

Federated Search and Z39.50 Is an Information Retrieval Standard

**Ms. Anupama Sonaji Sonwane*

Introduction:

Federated search is an integral component of an information portal, which provides the interface to diverse information resources. Once the user enters his or her search query in the search box of the Information Portal, the system uses federated search technology to send the search string to each resource that is incorporated into the portal. The use of several databases is really a frustrating and daunting task, So Federated searching is a very new concept that is gaining importance in libraries everywhere.

Z39.50 as a information retrieval standard it is formally known as the American National Standards Institute/National Information Standards. It is a network application standard that enables different computer systems, that run on different hardware and use different software, to interoperate and work together it is developed to overcome the problems associated with multiple database searching such as having to know the unique menus, command language z39.50 - ANSI/NISO z39.50 as a defines a standard way for two computers to communicate for the purpose of information retrieval. Z39.50 makes it easier to use large information databases by standardizing the procedures and features for searching and retrieving information. The Specifically z39.50 supports information retrieval in a distributed client and server environment where a computer operating as a client submits a search request to another computer acting as an information server. Software on the server performs a search on one or more databases and creates a set of records that meet the criteria of the search request as a result. The server returns records from the resulting set to the client for processing. The power of z39.50 is that it separates the user interface on the client side from the information servers the search engines and databases. Z39.50 provides a consistent view of information from a wide variety of sources and offers client implementers the capability to integrate information from a range of databases and servers.

What is Federated Search?

A historical perspective on the precursors to federated searching shows us that we have back to travelled a long way from the 1970's through to the 1990's. We have experienced and used and inflicted upon our users multiple hard copy volumes and indexes, acoustical modems, desktop access via the growth and development of PC's in the early 1980's, CD ROM's in the mid 1980's, and by the mid 1990's databases were making a migration to the Internet. Federated search is an information retrieval technology that allows the simultaneous search of multiple searchable resources. A user makes a single query request which is distributed to the search engines participating in the federation. The federated search then aggregates the results that are received from the search engines for presentation to the user. Users can then view search results in a single integrated list. In other words, users do no longer need to consult each information resource individually. Instead, they can search multiple library catalogs (OPACs), Web sites (e.g. Amazon.com, Google etc.), subscription and citation databases all at once. Federated search is an integral component of an Information Portal, which provides the interface to diverse information resources. Once the user enters his or her search query in the search box of the Information Portal, the system uses federated search technology to send the search string to each resource that is incorporated into the Portal. The individual information resources then send the Information Portal a list of results from the search query. Users can view the number of documents retrieved in each resource and link directly to each search result. In short, an Information Portal uses federated search technology to enable users to:

- 1) Search for information in multiple information resources through a single query.
- 2) View search results in a single list.
- 3) Link directly to each resource to expand the search.

Why Federated Searching:

There are several needs for the involvement of federated search as follows:

- The increased need for pin-pointed and information.
- It is very difficult to find out which database to search.
- It is very difficult to search all the databases even if you know the databases.
- It saves the time of the user, as faster searching is possible.

- The need to learn one simple interface rather than many complex interfaces.
- The search quality is high.
- The databases have simplest technical search.
- It helps in getting answers of quite complex and typical queries.
- The Search environment is robust type and centralized.
- It helps in locating best documents with the help of ranking.

Z39.50 as a Information retrieval standard:

Brief History of z39.50:

Early 80's; ie. NISO has established a standards committee in 1979 to work on an information retrieval protocol. Also in late 1970's funding by Council on Library Resources (CLR) supported the development of an experimental protocol as part of the Linked System Project (LSP) for searching bibliographic databases and transferring records among the Library of Congress, Online Computer Library Center (OCLC), The Research Libraries Group (RLG) and Washington Library Network. In 1988 the standard has been approved by the NISO and a group of z39.50 implementers began work to enhance and expand the utility of the standard. The z39.50 Implementers Group (ZIP) is responsible for the development of version 2 and 3. Version 4, development began in Autumn 1995.

Some current problems in Internet information retrieval

- There is only a little meta-information about the Internet information resources,
- There are a lot of different search syntaxes,
- There are a lot of different user interfaces, and
- Presentation of information (records) is not effective.
- Z39.50 has the capabilities to solve the problems mentioned above.

Clients and Servers versus Origins and Targets

Z39.50 follows what is known in Computing as a client/server model, where one computer submits a request to another computer which then services the request and returns some kind of answer. As we saw above with queries being sent to multiple databases simultaneously, there can be more than one server/target, although the user will normally send query from one client/origin at any given point of time.

Definition:

LYNCH, CLIFFORD A. :Z39.50 properly "Information Retrieval (Z39.50); Application Service Definition and Protocol Specification, ANSI/NISO Z39.50" – is a protocol which specifies data structures and interchange rules that allow a client machine (as called an "origin" in the standard) to search databases on a server machine (called a "target" in the standard) and retrieve records that are identified as a result of such a search.

Basic Structure of z39.50:

z39.50 is divided into eleven basic structural blocks, known as Facilities. These Facilities comprise Initialization, Search, Retrieval, Result-set-delete, Browse, Sort, Access Control, Accounting/Resource Control, Explain, Extended Services, and Termination. Each Facility is divided into one or more Services, and it is these Services that people usually talk about. A Service facilitates a particular type of operation between the Origin and the Target, and z39.50 applications select those Services which will be needed in order to fulfill their function. The most basic services are Initialization (Init.), Search, and Present, and all of these should be found in the majority of z39.50 applications.

1. Initialization

Initialization is the first step in any query process, and involves the Origin making itself known to the Target, and agreeing a few 'ground rules' for the manner in which subsequent queries will be handled. Systems which require passwords will also exchange authentication details at this stage.

2. Search

Search is where the majority of the work is done, as it is this Service which enables the Origin to submit queries to the Target. These queries may range from a very simple to complex Boolean queries (AND, OR, NOT, >, <, etc.).

3. Present

Present is used to control the manner in which results are returned to the user. Within Present, a user could ask for the first ten records of a large result set, or request that the data be returned in a different Record Syntax; for example UKMARC instead of USMARC,

Features Provided By Z39.50:

- **Initialization:**

When a session is first established between a client and server it provides means for initiating options that are to be used through the remainder. This includes the default character set, default language and protocol. It also provides the means of authenticating the user.

- **Searching :**

Provides means of searching one or more databases using a structural query using a well known search format. The query may contain boolean operators, fielded search terms, proximity searching, weight search terms, truncation specification, relation specifiers etc.

- **Presenting Records:**

An extensive means of accessing information from a set of search results is provided through the protocol. This includes requesting specific ranges of search results, specific elements in records, specific variants in records, search term highlighting etc.

- **Maintaining Multiple Search Results:**

Z39.50 provides the capability of creating, naming storing and retrieving from one or more search result sets. This also provides facility to client to apply a search criterion to previously created results set.

- **Browsing :**

Z39.50 provides the ability to browse a window of index term or fields within the database.

- **Sorting Of Results :**

Z39.50 offers a means to sort a set of search results based on any given sort criterion.

- **Controlling Access:**

Not only does Z39.50 enable authentication on per-session basis but it also allows authentication on a per-operation basis for the cases where the access to specific databases or records is controlled.

- **Controlling Resources:**

Z39.50 provides a means for clients to cancel a search or a presentation request in the middle of an operation while continuing to maintain an open session with the server. It also permits clients to request resource reports that include accounting information on the number of searches, retrievals etc performed by the user.

- **Extended Services :**

Z39.50 provides the ability to perform database maintenance operations, such as database updates, record insertion deletion etc. It also includes persistent result sets, queries and period queries.

How Z39.50 Works??

In this process the server is known as standard as the “**target**”.

A client is known as an “**origin**” in the standard. User selects the target library from the menu.

- Enters search terms, the query is then sent to server by client.
- The client's user interface will help to build the query and will translate it in the proper syntax.
- There is a primary negotiation between the client and server.
- This particular query will ask the server for records where the terms are found.
- The server translates the search request for the target's library's database and receives a response about the numbers of machines. The client's user interface will tell how many records are in the set but not transferred at this point.
- If the client decides to ask for the records in the result set a transfer is initiated. Client receives the records.

Z39.50 Versions 2 and 3:

- Version 2 was published in 1992 as ANSI/NISO Z39.50-1992.
- The standard is capable of building simple and complex search queries using wide range of query qualifiers.
- It can not be used conveniently for non-bibliographic records
- Version 3 was published in 1995.
- It incorporates many more features to support both bibliographic and nonbibliographic searching.

Z39.50 Client and Server Software on the Web:

There are several of Z39.50 Client/Server softwares on the web. Many are available as free downloads, where as some on evaluation. Following are few such available softwares:

- ZNavigation: a z39.50 Client software is designed to run on Windows 3.11 and Windows 95 PCS <http://www.sbu.ac.uk/~litc/caselib/software.html> [1.6 MB]
- Bookwhere Database: Free evaluation version. <http://www.bookwhere.com> [2963 KB]
- UFO/Fiat Lux: Windows 95/NT z39.50 client http://sun3.lib.uci.edu/~toyofuku/fiat_lux.htm
- A Search Module for Apache: Implements z39.50 search from within the Apache HTTP Server http://igsmatl.er.usgs.gov/apache/mod_search.html
- EndNote 3.0 - first z39.50 Mac Client (does Windows, too) <http://www.niles.com/>
- Netscape for Windows z39.50 plug-in (Alpha version) <http://www.markkelly.com/z3950/ciir.htm>
- SIRSI VIZION Pro : offers access to a series of z39.50 clients and servers <http://www.sirsi.com/Vizion/viziontoc.html>
- Isite: CNIDR, Center for Networked Information Discovery and Retrieval <http://www.cnidr.org/ir/isite.html>

Current Developments and New Initiatives:

1. Version 4 development has begun in ZIG. Most features relevant to the search of bibliographic data have already been included into version 3 ; the most important current work is ranked list query specification.
2. Test beds for Version-3, Scan, and explain facilities are under development.
3. Inclusion of Interlibrary Loan (ILL) services in z39.50 clients via the Extended Services Item Order service is being examined.
4. Various implementation projects, including the European German DBV-OSI II and EC's. OPAC Network in Europe (ONE), will produce high quality public domain z39.50 applications during 1996.
5. The Government Information Locator Service (GILS) is intended to help the public to locate and access information throughout the US government GILS will use z39.50 to provide a seamless connection between distributed GILS servers.
6. The Consortium for the Computer Interchange of Museum Information (CIMI) is examining z39.50 to assist CIMI in its goals of developing museum resources on digital networks and assuring standards-based open interchange of museum information.

Conclusion:

Federated searching reduces the time it takes to search and usually displays results in a common format. Most complete federated search solutions support multiple search protocols. z39.50 is often attacked on a variety of levels by those who see it as overly complex, old Fashioned, not sufficiently web-like, or simply no match for the latest 'great idea'. It is undoubtedly true that z39.50 has quirks and limitations, some of which have been outlined in this paper. It appears for the moment that, whatever its limitations, z39.50 remains the only effective means of enabling simultaneous queries upon distributed heterogeneous databases, and this remains something that the broader user community wants to use. Although it is not yet very popular, it has many perceived strengths and it is safe to say that within library circle it is considered the finest innovation since the OPAC and as a matter of fact z39.50 should soon be adopted on a larger scale. With information resource sharing becoming a major current feature, the z39.50 protocol can definitely not be ignored.

References:

1. Koganuramath M, Mallikarjun A, Kademani B S. z39.50 Information Retrieval Protocol.
2. Fay Turner. (1996) An overview of the z39.50 Information Retrieval Standard .
3. Michael L Smith. (1996) The z39.50 Protocol: literature review <http://istweb.syr.edu/~ist667/z3950.htm>
4. z39.50 : An overview of development and the future by Iltsis, Susannah. [http://www.deepwebtech.com/PDFs/Federated%](http://www.deepwebtech.com/PDFs/Federated%20search-engin)
5. <http://www.libraryjournal.com/lj/ljprintnetconnect/6>. <http://www.searchitright.com/federated-search-engin> e-article.asp
7. Wilson, K. (2004) Meta Search or Federated Search Services – Searching Multiple On-Line Resources. Online Currents, 19, 1-5.
8. Shamkant Deshmukh, Sonia Bhavsar, and Sandeep Bhavsar (2012). Open Source Software for Federated Search. *DESIDOC Journal of Library & Information Technology*, Vol. 32, No. 5, September 2012, pp. 427-430