Package 'fedregs'

October 8, 2018

Type Package

Title Text Analysis of the US Code of Federal Regulations

Version 0.1.0

Description The Code of Federal Regulations (CFR) annual edition is the codification
of the general and permanent rules published in the Federal Register by the departments
and agencies of the Federal Government of the United States of America. Simply, the
'fedregs' package facilitates word processing and sentiment analysis of the CFR using tidy
principles. Note: According to the Code of Federal Regulations XML Rendi-
tion User Guide Document:
``In general, there are no restrictions on re-use of information in Code of Federal Regulations
material because U.S. Government works are not subject to copyright. OFR and GPO do not
restrict downstream uses of Code of Federal Regulations data, except that independent providers
should be aware that only the OFR and GPO are entitled to represent that they are the providers
of the official versions of the Code of Federal Regulations and related Federal Register
publications.''

License GPL-3

BugReports https://github.com/slarge/fedregs/issues

Encoding UTF-8

LazyData true

Depends R (>= 3.1.0),

Imports dplyr (>= 0.7.4), magrittr (>= 1.5), xml2 (>= 1.2.0), purrr(>=
0.2.5), rvest(>= 0.3.2), stringi(>= 1.1.7), httr (>= 1.3.1),
tidytext (>= 0.1.9)

Suggests ggplot2 (>= 2.2.1), tidyr (>= 0.8.0), testthat (>= 2.0.0)

RoxygenNote 6.0.1

NeedsCompilation no

Author Scott Large [cre, aut]

Maintainer Scott Large <scott.large@noaa.gov>

Repository CRAN

Date/Publication 2018-07-28 16:40:06 UTC

R topics documented:

---

cfr_part   *Parse the Relevant Details for CFR urls.*

Description

cfr_part returns a data_frame year, title, volume, chapters, parts, and URL for each url

Usage

```
cfr_part(url, verbose = FALSE)
```

Arguments

url	A valid url for .xml CFR volumes. Ideally, from cfr_urls.
verbose	logical. Will return "helpful" messages regarding the status of the URL.

Details

cfr_part

Since we're after more refined data than a single volume, we need to figure out what chapters and parts are associated with each volume. This function parses the xml and scrapes the Table of Contents for the information held in each volume.

Value

Numeric (year, title, volume, and chapters) and characters (parts and URL).

Examples

```
part_vec <- cfr_urls(year = 2017, title_number = 50)
cfr_part(part_vec[1])
```

---

cfr_text   *Extract the Text for a Given Year, Title, Chapter, and Part*

Description

`cfr_text` returns a tibble of CFR text

Usage

```
cfr_text(year, title_number, chapter, part, token = "words",
  return_tidytext = TRUE, verbose = FALSE, ...)
```

Arguments

`year`	numeric between 1996 and 2017.
`title_number`	numeric between 1 and 50.
`chapter`	numeric or roman numeral.
`part`	numeric.
`token`	character. Unit for tokenizing. Currently
`return_tidytext`	
	logical. TRUE = tidytext, FALSE = raw data
`verbose`	logical. Will return "helpful" messages regarding the status of the URL.
`...`	Extra arguments passed on to tokenizers, such as n and k for "ngrams" and "skip_ngrams"

Details

cfr_text

This function is the main function of the `fedregs` package. It takes the title, chapter, part, and year and returns a tibble of raw text (return_tidytext = FALSE) or tidytext text (return_tidytext = TRUE). N.b., it has not been extensively tested on titles and chapters other than Title 50 chapter VI and part 648.

Value

a tibble with year, title_number, chapter, part, and text nested by subpart

Examples

```
regs <- cfr_text(year = 2017,
title_number = 50,
chapter = 6,
part = 648,
return_tidytext = TRUE,
verbose = TRUE)
head(regs)
```

---

cfr_urls   *URLs for .xml Code of Federal Regulations.*

Description

`cfr_urls` returns a character string of valid URLs associated with a year and CFR title.

Usage

```
cfr_urls(year, title_number, check_url = TRUE, verbose = FALSE)
```

Arguments

year	numeric (YYYY) between 1996 and 2017.
title_number	numeric between 1 and 50.
check_url	logical. Should the URLs be tested using `httr::http:error()`.
verbose	logical. Will return "helpful" messages regarding the status of the URL.

Details

cfr_urls

The Code of Federal Regulations (CFR) is divided into titles, chapters, parts, subparts, and sections. Each title within the CFR is divided into volumes. Unfortunately, each chapter isn't consistently in the same volume so `cfr_urls` function scrapes up all the valid URLs for a given title/year combination.

Value

Valid URLs are returned as a vector of character strings. Invalid URLs are returned as 'NA''s.

Examples

```
library(dplyr)
url_list <- expand.grid(years = 2015:2017,
  title = 50,
  KEEP.OUT.ATTRS = FALSE,
  stringsAsFactors = FALSE) %>%
  mutate(url = purrr::map2(years, title, cfr_urls, check_url = TRUE))
  head(url_list)
```

---

fedregs   fedregs *package*

Description

Text Analysis of the US Code of Federal Regulations

Details

See the README on [GitHub](#)

Index