

# Re: Need regexp to rejoin URL links broken by \n

---

*Source:* <http://coding.derkeiler.com/Archive/Perl/comp.lang.perl.misc/2005-06/msg01413.html>

---

- *From:* [gbacon@xxxxxxxxxx](mailto:gbacon@xxxxxxxxxx) (Greg Bacon)
  - *Date:* Wed, 22 Jun 2005 14:16:55 -0000
- 

In article <1119431715.381879.27690@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>, Tony <hawkmoon1972@xxxxxxxxxx> wrote:

: Can someone help me with a regular expression that removes \n's from  
: the middle of URL's?  
: [...]

My first thought was to suggest stripping all runs of whitespace and feeding the result to URI::Find, but then I realized that you're trying to reformat the message for human consumption.

Below is a cut at it:

```
$ cat try
#!/usr/local/bin/perl
```

```
use warnings;
use strict;
```

```
chomp(my $A = <<EOMessage);
Hello, this is an
email which has
been formatted to
fit a narrow
column. Here is a
URL: http://test.com/hello?test=option1&test2=option2. Thanks for
reading.
EOMessage
```

```
$A =~ s!(http://.+?.) !($a=$1) =~ tr\n//d; "$a\n"!se;
```

```
print $A, "\n";
```

```
$ ./try
Hello, this is an
email which has
```

Re: Need regexp to rejoin URL links broken by \n

been formatted to  
fit a narrow  
column. Here is a  
URL: <http://test.com/hello?test=option1&test2=option2>.  
Thanks for  
reading.

Using \. / as a terminator strikes me as being \*very\* brittle, but  
that only shows the truth of mjd's words: "Of course, this is a  
heuristic, which is a fancy way of saying that it doesn't work."

Hope this helps,  
Greg

--

It should be noted that government is never so zealous in suppressing  
crime as when that crime consists of direct injury to its own sources of  
revenue, as in tax evasion and counterfeiting of its currency.

-- Murray Rothbard

.

---

• **References:**

- ◆ **[Need regexp to rejoin URL links broken by \n](#)**  
◇ From: Tony

- Prev by Date: **[Re: Email Address Validation](#)**
- Next by Date: **[Re: Need regexp to rejoin URL links broken by \n](#)**
- Previous by thread: **[Need regexp to rejoin URL links broken by \n](#)**
- Next by thread: **[Re: Need regexp to rejoin URL links broken by \n](#)**
- Index(es):
  - ◆ **[Date](#)**
  - ◆ **[Thread](#)**