Noolaham Foundation

Project Report

Project title	Marumalarchi Digitization Project 2005-2008
Project number	NF/PG/2005/0001
Project location (District, Division, Village)	Colombo, Jaffna
Sector	Digital Archive of Print Publications
Implementing agency and contributors	Implementing agency - Noolaham Foundation, Contributors - Thelivatai Joseph, T. Thevathas, M. Mauran, P. Eelanathan and Kuganenthiran
Grant Agency and Contribution	Individuals of Noolaham Foundation.
Total Budget and Expenditure	N/A
Start and End date	Year 2005 to 2008
Responsible Stakeholder/s	Team leader - Mr. Kopinath Thillainathan, Ms. Sinduja Kopinath

Summary

Marumalarchi Digitization Project was one of the pioneer and earlier digitization initiatives which contributed towards Project Noolaham which evolved into Noolaham Foundation. Modeled on the Project Gutenberg approach, the Project Noolaham was started in 2005 of with a view to creating a fully volunteer contributed online e-book library for the Tamil speaking communities. The Marumalarchi Digitization Project initiative was started in 2004 and ran all the way to 2008. At the beginning of this project, the documents were typed. Adding e-books during the early part of his project was quite slow, due to time constrains. Later scanning was introduced through this Project and it became the accepted digitization method at Noolaham Foundation.

A total of 219 (55 books typed and 164 books scanned) books were digitized through this project and made available online through Noolaham Foundation's Digital Library www.noolaham.org. This project was a successful initiative and received special appreciation from Sri Lankan Tamil writers and publishers.

Introduction and Background

The digitization of Sri Lankan Tamil books was first initiated by Mr. R. Pathmanabalyer. In this effort, he digitized 40 Sri Lankan Tamil books in 1999 and 13 were added online through 'Project Madurai' in the latter part of 2000 (Project Madurai was started in 1998). Project Eelanool, which was initiated in 2004 by Mr. Muralitharan Mauran and Mr. Thillainathan Kopinath, was the first project solely for digitizing Sri Lankan Tamil books. To avoid

duplication and to encourage synergy effect, Project Noolaham was started incorporating Project Madurai ebooks and Eelanool ebooks. This project could be considered a test bed of Project Noolaham.

The implementation process: At the beginning, books were typed as text and converted to Unicode. After proof reading, they were converted to html format and were uploaded at Noolaham Digital Library. Later, the scanning technology was introduced and was adopted as the accepted digitizing method. Scanning is the method of digitizing printed texts so that the scanned image can be electronically edited, searched, stored more compactly, displayed on-line, and used in machine processes. Epson and HP scanners were used for this project. Adding to this process, various digitizing standards were tested on the ground to ensure the products are compatible with internationally acceptable digitizing and archiving standard. It has been identified, that TIFF [1] and BMP [2] formats are the ideal formats for these purposes.

This project was implemented in Colombo. Mr. Thillainathan Kopinath and Mrs. Sinduja Kopinath contributed to the digitization. The scanners were also bought by them. A number of individuals including Mr. Kuganenthiran, Mr. Thelivatai Joseph, Mr. Thambiaiya Thevathas, Mr. Muralitharan Mauran and Mr. Puvanenthiran Eelanathan gave the books and magazines to be digitized. Some publications were brought by individuals as well. After digitization, those publications were either returned to the contributors or donated to libraries.

Objective/s and Achievements or Results:

Two hundred and nineteen books and various kinds of documents were digitized through this project. The following books were uploaded on Noolaham Foundation's Digital Library www.noolaham.org.

Typed and digitized: (55 books)

15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 27, 29, 30, 31, 33, 35, 36, 38, 43, 44, 45, 52, 53, 67, 73, 75, 81, 82, 102, 103, 105, 106, 107, 108, 114, 119, 121, 122, 123, 144, 145, 146, 147, 148, 149, 150, 165, 166, 167, 168, 169, 170, 171, 172, 249

Scanned and digitized (164 books)

14, 15, 16, 18, 21, 23, 27, 29, 31, 33, 36, 52, 77, 103, 105, 107, 121, 122, 125, 144, 145, 146, 150, 165, 166, 167, 170, 172, 178, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296,

---

[1] TIF is an image format file for high-quality graphics. TIF files are also called .TIFF, which stands for "Tagged Image Format File. http://whatis.techtarget.com/fileformat/TIF-Tag-image-bitmap-file-TIFF

[2] The BMP file format, also known as bitmap image file or device independent bitmap (DIB) file format or simply a bitmap, is a raster graphics image file format. http://en.wikipedia.org/wiki/BMP_file_format

297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 324, 381, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1642, 1683, 1699, 1700, 1701, 1702, 1703, 1704, 1705, 1706, 1707, 1708, 1709, 1710, 1711, 1712, 1713, 1714, 1765, 1777, 1865, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, 1904, 1905, 1906, 1907, 3440, 3441

Constrains / Challenges and lesson learned:

- This project followed the Project Gutenberg and Project Madurai approach which depended fully on contribution by volunteers. But it was found that this model was not suitable for digitization projects which target a huge number of documents from all Sri Lankan Tamil speaking communities. Projects like this one require more contribution from volunteers for resource mobilization.

- When it comes to typing, proof reading was the biggest challenge and we were unable to ascertain the quality. The documentation aspect was also absent since we typed   the texts.

- Some of the typed or scanned books which were obtained through volunteer contribution were discarded as they could not meet the required standards.

Suggestions and Recommendations:

- Scanning was accepted as the standard method of documentation for Noolaham's initiatives.

- Tamil Optical Character Recognition has been a frontline research area in the field of human-machine interface for the last few years. When it is improved, it will save more time and solve some of the technical issues . We need to discuss with relevant experts and get their guidance to find the best solutions to the technical difficulties that we faced during this project.