

MODEL AND IMPLEMENTATION OF A SYSTEM FOR BIBLIOGRAPHIC RECORDS INTERCHANGE

Branko Milosavljević¹, Zora Konjović¹

Abstract. One of the primary functions of a library information system within a network of library institutions is shared cataloguing. Shared cataloguing is an interchange of bibliographic records. The records are usually created in national library institutions, and the local libraries fetch these records and supplement them with their local data. This paper presents a system for interchanging the bibliographic records created according to UNIMARC format. One additional option is the interchange of the records coded in XML form. The server-side software module for fetching records using a WWW browser is implemented. This way enables the local libraries to collect bibliographic records via the Internet and create their own local databases.

AMS Mathematics Subject Classification (1991): 68P20

Key words and phrases: UNIMARC, XML, shared cataloguing, Java servlets

1. Introduction

Bibliographic records are descriptions of the published documents, created according to a defined standard. They belong to different categories depending on the nature of the original document. For example, there exist descriptions of monographic, serial or geographic publications, as well as many other types. For each of these types one can define the way the bibliographic description of the published document should be created and stored. The most widely used standard for bibliographic records is UNIMARC format [1].

The library information system BASIS ver.3.01 is an Internet i.e. Intranet application in the Java environment. *Unicode standard* [2] is consequently implemented through the complete system BASIS. The paper [3] presents an overview of the development of the system BASIS ver. 3.01. Modeling and implementation of the UNIMARC text server are described in details in [4, 5]. The class diagram of the user search subsystem is given in [6]. This paper describes the architecture and implementation of the user search module, as well as implementation of the subsystem for bibliographic records interchange.

¹University of Novi Sad, Faculty of Technical Sciences, Trg D. Obradovića 6, 21000 Novi Sad, Yugoslavia

2. UNIMARC format

UNIMARC (*UNiversal MACHine Readable Cataloguing*) format is the standard whose aim is primarily to support national bibliographic institutions with the option of international interchange of the bibliographic data in machine readable form. This standard defines content marks (fields, indicators and subfields) which are used for creating the bibliographic records, as well as their logical and physical format. It covers monographic, serial and cartographic publications, music, audio records, images, video materials and reserves the place for adding computer files. The name *UNIMARC* format denotes the format prescribed by the standard, while the name *UNIMARC record* denotes one bibliographic unit described following the UNIMARC format.

Each UNIMARC record consists of a finite number of *fields*, whereby not all the fields are mandatory. If the contents of a particular field is not defined, this field does not appear in record. The fields are uniquely determined by three digit numbers. The field is the structure consisting of two *indicators* and a finite set of *subfields*. The subfields are also uniquely determined by the single alphanumeric character. Analogously to field, if the contents of some subfield is not defined, this subfield does not appear in the subfields set of the corresponding field. The content of the subfield is text. The subfield content is the basic information unit (*data element*) owned by UNIMARC record. Indicators (*first* and *second*) are used for more precise description of the contents contained within subfields.

Some fields are *repeatable*, i.e. they can appear more than once within a single record. There are also the fields which are *mandatory*, meaning that these fields must appear within the record. The fields that can appear as the values of subfields are called *secondary fields*. The subfields within a field can also be characterized as mandatory and repeatable.

UNIMARC format offers the possibility of extending the standard by additional fields definitions which are used in the national interchange (Block 9). Besides, for the description of different document types (monographs, serial publications, etc.) the different subfield sets are used.

UNIMARC format defines the specific way for storing the text consisting of the characters from more than one alphabet. On the other hand, the WWW technology supports the *Unicode* standard defining code layouts for simultaneous support of the most existing alphabets which are currently used. The contemporary WWW browsers are equipped with the possibility of displaying the text consisting of more than one alphabets by UTF-8 code layout prescribed by *Unicode*. The conversion of the text from the UNIMARC layout to *Unicode* layout and vice versa is possible for the alphabets supported by both standards.

3. Modeling the search system

The aim of the objects from *UserInfo* class is to keep information about the user session. The information contains a set of the document identifiers

found during the last search, data about the current state of the hits display and the set of the identifiers corresponding to the documents which are selected for transfer during the user session.

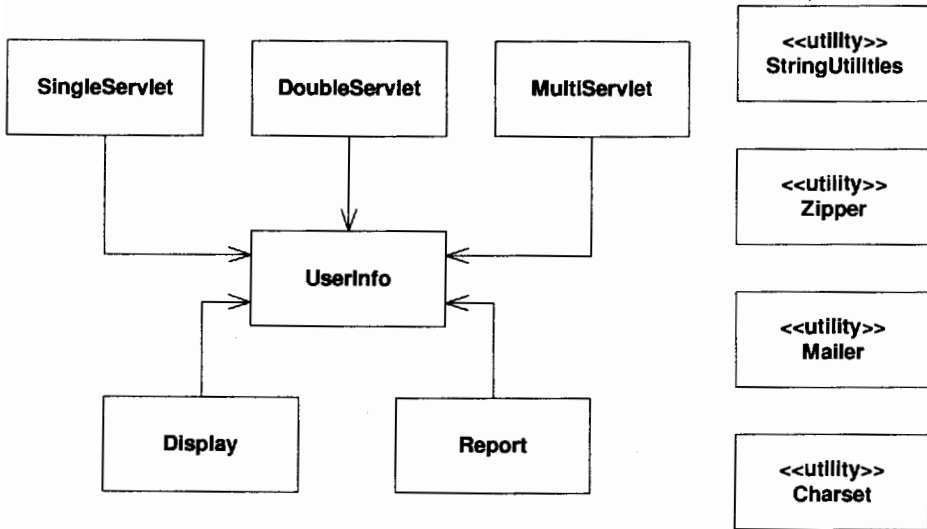


Figure 1 presents the UML class diagram for the user search subsystem.

The aim of the classes *SingleServlet*, *DoubleServlet* and *MultiServlet* is to interpret the queries issued by the HTML forms. Based on the data input through the HTML form, the search of the bibliographic database is carried out, the identifiers of the documents found are extracted and stored in the *UserInfo* class object connected to the current user session. The result of the execution which is presented to the user is the generated HTML page reporting the number of hits (documents).

From the HTML page reporting the number of hits (documents) the user can request the document preview. The servlet responsible for document preview is *Display*. This servlet supports two preview modes: *brief* and *detail*. The brief preview contains the basic data (author, title, publisher, place and year of publication), while the detail preview contains some additional data about the document (abstract, ISBN code, etc.). Initially, the user is in the brief preview mode, where he can navigate through the hit list spanning more than one pages (the number of hits that is shown on one page is limited). Each hit can be selected for the detail preview. In detail mode, the hits are also shown on more than one page and the navigation through the pages is also supported. Besides, the selection of hits for interchange purposes is supported. While moving through the hit list the user session is regularly updated.

The servlet *Report* performs copying of the documents. Based on the data from the user session and selected report format, the report is sent to the user by email service or by downloading the report file to the client computer. The file can be compressed as standard Zip format to speed up the process.

Since one of the available report formats is the UNIMARC format, this file can be used to insert the data into the new database. The process of bibliographic record interchange can be founded on this mechanism.

The class *StringUtilities* has certain number of methods for string manipulation. The class *Charset* has the methods converting the text coded using different coding layouts. The class *Mailer* implements the client side of the SMTP protocol in order to send the reports by email service. The class *Zipper* implements the algorithm for zip format file compression.

4. Implementation

The library information system WWW site is implemented in the Java environment on the basis of the model described in previous section. The servlet technology is used [7], as well as the Unicode standard for displaying multilingual text. The search is done on the database containing the bibliographic records which is also used by other modules of the system. Existence of some relational DBMS is assumed. The database is accessed by the servlet using standard JDBC interface. Communication with the database server is through SQL language. The architecture of the BISIS database is described in [4, 5, 8].

The WWW site is intended to serve the widest population of users. Along with the search and overview of the bibliographic records services, the site supports copying the records from the database for individual needs, as well as for storing these records in other databases.

The user enters his query through the fields of the HTML form. After entering the query, the database is searched and the search results are displayed. The query consists of one or more words which can contain the joker symbols. The allowed joker symbols are '?' (replacing single letter) and '*' (replacing the arbitrary character string). If the query contains more than one word, the search result will contain all the documents containing these words.

Regarding the search process, documents in the library system, are considered the structures consisting of one or more prefixes. Each prefix is determined by its name and can contain one or more contents. The contents of the prefix is some text. For example, AU is the prefix for the author(s). When requesting the search, the user has to select the prefix in which the given contents will be searched for.

After the search request was processed, the user gets the page (report) containing the number of records found. The user can get the list of these records by requesting the display option. One example containing the part of the search result list is shown in Figure 3. For each of the documents found one can get its detailed display containing some additional data about the particular document.

Library Information System search

Enter the search criterion

AU - Author	<input type="text"/>	AND
TI - Title	<input type="text"/>	AND
KW - Keywords	<input type="text"/>	AND
PU - Publisher	<input type="text"/>	AND
PY - Publishing year	<input type="text"/>	

Character set:

Figure 2. Search form page

Library Information System results

Document matches: 1 to: 20

Details	Retrieval	Documents
<input type="checkbox"/>	<input type="checkbox"/>	Surla, Dušan i dr.: <i>Planiranje trajektorije ravanskih mehanizama sa n stepeni slobode.</i> , [Novi Sad] 1989.
<input type="checkbox"/>	<input type="checkbox"/>	Kapor, D. V. i dr.: <i>Temperature Dependence of the Energy Gap of an Anisotropic Heisenberg Ferrromagnet in the-Complete Temperature Region.</i> PMF, Novi Sad 1985.
<input type="checkbox"/>	<input type="checkbox"/>	Surla, Katarina: <i>Numeričke i statističke metode u obradi eksperimentalnih podataka III.</i> Institut za matematiku, Novi Sad 1989.
<input type="checkbox"/>	<input type="checkbox"/>	Surla, Dušan: <i>Uvod u optimalno upravljanje i stabilnost dinamičkih sistema.</i> PMF: Institut za matematiku, Novi Sad 1985.
<input type="checkbox"/>	<input type="checkbox"/>	Surla, Katarina: <i>A Uniformly Convergent Spline Difference Scheme for</i>

Figure 3. Part of the search result list

Also, each record can be copied to the local computer by email or by downloading the file containing the selected records. The form used for copying the records is activated by the button 'CopyRecord' from the list page. The form is shown in Figure 4.

From Figure 4 one can see that two modes of record copying are currently supported: by email service or direct copy (download) of the file containing the records.

The image shows a screenshot of a Microsoft Internet Explorer browser window. The title bar reads "Search results - Microsoft Internet Explorer". The address bar shows "http://librv.in.ns.ac.yu/servlet/Display". The main content area features a logo for "Library Information System" and the word "results". Below this, there is a section titled "Get report about selected documents as:" with two radio button options: "e-mail: xxx@uns.ns.ac.yu" (which is selected) and "file". Underneath, there is a "Report type:" section with three radio button options: "Extended report" (selected), "XML (*)", and "UNIMARC". At the bottom of the form, there are three buttons: "Send report", "Selected documents", and "New search". A note at the bottom states: "(*) You can see Document Type Definition [here](#)." The browser's status bar at the bottom shows "Done" and "My Computer".

Figure 4. Record copy form

If the record copy by email is requested, one has to enter the email address in the corresponding field. As for the record copy option, there is also the possibility of selecting the of the records to be copied. Currently, three formats are supported:

- HTML format meaning the list of the records displayed the same way as they appear in the detail format on the WWW page of the BISIS;
- List of records in XML format;
- List of records represented by UNIMARC format.

If the user selects record downloading, the WWW browser is opening the window for selecting the file name and folder. The file is actually the Zip archive that can be unzipped by the appropriate software. The zip archive is used in order to decrease the file size and speed up the file downloading. The archive

consists of a single file containing the records in the desired format. After unzipping the archive, the file can be used for some other purposes.

If the selected format is UNIMARC, the file contains the records that can be imported to the database. Within the BISIS system importing is supported by the BISIS utility program IMPORT, which is part of the installation set.

XML as a format for record interchange is one possible solution to the problem of connecting the BISIS system to the systems having no direct support to the UNIMARC format. XML is general markup language for textual documents [9, 10]. This language is adopted by W3C consortium as one of the standards for structured documents interchange. The software for XML documents presentation is the part of the latest versions of the widely used WWW browsers (*Microsoft Internet Explorer 5, Netscape Navigator 6*).

5. Conclusion

The software system BISIS offers a flexible solution to the bibliographic records interchange between library institutions. Using public Internet access the user can find and copy desired records from the library databases. After copying the records he can easily add this record to the local database making it available for the later processing and search. The basis of the interchange is UNIMARC format. For the systems having no direct support to the UNIMARC format the option for record copying is based on some other standard formats. XML as an alternative format is getting wider support in the new versions of WWW browsers.

References

- [1] *UNIMARC Manual. Bibliographic format*, International Federation of Library Association and Institutions, IFLA Universal Bibliographic Control and International MARC Programme, New Providence, London 1994.
- [2] *The Unicode Standard*, Version 3.0, Unicode Consortium, <http://www.unicode.org>
- [3] Surla D., Konjović Z., Milosavljević B., Vidaković M., *Library Information System BISIS, ver 3.01*, 9th International Conference "Informatics in Education, Quality and Information Technologies", Zrenjanin 2000, pp. 494-504 (in Serbian)
- [4] Milosavljević, B., *Text server of UNIMARC Records*, M. Sc. Thesis, Faculty of Technical Sciences, Novi Sad 1999 (in Serbian)
- [5] Milosavljević, B., Konjović Z., *Search Specification in the Text Server for UNIMARC records*, Proceedings of Sym-op-is 99, Belgrade 1999 (in Serbian)
- [6] Milosavljević B., Konjović Z., *Architecture of the WWW Presentation for the Search in the Library Information System*, InfoScience, Beograd (in Serbian)(in press).
- [7] *Java Servlet Development Kit*, Version 2.2, Sun Microsystems, <http://java.sun.com>

- [8] Milosavljević B., Konjović Z., *Specification of the System for Indexing UNIMARC Records*, Proceedings of the YU Info 99, Kopaonik, Yugoslavia 1999 (in Serbian)
- [9] *Extensible Markup Language (XML) 1.0*, W3C Recommendation, WWW Consortium, <http://www.w3.org/XML/>
- [10] *XML Developer's Guide*, Microsoft Corporation, <http://www.msdn.microsoft.com/workshop/index/default.html>