Package ‘capesR’

December 19, 2024

Type Package

Title Access to CAPES Data

Version 0.1.0

Date 2024-12-17

Description Provides simplified access to the data from the Catalog of Theses and Dissertations of the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES, <https://catalogodeteses.capes.gov.br>) for the years 1987 through 2022. The dataset includes variables such as Higher Education Institution (institution), Area of Concentration (area), Graduate Program Name (program_name), Type of Work (type), Language of Work (language), Author Identification (author), Abstract (abstract), Advisor Identification (advisor), Development Region (region), State (state).

License GPL (>= 2)

Encoding UTF-8

LazyData true

Depends R (>= 4.0.0)

RoxygenNote 7.3.2

Imports arrow, dplyr, magrittr, rlang, stringr, utils

Suggests knitr, rmarkdown

VignetteBuilder knitr

URL <https://github.com/hugoavmedeiros/capesR>

Config/Needs/website tidyverse/tidytemplate

NeedsCompilation no

Author Hugo Vasconcelos Medeiros [aut, cre],
Dalson Figueiredo Filho [aut],
André Leite [aut]

Maintainer Hugo Vasconcelos Medeiros <hugo.medeiros@ufpe.br>

Repository CRAN

Date/Publication 2024-12-19 16:20:02 UTC

Contents

`capes_synthetic_df` *Synthetic CAPES Data*

Description

Aggregated data from the CAPES Catalog of Theses and Dissertations, containing summarized information by year, institution, area, program, type, region, and state (UF).

Usage

```
capes_synthetic_df
```

Format

A data frame with the following columns:

base_year Reference year of the data.

institution Higher Education Institution.

area Area of Concentration.

program_name Name of the Graduate Program.

type Type of work (e.g., Master's, Doctorate).

region Region of Brazil.

state Federative Unit (state).

n Total number of works.

Source

Synthetic data created from the CAPES Catalog of Theses and Dissertations.

Examples

```
data(capes_synthetic_df)
head(capes_synthetic_df)
```

`download_capes_data` *Download CAPES Data*

Description

Downloads CAPES theses and dissertations data files from OSF for selected years.

Usage

```
download_capes_data(years, destination = tempdir(), timeout = 120)

baixar_dados_capes(years, destination = tempdir(), timeout = 120)
```

Arguments

`years`	A vector with the desired years.
`destination`	The directory where the files will be saved (default: temporary directory).
`timeout`	The timeout in seconds for the download process (default: 120 seconds).

Value

A list of file paths for the downloaded or already existing files.

Examples

```
# Download data for the years 1987 and 1990
capes_files <- download_capes_data(c(1987, 1990))
```

`read_capes_data` *Read and filter data from the CAPES Catalog of Theses and Dissertations*

Description

This function combines data from multiple Parquet files and applies optional filters, including text-based searches.

Usage

```
read_capes_data(files, filters = list())

ler_dados_capes(files, filters = list())
```

Arguments

`files`	A vector or list of paths to Parquet files.
`filters`	A list of filters to apply (e.g., list(base_year = 1987, state = "SP", title = "education")).

Value

A ‘data.frame‘ containing the combined and filtered data.

Examples

```
# Download data for the years 1987 and 1990
capes_files <- download_capes_data(c(1987, 1990))
# Combine all selected data
combined_data <- read_capes_data(capes_files)
```

`search_capes_text` *Search for terms in text fields of the CAPES Catalog of Theses and Dissertations data*

Description

This function allows searching for specific terms in the text fields of a previously loaded ‘data.frame‘.

Usage

```
search_capes_text(data, term, field)

buscar_texto_capes(data, term, field)
```

Arguments

`data`	A ‘data.frame‘ containing the CAPES Catalog of Theses and Dissertations data.
`term`	A string, the term to search for.
`field`	A string, the name of the field to search in (e.g., "resumo", "titulo").

Value

A ‘data.frame‘ with rows matching the search or a message indicating no results were found.

Examples

```
# Download data for the years 1987 and 1990
capes_files <- download_capes_data(c(1987, 1990))
# Combine all selected data
combined_data <- read_capes_data(capes_files)
# Search data
results <- search_capes_text(
data = combined_data,
term = "Educação",
  field = "titulo"
)
```

`years_osf` *Identifiers (IDs) on OSF for the annual data of the Catalog of Theses and Dissertations from the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES)*

Description

A data frame containing the years and the corresponding IDs for downloading the files.

Usage

```
years_osf
```

Format

A data frame with the following columns:

year Year of the data (1987-2022).

osf_id OSF ID corresponding to the year.

Source

https://osf.io/

Examples

```
data(years_osf)
head(years_osf)
```

Index