

Ruby trunk - Bug #9159

[patch] use rb_fstring for internal strings

11/26/2013 06:30 PM - tmm1 (Aman Gupta)

Status:	Closed	
Priority:	Normal	
Assignee:		
Target version:	2.1.0	
ruby -v:	ruby 2.1.0dev (2013-11-26 trunk 43852)	Backport: 1.9.3: UNKNOWN, 2.0.0: UNKNOWN
Description		
I added rb_fstring wrappers around internal strings generated by iseqs, regexps and eval: https://github.com/tmm1/ruby/commit/9587fae171835ccf013661ba837f097754f170ef		
In our rails app, this reduces the number of long-lived strings on the heap by 30%.		
<pre>\$ ruby -rconfig/environment -e' GC.start; p ObjectSpace.count_objects[:T_STRING] ' 246100 \$ ruby -rconfig/environment -e' GC.start; p ObjectSpace.count_objects[:T_STRING] ' 173956</pre>		
If this patch is acceptable, I can commit it. make test and test-all pass.		
Related issues:		
Related to Ruby trunk - Bug #8148: [patch] reduce allocations due to __FILE__...		Rejected 03/22/2013

Associated revisions

Revision cbb56e30 - 11/26/2013 11:30 PM - tmm1 (Aman Gupta)

- compile.c: Use rb_fstring() to de-duplicate string literals in code. [ruby-core:58599] [Bug #9159] [ruby-core:54405]
- iseq.c (prepare_iseq_build): De-duplicate iseq labels and source locations.
- re.c (rb_reg_initialize): Use rb_fstring() for regex string.
- string.c (rb_fstring): Handle non-string and already-fstr arguments.
- vm_eval.c (eval_string_with_cref): De-duplicate eval source filename.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@43866 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 43866 - 11/26/2013 11:30 PM - tmm1 (Aman Gupta)

- compile.c: Use rb_fstring() to de-duplicate string literals in code. [ruby-core:58599] [Bug #9159] [ruby-core:54405]
- iseq.c (prepare_iseq_build): De-duplicate iseq labels and source locations.
- re.c (rb_reg_initialize): Use rb_fstring() for regex string.
- string.c (rb_fstring): Handle non-string and already-fstr arguments.
- vm_eval.c (eval_string_with_cref): De-duplicate eval source filename.

Revision 43866 - 11/26/2013 11:30 PM - tmm1 (Aman Gupta)

- compile.c: Use rb_fstring() to de-duplicate string literals in code. [ruby-core:58599] [Bug #9159] [ruby-core:54405]
- iseq.c (prepare_iseq_build): De-duplicate iseq labels and source locations.
- re.c (rb_reg_initialize): Use rb_fstring() for regex string.
- string.c (rb_fstring): Handle non-string and already-fstr arguments.
- vm_eval.c (eval_string_with_cref): De-duplicate eval source filename.

Revision 43866 - 11/26/2013 11:30 PM - tmm1 (Aman Gupta)

- compile.c: Use rb_fstring() to de-duplicate string literals in code. [ruby-core:58599] [Bug #9159] [ruby-core:54405]
- iseq.c (prepare_iseq_build): De-duplicate iseq labels and source locations.
- re.c (rb_reg_initialize): Use rb_fstring() for regex string.
- string.c (rb_fstring): Handle non-string and already-fstr arguments.
- vm_eval.c (eval_string_with_cref): De-duplicate eval source filename.

Revision 43866 - 11/26/2013 11:30 PM - tmm1 (Aman Gupta)

- compile.c: Use rb_fstring() to de-duplicate string literals in code. [ruby-core:58599] [Bug #9159] [ruby-core:54405]
- iseq.c (prepare_iseq_build): De-duplicate iseq labels and source locations.
- re.c (rb_reg_initialize): Use rb_fstring() for regex string.
- string.c (rb_fstring): Handle non-string and already-fstr arguments.
- vm_eval.c (eval_string_with_cref): De-duplicate eval source filename.

Revision 43866 - 11/26/2013 11:30 PM - tmm1 (Aman Gupta)

- compile.c: Use rb_fstring() to de-duplicate string literals in code. [ruby-core:58599] [Bug #9159] [ruby-core:54405]
- iseq.c (prepare_iseq_build): De-duplicate iseq labels and source locations.
- re.c (rb_reg_initialize): Use rb_fstring() for regex string.
- string.c (rb_fstring): Handle non-string and already-fstr arguments.
- vm_eval.c (eval_string_with_cref): De-duplicate eval source filename.

Revision 43866 - 11/26/2013 11:30 PM - tmm1 (Aman Gupta)

- compile.c: Use rb_fstring() to de-duplicate string literals in code. [ruby-core:58599] [Bug #9159] [ruby-core:54405]
- iseq.c (prepare_iseq_build): De-duplicate iseq labels and source locations.
- re.c (rb_reg_initialize): Use rb_fstring() for regex string.
- string.c (rb_fstring): Handle non-string and already-fstr arguments.
- vm_eval.c (eval_string_with_cref): De-duplicate eval source filename.

Revision fdcc476e - 11/27/2013 05:39 AM - tmm1 (Aman Gupta)

- test/ruby/test_eval.rb (class TestEval): Add test for shared eval filenames via rb_fstring().
- test/ruby/test_iseq.rb (class TestISeq): Add test for shared iseq labels via rb_fstring(). [Bug #9159]

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@43871 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 43871 - 11/27/2013 05:39 AM - tmm1 (Aman Gupta)

- test/ruby/test_eval.rb (class TestEval): Add test for shared eval filenames via rb_fstring().
- test/ruby/test_iseq.rb (class TestISeq): Add test for shared iseq labels via rb_fstring(). [Bug #9159]

Revision 43871 - 11/27/2013 05:39 AM - tmm1 (Aman Gupta)

- test/ruby/test_eval.rb (class TestEval): Add test for shared eval filenames via rb_fstring().
- test/ruby/test_iseq.rb (class TestISeq): Add test for shared iseq labels via rb_fstring(). [Bug #9159]

Revision 43871 - 11/27/2013 05:39 AM - tmm1 (Aman Gupta)

- test/ruby/test_eval.rb (class TestEval): Add test for shared eval filenames via rb_fstring().
- test/ruby/test_iseq.rb (class TestISeq): Add test for shared iseq labels via rb_fstring(). [Bug #9159]

Revision 43871 - 11/27/2013 05:39 AM - tmm1 (Aman Gupta)

- test/ruby/test_eval.rb (class TestEval): Add test for shared eval filenames via rb_fstring().
- test/ruby/test_iseq.rb (class TestISeq): Add test for shared iseq labels via rb_fstring(). [Bug #9159]

Revision 43871 - 11/27/2013 05:39 AM - tmm1 (Aman Gupta)

- test/ruby/test_eval.rb (class TestEval): Add test for shared eval filenames via rb_fstring().
- test/ruby/test_iseq.rb (class TestISeq): Add test for shared iseq labels via rb_fstring(). [Bug #9159]

Revision 43871 - 11/27/2013 05:39 AM - tmm1 (Aman Gupta)

- test/ruby/test_eval.rb (class TestEval): Add test for shared eval filenames via rb_fstring().
- test/ruby/test_iseq.rb (class TestISeq): Add test for shared iseq labels via rb_fstring(). [Bug #9159]

History

#1 - 11/26/2013 06:54 PM - tmm1 (Aman Gupta)

One example of strings de-duplicated by this patch are iseq labels. Before:

```
irb(main):001:0> mm = "method_missing"
irb(main):002:0> GC.start; ObjectSpace.each_object(String).select{ |o| o == mm }.size
=> 77
```

After:

```
irb(main):002:0> GC.start; ObjectSpace.each_object(String).select{ |o| o == mm }.size
=> 3
```

#2 - 11/26/2013 06:56 PM - tmm1 (Aman Gupta)

```
if (nd_type(node) == NODE_STR) {
  • hide_obj(node->nd_lit);
  • node->nd_lit = rb_fstring(node->nd_lit); ADD_INSN1(ret, nd_line(node), putobject, node->nd_lit);
}
```

Some internal strings are no longer hidden with this patch, since the underlying T_STRING object can be shared via the fstr table. Since the object is already marked immutable (via the frozen flag), hiding it is unnecessary.

#3 - 11/27/2013 08:26 AM - sam.saffron (Sam Saffron)

Seeing similar results for Discourse:

<https://gist.github.com/SamSaffron/7668043>

#4 - 11/27/2013 08:30 AM - tmm1 (Aman Gupta)

- Status changed from Open to Closed

- % Done changed from 0 to 100

This issue was solved with changeset [r43866](#).

Aman, thank you for reporting this issue.

Your contribution to Ruby is greatly appreciated.

May Ruby be with you.

-
- compile.c: Use rb_fstring() to de-duplicate string literals in code. [ruby-core:58599] [Bug [#9159](#)] [ruby-core:54405]
 - iseq.c (prepare_iseq_build): De-duplicate iseq labels and source locations.
 - re.c (rb_reg_initialize): Use rb_fstring() for regex string.
 - string.c (rb_fstring): Handle non-string and already-fstr arguments.
 - vm_eval.c (eval_string_with_cref): De-duplicate eval source filename.

#5 - 11/27/2013 08:53 AM - normalperson (Eric Wong)

Can somebody following this thread check out my patch for

<https://bugs.ruby-lang.org/issues/8998>

Theoretically it should be better than nothing (but not ideal), but I don't think any of my code uses string hash keys enough to matter...

Thanks.