Package 'GDELTtools'

February 19, 2015

Type Package

Title Download, slice, and normalize GDELT data

Version 1.2

Date 2013-02-17

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

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

Description The GDELT data set is over 60 GB now and growing 100 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 (>= 2.10), tools

Imports plyr, TimeWarp

Suggests testthat

NeedsCompilation no

Repository CRAN

Date/Publication 2014-02-27 18:04:13

R topics documented:

GDELTtools *Download, slice, and normalize GDELT data*

Description

The GDELT data set is over 60 GB now and growing 100 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 timeseries analysis.

Details

Package:	GDELTtools
Type:	Package
Version:	1.2
Date:	2014-02-17
License:	MIT + file LICENSE

`GetGDELT` is used to download and subset data.

`NormEventCounts` takes the output from `GetGDELT` and normalizes the counts appropriately for conducting time series analysis.

Author(s)

Stephen R. Haptonstahl	<srh@haptonstahl.org>
Thomas Scherer	<tscherer@princeton.edu>
Oskar N.T. Thoms	<othoms@princeton.edu>
Patrick Wheatley	<pvwheatley@gmail.com>

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

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. http://gdelt.utdallas.edu/

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

Description

Downloads all (missing) GDELT files. ** 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

local.folder	character, path to the file to be validated.
data.url.root	character, URL for the folder with GDELT data files.
force	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, 1979-2012. Presented at the 2013 meeting of the International Studies Association in San Francisco, CA. http://www.gdeltproject.org/

Examples

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

GetGDELT *Download and subset GDELT data*

Description

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

Usage

```
GetGDELT(start.date, end.date = start.date, filter,
  local.folder = tempdir(), max.local.mb = Inf, allow.wildcards = FALSE,
  use.regex = FALSE, data.url.root = "http://data.gdeltproject.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.
filter	list, named list encoding the values to include for specified fields. See Details.
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.
allow.wildcards	logical, must be TRUE to use * in filter to specify 'any character(s)'.
use.regex	logical, if TRUE then filter will be processed as a regular expression.
data.url.root	character, URL for the folder with GDELT data files.
verbose	logical, if TRUE then indications of progress will be displayed.

Details

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

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

Value

data.frame

Filtering Results

This is how you write the filter.

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, 1979-2012. Presented at the 2013 meeting of the International Studies Association in San Francisco, CA. http://www.gdeltproject.org/

Examples

```
## Not run:
test.filter <- list(ActionGeo_ADM1Code=c("NI", "US"), ActionGeo_CountryCode="US")
test.results <- GetGDELT(start.date="1979-01-01", end.date="1979-12-31",
```

```
  filter=test.filter)
table(test.results$ActionGeo_ADM1Code)
table(test.results$ActionGeo_CountryCode)
## End(Not run)

# Specify a local folder to store the downloaded files
## Not run:
test.results <- GetGDELT(start.date="1979-01-01", end.date="1979-12-31",
                         filter=test.filter,
                         local.folder="~/gdeltdata",
                         max.local.mb=500)
## End(Not run)
```

NormEventCounts *Scale event counts*

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.day, 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 Beieler	<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. http://www.gdeltproject.org/

Examples

```
## Not run:
GDELT.subset.data <- GetGDELT("2013-06-01", "2013-06-07", allow.wildcards=TRUE,
  filter=list(ActionGeo_CountryCode=c("AF", "US"), EventCode="14*"),
  local.folder="~/gdeltdata")
GDELT.normed.data <- NormEventCounts(x = GDELT.subset.data,
  unit.analysis="day",
  var.name="protest")
## End(Not run)
```

Index