Package 'readbulk'

August 25, 2020

Type Package

Version 1.1.3

Date 2020-08-24

Title Read and Combine Multiple Data Files

Description Combine multiple data files from a common directory.
The data files will be read into R and bound together, creating a
single large data.frame. A general function is provided along with
a specific function for data that was collected using the open-source
experiment builder 'OpenSesame' <https://osdoc.cogsci.nl/>.

Maintainer Pascal J. Kieslich <pascal.kieslich@gmail.com>

URL https://github.com/pascalkieslich/readbulk

BugReports https://github.com/pascalkieslich/readbulk/issues

License GPL-3

Depends R (>= 3.1.0)

Imports utils, plyr, jsonlite

Encoding UTF-8

LazyData TRUE

RoxygenNote 7.1.1

NeedsCompilation no

Author Pascal J. Kieslich [aut, cre] (<https://orcid.org/0000-0002-0853-9364>),
Felix Henninger [aut] (<https://orcid.org/0000-0002-7730-9511>)

Repository CRAN

Date/Publication 2020-08-25 15:50:03 UTC

R topics documented:

1

process_labjs   *Process lab.js data stored in survey software*

Description

Process lab.js data stored in survey software by converting them from JSON data to a data.frame in
long format. This function is currently in beta mode.

Usage

```
process_labjs(
  data,
  labjs_column,
  keep = NULL,
  remove_no_json = FALSE,
  verbose = FALSE,
  ...
)
```

Arguments

data	a data.frame.
labjs_column	a character string specifying the column in data that contains the lab.js data.
keep	a character vector specifying which additional columns in data should be kept.
remove_no_json	logical indicating whether rows with no JSON data should be kept. Only relevant if keep contains one or more variables.
verbose	logical indicating whether function should report its progress.
...	additional arguments passed on to fromJSON.

Value

A data.frame containing the processed data.

Author(s)

Felix Henninger (<mailbox@felixhenninger.com>)

Pascal J. Kieslich (<kieslich@psychologie.uni-mannheim.de>)

Examples

```
## Not run:
raw_data <- read.csv("raw_data.csv",encoding="UTF-8")

dataset <- process_labjs(data=raw_data,
  labjs_column = "FB01_01",
  keep=c("CASE","SERIAL"),
```

```
    verbose=TRUE)

  ## End(Not run)
```

---

read_bulk   *Read and combine multiple data files*

Description

Read and combine multiple data files. The files will be merged into one [data.frame](#).

Usage

```
read_bulk(
  directory = ".",
  subdirectories = FALSE,
  extension = NULL,
  data = NULL,
  verbose = TRUE,
  fun = utils::read.csv,
  ...
)
```

Arguments

directory	a character string. Name of the folder where the raw data are stored. If it does not contain an absolute path, the file name is relative to the current working directory. Defaults to current working directory.
subdirectories	logical indicating whether the directory contains subdirectories. If FALSE (the default), it is assumed that all raw data files are directly included in the directory. If TRUE, it is assumed that the raw data files are stored in folders within the directory. Alternatively, a vector of folder names that contain the raw data.
extension	an optional character string. If specified, only files ending with the specified extension will be merged.
data	A data.frame to which the new data will be added. This is optional, and an empty data.frame is used if none is provided.
verbose	logical indicating whether function should report its progress.
fun	the function used for reading the individual files. By default, this is [read.csv](#). Can be any data import function as long as it takes the file name as first argument.
...	additional arguments passed on to fun.

Details

read_bulk provides a wrapper around a specific data import function (read.csv by default) to load the individual data files. After loading, the different data files are merged using rbind.fill. This function can deal with varying column names across files, and still places data into the appropriate columns. If a column is not present in a specific file, it will be filled with NA.

Value

A data.frame containing the merged data.

One column in the data.frame (File) contains the name of the raw data file. If the subdirectories option is set, an additional column (Subdirectory) with the name of the subdirectory is added.

See Also

read.table for reading individual data files.

rbind.fill is responsible for merging files.

write.table for data export.

Examples

```
## Not run:
# Merge all files in the main folder "raw_data"
# (which is in the current working directory)
raw_data <- read_bulk(directory = "raw_data")

# Merge files with file extension ".csv"
raw_data <- read_bulk(directory = "raw_data",
  extension = ".csv")

# Merge all files stored in separate folders
# within the folder "raw_data"
raw_data <- read_bulk(directory = "raw_data",
  subdirectories = TRUE)

# Merge all raw data stored in the folders "Session1"
# and "Session2" within the folder "raw_data"
raw_data <- read_bulk(directory = "raw_data",
  subdirectories = c("Session1","Session2"))

# Merge tab separated data files and prevent
# character vectors from being converted to factors
raw_data <- read_bulk(directory = "raw_data",
  fun=read.delim,stringsAsFactors=FALSE)

## End(Not run)
```

read_opensesame	*Read and combine raw OpenSesame data*

Description

Read and combine multiple raw data files that were collected with OpenSesame (Mathot, Schreij, & Theeuwes, 2012). The files will be merged into one data.frame.

Usage

```
read_opensesame(
  directory = ".",
  subdirectories = FALSE,
  extension = NULL,
  data = NULL,
  verbose = TRUE
)
```

Arguments

directory	a character string. Name of the folder where the raw data are stored. If it does not contain an absolute path, the file name is relative to the current working directory. Defaults to current working directory.
subdirectories	logical indicating whether the directory contains subdirectories. If FALSE (the default), it is assumed that all raw data files are directly included in the directory. If TRUE, it is assumed that the raw data files are stored in folders within the directory. Alternatively, a vector of folder names that contain the raw data.
extension	an optional character string. If specified, only files ending with the specified extension will be merged.
data	A data.frame to which the new data will be added. This is optional, and an empty data.frame is used if none is provided.
verbose	logical indicating whether function should report its progress.

Details

OpenSesame generally produces an output .csv file for each participant in the experiment. This is handy during data collection, but for the analysis it is often useful to combine many such files into a single data.frame. This is the single task of the read_opensesame function, which loads all files from a given directory and attempts to combine them into a data.frame.

read_opensesame provides a wrapper around read_bulk to load the raw data files. After loading, the different data files are merged using rbind.fill. This function can deal with varying column names across files, and still places data into the appropriate columns. If a column is not present in a specific file, it will be filled with NA.

Value

A data.frame containing the merged raw data.

One column in the data.frame (`File`) contains the name of the raw data file. If the `subdirectories` option is set, an additional column (`Subdirectory`) with the name of the subdirectory is added.

References

Mathot, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324.

See Also

read_bulk for reading and combining multiple data files that have other file formats.

Examples

```
## Not run:
# Read single raw data file from OpenSesame
raw_data <- utils::read.csv("raw_data/subject-1.csv",encoding = "UTF-8")

# Merge all files in the main folder "raw_data"
# (which is in the current working directory)
raw_data <- read_opensesame(directory = "raw_data")

# Merge files with file extension ".csv"
raw_data <- read_opensesame(directory = "raw_data",
  extension = ".csv")

# Merge all files stored in separate folders
# within the folder "raw_data"
raw_data <- read_opensesame(directory = "raw_data",
  subdirectories = TRUE)

# Merge all raw data stored in the folders "Session1"
# and "Session2" within the folder "raw_data"
raw_data <- read_opensesame(directory = "raw_data",
  subdirectories = c("Session1","Session2"))

# Export merged data to a file using write.table
write.table(raw_data, file = "raw_data.csv",
  sep=",", row.names = FALSE)

## End(Not run)
```

Index