Package ‘MediaNews’

February 9, 2020

Type Package
Title Media News Extraction for Text Analysis
Version 0.1.0
Date 2020-01-23
Author Vatsal Aima
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 from India. The data (or news) used are publicly available to consumers.
License LGPL-3
Encoding UTF-8
LazyData true
OS_type windows
Imports lubridate (>= 1.7.4), rvest (>= 0.3.5), stats (>= 3.6.1), stopwords (>= 1.0), svMisc (>= 1.1.0), xml2 (>= 1.2.2)
RoxygenNote 7.0.1
NeedsCompilation no
Repository CRAN
Date/Publication 2020-02-09 17:10:12 UTC

R topics documented:

ClearText ......................................................... 2
TOI_News_Articles ............................................ 3
TOI_News_Dataset ............................................. 4
Index ......................................................... 6
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 saadc@ruby.com learn #world all about regular expressions from scratch. At first, #world detective you might find these confusing, or complicated, but after https://anaconda.com/anaconda-enters-new-chapter/ expressions tricky, scooby-doo doing practical hands-on exercises (done below) you should feel bcc: @MikeQuindazzi quite comfortable with it. In addition, we'll also cartoon-network learn about string manipulation functions in R. This formidable combination of #DL #4IR #Robots #ArtificialIntelligence string manipulation functions and regular expressions 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]])
TOI_News_Dataset

**IsDate**  
Boolean Value, to determine whether to perform filtering operations based on dates. If set to **TRUE** then Date operations will be applied (default **FALSE**).

**start_date**  
Date Value, provide the starting date FROM where the data should be extracted.  
**NOTE**: only provide `start_date` when `IsDate` is set **TRUE**.

**end_date**  
Date Value, provide the ending date TO where the data should be extracted.  
**NOTE**: only provide `end_date` when `IsDate` is set **TRUE**.

**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",  
IsDate = TRUE,  
start_date = Sys.Date()- 31,  
end_date = Sys.Date())

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

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


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

ClearText, 2

TOI_News_Articles, 2, 3, 5
TOI_News_Dataset, 2, 4, 4
TOI_News_Links (TOI_News_Dataset), 4