Paper: CE

The Role of Library Classification in Organizing the Web

Suman S.*

and

Debanshu Karmakar**

Documentation Research and Training Centre

Indian Statistical Institute

Banglore- 560 059

email: *suman_sa@yahoo.com**

email: *debanshu_k@yahoo.com***

Abstract

Information Retrieval in the Web remains a puzzling challenge despite the development of so many search engines and meta-search engines. Library classification schemes have been devised with the aim of organizing the collection to make the retrieval easier. As there is a shift in the focus of libraries the role of classification schemes are also changing. Can they be used in organizing the Web? What are the key advantages? And what are the current practices? The paper tries to explain these issues emphasizing on the use of Dewey Decimal Classification.

1   Introduction

"Classification aims at demonstrating the way in which the human intellect transforms the chaos of sense impressions into a cosmos of concepts". Human beings seem to have an innate need to organize entities. This need to organize large amount of knowledge and information led to the development of classification schemes and other organizational tools. The ultimate aim of all classification work is to make sense out of chaos by grouping similar things together. Library classification has mainly four purposes like order the fields of knowledge in a systematic way, bring related items together in the most helpful sequence, provide orderly access to the shelves, and lastly to provide an exact location for an item on the shelf.

Modern classification research must find order in even greater chaos, with many more intellects and a more complex cosmos, especially in the context of Internet — "a chaotic repository for the collective output of the world's digital printing presses" Lynch (1997). In this situation Library classification can create cohesion across diverse information stores by linking up conceptual contexts.

2   Web organization

Value and relevance of subject descriptor systems were being questioned during the early days of networked information services. The increasing number of automatic full indexing systems added to this kind of a criticism.

But as the amount of information stored on the Net was increasing, this conception got changed. "The more successful the Web, the greater the problem of information and resource discovery" (**Wallis and Burden**). Two distinct ways of finding resources on the Internet emerged (Dodd 1996). One was the use of robot or spider based search engines and the other approach started with producing 'hotlists', which would encourage users to browse the Web. This production of hierarchically arranged lists brought in the concept of Library classification schemes into the Web environment. Subject directories like Yahoo! and other quality controlled subject gateways started to understand how a browsing structure based on classification schemes would enhance searching the Net.

3   Use of library classification schemes

Classification is an intellectual activity requiring much time and energy and also an understanding of the scope and dimensions of the subject concerned. The advantages of using classification schemes to Internet services should be seen in connection with this point. A site that organizes information according to classification schemes may have the following advantages over others:

- Enhanced browsing and navigation
- More recall and precision through broadening and narrowing search terms
- Contextual representation of search terms
- Potentiality to permit multilingual access to a collection if required
- The partitioning and manipulation of a database
- Use of a common classification scheme could make search easier across databases.
- Users may find it easier to browse, because of the familiarity with the classification scheme (which they might have been using in libraries).

Classification schemes however could be criticized at many points:

- The division of logically related materials.
- Difficulty in assimilating new new areas of interests: classification schemes, since they are usually updated through formal processes by organized bodies, often find it difficult to react to or accept new areas of studies quickly.

4   Types of classification schemes

There are several different types of classification schemes around, varying in scope, methodology and other characteristics. The division could be based on subject coverage, language, geography, use, or even the structure of the scheme. Actually, none of these categories are dichotomic; a classification scheme can fit into more than one category. In reality, the most frequently used types of classification schemes are:

✓ Universal

- ✓ National
- ✓ Subject specific schemes, most often international
- ✓ Homegrown systems and
- ✓ Local adaptations of all types

4.1   UNIVERSAL SCHEMES

The term 'universal schemes' is used for schemes that are globally accepted, multilingual and multidisciplinary in nature. UDC, DDC, LC etc are the commonly used Universal schemes.

The use of a universal, multidisciplinary classification scheme in an Internet context could result in the following advantages:

- ✓ They can cover all subject areas.
- ✓ They are widely supported.
- ✓ Continuous updating.
- ✓ User familiarity.
- ✓ Multilingual access to a collection.
- ✓ Availability in machine-readable form. (since most of them are available in machine readable forms).

Universal classification schemes, however, are subject to several criticisms:

- ✓ Rigid or false ontology: The limitations of enumerated classification numbers had given birth to more flexible analytico-synthetic classification schemes like CC
- ✓ Delay in updating and adding new subject areas.
- ✓ Further no classification scheme is able to fully represent the Universe of Knowledge

4.2   NATIONAL SCHEMES

Apart from the advantages and disadvantages of universal classification schemes the national schemes [like Nederlandse Basisclassificatie (BC) and the Sveriges Allmäma Biblioteksförening (SAB) etc.] have some drawbacks as discussed below.

- ✓ Less or unfamiliar among international users.

   ✓ Multilingual capability is not a primary concern for national schemes, apart from countries with multiple languages.

   ✓ National schemes are likely to have a geographic bias

4.3   SPECIFIC SUBJECT SCHEMES

Specific subject schemes (like NLM, and Ei) are usually created for special collections or indexing and/or abstracting services in a scientific discipline. They do have the potential to provide a structure and terminology much closer to the discipline and can be more up-to-date, compared to universal schemes. They could be used mostly in subject-based services. These schemes also do have drawbacks like:

   ✓ They make co-operation between subject services from different subject areas more difficult.

   ✓ People from other subject areas may find difficulties in using the scheme.

   ✓ Some fringe topics, which could be found in subject specific resources, will not be adequately covered within these schemes.

4.4   HOME-GROWN SCHEMES

Some Web sites have tried to organize knowledge on the Internet by devising their own classification schemes like Yahoo. Home-grown schemes do have some theoretical advantages over library universal classification schemes:

- Relatively flexible and easy to change.
- They can very quickly absorb new areas of interests.

On the other hand, these schemes have a number of disadvantages like:

- Lack of consistency
- Unlikely to be well known.
- Not so economical.

5   DDC and the web

Dewey Decimal Classification System (DCC) originally produced in 1876 by Melvil Dewey for a small North American college library is currently one of the most popular Library Classification Schemes. DDC is being revised faster than any other Universal scheme and currently is in its 21st edition. DDC is distributed in Machine-readable Cataloguing (MARC) records produced by the Library of Congress (LC) and bibliographic utilities like OCLC and RLIN. DDC is available online (paid service) as WebDewey and Abridged WebDewey (DDC 21st ed.). DDC is also used in the national bibliographies of the UK, Canada, Australia, Italy and other countries. Research carried out by OCLC in the 1980s established that DDC is a suitable tool for browsing, first for library catalogues and then for the Internet (e.g. Markey 1989; Vizine-Goetz 1996).

5.1   STRENGTHS AND WEAKNESSES OF THE SCHEME

The general characteristics and advantages of universal schemes as given earlier are applicable in the case of DDC. Other characteristics to be noted are as follows:

✓ The scheme is revised more frequently than any other Universal Scheme.

✓ One more flexibility of DDC is that the numbers can be linked to other subject descriptive systems. For example the numbers linked to LCSH headings by most major bibliographic services to the extent that their bibliographic records contain LCSH headings together with DCC and LCC classification data. The USMARC record contains specific tags for several different classification schemes: DDC, LCC, UDC and NLM together with tags for subject headings including LCSH and MeSH. Selected new LCSH headings are individually   linked   to   DDC   numbers   and   are   made   available   via   URL: *http://www.oclc.org/oclc/press/961206.htm*. The USMARC format also allows for links to be made between DDC and other classification systems, including LCC, UDC and NLM.

✓ Digital Availability

✓ Copyright issues: Those who are using the classification can use the notation without restraints in library catalogues and WWW pages, but use of the other information in the schedules would require permission from Forest Press.

✓ It is quite clear that a Faceted Classification is more flexible than an Enumerative Classification Scheme. Though DDC was devised to be as an Enumerated Scheme, subsequent revisions have absorbed the structure and methodology of faceted classification and the use of facet analysis. As a result, subsidiary tables and 'divide like' devices that reflect and can express many aspects of complex topics have been expanded.

✓ Theoretically it could be said that DDC is more flexible than the Library of Congress Classification and certainly simpler than UDC.

If we enter into the schedule there are certain weaknesses like:

✓ **In Engineering (620 Class):** Certain concepts could not be represented uniquely, without overlapping. For example, the concept of "building" as a human act that involves design is spread between 624, 690 and 720. This overlapping always had been a problem with DDC since its inception in 1876.

✓ **In Arts (700 Class):** Same kind of overlapping problem is found in the entire 700 class especially in 780 and 790, which are extremely chaotic.

✓ **In Social sciences:** The 300s have seen much revision to try and iron out weaknesses, yet they still remain. For example, social groups are still classed separately from their culture. While the statistics of a subject are now classed at the subject with 021 appended, no number exists for the statistics of neo- or perinatal death or indeed, any mortality statistics with respect to a particular disease. The law schedule has seen major disagreement over whether jurisdiction or type of law should be classed first and, as a result, it allows 'options' in its interpretation. According to a comparative study undertaken by Biz/ed between the treatments of Business and Economics in DDC and UDC, it was proved that DDC treats the subject in the most suitable way. The recent shift in the subject towards market economies and all is being included in the DDC schedules.

6   Some examples of Web based information Services and gateways, which make use of Dewey Decimal System, are as follows:

6.1   SUBJECT GATEWAYS

Subject gateways are Web-based services that provide access to Internet resources. These kinds of services often offer hierarchical browse structures based on subject classification

schemes. A wide variety of different classification schemes have been used to provide browse access to Internet resources. This section will try to look into the various subject gateways, which make use of the Dewey Decimal Classification System.

✓ **ADAM**, the Art, Design, Architecture & Media Information Gateway, is a service being developed to help find useful, quality-assured information on the Internet. ADAM gives a searchable on-line catalogue describing Internet resources such as web sites or electronic mailing lists, in much the same way as a library catalogue describes. The browsable catalogue is arranged according to DDC 21st edition.

   URL: *http://adam.ac.uk/*

✓ **Biz/ed** is a subject gateway for business education, which offers an online catalogue of good quality Internet resources (Like SOSIG Biz/ed also uses the ROADS software for its gateway). Since its inception in 1996, this catalogue has been using an abridged version of DDC to classify resources, and to create browsable subject categories. It has used the business section of DDC to pick out a selection of numbers and classes that could be used to form the browsable sections on the site.

   URL: *www.bized.ac.uk*

✓ **BUBL** the BUBL Information Services for the UK Higher Education Community gives subject based information services and covers all the subject areas. Subjects could be browsed by alphabetical or DDC (Ed. 21) order.

   URL: *http://bubl.ac.uk*

7   Search engines, subject directories and other web based services

✓ **Blue Web'n Content Categories**. This service gives the facility to browse by subject areas where each subject area and its sub-divisions are marked by their respective DDC notations. DDC numbers are given as mere references and the site does not make use of DDC for organizing the resources.

   URL: *http://www.kn.pacbell.com/wired/bluewebn/categories.html*

✓ **Canadian Information by Subject**: The objective of the National Library of Canada's Canadian Information by Subject service is to provide links to information about Canada, arranged by subject, from Internet resources around the world. The subject arrangement is in the form of an inverted "Subject Tree" which is based on the structure of the Dewey Decimal Classification System. The subject tree is made up of ten main branches

numbered 0 to 9, each representing a broad discipline or field of study. Each of these main branches is divided into smaller and more specific branches, represented by longer numbers.

For example:

> 6 Technology
>
> 61 Medicine
>
> 610.69 Medical Personnel

URL: *http://www.nlc-bnc.ca/caninfo/esub.htm*

✓ **NetFirst** is an authoritative abstracted and indexed database from OCLC. It has used DDC to organize a browsing structure since October 1996. DDC notations had been present in their links from the start of the service, but have only recently been made available for browsing.

URL: *http://www.oclc.org/oclc/netfirst/*

✓ **PICK**: **Quality Internet Resources in Library and Information Science** by University of Wales Aberystwyth, Thomas Parry Library was using DDC based categories for browsing. However the service is no longer available.

✓ **WEBrary** (Online Ready Reference System) by Morton Grove Public Library has organized their Web links according to DDC. Each category is identified by respective DDC numbers.

URL: *http://www.nslsilus.org/mgkhome/orrs/webrary.html*

✓ Another important example of the use of DDC on the Net is the **Yahoo! Subject directory**. Yahoo has been following the broader DDC outlines of different subject areas in their subject categorization. It is true that Yahoo completely never follows the DDC 21st edition outline. In effect it is incorporating DDC with its own home grown classification scheme to present a vast list of browsable subject categories.

All the sites used DDC to order resources, with the exception of the **Blue Web'n**, a site specializing in Web resources for education. Its use of DDC is not as an organizer but as an adjunct to a list of subjects in alphabetical order. Each subject is given its 3-digit DDC number, merely as a reference.

Of these general-resources **Morton Grove Public Library** (also known as **WEBrary**) uses primarily 3-digit DDC numbers to group the resources they point to. Even at this general level of classification many numbers only lead to single resources. In the National Library of Canada's **Canadian Information by Subject** DDC numbers goes beyond the initial decimal point and the quantity of resources means that clumps of resources appeared under individual DDC numbers. Nevertheless the classification is not that deep.

8   Conclusion

The Web is a confused repository of diverse information, which lacks the coordination and organization of a traditional library concept. It has been practiced and proved that the use of traditional library tools and techniques could be a great help in taming the Net. Still picking up the right information from this 'flea market' remains a problem. There are a lot of advantages and disadvantages of the use of DDC on the Net. Other Universal scheme like UDC and LC also could have almost the same advantages and disadvantages. Still DDC always remains less complex and popular. As **Millwaine** had suggested, UDC based on the structure of DDC could complement the latter by providing greater details and more distinguishable facet indicators in the machine oriented world.

Lastly, we can say that enumerative schemes have their own disadvantages in the Networked environment also; and it can also be suggested that the use of a more flexible faceted scheme like Colon Classification may be more suitable in the Web environment. Or a better combination of different schemes could also be used.

9   References

1. **Diane (Vizine-Goetz).** Using Library Classificatioon Schemes for Internet Resources. OCLC Internet Cataloging Project Colloquium, OCLC Office of Research and Special Projects. *http://www.oclc.org/oclc/man/colloq/v-g.htm*

2. **Koch (T.). and Day (M.).** The role of classification schemes in Internet resource description and discovery, *DESIRE D3.2* (3), February 1997. ***http://www.ukoln.ac.uk/metadata/desire/classification/***

3. **Susan (Irwin).** Classification theory and the Internet: A move toward multidimensional classification. ***http://www.du.edu/LIS/collab/library/student/Irwin.htm***

4. **Savvy** searching. *Online information review,* 25 (2) 2001.

5. **Charlotte (Jenkins), Mike (Jackson), Peter (Burden) and Wallis (Jon).** Automatic Classification of Web resources using Java and Dewey Decimal Classification. *http://www.scit.wlv.ac.uk/seed/www7paper/*

6. **Wallis (Jon) and Burden (Peter).** Towards a Classification-based Approach to Resource Discovery on the Web. ***http://www.scit.wlv.ac.uk/wwlib/position.html***

7. **Day (Michael).** Renardus DDC classification mapping - a summary of work in progress. (UKOLN), on behalf of the Renardus DDC Mapping Working Group, 12 June 2001

8. **Lynch (Clifford).** The Internet is not a digital library. Searching the Internet. *Scientific American.* March 1997.

9. **Adeock (L.).** Building a Virtual Music Library. *Knowledge Organization*, 28(2) 2001.