

Ruby - Bug #9221

Time.parse performance becomes exponentially worse as string length grows

12/05/2013 10:53 PM - mpelzsherman (Michael Pelz-Sherman)

Status:	Closed	
Priority:	Normal	
Assignee:	tadf (tadayoshi funaba)	
Target version:	2.1.0	
ruby -v:	ruby-2.0.0-p247	
		Backport: 1.9.3: DONE, 2.0.0: DONE
Description		
See attached script. Output:		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.010443		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.017739		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.028127		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.049885		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.071379		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.112612		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.184517		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.293784		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.476253		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 0.786087		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 1.256976		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 2.019426		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 3.300646		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 5.37757		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 8.763601		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		
elapsed: 14.165842		
parsing 12:00 PM ffffffffffffffffff		
2013-12-05 12:00:00 -0500		

elapsed: 23.888907

...

Associated revisions

Revision 334294d0e2d42ba233672a66b8c670d6f933da2f - 12/09/2013 02:14 AM - nobu (Nobuyoshi Nakada)

date_parse.c: get rid of backtrack explosion

- ext/date/date_parse.c (parse_time): unset case-insensitive flag for [:alpha:], which already implies both cases, to get rid of backtrack explosion. [ruby-core:58876] [Bug #9221]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@44086 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 334294d0 - 12/09/2013 02:14 AM - nobu (Nobuyoshi Nakada)

date_parse.c: get rid of backtrack explosion

- ext/date/date_parse.c (parse_time): unset case-insensitive flag for [:alpha:], which already implies both cases, to get rid of backtrack explosion. [ruby-core:58876] [Bug #9221]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@44086 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 9a938987cb6b3ed3f0c7735f5b19b10f45694a3f - 12/10/2013 02:58 PM - nobu (Nobuyoshi Nakada)

date_strptime.c: get rid of backtrack explosion

- ext/date/date_strptime.c (date__strptime_internal): unset case-insensitive flag for [:alpha:], which already implies both cases, to get rid of backtrack explosion. [ruby-core:58984] [Bug #9221]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@44126 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 9a938987 - 12/10/2013 02:58 PM - nobu (Nobuyoshi Nakada)

date_strptime.c: get rid of backtrack explosion

- ext/date/date_strptime.c (date__strptime_internal): unset case-insensitive flag for [:alpha:], which already implies both cases, to get rid of backtrack explosion. [ruby-core:58984] [Bug #9221]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@44126 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision ef61d714f3d1ef8652935fc95776e7e0f93fdb07 - 12/20/2013 06:48 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 44086,44126: [Backport #9221]

```
* ext/date/date_parse.c (parse_time): unset case-insensitive flag
for [:alpha:], which already implies both cases, to get rid of
backtrack explosion. [ruby-core:58876] [Bug #9221]
```

```
* ext/date/date_strptime.c (date__strptime_internal): unset
case-insensitive flag for [:alpha:], which already implies both
cases, to get rid of backtrack explosion. [ruby-core:58984]
[Bug #9221]
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_0_0@44306 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision ef61d714 - 12/20/2013 06:48 PM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 44086,44126: [Backport #9221]

```
* ext/date/date_parse.c (parse_time): unset case-insensitive flag
for [:alpha:], which already implies both cases, to get rid of
backtrack explosion. [ruby-core:58876] [Bug #9221]
```

```
* ext/date/date_strptime.c (date__strptime_internal): unset
case-insensitive flag for [:alpha:], which already implies both
cases, to get rid of backtrack explosion. [ruby-core:58984]
[Bug #9221]
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_2_0_0@44306 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 322610cb37327860396979507c54bb4964fc237f - 01/29/2014 05:45 AM - U.Nakamura

merge revision(s) 44086,44126: [Backport #9221]

```
* ext/date/date_parse.c (parse_time): unset case-insensitive flag
for [:alpha:], which already implies both cases, to get rid of
backtrack explosion. [ruby-core:58876] [Bug #9221]
```

```
* ext/date/date_strptime.c (date__strptime_internal): unset
case-insensitive flag for [:alpha:], which already implies both
cases, to get rid of backtrack explosion. [ruby-core:58984]
[Bug #9221]
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_1_9_3@44746 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 322610cb - 01/29/2014 05:45 AM - U.Nakamura

merge revision(s) 44086,44126: [Backport #9221]

```
* ext/date/date_parse.c (parse_time): unset case-insensitive flag
for [:alpha:], which already implies both cases, to get rid of
backtrack explosion. [ruby-core:58876] [Bug #9221]
```

```
* ext/date/date_strptime.c (date__strptime_internal): unset
case-insensitive flag for [:alpha:], which already implies both
cases, to get rid of backtrack explosion. [ruby-core:58984]
[Bug #9221]
```

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/branches/ruby_1_9_3@44746 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

History

#1 - 12/06/2013 04:14 PM - duerst (Martin Dürst)

This runs in no time on ruby 2.0.0p247 (2013-06-27) [i386-mingw32] (same patch level!). Can you give more details about your environment?

#2 - 12/06/2013 11:40 PM - mpelzsherman (Michael Pelz-Sherman)

duerst (Martin Dürst) wrote:

This runs in no time on ruby 2.0.0p247 (2013-06-27) [i386-mingw32] (same patch level!). Can you give more details about your environment?

Model Name: MacBook Pro
Model Identifier: MacBookPro9,1
Processor Name: Intel Core i7
Processor Speed: 2.3 GHz
Number of Processors: 1
Total Number of Cores: 4
L2 Cache (per Core): 256 KB
L3 Cache: 6 MB
Memory: 16 GB
Boot ROM Version: MBP91.00D3.B08
SMC Version (system): 2.1f173
Serial Number (system): C02JP19HF1G3

#3 - 12/06/2013 11:41 PM - mpelzsherman (Michael Pelz-Sherman)

Running ruby-2.0.0-p247 [x86_64], installed using RVM.

#4 - 12/07/2013 11:07 AM - Anonymous

Reproduced on trunk:

```
irb(main):027:0> RUBY_DESCRIPTION
=> "ruby 2.1.0dev (2013-12-07 trunk 44044) [x86_64-darwin13.0]"
irb(main):028:0> Benchmark.measure { Time.parse("12:00 PM " + "f"*30) }.to_s
=> " 4.760000  0.000000  4.760000 ( 4.759983)\n"
irb(main):029:0> Benchmark.measure { Time.parse("12:00 PM " + "f"*31) }.to_s
=> " 7.770000  0.000000  7.770000 ( 7.787989)\n"
irb(main):030:0> Benchmark.measure { Time.parse("12:00 PM " + "f"*32) }.to_s
=> "12.390000  0.000000 12.390000 (12.405786)\n"
```

#5 - 12/07/2013 04:48 PM - hsbt (Hiroshi SHIBATA)

- Target version set to 2.1.0

#6 - 12/08/2013 07:56 PM - phasis68 (Heesob Park)

It can be reproduced on ruby 2.0.0-p247 i386-mingw32.

I modified a test code to extension library free version.

```
C:\work>ruby -v
ruby 2.0.0p247 (2013-06-27) [i386-mingw32]
```

```
C:\work>type test.rb
require 'benchmark'
str = "12:00 PM ffffffffffffffffffffffff"
pat=/((?:\d+s*\s*\d+(?:\s*\s*\d+(?:[.,]\d*)?)?)\d+sh(?:\s\d+m(?:\s*\d+s?)?)?)(?:\s*ap)?\d+s*ap(?:\s*((?:gmt|utc)?(?:-+)\d+(?:[.,]\d+)(?:\s*\d+)?)?)[[:alpha:]\s]+(?:standard|daylight)\stime\b[[:alpha:]]+(?:\sdst)?\b)?/i
p Benchmark.measure { pat.match(str) }
```

```
C:\work>ruby test.rb
#<Benchmark::Tms:0x24a3440 @label="", @real=3.303211, @cstime=0.0, @cutime=0.0, @stime=0.0, @utime=3.2969999999999997, @total=3.2969999999999997>
```

Interestingly, this issue is not reproduced on irb session or without filename.

```
C:\work>ruby < test.rb
#<Benchmark::Tms:0x24cf4e0 @label="", @real=0.0, @cstime=0.0, @cutime=0.0, @stime=0.0, @utime=0.0, @total=0.0>
```

```
C:\work>irb
DL is deprecated, please use Fiddle
irb(main):001:0> require 'benchmark'
=> true
irb(main):002:0> str = "12:00 PM ffffffffffffffffffffffff"
=> "12:00 PM ffffffffffffffffffffffff"
<].\s]+(?:standard|daylight)\stime\b[[:alpha:]]+(?:\sdst)?\b)?/i
=> /((?:\d+s*\s*\d+(?:\s*\s*\d+(?:[.,]\d*)?)?)\d+sh(?:\s\d+m(?:\s*\d+s?)?)?)(?:\s*ap)?\d+s*ap(?:\s*((?:gmt|utc)?(?:-+)\d+(?:[.,]\d+)(?:\s*\d+)?)?)[[:alpha:]\s]+(?:standard|daylight)\stime\b[[:alpha:]]+(?:\sdst)?\b)?/i
irb(main):004:0> p Benchmark.measure { pat.match(str) }
#<Benchmark::Tms:0x27059a0 @label="", @real=0.0, @cstime=0.0, @cutime=0.0, @stime=0.0, @utime=0.0, @total=0.0>
=> #<Benchmark::Tms:0x27059a0 @label="", @real=0.0, @cstime=0.0, @cutime=0.0, @stime=0.0, @utime=0.0, @total=0.0>
```

#7 - 12/08/2013 09:38 PM - akr (Akira Tanaka)

- Status changed from Open to Assigned

- Assignee set to tadf (tadayoshi funaba)

I assigned this issue to tadf because the pattern is written in ext/date/date_parse.c.

#8 - 12/09/2013 10:34 AM - phasis68 (Heesob Park)

I think this is an issue of Oniguruma(regular expressions library).

Here is a shortest code of producing this issue.

```
p /(?:[[:alpha:]]+s)?/i.match("f"*40)
```

The [[:alpha:]] bracket and case ignore flag combination shows bad performance.

The simple workaround is using [a-z] instead of [[:alpha:]].

#9 - 12/09/2013 10:59 AM - nobu (Nobuyoshi Nakada)

[[:alpha:]] doesn't equal to [a-z], the former matches unicode alphabet categories.

#10 - 12/09/2013 11:14 AM - nobu (Nobuyoshi Nakada)

- Status changed from Assigned to Closed

- % Done changed from 0 to 100

This issue was solved with changeset r44086.

Michael, thank you for reporting this issue.

Your contribution to Ruby is greatly appreciated.

May Ruby be with you.

date_parse.c: get rid of backtrack explosion

- ext/date/date_parse.c (parse_time): unset case-insensitive flag for [:alpha:], which already implies both cases, to get rid of backtrack explosion. [\[ruby-core:58876\]](#) [Bug #9221]

#11 - 12/09/2013 11:43 AM - phasis68 (Heesob Park)

I know [[:alpha:]] is not equal to [a-z].

But, it seems that the current time parsing is only for ascii string.

BTW, date_strptime.c has the same regular expression on line 567-572.

It should be modified also.

#12 - 12/10/2013 12:20 AM - nagachika (Tomoyuki Chikanaga)

- Status changed from Closed to Open

- Backport changed from 1.9.3: UNKNOWN, 2.0.0: UNKNOWN to 1.9.3: REQUIRED, 2.0.0: REQUIRED

Thank you for your notice. I'll re-open this ticket.

And I confirmed ruby_1_9_3 and ruby_2_0_0 have these regular expressions.

#13 - 12/10/2013 11:42 AM - headius (Charles Nutter)

Also affects JRuby: <https://github.com/jruby/jruby/issues/1319>

#14 - 12/10/2013 11:58 PM - nobu (Nobuyoshi Nakada)

- Status changed from Open to Closed

This issue was solved with changeset r44126.

Michael, thank you for reporting this issue.

Your contribution to Ruby is greatly appreciated.

May Ruby be with you.

date_strptime.c: get rid of backtrack explosion

- ext/date/date_strptime.c (date__strptime_internal): unset case-insensitive flag for [:alpha:], which already implies both cases, to get rid of backtrack explosion. [\[ruby-core:58984\]](#) [Bug #9221]

#15 - 12/21/2013 03:50 AM - nagachika (Tomoyuki Chikanaga)

- Backport changed from 1.9.3: REQUIRED, 2.0.0: REQUIRED to 1.9.3: REQUIRED, 2.0.0: DONE

r44086 and r44126 are backported to ruby_2_0_0 branch at r44126.

#16 - 01/29/2014 05:45 AM - usa (Usaku NAKAMURA)

- Backport changed from 1.9.3: REQUIRED, 2.0.0: DONE to 1.9.3: DONE, 2.0.0: DONE

backported into ruby_1_9_3 at r44746.

Files

time_parse.rb	237 Bytes	12/05/2013	mpelzsherman (Michael Pelz-Sherman)
---------------	-----------	------------	-------------------------------------