

HTML Seite einlesen

Mit dieser Function kann man (externe) HTML Seiten einlesen

```
//function to get the remote data
function url_get_contents ($url) {
    if (function_exists('curl_exec')){
        $conn = curl_init($url);
        curl_setopt($conn, CURLOPT_SSL_VERIFYPEER, true);
        curl_setopt($conn, CURLOPT_FRESH_CONNECT, true);
        curl_setopt($conn, CURLOPT_RETURNTRANSFER, 1);
        $url_get_contents_data = (curl_exec($conn));
        curl_close($conn);
    }elseif(function_exists('file_get_contents')){
        $url_get_contents_data = file_get_contents($url);
    }elseif(function_exists('fopen') &&
function_exists('stream_get_contents')){
        $handle = fopen ($url, "r");
        $url_get_contents_data = stream_get_contents($handle);
    }else{
        $url_get_contents_data = false;
    }
    return $url_get_contents_data;
}
```

So sieht es dann in Aktion aus:

```
$data = url_get_contents("http://www.google.com");
if($data){
    //Do Something....
}
```

eingelassenen HTML Quelltext anzeigen lassen:

```
echo "<pre>".htmlentities($data)."</pre>";
```

Das ganze kann man auch mit DOMDocument und XPath erledigen

```
<?php
// to retrieve selected html data, try these DomXPath examples:

$file = $DOCUMENT_ROOT. "test.html";
$doc = new DOMDocument();
$doc->loadHTMLFile($file);

$xpath = new DOMXPath($doc);

// example 1: for everything with an id
```

```
// $elements = $xpath->query("//*[@id]");

// example 2: for node data in a selected id
// $elements = $xpath->query("/html/body/div[@id='yourTagIdHere']");

// example 3: same as above with wildcard
$elements = $xpath->query("*/div[@id='yourTagIdHere']");

if (!is_null($elements)) {
    foreach ($elements as $element) {
        echo "<br/>[". $element->nodeName. "];

        $nodes = $element->childNodes;
        foreach ($nodes as $node) {
            echo $node->nodeValue. "\n";
        }
    }
}
?>
```

From:
<https://wiki.hennweb.de/> - HennWeb

Permanent link:
https://wiki.hennweb.de/doku.php?id=programmieren:php:snippets:externe_html_lesen&rev=1652343938

Last update: 12/05/2022 09:25

