

Uniscribing Issues and Win Myanmar Spell-Checker: an Implementation of Myanmar Character Set

Presented at the Workshop on Encoding Myanmar Character Set. Sponsored by Forever Group.

IBC, Yangon, Myanmar
Sunday, 26th August 2001

by **Zaw Htut**
System Developer
Win Myanmar Systems
<http://winmyanmar.bizhosting.com>

Member, Working Group for Myanmar Character Set Encoding, MCSA

Webmaster / Technical Director
WWW Media Co Ltd
INTERNET: <http://www.myanmars.net>
INTRANET: <http://www.e-application.com.mm/myanmarsnet>
EMAIL: info@www.com.mm

ယူနီကုတ် (Unicode)

- Unicode ဆိုသည်မှာ တစ်ကမ္ဘာလုံးရှိရှိသမျှ ဘာသာ (language) များမှ အက္ခရာများ အားလုံးကို ထည့်သွင်းထားသော information ကူးပြောင်းမှု လွယ်ကူစေရေး၊ တစ် ကမ္ဘာလုံး အသုံးပြုနိုင်သည့် software များ ပေါ်ပေါက်ရေး တို့အတွက် တီထွင်ထားသည့် စံနစ်ကြီးတစ်ခုဖြစ်။
- ISO 10646 structure အရ 16 bit ရော 32 bit ပါ define လုပ်ထားသဖြင့် Unicode တွင် အများဆုံး စာလုံးပေါင်း ၂,၁၄၇,၄၈၃,၆၄၈ လုံး ထည့်သွင်းနိုင်။
- ယခု 16 bit တွင် ပင် စာလုံးပေါင်း ၆၅၅၅၅ ထည့်နိုင်။ မြန်မာနှင့် မြန်မာတိုင်းရင်းသား ဘာသာ များအတွက် (ASCII နံပါတ်ကဲ့သို့) အက္ခရာ အမှတ် 1000-109F အတွင်း စာလုံးပေါင်း ၁၆၀ စာ နေရာ ရရှိထား။



The Unicode Standard, Version 3.0 စာမျက်နှာ ၄၄၀ တွင် ဖော်ပြထားပုံ

	Myanmar									
	100	101	102	103	104	105	106	107	108	109
0	က	ခ	ဂ	ဃ	င	စ	ဆ	ဇ	ဈ	ည
1	ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ
2	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ	ဎ	ဏ	တ
3	ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ
4	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ	ဎ	ဏ	တ
5	ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ
6	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ	ဎ	ဏ	တ
7	ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ
8	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ	ဎ	ဏ	တ
9	ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ
A	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ	ဎ	ဏ	တ
B	ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ
C	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ	ဎ	ဏ	တ
D	ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ
E	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ	ဎ	ဏ	တ
F	ဋ	ဌ	ဍ	ဎ	ဏ	တ	ထ	ဒ	ဌ	ဍ

အိုးပင်းတိုက်ပီ ဗွန် (OpenType Font)

- ယခင် TrueType Font များထက် ပိုအဆင့် မြင့်လာသော font အမျိုးအစား ဖြစ်ပြီး၊ အက္ခရာ တစ်ခုချင်းအတွက် vector definition တင် မကတော့ဘဲ၊ ပိုမိုရှုပ်ထွေးသည့် logic ကိုပါ ထည့်သွင်း၍ ရသည့် font technology ဖြစ်။
- ယူနီကုတ်အဖြစ် အသုံးပြုလာရမည့် အခါတွင် ပုံစံကွဲ မည်သို့ပင် ရှိစေကာမူ အက္ခရာ တစ်ခုစာအတွက်သာ ထည့်သွင်း encode လုပ်မည်ဖြစ်။ opentype font (OTF) technology ကိုမိမိ၍လည်း ဤသို့ လုပ်ခြင်းဖြစ်၍ OTF ကို မြန်မာစာ ကွန်ပျူတာစနစ်အတွက် မသုံးမဖြစ် အသုံးပြုရမည် ဖြစ်။

ယူနီကုတ်၏ အားသာချက်များ

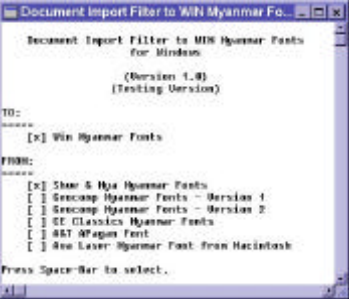
- အမှန်တကယ် စနစ်ကျသည့် မြန်မာစာသုံး ကွန်ပျူတာစနစ် စတင် ပေါ်ပေါက်လာမည်။
- search လုပ်ရန် များစွာ လွယ်လာမည်။
- application များကို မြန်မာမှုပြုရန် လွယ်ကူ လာမည်။
- spell-check လုပ်ရန်တွင်ပင် မြန်မာ-အင်္ဂလိပ် ရောရိုက်ထားသည့် စာသားကိုပင် ဘာသာ ၂ မျိုးလုံးဖြင့် စာလုံးပေါင်း တစ်ခါတည်း စစ်၍ ရနိုင်လာမည်။
- စာရိုက်သည့်အခါ အင်္ဂလိပ်မှ မြန်မာ၊ မြန်မာမှ အင်္ဂလိပ် အကြိမ်ကြိမ် font ပြောင်းရန် မလိုဘဲ၊ ခလုတ် တစ်ချက် နှိပ်ရုံဖြင့် keyboard layout ရော၊ language ပါ ပြောင်းနိုင်မည်။ ၂ ဘာသာ လုံးကို font တစ်မျိုးတည်းဖြင့် ရိုက်နိုင်မည်။
- word-wrapping အတွက် ဘာမှ စိတ်ပူစရာ မလိုတော့ဘဲ လွတ်လွတ်လပ်လပ် ရိုက်နိုင်မည်။ space bar နှိပ်၍ ကြိုလွန်း စိတ်လွန်းနေသည်ကို လိုက်ညီရန် လိုတော့မည် မဟုတ်ပါ။

ယူနီကုတ်၏ အားနည်းချက်များ

- 2-byte font ဖြစ်လာမည့်အတွက် မြန်မာဘာသာဖြင့် ရိုက်ထားသည့် document များ အရွယ် မှာ ၂ဆခန့် ဖောင်းပွလာခြင်း။
- Unicode သုံး၍ ရိုက်ထားသော စာများကို model အနိမ့် စက်အဟောင်းများသို့ ကူးပြောင်း၍ မရတော့ခြင်း။
- 2-byte character များမှာ debug လုပ်လျှင် 1 byte စီ ခွဲမြင် ရတတ်သည့်အတွက် application developer / programmer များ အတွက် အနည်းငယ် ခက်ခဲနိုင်ခြင်း။
- C#, C++, VB စသည့် programming language များကိုပင် မြန်မာဘာသာဖြင့် လုပ်၍ ရလာနိုင်ခြင်း။

စာများကို စနစ်တစ်ခုမှ တစ်ခုသို့ ကူးပြောင်းပေးသည့် ပရိုဂရမ်

အခြားစနစ်များသုံး၍ ရိုက်ထားသည့် စာများကို Win Myanmar Character Set အဖြစ် ကူးပြောင်း ပေးသည့် ပရိုဂရမ်ဖြစ်။ Unicode format အတွက်လည်း ဤပရိုဂရမ်မျိုး မရှိမဖြစ် လိုအပ်လာမည်ဖြစ်။



ဝင်းမြန်မာ Character Set ကို အသုံးပြုသော အခြားစနစ်များ

- (1)Nan fonts
- (2) A-Myanmar fonts
- (3)Aung San fonts
- (4)Bo-tayza fonts
- (5)Nanda font series

မြန်မာဘာသာ စာလုံးပေါင်း သတ်ပုံစစ်သည့် ပရိုဂရမ် (WinMyanmar Spell-Checker)



အသုံးဝင်မည့် လိပ်စာများ

Unicode Consortium နှင့်တွဲဖက်၍ Myanmar Character Set (Unicode 3) ထွက်ပေါ်လာရေး တွင် လုပ်ဆောင်ခဲ့သော ပုဂ္ဂိုလ်များကို လွယ်ကူစွာ ဆက်သွယ်နိုင်ရန် ဖော်ပြလိုက်ပါသည်။
 Dr Kyaw Thein <tech@mptmail.net.mm>
 U Thaug T In <kmd.ygn@mptmail.net.mm>
 U Thein Htut <geo@mptmail.net.mm>
 Dr Aung Maw <cetech@mptmail.net.mm>
 Zaw Htut <zhtut@winmyanmar.bizhosting.com>
 Kenneth Whistler <kenw@sybase.com>
 Michael Everson <everson@indigo.ie>
 Lee Colins <lcolins@apple.com>
 INTERNET: <http://www.unicode.org>

ဆွေးနွေးသင့်သည့် ကိစ္စများ

- စံသုံး ကွန်ပျူတာ လက်ကွက် (standard keyboard layout)
- စံသုံး ရိုက်နည်းရိုက်ဟန် (standard input method)
- အင်္ဂလိပ်ဘာသာဖြင့် မြန်မာစာလုံးများကို အသံထွက် မှန်ကန်စေရန် ပေါင်းသည့် စနစ် (romanizing Myanmar language)
- တိုင်းရင်းသား ဘာသာစကားများမှ စာလုံးများကိုလည်း Unicode တွင် ဖြည့်စွက် ထည့်သွင်းသွားရေး
- တက္ကသိုလ်များမှ ဘာသာဗေဒပညာရှင်များ ပါဝင် လုပ်ကိုင်မှု ရရှိနိုင်မည့် computerized translation ကဲ့သို့သော စီမံကိန်းကြီးများကို ကွန်ပျူတာ အသင်းချုပ်မှ အကူအညီပေးရေး