



Document Number: Korea K1466
ISO/IEC JTC1/SC2/WG2 N3095

2006-04-24

Universal Multiple Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Proposed modification to Clause 26.1 of ISO/IEC 10646:2003
Source: Kyongsok Kim, Dae Hyuk Ahn
Status: National Body Position (Rep. of Korea)
Action: For consideration by JTC1/SC2/WG2
Date: 2006-04-24

Relevant Documents:

WG2 N2996 (= K1456) Clause 26.1 of ISO/IEC 10646:2003
WG2 N2994 Request to clarify text on conjoining Jamo clause 26.1
WG2 N3006 Clarification concerning Hangul Syllables 26.1

1. We came to know that some people seem to misunderstand clause 26.1, although we don't think there is any ambiguity in clause 26.1.

Until recently, there has been no confusion as to clause 26.1 since ISO/IEC 10646-1:1993 was published.

Therefore, it is desirable to clarify the meaning of clause 26.1 and explain the original design idea of Hangul U011xx Hangul Jamo.

2. Specifically, as we understand it, some people seem to misunderstand that the following combinations are valid and that they should be interpreted as a new syllable:
 - 1) Jamo (U01100-U011FF) + Precomposed Hangul syllable (U0AC00-D7A3) OR
 - 2) Precomposed Hangul syllable (U0AC00-D7A3) + Jamo (U01100-U011FF) OR
 - 3) Jamo (U01100-U011FF) + Precomposed Hangul syllable (U0AC00-D7A3)) +
Jamo (U01100-U011FF)

--> A new syllable
(e.g. U0AC00 (syllable GA) + U01102 (Jamo N) --> U0AC04 (syllable GAN))
3. However, U0AC00 is a precomposed syllable, which is not intended to be concatenated with Hangul Jamo (U011xx) any further. In other words, for example,

U0AC00 is a complete syllable by itself and cannot be concatenated with Hangul Jamo any more to make up another syllable.

Hangul syllable GAN can be represented in one of two ways:

- Precomposed syllable: U0AC04 (GAN)
- Jamo block: U01100 (G) + U01161 (A) + U011AB (N)

However, the following sequence is invalid:

U0AC00 (syllable GA) + U011AB (Jamo N)

Examples of other invalid sequences are shown below:

- 1) U01100 (G) + U0AC00 (GA) + U011A8 (G) != GGAG (U0AE4D)
- 2) U01100 (G) + U0AC01 (GAG) != GGAG (U0AE4D)
- 3) U0AE4C (GGA) + U011A8 (G) != GGAG (U0AE4D)

This is the original design idea when we submitted Jamo proposal to WG2 at Seoul meeting in 1992. The proposal (article 26.1) from Korean National Body was accepted at that WG2 meeting and was included in ISO/IEC 10646-1:1993. The contents of article 26.1 have stayed there without any changes since then.

Until recently, there has been no confusion as to clause 26.1 since ISO/IEC 10646-1:1993 was published.

4. [Responses to WG2 N2994, 2005-09-14]

4.1 Terms (1)

The terms Choseong, Jungseong, Jongseong, syllable-initial (SI), syllable-peak (SP), and syllable-final (SF) characters are explained below.

- As specified in clause 26.1 of ISO/IEC 10646, Choseong is interchangeable with a syllable-initial (SI) character (or letter).
- Similarly, Jungseong is interchangeable with a syllable-peak (SP) character (or letter).
- Similarly, Jongseong is interchangeable with a syllable-final (SF) character (or letter).

4.2 SI (syllable-initial) letters

- An SI (syllable-initial) letter can be either simple or complex (2-complex, 3-complex).
- An SI simple letter is composed of one SI simple letter: e.g., U01100, U01102

- An SI 2-complex letter is composed of two SI simple letters: e.g., U01101, U01104
- An SI 3-complex letter is composed of three SI simple letters: e.g., U01122, U01123

Note. The numbers of code positions required for representing complex letters are explained below.

4.3 The same can be said of SP (syllable-peak) and SF (syllable-final) letters.

- Examples of SP simple letters: U01161, U01163
- Examples of SP 2-complex letters: U01162, U01164
- Examples of SP 3-complex letters: U0118B, U0118C

- Examples of SF simple letters: U011A8, U011AB
- Examples of SF 2-complex letters: U011A9, U011AC
- Examples of SF 3-complex letters: U011CC, U011CF

4.4 The number of code positions required for representing complex letters

- More than one code position can be used to represent one complex letter
- An SI complex letter which does not have a single corresponding code position can be represented using more than one code position.
- A 2-complex letter can be represented
 - . Either by one code position (e.g., U01101) or
 - . By two code positions (e.g., U01100 U01102, if there is such a 2-complex letter).
- Similarly, a 3-complex letter can be represented
 - . Either by one code position (e.g., U01122), or
 - . By three code positions (e.g., U01100 U01102 U01103, if there is such a 3-complex letter)
- The same can be said of SP or SF letters.

4.5 Terms (2)

The terms a syllable block, a complete syllable block, and an incomplete syllable (block) are explained.

- As specified in clause 26.1 of ISO/IEC 10646, a complete syllable block is composed of a Choseong and a Jungseong, and optionally a Jongseong.

An incomplete syllable is a string of one or more characters which does not constitute a complete syllable (for example, a Choseong alone, a Jungseong alone, a Jongseong alone, or a Jungseong followed by a Jongseong).

The reason we excluded a Choseong followed by a Jongseong sequence is that no such combination seems to have been used.

- "An incomplete syllable" stands for "an incomplete syllable block".
- A syllable block is either a complete syllable block or an incomplete syllable block.

4.6 Representing a syllable block

- A complete syllable block can be represented
 - . Either by Precomposed syllables (U0ACxx - U0D7xx) or
 - . By a combination of Jamo (U011xx).
 - . Note that a complete syllable block cannot be represented by combining a Precomposed syllable and Jamo.
- An incomplete syllable block can be represented by a combination of Jamo (U011xx).
- Old Hangeul can be represented only by a combination of Jamo (U011xx).

Note. In this document, we are only considering Jamo (U011xx) and Precomposed syllables (U0ACxx - U0D7xx), excluding the following code positions:

- Compatibility Hangeul Jamo: U031xx (U03131 - U0318E)
- Half-width Hangeul letters: U0FFxx (U0FFA0 - U0FFDC)
- Parenthesized/Circled Hangeul syllables/letters: U032xx (U03200-U0321C, U03260-U0327B, U0327E)

5. Korean Industrial Standard KS X ISO/IEC 10646:2005

- To avoid any ambiguity and confusion, we added normative Annex 1 to Korea Standard KS X ISO/IEC 10646:2005, which was adopted in Dec. 2005 as Korean Industrial Standard (KS).
- The text in English is shown below. A complete Annex 1 (including text in Hangeul) is shown in Attachment 4.

Annex 1 (Normative). Usage of SYL-IPF letters (U01100 ~ U011FF) and Wanseong syllable blocks (U0AC00 ~ U0D7A3)

1) Since Wanseong syllable blocks (U0AC00 ~ U0D7A3) are already complete syllable blocks, they cannot be concatenated with letters such as SYL-IPF letters (U01100 ~ U011FF) to make up other syllable blocks (see clause 26.1).

2) SYL-IPF letters (U01100 ~ U011FF) can be concatenated with only SYL-IPF letters (U01100 ~ U011FF) (see clause 26.1).

3) Old Hangeul (Hangul) must be represented using only SYL-IPF letters (U01100 ~ U011FF).

4) The above rules 1), 2) and 3) apply to all processes of handling data including, but not limited to, input, output, transmission and storage; the rules also apply not only to direct processes but also to processes involving intermediate steps.

Note. 'SYL-IPF' is an acronym of 'syllable-initial, -peak, and -final'.

6. Conclusions

We propose that Annex 1 of KS X ISO/IEC 10646:2005 be added to clause 26.1 of ISO/IEC 10646 to avoid any ambiguity and confusion.

* * *

26 Special features of individual scripts

26.1 Hangul syllable composition method

In rendering, a sequence of Hangul Jamo (from HANGUL JAMO block: 1100 to 11FF) is displayed as a series of syllable blocks.

Jamo can be classified into three classes: Choseong (syllable-initial character), Jungseong (syllable-peak character), and Jongseong (syllable-final character).

A complete syllable block is composed of a Choseong and a Jungseong, and optionally a Jongseong.

An incomplete syllable is a string of one or more characters which does not constitute a complete syllable (for example, a Choseong alone, a Jungseong alone, a Jongseong alone, or a Jungseong followed by a Jongseong).

An incomplete syllable which starts with a Jungseong or a Jongseong shall be preceded by a CHOSEONG FILLER (0000 115F).

An incomplete syllable composed of a Choseong alone shall be followed by a JUNGSEONG FILLER (0000 1160).

The implementation level 3 shall be used for the Hangul syllable composition method.

NOTE 1 Hangul Jamo are not combining characters.

NOTE 2 When a combining character such as HANGUL SINGLE DOT TONE MARK (0000 302E) is intended to apply to a sequence of Hangul Jamo it should be placed at the end of the sequence, after the Hangul Jamo character which completes the syllable block.

Attachment 2: Hangul Jamo (U011xx)

Table 36 - Row 11: Hangul Jamo

	110	111	112	113	114	115	116	117
0	ㄱ 1100	ㅌ 1110	ㅍ 1120	ㅊ 1130	ㅌ 1140	ㅍ 1150	ㄱ 1160	ㅌ 1170
1	ㄲ 1101	ㄴ 1111	ㄷ 1121	ㄹ 1131	ㅁ 1141	ㅂ 1151	ㅅ 1161	ㅇ 1171
2	ㄴ 1102	ㅇ 1112	ㅅ 1122	ㅈ 1132	ㅊ 1142	ㅌ 1152	ㅍ 1162	ㅊ 1172
3	ㄷ 1103	ㄴ 1113	ㅍ 1123	ㅊ 1133	ㅌ 1143	ㅍ 1153	ㅅ 1163	ㅡ 1173
4	ㄷ 1104	ㄴ 1114	ㅍ 1124	ㅊ 1134	ㅌ 1144	ㅍ 1154	ㅅ 1164	ㅣ 1174
5	ㄹ 1105	ㄴ 1115	ㅍ 1125	ㅊ 1135	ㅌ 1145	ㅍ 1155	ㅅ 1165	ㅣ 1175
6	ㅁ 1106	ㅂ 1116	ㅅ 1126	ㅈ 1136	ㅊ 1146	ㅌ 1156	ㅍ 1166	ㅊ 1176
7	ㅂ 1107	ㅅ 1117	ㅈ 1127	ㅊ 1137	ㅌ 1147	ㅍ 1157	ㅅ 1167	ㅊ 1177
8	ㅅ 1108	ㅈ 1118	ㅊ 1128	ㅌ 1138	ㅍ 1148	ㅅ 1158	ㅈ 1168	ㅊ 1178
9	ㅈ 1109	ㅊ 1119	ㅌ 1129	ㅍ 1139	ㅅ 1149	ㅈ 1159	ㅊ 1169	ㅊ 1179
A	ㅊ 110A	ㅌ 111A	ㅍ 112A	ㅊ 113A	ㅌ 114A		ㅅ 116A	ㅊ 117A
B	ㅇ 110B	ㄹ 111B	ㅅ 112B	ㅈ 113B	ㅊ 114B		ㅅ 116B	ㅊ 117B
C	ㅅ 110C	ㅈ 111C	ㅊ 112C	ㅌ 113C	ㅍ 114C		ㅅ 116C	ㅊ 117C
D	ㅈ 110D	ㅊ 111D	ㅌ 112D	ㅍ 113D	ㅅ 114D		ㅅ 116D	ㅊ 117D
E	ㅅ 110E	ㅈ 111E	ㅊ 112E	ㅌ 113E	ㅍ 114E		ㅅ 116E	ㅊ 117E
F	ㅅ 110F	ㅈ 111F	ㅊ 112F	ㅌ 113F	ㅍ 114F	ㅅ 115F	ㅈ 116F	ㅊ 117F

G = 00
P = 00

Table 37 - Row 11: Hangul Jamo

	118	119	11A	11B	11C	11D	11E	11F
0	ㅈ 1180	ㅊ 1190	ㅋ 11A0	ㄹ 11B0	ㅍ 11C0	ㅂ 11D0	ㅅ 11E0	ㅇ 11F0
1	ㅊ 1181	ㅑ 1191	ㅣ 11A1	ㄹ 11B1	ㅍ 11C1	ㅂ 11D1	ㅅ 11E1	ㅇ 11F1
2	ㅊ 1182	ㅑ 1192	ㅍ 11A2	ㄹ 11B2	ㅍ 11C2	ㅂ 11D2	ㅅ 11E2	ㅇ 11F2
3	ㅊ 1183	ㅑ 1193		ㄹ 11B3	ㅍ 11C3	ㅂ 11D3	ㅅ 11E3	ㅇ 11F3
4	ㅊ 1184	ㅑ 1194		ㄹ 11B4	ㅍ 11C4	ㅂ 11D4	ㅅ 11E4	ㅇ 11F4
5	ㅊ 1185	ㅑ 1195		ㄹ 11B5	ㅍ 11C5	ㅂ 11D5	ㅅ 11E5	ㅇ 11F5
6	ㅊ 1186	ㅑ 1196		ㄹ 11B6	ㅍ 11C6	ㅂ 11D6	ㅅ 11E6	ㅇ 11F6
7	ㅊ 1187	ㅑ 1197		ㄹ 11B7	ㅍ 11C7	ㅂ 11D7	ㅅ 11E7	ㅇ 11F7
8	ㅊ 1188	ㅑ 1198	ㅍ 11A8	ㄹ 11B8	ㅍ 11C8	ㅂ 11D8	ㅅ 11E8	ㅇ 11F8
9	ㅊ 1189	ㅑ 1199	ㅍ 11A9	ㄹ 11B9	ㅍ 11C9	ㅂ 11D9	ㅅ 11E9	ㅇ 11F9
A	ㅊ 118A	ㅑ 119A	ㅍ 11AA	ㄹ 11BA	ㅍ 11CA	ㅂ 11DA	ㅅ 11EA	
B	ㅊ 118B	ㅑ 119B	ㅍ 11AB	ㄹ 11BB	ㅍ 11CB	ㅂ 11DB	ㅅ 11EB	
C	ㅊ 118C	ㅑ 119C	ㅍ 11AC	ㄹ 11BC	ㅍ 11CC	ㅂ 11DC	ㅅ 11EC	
D	ㅊ 118D	ㅑ 119D	ㅍ 11AD	ㄹ 11BD	ㅍ 11CD	ㅂ 11DD	ㅅ 11ED	
E	ㅊ 118E	ㅑ 119E	ㅍ 11AE	ㄹ 11BE	ㅍ 11CE	ㅂ 11DE	ㅅ 11EE	
F	ㅊ 118F	ㅑ 119F	ㅍ 11AF	ㄹ 11BF	ㅍ 11CF	ㅂ 11DF	ㅅ 11EF	

G = 00
P = 00

Attachment 3: Precomposed Hangul Syllables (U0ACxx - U0D7xx)

Table 117 - Row AC: Hangul Syllables

	AC0	AC1	AC2	AC3	AC4	AC5	AC6	AC7	AC8	AC9	ACA
0	가 AC00	감 AC10	갠 AC20	갬 AC30	갈 AC40	각 AC50	갯 AC60	거 AC70	검 AC80	겐 AC90	갬 ACA0
1	각 AC01	갑 AC11	갯 AC21	갬 AC31	갯 AC41	갈 AC51	갬 AC61	걱 AC71	겁 AC81	갯 AC91	겐 ACA1
2	갯 AC02	갬 AC12	갬 AC22	갯 AC32	갬 AC42	갬 AC52	갬 AC62	겨 AC72	겉 AC82	갬 AC92	갬 ACA2
3	갬 AC03	갬 AC13	갠 AC23	갯 AC33	갬 AC43	갬 AC53	갬 AC63	겨 AC73	겉 AC83	겐 AC93	갬 ACA3
4	간 AC04	갬 AC14	갬 AC24	갬 AC34	갬 AC44	갬 AC54	갬 AC64	건 AC74	겉 AC84	갬 AC94	갬 ACA4
5	갬 AC05	갬 AC15	갬 AC25	갬 AC35	갬 AC45	갬 AC55	갬 AC65	겉 AC75	겉 AC85	갬 AC95	갬 ACA5
6	갬 AC06	갬 AC16	갬 AC26	갬 AC36	갬 AC46	갬 AC56	갬 AC66	겉 AC76	겉 AC86	갬 AC96	갬 ACA6
7	간 AC07	갬 AC17	갬 AC27	갬 AC37	갬 AC47	갬 AC57	갬 AC67	건 AC77	겉 AC87	갬 AC97	갬 ACA7
8	갈 AC08	각 AC18	갬 AC28	가 AC38	감 AC48	갠 AC58	갬 AC68	겉 AC78	겉 AC88	갬 AC98	겨 ACA8
9	갬 AC09	갈 AC19	갬 AC29	각 AC39	갬 AC49	갬 AC59	갬 AC69	겨 AC79	겉 AC89	갬 AC99	겨 ACA9
A	갬 AC0A	갬 AC1A	갬 AC2A	갬 AC3A	갬 AC4A	갬 AC5A	갬 AC6A	겉 AC7A	겉 AC8A	갬 AC9A	겨 ACA A
B	갬 AC0B	갬 AC1B	갬 AC2B	갬 AC3B	갬 AC4B	갬 AC5B	갬 AC6B	겉 AC7B	겉 AC8B	갬 AC9B	겨 ACA B
C	갬 AC0C	갬 AC1C	갬 AC2C	갬 AC3C	갬 AC4C	갬 AC5C	갬 AC6C	겉 AC7C	겉 AC8C	갬 AC9C	겨 ACA C
D	갬 AC0D	갬 AC1D	갬 AC2D	갬 AC3D	갬 AC4D	갬 AC5D	갬 AC6D	겉 AC7D	겉 AC8D	갬 AC9D	겨 ACA D
E	갬 AC0E	갬 AC1E	갬 AC2E	갬 AC3E	갬 AC4E	갬 AC5E	갬 AC6E	겉 AC7E	겉 AC8E	갬 AC9E	겨 ACA E
F	갬 AC0F	갬 AC1F	갬 AC2F	갬 AC3F	갬 AC4F	갬 AC5F	갬 AC6F	겉 AC7F	겉 AC8F	갬 AC9F	겨 ACA F

G = 00
P = 00

부속서 1 (규정). 첫가끝 조합형 글자(U01100 ~ U011FF)와 완성형 글자마디(U0AC00 ~ U0D7A3) 용법

- 1) 완성형 글자마디(U0AC00 ~ U0D7A3)는 이미 완전한 글자마디이기 때문에, 첫가끝 조합형 글자(U01100 ~ U011FF)등의 글자와 결합하여 다른 글자마디가 될 수 없다 (참고. 26.1 절).
- 2) 첫가끝 조합형 글자(U01100 ~ U011FF)는 첫가끝 조합형 글자(U01100 ~ U011FF)끼리만 조합할 수 있다 (참고. 26.1 절).
- 3) 옛한글은 첫가끝 조합형 글자(U01100 ~ U011FF)만으로 나타내야 한다.
- 4) 위의 1), 2), 3)은 입출력, 전송, 저장 등을 포함하여 자료가 처리되는 모든 과정에 적용되며, 직접적인 경우뿐만 아니라 중간 단계를 거칠 경우에도 모두 적용된다.

Annex 1 (Normative). Usage of SYL-IPF letters (U01100 ~ U011FF) and Wanseong syllable blocks (U0AC00 ~ U0D7A3)

- 1) Since Wanseong syllable blocks (U0AC00 ~ U0D7A3) are already complete syllable blocks, they cannot be concatenated with letters such as SYL-IPF letters (U01100 ~ U011FF) to make up other syllable blocks (see clause 26.1).
- 2) SYL-IPF letters (U01100 ~ U011FF) can be concatenated with only SYL-IPF letters (U01100 ~ U011FF) (see clause 26.1).
- 3) Old Hangeul (Hangul) must be represented using only SYL-IPF letters (U01100 ~ U011FF).
- 4) The above rules 1), 2) and 3) apply to all processes of handling data including, but not limited to, input, output, transmission and storage; the rules also apply not only to direct processes but also to processes involving intermediate steps.

Note. 'SYL-IPF' is an acronym of 'syllable-initial, -peak, and -final'.