

Library Discovery System in Bengali Script: An Experiment with VuFind

*Debabrata Barman**, *Parthasarathi Mukhopadhyay*

Department of Library and Information Science, University of Kalyani, West Bengal, India

Abstract

Owing to the improvement of web scale resource discovery systems on one hand and the Unicode-compliant Indic-scripts based retrieval systems on the other hand provided libraries unique opportunity in integrating these two parallel development processes. The services are called Web-scale Resource Discovery and generally considered as next generation library interface. VuFind is a highly customizable open source Web-scale Resource Discovery (WSRD) product that allows information repositories and integrates multiple services into a unified interface. It has powerful Apache Solr as search platform in order to showcase any digital content and also supports 25 languages in global implementation. Language issues have a vital role to play in library discovery system. This paper demonstrates the processes and procedures related to developing Bengali script based interfaces of VuFind. The product has been developed on the basis of rules and regulations of the VuFind developer community and now included in VuFind from release of the version 3.1 dated 26.09.2016.

Keywords: VuFind, Bengali script based OPAC, web-scale resource discovery, Unicode-compliant Indic-scripts

***Author for Correspondence** E-mail: b.deba87@gmail.com

INTRODUCTION

Today, academic libraries are using different software for different services. These libraries also maintain different databases for these purposes. As a result, users need to move from one software interface to another software for retrieving catalogue records, full-text resources, learning objects etc. This proposes a single-window search interface by using federated search mechanism for different retrieval systems presently in use for university library systems in West Bengal. Federated searching is information retrieval technique that flying from different resources, individual indexes and different database, that reason user trouble their valuable time for searching resources. In this situation, libraries need to move Google like single window sophisticated search system and its final form will be able to help users to retrieve required resources.

The Next Generation Library Catalogue (NGLC) is referred to Web Scale Resource Discovery (WSRD) system. The WSRD service is an evolutionary step forward in providing library users with a 'one-stop shop' where users can find sources for their research

[1]. Discovery service providing a single search box that user can search a library's online and physical content including articles, books, journals, newspaper articles, e-books, specialist collections and more.

In the sense there are two kinds of discovery tools available, one is open source and another is commercial. VuFind is a single window library resource discovery portal that users can use to search and browse all library resources, which is generally available in different retrieval platform. It is designed and developed by Villanova University and it is highly customizable open source product that allows information repositories and integrates multiple services into a unified interface. It has powerful Apache Solr as search platform, faceted results as users search query, author bibliographies for more about author details, support Zotero tools for reference management system in order to showcase any digital content and also supports 25 languages in global implementation [2].

Now, language issue has a vital role to play to retrieve resources from different databases. Library databases are no exceptions for it.

While Bengali language is very important aspect in Bengali-speaking regions; in this paper we demonstrate the Bengali language as expanded transliteration processes and procedures related to developing Bengali script based interfaces of VuFind and it will support Unicode-compliant search environment to retrieve Indic-script based resources.

INDIC SCRIPT

India is a land of “Unity in diversity” [3]. That shifts focus from unity based on a mere tolerance of physical, cultural, linguistic, social, religious, political, ideological and/or psychological differences [4]. India has several language families, the major ones being the Indo-Aryan languages spoken by 75% of Indians and the Dravidian languages spoken by

20% of Indians. Other languages spoken in India belong to the Austroasiatic, Sino-Tibetan, Tai-Kadai, and a few other minor language families and isolates. According to Census of India of 2001, India has 122 major languages and 1599 other languages. Only 18 languages are constitutionally recognized as written in a variety of scripts. Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konakani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, Tamil, Telugu and Urdu are 18 constitutionally recognized languages being represented by 10 different scripts, i.e. Arabic, Devnagari, Bengali, Gurumukhi, Gujarati, Oriya, Tamil, Telugu, Kannada and Malayalam [5].

The history of origin and developments of Bengali-script are often described in Figure 1.

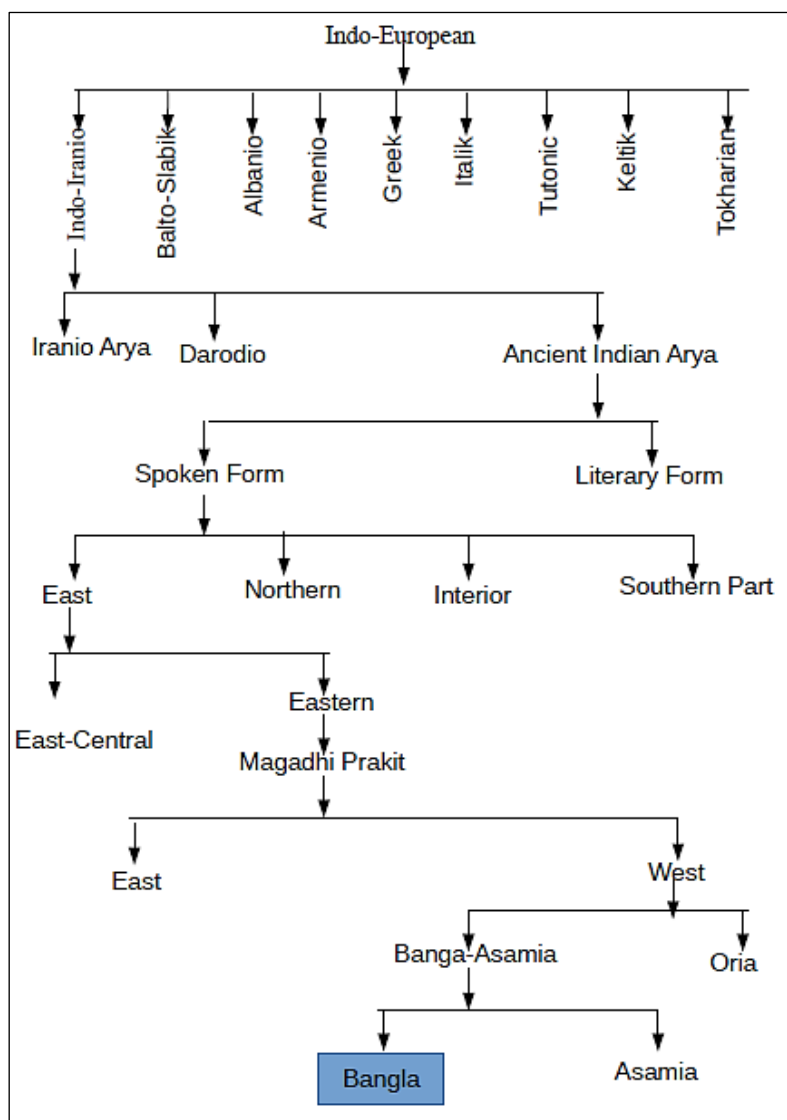


Fig. 1: Bengali Origin and Developments [5].

Today, Bengali language stands at seventh position in the world [6]. The Bengali speaking regions in India are East India, mainly West Bengal and Andaman Island. Also, Bengali is the primary and National language in Bangladesh and there are more than 210 million speakers in the world.

INDIAN SCRIPT CODE FOR INFORMATION INTERCHANGE: ISCII

The first Indian computer based character encoding standards for Indian languages and keyboard overlays had been formed in a specific frame by a standardized committee. The committee was formed under the department of official languages and C-DAC (Centre for Development of Advanced Computing) and then it merged under department of electronics. This project had been composed since 1970 [7].

During 1986–1988, the department of electronics declared that Indian Standard Code for Information Interchange (ISCII), and also called ISCII-83, based on ISO 8 bit proposals. Be that as it may, the issue is that all these ISCII based arrangements utilize 1 byte portrayal of characters. As just a greatest of 256 esteem can be put away. It implies that these innovations can work for making bi-content databases, one Indian dialect in the mix with English and cannot bolster the genuine multi-script condition [8].

UNICODE

The Unicode is a widespread character-encoding standard utilized for the portrayal of content for PC handling. It gives 2 byte character set that contains 16 bit for every character. Subsequently, an aggregate of 216 means 65536 characters can be spoken to by utilizing such a code. This single character set contains every one of the characters. The Unicode character set is rapidly getting to be plainly prominent among open source applications and it will begin another time of multilingual registering. The Unicode Consortium was joined in January 1991 to advance the Unicode standard as a global encoding framework for data trade [9].

India is multilingualism nation. Today Internet aged patrons look through a quarry in various dialects. Library Management System (LMS) is empowered to progressively change at the

purpose of patrons' desire. For instance, a cataloguer could index in English, and after that change, dialects to enter the record in Hindi, Bengali, Tamil and Marathi and so on. The completely useful multilingual framework trade of information and programs written in any dialect is conceivable with no uncommon trade program.

Bengali Script in Unicode

Bengali language script was added to the Unicode Standard in October 1991. Bengali script is a Unicode standard like language, which is derived from ancient Brahmi script. It is used for writing languages such as Bengali, Assamese, Bishnupriya Manipuri, Khasi, Mizo, Munda, Naga, and Santali. The Bengali script code pointed U+0981 to U+09CD. The Bengali characters A1-ED from the 1988 ISCII standard, as well as several Assamese ISCII characters in the U+09F0 column. The Unicode block for Bengali is U+0980–U+09FF (Figure 2) [10].

BENGALI SCRIPT ENCODING TOOLS

A Unicode font is a universally computer encoding system that contains a wide range of characters, letters, digits, glyphs, symbols, ideograms, logograms etc., which are allotted to the special number of estimation of each character in any stage [11]. In this examination, we recognize following programming devices: virtual console, monograph and web monster 'Google Indic Translator' to figure out and the procedure of Bengali dialect textual style to Unicode agreeable multilingual environment.

Virtual Keyboard 'Avro'

Avro is customizable, award winning, open-source, graphical keyboard layout changer for multiple platform based product. It easily creates Bengali font text content with artificial intelligence (AI) sensor technology (Figure 3) [12]. This product is designed and developed by Omicron Lab and its basic features are as below:

- Unicode and ANSI supported mechanism;
- Mouse-based Bengali typing window for windows user;
- Automatic Spelling Checker facility;
- Customizable mode-switching;
- Customizing keyboard layout;
- User editable dictionary support for phonetic typing (Figure 4);
- Keyboard macro support; and
- Unicode to Bijoy converter.

Bengali ^{[1][2]}																
Official Unicode Consortium code chart (PDF)																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+098x	ঐ	ঐঁ	ঐং	ঐঃ		অ	আ	ই	ঈ	উ	ঊ	ঋ	ৠ			এ
U+099x	ঐ			ও	ঐ	ক	খ	গ	ঘ	ঙ	চ	ছ	জ	ঝ	ঞ	ট
U+09Ax	ঠ	ড	ঢ	ণ	ত	থ	দ	ধ	ন		প	ফ	ব	ভ	ম	য
U+09Bx	র		ল				শ	ষ	স	হ			়	২	়া	়ি
U+09Cx	়ী	়ু	়ে	়ং	়ঃ			়ে	়ৈ			়ো	়ৌ	়ে	়ে	
U+09Dx								়ী					়	়		য়
U+09Ex	ঋ	ৠ	ঌ	঍			০	১	২	৩	৪	৫	৬	৭	৮	৯
U+09Fx	ৰ	ৱ	্	ৗ	্	৚	৛	ড়	ঢ়	৞	য়	ৠ	ৡ			

Fig. 2: Bengali Script in Unicode [10].

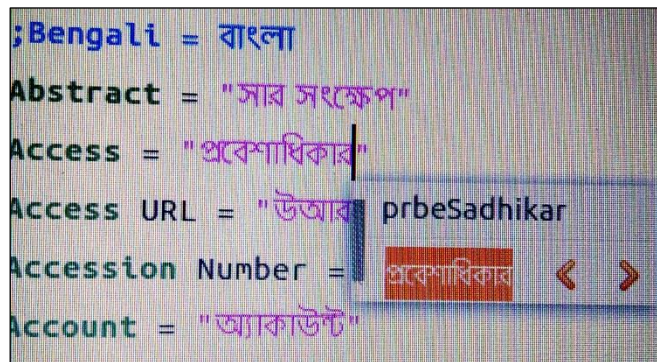


Fig. 3: Bengali Font Typing with Avro and Its Transliteration.

Legend:															
ক	k	ট	T	প	p	স	s	অ	o	ঔ	ৌ	OU	০	0	Case sensitive Not case sensitive (v) - Any vowel (c) - Suitable consonent - Accent Key
খ	kh	ঠ	Th	ফ	ph,f	হ	h	আ	।	ব	(ফলা)	w	১	1	
গ	g	ড	D	ব	b	ড়	R	ই	ি	়	- য ফলা	(c)y, Z	২	2	
ঘ	gh	ঢ	Dh	ভ	bh,v	ঢ়	Rh	ঈ	ী	্	- র ফলা	(c)r	৩	3	
ঙ	Ng	ণ	N	ম	m	য়	y	উ	ু	্	- রেফ	(v)rr (c)	৪	4	
চ	ch	ত	t	য	z	ৎ	t''	ঊ	ু	্	- হসত্ত	''	৫	5	
ছ	chh	থ	th	র	r	ং	ng	ঋ	ৠ	্	- দাড়ি	.	৬	6	
জ	j	দ	d	ল	l	ঃ	:	এ	ৈ	্	- টাকা	\$	৭	7	
ঝ	jh	ধ	dh	শ	sh,S	্	^	ঐ	ৌ	্	- ডট	. (NumPad)	৮	8	
ঞ	NG	ন	n	ষ	Sh			ও	ৌ	্	- কোলন	:'	৯	9	

Fig. 4: Avro Phonetic Keyboard Layout.

Book

In this transliteration mechanism process, another important tools is “Granthagar O Tathya Vijnaner Paribhashakosh: English-Bengali” [13]. Library and information science domain area’s core English terms are being

converted in standard Bengali terms, are selected from this book. Some examples and book details are:

Accession Number = পরিগ্রহণ সংখ্যা (bn.ini file line no. 4)

Call Number = ডাক সংখ্যা (bn.ini file line no. 32)

Series = মাল্লা (bn.ini file line no. 39)
Title: Granthagar O Tathya Vijnaner
Paribhashakosh: English-Bengali
Author: Bimal Kanti Sen
Publisher: Bengal Library Association
Year: 2013
Place: Kolkata
Edition: 2nd

Google Indic Transliteration

Internet giant Google has revealed lots of products for global users. Transliteration service is one of other services. It is able to translate text, speech, snap, images, sites, or real-time video from one language into another. Now, Google transliteration process is working on different mobile apps and OS and communicable in more than 100 languages [14].

RETRIEVAL TECHNIQUE

VuFind is a highly customizable open source Web-scale Resource Discovery (WSRD) product that allows information repositories and brands multiple services into a unified interface. It has powerful Apache Solr for search platform in order to showcase any digital content.

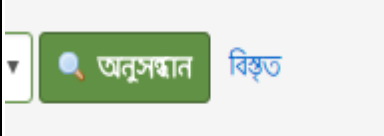
TRANSLITERATION PROCESS

VuFind software development group tendered the software globally to translate, modify,

change and redistribute. In this case, I have selected the transliteration part. In this part, VuFind software interface user can access in Bengali language. The Bengali language transliteration process mechanism has been drawn in step by step.

- **First:** download the master default English language (en.ini) file from VuFind language directory. (Download Location: (<https://github.com/VuFind-org/VuFind/blob/master/languages/en.ini>) (Figure 5).
- **Second:** the master en.ini file edit by the gedit text editor.
- **Third:** all English language phrases being translated into Bengali formal language with the help of above encoding tools. Left side column is language code file (.ini), which was edited by “gedit” text editor and another side is VuFind OPAC interface. Some following screenshots are (Table 1).
- **Fourth:** after completing all phrases, the files were stored in bn.ini file and kept in the actual location (repository hub). Some following examples are (Table 2).
- **Fifth:** Finally bn.ini file has been sent to the development group and they are uploaded in VuFind language directory and Bengali language enabled VuFind software is ready to download and access.

Table 1: Translation mechanism (left side column) and displayed in OPAC interface (right side column).

Sl. No.	In Text Editor (gedit) Screenshot	VuFind OPAC Bengali Interface Screenshot
1	Search = "অনুসন্ধান" Advanced = "বিস্তৃত"	
2	Search Options = "অনুসন্ধানের বিকল্পগুলি" Search History = "ইতিহাসে অনুসন্ধান করুন" Advanced Search = "বিস্তৃত অনুসন্ধান"	অনুসন্ধানের বিকল্পগুলি • ইতিহাসে অনুসন্ধান করুন • বিস্তৃত অনুসন্ধান
3	Search More = "আরও অনুসন্ধান" Browse the Catalog = "গ্রন্থ তালিকা ব্রাউজ করুন" Browse Alphabetically = "বর্ণানুক্রমে ব্রাউজ করুন" Course Reserves = "পাঠ্যক্রম আরক্ষিত" New Items = "নতুন উপাদানগুলি"	আরও অনুসন্ধান • গ্রন্থ তালিকা ব্রাউজ করুন • বর্ণানুক্রমে ব্রাউজ করুন • পাঠ্যক্রম আরক্ষিত • নতুন উপাদানগুলি

4	<p>User Account = "ব্যবহারকারীর অ্যাকাউন্ট"</p> <p>First Name = "নামের প্রথম অংশ"</p> <p>Last Name = "পদবি"</p> <p>Email Address = "ই-মেইল ঠিকানা"</p> <p>Desired Username = "কাম্পিফত ব্যবহারকারির নাম"</p> <p>Password = "পাসওয়ার্ড"</p> <p>Password Again = "আবার পাসওয়ার্ড দিন"</p>	<p>ব্যবহারকারীর অ্যাকাউন্ট</p> <p>প্রথম নাম: <input type="text"/></p> <p>শেষ নাম: <input type="text"/></p> <p>ই-মেইল ঠিকানা: <input type="text"/></p> <p>কাম্পিফত ব্যবহারকারির নাম: <input type="text"/></p> <p>পাসওয়ার্ড: <input type="text"/></p> <p>আবার পাসওয়ার্ড দিন: <input type="text"/></p> <p>← Back <input type="button" value="দাখিল করুন"/></p>
5	<p>An error has occurred = "একটি ত্রুটি ঘটেছে"</p> <p>An error occurred during execution = "অপেক্ষে হবে একটি পরে আরও ঠিক করা চলবে সময় একটি ত্রুটি ঘটেছে, অপর হবে একটি পরে আরও ঠিক করা।"</p> <p>Please contact the Library Reference Department for assistance = "সহায়তা জন্য গ্রন্থাগার বিভাগের সাথে যোগাযোগ করুন"</p>	<p>একটি ত্রুটি ঘটেছে</p> <p>চালানোর সময় একটি ত্রুটি ঘটেছে; অনুগ্রহ করে একটু পরে আবার চেষ্টা করুন।</p> <p>সহায়তার জন্য গ্রন্থাগার রেফারেন্স বিভাগের সাথে যোগাযোগ করুন</p> <p>b.deba87@gmail.com</p>

Table 2: bn.ini file (left side column) and full name of languages (right side column).

<ul style="list-style-type: none"> ar.ini bn.ini ca.ini cs.ini cy.ini da.ini de.ini el.ini en-gb.ini en.ini es.ini eu.ini fi.ini fr.ini ga.ini 	<pre>[Languages] en = "English" ; American spellings ;en-gb = "English" ; British spellings de = "German" es = "Spanish" fr = "French" it = "Italian" ja = "Japanese" nl = "Dutch" pt = "Portuguese" pt-br = "Brazilian Portugese" zh-cn = "Simplified Chinese" zh = "Chinese" tr = "Turkish" he = "Hebrew" ga = "Irish" cy = "Welsh" el = "Greek" ca = "Catalan" eu = "Basque" ru = "Russian" cs = "Czech" fi = "Finnish" sv = "Swedish" pl = "Polish"</pre>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

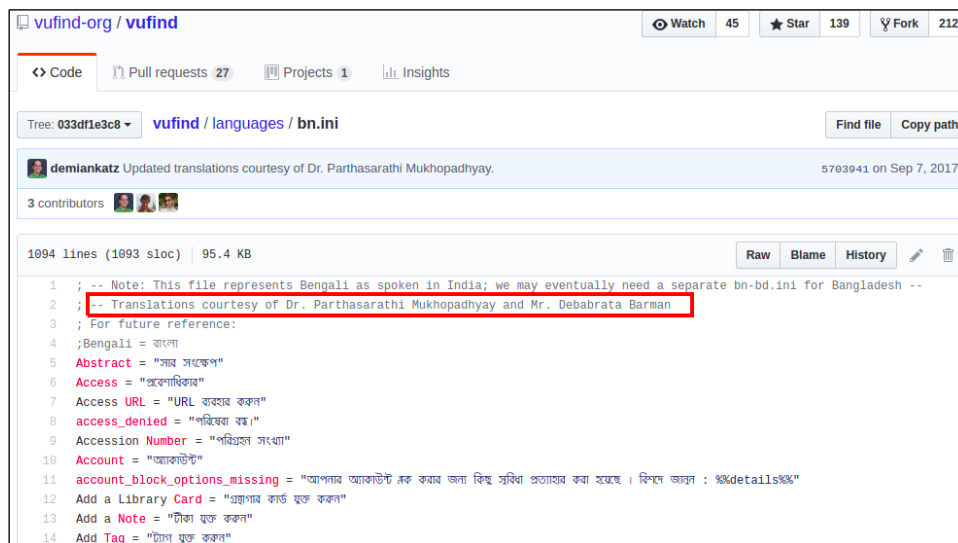


Fig. 5: Expanded Translation Support (<https://github.com/vufind-org/vufind/blob/master/languages/bn.ini>).

CONCLUSION

VuFind is open source, completely modular, and it looks identical to the library's website. It is enabled to users to search and browse library's all resources by replacing the traditional OPAC. In this paper, VuFind Bengali OPAC interface is a new oxygen for Bengali language information seeker. Traditional OPAC interface is not user friendly for retrieve Bengali literature related data, because of language biasness and different search interfaces. Besides, with the VuFind interface users can save their choice list, post their comments, reserve book list etc. so that, VuFind Bengali OPAC is a new milestone for Bengali literature seekers.

REFERENCES

1. Keene C. Discovery Services: Next Generation of Searching Scholarly Information. *Serials*. 2013; 24(2): 193–196p.
2. Mukhopadhyay Parthasarathi. Cross Collection Discovery System in Library: Designing a Framework. In: Halder SN, editor. *Exertion to Establish Knowledge Society: Responsibilities of Academic Libraries*. 2017; 36–42p.
3. Unity in Diversity. In Wikipedia: The Free Encyclopedia. Oct 16, 2016. https://en.wikipedia.org/w/index.php?title=Unity_in_diversity&oldid=744599771 Shaw (Accessed on 16th Oct, 2016).
4. Goel Ajay. *Important India*. "Unity in Diversity" in India. Jul 3, 2013. <http://www.importantindia.com/270/article-on-unity-in-diversity-in-india/> (Accessed on 15th Oct, 2016).
5. Languages of India. In Wikipedia: The Free Encyclopedia. Oct 16, 2016. https://en.wikipedia.org/w/index.php?title=Languages_of_India&oldid=744629531 (Accessed on 16th Oct, 2016).
6. Lane J. *The 10 Most Spoken Languages in the World* [Corporate]. 2017. <https://www.babble.com/en/magazine/the-10-most-spoken-languages-in-the-world> (Accessed on 21st Apr, 2017).
7. Das Gupta I, Nessa N. Enabling Indic Support in Library Information Systems: An Opensource Localizer's Perspective. *3rd International CALIBER*; Cochin: INFLIBNET Centre; 2005; 132–145p.
8. Mukhopadhyay P, Azim SA. Multi-script Information Retrieval System: A FLOSS based Prototype for Indic Scripts with Special Reference to Bengali Script. *The Conference on Information Management in Digital Libraries*; Kharagpur: Indian Institute of Technology. 2006; 305–316p.
9. Rahaman SS, Al-Mahmud Islam MR, Akhand MAH. *Design and Development of a Bengali Unicode Font Converter*. *IEEE Conference Publications*, 2013. Dhaka: IEEE; 2013; 1–4p. <https://doi.org/10.1109/ICIEV.2013.6572667>
10. Bengali (Unicode block). In Wikipedia: The Free Encyclopedia. Apr 12, 2017. [https://en.wikipedia.org/w/index.php?title=Bengali_\(Unicode_block\)&oldid=775132535](https://en.wikipedia.org/w/index.php?title=Bengali_(Unicode_block)&oldid=775132535) (Accessed on 12th Apr, 2017).
11. Bengali Language. In Wikipedia: The Free Encyclopedia. Oct 15, 2016. https://en.wikipedia.org/w/index.php?title=Bengali_language&oldid=744500711 (Accessed on 15th Oct, 2016).
12. Avro Keyboard. (). In Wikipedia: The Free Encyclopedia. Apr 01, 2017. https://en.wikipedia.org/w/index.php?title=Avro_Keyboard&oldid=773355570 (Accessed on 01st Apr, 2017).
13. Sen Bimal Kanti. *Granthagar O Tathya Vijnaner Paribhashakosh: English-Bengali*. 2nd Edn. Kolkata: BLA; 2013.
14. Google Translate. (). In Wikipedia: The Free Encyclopedia. Nov 15, 2017. https://en.wikipedia.org/w/index.php?title=Google_Translate&oldid=749610124 (Accessed on 15th Nov, 2017).

Cite this Article

Debabrata Barman, Parthasarathi Mukhopadhyay. Library Discovery System in Bengali Script: An Experiment with VuFind. *Journal of Advancements in Library Sciences*. 2018, 5(2): 20–26p.