4-25-2017

BURA: An Open Access Multilingual Information Retrieval and Representation System for Indian Higher Education and Research Institutions

Bijan Kumar Roy  
Dept. of Lib. & Inf. Sc., The University of Burdwan, West Bengal, India, bijankumarroy@yahoo.co.in

Subal Chandra Biswas  
Dept. of Lib. & Inf. Sc., The University of Burdwan, West Bengal, India, scbiswas_56@yahoo.co.in

Parthasarathi Mukhopadhyay  
Dept. of Lib. & Inf. Sc., University of Kalyani, West Bengal, India, psmukhopadhyay@gmail.com

Follow this and additional works at: http://digitalcommons.unl.edu/libphilprac

Part of the Library and Information Science Commons

http://digitalcommons.unl.edu/libphilprac/1541
BURA: An Open Access Multilingual Information Retrieval and Representation System for Indian Higher Education and Research Institutions

Dr. Bijan Kumar Roy¹, Dr. Subal Chandra Biswas² and Dr. Parthasarathi Mukhopadhyay³

¹Assistant Professor, Dept. of LIS, The University of Burdwan, WB, Email: bijankumarroy@yahoo.co.in
²Professor, Dept. of LIS, The University of Burdwan, WB, Email: scbiswas_56@yahoo.co.in
³Associate Professor, Dept. of LIS, University of Kalyani, WB, Email: psmukhopadhyay@gmail.com

Abstract

The paper describes the growth and development of open access repositories (OARs) in India. The paper proposes a Unicode-compliant information retrieval and representation (IRR) system viz. BURA (Burdwan University Research Archive) for Indian Universities. It has been developed using a number of open standards and open source software (OSS). This Unicode-compliant interface allows administrators to perform various system level operations as well as end users can browse and search resources in Bengali language. Also, describes the necessity of integrating Indic-script based SKOS-enabled subject access system (here DDC – Dewey Decimal Classification) into the proposed model in order to fulfil the subject search of the users. Finally, offers a single window search interface for harvesting metadata from multiple interoperable OARs.

Keywords: Institutional Repository, Open Access Repository, Multilingual, SKOS, digital library, open access, Federated searching.

1. Introduction

Advancement of information and communication technology (ICT) has brought many changes in library environment by transforming the traditional publishing system into electronic publishing system. Several open access (OA) publishing models have emerged and the way of information production, acquisition, dissemination and its preservation have also changed (Roy, Biswas & Mukhopadhyay, 2016a). The main objective of this OA movement is to solve the problem “serials crisis” faced by different libraries by providing Gratis OA (free to view) and Libre OA (free to reuse and rework) to all public funded research outputs. The open access repositories (OARs) as a green road to OA have emerged as a new publishing platform due to the problem of the present system of scholarly publishing. Jones, Andrew & MacColl (2006) have rightly identified the following elements that led to the development of OARs such as - e-thesis archives; faculty practice of e-prints on personal web pages; subject repositories; need from institutions for presentation of research outputs in various form and format; OA aims etc. In SPARC position paper, Crow (2002) defined an institutional repository (IR) as “a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and made accessible to end users both within and outside of the institution, with few if any barriers to access”.

India has more than 750 universities of some kind in each of the 29 states of India as well as three of the union territories (www.ugc.ac.in). It includes State, Central, Private and deemed to be Universities. All these institutes have been generating valuable knowledge resources in the form of theses, dissertations, project reports, courseware, pre-prints, etc through research and development activities. But seamless access to this scholarly information is the biggest problem of Indian Universities. In a study (The Open Citation Project, 2004), it was found that near about 80-85% of digital intellectual output of universities is never made accessible to the public. There is no such
mechanism for holding and sharing such public funded research results globally. Existing information retrieval (IR) systems are not able to access these public funded research outputs. In this context, a Web-enabled distributed IR system is required for perpetual access of these OA knowledge resources. The solution named BURA may become an alternative platform and one of the easiest ways to facilitate OA to scientific literatures available in different form and format. The objective of the paper is to describe key features and functions of the proposed model - BURA.

2. Literature Review

A good number of literatures including journal articles, seminar papers, reports, projects etc are now available on evaluation of India OARs (Choudhury & Choudhury, 2014; Das & Chatterjee, 2015; Roy, 2007; Roy, 2014b; Roy & Mukhopadhyay, 2011; Sengupta, 2012; Thaker & Oza, 2010). Roy, Biswas & Mukhopadhyay (2012a) provided an overview of Indian OARs and compared in respect of contents types, objects uploaded, software used etc. In another paper, they gave an overview of current state of OARs in Asian countries with special reference to SAARC countries (Roy, Biswas & Mukhopadhyay, 2012b). In their current study, they compared OARs against different criteria and showed ranking of Indian IDR in World’s repositories along with several policy issues (Roy, Biswas & Mukhopadhyay, 2013). Several experts have shared their practical experiences in developing and maintaining IR at their own institution using OSS (Doctor, 2007; Doctor & Ramachandran, 2008; Jayakanth et al., 2008; Krishnamurthy, 2005; Sutradhar, 2006; Anuradha, 2005; Narang et al., 2005; Singh & Pandita, 2005; Shewale, 2012). Cherukodan, Santhosh Kumar & Humayoon Kabir (2013) described design and development of a digital library at Cochin University of Science and Technology (CUSAT) using DSpace. Beena & Archana (2011) also described the real time experiences of managing and sharing of intellectual wealth of academia of CUSAT using OSS. Another study (Vijayakumar, Murthy & Khan, 2006) proposed a prototype model for Indian universities to preserve electronic theses and dissertations (ETDs). Krishnamurthy & Kemparaju (2011) reported the use of IR in Indian universities and research institutes. Khaparde & Ambedkar (2014) discussed growth and the development of ETDs in India along with policy framework for building ETD archives. Whereas, Jayakanth, Minj & Dastidar (2012) reported that many academic and research institutes have made it mandatory to set up IRs using OSS.

3. Open Access Repository Movement in India

The first IR was built in the School of Electronics & Computer Science at the University of Southampton, United Kingdom, in 2000 (http://eprints.ecs.soton.ac.uk/) and after then more than 3300 OARs have been developed in different countries (OpenDOAR, 2016; ROAR, 2016). The number of OARs is increasing rapidly not only in some developed countries like United States, United Kingdom but a few developing countries like Japan, Brazil, India are also in a good position. In April 2004, there were only four (4) IRs in India and this number had climbed to eighty (80) in 2016 (December) with an average increase of six (6) new repositories per year (OpenDOAR, 2016).
OpenDOAR recorded near about 3300 (as on December, 2016) OARs all over the World and India being an Asian country ranks 10\textsuperscript{th} position (after Poland) (Fig. 2) and 2\textsuperscript{nd} position in Asia (Fig. 3) only after Japan in terms of total number of repositories registered in OpenDOAR database. But as per ROAR database (ROAR, 2016), India having 105 repositories, out of 3600, possesses 6\textsuperscript{th} position in the World. Again if we consider all OARs in terms of percentage (\%) of OAI-PMH compliant repository, India possesses 4\textsuperscript{th} position in Asia\textsuperscript{21} and 5\textsuperscript{th} position in the World (OpenDOAR, 2016).

Table 1 (column - 2 shows year-wise addition of OARs and column - 3 shows cumulative growth of OARs) gives an overall growth of OARs in the country during last six years starting from December, 2010. It is found that the growth rate of OARs was quite low in all the years except 2010-2011 and 2012-2013. Though ROAR represents something different from OpenDOAR database and total number of OARs varies from OpenDOAR. In preparing this table, we have faced several technical and non-technical problems. Some of the OARs have not been registered in any databases like OpenDOAR or ROAR. Even some OARs have been placed in both the databases in different name. So, actual figure may vary and we have excluded those OARs from our study.

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of OARs</th>
<th>Cumulative Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010 (December)</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>2011 (December)</td>
<td>11</td>
<td>53</td>
</tr>
<tr>
<td>2012 (December)</td>
<td>03</td>
<td>56</td>
</tr>
<tr>
<td>2013 (December)</td>
<td>13</td>
<td>69</td>
</tr>
<tr>
<td>2014 (December)</td>
<td>02</td>
<td>71</td>
</tr>
<tr>
<td>2015 (December)</td>
<td>02</td>
<td>73</td>
</tr>
<tr>
<td>2016 (December)</td>
<td>06</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 1: Cumulative growth of OARs  
(Source: OpenDOAR database)

4. Need of OARs for Indian Higher Education and Research Institutions

Several experts (Bailey et al., 2006a; Crow, 2002; Kircz, 2005; Gibbons, 2004b) described the importance of maintaining IRs in higher education systems. The benefits of having IR in universities are multifaceted and is helpful for improving teaching and learning (Lynch, 2003; Day, 2003; Manuel & Oppenheim, 2007); maximizing its uptake, readership, usage and citation impact (Garfield, 2004; Holmes & Oppenheim, 2001; Oppenheim, 1996; Heijne, 2005); maximizing
availability, accessibility and functionality of the research products (Johnson, 2002; Pickton & Barwick, 2006; Lyte et al, 2009; Kim, 2011). Heery & Anderson (2005) reported that it facilitates records management whereas Prosper (2003) is in support by saying that it also act as a CV. Banks (2006) argued that OARs represent an exciting possibility for both the preservation and retrieval of grey literature of the universities. Bjork (2004) stated that IRs can be devoted to find alternative marketing channels for universities. So the prototype, named as BURA is required to - i) improve impact and citations of research outputs; ii) support OA trends in scholarly communication by showcasing research outputs; iii) greater visibility of research outputs; iv) global presence of local research outputs; v) preserve institutional research outputs safely for perpetual access; vi) act as a marketing tool; vii) measure research and teaching activities; and viii) speeds up and promote collaborative research.

5. National Initiatives towards OARs Movement

Arunachalam (2004) reported that MHRD has set up the ‘Indian National Digital Library in Engineering Sciences and Technology (INDEST) Consortium’ to subscribe electronic resources for the members Consortium. University Grants Commission (2005) drafted a policy framework entitled “UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations, 2005”, to strengthen national capability of producing electronic theses and dissertations and to maintain university-level and national level databases of theses and dissertations. Bangalore declaration (2006) drafted a model National Open Access Policy for Developing Countries. The main objective was to bring together policy makers and research scientists from major developing countries to agree a path forward towards adopting full OA to publicly-funded research publications. National Knowledge Commission (2007) “Working Group on Open Access and Open Educational Resources and Working Group on Libraries” have also strongly recommended for OA and advocated that, on policy level, all research articles published by Indian authors receiving any government or public funding must be made available under OA and should be archived in the standard OA format on his/her website. And ‘The Right to Information Act’, which came into effect in 2005, has had an impact on publicly financed research: since this act was passed, all citizens have had the right to know the results and social benefits of this type of research. In December 2014, India’s Ministry of Science and Technology, the Department of Biotechnology (DBT) and the Department of Science and Technology (DST) released a new OA Policy. This is a big step for the promotion of OA scholarly communications in India as DBT and DST are the nation’s two top most scientific departments (DBT and DST Open Access Policy, 2014). Recently, Indian Institute of Technology, Kharagpur started working on National Digital Library (NDL) project in order to create a national online educational repository for the nation (https://ndl.iitkgp.ac.in/). On the basis of the above studies, the following key events (Table 2) are identified (https://en.wikipedia.org/wiki/Open_access_movement_in_India) in support of OAM in India. This is a list of landmark events in the history of open access (OA).

Table 2. Key events in support of OA to knowledge movement in India

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>Conference on ‘Advances in Information Access and Science Communication’ organised at M S Swaminathan Research Foundation</td>
</tr>
<tr>
<td>2002</td>
<td>First Open Access Repository ePrints@IISc was established by the Indian Institute of Science (IISc)</td>
</tr>
<tr>
<td>2004</td>
<td>Indian National Science Academy signed the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities</td>
</tr>
<tr>
<td>2004</td>
<td>Workshop on ‘Open Access and Institutional Repositories’ under the aegis of the M. S. Swaminathan Research Foundation,</td>
</tr>
<tr>
<td>2005</td>
<td>UGC (Submission of Metadata and Full-text of Doctoral Theses in Electronic Format) Regulations</td>
</tr>
</tbody>
</table>
2006  Special session on OA was held at the 93rd Indian Science Congress
2006  NIT Rourkela mandated Open Access policy
2006  Bharathidasan University adopted OA self archiving Policy
2007  NKC recommends Open Educational Resources and Open Access
2009  ICRISAT mandates open access to all its scientific and scholarly publications
2011  CSIR Open Access Mandate was launched
2011  Shodhganga: A Reservoir of Indian Theses was launched by the INFLIBNET Centre
2011  One-day seminar on open access at the India International Centre, in New Delhi.
2011  Formed a voluntary group called ‘Open Access India’ by agricultural researchers
2012  National Data Sharing and Accessibility Policy (NDSAP) was launched
2013  Indian Council of Agricultural Research (ICAR) adopted its Open Access policy
2014  Department of Biotechnology & Department of Science and Technology adopted OA Policy
2015  UNESCO and CEMCA jointly launched an Open Access Curriculum for Researchers, and an Open Access Curriculum for Library Schools
2015  CSIR-URDIP launched a portal, "Listing of Open Access Databases" or LOADB

6. BURA: The Key Features

The purpose of BURA software framework is to provide free full text access to the public funded OA research outputs in different form (e.g. theses, dissertations, project reports, etc) and format (e.g. doc, pdf, etc) created by academicians, researchers of an organizations. The key features of the model are - i) compatible with open standards and open source software; ii) displays bibliographic records as well as full text; iii) sort by different search syntax such as author, title, publication etc; iv) support different metadata standards and languages; v) supports browsing and searching in Bengali language; vi) supported by subject access system (here DDC – Dewey Decimal Classification); vii) supported by federated searching mechanism; and viii) supported by web 2.0 tools.

6.1. Browsing and Searching in BURA

BURA (developed by DSpace software - http://www.dspace.org/) allows users to find contents in a number of ways. The end user can browse, search and access the collections using the hierarchies and also the alphabetic bar menu (Fig. 4). BURA also supports sophisticated searching including advanced search with the help of search operators (Boolean, positional and relational operators) both within the local repository and across the repositories of multiple institutions (Fig. 5).

![Fig. 5: Advanced search interface of BURA](image-url)
6.2. Controlled Vocabulary System

The need of SKOS-enabled Indic-script based subject access system/controlled vocabulary system is mandatory for any digital information retrieval system in India (Roy, Biswas & Mukhopadhyay, 2016e). An attempt has been made to use the DDC22 available in Bengali language and demonstrates the required mechanisms for system-level integration (Roy, 2014a). BURA allows users searching specific subject category using SKOS-enabled vocabulary control system (here DDC - Dewey Decimal Classification) (Fig. 6). The plus sign (+) indicates that category has sub-categories and/or links to resources under it. It is available both in English as well as in Bengali and user can navigate throughout the list and can select appropriate subject category.

6.3. Novelty in the Submission Interface

The default submission interface of BURA has been modified and DDC (22nd edition up to 3rd summary) has been incorporated in BURA software framework. It has also been modified in Bengali language to fulfil the subject approach of Bengali speaking users (Fig. 6). Bengali language equivalents are based on the available translation work (Saha, 2008).Submitter can select specific subject category from DDC both in English (Fig. 7) as well as in Bengali (Fig. 8). This Indic-script based subject access system supports browsing and searching specific subject category using DDC in English as well as in Bengali. Though, there is the provision of incorporating any national languages in the system.
Fig. 6: Subject searching using DDC

Fig. 7: Selection of Subject Categories from DDC in English
India has 418 languages of which 407 are living and 11 are extinct (Maitra, 2002) and 5% people can read and write English (Technology Development for Indian Languages Group, 2003). In another study, it was reported that only 5% to 10% of population in our country, India is aware of English language and can either read or write English (Vikas, 2005). So, near about 90% people are unable to access English-language based objects due to the absence of multilingual information retrieval and representation system. As stated earlier, India has more than 750 universities (including state, central, private and deemed to be universities) and have been generating valuable knowledge resources in different regional languages in the form of theses, dissertations, project reports, courseware, pre-prints, etc through research and development activities. But it is found that majority of Indian OARs cover only English documents and near about 75% contents are in English (Fig. 9). So, the processing and retrieval of non-English informational objects (25%) have been a challenging task to the existing traditional keyword based textual information retrieval and representation (IRR) systems. Here lies the importance of devising multilingual information retrieval system (Roy, Biswas & Mukhopadhyay, 2016d) and this Indic script-based IRR system (e.g. BURA) is capable of handling non-English knowledge objects and enables users to access such scholarly resources in their own native language (Fig. 10).

This multilingual interface (Fig. 10) allows administrators to perform various system level operations such as creation of community/collection, user registration, password management etc. It also allow users browsing and searching resources in Bengali language using different search options such as by author (here লেখক), by title (here আখ্যা), by subject (here বিষয়) and so on.
Advanced search mechanism can be accessed by simply clicking the corresponding link ‘Advanced Search’ (here বিস্তৃত অনুসন্ধান) (Fig. 11). User can limit his/her search using different search syntax.
Apart from the above, the model (Fig. 12) supports harvesting metadata as well as full text efficiently and effectively from any OAI-PMH compliant repositories globally. Even this model can also be applied in any other domains (Roy, Biwas & Mukhopadhyay, 2016b, 2016c).

7. Conclusion

The open access to knowledge movement unquestionably became much popular throughout the country specially in academic circles and finds a rational and convergent support from higher education and research institutions. Many academic and research organizations have already set up repository and have made it mandatory to deposit research outputs thereon. The increasing awareness among the academicians and researchers is a very good sign for the future of scholarly publications. Our Government and other national agencies are preparing mandate, draft law or by-
law similar to those of other developed countries like USA, UK and are adopting open standards, open technologies and best practices in designing OARs in the line of global recommendations in order to make our OARs searchable Worldwide. This proposed model can be integrated with national network and can be accessed either on their Intranet or the Internet. This Unicode-compliant multilingual IRR system can process any Indic script-based knowledge objects and thus may be helpful to those planning to managing non-English informational objects to their IR. This model has the capabilities to satisfy the subject approach of the users through subject access system like DDC. In addition, it may be an interoperable solution to the digital library developers who are planning to provide federated search services to the users through a unified search interface.

References


Roy, B. K., Biswas, S. C., & Mukhopadhyay, P. (2016b). Designing metadata harvesting framework...


