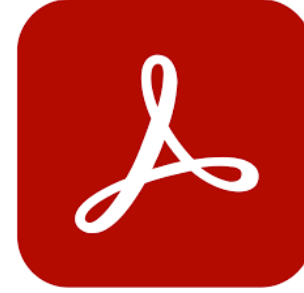




தமிழும் PDF ம்



த. சீனிவாசன்

கணியம் அறக்கட்டளை

tshrinivasan@gmail.com

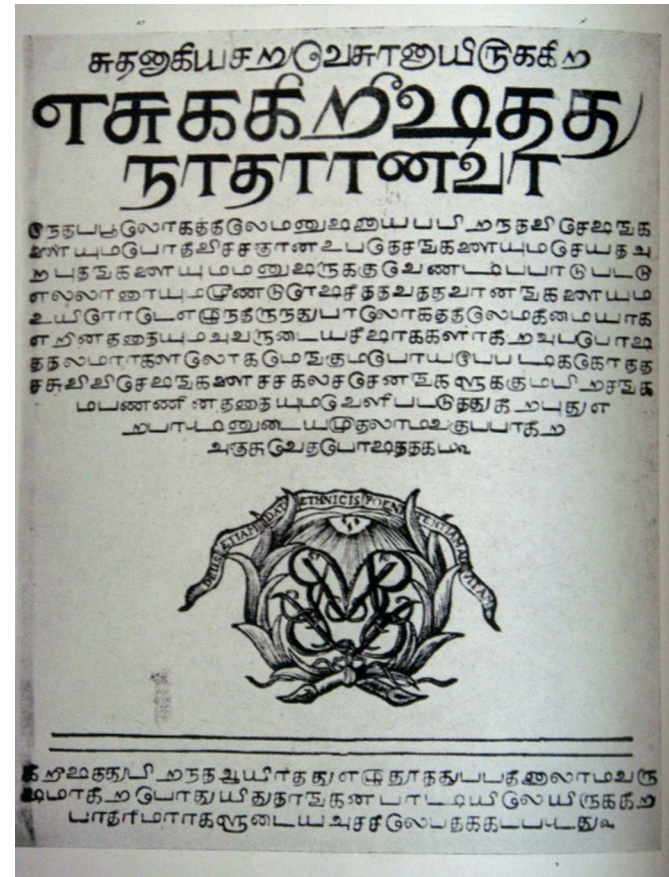
ஓலைச்சுவடி



Handwritten text in Tamil script on a palm leaf manuscript. The text is arranged in approximately 10 horizontal lines. The characters are finely etched into the surface of the leaf. The script is a traditional form of the Tamil language, likely used for religious or philosophical purposes. The leaf shows signs of age and wear, with some discoloration and a slightly uneven texture.

கல்வெட்டு

1715 ல் வெளியான தமிழ் விவிலியம்





The quick brown fox jumps over the lazy dog and all the seven colors of the rainbow together with heaven and earth and the firmament.

Graham



Desktop
Publishing

DTP மென்பொருட்கள்



L^AT_EX

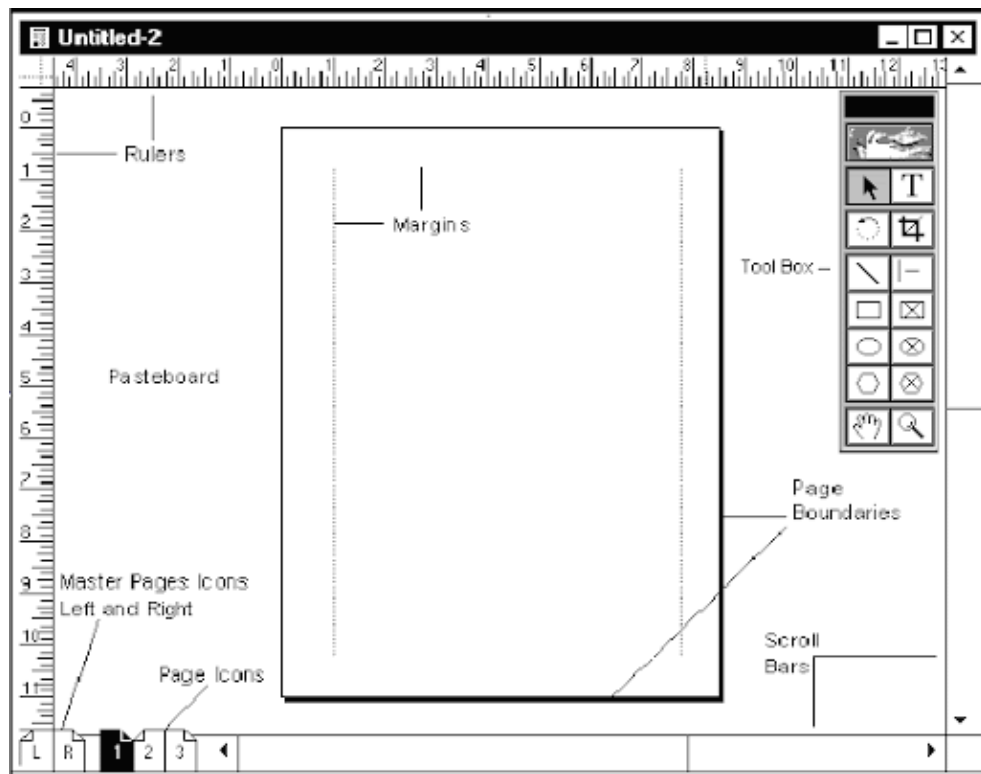




|

shutterstock.com · 1470838202





பல்வேறு தட்டச்சு முறைகள்





ASCII

ISCII

TSCII

Unicode

The ASCII code

American Standard Code for Information Interchange

ASCII control characters

DEC	HEX	Simbolo ASCII	
00	00h	NULL	(carácter nulo)
01	01h	SOH	(inicio encabezado)
02	02h	STX	(inicio texto)
03	03h	ETX	(fin de texto)
04	04h	EOT	(fin transmisión)
05	05h	ENQ	(enquiry)
06	06h	ACK	(acknowledgement)
07	07h	BEL	(timbre)
08	08h	BS	(retroceso)
09	09h	HT	(tab horizontal)
10	0Ah	LF	(salto de línea)
11	0Bh	VT	(tab vertical)
12	0Ch	FF	(form feed)
13	0Dh	CR	(retorno de carro)
14	0Eh	SO	(shift Out)
15	0Fh	SI	(shift In)
16	10h	DLE	(data link escape)
17	11h	DC1	(device control 1)
18	12h	DC2	(device control 2)
19	13h	DC3	(device control 3)
20	14h	DC4	(device control 4)
21	15h	NAK	(negative acknowle.)
22	16h	SYN	(synchronous idle)
23	17h	ETB	(end of trans. block)
24	18h	CAN	(cancel)
25	19h	EM	(end of medium)
26	1Ah	SUB	(substitute)
27	1Bh	ESC	(escape)
28	1Ch	FS	(file separator)
29	1Dh	GS	(group separator)
30	1Eh	RS	(record separator)
31	1Fh	US	(unit separator)
127	20h	DEL	(delete)

ASCII printable characters

DEC	HEX	Simbolo	DEC	HEX	Simbolo	DEC	HEX	Simbolo
32	20h	espacio	64	40h	@	96	60h	`
33	21h	!	65	41h	A	97	61h	a
34	22h	"	66	42h	B	98	62h	b
35	23h	#	67	43h	C	99	63h	c
36	24h	\$	68	44h	D	100	64h	d
37	25h	%	69	45h	E	101	65h	e
38	26h	&	70	46h	F	102	66h	f
39	27h	'	71	47h	G	103	67h	g
40	28h	(72	48h	H	104	68h	h
41	29h)	73	49h	I	105	69h	i
42	2Ah	*	74	4Ah	J	106	6Ah	j
43	2Bh	+	75	4Bh	K	107	6Bh	k
44	2Ch	,	76	4Ch	L	108	6Ch	l
45	2Dh	-	77	4Dh	M	109	6Dh	m
46	2Eh	.	78	4Eh	N	110	6Eh	n
47	2Fh	/	79	4Fh	O	111	6Fh	o
48	30h	0	80	50h	P	112	70h	p
49	31h	1	81	51h	Q	113	71h	q
50	32h	2	82	52h	R	114	72h	r
51	33h	3	83	53h	S	115	73h	s
52	34h	4	84	54h	T	116	74h	t
53	35h	5	85	55h	U	117	75h	u
54	36h	6	86	56h	V	118	76h	v
55	37h	7	87	57h	W	119	77h	w
56	38h	8	88	58h	X	120	78h	x
57	39h	9	89	59h	Y	121	79h	y
58	3Ah	:	90	5Ah	Z	122	7Ah	z
59	3Bh	;	91	5Bh	[123	7Bh	{
60	3Ch	<	92	5Ch	\	124	7Ch	
61	3Dh	=	93	5Dh]	125	7Dh	}
62	3Eh	>	94	5Eh	^	126	7Eh	~
63	3Fh	?	95	5Fh	-			

theASCIIcode.com.ar

Extended ASCII characters

DEC	HEX	Simbolo	DEC	HEX	Simbolo	DEC	HEX	Simbolo	DEC	HEX	Simbolo
128	80h	Ç	160	A0h	á	192	C0h	Ł	224	E0h	Ó
129	81h	ü	161	A1h	í	193	C1h	ł	225	E1h	õ
130	82h	é	162	A2h	ó	194	C2h	ŧ	226	E2h	ö
131	83h	â	163	A3h	ú	195	C3h	ı	227	E3h	ø
132	84h	ä	164	A4h	ñ	196	C4h	ı̇	228	E4h	ö
133	85h	à	165	A5h	Ñ	197	C5h	ı̈	229	E5h	õ
134	86h	á	166	A6h	ª	198	C6h	ı̊	230	E6h	µ
135	87h	ç	167	A7h	º	199	C7h	ı̋	231	E7h	þ
136	88h	è	168	A8h	¿	200	C8h	ı̌	232	E8h	Û
137	89h	ë	169	A9h	®	201	C9h	ı̍	233	E9h	Ü
138	8Ah	è	170	AAh	¬	202	CAh	ı̎	234	EAh	ú
139	8Bh	ï	171	ABh	½	203	CBh	ı̏	235	EBh	Û
140	8Ch	î	172	ACH	¼	204	CCh	ı̐	236	ECh	ý
141	8Dh	ï	173	ADh	ı	205	CDh	ı̑	237	EDh	ÿ
142	8Eh	Ä	174	A Eh	«	206	CEh	ı̒	238	EEh	ı
143	8Fh	Å	175	AFh	»	207	CFh	ı̓	239	EFh	ı
144	90h	É	176	B0h	⋮	208	D0h	ı̔	240	F0h	ı
145	91h	æ	177	B1h	⋮	209	D1h	ı̕	241	F1h	±
146	92h	Æ	178	B2h	⋮	210	D2h	ı̖	242	F2h	ı
147	93h	ô	179	B3h	⋮	211	D3h	ı̗	243	F3h	¼
148	94h	ò	180	B4h	⋮	212	D4h	ı̘	244	F4h	¶
149	95h	ò	181	B5h	⋮	213	D5h	ı̙	245	F5h	§
150	96h	ù	182	B6h	⋮	214	D6h	ı̚	246	F6h	÷
151	97h	ú	183	B7h	⋮	215	D7h	ı̛	247	F7h	ı
152	98h	ÿ	184	B8h	©	216	D8h	ı̜	248	F8h	ı
153	99h	Ö	185	B9h	⋮	217	D9h	ı̝	249	F9h	ı
154	9Ah	Ü	186	BAh	⋮	218	DAh	ı̞	250	FAh	ı
155	9Bh	ø	187	BBh	⋮	219	DBh	ı̟	251	FBh	ı
156	9Ch	£	188	BCh	⋮	220	DCh	ı̠	252	FCh	ı
157	9Dh	ø	189	BDh	¢	221	DDh	ı̡	253	FDh	ı
158	9Eh	x	190	BEh	¥	222	DEh	ı̢	254	FEh	ı
159	9Fh	f	191	BFh	¬	223	DFh	ı̣	255	FFh	ı

ISCII

Indian Script Code for Information Interchange

தகவற் பரிமாற்றத்திற்கான இந்திய நியமக் குறியீட்டு முறை (இஸ்கீ; Indian Script Code for Information Interchange, ISCII) என்பது **இந்திய** மொழிகளுக்கான ஒரு குறியீட்டு முறை, இது பெரும்பாலான மொழிகளையும், அதன் ஒலிபெயர்ப்பையும் குறிப்பிடத்தக்கது. பின்வரும் மொழிகளில் இஸ்கீ குறியீட்டு முறை: **அஸ்ஸாமி, பெங்காலி** (பங்களா) **ஸ்கிரிப்ட், தேவநாகரி, குஜராத்தி, குர்முகி, கன்னடம், மலையாளம், ஒரியா, தமிழ், மற்றும் தெலுங்கு**

இஸ்கீ பரவலாக சில அரசு நிறுவனங்களால் பயன்படுத்தப்பட்டு வருகிறது. ஆனாலும் அதைப் பரவலாக மக்கள் ஏற்கவில்லை. **இஸ்கீ**-யைவிட இப்போது **யுனிக்கோடே** பெரும்பாலும் பயன்படுத்தப்படுகின்றது.

TSCII

Tamil Standard Code for Information Interchange

தகவற் பரிமாற்றத்திற்கான தமிழ் நியமக் குறியீட்டு முறை என்பது **கணினி** சார் தேவைகளுக்கு **தமிழ்** எழுத்துக்களை பயன்படுத்துவதற்கென உருவாக்கப்பட்ட, எட்டு பிட் அடிப்படையில் அமைந்த ஆங்கிலம்-தமிழ் என இரு மொழிகளை கையாளத்தக்க ஒரு **எழுத்து குறிமுறை** நியமமாகும். இக்குறிமுறையின் பெயர் ஆங்கிலத்தில் சுருக்கமாக (திஸ்கீ) **TSCII** (Tamil Standard Code for Information Interchange) என்றவாறு குறிப்பிடப்படுகிறது. தமிழில் **திசுகி** எனவும் **தகுதரம்** (**தமிழ் குறியீட்டுத் தரம்**) எனவும் வழங்கப்படும் இதுவே முதன்முதல் தமிழில் உலகம் தழுவிய திறந்த இணைய உரையாடல் மூலம் தரப்படுத்தப்பட்டு உருவாக்கப்பட்ட குறிமுறை நியமமாகும்.

ASCII குறிமுறையில் அமைந்த ஆங்கில எழுத்துக்களின் மேலாக, **எழுத்துரு** ஒன்றினை பயன்படுத்தி தமிழ் எழுத்துக்களை கணினியில் பயன்படுத்திவந்த காலத்தில், தமிழ் - ஆங்கிலம் என இரு மொழிகளையும் பயன்படுத்தி ஆவணமொன்றினை தொகுப்பது மிகக் கடினமான வேலையாக இருந்தது. **தமிழ் எழுத்து குறிமுறை** வரலாற்றில், இவ்வாறான சிக்கல்மிக்க ஒருமொழி ASCII குறிமுறையின் போதாமையை களைந்து உருவாகிய அடுத்தக்கட்ட படிமுறை வளர்ச்சியே **தகுதரம்**. ASCII குறிமுறை அட்டவணையில் ஆங்கில எழுத்துக்கள், இன்றியமையாக குறியீடுகள் தவிர்ந்த ஏனைய இட ஒருக்கீடுகளில் தமிழ் எழுத்துக்களை பிரதியீடு செய்தலே **தகுதரம்** குறிமுறையின் அடிப்படையாகும். தொடக்கத்தில் ஆங்கில எழுத்துக்களுக்கான ஒதுக்கீட்டிலேயே தமிழ் எழுத்துக்களை உள்ளீடு செய்த முறைமையைக் காட்டிலும் இது முன்னேற்றகரமானதாகவும், இருமொழி கையாட்சிக்கு எளிமையானதாகவும் இருந்தது.

TSCII

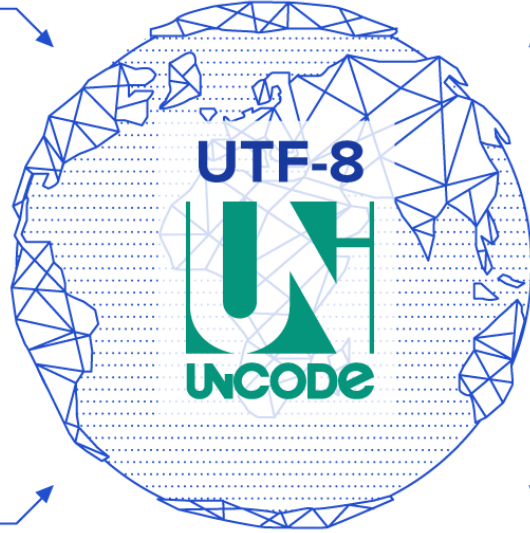
A possible 8-bit font encoding scheme for Tamil -v. 1.7
(Tamil standard Code for Information Interchange TSCII)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
0 (0-15)																	
1 (16-31)																	
2 (32-47)		!	"	#	\$	%	&	'	()	*	+	,	-	.	/	
3 (48-63)	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?	
4 (64-79)	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
5 (80-95)	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_	
6 (96-111)	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	
7 (112-127)	p	q	r	s	t	u	v	w	x	y	z	{		}	~		
8 (128-143)	஠	க	ஙீ	ஐ	ஷ	ஸ	ஹ	ஃ	ஐ	ஃ	ஸ்	ஹ்	ஃ	ஃ	உ	ந	ச
9 (144-159)	஠	‘	’	“	”	஠	எ	அ	஠	஠	஠	஠	஠	஠	஠	஠	஠
A (160-175)		஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	
B (176-191)	஠	எ	ஏ	ஐ	ஓ	ஓ	ஓ	ஃ	க	ங	ச	஠	஠	஠	஠	஠	
C (192-207)	ப	ம	ய	ர	ல	வ	ழ	ள	ற	ன	டி	டி	கு	சு	டு	஠	
D (208-223)	து	நு	பு	மு	யு	ரு	லு	வு	மு	஠	று	஠	஠	஠	஠	஠	
E (224-239)	தா	நா	பு	மு	யு	ரு	லா	வா	மு	஠	று	஠	஠	஠	஠	஠	
F (240-255)	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	஠	

123
ABC

吧博
什兩

ضلع
رکھ



ᱫᱷᱟ
ᱵᱤᱨ

थड़
कजड़

👤
👉
👇
🔴
🔵
🌟

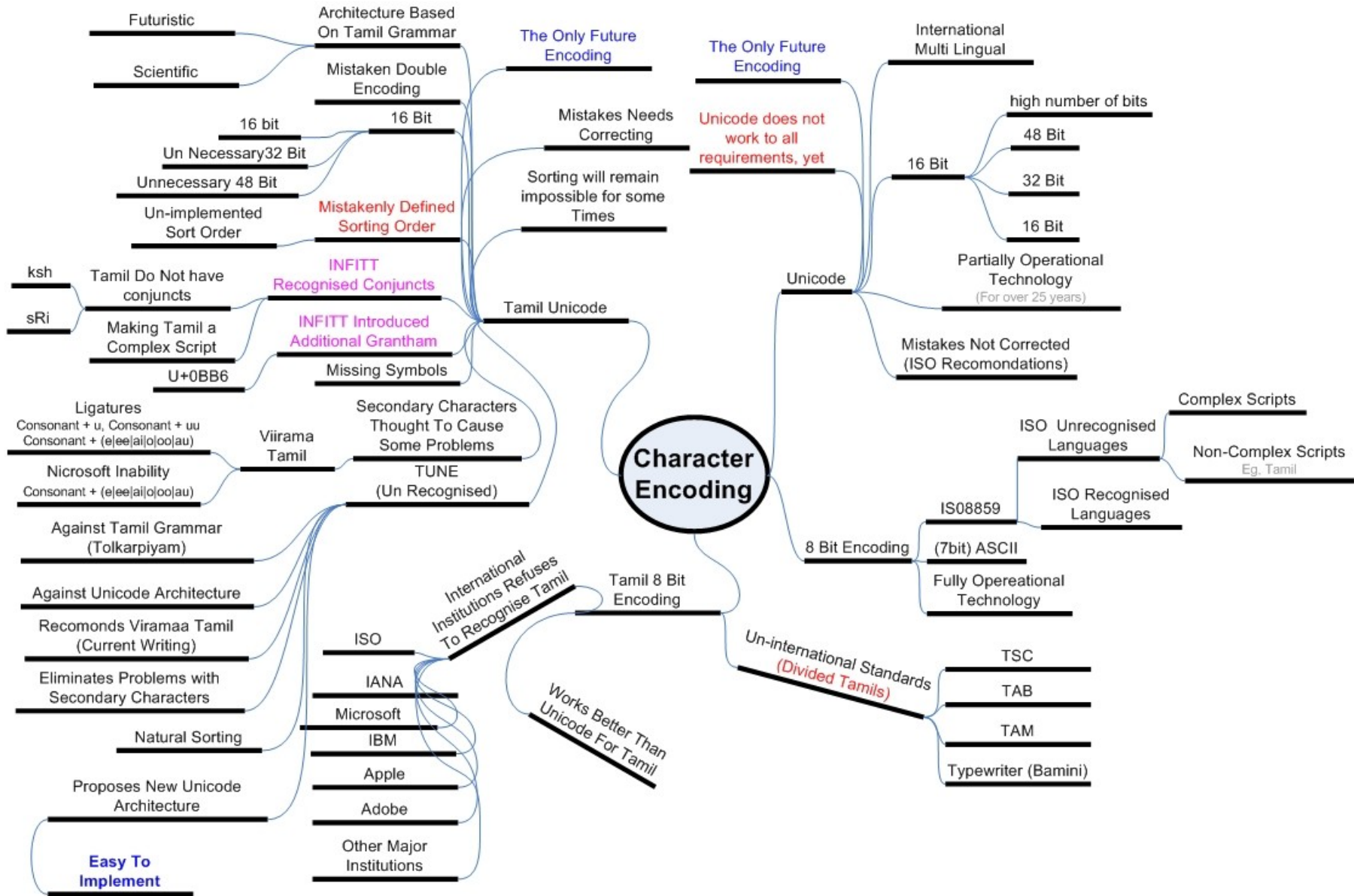
தமிழ் ஒருங்குறி

தமிழ் ^[1]																
Unicode.org chart (PDF)																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+0B8x			ஃ	ஃ		அ	ஆ	இ	ஈ	உ	ஊ				எ	ஏ
U+0B9x	ஐ		ஓ	ஓ	ஓள	க				ங	ச		ஐ		ஞ	ல
U+0BAx				ண	த				ந	ன	ப				ம	ய
U+0BBx	ர	ற	ல	ள	ழ	வ	ஸ	ஷ	ஸ	ஹ					ா	ி
U+0BCx	ீ	ஶ	ஶ				ெ	ே	ை		ொ	ோ	ெள	ஃ		
U+0BDx	ஔ							ள								
U+0BEx							஠	க	உ	ங	ச	ஞ	கா	எ	அ	கூ
U+0BFx	ய	ள	த	உ	மீ	ஶ	பூ	ஶ	ஷ	நீ						

Notes

1.^ As of Unicode version 6.1

https://ta.wikipedia.org/wiki/தமிழ்_ஒருங்குறி



ஒருங்குறியத்தமிழ்
85% தமிழ் எழுத்துக்களுக்கு இடமில்லை.
பிற எழுத்துக்களைப் பெற, தனியொரு இடைமென்பொருள் தேவை
கட்டளைகள், இரண்டாம்தர நிலை செயற்பாட்டில் (Level - 2 Implementation) தான் செயற்படும்.
தரவுத்தேக்கமும், செயற்பாட்டு நேரமும் 40%அதிகம்.
ISCI-1988 தரத்தை அடிப்படையாகக் கொண்டது. அதில் தமிழ் எழுத்துக்கள் இயற்கையான அகரவரிசையில் இல்லை. எனவே, அகரவரிசைப்படுத்த ஒத்திணைப்பு மென்பொருள்(collection Algorithm) தனியாகத்தேவை.
ஒரு உயிர்மெய் எழுத்தை உருவாக்க, 2,3புள்ளிகள் தேவை. அதனால் பாதுகாப்புச் சிக்கல்களும்(security vulnerabilities), பொருள் மயக்கச்சேர்க்கையையும்(ambiguous combination) உண்டாகின்றன. இவற்றைச் சரிசெய்ய இயல்பாக்க மென்மியம் தேவை.
எழுத்துக்கள் பிரியாமல் இருக்கவும் அல்லது பிரிந்து இருக்கவும் கணினிக்கு கூடுதல் பணிச்சுமை ஏற்படுகிறது.(zero - width joiner/ zero - width non-joiner)
தவறான எழுத்துக்கள் உருவாக்கப்படாமல் இருக்க, விலக்கெழுத்து அட்டவணை தேவை.
தகவற்பரிமாற்றத்திற்கு மட்டுமே உதவும்.
ஒருங்குறியச்சேர்த்திய விதிகளின் படி, எழுத்துக்கள்(characters)மட்டுமே குறியிடப்படல் வேண்டும். வரிவரைவுகள் (glyphs) குறியிடப்படக்கூடாது. எனவே, தற்போதுள்ள இந்த ஒருங்குறியத்தமிழ் ஒருங்குறியக்கூட்டமைப்பின் கோட்பாடுகளுக்கு முரணானது.
கோப்பைகளை சேமிப்பதில் அதிக இடம் தேவைப்படும்
அதிக இணைய தகவல் பரிமாற்ற தேவை
இந்த ஒருங்குறியத்தமிழ், பல இயக்கமென்மியங்களில் சரிவர செயற்படுவதில்லை.

யூனிகோடும் UTF-8 முறையும்


Nov 25th, 2003 by இரா. செல்வராசு

யூனிகோடு மற்றும் UTF-8 குறியீட்டு முறைகள் பற்றி எனக்கு இன்னும் கொஞ்சம் தெளிவு ஏற்பட்டிருக்கிறது. கணினிகளின் பயன்பாட்டுக்கு எப்படி அமெரிக்க ஆஸ்கி முறை ஆங்கில எழுத்துக்களுக்கு உரிய எண்களை அடையாளம் காட்டுகிறதோ, அதுபோல யூனிகோடு உலக மொழிகள் அத்தனையிலும் உள்ள எழுத்து வடிவங்களுக்கும் உரிய ஒரு எண்ணைக் கட்டிச் சேர்த்து வைக்கும் பெரிய அண்ணன். உதாரணத்திற்கு 65 என்றால் 'A', '97' என்றால் குட்டி 'a' – இது ஆஸ்கி. ஒரு பைட்டுக்கு ஒரு எழுத்து என்று ஆங்கிலத்திற்கு மட்டும் 128 இடங்களுடன் அது போதுமானதாய் இருந்தது. ஏன் மிச்சம் ஒரு பிட் கூட இருந்தது. (இன்னொரு 128 இடங்கள் = விரிவு ஆஸ்கி). ஆனால் எல்லா மொழிகளையும் உள்ளடக்க இது போதுமானதாய் இல்லை. சரி, ஒன்றிற்கு இரண்டு பைட்டுகள் என்று (16 பிட் = 65536 இடங்கள்) எடுத்துக் கொண்டால் இது சாத்தியமாகும் என்று யூனிகோடு முயற்சிகள் ஆரம்பமாயின. இப்போது சில சீன, கொரிய, ஜப்பானிய எழுத்துக்களின் தேவைக்காக 4 பைட் வரை (32 பிட்) எடுத்துக் கொள்ளலாம் என்ற அளவிற்கு வந்திருக்கின்றனர். ஆனால் பெரும்பாலான மொழிகளை இன்னும் இரண்டு பைட்டுக்களுக்குள் அடக்கி விடலாம்.

0 முதல் 65535 வரையில் சாத்தியமான எண்களில், தமிழ் எழுத்துக்களுக்கு என்று ஒதுக்கப்பட்ட இடங்கள் 2944 முதல் 3071 வரை. இதையே U+0B80 முதல் U+0BFF வரை என்றும் கூடச் சொல்லலாம். இது பதினாறாம் அடிப்படையில் இருக்கும் இலக்க முறை (hexa decimal system representation). இந்த எண்களை இரண்டு பைட்டுக்களில் அடக்கி விடலாம். பதினாறு பிட்டுக்கள் போதும்.

எல்லா யூனிகோடு எழுத்துக்களையும் இரண்டு பைட்டுக்களுக்குள் அடைத்து விட முடியுமா எனில் அதில் சில இடைஞ்சல்கள். முதலில் பழைய ஆஸ்கி எழுத்துக்கள். அவை நீண்ட காலமாக ஒரு பைட் கொண்டே பயன்படுத்தப்பட்டன. அதனால் அதே முறையில் இருந்தால் தான் வசதி. இரண்டாவது, புதிய 4 பைட்டுக்கள் தேவைப்படும் சீ.கொ.ஐ (சீன, கொரிய, ஜப்பானிய) எழுத்துக்களை இரண்டு பைட்டுக்களில் அடக்க முடியாது.

இந்தச் சிக்கல்களை நீக்க ஒரு குறியீட்டு முறை வேண்டும் என்று பிறந்தது தான் UTF-8 முறை.

Font	Glyphs	Character Repertoire																														
<p style="text-align: center; font-size: 48pt; font-weight: bold;">A</p>		<table border="1" style="width: 100%; text-align: center;"> <tr> <td data-bbox="1081 364 1193 482">´</td> <td data-bbox="1193 364 1306 482">μ</td> <td data-bbox="1306 364 1419 482">¶</td> <td data-bbox="1419 364 1532 482">·</td> <td data-bbox="1532 364 1644 482">.</td> <td data-bbox="1644 364 1757 482">¹</td> </tr> <tr> <td data-bbox="1081 482 1193 599">°</td> <td data-bbox="1193 482 1306 599">»</td> <td data-bbox="1306 482 1419 599">¼</td> <td data-bbox="1419 482 1532 599">½</td> <td data-bbox="1532 482 1644 599">¾</td> <td data-bbox="1644 482 1757 599">¿</td> </tr> <tr> <td data-bbox="1081 599 1193 716">À</td> <td data-bbox="1193 599 1306 716" style="background-color: #4a7c9c; color: white;">A</td> <td data-bbox="1306 599 1419 716">Â</td> <td data-bbox="1419 599 1532 716">Ã</td> <td data-bbox="1532 599 1644 716">Ä</td> <td data-bbox="1644 599 1757 716">Å</td> </tr> <tr> <td data-bbox="1081 716 1193 833">Æ</td> <td data-bbox="1193 716 1306 833">Ç</td> <td data-bbox="1306 716 1419 833">È</td> <td data-bbox="1419 716 1532 833">É</td> <td data-bbox="1532 716 1644 833">Ê</td> <td data-bbox="1644 716 1757 833">Ë</td> </tr> <tr> <td data-bbox="1081 833 1193 953">Ì</td> <td data-bbox="1193 833 1306 953">Í</td> <td data-bbox="1306 833 1419 953">Î</td> <td data-bbox="1419 833 1532 953">Ï</td> <td data-bbox="1532 833 1644 953">Ð</td> <td data-bbox="1644 833 1757 953">Ñ</td> </tr> </table>	´	μ	¶	·	.	¹	°	»	¼	½	¾	¿	À	A	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ
´	μ	¶	·	.	¹																											
°	»	¼	½	¾	¿																											
À	A	Â	Ã	Ä	Å																											
Æ	Ç	È	É	Ê	Ë																											
Ì	Í	Î	Ï	Ð	Ñ																											

Unicode

- **Unicode** provides a **unique *number*** for every character

- The **number 65** of Unicode represents the **(English) character A**
- The **number 35** of Unicode represents the **character #**
- The **number 948** of Unicode represents the **(Greek) character δ**
- The **number 24373** of Unicode represents the **(Chinese) character 張**

இந்த முடிபின்படி தமிழின் ஒவ்வொரு எழுத்துக்கும் ஒரு குறிப்புள்ளி கொடுக்கப் பட்டிருக்க வேண்டும். அப்படி இருந்திருந்தால் ஒரு உதாரணத்துக்கு 'அம்மா' என்ற சொல்லுக்கு மூன்று குறிப்புள்ளிகள் போதுமானதாக இருந்திருக்கும். 'அ'வுக்கு ஒன்று 'ம்'க்கு ஒன்று; மற்றும் 'மா'வுக்கு ஒன்று.

அப்படி எளிமையாக நேரடியாக (linear model) இருந்திருக்க வேண்டிய ஒன்று, தேவையின்றித் தமிழையும் இந்திய மொழிக் குடும்பத்தோடு (indic languages) சேர்த்த காரணத்தால் சற்றுக் குதறிப் போயிருக்கிறது.

இந்தக் குதறலுக்கு ஒரு எடுத்துக்காட்டைப் பார்ப்போம்.
தற்போதைய யூனிக்கோட்டுத் தமிழில் 'அம்மா'வை இப்படித்
தான் பிளந்து போட்டிருக்கிறார்கள்.

0B85: அ TAMIL LETTER A (Tamil)
0BAE: ம TAMIL LETTER MA (Tamil)
0BCD: ` TAMIL SIGN VIRAMA (Tamil)
0BAE: ம TAMIL LETTER MA (Tamil)
0BBE: ா TAMIL VOWEL SIGN AA (Tamil)

கணினிகளில் இன்று யூனிக்கோடு தமிழ் தெரிய இரண்டு விஷயங்கள் வேண்டும்.

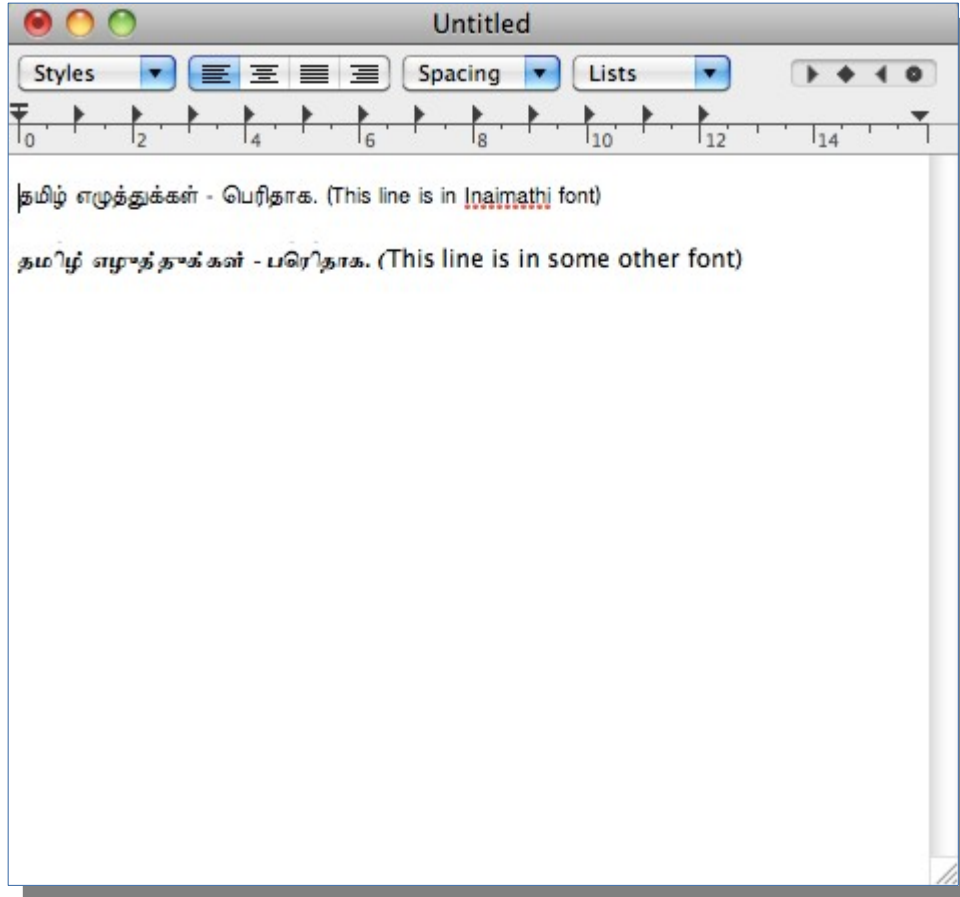
ஒன்று - ஒரு ஒருங்குறி எழுத்துரு (font):இது ஒவ்வொரு குறிப்புள்ளி அல்லது குறிப்புள்ளி வரிசைக்கு என்ன வடிவத்தை நல்க (render) வேண்டும் என்று அமைக்கப்பட்டிருக்கும் ஒரு அட்டவணை. அதில் சிறப்பு வடிவங்கள், கூட்டெழுத்து வடிவங்கள் இவை இருக்கலாம்.

இரண்டு - குறிப்புள்ளி வரிசையைப் புரிந்து கொண்டு முன்னும் பின்னுமாக வடிவங்களை மாற்றிப் போட்டுச் சரியான வடிவத்தைத் தரும் 'பலக்கிய எழுத்துவடிவ நல்கு எந்திரம்' (complex script rendering engine).

உதாரணத்துக்கு, '0BE7' '0BCA' என்னும் குறிப்புள்ளிகளுக்கான வடிவம்: 'க' மற்றும் 'ொ'. இவற்றில் இரண்டாவதாக உள்ள வடிவத்தை இரண்டாக உடைத்து ஒன்றைக் ககரத்தின் முன்னும் ஒன்றைப் பின்னும் போட்டுக் 'கொ' என்னும் வடிவத்தைக் கொண்டு வர வேண்டும். இந்த வேலையைத் தான் 'பலக்கிய எழுத்து நல்கு எந்திரம்' செய்கிறது. அதனால் தான் சரியான முறையில் யூனிக்கோட்டுத் தமிழைக் கணினிகளில் பார்க்க முடிகிறது.

Linux - Pango

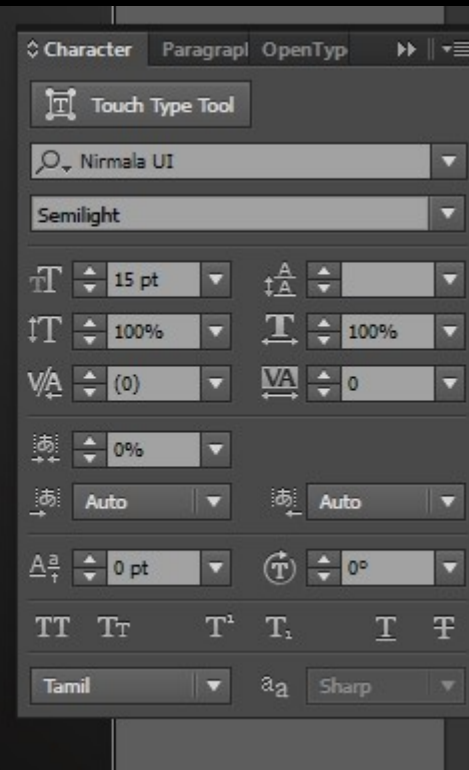
windows - uniscribe



மேலும், புகை,

இயற்கை மூலிகைகளில் இரந்த
ஈர்த்தடெக்கப்பட்ட வீரிய
உட்பொருட்களை உள்ளடக்கி
எந்த இரசாயன சேர்க்கைகளும்
இல்லாமல் உரவாக்கப்பட்ட
இந்தியாவின் முதல் சித்த
தயாரிப்பு.

மடியை வரேலிரந்த பலப்படத்தி,
பொடக மர்மம் மடி உகிர்கலகை



```
<?php
header('Content-type: image/png; charset=UTF-8');
$text = 'சுற்றிப்பார்க்க ';
$font = './fonts/UniIlla.ttf';
$im = imagecreatetruecolor(600, 300);
$bg_color = imagecolorallocate($im, 255, 255, 255);
$font_color = imagecolorallocate($im, 0, 0, 0);
imagefilledrectangle($im, 0, 0, 599, 299, $bg_color);
imaggottext($im, 20, 0, 10, 50, $font_color, $font, $font);
imaggottext($im, 20, 0, 10, 120, $font_color, $font, $text);
imagepng($im);
imagedestroy($im);
?>
```

./fonts/UniIlla.ttf

சுற்றிப்பார்க்க

Tamil font not displayed correctly #6018

🔒 Closed

muruga86 opened this issue on Feb 4, 2022 · 18 comments



muruga86 commented on Feb 4, 2022 · edited ↕

I am writing a small program in python using PIL library to render a Tamil text on an image using couple of Tamil fonts. However, certain letters in Tamil are not displayed properly. Some characters are not correctly or aesthetically printed on image like பி, ழ, த், து, யோ, கு, டி. Need to use any reshaper or any other thing in code. Could some one throw light on this issue?

Even tried layout_engine=ImageFont.LAYOUT_RAQM, but still same issue

Example: இனிய பிறந்தநாள் வாழ்த்துக்கள் யோஷினி குட்டி

Displayed as:

இனிய பிறந்தநாள் வாழ்த்துக்கள் யோஷினி கட்டி

Correct display:

இனிய பிறந்தநாள் வாழ்த்துக்கள் யோஷினி குட்டி

Code:



PDF ReportLab Tamil Forn Not Show Correctly ✎

■ வினா ■ Language and Input



It's been a while since we've seen Shaik — their last post was 2 years ago.



Shaik

Apr 2023

PDF கோப்பு REPORTLAB மற்றும் PYTHON பயன்படுத்தி உருவாக்கும் போது தமிழ் எழுத்துருக்கள் சரியாக தோன்றவில்லை, எப்படி சரி செய்வது?

முதலில் உள்ள வரி SUNTOMMY எழுத்துரு
இரண்டாவது வரி NOTOSANSTAMIL எழுத்துரு
UNICODE எழுத்துருக்கள் வேறு பயன்படுத்தினாலும் இப்படியேதான் உள்ளது...

ரோஜா பூ பூத்துள்ளது

ரோஜா பூ பூத்துள்ளது



Apr 2023

1 / 21

Apr 2023

May 2



✓ Solved by Shaik in post #19

#solved தற்காலிகமான தீர்வாக glyph களை தெலுங்கு யூனிகோடுகளின் இடத்தில் மாற்றம் செய்து கீழுள்ள பைதான் புரோகிராம் உதவியுடன் சரியான தமிழ் எழுத்தாக மாற்றம் செய்யப்படுகிறது [Screenshot from 2023-06-08 21-11-07] [Screenshot from 2023-06-08 21-44-46] நன்றி @tshrinivasan @SelvamS

```
temp
Close Edit View Bookmarks Plugins Settings Help
New Tab Split View
ocr4wikisource : tmux: client temp : bash Downloads : bc - : sudo su Books : bash tpl-map : bash (shrini)
shrini@dell-optiplex-9100:~/temp$ ls
indic-trans test-tamil.txt t.yaml
shrini@dell-optiplex-9100:~/temp$ cat test-tamil.txt
தாங்கள் முயற்சி க்கும் ta-phonetic மற்றும் ta-itrans எண் யெர்கள்
காட்டு ம் விகைற்றிகள் இனும் ஏனை "ta-" தெங்கும் தமிழ்
விகைற்றிகள் எல்லாம் m17n எனும் உள்ளிடல் முறையும் மெறி (Input Method
Engine - IME) இல் உள்ள பல்வேறு மெறிகளக்கானவிகைற்றிகளில்
உள்ளடங்கு பக்கம்

m17n ஐ தற்கால விகைற் வினியேசங்களில் இக்குவது ibus, scim மற்றும்
uim எனும் 3 உள்ளிடல் முறைமக்டமைபு (Input Method Frameworks) களில்
ஒற்றல் - தாங்கள் பயன்பு த்துவது ibus.

(m17n ச யமாக இபக்க சவு யது mlterm எனும் முறையிலும் e-macs எனும்
உறைய வ இமுகப்பு ( Textual User Interfac - TUI) உனனஉதை ரு த்தி
கெலியிலும். அற்றில் தமிழ் மற்றும் ஏனை தெற்காசி ய சி க்கல்
விரி யு ரு க்கள் உரு வாகக்கலுக்கு த்து வையிலதலால் அற்றில் தமிழ்
உள்ளிட்டும் பயன்பெனவது வேறு விடயம்)

தக்கைள் லேபு துப் பயர்களின்தேவைகக்கு துணைக வழிகாட்டு ம் கைகில்
தெர்பானவிகைற்றிகளின்தகக்கே உரு வக்கைள (Keyboard Layout) யராவுது
அக்கி வெரியிட வேண்டும் தாக்

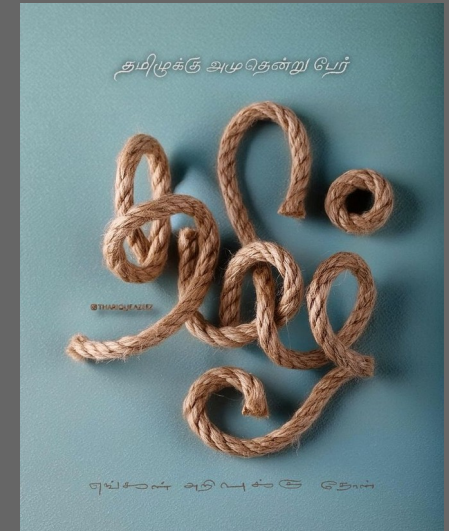
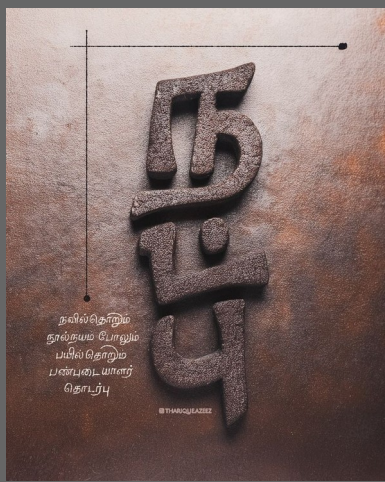
தற்பேசைக்குத் தாங்கள் அப்பிர விகைற்றிகளக்கானகேபு க்கைளedit
அவது kate லேபு உதை ரு த்தி யில் திறந்து அபு படைஅர - உயிர்மெய்
மற்றும் உயிர் எழுத்துக்கள் எண்ணற்றிரகாக map கெய்ப்புடும் அக்கில
விகைள் அவது விகைத்தெர்க்கைளறியலாம். பிணரும் கேபு க்கைத்
திறந்து பாரக்கவும் (உபு எடு களில் உள்ளவறு) :
shrini@dell-optiplex-9100:~/temp$
```

PDF – பயன்கள்



பல்வேறு எழுத்துருக்கள்

தமிழ் தமீழ் தமிழ் தமிழ்
தமிழ் தமீழ் தமிழ் தமிழ்
தமிழ் தமிழ் தமிழ் தமிழ்
தமிழ் தமிழ் தமீழ் தமிழ்
தமிழ் தமீழ் தமிழ் தமிழ்



வடிவமைப்பு



தமிழ் PDF - சிக்கல்கள்



தவறான உரை நகல்

COPY

CTRL+C

PASTE

CTRL+V

உலகமே உற்று நோக்கும் ஒரு அற்புதச் சுற்றுலாத் தலமாக அந்தமான் தீவுகள் உள்ளன. வாழ்க்கையில் ஒரு முறையேனும் அந்தமான் தீவிற்குச் சென்று வர வேண்டும். பூமியின் சொர்க்கம் என அந்தமான் தீவுகளை சிலர் கூறுகின்றனர். சொர்க்கம் என்பது எப்படிப்பட்டது என்பது யாருக்கும் தெரியாது. ஆனால் அந்தமான் தீவுகள் இயற்கை வளம் நிறைந்த தீவுகளைக் கொண்டுள்ளது. மனிதச் சுரண்டலுக்கு ஆட்படாத பல தீவுகள் இருக்கின்றன. எந்தப்பக்கம் சென்றாலும் கடலும், தீவுகளுமே தெரிகின்றன. நீலக்கடலுக்கு இடையே பச்சை நிறம் கொண்ட தீவுகள் ரசிக்கும்படியாக இருக்கின்றன. இங்கே அமைதியான மாசுபடாத, வெண்மணல் கொண்ட கடற்கரையைக் காணலாம். அரிய தாவரங்கள், விலங்குகள், பறவைகள் என பல்லுயிர் வளம் மிக்க பகுதியாக அந்தமான் தீவுகள் விளங்குகின்றன. கடற்கரை ஓரத்திலேயே பல வண்ணம் கொண்ட பவளப் பாறைகளையும், அதன் ஊடே நீந்திக் திரியும் பல வண்ண மீன்களையும் ரசிக்க ஏற்ற இடம். இயற்கை அழகோடு வரலாற்றுச் சிறப்பு மிக்க பல இடங்களும் உள்ளன. அந்தமான் பயணத்தில் முழுக்க, முழுக்க சுற்றுலா அனுபவத்தைப் பெறலாம்.

உலககேம உற்று கே ாக்கும் ஒரு அற்புதச் சுற்றுலாத் தலமாக அந்தமான் அழகு தீவுகள் உள்ளன .

வாழ்க்கைகையில் ஒரு முகை(கே)னும் அந்தமான் அழகு தீவிற்குச் செ-ன் அழகுறு வர கேவண்டும். பூமியின் அழகு

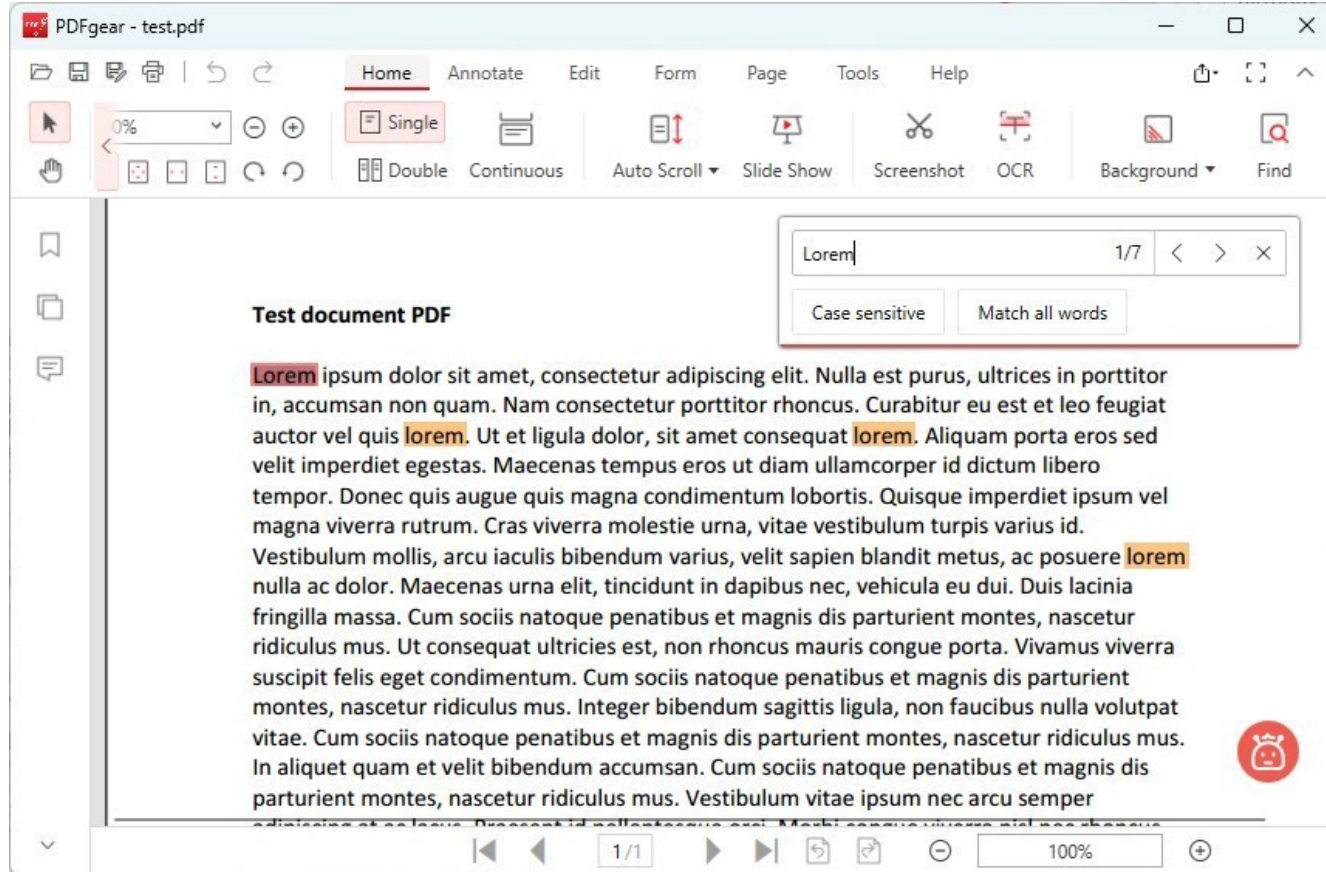
செ-ார்க்கம் என அந்தமான் அழகு தீவுகளை சிலர் கூறுகின் அழகு(னர் .
செ-ார்க்கம் என் அழகுபது எப்படிப்பட்டது என் அழகுபது)ாருக்கும்
செதரி)ாது . ஆனால் அந்தமான் அழகு தீவுகள் இ)ற்கைக வளம் நிகை(ந்த

தீவுகளைக் செகாண்டுள்ளது . மனிதச் சுரண்டலுக்கு ஆட்படாத பல தீவுகள் இருக்கின் அழகு(ன . எந்தப்பக்கம் செ-ன் அழகு(ாலும் கடலும் , தீவுகளுகேம செதரிகின் அழகு(ன . நீலக்கடலுக்கு இகைடகே) பச்சை- நி(ம் செகாண்ட தீவுகள் ரசிக்கும்படி)ாக இருக்கின் அழகு(ன . இங்கேக அகைமதி)ான மாசுபடாத ,

செவண்மணல் செகாண்ட கடற்கரைகை)க் காணலாம் .

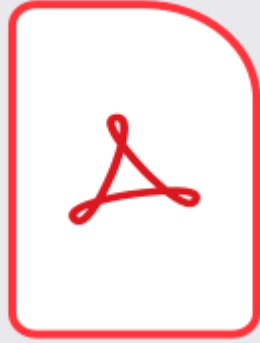
அரி) தாவரங்கள் , விலங்குகள் , ப(கைவகள் என பல்லுயிர் வளம் மிக்க பகுதி)ாக அந்தமான் அழகு தீவுகள் விளங்குகின் அழகு(ன

வார்த்தைகளை தேட முடியாது



தீர்வு





PDF/A

Comparison of Archive Formats



PDF/A



PDF



TIFF
without OCR

Device Independent			
100% self-contained			
Guaranteed WYSIWYG			
Archival standard approved by ISO			
Guaranteed color reproduction			
Contains XMP metadata			
Guaranteed accessibility across multiple platforms/systems			
Embedded fonts			
Free from proprietary constraints			
Able to capture documents' logical structure			
High quality output			
Searchable text			
Most compact file size			



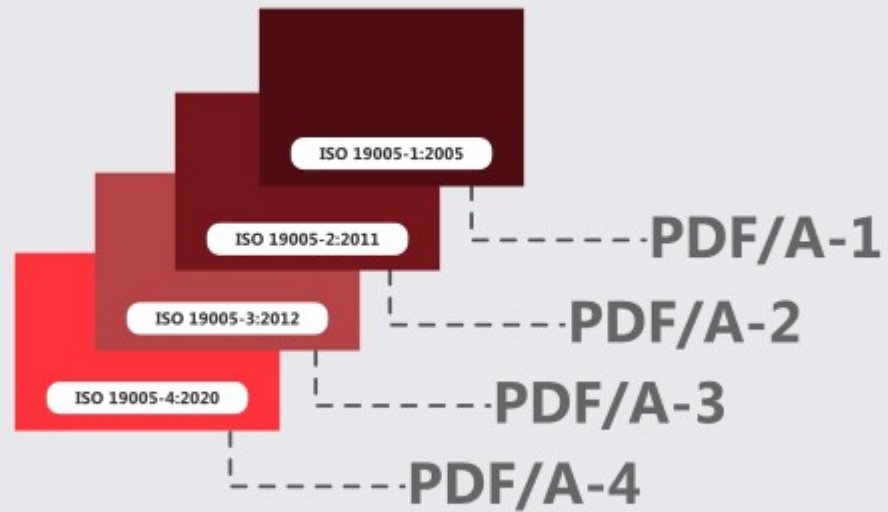
= Always



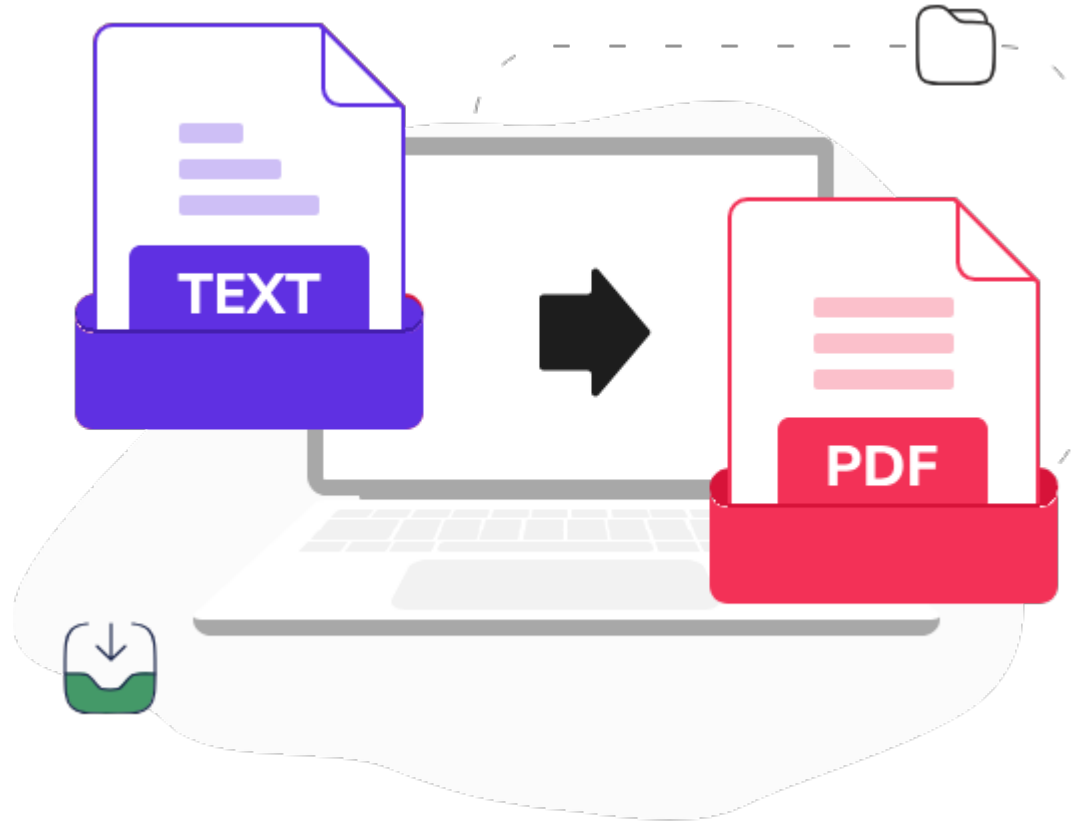
= Sometimes



= Never



Born Digital



PDF Options

General Initial View User Interface Links Security Digital Signatures

Range

- All
- Pages:
- Selection
- View PDF after export

Images

- Lossless compression
- JPEG compression
- Quality:
- Reduce image resolution

Watermark

- Sign with watermark

General

- Hybrid PDF (embed ODF file)
- Archive (PDF/A, ISO 19005)
- PDF/A-2b PDF/A-1b
- Tagged PDF (add document structure)
- Create PDF form
- Submit format:
- Allow duplicate field names

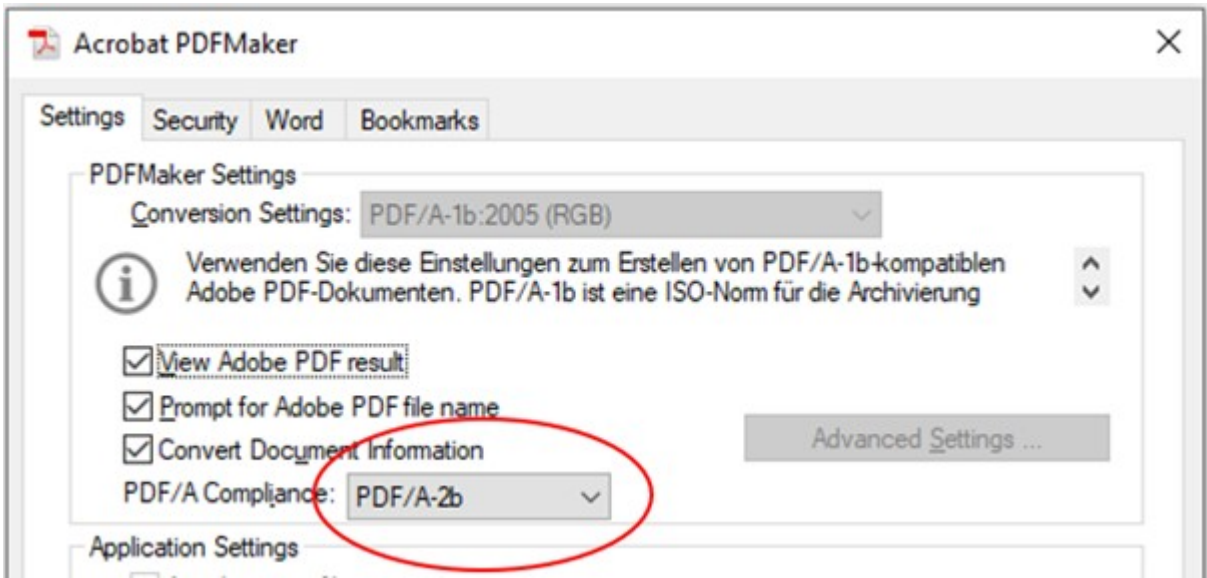
Structure

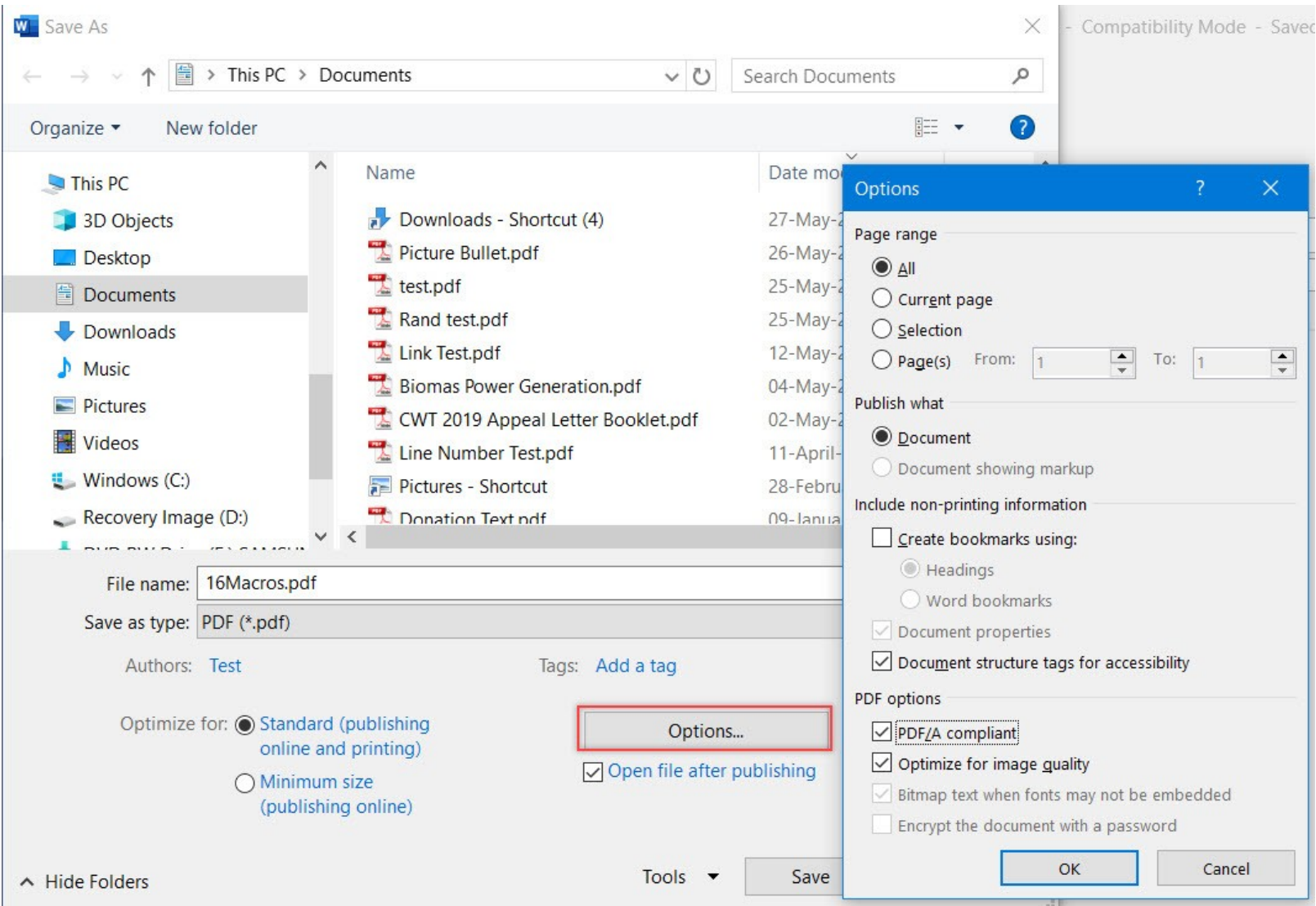
- Export bookmarks
- Export placeholders
- Comments as PDF annotations
- Export automatically inserted blank pages
- Use reference XObjects

Help

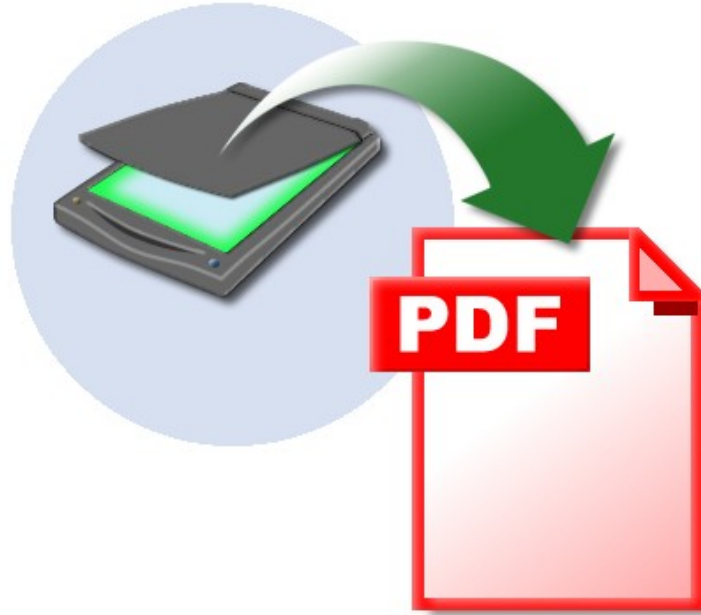
Cancel

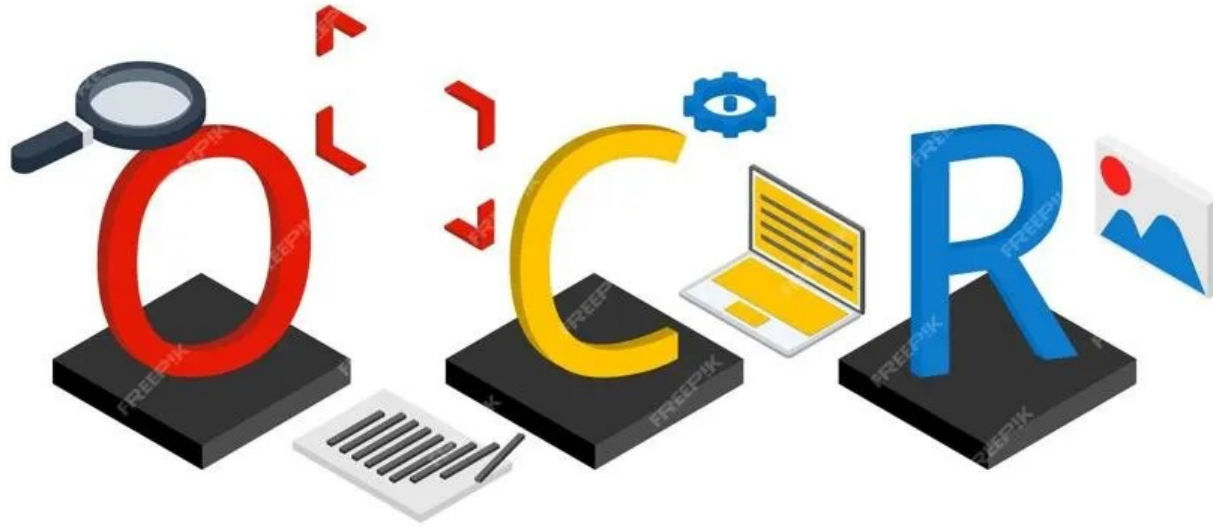
Export





Scan to PDF

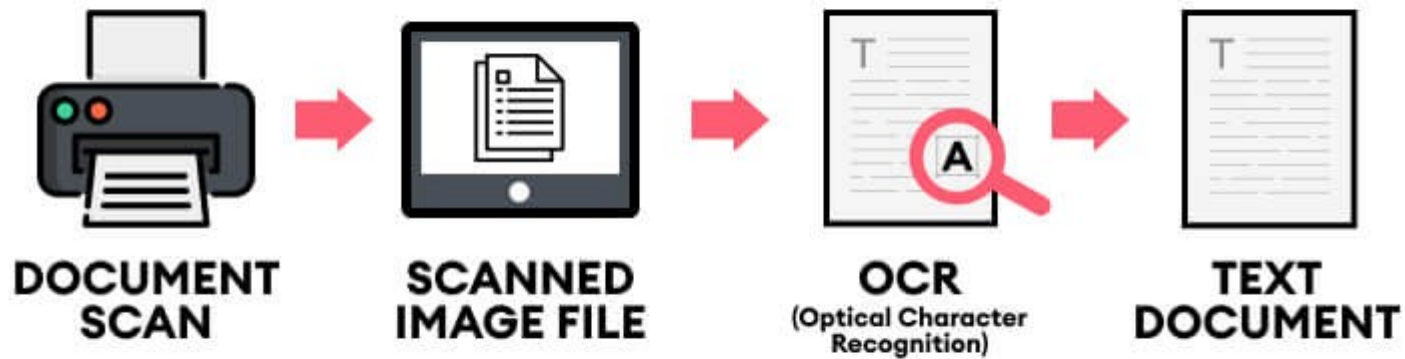




Optical


Character

Recognition



OCR - எழுத்துணரி


Scanned.pdf/png/tif/jpg



WHOLE FOODS MARKET - WESTPORT, CT 06880
399 POST RD WEST - (203) 227-6858

365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
	BROTH CHIC NP	2.19	F
	FLOUR ALMOND NP	1.99	F
CHKN	BRST BNLS SK NP	18.80	F
	HEAVY CREAM NP	3.39	F
	BALSVC REDUCT NP	6.49	F
	BEEF GRND 85/15 NP	5.04	F
	JUICE COF CASHEW C NP	8.99	F
	DOCS PINT ORGANIC NP	14.49	F
	HNY ALMOND BUTTER NP	9.99	F
****	TAX .00 BAL	101.33	

Searchable.pdf




WHOLE FOODS MARKET - WESTPORT, CT 06880
399 POST RD WEST - (203) 227-6858

365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
	BROTH CHIC NP	2.19	F
	FLOUR ALMOND NP	1.99	F
CHKN	BRST BNLS SK NP	18.80	F
	HEAVY CREAM NP	3.39	F
	BALSVC REDUCT NP	6.49	F
	BEEF GRND 85/15 NP	5.04	F
	JUICE COF CASHEW C NP	8.99	F
	DOCS PINT ORGANIC NP	14.49	F
	HNY ALMOND BUTTER NP	9.99	F
****	TAX .00 BAL	101.33	

Find (1/4) ✕

 ⚙️

Previous Next

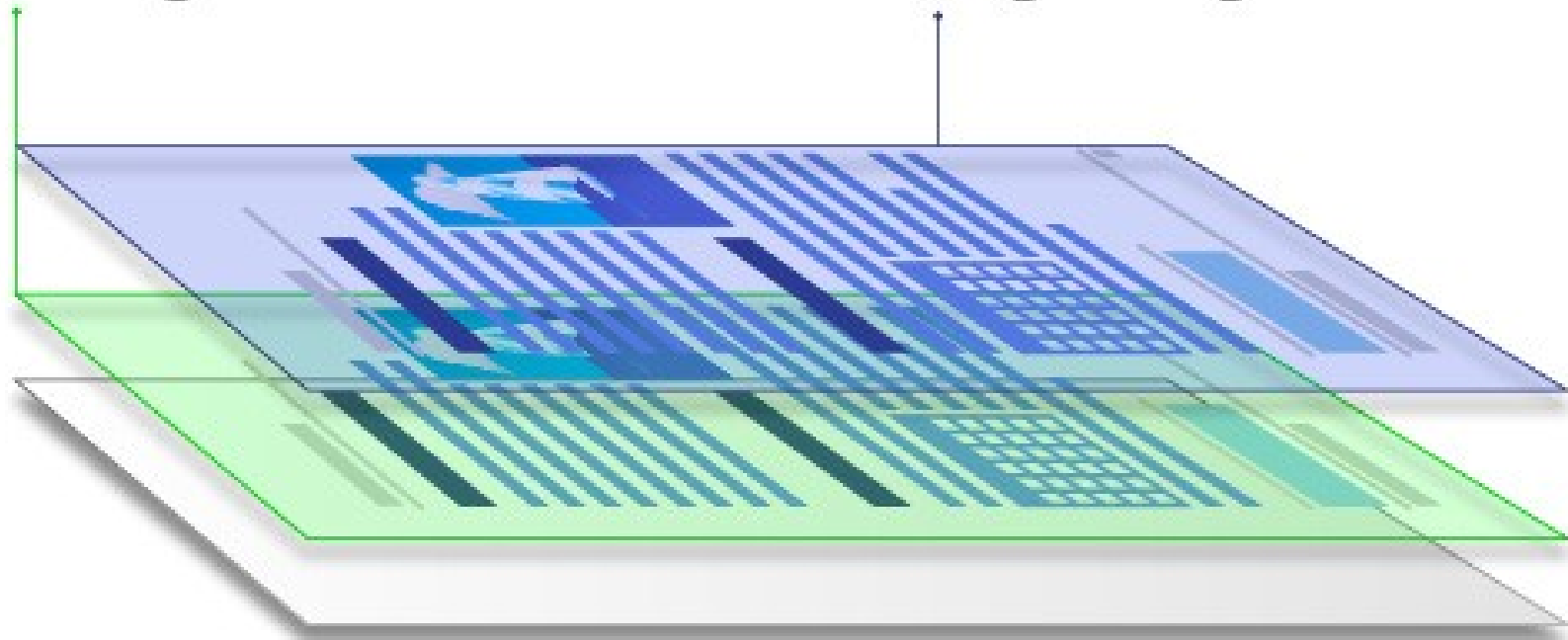


WHOLE FOODS MARKET - WESTPORT, CT 06880
399 POST RD WEST - (203) 227-6858

365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
365	BACON LS NP	4.99	F
	BROTH CHIC NP	2.19	F
	FLOUR ALMOND NP	1.99	F
CHKN	BRST BNLS SK NP	18.80	F
	HEAVY CREAM NP	3.39	F
	BALSVC REDUCT NP	6.49	F
	BEEF GRND 85/15 NP	5.04	F
	JUICE COF CASHEW C NP	8.99	F
	DOCS PINT ORGANIC NP	14.49	F
	HNY ALMOND BUTTER NP	9.99	F
****	TAX .00 BAL	101.33	

Recognized text

Page image

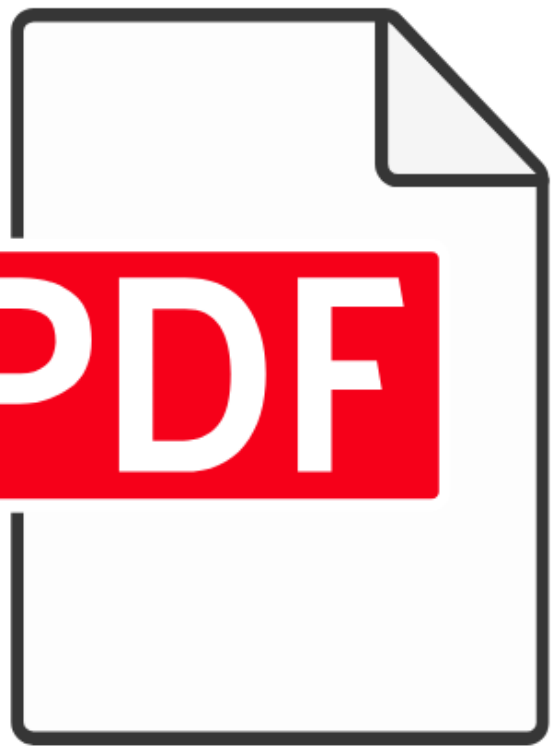


The logo for OCRmyPDF. The word "OCR" is in a large, bold, black sans-serif font, enclosed within a light blue rectangular box. A vertical blue line with circular endpoints at the top and bottom passes through the center of the "OCR" box. To the right of "OCR" is the word "my" in a smaller, black, lowercase sans-serif font.

OCRmy

A red rectangular box containing the letters "PDF" in a white, bold, sans-serif font. This box is positioned to the right of the "my" text and is partially overlapped by a document icon.

PDF





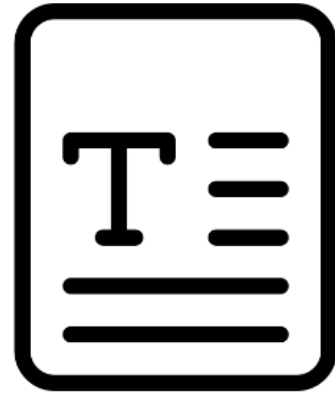
 Azure Pipelines succeeded  `pypl v9.2.0`  `homebrew v9.2.0`  `docs passing`  `python 3.6 | 3.7 | 3.8`

OCRmyPDF adds an OCR text layer to scanned PDF files, allowing them to be searched or copy-pasted.

Main features

- Generates a searchable [PDF/A](#) file from a regular PDF
- Places OCR text accurately below the image to ease copy / paste
- Keeps the exact resolution of the original embedded images
- When possible, inserts OCR information as a "lossless" operation without disrupting any other content
- Optimizes PDF images, often producing files smaller than the input file
- If requested deskews and/or cleans the image before performing OCR
- Validates input and output files
- Distributes work across all available CPU cores
- Uses [Tesseract OCR](#) engine to recognize more than [100 languages](#)
- Scales properly to handle files with thousands of pages
- Battle-tested on millions of PDFs

OCR on PDF files



Installing OCRmyPDF

pypi v16.4.2

The easiest way to install OCRmyPDF is to follow the steps for your operating system/platform. This version may be out of date, however.

These platforms have one-liner installs:

Debian, Ubuntu	<code>apt install ocrmypdf</code>
Windows Subsystem for Linux	<code>apt install ocrmypdf</code>
Fedora	<code>dnf install ocrmypdf tesseract-osd</code>
macOS (Homebrew)	<code>brew install ocrmypdf</code>
macOS (MacPorts)	<code>port install ocrmypdf</code>
LinuxBrew	<code>brew install ocrmypdf</code>
FreeBSD	<code>pkg install textproc/py-ocrmypdf</code>
Conda (WSL, macOS, Linux)	<code>conda install ocrmypdf</code>
Snap (snapcraft packaging)	<code>snap install ocrmypdf</code>

```
apt-get install tesseract-ocr-tam
```


Basic examples

Help!

ocrmypdf has built-in help.

```
ocrmypdf --help
```

Add an OCR layer and convert to PDF/A

```
ocrmypdf input.pdf output.pdf
```

Add an OCR layer and output a standard PDF

```
ocrmypdf --output-type pdf input.pdf output.pdf
```

Create a PDF/A with all color and grayscale images converted to JPEG

```
ocrmypdf --output-type pdfa --pdfa-image-compression jpeg input.pdf output.pdf
```

```
ocrmypdf -l tam+eng --output-type pdfa-3 input.pdf output.pdf
```

“தமிழ் யூனிகோடு அனைத்து எழுத்துரு தரப்பாடு”
(TACE 16)

பணிக்குழுவின் பரிந்துரைகள்

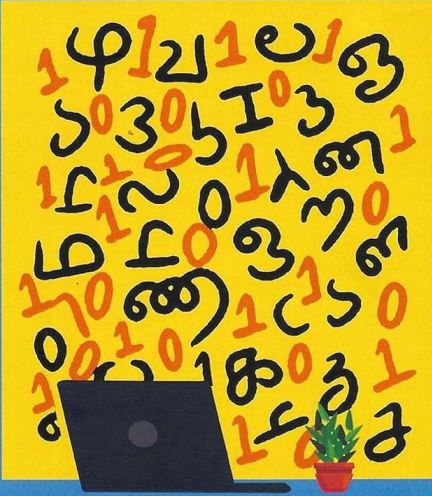
தமிழக அரசின் தகவல் தொழில்நுட்பத் துறைக்கு வழங்கப்படுகிறது

தமிழ் இணையப் பல்கலைக்கழகம்
Elnet மென்பொருள் நகரம், தரமணி, சென்னை.

மே 2008

ஒருங்குறியத்தமிழ்	ஒருங்குறிய அனைத்து எழுத்துரு தரப்பாடு(TACE-16)
85% தமிழ் எழுத்துக்களுக்கு இடமில்லை.	100% தமிழ் எழுத்துக்களுக்கும் இடமுண்டு.
பிற எழுத்துக்களைப் பெற, தனியொரு இடைமென்பொருள் தேவை	அத்தகைய இடைமென்பொருள் தேவையில்லை என்பதால், கணினியின் செயற்திறன் ஓங்கும்.
கட்டளைகள், இரண்டாம்தர நிலை செயற்பாட்டில் (Level - 2 Implementation) தான் செயற்படும்.	கட்டளைகள், முதல்தர நிலைச் செயற்பாட்டிலேயே (Level - 1 Implementation) செயற்படும்.
தரவுத்தேக்கமும், செயற்பாட்டு நேரமும் 40%அதிகம்.	தரவுத்தேக்கமும், செயற்பாட்டு நேரமும் வெகுக்குறைவு.
ISCI-1988 தரத்தை அடிப்படையாகக் கொண்டது. அதில் தமிழ் எழுத்துக்கள் இயற்கையான அகரவரிசையில் இல்லை. எனவே, அகரவரிசைப்படுத்த ஒத்திணைப்பு மென்பொருள்(collection Algorithm) தனியாகத்தேவை.	அகரவரிசையில் அமைந்துள்ளமையால், ஒத்திணைப்பு மென்மியம் தேவையில்லை.
ஒரு உயிர்மெய் எழுத்தை உருவாக்க, 2,3புள்ளிகள் தேவை. அதனால் பாதுகாப்புச் சிக்கல்களும்(security vulnerabilities), பொருள் மயக்கச்சேர்க்கையையும்(ambiguous combination) உண்டாக்கின்றன. இவற்றைச் சரிசெய்ய இயல்பாக்க மென்மியம் தேவை.	அத்தகையப் பாதுகாப்புச் சிக்கல்களும், பொருள்மயக்கச்சேர்க்கையையும் உண்டாவது இல்லை. அதனால் இயல்பாக்க மென்மியம் தேவையில்லை.
எழுத்துக்கள் பிரியாமல் இருக்கவும் அல்லது பிரிந்து இருக்கவும் கணினிக்கு சுடுதல் பணிச்சுமை ஏற்படுகிறது.(zero - width joiner/ zero - width non-joiner)	அத்தகையப் பணிச்சுமை ஏற்படுவதில்லை.
தவறான எழுத்துக்கள் உருவாக்கப்படாமல் இருக்க, விலக்கெழுத்து அட்டவணை தேவை.	விலக்கெழுத்து அட்டவணை தேவையில்லை.
தகவற்பரிமாற்றத்திற்கு மட்டுமே உதவும்.	தகவற்பரிமாற்றத்திற்கும், இன்னும்பிற மேம்பட்ட தமிழ்கணிமை ஆய்வுக்கும் , வளர்ச்சிக்கும் இன்றியமையாதது ஆகும்.
ஒருங்குறியச்சேர்த்திய விதிகளின் படி, எழுத்துக்கள்(characters)மட்டுமே குறியிடப்படல் வேண்டும். வரிவரைவுகள் (glyphs) குறியிடப்படக்கூடாது. எனவே, தற்போதுள்ள இந்த ஒருங்குறியத்தமிழ் ஒருங்குறியக்கூட்டமைப்பின் கோட்பாடுகளுக்கு முரணானது.	வரிவுரைவுகளோ அல்லது துணைக்குறியீடுகளோ உருவாவது இல்லை. அனைத்து எழுத்துக்களுக்கும் தனியிடம் கொடுக்கப்பட்டுள்ளது.
கோப்பைகளை சேமிப்பதில் அதிக இடம் தேவைப்படும்	கோப்பைகளை ஒருங்குறி தமிழ் குறியேற்றத்தை விட குறைந்த இடத்தைக் கொண்டு சேமிக்க முடியும்.
அதிக இணைய தகவல் பரிமாற்ற தேவை	ஒருங்குறி தமிழ் குறியேற்றத்தை விட குறைந்த இணைய தகவல் பரிமாற்றமே தேவை.
இந்த ஒருங்குறியத்தமிழ், பல இயக்கமென்மியங்களில் சரிவர செயற்படுவதில்லை.	இயக்கமென்மியங்களில் (system softwares) சரிவர செயற்படும்.

என்கோடு
உன்கோடு
யுனிகோடு
தனிகோடு



காசி ஆறுமுகம்

மேலும் படிக்க

<https://www.amazon.in/-/ta/dp/B08CTJDXP>

<https://blog.selvaraj.us/archives/29>

<https://blog.selvaraj.us/archives/253>

படங்கள் - மூலங்கள்

https://commons.wikimedia.org/wiki/File:Nixs_News_Tamil_Logo.png

https://commons.wikimedia.org/wiki/File:Tamil_palm-leaf_manuscript_of_Ci%E1%B9%9Fupa%C3%B1cam%C5%ABlam_2.jpg

https://commons.wikimedia.org/wiki/File:%E0%AE%A4%E0%AE%9E%E0%AF%8D%E0%AE%9A%E0%AF%88_%E0%AE%AA%E0%AF%86%E0%AE%B0%E0%AE%BF%E0%AE%AF_%E0%AE%95%E0%AF%8B%E0%AE%B5%E0%AE%BF%E0%AE%B2%E0%AF%8D_%E0%AE%95%E0%AE%B2%E0%AF%8D%E0%AE%B5%E0%AF%86%E0%AE%9F%E0%AF%8D%E0%AE%9F%E0%AF%81.jpg

https://www.worldhistory.org/Johannes_Gutenberg/#google_vignette

https://commons.wikimedia.org/wiki/File:Platen_Printing_press.jpg

<https://commons.wikimedia.org/wiki/File:Ascii-codes-table.png>

<https://web.archive.org/web/20070319230051/http://www.araichchi.net/kanini/unicode/t-unicode-mind-map.jpg>

<https://pdf.abbyy.com/blog/redacting-a-pdf-document/>

www.kalavipattil.com

