Web Cataloguing

M Ramshirish

Abstract

Increased Research and Development activities in all subject areas initiated massive growth of literature and the advent of Internet provided further impetus to the process. Open nature of Internet served as platform for people to share/ publish their ideas. This condition led to the chaotic state of web. The main reason behind it was the unorganized heap of information in the form of web documents. So, the need of the hour is cataloguing the web, thereby providing appropriate surrogate to locate the relevant piece of information. This paper discusses the need of web cataloguing and the important issues therein.

Keywords : Internet, Catalogue, Web Cataloguing, Metadata, Information Gateway

0. Introduction

We know that catalogue is a tool for locating the documents and acts as a surrogate that facilitates information retrieval. Catalogue actually contains bibliographic details such as author, title, publisher, copyright details etc.

The exponential growth of information on Internet and the World Wide Web, in recent years, has created a state of chaos. Because of this sudden growth, a plethora of materials on different subjects is instantaneously available on desktops in every nook and corner of the world. But these materials are not organized properly and differences in quality and stability also exist. So, it is very difficult to conceptualize, browse, search, filter or refer to any piece of information.

Although cataloguers are skilled at cataloguing all forms of media, the immense amount of resources on the web presents an entirely new set of issues and the challenge lies therein.

1. Cataloguing

Library Cataloguing is the process of preparing catalogue. As mentioned earlier, cataloguing is nothing but creating the surrogate to the entities of the collection of a library(s).

The main use of catalogue is to help the library users in finding/locating the items they want in the library. The catalogue serves following three purposes:

- To find out the availability of resources,
- To know the location of the resource, and
- To bring related resources together

To quote David M.Levy, "Cataloguing is not just a matter of "reading off" the properties the items have but of normalizing or regularizing the material to conform to standard categories of description and thereby making the properties in the act of describing them". [1]

In the context of web cataloguing, rules or guidelines specify how the content of the metadata format is entered. Most of the standards or guidelines as of date, were originally designed for the cataloguing of

traditional-mostly printed-collections. However now, many of them have been revised to take account of the resources based on newer technologies developed till date.   [2]

1.1   Cataloguing The Web

Now, if we come to the wide area of Internet and World Wide Web, then the situation changes a lot. The web is a dynamic and open resource. Here, any one can deposit any thing at any time and the content of the web pages also keeps on changing. The content of the web pages keeps changing in the time period of fortnights. It is mind boggling for cataloguers to cope with the challenge.

2.   **Efforts of Cataloguing the Web**   [3]

World Wide Web has created a difficult situation for the cataloging community. Now, the user is not only researcher but also general public on worldwide scale, so, the cataloguer of today is caught between the demands of the user and infinite amounts of Information. The users of present day need information of completely different nature. Subject holding is much larger on the web, and is also continuously changing, thus it requires constant maintenance. So, the need of the hour is cataloguing the web.

The efforts till date can be enlisted as: -

- First method of organizing the Information on the Internet was search engine. Search Engines had an initial information space, which was managed by the search indexes, they maintained through spiders. But, there weren't many pages to handle, so results returned were manageable. However, search indexes started faltering with the increasing size of the Internet.

- In addition to above, less and less of the Internet was being indexed and at the same time, user searches on common terms started returning so many hits. It resulted in increased recall and decreased precision.

  At this juncture, portals such as Yahoo came up to give a catalogue like structure. This structure allowed both browsing and searching. Although the result was relevant in most of the cases, but the exact precision of retrieval as facilitated by a catalogue was missing.

- Then, the idea of proper cataloguing of web started gaining ground. Here, the job gets defined for trained professional who knows how to order information so that the system designed not only performs in the short term but is also capable of coping with the problems in the long term.

Here, the concept of metadata came into being. While metadata has become a recent buzzword, the concept has long been familiar to librarians who use the terms "cataloguing" or "indexing". The MARC (Machine Readable Catalogue) record used by Librarians to describe resources is one form of metadata only.

Metadata can be encoded in the object or stored separately. Metadata or "meta tags" are embedded in the HTML header tag of a web page. These tags can be used by web site creators to help search engines find and rank a web site.   **[4]**

But, there are ethical and legal problems associated with meta tags on the unregulated Internet, where web pages may be submitted with keyword meta tags that do not accurately reflect the content of the site.

Problems of this type get solved in the Information gateways, whose scope is regulated and limited.

Subject gateways direct users to quality resources by linking to searchable and browsable catalogues of Internet based resources. The most crucial element of a subject gateway is the catalogue, or directory, of Internet resources.

The use of Internet gateways by libraries has transformed the notion of the catalogue as a retrieval tool for items held by a library, to a kind of virtual interlibrary loan. Liu points out that "as librarians have moved to gateways, catalogues have moved to become portals to content". **[4]**

However, there are many cataloguing issues associated with development of Information gateways.

3. Cataloguing Issues for the Information Gateways **[2]**

Information gateways need tools to facilitate the identification, location, access and use of resources.

Following issues are of concern for the purpose: -

3.1 Metadata formats

The role of cataloguing rules or guidelines is to specify how the content of metadata format is entered. Here, Cataloguing issues are related to the decisions that information gateways have to make about metadata formats.

- Types of descriptive information considered by an information gateway for inclusion of a document are:

a. Bibliographic type descriptive information.

   Here, details like title, its location (usually a URL) and Statement of responsibility are taken and these are to be taken primarily from the source itself.

b. Subject Information

   It includes terms taken from classification schemes, thesauri and subject heading lists etc.

c. Administrative metadata

   This includes any other information that may be useful to the management of the subject gateway like cataloguer, date of selection of resource etc.

3.2 Content Standards and developing cataloguing rules

Once the metadata format has been adapted and decision has been taken on the particular information that resource descriptions need to contain, then, guidelines are prepared for cataloguing.

Cataloguing guidelines contain: -

a. A list of all data elements
b. A brief explanation of what particular information each element is supposed to hold.
c. An explanation of how information should be entered into this element (the rule)
d. Some guidelines on the use of formats for dates, language codes, etc
e. Notes of (and links to) external standards used, e.g., classification schemes, name authorities.

3.3   Cataloguing tools and interfaces

The creation of Internet resource descriptions for information gateways largely takes place through an interface or cataloguing tool.

Information gateways need cataloguing interfaces that can be adapted for their particular needs that contain, for example, the subject schemes used by that particular gateway. As it difficult, so, including some help in the form of cataloguing rules and examples is a good idea.

3.4   Catalogue maintenance

It necessitates the use of automated tools like link checks, but may also entail some periodic checking of Information content because of the dynamic nature of information on Internet.

4.   Main Issues in Web Cataloguing

The main issues in Cataloguing the Web Resources can be enumerated as: **[4]**

4.1   The amount of data

Sky is the limit for the data on the web. And day-by-day it is getting difficult to find out a piece of information on the World Wide Web. An estimation in the year 2000 by NEC-RI and Inktomi puts the number of pages at over 1 billion indexable pages or about 10 million sites. Jackob Nielson states that "the web is expected to increase by a factor of 20, meaning thereby that there will be about 20 times more sites on the Internet (increasing from 10 million sites on in 1999 to 200 million sites in 2005).

In this situation, how is it possible to catalogue such a huge number of resources on the Web?

4.2   Lack of stability

There is no stability on the web, it keeps on changing. Existing sites change their location or are completely withdrawn, new sites appear constantly. And due to revision of sites, the material on the web page changes very frequently.

4.3   Lack of Controlled Vocabularies

Librarians have developed consistent cataloguing standards in the form of authority lists or controlled vocabularies, which allowed for locating, collocating, and retrieving the items. There is no standard vocabulary or authority control on the Web. And because of this open nature of the Internet, it is not possible to follow standards as well.

4.4   Lack of Quality Standards

Due to open nature of web, nature of publishing has changed. Now, it is free and equal for all. In general, anything that is accepted for publishing in print, is reviewed i.e., assessed by editors and publishes. But when a web page is to be published, it is merely uploaded to a server. So, any one can publish on web without caring for the quality standard.

The world of information processing has been revolutionized by the development and growth of web. The word 'anarchy' comes from the Greek *an-archos,* or rule by none. So, it doesn't appear to be an exaggeration when we say that the present web is nothing but anarchy.

However, the work is going on to put some control over the web using directories and metadata.

5.   Conclusion

Rapid growth of information and consequent increase in the number of users of information on World Wide Web is posing a new challenge for information organization. In this situation, only cataloguing appears as a means of sustainable navigation.

The Internet continues to transform the catalogue and cataloguing practice. Information professionals are looking for ways in which to organize and make accessible the resources on the web.

Few years from now, we will see the profound growth of subject-based portals. And at the same time, cataloguing professionals will play the leading role in changing the state of chaos to an ordered form, by designing techniques for web cataloguing and applying them.

6.   References

1. Levy, David.M. "Cataloguing the Digital Order". In The Second Annual Conference on the Theory and Practice of Digital Libraries June 11-13, 1995 - Austin, Texas, USA. < http://csdl.tamu.edu/DL95/papers/levy/levy.html > (November 27, 2003)

2. Michel, Day. "Cataloguing". DESIRE Information Gateways Hand book (Print Version). <http://www.desire.org/handbook/2-4.html> (November 27, 2003).

3. Mcleod, Bruce. "Factoring the Librarian back into World Wide Web cataloguing". <http://ausweb.scu.edu.au/aw2k/papers/mcleod/paper.html> (November 27, 2003).

4. Atkey, Susan. "The Moving Target: Issues in cataloguing the Web" <http://www.slais.ubc.ca/courses/libr500/02-03-wt1/www/S_Atkey/index.htm> (November 27, 2003).

5. Bowman, J.H. "Essential Cataloguing". London: Facet Publishing, 2003.

6. Dillon, Martin and Jul, Erik. "Cataloguing Internet Resources". Cataloguing and   Classification Quarterly, Vol:22, No:3/4(1996). 197-238p.

7. Gorman, Michael. "What is the future of Cataloguing and Cataloguers?". <http://www.ifla.org/IV/ifla63/63gorm.htm> (November 27, 2003).

8. "Cataloging Internet Resources: A Manual and Practical Guide"

    <http://www.oclc.org/support/documentation/worldcat/cataloging/internetguide/default.htm>

    (November 24, 2003).

9. Chapman, Ann, Day, Michael and Hiom, Debra. "Metadata : Cataloguing practice   and Internet subject-based information gateways"

    < http://www.ariadne.ac.uk/issue18/metadata/> (November 27, 2003).

About Author

Mr M.Ramshirish is ADIS Student at DRTC/ISI, Mysore Road, Bangalore 560 059, India and holds BLISc.
E-mail: mramshirish@yahoo.co.in, ramshirish@rediffmail.com