Package 'provExplainR'

October 14, 2022

Title Compare Provenance Collections to Explain Changed Script Outputs

Version 1.1.1

Date 2022-08-15

Copyright President and Fellows of Harvard College, Trustees of Mount Holyoke College

Description Inspects provenance collected by the 'rdt' or 'rdtLite' packages, or other tools providing compatible PROV JSON output created by the execution of a script, and find differences between two provenance collections. Factors under examination included the hardware and software used to execute the script, versions of attached libraries, use of global variables, modified inputs and outputs, and changes in main and sourced scripts. Based on detected changes, 'provExplainR' can be used to study how these factors affect the behavior of the script and generate a promising diagnosis of the causes of different script results. More information about 'rdtLite' and associated tools is available at <https://github.com/End-to-end-provenance/> and Barbara Lerner, Emery Boose, and Luis Perez (2018), Using Introspection to Collect Provenance in R, Informatics, <doi:10.3390/informatics5010012>.

License GPL-3 | file LICENSE

Encoding UTF-8

Imports dplyr, provParseR (>= 0.1.2), diffobj, digest, stringr

Suggests testthat

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

BugReports https://github.com/End-to-end-provenance/provExplainR/issues

RoxygenNote 7.2.0

NeedsCompilation no

Author Barbara Lerner [cre], Emery Boose [aut], Khanh Ngo [aut]

Maintainer Barbara Lerner <blerner@mtholyoke.edu>

Repository CRAN

Date/Publication 2022-08-15 20:20:02 UTC

R **topics documented:**

---

prov.explain   *Provenance comparison functions*

Description

prov.explain reads two provenance collections and finds differences between these two versions.

prov.diff.script visualizes the differences between two versions of a script that were previously executed.

Usage

```
prov.explain(dir1, dir2, save = FALSE)

prov.diff.script(dir1, dir2, first.script = NULL, second.script = NULL)
```

Arguments

dir1	path to first provenance directory
dir2	path to second provenance directory
save	if true saves the report to the file prov-explain.txt in the first directory
first.script	name of first script. If no value is passed in, it will use the main script
second.script	name of second script. If both first and second script name are NULL, it will use the main script form the second directory. If second script name is NULL, but first script name is not, it will use first script name.

Details

prov.explain and prov.diff.script are intended to help a user determine what has changed if multiple executions of a script lead to different results. prov.explain does this by comparing provenance collected using the rdtLite or rdt packages. prov.diff.script compares copies of the R scripts saved in provenance directories at the time that the scripts were executed.

The types of differences that prov.explain can find include:

- Environmental information identifying when the scripts were executed, the version of R, the computing systems, the tool and version used to collect the provenance, the location of the provenance file, and the hash algorithm used to hash data files.
- Versions of libraries loaded
- Versions of provenance tools
- Contents and names of main and sourced scripts

The prov.diff.script compares two versions of a script. Users must specify the name of the first script, the provenance directory associated with the first execution of the script, and the provenance directory associated with the second execution of the script. The name of the second script is optional. If it is omitted, the same script name is looked for in the second provenance directory

Examples

```
## Not run: prov.explain("first.test.dir", "second.test.dir")
```

Index