

Ruby - Feature #14618

Add display width method to String for CLI

03/19/2018 07:00 PM - aycabta (aycabta .)

Status:	Open
Priority:	Normal
Assignee:	
Target version:	

Description

Abstract

Unicode has display width data of characters, "Narrow" or "Wide".
For example, "A" is "Narrow", "𐀀" ("𐀀" ("\u{1f48e}")) is "Wide".
<http://unicode.org/reports/tr11/>
This data is very important for CLI tools.

Use-case

I'm developing Readline compatible library by pure Ruby implementation for Ruby core.
<https://github.com/aycabta/reline>

I'm discussing it with [@hsbt \(Hiroshi SHIBATA\)](#), and I think that the pure Ruby version should be used only when the native extension version doesn't exist.
ref. <https://bugs.ruby-lang.org/issues/11084>
The Readline library is very important for that IRB *always* provides Readline's features.
So display width method is needed by Ruby core.

Implementation approach

Uses the official data table

Unicode Consortium provides display width data as "EastAsianWidth.txt".
<http://www.unicode.org/Public/10.0.0/ucd/EastAsianWidth.txt>

This name is based on historical reasons.
This table is not exclusively for East Asian's characters in the present day, for example, Emoji.

Uses new Regexp feature (work in progress)

I propose new Unicode properties for Onigmo like Perl's one.
<https://github.com/k-takata/Onigmo/pull/102>

I think that this is a better approach if the proposal for Onigmo is merged because String#grapheme_clusters what is based on Unicode specification uses Onigmo's feature inside.

Cases of other languages or libraries

Python: unicodedata.east_asian_width (standard library)
https://docs.python.org/3.6/library/unicodedata.html#unicodedata.east_asian_width

Perl: "East_Asian_Width: *" of Unicode properties (regular expression in language)
<https://perldoc.perl.org/perluniprops.html>

Go: golang.org/x/text/width
<https://godoc.org/golang.org/x/text/width>

PHP: mb_strwidth (standard library)
<http://php.net/manual/en/function.mb-strwidth.php>

JavaScript: eastasianwidth (npm library)

<https://www.npmjs.com/package/eastasianwidth>

RubyGems: unicode-display_width gem
https://rubygems.org/gems/unicode-display_width

Related issues:

Related to Ruby - Feature #13241: Method(s) to access Unicode properties for ...

Open

Has duplicate Ruby - Feature #16006: String count and alignment that consider...

Rejected

History

#1 - 03/19/2018 09:02 PM - Anonymous

Dne 19.3.2018 v 20:00 aycabta@gmail.com napsal(a):

Use-case

I'm developing Readline compatible library by pure Ruby implementation for Ruby core.
<https://github.com/aycabta/reline>

I'm discussing it with [@hsbt \(Hiroshi SHIBATA\)](#), and I think that the pure Ruby version should be used only when the native extension version doesn't exist.

ref. <https://bugs.ruby-lang.org/issues/11084>

The Readline library is very important for that IRB *a/ways* provides Readline's features.

Just out of curiosity, why are you reimplementing readline in pure Ruby if there exists rb-readline [1], used by Ruby Installer.

Vít

[1] <https://github.com/ConnorAtherton/rb-readline>

#2 - 03/19/2018 11:55 PM - shevegen (Robert A. Heiler)

Martin Dürst is doing some unicode stuff; perhaps he can chime in when he has some time.

#3 - 03/21/2018 04:00 AM - aycabta (aycabta .)

My e-mail to the mailing list [ruby-core] is not synced with Redmine.
<http://blade.nagaokaut.ac.jp/cgi-bin/scat.rb/ruby/ruby-core/86213>

#4 - 04/09/2018 07:28 AM - duerst (Martin Dürst)

- Related to Feature #13241: Method(s) to access Unicode properties for characters/strings added

#5 - 07/16/2019 03:55 AM - shyouhei (Shyouhei Urabe)

- Has duplicate Feature #16006: String count and alignment that consider multibyte characters added