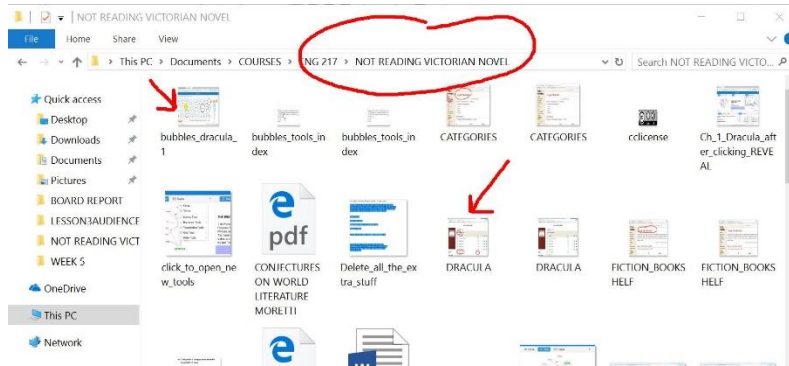


ENG 217¹ Doc McGrail Spring 2018

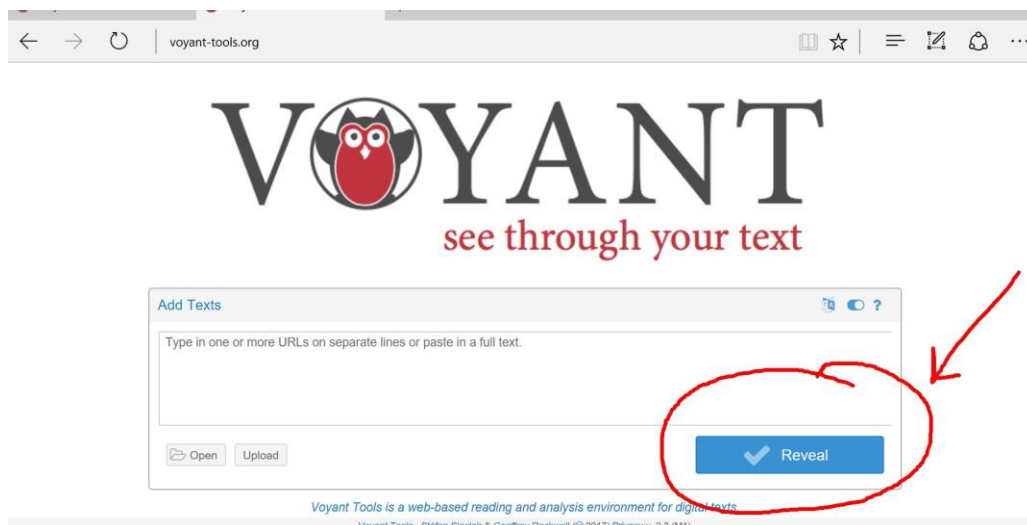
PART 2 Instructions for "Not Reading a Victorian Novel"

Distant Reading Project: **Due Dates: Sunday May 6 (Grace period Monday May 7 at midnight)**

Note: these instructions pick up where the Part 1 Instructions leave off.



1. Once you have made your *predictions*, it's time to run your big Victorian novel through some tools!
2. Go to Voyant-tools.org and you'll see the following box open up for you. Be sure you see the blue REVEAL button on the right. If you don't have java script enabled, you may not see this. You might have to try Internet Explorer or Safari instead of Google Chrome if you don't see the REVEAL button.



¹ This assignment relies on multiple other scholars' work with Voyant Tools, including [Professor Ryan Cordell](#), Francesca [Gianetti](#) and the documentation for Voyant and the [DWRL at UTexas](#). The screenshots are my own. Information and steps are adapted with permission under a Creative Commons license.

- We'll do two different operations with Voyant-tools.org. First make a large selection of text and paste it in the white box. (You could include 100 pages or Part 1 of your novel, or if your novel is short, the whole thing1)

I cut and pasted CH 1 of *Dracula* into the box and I got several different ways of visualizing the frequency of words in this chapter. On the left column, you see a Word Cloud, where the largest words are the most frequent (“driver” “horses” “know” bigger than “way” “road” “old”). You can see a trend graph on the right.

For your blog: make a screncapture of this entire screen and save it to your new folder.

The screenshot shows the Voyant Tools interface with the following components:

- Word Cloud:** A word cloud on the left side with 'driver', 'horses', and 'know' as the most prominent words.
- Text Viewer:** The central pane displays 'CHAPTER I JONATHAN HARKER'S JOURNAL (_Kept in s...)'.
- Trend Graph:** A line graph on the right showing 'Raw Frequencies' for the terms 'driver', 'horses', 'know', 'time', and 'said' across 10 document segments.
- Summary Panel:** Located at the bottom left, it provides corpus statistics: 'This corpus has 1 document with 5,776 total words and 1,473 unique word forms. Created now.', 'Vocabulary Density: 0.255', and 'Average Words Per Sentence: 25.6'.
- Contexts Panel:** Located at the bottom right, it shows a table of contexts for the selected terms.

Document	Left	Term	Right
1) CHA...	got on the coach the	dri...	had not taken his seat
1) CHA...	of the yard. Then our	dri...	, whose wide linen drawer...
1) CHA...	the haste meant, but the	dri...	was evidently bent on losing

Next look at the two charts on the bottom of your screen where you see blue boxes with “Summary” and “Contexts.” When I scroll down in the Summary dialogue box, I’ll find the **Most frequent words** in the corpus and next to each word the # of times that word showed up: [driver](#) (24); [horses](#) (19); [know](#) (16); [time](#) (15); [said](#) (14). (See below)

The screenshot shows the Summary panel with the following text:

This corpus has 1 document with 5,776 total words and 1,473 unique word forms. Created now.

Vocabulary Density: 0.255

Average Words Per Sentence: 25.6

Most frequent words in the corpus: [driver](#) (24); [horses](#) (19); [know](#) (16); [time](#) (15); [said](#) (14)

Next click on the Phrases Tab and see what the most frequent phrases are: (see below)

Summary Documents Phrases ?				
	Term	Count	Length	Trend
<input type="checkbox"/>	do you know what day it is	2	7	
<input type="checkbox"/>	i did not know what to do	2	7	
<input type="checkbox"/>	_mem i must ask the count	2	6	
<input type="checkbox"/>	i should be able to get	2	6	

256 Length Overlap

4. For your blog: make a screencapture of your results and take notes so you can save it to your new folder.

Now is a good time to **record your findings** using both screencaptures and also taking notes. I noticed, for example, that “the howling of wolves” was a phrase that showed up three times in Ch. 1 of *Dracula*. This phrase met my expectation for what I thought might happen in this novel: a creepy setting with natural elements taking on preternatural qualities.

- Now it’s time to play with some tools. Take your original file(s) and make sure you have small, medium and large sections of the novel saved so you can play around.
- You’ll notice that there are multiple tools that show up if you run your cursor over the boxes. I found “links” and “topics” by scrolling over the “?” symbol at the top of the Cirrus Box. If you can’t see it, here’s a screencast to show you here: <https://www.screencast.com/t/OzZntbS3>

The screenshot shows the Cirrus tool interface. On the left, a network diagram displays nodes for 'american', 'schools', 'right', 'jobs', 'new', 'states', 'economy', 'america', 'united', 'good', 'city', 'lefts', 'energy', 'york', and 'caterpillar'. On the right, a table lists topics and their scores:

Topic	Scores
work families it's protect folks government live g...	[Line graph]
world continue lead democracy stand freedom...	[Line graph]
just we're secure comprehensive hour reach wa...	[Line graph]
jobs years people time create oil parties busine...	[Line graph]
way deserve united example follow named desil...	[Line graph]
task trillion matter work basic improve programs...	[Line graph]
time military women law immigration president...	[Line graph]
energy gas good half modern natural climate pr...	[Line graph]

Below the table are search and run controls.

- For more tools and explanations, go to the Tools Index here to find out how some work: <http://docs.voyant-tools.org/tools/>

docs.voyant-tools.org/tools

Voyant Tools Documentation

HOME GETTING STARTED INTERFACE TOOLS INDEX USAGE & TRAINING ABOUT

TOOLS INDEX

Bubblelines visualizes the frequency and repetition of a word's use in a corpus. Each document in the corpus is represented as a horizontal line and divided into segments of equal length. Each selected word is represented as a bubble with the size of the bubble indicating the word's frequency in the corresponding segment of text. The larger the bubble's radius the more frequently the word occurs. [[more documentation](#)] [[use it](#)]

Bubbles represents the relative frequency of words in a corpus through a cloud of bubbles. Each word is represented by a bubble, where the size of the bubble is proportionate to the frequency with which the word occurs in the corpus. The larger the bubble's radius the more frequently the word occurs. [[more documentation](#)] [[use it](#)]

Cirrus is a word cloud displaying the frequency of words appearing in a corpus. Words occurring more frequently appear larger. [[more documentation](#)] [[use it](#)]

Corpus Grid provides an overview of a corpus, displaying each document's title, total number of words (word tokens), number of unique words (word types), and lexical density (the ratio of tokens to types). [[more documentation](#)] [[use it](#)]

You'll notice that there are multiple tools to try. You can first read about the tool and then click the words "use it" and put your text in the box and hit "REVEAL". I'm going to try the Bubbles tool by clicking on the words "use it" and then pasting the section of *Dracula* I want to see in the white box that pops up:



Bubbles represents the relative frequency of words in a corpus through a cloud of bubbles. Each word is represented by a bubble, where the size of the bubble is proportionate to the frequency with which the word occurs in the corpus. The larger the bubble's radius the more frequently the word occurs. [[more documentation](#)] [[use it](#)]

Here's the bubble

The screenshot shows the Voyant Tools interface. At the top, there's a browser window with the URL 'voyant-tools.org/?corpus=bb335dc6c887ec59e1f6751d7e3f8b56&view=Bubbles'. Below the browser, the Voyant Tools logo is visible. The main area displays a word cloud for 'The Project Gutenberg EBook of Dracula, by Bram...'. The word 'count' is highlighted in yellow, and 'know' is highlighted in black. Other words in the cloud include 'door', 'asked', 'white', 'round', 'light', 'hand', 'left', 'diary', 'heard', 'old', 'room', 'said', 'friend', 'night', 'man', 'began', 'went', 'like', 'great', 'come', 'got', 'eyes', 'just', 'ran', 'sort', 'looking', 'chapter', 'opened', 'way', 'came', 'face', 'till', 'time', 'long', 'looked', 'good', 'mina', 'place', 'things', 'little', 'fear', 'saw', 'strange', 'Jonatha'. To the right of the word cloud, there is a vertical list of words: 'know', 'chapter', 'said', 'time', 'great', 'long', 'count', 'door', 'went', 'white', 'like', 'old', 'night', 'round', 'eye', 'way', 'strange', 'come', 'heard', 'room'. At the bottom left, there is a progress bar showing '812 / 7299' and a 'speed' control. At the bottom right, it says 'Adapted from Letter Pairs by Martin Ignacio Bereciartua'.

8. Try out some tools and see what you find. If you are ambitious you could compare chapters or compare different novels. For each operation that you undertake, be sure you take a screenshot of your findings and try to speculate on what your findings mean.
9. **Expect dead ends and frustrating moments!** You can't expect this to work like a Google search. What you are looking for isn't already known—you are shining a flashlight in a dark forest and so who knows what you'll find? The goal of this exercise is to respond to Franco Moretti's call for *"maximum methodological boldness."* That means that what you create with Voyant tools—what you uncover and discover—is subject to all the vagaries of the unknown. The learning outcome for this exercise is *not the answer but the methodological journey*.
10. **For your blog: post your findings, screenshots and citations and tell us about your methodological boldness and discoveries!**

Assignment Requirements Rubric for Distant Reading Project

1. **Choose and cite.**
You choose a Victorian novel from the Fiction shelf in the Project Gutenberg archive and you save it as a .txt file. **You do not choose *Dracula*. (You can choose *Dracula* as a second text if you want to compare texts, but it can't be your primary text.)** You include all the citation information on your blog. (This is found on p. 1 of the book but you'll be deleting this archive info for your text analysis.)
2. **Make predictions.**
You write up your predictions and hypotheses for what you *think* you'll find.
3. **Provide screenshots of the overall view.**
You provide screen captures on your blog of the overall view (see #11 screenshot in these instructions above).
4. **You keep a record of your work and post it on your blog.**
You record the most frequent words and most frequent phrases in your blog and include screenshots of these.
5. **Try out at least 3 tools.**
You try out at least 3 other tools using the Voyant tools index <http://docs.voyant-tools.org/tools/> and you include your screenshots and your reflections on what you find.
6. **Connect and compare**

You connect what you found to Moretti's call for methods of "distant reading." You compare this exercise to your previous "close reading" of novels. Write at least one substantive paragraph about this experience of "not" reading or of "distant reading."

Grading Rubric:

You do all of the above on time (by May 8 at midnight) and post the link to your blog in Moodle= **Grade of B**

You do all of the above on time and you provide exceptionally insightful, thoughtful, thorough, witty or "methodologically bold" commentary/discoveries= **Grade of A**

You do some of the above and post the link to your blog in Moodle= **Grade of C**

You do some of the above but in a haphazard and below satisfactory manner and post the link to your blog in Moodle= **Grade of D**

You do very little of the above in a haphazard and below satisfactory manner but you still post the link to your blog in Moodle= **Grade of F** (remember an F is 50% which is better than a 0)