

RE: Hi, how to extract five texts on each side of an URI? I post my own perl script and its use.

RE: Hi, how to extract five texts on each side of an URI? I post my own perl script and its use.

Source: <http://coding.derkeiler.com/Archive/Perl/perl.beginners/2006-11/msg00383.html>

- *From:* cclarkson@xxxxxxxxxx (Charles K. Clarkson)
 - *Date:* Sat, 11 Nov 2006 21:37:48 -0600
-

Hui Wang <<mailto:whui05@xxxxxxxxxxxxxx>> wrote:

: Can anybody tell me how to improve the running speed of this
: program? Thanks.

I don't know if this is faster, but it is a more accurate solution. Your submitted code failed under some untested circumstances. I created another page similar to the CPAN page you used and fed it more complicated tests.

Chakrabarti placed relevance on distance from the link. I changed your report to reflect this relevance. Instead of squashing all text together, it now shows a report of text token relevance. This change allowed me to test more thoroughly as well. Here is the sample report for one link with multiple texts inside the anchor.

<http://www.clarksonenergyhomes.com/scripts/index.html>

-5: 3401 MB 280 mirrors
-4: 5501 authors 10789 modules
-3: Welcome to CPAN! Here you will find All Things Perl.
-2: Browsing
-1: Perl modules
0: Perl
0: scripts
+1: Perl binary distributions ("ports")
+2: Perl source code
+3: Perl recent arrivals
+4: recent
+5: Perl modules

You can find the modified code here (for a short time):

Script: <http://www.clarksonenergyhomes.com/chakrabarti.txt>

Module: <http://www.clarksonenergyhomes.com/chakrabarti.pm>

HTH,

RE: Hi, how to extract five texts on each side of an URI? I post my own perl script and its use.

RE: Hi, how to extract five texts on each side of an URI? I post my own perl script and its use.

Charles K. Clarkson

—

Mobile Homes Specialist

Free Market Advocate

Web Programmer

254 968-8328

<http://www.clarksonenergyhomes.com/>

Don't tread on my bandwidth. Trim your posts.

.