

Wednesday, June 27, 2007

HTML Purifier 2.0.0 - new version of the PHP HTML filter library

There are a lot of cool people on the PHP Developer Network forums. One of them is Edward Z. Yang.

On June 20th, [Edward released HTML Purifier 2.0.0](#). [HTML Purifier](#) is a standards compliant HTML filter written in PHP. It uses a whitelisting approach and outputs standards compliant code, even if originally scrambled into an unintelligible mess. It uses functionality in the background based on Tidy's behaviour (so your preferred DTD is adhered to while filtering).


It's purpose, in case you're not familiar with HTML filter libraries, is to filter HTML user input to ensure it only includes whitelisted elements and attributes, and absolutely no XSS. The site contains a page dedicated to tests against the infamous <http://hackers.org/xss.html> exploits. There's even a demo page for testing it against your own possible exploits.

For my own part, HTML Purifier is probably the finest HTML filtering library in PHP at the moment. Its design is top notch, it's a doddle to extend, and the API is intuitive for whitelisting (see the [Advanced API](#)). The library's website has a stock of documentation for users and developers - including some useful tips for improving performance. Go get a copy and give it a whirl.

Posted by Pádraic Brady in PHP General, PHP Security at 00:28

sure would be nice if something like this made it into pear. Anonymous on Jun 27 2007, 04:00

Now all HTMLPurifier needs to do is add support for CSS purification!

We use Xinha as a rich text editor on the client side however they are allowed just a little too much control some times with everything being done as CSS styles. Then there is allowing HTML comments from users... even after HTMLPurifier still lets them add a `style="font-size: 400px"` to their html and it doesnt strip it! 


Who's volunteering to write CSSPurifier next? 

Anonymous on Jun 27 2007, 05:20

That is a bit odd - HTML Purifier seems to be stripping the style attribute in my own code. You should post a report on the forum perhaps in case your markup isn't been parsed correctly - I couldn't replicate though. Anonymous on Jun 27 2007, 06:54


Its not that it wont strip the style, its that i can't lock down the style's to things that would screw the layout only. Using Xinha it likes to use `style=` for almost everything which means its difficult to filter out things that will mess up layouts and nothing else. Either i nuke all `style=` or i hope that noone does naughty things.

No good solutions that i know of just yet. Anonymous on Jun 27 2007, 06:57

Hi Cameron, if you're going to complain, please complain somewhere where I'm bound to see it! 

The CSS Purifier you are referring to is the vaporware "HTML Purifier 3.0". In that version, HTML Purifier will abandon `explode(';', $css)` and implement a legitimate token based parser for CSS.

Locking down styles is somewhat difficult to do, because how disruptive does text have to be before it disrupts layout? Is 16px to big?

How about 32px? 64px? Is blue text disruptive? Red text? Neon green? I have thought about the problem, see (<http://htmlpurifier.org/docs/proposal-filter-levels.txt>) at the bottom, but it still needs to go a long ways. For now, moderate your application and remove egregious styling. 

If you're interested in disabling specific CSS styles, or you know exactly what sort of behavior you'd like to restrict, pop on over to the HTML Purifier forums (<http://htmlpurifier.org/phorum>) and I'll probably be able to help you. Anonymous on Jun 28 2007, 02:48

I guess with a post like that i just need go complain and beg for the feature and hope it becomes a priority



Anonymous on Jun 28

2007, 06:33

Precisely! Anonymous on Jun 28 2007, 06:39