Ruby master - Feature #14802
Update Unicode data to Unicode Version 11.0.0
06/01/2018 05:16 AM - duerst (Martin Dürst)

Status: Closed
Priority: Normal
Assignee: duerst (Martin Dürst)
Target version: 2.6

Description
Unicode Version 11.0.0 will be published sometimes later this year, probably in late June. This is an issue to manage updating Ruby to Unicode 11.0.0. Details to follow.

Related issues:
Related to Ruby master - Feature #13685: Update Unicode data to Unicode Version... Closed
Blocked by Ruby master - Feature #14839: How to deal with capitalizing Georgian in Unicode 11.0.0 Closed
Blocked by Ruby master - Feature #15182: Update extended grapheme cluster impl... Closed
Blocked by Ruby master - Feature #15317: How to deal with obsolete property v... Closed
Blocks Ruby master - Feature #15321: Update Unicode data to Unicode Version 1... Closed
Blocked by Ruby master - Bug #15337: String#each_grapheme_cluster wrongly spl... Closed

Associated revisions
Revision 66295 - 12/09/2018 02:02 AM - duerst (Martin Dürst)
delete Unicode 10.0.0 related files, no longer needed [#14802]
This line, and those below, will be ignored--
D enc/unicode/10.0.0

Revision 66295 - 12/09/2018 02:02 AM - duerst (Martin Dürst)
delete Unicode 10.0.0 related files, no longer needed [#14802]
This line, and those below, will be ignored--
D enc/unicode/10.0.0

History
#1 - 06/01/2018 05:17 AM - duerst (Martin Dürst)
- Related to Feature #13685: Update Unicode data to Unicode Version 10.0.0 added

#2 - 06/01/2018 09:29 AM - shevegen (Robert A. Heiler)
All power to the emoji. \o/

#3 - 06/06/2018 05:38 AM - duerst (Martin Dürst)
Unicode Version 11.0.0 has been published, the official announcement can be found at http://blog.unicode.org/2018/06/announcing-unicode-standard-version-110.html.

#4 - 06/10/2018 01:04 AM - duerst (Martin Dürst)
- Blocked by Feature #14839: How to deal with capitalizing Georgian in Unicode 11.0.0 added

#5 - 09/27/2018 10:03 AM - naruse (Yui NARUSE)
Just a note, the definition of extended grapheme cluster is changed in Unicode 11 (Unicode® Standard Annex #29 UNICODE TEXT SEGMENTATION revision 33: https://www.unicode.org/reports/tr29/tr29-33.html)
This affects Regexp /\X/ which is hardcoded in node_extended_grapheme_cluster() in regparse.c.

```ruby
{ CRLF
  | Prepend
  { RI-sequence | Hangul-Syllable | !Control }
  { Grapheme_Extend | SpacingMark }*
  | . }
crlf
```

04/09/2020
naruse (Yui NARUSE) wrote:

Just a note, the definition of extended grapheme cluster is changed in Unicode 11

This is mentioned at http://www.unicode.org/versions/Unicode11.0.0/, so I was (vaguely) aware of it, but thanks for the reminder. I have created a subissue at #15182. I may have to get back to you for some help, but first I have to fight with #14902 :-(.

Some hints for future Unicode updates:

- Check early whether modification to algorithms,... are necessary.
- For tests, these are the main ones:
  test/test_unicode_normalize.rb
  test/ruby/enc
  test/ruby/test_m17n*
  test/ruby/test_regexp.rb
  test/ruby/test_string*
- There are also some specs involved, so make sure to check them, too.