

Package ‘BigDataPE’

July 21, 2025

Title Secure and Intuitive Access to 'BigDataPE' 'API' Datasets
Version 0.0.96
Date 2025-04-02
Description Designed to simplify the process of retrieving datasets from the 'Big Data PE' platform using secure token-based authentication. It provides functions for securely storing, retrieving, and managing tokens associated with specific datasets, as well as fetching and processing data using the 'httr2' package.
License MIT + file LICENSE
Encoding UTF-8
RoxygenNote 7.3.2
Depends R (>= 4.1.0)
Imports dplyr, tibble, httr2
URL <<https://github.com/StrategicProjects/bigdatape>>
Config/Needs/website tidyverse/tidytemplate
NeedsCompilation no
Author Andre Leite [aut, cre],
Hugo Vaconcelos [aut],
Diogo Bezerra [aut]
Maintainer Andre Leite <leite@castlab.org>
Repository CRAN
Date/Publication 2025-04-02 13:10:05 UTC

Contents

bdpe_fetch_chunks	2
bdpe_fetch_data	3
bdpe_get_token	4
bdpe_list_tokens	5
bdpe_remove_token	5
bdpe_store_token	6
parse_queries	6

bdpe_fetch_chunks	<i>Fetch data from the BigDataPE API in chunks</i>
-------------------	--

Description

This function retrieves data from the BigDataPE API iteratively in chunks. It uses `bdpe_fetch_data` as the base function and supports limits for the total number of records to fetch and the size of each chunk.

Usage

```
bdpe_fetch_chunks(
  base_name,
  total_limit = Inf,
  chunk_size = 50000L,
  query = list(),
  verbosity = 0L,
  endpoint = "https://www.bigdata.pe.gov.br/api/buscar"
)
```

Arguments

<code>base_name</code>	A string specifying the name of the dataset associated with the token.
<code>total_limit</code>	An integer specifying the maximum number of records to fetch. Default is <code>Inf</code> (all available data).
<code>chunk_size</code>	An integer specifying the number of records to fetch per chunk. Default is 50000
<code>query</code>	A named list of additional query parameters to filter the API results. Default is an empty list.
<code>verbosity</code>	An integer specifying the verbosity level for the API requests. Values are: - 0: No verbosity (default). - 1: Minimal verbosity, showing request status. - 2: Detailed verbosity, including request and response details.
<code>endpoint</code>	A string specifying the API endpoint URL. Default is "https://www.bigdata.pe.gov.br/api/buscar".

Value

A tibble containing all the data retrieved from the API.

Examples

```
## Not run:
# Store a token for the dataset
bdpe_store_token("dengue_dataset", "token")

# Fetch up to 500 records in chunks of 100
data <- bdpe_fetch_chunks("dengue_dataset", total_limit = 500, chunk_size = 100)
```

```
# Fetch all available data in chunks of 200
data <- bdpe_fetch_chunks("dengue_dataset", chunk_size = 200)

## End(Not run)
```

bdpe_fetch_data

Fetch data from the BigDataPE API

Description

This function retrieves data from the BigDataPE API using securely stored tokens associated with datasets. Users can specify pagination parameters (`limit` and `offset`) and additional query filters to customize the data retrieval.

Usage

```
bdpe_fetch_data(
  base_name,
  limit = Inf,
  offset = 0L,
  query = list(),
  verbosity = 0L,
  endpoint = "https://www.bigdata.pe.gov.br/api/buscar"
)
```

Arguments

<code>base_name</code>	A string specifying the name of the dataset associated with the token.
<code>limit</code>	An integer specifying the maximum number of records to retrieve per request. Default is <code>Inf</code> (all records). If set to a non-positive value or <code>Inf</code> , no limit will be applied.
<code>offset</code>	An integer specifying the starting record for the query. Default is 0. If set to a non-positive value or <code>Inf</code> , no offset will be applied.
<code>query</code>	A named list of additional query parameters to filter the API results. Default is an empty list.
<code>verbosity</code>	An integer specifying the verbosity level for the API requests. Values are: - 0: No verbosity (default). - 1: Minimal verbosity, showing request status. - 2: Detailed verbosity, including request and response details.
<code>endpoint</code>	A string specifying the API endpoint URL. Default is "https://www.bigdata.pe.gov.br/api/buscar".

Value

A tibble containing the data returned by the API.

Examples

```
## Not run:
# Store a token for the dataset
bdpe_store_token("dengue_dataset", "token")

# Fetch 50 records from the beginning
data <- bdpe_fetch_data("dengue_dataset", limit = 50)

# Fetch records with additional query parameters
data <- bdpe_fetch_data("dengue_dataset", query = list(field = "value"))

# Fetch all data without limits
data <- bdpe_fetch_data("dengue_dataset", limit = Inf)

## End(Not run)
```

bdpe_get_token*Retrieve the token associated with a specific dataset*

Description

This function retrieves the authentication token stored in an environment variable for a specific dataset. If the token is not found, it returns NULL and prints a message instead of throwing an error.

Usage

```
bdpe_get_token(base_name)
```

Arguments

base_name The name of the dataset (character).

Value

A string containing the authentication token, or NULL if the token is not found.

Examples

```
token <- bdpe_get_token("education_dataset")
```

bdpe_list_tokens	<i>List all datasets that have stored tokens in environment variables</i>
------------------	---

Description

This function returns a character vector of dataset names that have their tokens stored in environment variables. Specifically, it looks for variables that begin with the prefix "BigDataPE_".

Usage

```
bdpe_list_tokens()
```

Value

A character vector of dataset names with stored tokens. If no tokens are found, an empty vector is returned and a message is printed.

Examples

```
bdpe_list_tokens()
```

bdpe_remove_token	<i>Remove the token associated with a specific dataset</i>
-------------------	--

Description

This function removes the authentication token stored in an environment variable for a specific dataset. If the token is not found, it prints a message and does not throw an error.

Usage

```
bdpe_remove_token(base_name)
```

Arguments

base_name	The name of the dataset (character).
-----------	--------------------------------------

Value

No return value. If the token is found, it is removed. If not, a message is displayed.

Examples

```
bdpe_remove_token("education_dataset")
```

bdpe_store_token	<i>Store a token in an environment variable for a specific dataset</i>
------------------	--

Description

This function stores an authentication token for a specific dataset in a system environment variable. The environment variable name is constructed by converting the dataset name to ASCII (removing accents), replacing spaces with underscores, and prefixing it with "BigDataPE_".

Usage

```
bdpe_store_token(base_name, token)
```

Arguments

base_name	The name of the dataset (character).
token	The authentication token for the dataset (character).

Details

If a variable with that name already exists (and is non-empty), the function will stop and notify you to avoid overwriting. Adjust this behavior as needed.

Value

No return value, called for side effects.

Examples

```
bdpe_store_token("educação dataset", "your-token-here")
```

parse_queries	<i>Constructs a URL with query parameters</i>
---------------	---

Description

This function appends a list of query parameters to a base URL.

Usage

```
parse_queries(url, query_list)
```

Arguments

url	The base URL to which query parameters will be added.
query_list	A list of query parameters to be added to the URL.

Value

The complete URL with the query parameters appended.

Examples

```
parse_queries("https://www.example.com", list(param1 = "value1", param2 = "value2"))
```

Index

`bdpe_fetch_chunks`, [2](#)
`bdpe_fetch_data`, [3](#)
`bdpe_get_token`, [4](#)
`bdpe_list_tokens`, [5](#)
`bdpe_remove_token`, [5](#)
`bdpe_store_token`, [6](#)

`parse_queries`, [6](#)