

CONTENT MANAGEMENT IN WEB SITE WWW.ARIPUNE.ORG WITH SPECIAL REFERENCE TO THE BIBLIOGRAPHIC DATABASE OF SCIENTIFIC PAPERS PUBLISHED BY SCIENTISTS OF ARI

by

Kulkarni S. N. *

ABSTRACT

The paper is based on the practical experience gained during the process of web site creation to its launching on to the Internet. It attempts to describe the methodology followed for the content collection and its organization in the web site <www.aripune.org>, a web site of Agharkar Research Institute, Pune. It describes the content organization and management issues faced especially with the content 'author name' in the bibliographic database of Scientific Papers of the institute. It finally, suggests the need to prepare the Universal Style Manual for Authors and advocates to the Library and Information Science professionals to take more interest in the content management in web site.

Keywords: Content Management, IARI, Pune

* Chief Librarian, Agharkar Research Institute, Pune

0 Introduction

Agharkar Research Institute (ARI) is an autonomous grant-in-aid Research Institute aided by Department of Science and Technology (DST), Government of India. Established in 1946 as the Maharashtra Association for the Cultivation of Science. The institute was renamed as Agharkar Research Institute in 1992 to honour its founder Director, Late Professor S. P. Agharkar. The web site of the institute www.aripune.org was launched in 1998. It was then in plain HTML textual format and was without any databases. The focus was to present information about various scientific activities the institute. In the beginning of new millennium year, it was decided to reorganize the web site by adding new information (content). I was closely associated with the Content Management of the web site, as one of the members of web site Committee. The present paper discusses the various efforts undertaken for the effective Content Management in www.aripune.org with special reference to the bibliographic database of scientific papers published by scientists of ARI.

1 Content, Content Management, and Content Management System

These terms are very well and precisely defined on a web site [1] and are reproduced here to develop better understanding of the subject. To understand what is Content, Content Management (CM) and Content Management System (CMS), we have to first define

?? what it is that we are referring to when we talk about content

?? what do we understand under its management

?? what do we mean by a system

What is Content?

Content is in essence, any type or unit of digital information that is used to populate a page. It can be text, images, graphics, video, sound, etc. or in other words- anything that is likely to be published across an inter-, intra- and/or extranet.

What is Content Management (CM)?

Content management is effectively the management of the content described above, by combining rules, process, and/or workflow such that, centralized webmasters and decentralized web authors/editors can create, edit, manage and publish all the content of a web page in accordance with a given framework or requirements.

What is Content Management System (CMS)?

The system itself is definable as a tool or combination of tools that facilitates the efficient and effective production of the desirable web pages using the managed content.

To combine all three, we can say:

"A CMS is a tool that enables a variety of technical (centralized) and non-technical staff (de-centralized) to create, edit, manage, and finally publish a variety of content. These include text, graphics, video, sound, etc., and are being constrained by a centralized set of rules, process and, workflow that ensure a coherent, validated web site appearance."

2 Content Collection/Creation process for the web site aripune.org

The first pre-requisite of any content management system and/or web site is the collection/creation of content. If proper methods are used to collect the content for a web site, then most of the problems get evaporated at the initial stage. Otherwise, at a later stage, it becomes increasingly difficult to co-relate such content in a web site. It is therefore indispensable to prepare suitable formats for whatever content is to be collected. First, it helps in developing the standard terminology and assures uniformity. Secondly, it gives a clear idea to the content providers as to how the content is to be given.

All these points were considered for content collection and creation of the web site. When the content for your web site is defined, it is relatively easy to determine the technical needs of the web site. What user data do you want to collect and use? [4]. The scope of content collection activity was defined to cover the following areas/aspects 'about the institute, research activities, patents, publications, bio-data in brief, facilities available, library services, events and contact information'. Accordingly, the contents were collected from all the scientists of institute and from other sources.

3 Illustration of Content Collection Methodology

I) Content of Research Activity:

- ?? Topic of Research
- ?? Past accomplishments -Present scope of research-Future prospects

II) Content of bio -data of the scientists

- ?? A format used by INSA (Indian National Science Academy) yearbook was adopted.
- ?? And additional information on patents and publications (five only).

III) Content of Common/General Information

- ?? Annual reports -Brochures-Existing Published/ Authentic Sources
(Formats were prepared for Events, Seminars, etc.)

However, in case of some of the content providers, given content needed to be rewritten. The formats and methods used to collect the content were found to be very useful and further helped in integrating the content in a database format.

4 Bibliographic Database of Scientific Papers of ARI

The publication of scientific papers is a prime activity of any research institute and is one of the important parametric indicators of the progress made by that institute. ARI is no exception to this and publication of scientific papers is a regular outcome of its research activity. It was decided to add the database of scientific publications as the content in the web site. The purpose of the web site is to facilitate the exchange of scientific information and to generate new scientific collaborations with other institutes. At first glance, adding such content (i.e. database) in a web site appears to be an easy task. In reality, to incorporate such content in a web site requires careful study of it. Mere data are not sufficient or useful to get the desirable and satisfactory search results from the database content.

5 Nature of Content of Bibliographic Database of Scientific Papers of ARI

The data elements are author/s name, title of paper, journal title, issue number, volume number, year, page numbers, and keywords, if any. The data elements like author name, title of paper (key words only), and year are assumed to play an additional important function as that of search key elements. It is a common practice to browse the database using one or a combination of such search key(s). It is a well-established fact that, the author name is the primary search key for any bibliographic database. However, it was noticed from the bibliographic database of scientific papers of ARI, that large variations in author name exist in spite of the fact that the data were obtained from the original sources. Such variations with any content are always a matter of serious problem in content management.

Illustration of Variation in rendering of author name

- | | | |
|---------------------------|------|------------------------------|
| 1) Sunil Narayan Kulkarni | ---- | Full Name |
| 2) S. N. Kulkarni | ---- | Initials with Surname |
| 3) Sunil Kulkarni | ---- | Name |
| 4) S. Kulkarni | ---- | First Initial and Surname |
| 5) Sunil N. Kulkarni | ---- | First Name, Initial, Surname |

Further, variations were also found in the abbreviated journal titles e.g. Journal, Jour., J, Science, Sci, etc. The above illustration is sufficient to appreciate the problems posed in the content management of bibliographic databases. Some of these are listed here:

1. Creates confusion to all those who wish to search the database
2. Content components get scattered (for the same author)
3. Defeats the very idea of a unique search key
4. Multiple search keys become necessary to complete the search
5. Additional entries like "see also" may be required
6. Specific instructions need to be given to the users
7. Search results for (author + year) may remain incomplete
8. Cross referring may not be possible in every case
9. Online search time may increase or get wasted.

6 Content Management of Bibliographic Database of Scientific Papers of ARI

The content management in bibliographic database invites foremost attention to the content of author name over any other content in bibliographic database. It was discussed, with some of the authors to find out the reasons behind such variations in rendering author name. It was realized that the style manual (instructions to authors) of publishers, publishing bodies are in some way responsible for such variations, as they do not follow an identical style for author's name. The style manuals are required to be followed strictly by authors for submission of papers to the publishers. On the other hand, some of the publishers are not very strict about a particular style (for author's name only) and offer freedom to the authors by just indicating "Author Name" in their style manual. Such policies encourage authors to write their name in any style they choose.

In spite of all such variations, content needs to be organized and managed so as to get the desired search results. We were fortunate to identify these problems at an initial stage of content management. The necessary precautions and measures were undertaken while designing the database of publications. It involved some additional programming, data entry work, and of-course the time as well. Finally, we were successful in managing the content in the required manner and to get search results as desired.

7 Content Management System in web site www.aripune.org

A secure password oriented Control Panel System is employed to add, update, delete the content of web site. In other words, this system is called as the "Content Management System" of web site. It is totally menu driven system and provides facility to add-edit-delete every content file in the web site online. The changes in the content are incorporated instantly. It is easy, quick and does not require any sort of web programming knowledge. Most of the web sites nowadays use control panel system as a tool for efficient and effective way for content management.

The control panel system is one of the best suited systems to manage the content of database oriented web sites online. Frequent uploading and downloading of database may consume hours-and-hours of online time. Further, the web site may not be available to users during such process.

However, to incorporate major changes or for annual maintenance of databases and alike, ftp login-in is preferred. As a part of content management system, we have prepared a schedule

to add-edit-delete the content and also prepared formats to collect newly generated content so as to publish it on the web site. The ideal content management system takes care of the content based on its ultimate usage and life cycle. The content life cycle is given here:

“Collection/creation--preview--edit--approve--publish/distribute--archive/delete”

8 Findings and Conclusions

The findings and conclusions of this study are presented here.

1. Content collection needs proper planning.
2. Formats need to be prepared for each and every aspect of content to be collected.
3. Formats and proper methodology employed to collect the content help to rationalize it and assure uniformity.
4. All good writing is rewriting [2]. Content collected requires rewriting in some cases so as to maintain the uniformity. However, the originality of content should be preserved in any case.
5. Content collected needs to be organized and managed to suit its ultimate usage
6. Content management is required for bibliographic database and merely data are not sufficient or useful to get the desirable results.
7. Content management in bibliographic database deserves serious attention to author's data, as it is a fundamental content.
8. The software solutions for content management in a web site are always possible, however to identify such content is essential.
9. The content life cycle needs to be understood for better content management e.g. outdated contents that are to be deleted in time or replaced with the current.
10. A secure Content Management System for maintenance of complete web site is mandatory for coherent, validated web site appearance.

9 Recommendations

It is worthwhile here to make a few recommendations based on the above discussions and the practical experience gained during the process of web site creation to its launching on the server.

- 1) **It is necessary to promote an awareness about the concept of Content, Content Management and, Content Management System.**

There is certainly a need to create awareness about the concept of Content, Content Management and, Content Management System, amongst all those, who are involved in web site creation, management and publication. This will help to make the concerned persons more serious about the web site and then to make it more productive. At present several web sites are in existence with ceased content.

- 2) **Content management warrants, first to exercise the other options/solutions, than those based on software for managing the content**

The software technologies can always provide quick solutions to the problems of content management; however other feasible solutions need to be exercised first. Otherwise, the role of content management itself may be neglected or overlooked.

The content collection methodology, formats, and content itself need to be checked, verified first and then only software based solution may be attempted.

3) Universal Style Manual for Authors needs to be prepared and followed.

The web really only took off towards the end of 1993 and is still in early stages in its evolution [3]. It is a pressing need of (digital) time and information technology that, all the publishers in the world should follow a standard and unique pattern for an author name. This will offer the million-dollar solution at zero cost, as the thousands and thousands of bibliographic databases face a similar problem. In this context, it is necessary to create awareness among the publishers. It is not a difficult task if one considers the agreement reached by the European countries to bring in a common currency, the Euro dollar. Urgent efforts are required in this direction. In this era of InfoTech super highway the significance of free flow of (scientific) information worldwide has been accepted. However, the above mentioned problems may act as harmful speed-breakers and lead to greater confusion in future. A timely attention needs to be given to prepare a universally accepted style manual for authors; All such efforts will help diminish the content management problems in bibliographic databases on web sites. The common international standards are very well practiced in several fields. It is really a wonder therefore, that the publication industry which is an internationally well settled knowledge industry is unable to follow such a simple “knowledge management standard” as regards with the name of the author, the fundamental content creator of knowledge.

4) The Library and Information Science professionals need to take more interest in the content management in web site.

The content creator, content manager, content management system provider and content- or end-users are the pillars of any content-based web sites. The content creators are authors, publishers, etc. Content managers are those who are involved in collection, organization and management of the content. The content management system providers are those who develop the content management system. Lastly, user means any one who makes use of the content available on a web site i.e., end-user. The library and information professionals mainly come under the category of content managers by virtue of the nature of their work. It is therefore their professional responsibility to take a keen interest in research and development of the field. It is a new professional responsibility for them. These professionals handle a variety of contents every day related to different fields. Further they are well acquainted with several aspects of content in the form of information, data, reference, bibliography and so on. Now, there is change only in the medium of communication, from printed to electronic or more precisely, digital. In view of the re-engineering of library services it may be predicted that web designing and web hosting are going to part and parcel of library's work in near future [5]. These new challenges need to be accepted to ensure further development of their profession and to keep it dynamic.

10 Acknowledgements

I am grateful to Dr. Surendra Ghaskadbi, Scientist, ARI, for critically going through this article and giving valuable comments for its improvement. I thank Dr. V.S. Rao, Director, ARI, for his encouragement and guidance. Thanks are also due to members of the institute web committee and scientists for their cooperation.

11 References

1. <http://www.contentmanager.eu.com/history.htm>
2. DIANNA (BOORHER). E-Writing 21st Century Tools for Effective Communication. 2001. Pocket Books, New York. p 52.
3. ANDREW (FORD). Spinning the Web: How to Provide Information on the Internet. 1995. International Thomson Publishing, London. p 6.
4. MARY (HAGGARD). Survival Guide to Web Site Development. 2001. Printice-Hall of India, New Delhi. p 7.
5. PUJAR (S M) and MANJUNATH (G K). Domain Registration and Hosting of a Web site: The Ways and Means. Annals of Library and Information Studies. 48, 2; 2001. p 49-55.

12 Bibliography

1. PHIL (BRADELY) and ANNA (SMITH). World Wide Web: How to Design and Construct Home Pages. Aslib, London; 1995.
2. MARK (KERR). How to Promote Your Web Site Effectively. Aslib, London; 1999.
3. GANESHAND (NATARAJAN) and SADHYA (SHEKHAR). Knowledge Management: Enabling Business Growth. 2000. Tata Mcgraw-Hill, New Delhi.
4. SINGH (SHANKAR). World Wide Web Handbook For Librarians. 2000. Ess Ess Publications, New Delhi.
5. Instructions to Authors from several Indian and Foreign periodical.