<oo>→<dh> Digital humanities

Test #3: Regular Expressions

The task

For this test, we are asking you to up-convert a one-act play called *The bicyclers and three other farces* using regular expressions, which is available at http://dh.obdurodon.org/2018-02-16_regex-test.txt. Please copy and paste the text of the play into a plain text file in <oXygen/> and use the regex search and replace function to add structural tags to the play.

We encourage you to familiarize yourself with the format of the document (e.g., the division of scenes, the format of speeches and speaker names, etc.) before beginning your tagging. As was the case with the regex homework assignments, we recommend that you first search for any reserved characters and normalize the new line characters in the play, and then use regex to tag specific textual components. Our solution tagged the following:

- Each of the four scenes should be wrapped in `<scene>` tags, and the title of each scene (inside the `<scene>` element) should be wrapped in `<title>` tags.
- The setting description that precedes each individual scene should be wrapped in `<sceneDesc>` tags.
- The list of characters at the beginning of each scene should be wrapped, all together, in a single set of `<characters>` tags. (See also the bonus question, below.)
- All stage directions should be wrapped in `<stage>` tags.
- Each speech in the play, comprising both the speaker name and the spoken lines, should be wrapped in `<speech>` tags.
- Inside each `<speech>` element, the name of the speaker should be wrapped in `<speaker>` tags and the spoken text (all of it, together) should be wrapped in a single pair of `<line>` tags.

A sample speech might be tagged as follows:

```
<speech>
    <speaker>Mrs. Perkins</speaker>
    <lines><stage>foreseeing a quarrel</stage> Thaddeus! 'Sh! Ah, by-the-
    way, Mr. Bradley, where is Emma this evening? I never knew you to be
    separated before.</lines>
</speech>
```

Bonus: Within the list of character names at the beginning of each scene, wrap each item in the list in `<character>` tags, with the character's name in `<name>` tags and his or her character description in `<desc>` tags.

In our solution, we performed each of these steps with a single search and replace, but any solution that makes meaningful use of regex to tag your text is fine. **Remember to tell us explicitly when you use "dot matches all".**

You should pseudo mark-up whenever possible (for example,, the period that follows each speaker's name and the "CURTAIN" string that denotes the end of a scene). You should manually tag the play title and contents at the very beginning before you begin your up-conversion, or you should remove them until you've completed your up-conversion, tagging them at the end.

What to submit

Upload to CourseWeb only a step-by-step description of how you performed your up-conversion. We will then apply the steps to the plain text play to recreate your XML result. **Don't forget to save your tagged file as XML file after you finish your work and then reopen it in <oXygen/> to check for well-formedness!**