Classification of Library Documents by Using the Web

Badan Barman

Krishna Kanta Handiqui State Open University, Dispur, Guwahati - 781006, Assam, India

Email: badanbarman@gmail.com

Abstract

The automatic classification of library document by using the computer is still far away from being implementation. But, the emergence of web and merging of different databases of different libraries paved a new way for classification of library document. The study only includes the web tools that can be used freely or does not required user to "Sign Up" or "Log in". Sometime, some search in such database or tools will produce null results. In such cases, the classifier should turn his hand to the next option available. In case of doubt of the validity of the classification number obtained from the online tools, the classifier should verify the number by using the original DDC scheme. This paper will depicts such online tools that can be used for classification of the library document with efficiency.

Keywords: Automatic Classification, Classify, Dewey Browser, Dewey Decimal Classification, Library Classification, Library of Congress Online Catalogue, Web Classification

1. INTRODUCTION

The library classification system provides a system for organizing the knowledge embodied in books, CD, web, etc. It supplied a notation (in case of DDC, it is Arabic numerals) to the document. The Dewey Decimal Classification (DDC) system is a general knowledge organization tool that is continuously revised to keep pace with the development of knowledge. It is the most widely used classification scheme in the world. Libraries in more than 135 countries use the DDC to organize their collection. It is also used over the web for organizing the web resources for the purpose of browsing.

The DDC scheme was conceived by Melvil Dewey in 1873 to arrange the documents of the Amherest College Library. The first edition entitled "A Classification and subject index for cataloguing and arranging the books and pamphlets of a library" was published in 1876. It appeared in the form of a small book of 44 pages. The Decimal Classification Editorial Policy Committee (EPC) was established in 1937 to serve as an advisory body to the Dewey Decimal Classification. In 1988, Online Computer Library Center, Inc (OCLC) acquired the DDC. The editorial headquarters was located at the Library of Congress in the Decimal Classification Division. The editors prepare the proposed schedule revisions and expansions, and forward the proposals to EPC for review and recommended action. Nowadays, DDC is published by Online Computer Library Center, Inc in full and abridged editions. The abridged edition targets the general libraries having less than 20,000 titles. Both the full and abridged editions are available in print as well as in electronic version.

2. CLASSIFICATION OF DOCUMENTS BY USING THE WEB

The cost of DDC is very high. Every library in India and in other developing countries cannot afford to have a set of DDC as its own. But the classification of the documents in a library is a must. To meet this end, librarians can use some tools and techniques to have a class number of a document they have procured in their library. There are some excellent tools over the web that shares the class numbers. Some of these tools and techniques are discussed below. They will provide the readymade class number of a document and will save the time of the classifier. We may not require to follows these options if we have a set of DDC. We are to only

follow the options listed bellow in the event of not having a set of DDC. We can also follow these options to verify the class number obtained by consulting the DDC on our own.

3. CLASSIFY: AN EXPERIMENTAL CLASSIFICATION WEB SERVICE

OCLC Research experimental classification service launched "Classify" (http://classify.oclc.org/classify2/) which is targeted to support the assignment of class number and subject heading by using the web. The interface can be used both by a machine as well as human being. It provides access to more than 36 million collectively built records from a large pool of related resources. Each record in the database contains Dewey Decimal Classification (DDC) numbers, Library of Congress Classification (LCC) numbers, or National Library of Medicine (NLM) Classification numbers, and subject headings from the Faceted Application of Subject Terminology (FAST).

In the database of Classify (http://classify.oclc.org/classify2/) by inputting any one or in combination of some basic information related to the document, the class number or subject heading can be obtained. The inputted information may be of the following types:

i   ISBN: The Classifier can use the 10 or 13 digit ISBN. The ISBN should be used without hyphens in between. One can find more about ISBN over: http://www.isbn-international.org/

ii OCLC: Each bibliographic record in the WorldCat has a unique number that range from 1 to 9 digits in length. The classifier can also use this number to find out the information from the database. More about OCLC # is available over: http://www.worldcat.org/links/default.jsp

iii Barcode / The Universal Product Code (UPC): The classifier can use the 12 digits UPC number found in the document. One can know more about Barcode over: http://www.gs1us.org/

iv International Standard Serial Number (ISSN): The classifier can use the eight digits ISSN with or without hyphen (as it is appeared in the document). One can know more about ISSN over: http://www.issn.org

v   Title and / or Author: The classifier can also use full title of the document or some portion of it or its author or both the title and the author as a combined search.

vi Faceted Application of Subject Terminology (FAST): The classifier can also use the FAST controlled vocabulary that is based on the Library of Congress Subject Headings (LCSH). One can collect more information about FAST over: http://www.oclc.org/research/activities/fast/

If one go to (http://classify.oclc.org/) web address and enter the ISBN / ISSN or any standard number correctly in the interface it sometimes shows a "No data found for the input argument" error. But, if one use the title and some portion of the authors' name of the same document it shows the result. It happens probably because sometimes people perhaps do not enter those fields in the records of the database, while preparing it. Entering some portion of the title and the first author's surname (or sometimes the forename) of the document in the interface mostly leads to the relevant document and class number. The classifier can use this option as the first approach to obtain the class number of the document or its subject heading.
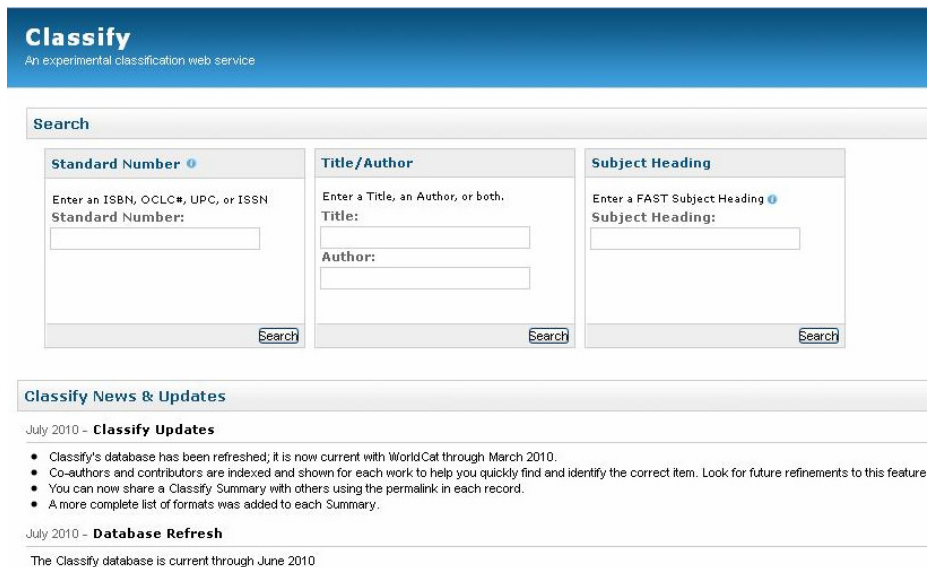
Fig. 1 Home page of classify of OCLC

4. DEWEYBROWSER

The DeweyBrowser (http://deweybrowser.oclc.org) provides access to approximately 2.5 million records from the OCLC Worldcat database. The DeweyBrowser helps to access those resources. The classifier can also use this interface to obtain readymade class number of a document in the library. Just the need is to make a search by entering the complete title of the document in the search box of the site to have its class number.
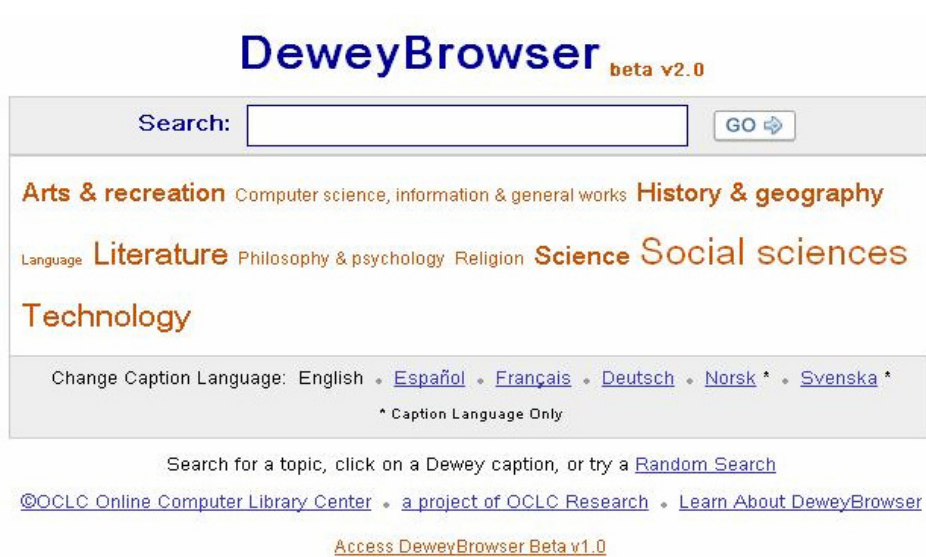

Fig. 2 Home page of DeweyBrowser

5. ISBNDB.COM

ISBNdb.com (http://isbndb.com/) is a database of books that is built by taking data from hundreds of libraries across the world. It is developed by Andrew Maltsev. He has also a company named Ejelta LLC (http://ejelta.com/), based in San Gabriel, CA. This ISBNdb.com is one of the outputs of the company. The classifier can enter the keywords, book title, author, publisher, topic or ISBN of the document in its search box to have its class number. After displaying the result by the interface, clicking on the most relevant title under the heading of "Books Matching ('your enter title')" and then consulting the "Dewey Class:" under "Classification:" heading will provide the classification number of the document he is looking for.

If the classifier don't find the heading "Classification:" or find the heading "Classification:" but don't find the "Dewey Class:" then he should move to the appropriate title under "Libraries this book has an entry in:". Now under the "MARC Record" he should consult the number against: 082 or 092: $a. This will be the classification number of the document he was looking for.

Fig. 3 Home page of ISBNdb.com

6. LIBRARY OF CONGRESS ONLINE CATALOGUE

To classify document by using Library of Congress Online Catalogue (http://catalog.loc.gov/), the classifier need to enter the address http://catalog.loc.gov/ in the address bar of the browser, and then click on "Alternative Interface to the LC Online Catalog (Z39.50)". It will lead the classifier to a new screen, from where he has to opt for "Advanced Search (multiple terms using Boolean operators)". In the new page he has arrived at (it will look just like the following) he can search for class numbers by entering different details about the document in the library. The search term may be the name of the author, title, series, ISBN, ISSN, publisher and many others to choose from. After submission of the details in the interface one need to click on "Submit Query" and then should navigate to "More on this record". Now, against the "Dewey No.:", one will find the class number of the document he is searching for. Please note that for some titles one will not be able to find DDC number in this database, as it was mainly designed by using the Library of Congress Classification number.
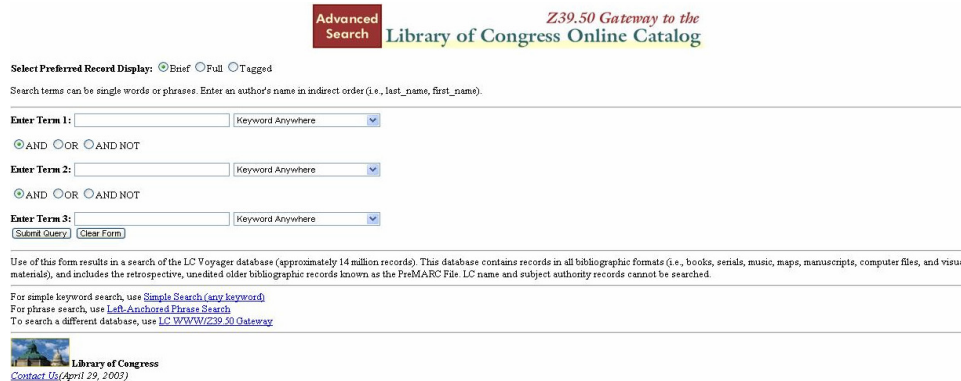
Fig. 4 Home page of Congress online catalogue

7. INDCAT: ONLINE UNION CATALOGUE OF INDIAN UNIVERSITIES

IndCat: Online Union Catalogue of Indian Universities (http://indcat.inflibnet.ac.in/indcat/) is unified Online Library Catalogues of books, theses and journals available in major university libraries in India. The union database contains bibliographic description, location and holdings information for books, journals and theses in all subject areas available in more than 123 university libraries across the country. A Web-based interface is designed to provide easy access to the merged catalogues. The IndCat is a major source of bibliographic information that can be used for inter-library loan, collections development as well as for copy cataloguing and retro-conversion of bibliographic records. The IndCat books database contains 61.4 Lakhs Unique Records. User can search over the platform to obtain the DDC numbers by using title, author, ISBN, and such other fields.
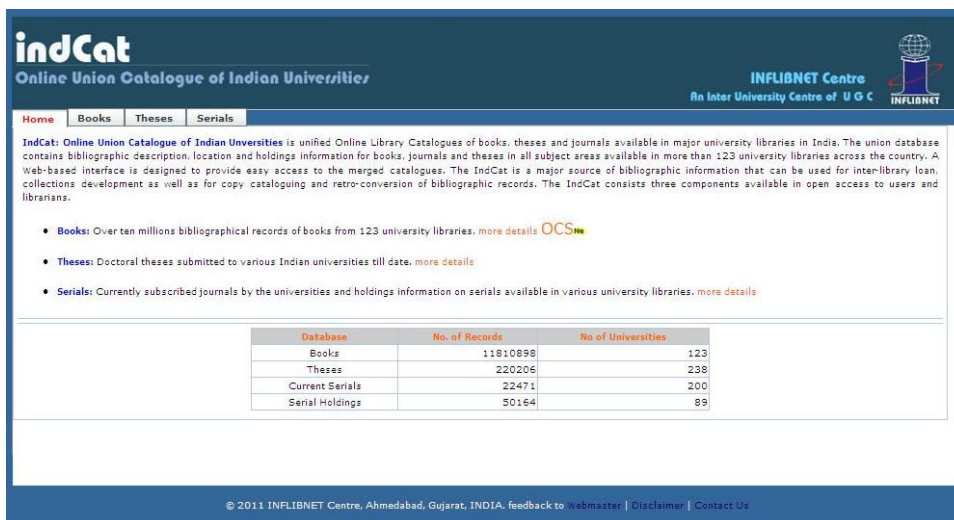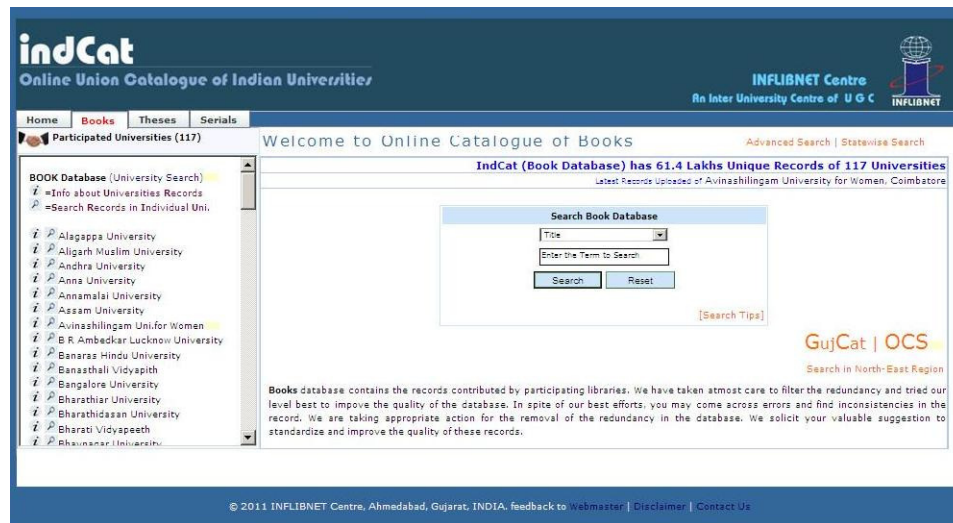

Fig. 5 Home page of IndCat

Fig. 6 Home page of IndCat book database

Besides the above methods, one can also obtain DDC numbers from MARC records. MARC tag 082 stands for DDC number. It also specifies the DDC edition used for assigning that number. For example, the title "Medical terminology: a word-building approach" has the MARC tag 082 00 |a 610.1 |2 22. Here, "610.1" is the DDC number of the document and "22" stands for the DDC edition.

8. CONCLUSION

In the above paragraphs, we discussed how to classify a document by using different online tools and techniques. The relevant matter included was Classify, Dewey Browser, ISBNdb.com, Library of Congress Online Catalogue, and IndCat. Each of these concepts has been exercised to give an idea about the use of the web for classification.

Sometimes a book itself may contain the classification number. In such cases, one can simply copy down that classification number from Cataloguing in Publication (CIP) data. The CIP will provide classification number, subject headings, and notes. This type of data is very common in the verso of the title page of many books published from U.S., Australia, British, and Canada.

REFERENCES

[1] Badan Barman, "Classify: An experimental classification web service from OCLC: My Experience" Message posted to http://lislinks.com/forum/topics/classify-an-experimental, 2010.

[2] "OCLC,Classify: An experimental classification web service", Available over web at: http://classify.oclc.org/classify2/, 2010.

[3] OCLC, Dewey Decimal Classification and Relative Index, 22nd Edition. Ohio: Online Computer Library Center, Inc., 2003.