

# Package ‘rodham’

May 4, 2016

**Type** Package

**Title** Fetch Hillary Rodham Clinton's Emails

**Version** 0.1.1

**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

**VignetteBuilder** knitr

## R topics documented:

edges_emails . . . . .	2
emails . . . . .	2
get_emails . . . . .	3
get_extractor . . . . .	4
rodham . . . . .	5
search_emails . . . . .	6
<b>Index</b>	<b>8</b>

---

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 <a href="#">search_emails</a> , defaults to emails (see <a href="#">emails</a> )
...	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\\_extractor](#) which attempts to download and extract the text to pdf extractor.

### 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\\_extractor](#)

### Examples

```
## Not run:
ext <- get_extractor()
# get emails released in august
emails_cuba <- get_emails(release = "August", save.dir = "C:/",
  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)
files <- list.files(emails_bengh)
content <- lapply(1:length(files), function(x){
  readLines(paste0(emails_bengh, "/", files[[x]]))
})

## End(Not run)
```

---

get\_extractor

*Get pdf to text parser (xpdf)*

---

### Description

Downloads and extracts pdf to text extractor, see details.

### Usage

```
get_extractor(dest = "C:/")
```

**Arguments**

dest                      Destination folder defaults to C: /

**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.exe

**Author(s)**

John Coene <jcoenep@gmail.com>

**See Also**

[get\\_emails](#)

---

rodham

*rodham: Hillary Rodham Clinton emails*

---

**Description**

rodham: Hillary Rodham Clinton emails

**Dataset**

- emailsdataset of 29444 emails

**Functions**

- [search\\_emails](#)
- [edges\\_emails](#)
- [get\\_emails](#)
- [get\\_extractor](#)

**Examples**

```
## Not run:
# get emails from internal data set
data("emails")

# build graph
edges <- edges_emails(emails)
g <- igraph::graph.data.frame(edges)
plot(g)

# plot communities
cm <- igraph::walktrap.community(g)
plot(cm, g)
```

```
# get extractor to extract content from emails
ext <- get_extractor()

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

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

## End(Not run)
```

---

search\_emails

Search Rodham's emails

---

## 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 <code>internal = TRUE</code> then matches pattern, if <code>internal = FALSE</code> 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 <code>data(emails)</code> ), if FALSE fetches the data through the Wall Street Journal API. <code>data(emails)</code> is equivalent to <code>internal TRUE</code>

## 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](#), [5](#)

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

get\_emails, [3](#), [5](#)

get\_extractor, [4](#), [4](#), [5](#)

rodham, [5](#)

rodham-package (rodham), [5](#)

search\_emails, [2](#), [5](#), [6](#)