

Ruby - Bug #21294

URI.extract is extracting invalid URIs with a mishmash of IPv6 notation with IPv4 address

04/29/2025 04:12 PM - Keeyan (Keeyan Nejad)

Status:Open Priority:Normal Assignee: Target version: ruby -v:ruby 3.4.3 (2025-04-14 revision d0b7e5b6a0) +PRISM [x86_64-linux]	Backport:3.2: UNKNOWN, 3.3: UNKNOWN, 3.4: UNKNOWN
Description The following is not a valid URI: http://[127.0.0.1]. So URI.extract should not extract it. It seems it is extracting it, though. So if you have code which extracts all URIs and then parses them, like the following, an error will be raised: <pre>require 'uri' URI.extract("Fake URL: http://[127.0.0.1]" , :http).each do uri # => ['http://[127.0.0.1]'] URI.parse(uri) # => raise URI::InvalidURIError end</pre> <pre>/home/keeyan/.local/share/mise/install/ruby/3.4.3/lib/ruby/3.4.0/uri/rfc3986_parser.rb:130:in 'URI::RFC3986_Parser#split': bad URI (is not URI?): "http://[127.0.0.1]" (URI::InvalidURIError) from /home/keeyan/.local/share/mise/install/ruby/3.4.3/lib/ruby/3.4.0/uri/rfc3986_parser.rb:135: in 'URI::RFC3986_Parser#parse' from /home/keeyan/.local/share/mise/install/ruby/3.4.3/lib/ruby/3.4.0/uri/common.rb:212:in 'URI. parse' from test.rb:4:in 'block in <main>' from test.rb:3:in 'Array#each' from test.rb:3:in '<main>'</pre> Instead, I believe URI.extract, should return an empty array.	
Related issues: Related to Ruby - Feature #2542: URI lib should be updated to RFC 3986 Closed01/01/2010	

History

#1 - 04/29/2025 04:46 PM - mame (Yusuke Endoh)

URI.extract is obsolete. You can confirm this by running the code in \$VERBOSE mode:

```
$ ruby -w -ruri -e 'URI.extract("Fake URL: http://[127.0.0.1]" , :http)'
-e:1: warning: URI.extract is obsolete
/home/mame/.rbenv/versions/ruby-dev/lib/ruby/3.5.0+0/uri/common.rb:268: warning: URI::RFC3986_PARSER.extract is
s obsolete. Use URI::RFC2396_PARSER.extract explicitly.
```

If you still need this functionality, you should use URI::RFC2396_PARSER.extract along with URI::RFC2396_PARSER.parse. URI::RFC2396_PARSER.parse can successfully parse http://[127.0.0.1]. However, please note that this behavior is based on an older RFC.

```
require 'uri'

URI::RFC2396_PARSER.extract("Fake URL: http://[127.0.0.1]" , :http).each do |uri| # => ['http://[127.0.0.1]']
  p URI::RFC2396_PARSER.parse(uri) # => #<URI::HTTP http://[127.0.0.1]>
end
```

#2 - 04/30/2025 07:58 AM - Keeyan (Keeyan Nejad)

Ah thank you @mame (Yusuke Endoh)! I wasn't aware it was obsolete. We can use URI::RFC2396_PARSER for our cases. Do you happen to know why extract is not being included in the newer parses? I had a look at the relevant PRs for in the URI repo, but couldn't find anything explaining the reasoning.

#3 - 05/07/2025 07:56 AM - mame (Yusuke Endoh)

- Related to Feature #2542: URI lib should be updated to RFC 3986 added