|  |  |  |
| --- | --- | --- |
|  |  | **Voyant as a text mining tool** |
|  |  |  |
| Objective |  | To learn how to use Voyant to explore the language used in digital humanities journals, conference proceedings and books. * Voyant allows us to see what words are most common and search for terms you would expect to see.
 |
|  |  |  |
| Material |  | [Journal of Digital Humanities](http://journalofdigitalhumanities.org/) |
|  |  |  |
| Questions |  | What do you notice about the results? What view of digital humanities do the texts seem to take? What do you find surprising? Did this exercise change or expand your view of digital humanities? How might this method be useful in humanities research, and what are its limitations?  |
|  |  |  |
| Steps to follow |  | 1. Go to [journalofdigitalhumanities.org](file:///C%3A%5CUsers%5Cknesbitt%5CAppData%5CLocal%5CMicrosoft%5CWindows%5CTemporary%20Internet%20Files%5CContent.Outlook%5C8ICX5HOT%5Cjournalofdigitalhumanities.org) and copy the URL of two journal volumes into Notepad.
2. Open Firefox (Voyant does not work currently with Internet Explorer).
3. Go to [voyant-tools.org](file:///C%3A%5CUsers%5Cknesbitt%5CAppData%5CLocal%5CMicrosoft%5CWindows%5CTemporary%20Internet%20Files%5CContent.Outlook%5C8ICX5HOT%5Cvoyant-tools.org).
4. Paste the URL for a volume into the Voyant window.
5. Click **Reveal** for Voyant to begin mining the text.
6. Click on the disk icon in the upper-right corner of the Corpus Reader screen.
7. When the Export window opens, click **OK** (leave radio button as is). A URL appears in the next window (note that the corpus ID number appears in the URL as everything to the right of the equal sign).
8. Copy the URL and paste it into Notepad so you can retrieve it a bit later.
9. Click **OK.**
10. Process a second corpus by going back to [voyant-tools.org](http://voyant-tools.org) and repeating steps 4 and 5.

. |
|  |  |  |
|  |  | Lorem ipsum dolor |
| Steps to follow |  | 1. Click **Words in the Entire Corpus** in the bottom left corner of the screen to compare the two corpuses.
2. You will see a list of the most common words in your second corpus.
3. Click on the arrow beside **Trend**.
4. Under **Columns**, click on **Difference** to activate this function. This adds a column that displays the difference between the two corpuses.

You can now add a list of stop words to your corpus. Stop words are words that typically carry little meaning and should be excluded from the results, words such as *the*, *a*, *in*, *to*, *from*, etc.).1. To add a stop words list, click on the options function (the gear icon at the extreme right of **Words in the Entire Corpus**).
2. In the **Stop Words List**, select **English (Taporware)** and then check **Apply Stop Words Globally**.
3. Now, in order to compare both corpuses, add the first corpus by pasting the corpus ID number (the portion of the URL to the right of the equal sign, for example, voyant-tools.org/?corpus=1392906089715.6130) into the **Corpus Comparison Box**.
4. Click **OK**.
5. You’ll notice how interesting the Cirrus word cloud has become! Now navigate over the words to see how many times particular words appear in the corpus.
6. To find out about additional Voyant tools, go to [docs.voyant-tools.org/tools/](file:///C%3A%5CUsers%5Cknesbitt%5CAppData%5CLocal%5CMicrosoft%5CWindows%5CTemporary%20Internet%20Files%5CContent.Outlook%5C8ICX5HOT%5Cdocs.voyant-tools.org%5Ctools%5C).
 |