

Forum:[Hack Forums](#)

Thread: [Extracting data from an html table](#)

Topkek_Is_Life - February 1, 2019, 12:35 pm

I am trying to convert my html table to a php array. (Table here: <http://goldenjoe.xyz/messages/>), but I seem to be having a problem. The rows are generated using a json file and the php script reads directly from the index.html on the server, which shows the tbody is empty (because they are generated).

php:

PHP Code

```
<?php
```

```
$htmlContent = file_get_contents("http://goldenjoe.xyz/messages/");
```

```
$DOM = new DOMDocument();
```

```
$DOM->loadHTML($htmlContent);
```

```
$Header = $DOM->getElementsByTagName('th');
```

```
$Detail = $DOM->getElementsByTagName('tr');
```

```
##Get header name of the table
```

```
foreach($Header as $NodeHeader)
```

```
{
```

```
    $aDataTableHeaderHTML[] = trim($NodeHeader->textContent);
```

```
}
```

```
//print_r($aDataTableHeaderHTML); die();
```

```
##Get row data/detail table without header name as key
```

```
$i = 0;
```

```
$j = 0;
```

```
foreach($Detail as $sNodeDetail)
```

```
{
```

```
    $aDataTableDetailHTML[$j][] = trim($sNodeDetail->textContent);
```

```
    $i = $i + 1;
```

```
    $j = $i % count($aDataTableHeaderHTML) == 0 ? $j + 1 : $j;
```

```
}
```

```
print_r($aDataTableDetailHTML); die();
```

```
?>
```

index.html:

Code <html>

<head>

<script type="text/javascript" src="http://ajax.googleapis.com/ajax/libs/jquery/1.6.2/jquery.min.js">

</script>

<script>

```
$(function() {
```

```

var messages = [];

$.getJSON('messages.json', function(data) {
    $.each(data.items, function(i, f) {
        var tblRow = "<tr>" + "<td>" + f.storage.url +
            "<td>" + f.message.headers.to + "<td>" + f.message.headers.subject + "<td>"
            $(tblRow).appendTo("#userdata tbody");
    });

});

});
</script>
</head>

<body>

<div class="wrapper">
<div class="profile">
    <table id= "userdata" border="2">
    <thead>
        <th>Url</th>
        <th>Target</th>
        <th>Subject</th>
    </thead>
    <tbody>

    </tbody>
    </table>

</div>
</div>

</body>
</html>

```

When I run the php I get the error

Code Notice: Undefined variable: aDataTableDetailHTML in
/home/gueruex/public_html/messages/test2.php on line 26

Which makes sense because the variable "aDataTableDetailHTML" is declared in the foreach loop, but that loop isn't running because it does not see any <td> tags in the index.html

When you load the actual site up though, the table loads and the <td> tags are present.

Spoiler ([Click to View](#))

<https://prnt.sc/mff7bd>

Does anyone here know of a workaround I could use. I've been trying to download the complete webpage and generate a separate html file to use, but wget and curl just download the index.html directly off the server.

+Haze - February 1, 2019, 12:50 pm

You just said you own the website. If that's the case, just create an API that outputs a JSON object and read it that way instead of parsing the HTML table.

Topkek_Is_Life - February 1, 2019, 3:07 pm

Thanks for the suggestion. I decided to do that instead of the html table and everything worked out!

Abruptum - February 1, 2019, 3:33 pm

PHP Code

```
<?php
$raw = file_get_contents("http://goldenjoe.xyz/messages/messages.json");
if($raw !== false){
    $json = json_decode($raw);
    foreach($json->items as $entry){
        // do something with each $entry in the list
        var_dump($entry);
    }
}
```

Squito - February 2, 2019, 1:01 pm

(Feb 1, 2019 01:35 PM)Topkek_Is_Life Wrote: [\[CITE\]](#) Which makes sense because the variable "aDataTableDetailHTML" is declared in the foreach loop, but that loop isn't running because it does not see any <td> tags in the index.html

Because PHP will see the HTML code you posted in OP, not the one you see in the browser, and in that one there is no <td>.