

PEAK RETRIEVAL, LLC

Information Retrieval Specialists

Z39.50 Protocol

ANSI/NISO Specification

PEAK RETRIEVAL, LLC

Z39.50 Protocol

© Peak Retrieval, LLC
Author: Ron Hardman
2815 Freewood Pt • Suite 7
Phone 720.339.5807

If you are new to the Z39.50 Specification, you may want to consider familiarizing yourself with key terms and definitions. You will find a complete list of terms at <http://lcweb.loc.gov/z3950/agency/markup/02.html>.

Table Of Contents

<i>History</i> _____	3
<i>Overview</i> _____	3
<i>Information Retrieval Service</i> _____	4
Information Retrieval Facilities _____	5
<i>Z39.50-International: Next Generation (ZING)</i> _____	7
<i>Final Note</i> _____	9

History

How does an organization connect to a large data source for the purpose of retrieving information? How about connecting to 100 sources of data through a single interface and reliably retrieve information from all of them. This problem is what the American National Standards Institute (ANSI), through the National Information Standards Organization (NISO), sought to resolve through the Z39.50 protocol.

A Standards Committee was formed by NISO in 1979, in 1984 the protocol was proposed, and in 1988 the first version of the Z39.50 protocol was approved called Z39.50-1988. Z39.50 came into compliance with the international standards ISO 10162 and ISO 10163-1 with the second revision, referred to as Z39.50-1992. In 1995, the international development community approved and adopted version 3 of the Z39.50 protocol, called Z39.50-1995, as the international standard. You will sometimes see references to Z39.50-1994 as version 3, but this is not the approved version. They were earlier drafts of Z39.50-1995 and are not complete. Z39.50-200x is in draft form, and has been recommended by NISO. At the time of the publication of this white paper the new version had yet to be approved.

The original Standards Committee, commissioned by NISO to develop Z39.50-1988, was disbanded after the first version was approved. The following year, the Z39.50 Maintenance Agency was created to carry on this task, and is managed by the Library of Congress. The success of version one, and adoption by the development community, brought about the formation of the Z39.50 Implementers Group (ZIG).

Through ZIG meetings, enhancements are proposed, refined, and passed to the Z39.50 Maintenance Agency for inclusion in the next version of the specification. Membership to ZIG is open to all parties. Enhancements continue to be proposed, although no major version has been approved since Z39.50-1995

Overview

The Z39.50-1995 protocol, referred to as version 3, specifies standards for information search and retrieval. The standards allow client environments to connect to multiple data sources transparently. It separates the client environment from the data sources being searched, and provides a standard for the communication between them.

Z39.50-1992, referred through the remainder of this text as version 2, is still in use by a number of implementers, and is fully supported by version 3 of the protocol. In fact, version 3 is a superset of version 2, and adds enhancements to the existing specification.

At its most basic, versions 2 and 3 consist of three types of Facilities:

Initialization Facility – Defines the initial series of messages between the client and server(s) that create the connection, and establish activity limits on the transactions to

follow. The server to control the connection, and restrict client requests may use *Access Control*.

Search Facility – The Z39.50 server receives the query initiated by the Z39.50 client, executes it against the data source, and gets the result set. It awaits further instruction from the client. The client to restrict certain data elements from being retrieved may use an *Element Set*.

Retrieval Facility – Initiated by the client, the result set can be retrieved or further refined. The “result set” may be named, and records can be referenced by position within the set. Named “result sets” can also be deleted by the client that created them. Response records can then be sent in the client’s *Preferred Syntax* such as USMARC.

Additional facilities to support advanced forms of information retrieval are included in these two versions of the specification, and will be reviewed in more detail later.

Adoption of the Z39.50 protocol has steadily increased with advances in Internet communication, and text database design. The protocol defines the communication between compliant client and server environments, not the transport method between them. With TCP/IP fully supported, adoption by organizations communicating over the Internet becomes more likely.

Information Retrieval Service

Procedures related to the client are referred to as the Z39.50 origin, and procedures related to the server are referred to as the Z39.50 target. All communication between origin and target is through an A-association, and one or more Z-association belonging to it. Termination of this communication is done by either the origin or target, and cannot be restarted once terminated.

Messages sent via the Z39.50 services are either requests for information, or responses to the requests. There are three types of services:

- **Confirmed** – both request and response
- **Non-Confirmed** – request with no response
- **Conditionally-Confirmed** – request with possible response

A request initiates an operation. This type of request is referred to as an *initiating request*, and it invokes one of eight operation types: Init, Search, Present, Delete, Scan, Sort, Resource-report, and Extended-services. The target may issue a terminating response that ends the operation. For example, a Search request initiates the Search operation, and the resulting response from the target terminates the operation.

When the origin sends a request to the target to search a database, or multiple databases, for matching records, it specifies the databases to be searched, and the query to execute. The

Z39.50 protocol supports the Type-1 query that uses either a single access point (unique, or non-unique key associated with the records being searched), or multiple access point clauses joined by operators such as AND, OR, etc.

Upon completion of the search, the result set (which may be named by the origin) is available for use. To provide a common representation of data retrieved from possibly multiple unrelated data sources, the specification uses a schema to provide a logical representation of the data in a mutually understood way. This representation is called an *abstract database record*, and may be combined with an element specification that results in another *abstract database record*. Once the record syntax is applied to the abstract database record, it is called a *retrieval record*, and is available for export to the origin.

Information Retrieval Facilities

Facilities are made up of logic groups of Information Retrieval Services. Some are made of a single service, but may include more in future versions of the protocol. There are 11 facilities defined in the version 3 specification.

Access Control Facility –

Made up of a single confirmed service initiated by the target. The purpose is for authentication of the origin. The target sends an Access-Control Request to the origin, and the origin must reply with an Access-Control Response. If there is no response, or the origin does not have necessary permissions for the requested operation, the target will take appropriate action including termination of the operation.

Accounting/Resource Control Facility – Consists of three services:

Resource-Control Service – Invoked by the target. Allows the target to initiate a Resource-Control request. There are a number of ways the target may use this service, but one common use is to report to the origin that agreed upon limits will be exceeded by an operation initiated by the origin.

Trigger-Resource-Control Service – Initiated by the origin. This service allows the origin to request that the target cancel the operation, or invoke the Resource-Control Service.

Resource-Report Service – Invoked by the origin. Allows requests from the origin to the target to initiate the Resource-Report operation, and send a Resource-Report related to a completed operation.

Browse Facility –

Includes a single confirmed service, called a Scan Service, which is initiated by the origin. The origin provides an attribute list containing attributes pointing to which term-list (names, titles, subjects, etc.) to scan. The origin also provides a starting term that tells the target where to begin scanning the term list. The Scan Service scans the ordered term list, beginning with the starting term. The target provides feedback to the

origin as to the status of the scan, including success, 5 levels of partial success, and failure.

Explain Facility –

This is the only facility that does not include any services of its own. It uses the services of the Retrieval and Search facilities to provide the origin with details about the target. Information on record syntax, attribute sets, query types supported, diagnostic sets, etc. The Explain database is searchable by the origin as any other database supported by the target.

Extended Services Facility –

Includes a single confirmed service called the Extended-Services service. This service is initiated by the origin, and provides access to the Extended Services Database (IR-Extend-1). Records in this database are actually task packages built from the Request-Parameter-Package parameter. The specification suggests that, although no task package is required for a task being executed immediately, that a task package exist when the task status is pending, active or aborted.

Initialization Facility –

The origin initiates the single confirmed service called init. This service allows the origin to create a Z-association to the target. When the origin invokes the init service, it proposes values for concurrent operations, use of named result sets, resource control settings, etc. It is up to the target to accept or reject this connection, and whether to accept the values proposed by the origin, or set the values of the initialization parameters itself. If the target responds to the origin with Result = 'accept', then the Z-association is created. If the values sent to the origin by the target are not acceptable, the origin may use the Close service to terminate the Z-association.

Result-set-delete Facility –

This facility contains a single confirmed service called the Delete Service, and it is initiated by the origin. The Delete Service allows the origin to request the target delete a named result set, or all result sets, for a given Z-association. The target then responds with the result of the operation.

Retrieval Facility –

Consists of two services:

Present Service – Initiated by the origin, it is a confirmed service that allows requests for response records related to database records in a specified result set. The origin specifies a range in a Present request, like records 1-100 in a result set, and then issue a new request for records 101-200.

Segment Service – A non-confirmed service invoked by the target. It occurs during a Present operation. Records within the range specified by the Present request are returned to the origin in single segment when possible, or in multiple segments when the records will not fit into a single segment. If the records fit into a single segment, the results are not sent as a segment at all however. They are sent as a Present Response by the target to the origin. If multiple segments are required, all segments are sent as segment requests except the last segment, which is returned as a Present Response.

In short, a Present Request is followed by zero or more Segment Requests, and ends with a Present Response.

Search Facility –

The Search Service is a confirmed service initiated by the origin. It includes parameters to indicate the database to search, and the query to be applied to the specified database as part of the Search request. The result set will either be held by the target for further operations, or returned as specified in the request as part of a search response.

Sort Facility –

Includes a single confirmed service called the Sort Service. It is initiated by the origin, and allows it to request that the target sort a result set according to a specified sequence of elements. Multiple result sets can also be merged, then sorted.

Termination Facility –

The Close service, a confirmed service, is unique as it can be invoked by the origin or target, and is not part of any operation. It allows either target or origin to immediately end any operation, or the Z-association. This service can only be used when adhering to version 3 of the protocol.

Z39.50-International: Next Generation (ZING)

<http://www.loc.gov/z3950/agency/zing/>

Zing includes a number of projects initiated by Z39.50 implementers. Some of the projects intend to take Z39.50 more mainstream, adding support for SOAP, XML and XML Schema, and leveraging CQL for searches. Other projects aim to keep the standard the way it is, but hide the complexity using object-oriented API's to connect to a subset of the standard. The following projects are under development:

Common Query Language (CQL) – CQL is a formal query language being developed by the Z39.50 Maintenance Agency as part of the ZING initiative. It provides a simple, yet powerful

query language intended for SRW and SRU, but can be used by other protocols as well. XCQL is also under development, and is an XML form of CQL.

Search/Retrieve for the Web (SRW) – SRW is a web-service based protocol that enables URL and SOAP based access mechanisms to support a number of different client technologies. It supports open standards of CQL and XCQL for queries, and uses XML, XML Schema, XPath and SOAP where appropriate. SRW has some standards that differ from the Z39.50 protocol. Read the SRW background description for more details on the differences (<http://www.loc.gov/z3950/agency/zing/srw/background.html>).

ZOOM – The Z39.50 object-oriented model provides an API to a subset of services of the Z39.50 protocol. It aims to broaden the appeal of the standard by simplifying access to the existing Z39.50 specification.

ez39.50: XER over SOAP – Aims to provide Z39.50 support over alternative internet transfer protocols. This ZING initiative is the agreement of how to implement XER (XML Encodign Rules) over SOAP.

ZeeRex (Z39.50 Explain, Explained and Re-Engineered in XML) – Seeks to solve two problems:

- 1) Finding Z39.50 resources - servers and their databases - for a client to use. It attempts to solve this by using a crawler, similar to Google, or Oracle's Ultra Search, to automatically build a repository of Z39.50 servers. Nicknamed Friends and Neighbors (F&N), the crawler gathers small XML records describing new Z39.50 servers from known servers. In this way, the list can be automatically updated simply by adding a reference of a new server to an existing and known server.
- 2) Once a server is known, how do you know what it can do – ZeeRex attempts to solve this problem by combining two past approaches, Explain Classic and Explain Lite, and then using the same crawler mechanism described in the first bullet to allow one database to explain another. It calls this implementation ZeeRex Proper. The specification defines a searchable Explain database, stored as an XML record. Each record represents a Z39.50 database somewhere, and just like F&N, as long as some server contains this information, the details of what the database can do is known. In this way, complete information can be obtained without having every server implement ZeeRex Proper.

Final Note

Important, and very complex topics have been presented at a very high level. I recommend reading more about the Z39.50 protocol at <http://lcweb.loc.gov/z3950/agency/>. Peak Retrieval has scheduled additional white papers detailing SRW and CQL in case studies, providing more in-depth information on the implementation of these standards.

Peak Retrieval, LLC provides development, training and documentation services in the area of Information Retrieval. We specialize in Oracle Text (formerly interMedia text), Ultra Search, SQL Server, XML, XML Databases, and Z39.50 implementations. Give us a call at (720)339-5807 or e-mail info@PeakRetrieval.com and let us know what you need.