

DOM manipulation in PHP

Asked 16 years, 4 months ago Modified 5 years, 11 months ago Viewed 9k times

 Part of [PHP](#) Collective



10



I am looking for good methods of manipulating HTML in PHP. For example, the problem I currently have is dealing with malformed HTML.

I am getting input that looks something like this:

```
<div>This is some <b>text
```

As you noticed, the HTML is missing closing tags. I could use regex or an XML Parser to solve this problem. However, it is likely that I will have to do other DOM manipulation in the future. I wonder if there are any good PHP libraries that handle DOM manipulation similar to how Javascript deals with DOM manipulation.

PHP

php

html

dom

Share

Improve this question

Follow

edited Jan 18, 2019 at 11:19



[Ijas Ameenudeen](#)

9,259 ● 3 ● 43 ● 55

asked Aug 26, 2008 at 1:47



[Eldila](#)

15.7k ● 23 ● 59 ● 62

I would run malformed HTML through HTML Purifier before doing anything else with it. – [TRiG](#)
Nov 2, 2010 at 15:15

4 Answers

Sorted by: Highest score (default)



11



PHP has [a PECL extension that gives you access to the features of HTML Tidy](#). Tidy is a pretty powerful library that should be able to take code like that and close tags in an intelligent manner.

I use it to clean up malformed XML and HTML sent to me by a classified ad system prior to import.

Share Improve this answer Follow

answered Aug 26, 2008 at 1:55



[ceejayoz](#)

180k ● 41 ● 308 ● 380



I've found PHP Simple HTML DOM to be the most useful and straight forward library yet. Better than PECL I would say.

6



I've written an article on [how to use it to scrape myspace artist tour dates](#) (just an example.) Here's a link to the [php simple html dom parser](#).



Share Improve this answer Follow



answered Feb 2, 2009 at 17:04



bedlam

357 ● 3 ● 9



The DOM library which is now built-in can solve this problem easily. The loadHTML method will accept malformed XML while the load method will not.

3



```
$d = new DOMDocument;  
$d->loadHTML('<div>This is some <b>text</b>');  
$d->saveHTML();
```



The output will be:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"  
"http://www.w3.org/TR/REC-html40/loose.dtd">  
<html>  
  <body>  
    <div>This is some <b>text</b></div>  
  </body>  
</html>
```

Share Improve this answer Follow

answered Nov 29, 2010 at 11:47



Decko

19.4k ● 2 ● 31 ● 39

I entered a similar answer with a bit more on string encoding and how to extract a document fragment [here](#) – Kevin Newman Apr 4, 2017 at 16:38



For manipulating the DOM i think that what you're looking for is [this](#). I've used to parse HTML documents from the web and it worked fine for me.

1

[Share](#) [Improve this answer](#) [Follow](#)

answered Aug 26, 2008 at 2:24



[Champo](#)

3,419 ● 30 ● 32

