Kaylee Stinebiser

Regex Exercise 1 -- Step by Step

Step 1:

After opening the document as a text document in oxygen and deleting the text above and below the sonnets that I don’t need for the exercise, I searched the document to make sure there weren’t any reserved characters I needed to mark for XML.

I did a Find on the &, <, and > symbols individually. None of them had results, so I didn’t need to mark them up for the XML document.

Step 2:

I removed the white spaces before each line. I typed this into the Find box: ^\_+ (instead of the \_, I used an actual space from the spacebar) and left the Replace box empty so it deleted all of the spaces I’d selected.

Step 3:

I removed the extra spaces between the sonnets.

Find box: \n\n

Replace box: \n

Step 4:

I manually entered the root element <shakeSonnets> around the entire document.

<shakeSonnets>

</shakeSonnets>

Step 5:

I wrapped each line with the line element tag.

Find box: ^.+$

Replace box: <line>\0</line>

Step 6:

I wrapped the sonnets in the sonnet element tag, but I used the closed-open approach so that the entire sonnet would be wrapped together and the Roman Numerals would become an attribute.

Find box: <line>[IXVLC]+</line>

Replace box: </sonnet><sonnet number=“\1”>

Step 7:

I manually removed the extra </sonnet> end-tag to the first sonnet in the list and added the missing <sonnet> start-tag to the very last sonnet.

Step 8:

I added a space between the sonnet end-tags and start-tags for better legibility, so that visually the sonnets are separated.

Find box: </sonnet>

Replace box: \0\n

Step 9:

For visual/legibility purposes, I added white spaces back before the lines, but this time, only the actual lines of the sonnet received white spaces. This creates a layered/tabbed visual in the XML document and I think it’s easier to read.

Find box: <line>

Replace box: \_\_\_\_\_\_<line> (6 spaces from the space bar before the tag)

Step 10:

I manually go back to the top of the document and remove the <line> tags over the document title and the author’s name. I replace these with:

<title>SONNETS</title>

<byLine>by William Shakespeare</byLine>

After that, I saved the document as an XML document and opened it. The box in the corner was green and happy with me!