
SIGSEGV in boost_html in file element.hpp

Posted by irok84 - 2008/05/24 13:15

Hi!

I have segmentation fault when i parsing contents of some sites (not all) by boost_html.

Program stoped in file element.hpp in method read on line 273:
ptr_element_t newElement = factory(elementName, this);

My fragment of code:

```
stringstream ost;  
ost
```

Re:SIGSEGV in boost_html in file element.hpp

Posted by softwareace - 2008/05/24 14:31

i'm actually in the process of doing some work in the boost.html library.

one of the things that will change will be the namespace (fyi) to boost.dom

Can you email me the links to the sites that are killing the library? i've tested the code on random sites, but am interested in sites that will kill the parser.

Re:SIGSEGV in boost_html in file element.hpp

Posted by irok84 - 2008/05/25 01:50

Example output from my program:

```
Reading: http://google.pl  
Reading: http://images.google.pl/imghp?hl=pl&tab=wi  
Reading: http://video.google.pl/?hl=pl&tab=vw  
Reading: http://groups.google.pl/grphp?hl=pl&tab=wg  
Reading: http://google.pl/dirhp?hl=pl&tab=wd  
Reading: http://mail.google.com/mail/?hl=pl&tab=wm  
Segmentation fault
```

other:

```
Reading: http://www.wp.pl  
Segmentation fault
```

Re:SIGSEGV in boost_html in file element.hpp

Posted by softwareace - 2008/05/25 12:53

This is a good problem!
And, I might have a solution pretty quick.

Re:SIGSEGV in boost_html in file element.hpp

Posted by irok84 - 2008/05/26 10:32

Ok. I'm waiting for solution

=====

Re:SIGSEGV in boost_html in file element.hpp

Posted by softwareace - 2008/05/26 11:03

I've been in the process of refactoring alot of the code, primarily the parser that reads the HTML/XML files. This problem is directly related to tag opening and tag closing.

The code chokes on the " in the javascript portions. I think in a few days the new code will be ready to download. If you find more examples of crashes, please post or email me the links, so I have a wider range to test against.

=====