Package ‘usdoj’

March 25, 2023

Title For Accessing U.S. Department of Justice (DOJ) Open Data

Version 1.1.0

Maintainer Steph Buongiorno <steph.buon@gmail.com>

Depends R (>= 3.4.4)

Imports anytime, dplyr, httr, jsonlite, stringr, tibble

URL https://github.com/ropengov/usdoj

BugReports https://github.com/ropengov/usdoj/issues

Description Fetch data from the <https://www.justice.gov/developer/api-documentation/api_v1> API such as press releases, blog entries, and speeches. Optional parameters allow users to specify the number of results starting from the earliest or latest entries, and whether these results contain keywords. Data is cleaned for analysis and returned in a dataframe.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.2

NeedsCompilation no

Author Steph Buongiorno [aut, cre] (<https://orcid.org/0000-0002-6965-0787>)

Repository CRAN

Date/Publication 2023-03-25 21:30:02 UTC

R topics documented:

doj_blog_entries ................................................. 2

doj_press_releases ............................................. 3

doj_speeches .................................................... 4

Index 5
doj_blog_entries  
Fetch U.S. DOJ Blog Entries

Description

For accessing U.S. Department of Justice (DOJ) open data blog entries.

Usage

doj_blog_entries(
    n_results = 50,
    search_direction = "DESC",
    keyword = NULL,
    clean = TRUE
)

Arguments

n_results The number of results to return. By default, 50 results are returned.
search_direction Set to ASC or DESC to fetch data starting from the most or least recent entries.
                 By default, data is sorted in descending order (the most recent entries).
keyword Only return results containing a keyword. Keyword can be a single word, a list
          or words, or a regex separated by the pipe delimiter
clean Set to FALSE to return data without additional cleaning. By default HTML tags,
        white space, line breaks, and other messy characters are removed. Epoch time
        is converted to year-month-day format.

Value

Dataframe containing API pull of U.S. DOJ blog entries.

Examples

# Return a dataframe with 50 of the latest records. Data is cleaned.
latest_records <- doj_blog_entries()
head(latest_records)

# Return a dataframe with 15 of the earliest records. Data is not cleaned.
earliest_records <- doj_blog_entries(n_results = 15, search_direction = "ASC", clean = FALSE)
head(earliest_records)
doj_press_releases

Fetch U.S. DOJ Press Releases

Description

For accessing U.S. Department of Justice (DOJ) open data press releases.

Usage

doj_press_releases(
  n_results = 50,
  search_direction = "DESC",
  keyword = NULL,
  clean = TRUE
)

Arguments

n_results The number of results to return. By default, 50 results are returned.

search_direction

  Set to ASC or DESC to fetch data starting from the most or least recent entries.
  By default, data is sorted in descending order (the most recent entries).

keyword

  Only return results containing a keyword. Keyword can be a single word, a list
  or words, or a regex separated by the pipe delimiter

clean

  Set to FALSE to return data without additional cleaning. By default HTML tags,
  white space, line breaks, and other messy characters are removed. Epoch time
  is converted to year-month-day format.

Value

Dataframe containing API pull of U.S. DOJ press releases.

Examples

# Return a dataframe with 50 of the latest records. Data is cleaned.
latest_records <- doj_press_releases()
head(latest_records)

# Return a dataframe with 15 of the earliest records. Data is not cleaned.
earliest_records <- doj_press_releases(n_results = 15, search_direction = "ASC", clean = FALSE)
head(earliest_records)
doj_speeches  

Fetch U.S. DOJ Speeches

Description
For accessing U.S. Department of Justice (DOJ) open data speeches.

Usage

```r
doj_speeches(
  n_results = 50,
  search_direction = "DESC",
  keyword = NULL,
  clean = TRUE
)
```

Arguments

- **n_results**
  The number of results to return. By default, 50 results are returned.
- **search_direction**
  Set to ASC or DESC to fetch data starting from the most or least recent entries. By default, data is sorted in descending order (the most recent entries).
- **keyword**
  Only return results containing a keyword. Keyword can be a single word, a list or words, or a regex separated by the pipe delimiter.
- **clean**
  Set to FALSE to return data without additional cleaning. By default HTML tags, white space, line breaks, and other messy characters are removed. Epoch time is converted to year-month-day format.

Value

Dataframe containing API pull of U.S. DOJ speeches.

Examples

# Return a dataframe with 50 of the latest records. Data is cleaned.
latest_records <- doj_speeches()
head(latest_records)

# Return a dataframe with 15 of the earliest records. Data is not cleaned.
earliest_records <- doj_speeches(n_results = 15, search_direction = "ASC", clean = FALSE)
head(earliest_records)
Index

doj_blog_entries, 2

doj_press_releases, 3

doj_speeches, 4