Recommendations from the UTC#103 Ad-hoc group on document L2/05-093 "Some comments on Unicode line-breaking, mirroring, and bidi line processing"

2005-05-13

Ad hoc members;
Asmus Freytag, Eric Mueller, Ken Whistler, Lisa Moore, Markus Scherer, Michel Suignard.

Line break table

Many visual rearrangements of the line break table are possible. For ease of maintenance the current scheme will be maintained. The suggested changes in properties, while seemingly minor, often reflect input from specific constituents.

Line break properties

The ad-hoc reviewed the proposed changes in linebreak class from the document (for details see the notes below) and arrived at the following recommendations:

Recommendation: The line break properties for the following seven characters should be changed
1735 and 1736 from AL to BA
17D9 from NS to AL
203D, 2047..2049 should become NS (from AL) to match 203C.

Recommendation: The gc and lb values of the following characters should be investigated: 3164 and FFA0

On 00A1 and 00BF the ad-hoc notes that these have lb value of AI because they occur in EA legacy character sets as wide characters. AI can only be resolved to AL.

Recommendation: A note should be placed into the text of UAX#14 to note that these could be tailored to OP, for their use in Spanish.

Some Notes:
In particular, the ad-hoc notes that the detailed behavior of C0 and C1 controls outside the common subset could be tailored for potentially better results in systems where such controls exist in data streams, but that the issue is not critical to burden the default implementation with it.

The LB assignment for Thai conforms to actual implementations.

The monogram, digram, tetragram characters would not be expected to occur in sequences, therefore, they do not need the ID property, even in EA contexts.

The ornament characters are explicitly treated as symbols and not as punctuation marks due to the fact that they were encoded as dingbats.

The quote marks were deliberately assigned ambiguous opening/closing semantics in 4.1.0 since they can be used ambiguously.

Mirroring

Changing the mirroring of a character destabilizes all existing documents using this character. However, incorrect or inconsistent mirroring can create a usability problem.

Recommendation: The ad-hoc recommends that a Public Review Issue be prepared to ask for the creation of a formal stability policy that would make the Mirrored property immutable. The choices offered should be:
1. Adopt such a policy
2. Adopt such a policy after making some final corrections
3. Do not adopt such a policy

Bidi

The ad-hoc recommends no action on the issues of leading spaces and paragraph embedding level after receiving e-mail comments suggesting that the issue is not severe enough to justifying adjustments to the algorithm.

Recommendation: The ad hoc recommends that the bidi committee to review the issue raised around the removal of SHY in the line break algorithm, and suggest that the bidi committee provide clear guidelines on how to handle SHY in a bidi context. Such guidelines could be added to the description of SHY in UAX#14 or to UAX#9 or both.