



Gutenberg Text Cleaner

Clean and analyze text from Project Gutenberg books

Project Gutenberg URL:

<https://www.gutenberg.org/files/2701/2701-0.txt>

Clean & Analyze Text

Example URLs to try:

[Pride and Prejudice by Jane Austen](#)

[Frankenstein by Mary Shelley](#)

[Alice in Wonderland by Lewis Carroll](#)

[Moby Dick by Herman Melville](#)



Analysis Results

1,188,130

Characters

213,184

Words

10,717

Sentences

4.5

Avg Word Length

19.9

Avg Sentence Length



Top 10 Most Common Words

1. the (14,411)
2. of (6,593)
3. and (6,378)
4. a (4,629)
5. to (4,578)
6. in (4,151)
7. that (2,926)
8. his (2,520)
9. it (2,371)
10. i (1,935)



Summary (First 3 sentences)

moby-dick or the whale. by herman melville contents etymology. extracts supplied by a sub-sub-librarian.



Cleaned Text (First 500 characters)

MOBY-DICK; or, THE WHALE. By Herman Melville CONTENTS ETYMOLOGY. EXTRACTS (Supplied by a Sub-Sub-Librarian). CHAPTER 1. Loomings. CHAPTER 2. The Carpet-Bag. CHAPTER 3. The Spouter-Inn. CHAPTER 4. The Counterpane. CHAPTER 5. Breakfast. CHAPTER 6. The Street. CHAPTER 7. The Chapel. CHAPTER 8. The Pulpit. CHAPTER 9. The Sermon. CHAPTER 10. A Bosom Friend. CHAPTER 11. Nightgown. CHAPTER 12. Biographical. CHAPTER 13. Wheelbarrow. CHAPTER 14...