

21

தமிழ்க் கணினி: சில சிந்தனைகள்

எஸ். ரங்கராஜன் (சுஜாதா)

இதுவரை நான்கு இணைய மாநாடுகள் முடிந்திருந்தும் தமிழ்மொழியை இணையத்துக்கும், மற்ற தகவல் தொழில்நுட்பங்களுக்கு ஏற்பவும் வடிவமைப்பதில் உள்ள குழப்பம் அதிகமாகியிருக்கிறதே தவிர குறையவில்லை. மற்ற இந்திய மொழிகளில் இந்தக் குழப்பம் இந்த அளவுக்கு இல்லை. ஏன் தமிழுக்கு மட்டும் இந்தச் சிக்கல்? காரணம், தமிழினின் தனித் தன்மையும், கலாச்சாரமும், உலகெங்கும் பரவியுள்ள தமிழர்களின் ஆர்வமிகுதியும்தான். 'தமிழன் என்றொரு இனம் உண்டு! தனியே அதற்கொரு குணம் உண்டு' என்று நாமக்கல் கவிஞர் சொன்னதில் 'தனியே' என்பதை அடிக்கோடிடவேண்டும். தமிழ் மொழியைக் காப்பாற்றுவதில் எந்தவிதமான ஒற்றுமையும் இன்றி ஆளாளுக்கு காப்பாற்றி வருவதால் வந்த விளைவு இது. இன்றைய உலகில் உள்ள மொழிகளில் தமிழ்மொழிக்கு மட்டும் தான் ஐந்து encoding standards – குறியீட்டுத் தரங்கள் உள்ளன. டாம், டாப், டிஸ்கி, இஸ்கி, யூனிகோடு (TAM, TAB, TSCII, ISCII, UNICODE) நான்கு தட்டச்சு முறைகளுக்கான விசைப்பலகை, இட ஒதுக்கீடுகள். இதனுடன் inscript என்கிற பெயரில் மற்ற இந்திய மொழிகள் சார்ந்த சற்றே வேறுபட்ட முயற்சியும் உள்ளது. பார்க்க பட்டியல்1

மொழி	தகுதரம்	தட்டச்சு முறை	லினக்ஸ் மென்பொருள்
தமிழ்	TAM, TAB, TSCII, ISCII, UNICODE	Phonetic, Anjal, New typewriter Old typewriter, inscript (முழுமையாகப் பரிசோதிக்கவில்லை)	KDE (TSCII – உள்ளிடப்பட்டு) Unicode இல் compile செய்யப்பட்டுள்ளது Open Office - UNICODE -NCST Mozilla - UNICODE - Thamiza
ஹிந்தி	ISCII, PASCII, UNICODE	Romanized, typewriter, DOE Phonetic	GNOME Desktop - 62.74% Dev-lib - 32.98 Open office - completed - UNICODE - NCST Mozilla - Completed - UNICODE - NCST
வங்காளம்	ISCII, UNICODE	Romanized, typewriter, DOE Phonetic	GNOME Desktop - 61.26% Dev-lib - 51.73%
கன்னடம்	ISCII, UNICODE	Romanized, typewriter, DOE Phonetic	GNOME Desktop - 7.08% Dev-lib - 7.67
மலையாளம்	ISCII, UNICODE	Romanized, typewriter, DOE Phonetic	GNOME Desktop - 39.97% Dev-lib - 53.93
தெலுங்கு	ISCII, UNICODE	Romanized, typewriter, DOE Phonetic	GNOME - Glossary translated 85%
மராத்தி	ISCII, UNICODE	Romanized, typewriter, DOE Phonetic	GNOME - Glossary translated

பட்டியல் 1

மிக சிக்கலான பட எழுத்து மொழியான சீன மொழியில் கூட இந்தக் குழப்பம் இல்லை. சீனர்கள் BIG 5 என்ற எழுத்துரு முறையை பயன்படுத்தி சிறந்த முறையில் தகவல் பரிமாற்றம் செய்துகொள்கிறார்கள். மற்ற இந்திய மொழிகளான ஹிந்தி, மராத்தி, மலையாளம், தெலுங்கு, குஜராத்தி போன்றவை விகவாசமாக மைய அரசு 1988ல் பரிந்துரைத்து, 1991ல் இந்திய தகுதரமாக்கிய (Indian Standard IS 13194 1991) இஸ்கியை, அதன் குறை நிறைகளோடு ஏற்றுக்கொண்டதால் நம் அளவுக்கு குறியீட்டுக் குழப்பம் இல்லை. திராவிட மொழிகளில் தமிழில் எழுத்துக்கள் குறைவு. வர்க்க எழுத்துகள் என்று சொல்லப்படும் கன எழுத்துக்களுக்கான வரி வடிவங்கள் தமிழில் இல்லை. தேவநாகரி போல நான்கு 'க', நான்கு 'ச', நான்கு 'த' வெல்லாம் நமக்கு இல்லை. ஒலிகள் உண்டு. எழுத்துகள் இல்லை. மேலும் நமக்கு கூட்டு எழுத்துக்களே இல்லை. அனைத்தும் தனித்தனியே எழுதுகிறோம். இதனால் எழுத்துக்களை ஒட்ட வைப்பதற்கு ஆணை தரும் தனியான குறியீடு தேவை இல்லை. மற்ற இந்திய மொழிகள் பிரம்மி சார்ந்த எழுத்து வடிவங்களைக் கொண்டிருப்பதால் அவைகளின் பொதுத் தன்மைக்காக பரிந்துரைக்கப்பட்ட இஸ்கியில் வீணாக சில சில இடங்கள் தமிழுக்கு ஒதுக்கப்பட்டன. இதுவே யூனி கோடுக்கும் கொடுக்கப்பட்டது. மேலும் தேவநாகரியின் நெடுங்கணக்கு வரிசை தமிழிலிருந்து வேறு பட்டது. இதனால் இடஒதுக்கீட்டில் தொடர்ச்சி இல்லாமல் sorting போன்ற வரிசைப்படுத்துவதற்கு கூடுதலாக மென்பொருள் நிரல்கள் எழுதி சமாளிக்கவேண்டியிருந்தது.

இதையெல்லாம் நாம் எல்லா இணைய மாநாடுகளில் விவாதித்த பின்னும் ஒருமித்த கருத்துள்ள முடிவுக்கு வரமுடியவில்லை. பட்டியல் ஒன்றில் தமிழ் எழுத்துக்களின் குறியீட்டு விவரம் தரப்பட்டுள்ளது - பட்டியல் 2.

LETTER	TAB	TAM	TSCII	UNICODE
அ	DC	DC	AB	0B85
ஆ	DD	DD	AC	0B86
இ	DE	DE	AD	0B87
ஈ	DF	DF	AE	0B88
உ	E0	E0	AF	0B89
ஊ	E1	E1	B0	0B8A
எ	E2	E2	B1	0B8E
ஏ	E3	E3	B2	0B8F
ஐ	E4	E4	B3	0B90
ஓ	E5	E5	B4	0B92
ஔ	E6	E6	B5	0B93
ஒள			B6	0B94
஠	E7	E7	B7	0B83
஡	E8	E8	B8	0B95
஢	E9	E9	B9	0B99
	EA	EA	BA	0B9A
ண	EB	EB	BB	0B9C

பூ	EC	EC	BC	0B9E
பூ	ED	ED	BD	0B9F
பூ	EE	EE	BE	0BA3
பூ	EF	EF	BF	0BA8
பூ	F0	F0	C0	0BAA
பூ	F1	F1	C1	0BAE
“	F2	F2	C2	0BAF
”	F3	F3	C3	0BB0
‘	F4	F4	C4	0BB2
’	F5	F5	C5	0BB5
லூ	F6	F6	C6	0BB4
ள	F7	F7	C7	0BB3
ழூ	F8	F8	C8	0BB1
ளூ	F9	F9	C9	0BA9
னூ	FB	FB	84	0BB7
அ	FC	FC	83	0B9C
றூ	FA	FA	85	0BB8
ஆ	FD	FD	86	0BB9
ஈ	A3	A3	A1	0BBE
ஈ	A4	A4	A2	0BBF
ஈ	A6	A6	A3	0BC0
TAMIL VOWEL U	AE-C2,1C7	AE-C2	CA,CC-DB,A4,99,9A	0BC1
TAMIL VOWEL UU	C3-DB,18A	C3-CF , D6-DB	A5,DC-EB,9B,9C	0BC2
ஊ	AA	AA	A5	0BC6
ஊ	AB	AB	A6	0BC7
ஊ	AC	AC	A7	0BC8
ஊ ஈ				0BCA
ஊ ஈ				0BCB
ஊ ஈ				0BCC
TAMIL AU LENGTH MARK			AA	0BD7
TAMIL SIGN VERMA				0BCD

TAMIL SIGN ANUSVAR				0B82
TAMIL ONE TO TEN				0BE7-0BF0
TAMIL HUNDRED				0BF1
TAMIL THOUSAND				0BF2
TAMIL DAY				0BF3
TAMIL MONTH				0BF4
TAMIL YEAR				0BF5
TAMIL CREDIT				0BF7
TAMIL DEBIT				0BF5
TAMIL AS ABOVE				0BF8
TAMIL RUPEE				0BF9
TAMIL NUMBER				0BFA
KSHA				
ற்				
கி - னி,டி ,ஸி ரூ ஹி		41-51, AE,52-55		
கீ - ளீ,ணீ - னீ, லீ,ஸீ - ஹீ		57-5A,61-6D		
கீ - ய்,ர் - ள்,ற்,ன்,ஜ் - ஹ்		82-8C,98-9C,9F,A1	EC-FD,89-8C	

நவீன கணிப்பொறி இயலில் இம்மாதிரியான வேறுபாடுகளைச் சமாளிப்பது அப்படி ஒன்றும் பெரிய காரியமல்ல. ஒரு குறியீட்டிலிருந்து மற்றதுக்கு மாற்றிக்கொள்ள சின்னச் சின்ன ஒட்டு நிரல்கள் patches எழுதி சமாளிக்க முடியும். If you can't cure it, endure it. இதனால் உலகில் எத்தனை தமிழர்கள் உள்ளனரோ, அத்தனை தகுதரங்கள் வந்தாலும் விரயமான மென்பொருள்கள் எழுதி சமாளிக்க முடிந்தது. இன்றும் எனக்கு திடீர் என்று புதுசாக கண்டுபிடித்தது போல தமிழ் மொழியைக் காப்பாற்றும் யோசனைகள் அஞ்சலில் வந்துகொண்டிருக்கின்றன. தமிழர்கள் மனதை மாற்றுவதும் மிகக் கடினமாக உள்ளது. அவர்கள் தனித் தன்மையும், பிடிவாதமும், மொழியின் முன்னேற்றத்திற்கு தடையாக உள்ளதை பலர் அறிவ தில்லை. உதாரணங்கள் பல இன்று உள்ளன.

கணிப்பொறியியல் மிக வேகத்தில் முன்னேறிக் கொண்டிருக்கும்போது இந்த இணைய மாநாட்டிலாவது செய்யவேண்டிய சில காரியங்களை கணிப்பொறிஇயல் அறிந்த தமிழ் எழுத்தாளன் என்கிற தகுதியில் முன்வைக்க விரும்புகிறேன். அண்மையில் திறந்த மூலம் Open Source முயற்சிகளை கவனித்ததில், கணிப்பொறியைப் பண்படுத்த ஆங்கில அறிவு தேவையில்லை. தமிழிலேயே வர்த்தகமும், தகவல் பரிமாற்றமும் செய்ய ஏதுவாக கேடிஇ KDE, போன்ற சூழ்நிலையைக் கொண்டுவர இயலும் அருமை யான வாய்ப்பு இருப்பதை உணர்ந் தேன். இதைத் தொடர்ந்து ஆர்வலர்கள் சிலர் சேர்ந்து ஒரு குழு அமைத்துள்ளோம். இதில் சேர அனுமதியோ, பணம் காசோ தேவையில்லை. ஆர்வமும் திறந்த மனதும் போதும். இதைத் தமிழக அரசின் இணைய பல்கலைக்கழகமும் ஆதரித்துள்ளது. இதில் நாங்கள் இது வரை செய்திருப்பதையும், இதில் சாதனைகள் புரிய செய்யவேண்டிய சில விஷயங்களையும் சொல் கிறோம்.

முதலில் சொற்கள். கலைச்சொற்கள். இவைகளை அமைப்பதில் லினக்ஸ் ஆர்வலர்கள் பலர் இஷ்டத் துக்கு மொழிபெயர்த்து அவைகளை இணையத்தில் தந்திருக்கிறார்கள். இதில் ஏற்படும் வேறுபாடுகள் சில திகைக்க வைக்கின்றன. இதற்கு சில உதாரணங்கள் பட்டியல் 3.

English	தற்போதுள்ள KDE	தமிழக அரசு கலை சொல் அகராதி
properties	சொத்துக்கள்	பண்புகள்
Application	நிரல்	பயன்பாடு
Compressed	அழுத்திய	இறுகிய

Script	கிறுவல்	எழுத்து
Mouse settings	எலி அமைப்புகள்	கட்டி அமைப்புகள்
scanner	பதாகை	வருடி
Remote login	சேய்மை	தொலைநிலை
Bibliographic data	உசாத்துணைத் தரவுகள்	நூலடைவு
format	வழுவுட்டி	வடிழிவமை
Romanised	காதலுடன்	ரோமனைஸ்டு

பட்டியல் 3

குறிப்பாக KDEக்கு உண்டான செய்திகளை தமிழ்ப்படுத்துவதில் அதிக சிரத்தை தேவைப்படுகிறது. உதாரணமாக, unlocks the current view, so that it becomes normal again என்பது வார்த்தைக்கு வார்த்தை மொழிபெயர்க்காமல் 'இப்போது தெரிவதை பழையபடி மாற்றுகிறது' என்று சொல்லும்போது தெளிவடைகிறது.

முதலில் எல்லோரும் ஒரே சொல் பட்டியலைப் பயன்படுத்துவதற்காக லினக்ஸ் சூழ்நிலைக்கேற்ப ஒரு பட்டியல் glossary தயாரித்து அந்தச் சொற்களைப் பதிப்பித்திருக்கிறோம். இதை மற்றவர்கள் பயன்படுத்தலாம். இவைகளை அமைப்பதில் எங்கள் அணுகுமுறை - நேரடியாக மொழிபெயர்க்காமல், அந்தஅந்த கட்டளைகளையும், வார்த்தைகளையும் திரையில் இயக்கிப்பார்த்து, 'அவைகள் என்ன செய்கின்றன' என்பதைக் கண்டு, அதற்கேற்ற தமிழ்ச் சொற்களை தமிழில் சுமார் நாற்பதாண்டுகள் கதை, கட்டுரைகளும், அறிவியலும் எழுதியவன் என்ற தகுதியில் கொடுத்திருக்கிறேன். இந்தச் சொற்களைத்தான் பயன்படுத்தவேண்டும் என்கிற கட்டாயமில்லை. இவைகளை முதலில் பயன்படுத்திவிட்டு, இவைகளை விடச் சிறந்த சொற்கள் கிடைத்தால் அவைகளுக்கு இடமளிக்க ஒப்பன் சோர்ஸ் அனுமதிக்கிறது. பழகப் பழகத்தான் சொற்கள் சாகவதம் பெறும். உதாரணமாக 'கோப்பு', 'இணையம்' போன்ற சொற்களை எல்லோரும் பயன்படுத்தத் தொடங்கிவிட்டனர். அன்றாடத் தமிழிலும் இவைகள் வரும்போது மெல்ல மெல்ல ஆங்கிலம் சார்ந்த நிலை மாறி, முழுவதும் தமிழிலேயே எல்லாக் காரியங்களும் செய்ய முடியும். செய்யவேண்டும்.

அடுத்த கட்டமாக OpenOfficeஐ தமிழில் முழுவதும் மொழிபெயர்த்துக்கொண்டிருக்கிறோம். இதில் NCST ஏற்கனவே செய்துள்ள விஷயங்களைப் பயன்படுத்துவதோடு, கலை சொற்களில் தேவையான மாற்றங்கள் செய்து வெளியிட உள்ளோம்.

லினக்ஸ் ஆர்வலர்கள் KDEயில் இதுவரை மொழிமாற்றம் செய்து முடிக்காத கோப்புகளை மொழிபெயர்த்துக்கொண்டிருக்கிறோம்.

தமிழில் தரக்கட்டுப்பாடு வராத நிலையில் லினக்லை தமிழ்ப்படுத்தும் தமிழ் லினக்ஸ் ஆர்வலர்கள் எதிர்காலத்தை மனதில் கொண்டு சில வழிமுறைகள் பின்பற்றலாம்.

Open Source மென்பொருள் முழுவதும் UNICODEடையே பயன்படுத்தவேண்டும். உதாரணமாக - Open Office - NCST குழுவும், Mozilla - தமிழா குழுவும் UNICODE-டையே பயன்படுத்துகிறது. யூனிகோடில் தமிழைப் பொறுத்த வரை சில வேண்டாத விஷயங்கள் உள்ளன. ஆனால், அதை மாற்றும் முயற்சிகள் மிகுந்த தாமதங்களுக்கு உட்பட்டவை. யூனிகோடு பயன்படுத்தி தமிழ் எழுத்துக்கள் அனைத்தையும் உருவமைப்பதற்கான நிரல்கள் சாத்தியமே. தற்போது தமிழின் வித வித எழுத்துருக்களுக்குள் ஒன்றை ஒன்று மாற்றியமைக்க இம்மாதிரி நிரல்களைத்தான் எழுதி வந்திருக்கிறோம். யூனிகோடு நம்முடைய அசிரத்தையினால் இஸ்கி சார்ந்த ஒதுக்கீட்டு முறையை ஏற்றுக் கொண்டுவிட்டது. இதை மாற்றும் முயற்சிகள் ஒரு பக்கம் நிகழ்ந்துகொண்டிருக்கையில் யூனிகோடை டிஸ்கிக்கோ டாப்க்கோ மாற்றி அமைப்ப தால் உண்டான நன்மைகளையும் சிக்கல்களையும் கவனிப்போம். நன்மைகள் தமிழின் எல்லா எழுத்துகளையும் கூடுதல் மென்பொருள் நிரல்கள் தேவையில்லாமல்

நேரடியாகக் காட்ட முடியும். இதற்கான விலை, மற்ற இந்திய மொழிகளுடன் உள்ள ஒற்றுமையை நாம் துறக்கவேண்டும். இந்தியத் துணைக் கண்டத்தில் சில மாநிலங்களில் இரண்டு மூன்று மொழிகள் பயிலும் எல்லை மாகாணங்களில் வேட்பாளர் பட்டியல் போன்றவைகளைத் தயாரிக்கும் வசதிகளை இழக்கவேண்டும். எப்படியும் மாற்றும் முயற்சிகள் யூனிகோடு மார்க்கமாகத்தான் நிகழவேண்டும். அப்போதுதான் மைக்ரோசாப்ட் போன்ற நிறுவனங்கள் மாறச் சம்மதிப்பார்கள். இது அதிக தாமதமானால் மாற்றவே முடியாத சூழ்நிலை ஏற்பட்டுவிடும்.

எல்லோரும் யூனிகோடுக்கு மாறிவிட்டால் நம்மிடையே ஒற்றுமையுடன், மற்ற இந்திய மொழிகளுடனும் ஒருமைப்படுத்த முடியும். எனவே இனி உள்ளிடும் தமிழ் எல்லாம் யூனிகோடில் இருப்பது புத்திசாலித்தனம் என்று படுகிறது.

சில மென்பொருள்களை UNICODEல் மொழிபெயர்க்க முடியாத நிலை இருப்பின், UNICODE support பெறுவதற்கான முயற்சி மேற்கொள்ளவேண்டும். உதாரணம் kde.org இல் UNICODE support.

மற்ற தகுதரங்கள் நம்மைவிட்டு எளிதில் விலகப் போவதில்லை. இந்த சூழ்நிலையில் UNICODE பயன்படுத்தி வடிவமைக்கப்பட்ட மென்பொருளைக் கொண்டு TAB, TSCII, TAM பயன்படுத்தி தகவல் பரிமாற்றம் செய்துகொள்ளும் வசதி இருக்கவேண்டும். இதற்கான முயற்சிகளை டீஸ்கி ஆர்வலர்கள் செய்து வருகிறார்கள்.

TSCII ல் உள்ளிட்டு, unicodeல் compile செய்வதற்கு பதில் unicodeலேயே உள்ளிடவேண்டும். இந்த முறையில் எந்தவிதமான இடர்ப்பாடுகளும் இருக்காது. மாறாக எல்லோரும் ஒரே தகுதரத்தைப் பயன்படுத்தும் ஒருமைப்பாடு நிகழும்.

உலகெங்கும் உள்ள தன்னார்வக்குழுக்கள் தொலைநோக்குடன் செயல்பட்டு UNICODEஐ வருங்கால மாக்கும் முயற்சிகளை மேற்கொள்ள வேண்டும். UNICODEல் தமிழுக்கு அளிக்கப்பட்ட விரயமான இடங்களை மறுபயன்படுத்துவதற்கான முயற்சிகளை UNICODE குழுவினரிடம் எடுத்துரைக்க வேண்டும். அவைகளை மாற்றும் முயற்சிகளில் நம்மிடை தேசம் கடந்த ஒற்றுமை தேவைப்படும். ஒவ்வொருவரும் ஒவ்வொரு யோசனை சொன்னால் அவர்கள் மேலும் குழம்பிப்போவார்கள். இந்த இணைய மாநாட்டில் யூனிகோடில் நாம் விரும்பும் மாற்றங்கள் பற்றிய ஒருமித்த கருத்துக்கு வந்தால் அதுவே ஒரு சாதனையாகும்.

ஐப்பானிய, கொரிய, சீன மொழிகளில் எல்லாம் நிகழ்வது தமிழ் மொழியில் நிகழும்போதுதான் தமிழ் நாடு முன்னேற்றமடையும். இல்லையேல் அடுத்த இணைய மாநாட்டிலும் இதைப் பற்றியே பேசிக் கொண்டிருப்போம். தகுதரங்கள் போதும்; மற்ற வேலைகளை கவனிப்போம்.