

Forum:[Hack Forums](#)

Thread: [Open a file HTML from PHP and get CONTENT](#)

MXD-Takeshi - January 18, 2019, 8:45 am

Hello everyone, I make this request perhaps a bit 'foolish but I find myself to have a static site, with over 3000 pages, all equal the only difference in small content, such as the title, image, text and other data, I wanted to know if it were possible create a script via php that would allow me to get these changes, from them generate a sql script to insert them into a database.

I have an idea of how to open files and more, but xml or ini, never happened to work on html and get data from it.

sorry for my bad english.
Thank you :)

Abruptum - January 18, 2019, 9:14 am

(Jan 18, 2019 09:45 AM)MXD-Takeshi Wrote: [\[CITE\]](#) Hello everyone, I make this request perhaps a bit 'foolish but I find myself to have a static site, with over 3000 pages, all equal the only difference in small content, such as the title, image, text and other data, I wanted to know if it were possible create a script via php that would allow me to get these changes, from them generate a sql script to insert them into a database.

I have an idea of how to open files and more, but xml or ini, never happened to work on html and get data from it.

sorry for my bad english.
Thank you :)

You could create Regular Expressions to fetch data from the loaded pages such as title and content.

Then you can push these information to the database and load the next html file.

Nachfolger - January 18, 2019, 9:15 am

You're trying to parse the HTML elements and grab the text from within them? Yes, PHP can do this with either an existing library or using regex (harder).

-- <https://stackoverflow.com/questions/18349130/how-parse-html-in-php>
-- <http://php.net/manual/en/domdocument.loadhtml.php>
-- Google

If this is a one time thing, I say skip using PHP and hand the job off to Python.

-- <https://stackoverflow.com/questions/11709079/parsing-html-using-python>
-- <https://stackoverflow.com/questions/372885/how-do-i-connect-to-a-mysql-database-in-python>
Or save it as a CSV file, which would require less work
-- <https://stackoverflow.com/a/1871557/4149581>

phlex - January 18, 2019, 10:17 am

You could use `file_get_contents()` & regex to get the values of the title/etc