Author: KIM, Kyongsook (GIM, Gyeongseog)
Date: 2002-06-10
Subject: Problems with the current CTT of ISO/IEC 14651 RE Hangeul

Summary: Some problems with the current CTT (Common Template Table) of ISO/IEC 14651 RE Hangeul are discussed and possible solutions are suggested.

1. Five categories of Hangeul letters/syllables in UCS

<table>
<thead>
<tr>
<th>Category</th>
<th>Letters</th>
<th>Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Hangeul IPF-Johab letters</td>
<td>240</td>
<td>--</td>
</tr>
<tr>
<td>- U+11xx (U+1100 - 11FF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) Hangeul Wanseong syllables</td>
<td>--</td>
<td>11,172</td>
</tr>
<tr>
<td>- U+A/B/C/Dxxx (U+AC00 - D7A3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) Hangeul Compatibility CV-Johab letters</td>
<td>94</td>
<td>--</td>
</tr>
<tr>
<td>- U+31xx (U+3131 - 318E)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4) Hangeul Half-width CV-Johab letters</td>
<td>52</td>
<td>--</td>
</tr>
<tr>
<td>- U+FFxx (U+FFA0 - FFDC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5) Hangeul Enclosed (Parenthesized /Circled) syllables/letters)</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>- U+32xx (U+3200-321C, 3260-327B)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* IPF=initial-peak-final, CV=consonant-vowel
2. Problems with the current CTT of ISO/IEC 14651 RE Hangeul

2.1 Hangeul Compatibility CV-Johab (U+31xx)

2.1.1 A Problem with the current CTT RE Compatibility CV-Johab

1) A problem with Letters
   - e.g., a letter "ㅏ": U+3131, which will be transformed by CTT as follows:
     --> <S1100>:<BASE>:<COMPAT>:<U3131> % HANGUL LETTER KIYEOK
   - In contrast, it is not the same as a letter "ㅏ" in IPF-Johab
     <U1100> <U1160>, which will be transformed by CTT as follows:
     --> <S1100> <S1160>; <BASE> <BASE>;<MIN> <MIN>; <U1100><U1160>
   - As a result, the two will not compare equal even at level 1, which is incorrect.

2) A problem with Syllables
   - e.g., a syllable "ㅏ": U+3164 3131 314F 3164 U+3131, which will be transformed by CTT as follows:
     <S1160><S1100><S1161><S1160>; <BASE><BASE><BASE><BASE>;
     <COMPAT><COMPAT><COMPAT><COMPAT>; <U3164><U3131><U314F><U3164>;
   - In contrast, it is not the same as a letter "ㅏ" in IPF-Johab
     <U1100> <U1161>, which will be transformed by CTT as follows:
     --> <S1100> <S1161>; <BASE> <BASE>;<MIN> <MIN>; <U1100><U1161>
   - As a result, the two will not compare equal even at level 1, which is not correct.

3) The relevant portion of CTT

   <U1100> <S1100>:<BASE>:<MIN>:<U1100> % HANGUL CHOSEONG KIYEOK
   <U1160> <S1160>:<BASE>:<MIN>:<U1160> % HANGUL JUNGSEONG FILLER

   <U3131> <S1100>:<BASE>:<COMPAT>:<U3131> % HANGUL LETTER KIYEOK
   <U314F> <S1161>:<BASE>:<COMPAT>:<U314F> % HANGUL LETTER A
   <U3164> <S1160>:<BASE>:<COMPAT>:<U3164> % HANGUL FILLER

2.1.2 A proposal solution RE Compatibility CV-Johab letters
   - Since filler characters U+115F and U+1160 in IPF-Johab and U+3164 in Compatibility CV-Johab have drastically different usage, a blind transformation in the current CTT produces incorrect results. In other words, CTT does not take into consideration the preceding letters and, therefore, cannot transform Compatibility CV-Johab
   - A reasonable solution will be to preprocess Compatibility CV-Johab letters so that they are transformed into IPF-Johab letters and therefore Compatibility CV-Johab letters should not be processed by CTT.
2.2 Hangeul Halfwidth CV-Johab letters (U+FFxx)
2.2.1 A Problem with the current CTT RE Halfwidth CV-Johab letters
   - The usage of Halfwidth CV-Johab is different from IPF-Johab or Compatibility CV-Johab letters.
   - The problem is somewhat similar to, though not identical with, that of Compatibility CV-Johab letters.
   - The details are not shown here.

2.2.2 A proposed solution RE Halfwidth CV-Johab letters
   - A reasonable solution will be to preprocess Halfwidth CV-Johab letters so that they are transformed into IPF-Johab letters and therefore Halfwidth CV-Johab letters should not be processed by CTT.

2.3 Enclosed (Parenthesized, Circled) Hangeul letters (U+32xx)
2.3.1 Problems with the current CTT RE Enclosed Hangeul letters
1) e.g.,
   <U3200> --> <S1100>;<BASE>;<COMPAT>;<U3200> % PARENTHESIZED HANGUL KIYEOK
   <U3260> --> <S1100>;<BASE>;<CIRCLE>;<U3260> % CIRCLED HANGUL KIYEOK

2) Independent Hangeul letter KIYEOK is represented in IPF-Johab as U+1110 1160
   - Therefore, two enclosed Hangeul letters will not be equal to independent letters at level 1.

2.3.2 A proposed solution RE Enclosed letters
   - A reasonable solution will be to change the lines in CTT as follows so that enclosed letters will be equal to independent letters at level 1:

   <U3200> --> <S1100><S1160>;<BASE><BASE>;<COMPAT><COMPAT>;<U3200>
   % PARENTHESIZED HANGUL KIYEOK
   <U3260> --> <S1100><S1160>;<BASE><BASE>;<CIRCLE><CIRCLE>;<U3260>
   % CIRCLED HANGUL KIYEOK
2.4 Old Hangeul complex letters not included in UCS
- Actually, I am not pointing out problems of CTT; rather, I explain how to
treat complex letters not included in UCS.
- Korean scholars claim that they found tens of old complex letters not
included in UCS.
- Fact: Although the collating sequence for Modern Hangeul letters are well
defined, no well-defined collating sequence for modern and old letters combined
exists.
- Complex letters are treated as one unit and, therefore, newly found Old
Hangeul complex letters must be defined as collating-element to sort properly:
collating-element `<Uxxxx_yyyy>' from `<Uxxx><Uyyyy>'
collating-element `<Uxxxx_yyyy_zzzz>' from `<Uxxx><Uyyyy><Uzzzz>'

2.5 Old Hangeul Tone marks (Bangjeom)
- As mentioned in 2.4 above, there is no widely accepted collating sequence
for Old Hangeul letters. Therefore, the way to treat tone marks is not well
defined yet either.
- We need a further investigation.

***