

Package ‘provSummarizeR’

August 18, 2022

Title Summarizes Provenance Related to Inputs and Outputs of a Script or Console Commands

Version 1.5.1

Date 2022-08-17

Description Reads the provenance collected by the 'rdtLite' or 'rdt' packages, or other tools providing compatible PROV JSON output, created by the execution of a script or a console session, and provides a human-readable summary identifying the input and output files, the scripts used (if any), errors and warnings produced, and the environment in which it was executed. It can also optionally package all the files into a zip file. The exact format of the PROV JSON file created by 'rdtLite' and 'rdt' is described in <https://github.com/End-to-end-provenance/ExtendedProvJson>. More information about 'rdtLite' and associated tools is available at <https://github.com/End-to-end-provenance/> and Lerner, Boose, and Perez (2018), Using Introspection to Collect Provenance in R, Informatics, [doi:10.3390/informatics5010012](https://doi.org/10.3390/informatics5010012).

Depends R (>= 3.5)

License GPL-3

Encoding UTF-8

Imports provParseR (>= 1.0)

Suggests digest, knitr, rdtLite, rdt, testthat

Additional_repositories <https://end-to-end-provenance.github.io/drat/>

URL <https://github.com/End-to-end-provenance>

BugReports <https://github.com/End-to-end-provenance/provSummarizeR/issues>

RoxygenNote 7.2.0

VignetteBuilder knitr

NeedsCompilation no

Author Emery Boose [cre],
Barbara Lerner [aut],
President and Fellows of Harvard College [cph],
Trustees of Mount Holyoke College [cph]

Maintainer Emery Boose <boose@fas.harvard.edu>

Repository CRAN

Date/Publication 2022-08-17 22:10:13 UTC

R topics documented:

prov.summarize 2

Index 5

prov.summarize *Provenance summarization functions*

Description

prov.summarize uses provenance collected from execution of an R script (prov.run) or from a console session (prov.init) and outputs a text summary to the R console.

Usage

```
prov.summarize(
  details = FALSE,
  check = TRUE,
  console = TRUE,
  save = FALSE,
  create.zip = FALSE,
  notes = TRUE
)
```

```
prov.summarize.file(
  prov.file,
  details = FALSE,
  check = TRUE,
  console = TRUE,
  save = FALSE,
  create.zip = FALSE,
  notes = TRUE
)
```

```
prov.summarize.run(
  r.script,
  details = FALSE,
  check = TRUE,
  console = TRUE,
  save = FALSE,
  create.zip = FALSE,
  notes = TRUE,
```

```
    ...
  )
```

Arguments

details	whether to display library, script, file, and message details
check	whether to check against the user's file system
console	whether to display results in the console
save	whether to save the provenance summary to the file prov-summary.txt in the current working directory
create.zip	whether to package the provenance data into a zip file stored in the current working directory
notes	whether to include notes
prov.file	the path to the file containing provenance
r.script	the name of a file containing an R script
...	extra parameters are passed to the provenance collector. See rdt's prov.run function or rdtLites's prov.run function for details.

Details

These functions use provenance collected by the rdtLite or rdt packages.

The provenance summary includes:

- The name of the script file executed.
- Environmental information identifying when the script was executed, the version of R, the computing system, the tool and version used to collect provenance, the location of the provenance data, and the hash algorithm used to hash files.
- The names of any scripts sourced.
- The names of any variables in the global environment that are used but not set by a script or a console session.
- Any URLs loaded.
- The names of files input or output.
- Any messages sent to standard output.
- Any errors or warnings.

If details = TRUE, additional details are displayed, including (1) libraries loaded by the user's code, loaded before the script started, or loaded by rdtLite or rdt, with version numbers; (2) timestamps, hash values, and stored copies for scripts and data files; and (3) source code locations for messages.

If details = FALSE, only libraries loaded by the user's code at the time of execution are displayed. Note that some libraries used by the script might have been loaded before the script was executed. Run provSummarizeR with details = TRUE to see a complete list of loaded libraries.

If check = TRUE, the user's file system is checked to see if input files, output files, and scripts (in their original locations) are unchanged, changed, or missing. The status of each file is marked as

follows: file unchanged [:], file changed [+], file missing [-], or file not checked []. Copies of the original files are stored on the provenance directory.

If `console = TRUE`, output is displayed in the console.

If `save = TRUE`, results are saved to the file "prov-summary.txt" or "prov-summary-details.txt" in the current working directory.

If `create.zip = TRUE`, the provenance data is saved as a zip file in the current working directory.

If `notes = TRUE`, notes are included for how to interpret the provenance summary.

For provenance collected from a console session, only the environment, library, pre-existing variables, URL, and file information appear in the summary.

Creating a zip file depends on a zip executable being on the search path. By default, it looks for a program named "zip". To use a program with a different name, set the value of the `R_ZIPCMD` environment variable. This code has been tested with Unix zip and with 7-zip on Windows.

Value

string containing provenance summary

Examples

```
## Not run: prov.summarize ()
testdata <- system.file("testdata", "prov.json", package = "provSummarizeR")
prov.summarize.file (testdata)
## Not run:
testdata <- system.file("testscripts", "console.R", package = "provSummarizeR")
prov.summarize.run (testdata)
## End(Not run)
```

Index

prov.summarize, [2](#)