

## ***JTC1/SC2/WG2 N3396***

**Date: February 8, 2008**

**L2/08-125**

<b>Title:</b>	<b>Comments on FCD 19757-7</b>
<b>From:</b>	<b>INCITS/L2 (US Tag to SC2)</b>
<b>To:</b>	<b>SC2</b>
<b>cc:</b>	<b>INCITS/V1 (US Tag to SC34)</b>

In document SC2 N3997, JTC 1/SC 34 solicits input from SC 2 on the FCD of ISO/IEC 19757-7, Information technology - Document Schema Definition Languages (DSDL) - Part 7: Character Repertoire Description Language (CRDL).

The US NB appreciates the opportunity offered by SC 34.

1. The name of U+000D is incorrectly spelled CARIAGE instead of CARRIAGE (multiple times)
2. In addition to the union, intersection and difference operations, the symmetric difference operation is often useful.
3. We do not understand the requirement for kernel and hull. The document would benefit from detailed examples that illustrate the benefit of these constructs.
4. The syntax appears to allow for the specification of a kernel that does not contain characters in the hull, which would be incoherent.
5. The representation of characters by themselves may not survive normalization. For example, suppose that you have a collection described by the single "<char>a&#x308;</char>". According to the Unicode standard, that is equivalent to "<char>&#xE4;</char>". Suppose that two implementations receive the XML format, where in transmission to the second one the text has been normalized according to NFC (as permitted by Unicode), then the two implementation's interpretations of the values would be different. This could be fixed by specifying that the interpretation of the contents of <char> is always the text after normalization to NFC, or that the text must be in NFC to start with. Failing that, at least a very strong warning should be supplied.
6. In section 6.4, you probably want to add something along the lines of "and has the same semantics" at the end of the first paragraph.
7. The regular expression in 6.4 does not include all the properties listed in requirement RL1.2 of Unicode Technical Standard #18, Unicode Regular Expressions.

8. In section 7.7, including some form of version of the registry would improve the reliability. For example, "<repertoire registry='CLDR' version='1.5' ...>"

9. The use of `minUcsVersion` and `maxUcsVersion` can lead to confusion. For example, there is no guarantee that the category of a character remain constant accross versions, so "`\p{Po}`" together with `minUcsVersion=2.0` and `maxUcsVersion=5.0` is ill-defined; is it those characters that were `Po` in any version between 2.0 and 5.0 or in all versions?