



Gutenberg Text Cleaner

Clean and analyze text from Project Gutenberg books

Project Gutenberg URL:

<https://www.gutenberg.org/files/1342/1342-0.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

702,482

Characters

127,518

Words

7,356

Sentences

4.5

Avg Word Length

17.4

Avg Sentence Length



Top 10 Most Common Words

1. the (4,646)
2. to (4,289)
3. of (3,830)
4. and (3,710)
5. her (2,251)
6. i (2,078)
7. a (2,019)
8. in (1,974)
9. was (1,870)
10. she (1,725)



Summary (First 3 sentences)

illustration george allen publisher 156 charing cross road london ruskin house illustration
_reading jane's letters. _ _chap 34. _ pride.



Cleaned Text (First 500 characters)

[Illustration: GEORGE ALLEN PUBLISHER 156 CHARING CROSS ROAD LONDON
RUSKIN HOUSE] [Illustration: _Reading Jane's Letters._ _Chap 34._] PRIDE. and
PREJUDICE by Jane Austen, with a Preface by George Saintsbury and Illustrations by Hugh
Thomson [Illustration: 1894] Ruskin 156. Charing House. Cross Road. London George
Allen. CHISWICK PRESS:--CHARLES WHITTINGHAM AND CO. TOOKS COURT,
CHANCERY LANE, LONDON. ...