Update Unicode data to Unicode Version 13.0.0

03/26/2021 06:48 AM - duerst (Martin Dürst)

Description

The newest version of Unicode is Unicode 13.0.0 since about one year (see [http://www.unicode.org/versions/Unicode13.0.0/](http://www.unicode.org/versions/Unicode13.0.0/)). We should finally update Ruby to use Unicode 13.0.0, and probably also backport the result to Ruby 3.0. This issue serves as the main issue for this update; if necessary, I'll create sub-issues.

Currently, I don't expect any major issues (stuff that would require rewriting code) for this upgrade, but I'll check again carefully.

One problem I know of is that Unicode changed the structure of their directories somewhat. This may affect the Ruby makefile infrastructure.

Related issues:

- Related to Ruby master - Feature #14802: Update Unicode data to Unicode Version 11.0.0 added
- Blocked by Bug #18022: Spec errors for rbconfig/unicode_[emoji_]version_spec: Using Ruby 2.7 even when on Ruby 3.1 added
- Related to Bug #18028: test/ruby/enc/test_emoji_breaks.rb does not deal with Unicode ranges in file emoji-sequences.txt added
- Related to Bug #18027: test/ruby/enc/test_emoji_breaks.rb does not use the file emoji-variation-sequences.txt added
- Related to Feature #18029: Update Unicode Emoji version to 13.1 added
- Related to Feature #18037: Update Unicode data to Unicode Version 14.0.0 added

History

#1 - 07/04/2021 02:10 AM - duerst (Martin Dürst)
- Related to Feature #14802: Update Unicode data to Unicode Version 11.0.0 added

#2 - 07/04/2021 09:52 AM - duerst (Martin Dürst)
- Blocked by Bug #18022: Spec errors for rbconfig/unicode_[emoji_]version_spec: Using Ruby 2.7 even when on Ruby 3.1 added

#3 - 07/05/2021 12:02 AM - duerst (Martin Dürst)
Details about moving files (see [https://www.unicode.org/Public/13.0.0/ucd/emoji/ReadMe.txt](https://www.unicode.org/Public/13.0.0/ucd/emoji/ReadMe.txt)):

The two files emoji-data.txt and emoji-variation-sequences.txt move from https://www.unicode.org/Public/emoji/12.1/ in Emoji version 12.1 to https://www.unicode.org/Public/13.0.0/ucd/emoji/ in Unicode Version 13.0.0.

We will reflect this in also changing the location of the files when building Ruby.

To get all the files, use

```
$> make update-unicode
```

In some cases, it will be necessary to temporarily uncomment the following line in common.mk:

```
# ALWAYS_UPDATE_UNICODE = yes
```

#4 - 07/08/2021 06:30 AM - duerst (Martin Dürst)
- Related to Bug #18028: test/ruby/enc/test_emoji_breaks.rb does not deal with Unicode ranges in file emoji-sequences.txt added

#5 - 07/08/2021 06:30 AM - duerst (Martin Dürst)
- Related to Bug #18027: test/ruby/enc/test_emoji_breaks.rb does not use the file emoji-variation-sequences.txt added

#6 - 07/08/2021 09:12 AM - duerst (Martin Dürst)
- Related to Feature #18029: Update Unicode Emoji version to 13.1 added

#7 - 07/12/2021 07:44 AM - duerst (Martin Dürst)
- Related to Feature #18037: Update Unicode data to Unicode Version 14.0.0 added
Closed (finally) with revision 21d83a823.