

# Package ‘textfeatures’

March 6, 2018

**Type** Package

**Title** Extracts Features from Text

**Version** 0.1.1

**Description** A tool for extracting some generic features (e.g., number of words, line breaks, characters per word, URLs, lower case, upper case, commas, periods, exclamation points, etc.) from strings of text.

**License** MIT + file LICENSE

**URL** <https://github.com/mkearney/textfeatures>

**BugReports** <https://github.com/mkearney/textfeatures/issues>

**Depends** R (>= 3.1.0)

**Imports** dplyr, rlang, tibble

**Suggests** knitr, roxygen2, testthat

**Encoding** UTF-8

**LazyData** yes

**RoxygenNote** 6.0.1.9000

**NeedsCompilation** no

**Author** Michael Wayne Kearney [aut, cre]  
(<<https://orcid.org/0000-0002-0730-4694>>)

**Maintainer** Michael Wayne Kearney <[kearneymw@missouri.edu](mailto:kearneymw@missouri.edu)>

**Repository** CRAN

**Date/Publication** 2018-03-06 18:48:55 UTC

## R topics documented:

textfeatures . . . . . 2

**Index** . . . . . 4

---

textfeatures	<i>textfeatures</i>
--------------	---------------------

---

### Description

Extracts features from text vector.

### Usage

```
textfeatures(x)
```

### Arguments

**x** Input data. Should be character vector, data frame, or grouped data frame (grouped\_df) with character variable of interest named "text". If grouped\_df is provided, then features will be averaged across all observations for each group.

### Value

A tibble data frame with extracted features as columns.

### Examples

```
## the text of five of Trump's most retweeted tweets
trump_tweets <- c(
  "#FraudNewsCNN #FNN https://t.co/WYUnHjjUjg",
  "TODAY WE MAKE AMERICA GREAT AGAIN!",
  paste("Why would Kim Jong-un insult me by calling me \"old,\" when I would",
    "NEVER call him \"short and fat?\" Oh well, I try so hard to be his",
    "friend - and maybe someday that will happen!"),
  paste("Such a beautiful and important evening! The forgotten man and woman",
    "will never be forgotten again. We will all come together as never before"),
  paste("North Korean Leader Kim Jong Un just stated that the \"Nuclear",
    "Button is on his desk at all times.\" Will someone from his depleted and",
    "food starved regime please inform him that I too have a Nuclear Button,",
    "but it is a much bigger & more powerful one than his, and my Button",
    "works!")
)

## get the text features of a character vector
textfeatures(trump_tweets)

## data frame with a character vector named "text"
df <- data.frame(
  id = c(1, 2, 3),
  text = c("this is A!\t sEntence https://github.com about #rstats @github",
    "doh", "The following list:\n- one\n- two\n- three\nOkay!?!")
)
```

```
## get text features of a data frame with "text" variable  
textfeatures(df)
```

# Index

textfeatures, [2](#)