**Ruby master - Feature #14802**

**Update Unicode data to Unicode Version 11.0.0**

06/01/2018 05:16 AM - duerst (Martin Dürst)

<table>
<thead>
<tr>
<th>Status:</th>
<th>Closed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
</tr>
<tr>
<td>Assignee:</td>
<td>duerst (Martin Dürst)</td>
</tr>
<tr>
<td>Target version:</td>
<td>2.6</td>
</tr>
</tbody>
</table>

**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 10.0.0 added
- Related to Ruby master - Feature #17750: Update Unicode data to Unicode Version 10.0.0 added
- Blocked by Ruby master - Feature #14839: How to deal with capitalizing Georgian in Unicode 11.0.0 added
- Blocked by Ruby master - Feature #15182: Update extended grapheme cluster implementation added
- Blocked by Ruby master - Feature #15317: How to deal with obsolete property v... added
- Blocks Ruby master - Feature #15321: Update Unicode data to Unicode Version 1... added
- Blocked by Ruby master - Bug #15337: String#each_grapheme_cluster wrongly spl... added

**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 regexp.c.

```c
{ CRLF
 | Prepend*  
 | RI-sequence | Hangul-Syllable | !Control }*
{ Grapheme_Extend | SpacingMark }*  
| . }```
#6 - 10/01/2018 02:24 AM - duerst (Martin Dürst)
- Blocked by Feature #15182: Update extended grapheme cluster implementation for Unicode 11 added

#7 - 10/01/2018 03:05 AM - duerst (Martin Dürst)
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 #14802 :-(.

#8 - 11/18/2018 09:36 AM - duerst (Martin Dürst)
- Blocked by Feature #15317: How to deal with obsolete property values in Unicode 11.0.0 added

#9 - 11/20/2018 06:58 AM - duerst (Martin Dürst)
- Blocks Feature #15321: Update Unicode data to Unicode Version 12.0.0 added

#10 - 11/24/2018 08:52 AM - duerst (Martin Dürst)
- Blocked by Bug #15337: String#each_grapheme_cluster wrongly splits "irin" added

#11 - 12/10/2018 04:46 AM - duerst (Martin Dürst)
- Status changed from Open to Closed

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.

#12 - 07/04/2021 02:10 AM - duerst (Martin Dürst)
- Related to Feature #17750: Update Unicode data to Unicode Version 13.0.0 added