

Hail Unicode

သင်ပုန်းကြီးအလိုက် ကုဒ်ပွိုင့်

မြန်မာစာကို သင်ပုန်းကြီးလို code point အလိုက် မစီထားရလို့ Backward compatible ဖြစ်အောင် မလုပ်ပဲ အမှန် ဖြစ်အောင် ပြင်လိုက်ရပါတယ်လို့ ရေးကြ ရှိကြတာတွေ တွေ့မိလို့ပါ။

ပထမဆုံး မြန်မာစာသင်ပုန်းကြီးအတိုင်းစီပြီးရင် sort လုပ်နိုင်ပြီဆိုတော့ အကုန်လုံးကို “က” ကြီးကနေ “အ” အထိ အစအဆုံးစီပြီးပြီ။ ဒီတော့ မြန်မာစာ (အမှန်က ဗမာစာ) တွေ အကုန် sort လုပ်နိုင်ပြီပေါ့။ ဗမာတွေအတွက်တော့ ပျော်စရာပဲ။ ဗမာလဲ “က” ရှမ်းလဲ “က” ဗမာလဲ “သ” ရှမ်းလဲ “သ” ဆိုတာမျိုး အကုန်တူနေတာမဟုတ်တော့ ရှမ်းစာ ကိုကော က ကနေ အ အထိ မြန်မာစီသလိုပဲ စီလို့ရမှာလား။ ရှမ်းဆိုတာက နမူနာ တစ်ခုပဲ ရှိသေးတယ်။ မြန်မာယူနီ ကုဒ်လို့ ခေါ်တဲ့ Range ထဲမှာပါတာ ဘာသာစကား ၂၀ ကျော်တယ်ဆိုပဲ။ မွန်တို့ ကရင်တို့ ပလောင်တို့ ပအိုဝ်တို့ အဲ တာတွေကကော အကုန်စီလို့ရပြီပေါ့။ ဗမာသင်ပုန်းကြီးနဲ့ပဲ စီလို့ရမှာလား ကျန်တာတွေ သူတို့ အက္ခရာအတိုင်း စီလို့ ရနိုင်မလား။ အဲဒီမှာ စပြီး စဉ်းစားရပြီ။

အင်္ဂလိပ်အက္ခရာမှာဆိုရင်ကော sort လုပ်ရင် ဘာလို့ a အကြီးအသေး အကုန်လုံးတူတူပြရတာလဲ။ code point အရ ဆိုရင် A အကြီးနဲ့ a အသေး Point မတူပါဘူး။ A အကြီးက U+0041 ဖြစ်ပြီး a အသေးဆိုရင်တော့ U+0061 ဖြစ်ပါ တယ်။ ထို့အတူ FULLWIDTH LATIN A ဟာ U+FF41 ဖြစ်ပြီး FULLWIDTH LATIN a ဟာ U+FF61 ဖြစ်ပါ တယ်။ CYRILLIC CAPITAL LETTER A U+0410 ဆိုပြီး ရှိပါသေးတယ်။ ဒါဟာ နမူနာဖြစ်ပြီး တခြားသော A တွေ ရှိ ပါသေးတယ်။ အဲတော့သူတို့ ကုဒ်ပွိုင့်ဟာ မတူရုံသာမက နေရာတော်တော်ကြီး ခြားနေတာပဲမဟုတ်လား အဲတာဆို ဘာလို့ တူတူပြပေးနိုင် Sort လုပ်နိုင်ရတာလဲ။ Comment မှာ အခြား A နမူနာတွေ ထပ်ထည့်ပေးလိုက်ပါတယ်။

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
X	Y	Z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
u	v	w	x	y	z	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î		
Ĭ	Đ	Ñ	Ò	Ó	Ô	Õ	Ö	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß	à	á	â	ã	ä	å	æ
ç	è	é	ê	ë	ì	í	î	ï	ð	ñ	ò	ó	ô	õ	ö	ø	ù	ú	û	ü	ý	þ

by Latin Script.

											!	"	#	\$	%	&	'	()	*	+
,	-	.	/	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?	@	A
B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
X	Y	Z	[\]	^	_	`	a	b	c	d	e	f	g	h	i	j	k	l	m
n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~					

by Latin Unicode Block



U+1D400 MATHEMATICAL BOLD CAPITAL A

အင်္ဂလိပ် အက္ခရာမှာ နံပါတ်တွေအနေနဲ့ U+0030 = 0 ဖြစ်ပြီး U+0031 = 1 ကနေ U+0039 = 9 ဖြစ်ပါတယ်။ ဆိုလိုချင်တာ 0030 - Zero ကနေ 0039 - Nine ဟာ 0041 A ရဲ့အရှေ့မှာ ရှိပါတယ်။ အဲတာကြောင့် ဂဏန်းတွေဟာ အရှေ့မှာရှိနေတာလား။ သို့သော် မြန်မာယူဒီကုဒ်ပွိုင့်တွေမှာတော့ ၀ ကနေ ၉ ဟာ ကကြီး ခခွေးတွေတင် သာမက တိုင်းရင်းသားစာတွေရဲ့ နောက်မှာ ရောက်နေတယ်။ က ကြီးဟာ U+1000 ဖြစ်ပေမယ့် သံညကတော့ U+1040 မှာစတယ်။ အဲတာကို ဘယ်လိုလုပ်ပြီး ကကြီးရဲ့ အရှေ့ရောက်နိုင်မှာလဲ။ ရှမ်းရဲ့ ၁ ၂ ၃ ၄ ဆို မြန်မာထပ်တောင် နောက်ရောက်သေးတယ်။ U+1090 မှာ ရှမ်း သံည စတယ်။ Sorting စီရင် အဆင်ပြေပါတော့မလား။ ကကြီးက U+1000 ဆိုပေမယ့် ရှမ်းတို့ ခမ်းတီးတို့အတွက် လိုလို့ နောက်မှ ထပ်ဖြည့်ထားတဲ့ Code တွေဆိုရင်တော့ U+AA60 ကနေ Extend A ဆိုပြီး အဝေးကြီးမှာ ထပ်ဖြည့်ထားတာပါ။ U+A9E0 ကနေမှ Extend B ဆိုပြီး ထပ်ဖြည့်ထားတာပါ။ ဟိုတစ်ပုံ ဒီတစ်ပုံကြီးဆို Code Point နဲ့ စီလို့ ရပါတော့မလား။

တိုင်းရင်းသား သင်ပုန်းကြီး

Unicode Point နံပါတ်အရ နောက်ကိုရောက်နေသော်လည်း Sort လုပ်တဲ့အခါ မြန်မာသင်ပုန်းကြီးအတိုင်း ဘယ်လို စီရပါသလဲ။ code point နဲ့ဆို မြန်မာ ခေါ်ဗမာပဲ အဆင်ပြေမယ်။ အဲတာတောင် ဂဏန်းတွေ အဆင်ပြေမှာ မဟုတ်သေးဘူး။ တိုင်းရင်းသားတွေကလဲ မြန်မာသင်ပုန်းကြီးအတိုင်း လိုက်စီပြီး ထားရမယ်လို့ ဖြစ်သွားမှာပေါ့။

က	ခ	ဂ	ဃ	င	စ	ဆ	ဇ	ဈ	ည	ဋ	ဌ	ဍ	ဎ	တ	ထ	ဒ	ဓ	န	ပ
ဖ	ဗ	ဘ	မ	ယ	ရ	လ	ဝ	သ	ဟ	ဠ	အ	ဏ	ဏ	ဉ	ဦ	ဧ	ဉ	ဩ	ိ
တ	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို	ို
၂	၃	၄	၅	၆	၇	၈	၉	၁	၂	၃	၄	၅	၆	၇	၈	၉	၁	၂	၃
၄	၅	၆	၇	၈	၉	၁	၂	၃	၄	၅	၆	၇	၈	၉	၁	၂	၃	၄	၅
က	ဟ	ဃ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ
င	င	င	င	င	င	င	င	င	င	င	င	င	င	င	င	င	င	င	င
ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ
ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ	ာ

Myanmar Unicode Point (by Script)

က ([k])	ခ ([kʰ])	ဂ ([g])	ဃ ([g])	င ([ŋ])	
ဇ ([ts])	ဆ ([sʰ])	ဆ ([z])	ဇ ([z])	ည ([z])	ဏ ([n])
တ ([t])	ထ ([tʰ])	တ ([d])	ထ ([d])	န ([n])	
တ ([t])	ထ ([tʰ])	တ ([d])	ထ ([d])	န ([n])	
ပ ([p])	ဖ ([pʰ])	ဖ ([f])	ပ ([b])	ဖ ([b])	မ ([m])
ယ ([j])	ရ ([r])	လ ([l])	ဝ ([w/v(p)])	ဆ ([θ] or [ð])	
	ဟ ([h])	လ ([l])	က ([ʔ])		

ရှမ်း သင်ပုန်းကြီး

အဲတာဆို ဘယ်လိုလုပ်ပြီး တိုင်းရင်းသားတွေက Sorting စီရမ္မာလဲ။ ဟုတ်ကဲ့ Unicode ဟာ Code Point တွေနဲ့တင် အလုပ်လုပ်နိုင်မှာ မဟုတ်ပါဘူး။ ICU ဆိုတာတွေ ရှိပါသေးတယ်။ (intensive care unit) အထူးကြပ်မတ်ကုသ ဆောင်လို့ထင် မြင်ချင်မြင်ပါလိမ့်ဦးမယ် (စတာပါ)။ တကယ်တော့ တခြားသော ICU တွေထဲက တစ်ခုခုကို သုံးပြီးမှ အဲဒီ Sorting တွေ Localization တွေ လုပ်လို့ရမှာပါ။ ICU ဆိုတာကတော့ International Components for Unicode ပါ။ သူက CLDR မှာထည့်ထားတဲ့ ဒေတာတွေကို ယူပြီး သုံးတာပါ။ သူ မပါပဲ Unicode ဆိုတာရဲ့ အနှစ်သာရ ဘာမှ မရနိုင်ပါဘူး။ CLDR ရဲ့ လုပ်ဆောင်ချက်တွေကို ဖြည့်သွင်းထားမှသာ Sort လုပ်ခြင်း Localization ပြုလုပ်ခြင်း System တွင် လုံလောက် ပြည့်ဝစွာ အသုံးပြုခြင်းတို့ကို ပြုလုပ်နိုင်မှာဖြစ်ပါတယ်။

CLDR (Unicode Common Locale Data Repository)

ဒီ CLDR က အချက်အလက်တွေကိုမှ သက်ဆိုင်ရာ ICU တွေကနေ ယူသုံးတာပါ။ ဒီအထဲမှာ ပါဝင်တဲ့ အချက်အလက်တွေကို တတ်နိုင်သမျှ ချရေးကြည့်ပါမယ်။ နားမလည်နိုင်ခဲ့ရင် ကျွန်တော့်ရဲ့ အရေးအသား ညွှန်ကြားမှု မှုကြောင့် ဆော့ရီးပါ။ ကိုယ့်ဘာသာ ဖတ်တာလောက်တော့ ကောင်းမှာ မဟုတ်ပါဘူး။

- Locale ဒေသဆိုင်ရာသုံး အချက်အလက်ပုံစံ
 - နေ့ရက်၊ နာရီ၊ ဒေသစံတော်ချိန်၊ နေ့စွဲ၊ ပြက္ခဒိန်
 - ဂဏန်း နံပါတ်များ / ဂဏန်း အသုံးအနှုန်းများ ငွေကြေးတန်ဖိုး
- Translation အတွက် အချက်အလက်
 - ဘာသာဗေဒ၊ နိုင်ငံ၊ ရေးသားမှုပုံစံ

- ငွေကြေး အခေါ်အဝေါ်အမှတ်အသား သင်္ကေတ နှင့် plurals တွက်နည်းစနစ်
 - ဥပမာ US DOLLAR / CENT - US\$ / Myanmar Kyat / PYA - MMK)
- ရက်သတ္တပတ် အခေါ်အဝေါ်များ၊ လအခေါ်အဝေါ်များ၊ ခေတ်အလိုက်အခေါ်အဝေါ် နေ့ရက် (အတို၊ အပြည့်အစုံ) (ပုံနမူနာများ ထည့်ပေးထားသည်။)
- Time Zone (UTC+06:30/MMT) (အချို့နိုင်ငံများတွင် မြို့အလိုက်ရှိနိုင်သည်။ မြန်မာပြည်တွင် +6:30 တစ်ခုတည်းသာ ရှိပါသည်။)
- အတွက်အချက် အခေါ်အဝေါ်များ အတို / အပြည့်အစုံ နှင့် plurals ပုံစံ / past နှင့် future တူညီမှု မရှိပါက အသုံးပြုသည့်ပုံစံ
 - ဥပမာ ၁ ကြိမ် / ၁၀ ကြိမ် / အကြိမ် ၂၀ (၂၀ ကြိမ်မဟုတ်ပါ။) / အကြိမ် ၃၀ / အကြိမ် ၁၀၀
 - ၁ ယောက် / ၁၀ ယောက် / အယောက် ၂၀ / အယောက် ၁၀၀
 - ၁ ပါး / ၂ ပါး / အပါး ၂၀ / အပါး ၁၀၀
- ဘာသာစကား နှင့် ရေးသားမှုပုံစံ အချက်အလက်
 - စာလုံးပုံစံနှင့် အရန်စာလုံးပုံစံများနှင့် အသုံးပြုရာ ဘာသာစကား အချက်အလက်
 - ဘာသာစကားများအတွက် အခေါ်အဝေါ်များအတွက် တွက်ချက်ပုံနည်းစနစ်
 - Gender အခေါ်အဝေါ်များ
 - Searching (ရှာဖွေရန်) နှင့် Sorting (စီရန်) အတွက် အချက်အလက်များ
 - ရေးသားရန် အတွက် အချက်အလက်များ ရေးနည်းစနစ် (ဆွဲချက် / L to R / R to L)
 - ဂဏန်းများကို စာလုံးပေါင်းခြင်းဆိုင်ရာ ရေးနည်းစနစ် (၀= သို့ည ၁ = တစ် / ပထမ / ဧက)
 - စာလုံးများ စာပုဒ်များ စာပိုဒ်များအလိုက် ဖြတ်တောက်ပေးသော (ဝဏ္ဏဖြတ်) စနစ်ဖြင့် အပိုဒ်လိုက် ဖြတ်ပြီး တကြောင်းဆင်းစေနိုင်ခြင်း ။
- နိုင်ငံ / တိုင်းပြည် အလိုက် အချက်အလက်
 - ဘာသာစကား (ဘာသာစကားနှင့် ရေးသားပုံ) ဒေသလိုက် ဘာသာစကားနှင့် ရုံးသုံး ဘာသာစကား
 - နိုင်ငံအလိုက် ဘာသာစကားနှင့် ဘာသာစကားအလိုက် နိုင်ငံ
 - ဥပမာများဖြစ်ပါသည်။
 - မြန်မာဘာသာစကား သည် Myanmar Script ကိုသုံးပြီး Myanmar Region တွင်သုံးသည်။
my_Mymr_MM
 - Marma (မရမာ) ဘာသာစကားသည် Myanmar Script ကို သုံးပြီး Bangladesh Region တွင်သုံးသည်။
rmz_Mymr_BD
 - ရှမ်းဘာသာစကားကို Myanmar Script ဖြင့် Myanmar Region / Thailand တွင်သုံးသည်။
shn_Mymr_MM shn_Mymr_TH
 - မွန်ဘာသာစကားကို Myanmar Script ဖြင့် Myanmar Region နှင့် Thai Region တွင် သုံးသည်။
mnw_Mymr_MM, mnw_Mymr_TH
 - ခမ်းတီး (Khamti) ဘာသာစကားကို Myanmar Script ဖြင့် India Region တွင် သုံးသည်။
kht_Mymr_IN
 - Kachin ဘာသာစကားကို Latin Script ဖြင့် Myanmar Region တွင် သုံးသည်။ kac_Latn_MM
 - Kayah Li ဘာသာစကားကို Kayah Li Script ဖြင့် Myanmar Region နှင့် Thailand Region တွင် သုံးသည်။ eky_Kali_MM, eky_Kali_TH
 - ပြက္ခဒိန် အတွက် အချက်အလက်များ
 - တပတ်တာ၏ အစရက် (အချို့က Monday ဖြင့်စပြီး အချို့က Sunday ဖြင့် စသည်။)
 - တယ်လီဖုန်း နံပါတ် ပုံစံစနစ် / ဖုန်း ကုဒ်
 - အိမ်လိပ်စာ လမ်းလိပ်စာ ပုံစံစနစ်
- အခြားသော
 - ဘာသာစကား / နိုင်ငံ ISO Code ချိတ်ဆက်မှုများ
 - ကီးဘုတ် ပုံစံ

- Translation Guidelines

- (အပေါ်တွင်ပါသော အချက်အလက်များအပါအဝင် နည်းလမ်းပေါင်း ၃၅ မျိုး ခန့်)

System	First day of week	First week of year contains			Can also be last week of previous year	Used by/in
ISO-8601	Monday	4 January	1st Thursday	4–7 days of year	yes	EU and most of other European countries, most of Asia and Oceania
(Islamic)	Saturday	1 January	1st Friday	1–7 days of year	yes	Much of the Middle East
(North American)	Sunday	1 January	1st Saturday	1–7 days of year	yes	Canada, USA, China, Korea, Japan, Israel, South Africa, most of Latin America

First Day of Weeks.

#	Narrow Names	Short Names	Long Names
0	S	Sun	Sunday
1	M	Mon	Monday
2	T	Tue	Tuesday
3	W	Wed	Wednesday
4	T	Thu	Thursday
5	F	Fri	Friday
6	S	Sat	Saturday

English Belize locale en_BZ

#	Narrow Names	Short Names	Long Names
0	J	Jan	January
1	F	Feb	February
2	M	Mar	March
3	A	Apr	April
4	M	May	May
5	J	Jun	June
6	J	Jul	July
7	A	Aug	August
8	S	Sep	September
9	O	Oct	October
10	N	Nov	November
11	D	Dec	December

English Belize locale. en_BZ

#	Narrow Names	Short Names	Long Names
0	D	dim.	dimanche
1	L	lun.	lundi
2	M	mar.	mardi
3	M	mer.	mercredi
4	J	jeu.	jeudi
5	V	ven.	vendredi
6	S	sam.	samedi

ပြင်သစ် fr_FR

#	Narrow Names	Short Names	Long Names
0	J	janv.	janvier
1	F	févr.	février
2	M	mars	mars
3	A	avr.	avril
4	M	mai	mai
5	J	juin	juin
6	J	jul.	juillet
7	A	août	août
8	S	sept.	septembre
9	O	oct.	octobre
10	N	nov.	novembre
11	D	déc.	décembre

ပြင်သစ် fr_FR

#	Narrow Names	Short Names	Long Names
0	D	dom	domingo
1	S	seg	segunda-feira
2	T	ter	terça-feira
3	Q	qua	quarta-feira
4	Q	qui	quinta-feira
5	S	sex	sexta-feira
6	S	sáb	sábado

ပေါ်တူဂီ - ဘရာဇီး pt_BR

#	Narrow Names	Short Names	Long Names
0	J	jan	janeiro
1	F	fev	fevereiro
2	M	mar	março
3	A	abr	abril
4	M	mai	maio
5	J	jun	junho
6	J	jul	julho
7	A	ago	agosto
8	S	set	setembro
9	O	out	outubro
10	N	nov	novembro
11	D	dez	dezembro

ပေါ်တူဂီ - ဘရာဇီး pt_BR

#	Narrow Names	Short Names	Long Names
0	တ	တနင်္ဂနွေ	တနင်္ဂနွေ
1	တ	တနင်္လာ	တနင်္လာ
2	အ	အင်္ဂါ	အင်္ဂါ
3	ဗ	ဗုဒ္ဓဟူး	ဗုဒ္ဓဟူး
4	က	ကြာသပတေး	ကြာသပတေး
5	သ	သောကြာ	သောကြာ
6	စ	စနေ	စနေ

မြန်မာ / ဗမာ - မြန်မာ my_MM

Formatting

#	Narrow Names	Short Names	Long Names
0	၀	၀န့်	၀န့်န့ဝါရီ
1	၁	၁ဖ	၁ဖဖော်ဝါရီ
2	၂	၂တ်	၂တ်
3	၃	၃ပြီ	၃ပြီ
4	၄	၄မ	၄မ
5	၅	၅န့်	၅န့်
6	၆	၆	၆လိုင်
7	၇	၇	၇ရတ်
8	၈	၈က်	၈က်တင်ဘာ
9	၉	၉ောက်	၉ောက်တိုဘာ
10	၁၀	၁၀	၁၀တင်ဘာ
11	၁၁	၁၁	၁၁တင်ဘာ

မြန်မာ / ဗမာ - မြန်မာ my_MM

hail@unicode:/ \$ cal 2017

2017 January							2017 February							March						
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7					1	2	3	4				1	2	3
8	9	10	11	12	13	14	5	6	7	8	9	10	11	5	6	7	8	9	10	11
15	16	17	18	19	20	21	12	13	14	15	16	17	18	12	13	14	15	16	17	18
22	23	24	25	26	27	28	19	20	21	22	23	24	25	19	20	21	22	23	24	25
29	30	31					26	27	28					26	27	28	29	30	31	
April							May							June						
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa
						1		1	2	3	4	5	6					1	2	3
2	3	4	5	6	7	8	7	8	9	10	11	12	13	4	5	6	7	8	9	10
9	10	11	12	13	14	15	14	15	16	17	18	19	20	11	12	13	14	15	16	17
16	17	18	19	20	21	22	21	22	23	24	25	26	27	18	19	20	21	22	23	24
23	24	25	26	27	28	29	28	29	30	31				25	26	27	28	29	30	
30																				
July							August							September						
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa
						1		1	2	3	4	5							1	2
2	3	4	5	6	7	8	6	7	8	9	10	11	12	3	4	5	6	7	8	9
9	10	11	12	13	14	15	13	14	15	16	17	18	19	10	11	12	13	14	15	16
16	17	18	19	20	21	22	20	21	22	23	24	25	26	17	18	19	20	21	22	23
23	24	25	26	27	28	29	27	28	29	30	31			24	25	26	27	28	29	30
30	31																			
October							November							December						
Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7					1	2	3						1	2
8	9	10	11	12	13	14	5	6	7	8	9	10	11	3	4	5	6	7	8	9
15	16	17	18	19	20	21	12	13	14	15	16	17	18	10	11	12	13	14	15	16
22	23	24	25	26	27	28	19	20	21	22	23	24	25	17	18	19	20	21	22	23
29	30	31					26	27	28	29	30			24	25	26	27	28	29	30
														31						

hail@unicode:/ \$ █

English Locale Calendar 2017

2016																								
Janvier							Février							Mars										
di	lu	ma	me	je	ve	sa	di	lu	ma	me	je	ve	sa	di	lu	ma	me	je	ve	sa				
					1	2		1	2	3	4	5	6				1	2	3	4	5			
3	4	5	6	7	8	9	7	8	9	10	11	12	13	6	7	8	9	10	11	12				
10	11	12	13	14	15	16	14	15	16	17	18	19	20	13	14	15	16	17	18	19				
17	18	19	20	21	22	23	21	22	23	24	25	26	27	20	21	22	23	24	25	26				
24	25	26	27	28	29	30	28	29						27	28	29	30	31						
31																								
Avril							Mai							Juin										
di	lu	ma	me	je	ve	sa	di	lu	ma	me	je	ve	sa	di	lu	ma	me	je	ve	sa				
					1	2		1	2	3	4	5	6	7				1	2	3	4			
3	4	5	6	7	8	9	8	9	10	11	12	13	14	5	6	7	8	9	10	11				
10	11	12	13	14	15	16	15	16	17	18	19	20	21	12	13	14	15	16	17	18				
17	18	19	20	21	22	23	22	23	24	25	26	27	28	19	20	21	22	23	24	25				
24	25	26	27	28	29	30	29	30	31						26	27	28	29	30					
Juillet							Août							Septembre										
di	lu	ma	me	je	ve	sa	di	lu	ma	me	je	ve	sa	di	lu	ma	me	je	ve	sa				
					1	2		1	2	3	4	5	6					1	2	3				
3	4	5	6	7	8	9	7	8	9	10	11	12	13	4	5	6	7	8	9	10				
10	11	12	13	14	15	16	14	15	16	17	18	19	20	11	12	13	14	15	16	17				
17	18	19	20	21	22	23	21	22	23	24	25	26	27	18	19	20	21	22	23	24				
24	25	26	27	28	29	30	28	29	30	31						25	26	27	28	29	30			
31																								
Octobre							Novembre							Décembre										
di	lu	ma	me	je	ve	sa	di	lu	ma	me	je	ve	sa	di	lu	ma	me	je	ve	sa				
						1					1	2	3	4	5				1	2	3			
2	3	4	5	6	7	8	6	7	8	9	10	11	12	4	5	6	7	8	9	10				
9	10	11	12	13	14	15	13	14	15	16	17	18	19	11	12	13	14	15	16	17				
16	17	18	19	20	21	22	20	21	22	23	24	25	26	18	19	20	21	22	23	24				
23	24	25	26	27	28	29	27	28	29	30						25	26	27	28	29	30	31		
30	31																							

fr_FR Locale calendar 2016

```
hail@unicode:/$ date +%d-%A-%B-%Y
23-Saturday-July-2016
hail@unicode:/$
```

en_US locale date

```
hail@unicode:/$ date +%d-%A-%B-%Y
23-samedi-juillet-2016
hail@unicode:/$
```

fr_CA locale date

Number	Mon		
	Numeral	Written	IPA
0	၀	သုည	sunɲaʔ
1	၁	မွဲ	mòa
2	၂	စါ	ba
3	၃	ပိ	pɔeʔ
4	၄	ပန်	pɔn
5	၅	မသုန်	pasɔn
6	၆	တြဲ	karao
7	၇	ထပုံ	həpɔh
8	၈	ခွါ	həcam
9	၉	ဒစိတ်	həcit
10	၁၀	စုံ	cɔh

မွန် - သချံာ် ဂဏန်း

ICU

အသုံးပြုနိုင်သော ICU Module / Plugin (နမူနာ)များ

- ICU4C - C , C++
- ICU4J - Java
- Related: (sample)
 - ICU-DOTNET - C#, DotNet
 - ICU-Lua - Lua
 - PICU - Perl
 - PHP Intl - Php / PHP 6 and + ICUC
 - PyICU - Python
 - ICU4R - Ruby

CLDR

v.1 2003-12-19

Now v.29 2016-03-16

ICU

v.3.2 2004-11-22 (CLDR 1.2)

Now v.57.1 2016-03-23

Unicode

v.1 1991

Now v.9 Beta 2016 / v.8 Release 2015

ဒါဟာ အပြည့်အစုံမဟုတ်သေးပါဘူး။ နမူနာပဲ ရှိပါသေးတယ်။ အဲလို အချက်အလက်တွေ ထည့်ပေးထားတဲ့ ICU / CLDR က အချက်အလက်တွေနဲ့ Sort လုပ်နိုင်မှာဖြစ်ပါတယ်။ Localization Project တွေမှာ အသုံးပြုနိုင်မှာ ဖြစ်ပါ

တယ်။ အဲလိုမဟုတ်ရင် Code Point ရထားတဲ့ စာလုံးပုံစံ တနေရာထက် ဘာမှ မပိုပါဘူး။ လုပ်ယူလို့ မရတာတော့ မဟုတ်ပါဘူး။ အဲလို CLDR ထဲမှာ မထည့်ထားခဲ့ရင် ယူသုံးလို့ရတဲ့ ICU အသင့်မရှိခဲ့ရင် လုပ်ယူရတာ ပင်ပန်းမှာပါပဲ။

Backward Compatible ဆိုတာ အလကား ထည့်ထားတာ မဟုတ်ပါဘူး။ စက်အဟောင်း အသစ်တွေမှာ ဖြစ်ပေါ်နိုင်တဲ့ Unicode Version မတူညီမှုတွေကို အလွယ်ကူဆုံး အထိရောက်ဆုံး တနည်းအနေနဲ့ ဖြေရှင်းပေးထားတာဖြစ်ပါတယ်။ ဥပမာ Model နိမ့် ဖုန်းထဲက Unicode Version အဟောင်းနဲ့ Laptop ထဲက Unicode နောက်ဆုံး Version ဆိုရင်တောင် အသစ်ကနေ အဟောင်းကိုတော့ ဖတ်လို့ရအောင် လုပ်ပေးထားတာဖြစ်ပါတယ်။ အဟောင်းကနေ အသစ်ကိုလဲ update ဖြစ်နေတဲ့ အချက်အလက်တွေတော့ အတိအကျ မရနိုင်ပဲ အနည်းငယ်သာ လွဲချော်နေပေမယ့် ဖတ်လို့ ရမှာဖြစ်ပါတယ်။

အဲဒီတော့ ICU / CLDR နဲ့ Sort / Transliteration / Translation တွေ အသုံးပြုတယ်။ System မှာ ဘာသာစကား အလိုက် ဒေသအလိုက် အချက်အလက်တွေသုံးတယ်။ Sorting အတွက် Code Point လေး ပြောင်းရုံနဲ့ မရဘူးဆိုတာ ရှင်းပြီလို့ ထင်ပါတယ်။

ပုံမှာပါတဲ့အတိုင်း ပထမဗားရှင်းမှာ မြန်မာအက္ခရာတွေကို နေရာချထားလိုက်ပြီးပြီ ဆိုကြပါစို့။ ဟုတ်ပြီ ဒုတိယဗားရှင်းလုပ်ကြပြီ။ ဒုတိယဗားရှင်းလုပ်တဲ့အခါကျမှ ဌ (ဌကြီး) ထည့်ဖို့လိုနေမှန်းသိကြရော။ အဲဒါဆို ဘယ်နေရာမှာထည့်ကြမလဲ။

[၁] Backward Compatibility ရအောင် အရင်ဗားရှင်းကိုလုံးဝ မထိခိုက်စေရပဲ (အ) နောက်မှာပဲထည့်လိုက်မလား။

ဒါမှမဟုတ်

[၂] ဖြစ်သင့်တဲ့ပုံစံအတိုင်း (အ) ကိုနောက်တစ်ကွက်ရွှေ့ပြီး (ဌ)(အ) ဆိုပြီးထည့်လိုက်မလား။

ဒီနေရာမှာ ထည့်စဉ်းစားရမယ့် အချက်လေးတွေရှိပါတယ်။ အဲဒါတွေက -

>> ပထမ ဗားရှင်းဟာ အလုံးစုံပြီးပြည့်စုံတဲ့ ဗားရှင်းမဟုတ်သေးဘူး ဖြည့်စွက်မှုတွေ အများကြီးလုပ်ရဦးမယ်လို့ အားလုံးကလည်း သဘောတူလက်ခံထားတယ်။

>> ဒုတိယဗားရှင်းကို လုပ်တဲ့အချိန်မှာ ပထမဗားရှင်းကို အသုံးပြုသူနည်းသေးတယ်လို့ ယူဆနိုင်ပါတယ်။

>>(အ)(ဌ) လို့ဒုတိယဗားရှင်းမှာ ထည့်လိုက်ရင် ကွန်ပျူတာထဲမှာ မြန်မာစာ အက္ခရာစဉ်တဲ့ ကိစ္စဟာ နည်းနည်းလေးပိုပြီး ရှုပ်ထွေးသွားနိုင်စရာရှိတယ်။

ဘယ်လိုလုပ်ကြမလဲ [၁] လား [၂] လား

U1000

က	ခ	ဂ	ဃ	င	စ	ဆ
ဇ	ဈ	ည	ဋ	ဌ	ဍ	ဎ
ပ	ဖ	ဗ	ဘ	မ	ယ	ရ
လ	ဝ	သ	ဟ	အ		
					ဝ	သ
၂	၃	၄	၅	၆	၇	၈
၉	၊	။	၌	၎	၎	၏

Unicode Chart အမှန်မဟုတ်၊ ဥပမာပြရန် ရှင်းအောင်ပြုပြင်ထား

Facebook တွင် မေးမြန်းထားသော နမူနာ မေးခွန်းတစ်ခု - အောက်က Comment များက ပိုပြီး အံ့အားသင့်ဖွယ် ဖြစ်သည်။

ပြည်ထောင်စု သမ္မတ မြန်မာနိုင်ငံတော် နှင့် မြန်မာစာ မြန်မာယူနီကုဒ်

ပြည်ထောင်စု သမ္မတ မြန်မာနိုင်ငံတော် ဆိုတာ တိုင်းရင်းသား/ဘာသာစကား ၁၃၅ မျိုးနဲ့ ပြည်နယ်နဲ့ တိုင်း ၁၄ ခု အဲဒီကနေမှတဆင့် အဆင့်ဆင့်ပြောရရင် နောက်ဆုံး ကျေးရွာပေါင်း ၆၅၀၀၀ ကျော်အထိ အကုန်လုံးပါဝင်ပါတယ်။ အားလုံးကို ကိုယ်စားပြုတာ ဖြစ်ပါတယ်။

မြန်မာစာ ဆိုရင်တော့ တိုင်းရင်းသားစာနဲ့ မြန်မာစာဆိုပြီး ခွဲပြောနေရပါတယ်။ အခုချိန်မှာ မြန်မာစာဟာ ဗမာစာပဲ ဖြစ်ပြီး အခြားတိုင်းရင်းသားစာတွေကို မဆိုလိုပါဘူး။ တကယ်တမ်း ဗမာစာ ဗမာစကားဟာလည်း မြန်မာနိုင်ငံ တိုင်းရင်းသားဘာသာစကားတွေထဲက တစ်ခုပါပဲ။ မြန်မာသင်ပုန်းကြီးဆိုတဲ့ စာအုပ်ကို ပြန်မြင်ယောင်ကြည့် လိုက်ပါ။ နိုင်ငံတော် ရုံးသုံး ဗမာစာကို မြန်မာစာလို့ ဆိုလိုတာဖြစ်ပါတယ်။ အဲဒီတော့ မြန်မာနိုင်ငံဆိုတာဟာ အကုန်လုံးကို ကိုယ်စားပြုပါတယ်။ မြန်မာစာဟာ ဗမာစာကိုပဲ ကိုယ်စားပြုပါတယ်။

မြန်မာယူနီကုဒ် ကကော တိုင်းရင်းသား ဘာသာစကား ၁၃၅ မျိုးလုံးကို ကိုယ်စားပြုမှာလား ? အပေါ်မှာ ပြောခဲ့သလိုပဲ ကျွန်တော်တို့ဟာ Burma , Burmese, Myanmar ဆိုတာတွေနဲ့တင် လိပ်ပတ်မလည်နိုင်ပါ ဘူး။ မြန်မာယူနီကုဒ်ဆိုရင် မြန်မာနိုင်ငံကို ကိုယ်စားပြုတာလား မြန်မာစာကို ကိုယ်စားပြုတာလား ခွဲခြားပြောမှ ဖြစ် ပါလိမ့်မယ်။

- မြန်မာနိုင်ငံကို ကိုယ်စားပြုရင်တော့ ဘာသာစကား ၁၃၅ မျိုးလုံး ပါရမှာပေါ့။
- မြန်မာစာ မြန်မာစကားကို ကိုယ်စားပြုတယ်ဆိုရင်တော့ ရုံးသုံး မြန်မာစာက အဓိက ဖြစ်နေမှာပေါ့။

တကယ်တော့ ၂ ခုလုံး မဟုတ်ပါဘူး။

**Myanmar Unicode Blocks , Myanmar Script

**မြန်မာယူနီကုဒ်ဟာ Myanmar Script (မြန်မာ အက္ခရာ / စာပေ ရေးသားပုံ) / Myanmar Unicode Blocks (Glyph အစု) ကို ကိုယ်စားပြုပါတယ်။ အပေါ်မှာ ပြောသလိုပဲ တချို့ဟာ Myanmar Script အောက်မှာ ရှိပါတယ်။ တချို့ ကတော့ Latin Script အောက်မှာ ရှိပါတယ်။ တချိန်က မြန်မာကိုယ်တိုင် India Script အောက်မှာ ရှိခဲ့ဖူးပါတယ်။



Wikipedia တွင် ဖော်ပြထားသော Myanmar/Burmese Script - (ကျိုက်ထီးရိုး)

မြန်မာအက္ခရာ (ခ) ဗမာ အက္ခရာတွေ ပုံစံတွေရဲ့ အောက်မှာ ပုံစံတူတဲ့ မြန်မာ (ခ) ဗမာ၊ ရှမ်း၊ ကရင်၊ ပအိုဝ်၊ ပလောင် အစရှိတာတွေကို တူတာတွေကို ယူ လိုတာတွေကို ထပ်ဖြည့်ပြီး Myanmar Unicode Block မှာ အသုံးပြုနိုင်အောင် ပြုလုပ်ထားတာဖြစ်ပါတယ်။ ဥပမာ ရှမ်း Code Point တွေဟာ Myanmar Unicode Block ထဲမှာ ဖြစ်ပြီး Script အနေနဲ့ Myanmar Script အောက်မှာပါ။ Kachin ဟာ Latin Script အောက်မှာ ရှိပြီး ကိုးကန့်ဟာ China Script တွေ သုံးပါတယ်။ လောလောဆယ် မြန်မာယူနီကုဒ် Blocks အောက်မှာ Myanmar Script အောက်မှာ ဘာသာစကား အမျိုး ၂၀ အထက် သုံးထားတယ်လို့ ဆိုပါတယ်။ Korea နဲ့ Japan မှာလဲ တရပ်စာလုံးတွေကို အသုံးပြုပြီး Korea လို Hanji / Japan လို Kanji လို့ ခေါ်ဆိုကြပါတယ်။

အဲဒီတော့ ကချင်သာမက ဂျိန်းဖော(Jingpho) လေရှီ(Lashi)အစရှိတဲ့ မျိုးနွယ်တွေရဲ့ Unicode ဟာလဲ Latin Script ထဲမှာပါပါတယ်။ ရိုးရိုး QWERTY Keyboard (ပုံမှန် English Keyboard) ပဲ သုံးပြီး ရိုက်ကြပါတယ်။ အဲဒီအတွက် မြန်မာယူနီကုဒ်ထဲမှာ မပါပါဘူး။ မြန်မာနိုင်ငံမှ ယူနီကုဒ် အသုံးပြုသော တိုင်းရင်းသား လို့ပြောရင်တော့ ရမယ် ထင် ပါတယ်။

အပေါ်က ICU / CLDR အချက်အလက်တွေအရ သုံးတယ်ဆိုတာရှင်းပြီးသားဖြစ်လို့ ဥပမာအားဖြင့် မြန်မာ (ခ) ဗမာ “သ”၊ ရှမ်း “သ”၊ မွန် “သ”၊ ကရင် “သ” စတာတွေရဲ့ တန်ဖိုးတွေဟာ Code Point ကလွဲလို့ Sorting / Translation / Date / Day တွေမှာက အစ ထပ်တူမကျတာကို ထပ်မရှင်းတော့ပါဘူး။

****Code Point, Blocks, Script, Language, Country**

****လောလောဆယ်ရှိနေတဲ့ Myanmar Blocks ထဲက Myanmar Script ထဲက Language တွေ (တချို့ ဘာသာစကား တွေဆိုရင် မြန်မာနိုင်ငံထဲမှာ အသုံးနည်းပြီး တခြားနိုင်ငံမှာပဲ အသုံးများတဲ့ ဘာသာစကားတွေလဲ ရှိပါတယ်။) တိုင်းရင်းသားစာတွေအတွက် CLDR ဘယ်လောက် အထိစုံပြီလဲဆိုတာလဲ ကျွန်တော်မသိပါဘူး။ မြန်မာယူနီကုဒ်ထဲ**

မှာ မပါပေမယ့် မြန်မာနိုင်ငံ တိုင်းရင်းသားစာရင်းထဲမှာ ပါနေတဲ့ Kachin / Jingpho / Lashi / Kokang လိုမျိုး ဘာသာစကားတွေရဲ့ (Script တွေကတော့ လုပ်ပေးစရာ မလိုတော့ပါဘူး။) အချက်အလက်တွေကို လုပ်ပေးမှာလား မလုပ်ပေးဘူးလားဆိုတာတော့ ကျွန်တော်လဲ မသိပါဘူး။

နိုင်ငံအများစုမှာ နေကြတဲ့ လူမျိုး ဘာသာစကားတွေအတွက်ဆိုရင် တခြားနိုင်ငံတွေက လုပ်ပေးထားကြတာတွေ ရှိပြီးသားဖြစ်ပါတယ်။ တခြားနိုင်ငံက လုပ်ပေးရင်လဲ မြန်ပါတယ်။ နည်းပညာ ကွာလို့မဟုတ်ပါဘူး။ နိုင်ငံကွာလို့ပဲ ဖြစ်ပါတယ်။

နိုင်ငံတော် အဆင့်အနေနဲ့ လုပ်တယ်ဆိုရင်တော့ တိုင်းရင်းသားတိုင်းအတွက် ကြိုးစားပေးရမှာပါ။ ဘယ်လို ရည်မှန်းချက်ချထားလဲဆိုတာလဲ ကျွန်တော် လိုက်ရှာ မဖတ်ဘူး မမေးဘူးတော့ မသိပါဘူး။ ထို့အတူပဲ ဘာတွေ ဆက်လုပ်မယ်လဲ ကျွန်တော် မသိပါဘူး။

ကျွန်တော်သိတာကတော့ တိုင်းရင်းသားအတွက် အနာဂတ်မှာ သုံးနိုင်မယ်ဆိုတဲ့ Sorting / Localization / Translation တွေအတွက် CLDR တွေကို မြန်စေချင်ပါတယ်။ တကယ်လို့ လုပ်နေကြသူတွေရှိရင်လဲ ကိုယ့်ဘာသာ လုပ်လိုက်မယ် တွေးထားတာ ကိုယ့်ဘာသာလုပ်နေတာထက် တိုင်းရင်းသားတွေနဲ့ တိုင်းရင်းသားစာပေ လေ့လာ လိုက်စားသူ သုတေသနသမား တွေထံကနေ

- ဘာတွေ လိုအပ်တယ်။
- ဘယ်လိုပုံစံနဲ့ လိုအပ်တယ်။
- ဘာကြောင့် လိုအပ်တယ်။
- ဘာကြောင့် အတည်ပြုဖို့ လိုအပ်တယ်။
- ဘာကြောင့် အရေးကြီးတယ်။
- အချိန်ဘယ်လောက် အတွင်း တင်ပြမယ်။
- အချက်အလက်ရပြီး အချိန်ဘယ်လောက်အတွင်း
 - အများပြည်သူ ကြည့်ရှုလေ့လာနိုင်မယ်
 - အထောက်အထားနဲ့ ကန့်ကွက်နိုင်တယ် ပြုပြင်နိုင်တယ် ဆိုတာကို

သေချာ ဖိတ်ခေါ်တာ စာပို့တာ ဆွေးနွေးတာလုပ်ပြီး အချက်အလက် တောင်းခံသင့်ပါကြောင်း ထင်မြင်မိပါတယ်။ အဲတာတွေလည်း လုပ်ထားလား မလုပ်ထားလား မသိတဲ့အတွက် လုပ်ထားလက်စ Project တွေလဲ ဖြစ်ကောင်းဖြစ် နိုင်ပါတယ်။

Open Source Community

ကျွန်တော် သိသလောက်တော့ ဒီမှာ Private Sector ကနေ Font ရေးသားကြသူတွေ ရှိပါတယ်။ အစိုးရအနေနဲ့ (NLP ကနေ) ရေးသား ထုတ်ဝေပေးတာရှိပါတယ်။ နောက် နိုင်ငံရပ်ခြားကနေ မြန်မာစာရင်း ရေးသားပေးနေတာတွေ ရှိပါ တယ်။ အချို့က အခမဲ့ ဖြန့်ဝေပါတယ်။ အချို့က ရောင်းပါတယ်။ (ဒီနေရာမှာ ရောင်းစားတယ်လို့ အပြစ်မြင်ခြင်း မဟုတ်ပါ။ Software Developer များနည်းတူ Font Developer များ Designer များရဲ့ မူပိုင်ခွင့်နဲ့ ဉာဏ် မူပိုင်ခွင့်ကို လေးစားပါတယ်။) Open ပေးထားတာတွေအပေါ် ကူညီတဲ့သူများလာရင် ပိုများ စိတ်ဝင်စားမလား လေ့လာကြမ လား တွေးမိတာကို ပြောချင်တာပါ။

အဲဒီနေရာမှာ တချို့တွေဟာ အရမ်းကို ဒီ Unicode နဲ့ Font အပေါ်မှာ လေ့လာမှု သက်တမ်းကြာရှည်တဲ့ လူတွေဖြစ်ပါ တယ်။ သူတို့အီက အခက်အခဲကို ဘယ်လို ဖြတ်ကျော်ခဲ့တယ်။ ဘယ်လို ရည်ရွယ်ချက်နဲ့ ဒီလိုတွေ လုပ်ပေးခဲ့ကြ တယ်ဆိုတာကို ကောင်းတာလေးတွေ ရွေးယူနိုင်ပါတယ်။ သူတို့ရဲ့ Source တွေ Software တွေကို တချို့တွေက Open Source အနေနဲ့ တင်ပေးထားကြပါတယ်။ ဥပမာ OFL နဲ့သော်ငှား GNU/GPL သို့မဟုတ် MIT သို့မဟုတ်

Apache လိုင်စင်နဲ့လည်း ဖြစ်နိုင်ပါတယ်။ အဲတာတွေကို Repository List တွေ စုပြီး တင်ထားပေးတဲ့နေရာ ရှိသင့်တယ်လို့ ထင်ပါတယ်။ အဲလို စုထားတာ ရှိမရှိတော့ ကျွန်တော်မသိပါဘူး။ NLP က Github လား Google Code လားတော့ မသိ မြင်ဖူးလိုက်ပါတယ်။

၁၉၉၉ ခုနှစ်မှ ၂၀၀၂ ခုနှစ်အထိ မြန်မာစာစနစ်အတွက် SIL ထုတ် Padauk ဖောင့်သာရှိ။

၂၀၀၂ ခုနှစ် ဒီဇင်ဘာလတွင် မြစေတီ ထွက်ပေါ်လာ

၂၀၀၃ ခုနှစ် ဇန်နဝါရီလ etrademyanmar.com တွင် မြစေတီစသုံး

၂၀၀၃ ခုနှစ် အောက်တိုဘာ မှာ Myanmar NLP ကို စတင်ဖွဲ့စည်း

၂၀၀၅ ခုနှစ် အောက်တိုဘာ မှာ Myanmar1 (စမ်းသပ်) ထွက်ပေါ်

၂၀၀၆ ခုနှစ် ဇူလိုင်လမှာ Myanmar2 (စမ်းသပ်) ထွက်ပေါ်(ပင့်ရစ်ဆွဲထိုးတွေ ပါလာပါတယ်။)

၂၀၀၇ ခုနှစ် ဒီဇင်ဘာလမှာ Myanmar3 ထွက်။ (မြန်မာအတွက် Dotted Circle ထည့်ပေးထားပါတယ်။)

၂၀၁၁ ခုနှစ် ဇူလိုင်လမှာ Guide , Keyboard (၂၀၁၁ ခုနှစ် ဇွန်လတွင် ထွက်ရှိသော မြန်စံ စနစ်), Font တို့ကို အစိုးရဌာနတွေကို ဖြန့်ချိ

အခုအချိန်မှာ

Myanmar NLP စတင်ဖွဲ့စည်းခဲ့ချိန်မှ စတင်ရေတွက်သော် ၁၃ နှစ်

Myanmar NLP မပေါ်ပေါက်မီ MITSC စတင်ခဲ့သည့် ၁၉၉၈ မှ စတင်ရေတွက်သော် ၁၈ နှစ်

Myanmar Unicode ကို Michael Everson စတင် အဆိုတင်သွင်းသည့်

၁၉၉၆ မှ စတင်ရေတွက်သော် အနှစ် ၂၀ ပြည့်ခဲ့ပြီဖြစ်ပါတယ်။

အနှစ် ၂၀ အတွင်းမှ အချက်အလက်များကို တနေရာတည်းမှာ Link တွေပဲ ဖြစ်ဖြစ် စုစည်းထားနိုင်မယ်ဆိုရင် လက်ရှိ Unicode ကို အထောက်အပံ့ပေးနိုင်မည့် ဆောင်ရန် ရှောင်ရန်များ၊ နောက်ထပ် ထွက်ပေါ်လာမည့် Unicode နှင့် သက်ဆိုင်ရာ နည်းပညာများအတွက် အထောက်အပံ့ ဖြစ်စေနိုင်မယ်လို့ ယုံကြည်မိပါတယ်။ နောက်ထပ် ဝင်လာမယ့် မျိုးဆက်သစ်တွေက Unicode အတွက် တတတ်တအား ကူညီမယ်ဆိုလဲ အဆင်သင့် ဖြစ်နေပါလိမ့်မယ်။ ဒီလိုမှ မဟုတ်ရင် အကုန်လုံး စမ်းတဝါးဝါးနဲ့ အစအဆုံး လိုက်လေ့လာနေရပါလိမ့်မယ်။ ဒါပေမယ့် ဒါဟာ အသုံးပြုသူတွေအတွက် ပြောတာမဟုတ်ပါဘူး။ တီထွင်ဖန်တီးသူ ရေးသားသူတွေအတွက်ပြောတာပါ။

ယင်း မြန်မာအက္ခရာ စာလုံးများအတွက် ယူနီကုဒ် point များမှာ (1408)16 မှ (14FF)16 ထိ (၁၂၈)ကွက်သာ လျာထားသတ်မှတ်ပေးထားပြီး အယ်ဗာဆန်၏အဆိုပြုချက်၌ (အ၊ အာ၊ ဣ၊ ဤ... စသည်) သရ Vowel များကို ဦးစွာနေရာပေး စီစဉ်ထားသောကြောင့် "အ" အက္ခရာမှာ ဗျည်းအက္ခရာ "က" ၏ ရှေ့ ဦး ရောက်ရှိနေရာယူထားသည်ကို တွေ့နိုင်ပါသည်။ မြန်မာနိုင်ငံ မြန်မာစာအဖွဲ့ဦးစီးဌာနမှ ပြဋ္ဌာန်းထားသော မြန်မာစာလုံးပေါင်းသတ်ပုံကျမ်းပါ ဝလီစဉ်နည်း (Sorting) နှင့် ကိုက်ညီမှု လုံးဝမရှိကြောင်း တွေ့ရသဖြင့် ယင်းစနစ် နှင့်အပြိုင် အဆိုပြုချက် (proposal) တစ်ခုကို ISO/IEC JTC1/S C2/WG2 အစည်းအဝေးအစီ မီခင်မြန်မာနိုင်ငံမှ တင်ပြနိုင်စေရန် မြန်မာနိုင်ငံကွန်ပျူတာပညာပွဲပြီး ရေးကောင်စီက ` ` မြန်မာစာသုံး ကွန်ပျူတာစနစ် ဖြစ်မြောက်ရေး ကော်မတီ Myanmar IT Standardization Committee (MITSC) ကို ပညာရှင် (၉) ဦးဖြင့် ဖွဲ့စည်းပေးခဲ့ပါသည်။ MITSC အဖွဲ့ ဝင်များမှာ ဒေါက်တာကျော်သိန်း၊ ဦးစံလွင်၊ ဦးထွန်းတင်၊ ဦးသော်ကောင်း၊ ဦးခင်မောင်လွင်၊ ဦးသောင်းတင်၊ ဦးသိန်းထွန်း၊ ဒေါက်တာအောင်မော်နှင့် ဦးမြိုးမောင်မောင်တို့ ဖြစ်ကြပါသည်။ ဖြစ်မြောက်ရေး ကော်မတီသည် အပြိုင်အဆိုပြုချက်တစ်ခု တင်ပြရန်အတွက် မြန်မာနိုင်ငံ မြန်မာစာအဖွဲ့ဦးစီးဌာန၏ မြန်မာစာလုံးပေါင်းသတ်ပုံကျမ်းကို ကိုးကားလျက် မြန်မာစာ အဖွဲ့ဦးစီးဌာန၏ ထောက်ခံချက်နှင့်တကွ အပြိုင်အဆိုပြုချက်တစ်ခုကို ဂျပန်နိုင်ငံ မှ ISO/IEC JTC1/S C2/WG2 အဖွဲ့ဝင် Mr. Takayuki အကူအညီဖြင့် ISO သို့ ပေးပို့တင်ပြခဲ့ရာ ၁၉၉၈ ခုနှစ် စက်တင်ဘာလ (၂၁) မှ (၂၉) ရက်အထိ လန်ဒန်မြို့၌ ကျင်းပပြုလုပ်သော ISO/IEC JTC1/S C2/WG2 တွင် လေ့လာသူအဖြစ်တက် ရောက်ခွင့်ရရှိခဲ့ပါသည်။ ဖြစ်မြောက်ရေး ကော်မတီဥက္ကဋ္ဌ ဒေါက်တာကျော်သိန်း ဦးဆောင်လျက် ဦးသောင်းတင်၊ ဦးသိန်းထွန်း၊ ဦးခင်မောင်လွင်၊ ဒေါက်တာ အောင်မော် ပညာရှင်(၅)ဦးသည် ယင်းအစည်းအဝေးသို့ တက်ရောက်ခဲ့ကြပါသည်။ ISO အဖွဲ့ဝင်မဟုတ်၍ ပြောပိုင်ခွင့်၊ ဆွေးနွေးခွင့်ကို မရရှိခဲ့သော်လည်း နောက်ပိုင်းတွင် ပထမအဆိုရှင် အယ်ဗာဆန်အပါအဝင် ယူနီကုဒ် မှ ပုဂ္ဂိုလ်များ၊ WG2 မှ ပညာရှင်များ၏ ထောက်ခံတင်ပြချက်အရ MITSC တင်သွင်းသော အဆိုပြု (ယပင့်၊ ရရစီ၊ ဝဆွဲ၊ ဟထိုး၊ မောက်ချနှင့် အသတ်ရှေ့ထိုးတို့ အား ပြင်ဆင်ချက်ဖြင့်) ကို လက်ခံခဲ့ကာ နောက် (၆) လ တစ်ကြိမ် အစည်းအဝေးများ၌ Amendment 26 ဖြင့် ပြင်ဆင်ခွင့် ရရှိခဲ့ပြီး နောက် (၆) လ တစ်ကြိမ် အစည်းအဝေးများဖြင့် အတည်ပြုချက်ရယူရန်ကိုပါ ဆုံးဖြတ်ခဲ့ပါသည်။

မြန်မာအက္ခရာ စာလုံးများအတွက် သတ်မှတ်နေရာကို လည်း U+1000 မှ U+109F အထိ (၁၆၀) ကွက်ထိ တိုးချဲ့ကာ Table 240-Row 10: Myanmar ဟူ၍ သတ်မှတ်နိုင်ခဲ့ပြီး၊ မြန်မာစာကို Burmese အစား Myanmar ဟု ISO မော်ပြုချက်များ ၌ သုံးစွဲရန်ကိုပါ သဘောတူခဲ့ပါသည်။

၁၉၉၉ ခုနှစ် မတ်လတွင် ဂျပန်နိုင်ငံ ဖူကူအိုကာမြို့၌ ကျင်းပခဲ့သော ISO/WG2 အစည်းအဝေးကို မြန်မာနိုင်ငံ MITSC မှ ကိုယ်စားလှယ်များတက်ရောက် ခဲ့ရာ ယင်းအဆိုကို ပထမအဆင့်အပြီးသတ် Final Draft Admendment (FDAM) အဖြစ် အတည်ပြုနိုင်ခဲ့ပါသည်။

MITSC ၏ ကြိုးပမ်းချက်များ

ဥပမာများနှင့် စကားများခြင်း

စာရေးဆရာ တစ်ယောက်ယောက်ရေးထားတဲ့ ဇာတ်လမ်းတိုလေး မှတ်မိသလောက် ပြန်ရေးပြပါမယ်။
အိမ်နီးနားခြင်း ဒေါ်ကျင်ဆီကို အမှာစကားလေး ပြောပေးဖို့ ဦးကြည်ကနေ ဒေါ်နီတို့ သားအမိကို အကူအညီတောင်းခဲ့ပါတယ်။ ဒေါ်ကျင်နဲ့တွေ့တော့ ဒေါ်နီကနေ “မနောက နေ့လည်လောက်က ဦးကြည်က မှာသွားတယ်” လို့စပြောပါ

တယ်။ ဒီမှာ ဒေါ်နီသမီး မိညိုက “အမေရယ် မနော့က နေ့လည် မဟုတ်ပါဘူး။ မနက်ပိုင်းကပါ။ မှားနေတယ်” လို့ ဝင်ထောက်တယ်။ ဒေါ်နီကလဲ “နေ့လည်ကပါဟဲ့ ဘုရားဆွမ်းတော် စွန့်ပြီးမှ ရောက်လာတာ ငါမှတ်မိတယ်။” ဆိုပြီး ပြန်သက်သေထူတယ်။ မိညိုကလဲ ခေသူမဟုတ် “မနက်တောင် ဆွမ်းမကျက်သေးလို့ ဘုန်းကြီးတောင် ဆွမ်းမလောင်းရသေးဘူး။ မနက်ပိုင်းကြီး အမေက လူကြီးဖြစ်ပြီး ငြင်းတယ်။” လို့ ခံပက်တယ်။ အမေ လုပ်သူကလဲ “ဟဲ့ အရိုင်းအစိုင်းမ လူကြီးကို ဒါမျိုး ပြောရသလား သွားစမ်း ဝင်မပြောနဲ့ နောက်ဖေးမှာ ပန်းကန်သွားဆေးချေ” ဆိုပြီး နီးတဲ့ကွမ်းအစ်အဖုံးနဲ့ ကောက်ထုပါလေရော။ သမီးဖြစ်သူကလဲ သူပြောတာ မှန်ရဲ့သားနဲ့ ကလေးဆိုပြီး အနိုင်ကျင့်တယ်လူကြီးဖြစ်ပြီး လိမ်တယ်ဆိုပြီး ဖြစ်တောက် ဖြစ်တောက်နဲ့ မကျေနပ်တော့ဘူး။

အဲဒီမှာ တကယ့်အချက်အလက်က မနက်တွေ ညနေတွေမဟုတ်ပါဘူး။ ဦးကြည်က ဒေါ်ကျင်ကို ဘာပြောခဲ့သလဲဆိုတာပဲဖြစ်ပါတယ်။ ဒါပေမယ့် ကျွန်တော် မြင်တာတော့ ဥပမာတွေထဲမှာပါတဲ့ မနက်နဲ့ နေ့လည် ကိုပဲ ငြင်းနေတာတွေပဲ တွေ့ရတယ်။

(ဒီနေရာမှာတော့ စကားပြောနဲ့ မရေးတော့ပါဘူး ခွင့်လွှတ်ပါ။)

“သေချင်တဲ့ကျား တောပြောင်း” နှင့် “တရွာမပြောင်း သူကောင်းမဖြစ်” ဆိုသည့် စကားပုံ နှစ်ခုကိုကြည့်ပါ။ တစ်ခုမှာ နေရာရွေ့မှ ဒုက္ခရောက်ရခြင်းဖြစ်ပြီး အခြားတစ်ခုမှာ နေရာမရွေ့လျှင် ကောင်းမလာနိုင်တော့ဟူသည့် အနက်အဓိပ္ပါယ်ရပါသည်။ စကားပုံများ ဥပမာများသုံးမည်ဆိုပါက နောက်ခံ အခြေအနေပေါ်တွင်လည်း မူတည်နိုင်ပါသည်။ သေချင်တဲ့ကျားသည် အဆင်ပြေနေသော တောမှ ပြောင်းကာ အသတ်ခံရသည်။ ပတ်ဝန်းကျင် အသိုင်းအဝိုင်းမှ (မိမိ အမှားတခုကြောင့်သော်၎င်း၊ အမြင်စွဲကြောင့်သော်၎င်း) အထင်အမြင် သေးနေသည့် နေရာမှ မပြောင်းရွေ့နိုင်လျှင် မည်မျှပင် ကြိုးစားစေကာမူ လူကောင်း လူတော်ဟု အမြင်မခံရနိုင်ကြောင်း ပြောထားခြင်းသာ ဖြစ်ပါသည်။ နောက်ခံ အခြေအနေ မတူညီပါ။

သို့ဖြစ်ရာ ဇော်ဂျီအတွက်ပေးသော ဥပမာများ၊ ယူနီကုဒ်အတွက် ပေးသော ဥပမာများသည် အပြည့်အဝ တိုက်ရိုက် အဆင်ပြေမည် အသုံးတည့်မည်မဟုတ်ပါ။

Unicode အကြောင်း Zawgyi အကြောင်း ဥပမာတွေမှာ ဥပမာတွေကို ငြင်းနေရတာနှင့်ပင် အချိန်တွေ ကုန်နေကြရပါသည်။ တကယ့် အကြောင်းရင်းသို့ရောက်တာ မတွေ့ရပါ။ တခါတလေ အဓိကအချက်စီသို့ ရောက်ပါသော်လည်း သေချာဆွေးနွေးတာ လက်ခံတာ အဖြေရှာတာ မမြင်မိပါ။ အဖွဲ့အစည်းစိတ်ဓါတ်အပြည့်နဲ့ အင်တိုက်အားတိုက် ပြောကြသည်မှာ ဥပမာ၏ အကြောင်းကိုသာ ဖြစ်နေပါသည်။

ထိုကဲ့သို့ ငြင်းခုံနေကြသော အချိန်များတွင် တိုင်းရင်းသား CLDR အတွက် စာတမ်းတစ်ခု ရေးသားလျှင်သော်၎င်း မည်ကဲ့သို့ CLDR တင်ရသည် အတည်ပြုရသည်ကိုသော်၎င်း တိုင်းရင်းသားများကို ရှင်းပြခဲ့ကြမည်စုဆောင်းခဲ့ ရှာဖွေခဲ့ကြမည် ဆိုပါလျှင် ၁၉၉၆ မှ စတင်ခဲ့ပြီး နှစ် ၂၀ သက်တမ်းရှိသော Myanmar Unicode သည် ယခုထက် (လူသုံးများစေခြင်းသက်သက်ထက်) ပိုမို အဆင်ပြေနေလိမ့်မည်ဟု မျှော်လင့်ရပါကြောင်း

ဆရာတွေ ဦးဆောင်လုပ်ရမှာပါ။

ကျွန်တော်တို့က end user တွေပါ။

အခုတော့ လက်ညှောင်းလို့ ဒီလောက်နဲ့ပဲ ရပ်လိုက်ပါတော့မယ်။
စိတ်ပါရင်တော့ နောက်ထပ်ရေးချင် ရေးပါလိမ့်ဦးမယ်။

[Ko Ko Ye](#)

2016, July, 24 (Sunday)

<https://www.facebook.com/notes/682524079057364/>