

Package ‘GDELTtools’

November 26, 2021

Type Package

Title Download, Slice, and Normalize GDELT V1 Data

Version 1.5

Date 2021-11-28

Author Stephen R. Haptonstahl, Thomas Scherer, Timo Thoms, and Patrick Wheatley

Maintainer Stephen R. Haptonstahl <srh@haptonstahl.org>

Description The GDELT V1 Event data set is over 37 GB now and growing 250 MB a month. The number of source articles has increased over time and unevenly across countries. This package makes it easy to download a subset of that data, then normalize that data to facilitate valid time series analysis.

License MIT + file LICENSE

Depends R (>= 4.0), utils

Imports plyr, TimeWarp, dplyr

Suggests testthat

RoxygenNote 7.1.2

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2021-11-26 14:00:06 UTC

R topics documented:

GetAllOfGDELT	2
GetGDELT	3
NormEventCounts	5
Index	7

`GetAllOfGDELT`*Download all the GDELT files to a local folder*

Description

Downloads all GDELT files not already present locally. **** This takes a long time and a lot of space. ****

Usage

```
GetAllOfGDELT(  
  local_folder,  
  data_url_root = "http://data.gdeltproject.org/events/",  
  force = FALSE  
)
```

Arguments

<code>local_folder</code>	character, path to the file to be validated.
<code>data_url_root</code>	character, URL for the folder with GDELT data files.
<code>force</code>	logical, if TRUE then the download is carried out without further prompting the user.

Value

logical, TRUE if all files were downloaded successfully.

Author(s)

Stephen R. Haptonstahl <srh@haptonstahl.org>

References

GDELT: Global Data on Events, Location and Tone, 1979-2013. Presented at the 2013 meeting of the International Studies Association in San Francisco, CA. <https://www.gdeltproject.org/>

Examples

```
## Not run:  
GetAllOfGDELT("~/gdeltdata")  
## End(Not run)
```

 GetGDELТ

Download and subset GDELТ data

Description

Download the GDELТ files necessary for a data set, import them, filter on various criteria, and return a data.frame.

Usage

```
GetGDELТ(
  start_date,
  end_date = start_date,
  row_filter,
  ...,
  local_folder = tempdir(),
  max_local_mb = Inf,
  data_url_root = "http://data.gdelтproject.org/events/",
  verbose = TRUE
)
```

Arguments

start_date	character, just about any human-readable form of the earliest date to include.
end_date	character, just about any human-readable form of the latest date to include.
row_filter	<data-masking> Row selection. Expressions that return a logical value, and are defined in terms of the variables in GDELТ. If multiple expressions are included, they are combined with the & operator. Only rows for which all conditions evaluate to TRUE are kept.
...	<tidy-select>, Column selection. This takes the form of one or more unquoted expressions separated by commas. Variable names can be used as if they were positions in the data frame, so expressions like x:y can be used to select a range of variables.
local_folder	character, if specified, where downloaded files will be saved.
max_local_mb	numeric, the maximum size in MB of the downloaded files that will be retained.
data_url_root	character, URL for the folder with GDELТ data files.
verbose	logical, if TRUE then indications of progress will be displayed_

Details

Dates are parsed with dateParse in the TimeWarp package. Years must be given with four digits.

If local_folder is not specified then downloaded files are stored in tempdir(). If a needed file has already been downloaded to local_folder then this file is used instead of being downloaded. This can greatly speed up future downloads.

Value

data.frame

Filtering Results

The `row_filter` is passed to `filter`. This is a very flexible way to filter the rows. It's well worth checking out the `filter` documentation.

Selecting Columns

The `...` is passed to `select`. This is a very flexible way to choose which columns to return. It's well worth checking out the `select` documentation.

Author(s)

Stephen R. Haptonstahl	<srh@haptonstahl.org>
Thomas Scherer	<tscherer@princeton.edu>
John Beieler	<jub270@psu.edu>

References

GDELT: Global Data on Events, Location and Tone, 1979-2013. Presented at the 2013 meeting of the International Studies Association in San Francisco, CA. <https://www.gdeltproject.org/>

Examples

```
## Not run:
df1 <- GetGDELT(start_date="1979-01-01", end_date="1979-12-31")

df2 <- GetGDELT(start_date="1979-01-01", end_date="1979-12-31",
  row_filter=ActionGeo_CountryCode=="US")

df3 <- GetGDELT(start_date="1979-01-01", end_date="1979-12-31",
  row_filter=Actor2Geo_CountryCode=="RS" & NumArticles==2 & is.na(Actor1CountryCode),
  1:5)

df4 <- GetGDELT(start_date="1979-01-01", end_date="1979-12-31",
  row_filter=Actor2Code=="COP" | Actor2Code=="MED",
  contains("date"), starts_with("actor"))

# Specify a local folder to store the downloaded files
df5 <- GetGDELT(start_date="1979-01-01", end_date="1979-12-31",
  row_filter=ActionGeo_CountryCode=="US",
  local_folder = "~/gdeltdata")

## End(Not run)
```

NormEventCounts	<i>Scale event counts</i>
-----------------	---------------------------

Description

Scale event counts based on the unit of analysis.

Usage

```
NormEventCounts(x, unit_analysis, var_name = "norming_vars")
```

Arguments

x	data.frame, a GDELT data.frame.
unit_analysis	character, default is country_day; other options: country_month, country_year, day, month, year
var_name	character, base name for the new count variables

Details

For unit_analysis, day and country-day put out a data set where date is of class 'date'. All other options put out a data set where year or month is integer (this needs to be unified in a later version).

Value

data.frame

Author(s)

Oskar N.T. Thoms	<othoms@princeton.edu>
Stephen R. Haptonstahl	<srh@haptonstahl.org>
John Beielor	<jub270@psu.edu>

References

GDELT: Global Data on Events, Location and Tone, 1979-2012. Presented at the 2013 meeting of the International Studies Association in San Francisco, CA. <https://www.gdeltproject.org/>

Examples

```
## Not run:  
GDELT_subset_data <- GetGDELT("2013-06-01", "2013-06-07",  
  (ActionGeo_CountryCode=="AF" | ActionGeo_CountryCode=="US") & EventCode>=140 & EventCode<150,  
  local_folder="~/gdelldata")
```

```
GDELT_normed_data <- NormEventCounts(x = GDELT_subset_data,  
  unit_analysis="day",  
  var_name="protest")  
## End(Not run)
```

Index

filter, [4](#)

GetAllOfGDELT, [2](#)

GetGDELT, [3](#)

NormEventCounts, [5](#)

select, [4](#)