

Z39.50 Interoperability Testing Framework for Online Library Catalogs Using Radioactive MARC Records

Introduction

In a first phase of the Z39.50 Interoperability Testbed, a large dataset of MARC records was used. In this work, we are exploring how a set of special, diagnostic MARC records can be developed and used to identify interoperability problems between a Z39.50 client and a Z39.50 server providing access to a database of bibliographic records supporting the search and retrieval functions of an online library catalog. We refer to these special, diagnostic records as radioactive MARC records. This document describes a framework for interoperability testing using these radioactive MARC records. It discusses the various components and identifies the tasks related to developing and implementing the components.

Search Requirements To Be Tested

The Z39.50 profiles for library application (Bath and the U.S. National) specify searches that Z39.50 clients and Z39.50 servers (and the underlying information retrieval system) should support if vendors' systems are conformant with the profiles. Each profile specifies levels of conformance, where each level requires the support of a set of searches. For this work, we will base our searches on the U.S. National Profile for Library Applications, approved by NISO as an American National Standard in 2003. Searches defined in conformance levels 0 and 1 will be tested. The following is a list of these searches. For details on the Z39.50 attribute combinations for each search, refer to the profile.

Level 1 Searches

- Author Search – Keyword
- Title Search – Keyword
- Subject Search – Keyword
- Any Search – Keyword

Level 2 Searches

- Author Search – Keyword with Right Truncation
- Author Search – Exact Match
- Author Search – First Words in Field
- Author Search – First Characters in Field
- Title Search – Keyword with Right Truncation
- Title Search – Exact Match
- Title Search – First Words in Field
- Title Search – First Characters in Field
- Subject Search – Keyword with Right Truncation
- Subject Search – Exact Match
- Subject Search – First Words in Field
- Subject Search – First Characters in Field
- Any Search – Keyword with Right Truncation
- Standard Identifier Search
- ISBN Search
- ISSN Search
- Remote System Record Number Search
- Date of Publication Search
- Language Search
- Format of Material Search – Keyword

These searches span a range of access points (e.g., author, title, subject, ISBN, ISSN, etc.), address precision oriented searches (e.g., exact match, first words in fields, first characters in fields) and recall oriented searches (e.g., keyword, keyword with right truncation, phrase with right truncation), and searches that refine results (e.g., limiting by date, format of material, language).

Test Searches

The searching requirements will inform the development of test searches. One or more test searches may be necessary to adequately test the search types listed above. Each test search will define the type of search, the search term, and the expected results to be retrieved from the system being tested. Searches will be sent from the testbed client to ensure that appropriate Z39.50 attribute combinations are used to characterize the search term. The attribute combinations are defined in the U.S. National Profile.

The search term in the test searches will be aligned with specific data from specific fields/subfields in the radioactive MARC record. The retrieved records for each test search should be one or more of the radioactive MARC records, assuming the system being tested is in conformance with the U.S. National Profile.

The Radioactive MARC Records

One of the key research activities in this project is the development of a set of radioactive MARC records. This is particularly challenging because:

- MARC provides for rich content designation (e.g., various types of title information can be recorded in many different areas of the record)
- MARC records represent different formats of materials (e.g., books, computer files, etc.), and different formats may require the use of specific content designation not used in other formats

The number of required radioactive MARC records is not known at this time, but we anticipate creating at least one MARC record for each of the following formats of materials:

- Books, Pamphlets, and Printed Sheets
- Cartographic Materials
- Electronic Resources
- Continuing Resources
- Manuscripts (including manuscript collections)
- Music (Notated and manuscript music)
- Sound Recordings (musical and non-musical)
- Motion pictures and video-recordings (including digital and non-digital)
- Graphic materials (includes mixed materials, with or without archival control)
- Three Dimensional Artifacts and Realia

Multiple records for each format may be necessary to provide adequate coverage for the anticipated interoperability tests.

The MARC records will use the fields/subfields appropriate to each of these formats (since some MARC fields/subfields only apply to specific formats of materials). In each of the records, there will be patterns of alpha/numeric characters encoded in appropriate fields/subfields. The specifications for the character strings and how they will be generated to code into the records will be determined in discussions with project staff. An example of the concept follows:

The 245\$a (title information) for two records might look like the following:

```
245 $aXysos sifys asdsai-aduus  
245 $aYszaqs sifys asdsai-aduus
```

A title keyword search with the search term “sifys” should retrieve both of these records. A title exact match for the term “Xysos sifys asdsai-aduus” should only retrieve the first record.

Indexing Policies

During the first phase of the Z39.50 Interoperability Testbed Project, we developed indexing guidelines for several of the profile-defined searches (i.e., author, title, and subject keyword searches). When developing these guidelines, we identified approximately 500 subfields that could potentially store relevant data.

Since the testing planned for the current work encompasses searches beyond the keyword searches used in the first phase, it will be necessary to produce indexing guidelines for the additional searches. The indexing guidelines will be implemented in the Z-Interop Z39.50 server and underlying information retrieval system reference implementation. This is necessary to ensure baseline information about searching and retrieval behavior prior to conducting interoperability testing with other systems.

Testing Procedures

The intention of this work is to develop a set of automatic procedures to send the test searches and retrieve results. In the first phase of the project, sending the test searches was manually done by project staff. Automatic procedures can increase the efficiency of interoperability testing. These procedures should enable an increased number of test searches to be used. This will be important to assess the indexing policies in operation on the systems that will be tested. For example, we know that title information can be stored in different MARC subfields. The radioactive MARC records will have different patterns of characters in each of those subfields, and we can send individual searches with the specific string of characters that are stored in each of the subfields to see if the record is returned. If it is, we can conclude:

- The Z39.50 server is processing the search properly and sending it to the information retrieval system
- The information retrieval system does index the particular subfield.

The automated testing procedures will also assist in the management and analysis of results to reduce the manual effort that was required in the first phase of the testbed project. For example, the procedures should record any Z39.50 server diagnostic messages to understand if the server does not support particular types of searches; the procedures should record the search sent and the record retrieved; the procedures should enable compilation of results in an easy to use format for generating reports.

Availability of Radioactive MARC Records

For systems that choose to go through interoperability testing, the radioactive MARC records must be available for access and loading into the local system. The records will be stored on a password protected website from which system managers will be able to download the records. The records will be in conformant MARC 21 format, and system managers should be able to put the records into normal batch loading or other input queues.

Analysis and Reporting

Upon completion of interoperability testing, the results will be analyzed using manual and automatic methods. Wherever possible, automatic analysis should be used to streamline the process of analysis. Interoperability testing reports should include the following:

- Assessment of the level of Z39.50 interoperability
- Assessment of search functionality available
- Assessment of indexing policies.

Completed reports should be easy to understand by the managers of the system tested, and should enable them to take specific actions that might improve interoperability.

Tools and Resources Required

A number of tools will be required to set up and operate the testbed using the radioactive MARC records. The section describes anticipated tools and resources.

Database of MARC Content Designation

Because of the rich content designation available in MARC for bibliographic records, and because some of the content designation is intended for use when describing specific formats of materials, we will create a MARC Content Designation Database (MARC DB). This database will store information from the MARC 21 document and include the following:

- Field tags
- Indicator values, names, and definitions associated with field tags
- Subfield codes, names, and definitions associated with field tags
- Format applicable for each subfield code.

The MARC DB will enable queries to identify all appropriate content designation for the 10 formats of materials discussed above. In addition, the MARC DB will be used when generating reports of testing results to efficiently pull content designation names and definitions to make the reports easier to understand.

[Status: Under development by TxCDK]

MARC Record Creation Tool

To create the radioactive MARC records, we will use a basic MARC record creation tool or cataloging tool. The tool must allow the use of all available MARC content designation (either through a template or by editing). It would be best to use a tool that provides basic validation of MARC input. If possible, the tool should allow import from a CSV or other form of text file to get data into the record and then produce a conformant MARC 21 record.

[Status: To be identified and evaluated by TxCDK]

Radioactive MARC Records

A set of specialized, diagnostic MARC records will be created for use in interoperability testing.

[Status: Provided by TxCDK]

Z39.50 Client for Testing

A Z-Interop Z39.50 client will be used to send the test queries. The client must be conformant with Levels 0 and 1 of the U.S. National Profile.

[Status: Provided by Index Data?]

Z39.50 Server and Information Retrieval System Reference Implementation

To ensure the reliability of all aspects of the interoperability testing, we will use a reference implementation of a Z39.50 server and information retrieval system. Current plans are to use the existing Z-Interop reference implementation (the SIRSI Unicorn system). Current and new indexing guidelines will inform the indexing policies for the reference implementation. The radioactive MARC records will be loaded into the system and made available to searching. Searching and retrieval procedures will be tested and verified.

[Status: Probably needs some checking, but should be operational with a low level of effort]

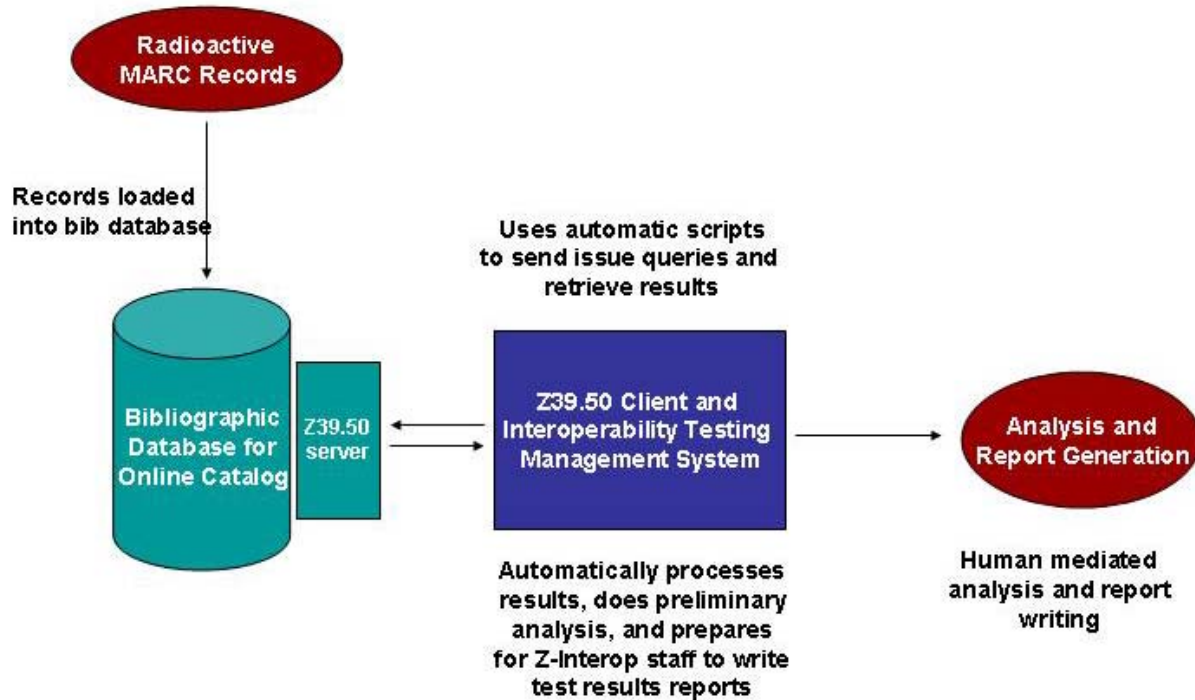
Automatic Scripts for Testing and Reporting

Scripts will be used to send test searches from the Z-Interop client to systems.

[Status: Provided by Index Data?]

Visual Overview of Radioactive MARC Record Interoperability Testing

The basic interoperability testing framework can be graphically illustrated to highlight key components and processes involved as in the following.



Summary

This document is intended to initiate and guide the work for developing an approach for interoperability testing using radioactive MARC records. It has identified the components for interoperability testing and some of the tools and resources needed. Each of these components and processes will be described in more detail in separate documents. Additionally, a work plan document will identify the work areas, activities, and tasks to further guide the work and provide project management oversight for the work.