L2/11-051
2011-02-03

Title: **Response to UTC Suggestions Regarding Thai in UAX#29**
Author: **Martin Hosken**
Action: **For consideration by UTC**
Date: **2011-02-03**

Proposal: This document proposes the following changes to UAX#29, with immediate effect:

- Remove all characters in the includes list for prepending characters: U+0E40-U+0E44, U+0EC0-U+0EC4, U+AAB5, U+AAB6, U+AAB9, U+AABB, U+AABC.
- Remove listed Thai and Lao characters from Spacing Marks (U+0E30, U+0E32, U+0E33, U+0E45, U+0EB0, U+0EB2, U+0EB3)
- Remove Thai examples from Table 1a, Extended Grapheme Clusters.

Introduction: This document follows on from L2/10-460 and contains the same proposal.

Rationale: The response to L2/10-460 was:

> Respond to Martin Hosken: "We believe this has been addressed by the legacy grapheme clusters specified in version 6.0 of UAX #29. If this is not the case, please describe the changes that you would like in the 6.0 version. If you are referencing particular implementations like CLDR, then you might address yourself to those implementations. We will update UAX #29 to illustrate this."

While this does allow users of the annex the option of doing the right thing with regard to the scripts, it does not discourage them from doing the wrong thing. UAX#29 states:

> The extended grapheme cluster boundaries are recommended for general processing, while the legacy grapheme cluster boundaries are maintained for backwards compatibility with earlier versions of this specification.

And this clearly states that the default behaviour for these scripts, therefore, should link prepending characters to the following consonant for these scripts. But that behaviour is not wanted. No argument has been given as to why this new behaviour (of linking prepended characters with the following consonant) is of benefit to these scripts. So instead of making the required and default behaviour something that must be tailored in for every language that uses the script, the default behaviour should be made the default behaviour.