Study of 13-17th century Marathi manuscripts: author generated metadata and its mapping with Dublin Core

Shubhada Nagarkar, University of Pune, India.
shubha@bioinfo.ernet.in

Harsha Parekh, SNDT Women's University, India
harsha_parekh@vsnl.com

V.L. Manjul, Bhandarkar O.R. Institute, India.
vasudeomanjul@hotmail.com

Abstract

This paper reports on a study that examined the author generated metadata for 13th –17th century manuscripts. For the study 30 different handwritten Marathi (local language of Maharashtra State, India) manuscripts by Marathi poet saints were analyzed critically especially the start and end note areas (colophon area) from which one can easily identify the author, subject, geographical area, time period of the manuscript, etc. An attempt is made to map these author-generated metadata with Dublin Core elements set. Translated DC elements into Marathi are also used for mapping purpose. The results indicate that authors created good quality metadata to classify the manuscript easily. This research suggests that metadata created by poet saints from 13th – 17th century can be easily mapped with DC elements. There are key metadata elements found in manuscripts that need to be incorporated. The study is very significant as the digitization of these manuscripts is under progress and the basic metadata will be useful for "resource discovery" once the data will be made available on the Internet.
Keywords: Marathi metadata, Marathi manuscripts, Dublin Core registry, 13-17th century manuscripts

1. Introduction

In the 1400s and 1500s, the Maharashtra region went through a religious revival influenced by the bhakti Saints. During the period much of spiritual and religious information was generated and disseminated through personal contacts and orally in public places especially in temples. The author himself or some other person documented this information and preserved it in the form of handwritten manuscritps. Saint poets like Dnyaneshwar, Tukaram, Namdeo, Eknath, Kanhopatra and Chokha Mela who emphasised the devotion aspect of Hinduism, broke new ground by using the vernacular (Marathi) instead of Sanskrit in their songs. Their poetry is popular even today. Besides their religious singificance these manuscripts help in understanding the social and cultural changes in Maharashtra; they also represent the development of the Marathi language.

There are approximately 200,000 Sanskrit and Marathi manuscripts available in Pune city area. About 5000 of these are preserved in the Jayakar Library of University of Pune. Digitization of these manuscripts is under progress since the University of Pune is collaborating with Carnegie Mellon University, USA and Indian Institute of Science, Bangalore under the project "Universal Digital Library" (http://www.ulib.org). This program involves the creation of digitial documents and its appropriate metadata. The digital documents are maintained as image (tiff) files since the development of OCR techniques for Marathi is still under progress. Metadata is in Marathi using unicode (mangal.tif) font. For this purpose catalogues of sanskrit mansucripts that are available are referred to. It has been noticed that a typical Marathi manuscript has a start note which describes title and content, then actual chapters, chapter number at the end of the chapter, and end note area (colophone) describes author, place, day, time, name of commentator, etc. In such a situation, start and colophon of Marathi manuscripts is useful to identify, classify and catalogue the manuscripts. Dublin core metadata standards have been selected for the cataloguing purpose and hence will help for "resource discovery" on the net after its digitization. We have considered only bibliographic metadata to map with DC elements. Descriptive and administrative metadata will be also mapped in near future.

There are two ways currently practiced for generation of metadata, firstly by machine, secondly by human effort. Metadata by human agents is more reliable than machine generated metadata particularly for manuscripts. A study by Greenberg with others found that authors can create good metadata [1]. Marathi metadata has to be drawn from manuscripts themselves. Our attempt has been to look at different sections particularly colophon area of Marathi manuscripts and locate the most obvious places of each DC element. Most of the manuscripts have start note and colophon which

can be considered as "author generated metadata" (figure 1).
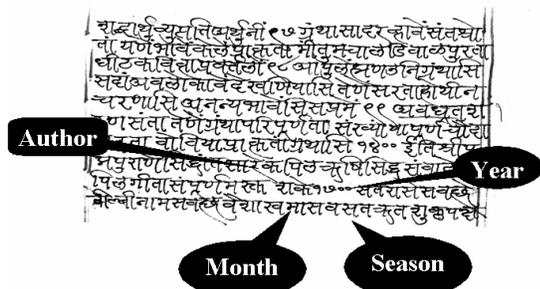
Figure 1. Colophon area with embedded metadata elements

Our attempt here is to compare this area with Dublin Core metadata elements so that anyone who would like to catalogue manuscripts using DC standard can create metadata easily. We have used translated DC elements into Marathi (http://bioinfo.ernet.in/library/dc/ele.txt), which are registered, at DCMI registry (http://www.dublincore.org/dcregistry/index.html).

2. Objectives of the study

1. To study the start and end note area of each Marathi manuscript
2. To map the colophon areas with Dublin core metadata elements.
3. To examine the more qualifiers under DC elements when manuscripts digitized.

3. Preliminary findings

A preliminary analysis of 30 digitized manuscripts (which will be made available at 144.16.85.66) indicates that various DC elements often found at certain specific positions in the manuscripts as indicated in Table 1.

Table 1 – DC elements and their places in Manuscripts

DC Elements	Description/place/location in manuscript
Title	Title page, start note and in colophon area
Creator	Author name in start and end note
Subject	Colophon area and sometimes after title
Description	Briefly described in end note area
Publisher	Author and copier/commentator name on title page, in colophon area, start note
Contributor	In endnote area copier, depositor
Date	Year in "Sake", Day and Time (explained below under key metadata elements)
Type	Colophon area mentioned as original or criticism or commentary etc.
Format	Foliage (on each page right side) number of lines per page, and total no. of pages at the end

Identifier	Accession number
Source	In end note area (Name of original manuscript if it is commentary or copy)
Relation	Colophon area if commentary or copy
Coverage	In colophon area/end note. In absence of geographical area note Illustrations/images (Images) can be used to identify the geographical area, cast, state, etc.
Rights	Endnote explains owner of the manuscript mostly author but also name of owner and the place where it deposited has mentioned.

Language of each manuscript is Marathi and will be entered into the database like MARATHI MR as per Indian ISO standard 639 (http://www.oasis-open.org/cover/iso639a.html)

3.1. Key metadata elements

There are cases where start and colophon areas are absent or damaged, in such cases the illustrations are used to identify the context of the manuscript like geographic region, historical period, religious sect in which the manuscript is created. Since the date in the Marathi manuscripts is recorded according to Saka system of Hindu calendar (1922 for this year) this needs to be converted as per the Gregorian calendar. However scholars still approach the documents using the Saka date. It therefore might be suggested that the date element have additional qualifiers for different systems. Marathi is written in more than one script (Devanagari and modi) therefore the script qualifier should be added under the language element.

4. Conclusion

Marathi manuscripts from this period are distributed across several institutional and personal collections. Our experience of identifying different DC elements from specific regions of the manuscripts will help others to generate metadata for digitized manuscripts. Major DC elements are found embedded in manuscripts themselves in colophon area. To establish bibliographic control of Marathi manuscripts, the author-generated description of their own writings can be rich sources of metadata for today's cataloguer working in electronic environment. This project further illustrates the feasibility of using DC elements in diverse applications.

References

[1] Greenberg, J., Pattuelli M.C., Parsia B. & Davenport R. W. (2001) Author-generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization *Journal of Digital information*, 2(2): Web site: http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/