To: UTC and script ad-hoc

Title: Recommendation for UTR #53 to become a UAX

From: Roozbeh Pournader, Ken Whistler, Lorna Evans, Bob Hallissy

Date: 2 October 2023

Introduction

UTC # 176 gave an action item:

> [176-A122] Action Item for Roozbeh Pournader and Lorna Evans, SAH: Write a document recommending UTR #53 become a UAX, listing related actions that are needed.

From the tr-template.html, a UTR is:

> ***A Unicode Technical Report (UTR)*** *contains informative material. Conformance to the Unicode Standard does not imply conformance to any UTR. Other specifications, however, are free to make normative references to a UTR.*

And a UAX is:

> ***A Unicode Standard Annex (UAX)*** *forms an integral part of the Unicode Standard, but is published online as a separate document. The Unicode Standard may require conformance to normative content in a Unicode Standard Annex, if so specified in the Conformance chapter of that version of the Unicode Standard. The version number of a UAX document corresponds to the version of the Unicode Standard of which it forms a part.*

UTR#53 was first published in 2018. It has been through several revisions since that time. However, those revisions have not been technical changes except with regard to additions to Unicode and the MCM.

The authors believe it is now time to advance UTR#53 to become an integral part of the Unicode Standard as a UAX. In addition, we believe that MCM should become a formal binary property of the standard.

Rationale for moving from a Technical Report to an Annex

- Displaying Unicode Arabic text without knowing the order that should be used for combining marks is an underspecified task and would require research from implementors. The Unicode Standard should be clear about how to do that, and making #53 part of the Unicode Standard makes it clear that the Unicode Standard itself specifies the order of display for Arabic combining marks.
- Representing Arabic text with multiple combining marks, common in religious or pedagogical texts, is error-prone without knowing how it would get displayed. By making #53 part of the standard, people creating Unicode text can be confident that they are representing their text in Unicode correctly.
- The content of #53 was originally proposed to be a part of the Core Spec. The UTC thought that a separate publication would make it more reviewable. Now that it's mature enough, we can make it an integral part of the standard.
- Implementations of the Unicode Standard need to know where to display any newly encoded Arabic combining mark, and cannot be expected to wait for an update to #53 in order to figure that out. (This means that at the very least, #53 should be published simultaneously with any new version of the Unicode Standard that includes a new Arabic combining mark.)

Suggested decisions and actions

1. Request a decision from the UTC to change the status from a UTR to a UAX.

Recommendation for UTR #53 to become a UAX

a. **Recommendation:** We recommend the UTC make the following disposition:
The UTC authorizes the UTR #53 to become a UAX.

2. Request a decision from the UTC to open a PRI for a proposed update UAX #53, for 16.0.
    a. **Recommendation:** We recommend the UTC make the following disposition:
    The UTC authorizes a PRI on changing the status of UTR #53 to UAX #53.

3. Request UTC add a new normative, binary character property, `Modifier_Combining_Mark` (short name: MCM) to the UCD, for 16.0.
    a. Recommendation: We recommend the UTC make the following disposition:
    The UTC authorizes the addition of a new normative, binary character property, Modifier_Combining_Mark (short name: MCM) to the UCD, for 16.0.

4. Action Items:
    a. **Action Item** for Lorna Evans, Roozbeh Pournader, and Bob Hallissy, SAH to prepare a proposed update UAX #53, for 16.0.
    b. **Action Item** for Rick McGowan, UTC to post the PRI for PU-UAX #53, to close 2024-xx-xx.
    c. **Action Item** for Rick McGowan, UTC to adjust the `reports/index.html` page for the changed status for 16.0.
    d. **Action Item** for Ken Whistler, EDC to add UAX #53 to the list of annexes in the core spec for 16.0.
    e. **Action Item** for Peter Constable, RMG to add UAX #53 to BRS and other release process documents for 16.0.
    f. **Action Item** for Markus Scherer, PAG. Add MCM property values to `PropList.txt`, for 16.0.
    g. **Action Item** for Markus Scherer, PAG. Add property and property value aliases for MCM to `PropertyAliases.txt` and `PropertyValueAliases.txt`, for 16.0
    h. **Action Item** for Ken Whistler, PAG. Add documentation of MCM to property lists in UAX #44, for 16.0
    i. **Action Item** for Ken Whistler, EDC. Add AMTRA to the list of Unicode algorithms in the core spec, for 16.0

Informational notes:

Ordinarily, for a new UCD property, we would also need to record actions for UAX #42 and UCDXML, but the status of UAX #42 is up in the air for 16.0, so we believe these can be omitted for the moment until that is resolved.

For a small binary property like this, it probably does not warrant a new data file, when it could clearly just be added as another small section of PropList.txt

Note that when we add a property MCM this way, that will imply adjusting the text of Section 3.1 of UTR #53 for the proposed update, as the list of characters will no longer be specified normatively in Section 3.1, but by the data file. The text should list examples of MCM=Yes characters, including those most prominently discussed in the examples, but then should say, See PropList.txt in the UCD for the full enumeration of the set of characters affected by AMTRA.

Recommendation for UTR #53 to become a UAX