

Package ‘SNSFdatasets’

December 8, 2024

Type Package

Title Download Datasets from the Swiss National Science Foundation (SNF, FNS, SNSF)

Version 0.1.1

Date 2024-01-31

Description Download and read datasets from the Swiss National Science Foundation (SNF, FNS, SNSF; <<https://snf.ch>>). The package is lightweight and without dependencies. Downloaded data can optionally be cached, to avoid repeated downloads of the same files. There are also utilities for comparing different versions of datasets, i.e. to report added, removed and changed entries.

License GPL-3

URL <http://enricoschumann.net/R/packages/SNSFdatasets/> ,
<https://git.sr.ht/~enricoschumann/SNSFdatasets> ,
<https://github.com/enricoschumann/SNSFdatasets>

LazyLoad yes

ByteCompile yes

NeedsCompilation no

Author Silvia Martens [ctb] (<<https://orcid.org/0009-0001-7554-3195>>),
Enrico Schumann [aut, cre] (<<https://orcid.org/0000-0001-7601-6576>>)

Maintainer Enrico Schumann <es@enricoschumann.net>

Contents

SNSFdatasets	2
Index	4

Description

Download datasets from the Swiss National Science Foundation (SNF, FNS, SNSF) in CSV format.

Usage

```
fetch_datasets(dataset,
               dest.dir = NULL,
               detect.dates = TRUE, ...)

compare_datasets(filename.old, filename.new,
                 match.column = "GrantNumber", ...)

read_dataset(filename, detect.dates = TRUE, ...)
```

Arguments

dataset	a character vector. When of length greater than one, datasets are only downloaded, but not read. Currently supported are: <ul style="list-style-type: none"> • Grant • GrantWithAbstracts • Person • OutputdataScientificPublication • OutputdataUseInspired • OutputdataPublicCommunication • OutputdataCollaboration • OutputdataAcademicEvent • OutputdataAward • OutputdataDataSet • OutputdataKnowledgeTransferEvent
dest.dir	a directory; if NULL, a <code>tempdir</code> is used
detect.dates	logical: if TRUE, columns consisting of entries such as <code>2000-10-31T00:00:00Z</code> are converted to <code>Date</code> ; empty rows in such columns are ignored and become <code>NA</code>
filename.old	string: the filename
filename.new	string: the filename
filename	string: the filename
match.column	string: the name of the column to use for matching entries in old and new file
...	arguments to be passed to <code>download.file</code> (for <code>fetch_datasets</code>)

Details

`fetch_datasets` downloads datasets in CSV format from the SNSF's website and stores them, with a date prefix, in directory `dest.dir`. If the latter is `NULL`, a temporary directory is used (through `tempdir`); but much better is to use a more-persistent storage location. If a file with today's date exists in `dest.dir`, that file is read, and nothing is downloaded. If more than one dataset is specified, those files are downloaded (if not current in `dest.dir`) but not read.

For downloading, function `download.file` is used. If it fails, `fetch_datasets` returns `NULL`. Settings can be passed via `....`. See `download.file` for options; in particular, see the hints about `timeout`.

`compare_datasets` will match old and new dataset via the specified `match.column` and report

- added lines (in new, but not in old file),
- removed lines (in old, but not in new file), and
- changed lines (in both files, but with differing content).

`read_dataset` is a simple wrapper of `read.table` with appropriate settings.

Value

A `data.frame` for `fetch_datasets` and `read_dataset`. For `compare_datasets`, a `list` of three components named `added`, `removed` and `changed`.

Author(s)

Silvia Martens and Enrico Schumann

References

<https://data.snf.ch/datasets>

See Also

`download.file`; `options` (`timeout`, in particular)

Examples

```
## requires internet connection, and file may be large
dataset <- "OutputdataAward"

SNSF.dir <- tempdir() ## This is just an example.
                ## In practice it's much more useful to
                ## store files in a persistent location,
                ## such as '~/Downloads/SNSFdatasets'.

data <- fetch_datasets(dataset = dataset, dest.dir = SNSF.dir)

## all award titles
table(data[["Award_Title"]])
```

Index

`compare_datasets` (SNSFdatasets), [2](#)

`data.frame`, [3](#)

`Date`, [2](#)

`download.file`, [2](#), [3](#)

`fetch_datasets` (SNSFdatasets), [2](#)

`list`, [3](#)

`NA`, [2](#)

`options`, [3](#)

`read.table`, [3](#)

`read_dataset` (SNSFdatasets), [2](#)

SNSFdatasets, [2](#)

SNSFdatasets-package (SNSFdatasets), [2](#)

`tempdir`, [2](#), [3](#)