

**Title:** A new test file for Bidi  
**Author:** Roozbeh Pournader (Google)  
**Date:** 2013-01-31  
**Action:** For UTC's consideration

## **Background**

At UTC #132, I was asked to provide a sample bidi test file, named BidiCharacterTest.txt. This was intended as an end to end test, to make sure characters like U+002F SOLIDUS are assigned the correct bidi class and the final text is reordered properly. The file can also be used to provide test data for UBA changes suggested by Microsoft for paired brackets.

This document provides sample lines for BidiCharacterTest.txt for discussion and potential approval of the UTC. Considering the timing of the changes to the UBA, the first version of such file would be provided by Microsoft, containing bracket test data.

The file shouldn't try to be thorough and exhaustive like BidiTest.txt, but selective in only providing interesting cases for implementations and the edge cases for testing. BidiTest.txt will continue to exist for exhaustive tests.

## **File format**

Here is my suggestion for the file format, based on the similar Unicode data files:

```
# BidiCharacterTest-6.3.0-txt
# Date: ...
#
# Unicode Character Database
# Copyright (c) 2013 Unicode, Inc.
# For terms of use, see http://www.unicode.org/terms\_of\_use.html
# For documentation, see http://www.unicode.org/reports/tr44/
#
# Each line contains four fields, separated by a semicolon.
#
# Field 0: The source string, as a list of codepoints, separated by
```

```

#   space.
#
# Field 1: The resolved paragraph embedding level according to
#   steps P2 and P3 of the UBA.
#
# Field 2: The list of resolved levels after running the algorithm.
#   For characters removed in step X9 of UBA, the level is shown by a
#   lowercase "x".
#
# Field 3: The order of characters, when reordered according the UBA,
#   with the leftmost character first. Characters removed in step X9
#   are not listed in this field.
#
061C;1;1;0
05D0 2067 202A 0041;1;1 1 x 4;3 1 0
062A 0031 002F 0032;1;1 2 2 2;1 2 3 0

```