University of Kashmir, India

From the SelectedWorks of NADIM AKHTAR KHAN

2010

Designing Digital Library of Perso-Arabic Script: An Experiment.

Nadim Akhtar Khan
Rosy Jan
Sheikh Shazia

Available at: https://works.bepress.com/nadim_khan/9/
Designing Digital Library of Perso-Arabic Script: An Experiment

Nadim Akhtar Khan
Assistant Professor, Department of Library and Information Science, University of Kashmir, e-mail: nadim@kashmiruniversity.ac.in

Rosy Jan
Assistant Professor, Department of Library and Information Science, University of Kashmir, e-mail: rosy@kashmiruniversity.ac.in

Sheikh Shazia
Student, Department of Library and Information Science, University of Kashmir

Abstract
The Greenstone digital library software is a comprehensive system for building and distributing digital library collections providing a new way for organizing and publishing information on the Internet. The paper describes how multilingual digital library collections can be created and customized using various features available in Greenstone Digital Library Software. It is an attempt towards creating and managing digital library collection in different scripts for M. Phil and Ph. D theses submitted to University of Kashmir in Arabic, Persian, Kashmiri and Urdu.

Keywords: Multilingual library, Greenstone, Open source software, Theses, Unicode

Introduction
In Digital Library sector, most of the research and development activities are concentrated in English and European languages. Since more than half of the world literature being published in languages other than English, the question of multilingual access and multilingual information retrieval is becoming increasingly important. A multilingual digital library need all functions operate simultaneously in as many languages as desired and whose search & retrieve functions are language independent. Thus libraries world over are now looking forward to set up digital libraries using Multilingual Digital Library software. Greenstone is one such suite of software for building and
distributing multilingual digital library collections that provides a way of organizing information and publishing it on the Internet or on removable media.

Jammu & Kashmir state being an important place along silk route has been greatly influenced by oriental literature and languages, thus much of the Literature available in libraries and archives is in non-roman script. The bibliographic records pertaining to such materials are still being transliterated and as such much of their use is affected. University of Kashmir stands no exception to this scenario. Much research has being carried out in Arabic, Persian and Kashmiri languages and such thesis need to be digitized and provided access in their native scripts. The present work is a step towards creating a bibliographical collection of such theses in their native script taking full advantage of multilingual support available in Greenstone

Problem

English/European languages which are written from left to right have been always been used for information processing and retrieval world over. The representation of other documents in their native script has always been problematic. Languages like Arabic, Persian, Urdu, and Kashmiri need a special attention, as much of the literature in these languages is transliterated using English language. The present work is an experiment to create a digital collection of bibliographic details of select Non-Roman script Ph. D and M. Phil theses submitted to University of Kashmir in the departments of Arabic, Kashmiri, Persian and Urdu in their native scripts using Greenstone Digital Library Software. The experiment can pave way for creating digital library of bibliographic and full records of the literature available in the state in various scripts taking advantage of Unicode supportive fonts.

Scope

The scope of the present work is confined to create a digital collection of bibliographic details of Ph.D and M. Phil theses submitted to the University of Kashmir in the departments of Arabic, Kashmiri, Persian and Urdu since their inception till date. The bibliographic details collected include the name of the Scholar, name of the Supervisor, Title of the work, Year of submission, nature of the Degree (PhD or M. Phil), name of the Department and a brief Abstract.
Objectives

The objectives of the study are as under:

➢ To collect the bibliographic details of theses available in Persian, Arabic and Kashmiri languages

➢ To make the MS-Word file for each bibliographic record in native script

➢ To create new set of metadata elements for identifying each record like Name of Scholar, Name of Supervisor, Title of Work, Year of submission, nature of the Degree (PhD or M. Phil), name of the Department and a brief Abstract.

➢ To create the digital collections of the theses using Greenstone Digital Library Software.

➢ To assess the suitability of Greenstone Digital Library Software for developing digital library collection in Urdu, Arabic and Kashmiri language.

➢ To search and browse the collection of Urdu, Arabic, Persian and Kashmiri theses up to utmost satisfaction.

Methodology

This digital collection of MPhil and PhD theses submitted to the departments of Arabic, Kashmiri, Persian and Urdu consist of around 211 records were used as the base for the work of creating the digital collection using the Greenstone Digital Library Software. The work was carried out in the following steps:

➢ Each thesis was consulted individually so as to gather the bibliographic details such as the name of the Scholar, name of the Supervisor, Title of the work, Year of submission, nature of the Degree (PhD or M. Phil), name of the Department and a brief Abstract.

➢ The details of each thesis were recorded in MS Word i.e., word files were created for recording the bibliographic details of the M. Phil and PhD. theses.

➢ The collections of Urdu and Persian theses were written using the Open type font Nafees Pakistani Naskh v2.0 and Nafees Tahreer Naskh v1.0.

➢ The Arabic and Kashmiri records were typed using Arabic transparent and Kashmiri font respectively.
The word files were included in the Greenstone Digital Library Software (Version: 3s 2009.04.07).

Metadata elements like name of the Scholar, name of the Supervisor, Title of the work, Year of submission, nature of the Degree (PhD or M. Phil), name of the Department and a brief Abstract were provided for each collection.

Literature review

A vast amount of literature has been published discussing the topic of the digital libraries and Open Source Software (OSS). As per Liew (2009), since the launch of the Digital Library Initiative (DLI) projects in 1996, there has been substantial attention accorded to digital libraries research and development. Concurrent with this interest has been a significant emphasis on the digital library (DL) phenomenon in academic research. Over the past 11 years, there has been a tremendous growth in the number of journals dedicated to discussion of digital libraries (e.g. International Journal on Digital Libraries, D-Lib Magazine) and various conferences aimed specifically for both professionals and academics working on digital libraries (e.g. The International conferences on digital libraries (ICDL), the European conferences on digital libraries (ECDL), the International conferences on Asian digital libraries (ICADL), the Joint ACM/IEEE conferences on digital libraries).

Witten & Bainbridge (2001) describes the Greenstone Digital Library Software as a comprehensive, open-source system for the construction and presentation of information collections. He further says that collections built with Greenstone offer effective full-text searching and its metadata-based browsing facilities are attractive, easy to use, and easily maintainable and can be augmented and rebuilt entirely automatically. Witten & Bainbridge (2004) expressed that Greenstone digital library software provides a flexible way of organizing information and publishing it on the Internet or removable media such as CDROM. Its aim is to empower users, particularly in universities, libraries and other public service institutions, to build their own digital libraries. Witten & Bainbridge (2005) introduced Greenstone and explain how librarians use it to create and customize digital library collections. Discussing its features they point out that through an end-user interface, users may add documents and metadata to collections, create new collections whose structure mirrors existing ones,
and build collections and put them in place for users to view. With Greenstone entering its second
decade, *Witten (2007)* takes a retrospective look at its development addressing the challenges that have
been faced, and the lessons that have been learned in developing and deploying a comprehensive open-
source system for the construction of digital libraries internationally with particular emphasis on
developing countries.

*Wan (2006)* provides quite a few examples in his article to illustrate how information
visualization techniques and applications, such as Digital Library Visualizer, Fluid Treemap Interface,
ActiveGraph, 3-D Vase Museum, Collaborative Visual Interfaces, and AquaBrowser, could be
integrated into digital libraries to help users perceive information better and assist librarians in
discussed the meaning of char set, encoding & fonts, and trained the participants in creating interface in
the native languages for GSDL, setting default interface language, handling multilingual content,
Unicode text, Non-Unicode text & the limitations involved therein.

**The Experimental Multilingual Digital Library**

The provision for creation of metadata sets based on native scripts has explored ways for
browsing and searching digital collections in respective languages. The Fig 1.1 displays various titles
available in kashmiri collection. Fig 1.2 and 1.3 demonstrate full record available under Mphil theses in
kashmiri and Arabic languages respectively. Fig 1.4 displays collection in Urdu Language and 1.5
displays collection as browsed under supervisor.
Fig 1.1. Screen displaying the list of titles included in Kashmiri collection.
Fig 1.2 Details of one of the records for M.Phil thesis in Kashmiri Language

[[Image of text in Kashmiri]]

- **Title:** موضوع
- **Author:** نظر کیاری
- **Department:** مقدمات مباحث
- **Year:** 2006
- **Library:** ایم فل
 الموضوع: أحمد فارس الشماق ومساهماته في تطوير الأدب العربي.

الباحث: ظهير أحمد العلامة

تحت إشراف: الدكتور شاكر نسيب

لعام: 2004

قسم اللغة: العربية

قل شهادة: أمه فل

استخراج:

هذا الموضوع يحلو للاستدلال، إذ أن أحمد فارس الشماق في تطوير الأدب العربي، كما يحث الباحث على بداية نموذج، نموذج على عام 1798، وطلبه من هذه السنة، هي حملة نقدية وعلمية. هذه الجاهل قد انتشرت العرب باللسان العربي. وفي زمن هذا الحالة، كان أحمد فارس الشماق من رواد حركة التجديد في الأدب العربي، وتحت إشراف، في نموذج حقل اسم "الزنك" أو "الزنك" الأدبي. وتحت إشراف، هذه الحالة في نموذج تجديد الأدب العربي، وتحت إشراف، هذه الحالة في نموذج تجديد الأدب العربي، وتحت إشراف، هذه الحالة في نموذج تجديد الأدب العربي، وتحت إشراف، هذه الحالة في نموذج تجديد الأدب العربي، وتحت إشراف، هذه الحالة في نموذج تجديد الأدب العربي.
Fig 1.4 Details of one of the records for M.Phil thesis in Urdu Language

عنوان
"انگر" اور تفسیم مئکا کرب
مقدمات
محمد شفیع دار
تکرار
دکتر محمد حسین امین
شعب: اردو
سال: 2006
مکان: نارو: ایم فن
تجلد:

به تحقیق مقدماتی 4 عنوانات کی تناول پر بنا یہ خلاصه میں ایک تعریف جس میں اندازہ کی سانی حدید مستور کی شخصیت اور آن کیں کہ کیا، یہاں بھی مقبول میں نئے عناوین میں اور انتخابات میں نئے عناوین میں ںبہا شاہ شاہ نومال ہے اور توازن ہے کی انتخابات میں دوز وراثت کا تعلیم شاہ شاہ میں مسلمان نتائج "انگر" کا تعلیم شاہ شاہ میں مسیح کا یہاں بھی چیس میں سرکاری دو نتائج
کی نتائج "انگر" کا بور جانے میں "انگر" میں تفسیم کی اہمیت کی پیش کشا "انگر" میں کردار
تکرار "انگر" میں مسلمان گئی اور میں پیش کی کی نویس
Fig 1.5. Persian Collection as browsed under Supervisor’s name.

Findings and Conclusion

With the its state of art technology Greenstone Digital Library software technology is able to process multilingual script materials and also provides facilities for browsing the collection by setting the search preferences through various languages. Unicode transformation format (UTF-8) should be chosen from preferences page of the web explorer to view Non-Roman materials. Additionally, we need to have particular font installed in the fonts folder of the operating system. Managing Kashmiri script materials is still a difficult task as no established Keyboard support is still available. So Perso-Arabic font was used for creating Kashmiri Collection. Urdu script has been successfully written using Open type font Nafees Pakistani Naskh v2.0 and Nafees Tahreer Naskh v1.0. The scope of the work can
further be enhanced by making attempts on Information Visualization so as to make information retrieval more user friendly with more visible display of data. True type fonts have not yet been fully able to be used for representation of many non-roman scripts, therefore attempts need to be made for the same.
References


