Softwares for Cataloguing Digital Resources with Special Emphasis on MARCit

T S Prasanna & Madhuresh Singhal
National Centre for Science Information
Indian Institute of Science
Bangalore –12.
Email: prasanna@ncsi.iisc.ernet.in & madhuresh@ncsi.iisc.ernet.in

Abstract

This paper gives an overview of a few Cataloguing Softwares, which are capable of cataloguing digital resources. Discusses usefulness of the software for cataloging digital resources, availability of the software, their features in terms of platforms, access cost, supported modules and their limitations. The study has been made based on the testing of the software and their usage. MARCit is one such software for cataloguing Internet resources with lots of features like user-friendly interfaces, use of MARC format for cataloguing, interoperable with other local library softwares. It supports USMARC, can be integrated with any browser, and can also work with ftp sites. It does automatic filling of certain fields in the catalogue based on the browser properties. We have selected and reviewed some other software also. These are particularly useful to catalogue different types of resources. They enhance the search, browse and retrieval capability of the catalogued resources as well as provide single click access to all the information and also can be used for printing spine labels, catalogue cards, etc. Also they can be used as a stand-alone or they can be combined with the existing library packages.

Keywords

Cataloguing Software, MARCit, All Media Library, MARC Format, Digital Resources, Internet Resources.

Introduction

Historically, librarians have organized the world's information. For centuries, they have successfully managed, classified, and filtered information of many types by creating surrogates. This is true of traditional materials as well as new electronic resources. But now as the amount of accessible electronic information increases, the cost of accessing this information will increase and the communities unfamiliar with library science are beginning to grapple with the problem of metadata and the organization of large collections of data. So there has been a general push to apply and develop techniques to make these resources searchable and more widely accessible. Now, the requirement is that every electronic item should have a catalog entry or its equivalent. But all electronic resources can never be humanly cataloged. It's just too expensive and also librarians are

so overburdened that they can barely keep up with their traditional workload, let alone begin to catalog and organize the vast amounts of information available electronically. Clearly, automated tools to apply library science ideas like classification and filtering to electronic resources at high speed and low cost are needed.

The whole scenario of information management can be divided into two worlds. These two worlds -- the seemingly unorganized Web and the organized world of libraries -- have much to offer one another. The Web can offer automated tools for searching raw information and the library world can offer experience organizing and understanding information of all types. By combining their talents and techniques, these two communities can bring powerful resources to bear on the problems of accessing, maintaining, and supplying electronic information. But this is not an easy task. Even though using latest technologies, searching the raw content of every document still seems to be almost impossible since it's not uncommon to retrieve hundreds of documents for a given search.

Again the question arises is, if it is so easy to use software tools for all these purpose, why we were not using it till yet. The main cause behind it may be the development of standards and the creation of concepts. These standards and concepts have only just become stable enough to begin to use. So this is the right time to start using the software for this purpose. This article is an attempt to make aware the library community about the softwares with their features and to use them effectively.

MARCit (http://www.marcit.com/)

MARCit is an aid to easy cataloging of web sites, working with a web browser. Cataloging a site involves viewing the site with the browser, clicking on the MARCit icon, adding other information to the cataloging screen, saving the cataloging to a file, and importing the file into a library automation system. A user's manual, which can be downloaded, is very helpful and written so that it can be understood by a non-technical person. The web browser opens when MARCit is opened. Then, while viewing a web site, one can click on the MARCit icon. The URL and title information appear in the MARCit screen directly from the web site. Several fields are cataloged automatically; other fields can be filled in manually. The software allows the cataloger to rate the site. Different records can be saved to different files. A saved file can be emptied, renamed, and deleted without leaving MARCit. The method used for importing the file into the library's online catalog depends on the automation system used.

After cataloguing Internet resources, if the user searches for items on "Monalisa" for example, s/he might find a web site dedicated to Monalisa's art. The user can then access the site and get the information needed. The user doesn't have to wade through the

millions of hits a search on Monalisa might yield with a search engine. MARCit makes cataloging such sites simple. The librarian who catalogs it can also rate each MARC record. Users seeking the best information available can use the ratings as a valuable guide to quality sites. MARCit has been designed taking into consideration that accessing the information would be easier and more efficient if a librarian has designed it.

How MARCit Works

1. Open the site to be catalogued in the browser.
2. Click on the floating MARCit icon.
3. MARCit automatically grabs the site's title and URL from the browser and enters them into the spaces provided. Then catalog entries for all the fields should be entered in the space provided, including the mandatory MARC fields.
4. Enter as much or as little information into the MARCit fields.
5. Click the Save button. A MARC record of the site is automatically created and saved into a file that can be imported into library automation system database.

Fields Catalogued

The fields that can be catalogued are

LDR	Leader
001	Control Number
003	Control Number Identifier
005	Date and Time of Last Transaction
007	Physical Description Fixed Field
008	Fixed-Length Data Elements
040	Cataloging Source
245 a	Title Proper
245 h	General Material Designation
256 a	File Characteristics
514 a	Data Quality Note
516 a	Note: Type
538 a	System Detail Note (Mode of Access)
856 u	Uniform Resource Locator

MARC Format Adherence

MARCit adheres to the July 1997 Library of Congress MARC Standards. The comprehensive MARC records it generates comply fully with these standards and include all mandatory fields. Indicators are set automatically. MARCit even sets whether the site

is a web site (http) or an FTP site automatically in the 856 fields, in accordance with the MARC Standards. MARCit also automatically handles one main author entry, which can be either personal or corporate, but not both, and several other author entries - both personal and corporate. MARCit also sets the correct MARC fields, subfields and indicators for different types of subjects and subject divisions. It automatically checks ISBNs entered. It makes sure whether title is entered. MARCit does all of this automatically, ensuring that the records are of the highest quality.

Hardware Requirements

Computer running Windows XX or NT, 5 MB hard drive space, 16 MB RAM, Netscape Navigator v.3.0 or later OR Microsoft Internet Explorer v.3.0 or later and Internet connection.

Features

- MARCit will work with any library automation system that can import USMARC records from a PC-based hard or floppy disk.
- One can specify Dewey, LC, or local as primary call number. One can also choose to catalog using all three call numbers. These and other default parameters can be changed at any time.
- The records MARCit creates fully conform to the MARC standard incorporating Update No. 3 (ISBN 0-8444-0941-3), dated July 1997, to the USMARC Format for Bibliographic Data (ISBN 0-8444-0809-3).

MARCit is commercial software but the evaluation version is also available. We have tested the evaluation version (7.11) but the limitation is number of records. It is limited by ten records. After saving 10 records, one can no longer create and save MARC records with MARCit. The MARCit evaluation copy is approximately 2 Megabytes in size. The screenshot of front screen is as given below.

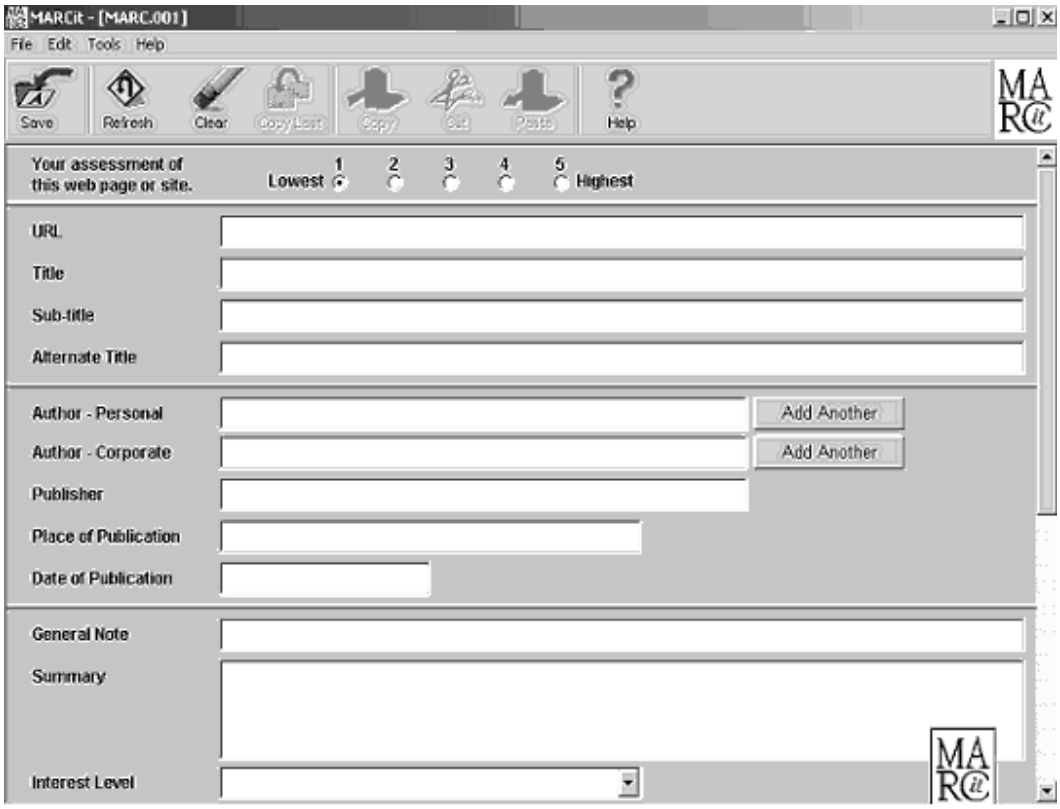

Fig. 1 – Screenshot of MARCit

All Media Library 7.11 (http://www.Latshaw.com)

File Size: 6.3 MB
License: Shareware
OS: Win95, 98,NT, 2000

All Media Library (AML) is a database for tracking all forms of media such as CDs, Videos, Books, etc. The program catalogs all library items and features a powerful search facility enabling us to keep track of people who borrow from our library, know what items are on loan, who has borrowed them, when they were borrowed, and who borrowed last.

Built in inventory reports allow to know exactly what is available in the library. AML will run on a network and also works with barcode readers. AML is being used by home collectors; major corporations like Ford Motor Company, Memorex, Digital Equipment, and even the U.S. Dept. of Interior.

Features of the latest version:

- Added ability to print labels for select records only in addition to the entire list.

- In addition to storing all of the media records in the library database, AML also stores other lists of information like the list of borrowers.
- When copying and pasting records, AML no longer fills the entire fields with extra spaces.
- It uses inbuilt Microsoft Jet Relational Database and the Library size is limited to 1 GB.
- About 255 simultaneous network users can edit a given library at a time.

Hardware Requirements

- PC running Windows 95/98/NT4/2000 or later operating systems
- 24MB RAM (128 recommended)
- 10MB available Hard Drive space (more for larger collections, 1GB max library size)

Recommended for speed enhancement of large libraries (over 20,000 titles)

- 300Mhz or faster Pentium, Pentium Pro, Pentium II, or higher processor
- 128MB RAM or more
- 2GB free hard drive space (room for max library size and room for compacting)

Network Requirements (Does not apply to use on a single machine)

- 10Mbps(or faster) Network with file share ability (client/server or peer to peer)
- Running either TCP/IP, IPX, or NetBEUI protocol

Fields covered in AML

TITLE: The title of the item limited to 255 characters.

ARTIST (S): The artist field is used for recording the artist, creator, manufacturer, etc. This field is limited to 255 characters.

CATEGORY: The category field is used for placing the item into a category. If one can't find an accurate category in the list, they can edit the list of categories at anytime by clicking the Media Types/Categories button on the toolbar.

MEDIA TYPE: The media type field is for recording & displaying the type of media (Tape, CD, etc.). If there is any Media Type that is not in the list, one can edit the list of media types by clicking the Media Types/Categories button from the toolbar at the top.

***CATALOG NUMBER*:** Catalog Number or Barcode Number can be used here. This field is limited to 255 characters.

MEDIA DATE: The Media Date field is used to represent the creation or copyright date of an item.

PIECES (Formerly "Quantity in Set"): If there is a title with more than one piece of media holding it, enter the quantity here. Note: If there is multiple **copies** of a title, make a separate record for each, assigning a unique catalog number to each.

VALUE: The value field is where one can enter what the item is worth. This field is limited to numeric values.

LOCATION: The LOCATION field is used for recording WHERE the item is physically located.

INDUSTRY RATING: This field is used primarily for movies or videos that have been rated with an industry rating.

PERSONAL RATING: The personal rating field can come in very handy.

NOTES: The final, and probably most useful field is the notes field. This is where one can place detailed information about the items.

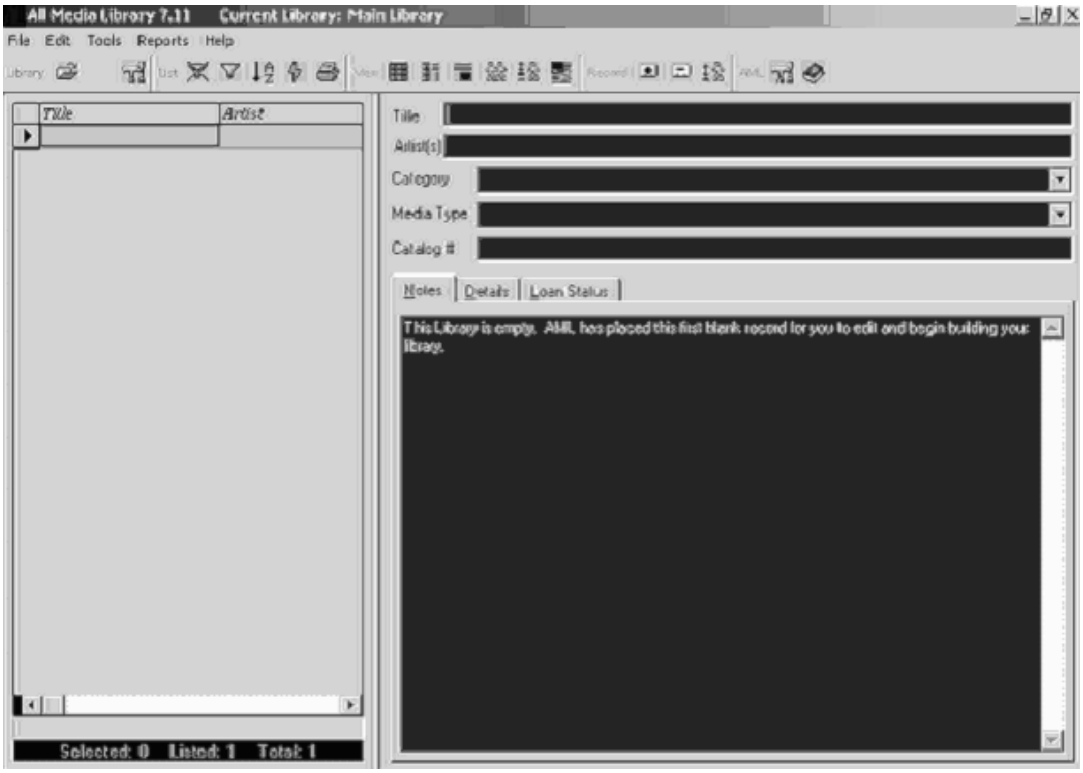

Fig.2 – Screenshot of All Media Library

List of Other Softwares/Services

1. **Klarity Software (http://www.klarity.com.au/)**
   Klarity is a commercial software program developed to automatically categorise documents based on the concepts found in the text. It can analyse and process large collections of documents and generate meta-data to assist in the categorisation and locating of information. The output can be customised to various formats including META tags, RDF and ASCII output for upload to databases. Klarity has been developed so that it can run on multiple platforms viz. Linux, Solaris and Windows NT.

2. **Scorpion Project (http://orc.rsch.oclc.org: 6109/)**
   Scorpion is a project of the OCLC Office of Research exploring the indexing and cataloging of electronic resources. Since subject information is key to advanced retrieval, browsing, and clustering, the primary focus of Scorpion is the building of tools for automatic subject recognition based on well-known schemes like the Dewey Decimal System**.**

3. **Dublin Core Metadata Editor (http://www.ukoln.ac.uk/metadata/dcdot/)**
   This service will retrieve a Web page and automatically generate Dublin Core metadata, either as HTML <meta> tags or as RDF/XML, suitable for embedding in the <head>...</head> section of the page. The generated metadata can be edited using the form provided and converted to various other formats (USMARC, SOIF, IAFA/ROADS, TEI headers, GILS, IMS or RDF) if required.

4. **MANTIS (http://orc.rsch.oclc.org:6464/)**
   MANTIS is a research toolkit developed at OCLC for building arbitrary Web-based cataloging systems. Mantis has been packaged for external use in SiteSearch Release 4.1**.**

5. **The MetaWeb Project (http://www.dstc.edu.au/Research/Projects/metaweb/)**
   The aim of the Metadata Tools and Services project - known as MetaWeb - is to develop indexing services, user tools, and metadata element sets in order to promote the use of, and exploitation of metadata on the Internet. The project partners have committed to a broad set of activities to achieve this aim.

Conclusion

As the amount of accessible electronic information increases, the cost of accessing this information will increase. That is, even though users can now use free search services to find items of interest, they will increasingly spend their valuable time wading through masses of irrelevant documents to get the information they need. So there is a requirement of cataloguing of Internet resources in such a way that without wasting the time, one can find the high quality information easily. However one can't deny the fact that software can be of great help in managing the Internet resources but they cannot replace human cataloguing. There are many aspects of human cataloguing that are difficult if not impossible to automate. We just need to have a balance in both aspects. So finally we can say that the software tools can work as an aid in cataloguing and managing the Internet resources.

References

1. http://www.marcit.com/
2. http://www.latshaw.com
3. http://www.klarity.com.au/
4. http://orc.rsch.oclc.org: 6109/
5. http://www.ukoln.ac.uk/metadata/dcdot/
6. http://orc.rsch.oclc.org:6464/
7. http://www.dstc.edu.au/Research/Projects/metaweb/
8. http://www.dublincore.org/