Add "graphemes" and "each_grapheme" aliases

graphemes sounds like an element in the grapheme cluster. How about each_grapheme_cluster? If everyone gets used to the grapheme as an alias of grapheme cluster, we'd love to add an alias each_grapheme.

Matz.

Languages that have added grapheme cluster support seem to be almost exclusively opting for the shorter "graphemes" alias as a part that stands for the whole.

- JavaScript/TypeScript grapheme-splitter library: splitGraphemes
- PHP: grapheme_extract
- Zig ziglyph library: Graphemeliterator
- Golang uniseg library: NewGraphemes
- Matlab: splitGraphemes
- Python grapheme library: graphemes
- Elixir: graphemes
- Crystal uni_text_seg library: graphemes
- Nim nim-graphemes library: graphemes
- Rust unicode-segmentation library: graphemes

Now that some time has passed and the "graphemes" alias for "grapheme clusters" has been fairly widely adopted by languages and libraries, I'd like to go ahead and propose a graphemes alias for grapheme_clusters and an each_grapheme alias for each_grapheme_cluster.

Related issues:
Related to Ruby master - Feature #13780: String#each_grapheme Closed

History

#1 - 02/01/2022 08:26 PM - mame (Yusuke Endoh)
- Related to Feature #13780: String#each_grapheme added

#2 - 02/01/2022 08:27 PM - mame (Yusuke Endoh)
- Description updated

(I have added to the description an url to matz's original statement)

#3 - 03/09/2022 07:07 AM - znz (Kazuhiro NISHIYAMA)
- Subject changed from Add "graphemes" and "each_grapheme aliases to Add "graphemes" and "each_grapheme" aliases

#4 - 03/17/2022 08:14 AM - nobu (Nobuyoshi Nakada)
How about letters and each_letter?

#5 - 03/17/2022 08:56 AM - matz (Yukihiro Matsumoto)
- Status changed from Open to Closed

For the record, "Grapheme" and "Grapheme cluster" are different concepts. If we call them "grapheme", it's kind of like calling "Wikipedia" as "Wiki". Until Unicode consortium defines a shorter name for them or the convention calling them "grapheme" become popular as common sense, we don't provide such aliases. So my opinion has not been changed since.
Short answer: "not yet".

Matz.

#6 - 03/17/2022 06:18 PM - Dan0042 (Daniel DeLorme)
nobu (Nobuyoshi Nakada) wrote in #note-4:

   How about letters and each_letter?

I like the general idea, but to me "letters" mean \p{L}
[retracted part]

Or how about characters and each_character?