Feedback on PRI 185: Revision of UBA for improved display of URL/IRIs

By Jonathan (Jony) Rosenne, July 26, 2011.

The PRI suffers, in my view, from the following issues: feasibility and stability. I offer a counter proposal.

Introduction

The introduction concentrates on mixed script cases. It misses the obvious:

Memory order:

    HTTP://EF.GH/IJ/KL

Display order:

    LK/JI/HG.FE//:PTTH

This should be considered the primary example for bidi URIs. The mixed script items are less important edge cases, which, though they certainly must be addressed, should not be allowed to drive the discussion.

In the interim phase, until the bidi countries establish ccTLDs in their own scripts, as some have done, see http://www.icann.org/en/topics/idn/fast-track/string-evaluation-completion-en.htm (with the notable omission of Israel), one may expect a significant number of mixed script URLs, but once the ccTLD process moves forward their use will diminish quickly, because they just do not make sense from the end user's point of view.

I assume that the schemes, such as http or file, will get suitable translations to the RTL script used.

Feasibility

The proposal, as hinted in the background document, suffers from grave feasibility problems.

This issue is actually questioned in the PRI and in the background document. My answer is a clear and simple NO.

The very success of Unicode in general and the UBA in particular and their widespread adoption across innumerable platforms using a large number of implementations ensure that any change will take years to accomplish for most implementations and for some may never even be attempted.

Until the emergence of Unicode and the UBA, the implementation of Hebrew and Arabic was considered difficult and expensive, and many vendors thought the cost could not be justified and just did not do it. With the advent of UBA, a number of major and influential vendors decided to

invest in bidi and implement Arabic and Hebrew in their products, and other vendors either based their products on the available platforms or implemented bidi on their own. Many commercially available implementations are still far from perfect, but the vendors do not feel any further investment should be given priority and that we – the users – can live with it.

The proposal is for a complex change in the algorithm. Just take a look at the proposed, admittedly incomplete, BNF. Most of the smaller vendors will not even take a second look at it.

Another obstacle to feasibility is user expectations. Users currently are often confounded by the results of mixed script text, whether the UBA is perfectly or imperfectly implemented, and the addition of the PRI will only increase the confusion, even were it to have been implemented perfectly and simultaneously across all platforms.

Stability

The proposed change violates the spirit, even if not the letter, of the Unicode stability guarantee. After the change, bidi text which is currently displayed in a certain manner may be displayed differently.

Counter Proposal

I cannot see why pasting a URI into a bidi document should be treated differently than any other cut and paste operation. In any case the directionality surrounding the cut text and the directionality surrounding the point of parsing may be different and current implementations do address this issue.