

Enabling Indic Support in Library Information Systems

(An Opensource Localizer's Perspective)

By

Indranil Das Gupta <indradg@icbic.com>
Najmun Nessa <najmunnessa@yahoo.com>

About us :

Indranil Das Gupta

Opensource Technology Consultant

Developer-Member, Ankur Bangla Group (www.bengalinux.org)

Specializing in Localized Low-Cost Computing (www.l2c2.org)

Najmun Nessa

Asst. Librarian, West Bengal University of Technology (www.wbut.net)

Founding Member, Indian Koha Interest Group.

L10N Support & Libraries : Socio-Economic Impact

- ✓ Universal access to information & (digital) knowledge stores
- ✓ Preservation of multi-lingual content (traditional knowledge)
- ✓ Address non-English speaking segments
- ✓ Empowerment
- ✓ Build bridges across the Digital Divide

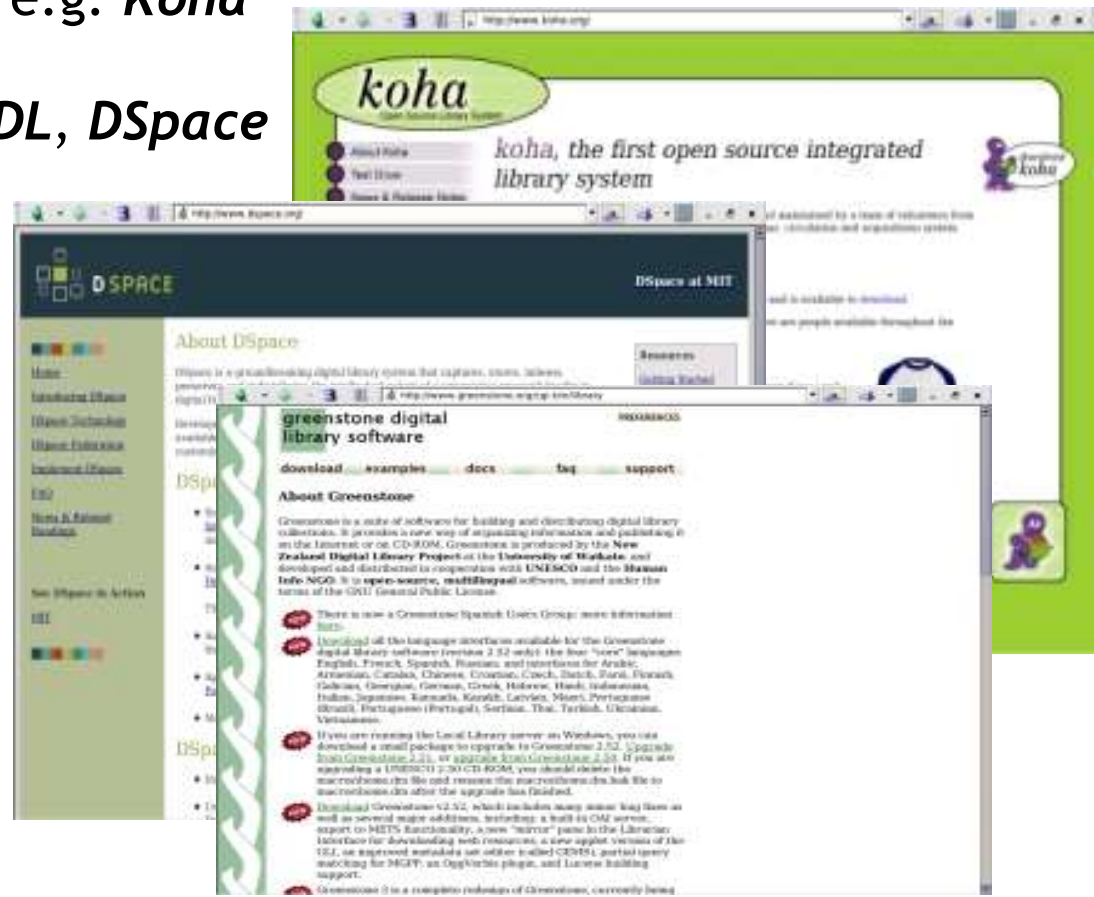
The Old and the New : A Paradigm Shift

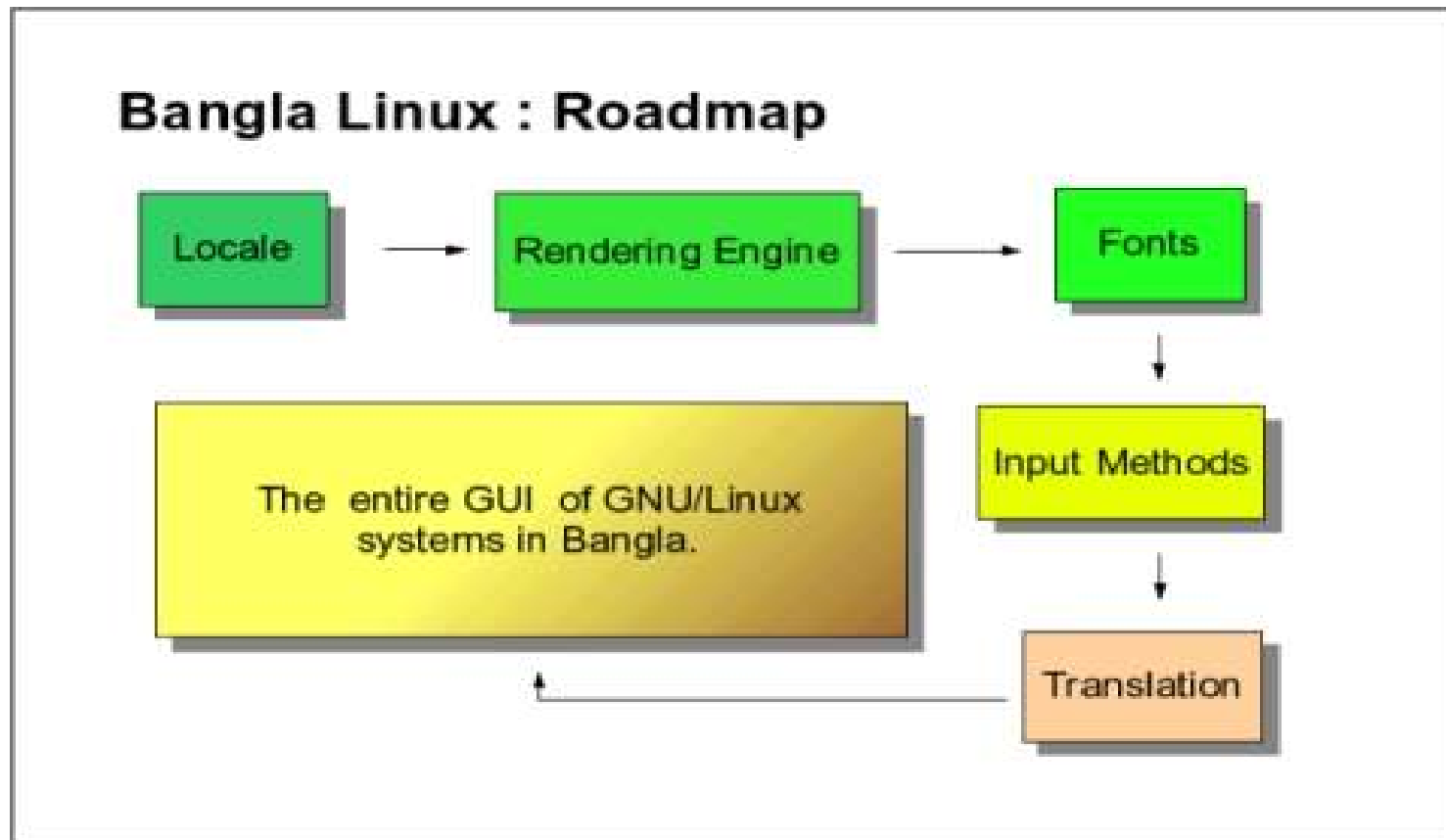
- ✓ Lack of (de-facto, universally accepted) standards
 - Proprietary fonts, encodings etc.
 - Lack of support at Operating System & Application-level
 - Non-participatory (little interaction between policy/standards makers, developers and users)

- ✓ Universal standards based on/around Unicode.
 - Open standards (OpenType Fonts, Unicode, Transport Protocols, Language Specifications etc)
 - Support at Operating System & application level
 - Well-established, participatory Opensource processes
 - Vibrant user-developer community

Localizing Library Information Systems

- ✓ Integrated Library Systems (ILS) e.g. *Koha*
- ✓ Digital Library Software e.g. *GSDL*, *DSpace*





The Obstacles in mass Unicode Adoption

- ✓ Lack of availability of OpenType fonts
- ✓ Input methods (keyboard drivers)
- ✓ Rendering Engine bugs
- ✓ Inaccurate/incomplete character-sets
- ✓ Web browsers (Mozilla/FireFox)
- ✓ Lack of collation sequence data
- ✓ Porting Tools
- ✓ **Lack of awareness!**

The Obstacles in mass Unicode Adoption (an example)

সাহায্য <- incorrect

জুদ <- incorrect

সাহায্য <- correct

জন্ম <- correct

The va-nhala issue #1

The ha-nhala issue

অ্যা <- incorrect

অ্যা <- correct

The va nhala issue #2

Localizing Koha - The Prerequisite Checklist

“Koha was built using Perl scripting language, MySQL Relational Database Management System, and Apache Web Server, running on the GNU/Linux Operating System; however, it has been ported to other operating systems (including Windows) and should be compatible with any system running an SQL database, a web server, and Perl.”

Joshua Ferraro, Koha Developer

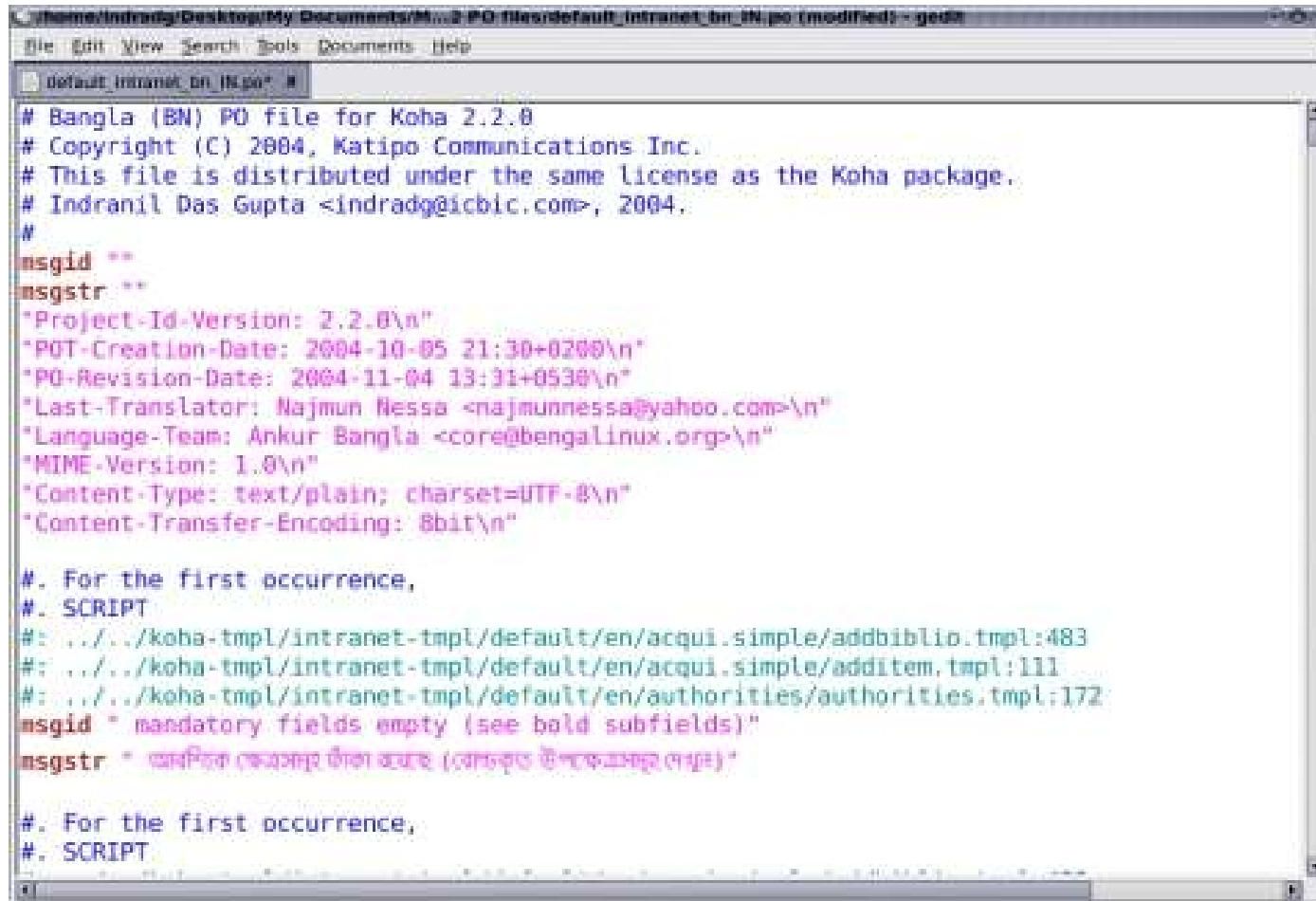
- ✓ The RDBMS must supports Unicode and UTF-8
- ✓ Server-side scripting engine must support Unicoded strings / regexp
- ✓ The web server must handle UTF-8 as a native CHARSET
- ✓ The web browser must support and accurately render the Indic Script

Localizing Koha - The Checks and Measures

- ✓ Use of latest version of MySQL for its implicit Unicode capabilities
- ✓ Apache web-server tested using **AddCharset** Directive set to UTF-8 and trapping the http headers for a test page with its **CHARSET** attribute set to UTF-8
- ✓ Database Unicode compliance checked using phpMyAdmin tool
- ✓ The browser's Unicode rendering got thoroughly tested during these.
- ✓ Perl did not need any testing as its well-known for its UTF handling

Localizing Koha - Translation (work-in-progress)

- ✓ Koha uses the de-facto GNU gettext package for i18n / l10n

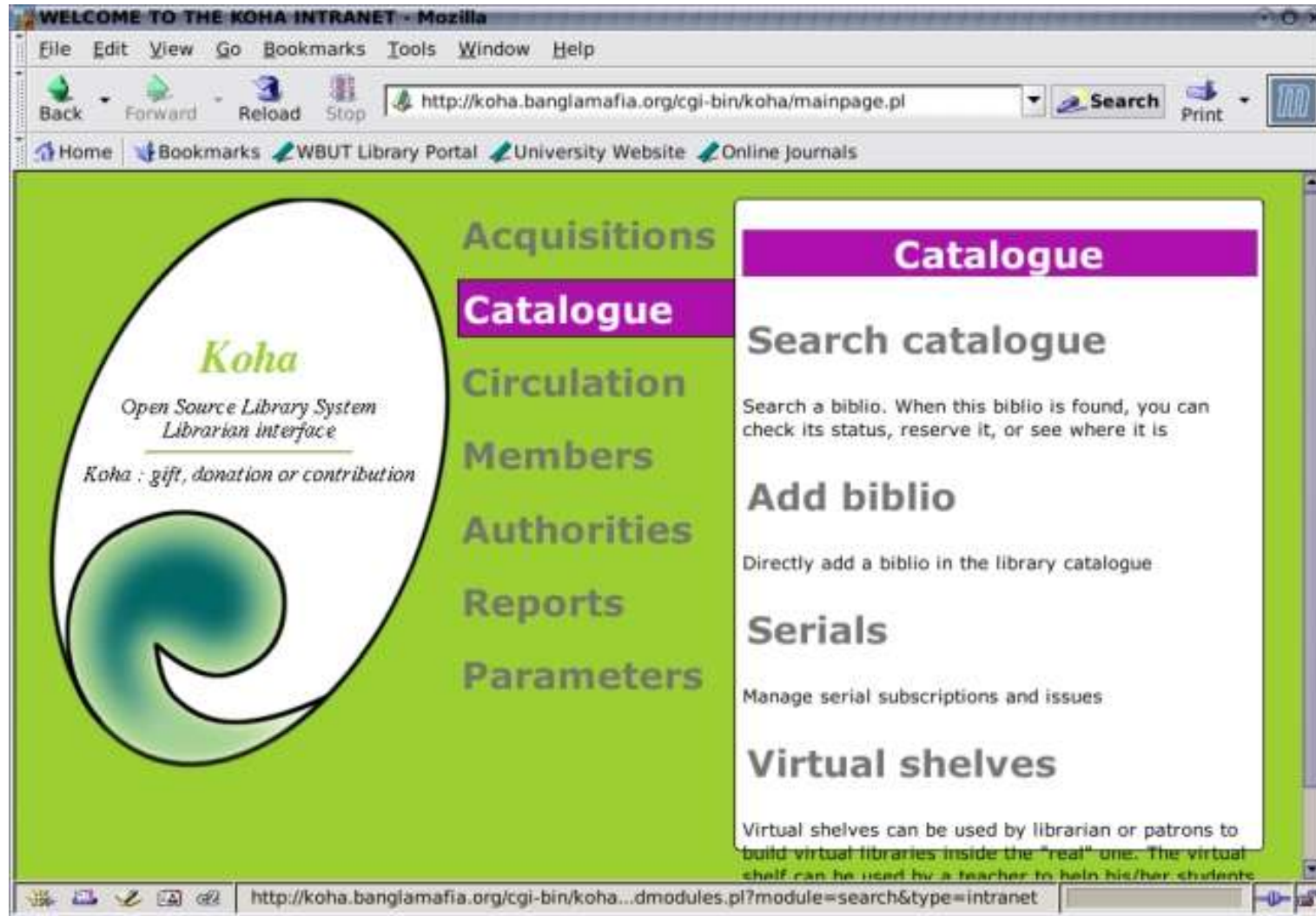


```
C:\home\Indranil\Desktop\My Documents\My...2 PO files\default_intranet_bn_IN.po (modified) - gedit
File Edit View Search Tools Documents Help
default_intranet_bn_IN.po *
# Bangla (BN) PO file for Koha 2.2.0
# Copyright (C) 2004, Katipo Communications Inc.
# This file is distributed under the same license as the Koha package.
# Indranil Das Gupta <indradg@icbic.com>, 2004.
#
msgid ""
msgstr ""
"Project-Id-Version: 2.2.0\n"
"POT-Creation-Date: 2004-10-05 21:30+0200\n"
"PO-Revision-Date: 2004-11-04 13:31+0530\n"
"Last-Translator: Najmun Nessa <najmunnessa@yahoo.com>\n"
"Language-Team: Ankur Bangla <core@bengalinux.org>\n"
"MIME-Version: 1.0\n"
"Content-Type: text/plain; charset=UTF-8\n"
"Content-Transfer-Encoding: 8bit\n"

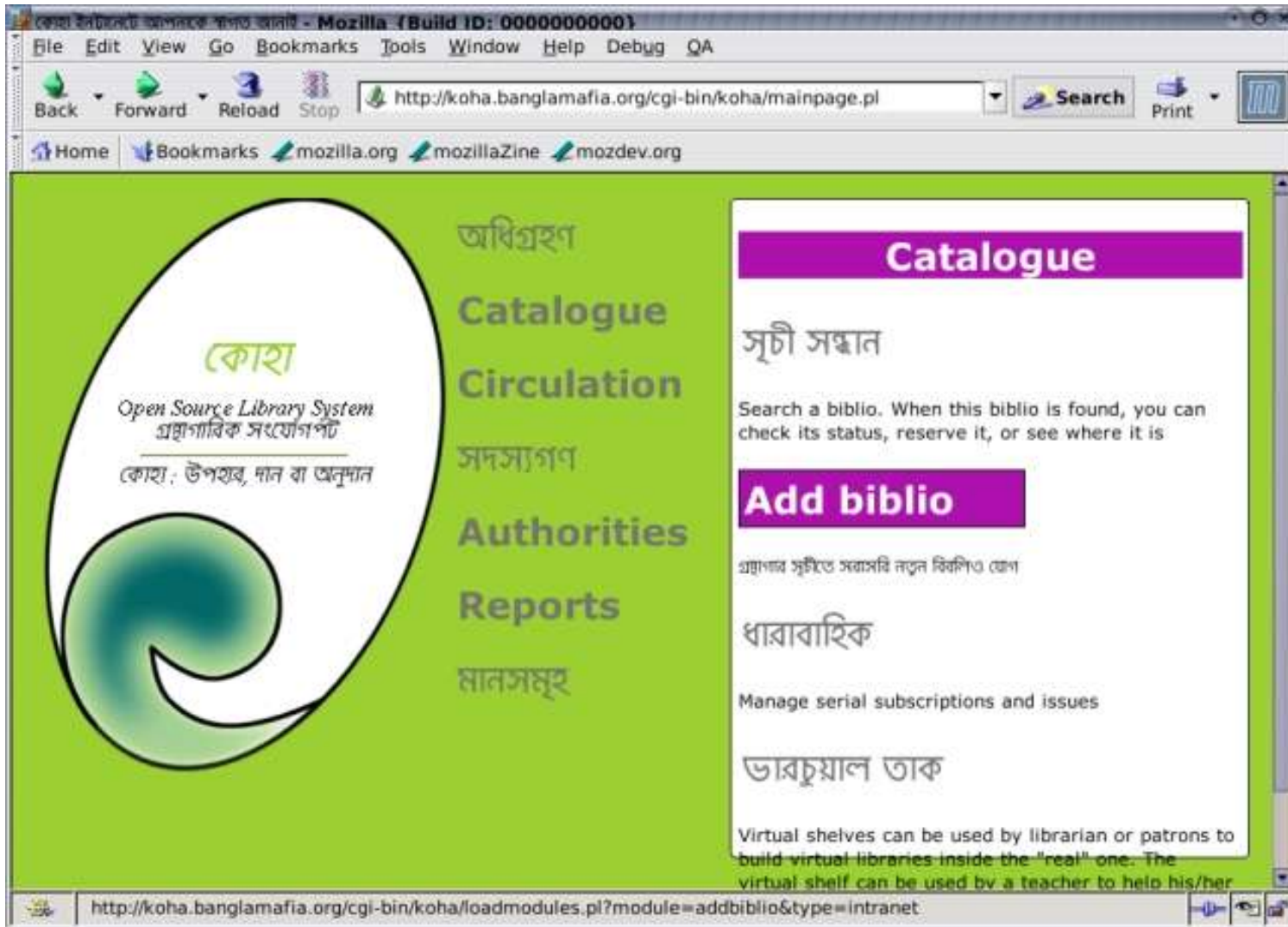
#. For the first occurrence,
#. SCRIPT
#: ../../koha-tmpl/intranet-tmpl/default/en/acqui.simple/addbiblio.tmpl:483
#: ../../koha-tmpl/intranet-tmpl/default/en/acqui.simple/additem.tmpl:111
#: ../../koha-tmpl/intranet-tmpl/default/en/authorities/authorities.tmpl:172
msgid "mandatory fields empty (see bold subfields)"
msgstr "আবশ্যিক ক্ষেত্রসমূহ খালি রয়েছে (বোল্ডকৃত উপক্ষেত্রসমূহ দেখুন)"

#. For the first occurrence,
#. SCRIPT
```

Localizing Koha - A work-in-progress (Koha's Admin Home Page)



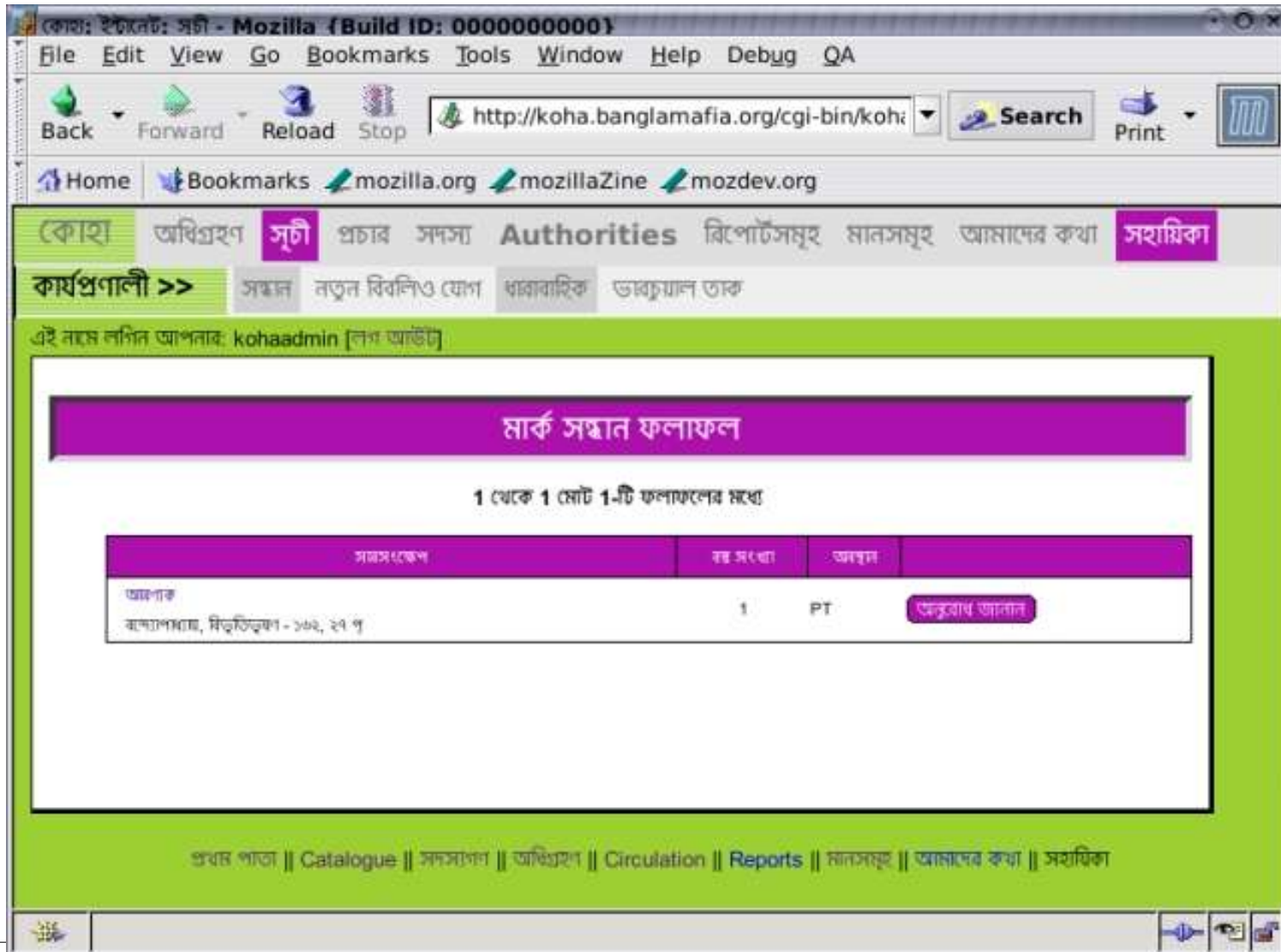
Localizing Koha - A work-in-progress (Koha's Admin Home Page)



Localizing Koha - A work-in-progress (Catalog Search Interface)

The screenshot displays the Koha catalog search interface in Bengali. The browser window title is "কোহা: ইন্ট্রানট: সচী - Mozilla (Build ID: 0000000000)". The address bar shows the URL: <http://koha.banglamafia.org/cgi-bin/koha/search.marc/search.pl?type=intranet>. The interface includes a navigation menu with items like "কোহা", "অধিগ্রহণ", "সচী", "প্রচার", "সঙ্গ", "Authorities", "রিপোর্টসমূহ", "মানসমূহ", "আমাদের কথা", and "সহায়িকা". Below the menu, there are tabs for "কার্যপ্রণালী >>" and "সময়", "নতুন বিবিসিও যোগ", "ধারাবাহিক", "ডাটাবেস ডাক". A message states: "এই নামে লগিন আপনার: kohaadmin [লগ আউট]". The main content area is titled "সচী সন্ধান" (Catalog Search) and contains two sections: "এর ভিত্তিতে সন্ধান করুন" (Search by) and "Other options". The "Search by" section has input fields for "যে কোন শব্দ", "আখা", "গ্রন্থকাল", "বিষয়", "বছর", and "গ্রন্থালয় শাখা". The "Other options" section has input fields for "বন্ধুর ক্রমক্রম", "ডাক সংখ্যা", "আই-এস-বি-এন", and "প্রকাশক". Below these sections is a section titled "আরও ফিল্টারসমূহ" (More filters) with three rows of input fields and dropdown menus. The browser status bar shows the URL: <http://koha.banglamafia.org/cgi-bin/koha/members/members-home.pl>.

Localizing Koha - A work-in-progress (Catalog Search Results)



কোহা: ইন্টারনেট: সূচী - Mozilla (Build ID: 0000000000)

File Edit View Go Bookmarks Tools Window Help Debug QA

Back Forward Reload Stop http://koha.banglamafia.org/cgi-bin/koha Search Print

Home Bookmarks mozilla.org mozillaZine mozdev.org

কোহা অধিগ্রহণ সূচী প্রচার সদস্য Authorities রিপোর্টসমূহ মানসমূহ আমাদের কথা সহায়িকা

কার্যপ্রণালী >> সন্ধান নতুন বিবলিও ফোন ধারাবাহিক ভারচুয়াল তাল

এই নহেম লগিন আপনাব: kohaadmin [লগ আউট]

মার্ক সন্ধান ফলাফল

1 থেকে 1 মোট 1-টি ফলাফলের মধ্যে

সমসংক্ষেপ	বই সংখ্যা	অনুসূচ	
আগণক বন্দ্যোপাধ্যায়, মিত্ততিত্বয় - ১০২, ২৭ পৃ	1	PT	অনুরোধ আনান

প্রথম পাতা || Catalogue || সদস্যগণ || অধিগ্রহণ || Circulation || Reports || মানসমূহ || আমাদের কথা || সহায়িকা

DSpace - A sneak peek

- ✓ Creating multi-lingual (Indic) Repositories
- ✓ Browsing the Indic Content
- ✓ Searching the repository

Logged in as
indradg@banglamaf...
(Logout)

Search DSpace:

 Go

[Advanced Search](#)

[Home](#)

Browse

[Communities
& Collections](#)

[Titles](#)

[Authors](#)

[By Date](#)

Sign on to:

[Receive email
updates](#)

[My DSpace](#)
authorized users

[Edit Profile](#)

[Help](#)

[About DSpace](#)

[Digital Library @ Pailan Group of Institutes](#) >

Communities and Collections

Shown below is a list of communities and the collections and sub-communities within them. Click on a name to view that community or collection home page.

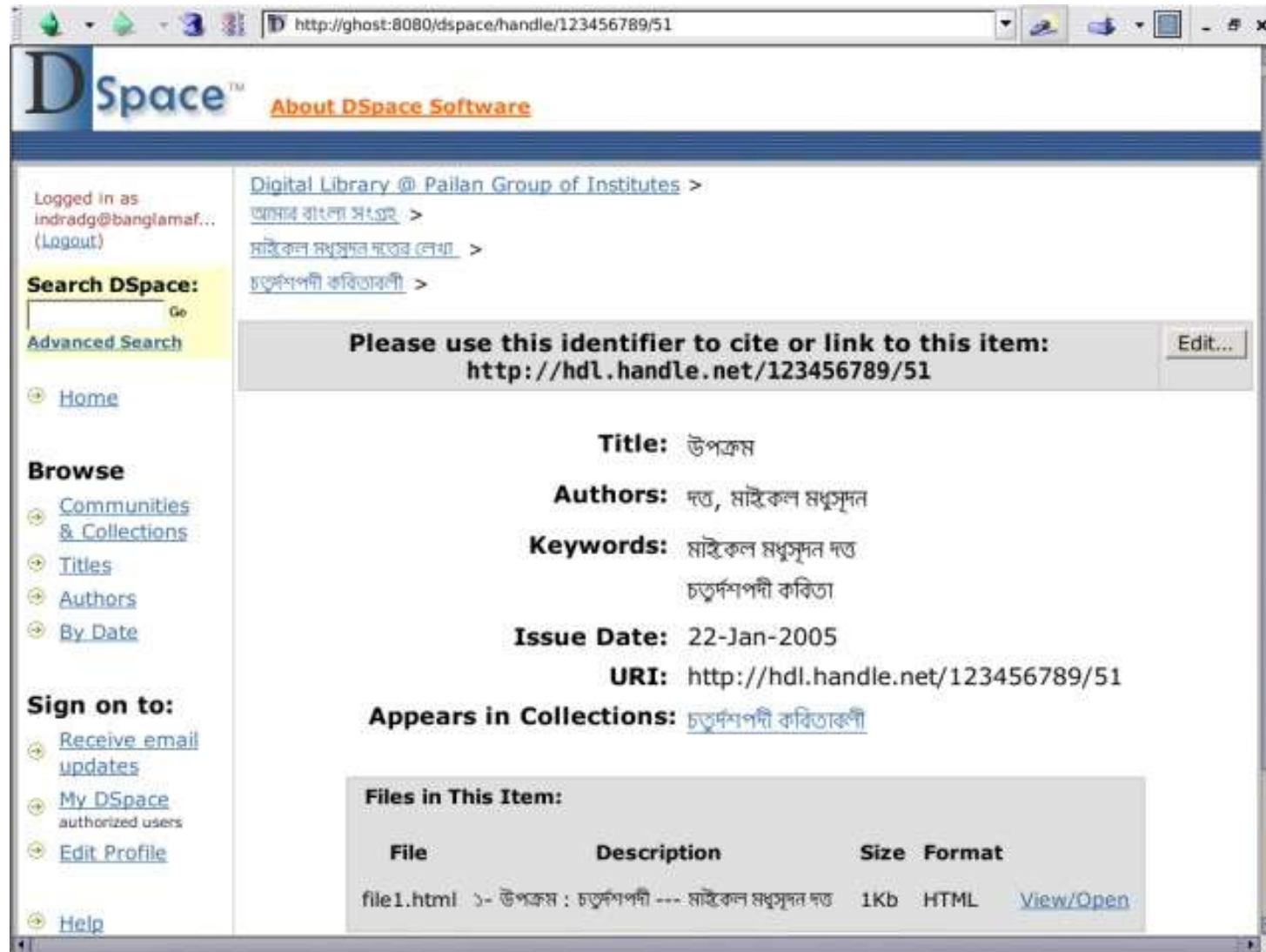
- [আমার বাংলা সংগ্রহ](#)
 - [সুকারের কবিতা](#)
 - [মাইকেল মধুসূদন দত্তের লেখা](#)
 - [চতুর্দশপদী কবিতাবলী](#)
- [रविशकर श्रीवास्तव की रचना](#)
 - [गज़लों का रचना संसार](#)
- [ICONIP 2004](#)
 - [Computational Neuroscience](#)
- [Pailan Digital Library](#)
 - [Masters in Business Administration \(MBA\)](#)
 - [Masters in Computer Applications \(MCA\)](#)
- [Test Tamil Community](#)
 - [யாமறிகந்த மொழிகளிலே தமிழ்மாழி போலடு](#)

Admin Tools

[Create Top-Level Community...](#)

[Admin Help...](#)

DSpace - A sneak peek (Browsing Indic Content)



The screenshot shows a web browser window displaying a DSpace digital library record. The browser's address bar shows the URL: <http://ghost:8080/dspace/handle/123456789/51>. The page header includes the DSpace logo and a link to "About DSpace Software".

On the left side, there is a navigation menu with the following options:

- Logged in as [indradg@banglamaf...](#) (Logout)
- Search DSpace:** [Search box] Go
- Advanced Search**
- [Home](#)
- Browse**
 - [Communities & Collections](#)
 - [Titles](#)
 - [Authors](#)
 - [By Date](#)
- Sign on to:**
 - [Receive email updates](#)
 - [My DSpace](#) authorized users
 - [Edit Profile](#)
- [Help](#)

The main content area displays the following information:

Digital Library @ Pailan Group of Institutes >
[আমার বাংলা সংগ্রহ >](#)
[মাইকেল মধুসূদন দত্তের লেখা >](#)
[চতুর্দশপদী কবিতাবলী >](#)

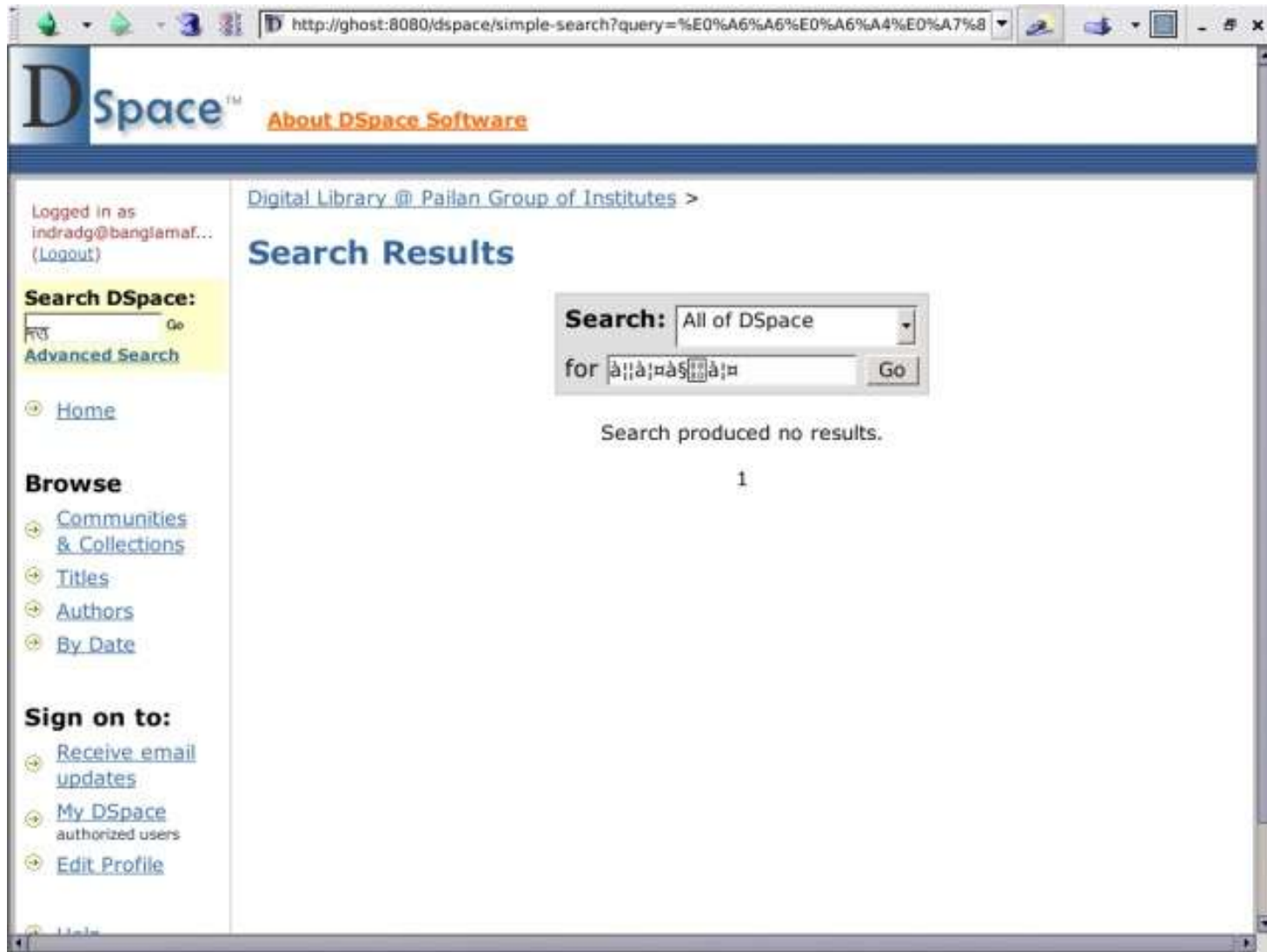
Please use this identifier to cite or link to this item:
<http://hdl.handle.net/123456789/51> [Edit...](#)

Title: উপক্রম
Authors: দত্ত, মাইকেল মধুসূদন
Keywords: মাইকেল মধুসূদন দত্ত
চতুর্দশপদী কবিতা
Issue Date: 22-Jan-2005
URI: <http://hdl.handle.net/123456789/51>
Appears in Collections: [চতুর্দশপদী কবিতাবলী](#)

Files in This Item:

File	Description	Size	Format	
file1.html	১- উপক্রম : চতুর্দশপদী --- মাইকেল মধুসূদন দত্ত	1Kb	HTML	View/Open

DSpace - A sneak peek (Search Interface - Encoding Error)



Indic/Bengali Localization Resources:

1. Indic Coordination/Hindi L10N : <http://www.indlinux.org>
2. Ankur Bangla Project : <http://www.bengalinux.org>
3. Free Bangla Fonts Project : <http://www.nongnu.org/freebangfont/>
4. GRIND Project : <http://grind.randomink.org>
5. Indian Koha Interest Group
http://mail.randomink.org/mailman/listinfo/kig-india_randomink.org

This presentation can be downloaded from:
<http://www.randomink.org/indradg/caliber2005.pdf>

Questions?

Thank You!