Editing/Formatting

Remove 3 or more linebreaks (multiple linebreaks)?

A user wrote:

I hope your plans include a feature to remove repeated (multiple, 3 or more in sequence) end of paragraph characters, during a "text cleanup". Of course I need two paragraph marks, one to end the paragraph, and one to provide one line space between paragraphs. In the editing work that I do, that is the single most irritating problem that I encounter, where the author puts more than two paragraph marks.

Your wish is granted! (poof!)

The RegEx Find & Replace can handle this easily. Well, it WOULD be easy if I were better at RegEx. But I figured it out.

Set as follows:

```
Find:  (\x0D\x0A){3,}      [x] RegEx?
Repl:  \x0D\x0A\x0D\x0A    [x] RegEx?
                           [ ] Line By Line?
```

Explanation:
\x0D is a carriage return, \x0A is a linefeed, expressed in hexidecimal notation, understandable by the RegEx parser.
By placing in round braces (), that groups them together so that the iterator {} can treat as a whole. 3 means "at least 3 of these", where "these" are the preceding pattern, which is treated as a whole due to the round braces. The comma after the 3 separates min/max. Here min is 3, and max is unlimited. I could have written {3,999} instead, for example. The max is optional, but the comma isn't - otherwise it would do "only 3".
We're replacing with 2 carriage-return/linefeeds.

We're turning off the line-by-line option, so that the linebreaks are visible to the RegEx (otherwise it would not see them).

For this and more RegEx cleanup examples, see

http://www.thornsoft.com/sf/TML_help/Threpac_cleanupexamples.htm

Unique solution ID: #1006
Author: Chris Thornton
Last update: 2008-11-24 10:05

(c) 2023 Chris Thornton <webmaster@thornsoft.com> | 2023-12-07 17:33
URL: http://ftp.thornsoft.com/faq/index.php?action=artikel&cat=3&id=7&artlang=en