How to Extract All URLs from a Web Page using PHP

codexworld.com/extract-all-urls-from-web-page-using-php/

26.07.2016

Extract URLs from the website is used in many cases, generating a sitemap from website URL is one of them. You can easily get all URLs from a web page using PHP. Here we'll provide short and simple code snippets to extract all URLs from a web page in PHP.

The following PHP code helps to get all the links from a web page URL. The `file_get_contents()` function is used to get webpage content from URL. Fetched web page content is stored in `$urlContent` variable. All the URLs or links are extracted from web page HTML content using `DOMDocument` class. All links will validate using `FILTER_VALIDATE_URL` before return and print if it is a valid URL.

```php
$urlContent = file_get_contents('http://php.net');

$dom = new DOMDocument();
@$dom->loadHTML($urlContent);
$xpath = new DOMXPath($dom);
$hrefs = $xpath->evaluate("/html/body//a");

for($i = 0; $i < $hrefs->length; $i++){
    $href = $hrefs->item($i);
    $url = $href->getAttribute('href');
    $url = filter_var($url, FILTER_SANITIZE_URL);
    // validate url
    if(!filter_var($url, FILTER_VALIDATE_URL) === false){
        echo '<a href="'.$url.'">'.$url.'</a><br />';
    }
}
```

Are you want to get implementation help, or modify or extend the functionality of this script?

Submit paid service request

Download Source CodeView Demo