


<oo> → <dh> Digital humanities

Maintained by: David J. Birnbaum (djbitt@gmail.com) 

Last modified: 2015-07-29T13:11:20+0000

Regex assignment #3

The text

Oscar Wilde's *The importance of being Earnest* is available in plain text from Project Gutenberg at <http://www.gutenberg.org/cache/epub/844/pg844.txt>. Download the text and manually remove the Project Gutenberg boilerplate from the beginning and end, so that all that remains is the text as Oscar Wilde wrote it.

The task

Your task is to prepare an XML-encoded digital edition of this play from the plain text using search and replace operations to introduce the markup. The specific markup you use is up to you, but as is appropriate for a play, you will want your XML to identify at least acts, scenes, speeches, speakers, and stage directions. Note that your goal is to use search and replace operations, with or without regular expressions, to create *descriptive* well-formed XML markup (rather than, for example, to create a presentational HTML editon). You should not use manual tagging except in situations that occur so rarely that they don't justify search and replace operations or stylesheet transformations (such as tagging the title of the play or creating a root element).

When you have completed your tagging, you should upload the XML document you create along with a separate page describing any global search and replace operations you used (through the search and replace dialog box) to introduce markup.

There is no single target output for this assignment. Any well-formed markup you create that is appropriate and sensible for the play is fine.