Ruby trunk - Feature #6482

Add URI requested to Net::HTTP request and response objects

05/23/2012 11:48 AM - drbrain (Eric Hodel)

Status:	Closed	
Priority:	Normal	
Assignee:	naruse (Yui NARUSE)	
Target version:	2.0.0	

Description
=begin This patch adds the full URI requested to Net::HTTPRequest and Net::HTTPResponse. The goal of this is to make it easier to handle Location, Refresh, meta-headers, and URIs in retrieved documents. (While the HTTP RFC specifies the Location must be an absolute URI, not every server follows the RFC.) In order to process redirect responses from bad servers or relative URIs in requested documents the user must create an object that contains both the requested URI and the response object to create absolute URIs. This patch reduces the amount of boilerplate they are required to write. Only the (({request_uri})) is used from the URI given when creating a request. The URI is stored internally and updated with the host, port and scheme used to make the request at request time. The URI is then copied to the response object for use by the user. To preserve backwards compatibility the new behavior is optional. This allows requests with invalid URI paths like (({Net::HTTP::Get.new '/f%'})) to continue to work. Users of string paths will not be able to retrieve the requested URI. This patch is for support of #5064 =end

Related issues:	
Related to Ruby trunk - Feature #5064: HTTP user-agent class	**Assigned**

Associated revisions

Revision 570b7669 - 12/21/2012 08:36 PM - drbrain (Eric Hodel)

- lib/net/http.rb: Requests may be created with a URI which sets the Host header. Responses contain the requested URI for easier redirect following. [ruby-trunk - Feature #6482]
    - lib/net/http/generic_request.rb: ditto.
    - lib/net/http/response.rb: ditto.j
    - NEWS (net/http): Updated for above.
    - test/net/http/test_http.rb: Tests for above.
    - test/net/http/test_http.rb: ditto.
    - test/net/http/test_httpresponse.rb: ditto.

git-svn-id: svn+ssh://ci.ruby-lang.org/ruby/trunk@38546 b2dd03c8-39d4-4d8f-98ff-823fe69b080e

Revision 38546 - 12/21/2012 08:36 PM - drbrain (Eric Hodel)

- lib/net/http.rb: Requests may be created with a URI which sets the Host header. Responses contain the requested URI for easier redirect following. [ruby-trunk - Feature #6482]
    - lib/net/http/generic_request.rb: ditto.
    - lib/net/http/response.rb: ditto.j
    - NEWS (net/http): Updated for above.
    - test/net/http/test_http.rb: Tests for above.
    - test/net/http/test_http.rb: ditto.
    - test/net/http/test_httpresponse.rb: ditto.

Revision 38546 - 12/21/2012 08:36 PM - drbrain (Eric Hodel)

- lib/net/http.rb: Requests may be created with a URI which sets the Host header. Responses contain the requested URI for easier redirect following. [ruby-trunk - Feature #6482]
    - lib/net/http/generic_request.rb: ditto.
    - lib/net/http/response.rb: ditto.j
    - NEWS (net/http): Updated for above.
    - test/net/http/test_http.rb: Tests for above.
    - test/net/http/test_http.rb: ditto.
    - test/net/http/test_httpresponse.rb: ditto.

Revision 38546 - 12/21/2012 08:36 PM - drbrain (Eric Hodel)

- lib/net/http.rb: Requests may be created with a URI which sets the Host header. Responses contain the requested URI for easier redirect following. [ruby-trunk - Feature #6482]
    - lib/net/http/generic_request.rb: ditto.
    - lib/net/http/response.rb: ditto.j
    - NEWS (net/http): Updated for above.
    - test/net/http/test_http.rb: Tests for above.
    - test/net/http/test_http.rb: ditto.
    - test/net/http/test_httpresponse.rb: ditto.

History

#1 - 05/23/2012 11:48 AM - drbrain (Eric Hodel)

- File net.http.request_response_uri.patch added

Forgot patch

#2 - 05/27/2012 01:07 AM - mame (Yusuke Endoh)

Hello, drbrain

Are you willing to be a net/http(s) maintainer?
I think you deserve it.

Matz, do you accept him if he is willing?

--
Yusuke Endoh mame@tsg.ne.jp

#3 - 05/28/2012 08:53 PM - naruse (Yui NARUSE)

2012/5/27 mame (Yusuke Endoh) mame@tsg.ne.jp:

> Issue #6482 has been updated by mame (Yusuke Endoh).
>
> Hello, drbrain

Are you willing to be a net/http(s) maintainer?
I think you deserve it.

Matz, do you accept him if he is willing?

You seem forget [ruby-core:43912].

--
NARUSE, Yui naruse@airemix.jp

#4 - 05/28/2012 08:53 PM - mame (Yusuke Endoh)

Oops, sorry. Please update the maintainer list of redmine wiki.

2012/5/28 NARUSE, Yui naruse@airemix.jp:

> 2012/5/27 mame (Yusuke Endoh) mame@tsg.ne.jp:
>
>> Issue #6482 has been updated by mame (Yusuke Endoh).
>>
>> Hello, drbrain
>>
>> Are you willing to be a net/http(s) maintainer?
>> I think you deserve it.
>>
>> Matz, do you accept him if he is willing?
>
>
> You seem forget [ruby-core:43912].
>
> --
> NARUSE, Yui naruse@airemix.jp

--
Yusuke Endoh mame@tsg.ne.jp

#5 - 05/30/2012 06:23 AM - drbrain (Eric Hodel)

On May 28, 2012, at 04:37, NARUSE, Yui wrote:

> 2012/5/27 mame (Yusuke Endoh) mame@tsg.ne.jp:
>
>> Issue #6482 has been updated by mame (Yusuke Endoh).
>>
>> Hello, drbrain
>>
>> Are you willing to be a net/http(s) maintainer?
>> I think you deserve it.
>>
>> Matz, do you accept him if he is willing?
>
>
> You seem forget [ruby-core:43912].

I prefer submitting patches that NARUSE Yui reviews for me. I am glad Yui is net/http maintainer.

#6 - 05/31/2012 11:10 PM - mame (Yusuke Endoh)

- Status changed from Open to Assigned

- Assignee set to naruse (Yui NARUSE)

#7 - 06/07/2012 04:35 AM - naruse (Yui NARUSE)

I'm still considering this, but current thought is

The direction of this seems correct.
On HTTP/1.1 requires Host field in the header.

This is just needed by persistence connection.
When you connect a server and communicate two or more hosts on the server with one connection,
the Host information must be retrieved from each request,
and each response should have its own uri.

This means all request/response should have its own URI information.
So current patch's return the given URI seems not ideal.

#8 - 06/07/2012 10:02 AM - drbrain (Eric Hodel)

- File net.http.request_response_uri.2.patch added

naruse (Yui NARUSE) wrote:

> I'm still considering this, but current thought is
>
> The direction of this seems correct.
> On HTTP/1.1 requires Host field in the header.
>
> This is just needed by persistence connection.
> When you connect a server and communicate two or more hosts on the server with one connection,
> the Host information must be retrieved from each request,

I have updated the patch to obey the Host header when setting the URI, and to set the Host header from the URI when creating the request (unless overridden by initheader).

> and each response should have its own uri.
>
> This means all request/response should have its own URI information.
> So current patch's return the given URI seems not ideal.

Each response has a separate URI instance from the request due to use of dup. I've added extra assertions in test_http.rb to the revised patch to cover this.

By "all request/response should have its own URI information" do you mean "The request URI should not be edited"? This does not seem to match the current behavior of req['Host'] as it must be manually cleared in order to reuse the request with a different host.

What should this output:

require 'net/http'

uri = URI 'http://example/'
req = Net::HTTP::Get.new uri

res = Net::HTTP.start 'other.example' do |http|
http.request req
end

puts "req URI: #{req.uri}"
puts "req Host: #{req['Host']}"

With the updated patch, req.uri is http://example

With my original patch, req.uri is http://other.example

Unpatched, net/http shows "other.example" for the Host, "example" with the latest patch.

#9 - 06/09/2012 12:18 AM - naruse (Yui NARUSE)

drbrain (Eric Hodel) wrote:

> naruse (Yui NARUSE) wrote:
>
> > and each response should have its own uri.
> >
> > This means all request/response should have its own URI information.
> > So current patch's return the given URI seems not ideal.
>
> Each response has a separate URI instance from the request due to use of dup. I've added extra assertions in test_http.rb to the revised patch to cover this.
>
> By "all request/response should have its own URI information" do you mean "The request URI should not be edited"?

No for scheme and port.

> This does not seem to match the current behavior of req['Host'] as it must be manually cleared in order to reuse the request with a different host.

Try following:
require 'net/http'
req = Net::HTTP::Get.new '/'
puts "req Host: #{req['Host']}"
res = Net::HTTP.start 'redmine.ruby-lang.org' do |http|
http.request req
end
puts "req Host: #{req['Host']}"
res = Net::HTTP.start 'bugs.ruby-lang.org' do |http|
http.request req
end
puts "req Host: #{req['Host']}"

The host part of a URI for initialize seems to be the same thing of req['Host'].

#10 - 07/21/2012 08:16 AM - drbrain (Eric Hodel)

- File net.http.request_response_uri.3.patch added

=begin
naruse (Yui NARUSE) wrote:

> drbrain (Eric Hodel) wrote:
>
> > This does not seem to match the current behavior of req['Host'] as it must be manually cleared in order to reuse the request with a different
> > host.
>
>
> Try following:
> […]
>
> The host part of a URI for initialize seems to be the same thing of req['Host'].

I think I don't understand. My patch uses the host part of URI for initialize to set req['Host']. Also, if you set req['Host'] the URI is updated correctly.
Which server you connect to doesn't seem to matter.

I don't see the request Host header matching the connection host address with current net/http:

$ svnversion
36482
$ cat test.rb
require 'net/http'
req = Net::HTTP::Get.new '/'
puts "req Host: #{req['Host']}"
res = Net::HTTP.start 'redmine.ruby-lang.org' do |http|
puts "con Host: #{http.address}"
http.request req
end
puts "req Host: #{req['Host']}"
res = Net::HTTP.start 'bugs.ruby-lang.org' do |http|
puts "con Host: #{http.address}"
http.request req
end
puts "req Host: #{req['Host']}"

$ make runruby
./miniruby -I./lib -I. -I.ext/common ./tool/runruby.rb --extout=.ext -- --disable-gems ./test.rb
req Host:
con Host: redmine.ruby-lang.org
req Host: redmine.ruby-lang.org
con Host: bugs.ruby-lang.org
req Host: redmine.ruby-lang.org

My latest patch has identical behavior:

$ patch -p0 < net.http.request_response_uri.3.patch
[…]
$ make runruby
./miniruby -I./lib -I. -I.ext/common ./tool/runruby.rb --extout=.ext -- --disable-gems ./test.rb
req Host:
con Host: redmine.ruby-lang.org
req Host: redmine.ruby-lang.org
con Host: bugs.ruby-lang.org

req Host: redmine.ruby-lang.org

Identical test using URI instead of string path:

```
$ cat test.rb
require 'net/http'
u = URI("http://redmine.ruby-lang.org/")
req = Net::HTTP::Get.new u
puts "req Host: #{req['Host']}"
puts "req URI:  #{req.uri}"
res = Net::HTTP.start 'redmine.ruby-lang.org' do |http|
puts "con Host: #{http.address}"
http.request req
end
puts "req Host: #{req['Host']}"
puts "req URI:  #{req.uri}"
res = Net::HTTP.start 'bugs.ruby-lang.org' do |http|
puts "con Host: #{http.address}"
http.request req
end
puts "req Host: #{req['Host']}"
puts "req URI:  #{req.uri}"

$ make runruby
./miniruby -I./lib -I. -I.ext/common  ./tool/runruby.rb --extout=.ext  -- --disable-gems ./test.rb
req Host: redmine.ruby-lang.org
req URI:  http://redmine.ruby-lang.org/
con Host: redmine.ruby-lang.org
req Host: redmine.ruby-lang.org
req URI:  http://redmine.ruby-lang.org/
con Host: bugs.ruby-lang.org
req Host: redmine.ruby-lang.org
req URI:  http://redmine.ruby-lang.org/
```

=end

#11 - 08/24/2012 04:40 PM - naruse (Yui NARUSE)

Let me summarize (because I forgot the detail)...

An HTTP request has Host header.
It is usually used for NameVirtualHost.

Current net/http uses req['Host'] as Host header if explicitly set.
If not set, the hostname used for TCP connection is set to req['Host'] and used.

This topic is about initializing HTTPRequest with URI.
The problem now discussing is the relation between the URI and Host header (req['Host']).

5.1.2 of RFC 2616 says

> The most common form of Request-URI is that used to identify a
> resource on an origin server or gateway. In this case the absolute
> path of the URI MUST be transmitted (see section 3.2.1, abs_path) as
> the Request-URI, and the network location of the URI (authority) MUST
> be transmitted in a Host header field. For example, a client wishing
> to retrieve the resource above directly from the origin server would
> create a TCP connection to port 80 of the host "www.w3.org" and send
> the lines:
>
> ```
> GET /pub/WWW/TheProject.html HTTP/1.1
> Host: www.w3.org
> ```

Note that the "above" means http://www.w3.org/pub/WWW/TheProject.html

So a URI for initialization overwrites requesting Host header.

#12 - 11/24/2012 08:37 AM - mame (Yusuke Endoh)

- Target version changed from 2.0.0 to 2.6

#13 - 11/24/2012 08:37 AM - mame (Yusuke Endoh)

- Target version changed from 2.6 to 2.0.0

#14 - 12/20/2012 02:11 PM - drbrain (Eric Hodel)

- File net.http.request_response_uri.4.patch added

Ok, here is a patch that uses host from URI over connection host.

#15 - 12/21/2012 10:20 PM - naruse (Yui NARUSE)

drbrain (Eric Hodel) wrote:

> Ok, here is a patch that uses host from URI over connection host.

OK, commit it

#16 - 12/22/2012 05:36 AM - drbrain (Eric Hodel)

- Status changed from Assigned to Closed

- % Done changed from 0 to 100

This issue was solved with changeset r38546.
Eric, thank you for reporting this issue.
Your contribution to Ruby is greatly appreciated.
May Ruby be with you.

- lib/net/http.rb: Requests may be created with a URI which sets the Host header. Responses contain the requested URI for easier redirect
  following. [ruby-trunk - Feature #6482]
  - lib/net/http/generic_request.rb: ditto.
  - lib/net/http/response.rb: ditto.j
  - NEWS (net/http): Updated for above.
  - test/net/http/test_http.rb: Tests for above.
  - test/net/http/test_http.rb: ditto.
  - test/net/http/test_httpresponse.rb: ditto.

Files

net.http.request_response_uri.patch	12.9 KB	05/23/2012	drbrain (Eric Hodel)
net.http.request_response_uri.2.patch	14.1 KB	06/07/2012	drbrain (Eric Hodel)
net.http.request_response_uri.3.patch	9.94 KB	07/21/2012	drbrain (Eric Hodel)
net.http.request_response_uri.4.patch	10.5 KB	12/20/2012	drbrain (Eric Hodel)