

Package ‘pubmedR’

April 15, 2026

Title Gathering Metadata About Publications, Grants, Clinical Trials
from 'PubMed' Database

Version 1.0.0

Description A set of tools to extract bibliographic content from 'PubMed' database using 'NCBI' REST API <<https://www.ncbi.nlm.nih.gov/home/develop/api/>>. It includes functions to search, download, and convert 'PubMed' bibliographic records into data frames compatible with the 'bibliometrix' package. Features include programmatic query building, batch downloading by PMID, citation enrichment via 'NCBI' E-Link, and robust error handling with automatic retry logic.

License GPL-3

URL <https://github.com/massimoaria/pubmedR>

BugReports <https://github.com/massimoaria/pubmedR/issues>

Encoding UTF-8

Imports rentrez, XML

Suggests bibliometrix, knitr, rmarkdown, testthat (>= 3.0.0), withr

Config/testthat/edition 3

RoxygenNote 7.3.3

VignetteBuilder knitr

NeedsCompilation no

Author Massimo Aria [aut, cre] (ORCID:
<<https://orcid.org/0000-0002-8517-9411>>)

Maintainer Massimo Aria <massimo.aria@gmail.com>

Repository CRAN

Date/Publication 2026-04-15 07:10:08 UTC

Contents

pmApi2df	2
pmApiRequest	3

pmCitedBy	4
pmCollect	5
pmEnrichCitations	7
pmFetchById	8
pmQueryBuild	9
pmQueryTotalCount	11
pmReferences	12
Index	14

pmApi2df

Convert xml PubMed bibliographic data into a dataframe

Description

It converts PubMed data, downloaded using Entrez API, into a dataframe

Usage

```
pmApi2df(P, format = "bibliometrix")
```

Arguments

P	is a list following the xml PubMed structure, downloaded using the function pmApiRequest.
format	is a character. If format = "bibliometrix" data will be converted in the bibliometrix compatible data format. If format = "raw" data will save in a dataframe without any other data editing procedure.

Value

a dataframe containing bibliographic records.

To obtain a free access to NCBI API, please visit: <https://pmc.ncbi.nlm.nih.gov/tools/developers/>

To obtain more information about how to write a NCBI search query, please visit: <https://pubmed.ncbi.nlm.nih.gov/help/#search-tags>

See Also

[pmApiRequest](#)

[pmQueryTotalCount](#)

Examples

```
# Example: Querying a collection of publications

query <- "bibliometric*[Title/Abstract] AND english[LA]
        AND Journal Article[PT] AND 2000:2020[DP]"
D <- pmApiRequest(query = query, limit = 100, api_key = NULL)
M <- pmApi2df(D)
```

pmApiRequest	<i>Gather bibliographic content from PubMed database using NCBI entrez APIs</i>
--------------	---

Description

It gathers metadata about publications from the NCBI PubMed database. The use of NCBI PubMed APIs is entirely free, and doesn't necessarily require an API key. The function `pmApiRequest` queries NCBI PubMed using an entrez query formulated through the Entrez query language or the helper function `pmQueryBuilder`.

Usage

```
pmApiRequest(query, limit, api_key = NULL, batch_size = 200)
```

Arguments

query	is a character. It contains a search query formulated using the Entrez query language.
limit	is numeric. It indicates the max number of records to download.
api_key	is a character. It contains a valid API key for the NCBI E-utilities. Default is <code>api_key=NULL</code> . The API key can also be set via the environment variable <code>PUBMED_API_KEY</code> or <code>ENTREZ_KEY</code> .
batch_size	is numeric. The number of records to download per API request. Default is 200.

Details

Official API documentation is <https://www.ncbi.nlm.nih.gov/books/NBK25500/>.

Value

a list D composed by 5 objects:

data	It is the xml-structured list containing the bibliographic metadata collection downloaded from the P
query	It a character object containing the original query formulated by the user.
query_translation	It a character object containing the query, translated by the NCBI Automatic Terms Translation syst

records_downloaded It is an integer object indicating the total number of records downloaded and stored in "data".
 total_count It is an integer object indicating the total number of records matching the query (stored in the "query")

To obtain a free access to NCBI API, please visit: <https://pmc.ncbi.nlm.nih.gov/tools/developers/>

To obtain more information about how to write a NCBI search query, please visit: <https://pubmed.ncbi.nlm.nih.gov/help/#search-tags>

See Also

[pmQueryTotalCount](#)

[pmApi2df](#)

[pmQueryBuild](#)

Examples

```
query <- "bibliometric*[Title/Abstract] AND english[LA]
        AND Journal Article[PT] AND 2000:2020[DP]"
D <- pmApiRequest(query = query, limit = 100, api_key = NULL)
```

pmCitedBy

Find articles that cite a given PubMed article

Description

It retrieves the PMIDs of articles that cite a given PubMed article, using the NCBI E-Link service (PubMed Cited by).

Usage

```
pmCitedBy(pmid, api_key = NULL)
```

Arguments

pmid is a character or numeric. A single PubMed identifier (PMID).
 api_key is a character. It contains a valid API key for the NCBI E-utilities. Default is api_key=NULL. The API key can also be set via the environment variable PUBMED_API_KEY or ENTREZ_KEY.

Details

This function uses the NCBI E-Link endpoint with linkname "pubmed_pubmed_citedin" to find articles in PubMed that cite the given article.

Note: Citation data in PubMed is based on PubMed Central (PMC) and may not be as comprehensive as commercial citation databases (e.g. Web of Science, Scopus).

Value

a list containing:

pmid	The queried PMID.
cited_by	A character vector of PMIDs that cite the queried article.
count	The number of citing articles found.

See Also

[pmReferences](#)

[pmFetchById](#)

Examples

```
# Find articles that cite PMID 25824007
cites <- pmCitedBy(pmid = "25824007")
cites$count
cites$cited_by
```

pmCollect

Collect and process PubMed bibliographic data in one step

Description

A convenience wrapper that executes the full pubmedR workflow: query building, record count check, metadata download, conversion to data frame, and (optionally) citation enrichment via NCBI E-Link.

Usage

```
pmCollect(
  query = NULL,
  terms = NULL,
  fields = "Title/Abstract",
  language = NULL,
  pub_type = NULL,
  date_range = NULL,
  mesh_terms = NULL,
  limit = 2000,
  enrich = FALSE,
  format = "bibliometrix",
  api_key = NULL,
  batch_size = 200,
  verbose = TRUE
)
```

Arguments

query	is a character. A PubMed search query in Entrez syntax. Alternatively, if terms is provided, the query is built automatically using pmQueryBuild and this argument is ignored.
terms	is a character or character vector or NULL. Search terms passed to pmQueryBuild . When provided, a query is built automatically and the query argument is ignored. Default is NULL.
fields	is a character or character vector. PubMed search tags used when building the query from terms. Default is "Title/Abstract".
language	is a character or NULL. Language filter for query building. Default is NULL.
pub_type	is a character or NULL. Publication type filter for query building. Default is NULL.
date_range	is a character vector of length 2 or NULL. Date range in format c("YYYY", "YYYY"). Default is NULL.
mesh_terms	is a character or character vector or NULL. MeSH terms for query building. Default is NULL.
limit	is numeric. Maximum number of records to download. Default is 2000.
enrich	is logical. If TRUE, citation counts (TC) and cited references (CR) are added via pmEnrichCitations . Default is FALSE because enrichment makes 2 API calls per article and can be slow for large collections.
format	is a character. Output format passed to pmApi2df . Either "bibliometrix" (default) or "raw".
api_key	is a character or NULL. NCBI API key. Can also be set via the environment variable PUBMED_API_KEY or ENTREZ_KEY. Default is NULL.
batch_size	is numeric. Records per API request. Default is 200.
verbose	is logical. If TRUE (default), prints progress messages.

Details

This function chains together the core pubmedR functions in the recommended order:

1. **Query:** If terms is provided, builds the query with [pmQueryBuild](#); otherwise uses the query string directly.
2. **Count:** Checks the total number of matching records with [pmQueryTotalCount](#).
3. **Download:** Fetches metadata with [pmApiRequest](#).
4. **Convert:** Transforms XML to a data frame with [pmApi2df](#).
5. **Enrich** (optional): Adds citation data with [pmEnrichCitations](#).

Value

a data frame containing bibliographic records, compatible with the [bibliometrix](#) package when `format = "bibliometrix"`.

See Also

[pmQueryBuild](#), [pmQueryTotalCount](#), [pmApiRequest](#), [pmApi2df](#), [pmEnrichCitations](#)

Examples

```
# Using a raw query string
M <- pmCollect(
  query = "bibliometric*[Title/Abstract] AND english[LA] AND 2020:2024[DP]",
  limit = 50
)

# Using the query builder parameters
M <- pmCollect(
  terms = "bibliometric*",
  language = "english",
  pub_type = "Journal Article",
  date_range = c("2020", "2024"),
  limit = 50
)

# With citation enrichment (slower, requires extra API calls)
M <- pmCollect(
  terms = "bibliometric*",
  date_range = c("2023", "2024"),
  limit = 10,
  enrich = TRUE
)
```

pmEnrichCitations *Enrich a PubMed dataframe with citation data*

Description

It adds cited references (CR field) and citation counts (TC field) to a dataframe created by [pmApi2df](#), using NCBI E-Link data.

Usage

```
pmEnrichCitations(df, api_key = NULL)
```

Arguments

df is a dataframe. A bibliometric dataframe produced by [pmApi2df](#).

api_key is a character. It contains a valid API key for the NCBI E-utilities. Default is `api_key=NULL`. The API key can also be set via the environment variable `PUBMED_API_KEY` or `ENTREZ_KEY`.

Details

This function iterates over each record in the dataframe and queries NCBI E-Link to retrieve: (1) The PMIDs of references cited by each article (populates CR field), and (2) The count of articles citing each article (populates TC field).

Note: This process makes two API calls per article and can be slow for large datasets. An API key is strongly recommended.

Value

The input dataframe with updated CR (Cited References) and TC (Times Cited) fields.

See Also

[pmCitedBy](#)
[pmReferences](#)
[pmApi2df](#)

Examples

```
query <- "bibliometric*[Title/Abstract] AND english[LA]
          AND Journal Article[PT] AND 2000:2020[DP]"
D <- pmApiRequest(query = query, limit = 10, api_key = NULL)
M <- pmApi2df(D)
M <- pmEnrichCitations(M)
```

pmFetchById

Fetch PubMed records by PMID

Description

It downloads metadata for a set of PubMed articles identified by their PMID (PubMed Identifier). This is useful for retrieving specific known articles, updating existing datasets, or downloading records identified through other sources.

Usage

```
pmFetchById(pmids, api_key = NULL, batch_size = 200)
```

Arguments

pmids	is a character or numeric vector. A vector of PubMed identifiers (PMIDs).
api_key	is a character. It contains a valid API key for the NCBI E-utilities. Default is api_key=NULL. The API key can also be set via the environment variable PUBMED_API_KEY or ENTREZ_KEY.
batch_size	is numeric. The number of records to download per API request. Default is 200.

Details

The function uses the NCBI E-utilities efetch endpoint to retrieve records directly by their PMIDs, without requiring a search query. Records are downloaded in batches to respect API rate limits.

The output is compatible with [pmApi2df](#) for conversion to a dataframe.

Value

a list following the same structure as [pmApiRequest](#) output, containing:

data	The xml-structured list containing the bibliographic metadata.
query	A character string describing the PMID-based query.
query_translation	Same as query for PMID-based searches.
records_downloaded	The total number of records downloaded.
total_count	The total number of PMIDs requested.

See Also

[pmApiRequest](#)

[pmApi2df](#)

Examples

```
# Download specific articles by PMID
pmids <- c("34813985", "34813456", "34812345")
D <- pmFetchById(pmids = pmids)
M <- pmApi2df(D)
```

pmQueryBuild

Build a PubMed search query programmatically

Description

It helps to build a valid PubMed search query using the Entrez query language, combining multiple search terms with Boolean operators.

Usage

```
pmQueryBuild(
  terms = NULL,
  fields = "Title/Abstract",
  language = NULL,
  pub_type = NULL,
  date_range = NULL,
  mesh_terms = NULL,
```

```

    author = NULL,
    journal = NULL,
    operator = "AND"
)

```

Arguments

terms	is a character or character vector. Search terms to look for in title and abstract fields.
fields	is a character or character vector. PubMed search tags to apply. Default is <code>c("Title/Abstract")</code> . Common fields include: "Title/Abstract", "Title", "Author", "MeSH Terms", "Affiliation", "Journal".
language	is a character or NULL. Language filter (e.g. "english", "french"). Default is NULL (no filter).
pub_type	is a character or NULL. Publication type filter (e.g. "Journal Article", "Review", "Clinical Trial"). Default is NULL (no filter).
date_range	is a character vector of length 2 or NULL. Date range in format <code>c("YYYY", "YYYY")</code> or <code>c("YYYY/MM/DD", "YYYY/MM/DD")</code> . Default is NULL (no filter).
mesh_terms	is a character or character vector or NULL. MeSH (Medical Subject Headings) terms. Default is NULL (no filter).
author	is a character or character vector or NULL. Author names. Default is NULL.
journal	is a character or character vector or NULL. Journal names or abbreviations. Default is NULL.
operator	is a character. Boolean operator to combine multiple terms. One of "AND", "OR". Default is "AND".

Details

The function constructs a query string compatible with NCBI's Entrez search system. Multiple terms within the same parameter are combined with the specified operator, while different parameters (terms, language, pub_type, etc.) are combined with AND.

For more information about PubMed search tags, visit: <https://pubmed.ncbi.nlm.nih.gov/help/#search-tags>

Value

a character string containing the formatted PubMed query.

See Also

[pmQueryTotalCount](#)

[pmApiRequest](#)

Examples

```
# Simple query
q <- pmQueryBuild(terms = "bibliometrics", language = "english",
                  pub_type = "Journal Article", date_range = c("2000", "2023"))

# Multiple terms
q <- pmQueryBuild(terms = c("machine learning", "deep learning"),
                  operator = "OR", language = "english")

# MeSH terms query
q <- pmQueryBuild(mesh_terms = "COVID-19", pub_type = "Review",
                  date_range = c("2020", "2024"))

# Author search
q <- pmQueryBuild(terms = "bibliometrics", author = "Aria M")
```

pmQueryTotalCount	<i>Count the number of documents returned by a query</i>
-------------------	--

Description

It counts the number of documents that a query returns from the NCBI PubMed database.

Usage

```
pmQueryTotalCount(query, api_key = NULL)
```

Arguments

query	is a character. It contains a search query formulated using the Entrez query language.
api_key	is a character. It contains a valid API key for the NCBI E-utilities. Default is api_key=NULL. The use of NCBI PubMed APIs is entirely free, and doesn't necessarily require an API key. The API key can also be set via the environment variable PUBMED_API_KEY or ENTREZ_KEY.

Value

a list. It contains three objects:

total_count	The total number of records returned by the query
query_translation	The query translation by the NCBI Automatic Terms Translation system
web_history	The web history object. The NCBI provides search history features, which is useful for dealing with la

To obtain a free access to NCBI API, please visit: <https://pmc.ncbi.nlm.nih.gov/tools/developers/>

See Also[pmApiRequest](#)[pmApi2df](#)**Examples**

```
query <- "bibliometric*[Title/Abstract] AND english[LA]
        AND Journal Article[PT] AND 2000:2020[DP]"
D <- pmQueryTotalCount(query = query, api_key = NULL)
```

pmReferences

Find references cited by a given PubMed article

Description

It retrieves the PMIDs of articles that are cited by (referenced in) a given PubMed article, using the NCBI E-Link service.

Usage

```
pmReferences(pmid, api_key = NULL)
```

Arguments

pmid	is a character or numeric. A single PubMed identifier (PMID).
api_key	is a character. It contains a valid API key for the NCBI E-utilities. Default is api_key=NULL. The API key can also be set via the environment variable PUBMED_API_KEY or ENTREZ_KEY.

Details

This function uses the NCBI E-Link endpoint with linkname "pubmed_pubmed_refs" to find articles in PubMed that are referenced by the given article.

Note: Reference data is extracted from PubMed Central (PMC) full-text articles and is only available when the full text is deposited in PMC. Not all PubMed articles have reference data available.

Value

a list containing:

pmid	The queried PMID.
references	A character vector of PMIDs referenced by the queried article.
count	The number of references found.

See Also

[pmCitedBy](#)
[pmFetchById](#)

Examples

```
# Find references of PMID 25824007
refs <- pmReferences(pmid = "25824007")
refs$count
refs$references
```

Index

`pmApi2df`, [2](#), [4](#), [6–9](#), [12](#)
`pmApiRequest`, [2](#), [3](#), [6](#), [7](#), [9](#), [10](#), [12](#)
`pmCitedBy`, [4](#), [8](#), [13](#)
`pmCollect`, [5](#)
`pmEnrichCitations`, [6](#), [7](#), [7](#)
`pmFetchById`, [5](#), [8](#), [13](#)
`pmQueryBuild`, [3](#), [4](#), [6](#), [7](#), [9](#)
`pmQueryTotalCount`, [2](#), [4](#), [6](#), [7](#), [10](#), [11](#)
`pmReferences`, [5](#), [8](#), [12](#)