Date: 1998-03-10

Title: Revised proposal for encoding the Thaana script in the BMP of ISO/IEC 10646

Source: Michael Everson, Everson Gunn Teoranta (IE)

**Status:** Expert Contribution

Action: For consideration by JTC1/SC2/WG2

This document is a revision of N1519, based on feedback received from Maldivian expert Hassan Ahmed Maniku. It contains the proposal summary.

# A. Administrative

#### 1. Title

Proposal for encoding the Thaana script in ISO/IEC 10646

#### 2. Requester's name

Michael Everson, Everson Gunn Teoranta (WG2 member for Ireland)

# 3. Requester type

Expert contribution

#### 4. Submission date

1998-03-10

## 5. Requester's reference

http://www.indigo.ie/egt/standards/dv/thaana.html

# 6a. Completion

This is a complete proposal.

# 6b. More information to be provided?

No

# **B.** Technical -- General

# 1a. New script? Name?

Yes. Thaana.

#### 1b. Addition of characters to existing block? Name?

No

## 2. Number of characters

50

# 3. Proposed category

Category A

#### 4. Proposed level of implementation and rationale

Thaana requires Level 2 implementation as Indic scripts do.

#### 5a. Character names included in proposal?

Yes

#### 5b. Character names in accordance with guidelines?

Yes

# 5c. Character shapes reviewable?

Yes (see below)

#### 6a. Who will provide computerized font?

Michael Everson, Everson Gunn Teoranta

#### 6b. Font currently available?

Michael Everson, Everson Gunn Teoranta

#### 6c. Font format?

TrueType

## 7a. Are references (to other character sets, dictionaries, descriptive texts, etc.) provided?

Yes. An 8-bit font was made available to me as a source encoding. It has de-facto use status. There appear to be no standards. (see below)

# 7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Not provided here. Than is well-known. There is an online newspaper in Dhivehi (Haveeru Medhu Haftha, http://www.haveeru.com/), which displays text with gifs.

### 8. Does the proposal address other aspects of character data processing?

Yes (see below)

# C. Technical -- Justification

## 1. Contact with the user community?

Yes. The Hassan Ahmed Maniku, Maldivian Students' Association (GB), the National Centre for Linguistic and Historical Research (MV), Ministry for Education, (MV); Ministry for Information and Culture, (MV).

# 2. Information on the user community?

226,000 people live in the Maldives. Thaana is the national script.

# 3a. The context of use for the proposed characters?

Thaana script is commonly used to write Dhivehi.

#### 3b. Reference

See bibliography below

## 4a. Proposed characters in current use?

Yes

#### 4b. Where?

In the Maldives.

#### 5a. Characters should be encoded entirely in BMP?

Yes. Positions U+0780 .. U+07BF are proposed for the encoding.

#### 5b. Rationale

Thaana is a Category A script.

# 6. Should characters be kept in a continuous range?

Yes

# 7a. Can the characters be considered a presentation form of an existing character or character sequence?

No

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?

No

# 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** 

# E. Proposal

# User community

The Thaana script is used to write the modern Dhivehi language of the Republic of Maldives, a group of atolls in the Indian Ocean, circa 650 km SW of Sri Lanka, about 4°N 73°E. 226,000 people live in the Maldives.

# **Processing**

Thaana is written from right to left and partakes of features of both the Indic and Arabic script varieties. Consonants have no inherent a vowel sound, and are always written with either a vowel sign (U+xxA6 -> U+xxAF) or a null "vanishing vowel" sign (SUKUN, U+xxB0) above or below them. On THAANA LETTER ALIFU (U+xx87) the null vowel sign is a glottal stop.

Loanwords from Arabic may be written in the Arabic script (a custom not so current today) or transcribed by means of dots on existing Thaana letters. The use of modified Thaana letters dates from the middle of the twentieth century and is promoted by the government.

Both Arabic and European digits are used, according to Nakanishi, who shows the digits in *nastalîq* style (U+0660 -> U+0669, not U+06F0 -> U+06F9). Nakanishi's sample text, from a periodical entitled *noonlait* ('Moonlight'), is dated 29.12.1975 in European digits, and 1395 in Arabic digits. Numbers have LR display directionality in Thaana, and Arabic numeric punctuation is uded with digits, whether Arabic or European.

The ARABIC COMMA (U+060C), ARABIC SEMICOLON (U+061B), and ARABIC QUESTION MARK (U+061F) are used. According to Nakanishi, comma can be represented by FULL STOP (U+002E) and full stop by two FULL STOPs (U+002E + U+002E, or U+2025). I have seen no texts employing this practice, but as the double FULL STOP can be represented by characters already in UCS there is no need to add specific codes for it.

The THAANA REYTU SIGN is like COMMERCIAL AT. According to Hussain Zahid of the Maldivian Students' Association:

The "O" sign is commonly seen used in bills/invoices. Hence the character is usually pronounced "reyt", as in English. It is supposedly borrowed from English. But with the recent introduction of the internet in the Maldives, this sign (@) is likely to be pronounced as "at" in the future.

It may be possible to unify this character with U+0040. If it is understood by Maldivians to be *identical* to the @ sign in e-mail addresses, then it is imperative that the two be unified.

Transliteration of the character names in this proposal follows the usual method employed in the Maldives (and is different from the LC romanization). The character name at U+xx94, YAA, is found in some sources (Daniels & Bright 1996, Maniku 1990) as YAVIYANI, but Hassan Ahmed Maniku said that the former is more common today. A list of character names in Thaana script and in the romanization used here is appended to this proposal. Characters are listed in Thaana alphabetical order from HAA to TTAA for the Thaana letters, followed by the extended characters in Arabic alphabetical order from HHAA to WAAVU. The practice of Dhivehi dictionaries is at present unknown.

# **Unicode Character Properties**

```
Spacing letters, category "Lo", bidi category "R" (strong right to left)
```

```
000
      xx80
             THAANA LETTER HAA
001
      xx81
             THAANA LETTER SHAVIYANI
002
      xx82
             THAANA LETTER NOONU
003
      xx83
             THAANA LETTER RAA
004
      xx84
             THAANA LETTER BAA
005
            THAANA LETTER LHAVIYANI
      xx85
006
      xx86
            THAANA LETTER KAAFU
007
            THAANA LETTER ALIFU
      xx87
008
      xx88
            THAANA LETTER VAAVU
009
      xx89
            THAANA LETTER MEEMU
010
      xx8A
            THAANA LETTER FAAFU
011
      xx8B
            THAANA LETTER DHAALU
012
      xx8C
            THAANA LETTER THAA
013
      xx8D
            THAANA LETTER LAAMU
014
      xx8E
            THAANA LETTER GAAFU
015
      xx8F
             THAANA LETTER GNAVIYANI
016
      xx90
             THAANA LETTER SEENU
017
      xx91
             THAANA LETTER DAVIYANI
018
      xx92
             THAANA LETTER ZAVIYANI
019
      xx93
             THAANA LETTER TAVIYANI
020
      xx94
            THAANA LETTER YAA
021
      xx95
            THAANA LETTER PAVIYANI
022
      xx96
             THAANA LETTER JAVIYANI
023
      xx97
             THAANA LETTER CHAVIYANI
024
      xx98
             THAANA LETTER TTAA
025
      xx99
             THAANA LETTER HHAA
026
      xx9A
            THAANA LETTER KHAA
027
      xx9B
            THAANA LETTER THAALU
028
      xx9C
            THAANA LETTER ZAA
029
      xx9D
             THAANA LETTER SHEENU
030
      xx9E
             THAANA LETTER SAADHU
031
      xx9F
             THAANA LETTER DAADHU
032
      xxA0
             THAANA LETTER TO
033
      xxA1
             THAANA LETTER ZO
034
      xxA2
            THAANA LETTER AINU
035
      xxA3
            THAANA LETTER GHAINU
036
      xxA4
             THAANA LETTER OAAFU
037
             THAANA LETTER WAAVU
      xxA5
```

Non-spacing marks, category "Mn", bidi category "R" (strong right to left); combining priorities in parentheses:

```
    038 xxA6 THAANA VOWEL SIGN ABAFILI (230)
    039 xxA7 THAANA VOWEL SIGN AABAAFILI (230)
    040 xxA8 THAANA VOWEL SIGN IBIFILI (220)
    041 xxA9 THAANA VOWEL SIGN EEBEEFILI (220)
    042 xxAA THAANA VOWEL SIGN UBUFILI (230)
```

```
043 xxAB THAANA VOWEL SIGN OOBOOFILI (230)
044 xxAC THAANA VOWEL SIGN EBEFILI (230)
045 xxAD THAANA VOWEL SIGN EYBEYFILI (230)
046 xxAE THAANA VOWEL SIGN OBOFILI (230)
047 xxAF THAANA VOWEL SIGN OABOAFILI (230)
048 xxB0 THAANA VOWEL SIGN SUKUN (230)
```

Symbols, category "So"

```
049 xxB1 THAANA REYTU SIGN
```

Four columns are required to encode Thaana. The Thaana block is divided into the following ranges:

$U+xx80 \rightarrow U+xx97$	Consonant letters
$U+xx98 \rightarrow U+xxA5$	Extended Thaana letters
$U+xxA6 \rightarrow U+xxB0$	Non-spacing vowel signs
U+xxB1	Punctuation sign
$U+xxB2 \rightarrow U+xxBF$	currently unassigned

#### References

Barry, Randall K. 1997. *ALA-LC romanization tables: transliteration schemes for non-Roman scripts*. Washington, DC: Library of Congress Cataloging Distribution Service. ISBN 0-8444-0940-5

Daniels, Peter T., and William Bright, eds. 1996. *The world's writing systems*. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0

Faulmann, Carl. 1990 (1880). Das Buch der Schrift. Frankfurt am Main: Eichborn. ISBN 3-8218-1720-8

Maniku, Hassan Ahmed, and J. B. Disanayaka. 1990. *Say it in Maldivian (Dhivehi)*. Colombo: Lake House Investments Ltd.

Nakanishi, Akira. 1990. *Writing systems of the world: alphabets, syllabaries, pictograms*. Rutland, VT: Charles E. Tuttle. ISBN 0-8048-1654-9

Unicode Consortium. 1992. Unicode Technical Report #3: exploratory proposals.

**TABLE xxx - Row xx: THAANA** 

	xx8	xx9	xxA	xxB
0		-	Ģ	ំ
1	مو	٤	ģ	(3)
2	Ju	E	A	
3	æ	2	À	
4	•	,,	Ë	
5	*	3.	ġ	
6	V	2	ं	
7	1	56	ं	
8	9	Š	Ş	
9	7	1.	्र	
Α	3	j	ं	
В	<b>ع</b> ر	فر	<b>??</b>	
С	ç	منو	ំ	
D	,	ئو	ं	
Е	8	ب	<b>×</b> ·	
F	C	نو	្	

G = 00 P = 00

**TABLE xxx - Row xx: THAANA** 

dec	hex	Name
000 001 002 003 004 005 006 007 008 009 010 011 012 023 024 025 026 027 028 029 030 031 032 033 034 035 036 037 038 039 040 041 042 043 045 049 050 051 052 053 054 055 056 057 057 058 057 058 059 069 069 069 069 069 069 069 069 069 06	801 233 445 667 889 ABCDEF 01 23 345 667 889 ABCDEF 01 23 345 667 889 ABCDEF 01 23 345 667 889 ABCDEF	THAANA LETTER HAA THAANA LETTER SHAVIYANI THAANA LETTER RAA THAANA LETTER CHAVIYANI THAANA LETTER CHAVIYANI THAANA LETTER ALIFU THAANA LETTER MEEMU THAANA LETTER PAAFU THAANA LETTER DHAALU THAANA LETTER DHAALU THAANA LETTER GNAVIYANI THAANA LETTER TAYA THAANA LETTER TAYA THAANA LETTER TAYA THAANA LETTER TAYA THAANA LETTER TAYIYANI THAANA LETTER TAY THAANA LETTER TO THAANA LETTER TO THAANA LETTER TO THAANA LETTER SAADHU THAANA LETTER SA THAANA LETTER TO THAANA LETTER WAAVU THAANA LETTER OHADHU THAANA LETTER OHADHU THAANA LETTER WAAVU THAANA VOWEL SIGN BEBEFILI THAANA POWEL SIGN BEBEFILI THABANA POWEL SIGN BEBEFILI THABANA POWEL SIGN BEBEFILI THABANA POWEL SIGN BEBEFILI THABA

"	THAANA LETTER HAA	THAANA LETTER TTAA
, תפונית	THAANA LETTER SHAVIYANI	THAANA LETTER HHAA
2 22 00,000	THAANA LETTER NOONU	THAANA LETTER KHAA
2	THAANA LETTER RAA	THAANA LETTER THAALU
* *	THAANA LETTER BAA	THAANA LETTER ZAA
מ'פונית	THAANA LETTER LHAVIYANI	THAANA LETTER SHEENU
34	THAANA LETTER KAAFU	THAANA LETTER SAADHU
2.4	THAANA LETTER ALIFU	THAANA LETTER DAADHU
370	THAANA LETTER VAAVU	THAANA LETTER TO
2	THAANA LETTER MEEMU	THAANA LETTER ZO
77 34 33	THAANA LETTER FAAFU	THAANA LETTER AINU
رر درو درو	THAANA LETTER DHAALU	THAANA LETTER GHAINU
7,3	THAANA LETTER THAA	THAANA LETTER QAAFU
<i>\$</i>	THAANA LETTER LAAMU	THAANA LETTER WAAVU
7 <i>7</i>	7,00	THAANA ABAFILI
35	THAANA LETTER GAAFU	THAANA AABAAFILI
שפתית	THAANA LETTER GNAVIYANI	THAANA IBIFILI
سومتر د	THAANA LETTER SEENU	THAANA EEBEEFILI
MUSE	THAANA LETTER DAVIYANI	THAANA UBUFILI
mnge	THAANA LETTER ZAVIYANI	THAANA OOBOOFILI
שפתת	THAANA LETTER TAVIYANI	THAANA EBEFILI
'n	THAANA LETTER YAA	THAANA EYBEYFILI
rn93.	THAANA LETTER PAVIYANI	THAANA OBOFILI
ع و مرسر	THAANA LETTER JAVIYANI	THAANA OABOAFILI
mngs	THAANA LETTER CHAVIYANI	THAANA SUKUN
	246 87	THAANA REYTU SIGN