any".

The effect that is seen by the end user varies from not finding what they were looking for, even when they know it exists on the target server to getting the same result, regardless of whether your query is a title search, an author search or a subject search.

It is little wonder then, that the use of Z39.50 has such a poor reputation in some places. The pity is that it is not the standard itself that is poor, it is the poor implementation of the standard by some of the major ILS vendors.

WHAT IS THE BATH PROFILE?

A realistic view of the problems surrounding Z39.50 is simply that there are several hundred possible combinations of the six attributes. Implementers have thrown their hands in the air in horror at the prospect of being able to accommodate all of them. The standard does not really specify which combinations of attributes are most likely to be used. That piece of crystal ball gazing had hitherto been left to the programmers. Recognizing that they could only manage to handle properly a very small subset of the possible combinations, the programmers resorted to choosing their own defaults rather than send an error diagnostic as the response to the majority of possible combinations.

The Bath profile relieves the problem by defining which combinations of attributes a Bath compliant server should be able to handle, and the nature of the response to each of those combinations.

The Z39.50 Implementers Group and others have developed a profile of minimum attributes required to support effective use of the Z39.50 standard for a variety of library functions, including cataloguing, inter library loans, reference, and acquisitions. This profile has been named the „Bath Profile", after the site of a 1999 conference in England. The Bath Profile is an Internationally Registered Profile (IRP) which is used as the core set of requirements for international Z39.50 resource sharing.

The Bath Profile is modular, so Z39.50 communities can implement all or only specific functions. Related functions and requirements are grouped by functional area concepts. Within each functional area, several conformance levels are defined. Level 0 compliance achieves high recall, or the most hits. Level 1 achieves a higher level of precision.

Functional Area A – Basic Bibliographic Search and Retrieval

This area typically applies to Library Catalogues.

19 searches defined for author, title, subject, any, date and standard identifier including keyword, keyword right truncated, exact, first words in field and first characters in field.

6 scan/browse defined for author, title, subject and any combined with phrase or keyword.

MARC21/UNIMARC and SUTRS/XML record syntaxes.

Functional Area B – Bibliographic Holdings Search and Retrieval

This area of the profile facilitates the search and retrieval of item and serials holdings information. Level 0 includes holdings capabilities as part of the MARC record. Level 1 specifies XML record syntax and holdings search.

Functional Area C – Cross-Domain Search and Retrieval

This area facilitates the simultaneous search of library catalogues, archives, museums and the internet using a subset of searches/functionality specified in functional Area A.

WHY THE BATH PROFILE MAKES Z39.50 WORK

The reason that the Bath Profile makes the standard Z39.50 work in practice is two-fold:

On the one hand, programmers will not now have to consider what they have to do to handle the „other" combinations. The majority of queries that are going to be encountered have been defined. If other combinations arrive and they cannot be handled, there is no disgrace in sending an error diagnostic to say that the particular attribute value is not supported by this server.

As well as taking the pressure off the programmers in terms of which attributes and combinations of attributes they are expected to handle, the fruits of their labors can be measured. A model system is being set up with a model database. The results of searches using the profile against the database is being tested „by hand" such that the response to queries from a Bath compliant client to a Bath compliant server is known. The „Z39.50 Inter-operability Testbed Project" is being set up at the School of Library and Information Sciences at the University of North Texas. The project is being organized by Dr Bill Moen and is funded by US Federal Institute of Museum and Library Services.

Vendors wishing to have the compliance of their Bath profile implementation independently verified will be able to do so. The guess work is being taken out of the situation. ILS Vendors will be able to hold up their Z39.50 server implementations and say, „this is independently certified as Bath compliant". End users will know that when they submit a query to a Bath compliant server

with a Bath compliant client they will be able to trust the response they get back.

NATIONAL AND REGIONAL PROFILES.

National and regional organisations are able to build on the international standard, as in the example of the „TZIG" project implemented by Texas State Library and other Texas libraries. The TZIG project also includes Recommendations for Indexing MARC 21 Records. Other projects include DanZIG, ONE-2, CENL and the U.S. National Profile being developed by NISO to establish national bibliographic and holdings requirements. The U.S. National Profile is scheduled for adoption by ballot late in 2001.

WHAT WE WILL HAVE WHEN Z39.50 IS WORKING PROPERLY

When ILS vendors have implemented the Bath profile, end users will reliably be able to discover resources by providing broadcast searching across a variety of information resources, such as catalogues, museums, citation and full-text databases.

That in turn will support resource sharing for Inter Library Loans, document delivery and copy cataloguing.

WHO IS ENDORSING THE BATH PROFILE?

At the time of writing, the Bath Profile is supported and endorsed by:

- Atlantic Scholarly Information Network
- CIC (11 Major US Universities)
- Conference of European National Librarians
- Texas Library Association
- International Coalition Of Library Consortia
- National Library of Canada
- Saskatchewan Provincial Library
- City of Ottawa Regional Libraries

- OCLC

- SIRSI Corporation.

WHAT LIBRARIES CAN DO TO MAKE SURE THE BATH PROFILE WORKS

History has demonstrated that ILS vendors do respond to the demands of the market. If libraries want to be able to enjoy the benefits of resource discovery and sharing, they simply need to put pressure on their ILS vendor to support the Bath Profile. It is already common practice in the USA that research libraries seeking to replace their existing ILS system are making Bath compliance a mandatory requirement in the Request for Proposals being sent to ILS vendors.

The hard work has been done. The standard has been defined and the ZIG have defined a working profile of attribute combinations. The benefits are available to Libraries if they want them.

WHAT MIGHT HAPPEN NEXT?

A group of Z39.50 implementors has been informally discussing ways to evolve Z39.50 to a more mainstream protocol, attractive to information providers, vendors, and users. These discussions began at the December 2001 ZIG meeting and have continued since. The group met June 29-30 to define specifications for a new web service definition based on Z39.50 together with web technologies: XML, URI, SOAP (RPC), and HTTP. The specification will be called ZNG, „Z39.50 Next Generation". (Earlier, it had been using the code-name ZML: „Z39.50 over XML".)

This is a proof-of-concept initiative whose basic goal is to develop a standard search and retrieve service enabling development of value-added applications such as the scholar's portal that will integrate access to various networked resources. More specifically, the goal is to lower the barriers to implementation while preserving the existing intellectual contributions of Z39.50 that have accumulated over nearly 20 years, discarding those aspects no longer useful or meaningful.

More information about ZNG can be found at: <http://www.loc.gov/z3950/agency/zng.html>.

Acknowledgements

REFERENCES

National Information Standards Organization. (1995). ANSI/NISO Z3950-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Bethesda, MD: NISO Press. Electronic version of Z39.50 available at the Z39.50 Maintenance Agency. Available: <http://lcweb.loc.gov/z3950/agency>.

Carrol Lunau, Paul Miller, and William E. Moen (Editors).The Bath Profile:An International Z39.50 Specification for Library Applications and Resource discovery. Release 1.1. Available: <http://www.nlc-bnc.ca/bath/bp-current.htm>.

Husby, Ole. (1997, January 9). BIB-1 profile for ONE Available: <http://www.bibsys.no/one-wg/bib-1.profile.html>.

William E. Moen. Resource Discovery Using Z39.50. Available: <http://www.unt.edu/wmoen/publications.htm>.

Slavko Manojlovich, Memorial University Of Newfoundland. Bath / Z Texas Profile Compliance Test Reports. Available: <http://nofish.library.mun.ca/bathtest/report.htm>.

Willaim E Moen. An Applied Research and Demonstration Project to Establish and Demonstrate a Z39.50 Interoperability Testbed. Available: <http://www.unt.edu/zinterop/>.
The Open Archives Initiative. XML using the MARC DTD. Available: <http://www.dlib.vt.edu/projects/OpenArchives/oa_marc.html>.

Danish Z39.50 Implementers Group. (1999, March 4). DanZIG Z39.50 Profile. Available: <http://www.bs.dk/danzig/profil.htm>.

Texas Z39.50 Implementors Group. (1999, April). Z Texas Profile: A Z39.50 Profile for Library Systems Applications in Texas, Release 1.0. Available: <http://www.tsl.state.tx.us/ld/projects/z3950/TZIGProfile99Apr20.htm>.