Z39.50 for Finding It All

William E. Moen
<wemoen@unt.edu>
School of Library and Information Sciences
Texas Center for Digital Knowledge
University of North Texas
Denton, TX  72603
&
Slavko Manojlovich
in absentia
Overview

- Framing the ‘finding it all” question
- Finding it all: A resource discovery service project
- Update on the Z39.50 profiles
- Z39.50 next generation
Framing the “finding it all” question

- Finding
  - I can find only what’s been exposed

- Exposing
  - How content is exposed determines the how and if I can find it

- Common agreements on exposing and finding
  - Protocols
  - Syntaxes
  - Semantics
    - Database
    - Language

- Goal is to connect users to information

- Strategy is interoperability on multiple levels (protocol, syntax, semantics, organization, etc.)
What/how to expose? How to find?

Content as:
- Text
- Images
- Data
- Audio
- Multimedia
- ...

How to expose

How to find

Users
User Groups

Users
User Groups

Users
User Groups

Users
User Groups

Users
User Groups

Users
User Groups
Exposing/Finding Option 1

Diagram showing the flow of content and search interfaces, with connections to user groups.
Exposing/Finding Option 2

World’s Largest Library Database

Search Interface

- Users
  - User Groups

- Users
  - User Groups

- Users
  - User Groups

- Users
  - User Groups

- Users
  - User Groups
Exposing/Finding Option 3
Exposing/Finding Option 4

- Standards-based search/retrieval interface
- Customized method
- Single Search Interface
- Proprietary or non-standard search/retrieval interfaces

Users
- User Groups
- User Groups
- User Groups
- User Groups
Exposing/Finding Option 5

Standards-based search/retrieval interface

Proprietary or non-standard search/retrieval interfaces
Resource discovery service

- An approach for **finding**

- Discovery services
  - Extends the reach and range of a user across
    - Organizational, collections, and format boundaries
  - A variety of tools and approaches for
    - Discovering the existence of appropriate resources
    - Identifying and selecting useful resources

- Search and retrieval services are core
  - Single repository searching
  - Broadcast searching across similar repositories
  - Integrative searching across diverse resources
Resource discovery service

Content or Resource Collections
1. Commercial Databases
2. E-books
3. Electronic Archive of Government Information
4. Online Library Catalogs

Representations of Resources
1. Proprietary Records
2. Indices of Full-text Files
3. Dublin Core Metadata Records
4. MARC Records

LOT Search & Retrieval Interface
1. Broadcast Searching
   a. Similar Resources
   b. Dissimilar Resources
2. Common Retrieval Display

Middleware
1. Translates Searches
2. Formats Retrievals
3. Interacts with Multiple IR Systems

Information Retrieval Systems
Exposing content through metadata

- IR System
- Metadata
- Information Retrieval Systems

- IR System
- Metadata
- Resource Descriptions

- IR System
- Metadata

- IR System
- Metadata

- IR System
- Metadata

Images
Licensed Online Databases
Web Resources
Books

Information Resources
Z39.50 resource discovery
Factors affecting interoperability

- Multiple and disparate systems
  - operating systems, information retrieval systems, etc.
- Multiple protocols
  - Z39.50, HTTP, SOAP, etc.
- Multiple data formats, syntax, metadata schemes
  - MARC 21, UNIMARC, XML, / ISBD/AACR2-based, Dublin Core
- Multiple vocabularies, ontologies, disciplines
  - LCSH, MESH, AAT
- Multiple languages, Multiple character sets
- Indexing, word normalization, and word extraction policies
Z39.50 profiles update

- Bath Profile
- U.S. National Profile
- Texas Profile

A Z39.50 profile:

- Identifies searching requirements (tasks)
- Defines the searches (semantics and behavior)
- Specifies Z39.50 query to represent the search
  - Standard combination of Z39.50 attribute types and values
  - Clients must send all attribute type values specified for search
  - Servers must be able to process all values
- Requires support for specific formats for interchanging retrieval records
The Bath Profile

The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery

Release 2 (Draft 3, Oct. 2002) defines four Functional Areas

- Functional Area A: Basic Bibliographic Search and Retrieval, with Primary Focus on Library Catalogues
- Functional Area B: Bibliographic Holdings Search and Retrieval
- Functional Area C: Cross-Domain Search and Retrieval
- Functional Area D: Authority Record Search and Retrieval in Online Library Catalogues

FOR MORE INFORMATION, VISIT THE BATH MAINTENANCE AGENCY WEBSITE…

http://www.nlc-bnc.ca/bath/
Bath Functional Areas B & D

- **Area B -- Holdings Information**
  - Address the challenge of search and retrieval of bibliographic holdings information
    - Locations Only
    - Locations, Summary Information and Count if available
    - Summary Copy Level Holdings
  - Use of XML as Record Syntax

- **Area D – Authority Record Search/Retrieval**
  - Defines one conformance level
  - Defines 14 searches
U.S. National Z39.50 Profile for Library Apps.

- National Information Standards Organization (NISO) standards effort
- **U.S. National Profile:**
  - Addresses cross-catalog searching and holdings information interchange
  - Bath Profile is foundation for U.S. National Profile
  - Responds to national requirements
- Updated and harmonized with Bath Profile (A & B)
- Draft standard ready by end of 2002

**FOR MORE INFORMATION, VISIT THE PROJECT WEBSITE...**

http://www.unt.edu/zprofile
Z Texas Profile for Library Applications

- Release 3.0 (Draft November 2002)
- Harmonized with Bath & U.S. National Profiles
- Adds new functional area: **Functional Area E: Basic Search and Retrieval of Database Citation Records**
  - Search and retrieval of bibliographic information found within citation, full-text, and abstracting and indexing databases
  - Level 0: 4 searches; Level 1: 14 searches
  - Requires XML for retrieval records

For more information, visit the project website...

[http://www.tsl.state.tx.us/ld/projects/z3950](http://www.tsl.state.tx.us/ld/projects/z3950)
Z39.50 profiles are not enough

Profiles can:

- Identify searching requirements (tasks)
- Define the searches (semantics and behavior)
- Specify Z39.50 query to represent the search and formats of retrieval records

Also needed are:

- Agreements on indexing
- Common search functionality
- Methods and testbed for interoperability testing
- Conformance to profiles by vendors and libraries
ZING – Z\(^{39.50}\) International Next Generation

- Make intellectual/semantic content of Z39.50 more broadly available
- Make Z39.50 more attractive by lowering barriers to implementation
  - Use of XML – to represent and encode data
  - Use of HTTP – for transport
  - Use of SOAP – for interaction between client and server based on Remote Procedural Call (RPC)
- Several initiatives: ZOOM, ez39.50, ZeeRex, SRW/U

For more information, visit the project website...

http://www.loc.gov/z3950/agency/zing/zing.html
Search/Retrieve Web Service – Overview

- Builds on Z39.50 and web technologies
- Retains Z39.50 concepts:
  - Results sets
  - Abstract access points
  - Abstract record schemas
  - Explain & Diagnostics
- Web technologies: XML, SOAP/RPC, HTTP
- Combines several Z39.50 features into two “operation types”
  - search/Retrieve
  - Explain
- Provides a standards-based method for representing and issuing queries, and retrieving records
Classic Z39.50 and SRW models

**Classic Z39.50**

- **System A**
  - User Interface / client
  - Z39.50 client software
  - Query formulation

- **System B**
  - Z39.50 server software
  - Info server
  - Database / search engine

**SRW**

- **System A**
  - User Interface / client
  - SRW Client
  - Query formulation

- **System B**
  - SRW Server
  - Database
  - Database / search engine

Network Access via the Internet
Differences between SRW and Classic Z39.50

- No explicit concept of connection, session, or state
- Results sets named by server and may provide result set id for subsequent reference
- Single record syntax – XML
- Multiple XML schemas
- String queries using Common Query Language (CQL)
- Expressive query using XCQL (represented in XML schema)
- Flat access points/indexes rather than Z39.50 attributes
- Explain is database oriented
  - Information about supported indexes
  - Information about supported schemas
- XML instead of ASN.1 for abstract syntax and encoding
SRW searchRetrieve service definition

- Request parameters
  - query (mandatory)
  - sortSpec (optional)
  - startRecord (optional)
  - maximumRecords (optional)
  - recordSchema (optional)

- Response parameters
  - numberOfRecords (mandatory)
  - resultSetID (optional)
  - rsIdleTime (optional)
  - records (optional)
  - diagnostics (optional)
search/Retrieve query

- SRW query is logically one or more operands linked by Boolean operators
- Five categories of queries:
  1. single search clause
  2. two or more search clauses linked by Boolean
  3. search clauses and result sets linked by Boolean
  4. two or more result sets linked by Boolean
  5. single result set
Common Query Language (CQL)

- **Search clause**
  - always includes a term
    - simple terms consist of strings enclosed in double quotes
  - may include index name and relation
    - index name includes base name and may include prefix
      - title, subject
      - dc.title, dc.subject
    - relation
      - = used when 1) asking for result set; 2) for numeric equality; 3) for word adjacency when term is an ordered list of words
      - exact used for string matching
      - all when term is list of words to indicate all words must be found
      - any when term is list of words to indicate any words must be found
Other components of CQL

- Boolean operators: and, or, not, prox
- Prox
  - unit (word, sentence, paragraph, element)
  - relation (<, >, <=, >=, =, <>)
  - distance (integer)
  - ordering (ordered or unordered)
- Masking rules and special characters
  - single asterisk (*) to mask zero or more characters
  - single question mark (?) to mask a single character
  - carat/hat (^) to indicate anchoring, left or right
  - backslash (\) is used to escape other masking characters
searchRetrieve response

- Records returned in response
- All records in XML syntax
- According to one or more XML schemas (semantics)
- Dublin Core may be the one required syntax
- MARC in XML
SRW and Classic Z39.50

- Lightweight SRW clients
- Gateways between SRW and classic Z39.50
- SRW queries should map to Z39.50 queries
  - CQL $\rightarrow$ Z39.50 attribute combinations
- Retrieval records returned to SRW gateway
  - Transformed from a Z39.50 record syntax to XML

For more information, visit the project website...

http://www.loc.gov/z3950/agency/zing/zing.html
http://www.loc.gov/z3950/agency/zing/srwu/implementors.html
Can Z39.50 find it all?

- No
  - Not all content is exposed through Z39.50 servers
- But...
  - Content can be exposed more systematically through Z39.50 because of profiles
  - Content may be exposed more easily through SRW
    - OAI repositories
    - Commercial online databases
    - Local databases of metadata
Increasing standards-based finding it all!

- Content
- Classic Z39.50 Servers
- SRW
- Gateway
- Single Search Interface
- Users
  - User Groups
- OAI Repository
- SRW Servers
- Proprietary or non-standard search/retrieval interface
References

- The Bath Profile Maintenance Agency
  - http://www.nlc-bnc.ca/bath/
- U.S. National Profile
  - http://www.unt.edu/zprofile/
- Texas Profile
  - http://www.tsl.state.tx.us/ld/projects/z3950
- ZLOT Project
  - http://www.unt.edu/zlot
- ZING
  - http://www.loc.gov/z3950/agency/zing/zing.html
- Z39.50 Interoperability Testbed
  - http://www.unt.edu/zinterop/