

Package ‘poltextr’

October 24, 2020

Type Package

Title Easy Access Political Texts in R

Date 2020-07-16

Version 0.1.0

URL <https://github.com/lin-jennifer/poltextr>

Description This package contains corpuses of political texts that can be called into R with ease, along with helper scrape functions that can scrape texts off select websites based on the site that the function is designed. In the package documentation, anything that starts with ``Scraper" denotes a function, all else are political texts that are stored as ``.RData" files.
New texts will be added.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

VignetteBuilder knitr

Suggests knitr,
rmarkdown,
testthat

Imports rvest,
magrittr,
xml2

Depends R (>= 2.10)

R topics documented:

DemocratPlatforms	2
dnc16	2
dnc16_meta	3
dnc20	4
ExportingData	5
rnc16	5
rnc16_meta	6
rnc20	6
ScraperDebates	7
ScraperFunctions	8

Index**10**

DemocratPlatforms *Democrat Party Platforms*

Description

Party Platforms for the Democrat party from 2000 to 2016

Usage

DemocratPlatforms

Format

A data frame

X Line ID

head Heading of the Section

Year Year of the Election

Party Democrat

Date Date of the platform, as listed on UCSB site

Candidate Nominee for the election

topic General Topic of the section – Economy, Society, Foreign Affair, Government, Introduction

text Actual text minus headings

Source

<https://www.presidency.ucsb.edu/documents/presidential-documents-archive-guidebook/national-political-party-platforms>

Examples

Not run: DemocratPlatforms

dnc16

Democrat National Convention 2016

Description

Captures the speeches given at the 2016 Democrat National Convention

Usage

dnc16

Format

a Large Corpus with 155 texts

Source

Varies. <https://www.politico.com/story/2016/07/dnc-2016-schedule-of-events-and-speakers-225617>

Examples

```
## Not run: dnc16
```

dnc16_meta	<i>Metadata File for the 2016 DNC</i>
------------	---------------------------------------

Description

Metadata File for the 2016 DNC

Usage

```
dnc16_meta
```

Format

a data frame with 155 rows and 7 variables

Last Last Name of Speaker

First First Name of Speaker

Filename Corresponding File name to [dnc16](#)

Day Day in the convention the speech was given

Type speech or benediction

Source Provider of the document

Convention Type of convention – DNC

Examples

```
dnc16_meta
```

dnc20

Democrat National Convention 2020

Description

Captures the speeches given at the 2020 Democrat National Convention

Usage

dnc20

Format

a Data Frame with 1693 observations

Speaker Last Name of Speaker

Day Day in the convention the speech was given

Date Date of the Speech

Convention Type of convention – DNC

text Text of speech

Details

Due to the online nature of the convention, the speeches included in this data set are those included in the 9-11 PM EST prime time livestream

Links to the original speech transcripts from Rev are as follows:

Night 1: <https://www.rev.com/blog/transcripts/democratic-national-convention-dnc-night-1-transcript>

Night 2: <https://www.rev.com/blog/transcripts/democratic-national-convention-dnc-2020-night-2-transcript>

Night 3: <https://www.rev.com/blog/transcripts/democratic-national-convention-dnc-night-3-transcript>

Night 4: <https://www.rev.com/blog/transcripts/2020-democratic-national-convention-dnc-night-4-transcript>

Source

Varies. Nightly Transcripts from Rev

Examples

```
## Not run: dnc20
```

ExportingData

Export Functions for General Texts

Description

Export to plain text functions for plain text with settings pre-configured globally

Usage

```
exportDoc(df, dest)
```

Arguments

df	Data frame to export
dest	Location and file name to save the data

rnc16

Republican National Convention 2016

Description

Captures the speeches given at the 2016 Republican National Convention

Usage

```
rnc16
```

Format

a Large Corpus with 65 texts

Source

Varies. <https://www.politico.com/story/2016/07/rnc-2016-schedule-of-events-and-speakers-225704>

Examples

```
## Not run: rnc16
```

rnc16_meta	<i>Metadata File for the 2016 RNC</i>
------------	---------------------------------------

Description

Metadata File for the 2016 RNC

Usage

rnc16_meta

Format

a data frame with 65 rows and 7 variables

Last Last Name of Speaker

First First Name of Speaker

Filename Corresponding File name to [rnc16](#)

Day Day in the convention the speech was given

Type speech or benediction

Source Provider of the document

Convention Type of convention – RNC

Examples

rnc16_meta

rnc20	<i>Republican National Convention 2020</i>
-------	--

Description

Captures the speeches given at the 2020 Democrat National Convention

Usage

rnc20

Format

a Data Frame with 1056 observations

Speaker Last Name of Speaker

Day Day in the convention the speech was given

Date Date of the Speech

Convention Type of convention – DNC

text Text of speech

Details

Due to the online nature of the convention, the speeches included in this data set are those included in the 9-11 PM EST prime time livestream

Links to the original speech transcripts from Rev are as follows:

Night 1: <https://www.rev.com/blog/transcripts/2020-republican-national-convention-rnc-night-1-tra>

Night 2: <https://www.rev.com/blog/transcripts/2020-republican-national-convention-rnc-night-2-tra>

Night 3: <https://www.rev.com/blog/transcripts/2020-republican-national-convention-rnc-night-3-tra>

Night 4: <https://www.rev.com/blog/transcripts/2020-republican-national-convention-rnc-night-4-tra>

Source

Varies. Nightly Transcripts from Rev

Examples

```
## Not run: rnc20
```

ScraperDebates	<i>Scraper Functions for Debates</i>
----------------	--------------------------------------

Description

Functions to scrape Debates from <https://www.presidency.ucsb.edu> or <https://www.debates.org/>

Usage

```
scrapeDebUCSB(url)
```

```
scrapeDebOrg(url)
```

Arguments

`url` A URL in quotes

Details

Use `scrapeDebUCSB` for texts from <https://www.presidency.ucsb.edu> and use `scrapeDebOrg` for texts from <https://www.debates.org/voter-education/debate-transcripts/>. Since the websites are set up a bit differently, the functions are adjusted to account for the difference CSS labels of the main text. Both return a text string which can then be manipulated as desired.

Value

text string in the format `SPEAKERA: text SPEAKERB: text` with line breaks between the speakers

Examples

```
# Debates from https://www.presidency.ucsb.edu
url <- "https://www.presidency.ucsb.edu/documents/presidential-debate-washington-university-st-louis-missouri"
Debate <- scrapeDebUCSB(url) # Scrape 2016 Presidential Debate

# Debates from Debates.org
url <- "https://www.debates.org/voter-education/debate-transcripts/september-26-2016-debate-transcript/"
Debate <- scrapeDebOrg(url)
```

ScraperFunctions

Scraper Functions for General Texts

Description

Functions to scrape political texts from <https://www.presidency.ucsb.edu>

Usage

```
scrapePlat(url, content)

scrapeInaug(n, content)

scrapeSOTU(url, content)

scrapeUCSB(n, content)
```

Arguments

url	A URL in quotes
content	CSS content that you want to scrape from site
n	number, from end of URL for Inaugural Addresses ONLY

Details

scrapePlat() is designed to scrape party platforms but has the potential to work for a general text on the <https://www.presidency.ucsb.edu> site. This returns a text string, which can then be converted to a data frame and manipulated as desired.

scrapeInaug() is like scrapePlat() but is designed for Inaugural Addresses. Each address URL for the UCSB presidency page is designed with "https://www.presidency.ucsb.edu/documents/inaugural-address-2012-01-20" such that the last "n" is the unique identifier for a given address. Therefore, scrapeInaug(15, '.field-docs-content') will give you President Obama's 2012 address. The "n" value can be located by reading the last number on the URL

Examples

```
# Party Platforms
url <- "https://www.presidency.ucsb.edu/documents/2016-democratic-party-platform"
DemPP2016 <- scrapePlat(url, '.field-docs-content')

# Inaugural Address
```



```
# Source: https://www.presidency.ucsb.edu/documents/inaugural-address-15
# n = 15
Obama2012 <- scrapeInaug(15, '.field-docs-content')
```

```
# State of the Union
```

```
url <- "https://www.presidency.ucsb.edu/documents/address-before-joint-session-the-congress-the-state-the-u
ObamaSOTU16 <- scrapeSOTU(url, '.field-docs-content')
```

```
# Any Text on UCSB
```

```
# https://www.presidency.ucsb.edu/node/321069
```

```
# n = 321069
```

```
ObamaFarewell <- scrapeUCSB(321069, '.field-docs-content')
```

Index

* datasets

- DemocratPlatforms, [2](#)
- dnc16, [2](#)
- dnc16_meta, [3](#)
- dnc20, [4](#)
- rnc16, [5](#)
- rnc16_meta, [6](#)
- rnc20, [6](#)

- DemocratPlatforms, [2](#)
- dnc16, [2](#), [3](#)
- dnc16_meta, [3](#)
- dnc20, [4](#)

- exportDoc (ExportingData), [5](#)
- ExportingData, [5](#)

- rnc16, [5](#), [6](#)
- rnc16_meta, [6](#)
- rnc20, [6](#)

- scrapeDebOrg (ScraperDebates), [7](#)
- scrapeDebUCSB (ScraperDebates), [7](#)
- scrapeInaug (ScraperFunctions), [8](#)
- scrapePlat (ScraperFunctions), [8](#)
- ScraperDebates, [7](#)
- ScraperFunctions, [8](#)
- scrapeSOTU (ScraperFunctions), [8](#)
- scrapeUCSB (ScraperFunctions), [8](#)