

Package ‘datacollector’

June 8, 2017

Title Data Collector

Version 0.0.0.9000

Description this package collects data from web and social media

Depends R (>= 3.2.3)

License What license is it under?

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

R topics documented:

collectContentByKey	2
collectDataFromFacebook	2
collectDataFromGooglePlus	3
collectDataFromPage	3
collectDataFromTwitter	4
collectDataFromWebSite	4
collectEmails	5
collectMedias	5
collectPhones	6
collectSocialLinks	6
collectTopics	7
collectUsers	7
countData	8
downloadDocuments	8
downloadImages	9
downloadVideos	9
saveByKeyWord	10
selectData	10

Index	11
--------------	-----------

collectContentByKey	<i>Collect by keyword links from a web page this function extract webpage content by keyword, and return the blocs containing that keyword</i>
---------------------	--

Description

Collect by keyword links from a web page this function extract webpage content by keyword, and return the blocs containing that keyword

Usage

```
collectContentByKey(url, key)
```

Arguments

url	url of the target web page
key	keyword to serch for

Value

list of bloc content containing that keyword

collectDataFromFacebook	<i>Get facebook graph edge of a specific node</i>
-------------------------	---

Description

this function implements facebook graph API to collect public data using nodes and edges

Usage

```
collectDataFromFacebook(path, node, edge, user = FALSE)
```

Arguments

path	Path to save the data
node	Facebook graph node id, eg(Page id, User id, Post id)...
edge	Graph edge name, eg(comment,posts,feed,likes) ...

`collectDataFromGooglePlus`*Get google plus data*

Description

this function fetch activities from google plus by submitting a filtered search in save it to a json file in the directory path provided

Usage

```
collectDataFromGooglePlus(path, query = "", lang = "en")
```

Arguments

path	path to save the data
query	text query to search for activities
lang	language code of the activities ...

Value

void

`collectDataFromPage` *Collect data from a web page*

Description

This function scraps data from a web page and save it into an excel sheet Data such : Email, phones numbers, medias, links, social media links, and html tables

Usage

```
collectDataFromPage(path, url)
```

Arguments

path	Path to the excel file
url	url of the target web page

Value

void

 collectDataFromTwitter

Get Twitter data

Description

this function fetch tweets from Twitter by submitting a filtered search in save it to a json file in the directory path provided

Usage

```
collectDataFromTwitter(path, lang = "en", exact = "", allword = "",
  hashtags = "", noneWords = "", oneOf = "", accounts = "",
  attitude = TRUE, question = FALSE)
```

Arguments

path	path to save the data
lang	language code of the tweets ...
attitude	postive attitude : True, negative attitude : False
question	question exist in tweet : True else false.
allwords	filter tweets by all words in the string seperated by space
exactphrase	filter tweets by exact string
hashtag	filter tweets by hashtags in this string seperated by space ...
oneof	filter tweets by one of the words in this string seperated by space ...
noneof	filter tweets by excluding all the words in this string seperated by space ...

Value

void

 collectDataFromWebSite

Collect data from the whole website

Description

this function crawls the whole website for the requested data and save each scraped page into an excel sheet, and download the availble media in a sepcific directory

Usage

```
collectDataFromWebSite(url, excelpath, downloadpath)
```

Arguments

url	url of the target web page
excelpath	Path to the excel file
downloadpath	path to download directory

Value

void

collectEmails	<i>Collect emails from a web page this function extract all emails from a web page</i>
---------------	--

Description

Collect emails from a web page this function extract all emails from a web page

Usage

```
collectEmails(url)
```

Arguments

url	url of the target web page
-----	----------------------------

Value

list of emails

collectMedias	<i>Collect media links from a web page this function extract all medias such images, videos, audio, documents.. from a web page</i>
---------------	---

Description

Collect media links from a web page this function extract all medias such images, videos, audio, documents.. from a web page

Usage

```
collectMedias(url, media)
```

Arguments

url	url of the target web page
media	media type could be : image,video,document,audio

Value

list of medias

collectPhones	<i>Collect phone numbers from a web page this function extract all phone numbers from a web page</i>
---------------	--

Description

Collect phone numbers from a web page this function extract all phone numbers from a web page

Usage

```
collectPhones(url)
```

Arguments

url	url of the target web page
-----	----------------------------

Value

list of phone numbers

collectSocialLinks	<i>Collect SocialLinks this function extract all social medias links from a web page</i>
--------------------	--

Description

Collect SocialLinks this function extract all social medias links from a web page

Usage

```
collectSocialLinks(url)
```

Arguments

url	url of the target web page
-----	----------------------------

Value

list of social media links

collectTopics	<i>Collect topics from Social medias such twitter and google+</i>
---------------	---

Description

this function collect data about topics published in twitter (Tweets) and Google+ (activities)

Usage

```
collectTopics(path, query)
```

Arguments

path	Path to storage file directory
query	Search query

Value

void

collectUsers	<i>Collect data about users of facebook and google+</i>
--------------	---

Description

this function collect and fetch data about facebook and google+ users

Usage

```
collectUsers(path, query)
```

Arguments

path	Path to storage file directory
query	Search query

Value

void

countData	<i>Count data rows in a json file</i>
-----------	---------------------------------------

Description

This function excutes an SQL select count(*) query over a json file and return data row count'

Usage

```
countData(path, item = "*")
```

Arguments

path	Path to the json file
item	columns to count

Value

number of rows in the target file

downloadDocuments	<i>Download documents from a target page</i>
-------------------	--

Description

This functions extract all document files and download it

Usage

```
downloadDocuments(path, url)
```

Arguments

path	Path to the download directory
url	url of the target web page

Value

void

downloadImages	<i>Download images from a target page</i>
----------------	---

Description

This functions exctract all images files and download it
This functions exctract all images files and download it

Usage

downloadImages(path, url)

downloadImages(path, url)

Arguments

path	Path to the download directory
url	url of the target web page
path	Path to the download directory
url	url of the target web page

Value

void
void

downloadVideos	<i>Download videos from a target page</i>
----------------	---

Description

This functions exctract all videos files and download it

Usage

downloadVideos(path, url)

Arguments

path	Path to the download directory
url	url of the target web page

Value

void

saveByKeyWord	<i>Search content by key</i>
---------------	------------------------------

Description

return html blocs by keyword search and save into excel sheet

Usage

```
saveByKeyWord(path, url, key)
```

Arguments

path	Path to the excel file
url	url of the target web page
key	key to search for

Value

void

selectData	<i>Filter data from json file</i>
------------	-----------------------------------

Description

This function excutes an SQL select query over a json file and return filtered data'

Usage

```
selectData(path, item = "id", file = "")
```

Arguments

path	Path to the json file
item	column to select

Value

array of result query

Index

`collectContentByKey`, [2](#)
`collectDataFromFacebook`, [2](#)
`collectDataFromGooglePlus`, [3](#)
`collectDataFromPage`, [3](#)
`collectDataFromTwitter`, [4](#)
`collectDataFromWebSite`, [4](#)
`collectEmails`, [5](#)
`collectMedias`, [5](#)
`collectPhones`, [6](#)
`collectSocialLinks`, [6](#)
`collectTopics`, [7](#)
`collectUsers`, [7](#)
`countData`, [8](#)

`downloadDocuments`, [8](#)
`downloadImages`, [9](#)
`downloadVideos`, [9](#)

`saveByKeyWord`, [10](#)
`selectData`, [10](#)