

DIGITISATION OF MARATHI MANUSCRIPTS

By

Dr. (Mrs.) N.J. Deshpande*

Mr. B.M. Panage**

ABSTRACT

The present paper discusses the importance of manuscript collection and the necessity to preserve it for the future research and references. It highlights on the existing precious manuscript collection of Jayakar library. The paper tries to convert document in the digital form to preserve and maintain it. More emphasis is given on conversion of Marathi manuscripts available in the Jayakar library. It also discusses various methods of capturing data and its advantages and disadvantages. The central theme of the paper is to find out, how much space is required to store one image in various file formats like bitmap, tif and jpeg, which may act as a guideline for preservation and maintenance of manuscripts.

* Librarian and Head, Dept. of LIS, Jayakar Library, University of Pune, Pune - 411 007.

E-mail : njd@lib.unipune.ernet.in

** Assistant Librarian, Jayakar Library, University of Pune, Pune - 411007. E-mail :

bmpanage@rediffmail.com

0. Introduction

Manuscripts are precious part of our cultural heritage. Great ancient scholars dedicated their lives for creating written records of knowledge. India's most valued and revered gift to humanity is its profound and timeless heritage. This heritage encompasses almost every aspect of human enquiry, exploration and existence covering philosophy, religion, language, literature, metaphysics, art and dance and so on . Today this heritage is scattered in texts in libraries and in individual positions. This precious gift is slowly decaying and vanishing due to the improper handling . Indeed, preservation of this heritage presents a great challenge before us . However, fortunately the information technology is offering many solutions not only for preservation , but also for enhancement and for its wide scale access.

Knowing the fact that the treasure of manuscripts still lying scattered in the nooks and corners of India, Dr. Raghavan suggested University Grant Commission to appoint a manuscript committee. Accordingly, U.G.C. appointed a manuscript committee under the chairmanship of Dr.V. Raghvan in 1959. This committee visited many places and found that the preservation and other type of work related to manuscripts was not satisfactory and made following suggestion, " unless a systematic policy for the collection , preservation and utilization of manuscripts was pursued in right way, there is a danger of

manuscripts being taken out of the country by foreign agencies, as also destruction because of ignorance and negligence."

Many efforts are made to preserve the manuscripts and nowadays, many agencies have carried out digitisation projects . Some projects are completed, while some are at the initial stage. It is found that many Indic manuscripts are available in foreign universities and they have developed their digitized collection of manuscripts. One of such universities is University of Pennsylvania.

1. Manuscript Collection of Jaykar Library

Jayakar library has a very rich collection of manuscripts and at present this library holds around 5000 manuscripts. The present collection of manuscripts is enriched by acquiring it from various parts of India, These manuscripts are purchased or received as donations during the last five decade. This collection extends over important branches of learning and contains titles of many unpublished work. Jayakar Library has published the descriptive catalogue of manuscripts available in its collection, which contains Sanskrit, Marathi and Hindi manuscripts.

There are about 700 Marathi manuscripts, out of which three collections are eminent, namely Ketkar collection, Dattavarada Vitthala and Mahanubhava collection..

2. Digitisation (definition)

"Digitisation refers to the process of translating a piece of information such as a book, sound recording, picture or video into bits. Bits are the fundamental units of information in a computer system. Turning information into these binary digits is called digitisation." Digitisation is one of the hot topics in librarianship today. To build a 'digital library' requires that the content of a collection be available electronically. The rhetoric of the information highway has provided the impetus to convert many existing paper-based (or sound , video) collection into new digital media. The assumption is that digital collections will be more accessible to a broader range of users, presumably through networking techniques, and new efficiencies are to be gained in resource sharing and for preservation.

3. Digitisation Process

Digitisation requires a basic process, which involves different sets of hardware and software technologies at each step. Determining the appropriate technology is directly linked to the anticipated use and purpose of the material being digitized. For digitizing the text and other material, following five methods can be used.

- a) Manual data entry Scanning
- b) Optical character recognition (OCR)
- c) Excalibur Technologies and pattern recognition technologies
- d) Document imaging

a) The simplest method of converting an image of a page (or the real page of text) into digital text is to enter it manually. This is usually a time consuming method but very useful from the point of view of information retrieval.

b) In the second method, scanners are used to take digital pictures of objects. Scanners can be simple desk top machines or very large and complex systems that process thousands of documents.

c) Another simple digitization process is of OCR i.e. scanning printed pages to build a digital database of text. This process uses OCR (Optical Character Recognition) software, which takes a picture of the page and then turns it into digital text, which can be edited or fully indexed. OCR software must distinguish between black and white areas of text.

d) Excalibur Technologies and Pattern Recognition Technologies are the next generation of OCRs, represented by Pixie, a product being developed by Excalibur Technologies. This software uses a technology called Adaptive Pattern Recognition, which attempts to mimic aspects of the neural patterns of the brain. The software can be taught to recognize variations and relationships in pattern, such as patterns of text rather than readable text. The retrieval of search terms uses what Excalibur calls "fuzzy matching".

e) Document Imaging, a simple method of capturing text, involves taking an electronic picture of each page of text with the same type of scanner as one would use for OCR. However, the difference is that the images are stored as graphic files rather than text files. A similar technology is used for fax transmission. Each page is stored as one picture. The text on the page cannot be edited or indexed.

4. Methodology

This paper discusses in detail, how to convert the Marathi manuscripts available in Jayakar Library into digital form. One of the aspect of digitization is preservation and the authors have tried to preserve the manuscripts with the help of first and second methods which are mentioned above.

Following three steps are used for digitization of manuscript :-

1. Scanning the original manuscript folios and preserving the image.
2. Making manual data entry using Marathi software.
3. Preparing an index and translating the original text into English for foreign scholars.

5. Collection of Dattavarada Vitthala

Dattavarada Vitthala (Sake 1670-1720) was a Marathi poet, who lived 200 years ago. He was a contemporary of the Peshwas Nanasahab and Madhavrao. Most of his works are unpublished. These manuscripts are discovered by the late Shri. Kashinatha Panduranga Parakhi. These manuscripts were evaluated by M.M. Datto Vaman Potdar

and proved to be a valuable material for research in Marathi. There are around 52 manuscripts which are available in the Jayakar library, out of which, first ten manuscripts are selected for this pilot study. The details of these manuscripts are shown in the table given on next page.

Serial Number Accession Number Title 2081 2515 Adhyatma Ramayan Balakanda
2082 2533 Atmaprabhoda 2083 2524 (Bhagavat) Gitasara 2084 2516 Bhagavata-
Caturtha(4th) skanda tika 2085 2517 Bhagavata Pancama(5th) skanda tika 2086 2518
Bhagavata-Saptama(7th) skanda tika 2087 2552 Bhavanidasaka stotra caranavyatha,
Martanda dasakastotra and other works 2088 2544 Dattatraya lilavighraha 2089 2541
Dhyanacaurhasi 2090 2542 Ganesh panchayatana Pancastaka stotra

The scholar or user can access any one the manuscripts from the above table. He can select the manuscripts just by clicking on the accession number or on the title of the manuscripts. After clicking on the particular manuscripts, following information will be displayed on the screen. The description of theses manuscripts is based on the descriptive catalogue of manuscripts which are published by Jayakar Library, University of Pune.

Dhyanacaurhasi

Sr.No. 2089

Acc.No. 2541

Title Dhyanacaurhasi

Author Dattavarada Vitthala

Commentator

Material Paper

Script Devnagari

Size in Cms. 23 x 13

Folios 19

Lines per Page 10

Letters per line 21

Extent C

Condition and Age G

Additional Particulars

Further information can be accessed by folio number (page number) . i.e. if you want to see the information of folio number1 , then just click on that folio.

Folio	1	Folio	2	Folio	3	Folio	4	Folio
5								
Folio	6	Folio	7	Folio	8	Folio	9	Folio
10								
Folio	11	Folio	12	Folio	13	Folio	14	Folio
15								
Folio	16	Folio	17	Folio	18	Folio	19	

As soon as you select the folio number to see the information of that folio, you can see the original page of manuscript, which is scanned and will be displayed in the following manner

Technical details :- The above image is scanned using the following details

- a) Resolution - 100
- b) Sharpen level - low
- c) Pixels - 522 x 930

The total number of bytes required to store one page of manuscripts (23cm x 13cm) is 1,391 KB, if the image is saved in bit map structure. The same page requires 1,397 KB if it is saved using tif formats and 96 KB are required, if it is saved in jpeg format. In short, a single floppy disk (1.44MB) is required to store one image which is captured by the scanner.

User can select any folio to see the original contents of the manuscript. The original image of the page, which is scanned using some standard scanner, will be displayed. And at the same time, English translation of the above folio will be displayed at the end.

6. Conclusion

The UNESCO project entitled ' The Memory of the World' was built on the premise that the cultural society has the responsibility to preserve information about the history and make it available also for future generations. It aims to stimulate a responsible approach to the sources from which our historical consciousness grows and to contribute to the general availability of information about our history and culture.

The abundance of information of average and below average quality generates paradoxically the demand for new, unusual, exotic and uneasily available information. This explains the growing interest in old manuscripts. However manuscripts are not available easily, so digitization is the solution for preservation and access of manuscripts. It is observed that at present, scanning is the suitable alternative for storing manuscripts. This maneuver will serve as a sort of guideline for the preservation of such type of manuscripts. Indeed, this is a challenging and promising task but one has to undertake such kind of activity which will not only help librarians, library professionals but the entire humanity as a whole !

7. References

1. Hampson, Andrew, " Managing a digitization project" Aslib managing information, 5: 10, December 1998. pp.25-32.
2. Hampson, Andrew, " Scanning in the right direction" Library Technology 4(6) November 1999. pp.79-80.

3. Kuny, Terry, " An introduction to digitization technologies and issues" Network notes no.14,
National Library of Canada, October 1, 1995.
4. Mahajan, S.G. " Descriptive catalogue of manuscripts available in the Jayakar Library, University of Pune, vol. I part II Marathi manuscripts 1986.
5. University Grant Commission, Manuscript Committee (Chairman Dr. V. Raghvan), Manuscripts catalogues. Bangalore, 1963 p.4
6. University of Pennsylvani, Sanskrit manuscripts available at <http://www.library.upenn.edu/etext/sasia/ski-mss/index>.