

Re: Open and process remote page

Source: <http://coding.derkeiler.com/Archive/PHP/comp.lang.php/2005-12/msg00571.html>

- *From:* Erwin Moller <since_humans_read_this_I_am_spammed_too_much@xxxxxxxxxxxxxxxxxxxxx>
 - *Date:* Fri, 09 Dec 2005 18:21:56 +0100
-

William Hudson wrote:

- > I want to be able to ask users for a URL, open that page, change some of
- > the contents and then display that page as if they had typed the URL into
- > a browser. I have toyed with some of the php functions for opening URLs,
- > but what I am not clear on is how much work my script will have to do (do
- > I need to fully emulate a browser, for example).
- >
- > The net effect I am after is very similar to the page translation feature
- > that Google offers. Does anyone have any examples of this kind of
- > technique. Any ideas how much work is involved? (My 'translation' is
- > pretty trivial, so really it is mostly a question of how much work to
- > display the remote page).
- >
- > Thanks in advance!

Hi William,

Some thoughts:

– Opening a remote URL is very easy in PHP as you probably found out.
(Just fopen and offer an URL, PHP will in most cases wrap the whole complex request into a handle that can be treated as a (readonly) file.)

beware however of the paranoid webdesigner.
Many people have this twisted idea that they want to offer content to the world, but try to make it difficult for you to read the source.
Often Javascript is used to make things more difficult.
(Beats me why, but they come in masses.)

If you only want normal plain HTML–pages, I think you can just fopen, replace the stuff you want, and deliver that (in a frame eg., or whatever you like).

Also be aware of redirects by the server. (page moved)
I have seen a few situation where PHP doesn't handle that very well.
Or maybe it was the webserver sending something strange, I do not remember for sure, I only remember that PHP and redirects with fopen–wrapper around an URL had some issues.

Re: Open and process remote page

Beside the above possible traps, I do not expect you will find a lot of trouble. I once wrote something similar, be it more simple than what you are doing, and it was all very straightforward.

You could also get yourself in trouble (with regexp. or substringsearching, etc) when trying to replace some pieces in the HTML when the HTML is not coded as it 'should' be: Think about missing end-tags and the like. Browsers are very forgiving, but the programmers of the browsers had headaches before their program was forgiving enough. :-)

But maybe you can get away with just replacing stuff you understand, and let the remaining HTML as it was. Then the browser can display it the way it was ment. (probably).

just my 2 cents.

Good luck.

Regards,
Erwin Moller

>
> Regards,
>
> William

.

• *Follow-Ups:*

- ◆ [Re: Open and process remote page](#)
◇ From: Ian

• *References:*

- ◆ [Open and process remote page](#)
◇ From: William Hudson

- Prev by Date: [Re: Open and process remote page](#)
- Next by Date: [Re: problem with ftp put and ftp get](#)
- Previous by thread: [Re: Open and process remote page](#)
- Next by thread: [Re: Open and process remote page](#)
- Index(es):
 - ◆ [Date](#)
 - ◆ [Thread](#)