



ISO/IEC JTC1/SC2/WG2 N3128

2006-08-11

2006-09-25

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

Doc Type: Working Group Document
Title: Comments on Document N3113
Source: US NB and Unicode Consortium (contact: Asmus Freytag)
Status: Combined liaison and NB contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Related: N3113

In document N3113 the Korean NB proposes a ‘new model’ to handle Jamo composition. This document has been reviewed by the Unicode Technical Committee and the US National Body. This review has identified several issues.

Normalization

Document N3113 requests a change to Normalization. A very large number of important implementations and protocols, such as IETF protocols are based on Normalization. These implementations and protocols require very strict **stability** of the normalization specification. Therefore, in the Unicode context Normalization is subject to a formal *stability policy*, published at http://www.unicode.org/policies/stability_policy.html. The Unicode Consortium and the US National Body urge WG2 to not accept any changes in ISO/IEC 10646 that would adversely affect the stability of Normalization and they would definitely oppose any such changes to the Normalization specification.

Restrictions on valid sequences

Document N3113 proposes limitations on ‘valid’ sequences of characters. This is contrary to the practice of ISO/IEC 10646 and the Unicode Standard, which allows arbitrary sequences of characters, independent of whether they represent well-formed, or even useful, textual data in any particular language. Therefore, adding any such restrictions would be out of scope for ISO/IEC 10646. However, they may be present in any higher-level protocol.

Some of the restrictions proposed by document N3113 would not be stable under all normalization forms. In particular, any sequence consisting of only conjoining Jamos would be changed into a sequence containing precomposed syllables, or even a mixture of both conjoining Jamo and precomposed Hangul Syllables. In contrast to the ‘new

model' proposed in N3113, the description of *Korean Syllable Block* in the Unicode Standard is carefully designed to be stable under normalization. (Any Korean syllable block, when normalized, remains a Korean syllable block).

Additional Characters

Document N3113 makes reference to 121 new characters that are proposed for completion of the set of Jamos needed for the representation of Old Hangul Syllables, as well as certain additional filler characters.

Document N3113 does not contain an actual list of the proposed 121 Jamo characters. For that reason, neither the UTC nor the US NB were able to review these in time for the Tokyo meeting of WG2.

Filler Characters

The UTC and US NB do not see the need to add any 'filler' characters. In particular the function of, and rationale for a "high filler" is unclear. As fillers are invisible characters they require careful consideration and review before they can be accepted.

Decompositions

Existing Old Hangul complex Jamo characters do not have a canonical decomposition in the Unicode Standard. Because of normalization stability, no decompositions can be added for these characters. For consistency, the same would apply to any proposed additional Old Hangul Complex characters such as the 121 mentioned in N3113. Therefore, if it is desired to relate these characters to their constituents, it would be necessary to capture that information in some other way, for example by adding information on Jamo folding to UTS#30, "Character Foldings" (<http://www.unicode.org/reports/tr30/>).

Current implementations

UTC and US NB questioned their membership and determined that there exist implementations of fonts and layout systems that can represent Old Hangul with the currently encoded characters in the standard.

Updating Clause 26

Document N3133 proposes modifications to clause 26.1. As implied by the other comments made in this document, the proposed language in that document is not an acceptable starting point for any changes to that clause. However, the US NB and UTC agree that the specifications contained in Clause 26.1 are incomplete. In particular, they do not address the use of filler characters with sufficient detail. They also do not address the existence of 11,172 precomposed Hangul syllables and the relation to conjoining Jamos under normalization. Finally, clause 26 does not describe the Korean Syllable block as described in Unicode.

To be meaningful, any updated text for clause 26 needs to address all of these issues. Otherwise, it would be preferable to retain the status quo in 10646.