ISO/IEC JTC1/SC22/WG20 N953

Korea JTC1/SC22 (SC22-WG20) Doc #: K152

Author: KIM, Kyongsok (GIM, Gyeongseog)
Date: 2002-06-06
Subject: A summary of a paper by Kim, K: New Canonical Decomposition and
Composition processes for Hangeul

Abstract

   In this paper, the author proposed new canonical decomposition process (NFD:
Normalization form D) and composition process (NFC: Normalization Form C) for
Hangeul, Korean script. The new processes can handle both Modern and Old
Hangeul in a consistent way. Therefore, the new processews are more general
than the ones described in UAX #15 [UAX15], which handle only Modern Hangeul
syllables.
   Since NFD and NFC will play an important role in sorting and searching,
Internet-related applications such as Multilingual Domain names, databases,
etc., we need to establish good NFD and NFC for Hangeul before they are widely
used.
   The author hopes that NFD and NFC proposed in this paper will be reviewed and
reflected in ISO/IEC 14651 International String Ordering and Comparison
[ISO14651], Unicode Standard Annex #15 [UAX15], Nameprep for IDN
(Internationalized Domain Name) [Nameprep[, and other related documents.

The main idea of new Hangeul decomposition/composition processes

   Canonical decomposition and composition processes for Hangeul, as shown in UAX
#15, have limitations.
 1) UAX #15 decomposition process only decomposes 11,172 Modern Hangeul
syllables.
      Therefore, UAX #15 decomposition process cannot decompose complex (obth
Modern and Old) letters.
 2) UAX #15 composition process only composed Modern letters into syllables.
      Therefore, UAX #15 composition process cannot compose letters into Old
complex letters.

Therefore, we propose new canonical decomposition and composition processes for Hangeul, Korean script. These new processes solve the limitations of UAX #15 decomposition/composition.

A new decomposition process decomposes Hangeul letters/syllables into simple letters. Then a new composition process composes simple letters into complex letters (whenever possible), but not into syllables. These are the main idea of new decomposition and composition processes.

Tables 3 and 4 compare UAX #15 decomposition/composition and new ones.

Table 3. Comparison of UAX #15 Decomposition and a new Decomposition.

Decomposition	input	output
UAX #15	11,172 Modern syllables only	1) Modern complex letters (whenever possible) 2) Modern simple letters
a new process	1) Modern syllables 2) Modern complex letters 3) Old complex letters	1) Modern simple letters 2) Old simple letters

Table 4. Comparison of UAX #15 Composition and new Composition.

Composition	input	output
UAX #15	1) Modern complex letters 2) Modern simple letters	11,172 Modern syllables
a new process	1) Modern simple letters 2) Old simple letters	1) Modern complex letters (whenever possible) 2) Modern simple letters 3) Old complex letters (whenever possible) 4) Old simple letters

The full paper can be found at
    http://asadal.cs.pusan.ac.kr/hangeul/i18n/csi2001ab.PDF
* * *