

DOC TYPE: Expert contribution
TITLE: **Proposal to add characters for use with Thaana Script (Divehi language)**
SOURCE: Paul Nelson (Redmond, WA, USA), Paul Beam (Canada), Mohamed Abdul Gadir (Maldives)
PROJECT:
STATUS: Proposal
ACTION ID: FYI
DUE DATE: --
DISTRIBUTION: Worldwide
MEDIUM: Paper and web
NO. OF PAGES:

A. Administrative

1. Title	Proposal to add characters to complete Thaana/Divehi support
2. Requester's name	Paul Nelson (Redmond, WA, USA), Paul Beam (Canada), Mohamed Abdul Gadir (Maldives)
3. Requester type	Expert request.
4. Submission date	2000-04-07
5. Requester's reference	
6a. Completion	This is a complete proposal.
6b. More information to be provided?	Only as required for clarification.

B. Technical – General

1a. New script? Name?	No.
1b. Addition of characters to existing block? Name?	Yes. Thaana, Currency Symbols, and Arabic Presentation Forms-A
2. Number of characters	5
3. Proposed category	
4. Proposed level of implementation and rationale	

5a. Character names included in proposal?	Yes.
5b. Character names in accordance with guidelines?	Yes.
5c. Character shapes reviewable?	Yes.
6a. Who will provide computerized font?	Paul Nelson/Paul Beam.
6b. Font currently available?	Paul Nelson/Paul Beam.
6c. Font format?	OpenType.
7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?	Yes.
7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?	Yes.
8. Does the proposal address other aspects of character data processing?	

C. Technical – Justification	
1. Contact with the user community?	Yes. Paul Beam and Gadir have worked with computerization of Thaana for many years.
2. Information on the user community?	Native.
3a. The context of use for the proposed characters?	The Unicode 3.0 inclusion of Thaana missed characters that are commonly used by the Divehi language in the Maldives
3b. Reference	
4a. Proposed characters in current use?	Yes.
4b. Where?	Native speakers in the Maldives and worldwide.
5a. Characters should be encoded entirely in BMP?	Already in BMP and in accordance with Roadmap.
5b. Rationale	
6. Should characters be kept in a continuous range?	Yes. This greatly facilitates computational usage.
7a. Can the characters be considered a presentation form of an existing character or	No.

character sequence?	
7b. Where?	
7c. Reference	
8a. Can any of the characters be considered to be similar (in appearance or function) to an existing character?	No.
8b. Where?	
8c. Reference	
9a. Combining characters or use of composite sequences included?	Yes.
9b. List of composite sequences and their corresponding glyph images provided?	Yes.
10. Characters with any special properties such as control function, etc. included?	No.

D. SC2/WG2 Administrative	
To be completed by SC2/WG2	
1. Relevant SC 2/WG 2 document numbers:	
2. Status (list of meeting number and corresponding action or disposition)	
3. Additional contact to user communities, liaison organizations etc.	
4. Assigned category and assigned priority/time frame	
Other Comments	

Divehi has several commonly used characters that were not encoded with the initial group for Thaana support. These characters are currently found on Divehi typewriters and with MSL software that supports Divehi. The Maldivian National Centre for Linguistic and Historical Research requested the THAANA REYTU and RUFIIYAA SIGN.

The characters we are proposing to add to the Unicode Standard are listed below by category type. We have also included a proposed location for the placement of these characters.

Currency Symbol

This is the standard symbol for the Maldivian currency. It is actually a ligature of THAANA RAA + FULL STOP.

Suggested Unicode	Glyph	Name	Example
20B0	ރ	RUFUYAA SIGN	


Thaana Block Additions - Divehi Symbols and Letter

The THAANA NAA is used in the southern dialect spoken by people of Addu Atoll. The people from Addu Atoll are respected because they comprise a large percentage of the merchants in Maldives and so have much of its wealth. They are, however, also looked down upon because they are not from Malé, and because they tried to secede to form their own country. Of the four Dhivehi dialects, the only one officially written is the Malé dialect (the dialect of the ruling class); the Addu dialect is the only Dhivehi dialect having a character not found in the Malé dialect: that being the THAANA NAA character. Although the Addu dialect is not officially written (even if you ask people from Addu), in fact it is used somewhat. We feel it important not to overlook the THAANA NAA letter for social or political reasons, but rather to recognize it so it may be written.

The THAANA REYTU is used in Divehi to represent the COMMERCIAL AT, or rate, symbol in commerce. For example, 10 fish @ \$1.50 = \$15.00. This symbol is not in wide use yet, but was introduced by the Linguistic Centre that desires it to become the norm.

The DIVEHI NISHAAN is actually the seal of the Maldives. It is frequently used in correspondence. One can argue that this type of glyph could be done through PUA. However, given that this is found in normal correspondence with use of the Thaana script, it would be beneficial for plain text documents to render the correct glyph...without the requirement of a font that has the correct PUA character.

Suggested Unicode	Glyph	Name	Example
07B1	ނ	THAANA NAA	See Naa sample below

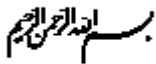
07BE		THAANA REYTU	
07BF		DIVEHI NISHAAN	See sample letter shown below

Ligatures

This ligature is commonly used in correspondence and publications. As the Maldives is a Muslim nation, the ‘Bismi’ is normally found at the top of letterhead. Current typewriters and MLS software allows the user to enter the ‘Bismi’ as one character.

While the policy of not adding additional Arabic presentation forms to the standard is important, this has a justifiable reason for being added. First, the Thaana script does not contain any Arabic characters. Yet, the ‘Bismi’ and ‘Allah’ Arabic presentation forms are frequently used. Second, as stated in the previous paragraph, all current typewriters and non-Unicode software uses the ‘Bismi’ and ‘Allah’ as one keystroke each. ‘Allah’ is currently encoded at U+FD2.

The alternatives of not adding the ‘Bismi’ to Unicode/ISO-10646 are to either force the user to place this commonly used character in PUA and loose correct plain text handling, or force all users install an Arabic keyboard and learn how to input Arabic characters.

Suggested Unicode	Glyph	Name	Example
FDFD		ARABIC LIGATURE BISMALLAH = <isolated> + 0628 + 0633 + 0645 + 0020 + 0627 + 0644 + 0644 + 0647 + 0020 + 0627 + 0644 + 0631 + 062D + 0645 + 0646 + 0020 + 0627 + 0644 + 0631 + 062D + 064a + 0645	See sample letter shown below

References

Sample letter showing ‘Bismi’ (ARABIC LIGATURE BISMALLAH) and ‘Nishaan’.

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ



ދިވެހިރާއްޖޭގެ ޖުމްހޫރިއްޔާ ޖެހިގެން ދެންނެވި ޖެހިގެން

މަޢުލޫމާތު

ދެންނެވި ޖެހިގެން.

"اللهم صل على محمد وآل محمد وصلى على خير خلقك محمد بن عبد

الله وآله وصحبه وسلم من بعدك يا ذا الجلال والإكرام اللهم

صلى على خير خلقك محمد بن عبد الله وآله وصحبه وسلم

اللهم صل على خير خلقك محمد بن عبد الله وآله وصحبه وسلم

اللهم صل على خير خلقك محمد بن عبد الله وآله وصحبه وسلم

اللهم صل على خير خلقك محمد بن عبد الله وآله وصحبه وسلم"

Transcription of the beginning of the letter from Abdullah Hameed (ޖުމްހޫރިއްޔާ ޖެހިގެން) on page 11 of *Jameeluge Mazumoonutah II*, March 1993, by the Maldivian National Centre for Linguistic and Historical Research

This letter is typical in that it starts with the “bismi” top and center. The “nishaan” (national shield) is under the “bismi,” typical of government correspondence.

An example of THAANA NAA as seen in a document from the National Centre for Linguistic and Historical Research.

