

ISO
INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION

ISO/IEC JTC1/SC2/WG2
Universal Multiple-Octet Coded Character Set (UCS)

ISO/IEC JTC1/SC2/WG2 N 1942

Date: 1998-12-08

TITLE: Hangul syllable name rules, proposed for 10646 2nd Edition
SOURCE: Bruce Paterson
STATUS: Personal contribution
ACTION: For review and confirmation by WG2
DISTRIBUTION: Members of JTC1/SC2/WG2

Character name tables in Amd.5 Hangul Syllables occupy 88 pages. Since the character names for Hangul syllables can be generated by algorithm, it would be possible in 10646 2nd Edition to omit those pages, and provide instead a description of the rules for constructing the name of any Hangul syllable character. The rules can also be extended to apply to the annotations which appear with most of the names. Further, if the character names (and annotations) are omitted then no action will be required on the Defect Report on Hangul syllable names and annotations, WG2 N1806.

This paper proposes a text which describes the rules. It makes use of decimal arithmetic only, avoiding operations in hexadecimal arithmetic which might be unfamiliar to some users. The notation used for these operations is similar to that in Annexes Q and R (UTF-16 and UTF-8).

In Clause 25 "Code tables and lists of character names" add a new subclause title for the existing text, as follows:

25.1 General

Add a new subclause:

25.2 Character names and annotations for Hangul syllables

at the end of the clause.

(The complete clause 25 as amended is shown overleaf.)

25 Code tables and lists of character names

25.1 General

An overview of the Basic Multilingual Plane is shown in figure 3. Detailed code tables and lists of character names for the Basic Multilingual Plane are shown on the following pages and in applicable Amendments.

Guidelines to be used for constructing names of characters are given in annex K for information. In some cases, a name of a character is followed by additional explanatory statements not part of the name. These statements are in parentheses and not in capital letters except for the initials of the word, where required.

25.2 Character names and annotations for Hangul syllables

Names for the Hangul syllable characters in code positions (hex) 0000 AC00 - 0000 D7A3 are derived from their code position numbers by the numerical procedure described below. Lists of names for these characters are not provided.

1. The code position number of a Hangul syllable character is of the form 0000 $h_1h_2h_3h_4$ where h_1 , h_2 , h_3 , and h_4 are hexadecimal digits; h_1h_2 is the Row number within the BMP and h_3h_4 is the cell number within the row. The code position number lies within the range 0000 AC00 to 0000 D7A3.

2. Derive the decimal numbers d_1 , d_2 , d_3 , d_4 that are numerically equal to the hexadecimal digits h_1 , h_2 , h_3 , h_4 respectively.

3. Calculate the character index C from the formula:

$$C = 4096 \times (d_1 - 10) + 256 \times (d_2 - 12) + 16 \times d_3 + d_4$$

Note: If $C < 0$ or $> 11,171$ then the character is not a Hangul syllable.

4. Calculate the syllable component indices I , P , F from the following formulae:

$$I = C / 588 \quad (\text{Note: } 0 \leq I \leq 18)$$

$$P = (C \% 588) / 28 \quad (\text{Note: } 0 \leq P \leq 20)$$

$$F = C \% 28 \quad (\text{Note: } 0 \leq F \leq 27)$$

where "/" indicates integer division (i.e. x / y is the integer quotient of the division), and "%" indicates the modulo operation (i.e. $x \% y$ is the remainder after the integer division x / y).

5. Obtain the Latin character strings that correspond to the three indices I , P , F from columns 2, 3, and 4 respectively of Table 1 below (for $I = 11$ and for $F = 0$ the corresponding strings are null). Concatenate these three strings in left-to-right order to make a single string, the syllable-name.

6. The character name for the character at position 0000 $h_1h_2h_3h_4$ is then:

HANGUL SYLLABLE s - n

where " s - n " indicates the syllable string derived in step 5.

Example.

For the character in code position D4DE:

$$d_1 = 13, d_2 = 4, d_3 = 13, d_4 = 14.$$

$$C = 10462$$

$$I = 17, P = 16, F = 18.$$

The corresponding Latin character strings are:

P , WI , BS .

The syllable name is $PWIBS$, and the character name is:

HANGUL SYLLABLE $PWIBS$

Annotations for the Hangul syllable characters in code positions (hex) 0000 AC00 - 0000 D7A3 are also derived from their code position numbers by a similar numerical procedure described below.

7. Carry out steps 1 to 4 as described above.

8. Obtain the Latin character strings that correspond to the three indices I , P , F from columns 5, 6, and 7 respectively of Table 1 below (for $I = 11$ and for $F = 0$ the corresponding strings are null). Concatenate these three strings in left-to-right order to make a single string, and enclose it within parentheses to form the annotation.

Example.

For the character in code position D4DE:

$$d_1 = 13, d_2 = 4, d_3 = 13, d_4 = 14.$$

$$C = 10462$$

$$I = 17, P = 16, F = 18.$$

The corresponding Latin character strings are:

ph , wi , ps ,

and the annotation is $(phwips)$.

Table 1: Elements of Hangul syllable names and annotations

Index number	Syllable name elements			Annotation elements		
	I string	P string	F string	I string	P string	F string
0	G	A		k	a	
1	GG	AE	G	kk	ae	k
2	N	YA	GG	n	ya	kk
3	D	YAE	GS	t	yae	ks
4	DD	EO	N	tt	eo	n
5	R	E	NJ	r	e	nc
6	M	YEO	NH	m	yeo	nh
7	B	YE	D	p	ye	t
8	BB	O	L	pp	o	l
9	S	WA	LG	s	wa	lk
10	SS	WAE	LM	ss	wae	lm
11		OE	LB		oe	lp
12	J	YO	LS	c	yo	ls
13	JJ	U	LT	cc	u	lth
14	C	WEO	LP	ch	weo	lph
15	K	WE	LH	kh	we	lh
16	T	WI	M	th	wi	m
17	P	YU	B	ph	yu	p
18	H	EU	BS	h	eu	ps
19		YI	S		yi	s
20		I	SS		i	ss
21			NG			ng
22			J			c
23			C			ch
24			K			kh
25			T			th
26			P			ph
27			H			h