

Editing/Formatting

Remove 3 or more linebreaks (multiple linebreaks)?

A user writes:

I hope your filters 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 (good!)
The RegEx Find & Replace can handle this easily. (Well, it WOULD be easy if I were better at RegEx. But I figured it out.
Get as follows:

```
Replace: (\s{3,}) with $1  
RegEx: \s{3,}
```

Explanation
\s is a whitespace return, \s{3,} is a bracketed, expressed in hexadecimal notation, understandable by the RegEx parser.
By placing in round brackets (), that groups them together so that the iterator () can treat as a whole. 3 means "at least 3 of these", where "three" are the preceding pattern, which is treated as a whole due to the round brackets. The comma after the 3 separates minimum. 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 paragraph marks/linebreaks.
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.thorntons.com/faq/faq784.html>

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