Drivers for Unicode Character System

V. R. Wable
BE Computer

GCoE Avasari

A. V. Chechare
BE Computer

GCoE Avasari

R. S. Lawande.
BE Computer

GCoE Avasari

V. R. Jadhav.
BE Computer

GCoE Avasari

ABSTRACT

As far as current reality and emerging trends in global management practices is concern, the use and role of information technology is changed very drastically. At the earlier phase, the language or media of storing data on computer was restricted to English language because of ASCII(American Standard for Information Interchange) standards developed by Bell laboratories.But as the computerization or automation increased in the field of business, banking, insurance, e-governance, e-books, e-commerce etc. it creates a new challenge to store different types of data(text/audio/video/images etc) in different volumes(quantity), also the network communication. This challenge spawns different areas in IT industry like data warehousing, data mining, www, networking, computer security etc. *But the present study focuses on the area which can influence on the computer users across the world.*The objective of the study is to explain the algorithm by which any computer user can communicate with computer in his/her mother tongue like (Marathi/Hindi/Sanskrit/Gujarati hundreds of Languages across the world)This is implemented by using UNICODE(Universal Code) which was introduced in 1988. But UNICODE could not become so popular as it was expected. We have implemented an algorithm and program in C#.net by which Computer user can customize the keystrokes of keyboards to match the keystroke of their mother tongue. He can create multiple keyboard layouts for many languages.This is immensely helpful for all the users from KG to PG and research, because language will not be the barrier for learning the use and usage of Computer.

1. INTRODUCTION

This research paper is representing a sub-theme in "Current Reality and Emerging Trends in Global Management Practices".This research paper has considered both the terms information and communication. Both of these two terms are depends upon Language, which is a media of communication. Communication might be in human being to human being or human being to machine i.e. Computer Till now, most of the computer users uses English for the communication (i.e. storing and retrieving the data on computer) It implies that user should have a knowledge of English. The ordinary computer user could not do a programming work or any other type of work in various Indian regional languages like Hindi/Gujarati/Tamil etc.Generally Desktop and Menus of Computer are in English, So it is required to have sufficient knowledge of English to understand the meanings of messages displayed on the screen. And this is the main drawback in man-machine communication. Now it is necessary to have an opportunity to learn and use a computer in various Indian languages.This gap can be bridged with help of

this topic in which we are going to explain an algorithm which will help to ordinary user to operate a computer in his regional languages(e.g. Marathi/Hindi/Gujarati etc). He can choose keyboard settings by his/her choice.

This paper highlights the following areas:
1. Various alternatives to use local Indian languages
2. Use of UNICODE character set in computer operations
3. Algorithm and flow of execution [Designing Keyboard driver for any regional language]
4. Screen dumps generated by the software for Marathi keyboard Driver
5. Advantages and Future Expansions
6. Conclusion

2. VARIOUS ALTERNATIVES TO USE LOCAL INDIAN LANGUAGES

a. FONT :- Fonts have often been indiscriminately mapped to the same set of bytes e.g. 0x00 to 0xff are ofter used for both character and dingbats.

Drawbacks
  I. It is required to be installed in client machine.
  II. It uses bitmap or set of bytes i.e. expensive as far as memory is concern.
  III. Use of multiple inconsistent character codes because of conflicting national and industry standards.

b. GIST:- It is an abbreviation of Graphics and Intelligence Based Script Technology. It is a hardware based solution for Indian languages developed by C-DAC.

Disadvantages
  I. GIST can handle only one Indian language at a time.It cannot be used without GIST card (hardware).
  II. It cannot be used for multi-lingual documents of languages

c. UNICODE:- The concept of Unicode is introduced to solve the above problems in supporting multi-lingual usage.

The UNICODE standards was designed to be Universal :- Same for IBM/ISO/MAC etc.

  I. Efficient:- Plain text is simple to parse, Quick and sorting process
  II. Uniform:- A fixed character set allows for efficient sorting, searching, display and editing of text.
  III. Unambiguous :- Unique identification to each character

3. USE OF UNICODE CHARACTER SET IN COMPUTER OPERATIONS

Some Ranges of languages are shown in fig(a) while fig(b) shows the available UNICODE character set of Devnagari (i.e. Marathi/Hindi/Sanskrit/Nepali).
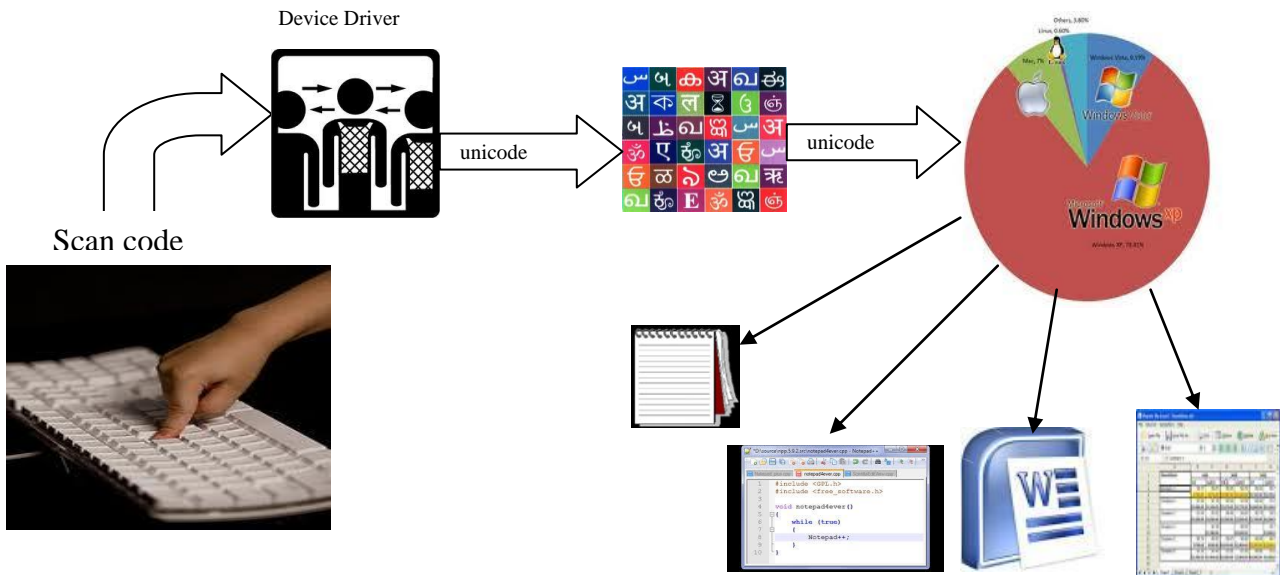
Unicode Chart

Range	Decimal	Name
0x0000-0x007F	0-127	Basic Latin
0x0080-0x00FF	128-255	Latin-1 Supplement
0x0100-0x017F	256-383	Latin Extended-A
0x0180-0x024F	384-591	Latin Extended-B
0x0250-0x02AF	592-687	IPA Extensions
0x02B0-0x02FF	688-767	Spacing Modifier Letters
0x0300-0x036F	768-879	Combining Diacritical Marks
0x0370-0x03FF	880-1023	Greek
0x0400-0x04FF	1024-1279	Cyrillic
0x0530-0x058F	1328-1423	Armenian
0x0590-0x05FF	1424-1535	Hebrew
0x0600-0x06FF	1536-1791	Arabic
0x0700-0x074F	1792-1871	Syriac
0x0780-0x07BF	1920-1983	Thaana
0x0900-0x097F	2304-2431	Devanagari
0x0980-0x09FF	2432-2559	Bengali
0x0A00-0x0A7F	2560-2687	Gurmukhi

Fig(b)

Fig(a)

4. ALGORITHM AND FLOW OF EXECUTION [Keyboard driver or any regional language]

5. SCREEN DUMPS GENERATED BY THE SOFTWARE FOR MARATHI KEYBOARD DRIVER

Fig. (d)

Fig(e)

Fig(f)

Fig(g)

These screen dumps are showing the present status of keyboard and other help. Secondly he/she can customize the keystrokes of keyboard.

मराठी जोडशब्द सराव

विशेष सूचना :- कोनताही जोदशब्द टाइप करताना पहिल्या अक्षराचा पाय मोडावा.
(पाय मोडन्यासाठी ~ चा वापर करावा) हे बटन कीबोर्डवर टॅब बटनच्या वर व १ किंवा । च्या शेजारी आहे.
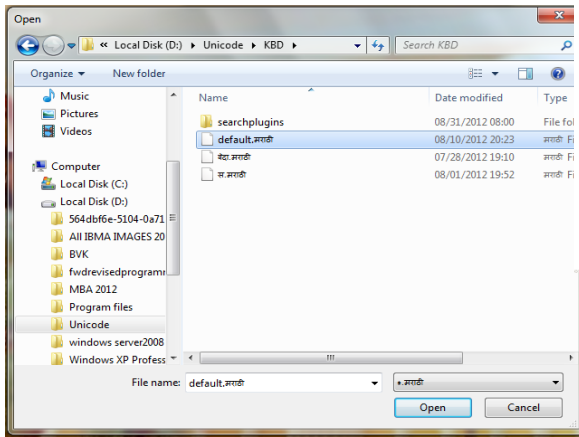
स्वर : -

अ	आ		ई	उ	ऊ	ए	ऐ		ओ	औ	अं	अः
a	x		e	E	[	]	{		}	V	Y	aM	a:

व्यंजणे : -

क	का	कि	की	कु	कू	के	कै		को	कौ	कं	कः
k	kA	ki	kI	ku	kU	ko	kO		k#	k$	kM	k:

ख	खा	खि	खी	खु	खू	खे	खै		खो	खौ	खं	खः
K	KA	Ki	KI	Ku	KU	Ko	KO		K#	K$	KM	K:

ग	गा	गि	गी	गु	गू	गे	गै		गो	गौ	गं	गः
g	gA	gi	gI	gu	gU	go	gO		g#	g$	gM	g:

Fig(h)

विशेष सूचना :- कोनताही जोदशब्द टाइप करताना पहिल्या अक्षराचा पाय मोडावा.
(पाय मोडन्यासाठी ~ चा वापर करावा) हे बटन कीबोर्डवर टॅब बटनच्या वर व १ किंवा । च्या शेजारी आहे.
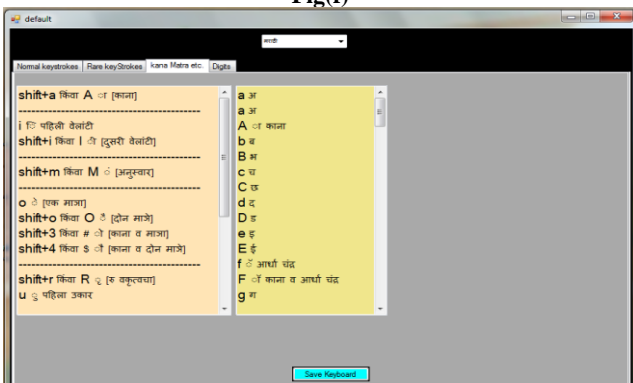
१. प्रतिक = प ् र त ि क
२. श्रीनिवास = श ् र ी न ि व ा स
३. स्वागत = स ् व ा ग त
४. कॉम्प्युटर प्रोग्रामिंग = क ॉ म ् प ् य ु ट र   प ् र ो ग ् र ा म ि ं ग
५. प्रकृती = प ् र क ृ त ी
६. वस्त्रालय = व स ् त ् र ा ल य
७. स्पर्श परिक्षा = स ् प र ् ध ण   प र ि क ् ष ण
८. नेतृत्व = न े त ृ त ् व
९. संस्कृत = स ं स ् क ृ त
१०. व्यक्तीपरत्वे निष्कर्ष = व ् य क ् त ी प र त ् व े   न ि ष ् क र ् ष
११. राशी भविष्य वर्ष २०१२ = र ा श ी   भ व ि ष ् य   व र ् ष   २ ० १ २
१२. कार्य कौशल्य = क ा र ् य   क ौ श ल ् य
१३. सुखस्थानातील शुक्र = स ु ख   स ् थ ा न ा त ी ल   श ु क ् र
१४. ज्ञानेश्वरी अध्याय = ज ् ञ ा न े श ् व र ी   अ ध ् य ा य
१५. औरत = औ र त
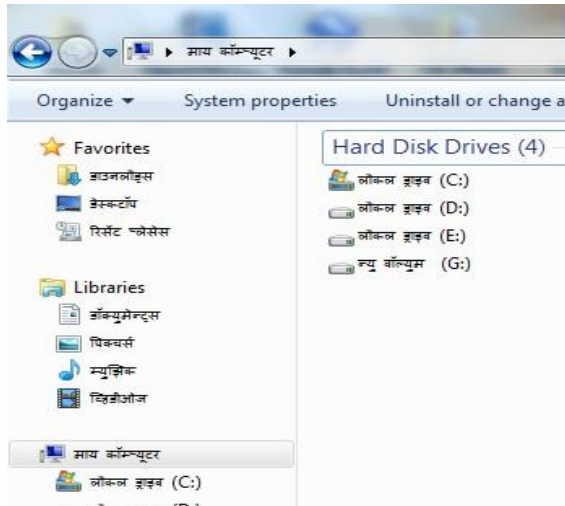१६. ओसरी = ओ स र ी

Fig(i)

Fig (j) shows that the whole desktop can be converted to Regional language

6. ADVANTAGES AND FUTURE EXPANSIONS

a.   The same concept can be adopted by other languages

b.   Operating Systems Desktop can be customized according to regional language selection .Keyboard Layout can be displayed visually. Language dependent versions of OS and editors can be designed
                                    .

7. CONCLUSION

As information is present in various languages and scripts, the importance of Unicode cannot be overemphasized. Many publishers, web sites, universities and government have been handling information in various languages and when the information is to be digitized, Unicode becomes essential and handy. Not it becomes mandatory that the operating systems, DBMS systems and web hosting software must have Unicode support, for better searching and editing. India is a one of the big countries in the world in which various types of languages are used, all these language contains rich literature and scientific information For digitization the communication should be in mother tongue and can be achieved by our software.

8. ACKNOWLEDGEMENT

We would like to thank all the people around us who had supported us and still supporting us throughout the research, our family ,friends and well-wishers and last but not least the almighty god for giving us strength.

9. REFERENCES:

[1]   Unicode Consortium www.unicode.org

[2]   UNICODE Code Charts www.unicode.org/charts/

[3]   Unicode Standard www.unicode.org/standard/standard.html

[4]   LANGUAGES   and   Scripts   of   India http://www.cs.colostate.edu/~malaiya/scripts.html

[5]   Unicode 6.2.0 www.unicode.org/versions/Unicode6.2.0/

[6]   Microsoft   Home   Page | Devices   and   Services www.microsoft.com/

[7]   MSDN – Explore Desktop, Web, Cloud, and – MSDN Microsoft http://msdn.microsoft.com/en-in/

[8]   http://msdn.microsoft.com/en-us/library/ms123401.aspx

[9]   Murach's C# 2008 By Mike Murach and Associates

[10] C# Programming Language (Covering C# 4.0), The (4th Edition) By Anders Hejlsberg, Mads Torgersen, Scott Wiltamuth, and Peter Golde

[11] Effective C#: 50 Specific Ways to Improve Your C# (2nd Edition) By Bill Wagner

[12] C# in Depth, Second Edition By Jon Skeet

[13] Essential C# 4.0 (3rd Edition) (Microsoft .NET Development Series) By Mark Michaelis

[14] Professional C# 4.0 and .NET 4 By Bill Evjen, Karli Watson, Christian Nagel, and Morgan Skinner

[15] SamsTeach Yourself Visual C# 2010 in 24 Hours: Complete Starter Kit By Scott Dorman

[16] C# 4.0 in a Nutshell: The Definitive Reference By Joseph Albahari, Ben Albahari