Text Analysis 101: Voyant Tools

Introduction:

This guide provides an introduction to Voyant Tools, an open-source, web-based application for performing text and data mining. Developed by Stéfan Sinclair at McGill University and Geoffrey Rockwell at the University of Alberta, Voyant Tools was created to support scholarly reading and interpretation of texts.

Created with digital humanities scholars in mind, Voyant Tools provides lightweight text analytics such as word frequency lists, frequency distribution plots, and KWIC (Key Word in Context) analysis. To learn more about Voyant Tools, please visit the Voyant Tools repository on GitHub or the Voyant Tools Help Guide.

Getting Started:

Voyant Tools can be accessed online at https://voyant-tools.org. You can also install the Voyant Server as a stand-alone version on your computer. This has several potential advantages, including optimal performance, reliability, security, and privacy.

To download the Voyant Server to your computer, you will need to have Java installed on your computer first. Once you have Java installed and set up on your computer, you will need to go to the latest releases page of Voyant Tools and click on the VoyantServer2_4-M28.zip file to download the files needed to set up the server. This is a large zip file of about 200MB – it includes large data models for language processing. The .zip archive file will need to be decompressed before you can install it.

For more detailed instructions on how to install the Voyant Server, see the Voyant Server Github Repository. If you are unsure about how to unzip the VoyantServer2_4-M28.zip file, see Microsoft's instructions for PC users and Apple's instructions for Mac users on how to decompress zip files.

Acceptable File Formats:

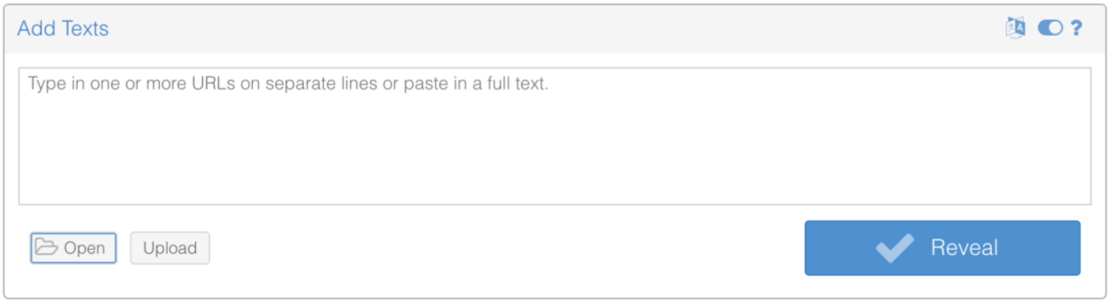
Voyant Tools is a web-based text reading and analysis environment. It allows you to upload a variety of text formats to analyze, including TXT, HTML, XML, PDF, RTF, and MS Word documents. You can create your own collection of texts, or you can use one of the sample corpuses available in Voyant Tools.

Uploading Texts:

There are four ways of uploading texts to Voyant Tools:

VOYANT

see through your text

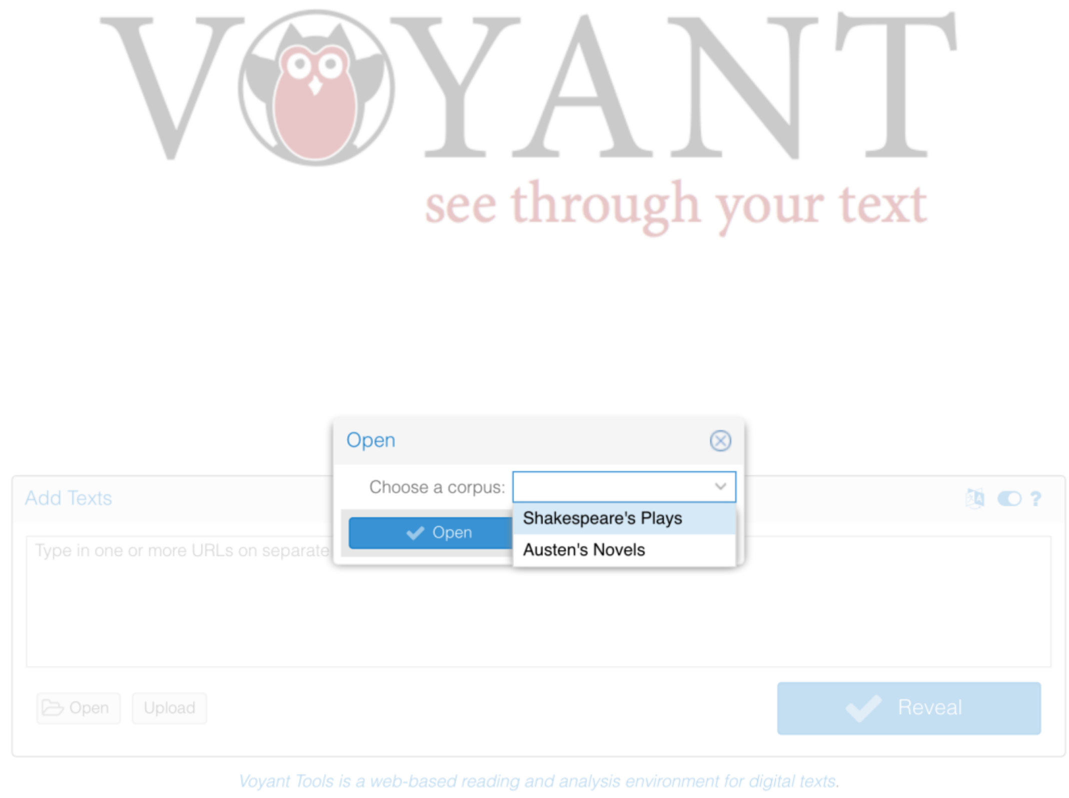


Add Texts

Type in one or more URLs on separate lines or paste in a full text.

Voyant Tools is a web-based reading and analysis environment for digital texts.

1. Open an existing corpus by Voyant Tools. Click on the “Open” button under the text box and select one of the sample corpuses provided by Voyant Tools.



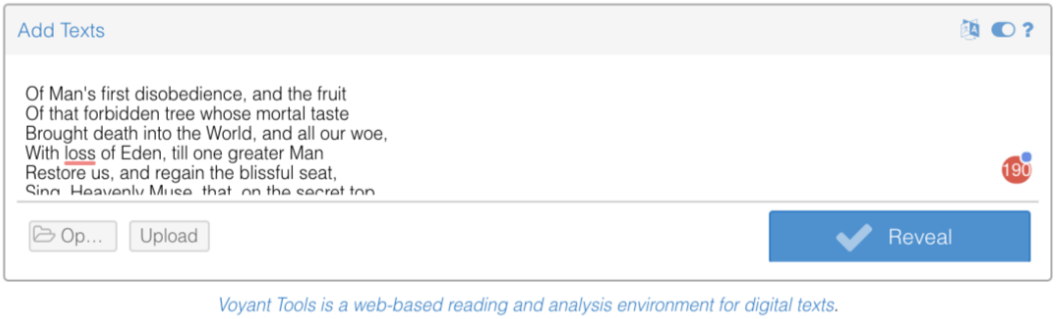
Voyant Tools is a web-based reading and analysis environment for digital texts.

You will see that Voyant Tools provides two sample corpuses: "Shakespeare’s Plays" and "Jane Austen’s Novels." These sample corpuses were taken from Project Gutenberg’s collections. The corpus, "Shakespeare's plays" includes all of William Shakespeare’s 37 plays. The corpus "Austen's Novels" includes the following texts by Jane Austen:

- *Emma*
- *Love And Friendship*
- *Lady Susan*
- *Mansfield Park*
- *Northanger Abbey*
- *Sense and Sensibility*
- *Persuasion*

- *Pride and Prejudice*

2. Type or paste text into the main text box (this creates a corpus with one document)

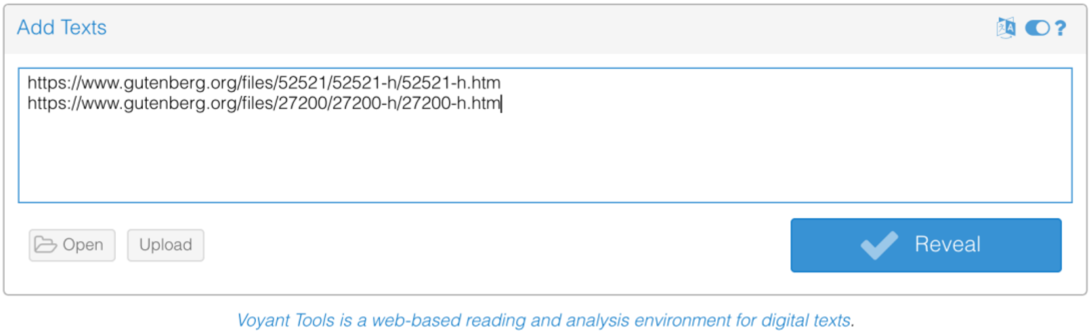


Of Man's first disobedience, and the fruit
Of that forbidden tree whose mortal taste
Brought death into the World, and all our woe,
With loss of Eden, till one greater Man
Restore us, and regain the blissful seat,
Sing, Heavenly Muse, that, on the secret top

Voyant Tools is a web-based reading and analysis environment for digital texts.

3.Type or paste one or more URLs into the main text box (one URL per line)

VOYANT
see through your text



```
https://www.gutenberg.org/files/52521/52521-h/52521-h.htm
https://www.gutenberg.org/files/27200/27200-h/27200-h.htm
```

Voyant Tools is a web-based reading and analysis environment for digital texts.

4. Click the “Upload” button under the text box to load files from your computer. You will need to select all of the files at one time (Press the Shift key on your computer and click on each item) or zip your files into a compressed folder and upload the zip file to Voyant.

Tutorial:

Text Analysis on Paradise Lost Corpus

Now that we have gone over how to access Voyant Tools and upload texts, let's examine our own corpus. In this tutorial, we will use Voyant Tools to look at word frequencies, collocates, trends, and keywords in context across a sample corpus of Paradise Lost.

This sample corpus was collected from Project Gutenberg and includes the

Follow Steps 1-10 of this tutorial to examine Paradise Lost in Voyant Tools.

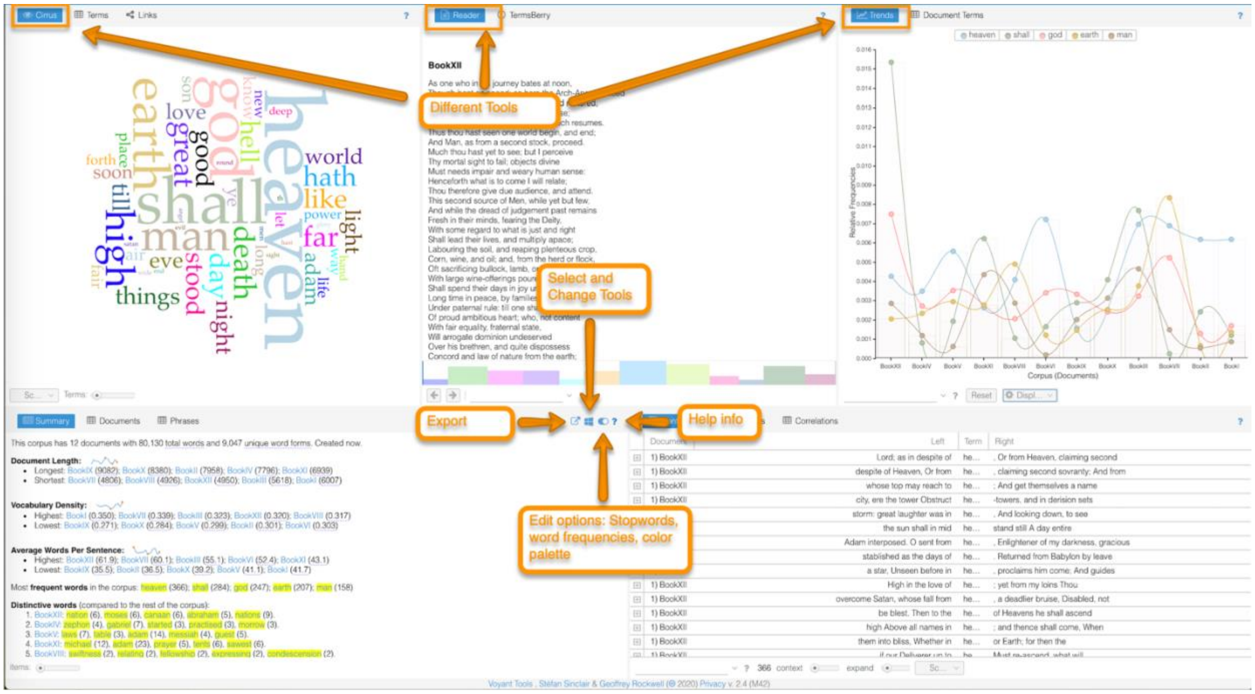
Step 1: Uploading Your Corpus

First, download the ParadiseLost.zip to your computer. This zip file contains our sample corpus of Paradise Lost’s 12 Books in plain text files (.txt).

In Voyant Tools, click upload and select the ParadiseLost.zip file that you just saved to your computer and click "Reveal." The contents of the zip file should load automatically within the Voyant Tools interface.

+Step 2: Voyant Tools Interface

The Voyant Tools interface displays five panels with different text analysis tools.



BookXII

As one who in journey bates at noon,
ch resumes.
Thus thou hast seen one world begin, and end;
And Man, as from a second stock, proceed.
Much thou hast yet to see; but I perceive
Thy mortal sight to fail; objects divine
Must needs impair and weary human sense:
Henceforth what is to come I will relate;
Thou therefore give due audience, and attend.
This second source of Men, while yet but few,
And while the dread of judgement past remains
Fresh in their minds, fearing the Deity,
With some regard to what is just and right
Shall lead their lives, and multiply apace;
Labouring the soil, and reaping plenteous crop,
Corn, wine, and oil; and, from the herd or flock,
Oft sacrificing bullock, lamb, o
With large wine-offerings poure
Shall spend their days in joy u
Long time in peace, by familie
Under paternal rule: till one sh
Of proud ambitious heart; who, not content
With fair equality, fraternal state,
Will arrogate dominion undeserved
Over his brethren, and quite dispossess
Concord and law of nature from the earth;

This corpus has 12 documents with 80,130 total words and 9,047 unique word forms. Created now.

Document Length:
- Longest: BookIX (9082); BookX (8380); BookII (7958); BookIV (7796); BookXI (6939)
- Shortest: BookVII (4806); BookVIII (4926); BookXII (4950); BookIII (5618); BookI (6007)

Vocabulary Density:
- Highest: BookI (0.350); BookVII (0.339); BookIII (0.323); BookXII (0.320); BookVIII (0.317)
- Lowest: BookIX (0.271); BookX (0.284); BookV (0.299); BookII (0.301); BookVI (0.303)

Average Words Per Sentence:
- Highest: BookXII (61.9); BookVII (60.1); BookIII (55.1); BookVI (52.4); BookXI (43.1)
- Lowest: BookIX (35.5); BookII (36.5); BookX (39.2); BookV (41.1); BookI (41.7)

Most **frequent words** in the corpus: heaven (366); shall (284); god (247); earth (207); man (158)

Distinctive words (compared to the rest of the corpus):
1. BookXII: nation (6), moses (6), canaan (6), abraham (5), nations (9).
2. BookIV: zephon (4), gabriel (7), started (3), practised (3), morrow (3).
3. BookV: laws (7), table (3), adam (14), messiah (4), guest (5).
4. BookXI: michael (12), adam (23), prayer (5), tents (6), sawest (6).
5. BookVIII: swiftness (2), relating (2), fellowship (2), expressing (2), condescension (2).

Document	Left	Term	Right
1) BookXII	Lord; as in despite of	he...	, Or from Heaven, claiming second
1) BookXII	despite of Heaven, Or from	he...	, claiming second sovranty; And from
1) BookXII	whose top may reach to	he...	; And get themselves a name
1) BookXII	city, ere the tower Obstruct	he...	-towers, and in derision sets
	storm: great laughter was in	he...	, And looking down, to see
	the sun shall in mid	he...	stand still A day entire
	Adam interposed. O sent from	he...	, Enlightener of my darkness, gracious
	stablished as the days of	he...	, Returned from Babylon by leave
	a star, Unseen before in	he...	, proclaims him come; And guides
1) BookXII	High in the love of	he...	; yet from my loins Thou
1) BookXII	overcome Satan, whose fall from	he...	, a deadlier bruise, Disabled, not
1) BookXII	be blest. Then to the	he...	of Heavens he shall ascend
1) BookXII	high Above all names in	he...	; and thence shall come, When
1) BookXII	them into bliss, Whether in	he...	or Earth; for then the
1) BookXII	if our Deliverer up to	he...	Must re-ascend, what will

366 context expand

Voyant Tools , Stéfan Sinclair & Geoffrey Rockwell (© 2020) Privacy v. 2.4 (M42)

By default, the following text analysis tools are available to you:

- **Cirrus:** The Cirrus tool is a word cloud that visualizes the top frequency words of a corpus or document. The word cloud positions the words such that the terms that occur the most frequently are positioned centrally and are sized the largest. As the algorithm goes through the list and continues to attempt to draw words as close as possible to the center of the visualization, it will also include small words within spaces left by larger words that do not fit together snugly. It's important to understand that the color of words and their absolute position are not significant (if you resize the window or reload the page, words may appear in a different location).

- **Reader:** The Reader Tool is where the text is displayed for reading. You can scroll down within the text reader to fetch more content. You can also hover over a word to show its frequency in the document. Additionally, you can click on a word or search for it in the search box to see how often it appears in your corpus.

- **Trends:** The Trends tool (also known as the Terms Frequency Chart) provides distribution plots that represent the frequencies of terms across texts in your corpus. Each series in the graph is colored according to the word it represents. At the top of the graph a legend displays which words are associated with certain colors. You can click on the words in the legend to toggle their visibility. Hovering over any point in the graph causes a callout box to appear with information about the point, including the word and the frequency.

- **Summary:** The Summary tool displays the number of documents in the corpus and the total number of words and unique words (multiple occurrences of words) in the corpus. The next part of the summary displays the document length of the corpus. It shows the longest and shortest documents by the number of words in the corpus. In parentheses after each title in the corpus, you will see the number of words. The next section provides the documents with the top vocabulary densities (the ratio of the number of words in the document to the number of unique words in the document) and the documents with the lowest vocabulary densities. Following this section is an approximation of the average number of words per sentence, both the highest and lowest values. Next, there are the five most frequent words in the corpus which are indicated to the right of the corpus.

- **Contexts:** The Contexts (or Keywords in Context) tool shows each occurrence of a keyword with its surrounding text (the context). The table view shows the following three columns:

  - Document: this shows which keyword and contexts occur
  - Left: contextual words to the left of the keyword (note that sorting by this column treats words in reverse order, right to left from the keyword)
  - Term: the keyword matching the default or user-provided term query
  - Right: contextual words to the right of the keyword

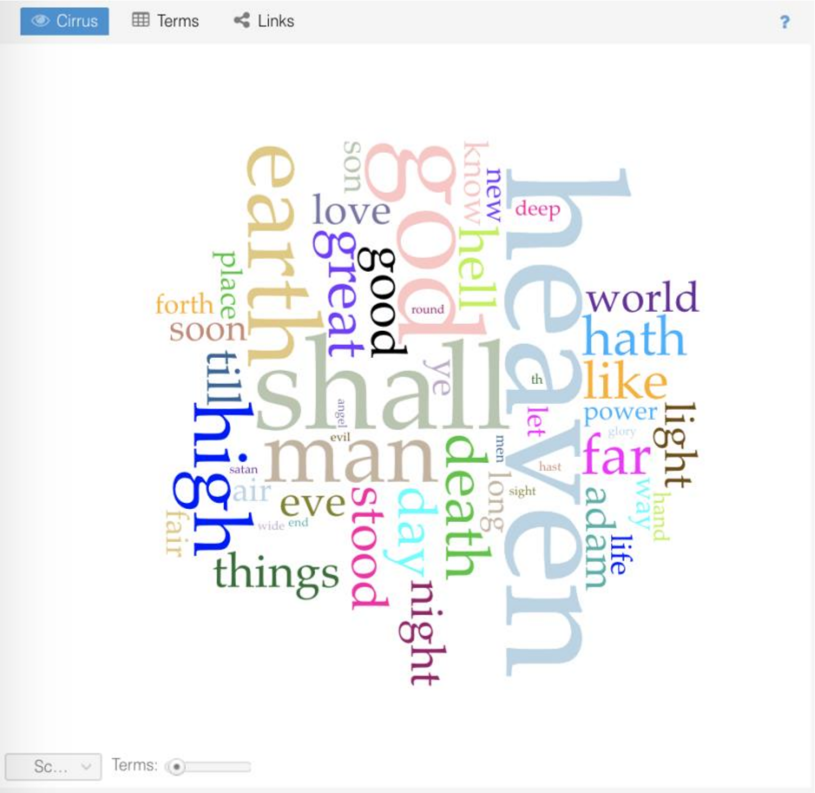
To select an alternative tool, you will need to hover over the gray bar at the top of the trends or cirrus window by the question mark symbol: ? until a menu of icons appears: ↗ ⊞ ◐ ?

If you select the window button: ⊞ , you will see a variety of other tools that can perform different visualizations and text analysis. There are many tools to choose from. To learn more about these tools and their functionalities, visit the Voyant Tools help guide.

Step 3: Word Clouds and Stopwords

Cirrus Tool: Word Cloud

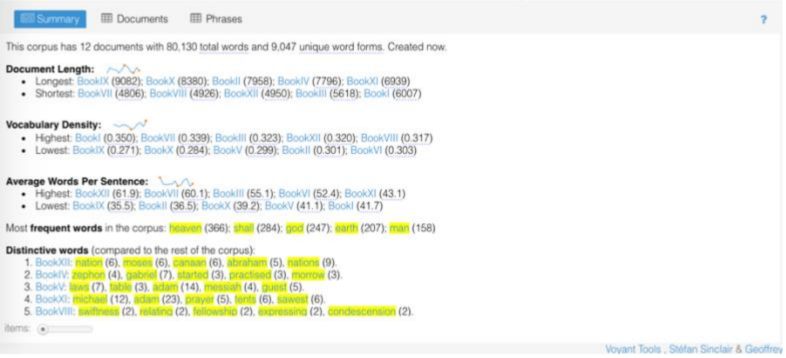
To begin, let's take a look at the Cirrus Tool. In this word cloud, you will see the most frequent words in the Paradise Lost corpus. The most common words are the largest words in the cloud.



The Cirrus tool has a slider near the bottom (with the label "Terms") that allows you to adjust the number of words displayed. By default, the minimum value is 25 and the maximum value is 500, and the slider adjusts by increments of 25.

Summary Tool

You can also see the most frequent words listed in the Summary tool with the number of times the word is mentioned in the entire Paradise Lost corpus.



This corpus has 12 documents with 80,130 total words and 9,047 unique word forms. Created now.

Document Length:
- Longest: BookIX (9082); BookX (8380); BookII (7958); BookIV (7796); BookXI (6939)
- Shortest: BookVII (4806); BookVIII (4926); BookXII (4950); BookIII (5618); BookI (6007)

Vocabulary Density:
- Highest: BookI (0.350); BookVII (0.339); BookIII (0.323); BookXII (0.320); BookVIII (0.317)
- Lowest: BookIX (0.271); BookX (0.284); BookV (0.299); BookII (0.301); BookVI (0.303)

Average Words Per Sentence:
- Highest: BookXII (61.9); BookVII (60.1); BookIII (55.1); BookVI (52.4); BookXI (43.1)
- Lowest: BookIX (35.5); BookII (36.5); BookX (39.2); BookV (41.1); BookI (41.7)

Most **frequent words** in the corpus: heaven (366); shall (284); god (247); earth (207); man (158)

Distinctive words (compared to the rest of the corpus):
1. BookXII: nation (6), moses (6), canaan (6), abraham (5), nations (9).
2. BookIV: zephon (4), gabriel (7), started (3), practised (3), morrow (3).
3. BookV: laws (7), table (3), adam (14), messiah (4), guest (5).
4. BookXI: michael (12), adam (23), prayer (5), tents (6), sawest (6).
5. BookVIII: swiftness (2), relating (2), fellowship (2), expressing (2), condescension (2).

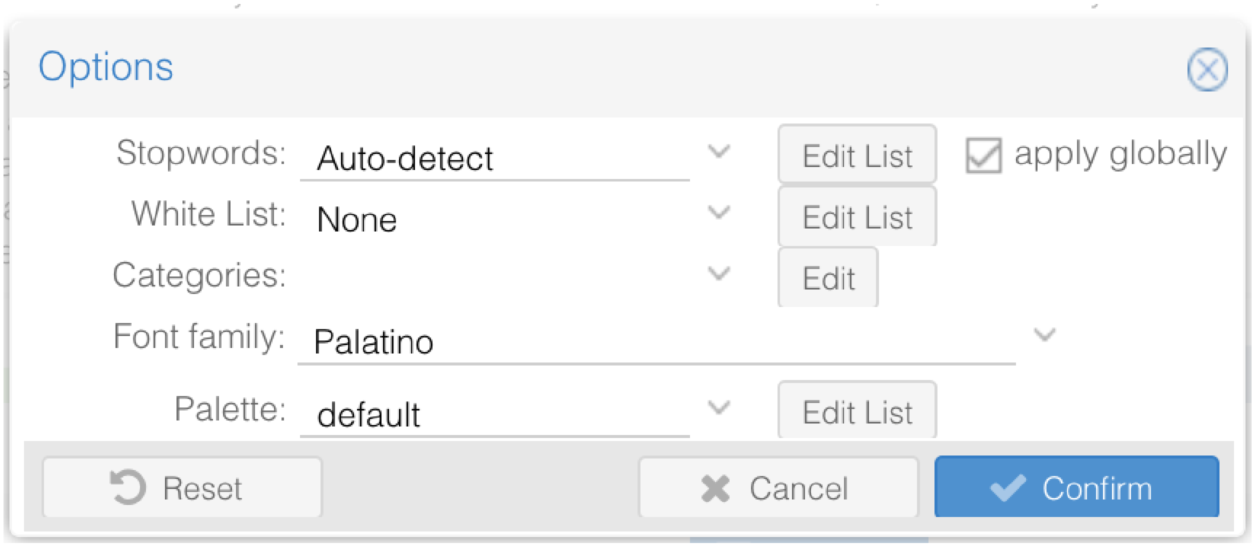
Stopwords

Voyant Tools has a list of common stopwords that have been removed from your corpus, such as the words: "an," "and," "or," "but," etc.

If you would like to remove other stopwords or words that don't add much meaning to your analysis, you can filter them out. For example, let's filter out the words, "said,"went," and "came" from your word cloud.

Hover over the gray bar at the top of the word cloud window until a menu of icons appears. Click on the blue options icon: .

In the Options pop-up window, you can also modify the tool's settings. Here is a list of options available to you:



Options

Stopwords: Auto-detect

White List: None

Categories:

Font family: Palatino

Palette: default

- Stopwords: you can define a set of stopwords to exclude – see the stopwords guide for more information

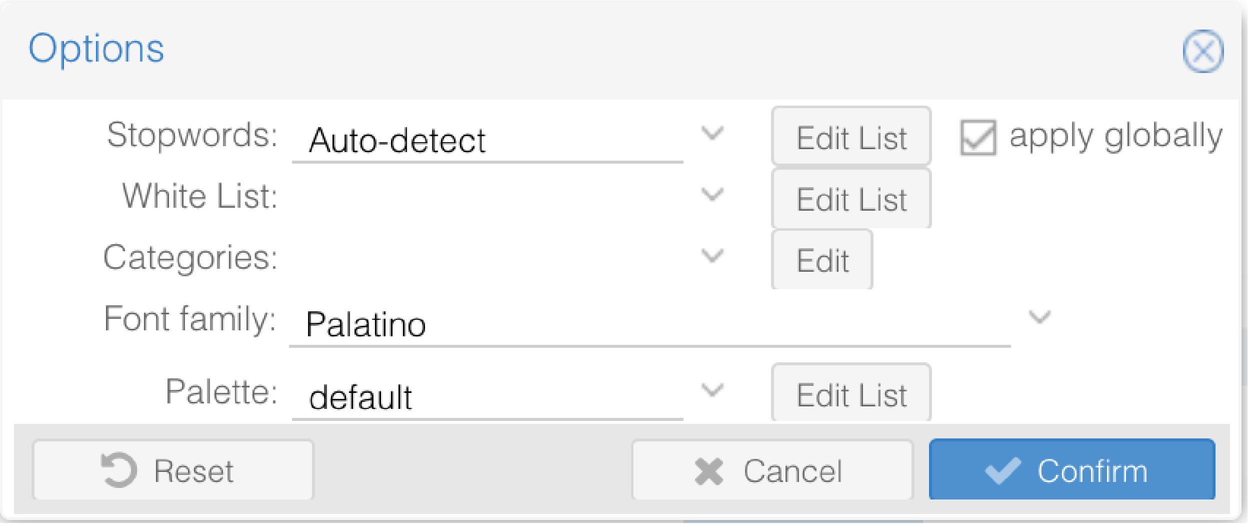
- White List: you can define a set of allowed words (the opposite of a stopwords list), only terms in this list will be shown in Cirrus (note that the stopwords list is still active, so you may want to choose "None" from the stopwords menu to deactivate it)

- Categories: you can specify categories based on the frequencies of words.

- Font Family: you can determine which font is used by Cirrus, a set of web-safe fonts is provided. Here you can also specify a font installed on your computer, but of course, it may not be available on other computers (in which case a default font is used)

- Palette: you can edit the colour palette

To filter out stopwords, click on the "Edit List" button to the right of the "Stopwords: Auto-detect" dropdown menu.



Options

Stopwords: Auto-detect

White List:

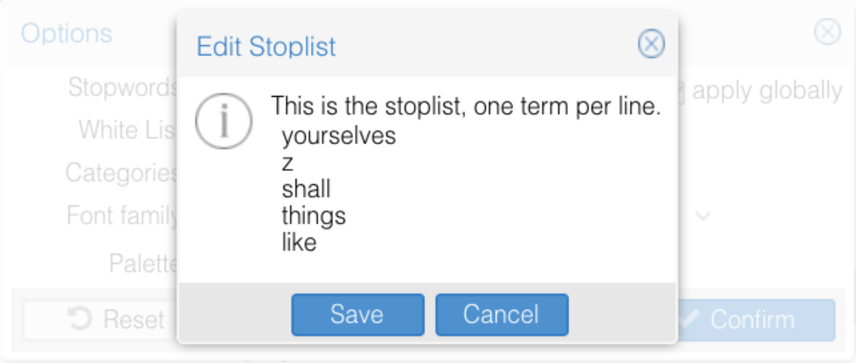
Categories:

Font family: Palatino

Palette: default

☑ apply globally

In the Edit Stoplist window that pops up, type in “shall,” “things,” “like” and any other words that you would like to filter out, then click Save.

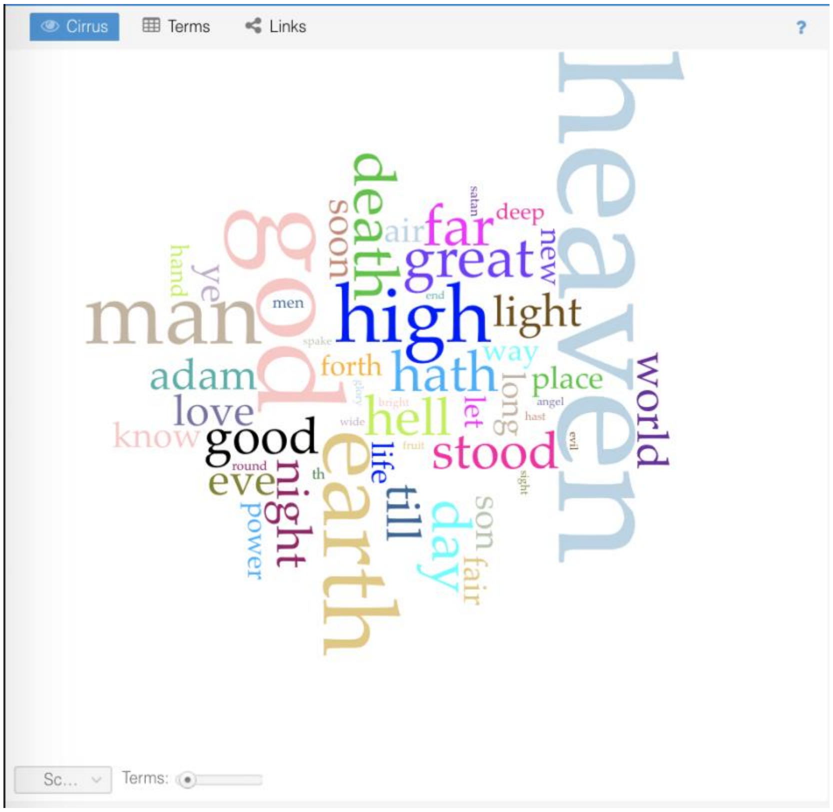


Edit Stoplist

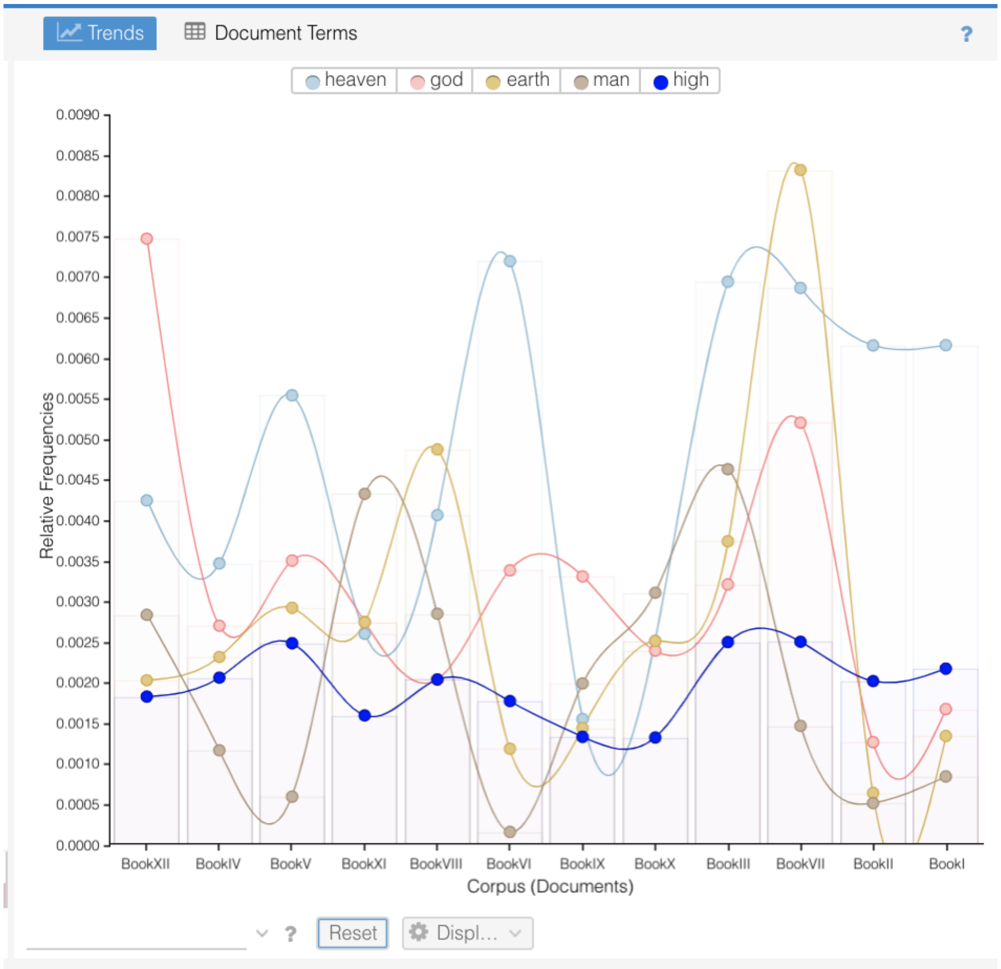
This is the stoplist, one term per line.

yourselves
z
shall
things
like

You will be taken back to the Options pop-up-window. Click Confirm. The word cloud should automatically update with the words, "shall," "things," and "like" filtered out of the word cloud.

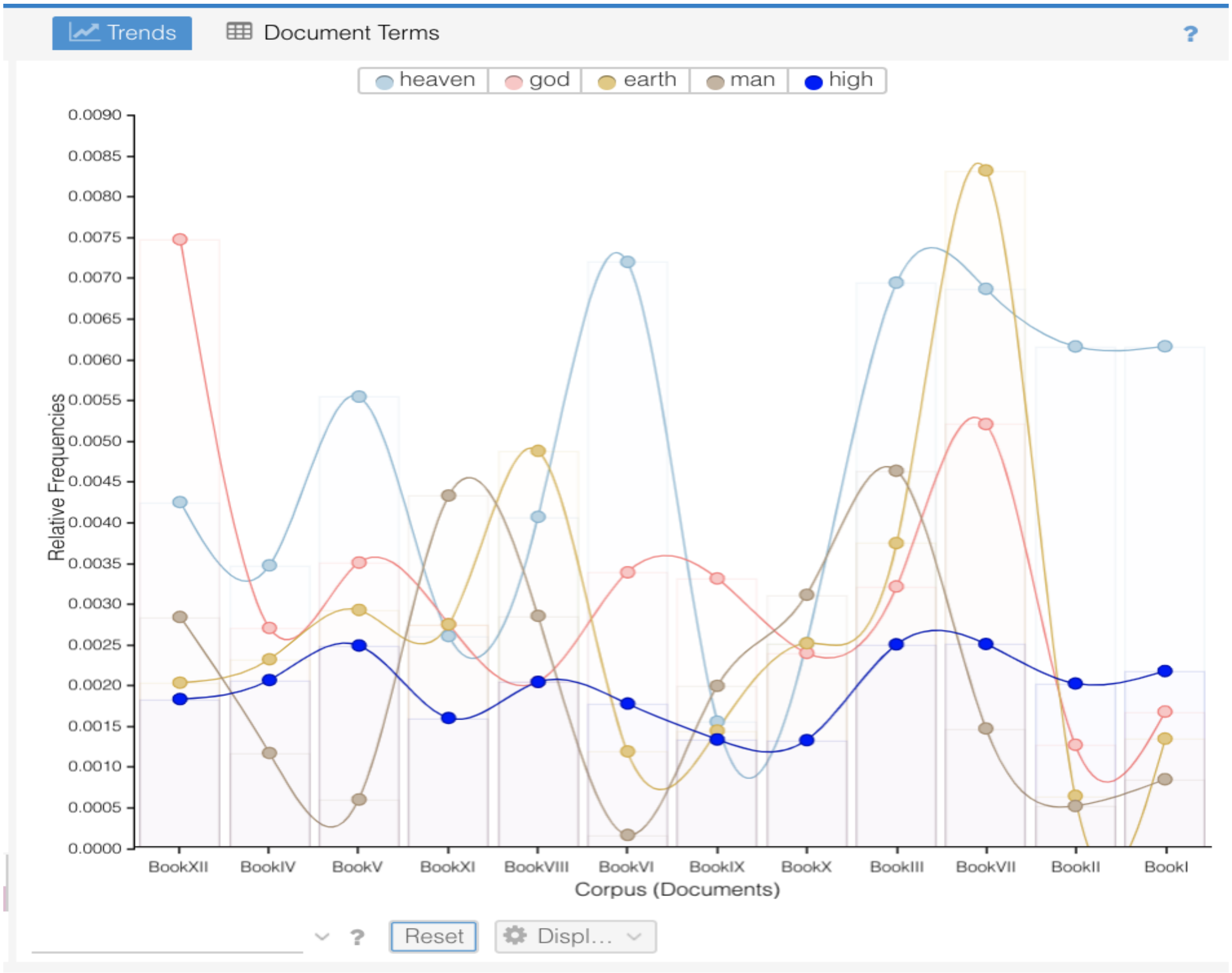


The words will only filter out in the word cloud. You will need to click the Reset button in the Trends tool for the rest of the tools in Voyant to update with the updated stopwords list filtered out.



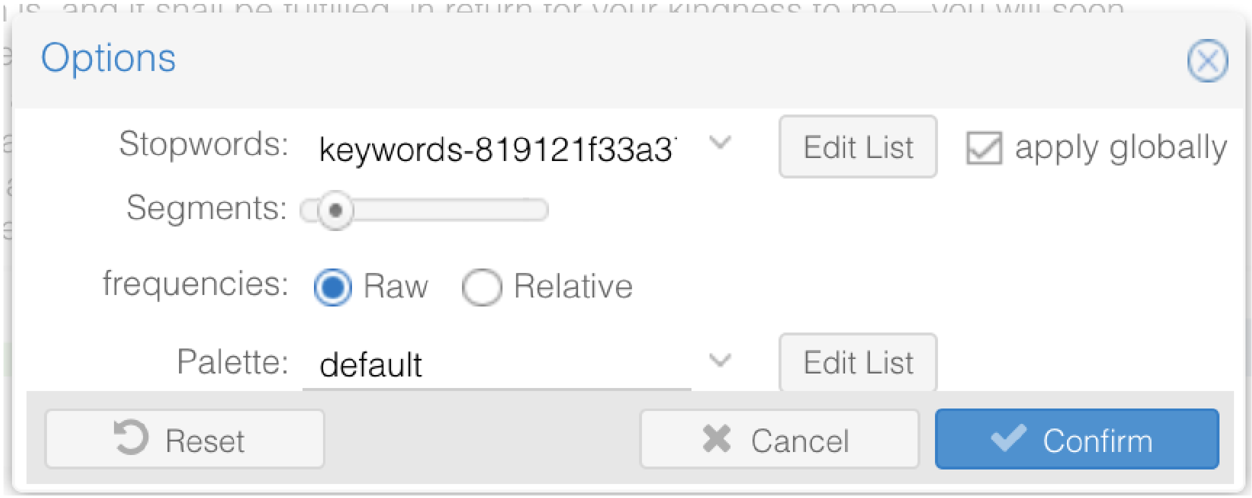
Step 4: Trends and Frequency of Words

The Trends tool, also known as the Term Frequencies Chart, shows a line graph of the most frequent words used in your corpus. Each series in the graph is colored according to the word it represents. At the top of the graph a legend displays which words are associated with certain colors. You can click on words in the legend to toggle their visibility. Hovering over any point in the graph causes a callout box to appear with information about the term selected and its frequency.



By default, the trends tool shows the Relative Frequencies of words in your corpus. To view the absolute count for each document in your corpus, you will need to select Raw Frequencies. Click on the blue options icon ● in the grey menu bar.

In the Options pop-up window, click on Raw Frequencies, then hit Confirm. The Trends tool should update automatically with the absolute count for each of the top words in your corpus.



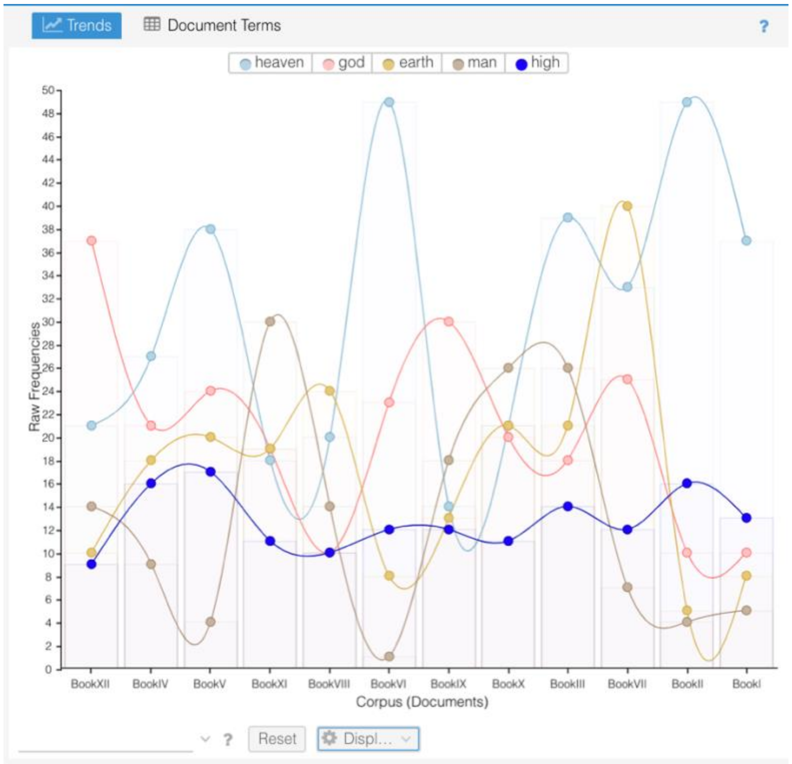
Options

Stopwords: keywords-819121f33a3 Edit List ☑ apply globally

Segments:

frequencies: ● Raw ○ Relative

Palette: default Edit List

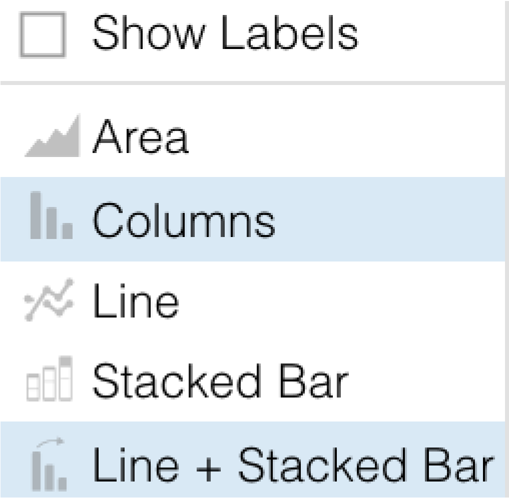


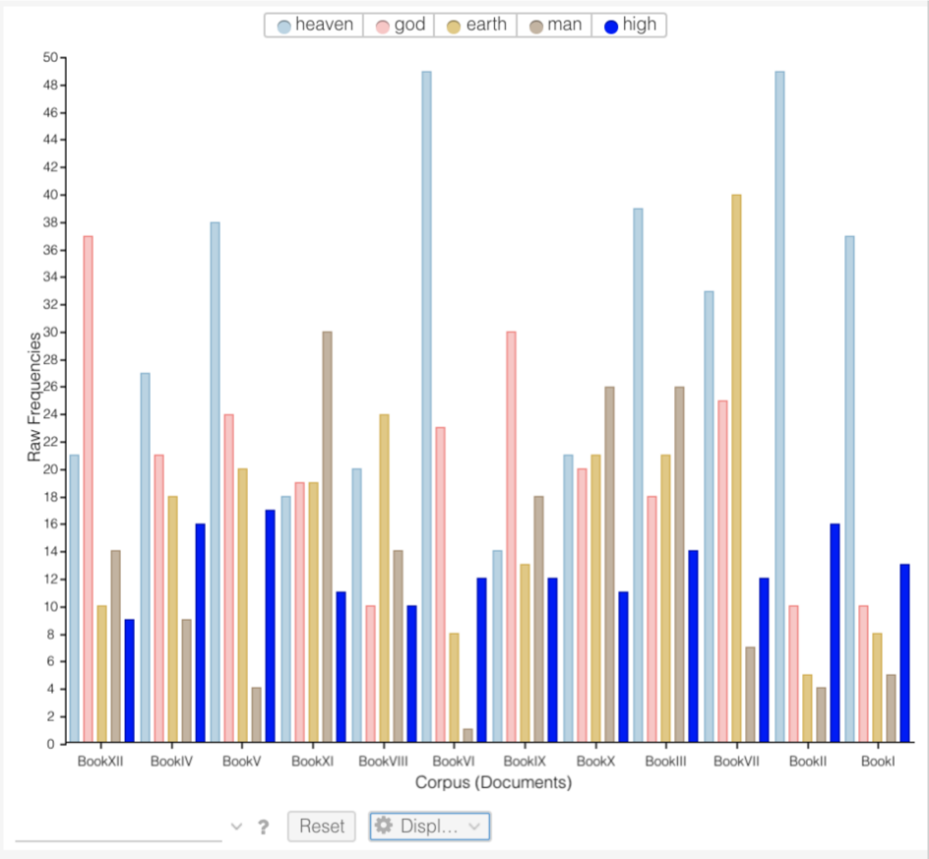
You can click the Reset button to return to the defaults for the tool at any time. The Display button has two components, Show Labels and Chart Mode.

- Show Labels: determines if each item in the chart should be labeled
- Chart Mode:
    - Area: an area chart (labels aren't available for this kind of chart)

- Columns: each item is its own column for each category
- Line: line chart across categories
- Stacked Bar: stacked bar chart (values are shown in columns)
- Line & Stacked Bar (default): superimposed line and stacked bar chart

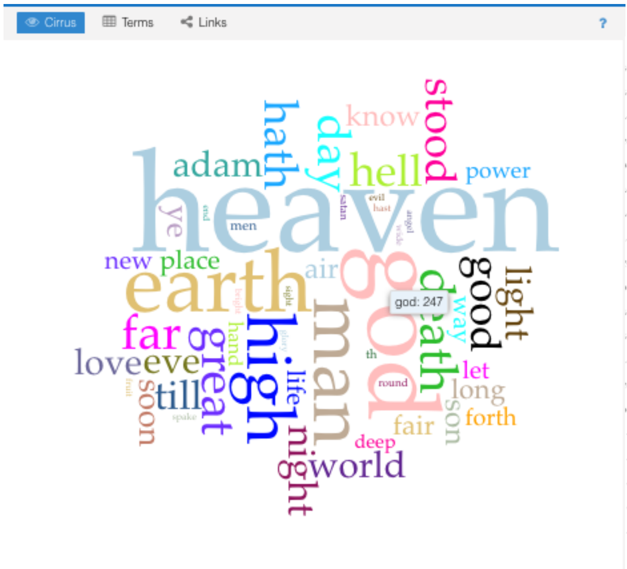
Let's change the display to Columns. Select the Display dropdown menu, then click on Columns. Your chart should automatically update with a bar chart.

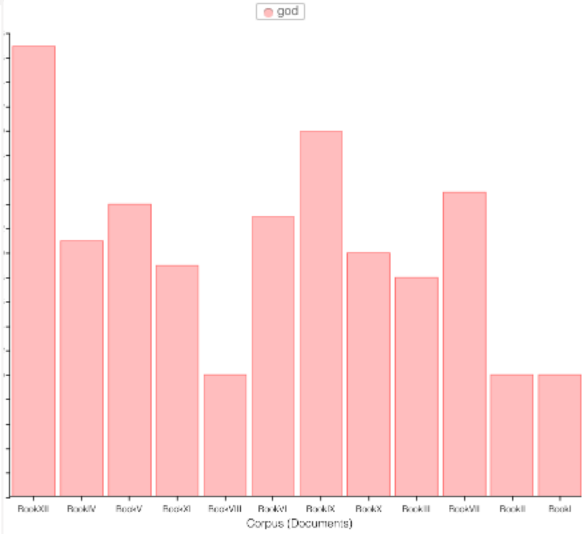




To select a particular word in Trends, click on that word in the word cloud Cirrus tool. The Trends tool will update with the frequency of that selected word across your corpus.

For example, click on the word, “god” in the word cloud and then take a look at the Trends column chart. You will see that the word, “god” shows up.



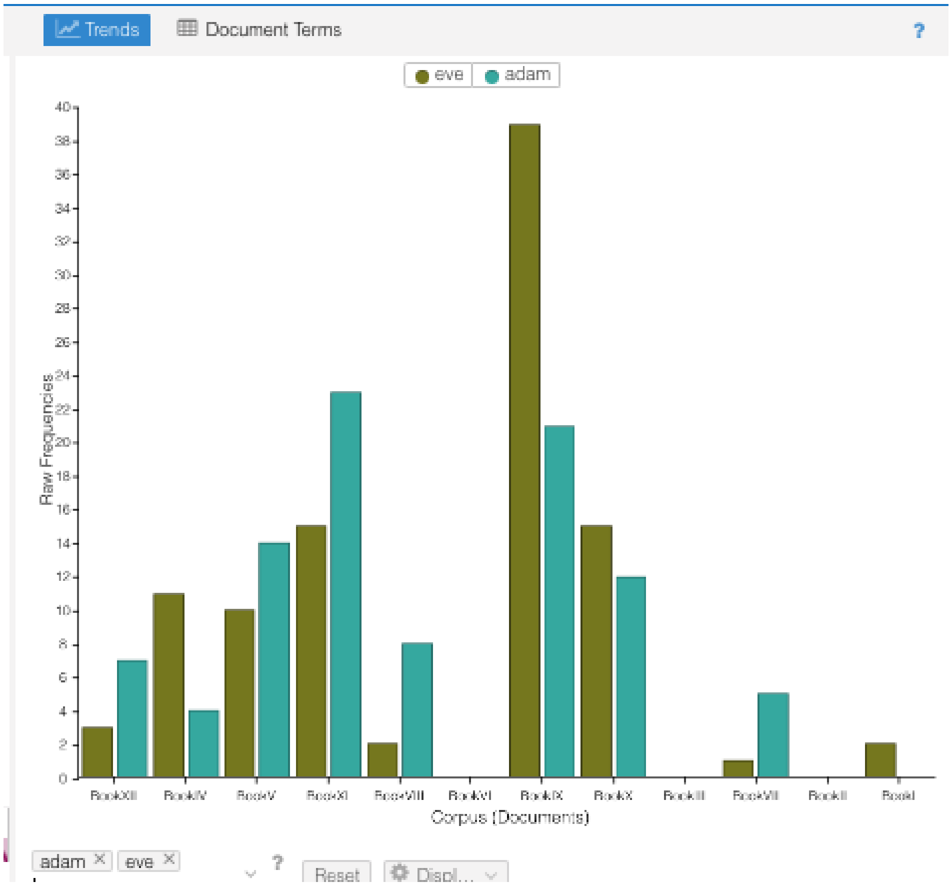


Step 5: Search Queries

If you would like to look at a particular term in your corpus, the Trends tool has a search box, where you can specify an advanced search query. You can search for multiple words in the search box.

Let's search for the words, "adam" and "eve" together and see how often they appear in your corpus.

Type the word "adam" and "eve" as separate words, then click enter. You will see the Trends tool update with the number of times these words show up in your corpus. Keep in mind that Voyant Tools does not capture Adam with a capital “A.” Every word is converted to lowercase, so you will have to use “adam” and “eve” instead.



As you can see from examining the above Trends chart, the adam shows up 94 times and eve shows up 98 times. There are several tools in the search box that allow you to do an advanced search query. Here are some examples of other advanced keyword searches you can do:

- eve: match exact term eve
- eve*: match terms that start with the prefix eve and then a wildcard as one term
- ^eve*: match terms that start with eve as separate terms(ever, even, evening, etc.)
- *ve: match terms that end with the suffix ve as one term
- ^*ve: match terms that end with suffix ve as separate terms(leave, live, serve, etc.)
- eve, adam: match each term separated by commas separate terms
- eve\|adam: match terms separated by pipes as a single term

- “eve and adam”: as an exact phrase(word order matters)
- “eve and adam”~0: phrase(word order doesn't matter but 0 words in between)
- “adam and eve”~5: match adam near eve(within 5 words)

Step 6: Words in Context

Word frequencies can only take you so far with your analysis. Let's move on and examine some keywords in context.

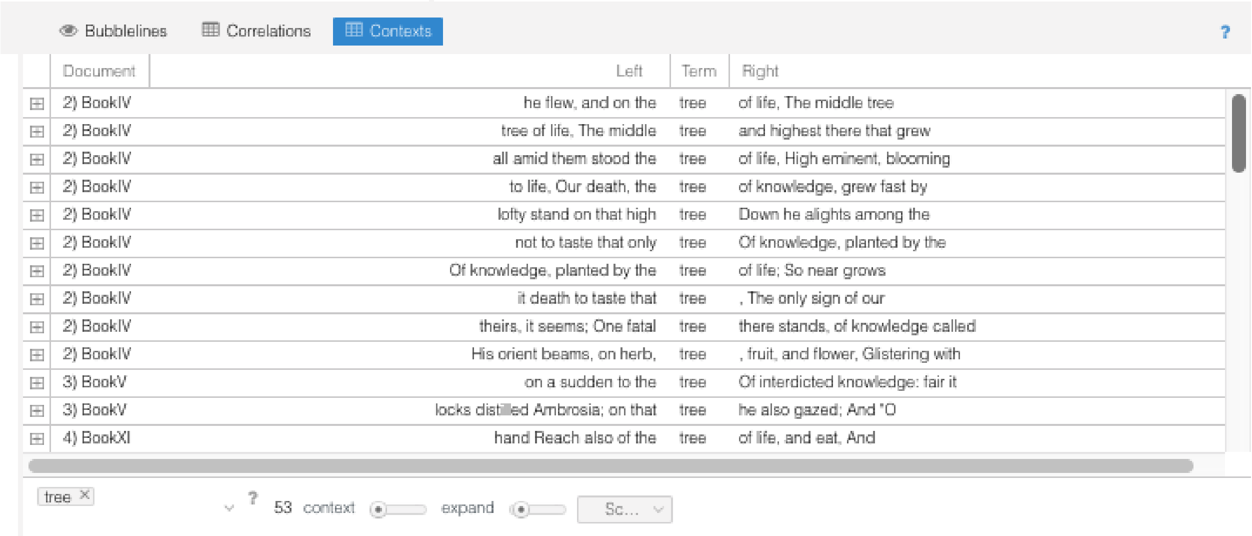
The Context Tool

The Contexts (or Keywords in Context) tool shows each occurrence of a keyword with a bit of surrounding text (the context). The table view shows the following three columns:

- *Document*: this displays which keyword and contexts occur together
- *Left*: this displays contextual words to the left of the keyword (note that sorting by this column treats words in reverse order, right to left from the keyword)
- *Term*: this displays the keyword matching the default or user-provided term query
- *Right*: this displays the contextual words to the right of the keyword

By default, contexts are shown for the most frequent words in your corpus. If you would like to search a different word, you can specify a term in the Context tool's search box.

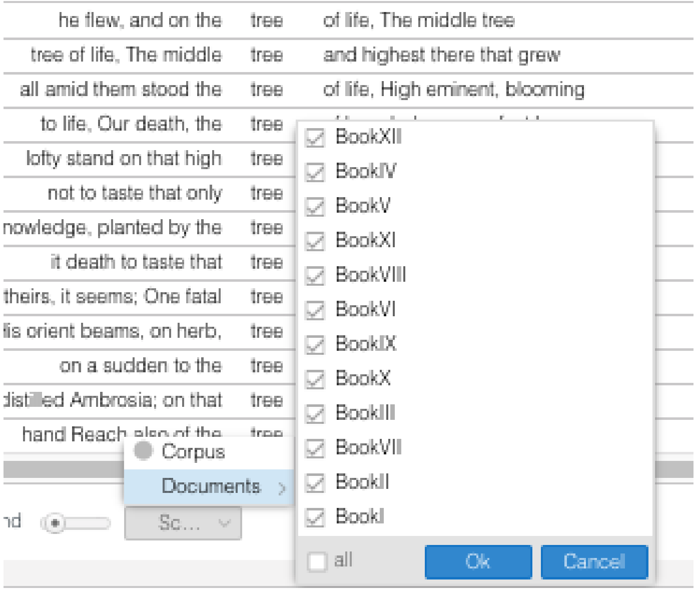
Let's examine the frequency of the word, "tree" in context. In the search box, type "tree" and click enter. The term will appear with its surrounding words (the context), organized by book.



	Document	Left	Term	Right
⊞	2) BookIV	he flew, and on the	tree	of life, The middle tree
⊞	2) BookIV	tree of life, The middle	tree	and highest there that grew
⊞	2) BookIV	all amid them stood the	tree	of life, High eminent, blooming
⊞	2) BookIV	to life, Our death, the	tree	of knowledge, grew fast by
⊞	2) BookIV	lofty stand on that high	tree	Down he alights among the
⊞	2) BookIV	not to taste that only	tree	Of knowledge, planted by the
⊞	2) BookIV	Of knowledge, planted by the	tree	of life; So near grows
⊞	2) BookIV	it death to taste that	tree	, The only sign of our
⊞	2) BookIV	theirs, it seems; One fatal	tree	there stands, of knowledge called
⊞	2) BookIV	His orient beams, on herb,	tree	, fruit, and flower, Glistering with
⊞	3) BookV	on a sudden to the	tree	Of interdicted knowledge: fair it
⊞	3) BookV	locks distilled Ambrosia; on that	tree	he also gazed; And "O
⊞	4) BookXI	hand Reach also of the	tree	of life, and eat, And

You can see the word, "tree" used throughout the entire corpus by clicking on the Scale drop-down menu and then selecting "corpus."

If you would like to look at one book at a time, you can click the Scale drop-down menu again and select Documents. This will output a list of documents in your corpus. You can then checkmark the book you would like to view in context.

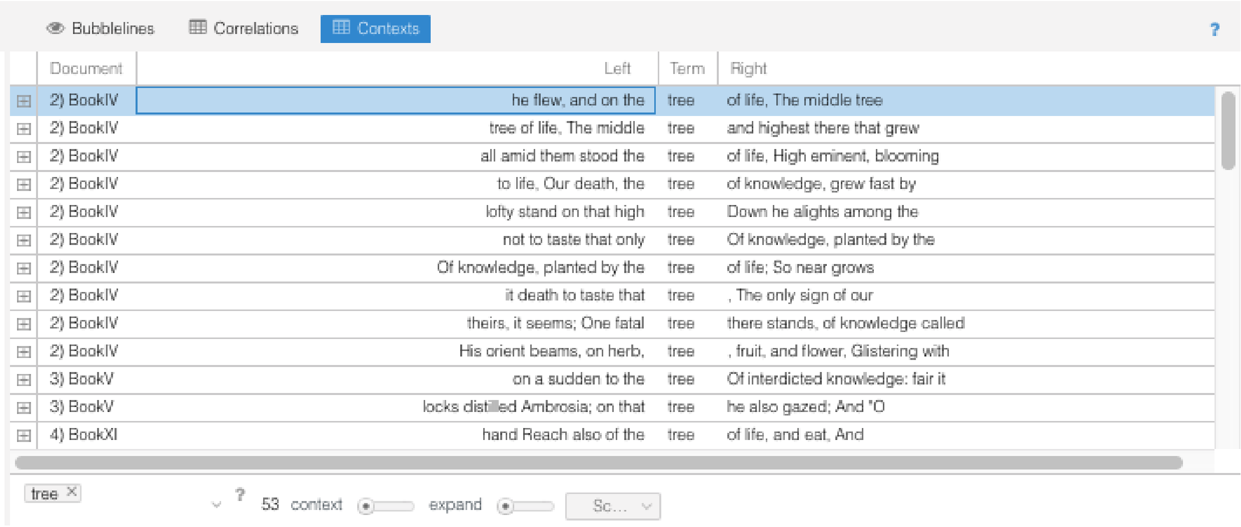


he flew, and on the	tree	of life, The middle tree
tree of life, The middle	tree	and highest there that grew
all amid them stood the	tree	of life, High eminent, blooming
to life, Our death, the	tree	
lofty stand on that high	tree	
not to taste that only	tree	
nowledge, planted by the	tree	
it death to taste that	tree	
theirs, it seems; One fatal	tree	
lis orient beams, on herb,	tree	
on a sudden to the	tree	
distilled Ambrosia; on that	tree	
hand Reach also of the	tree	

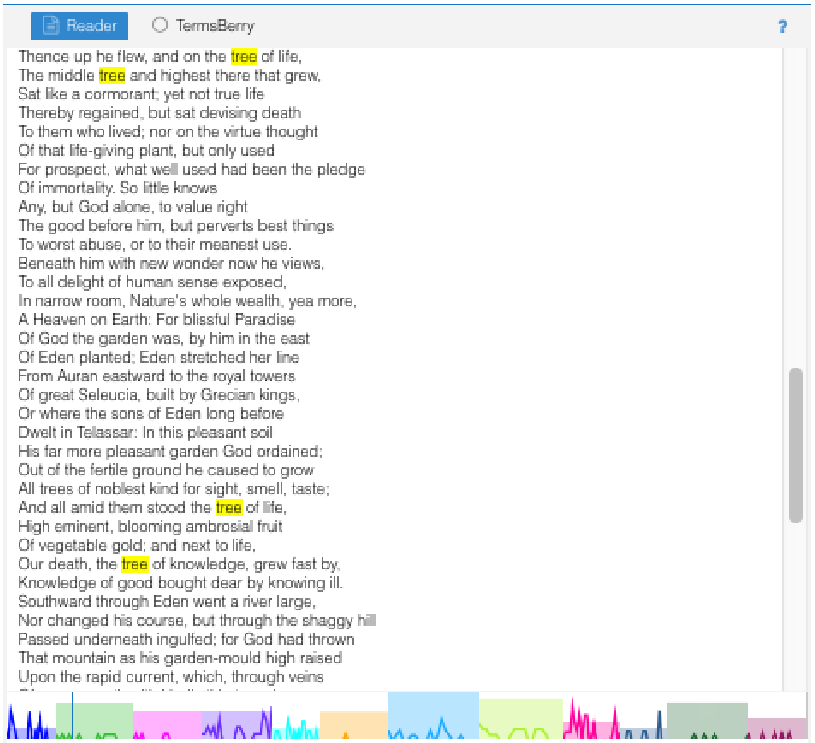
In addition to selecting specific texts, you can also expand your context search using the context slider. The context slider displays the words to the Left and Right of your search term. By default, the context is set to 5 words per side. The value specifies the number of words to consider on each side of the keyword (so the total window of words is double).

The context slider can have a maximum of 50. Similarly, there's an "expand" slider which determines how many words to show when you expand any given row (by clicking the plus icon in the left-most column) you can increase the number of words. The default is 50, the minimum is 5, and the maximum is 500 words.

Let's click on the first row in the Contexts Tool. When you select a specific row in the Contexts Tool, the Reader Tool will update with your search term highlighted in yellow. It will also display where that row of text appears in your corpus.



	Document	Left	Term	Right
⊞	2) BookIV	he flew, and on the	tree	of life, The middle tree
⊞	2) BookIV	tree of life, The middle	tree	and highest there that grew
⊞	2) BookIV	all amid them stood the	tree	of life, High eminent, blooming
⊞	2) BookIV	to life, Our death, the	tree	of knowledge, grew fast by
⊞	2) BookIV	lofty stand on that high	tree	Down he alights among the
⊞	2) BookIV	not to taste that only	tree	Of knowledge, planted by the
⊞	2) BookIV	Of knowledge, planted by the	tree	of life; So near grows
⊞	2) BookIV	it death to taste that	tree	, The only sign of our
⊞	2) BookIV	theirs, it seems; One fatal	tree	there stands, of knowledge called
⊞	2) BookIV	His orient beams, on herb,	tree	, fruit, and flower, Glistering with
⊞	3) BookV	on a sudden to the	tree	Of interdicted knowledge: fair it
⊞	3) BookV	locks distilled Ambrosia; on that	tree	he also gazed; And "O
⊞	4) BookXI	hand Reach also of the	tree	of life, and eat, And



Thence up he flew, and on the tree of life,
The middle tree and highest there that grew,
Sat like a cormorant; yet not true life
Thereby regained, but sat devising death
To them who lived; nor on the virtue thought
Of that life-giving plant, but only used
For prospect, what well used had been the pledge
Of immortality. So little knows
Any, but God alone, to value right
The good before him, but perverts best things
To worst abuse, or to their meanest use.
Beneath him with new wonder now he views,
To all delight of human sense exposed,
In narrow room, Nature's whole wealth, yea more,
A Heaven on Earth: For blissful Paradise
Of God the garden was, by him in the east
Of Eden planted; Eden stretched her line
From Auran eastward to the royal towers
Of great Seleucia, built by Grecian kings,
Or where the sons of Eden long before
Dwelt in Telassar: In this pleasant soil
His far more pleasant garden God ordained;
Out of the fertile ground he caused to grow
All trees of noblest kind for sight, smell, taste;
And all amid them stood the tree of life,
High eminent, blooming ambrosial fruit
Of vegetable gold; and next to life,
Our death, the tree of knowledge, grew fast by,
Knowledge of good bought dear by knowing ill.
Southward through Eden went a river large,
Nor changed his course, but through the shaggy hill
Passed underneath ingulfed; for God had thrown
That mountain as his garden-mould high raised
Upon the rapid current, which, through veins

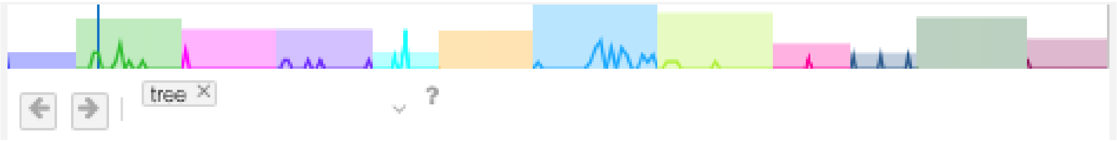
The Reader Tool

The Reader tool displays your entire corpus and is composed of two visual components: the text reader and the prospect viewer.

The text reader displays the entire text of your corpus. With the text reader you can:

- scroll down within the text reader to fetch more content
- hover over a word to show its frequency in the document
- click on a word to search for it in the Reader (and other tools if applicable)

The prospect viewer shows an overview of the entire corpus. This is especially useful when there are multiple documents in a corpus. The bars represent each document in the order they appear in the corpus.



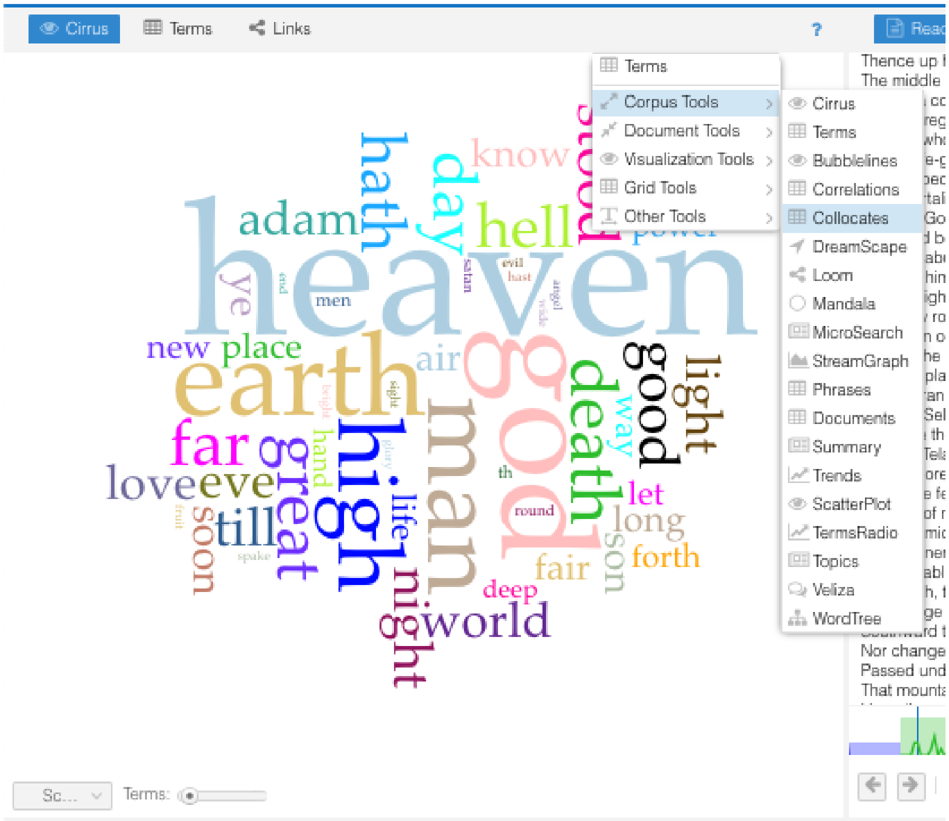
The relative length of the document is represented both vertically and horizontally (in other words, the taller and wider a document is shown, the longer it is). When you've searched for a particular term, an inner sparkline is shown at the top of the bars – this shows the relative frequency of terms. By default, each document is broken into segments of 25 equal parts for the sparkline.

There's also a thin vertical blue bar that indicates the current position of the text reader in the corpus. You can click anywhere along the prospect viewer to jump to another location. To go forward or backward in the text, use the arrows next to the search box.

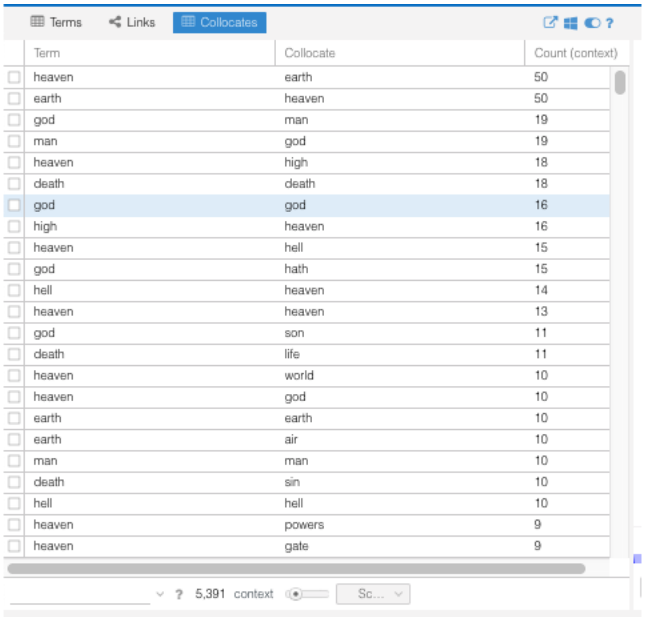
Step 7: Examining Collocates

While the Context Tool allows you to see what words surround a keyword in your corpus, the Collocates Tool shows you which terms appear most frequently in proximity to one another. Let's take a look at what terms appear in close proximity to the word, "tree" within your Paradise Lost corpus.

To select the Collocates Tool, hover over the question mark symbol: ? in the grey bar at the top of the Cirrus Tool until a menu of icons appear: ↗ ⊞ ◐ ?. Click on the window menu icon: ⊞ , and select the "Corpus Tools." Now scroll down through the list of tools and click on "Collocates."



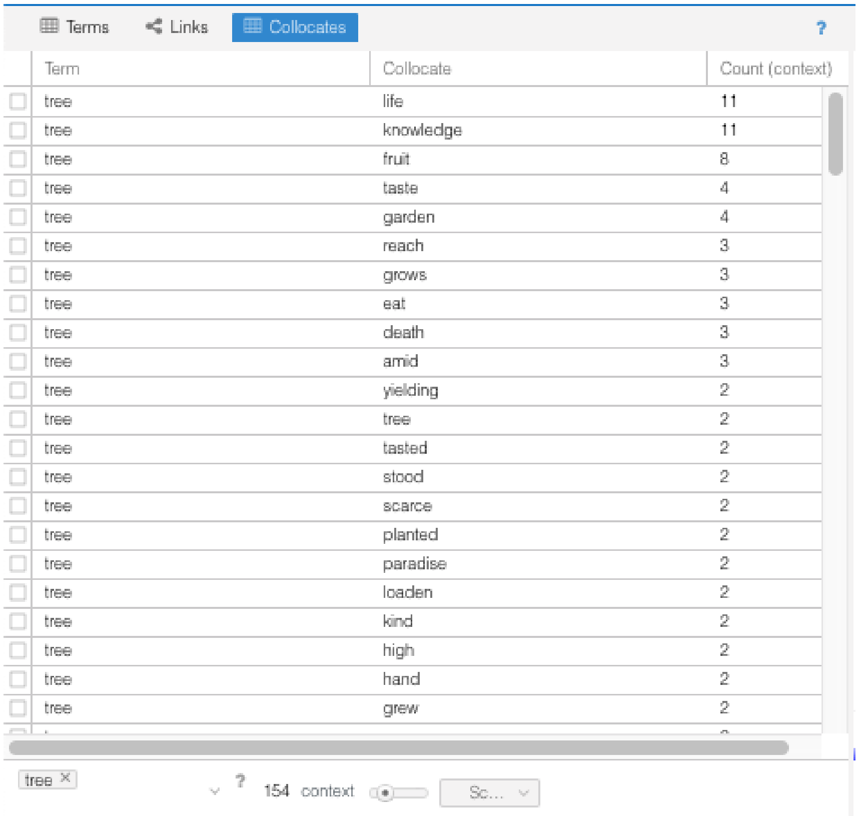
Once selected, the Collocates Tool should automatically appear and replace your Cirrus Tool (wordcloud).



Term	Collocate	Count (context)
heaven	earth	50
earth	heaven	50
god	man	19
man	god	19
heaven	high	18
death	death	18
god	god	16
high	heaven	16
heaven	hell	15
god	hath	15
hell	heaven	14
heaven	heaven	13
god	son	11
death	life	11
heaven	world	10
heaven	god	10
earth	earth	10
earth	air	10
man	man	10
death	sin	10
hell	hell	10
heaven	powers	9
heaven	gate	9

? 5,391 context Sc...

Now type in the word, "tree" into the empty search box of the Collocates Tool and click enter. You should see a table view of the most frequent words that appear in close proximity to the word, "tree."



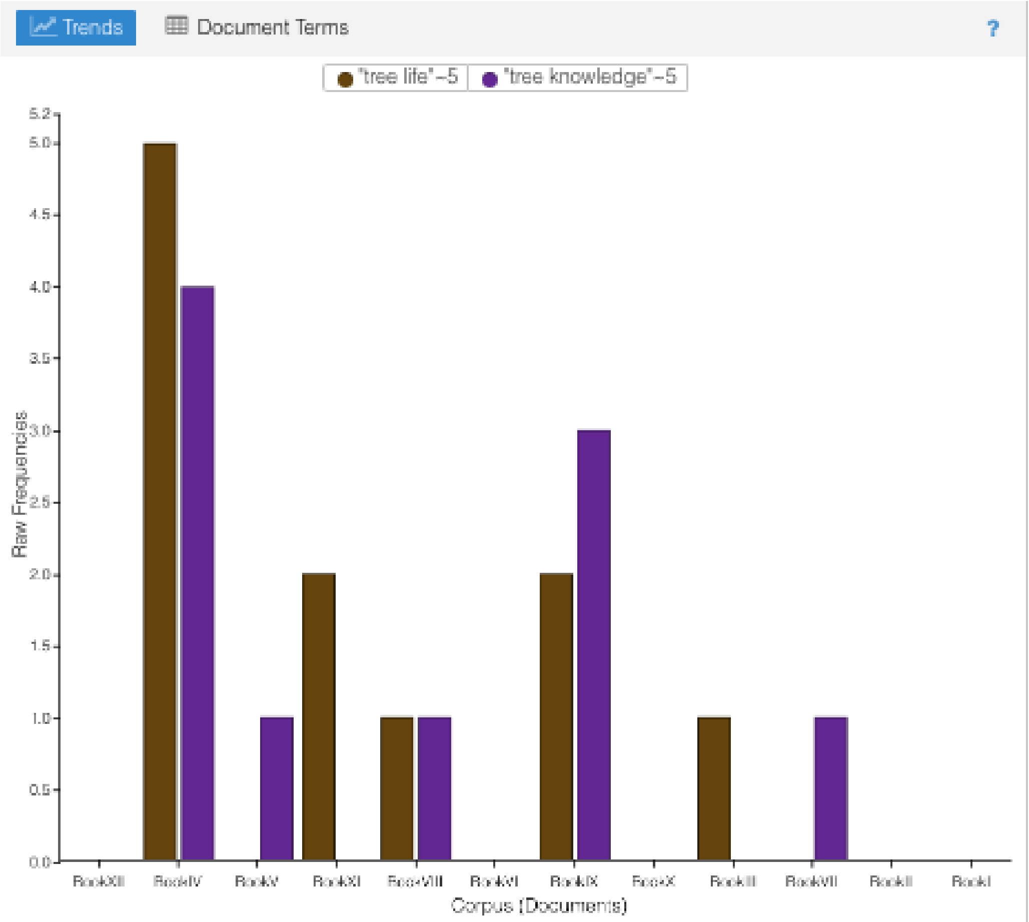
Term	Collocate	Count (context)
tree	life	11
tree	knowledge	11
tree	fruit	8
tree	taste	4
tree	garden	4
tree	reach	3
tree	grows	3
tree	eat	3
tree	death	3
tree	amid	3
tree	yielding	2
tree	tree	2
tree	tasted	2
tree	stood	2
tree	scarce	2
tree	planted	2
tree	paradise	2
tree	loaden	2
tree	kind	2
tree	high	2
tree	hand	2
tree	grew	2

By default, the table view shows the following three columns:

- Term: this displays the keyword (or keywords) being searched.
- Collocates: these are the words found in the proximity of each keyword(s).
- Count (context): this displays the frequency count of the collocate occurring in proximity to the keyword.

As you can see, the word, "tree" shows up most frequently (11 times to be exact) with the collocate, "life." Variations of this collocate appear in the corpus, with "tree" and “knowledge” also occurring 11 times, and “tree” and “fruit” occurring 8 times.

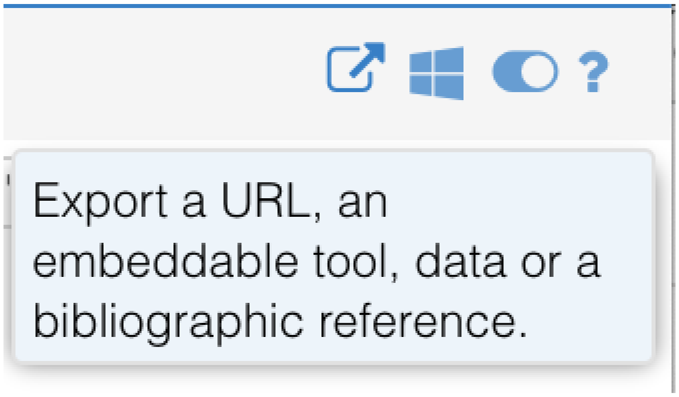
If you would like to compare collocates across your corpus, checkmark the collocates you would like to view. Let's compare the collocates, “tree” and “life,” and “tree” and “knowledge.” Notice that the trends tool automatically updates with the frequencies of the two collocates you've selected.



You will see in the Trends tool the following collocate frequencies: “tree” shows up within 5 words of “life” 5 times, while “tree” shows up within 5 words of “knowledge” 4 times in Book IV. You can also see that these collocate frequencies take place in Book V, Book XI, Book VIII, Book IX, Book III, and Book VII.

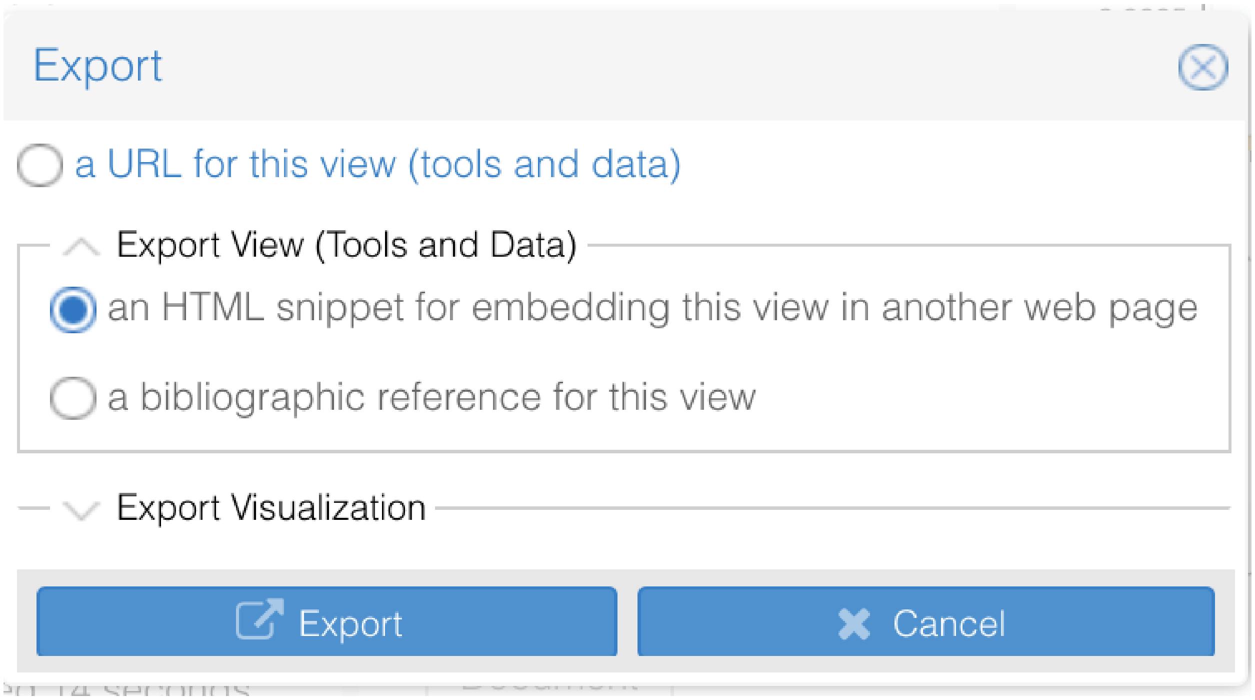
Step 8: Embedding Voyant Tools

Voyant is designed to function as a standalone environment (voyant-tools.org) or as a set of more independent modules that can be embedded into remote sites. The export feature allows you to embed Voyant Tools in other sites. Hover over the grey bar until a menu of icons appears:



Export a URL, an embeddable tool, data or a bibliographic reference.

Click the Export icon: ↗. In the Export window, you will be given three to four different choices when exporting:



Export

○ a URL for this view (tools and data)

Export View (Tools and Data)

● an HTML snippet for embedding this view in another web page

○ a bibliographic reference for this view

Export Visualization

Export: Voyant Tools

- *a URL for this view* (tools and data), default option:

Export View: Tools and Data

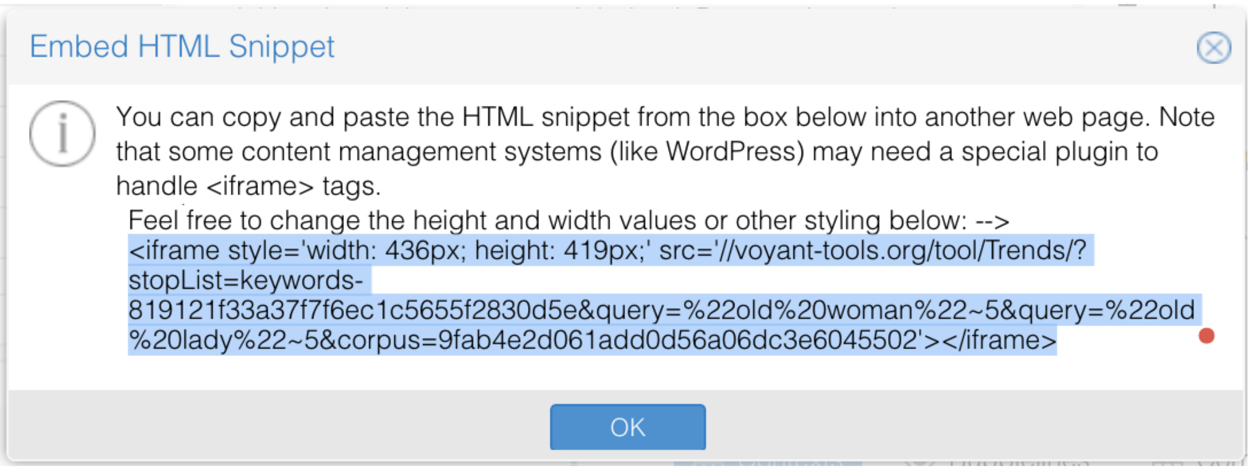
- *an HTML snippet for embedding this view in another web page*, under “Export View (Tools and Data)”: Returns a snippet of HTML code, which can be used to embed this session of Voyant Tools into a webpage.
- *a bibliographic reference for this view*, under “Export View (Tools and Data)”: Returns a bibliographic reference for this session of Voyant Tools.

Export Visualization:

- *export a PNG image of this visualization*, under “Export Visualization,” only accessible when exporting a specific tool: creates a PNG image of the current tool, or creates a snippet of HTML code which contains the image.

You will need to generate an HTML snippet to embed the current corpus and tool. Click the Export icon, expand the Export View section, select the *HTML snippet* radio button, and then click the Export button.

You should get a pop-up window with an HTML snippet of code you can use to embed your selected tool and corpus into another website. Copy and paste the iframe code that appears in the box into another website.



You can copy and paste the HTML snippet from the box below into another web page. Note that some content management systems (like WordPress) may need a special plugin to handle <iframe> tags.

```
Feel free to change the height and width values or other styling below: -->
<iframe style='width: 436px; height: 419px;' src='//voyant-tools.org/tool/Trends/?
stopList=keywords-
819121f33a37f7f6ec1c5655f2830d5e&query=%22old%20woman%22~5&query=%22old
%20lady%22~5&corpus=9fab4e2d061add0d56a06dc3e6045502'></iframe>
```

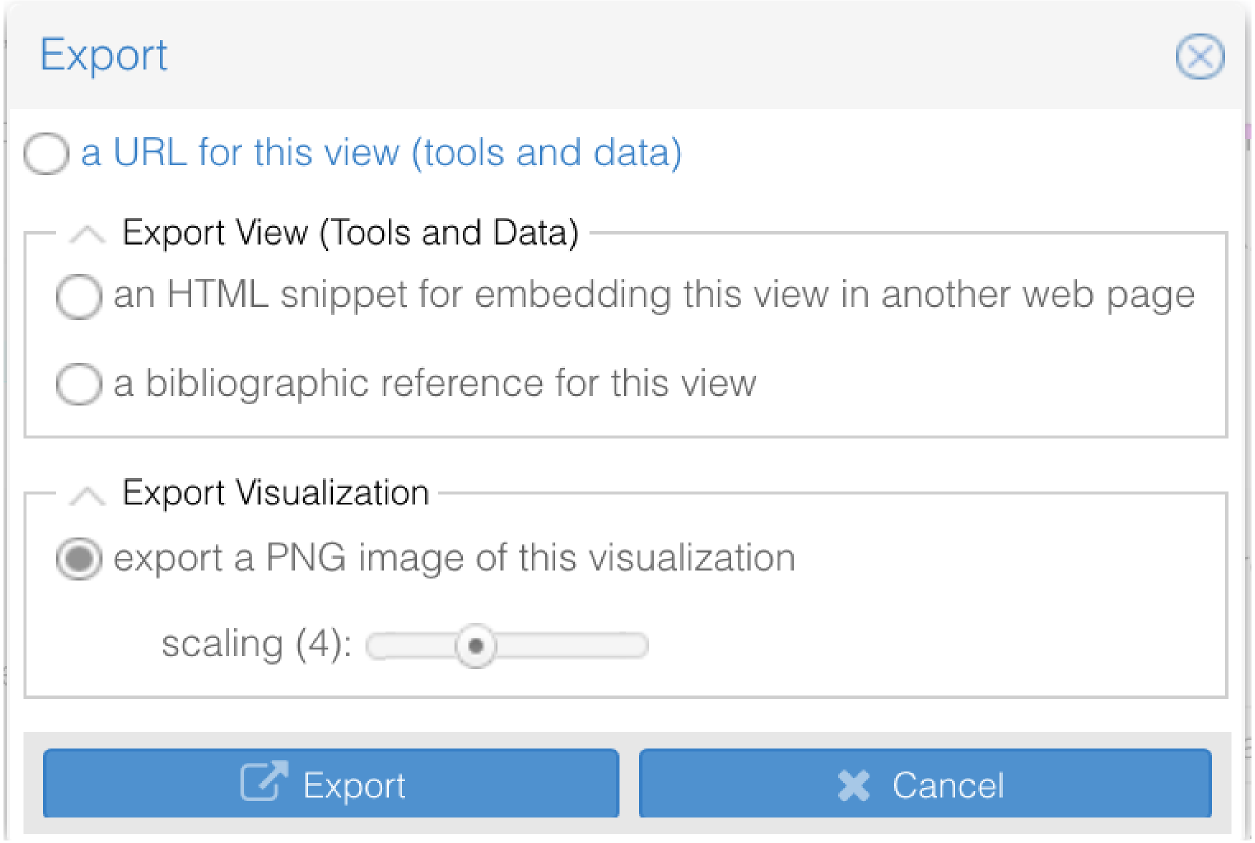
Step 9: Exporting Images from Voyant Tools

If you would like to export an image from Voyant Tools, select *export a PNG image of this visualization* under "Export Visualization." You will see the following text in the export window:

Export Visualization:

- *export a PNG image of this visualization*, under “Export Visualization,” only accessible when exporting a specific tool: creates a PNG image of the current tool, or creates a snippet of HTML code which contains the image.

After you have selected this option, you will also need to change the scale of the PNG using the scale slider bar.



Export

○ a URL for this view (tools and data)

Export View (Tools and Data)

○ an HTML snippet for embedding this view in another web page

○ a bibliographic reference for this view

Export Visualization

● export a PNG image of this visualization

scaling (4):

This scale slider bar will increase or decrease the size of your image file. Let's set the scale of your image to (3.5) and click Export. Your image should begin downloading immediately.

Step 10: Voyant Tools Resources

This tutorial only scratches the surface of what you can do with Voyant Tools and the types of text analytics you can perform. To learn more about Voyant Tools, see the following resources:

- Voyant Help Guide
- Voyant Tools Documentation
- Voyant Tools GitHub
- Voyant Tools Twitter: @VoyantTools
- Voyant Tools: A Tutorial for Text Analysis