

Transliteration Generator for Devnagari using Mangal Font

Arti P. Khadtare¹, Dr. Suhas Raut², M. S. Otari³

^{1,2 & 3} N.K. Orchid College of Engineering and Technology, Solapur

ABSTRACT

Transliteration generation is a part of language processing. A transliteration generator is a program to generate the transliteration for an input word given in Marathi (Devnagari). In current technique the transliteration generated for an input word depends on font used for typing the text. In proposed system we provide the transliteration generator for Mangal font. This transliteration of Marathi (Devnagari) words is used for further text processing such as translation (i.e. translating text from one language into another), comprehension etc. or other NLP (Natural Language Processing) applications.

Keywords: Transliteration, NLP, Devnagari fonts.

1. INTRODUCTION

Now a day's use of regional languages in computing is increased. For Maharashtra, Marathi is the language used by many people. To be able to write in Marathi (Devnagari) user has to install Devnagari font. The fonts available for typing in Devnagari are Aakar, Aakriti, Akshar, Shivaji, Mangal, Kruti, Ganga, Gauri, Kiran etc. User can use local language setting and on-screen keyboard feature available with operating system to write in Marathi (Devnagari) language. But the key mapping of standard keyboard for Marathi is different for different fonts. As user may not always have this keyboard mapping readily available with him. For Marathi language user it is difficult to spell Marathi words if he doesn't have enough English knowledge. For fonts like Shivaji, to type the Marathi word requires to type the word in equivalent English letters i.e. transliteration. For e.g. to type the word 'ज', user has to type two letters 'j' and 'a'. So user should know which alphabet represents exactly which word i.e. keyboard mapping for Shivaji font. And if he is not aware of it then it can be troublesome. But in case of Mangal font user can directly type Marathi words using the keyboard, because keyboard mapping required for Mangal font is available on on-screen keyboard and user doesn't need to know how to generate exact spelling because by setting local language to Marathi he can directly type in Devnagari. But internal representation in such a font is also in Marathi unlike Shivaji in which it is in form of transliteration. So for user to be able to enter words in Marathi without bothering about keyboard mapping, we have designed transliteration generator. This will convert the word entered in Marathi (Devnagari) to its equivalent transliteration.

2. RELATED WORK

From an information-theoretical point of view, systematic transliteration is a mapping from one system of writing into another, word by word, or ideally letter by letter. Most transliteration systems are one-to-one, so a reader who knows the system can reconstruct the original spelling. Transliteration of single words is often an informal non-systematic process; many variants of the same word are often used [8]. The International Alphabet of Sanskrit Transliteration (IAST) is a subset of the ISO 15919 standard, used for the transliteration of Sanskrit and Pali into Roman script with diacritics. IAST is a widely used standard. The Hunterian system is the "national system of romanization in India" and the one officially adopted by the Government of India. Compared to IAST, Harvard-Kyoto looks much simpler. It does not contain any of the diacritic marks that IAST contains. Instead of diacritics, Harvard-Kyoto uses capital letters. The use of capital letters makes typing in Harvard-Kyoto much easier than in IAST but produces words with capital letters inside them. ITRANS is an extension of Harvard-Kyoto [9]. There are many software's available which convert word typed using transliteration to Marathi text for example Baraha Keyboard where a user can type transliteration of Marathi word using English letters but the output will be Marathi text. In Shivaji font, for example, to write "ज" user has to enter "pUjaa". Here to represent "ज" user has to enter two alphabets 'j' and 'a' to make it complete word "ज", as it is a consonant. But for common user who knows Marathi but is not proficient in English and user who knows English but is not proficient in Marathi it can be a difficult task. But none of these provide user the reverse process i.e. entering Marathi (Devnagari) words which will be converted to its respective English letter.

3. TRANSLITERATION GENERATOR FOR MANGAL FONT

3.1 WORKING OF PROPOSED SYSTEM

In the proposed system we are going to replace the entered Marathi word with its equivalent English alphabet. For this we are going to create a list which will contain Marathi word and an English alphabet which represents it. As soon as user enters Marathi word from keyboard program will look up in list for entry and will replace it in string with equivalent English word. For performing exact match we will use Unicode representation of words to compare. Figure.1 shows the in general processing of proposed system. In our system we have used representation similar to Harvard-Kyoto i.e. we are using capital letters to represent words, e.g. 'ध'- Dh, 'ड'-D. But we are eliminating the 'a' in the end of a consonant to keep transliteration simple which will help in further processing of text in different application [1] - [7]. Here we are using Shivaji font to compare how the proposed system for Mangal font helps and reduces overhead of user. Fig 2.a and 2.b shows the different transliteration along with transliteration generated by proposed system.

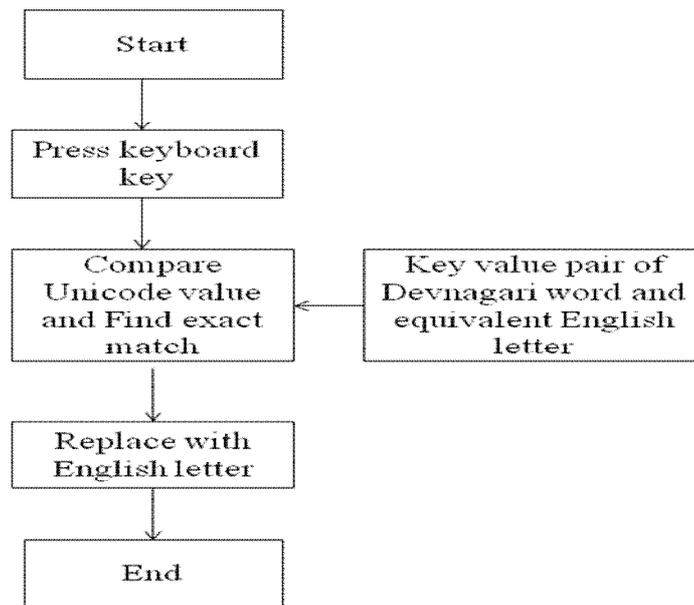


Figure 1: Processing of proposed transliteration generator

Devanagari	IAST	Harvard-Kyoto	ITRANS	Proposed System
अ	a	a	a	a
आ	ā	A	A/aa	aa/A
इ	i	i	i	i
ई	ī	I	I/ii	ii/I
उ	u	u	u	u
ऊ	ū	U	U/uu	uu/U
ए	e	e	e	e
ऐ	ai	ai	ai	ai
ओ	o	o	o	o
औ	au	au	au	au
अं	ṁ	M	M/.n/.m	M
अः	ḥ	H	H	H

Figure 2.a Transliteration of Vowels [2]

Devanagari	IAST	Harvard-Kyoto	ITRANS	Proposed System
क	ka	Ka	ka	K
ख	kha	Kha	kha	kh
ग	ga	Ga	ga	g
घ	gha	Gha	gha	gh
ङ	ṅa	Ga	~Na	Da
च	ca	Ca	cha	c
छ	cha	Cha	Cha	Ch
ज	ja	Ja	ja	j
झ	jha	Jha	jha	jh
ट	ṭa	Ta	Ta	T
ठ	ṭha	Tha	Tha	Th
ड	ḍa	Da	Da	D
ढ	ḍha	Dha	Dha	Dh
ण	ṇa	Na	Na	N
त	ta	Ta	ta	t
थ	tha	Tha	tha	th

Devanagari	IAST	Harvard-Kyoto	ITRANS	Proposed System
द	da	da	da	d
ध	dha	dha	dha	dh
न	na	na	na	n
प	pa	pa	pa	p
फ	pha	pha	pha	ph
ब	ba	ba	ba	b
भ	bha	bha	bha	bh
म	ma	ma	ma	m
य	ya	ya	ya	y
र	ra	ra	ra	r
ल	la	la	la	l
व	va	va	va/wa	v
श	śa	za	sha	sh
ष	ṣa	Sa	Sha	Sh
स	sa	sa	sa	s
ह	ha	ha	ha	h

Figure 2.b Transliteration of Consonants [2]

3.2 WORKING WITH SHIVAJI FONT

In Shivaji font, it is necessary for user to type five letters ‘javaL’ to spell the word ‘जवळ’ i.e. as ‘j’ and ‘v’ are consonant to complete them vowel ‘a’ has to be added but this is not the case for word ‘L’. Here though user is typing using English letters the output or text that is displayed on screen is in Devnagari. That is user must have knowledge about how to spell the Marathi (Devnagari) words using English letters on keyboard. The problem that arises here is that this transliteration may vary according to users understanding. User may try to type ‘Ka’ or ‘kha’ to represent ‘ख’ but here only single capital letter ‘K’ is enough to represent ‘ख’.

~	ॠ	ॡ	@	ॢ	#	ॣ	\$	।	%	॥	^	०	&	१	*	२	(३)	४	-	५	+	६	Back
1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	0	0	-	५	=	६	Space	
Tab	Q	W	E	R	T	Y	U	I	O	P	{	}		5											
	q	w	e	r	t	y	u	i	o	p	[]	\	/											
Caps	A	S	D	F	G	H	J	K	L	:	"	"	Enter												
Lock	a	s	d	f	g	h	j	k	l	:	"	"													
Shift	Z	X	C	V	B	N	M	<	>	?	Shift														
	z	x	c	v	b	n	m

Figure 3. keyboard layout for Shivaji font [10]

3.3 WORKING WITH MANGAL FONT

For Mangal font, as shown in figure 3.2, user has to type only three letters ज, व, ङ to type the word ‘जवळ’ in Devnagari keyboard layout i.e. he do not have to bother about which English alphabet represents which Devnagari letter because he can directly press the key for Devnagari word from keyboard. Here user is typing in Devnagari letters and the output or text displayed on screen is also in Devnagari form. The proposed system will then convert the word to its equivalent transliteration ‘jvL’ by using unicode value comparison.

