Subject: UCA 6.2 Conformance Test change
From: Mark Davis
Live Doc: http://goo.gl/f1uRc

Proposal

Change the UCA conformance test and the CollationTest.html documentation to use NFD comparison instead of code point comparison.

Background

Richard Wordingham recently noticed an issue in the UCA collation test file. The test follows the documentation in CollationTest.html, which reads as follows.

1. Read the next line.
2. Parse each sequence up to the semicolon, and convert it into a Unicode string.
3. **Compare that string with the string on the previous line, according to the UCA implementation.**
4. If the last string is less than the current string, continue to the next line (step 1).
5. If the last string is greater than the current string, then stop with an error.
6. **Compare the strings in code point order.**
7. **If the last string is greater than the current string, then stop with an error**
8. Continue to the next line (step 1).

However, line 6 is disconcerting, because it is not actually what is used in UCA for a tie-breaker, which has:

> **S3.10** If a semi-stable sort is required, then after all the level weights have been added, append a copy of the NFD version of the original string. This strength level is called the *identical level*. (See also *Appendix A, Deterministic Sorting*.)

Part of the problem is that the conformance test does not say what level it is testing.

While not formally incorrect, it would be much less confusing for users of the conformance test file if we matched S3.10. This would require a (simple) change to the conformance test file, and a change to the documentation.

This would take a change to step 3, to read:

3. **Compare that string with the string on the previous line, according to the UCA implementation, with strength = identical level (using S3.10).**

Remove steps 6, 7.

6. **Compare the original strings in code point order.**
7. **If the last string is greater than the current string, then stop with an error**