Package: DataClean (via r-universe)

October 13, 2024

Type Package

Title Data Cleaning

Version 1.0

Date 2016-03-20

Author Xiaorui(Jeremy) Zhu

Maintainer Xiaorui(Jeremy) Zhu <zhuxiaorui1989@gmail.com>

Depends R (>= 3.1.0)

Imports xlsx, XML

Description Includes functions that researchers or practitioners may use to clean raw data, transferring html, xlsx, txt data file into other formats. And it also can be used to manipulate text variables, extract numeric variables from text variables and other variable cleaning processes. It is originated from a author's project which focuses on creative performance in online education environment. The resulting paper of that study will be published soon.

License GPL-3

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2016-03-25 09:10:37

Contents

`consolida` *An internal function for data merging.*

Description

This is a function that use to match original data and addin data with identified variable.

Usage

```
consolida(row, data, mergeVar)
```

Arguments

`row`	One sample that is already divided from the original file.
`data`	The "addin" file.
`mergeVar`	The variable that use to merge.

Details

This function is for internal use only, so no need to export it. It figures out the ID in the "addin" file then merge variables in addin file to the original file. This function is used for further "lapply" porcess.

Value

is single line contains original variables and addin variables.

Author(s)

Xiaorui.Zhu

`getSfilesPath` *Collecting paths of some specified files that you want to import or read.*

Description

If you want to collect all files under certain folder, this function should be the perfect one. It will collect all files with certain name. Then this function will return a list will all paths of those files so that further import or read is feasible.

Usage

```
getSfilesPath(root.path, filename)
```

Arguments

`root.path`	is the root path including all folders and files that you would like to search.
`filename`	is the name of files that you want to collect.

Value

The whole paths of all files that meet the criteria were saved as a list.

Examples

```
getSfilesPath(root.path = R.home(), filename = "?.exe")
```

`htmltodata` *htmltodata*

Description

"htmltodata" function is used to transfer information from html files to R or xlsx files

Usage

```
htmltodata(path)
```

Arguments

`path`	is the path of the file that you want to import into R and then export.

Value

The return data are a list include all text results of submitters' answers.

Author(s)

Xiaorui(Jeremy) Zhu

MergerXLSX *A function to merger xlsx files by a same variable.*

Description

This is a function that can be used to merger xlsx file using identified variables.

Usage

```
MergerXLSX(original_file, addin_file, mergeID)
```

Arguments

`original_file` The name of original file. This file contains all original data. It should be a "xlsx" file and saved in the same working folder. This input must be a character string of file name if it is saved in working directory, or it should include saving path of file.

`addin_file` The file that need to be merged. It should be "xlsx" file and saved in the same working folder.

`mergeID` The merger variable name in both files. The variable name should be same in two files.

Details

This function need three parameters. First is name of the original file that contains original data. Second is name of file that need to be merged. Third is the identifiable variable name that in both files.

Value

Return data are all original data with addin variables.

Author(s)

Xiaorui (Jeremy) Zhu

References

Author's Github https://github.com/XiaoruiZhu. If you have trouble with rJava or xlsx, please check http://stackoverflow.com/questions/7019912/using-the-rjava-package-on-win7-64-bit-with-r for further information to fix it.

Examples

```
# file1 <- "C:/data.xlsx"
# file2 <- "C:/data2.xlsx"
# merged <- MergerXLSX(file1, file2, mergeID)
```

Index