Inter-Institutional Database Unification using Z39.50 and Dublin Core
~Resource Sharing System for Humanity Researches~

2003.11.09 @PNC 2003 Bangkok

Shoichiro Hara
(National Institute of Japanese Literature)
1-16-10, Yutaka-cho, Shinagawa-ku, Tokyo 142-8585, Japan
hara@nijl.ac.jp
(http://world.nijl.ac.jp/~kiban-s/)
I am a graduate student of Osaka City University. My major is “Literature and Gender.” First of all, I tried to find as many materials as possible from the Web. When I accessed the NIJL Home Page, I found that not only catalogue database but also other databases such as full texts, images, archives have related data, and found that other institutes also have materials. But I do not well which institute has what databases, then I have retrieved databases one by one. And each database system has its own operation. It was time-consuming jobs! Are there tools that can access databases of several institutes simultaneously and can retrieve information by one operation?
Objects

• Each Institute creates information retrieval systems that hide differences of
  – Database Locations
  – Database Structures
  – Retrieval Methods

• But do not change existent database systems
• Every institutes should be equal partner
• As the result
  – Users see databases on the Web as one database
What are the Problems?

Most Databases are Heterogeneous…

- **Non-portability**
  - Hardware Dependent, Software Dependent

- **Similar but Different Databases**
  - Historical Background, Different Purposes

- **Incompatibility**
  - Different Operations, Non-Interoperability
Example of Non-portability

Example of Non-portability

Binary

<table>
<thead>
<tr>
<th>C1</th>
<th>D4</th>
<th>D6</th>
<th>C1</th>
<th>01</th>
<th>B4</th>
<th>00</th>
<th>07</th>
<th>01</th>
<th>D0</th>
<th>85</th>
<th>F0</th>
<th>F8</th>
<th>F1</th>
<th>F5</th>
<th>F4</th>
</tr>
</thead>
</table>

means

EBICDIK A M O A Dir Tags Data (Call Number)

=436 =7 =464
Similar but Different Databases

### DB1

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Address</th>
<th>Birth</th>
<th>Weight</th>
<th>Height</th>
<th>BP</th>
<th>T.Chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoichiro Hara</td>
<td>M</td>
<td>1-16-10, Yutakacho, Shinagawaku, Tokyo, 142-8585, Japan</td>
<td>1957/5/20</td>
<td>170</td>
<td>65</td>
<td>140/85</td>
<td>220</td>
</tr>
</tbody>
</table>

### DB2

<table>
<thead>
<tr>
<th>First Name</th>
<th>Last Name</th>
<th>Sex</th>
<th>DateOfBirth</th>
<th>Street</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoichiro</td>
<td>Hara</td>
<td>Male</td>
<td>20-May-57</td>
<td>1-16-10, Yutakacho, Shinagawaku</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prefecture</th>
<th>Nation</th>
<th>ZIP</th>
<th>Hint</th>
<th>Weight</th>
<th>Chol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokyo</td>
<td>Japan</td>
<td>142-8585</td>
<td>170</td>
<td>65</td>
<td>220</td>
</tr>
</tbody>
</table>
Incompatibility

EXAMPLE:
Retrieve records containing term “TOSANIKKI”

**By Hitachi Mainframe Computer**
1 / B : TOSANIKKI

**By SQL Database**
SELECT author FROM books WHERE title LIKE ‘TOSANIKKI’

**By PAT**
region “title” including “TOSANIKKI”
region “book” including %
region “author” within %
Solutions

• Introducing Standards for
  – Data Description for Portability
  – Mutual Data Structure for Different Structures
  – Standard Data Retrieval for Compatibility

• Standardization NOT by
  – Compulsion
  – Authority
Schema of 3 Steps Standardization

Data Description Standard

Data Structure Standard

Data Retrieval Standard

Existant Databases

Existant Methods
3 Steps Standardizations

1. First Step: Data Description
   - SGML/XML

2. Second Step: Mutual Data Structure
   - Metadata (Dublin Core, EDI, EAD, TEI etc)

3. Third Step: Data Retrieval Compatibility
   - Protocol (Z39.50 etc)
Step 1: Data Description by SGML/XML

- <?xml version="1.0" encoding="Shift_JIS" ?>
- <!DOCTYPE classics SYSTEM "../dtd/xml-n-classics2.dtd">
- <?xml:stylesheet type="text/xsl" href="../xslt/classic/honmon_limit.xsl" ?>
- <classics>
  <text>
  <body>
  <div0>
    <div0.title>源氏物語</div0.title>
    <div1>
      <div1.title>きりつぼ</div1.title>
      <filename id="p27" />
    </div1>
  </div0>
  </body>
</classics>
Step2 Example of Metadata

<!DOCTYPE record-list SYSTEM "dc-history">
<record-list>
  <dc-record>
    <title>鈴木家</title><title>鈴木太郎</title>
    <creator>千葉県海上郡海上町史編纂委員会</creator>
    <subject>海上町史料所在目録 第三集</subject>
    <subject>千葉県海上郡海上町史編纂委員会</subject><subject>海上町関係史</subject>
    <subject>鈴木家</subject><subject>江戸前</subject>
    <subject>下総国 香取郡 鈴木村</subject><subject>相模</subject>
    <description>海上町関係史料</description>
    <publisher>千葉県海上郡海上町史編纂委員会</publisher>
    <date>1981</date>
    <type>史料所在目録データベース</type>
    <format>SGM テキスト</format>
    <identifier>1201724</identifier>
    <source>nlj.ac.jp</source>
    <language>ja</language>
    <rights>千葉県海上郡海上町史編纂委員会</rights>
    <rights>国文学研究資料館</rights>
  </dc-record>
</record-list>
Metadata Creation
-Hierarchical Metadata Model-

Area A
- Database 1
- Database 2

Individual Filter
- Area A Oriented Metadata
- Filter for Area A

Area B
- Database 3

Individual Filter
- Area B Oriented Metadata
- Filter for Area B

DC Metadata Database
Z39.50 Server
Step3 : Standardized Data Retrieval

How to Link Heterogeneous Systems?

1. One Data Model
   - Each institute has its own server
   - Every server uses the same data model
   - But, no flexibilities

2. Data Clearing House
   - Each institute does not need additional system
   - Upload every index data to the center server
   - Who maintains system, how to share costs?

   - Dublin Core Metadata and Z39.50
   - No center systems
How to Link Heterogeneous System?
Federation System by Dublin Core + Z39.50

- UC Berkeley
- ECAI Clearing House
- Inter University
- NMHF
- Universities
- Osaka City Univ.

- Z39.50 Gateway
- DC Meta Data Model
- Images
- Domain Specific SGML
- or XML Data Bases
- Standard Data Description
- Standard Data Model
- Meta Database
- Uploading
- Retrieving
- NIJL Data Clearing House

2003/11/09
Basic Components of NIJL Dana Sharing System

1. Each Database Server
   - HiRDB, Bibliotheca (Hitachi Co.) etc.

2. SGML/XML tools

3. Dublin Core Meta Data Server
   - Bibliotheca (Hitachi Co.)

4. Z39.50
   - Z39.50 Server Process
   - Z39.50 Web Gateway
   - Z39.50 Gateway Process
Structure of NIJL Z39.50 System

- Z39.50 Server Process
  - Databases: DB1, DB2, DB3

- Z39.50 Gate Way

- Z39.50 Client
- Web Client

- Local DB

- Z39.50 Server API
- Local DB Access Program

- Z39.50 Gateway Process
Participating Institutes

• The Graduate University for Advanced Studies
  – National Institute of Japanese Literature
  – National Museum of Ethnology
  – International Research Center for Japanese Studies
  – National Museum of Japanese History
  – The National Institute of Informatics
  – Research Institute for Humanity and Nature

• The Historiographical Institute, The University of Tokyo

• Osaka City University

• Keio University

• ECAI Clearing House

• (Institute of South East Studies, Kyoto University)

• (University of Shimane)
# DC-Z39.50 Data Retrieval

## 詳細検索

ブラウザの戻るボタンは使用しないで下さい。検索終了時は 検索先選択 をクリックして下さい。

### 接続状態

<table>
<thead>
<tr>
<th>状態</th>
<th>ホストアドレス</th>
<th>ポート番号</th>
<th>データベース名</th>
<th>備考</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>72.18.54.2</td>
<td>211</td>
<td>de-micro</td>
<td>Connection OK</td>
</tr>
<tr>
<td>O</td>
<td>72.18.54.2</td>
<td>212</td>
<td>de-moko</td>
<td>Connection OK</td>
</tr>
<tr>
<td>O</td>
<td>72.18.54.2</td>
<td>218</td>
<td>de-mori</td>
<td>Connection OK</td>
</tr>
<tr>
<td>O</td>
<td>72.18.54.2</td>
<td>214</td>
<td>de-mori</td>
<td>Connection OK</td>
</tr>
<tr>
<td>O</td>
<td>72.18.54.2</td>
<td>216</td>
<td>de-mori</td>
<td>Connection OK</td>
</tr>
</tbody>
</table>

## 検索条件入力

- **検索対象**
  - DC-Title
  - DC-Creator

### キーワード

<table>
<thead>
<tr>
<th>キーワード</th>
<th>田中先生</th>
</tr>
</thead>
</table>

### 項目内条件

<table>
<thead>
<tr>
<th>項目内条件</th>
<th>AND</th>
</tr>
</thead>
</table>

## 検索履歴

<table>
<thead>
<tr>
<th>選択</th>
<th>集合番号</th>
<th>ヒット件数</th>
<th>検索内容</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>2</td>
<td>35</td>
<td>@attr Use=DC-Title “田中先生”</td>
</tr>
</tbody>
</table>
Further Approach
- Next Z39.50 -

• **WEB Oriented**
  – Remote Procedure Call
  – Portability

• **Light Protocol**
  – Only for Data Retrieval

• **Introducing **SOAP** (Simple Object Access Protocol)**
  – How to Treat ASN.1?
**SOAP System**

- **SOAP Search.NET Client**
  - .NET Framework
  - Java2 SDK
  - Apache Tomcat
  - Apache-AXIS
  - SOAP Search Web Service

- **DB Access**
  - JNI
  - Database
  - DB I/F

- **NIJL E-Library Server (FLORA 730)**
  - Windows NT Server 4.0
  - Windows Terminal
  - Windows NT/2000/XP

- **Server: BASE2**
  - Solaris 8

- **SOAP Client**
  - SOAP Server
Example of Soap Message

(Sender)

<Search>
  <Page> Page Number </Page>
  <DbName> Database Name </DbName>
  <Reference>
    <Seq>
      <Field> Data Filed Name </Field>
      <Contents> Key Word </Contents>
    </Seq>
  </Reference>
</Search>
Example of Soap Message (Receiver)

<Result>
  <Search>
    <DbName> Database Name </DbName>
    <Reference>
      <Seq>
        <Field>Field Name</Field> <Contents>Key Word</Contents>
      </Seq>
    </Reference>
  </Search>
  <Info> <Count>Hit Counts</Count> </Info>
  <Data>
    <Rec>
      <ItemNo>Item Number</ItemNo> <Title>Title</Title>
      <Contents>Contents of Result</Contents>
    </Rec>
  </Data>
</Result>
Thank You for your Kind Attention

Related URL

National Institute of Japanese Literature
http://www.nijl.ac.jp/

Collaboration Project: http://world.nijl.ac.jp/~kiban-s/

ECAI (Electronic Cultural Atlas Initiative)
http://ecai.org

PNC (Pacific Neighborhood Consorcium)
http://pnc-ecai.oui.ac.jp

Resource Sharing Group

Contact E-mail: hara@nijl.ac.jp