**Computational Text Analysis**: Scholars who have pioneered and pushed computational analyses of digitized texts have made it possible for undergraduate scholars of any discipline to engage in a new kind of close reading, a computational sweep for word frequencies and other measures of a single or a library of documents.

Complete the preparatory assignment below and post your results to Canvas, before we meet as a group of 4 to talk about your plans for the final project.

**Tracking wording with Rolling Window Analysis:**

Upload → Scrub Data → Visualize using a Rolling Window

1. Go to: http://lexos.wheatoncollege.edu/upload
2. Click on Browse and upload the raw text file, Melville\_vs\_Austen.txt
3. Under the Prepare menu, select Scrub (perhaps keep hyphens? internal apostrophes?) to “clean up” or filter your data, select Apply Scrubbing [***make a note of what you chose***]
4. Under the Visualize menu, select Rolling Windows and make the following choices.
	* For this document “Melville\_vs\_Austen”
	* Perform a Rolling Average
	* Search for two word strings: the, to,
	* in a window of 10000
	* Count by words (tokens)
	* Click on Get Graph
5. The graph plots the frequency (y-axis) of each word (“the” and “to”) in “sliding” windows of 10,000 words, the left-side of the window moving one word at a time to the right (x-axis).
6. A “milestone marker” was inserted into the text. Turn on the option “Document has Milestones” and enter the milestone: foobar
7. The line that appears on the graph marks the boundary between Moby Dick and Pride and Prejudice
8. Now, have some fun testing hypotheses! Have you read either of these books or watched movies based on them?
	* Try some other words (e.g. whale, darcy or hate, love) to look for differences before and after the milestone boundary line. Pick two words that you think will be different between the two sources, and two keywords that you don’t expect to be different between the sources.
	* ***Make a note of the words you tried and screen shot the rolling windows graphs that were produced***
9. Post your results as an Announcement on Canvas

On Canvas, you can also find a more extensive set of Directions to Lexos (written by the computer scientist that designed the platform) if you are interested in exploring other functions of the program. I would recommend exploring the Top Words function.