

Appendix D: Corrections to Volume 1

Appendix D contains corrections to Volume 1. The corrections listed here are limited to those cases where an error in content could cause difficulty to implementors of the Unicode standard. A page-by-page list of minor typographical, spelling, and punctuation errors is not included. All such errors will be corrected in version 1.1.

This appendix contains:

- Corrections to the Bidirectional Algorithm
- A clarification of the Asian standards mapping table
- A page-by-page list of significant errata

Bidirectional Errata

The following section corrects a number of typographic errors or omissions in the Bidirectional Algorithm of version 1.0 of the Unicode standard. (See Volume 1, Appendix A.) Each item provides corrected text to be inserted at the location indicated. Corrected text in this appendix may be boldfaced for clarity.

Page 614

Before the heading “Resolving Embedding Levels,” insert the following text:

Logically speaking, non-spacing marks attach to base characters before the bidirectional algorithm is applied. A non-spacing mark applies to the preceding base character in the backing-store (except in some cases of Indic character reordering). Hence, the glyph representing a non-spacing mark does not necessarily attach to the glyph which is visually on its left. Depending on the direction of the text, it may, for example, attach to the glyph on the left, or on the right, or above.

Page 614

The characters RLE and LRE should be added to the entry under B1. The characters RLE should be added to the entry under B2. The final paragraph should include the term “global direction.” The corrected rules should read:

- B1. In the text, find the first strong directional character, RLE, LRE, RLO, or LRO.
(Because block separators delimit text in this algorithm, this will generally be the first strong character after a block separator or at the very beginning of the text.)
- B2. If the first strong directional character in the text is right-to-left, RLE, or RLO, then set the base level to one, otherwise set it to zero.

The direction of the base embedding level is called either the *base direction*, the *global direction*, or the *block direction*.

Page 616

T5 should read:

- T5. All explicit directional embeddings and overrides are completely terminated at block separators. Return to the state as of B1.

All overrides and resolution of numbers and neutrals take effect within the bounds of an embedding. That is, nothing within an embedding or override will affect the character direction of codes outside of that embedding, and vice versa. The one exception is in resolving neutrals (see N4).

Page 617

Rule N3 does not clearly specify the behavior of neutrals mixed with numbers that are not in the format of the subrules in N3. Some of this behavior can be derived from the example following N4, which has a list of numbers separated by numerals. However, for clarity rule N3 is to be replaced by the following:

- N3. For the purposes of resolving neutrals,
- (a) English numbers are treated as though they were the type of the previous strong letter.
 - (b) If there is no previous strong letter, English numbers are treated as though they had the base direction.
 - (c) Arabic numbers are treated as though they were R but do not affect the treatment of English numbers as in (a) and (b).

The following are examples:

R n EN n R	=>	R R EN R R
R n EN n L	=>	R R EN e L
L n EN n R	=>	L L EN e R
L n EN n L	=>	L L EN L L
R n AN n R	=>	R R AN R R
R n AN n L	=>	R R AN e L
L n AN n R	=>	L e AN R R
L n AN n L	=>	L e AN e L

After Rule N4, the first example should read:

```
R N [LRO <text> PDF] N L => R e LRO <text> PDF L L
R N [RLE <text> PDF] N L => R R RLE <text> PDF e L
```

The third example should read:

```
Storage:      IT IS A bmw 500, OK.
Visual:      KO ,bmw 500 A SI TI.
```

Clarification of Asian Standard Mappings

The values in the XJIS column contain mappings of non-ideographic characters from the JIS X 0208 and JIS X 0212 standards. For characters from the JIS X 0208 standard, code points are given in the so-called Shift-JIS encoding. The relationship between this Shift-JIS encoding and <actual row, column> notation of the JIS standard is the same as that given by the RowCellToShiftJis() function in “Unicode Encoding to Macintosh Character Mappings” in Section 6.2 of Volume 1, Vendor Mapping Tables. Characters from JIS X 0212 are given in a hexadecimal encoding rrcr with the first two hexadecimal digits providing the row number and the other two digits providing the column number according to this equation:

$$rrcc = (\text{row} + 0x20) * 256 + (\text{cell} + 0x20)$$

Hexadecimal numbers starting with digit 2 refer to characters from JIS X 0212. All other characters refer to characters from JIS X 0208, with the exception of numbers starting with 00, which refer to single-byte characters found in common implementations of Shift-JIS encodings, and which are not part of the JIS standards referenced. Note the Shift-JIS encodings are vendor-specific implementations of the JIS standards, and should not be confused with the JIS standards themselves.

Other Errata

This section contains a list of errors that were considered to be significant enough that they might have become causes of misinterpretation of the Unicode standard. Minor typographical errors are not included.

Back Cover

Digital Equipment Corporation was inadvertently omitted from the last paragraph of the back cover text as an active member of the Unicode consortium. The Consortium regrets the omission.

Page 4

The five scripts that are listed as included in draft form (Ethiopian, Burmese, Khmer, Sinhala, and Mongolian) were omitted from Volume 1 at the last minute.

Page 14

The sentence “5.5K of user space has been allocated in the range U+E800→U+FDFF” should read “5.5K of user space has been allocated in the range U+E000→U+F7FF.”

Page 68

U+0E46 THAI MAY YAMOK should be THAI MAI YAMOK.

Page 69

Thai Encoding Structure: U+0E01 should be U+0E00.

Page 69

U+0E24 and U+0E26 have been reclassified as vowels.

Page 102

Mandarin tone marks: The Unicode value for the fourth tone, given as U+2CB0 MODIFIER LETTER GRAVE should be U+02CB. The Unicode value for the light tone, given as U+0209 SPACING DOT ABOVE should be U+02D9.

Page 106

A standard romanization of the Japanese era names “Showa,” “Taishou” and “Meiji” are given in the CJK Squared Words Character Block: 337C Syouwa, 337D Taisyuu, and 337E Meiji.

Page 112

The correct name of the original Japanese character set should be JIS C 6226-1978, not JIS X 0208-1978.

Pages 114, 116, 648

The correct name of the Korean character set should read KS C 5601-1989, not KS C 5601-1987.

Pages 153, 332, 333, 584

As a result of the merger with ISO IS 10646, U+3004 is now a duplicate of U+4EDD and should not be used. U+4EDD should be used instead.

Page 230, 232

The glyph representing BENGALI LETTER YYA U+09DF is depicted using U+09B7 BENGALI LETTER SSA + U+09BC BENGALI SIGN NUKTA. The correct glyph should be: U+09AF BENGALI LETTER YA + U+09BC BENGALI SIGN NUKTA.

Page 262, 264

U+0E4E THAI YAMAKKAN is to be treated as a non-spacing mark. The glyph should be represented in the block chart as floating above a dotted circle, similar to U+0E4D THAI NIKKHAHIT.

Page 328

Unicode value 27B0 is currently unassigned (the cell should be blank). The Names List value (unassigned) is correct.

Page 470

The characters coded A2 and A1 in ISO 8859-7 (Greek) were originally mapped to U+0371 GREEK NON-SPACING DASIA PNEUMATA and U+0372 GREEK NON-SPACING PSILI PNEUMATA. These mappings were relinquished in favor of spacing forms, as part of the Unicode standard/ISO IS 10646 merger. The new mappings for these ISO 8859-7 characters should be:

02BC	A2	MODIFIER LETTER APOSTROPHE
02BD	A1	MODIFIER LETTER REVERSED COMMA

Page 517

The Apple logo occupies code point F7FF, not FDFE.

