Document Number: Korea K1456

ISO/IEC JTC1/SC2/WG2 N 2996

2005-09-15

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

Doc Type: **Working Group Document**
Title: **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 and UTC**
Date: **2005-09-15**

1. Recently 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.

   At this point, it seems desirable to clarify the meaning of clause 26.1 and explain the original design idea of Hangul 11xx Hangul Jamo.

2. Specifically, as we understand it, some people seem to misunderstand that the following structures are valid and they make up a new syllable:

   1) Jamo (U1100-U11FF) + Precomposed Hangul syllable (UAC00-D7A3) OR

   2) Precomposed Hangul syllable (UAC00-D7A3) + Jamo (U1100-U11FF) OR

   3) Jamo (U1100-U11FF) + Precomposed Hangul syllable (UAC00-D7A3)) + Jamo

      (U1100-U11FF)

      --> A new syllable

   (e.g. UAC00 (syllable GA) + U1102 (Jamo N) --> UAC04 (syllable GAN))

3. However, UAC00 is a precomposed syllable, which is not intended to be concatenated with Hangul Jamo any further. In other words, UAC00 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: UAC04 (GAN)

1

- Conjoining Jamo block: U1100 (G) + U1161 (A) + U11AB (N)

However, the following sequence is invalid:

UAC00 (syllable GA) + U11AB (Jamo N)

Examples of other invalid sequences are shown below:

1) U1100 (G) + <u>UAC00 </u>(GA) + U11A8 (G)　　　!= GGAG (UAE4D)

2) U1100 (G) + <u>UAC01 </u>(GAG)　　　　　　　!= GGAG (UAE4D)

3) <u>UAE4C</u> (GGA) + U11A8 (G)　　　　　　!= GGAG (UAE4D)

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 ac-accepted at that WG2 meeting and 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. Some explanation about Choseong, Jungseong, Jongseong, syllable-initial, syllable-peak, and syllable-final

- A SI (syllable-initial) letter can be either simple or complex (2-complex, 3-complex).

- A SI simple letter is composed of one SI simple letter: e.g., U1100, U1102

- A SI 2-complex letter is composed of two SI simple letters: e.g., U1101, U1104

- A SI 3-complex letter is composed of three SI simple letters: e.g., U1122, U1123

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

- Examples of SP simple letter: U1161, U1163

- Examples of SP 2-complex letter: U1162, U1164

- Examples of SP 3-complex letter: U118B, U118C

- Examples of SF simple letter: U11A8, U11AB

- Examples of SF 2-complex letter: U11A9, U11AC

- Examples of SF 3-complex letter: U11CC, U11CF

5. The number of code positions required for representing a letter.

  - A SI complex letter which does not have a "single" corresponding code position can be represented using more than one code position.

  - The same can be said of SP or SF letters.

Attachment: Article 26.1 of ISO/IEC 10646:2003

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.

***