Development and Application of Dublin Core Metadata Standards in Marathi

Shubhada Nagarkar
Sr. Technical Assistant
Bioinformatics Center, University of Pune, Pune – 411 007
shubha@bioinfo.ernet.in

Harsha Parekh
Librarian and Prof. of Library Science
SNDT Women's University, Mumbai - 400 020
harsha_parekh@vsnl.com

Abstract

To enable Dublin Core metadata to become a global standard for locating information on the Web, it is essential that such metadata be provided in different languages and in a variety of scripts. This would enable search and retrieval of Web documents in all languages. Considering the potentials of DC metadata in resource discovery on the Internet, the current paper reports an attempt to translate Dublin Core metadata elements into Marathi language (the local language of Maharashtra State in India), render them in Devnagari script and apply them to a significant section of Marathi literature – viz. the writings of one of the "saint-poets" of the Middle Ages.

Our efforts are to create a web based database and to assign Dublin Core international metadata standards rendered in Marathi for cataloguing the literature of one of the prominent "saint-poets" - Chokha Mela - available in print and electronic format as well as on Internet. This is conceived as part of a larger project of organizing all the literature of the "saint-poets".

We have chosen a group of "saint-poets" in Maharashtra from the 13th century (e.g. Jnandev 1275-96) to the 17th century (e.g. Tukaram 1608-90) who helped establish the 'Bhakti' (devotional) school of Hinduism in Western India and assisted in its spread from southern to northern India. Even today, the lives of the saint-poets and their literature continue to inspire a large section of the Marathi speaking population and provide them with emotional solace. As such, their writings constitute an important segment of Marathi literature and researchers from all over the world are engaged in studying it. The original writings in Sanskrit (including the early handwritten manuscripts) and commentaries in Marathi on them are scattered at different places in India and are now beginning to make an appearance on the Web.

Objectives of the study

- To translate the DCMES into Marathi (version 1.1).
- To test the translated elements with cultural heritage literature – the literature of the medieval poet saints.
- To find out limitations and add new qualifiers if needed.
- To send translated DC into international registries.

Project description

The project of translation of metadata into Marathi and its application for saint literature (database available at http://bioinfo.ernet.in/~shubha/dc/main.htm) is divided into following phases:

1. Translation: Each of the DC element (with definitions and comments) was translated into Marathi. For this translation work we referred to various dictionaries and subject headings and consulted linguists, grammarians and literature experts. Table 1 shows the basic 15 elements and their translation into Marathi.

2. Rendering them into Devnagari script: Once the translation of the DC metadata elements was completed, the work of actual implementation began. Initially we have used Shiv01.ttf Marathi font, which is available free on the net and is phonetic as well. The work done using Shiv01.ttf font, is based on ASCII character and could not be converted directly into UTF-8 which is widely accepted Unicode standard. However, we are experimenting with the tools available at C-DAC, Pune, (www.cdacindia.com) which help to convert ASCII based fonts into ISCII based ones and then into UTF-8.

Table 1. DC elements into Marathi (Version 1.1)

Title	SaIYa-k	Type	p`kar
Creator	inama-ata	Format	rcanaa
Subject	ivaYaya	Identifier	AaoLK
Description	vaNa-na	Source	]gama
Publisher	p`kaSak	Language	BaaYaa
Contributor	sahBaagaI / sahayyak	Relation	naato / saMbaMQa
Date	idnaaMk / tarI#a	Coverage	vyaaPtI
		Rights	h@k

3. Converting it into UTF-8 character coding: We also found that Microsoft has developed a Marathi Unicode font - Mangal which is available with Windows 2000 and Windows XP. We faced problems in displaying the Unicode font. Efforts are underway to use Microsoft's Web Embedding Font Technology (WEFT) to convert the Unicode font into - Embeded Open Type (EOT). Further we are working to use Mangal.eot with special java script that enables users to see the Devanagari script on any platform viz. Windows 95, 98, 2000 and NT with any version of IE 4.0 + and Netscape communicator version 4.3+.

4. Metadata registry: The Marathi DC elements are being sent to the DC registry at OCLC, USA (http://wip.dublincore.org:8080/dcregistry/index.htm) and to the registry maintained at University of Library and Information Science (ULIS), Open Metadata Registry, Tsukuba (http://avalon.ulis.ac.jp/registry/).

Discussions

Our primary attempt here was to translate the DC elements into Marathi. DC translation into Marathi was fairly simple and we found appropriate terms in Marathi. A blind reverse translation indicated that in almost all cases the Marathi term for the DC element was correctly translated back to the English term as well as properly understood. In some cases (e.g. identifier and relation) the understanding and the reverse translation indicated that it was necessary to provide the context by the detail qualifiers. Thus at the end of the process we have come up with DCMES Marathi translation (Unicode based) version 1.1(http://bioinfo.ernet.in/~shubha/element.txt). We are continuing to test these elements with a wider range of materials – including hand-written palm leaf manuscripts, music, films, etc. As the database grows to include this material on Chokha Mela and expands to cover other saint-poets, we will be able to test, improve and enhance the translation of DC metadata. Updated versions of this translation in Unicode fonts will be made available on the said web site and with the registries.