

# Exploratory Tools



**Northeastern University**  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point  
during the presentation!*

# Word Counter

- <https://databasic.io/en/wordcounter/>
- A user-friendly basic word counting tool
- It allows you to count words, bigrams, and trigrams in plain text files and to download spreadsheets with your results
- The max file upload is 10MB
- Can be run with and without **stopwords**



# Word Counter Examples

## TOP WORDS ⬇

| Word   | Frequency |
|--------|-----------|
| could  | 356       |
| one    | 314       |
| would  | 308       |
| us     | 292       |
| little | 269       |
| time   | 247       |
| life   | 218       |
| father | 210       |

Shows the top words in the text.

Stopwords aren't removed for the bigrams and trigrams because they need context.

## BIGRAMS ⬇

| bigram <sup>2</sup> | Frequency |
|---------------------|-----------|
| of the              | 885       |
| in the              | 726       |
| i was               | 497       |
| it was              | 371       |
| to the              | 319       |
| on the              | 309       |
| and the             | 296       |
| i had               | 272       |

## TRIGRAMS ⬇

| trigram <sup>3</sup> | Frequency |
|----------------------|-----------|
| i did not            | 89        |
| that i was           | 61        |
| i could not          | 47        |
| i do not             | 45        |
| it was a             | 40        |
| and i was            | 40        |
| there was no         | 38        |
| it was not           | 35        |

It is interesting how many of the trigrams are negations!



**Northeastern University**  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*

# Word Trees

- <https://www.jasondavies.com/wordtree/>
- A word tree depicts multiple parallel sequences of words
- This is a good way to see patterns in word usage, based on words that appear before and after a term or terms of interest.
- There are some restrictions in size: fewer than 1 million words should work

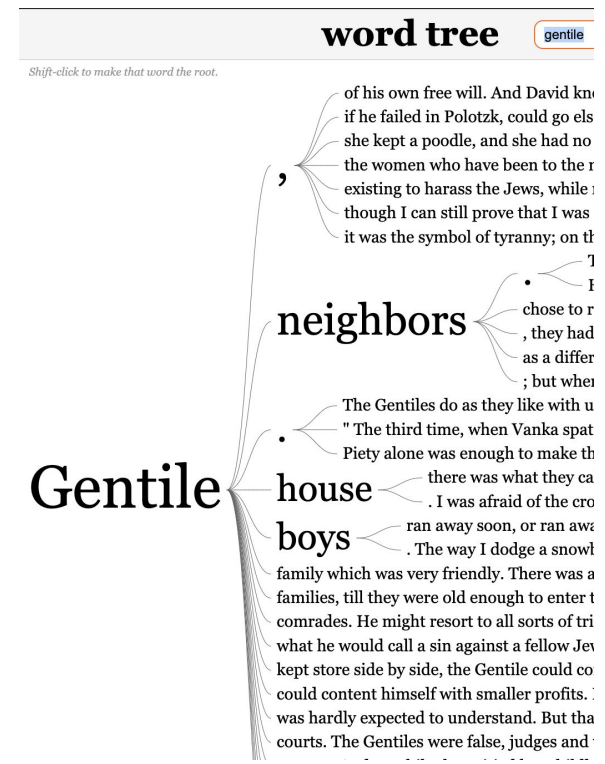
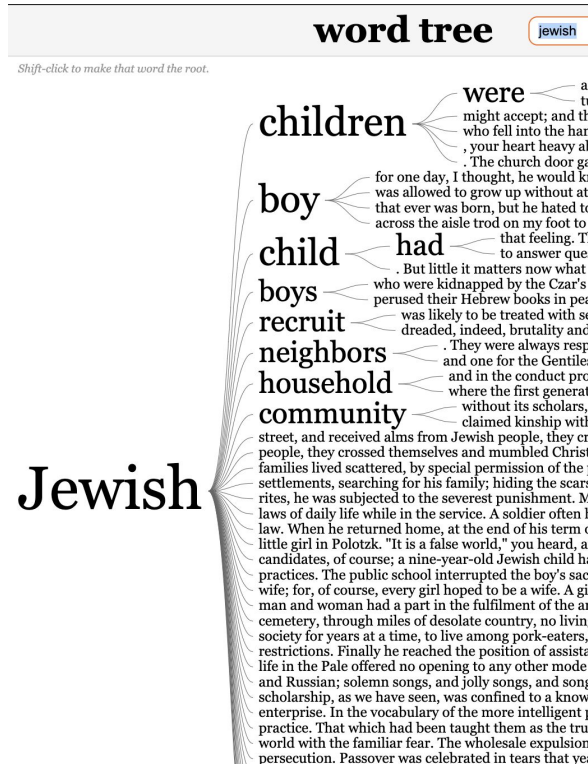


# Word Tree Examples

Reflects the focus of the book as on the Jewish community, while Gentiles are more likely neighbors.

Even though written by a woman, the book references Jewish and Gentile boys more than girls.

The punctuation following Gentile suggests it is often the end of the sentence.



**Northeastern University**  
NULab for Texts, Maps, and Networks

*Feel free to ask questions at any point during the presentation!*

# Word Tree: Reverse Trees

When words are commonly followed by punctuation, it is worth reversing the tree to see the words that often precede it. To do this, click “reverse tree” next to the search bar.



**Northeastern University**  
*NULab for Texts, Maps, and Networks*

*Feel free to ask questions at any point during the presentation!*