Need for Language Technology in Developing ETD Packages

Laila T. Abraham

Abstract—Most of the dissertations from Indian universities are prepared in Indian languages. Hence without digital library packages that can process Indian scripts, full text retrieval or piracy checking will not be possible. Existing popular international packages can process only scripts of international languages. Packages that can process Indian local scripts can be developed only indigenously. The only successfully tested package in this respect is Nithya D Arch, the software used for digitizing PhD theses of Mahatma Gandhi University, Kerala. The language technology adopted in this package is one of the key factors that can contribute to Open Access and Piracy Checking in Indian research environment. It is a unique venture among the initiatives for Open Access and quality control of research in Indian scenario. This paper evaluates the status of ETD archives in India in the above context with special reference to Mahatma Gandhi University's open access digital archive of PhD Theses.

Keywords: ETD Packages, Language Technology, Digital Archives, Nithya D Arch, Ph.D. Theses, Unicode, Full Text Search, Multilingual Search

INTRODUCTION

Universities are referred to as the seat of higher learning and excellence responsible for the generation and dissemination of knowledge by means of public funded research and its output. Research was earlier considered to be the achievement of a research scholar for his/her own private purposes. But with the arrival of industrial revolution, research became a means to find solutions to the meet the basic needs of human beings such as to increase food productivity, to discover medicines for incurable diseases etc. For this most of the developing countries have changed their strategy by generating technology to produce consumables instead of using imported products. This has helped them to have a footing among other nations in the world (Laila T. Abraham, 2015).

When nations began to depend more on research findings, society began to accept research document as an output of public fund project. It can be seen that before the arrival of Information and Communication Technology (ICT), the research outputs of institutions were underutilized because of the issues relating to its accessibility. The developments in ICT have now broken these restrictions and research output has become globally available for its effective application in the related areas.

This paper deals with the advantages in ETD packages which have incorporated language technology for Indian scripts, with a special reference to Mahatma Gandhi University, Kerala. 1

ETD ACTIVITIES AT NATIONAL LEVEL

Many institutions in India have attempted digitization activities of doctoral dissertations in India since late nineteenth century. The first OA Repository of Doctoral Theses was

^{*}Corresponding Author: Laila T. Abraham (Email: lailababu@gmail.com) University Librarian (I/ c), Mahatma Gandhi University

envisaged in India in 1996 by Dr. A.M Michael, Vice Chancellor of Kerala Agricultural University (KAU). This was a sub-set of the project named Kerala Agricultural University Library and Information System (KAULIS). With support from ARIS programme under ICAR, KAU digitized 400 of its 3000 Theses in 1997-98(Gireesh Kumar, T.K. and Jayapradeep M. 2011). At that time, even the term digital library was not heard in India. Special packages for digital libraries were not available in India or abroad at that time except Unix based Tech Lib Plus, which can be customized for the purpose, but was beyond the financial powers and available expertise of Indian university libraries. KAU used MS Access to prepare a simple catalogue of the PhD Theses, digitized the theses and prepared e-books of the theses and linked the pdf file-the e-thesis to the concerned record in MS Access. Full text search was also possible in this archive, which used some possibilities of Adobe package. A version of the digital library was prepared in Basis Plus TechLib Plus also by exporting the data and short period after it was accessible in the Intranet and Internet using the space freely allowed to the university by VSNL. The next ETD repository in India was developed in 1999 by the Indian Institute of Technology in Mumbai. University Grants Commission, Indian Council of Agricultural Research and Council of Scientific and Industrial Research are the agencies who have taken steps to encourage ETD activities in India. The next efforts in the ETD came out in the form of a national repository, namely, 'Vidyanidhi at http://www.vidyanidhi.org.in with the participation of many universities, academic institutions and other stake holders. It was followed by the MG University's official PhD theses repository http://mgutheses.org in 2008.

The next and the most significant venture in this direction is the national repository; 'Shodhganga' launched by INFLIBNET. It is ahead with its noble impetus by collecting and providing access to full text of PhD Theses awarded by all the universities who have signed MoU for online submission in to the Union database 'Shodhganga'.

INFLIBNET AND ETD INITIATIVES

UGC Regulations 2005 (Submission of Metadata and Full text of Doctoral Theses in Electronic Format) has been released to build up online PhD Theses repositories at university and nation level in order to strengthen the ETD potential of the country. For the purpose, universities were requested to develop university-level databases of theses and dissertations. This resulted in the development of open archives of digital theses among Indian universities.

Now, there are a total number of 693 universities in India with 45 central, 325 state, 128 deemed and 195 private universities. The total output of their research findings comprise the knowledge generated in the nation. To bring a bibliographical control and pool of these resources, INFLIBNET has developed the national knowledge repository of electronic theses namely 'Shodhganga'.

In 2009, UGC had released regulation for the standardization for the submission of online theses into 'Shodhganga' and envisaged the mandate for every university to sign MoU with INFLIBNET for their submission of online doctoral dissertations along with its metadata in to the repository. MG University was the first to contribute theses to this project. It transferred about 800 theses from its repository http://mgutheses.org with

which the Shodhganga started to function. Till date, 217 universities have signed MoU with INFLIBNET and the total number of theses uploaded is 40,186. 'Shodhgangotri'is another initiative of INFLIBNET to deposit electronic form of approved synopsis submitted by the research scholars for their PhD registration in universities. At present, 219 universities have submitted the synopsis and the total number of upload is more than 1900.

UGC as part of promoting ETD activities had released a grant of Rs. 1500,000 to selected university for setting up an ETD Lab and digitizing PhD theses awarded by the University. For taking measures towards checking plagiarism, INFLIBNET has provided anti-plagiarism software(s) namely, iThenticate and Turnitin to 100 universities that had signed MoU with the Centre for e-submission in 'Shodhganga' on trial basis for one year from 01 February 2014 to 31 May 2015 (Laila T. Abraham, 2015).

ETD INITIATIVES IN KERALA

ETD initiatives have been taken up by a few universities in Kerala. Total number of universities in Kerala is 15 where only 3 universities such as Mahatma Gandhi University (MGU),, Cochin University of Science and Technology (CUSAT) and University of Calicut have developed their own ETD projects. Mahatma Gandhi University is the first university in India to digitize its PhD theses and archive it and make available in the public domain.

Sl.	Name of University	Present Status of ETD Projects	Software
No.			Used
1	University of Kerala	Signed MoU with INFLIBNET Centre. Under the process	-
		of inviting tender for digitization of PhD theses.	
		Number of PhD Theses awarded: 4050+	
2	Mahatma Gandhi	MGU Digital Archive of PhD Theses library is available	Indigenous
	University	at http://www.mgutheses.in	s/w -Nithya
	-	Has completed the digitization of all the PhD Theses	D'Arch
		awarded. Running upto date.	
		Total contribution to Shodhganga project-2052	
3	Cochin University of	CUSAT Digital Archive of PhD Theses is available at	Open Source
	Science & Technology	http://dyuthi.cusat.ac.in	s/w -DSpace
		This is a database comprising of PHD Theses,	
		Conference Proceedings etc.	
		Total contribution to Shodhganga project-1449. No.of	
		PhD Theses awarded- 2060+	
4	University of Calicut	List of PhD Theses can be viewed under Directorate of	Open Source
		Research. Access to public is denied. Total contribution	s/w -DSpace
		to Shodhganga project- 642 Number of PhD Theses	
		available: 2560+	
5	Kannur University	Signed MoU with INFLIBNET Centre. ETD not	
		operational.	
6	Sree Sankaracharaya	Signed MoU with INFLIBNET Centre. ETD not	
	University of Sanskrit	operational.	

PRESENT STATUS OF THE ETD PROJECTS OF VARIOUS UNIVERSITIES IN KERALA

MAHATMA GANDHI UNIVERSITY AND ETD INITIATIVES

The digitization process of PhD theses in the University was done in two phasesdigitization of hard copies in the first phase and soft copies in the second phase. The work in the first phase was initiated during March 2008. The retrospective digital conversion was entrusted with M/s. Beehive Digital Concepts Pvt. Ltd., Kochi, Kerala. The second phase of digitization is being done in the Union Catalogue Section of the library. The digital repository of the university is available on the site www.mgutheses.org. The site is maintained in a remote server situated in United States. The first phase was completed within 6 months and the total number of theses uploaded was 1122. The need for using a digital library package which can process Indian scripts to enable full text search and incorporating repositories with Indian language dissertations in piracy checking systems was first identified in India by a team of experts lead by Dr. R. Raman Nair, the former University Librarian of MGU. His initiative to use such solutions in mgutheses.org was the first bold attempt in India to test a solution relevant to our environment; while numerous packages produced by international organizations and multinational companies are readily available and are commonly used in India for attempting ETD repositories.

LANGUAGE TECHNOLOGY IN ETD PACKAGES DEVELOPED USING OPEN SOURCE SOFTWARES

In the present scenario studies and research are being conducted in India in more than 20 regional languages also like Hindi, Malayalam, Kannada, Telugu, Marathi etc. Similarly, the output of the findings are brought out in diverse forms of resources like texts, recordings, dictionaries, annotations, software(s), protocols, data models, file formats, newsgroups and web indexes (Richard Jones, 2004). The categories of people who depend on these types of resources are linguists, teachers, researchers and speakers and politicians, software developers, publishers and the promoters and sponsors of language technology who need to access and organize these language resources and professional associations, government funding agencies and non-governmental organizations.

Different means are there for these groups of people to access required information. First method is to store the large resources in electronic format. For that, Extensible Markup Language (XML) and Unicode provide facilities to embody structured data with longevity. Second method is to publish the language resources in internet, which is the most efficient and practical way. Thirdly, providing a standard classification model known as 'Dublin Core Metadata Set' along with information transmission method offered by the Open Archives Initiative will help to develop a Union Catalogue of large repositories and digital archives. This technology is compatible to provide a reliable user-base for the fast growing language resources and its retrieval (Bird & Simons, 2001).

This combination of Dublin Core Metadata-OAI can be considered as a bridge connecting language technology developers of ETD packages and the information seekers of language resources. So, in the development of digital archives of PhD theses, the introduction of language technology has a significant role. More than 40% of the researches in India are conducted in regional languages. But most of the internationally acclaimed packages(s) like Dspace, Greenstone, EPrint etc. are not compatible to Unicode script and hence do not permit full text search facility in regional languages (Hussain, Raman Nair, & Raveendran, 2002). Due to this reason, checking of plagiarism is not possible for theses published in regional languages in India.

LANGUAGE TECHNOLOGY INCORPORATED IN NITHYA D ARCH

The software used for the digital archives of Mahatma Gandhi University is Nithya D Arch. This is an indigenous package developed by Centre for Informatics Research and Development (CIRD) with cooperation of many experts and firms like Swathatra Malayalam Computing, Centre for Informatics Research and Development, Shri K.H. Hussain etc. This is a unique package with regard to its multilingual full text search facility and compatibility to Unicode Standards (Raman Nair & Hussain, 2010). Its user friendly interface and OAI-PMH (Open Access Initiative-Protocol for Metadata Harvest) compliant features permits the search and retrieval facility and metadata mining by OAI Projects of the package laudable (Peter Suber, and Raman Nair, R. and Hussain, K.H. & Hussain, 2009).

Some of the other significant features of the package are:

- There is no restriction in the number and language of theses that can be hosted
- PhD Theses in all languages like English, Malayalam, Hindi, Sanskrit, Tamil and Kannada can be hosted
- Multilingual data input is possible
- Permits multilingual and multi-keyword search using Boolean Search mechanism
- PhD theses are not password protected and so no download is required to read and refer
- Compatible to all browsers/ operating systems
- No performance loss is there while handling simultaneous queries at any time
- Hosted on a live and dedicated Linux server
- The project is built on PHP programming language with MySQL back end
- Specific pages/ sections together with the entire theses can be retrieved
- Due to its open database structure, future migration to any different OAI-PMH compliant packages is possible.

Nithya D Arch used by Mahatma Gandhi University for developing its ETD package is the only indigenous software compatible with Unicode. It has been tested for full text retrieval in most of the Indian languages (Hussain, Vijayakumaran Nair, Chitrajakumar, Ravindran Asari and Raman Nair, 2005). Nithya D Arch permits multilingual and full text search in the theses archives using Indian scripts also.



262 ♦ ETD 2015: 18th International Symposium on Electronic Theses and Dissertations

Fig. 1: Home Page of MGU Theses Archive at http://mguthese.org

FULL TEXT NAVIGATION: MULTILINGUAL SEARCH IN NITHYA D ARCH

During search, after the full text is opened, user can navigate through the entire theses back and forth. Bookmarks displayed in the right side of the window helps to move directly to the required part and chapters. As Nithya D Arch is Unicode compliant, it permits the archiving, search and retrieval of dissertations in any Indian script (Jasimudeen, Maghesh Rajan and Suresh Kumar, 2012). This is the only database management system for full text records in India that is possessing multilingual search provisions. Presently, the archive contains theses in English, Malayalam and Hindi scripts. A visual keyboard is provided in the search module to input queries in Indian scripts. ETD Package of MGU can become a successful model for archiving regional language dissertations in India.

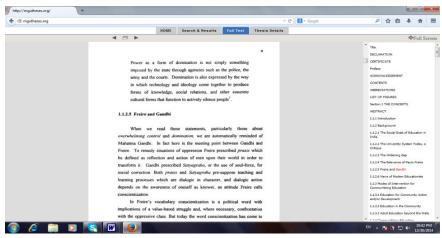


Fig. 2: Full Text Search using MGU Theses Archive

Need for Language Technology in Developing ETD Packages ◆ 263

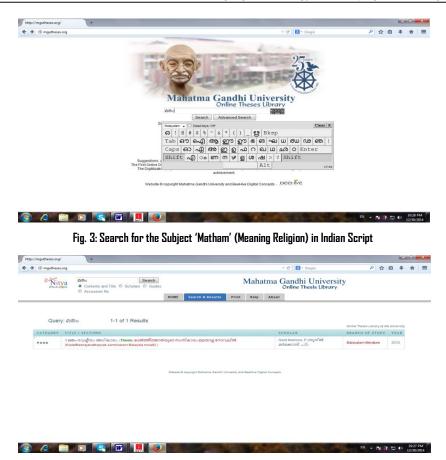


Fig. 4: The List of Items Retrieved from Search in Indian Script

By clicking the particular theses from the list of items retrieved, we reach page 87 where religion is discussed as in Fig. 5.

http://mgutheses.org/ +		
 Imputheses.org 	≂ C 🔣 - Goog	P ☆ 白 非 合 目
	HOME Search & Results Full Text Thesis Details	
	4 B >	Full Screen
		° 083-860
		സാക്ട്രപത്രം
		\$00001000-capto
		mml
		സ്ഥര്ചണം
		DESC40+
	സാണ്കാരത്തിന്റെ തിരധാതകഘടകങ്ങളായിവരുന്ന (Constituent	ആമുയം
	elements) മതവും രാഷ്ട്രീയവും കുടിപ്പേര്ന്നുണ്ടാകുന്ന സമസ്യകളെ	1 4555000 800000
	പ്രഥമാകരങ്ങളുടെ അടിസ്ഥാനത്തിൽ ഇവിടെ പറനം നടത്തുന്നു. ഈ	🗍 സാംസ്കാരിക ജീവിരം
	ഘട്ടത്തിൽ കടൽത്തിരജനതയുടെ ജീവിതത്തിൽ മത,മാഷ്ട്രിയങ്ങളുടെ	2 സാംസ്കാരിക സൂചകങ്ങള്
	ഇണപേരവും വേർപിരിയവും അധികാരസ്ഥാപനങ്ങളുടെ ഇടപെടലും	ആചാരങ്ങള് വിദ്ധാനങ്ങള്
	സാംസ്കാരിക സൃഷ്ടിക്ക് എങ്ങതെയെല്ലാം കാരണമാകുന്നുമെന്ന	3 ശേം രാഷ്ട്രീയം അധികാരം
	അന്തേഷണം പ്രസംതമാകുന്നു. വർഗീമകലാപങ്ങൾ, രാഷ്ട്രീമ സമരങ്ങൾ,	4 വിജ്ഞാനവിനിമയവും സമ്പങ്ങും
	അടിച്ചമർത്തലുകൾ തുടങ്ങിയവയൊക്കെ പ്രത്യേക പരിഗണനായ്ക്ക്	5 സാംസ്കാരിക ജീവിരവും
	വിദ്ധത്താക്കും ചെയ്യുന്നു.	«ស្រែង-CB
	3.5x	6 ഉപതംഹാരം
	മതം അതിന്റെ വിശാലമായ അർത്വത്തിൽ സംസ്കാരത്തകാൾ	അനുഭവസം
	any procession incompany and a second procession in the procession of the second procession of t	ശ്രന്ഥസൂചി
	ഓഗവുമാണ്. ജീവിതത്തിന്റെ യഥാർത്ഥ അർത്ഥവും ലക്ഷ്യവും മനുഷ്യത്	
	കാണിച്ചുകൊടുക്കുന്ന ഒൻ എന്ന അർത്തത്തിൽ മതം സംസ്കാരത്തിന്റെ	
	സങ്കയാണ്, എന്നാരി ആന്തരികമായ അറുഭവങ്ങിന്റെ ഭാതികമായ രൂപം	
	എന്ന നിലയിൽ അത് സംസ്കാരത്തിന്റെ വെറുമോലു ഘടകവുമാണ്. പത്മമായ	
	സതൃത്തിന്റെ ആത്തരികൊയ യാധാർത്തും എന്ന തിവയയിൽ മതം ഒടിക്കരും	
	സംസ്കാരങ്ങില് എതില്ലേ. എന്നാൽ മതം വെറുമൊരു ആപാരോയി	-

Fig. 5: A Specific Page in Indian Script where the Keyword Searched is Retrieved

RELEVANCE OF THE STUDY

It is seen that most of the universities in India have digitized its PhD Theses collection and has signed MoU with INFLIBNET for the online submission of the same into Shodhganga. Most of these universities' digital archives are in "space which is an open source repository software package typically used for creating open access repositories for scholarly and/ or published digital content. It serves as a digital archives system, focussed on the long-term storage, access and preservation of digital content.". In spite of its advantages, it does not permit multilingual search facility as it is not compatible to regional scripts existing worldwide. Due to this, many of the findings of research in India and abroad remain unrevealed to the world. Nithya D Arch, software used by Mahatma Gandhi University is the only indigenous software permitting multilingual search facility and Unicode compatibility. The study on the need of language technology in developing ETD packages have become very important due to the present quality loss in PhD research revealed by numerous pirated theses.

OBJECTIVES OF THE STUDY

- To find the frequency of use of the digital archives of PhD Theses at www.mgutheses.org
- To find the level of usefulness of the digital archives of PhD Theses
- To find out the most useful search facility of the Digital Archive
- To identify the usefulness of the unique features of Nithya D Arch
- To find the advantages of incorporating Multilingual facility in Nithya D Arch
- To find out Nithya's power to process Indian languages

METHODOLOGY

Survey method is used for collecting data. Questionnaires were distributed among 35 research scholars belonging to different disciplines in the university departments. Out of this, 30 responses were received.

SCOPE AND LIMITATION

The study was conducted among randomly selected 30 research scholars due to the limitation of time. The findings of the study have been generalized among the other research scholars also to reach at conclusions.

DATA ANALYSIS

FREQUENCY OF USE OF DIGITAL ARCHIVES OF THE PHD THESES (WWW.MGUTHESES.ORG)

From the study it is found that all the researchers are in the habit of using the theses archives like 46.6% (14 Nos.) daily, 33.3% (10 Nos.) weekly and 20% (6 Nos.) occasionally.

Sl. No.	Use of Digital Archives of theses	Response No. & Percentage
	Daily	14 (46.6%)
	Weekly	10 (33.3%)
	Occasionally	6 (20%)
	Never	0

Table 2

LEVEL OF USEFULNESS OF THE DIGITAL ARCHIVES OF PHD THESES

Table 3

Sl. No.	Level of Usefulness	Response No. & Percentage		
1	Highly Useful	21 (70%)		
2	Useful	9 (30%)		
3	Somewhat useful	0		
4	Not Useful	0		
5	Very Less Useful	0		

Regarding the level of usefulness of the archives, it is highly useful to 70% (21 Nos.) and Useful to 30% (9 Nos.). They have no different opinion.

TO FIND OUT THE MOST USEFUL SEARCH FACILITY OF THE DIGITAL ARCHIVE

Table 4

Sl. No.	Most Useful Feature of Digital Archive	Response No. & Percentage		
1	Title Search	8 (26.6%)		
2	Scholar Search	7 (23.3%)		
3	Guide Search	2 (6.6%)		
4	Keyword Search	13 (43.3%)		

Regarding the most useful search facility of Digital Archive, 43.3% (13 Nos.) opted for Keyword Search, 26.6% (8 Nos.) for Title Search, 23.3% (7 Nos.) for Author Search and 6.6% (2 Nos.) for Guide Search.

To Identify the Usefulness of the Unique Features of Nithya D Arch

Table 5

Sl. No.	Features	Highly Useful	Useful	Somewhat Useful	Not Useful	Less Useful	Response No. & Percentage
1	Full text search facility	13	-	-	-	-	13(43.3%)
2	Multilingual search facility	10	-	-	-	-	10(33.3%)
3	Bookmarking	7	-	-	-	-	7(23.3%)

The study examined the usefulness of the following Unique features of Nithya D Arch. It is found that all facilities were responded as highly useful. Among these, Full Text search was opted by 43.3% (13 Nos.), Multilingual search by 33.3% (10 Nos.) and Bookmarking by 23.3% (7 Nos.). This shows most of the research scholars are fully satisfied with the unique facilities incorporated in the above site.

Sl. No.	Facilities	Response No. & Percentage
1	Enable search using regional language script	12 (40%)
2	Enable access theses available in all languages	14 (46.6%)
3	Enable to view theses in the original language script	12 (40%)
4	Enable to check plagiarism	3 (10%)
4	All the above	15(50%)

Table 6

TO FIND THE ADVANTAGES OF INCORPORATING MULTILINGUAL FACILITY IN NITHYA D ARCH

The number of responses and percentage exceeds the total number of respondents i.e. 30, as most of them have given more than one preference. The response to the above advantages of the multilingual facility in Nithya D Arch is opted in the following sequence: 40% (12 Nos.) favour search facility using regional language, 46% (14 Nos.) favour its ability to access theses available in all languages, 40% (12 Nos.) favour its ability to view theses in the original language script, 10% (3 Nos.) favour ability to check plagiarism and 50% (15 Nos.) favours all the above advantages. The need of language technology in developing ETD packages is clear from this study.

TO FIND OUT NITHYA'S POWER TO PROCESS INDIAN LANGUAGES

Sl. No Highly Useful Somewhat Not Response Rating the Less Power to Useful (U) Useful (SU) Useful Useful No. & Process (HU) (NU) (LU) Percentage Indian Languages 1 Malayalam HU-12(40%) 12 4 --U-4(13.3%) Script 2 Hindi Script 11 3 HU--11(36.6%) U-3(10%)

This study shows that the respondents are taking advantage of the Unicode compatibility in Nithya D Arch. 40% (12 Nos.) find Malayalam script usage highly useful and 13.3% (4 Nos.) useful whereas Hindi script is highly useful to 36.6% (11 Nos.) and useful to 10% (3 Nos.).

FINDINGS

The study reveals the following:

- Due to the development of ETDs, the scholarly findings of Universities have been exposed globally for use by a large spectrum of clients.
- From the study it is found that all the researchers of Mahatma Gandhi University are in the habit of using University's theses archives. The frequency varies from daily, weekly, occasionally etc. depending upon the intensive nature of their research.

Table 7

- Majority of the researchers found the MGU theses archive to be highly useful. They have no different opinion.
- Among the different search facilities in theses archive, most of the research scholars preferred Keyword Search facility. Title and Author search are almost equal in preference. Guide search is less preferred.
- The unique features of Nithya D Arch embody the structured user base for the ETD package. All the facilities like Full Text search, multilingual search facility, bookmarking were responded as highly useful by the researchers.
- Nithya D Arch is the only indigenous software which is compatible to Unicode technology. Due to this, MGU Theses archive has many advantages like multilingual search facility, ability to access theses available in all Indian languages, to view theses in the original Indian script and to check plagiarism. The researchers have favoured all the advantages of Nithya D Arch since they have utilized all the features effectively. From this study the need for incorporating language technology is evident.
- The study shows that as more theses have been produced in Malayalam regional language usage of Malayalam script for search and retrieval is more when compared to that of Hindi script use. The study shows the high rating of Nithya D Arch in the power to process Indian scripts in the MGU theses archive.

RECOGNITIONS FOR NITHYA D ARCH

In 2010, MGU Theses website has been awarded with the E-Learning award of Government of Kerala. INTUITE, a UK based consortium of universities selected the site as one of the Very Best Web Resources for Education and Research and has included it in its system. UNESCO Libraries Portal, an international gateway to information for librarians and library users has included the website of MG University Library along with the Academic and Research Libraries of Asia-Pacific. H.E. Dr. APJ Abdul Kalam while launching the archive had stated the 'the project is an important step towards democratization of knowledge and that he was sure the model of digital archive created by MG University will be emulated by other Indian universities leading to transparency in the education system. The efforts of MGU in presenting the archives to the public domain through Open Access was appreciated by Dr. Richard M. Stallman, the father of Free Software Movement and stated that he was 'always for the freedom to redistribute such research documents and thereby widen the reach of public funded research activities.' He commented that 'it is an important Open Access Initiative in India.'

CONCLUSION

In India more than 40% of research is conducted in regional languages and as many of the popular ETD packages do not support multilingual search facility, the findings of these researches fail to reach the end user. Nithya D Arch package used by MGU is an exception in this regard. It is transparent and liberal in the inclusion of many unique features like Unicode compatibility and multilingual search facility, full text navigation etc. along with user friendly interface and OAI-PMH (Open Access Initiative-Protocol for Metadata Harvest) compliant features which help the package to make effective search and retrieval facility and metadata mining. The need for Digital Library packages that can process Indian languages to enable information retrieval and piracy checking as stressed by a few earlier comparative studies is evident in this study also.

REFERENCES

- [1] Bird, S., & Simons, G. (2001). OLAC Overview. Retrieved July 29, 2015, from http://www.language-archives.org/documents/overview.html
- [2] Gireesh Kumar,T.K. & Jayaprdeep M. (2011). Electronic theses and dissertations(ETD) initiatives to provide open access to public funded research in India.
- [3] Hussain, K. H., Raman Nair, R., & Raveendran, A. K. (2002). Importance of search and retrieval in CD-ROM full text publishing: Experiments using PDF documents and "Nitya" archival system. *Information Studies*. Retrieved from http://eprints.rclis.org/handle/10760/6219
- [4] Hussain, K. H., Vijayakumaran Nair, P., Chitrajakumar, R., Ravindran Asari, K., & Raman Nair, R. (2005). Creation of digital archives in Indian languages using CDS/ISIS: development of M-ISIS (Malayalam ISIS) and "Nitya" [Journal article (Print/Paginated)]. Retrieved July 29, 2015, from http://eprints.rclis.org/7848/
- [5] Jasimudeen, S., Maghesh Rajan, M., & Suresh Kumar, T. V. (2012). Multilingual Searching in Digital Archives: A Case Study of Mahatma Gandhi University On-line Theses Archive. School of Computer Sciences, Mahatma Gandhi University. Retrieved from http://eprints.rclis.org/19679/
- [6] Laila T Abraham. (2015). Open Access to Research findings: An overview of ETD initiatives in Mahatma Gandhi University, Kerala.
- [7] Peter Suber, and Raman Nair, R. and Hussain, K. H, P., & Hussain, K. H. (2009). Open Access to public funded research: a discussion in the context of Mahatma Gandhi University digital archives of doctoral dissertations. In A. K. Rai, P. Chand, S. R., A. A., & J. Arora (Eds.), (pp. 53–67). INFLIBNET Centre. Retrieved from http://eprints.rclis.org/13531/
- [8] Raman Nair, R., & Hussain, K. H. (2010). Nitya Archive: Software for Full Text Digital Libraries in Indian Languages. In R. K. Pachouri (Ed.), (pp. 515–523). Tata Energy Research Institute. Retrieved from http://eprints.rclis.org/15484/
- [9] Richard Jones. (2004). DSpace Vs. ETD-db: Choosing Software to Manage Electronic Theses and Dissertations | Ariadne. Retrieved July 29, 2015, from http://www.ariadne.ac.uk/issue38/jones
- [10] http://www.vidyanidhi.org.in
- [11] http://dyuthi.cusat.ac.in