

Dated: September 14, 2003

Title:	Proposal to add TAMIL LETTER SHA
Source:	International Forum for Information Technology in Tamil (INFITT)
Action:	For consideration by UTC and ISO/IEC JTC 1/SC 2/WG 2
Distribution:	ISO/IEC JTC 1/SC 2/WG 2 members and Liaison organizations

1. Introduction

Currently Unicode Tamil range does not include the Tamil grantha[1] character SHA. This document proposes addition of this character in Unicode character set, within the current Tamil range U+0B80 to U+0BFF.

ISO/IEC JTC 1/SC 2/WG 2 Proposal summary form N2352-F accompanies this document.

1.1. Character proposed in this document

Shape of proposed character	Proposed name	Position
ஃ	TAMIL LETTER SHA	0BB6

1.2. About INFITT and INFITT WG02

International Forum for Information Technology in Tamil (INFITT) is a non-governmental organization, with cooperation from various state and national governments in the Tamil speaking countries. It also has participation from industry representatives and individuals with expertise in Information Technology and Tamil language.

Working Group 2 (WG02) of INFITT deals with definition and implementation of Unicode in Tamil. WG02 also has a Liaison relationship with Unicode Consortium. Members of this working group are derived from various nations around the world, and represent interests of IT industry, governments, Tamil scholars, educators and other Tamil IT enthusiasts. This proposal has been prepared with the consensus of members of this working group.

1.3. Demographics

Tamil is the state language of the state of Tamil Nadu in India. It is also a national language in Sri Lanka and Singapore (Source: CIA World Factbook - <http://www.cia.gov/cia/publications/factbook/index.html>).

Total population of Tamil speakers around the world is estimated to be around 74 million (1999 estimate) as per the Ethnologue index prepared by Summer Institute of Linguistics. See the Ethnologue page for Tamil http://www.ethnologue.com/show_language.asp?code=TCV which provides details about the population and distribution of Tamil diaspora.

2. TAMIL LETTER SHA

This document proposes to encode Tamil grantha character SHA in unicode Tamil range at U+0BB6.

2.1. Usage of Tamil letter SHA

Devanagari and other Indian scripts have a character named SHA in their Unicode block. However in Tamil the corresponding slot is left vacant. While not in popular use, there does exist a character for letter SHA in Tamil. This document proposes to add Tamil letter SHA in Unicode Tamil range at U+0BB6.

American Library Association-Library of Congress (ALA-LC) Romanization scheme includes this character in their list. (<http://lcweb.loc.gov/catdir/cpso/romanization/tamil.pdf>)

Digital South Asia Library of Univ of Chicago and other American Universities also use the same ALA-LC scheme that recognizes this "sa" as one of the grantha characters repertoire. (<http://www.lib.uchicago.edu/e/su/southasia/transliteration-tamil.html>)

See also "Transliteration of Indic scripts; How to use ISO 15919", (<http://homepage.ntlworld.com/stone-catend/trind.htm>) Table 3, shown as example 5 under section 3.

2.2. Data Processing for SHA

SHA behaves similar to other Tamil grantha letters, when combined with vowel signs. Specifically, U and UU vowels signs for SHA are similar to other grantha letters. In terms of other Unicode properties, the proposed character has properties similar as U+0BB8 (TAMIL LETTER SA). Property values are:

```
0BB6;TAMIL LETTER SHA;Lo;0;L;;;;;N;;;;;
```

SHA is usually sorted after the Tamil letter NNNA and before all the grantha characters. However, minor differences might exist in this behaviour between various sort orders.

Inclusion of character SHA in Tamil range will also help in transliteration from other Indic scripts into Tamil.

2.3. Ligature SRI[2]

While the base character SHA is not widely used in modern literature and communications, its ligature form SRI is used widely. The character sequence of SRI is as below:

```
Proposed TAMIL LETTER SHA      +  
TAMIL LETTER RA (U+0BB0)       +  
TAMIL VOWEL SIGN II (U+0BC0)
```

This usage corresponds with the sequence of SRI in other Indic scripts.

Current implementations of Unicode define SRI in Tamil as follows:

```
TAMIL LETTER SA (U+0BB8)       +  
TAMIL LETTER RA (U+0BB0)       +  
TAMIL VOWEL SIGN II (U+0BC0)
```

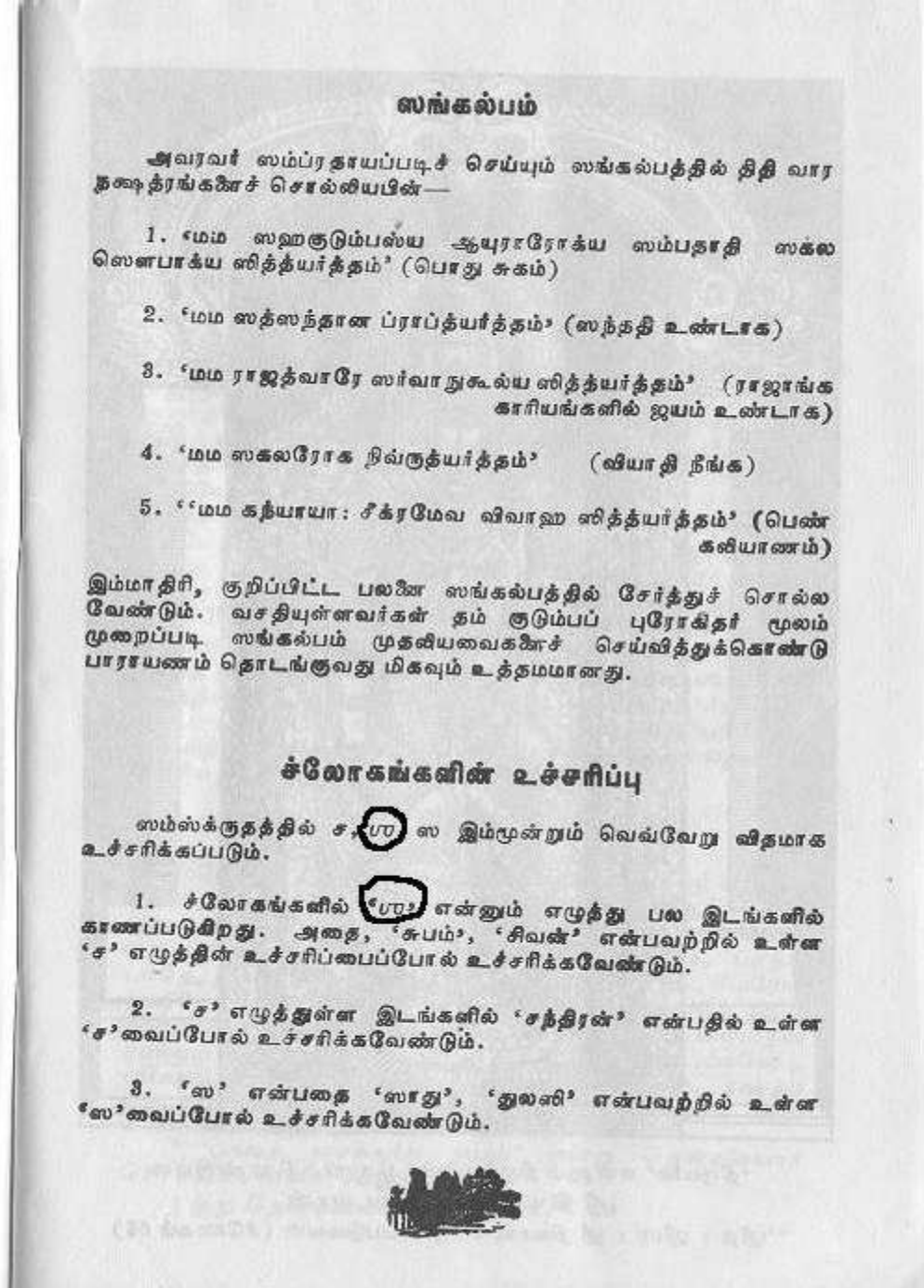
This is incorrect combination of SRI. This error is due to lack of TAMIL LETTER SHA in current Unicode repertoire.

Proposed character SHA may also form various other ligatures in combination with MA, YA, RA, and VA. However, these ligatures are archaic and are not widely recognized. Contemporary publications only use disjoint forms.

3. Published Usage Examples

Example 1: Pronunciation of SHA

This example explains the pronunciation of the proposed character SHA.



Example 2: Transliteration of SHA

This example gives the transliteration of proposed character SHA in Devanagari.

தாள கால அடையாளங்களின் விபரங்கள்		ஸம்ஸ்கிருத தெலுங்கு உச்சரிப்புக்களை காட்டும் எண்கள்	
காலங்கள்	அகாரம்		
1. ஸா ;	(1)	क क	ख ख
2. ஸா	(१)	ख क _२	न न
3. ஸ	(१)	ग क _३	प प
4. ஸ	(१)	घ क _४	फ प _२
5. ஸ	(१)	ङ ङ	ब प _३
		च च	भ प _४
		छ छ	म म
		ज ज	य य
		झ झ	र र
		ञ ञ	ल ल
		ट ट	व व
		ड ड	श शा
		ढ ढ	ष ष
		ण ण	स स
		त त	ह ह
		थ थ	ळ ळ
		द द	ळ ळ
			झ झ

இந்த கமா எந்தக் காலத்தில் உபயோகப்படுகிறதோ, அந்தக் காலத்தில் பாதியளவாகும்.

.....இந்தப் புள்ளிகள் ஸாஹித்யங்களில் வரும் அகாராதிக்கு அடையாளம்.

இது கமகமாகப் பாடும் அடையாளம்.

(ச = ரி) சுத்த ரிஷபம்	(ச = நை) சுத்த நைவதம்
(ச = ரி) *சதுச்சுருதி ரிஷபம்	(ச = நை) சதுச்சுருதி நைவதம்.
(ஷ = ரி) ஷட்சுருதி ரிஷபம்	(ஷ = நை) ஷட்சுருதி நைவதம்
(ச = கா) சுத்த காந்தாரம்	(ச = நி) சுத்த நிஷாதம்
(ஸா = கா) ஸாதாரண காந்தாரம்	(கை = நி) கைகிளி நிஷாதம்.
(அ = கா) அந்தர காந்தாரம்	(கா = நி) காசலி நிஷாதம்
(சு = ம) சுத்த மத்யமம்	ஆ: ஆரோஹணம்.
(ப்ரதி = ம) ப்ரதி மத்யமம்	அவ: அவரோஹணம்.

N.B.—*வேங்கடமகிழவர்கள் ஸம்ப்ரதாயப்படி சதுச்சுருதிரிஷபத்தை பஞ்சச்சுருதி ரிஷபம் சதுச்சுருதி நைவத்தை பஞ்சச்சுருதி நைவதமென்றும் சொல்லவேண்டும்.

First Edition	... 1950
Second Edition	... 1957
Third Edition	... 1967
Fourth Edition	... 1977

தி இண்டியா பிரிண்டிங் ஒர்க்ஸ், மைலாப்பூர், சென்னை-600004

Example 3: Shankarar

Name of Shankarar is spelt with this proposed SHA. The trasnliteration of the verse also uses SHA.

2. ஸ்ரீஸ்ரங்கரரின் விநயபாவம்-
ஸுப்ரஹ்மண்யோபாஸநம்
மஹாபாண்டித்யத்தை அடையவேறுது

न जानामि शब्दं न जानामि चार्थं
न जानामि पद्यं न जानामि गद्यम् ।
चिदेका षडास्या हृदि द्योतते मे
मुखान्निस्सरन्ते गिरश्चापि चित्रम् ॥ २ ॥

ந ஜாநாமிஸ்ரப்தம் ந ஜாநாமி சார்த்தம்
ந ஜாநாமி பத்யம் ந ஜாநாமி கத்யம்
சிதேகா ஷடாஸ்யா ஹ்ருதி த்யோததே மே
முகாந்நிஸ்ஸரந்தே கிரஸ்சாபி சித்ரம்

INSPIRATIONS OF ACHARYA

I know not Sabda, I know not Artha,
I know not Padyam, I know not Gadyam.
But there shines in My Hrudayam six faced luminous light,
Brilliant words flow out from my mouth extempore

நான் சப்தத்தை அறிந்தவனல்ல; பொருளையும் அறிந்த
வனல்ல; நான்குவிதமான பதங்களோடுகூடின ச்லோகத்
தையும் அறிந்தவனல்ல; சூர்ணிகைகளையும் அறிந்தவ
னல்ல; என்னுடைய மனதில் ஒரு ஒளியானது ஆறுமுகத்
தோடு கூடினதாய்க்கொண்டு ப்ரகாசித்துக்கொண்டிருக்க்
றது. அக்காரணத்தால் என்னுடைய முகத்தினின்றும் ஆ
சர்யமான வாக் விலாஸங்களும், வெளியில் வந்து ப்ரக
சிக்கின்றன.

Example 4: Vishnu sahasranaamam

First verse of Vishnu sahasranaamam uses SHA in several places.

ஓம்

ஸ்ரீ விஷ்ணவே பரமாத்மநே நம :

ஸ்ரீ விஷ்ணு ஸஹஸ்ரநாம ஸ்தோத்ரம்

1*

ஹரி : ஓம்

ஸூக்லாம்பரதரம் விஷ்ணும் ஸாசிவீர்ணம் சதுர்ப்புஜம் |
ப்ரஸந்நவதநம் த்யாயேத் ஸர்வவிக்நோபஸாந்தயே || 1

யஸ்ய த்விரதவக்த்ராத்யா : பாரிஷத்யா : பரஸ்ஸாதம் |
விக்நம் நிக்நந்தி ஸததம் விஷ்வக்ஸேநம் தமாஸரயே || 2

வ்யாஸம் வஸிஷ்டநப்தாரம் ஸக்தே : பௌத்ரமகல்மஷம் |
பராஸராத்தமஜம் வந்தே ஸூகதாதம் தபோநிதிம் || 3

வ்யாஸாய விஷ்ணுரூபாய வ்யாஸரூபாய விஷ்ணவே |
நமோ வை ப்ரஹ்மநிதயே வாஸிஷ்டாய நமோ நம: || 4

அவிகாராய ஸூத்தாய நித்யாய பரமாத்மநே |
ஸதைகரூபரூபாய விஷ்ணவே ஸர்வ ஜிஷ்ணவே || 5

யஸ்ய ஸ்மரணமாத்ரேண ஜன்மஸம்ஸாரபந்தநாத் |
விமுச்யதே நமஸ்தஸமை விஷ்ணவே ப்ரபவிஷ்ணவே || 6

ஓம் நமோ விஷ்ணவே ப்ரபவிஷ்ணவே ||

* இந்த 1-வது பகுதியிலுள்ள சீலோகங்கள் மஹாபாரதத்திலுள்ள ஸஹஸ்ரநாமப் பகுதியில் இல்லை. பாஷ்யகாரர்களும் இவற்றிற்கு அர்த்தம் எழுதவில்லை. ஆயினும் வெகுசாமமாகப் பெரியோர்களால் பாராயணம் செய்யப்படுகின்றன.

Example 5: Transliteration of Indic Scripts: How to use ISO 15919

Transliteration of proposed character SHA is shown as "s<acute>a" in the table below.

Latin	Dev.	Gur.	Guj.	Ben.	Ori.	Tam.	Mal.	Kan.	Tel.	Sin.	Notes
pa	प	ਪ	પ	প	ପ	ப	പ	ಪ	ప	ପ	
pha	फ	ਫ	ફ	ফ	ଫ		ഫ	ಫ	ఫ	ଫ	
ba	ब	ਬ	બ	ব	ବ		ബ	ಬ	బ	ବ	
bha	भ	ਭ	ભ	ভ	ଭ		ബ	ಬ	బ	ଭ	
ma	म	ਮ	મ	ম	ମ	ம	മ	ಮ	మ	ମ	
m̄ba										Ṁ	9.
ra	(र)					ற	റ	(ര)	(ర)	(ര.)	10.
ra							(ര)				
na	(न)					ണ	(ണ)			(ന.)	
la	(ळ)					ഴ	ഴ	(ല)	(ల)	(ല.)	
ya	य	ਯ	ય	য	ଯ	ய	യ	ಯ	య	ଯ	
yā	(य़)			য়	ୟ						11.
ra	र	ਰ	ર	র(ব)	ର	ர	റ	ര	ర	ର	
r̄	=										
la	ल	ਲ	લ	ল	ଲ	ல	ല	ല	ల	ল	
ḷa	ळ	ਲ	ਲ		ଲ	ள	ള	ഏ	ఱ	ঐ	
va	व	ਵ	વ	(ব)	(ব)	வ	വ	വ	వ	ব	
śa	श	ਸ	શ	শ	ଶ	(ശ)	ശ	ഷ	ష	শ	
ṣa	ष		ષ	ষ	ষ	(ഷ)	ഷ	ജ	ష	ষ	
sa	स	ਸ	સ	স	স	(ஸ)	സ	സ	స	স	
ha	ह	ਹ	હ	হ	ହ	(ஹ)	ഹ	ജ	హ	হ	
'	ऽ		ટ	৳	୳		ട	ഃ	ఱ		
qa	क	ਕ	ક	ক							
kha	ख	ਖ	ખ	খ							
ḡa	ग	ਗ	ગ	গ							
za	ज	ਜ	જ	জ				ജ	జ		
fa	फ	ਫ	ફ	ফ				ജ		ଫ	
wa				ব	৳						

Example 6: Ligature SRI

This wedding invitation contains the letter SRI (also transliterated in English as SHRI or SHREE).



Notes:

[1] Grantha

Grantha characters are part of Tamil script. Historically these characters are not considered to be part of proper Tamil Language, but are used to write non-Tamil words and names. These were first introduced by rulers of Pallava dynasty during 6-7th centuries A.D. These characters are widely used by Tamils in Indian subcontinent and the Tamil diaspora in other countries.

[2] Sri

Also written as SHRI or SHREE. Sri in Sanskrit means wealth. Hindu goddess of wealth is called "SRIDEVI". The prefix SRI is often added to many gods and goddesses and also while formally addressing elders.

ISO/IEC JTC 1/SC 2/WG 2

**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646**

Please fill all the sections A, B and C below.

(Please read Principles and Procedures Document for guidelines and details before filling this form.)
See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html> for latest Form.
See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for latest Principles and Procedures document.
See <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest roadmaps.

(Form number: N2352-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09))

A. Administrative

1. Title: Addition of Tamil character SHA
2. Requester's name: International Forum for Information Technology
 in Tamil (INFITT)
3. Requester type (Member body/Liaison/Individual contribution): Liaison
4. Submission date: 2003-03-10
5. Requester's reference (if applicable): WG02 SHA Proposal
6. (Choose one of the following:)
- This is a complete proposal: YES
- or, More information will be provided later:

B. Technical - General

1. (Choose one of the following:)
- a. This proposal is for a new script (set of characters): NO
Proposed name of script:
- b. The proposal is for addition of character(s) to an existing block: YES
Name of the existing block: Tamil Range: 0B80-0BFF
2. Number of characters in proposal: 1
3. Proposed category (see section II, Character Categories): A
4. Proposed Level of Implementation (1, 2 or 3)
(see clause 14, ISO/IEC 10646-1: 2000): 1
- Is a rationale provided for the choice? YES
- If Yes, reference: See attached proposal
5. Is a repertoire including character names provided? YES
- a. If Yes, are the names in accordance with the 'character naming
guidelines in Annex L of ISO/IEC 10646-1: 2000? YES
- b. Are the character shapes attached in a legible form suitable for review?
 YES
6. Who will provide the appropriate computerized font (ordered preference:
True Type, or PostScript format) for publishing the standard?
Mr.Muthu Nedumaran (muthu@murasu.com)
If available now, identify source(s) for the font (include address,
e-mail, ftp-site, etc.) and indicate the tools used:
7. References:
- a. Are references (to other character sets, dictionaries, descriptive
texts etc.) provided? YES
- b. Are published examples of use (such as samples from newspapers,
magazines, or other sources) of proposed characters attached? YES
8. Special encoding issues:
Does the proposal address other aspects of character data processing

(if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

_____ YES. See attached proposal. _____

9. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script.

Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database

<http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? _NO_
If YES explain _____
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? _____ YES _____
If YES, with whom? _INFITT WG02_
If YES, available relevant documents: _See section 1.2 of attached_
proposal
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? _____ YES _____
Reference: _See section 1.3 of attached proposal_
4. The context of use for the proposed characters (type of use; common or rare) _____ Common _____
Reference: _See attached proposal_
5. Are the proposed characters in current use by the user community? _YES_
If YES, where? Reference: _See attached proposal_
6. After giving due considerations to the principles in *Principles and Procedures document* (a WG 2 standing document) must the proposed characters be entirely in the BMP? _____ YES _____
If YES, is a rationale provided? _____ YES _____
If YES, reference: _See section 2.1 of attached proposal_
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? _____ YES _____
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? _____ NO _____
If YES, is a rationale for its inclusion provided? _____
If YES, reference: _____
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? _NO_
If YES, is a rationale for its inclusion provided? _____
If YES, reference: _____
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? _____ NO _____
If YES, is a rationale for its inclusion provided? _____
If YES, reference: _____
11. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? _____ YES _____
If YES, is a rationale for such use provided? _____ YES _____
If YES, reference: _See attached proposal section 2.3_
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? _____ YES _____
If YES, reference: _See section 2.3 and example 6_
12. Does the proposal contain characters with any special properties such as

control function or similar semantics? _____ NO _____
 If YES, describe in detail (include attachment if necessary) _____
 13. Does the proposal contain any Ideographic compatibility character(s)? _NO
 If YES, is the equivalent corresponding unified ideographic character(s)
 identified? _____
 If YES, reference: _____

A.1 Submitter's Responsibilities

The national body or liaison organization (or any other organization or an individual) proposing new character(s) or a new script shall provide:

1. Proposed category for the script or character(s), character name(s), and description of usage.
2. Justification for the category and name(s).
3. A representative glyph(s) image on paper:
 If the proposed glyph image is similar to a glyph image of a previously encoded ISO/IEC 10646 character, then additional justification for encoding the new character shall be provided.
Note: Any proposal that suggests that one or more of such variant forms is actually a distinct character requiring separate encoding, should provide detailed, printed evidence that there is actual, contrastive use of the variant form(s). It is insufficient for a proposal to claim a requirement to encode as characters in the Standard, glyphic forms which happen to occur in another character encoding that did not follow the Character-Glyph Model that guides the choice of appropriate characters for encoding in ISO/IEC 10646.
Note: WG 2 has resolved in Resolution M38.12 not to add any more Arabic presentation forms to the standard and suggests users to employ appropriate input methods, rendering and font technologies to meet the user requirements.
4. Mappings to accepted sources, for example, other standards, dictionaries, accessible published materials.
5. Computerized/camera-ready font:
 Prior to the preparation of the final text of the next amendment or version of the standard a suitable computerized font (camera-ready font) will be needed. Camera-ready copy is mandatory for final text of any pDAMs before the next revision. Ordered preference of the fonts is True Type or PostScript format. The minimum design resolution for the font is 96 by 96 dots matrix, for presentation at or near 22 points in print size.
6. List of all the parties consulted.
7. Equivalent glyph images:
 If the submission intends using composite sequences of proposed or existing combining and non-combining characters, a list consisting of each composite sequence and its corresponding glyph image shall be provided to better understand the intended use.
8. Compatibility equivalents:
 If the submission includes compatibility ideographic characters, identify the equivalent unified CJK Ideograph character(s).
9. Any additional information that will assist in correct understanding of the different characteristics and linguistic processing of the proposed character(s) or script.