

# User Guide

## The LNU Lost Generation Web Interface:

---

WHAT IS THE LNU LOST GENERATION WEB INTERFACE? .....	1
GETTING STARTED.....	1
CURRENT CONTENT .....	1
CONCORDANCER .....	2
<i>Saving Results in Google Sheets</i> .....	4
TOKENS AND LEMMA .....	6
N-GRAMS.....	8
CONTACT.....	8

## What is the LNU Lost Generation Web Interface?

The LNU Lost Generation Web Interface (LostGen Web) is an online interface that allows users to access materials via a computer-searchable text collection, or corpus. The corpus can be explored with a number of tools provided by LostGen Web, allowing users to gain insights into the literary styles of the Lost Generation authors. The goal of LostGen Web is to provide an interface for teachers to teach basic computational analysis to their students along with a number of pre-made lesson plans.

## Getting Started

<https://losgen.streamlit.app/>

- Before getting started, make sure you are running the web program on either the **Google Chrome** or **Firefox** browser. LostGen Web will not run on Safari or Internet Explorer.
- Occasionally, when accessing the webpage, you may see a popup that reads “This app has gone to sleep due to inactivity. Would you like to wake it back up?” in this case, press the button “Yes, get this app back up!”

## Current Content

The first screen, “Current Content,” shows all of the files currently compiled in the corpus tool. You can use the search function to see what specific texts are included in the corpus.

For example, to find out what works by F. Scott Fitzgerald are included in the corpus:

1. Type “Fitzgerald” into the search bar

2. Click ▾ in the section titled “select search column.”
3. Click “author”
4. Click “find.”

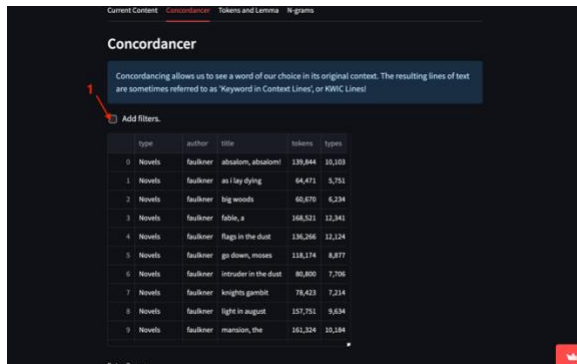
A new box is generated below listing all of the works by F. Scott Fitzgerald that are currently included in the corpus. The results are organized by type (Novel, Other, and Short Stories) and include the tokens (words per file) and types (number of unique words) of each given text.

## Concordancer

The second tab houses the concordancer. A concordance is a line of text taken from a corpus and is often used to show the context surrounding a keyword. The keyword is always placed in the middle of the concordance line, also known as the *Keyword in Context line* or KWIC line.

For example, to find the KWIC lines for the word “bird” in F. Scott Fitzgerald’s works:

1. Select the box next to the “add filters” function.

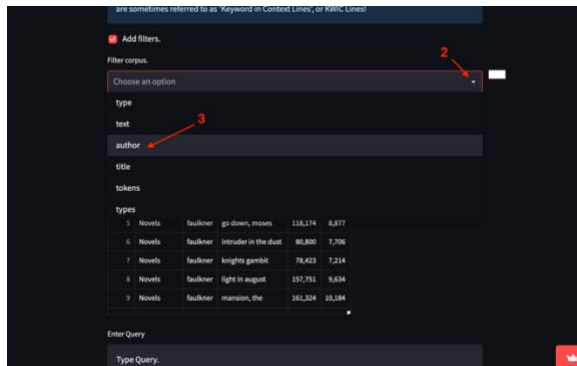


2. Click ▾ in the “filter corpus” box.
3. Select “author” and click ▾ again to close the selection tool.

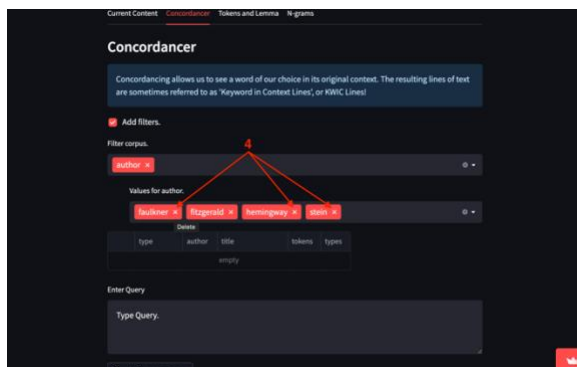
### The More You Know!

Text types that fall under the category of "other" include:

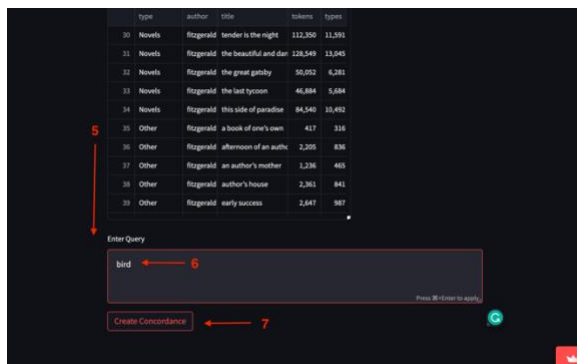
- Letters
- Diary Entries
- Speeches
- Articles



4. A new box labeled “values for author” has appeared below the “filter corpus” box. Click ✕ next to “Faulkner,” “Hemmingway,” and “Stein” to remove them from your search results.



5. Scroll down to the “enter query” box.
6. Type *bird* into the box.
7. Click “create concordance.” A loading symbol appears as the concordancer runs your query; your results may take a few moments to generate.

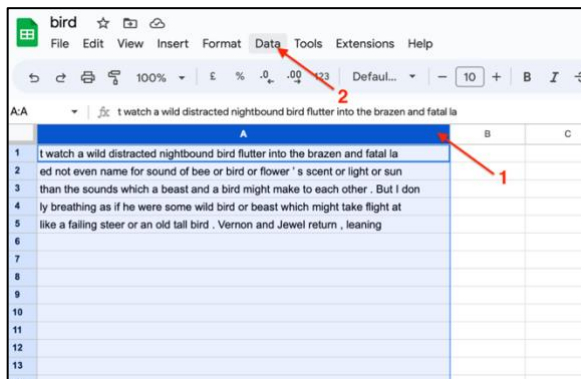


The resulting KWIC lines will show the query word, *bird*, with 36 characters (including spaces) on either side. To save the results, copy and paste the KWIC lines into Excel or Google Sheets, you will find more information about saving and formatting the KWIC lines in the next two sections.

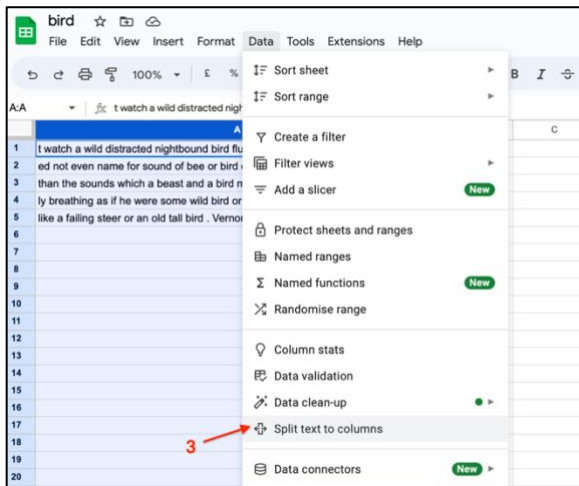
## Saving Results in Google Sheets

When you copy and paste your results into a new Google Sheets document, all of the results will be in one column. To split the text into separate columns and make the results more readable:

1. Select the column by clicking on the title “A.”
2. Click on the “Data” tab.



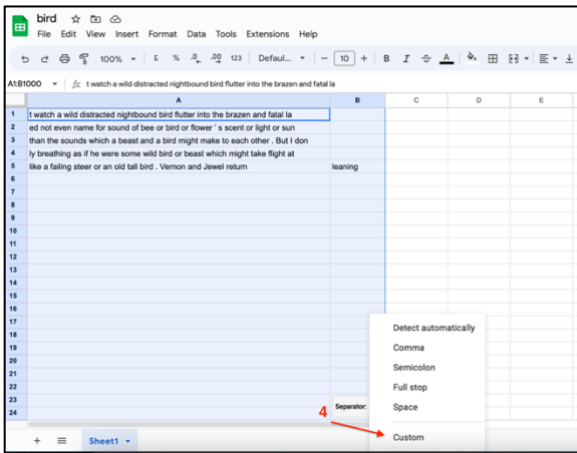
3. Select “Split text to columns.” The Separator function will attempt to automatically detect where you want the columns to be split, it usually is not correct.



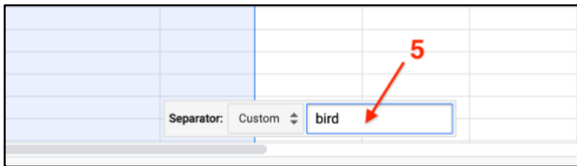
4. On the small separator popup at the bottom of the page, click on “Detect automatically” and a drop-down menu will appear. Click “Custom.”

### The More You Know!

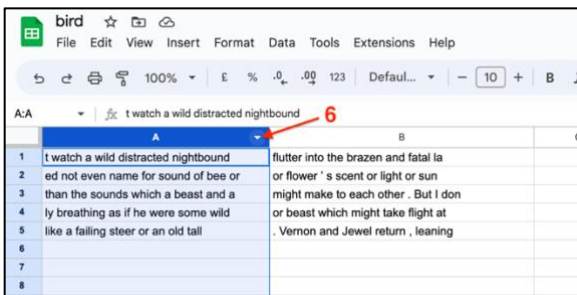
You don’t have to be an expert to perform magic on Google Sheets or Microsoft Excel. Both programs have extensive resources online to help you with any questions you have. By typing your questions into google, you can find numerous links with detailed explanations for how to perform different tasks.



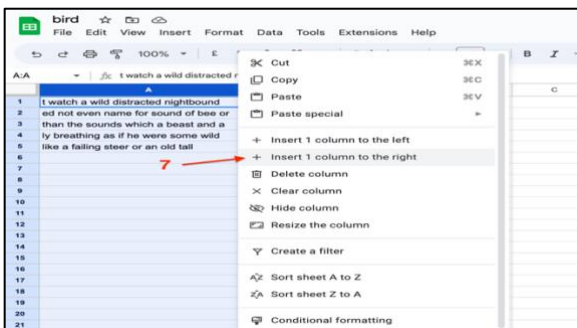
5. Type *bird* into the search bar. The text will now be broken into two columns, column “A” contains all of the text before the word *bird*, and column “B” contains all the words after the word *bird*.



6. To make reading easier, you may want to add a middle column containing the word *bird*. Hover your mouse over column “A” and a ▾ will appear. Click on the ▾ and a to view the dropdown menu.

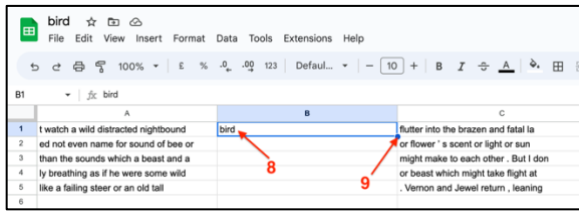


7. Select “Insert 1 column to the right.” A new column appears between the two text columns.

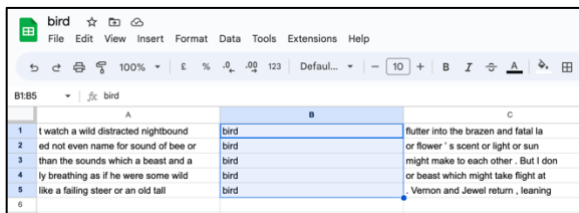
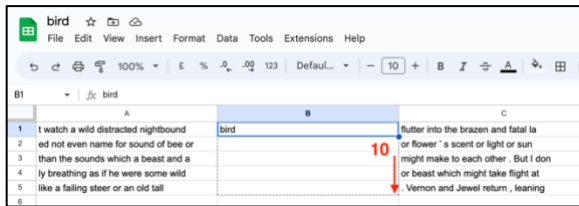


8. Type *bird* into the first cell of column “B.”

9. Click out of the cell and then hover your mouse over the previous cell. A blue dot will appear at the bottom right corner of the cell.



10. Click and drag the blue dot to the final row of data, this will make *bird* visible in all the cells in column “B.”



## Tokens and Lemma

The third tab, “Tokens and Lemma,” shows us the frequency of certain words in chosen texts. These frequencies can inform further corpus research such as collocation or concordance analysis.

**Tokens** are the smallest unit that a corpus consists of such as words, punctuation, and even abbreviations. In the LostGen Web, tokenization only includes words. To find all of the tokens in F. Scott Fitzgerald's work:

1. Select the box next to the “add filters” function.
2. Click ▾ in the “filter corpus” box.
3. Select “author” and click ▾ again to close the selection tool.
4. A new box labeled “values for author” has appeared below the “filter corpus” box. Click ✕ next to “Faulkner,” “Hemmingway,” and “Stein” to remove them from your search results.

5. Scroll down and click the button labeled “Tokenize.” The program will generate a frequency list, this may take a moment.

5.1 The frequency list is generated in random order. To see the list in

**alphabetical** order, click on the title “Token” once, this will show the list in alphabetical order from A-Z. Click the title “Token” a second time to show the list in reverse alphabetical order from Z-A. Clicking the title “Token” again will return the list to its original order.

5.2 To see the list in **numerical order**, click on the title “Frequency” once, this will show the list from lowest frequency to highest frequency. Click the title “Frequency” a second time to show the list from highest frequency to lowest frequency. Clicking the title “Frequency” again will return the list to its original order.

**Lemmas** are the dictionary form of a word, this means that results for “be” will include counts of “be,” “been,” “being,” “is,” “are,” “was,” and “were.” (Whereas in a token search these words would all be separate counts) To find all of the lemmas in F. Scott Fitzgerald's work:

1. Select the box next to the “add filters” function.
2. Click ▾ in the “filter corpus” box.
3. Select “author” and click ▾ again to close the selection tool.
4. A new box labeled “values for author” has appeared below the “filter corpus” box. Click ✕ next to “Faulkner,” “Hemmingway,” and “Stein” to remove them from your search results.
5. Scroll down and click the button labeled “Lemmatize.” The program will generate a frequency list, this may take a moment.

5.1 The frequency list is generated in random order. To see the list in

**alphabetical** order, click on the title “Lemma” once, this will show the list in alphabetical order from A-Z. Click the title “Lemma” a second time to show the list in reverse alphabetical order from Z-A. Clicking the title “Lemma” again will return the list to its original order.

5.2 To see the list in **numerical order**, click on the title “Frequency” once, this will show the list from lowest frequency to highest frequency. Click the title “Frequency” a second time to show the list from highest frequency to lowest frequency. Clicking the title “Frequency” again will return the list to its original order.



## N-grams

The final tab houses the N-gram extractor. N-grams are the groups of words most commonly found in a text or around a specific term. The “N” in n-gram is a variable that stands for the number of words in a specified group, for instance, a bigram is a two-word n-gram, a trigram is a three-word n-gram, and so on.

For example, to look for a trigram (a three-word n-gram) containing the word *bird*.

1. Select the box next to the “add filters” function.
2. Click ▽ in the “filter corpus” box.
3. Select “author” and click ▽ again to close the selection tool.
4. A new box labeled “values for author” has appeared below the “filter corpus” box. Click ✕ next to “Faulkner,” “Hemmingway,” and “Stein” to remove them from your search results.
5. Scroll down to the box titled “n-gram length” and type in the number 3 into the box.
6. In the bar titled “optional: n-grams must include” type in the word *bird*. If you are curious to see the most common three-word clusters in F. Scott Fitzgerald’s work, you can skip this step.
7. Select “get n-grams.” A loading symbol appears as the n-gram extractor runs your query; your results may take a few moments to generate.

If the word you are looking for can be part of a larger word, the N-gram extractor will identify all words containing the smaller word. For example, if you search for the term *tree*, the results will include tree, trees, street, and streets. To prevent this, type a space before and after the word, written [ *tree* ], the N-gram extractor will then only show you the results relating to a singular tree.

## Contact

For help or further access to the corpus, contact Daniel Ihrmark at [daniel.o.sundberg@lnu.se](mailto:daniel.o.sundberg@lnu.se).