

Package ‘rodham’

January 19, 2017

Type Package

Title Fetch Hillary Rodham Clinton's Emails

Version 0.0.3

Maintainer John Coene <jcoenep@gmail.com>

Description Fetch and process Hillary Rodham Clinton's ``personal" emails.

License MIT + file LICENSE

LazyData TRUE

Depends R (>= 2.10)

Imports jsonlite,
splitstackshape,
plyr,
methods,
utils

RoxygenNote 5.0.1

Suggests testthat,
covr,
knitr,
rmarkdown,
igraph,
lintr

VignetteBuilder knitr

R topics documented:

edges_emails	2
emails	2
get_emails	3
get_xpdf	5
search_emails	5
Index	7

edges_emails	<i>Network the treacherous</i>
--------------	--------------------------------

Description

Builds edge list from emails using from and to, edge source and target respectively.

Usage

```
edges_emails(emails = emails, ...)
```

Arguments

emails	Data frame of emails as returned by search_emails , defaults to emails (see emails)
...	any additional column to keep as meta-data

Author(s)

John Coene <jcoenep@gmail.com>

See Also

[search_emails](#)

Examples

```
## Not run:
emails <- search_emails()

edges <- edges_emails(emails)

## End(Not run)
```

emails	<i>Hillary Rodham Clinton emails</i>
--------	--------------------------------------

Description

A dataset containing 29444 emails from/to Hillary Rodham Clinton sent/received between 2009-08-14 and 2014-08-13.

Usage

```
emails
```

Format

A data frame with 29444 rows and 9 variables:

docID Primary key
docDate Date when document was sent or received
to Who the emails was sent to
from Who the email is received from
originalTo From whom the email originally comes from
originalFrom To whom the email was originally sent to
subject Subject of the email
interesting Rating, relevancy of email
not_interesting Rating, irrelevancy of email

Source

<http://graphics.wsj.com/hillary-clinton-email-documents/api/search.php?subject=&to=&from=&start=&end=&sort=docDate&order=desc&docid=&limit=27159&offset=0>

get_emails	<i>Get emails and its contents</i>
------------	------------------------------------

Description

Get the content of Hillary Rodham Clinton's emails by release.

Usage

```
get_emails(release, save.dir = getwd(), extractor)
```

Arguments

release	Name of the batch of release of emails; see details.
save.dir	Directory where to save the extracted text defaults to getwd()
extractor	Full path to pdf extractor (text to pdf), see details.

Details

Below are the valid values for release; follows the **WSJ** naming convention.

- Benghazi
- June
- July
- August
- September
- October
- November
- January 7

- January 29
- February 19
- february 29
- December
- Non-disclosure

The extractor argument is the full path to your pdftotext.exe extractor; visit [xpdf](#) to download or try [get_xpdf](#) which attempts to download and unzip the text to pdf extractor. See examples.

Value

Fetches email zip file from the WSJ and extract text files in `save.dir`, returns full path to directory that contains parsed txt files.

Author(s)

John Coene <jcoenep@gmail.com>

See Also

[get_xpdf](#)

Examples

```
## Not run:
# get xpdf extractor
ext <- get_xpdf()

# create
dir.create("./emails")

# get emails released in august
emails_aug <- get_emails(release = "August", save.dir = "./emails",
                        extractor = ext)

# use manually downloaded extractor
ext <- "C:/xpdfbin-win-3.04/bin64/pdftotext.exe"

# get emails related to Benghazi released in December
emails_bengh <- get_emails(release = "Benghazi", extractor = ext,
                          save.dir = "./emails")

files <- list.files(emails_bengh)
content <- lapply(1:length(files), function(x){
  readLines(paste0(emails_bengh, "/", files[[x]]))
})

## End(Not run)
```

get_xpdf	<i>Get pdf to text extractor (xpdf)</i>
----------	---

Description

Downloads and extracts pdf to text extractor, see details.

Usage

```
get_xpdf(dest = getwd())
```

Arguments

dest	Destination folder defaults to getwd()
------	--

Details

If function fails you can download the extractor manually from <http://www.foolabs.com/xpdf/> (tested on Windows only)

Value

Returns full path to pdftotext executable

Author(s)

John Coene <jcoenep@gmail.com>

See Also

[get_emails](#)

search_emails	<i>Search Rodham's emails</i>
---------------	-------------------------------

Description

Search Hillary Rodham Clinton's *personal* emails.

Usage

```
search_emails(subject = NULL, to = NULL, from = NULL, start = NULL,  
end = NULL, internal = TRUE)
```

Arguments

subject	Filter by subject, defaults to NULL(no filter). If internal = TRUE then matches pattern, if internal = FALSE then looks for exact match.
to	Filter by Receiver, defaults to NULL(no filter).
from	Filter by Sender, defaults to NULL(no filter).
start	Filter by date range, defaults to NULL(no filter).
end	Filter by date range, defaults to NULL(no filter).
internal	if TRUE (default) searches the internal data set (see data(emails)), if FALSE fetches the data through the Wall Street Journal API. data(emails) is equivalent to internal TRUE

Details

There are a total of 29444 emails ranging from 2009-08-14 to 2014-08-13, please consider leaving internal to TRUE to not hammer the Wall Street Journal's API. internal = TRUE is equivalent to [emails](#).

Author(s)

John Coene <jcoenep@gmail.com>

Examples

```
## Not run:
emails <- search_emails()

# only emails on cuba
emails <- search_emails(subject = "Cuba")

# only emails from Jake Sullivan since 2014
j_s <- search_emails(from = "Jake Sullivan", start = as.Date("2014-01-01"))

## End(Not run)
```

Index

*Topic **datasets**

emails, [2](#)

edges_emails, [2](#)

emails, [2](#), [2](#), [6](#)

get_emails, [3](#), [5](#)

get_xpdf, [4](#), [5](#)

search_emails, [2](#), [5](#)