

## TYPO3 Core - Bug #77642

### preg\_match: Compilation failed: regular expression is too large at offset 27

2016-08-25 16:47 - Tobias Schaefer

<b>Status:</b>	New	<b>Start date:</b>	2016-08-25
<b>Priority:</b>	Must have	<b>Due date:</b>	
<b>Assignee:</b>		<b>% Done:</b>	0%
<b>Category:</b>	Frontend	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>		<b>Complexity:</b>	
<b>TYPO3 Version:</b>	11	<b>Is Regression:</b>	No
<b>PHP Version:</b>	7.2	<b>Sprint Focus:</b>	
<b>Tags:</b>			

#### Description

If the cropHTML function (typo3/sysex/frontend/Classes/ContentObject/ContentObjectRenderer.php) is called to crop at 1074 (or more) characters it fails with this error message:

```
PHP Warning: preg_match(): Compilation failed: regular expression is too large at offset 27 in
typo3/sysex/frontend/Classes/ContentObject/ContentObjectRenderer.php
```

Increasing pcre.backtrack\_limit or pcre.recursion\_limit in php.ini doesn't help.

To reproduce the bug you can use the news extension. Create a news with a teaser text of more the 1073 characters and set up the list plugin to crop the teaser text at 1074 characters. Cropping at 1073 characters will work.

Here you find a discussion about this problem:

<http://stackoverflow.com/questions/31172837/regular-expression-is-too-large-error-in-php>

Possible solutions:

- If you are trying to match/parse HTML, I would recommend using DOMDocument to parse the HTML and then walk the DOM tree or build an XPATH to find what you're looking for.
- Shorten the Regular Expression by using DEFINE for any redundant sub-expressions (see below).
- Split your regular expression at | and process the resulting sub-expressions separately. If the regex is essentially numerous keywords separated by |, then converting to a strtok or a loop with strpos may be a better & faster choice.

TYPO3: 6.2.26

PHP: 5.5.14

Linux: SLES 12 SP1

Cheers,  
Tobias

#### History

##### #1 - 2016-10-18 10:44 - Patrick Broens

Seems to be still a problem in TYPO3 version 7.6.11. I can reproduce it in that version.

##### #2 - 2016-12-01 15:29 - DANIEL Rémy

I have this problem too: I'm am calling cropHTML with max 1500 chars.

Inside cropHTML, the pattern which try to find html entities ( #(&[^\s;]{2,8};|){0,X}#uis ) blows up because of pcre LINK\_SIZE.

This Stackoverflow post explains very well this limit: <http://stackoverflow.com/a/33988643/1053453>

But I am not a regex guru, so I don't know if this pattern could be optimised, or if another approach needs to be taken.

##### #3 - 2018-09-01 18:30 - Riccardo De Contardi

I think it is still valid in TYPO3 8.7.19

My Test with a fresh 8.7.19 TYPO3 Installation:

1) On TS Setup write:

```
page.120 = TEXT
page.120.value (
```

```
//Write here a very long text with 2000+ characters, I omit it here :)  
)  
page.120.cropHTML = 964| ...
```

2) Go to frontend and refresh

## Results:

1) In frontend, the text is not cropped

2) in TYPO3 Log module, you got 2 warnings:

```
Core: Error handler (FE): PHP Warning: preg_match(): Compilation failed: regular expression is too large at of  
fset 26 in /TYPO3-dists/typo3_src-8.7.19/typo3/sysex/typo3/sysex/frontend/Classes/ContentObject/ContentObjectRenderer.php  
line 3752
```

```
Core: Error handler (FE): PHP Warning: preg_match(): Compilation failed: regular expression is too large at of  
fset 26 in /TYPO3-dists/typo3_src-8.7.19/typo3/sysex/typo3/sysex/frontend/Classes/ContentObject/ContentObjectRenderer.php  
line 3746
```

With several trials I found that the "magic number" is 964; if you write `page.120.cropHTML = 963| ...` the crop works fine. I don't know if that depends on my environment settings.

### #4 - 2020-03-25 18:22 - Susanne Moog

- Category set to Frontend

### #5 - 2020-06-28 15:24 - Christian Toffolo

- Priority changed from *Should have* to *Must have*

- *TYPO3 Version* changed from 6.2 to 11

- *PHP Version* changed from 5.5 to 7.2

Still present in TYPO3 9 LTS and probably also in 10 LTS and current master since the code of method `cropHTML` didn't change.