Package ‘datasailr’

November 8, 2021

Type Package
Title Row by Row Data Processing Tool, Using ‘DataSailr’ Script
Version 0.8.10
Date 2021-10-27
Maintainer Toshihiro Umehara <toshi@niceume.com>
Description A row by row data processing tool. You can write data processing code in ‘DataSailr’ script which is specially intended for data manipulation. The package uses ‘libsailr’ (C/C++ library) for its ‘DataSailr’ code parsing and its execution.
Encoding UTF-8
License GPL (>= 3)
Biarch true
Imports Rcpp (>= 0.12.18)
Suggests RUnit
LinkingTo Rcpp
SystemRequirements GNU make, C++11
URL https://datasailr.io
BugReports https://github.com/niceume/datasailr
NeedsCompilation yes
Author Toshihiro Umehara [aut, cre],
  Troy Hanson [cph, ctb] (uthash),
  Howard Hinnant [cph, ctb] (date.h),
  Adrian Colomitchi [cph, ctb] (date.h),
  Florian Dang [cph, ctb] (date.h),
  Paul Thompson [cph, ctb] (date.h),
  Tomasz Kamiński [cph, ctb] (date.h),
  Nemanja Trifunovic [cph, ctb] (utfcpp),
  Kim Grasman [cph, ctb] (getopt_port),
  Jon Clayden [cph, ctb] (ore package),
  K.Kosako [cph, ctb] (onigmo author),
  K.Takata [cph, ctb] (onigmo author),
  ...
Description
A row by row data processing tool. You can write data manipulation code in 'DataSailr' script which is specially intended for the data manipulation. The package uses 'libsailr' (C/C++ library) for its 'DataSailr' code parsing and its execution.

Details
The main function, datasailr::sail(), takes dataframe as its 1st argument and 'DataSailr' script as its 2nd argument, and returns the processing results as dataframe. This package works with 'libsailr' library, which conducts arithmetic calculations and string/character manipulations following 'Sailr' script (which is a base version of 'DataSailr' script). 'DataSailr' package passes both 'DataSailr' script and values of each row of dataframe to the 'libsailr', and 'libsailr' processes those values of each row. The processing results are stored and are returned as dataframe.

Author(s)
Toshi Umehara, toshi@niceume.com. Maintainer: Toshi Umehara <toshi@niceume.com>

See Also
sail.
adult_total

Examples

```r
library(datasailr)
data(mtcars)

code_example = 'powerful = .
  if( 145 >= hp & hp > 0 ){
    powerful = 0
  }else if( hp > 145 ){  
    powerful = 1
  }else{
    powerful = 2
  }

  germany = re/(^Merc|^Porsche|^Volvo)/
  japan = re/(^Mazda|^Honda|^Toyota)/

  if ( _rowname_ =~ germany ){
    country = "Germany"
  }else if( _rowname_ =~ japan ){
    country = "Japan"
  }else{
    country = "Other"
  }

  company = rexp_matched(1)

  sail(mtcars, code_example)
```

Census income dataset

Description

"Census Income" dataset from UCI machine learning repository.

Usage

data(adult_total)

Format


Source

https://archive.ics.uci.edu/ml/datasets/Adult

References

https://archive.ics.uci.edu/ml/datasets/Adult

---

sail

Run 'DataSailr' script and manipulate dataframe

---

Description

datasailr::sail() is the main function of this 'DataSailr' package. This function takes a data.frame as the 1st argument and a 'DataSailr' script as the 2nd argument.

Usage

sail( df, code, fullData = TRUE, rowname = "_rowname_", stringsAsFactors = FALSE)

Arguments

df data.frame to be processed.

code 'DataSailr' script that instructs how to manipulate values and strings.

fullData When fullData is set FALSE, only the columns corresponding to left-hand-side(LHS) variables in 'DataSailr' scripts are returned. (e.g. In 'DataSailr' script, bmi = weight / height / height , the variable bmi is the LHS.) If set TRUE, all the columns are returned.

rowname Dataframe's row names are accessed via the variable name specified by this option. The default variable name is _rowname_, meaning that _rowname_ represents dataframe's row names in 'DataSailr' script, though this does not allow assignment. Even if you assign some value to this variable, they are not reflected in the row names of the resultant dataframe.

stringsAsFactors Columns of character vectors (= string vectors) that appear as left-hand-side(LHS) variables in 'DataSailr' scripts can be returned as either character vectors or factor vectors. If stringsAsFactors is set to FALSE(default), character columns are returned as character vectors. If stringsAsFactors is set to TRUE, character columns are returned as factor vectors.
Value

If `fullData` argument is set `FALSE`, return value is a `data.frame`. If `fullData` is `TRUE`, the result is created by `cbind()`, which concatenates the original input and newly created dataframe. For example, if the original input is `data.table`, the return value is `data.table`.

Examples

```r
library(datasailr)
data(iris)

iris_code = '
iris_type = .
if ( Species == "setosa" ) {
  type = 1
  type_char = "A"
} else if( Species == "versicolor" ) {
  type = 2
  type_char = "B"
} else if(Species == "virginica" ) {
  type = 3
  type_char = "C"
}

sail(iris, iris_code)
```
Index

* datasets
  adult_total, 3

* package
  datasailr-package, 2

adult_total, 3

datasailr-package, 2

sail, 2, 4