Package ‘europepmc’

October 13, 2022

Title  R Interface to the Europe PubMed Central RESTful Web Service

Version  0.4.1

License  GPL-3

Date  2021-08-20


BugReports  https://github.com/ropensci/europepmc/issues

Description  An R Client for the Europe PubMed Central RESTful Web Service (see <https://europepmc.org/RestfulWebService> for more information). It gives access to both metadata on life science literature and open access full texts. Europe PMC indexes all PubMed content and other literature sources including Agricola, a bibliographic database of citations to the agricultural literature, or Biological Patents. In addition to bibliographic metadata, the client allows users to fetch citations and reference lists. Links between life-science literature and other EBI databases, including ENA, PDB or ChEMBL are also accessible. No registration or API key is required. See the vignettes for usage examples.

LazyLoad  yes

VignetteBuilder  knitr

Depends  R (>= 3.00)

Imports  httr, jsonlite, plyr, dplyr, progress, urltools, purrr, xml2, tibble, tidyr, rlang

RoxygenNote  7.1.1

Suggests  testthat, knitr, rmarkdown, ggplot2

Encoding  UTF-8

NeedsCompilation  no

Author  Najko Jahn [aut, cre, cph], Maëlle Salmon [ctb]

Maintainer  Najko Jahn <najko.jahn@gmail.com>

Repository  CRAN

Date/Publication  2021-09-02 07:20:05 UTC
**R topics documented:**

epmc_annotations_by_id .......................... 2  
epmc_citations ................................. 3  
epmc_db ........................................ 4  
epmc_db_count ................................. 6  
epmc_details .................................... 7  
epmc_ftxt ....................................... 8  
epmc_ftxt_book .................................. 9  
epmc_hits ....................................... 9  
epmc_hits_trend ............................... 10  
epmc_lablinks .................................. 11  
epmc_lablinks_count ......................... 13  
epmc_profile ..................................... 14  
epmc.refs ....................................... 14  
epmc_search ..................................... 15  
epmc_search ..................................... 17  
epmc_search_by_doi ........................................ 18  
epmc_search_by_doi_ ............................ 19  
europepmc ...................................... 19  

---

epmc_annotations_by_id

*Get annotations by article*

**Description**

Retrieve text-mined annotations contained in abstracts and open access full-text articles.

**Usage**

epmc_annotations_by_id(ids = NULL)

**Arguments**

- **ids**, character vector with publication identifiers following the structure "source:ext_id", e.g. "MED:28585529".

**Value**

returns text-mined annotations in a tidy format with the following variables

- **source** Publication data source
- **ext_id** Article Identifier
- **pmcid** PMCID that locates full-text in Pubmed Central
- **prefix** Text snipped found before the annotation
### Examples

```
## Not run:
  annotations_by_id("MED:28585529")
# multiple ids
  annotations_by_id(c("MED:28585529", "PMC:PMC1664601"))

## End(Not run)
```

---

**epmc_citations**

Get citations for a given publication

**Description**

Finds works that cite a given publication.

**Usage**

```
epmc_citations(ext_id = NULL, data_src = "med", limit = 100, verbose = TRUE)
```

**Arguments**

- **ext_id** character, publication identifier
- **data_src** character, data source, by default Pubmed/MedLine index will be searched. The following three letter codes represent the sources Europe PubMed Central supports:
  - **agr** Agricola is a bibliographic database of citations to the agricultural literature created by the US National Agricultural Library and its co-operators.
  - **cba** Chinese Biological Abstracts
  - **ctx** CiteXplore
  - **eth** EthOs Theses, i.e. PhD theses (British Library)
  - **hir** NHS Evidence
  - **med** PubMed/Medline NLM
  - **nbk** Europe PMC Book metadata
epmc_db

Retrieve external database entities referenced in a given publication

Description

This function returns EBI database entities referenced in a publication from Europe PMC RESTful Web Service.

Usage

epmc_db(
    ext_id = NULL,
    data_src = "med",
    db = NULL,
    limit = 100,
    verbose = TRUE
)

Arguments

ext_id character, publication identifier
data_src character, data source, by default Pubmed/MedLine index will be searched. The following three letter codes represent the sources Europe PubMed Central supports:

agr Agricola is a bibliographic database of citations to the agricultural literature created by the US National Agricultural Library and its co-operators.
cba Chinese Biological Abstracts
ctx CiteXplore
eth EthOs Theses, i.e. PhD theses (British Library)
epmc_db

hir  NHS Evidence
med  PubMed/Medline NLM
nbk  Europe PMC Book metadata
pat  Biological Patents
pmc  PubMed Central

db  character, specify database:

'ARXPR'  Array Express, a database of functional genomics experiments
'CHEBI'  a database and ontology of chemical entities of biological interest
'CHEMBL'  a database of bioactive drug-like small molecules
'EMBL'  now ENA, provides a comprehensive record of the world's nucleotide sequencing information
'INTACT'  provides a freely available, open source database system and analysis tools for molecular interaction data
'INTERPRO'  provides functional analysis of proteins by classifying them into families and predicting domains and important sites
'OMIM'  a comprehensive and authoritative compendium of human genes and genetic phenotypes
'PDB'  European resource for the collection, organisation and dissemination of data on biological macromolecular structures
'UNIPROT'  comprehensive and freely accessible resource of protein sequence and functional information
'PRIDE'  PRIDE Archive - proteomics data repository

limit  integer, number of results. By default, this function returns 100 records.
verbose  logical, print some information on what is going on.

Value

Cross-references as data.frame

Examples

## Not run:
epmc_db("12368864", db = "uniprot", limit = 150)
epmc_db("25249410", db = "embl")
epmc_db("14756321", db = "uniprot")
epmc_db("11805837", db = "pride")

## End(Not run)
epmc_db_count

Retrieve the number of database links from Europe PMC publication database

Description

This function returns the number of EBI database links associated with a publication.

Usage

epmc_db_count(ext_id = NULL, data_src = "med")

Arguments

ext_id character, publication identifier
data_src character, data source, by default Pubmed/MedLine index will be searched.

Details

Europe PMC supports cross-references between literature and the following databases:

'ARXPR' Array Express, a database of functional genomics experiments
'CHEBI' a database and ontology of chemical entities of biological interest
'CHEMBL' a database of bioactive drug-like small molecules
'EMBL' now ENA, provides a comprehensive record of the world’s nucleotide sequencing information
'INTACT' provides a freely available, open source database system and analysis tools for molecular interaction data
'INTERPRO' provides functional analysis of proteins by classifying them into families and predicting domains and important sites
'OMIM' a comprehensive and authoritative compendium of human genes and genetic phenotypes
'PDB' European resource for the collection, organisation and dissemination of data on biological macromolecular structures
'UNIPROT' comprehensive and freely accessible resource of protein sequence and functional information
'PRIDE' PRIDE Archive - proteomics data repository

Value

data.frame with counts for each database
epmc_details

Examples

## Not run:
epmc_db_count(ext_id = "10779411")
epmc_db_count(ext_id = "PMC3245140", data_src = "PMC")

## End(Not run)

---

epmc_details  Get details for individual records

Description

This function returns parsed metadata for a given publication ID including abstract, full text links, author details including ORCID and affiliation, MeSH terms, chemicals, grants.

Usage

epmc_details(ext_id = NULL, data_src = "med")

Arguments

- **ext_id**: character, publication identifier
- **data_src**: character, data source, by default Pubmed/MedLine index will be searched. Other sources Europe PubMed Central supports are:
  - agr: Agricola is a bibliographic database of citations to the agricultural literature created by the US National Agricultural Library and its co-operators.
  - cba: Chinese Biological Abstracts
  - ctx: CiteXplore
  - eth: EthOs Theses, i.e. PhD theses (British Library)
  - hir: NHS Evidence
  - med: PubMed/Medline NLM
  - pat: Biological Patents
  - pmc: PubMed Central
  - ppr: Preprint records

Value

- list of data frames
Examples

```r
## Not run:
ePMC_details(ext_id = "26980001")
ePMC_details(ext_id = "24270414")

# PMC record
ePMC_details(ext_id = "PMC4747116", data_src = "pmc")

# Other sources:
# Agricolo
ePMC_details("IND43783977", data_src = "agr")
# Biological Patents
ePMC_details("EP2412369", data_src = "pat")
# Chinese Biological Abstracts
ePMC_details("583843", data_src = "cba")
# CiteXplore
ePMC_details("C6802", data_src = "ctx")
# NHS Evidence
ePMC_details("338638", data_src = "hir")
# Theses
ePMC_details("409323", data_src = "eth")
# Preprint
ePMC_details("PPR158112", data_src = "ppr")

## End(Not run)
```

description

This function loads full texts into R. Full texts are in XML format and are only provided for the Open Access subset of Europe PMC.

Usage

```r
ePMC_ftxt(ext_id = NULL)
```

Arguments

- `ext_id` character, PMCID. All full text publications have external IDs starting 'PMC_'

Value

`xml_document`
epmc_ftxt_book

Examples

```r
## Not run:
epmc_ftxt("PMC3257301")
epmc_ftxt("PMC3639880")

## End(Not run)
```

---

epmc_ftxt_book  Fetch Europe PMC books

Description

Use this function to retrieve book XML formatted full text for the Open Access subset of the Europe PMC bookshelf.

Usage

```r
epmc_ftxt_book(ext_id = NULL)
```

Arguments

- `ext_id` character, publication identifier. All book full texts are accessible either by the PMID or the 'NBK' book number.

Value

xml_document

Examples

```r
## Not run:
epmc_ftxt_book("NBK32884")

## End(Not run)
```

---

epmc_hits  Get search result count

Description

Search over Europe PMC and retrieve the number of results found

Usage

```r
epmc_hits(query = NULL, ...)
```
epmc_hits_trend

Get the yearly number of hits for a query and the total yearly number of hits for a given period

Description

Get the yearly number of hits for a query and the total yearly number of hits for a given period

Usage

epmc_hits_trend(query, synonym = TRUE, data_src = "med", period = 1975:2016)

Arguments

query query in the Europe PMC syntax
synonym logical, synonym search. If TRUE, synonym terms from MeSH terminology and the UniProt synonym list are queried, too. Disabled by default.
data_src character, data source, by default Pubmed/Medline index (med) will be searched. The following three letter codes represent the sources, which are currently supported:
  agr  Agricola is a bibliographic database of citations to the agricultural literature created by the US National Agricultural Library and its co-operators.
  cba  Chinese Biological Abstracts
  ctx  CiteXplore
  eth  EthOs Theses, i.e. PhD theses (British Library)
  hir  NHS Evidence
  med  PubMed/Medline NLM
  nbk  Europe PMC Book metadata
  pat  Biological Patents

See Also
epmc_search

Examples

## Not run:
epmc_hits('abstract:"burkholderia pseudomallei"')
epmc_hits('AUTHORID:"0000-0002-7635-3473"')

## End(Not run)
**epmc** PubMed Central

**ppr** Preprint records

**period** a vector of years (numeric) over which to perform the search

**Details**

A similar function was used in [https://masalmon.eu/2017/05/14/evergreenreviewgraph/](https://masalmon.eu/2017/05/14/evergreenreviewgraph/) where it was advised to not plot no. of hits over time for a query, but to normalize it by the total no. of hits.

**Value**

a data.frame (dplyr tbl_df) with year, total number of hits (all_hits) and number of hits for the query (query_hits)

**Examples**

```r
## Not run:
# aspirin as query
ePMC_hits_trend('aspirin', period = 2006:2016, synonym = FALSE)
# link to cran packages in reference lists
ePMC_hits_trend('REF:cran.r-project.org*', period = 2006:2016, synonym = FALSE)
# more complex with publication type review
ePMC_hits_trend('((REF:cran.r-project.org*) AND (PUB_TYPE:Review OR PUB_TYPE:review-article))',
period = 2006:2016, synonym = FALSE)
## End(Not run)
```

---

**epmc_lablinks** Get links to external sources

**Description**

With the External Link services, Europe PMC allows third parties to publish links from Europe PMC to other webpages or tools. Current External Link providers, which can be selected through Europe PMC’s advanced search, include Wikipedia, Dryad Digital Repository or other open services. For more information, see [http://europepmc.org/labslink](http://europepmc.org/labslink).

**Usage**

```r
ePMC_lablinks(
  ext_id = NULL,
  data_src = "med",
  lab_id = NULL,
  limit = 100,
  verbose = TRUE
)
```
Arguments

ext_id: publication identifier

data_src: data source, by default Pubmed/MedLine index will be searched. The following three letter codes represent the sources Europe PubMed Central supports:

agr: Agricola is a bibliographic database of citations to the agricultural literature created by the US National Agricultural Library and its co-operators.

cba: Chinese Biological Abstracts

ctx: CiteXplore

eth: EthOs Theses, i.e. PhD theses (British Library)

hir: NHS Evidence

med: PubMed/Medline NLM

nbk: Europe PMC Book metadata

pat: Biological Patents

pmc: PubMed Central

lab_id: character vector, identifiers of the external link service. Use Europe PMC’s advanced search form to find ids.

limit: Number of records to be returned. By default, this function returns 100 records.

verbose: print information about what’s going on

Value

Links found as nested data_frame

Examples

## Not run:
# Fetch links
e PMC lablinks("24007304")
# Link to Altmetric (lab_id = "1562")
e PMC lablinks("25389392", lab_id = "1562")

# Links to Wikipedia
e PMC lablinks("24007304", lab_id = "1507")

# Link to full text copy archived through the institutional repo of Bielefeld University
e PMC lablinks("12736239", lab_id = "1056")

## End(Not run)
epmc_lablinks_count  

Summarise links to external sources

Description

With the External Link services, Europe PMC allows third parties to publish links from Europe PMC to other webpages or tools. Current External Link providers, which can be selected through Europe PMC’s advanced search, include Wikipedia, Dryad Digital Repository or the institutional repo of Bielefeld University. For more information, see [http://europepmc.org/labslink](http://europepmc.org/labslink).

Usage

epmc_lablinks_count(ext_id = NULL, data_src = "med")

Arguments

- **ext_id**: publication identifier
- **data_src**: data source, by default Pubmed/MedLine index will be searched. The following three letter codes represent the sources Europe PubMed Central supports:
  - **agr**: Agricola is a bibliographic database of citations to the agricultural literature created by the US National Agricultural Library and its co-operators.
  - **cba**: Chinese Biological Abstracts
  - **ctx**: CiteXplore
  - **eth**: EthOs Theses, i.e. PhD theses (British Library)
  - **hir**: NHS Evidence
  - **med**: PubMed/Medline NLM
  - **nbk**: Europe PMC Book metadata
  - **pat**: Biological Patents
  - **pmc**: PubMed Central

Value

data.frame with counts for each database

Examples

```r
# Not run:
epmc_lablinks_count("24023770")
epmc_lablinks_count("PMC3986813", data_src = "pmc")

# End(Not run)
```
epmc_profile  
Obtain a summary of hit counts

Description

This function returns the number of results found for your query, and breaks it down to the various publication types, data sources, and subsets Europe PMC provides.

Usage

epmc_profile(query = NULL, synonym = TRUE)

Arguments

query  
character, search query. For more information on how to build a search query, see [http://europepmc.org/Help](http://europepmc.org/Help)

synonym  
logical, synonym search. If TRUE, synonym terms from MeSH terminology and the UniProt synonym list are queried, too. Enabled by default.

Examples

```r
## Not run:
epmc_profile('malaria')
# use field search, e.g. query materials and reference section for
# mentions of "ropensci"
epmc_profile(METHODS:"ropensci")
## End(Not run)
```

epmc_refs  
Get references for a given publication

Description

This function retrieves all the works listed in the bibliography of a given article.

Usage

epmc.refs(ext_id = NULL, data_src = "med", limit = 100, verbose = TRUE)
epmc_search

Arguments

- ext_id: character, publication identifier
- data_src: character, data source, by default Pubmed/MedLine index will be searched. The following three letter codes represent the sources Europe PubMed Central supports:
  - agr: Agricola is a bibliographic database of citations to the agricultural literature created by the US National Agricultural Library and its co-operators.
  - cba: Chinese Biological Abstracts
  - ctx: CiteXplore
  - eth: EthOs Theses, i.e. PhD theses (British Library)
  - hir: NHS Evidence
  - med: PubMed/Medline NLM
  - nbk: Europe PMC Book metadata
  - pat: Biological Patents
  - pmc: PubMed Central
- limit: integer, number of results. By default, this function returns 100 records.
- verbose: logical, print some information on what is going on.

Value

returns reference section as tibble

Examples

```r
## Not run:
epmc.refs("PMC3166943", data_src = "pmc")
epmc.refs("25378340")
epmc.refs("21753913")
## End(Not run)
```

description

This is the main function to search Europe PMC RESTful Web Service (http://europepmc.org/RestfulWebService). It fully supports the comprehensive Europe PMC query language. Simply copy & paste your query terms to R. To get familiar with the Europe PMC query syntax, check the Advanced Search Query Builder https://europepmc.org/advancesearch.
Usage

epmc_search(
  query = NULL,
  output = "parsed",
  synonym = TRUE,
  verbose = TRUE,
  limit = 100,
  sort = NULL
)

Arguments

query character, search query. For more information on how to build a search query, see http://europepmc.org/Help

output character, what kind of output should be returned. One of 'parsed', 'id_list' or 'raw'. As default, parsed key metadata will be returned as data.frame. 'id_list' returns a list of IDs and sources. Use 'raw' to get full metadata as list. Please be aware that these lists can become very large.

synonym logical, synonym search. If TRUE, synonym terms from MeSH terminology and the UniProt synonym list are queried, too. In order to replicate results from the website, with the Rest API you need to turn synonyms ON!

verbose logical, print progress bar. Activated by default.

limit integer, limit the number of records you wish to retrieve. By default, 100 are returned.

sort character, relevance ranking is used by default. Use sort = 'cited' for sorting by the number of citations, or sort = 'date' by the most recent publications.

Value
tibble

See Also

http://europepmc.org/Help

Examples

## Not run:
#Search articles for 'Gabi-Kat'
my.data <- epmc_search(query='Gabi-Kat')

#Get article metadata by DOI
my.data <- epmc_search(query = 'DOI:10.1007/bf00197367')

#Get article metadata by PubMed ID (PMID)
my.data <- epmc_search(query = 'EXT_ID:22246381')

#Get only PLOS Genetics article with EMBL database references
my.data <- epmc_search(query = 'ISSN:1553-7404 HAS_EMBL:y')
# Limit search to 250 PLOS Genetics articles
my.data <- epmc_search(query = 'ISSN:1553-7404', limit = 250)

# exclude MeSH synonyms in search
my.data <- epmc_search(query = 'aspirin', synonym = FALSE)

# get 100 most cited articles from PLOS ONE published in 2014
epmc_search(query = '(ISSN:1932-6203) AND FIRST_PDATE:2014', sort = 'cited')

# print number of records found
attr(my.data, "hit_count")

# change output

## End(Not run)

---

**Description**

In general, use `epmc_search` instead. It calls this function, calling all pages within the defined limit.

**Usage**

```r
epmc_search_(
  query = NULL,
  limit = 100,
  output = "parsed",
  page_token = NULL,
  ...
)
```

**Arguments**

- **query**: character, search query. For more information on how to build a search query, see [http://europepmc.org/Help](http://europepmc.org/Help)
- **limit**: integer, limit the number of records you wish to retrieve. By default, 25 are returned.
- **output**: character, what kind of output should be returned. One of `parsed`, `id_list` or `raw` As default, parsed key metadata will be returned as data.frame. `id_list` returns a list of IDs and sources. Use `raw` to get full metadata as list. Please be aware that these lists can become very large.
- **page_token**: cursor marking the page
- **...**: further params from `epmc_search`
epmc_search_by_doi  Search Europe PMC by DOIs

Description

Look up DOIs indexed in Europe PMC and get metadata back.

Usage

epmc_search_by_doi(doi = NULL, output = "parsed")

Arguments

doi, character vector containing DOI names.
output character, what kind of output should be returned. One of 'parsed', 'id_list' or 'raw' As default, parsed key metadata will be returned as data.frame. 'id_list' returns a list of IDs and sources. Use 'raw' to get full metadata as list. Please be aware that these lists can become very large.

Examples

## Not run:
# single DOI name
epmc_search_by_doi(doi = "10.1161/strokeaha.117.018077")
# multiple DOIname in a vector
my_dois <- c(
    "10.1159/000479962",
    "10.1002/sctm.17-0081",
    "10.1161/strokeaha.117.018077",
    "10.1007/s12017-017-8447-9")
epmc_search_by_doi(doi = my_dois)
# full metadata
epmc_search_by_doi(doi = my_dois, output = "raw")

## End(Not run)
epmc_search_by_doi_  Search Europe PMC by a DOI name

Description

Please use epmc_search_by_doi instead. It calls this method, returning open access status information from all your requests.

Usage

epmc_search_by_doi_(doi, .pb = NULL, output = NULL)

Arguments

doi  character vector containing DOI names.
.pb  progress bar object
output  character, what kind of output should be returned. One of ’parsed’, ’id_list’ or ’raw’ As default, parsed key metadata will be returned as data.frame. ’id_list’ returns a list of IDs and sources. Use ’raw’ to get full metadata as list. Please be aware that these lists can become very large.

Examples

## Not run:
epmc_search_by_doi_("10.1159/000479962")

## End(Not run)

europepmc  europepmc - an R client for the Europe PMC RESTful article API

Description

What is europepmc?:
europepmc facilitates access to Europe PMC RESTful Web Service. Europe PMC covers life science literature and gives access to open access full texts. Coverage is not only restricted to Europe, but articles and abstracts are indexed from all over the world. Europe PMC ingests all PubMed content and extends its index with other sources, including Agricola, a bibliographic database of citations to the agricultural literature, or Biological Patents. Besides searching abstracts and full text, europepmc can be used to retrieve reference sections and citations, text-mined terms or cross-links to other databases hosted by the European Bioinformatics Institute (EBI).

Index

epmc_annotations_by_id, 2
epmc_citations, 3
epmc_db, 4
epmc_db_count, 6
epmc_details, 7
epmc_ftxt, 8
epmc_ftxt_book, 9
epmc_hits, 9
epmc_hits_trend, 10
epmc_lablinks, 11
epmc_lablinks_count, 13
epmc_profile, 14
epmc.refs, 14
epmc_search, 10, 15, 17, 18
epmc_search_, 17
epmc_search_by_doi, 18, 19
epmc_search_by_doi_, 19
europepmc, 19