Package 'clintrialx'

March 12, 2025

Type Package

Title Connect and Work with Clinical Trials Data Sources

Version 0.1.1

Maintainer Indraneel Chakraborty <hello.indraneel@gmail.com>

Description Are you spending too much time fetching and managing clinical trial data? Struggling with complex queries and bulk data extraction? What if you could simplify this process with just a few lines of code? Introducing 'clintrialx' - Fetch clinical trial data from sources like 'ClinicalTrials.gov' <https://clinicaltrials.gov/> and the 'Clinical Trials Transformation Initiative - Access to Aggregate Content of ClinicalTrials.gov' database <https://aact.ctti-clinicaltrials.org/>, supporting pagination and bulk downloads. Also, you can generate HTML reports based on the data obtained from the sources!

License Apache License 2.0

Encoding UTF-8

Depends R (>= 4.0.0)

Imports httr, lubridate, readr, dplyr, progress, RPostgreSQL, tibble, DBI, rmarkdown

Suggests knitr

VignetteBuilder knitr

RoxygenNote 7.3.2

URL http://www.indraneelchakraborty.com/clintrialx/, https://github.com/ineelhere/clintrialx

NeedsCompilation no

Author Indraneel Chakraborty [aut, cre] (<https://orcid.org/0000-0001-6958-8269>)

Repository CRAN

Date/Publication 2025-03-11 23:10:15 UTC

Contents

---

aact_check_connection   *Check database connection*

Description

Check database connection

Usage

```
aact_check_connection(con)
```

Arguments

con   Database connection object

Value

A data frame with distinct study types

Examples

```
## Not run:
# Set environment variables for database credentials in .Renviron and load it
# readRenviron(".Renviron")

# Connect to the database
con <- aact_connection(Sys.getenv('user'), Sys.getenv('password'))

# Check the connection
aact_check_connection(con)

## End(Not run)
```

---

aact_connection *Connect to AACT PostgreSQL database*

Description

Connect to AACT PostgreSQL database

Usage

```
aact_connection(user, password)
```

Arguments

user	Database username
password	Database password

Value

A connection object to the AACT database

Examples

```
## Not run:
# Set environment variables for database credentials in .Renviron and load it
# readRenviron(".Renviron")

# Connect to the database
con <- aact_connection(Sys.getenv('user'), Sys.getenv('password'))

## End(Not run)
```

---

aact_custom_query *Run a custom query*

Description

Run a custom query

Usage

```
aact_custom_query(con, query)
```

Arguments

con	Database connection object
query	SQL query string

Value

A data frame with the query results

Examples

```
## Not run:
# Set environment variables for database credentials in .Renviron and load it
# readRenviron(".Renviron")

# Connect to the database
con <- aact_connection(Sys.getenv('user'), Sys.getenv('password'))

# Run a custom query
query <- "SELECT nct_id, source, enrollment, overall_status FROM studies LIMIT 5;"
results <- aact_custom_query(con, query)

# Print the results
print(results)

## End(Not run)
```

---

ctg_bulk_fetch   *Bulk Fetch Clinical Trial Data from ClinicalTrials.gov API*

Description

This function retrieves clinical trial data in bulk from the ClinicalTrials.gov API based on specified parameters. It handles pagination and returns a combined dataset.

Usage

```
ctg_bulk_fetch(
  condition = NULL,
  location = NULL,
  title = NULL,
  intervention = NULL,
  status = NULL
)
```

Arguments

condition	Character string specifying the condition to search for.
location	Character string specifying the location to search in.
title	Character string specifying the title to search for.
intervention	Character string specifying the intervention to search for.
status	A character vector specifying the recruitment status of the trials. Allowed values are: Valid values include:

- ACTIVE_NOT_RECRUITING - Studies that are actively conducting but not recruiting participants.
- COMPLETED - Studies that have completed all phases.
- ENROLLING_BY_INVITATION - Studies that are enrolling participants by invitation only.
- NOT_YET_RECRUITING - Studies that have not yet started recruiting.
- RECRUITING - Studies that are actively recruiting participants.
- SUSPENDED - Studies that are temporarily halted.
- TERMINATED - Studies that have been terminated before completion.
- WITHDRAWN - Studies that have been withdrawn before enrollment.
- AVAILABLE - Studies that are available.
- NO_LONGER_AVAILABLE - Studies that are no longer available.
- TEMPORARILY_NOT_AVAILABLE - Studies that are temporarily not available.
- APPROVED_FOR_MARKETING - Studies that have been approved for marketing.
- WITHHELD - Studies that have data withheld.
- UNKNOWN - Studies with an unknown status.

Value

A data frame containing the fetched clinical trial data.

Examples

```
## Not run:
trials <- ctg_bulk_fetch(location="india")

## End(Not run)
```

---

ctg_count *Get Count of Clinical Trials from ClinicalTrials.gov*

Description

This function retrieves the count of clinical trials from ClinicalTrials.gov based on specified parameters.

Usage

```
ctg_count(
  condition = NULL,
  location = NULL,
  title = NULL,
  intervention = NULL,
  status = NULL
)
```

Arguments

`condition`	A character string specifying the condition being studied (default: NULL).
`location`	A character string specifying the location of the trials (default: NULL).
`title`	A character string specifying keywords in the study title (default: NULL).
`intervention`	A character string specifying the type of intervention (default: NULL).
`status`	A character vector specifying the recruitment status of the trials. Allowed values are: Valid values include:

- `ACTIVE_NOT_RECRUITING` - Studies that are actively conducting but not recruiting participants.
- `COMPLETED` - Studies that have completed all phases.
- `ENROLLING_BY_INVITATION` - Studies that are enrolling participants by invitation only.
- `NOT_YET_RECRUITING` - Studies that have not yet started recruiting.
- `RECRUITING` - Studies that are actively recruiting participants.
- `SUSPENDED` - Studies that are temporarily halted.
- `TERMINATED` - Studies that have been terminated before completion.
- `WITHDRAWN` - Studies that have been withdrawn before enrollment.
- `AVAILABLE` - Studies that are available.
- `NO_LONGER_AVAILABLE` - Studies that are no longer available.
- `TEMPORARILY_NOT_AVAILABLE` - Studies that are temporarily not available.
- `APPROVED_FOR_MARKETING` - Studies that have been approved for marketing.
- `WITHHELD` - Studies that have data withheld.
- `UNKNOWN` - Studies with an unknown status.

Default is NULL.

Value

A number representing the total count of clinical trials matching the specified parameters.

Examples

```
ctg_count(
  condition = "Cancer",
  location = "India",
  title = NULL,
  intervention = "Drug",
  status = "RECRUITING"
)
```

---

ctg_data_report *Generate a Comprehensive Clinical Trial Data Report*

Description

This function creates a detailed, visually appealing HTML report from clinical trial data. It automates the process of data analysis and visualization, providing insights into various aspects of clinical trials such as study status, enrollment, duration, and funding sources.

Visit here for an example report - https://www.indraneelchakraborty.com/clintrialx/report.html.

Usage

```
ctg_data_report(
  ctg_data,
  title = "Clinical Trial Data Report",
  author = "Author Name",
  output_file = "./report.html",
 color_palette = c("#1f77b4", "#ff7f0e", "#2ca02c", "#d62728", "#9467bd", "#8c564b"),
  theme = "cerulean",
  include_data_quality = TRUE,
  include_interactive_plots = TRUE,
  custom_footer = NULL
)
```

Arguments

ctg_data	A data frame containing clinical trial data. Required columns include:

  - Study Status: Current status of each study (e.g., "Completed", "Ongoing")
  - Enrollment: Number of participants in each study
  - Start Date: The date each study began
  - Completion Date: The date each study ended or is expected to end
  - Phases: The phase of each clinical trial (e.g., "Phase 1", "Phase 2")
  - Funder Type: The type of organization funding each study
  - Study Type: The type of each study (e.g., "Interventional", "Observational")

title	Character string. The title of the report. Default is "Clinical Trial Data Report".
author	Character string. The name of the report author. Default is "Author Name".
output_file	Character string. The file path where the HTML report will be saved. Default is "./report.html". You can specify a different path if needed.
color_palette	Character vector. A set of colors to be used in the report's visualizations. Default is a preset palette of 6 colors. You can provide your own color codes for customization.

theme   Character string. The Bootstrap theme for the HTML report. Default is `"cerulean"`.
                         Other options include `"default"`, `"journal"`, `"flatly"`, `"readable"`, `"spacelab"`,
                         `"united"`, `"cosmo"`, `"lumen"`, `"paper"`, `"sandstone"`, `"simplex"`, and `"yeti"`.

include_data_quality
                         Logical. Whether to include a data quality assessment section. Default is `TRUE`.
                         Set to `FALSE` if you want to skip this section.

include_interactive_plots
                         Logical. Whether to generate interactive plots using plotly. Default is `TRUE`. Set
                         to `FALSE` for static plots, which may be preferred for certain use cases.

custom_footer   Character string or `NULL`. A custom footer for the report. If `NULL` (default), a
                         standard footer crediting the ClinTrialX package is used.

Details

The function performs these key steps:

1. Package Management:

   - Checks for required packages and offers to install any that are missing.

   - Required packages: `rmarkdown`, `ggplot2`, `plotly`, `dplyr`, `lubridate`, `reactable`, `scales`,
     `RColorBrewer`, `htmltools`.

2. Report Generation:

   - Creates a temporary R Markdown file with the report content.

   - Includes an executive summary with key statistics.

   - Provides an interactive data table for easy exploration of the dataset.

3. Data Visualization:

   - Study Status Distribution: Bar chart showing the count of studies in each status.

   - Enrollment by Study Phase: Box plot displaying enrollment numbers across different study
     phases.

   - Study Duration Timeline: Scatter plot showing the relationship between study start dates and
     durations.

   - Funding Sources and Study Types: Stacked bar chart illustrating the proportion of study types
     for each funder type.

4. Optional Sections:

   - Data Quality Assessment: Bar chart showing the percentage of missing data for each variable
     (if enabled).

   - Interactive Plots: Uses plotly to create interactive versions of all plots (if enabled).

5. Report Finalization:

   - Renders the R Markdown file to an HTML report.

   - Cleans up temporary files.

Value

This function doesn't return a value, but generates an HTML report at the specified location. It prints a message with the path to the generated report upon successful completion.

Tips for Users

- Ensure your data frame has all required columns before using this function.
- Experiment with different themes to find the most suitable look for your report.
- If you encounter any package installation issues, you may need to install them manually.
- For large datasets, setting include_interactive_plots = FALSE may improve performance.
- Custom color palettes can be used to match your organization's branding.
- The generated report is self-contained and can be easily shared or published on the web.

See Also

https://www.indraneelchakraborty.com/clintrialx/ for more information about the Clin-TrialX package.

---

ctg_get_fields	*Query ClinicalTrials.gov API*

Description

This function sends a query to the ClinicalTrials.gov API and returns the results as a tibble. Users can specify various parameters to filter the results, and if a parameter is not provided, it will be omitted from the query.

Usage

```
ctg_get_fields(
  condition = NULL,
  location = NULL,
  title = NULL,
  intervention = NULL,
  status = NULL,
  page_size = 20
)
```

Arguments

condition   A character string specifying the medical condition to search for. This will filter the results to studies related to the given condition.

location   A character string specifying the location (e.g., city or country) to search in. This will filter the results to studies conducted in the specified location.

title	A character string specifying keywords to search for in study title. This will filter the results to studies with title that include the specified keywords.
intervention	A character string specifying the intervention or treatment to search for. This will filter the results to studies involving the specified intervention.
status	A character vector specifying the overall status of the studies. Valid values include:

- ACTIVE_NOT_RECRUITING - Studies that are actively conducting but not recruiting participants.
- COMPLETED - Studies that have completed all phases.
- ENROLLING_BY_INVITATION - Studies that are enrolling participants by invitation only.
- NOT_YET_RECRUITING - Studies that have not yet started recruiting.
- RECRUITING - Studies that are actively recruiting participants.
- SUSPENDED - Studies that are temporarily halted.
- TERMINATED - Studies that have been terminated before completion.
- WITHDRAWN - Studies that have been withdrawn before enrollment.
- AVAILABLE - Studies that are available.
- NO_LONGER_AVAILABLE - Studies that are no longer available.
- TEMPORARILY_NOT_AVAILABLE - Studies that are temporarily not available.
- APPROVED_FOR_MARKETING - Studies that have been approved for marketing.
- WITHHELD - Studies that have data withheld.
- UNKNOWN - Studies with an unknown status.

| page_size | An integer specifying the number of results per page. The default value is 20. The maximum allowed value is 1,000. If a value greater than 1,000 is specified, it will be coerced to 1,000. If not specified, the default value will be used. |

Details

This function can return up to 1,000 results.

The function constructs a query to the ClinicalTrials.gov API using the provided parameters. It supports filtering by condition, location, title keywords, intervention, and overall status. The function handles the API response, checks for errors, and parses the results into a tibble.

Value

A tibble containing the query results. Each row represents a study, and the columns correspond to the study details returned by the API.

Examples

```
# Query for studies related to "diabetes" in "Kolkata" with the status "RECRUITING"
ctg_get_fields(condition = "diabetes", location = "Kolkata",
                            status = "RECRUITING")
```

```
# Query for studies with "vaccine" in the title and the status "COMPLETED"
ctg_get_fields(title = "vaccine", status = "COMPLETED", page_size = 50)
```

---

ctg_get_nct   *Fetch Clinical Trial Data Based on NCT ID*

Description

Retrieves data for one or more clinical trials from the ClinicalTrials.gov API based on their NCT ID(s).

Usage

```
ctg_get_nct(nct_ids, fields = NULL)
```

Arguments

nct_ids	A character vector of one or more NCT IDs (e.g., "NCT04000165") for the clinical trials to fetch.
fields	A character vector specifying the fields to retrieve. If NULL (default), all available fields are fetched. If specified, it must be a subset of the available fields.

Details

This function allows you to specify one or more NCT IDs and optionally select specific fields of interest. It fetches the relevant data and returns it as a tibble.

The function constructs a request for each NCT ID, specifying the desired fields. It uses a progress bar to show the progress of fetching data for multiple trials. The data is returned as a tibble with columns corresponding to the requested fields. If any fetches fail or if the API response contains columns not requested, warnings will be issued.

Ensure that the `fields` parameter contains valid field names as specified in the guide below. Invalid fields will result in an error.

Value

A tibble containing the clinical trial data with columns matching the requested fields.

Field Names Guide

The following are the available fields you can request from ClinicalTrials.gov: NCT Number, Study Title, Study URL, Acronym, Study Status, Brief Summary, Study Results, Conditions, Interventions, Primary Outcome Measures, Secondary Outcome Measures, Other Outcome Measures, Sponsor, Collaborators, Sex, Age, Phases, Enrollment, Funder Type, Study Type, Study Design, Other IDs, Start Date, Primary Completion Date, Completion Date, First Posted, Results First Posted, Last Update Posted, Locations, Study Documents

Examples

```
# Fetch data for a single NCT ID
trial_data <- ctg_get_nct("NCT04000165")
trial_data

# Fetch data for multiple NCT IDs
multiple_trials <- ctg_get_nct(c("NCT04000165", "NCT04002440"))
multiple_trials

# Fetch data for multiple NCT IDs with specific fields
specific_fields <- ctg_get_nct(
  c("NCT04000165", "NCT04002440"),
  fields = c("NCT Number", "Study Title", "Study Status")
)
specific_fields
```

---

hello   *Print a Welcome Message*

Description

This function returns a welcome message for ClinTrialX.

Usage

```
hello()
```

Value

A character string containing the welcome message.

Examples

```
hello()
```

---

version_info   *Get API Version Information*

Description

This function retrieves version information from specified clinical trials API sources.

Usage

```
version_info(source = "clinicaltrials.gov")
```

Arguments

source   A character string specifying the source to query. Currently, "clinicaltrials.gov" and "aact" are supported.

Value

A list containing API version and data timestamp for clinicaltrials.gov, or NULL for aact with a message printed.

References

ClinicalTrials.gov API - https://clinicaltrials.gov/api/v2/version AACT - https://aact.ctti-clinicaltrials.org/release_notes

Examples

```
version_info()
version_info("clinicaltrials.gov")
version_info("aact")
```

Index