

Ruby master - Feature #13936

Make regular expressions debugable

09/25/2017 01:27 AM - duerst (Martin Dürst)

Status:	Open
Priority:	Normal
Assignee:	
Target version:	
Description	
<p>Ruby has all kinds of features that allow a programmer to look at internals, in particular for debugging. However, one important part of Ruby, regular expressions, don't have any such features yet.</p> <p>Onigmo, the regular expression engine used in MRI, has compile-time switches to output various kinds of debugging information.</p> <p>This is a general proposal to gauge interest in the ability to debug regular expressions, e.g. by looking at the parse tree, the instruction sequence, and execution-time information. Because such information can be very large, in particular execution-time information, we have to make sure that the interface is designed carefully, but I'd like to concentrate on the general desirability/usefulness (or the absence of it) for such a feature.</p> <p>If there is positive feedback, I plan to implement the necessary features in Onigmo proper, and then add the necessary API methods (and maybe options,...) to Ruby.</p>	

History

#1 - 09/25/2017 09:06 AM - naruse (Yui NARUSE)

It needs to keep maintainability with original Onigmo/Oniguruma.

#2 - 10/03/2017 11:58 PM - shevegen (Robert A. Heiler)

I am inclined to agree with Martin. In general introspection is awesome; I remember many years ago having used Steve Dekorte's Io language; the syntax was not so nice of his language, but the introspection was really nice (I think that was back when ruby did not have `method_location` or `source_location` or whatever was the name either, so probably 1.8.x days; banister wrote that `show-method` thing in pry via the `gem method_source`, since then I can retrieve the source code to any or most ruby methods, so that is no longer a huge priority to me. But I think that in general, introspection is really great - I also love "pp"; and to quote tenderlove, "I am a puts"-debugger, that is also true, well, actually a pp-debugger :)).

I remember that I ran into encoding-related problems with regexes which were not trivial to resolve for me back then.

I am not familiar with the internals so I can't say much about it, but I am all for better introspection; and also being able to easily set which encoding a regexp is in (I have not tried again but back when I ran into these problems, it was not so trivial to resolve).

The documentation says to use any of these:

```
/pat/u - UTF-8
/pat/e - EUC-JP
/pat/s - Windows-31J
/pat/n - ASCII-8BIT
```

But I would ideally prefer to have Regexp also use methods similar to class String, in particular `.force_encoding()` which I tend to use a lot (whenever I am not using UTF-8).

If anyone wants to have a look at the docu of Regexp, here the quick link to the main documentation:

<https://ruby-doc.org/core/Regexp.html>

#3 - 10/04/2017 02:26 AM - MSP-Greg (Greg L)

Trunk version (with TOC):

https://msp-greg.github.io/ruby_trunk/Core/Regexp.html

#4 - 10/04/2017 06:53 AM - duerst (Martin Dürst)

shevegen (Robert A. Heiler) wrote:

I remember that I ran into encoding-related problems with regexes which were not trivial to resolve for me back then.

I am not familiar with the internals so I can't say much about it, but I am all for better introspection; and also being able to easily set which encoding a regexp is in (I have not tried again but back when I ran into these problems, it was not so trivial to resolve).

But I would ideally prefer to have Regexp also use methods similar to class String, in particular `.force_encoding()` which I tend to use a lot (whenever I am not using UTF-8).

I think this is a valid concern, but not directly related to the issue here. Please create a separate feature request.

#5 - 10/07/2017 01:25 AM - kernigh (George Koehler)

Do other languages, like Perl, have a feature for debugging regular expressions?