

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. Thaana 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 *nastaliq* style (U+0660 -> U+0669, not U+06F0 -> U+06F9). Nakanishi's sample text, from a periodical entitled  *moonlait* ('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 used 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 "⓪" 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	xx85	THAANA LETTER LHAVIYANI
006	xx86	THAANA LETTER KAAFU
007	xx87	THAANA LETTER ALIFU
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 QAAFU
037	xxA5	THAANA LETTER WAAVU

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 -> U+xx97	Consonant letters
U+xx98 -> U+xxA5	Extended Thaana letters
U+xxA6 -> U+xxB0	Non-spacing vowel signs
U+xxB1	Punctuation sign
U+xxB2 -> 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				
1				
2				
3				
4				
5				
6				
7				
8				
9				
A				
B				
C				
D				
E				
F				

G = 00
P = 00


TABLE xxx - Row xx: THAANA

dec	hex	Name
000	80	THAANA LETTER HAA
001	81	THAANA LETTER SHAVIYANI
002	82	THAANA LETTER NOONU
003	83	THAANA LETTER RAA
004	84	THAANA LETTER BAA
005	85	THAANA LETTER LHAVIYANI
006	86	THAANA LETTER KAAFU
007	87	THAANA LETTER ALIFU
008	88	THAANA LETTER VAAVU
009	89	THAANA LETTER MEEMU
010	8A	THAANA LETTER FAAFU
011	8B	THAANA LETTER DHAALU
012	8C	THAANA LETTER THAA
013	8D	THAANA LETTER LAAMU
014	8E	THAANA LETTER GAAFU
015	8F	THAANA LETTER GNAVIYANI
016	90	THAANA LETTER SEENU
017	91	THAANA LETTER DAVIYANI
018	92	THAANA LETTER ZAVIYANI
019	93	THAANA LETTER TAVIYANI
020	94	THAANA LETTER YAA
021	95	THAANA LETTER PAVIYANI
022	96	THAANA LETTER JAVIYANI
023	97	THAANA LETTER CHAVIYANI
024	98	THAANA LETTER TAA
025	99	THAANA LETTER HAA
026	9A	THAANA LETTER KHA
027	9B	THAANA LETTER THAALU
028	9C	THAANA LETTER ZAA
029	9D	THAANA LETTER SHEENU
030	9E	THAANA LETTER SAADHU
031	9F	THAANA LETTER DAADHU
032	A0	THAANA LETTER TO
033	A1	THAANA LETTER ZO
034	A2	THAANA LETTER AINU
035	A3	THAANA LETTER GHAINU
036	A4	THAANA LETTER QAAFU
037	A5	THAANA LETTER WAAVU
038	A6	THAANA VOWEL SIGN ABAFILI
039	A7	THAANA VOWEL SIGN AABAAFILI
040	A8	THAANA VOWEL SIGN IBIFILI
041	A9	THAANA VOWEL SIGN EEBEEFILI
042	AA	THAANA VOWEL SIGN UBUFILI
043	AB	THAANA VOWEL SIGN OOBEOFILI
044	AC	THAANA VOWEL SIGN EBEFILI
045	AD	THAANA VOWEL SIGN EYBEYFILI
046	AE	THAANA VOWEL SIGN OBOFILI
047	AF	THAANA VOWEL SIGN QABOAFILI
048	B0	THAANA VOWEL SIGN SUKUN
049	B1	THAANA REYTU SIGN
050	B2	(This position shall not be used)
051	B3	(This position shall not be used)
052	B4	(This position shall not be used)
053	B5	(This position shall not be used)
054	B6	(This position shall not be used)
055	B7	(This position shall not be used)
056	B8	(This position shall not be used)
057	B9	(This position shall not be used)
058	BA	(This position shall not be used)
059	BB	(This position shall not be used)
060	BC	(This position shall not be used)
061	BD	(This position shall not be used)
062	BE	(This position shall not be used)
063	BF	(This position shall not be used)


 THAANA LETTER HAA

 THAANA LETTER SHAVIYANI

 THAANA LETTER NOONU

 THAANA LETTER RAA

 THAANA LETTER BAA

 THAANA LETTER LHAVIYANI

 THAANA LETTER KAAFU

 THAANA LETTER ALIFU

 THAANA LETTER VAAVU

 THAANA LETTER MEEMU

 THAANA LETTER FAAFU

 THAANA LETTER DHAALU

 THAANA LETTER THAA

 THAANA LETTER LAAMU

 THAANA LETTER GAAFU

 THAANA LETTER GNAVIYANI

 THAANA LETTER SEENU

 THAANA LETTER DAVIYANI

 THAANA LETTER ZAVIYANI



























 THAANA LETTER TAVIYANI

 THAANA LETTER YAA

 THAANA LETTER PAVIYANI

 THAANA LETTER JAVIYANI

 THAANA LETTER CHAVIYANI


 THAANA LETTER TTAA

 THAANA LETTER HHAA

 THAANA LETTER KHAA

 THAANA LETTER THAALU

 THAANA LETTER ZAA

 THAANA LETTER SHEENU

 THAANA LETTER SAADHU

 THAANA LETTER DAADHU

 THAANA LETTER TO

 THAANA LETTER ZO

 THAANA LETTER AINU

 THAANA LETTER GHAINU

 THAANA LETTER QAAFU

 THAANA LETTER WAAVU

 THAANA ABAFILI

 THAANA AABAAFILI

 THAANA IBIFILI

 THAANA EEBEEFILI

 THAANA UBUFILI

 THAANA OOBOOFILI

 THAANA EBEFILI

 THAANA EYBEYFILI

 THAANA OBOFILI

 THAANA OABOAFILI

 THAANA SUKUN

 THAANA REYTU SIGN