

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Международная организация по стандартизации

Doc Type: Working Group Document

Title: Proposal for encoding the Philippine scripts in the BMP of ISO/IEC 10646
Source: Michael Everson, EGT (IE)
Status: Expert Contribution
Action: For consideration by JTC1/SC2/WG2
Date: 1998-05-25

This document is based on the proposal written by Rick McGowan and published in UTR#3. It contains the proposal summary.

A. Administrative

1. Title

Proposal for encoding the Philippine scripts in the BMP of ISO/IEC 10646

2. Requester's name

Michael Everson, EGT (WG2 member for Ireland)

3. Requester type

Expert contribution

4. Submission date

1998-05-25

5. Requester's reference**6a. Completion**

This is a complete proposal.

6b. More information to be provided?

No

B. Technical -- General

1a. New script? Name?

Yes. Four related scripts: Tagalog, Hanunóo, Buhid, and Tagbanwa, all contained in a single block, "Philippine scripts".

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

No

2. Number of characters

86

3. Proposed category

Category A

4. Proposed level of implementation and rationale

The Philippine scripts require Level 2 implementation as other 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?

Hector Santos via Michael Everson

6b. Font currently available?

Yes

6c. Font format?

TrueType

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?

Hardcopy is provided for WG2 distribution.

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

Yes (see below)

C. Technical -- Justification

1. Contact with the user community?

Yes: Hector Santos. See <http://www.bibingka.com/dahon>

2. Information on the user community?

See below.

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

Tagalog was formerly (until the mid 1700s) used to write Tagalog and Ilocano languages; Hanunóo, Buhid, and Tagbanwa are used today to write the Hanunóo, Buhid, and Tagbanwa languages.

3b. Reference

See bibliography below

4a. Proposed characters in current use?

Yes

4b. Where?

In the Philippines. The scripts also enjoy some use in North America by Philippine communities for various purposes.

5a. Characters should be encoded entirely in BMP?

Yes.

5b. Rationale

Contemporary use.

6. Should characters be kept in a continuous range?

Yes, they should be encoded in a single block as presented here.

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. There is some similarity between the four scripts but the differences are striking enough to warrant against their unification.

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, it is provided in the hardcopy samples to be circulated by WG2. Vowel signs combine with all letter characters; in some cases unique logotypes result from the combination. This is not a problematic issue.

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

Tagalog is a script of the Philippines. It was formerly used to write the Tagalog, Bisaya, Ilocano, and other languages. The Tagalog language now utilizes the Latin script. The Tagalog script is distantly related to the scripts of the southern Indian subcontinent, but the exact route by which they were brought to the Philippines is not certain. It seems that they may have been transported by way of the palaeographic scripts of Western Java between the 10th and 14th centuries. Written accounts of the Tagalog script by Spanish missionaries, and documents in Tagalog, are known from about the period of initial Spanish incursion (mid-1500s). The Tagalog script had fallen out of normal use by the mid 1700s. It has three living descendants: the Hanunóo, Buhid (also called Mangyan), and Tagbanwa (also called Bisaya) scripts, also part of this proposal.

Structure

Vowel signs are used in a manner similar to that employed by other Brahmic-derived the scripts. The vowel I is written with a mark above, and the vowel U with an identical mark below the associated consonant. The mark is known as *kudlit* or *tulbok* in Buhid and *ulitan* in Tagbanwa. The script has only the two vowel signs I and U, which are also used respectively to stand for the vowels E and O. Though all languages normally written with this script have syllables possessing final consonants, they cannot normally be expressed in the script. Reforms to express final consonants with a virama character were proposed for the Tagalog script, but were rejected by native users who considered the script adequate. A similar reform for the Hanunóo script seems to have been better received. These signs were not proposed for all of the scripts; because they are found in existing character sets, they are encoded here for Tagalog and Hanunóo. Other reforms, such as the addition of E and O vowel signs or the letter FA, have not been included here as they are wanting attestation (even in existing character sets). There is room in the table for their later addition should it prove necessary.

The Philippine scripts are read from left to right in horizontal lines running from top to bottom. They may be *written* either in that manner, or in vertical lines running from bottom to top, moving from left to right. In the latter case, the letters are written sideways so they may be *read* horizontally. This method of

writing is probably due to the medium and writing implements used. Text is often scratched with a sharp instrument onto beaten strips of bamboo which are held pointing away from the body and worked from the proximal to distal ends, from left to right.

UTR#3 states: “The alphabetical order of Tagalog is known from Tagbanwa speakers and is described in folktales. This order is used in the accompanying charts. The two vowel signs are added at the end of the alphabet.” The names list in UTR#3, however, is (except for the vowel signs) given in Latin alphabetical order (*a, i, u, ba, da, ga, ha, ka, la, ma, na, nga, pa, sa, ta, wa, ya, -i, -u*). Daniels & Bright give another ordering, “based on the 16th-century Tagalog sequence” (*a, i, u, ha, pa, ka, sa, la, ra, ta, na, ba, ma, ga, da, ya, nga, wa*). This proposal gives the characters in the traditional Brahmic order, which is followed in many sources, including Santos 1994 and 1995 (source of the fonts used in this proposal). The accompanying chart is divided into four segments, from left to right: Tagalog, Hanunóo, Buhid, Tagbanwa. Each of these 2-column segments should be given a separate collection ID in ISO/IEC 10646.

Processing

The Philippine scripts are written from left to right and follow the usual Indic pattern. Consonants have an inherent /a/ vowel sound, and can be written with either a vowel sign or (in the case of Tagalog and Hanunóo) a null "vanishing vowel" sign. In some cases, the vowel signs simply rest over or under the consonants. In Hanunóo and Buhid, however, special conjoined glyphs are formed. The Philippine scripts are straightforward and simple to encode.

Unicode Character Properties

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

x00-x0D, x0e-x11, x20-x31, x40-x51, x60-x6c, x6e-x70

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

x12, x32, x52, x72 (230)

x13-x14, x33-x34, x53, x73 (220)







Symbols, category "So"

x16-x18, x35-x36, x55-x56, x75-x76

References

- 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
- Haarmann, Harald. 1990. *Die Universalgeschichte der Schrift*. Frankfurt: Campus. ISBN 3-593-34346-0
- Nakanishi, Akira. 1990. *Writing systems of the world: alphabets, syllabaries, pictograms*. Rutland, VT: Charles E. Tuttle. ISBN 0-8048-1654-9
- Santos, Hector. 1994. *The Tagalog script*. (Ancient Philippine Scripts Series; 1). Los Angeles: Sushi Dog Graphics.
- Santos, Hector. 1995. *The living scripts*. (Ancient Philippine Scripts Series; 2). Los Angeles: Sushi Dog Graphics.
- Unicode Consortium. 1992. *Unicode Technical Report #3: exploratory proposals*.
- Wolf, Edwin, II, ed. 1947. *Doctrina christiana: the first book printed in the Philippines, Manila 1593. A facsimile of the copy in the Lessing J. Rosenwald Collection*. Washington, DC: Library of Congress.

TABLE xxx - Row xx: PHILIPPINE SCRIPTS

	x0	x1	x2	x3	x4	x5	x6	x7
0	 000	 016	 032	 048	 064	 080	 096	 112
1	 001	 017	 033	 049	 065	 081	 097	113
2	 002	 018	 034	 050	 066	 082	 098	 114
3	 003	 019	 035	 051	 067	 083	 099	 115
4	 004	 020	 036	 052	 068	084	 100	116
5	 005	021	 037	 053	 069	 085	 101	 117
6	 006	 022	 038	 054	 070	 086	 102	 118
7	 007	023	 039	055	 071	087	 103	119
8	 008	024	 040	056	 072	088	 104	120
9	 009	025	 041	057	 073	089	 105	121
A	 010	026	 042	058	 074	090	 106	122
B	 011	027	 043	059	 075	091	 107	123
C	 012	028	 044	060	 076	092	 108	124
D	013	029	 045	061	 077	093	 109	125
E	 014	030	 046	062	 078	094	 110	126
F	 015	031	 047	063	 079	095	 111	127

G = 00
P = 00

TABLE xxx - Row xx: PHILIPPINE SCRIPTS

dec	hex	Name
000	00	TAGALOG LETTER A
001	01	TAGALOG LETTER I
002	02	TAGALOG LETTER U
003	03	TAGALOG LETTER KA
004	04	TAGALOG LETTER GA
005	05	TAGALOG LETTER NGA
006	06	TAGALOG LETTER TA
007	07	TAGALOG LETTER DA
008	08	TAGALOG LETTER NA
009	09	TAGALOG LETTER PA
010	0A	TAGALOG LETTER BA
011	0B	TAGALOG LETTER MA
012	0C	TAGALOG LETTER YA
013	0D	(This position shall not be used)
014	0E	TAGALOG LETTER LA
015	0F	TAGALOG LETTER WA
016	10	TAGALOG LETTER SA
017	11	TAGALOG LETTER HA
018	12	TAGALOG VOWEL SIGN I
019	13	TAGALOG VOWEL SIGN U
020	14	TAGALOG SIGN VIRAMA
021	15	(This position shall not be used)
022	16	TAGALOG DOUBLE PUNCTUATION
023	17	(This position shall not be used)
024	18	(This position shall not be used)
025	19	(This position shall not be used)
026	1A	(This position shall not be used)
027	1B	(This position shall not be used)
028	1C	(This position shall not be used)
029	1D	(This position shall not be used)
030	1E	(This position shall not be used)
031	1F	(This position shall not be used)
032	20	HANUNOO LETTER A
033	21	HANUNOO LETTER I
034	22	HANUNOO LETTER U
035	23	HANUNOO LETTER KA
036	24	HANUNOO LETTER GA
037	25	HANUNOO LETTER NGA
038	26	HANUNOO LETTER TA
039	27	HANUNOO LETTER DA
040	28	HANUNOO LETTER NA
041	29	HANUNOO LETTER PA
042	2A	HANUNOO LETTER BA
043	2B	HANUNOO LETTER MA
044	2C	HANUNOO LETTER YA
045	2D	HANUNOO LETTER RA
046	2E	HANUNOO LETTER LA
047	2F	HANUNOO LETTER WA
048	30	HANUNOO LETTER SA
049	31	HANUNOO LETTER HA
050	32	HANUNOO VOWEL SIGN I
051	33	HANUNOO VOWEL SIGN U
052	34	HANUNOO SIGN PAMUDPOD
053	35	HANUNOO SINGLE PUNCTUATION
054	36	HANUNOO DOUBLE PUNCTUATION
055	37	(This position shall not be used)
056	38	(This position shall not be used)
057	39	(This position shall not be used)
058	3A	(This position shall not be used)
059	3B	(This position shall not be used)
060	3C	(This position shall not be used)
061	3D	(This position shall not be used)
062	3E	(This position shall not be used)
063	3F	(This position shall not be used)
064	40	BUHID LETTER A
065	41	BUHID LETTER I
066	42	BUHID LETTER U
067	43	BUHID LETTER KA
068	44	BUHID LETTER GA
069	45	BUHID LETTER NGA
070	46	BUHID LETTER TA
071	47	BUHID LETTER DA
072	48	BUHID LETTER NA
073	49	BUHID LETTER PA
074	4A	BUHID LETTER BA
075	4B	BUHID LETTER MA
076	4C	BUHID LETTER YA
077	4D	BUHID LETTER RA
078	4E	BUHID LETTER LA
079	4F	BUHID LETTER WA
080	50	BUHID LETTER SA
081	51	BUHID LETTER HA
082	52	BUHID VOWEL SIGN I
083	53	BUHID VOWEL SIGN U
084	54	(This position shall not be used)
085	55	BUHID SINGLE PUNCTUATION
086	56	BUHID DOUBLE PUNCTUATION
087	57	(This position shall not be used)
088	58	(This position shall not be used)

Group 00

dec	hex	Name
089	59	(This position shall not be used)
090	5A	(This position shall not be used)
091	5B	(This position shall not be used)
092	5C	(This position shall not be used)
093	5D	(This position shall not be used)
094	5E	(This position shall not be used)
095	5F	(This position shall not be used)
096	60	TAGBANWA LETTER A
097	61	TAGBANWA LETTER I
098	62	TAGBANWA LETTER U
099	63	TAGBANWA LETTER KA
100	64	TAGBANWA LETTER GA
101	65	TAGBANWA LETTER NGA
102	66	TAGBANWA LETTER TA
103	67	TAGBANWA LETTER DA
104	68	TAGBANWA LETTER NA
105	69	TAGBANWA LETTER PA
106	6A	TAGBANWA LETTER BA
107	6B	TAGBANWA LETTER MA
108	6C	TAGBANWA LETTER YA
109	6D	(This position shall not be used)
110	6E	TAGBANWA LETTER LA
111	6F	TAGBANWA LETTER WA
112	70	TAGBANWA LETTER SA
113	71	(This position shall not be used)
114	72	TAGBANWA VOWEL SIGN I
115	73	TAGBANWA VOWEL SIGN U
116	74	(This position shall not be used)
117	75	TAGBANWA SINGLE PUNCTUATION
118	76	TAGBANWA DOUBLE PUNCTUATION
119	77	(This position shall not be used)
120	78	(This position shall not be used)
121	79	(This position shall not be used)
122	7A	(This position shall not be used)
123	7B	(This position shall not be used)
124	7C	(This position shall not be used)
125	7D	(This position shall not be used)
126	7E	(This position shall not be used)
127	7F	(This position shall not be used)

Plane 00

Row XX