Package ‘janeaustenr’

October 13, 2022

Title Jane Austen's Complete Novels
Version 1.0.0
Description Full texts for Jane Austen's 6 completed novels, ready for
text analysis. These novels are "Sense and Sensibility", "Pride and
Prejudice", "Mansfield Park", "Emma", "Northanger Abbey", and
"Persuasion".
License MIT + file LICENSE
URL https://github.com/juliasilge/janeaustenr
BugReports https://github.com/juliasilge/janeaustenr/issues
Depends R (>= 3.5)
Suggests dplyr, testthat
Config/testthat/edition 3
Encoding UTF-8
LazyData true
RoxygenNote 7.2.1
NeedsCompilation no
Author Julia Silge [aut, cre] (<https://orcid.org/0000-0002-3671-836X>)
Maintainer Julia Silge <julia.silge@gmail.com>
Repository CRAN
Date/Publication 2022-08-26 22:32:06 UTC

R topics documented:

austen_books ......................................................... 2
emma ................................................................. 2
mansfieldpark ......................................................... 3
northangerabbey ..................................................... 3
persuasion ........................................................... 4
prideprejudice ....................................................... 4
sensesensibility .................................................... 5

Index 6
austen_books  
*Tidy data frame of Jane Austen’s 6 completed, published novels*

Description

Returns a tidy data frame of Jane Austen’s 6 completed, published novels with two columns: `text`, which contains the text of the novels divided into elements of up to about 70 characters each, and `book`, which contains the titles of the novels as a factor in order of publication.

Usage

```
austen_books()
```

Details

Users should be aware that there are some differences in usage between the novels as made available by Project Gutenberg. For example, "anything" vs. "any thing", "Mr" vs. "Mr.", and using underscores vs. all caps to indicate italics/emphasis.

Value

A data frame with two columns: `text` and `book`

Examples

```
library(dplyr)

austen_books() %>%
  group_by(book) %>%
  summarise(total_lines = n())
```

emma  
*The text of Jane Austen’s novel “Emma”*

Description

A dataset containing the text of Jane Austen’s 1815 novel “Emma”. The UTF-8 plain text was sourced from Project Gutenberg and is divided into elements of up to about 70 characters each. (Some elements are blank.)

Usage

```
emma
```
**mansfieldpark**

**Format**
A character vector with 15297 elements

**Source**
http://www.gutenberg.org/ebooks/158

---

**Description**
A dataset containing the text of Jane Austen’s 1814 novel "Mansfield Park". The UTF-8 plain text was sourced from Project Gutenberg and is divided into elements of up to about 70 characters each. (Some elements are blank.)

**Usage**
mansfieldpark

---

**northangerabbey**

**Format**
A character vector with 14768 elements

**Source**
http://www.gutenberg.org/ebooks/141

---

**Description**
A dataset containing the text of Jane Austen’s novel "Northanger Abbey", published posthumously in 1818. The UTF-8 plain text was sourced from Project Gutenberg and is divided into elements of up to about 70 characters each. (Some elements are blank.)

**Usage**
northangerabbey

---

**Format**
A character vector with 7840 elements

**Source**
http://www.gutenberg.org/ebooks/121
persistence

The text of Jane Austen's novel "Persuasion"

Description
A dataset containing the text of Jane Austen's novel "Persuasion", published posthumously in 1818. The UTF-8 plain text was sourced from Project Gutenberg and is divided into elements of up to about 70 characters each. (Some elements are blank.)

Usage
persuasion

Format
A character vector with 8328 elements

Source
http://www.gutenberg.org/ebooks/105

persistence

The text of Jane Austen's novel "Pride and Prejudice"

Description
A dataset containing the text of Jane Austen's 1813 novel "Pride and Prejudice". The UTF-8 plain text was sourced from Project Gutenberg and is divided into elements of up to about 70 characters each. (Some elements are blank.)

Usage
prideprejudice

Format
A character vector with 12447 elements

Source
http://www.gutenberg.org/ebooks/1342
sensesensibility

sensesensibility      The text of Jane Austen's novel "Sense and Sensibility"

Description

A dataset containing the text of Jane Austen’s 1811 novel "Sense and Sensibility". The UTF-8 plain text was sourced from Project Gutenberg and is divided into elements of up to about 70 characters each. (Some elements are blank.)

Usage

sensesensibility

Format

A character vector with 12262 elements

Source

http://www.gutenberg.org/ebooks/161
Index

* datasets
  emma, 2
  mansfieldpark, 3
  northangerabbey, 3
  persuasion, 4
  prideandprejudice, 4
  senseandresensibility, 5

austen_books, 2
emma, 2
mansfieldpark, 3
northangerabbey, 3
persuasion, 4
prideandprejudice, 4
senseandresensibility, 5