Unicode Based Multilingual Catalogue Module: A New Feature of SOUL

Divyakant R Vaghela   Dinesh L Rayka   Yatrik Patel   Rajesh Chandrakar

Abstract

The INFLIBNET Centre has SOUL software developed for the automation and management of the Indian academic libraries, which has presence across the country in around 1500 libraries. INFLIBNET Centre being a nodal agency for Indian universities and keeping in view the requirement of the modernization of the their libraries with ICT facilities, Centre has developed this software and made available to the libraries in the nominal cost. With the development of the technology and the requirement changes of the library with the technology, it was due for the INFLIBNET Centre to make some necessary changes with the SOUL software. This paper introduces the new catalogue module of the SOUL software with the very new features such as MARC21, Unicode and MARCXML etc abreast with the new applications of the ICT. The Centre happily announces the new name of the SOUL software from the "Software for University Libraries" to the "Software for Library Automation and Management".

Keywords: INFLIBNET, SOUL, MARC21, Unicode, MARCXML, OPAC

1.   Introduction

SOUL (Software for Library Automation and Management), formerly it was known as Software for University Libraries, is integrated library management software developed by the INFLIBNET Centre, Ahmedabad. The software adheres to most of the international standards used in the library management software. The software has six modules i.e. Acquisition, Cataloguing, Circulation, Serial Control, OPAC and Administration including the WebOPAC.[1] It is developed based on the requirements of the Indian academic universities and mostly suitable for the Indian academic system including the college and technical R&D institution's library. At the moment, the software has made presence at more then thousand libraries across the country including few libraries at SARC countries.

Keeping in view the requirements of the Indian academic libraries, where library has collection of different Indian languages. According to the collections of the library, the automation software must have the feature of cataloguing the Indian languages documents. Not only the software should accommodate the Indian collections but also it should accommodate the foreign language collections as most of the academic libraries of the country have instances of foreign collections. Thus, the INFLIBNET Centre stressed upon this important requirement of the software package and moved towards the technology that incorporates this feature. In the beginning, at the launch of the first version of SOUL, ISCII (Indian Script Code for Information Interchange) based product ISM (Indian Script Manager) developed by C-DAC, Pune was integrated with the SOUL software for cataloguing

of the Indian language documents. The utility is not a full-fledged solution for multilingual which had number of drawbacks. After the implementation of the new technology i.e. Unicode, the drawbacks that was available with the ISM based catalogue have been shorted out. This paper introduces the new catalogue module of the SOUL software developed using Unicode for multilingual support and describes about the features of the module.

2.   What is Unicode?

The Unicode is a multilingual standard for international encoding of the characters, unanimously adopted by the industry leaders on the behest of the Unicode consortium, which was incorporated in January 1991, under the name Unicode, Inc. It provides a unique number for every characters, no matter what the platform, no matter what the program, and no matter what the language is.[2] Unicode is not hardware or software, it is a formal standard. Unicode enables a single software product or a single website to be targeted across multiple platforms, languages and countries without re-engineering.[3] It allows data to be transported through many different systems without corruption. Unicode has three encoding systems i.e. UTF8, UTF16 and UTF32 and the features of the Unicode are[3]:

- ♦ **No overlap**
- ♦ **Determinate boundaries**
- ♦ **Pass-through**
- ♦ **Simple iteration**
- ♦ **Slow indexing**
- ♦ **Frequency**

Unicode 5.0.0 is the latest version of the Unicode Standard, which supersedes all previous versions.

2.1   Features of Unicode 5.0.0

i)   The Unicode Consortium has announced the release of a significant update of its widely-used Unicode Character Database (UCD). Unicode 5.0.0 is a major version of the Unicode Standard and supersedes all previous versions.

ii) The new version, Version 5.0.0, defines more than 99,000 characters for the languages of the world, and provides the detailed properties needed for computer software implementations. This latest level of the UCD contains all the information needed to update software to support the characters and algorithms that are the foundation for all modern computer programs — including the latest data for Unicode security mechanisms, collation, and locales.

iii) 1,369 new character assignments have been made to the Unicode Standard, Version 5.0 (over and above what was in Unicode 4.1.0). These additions include new characters for Cyrillic, Greek, Hebrew, Kannada, Latin, math, phonetic extensions, symbols, and five new

scripts: Balinese, N'Ko, Phags-pa, Phoenician, and Sumero-Akkadian Cuneiform.

iv) Unicode Character Database now includes Graphic (98,884), Format (140), Control (65), Symbols for private Use (137,468), Surrogate (2,048), Noncharacter (66) and Reserved (875,441).

More information about the Unicode and this version is available at the website on http://www.unicode.org/.

3.   SOUL Catalogue Module

The Catalogue Module is one of the modules of the SOUL software which has five more different modules integrated in it those are Acquisition, Circulation, Serial Control, OPAC and Administration. The Catalogue module is used for retrospective conversion of books; technical processing of books received from Acquisition Section, printing a range of records for verification, searching by title and accession numbers, authority files for publishers etc. One of the unique features of SOUL is accessing to authority files on screen as well as selection box in various fields of records. Once the proposed VSAT Network by INFLIBNET is commissioned this module will take care of automatic replication of data into union catalogue. This will avoid exchange of library data on physical media i.e. floppy, tape etc.

This module covers the following areas of the library functions:

- ◆ Cataloguing Process
- ◆ Data Entry
    - • Predefined data entry or Template based data entry of Books, Serials etc.
- ◆ Catalogue Search
    - • Simple Search for a particular Title, Author or subject wise entry
    - • Boolean Search which displays the records by giving more than one parameter
    - • Advance Boolean Search which searches the records having particular tag entry
- ◆ Data Export/ Import
    - • Data Export from MARC 21 format to database
    - • Data Import in text file form the database in MARC records
    - • Data convert from CCF Standard to MARC Standard and get it into database
- ◆ User Services
    - • Current awareness services
    - • Bibliographic services

3.1 Silent Features of Catalogue Modules

The catalogue module of the SOUL has got various unique features, which are popular to its users, those are:

1) Easy to use and menu driven windows
2) Inbuilt OPAC
3) Provides user reports
4) Evaluates staff performance
5) In-built barcode generation features, it does not require separate software and separate printer
6) Provides Bibliographic and Current Awareness Services
7) Prepares Catalogue card
8) Display has got very unique feature of tag wise display, details display and catalogue card display.
9) Adheres to the CCF and MARC 21 bibliographic standards including the exchange of records in to ISO2709 format.
10) Accommodates record of different formats through export and import facility of the software
11) Supports cataloguing of electronic documents, non book materials, theses and dissertations of various degrees including the magazines and the newspapers.
12) Facilitates stock verifications
13) Maintains various authority files
14) Features WebOPAC

Apart from those described feature, recently following new features have been added to the Catalogue module of the software, those are:

♦ **Unicode Based cataloguing of the document**: it gives facility to catalogue any kind of documents available at the library including non-book materials and electronic documents in any languages available in the world. It is implemented based on the Unicode based multilingual feature available with the Operating Systems. An example of the multilingual catalogue is shown in the fig. 1.

♦ **Full MARC 21 Support**: full features of the MARC21 has been implemented from the website of Library of Congress http://www.loc.gov/marc/.

♦ **Database Independent:** it is open for any database MSSQL, ORACLE or FOXPRO for this application.

♦ **Template Based Entry:** The module has default template prepared for cataloguing the various documents of the library collection. Most important feature of the module is that it provides platform to the users for creating their own templates based on their requirements

of the fields to be entered. This can be prepared as many as user wants. Fig.2 is the screen shots for the template preparation;

◆ **Supports ISO2709 format for data exchange**: data can be exported into ISO 2709 format and can be imported from ISO2709 format. It has feature of accepting data from CCF based ISO 2709 format.

◆ **Adoptability:** data transfer from old SOUL to new SOUL has been taken care with 100% data transfer.

◆ **Customized User Report:** The report of the catalogue module is empowered to the users. The user can generate report as per their choice. It is so simple that user can customize it as per the library requirement.

Fig.1: Screen Shots for the Multilingual Catalogue Entry
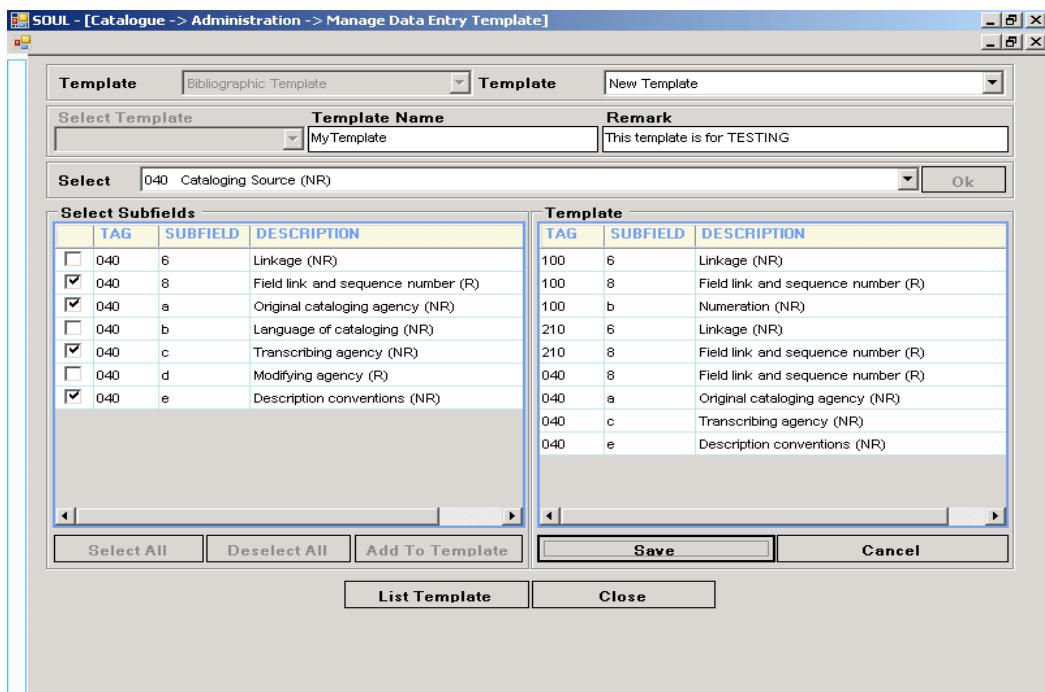
Fig.2: Screen Shots for the Template Preparation

◆   **OPAC Search**

There are three options for OPAC (Online Public Access Catalog) search.

1. **Simple Search**

   In this Simple Search, User can search by any one parameter like Title, Author, Title + Subject, Title + Author, Publisher, Accession No., Corporate Name, Meeting Name, Subject. The research of the search is available in BASIC, MARC and FULL records display formats.

2. **Boolean Search**

   In this Boolean Search, there is a facility to search by any three parameters given in a simple search using AND, OR and NOT queries and display of records is available in BASIC, FULL and MARC format.

3. **Advanced Boolean Search**

   In Advanced Search, there is a facility to search records using three parameters i.e. AND, OR and NOT, where tags of MARC are used as parameters.
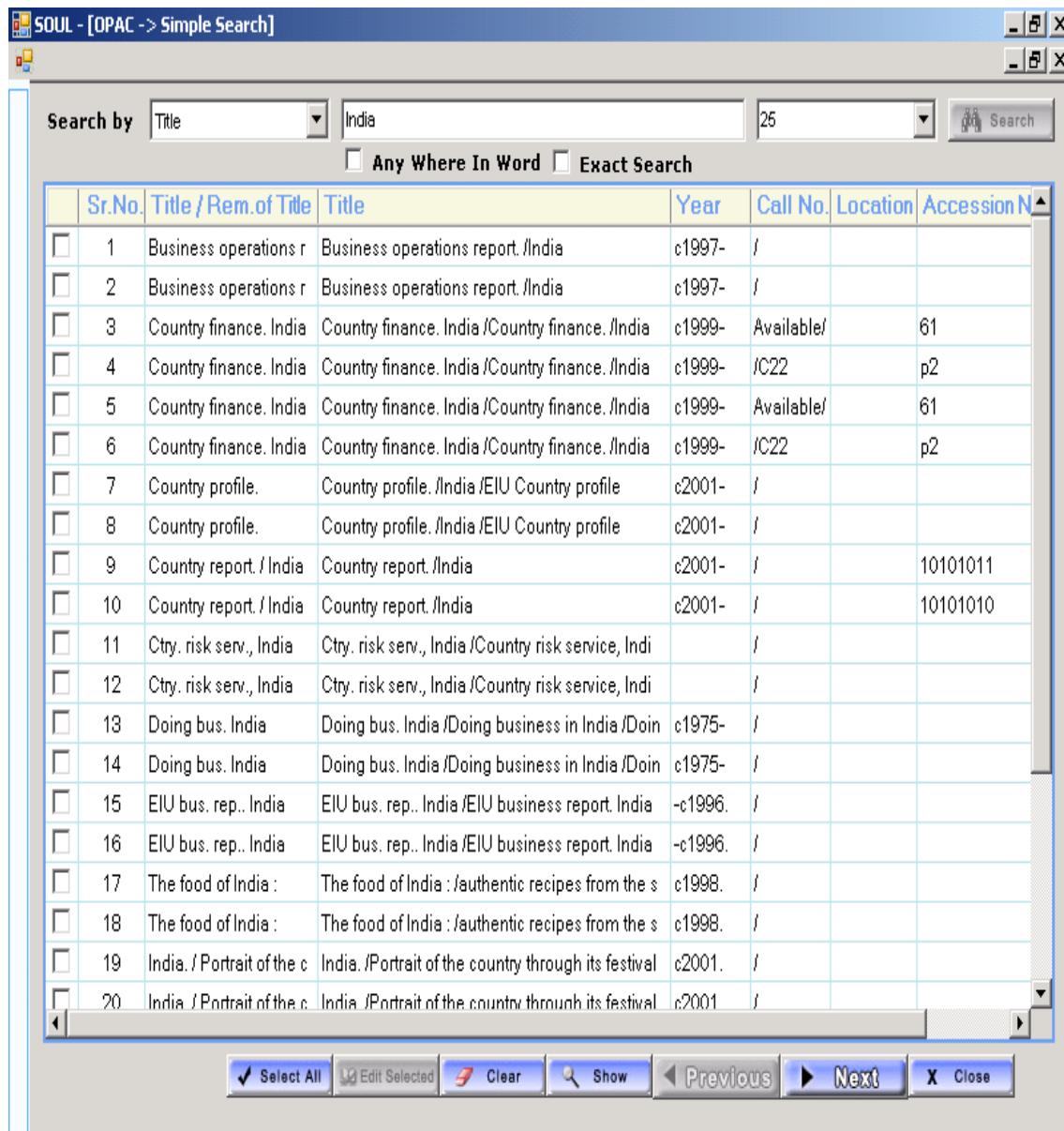
Fig. 3: Screen shots of the OPAC Catalogue Search

◆ **Export to MARCXML**

The new catalogue module of the SOUL has the feature of transferring data in to MARCXML. The development of an XML version of MARC21 was critical for format

interchange. The economically deep commitment to MARC data elements, proliferation of schemas beyond the library community control, and the rapidly growing XML tool environment mandated as evolutionary path into XML for MARC 21. And the team of the SOUL development succeeded in doing it. The structure of the MARCXML schema is given below;

1. The highest-level elements are <collection> and <record>, to enable a group of records to be assembled into a package and to clearly define a single record.

2. Field tags and indicators that are found in MARC 21 are treated as attributes. This allowed the first level number of elements to be radically reduced in MARCXML. Only 3 basic format related elements are needed: <leader>, <controlfield> and <datafield>. The example of a data field is describes as below:

    Example: <datafield tag="245" ind1="1" ind2="0">

(3) A <datafield> is described with the <subfield> (which is repeatable for each subfield within the field), and <subfield> has an attribute for specifying the subfield code for different subfields available within the field. An example is given below:

    Example: <datafield tag="082" ind1="0" ind2="0">

    <subfield code="a">796.6/4/0943</subfield>

    <subfield code="2">20</subfield>

    </datafield>

(4) The content of each <controlfield>, MARC tags, 001-009, is treated as a string. It means that the 008 tag of the MARC21, for which coded elements are defined by character position till 40 characters. The full 40 bytes of the 008 are transformed to XML including blanks. Fortunately XML has a simple schema specification that keeps blanks from being compressed or otherwise tampered with, essential for treating some control fields as strings. An example of control field 008 is described here:

    Example: <controlfield tag="008">931129s1994 wauab 001 0 eng </controlfield>

(5) The <leader> of the MARC record is also treated as a string like control fields. Some of the information in the leader is only relevant to an ISO 2709 record; it is simply carried over where it is easy to discard in subsequent transformations. An example of leader is given below:

    Example: <leader>00637cam 2200193 a 4500</leader>

An Example of Full MARCXML Record

A multilingual example of full record in MARCXML format is given below

```xml
<?xml version="1.0" encoding="utf-8" ?>
- <marc:collection xmlns:marc="http://www.loc.gov/MARC21/slim" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
  - <marc:record>
     <marc:leader>00749cam a2200241u 4500</marc:leader>
     <marc:controlfield tag="001">9835185</marc:controlfield>
     <marc:controlfield tag="005">20040714172412.0</marc:controlfield>
     <marc:controlfield tag="008">080205s1948 xx 000 0 HINo</marc:controlfield>
   - <marc:datafield tag="000" ind1="" ind2="">
       <marc:subfield code="">00749cam a2200241u 4500</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="001" ind1="" ind2="">
       <marc:subfield code="">9835185</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="005" ind1="" ind2="">
       <marc:subfield code="">20040714172412.0</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="008" ind1="" ind2="">
       <marc:subfield code="">080205s1948 xx 000 0 HINo</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="010" ind1="" ind2="">
       <marc:subfield code="a">sa 61000083</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="035" ind1="" ind2="">
       <marc:subfield code="9">(DLC)sa 61000083</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="040" ind1="" ind2="">
       <marc:subfield code="a">DLC</marc:subfield>
       <marc:subfield code="c">CarP</marc:subfield>
       <marc:subfield code="d">DLC</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="043" ind1="" ind2="">
       <marc:subfield code="a">a-ii---</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="050" ind1="0" ind2="0">
       <marc:subfield code="a">DS407</marc:subfield>
       <marc:subfield code="b">.G316</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="100" ind1="1" ind2="">
       <marc:subfield code="a">गान्धि</marc:subfield>
       <marc:subfield code="d">01-01-1883</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="245" ind1="1" ind2="0">
       <marc:subfield code="a">आदर्शू भारत कि स्परेखा</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="260" ind1="" ind2="">
       <marc:subfield code="c">[1948]</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="300" ind1="" ind2="">
       <marc:subfield code="a">p. cm.</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="500" ind1="" ind2="">
       <marc:subfield code="a">हिन्दी</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="651" ind1="0" ind2="">
       <marc:subfield code="a">भारत</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="700" ind1="1" ind2="">
       <marc:subfield code="a">प्रभु, रामाचन्द्रा किश्ना,</marc:subfield>
       <marc:subfield code="d">01-01-1883</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="906" ind1="" ind2="">
       <marc:subfield code="a">0</marc:subfield>
       <marc:subfield code="b">cbc</marc:subfield>
       <marc:subfield code="c">premunv</marc:subfield>
       <marc:subfield code="d">u</marc:subfield>
       <marc:subfield code="e">ncip</marc:subfield>
       <marc:subfield code="f">19</marc:subfield>
       <marc:subfield code="g">y-gencatlg</marc:subfield>
     </marc:datafield>
   - <marc:datafield tag="991" ind1="" ind2="">
       <marc:subfield code="b">c-GenColl</marc:subfield>
       <marc:subfield code="h">DS407</marc:subfield>
       <marc:subfield code="i">.G316 (Orien Hind)</marc:subfield>
       <marc:subfield code="t">Copy 1</marc:subfield>
       <marc:subfield code="w">??????</marc:subfield>
     </marc:datafield>
   </marc:record>
</marc:collection>
```

4.   Conclusion

The Software for Library Automation and Management, the new version of the SOUL software hopes to solve all the problems of existing users of the SOUL catalogue module and look forward to have a concrete feedback from the users for better improvement.

References

1.   SOUL (Software for Library Automation and Management). Available at http://www.inflibnet.ac.in/soul/soul.htm (Accessed on 03/02/2008)

2.   What is Unicode. Available at http://www.unicode.org/standard/WhatIsUnicode.html (Accessed on 01/02/2008).

3.   Chandrakar, Rajesh. Multi-Script Bibliographic Database: an Indian perspective. Online Information Review. West Yorkshire, England: MCB University Press, 2002,Vol. 26, no. 4. P. 246–251.

About Authors

Mr. Divyakant R Vaghela, Project Assistant, INFLIBNET Centre, Ahmedabad.
E-mail: divyakantbece2020@gmail.com

Mr. Dinesh L Rayka, Project Assistant, INFLIBNET Centre, Ahmedabad.
E- mail: rayka_dinesh@yahoo.co.in,

Mr. Yatrik Patel, Scientist B (CS), INFLIBNET Centre. Ahmedabad
E-mail : yatrik@inflibnet.ac.in.
Mr. Rajesh Chandrakar, Scientific & Technical Officer, INFLIBNET Centre, Ahmedabad.
E-mail: rajesh@inflibnet.ac.in