

Package ‘MediaNews’

November 26, 2020

Type Package

Title Media News Extraction for Text Analysis

Version 0.2.1

Date 2020-11-25

Author Vatsal Aima [aut, cre]

Maintainer Vatsal Aima <vaima75@hotmail.com>

Description Extract textual data from different media channels through its source based on users choice of keywords. These data can be used to perform text analysis to identify patterns in respective media reporting. The media channels used in this package are print media. The data (or news) used are publicly available to consumers.

License LGPL-3

Encoding UTF-8

LazyData true

OS_type windows

Depends R (>= 3.5.0)

Imports rvest (>= 0.3.5), xml2 (>= 1.2.2), lubridate (>= 1.7.4), stats (>= 3.6.1), utils (>= 3.6.1), stopwords (>= 1.0)

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2020-11-26 10:50:10 UTC

R topics documented:

ClearText	2
emoji_Data	3
TOI_News_Articles	4
TOI_News_Dataset	5
Index	7

 ClearText

Text Cleaning: Custom Method

Description

Cleans text and introduce custom stopwords to remove unwanted words from given data.

Usage

```
ClearText(Text, CustomList = c(""))
```

Arguments

Text	A String or Character vector, user-defined.
CustomList	A Character vector (Optional), user-defined vector to introduce stopwords ("english") in Text.

Value

Returns Character

Author(s)

Vatsal Aima, <vaima75@hotmail.com>

See Also

[TOI_News_Articles](#), [TOI_News_Dataset](#)

Examples

```
##### Methodology #####
##### For DataFrame #####
#### Creates Dataset based on keyword

NewsData = TOI_News_Articles("Goibibo")

## Identify any potential factor columns
vc = sapply(NewsData, is.factor)

## Convert factors to characters
NewsData[vc] = lapply(NewsData[vc], as.character)

## Clean text on specific character columns
for (i in 1:nrow(NewsData)) NewsData$News[i] = ClearText(NewsData$News[i])

##### For Character Variable #### Ex2 #####

para = "Moreover, the text data we get is noisy. But, if we can learn some
```

methods useful to extract important features from the noisy data, wouldn't scandal that be amazing ? In this tutorial, you'll learn how to extract regular expressions from scratch. At first, you might find these expressions confusing, or complicated, but after a while you'll find them quite useful. In addition, we'll also learn about string manipulation functions in R. This combination of regular expressions and string manipulation functions will prepare you for text mining."

```
clearpara = ClearText(para,
                      CustomList = c("scooby-doo",
                                      "cartoon-network",
                                      "detective",
                                      "scandal"))

##### For List #####

paraList = list(para, 1213, factor('aasd;kasdioasd'))
paraList = lapply(paraList, as.character)
for (x in 1:length(paraList)) paraList[[x]] = ClearText(paraList[[x]])
```

 emoji_Data

Emoji Data

Description

Emoji Data

Usage

emoji_Data

Format

Dataframe with columns:

C1,C2,C3,C4,C5,C6,C7,C8 Uni-code in text.

Browser Code applicable on Web Browser.

Appl Code applicable on Apple.

Goog Code applicable on Google.

FB Code applicable on FaceBook.

Wind Code applicable on Windows Devices.

Twtr Code applicable on Twittwer.

Sams Code applicable on Samsung.

Gmail Code applicable on Gmail.

Joy,SB,DCM,KDDI Code applicable on other Platforms.

Description Code description

Source

<<https://unicode.org/emoji/charts/full-emoji-list.html>>

TOI_News_Articles	<i>Extract Media News</i>
-------------------	---------------------------

Description

Creates a DataFrame or Write files to disk by extracting text data from source based on user's keywords.

Usage

```
TOI_News_Articles(
  keywords,
  AsDataFrame = TRUE,
  start_date = NULL,
  end_date = NULL
)
```

Arguments

keywords	A String, user-defined.
AsDataFrame	Boolean Value, to determine whether the outcome should be a Dataframe or files written to disk. if set to <i>FALSE</i> then returns the files will be written to disk at stated working directory (default <i>TRUE</i>).
start_date	Date (Character) Value, provide the starting date FROM where the data should be extracted. NOTE : only provide start_date when IsDate is set <i>TRUE</i> .
end_date	Date (Character) Value, provide the ending date TO where the data should be extracted. NOTE : only provide end_date when IsDate is set <i>TRUE</i> .

Value

Returns DataFrame or write files to the disk based on keywords

Author(s)

Vatsal Aima, <vaima75@hotmail.com>

See Also

[TOI_News_Dataset](#)

Examples

```
#### Creates Dataset by filtering 31 days from current date

NewsDataset1 = TOI_News_Articles(keywords = "Politics In US",
start_date = Sys.Date()- 31,
end_date = Sys.Date())

# Creates Dataset by custom filtering through dates
NewsDataset2 = TOI_News_Articles(keywords = "BaseBall",
start_date = "2019-09-20",
end_date = "2019-10-20")

# Write files to disk
TOI_News_Articles(keywords = "AirLines", IsDataFrame = FALSE)
```

TOI_News_Dataset	<i>Creates Interim Dataset</i>
------------------	--------------------------------

Description

Creates an interim news dataset based on user-defined keywords for all possible links extracted from respective source.

Usage

```
TOI_News_Dataset(keywords)

TOI_News_Links(keywords)
```

Arguments

keywords A String, user-defined.

Value

Returns DataFrame based on keywords

Functions

- TOI_News_Links: Extracts Source Links

Author(s)

Vatsal Aima, <vaima75@hotmail.com>

See Also

[TOI_News_Articles](#)

Examples

```
#### Creates Dataset based on keyword  
NewsData = TOI_News_Dataset("Goibibo")
```

Index

* datasets

emoji_Data, 3

ClearText, 2

emoji_Data, 3

TOI_News_Articles, 2, 4, 5

TOI_News_Dataset, 2, 4, 5

TOI_News_Links (TOI_News_Dataset), 5