

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹

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

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	Additional Arabic letters needed for Arwi Script		
2. Requester's name:	Seyed Mohamed Buhari, University Brunei Darussalam, Brunei Darussalam		
3. Requester type (Member body/Liaison/Individual contribution):	Individual Contribution		
4. Submission date:	15 th April 2009		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:			
(or) More information will be provided later:	Yes		

B. Technical – General

1. Choose one of the following:			
a. This proposal is for a new script (set of characters):			
Proposed name of script:			
b. The proposal is for addition of character(s) to an existing block:	Yes		
Name of the existing block:	Arabic		
2. Number of characters in proposal:	3		
3. Proposed category (select one from below - see section 2.2 of P&P document):			
A-Contemporary	<input type="checkbox"/>	B.1-Specialized (small collection)	<input checked="" type="checkbox"/>
C-Major extinct	<input type="checkbox"/>	D-Attested extinct	<input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input type="checkbox"/>	E-Minor extinct	<input type="checkbox"/>
G-Obscure or questionable usage symbols	<input type="checkbox"/>		
4. Is a repertoire including character names provided?			
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?			
b. Are the character shapes attached in a legible form suitable for review?	Yes		
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?			
Seyed Mohamed Buhari, University Brunei Darussalam, Brunei Darussalam			
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:			
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	YES		
7. 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)?			
NO			

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

¹ Form number: N3152-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05)

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?	<i>Roozbeh Pournader, Book Publisher (Shahul Hameed & Sons, Chennai, India)</i>
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	YES
Reference:	
4. The context of use for the proposed characters (type of use; common or rare)	common
Reference:	
5. Are the proposed characters in current use by the user community?	YES
If YES, where? Reference:	<i>Certain Religious books are still printed in Arwi</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	YES
If YES, is a rationale provided?	YES
If YES, reference:	<i>Only few characters are needed to be added</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	NO
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?	YES
If YES, is a rationale for its inclusion provided?	YES
If YES, reference:	<i>These are characters by itself and not diacritics added to a character</i>
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?	NO
If YES, is a rationale for such use provided?	
If YES, reference:	
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference:	
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:	

ARWI


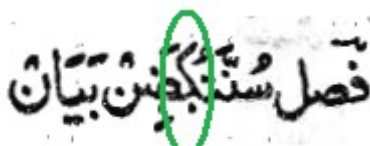
This proposal contains 2 Arwi (called as ArabuTamil) characters and one diacritic. Further research has to be done on the provided possible ligature. Arwi script is used to write Tamil Language using Arabic Script, with the inclusion of few extra characters and diacritics. Arwi script is used by Tamil speaking Muslim communities in Tamil Nadu, Malaysia, Singapore, etc. Books written using Arwi script on various topics like History, Religion, etc are being in use in Tamil Nadu.

Most of the characters needed for Arwi¹ are already present in the Arabic range of Unicode characters. This proposal includes the characters and diacritics that ought to be included into the Unicode Arabic range. These characters cannot be considered as a glyph variant of the existing characters in Arabic.



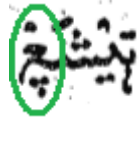
Bibliography

1. How to read and write Tamil characters [Accessed on: 22nd April 2008]: <http://www.xs4all.nl/~wjsn/tamil.htm> and History of Writing [Accessed on: 22nd April 2008]: <http://www.xs4all.nl/~wjsn/tekst/taalschriften.htm#QXQ>
2. Arwi language – Wikipedia, the free encyclopedia [Accessed on: 22nd April 2008]: http://en.wikipedia.org/wiki/Arwi_language
3. Tschacher. T., "How to die before dying? Sharia and Sufism in a 19th century Arabic-Tamil Poem", Panel 38, 18th European Conference on Modern South Asian Studies , at Lund, Sweden, 6 – 9 July 2004. <http://www.sasnet.lu.se/panelabstracts/38.html>
4. Shuayb Alim, "Arabic, Arwi and Persian in Sarandib and Tamil Nadu", Madras, 1993.
5. Sri Lanka – Ethnic Groups [Accessed on: 22nd April 2008]: <http://countrystudies.us/sri-lanka/38.htm>
6. Arwi [Accessed on: 22nd April 2008]: <http://www.armu.com/armu/works/archives/12dec1998/amc1.html>
7. Tschacher, Arwi (Arabic-Tamil) – An Introduction [Accessed on: 23rd April 2008]: <http://web.archive.org/web/20040822180630/www.fas.nus.edu.sg/journal/kolam/vols/kolam5&6/1AOldLit/Arwi.htm>
8. Nuhman MA, "Sri Lankan Muslims: Ethnic Identity within Cultural Diversity", International Centre For Ethnic Studies, Colombo, Sri Lanka, 2007.
9. Mohan V., "Muslims of Sri Lanka", Aalekh Publishers, Jaipur, India, 1985.
10. Arwi Books: Simthu Sibyaan, Adaabu Maalai, Thakkasuruth, etc.

KAF (0643) with DOT BELOW

Final Form	Middle Form
	

AIN (0639) with THREE DOTS BELOW

Initial Form	Middle Form	Final Form
		

Inverted DAMMA (064F) Below

Usage: This inverted Damma is used similar to Fatha or Damma (064B-0652). As an example, if inverted Damma is used with ARABIC TEH it becomes equivalent to the Tamil representation of TAEH.

حَيْفَبَاثُ وَيَشْنِي

Suggested code positions and complete names:

U+065F ARABIC INVERTED DAMMA BELOW

U+08A0 ARABIC LETTER AIN WITH THREE DOTS BELOW

U+08A1 ARABIC LETTER KAF WITH DOT BELOW

Information for ArabicShaping.txt:

08A0; AIN WITH 3 DOTS BELOW; D; AIN

08A1; KAF WITH DOT BELOW; D; KAF

Information for UnicodeData.txt:

065F;ARABIC INVERTED DAMMA BELOW;Mn;220;NSM;;;;;N;;;;;

08A0;ARABIC LETTER AIN WITH THREE DOTS BELOW;Lo;0;AL;;;;;N;;;;;

08A1;ARABIC LETTER KAF WITH DOT BELOW;Lo;0;AL;;;;;N;;;;;

Characters/diacritics Generated:

KAF (0643) with DOT BELOW (Four Forms)

ك كس كل س ك

AIN (0639) with THREE DOTS BELOW (Four Forms)

ع عس عل س ع

Inverted DAMMA (064F) Below

لس

Further Research: As per the author, it seems that further research ought to be done on the possible diacritic (Arabic Maddah + Fatha) and character (Arabic SAD with DOT below). I could find samples for the diacritic but not for the character.

Diacritic – Similar to 0653 Arabic Maddah Above + 064E Fatha

Usage: This diacritic is used to indicate the joint pronunciation of Maddah with Fatha. Sometimes, instead of using this diacritic, people just use Maddah with a Fatha above. This could be considered as a ligature.

شَدَّ نَكْتُ نَوَجْمَر

Diacritic – Similar to 0653 Arabic Maddah Above + 064E Fatha

لس

Reference: A mapping between Tamil and Arwi Scripts

அ	ا	க	ك or ك dot below	ய	ي	ஃ	ه
ஆ	ا	ங	غ	ர	ر	ஔ	و
இ	ا	ச	ج	ற	ر	ெ	ه
ஈ	اي	ஐ	ح	ல	ل	ே	ه
உ	ا	ஞ	غ or ح	ள	ض or ص with dot below	ை	ه
ஊ	أو	ட	د	ழ	ض	ொ	ه
எ	ه or ه	ண	ح	வ	و	ோ	و
ஏ	ه or ه	த	د or ث	ஷ	ص	ெள	و
ஐ	اي	ந	ن	ஸ	ش	்	ه
ஒ	ا	ன	ن	ஹ	ح	ள	
ஓ	أو	ப	ب	ா	ا	ஃ	
ஔ		ம	م	ி	ه	து	ت
				ீ	ه	ச்ச	ج
						ட்ட	د