

# Package ‘biorxivr’

May 8, 2020

**Type** Package

**Title** Search and Download Papers from the bioRxiv Preprint Server

**Version** 1.0.2

**Maintainer** Silvia Liverani <liveranis@gmail.com>

**LazyLoad** yes

**LazyData** yes

**Description** The bioRxiv preprint server (<<https://www.biorxiv.org>>) is a website where scientists can post preprints of scholarly texts in biology. Users can search and download PDFs in bulk from the preprint server. The text of abstracts are stored as raw text within R, and PDFs can easily be saved and imported for text mining with packages such as 'tm'.

**License** MIT + file LICENSE

**Depends** XML, RCurl

**Suggests** knitr, rmarkdown

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Edmund Hart [aut, cph],  
Silvia Liverani [aut, cre, cph]

**Repository** CRAN

**Date/Publication** 2020-05-08 16:20:07 UTC

## R topics documented:

bx_download . . . . .	2
bx_extract . . . . .	2
bx_extract_single . . . . .	3
bx_search . . . . .	4

<b>Index</b>	<b>6</b>
--------------	----------

---

bx_download	<i>Download PDFs of papers found by bx_search</i>
-------------	---

---

**Description**

Download PDFs of all the papers in your search results.

**Usage**

```
bx_download(bxso, directory, create = TRUE)
```

**Arguments**

bxso	search results from bx_search()
directory	The location you want to download the PDFs to.
create	TRUE or FALSE. If true create the directory if it does not exist. Default is TRUE.

**Value**

The PDFs are downloaded into the directory selected. Nothing is returned to R.

**Examples**

```
bxEco <- bx_search("ecology", limit=1)
bx_download(bxEco, tempdir())
```

---

bx_extract	<i>Extract data from search results</i>
------------	---

---

**Description**

Extract information about each paper in search results, returning author info, abstract and download metrics

**Usage**

```
bx_extract(bxso)
```

**Arguments**

bxso	Search results from bx_search().
------	----------------------------------

**Details**

The metrics field will return a month by month data frame with abstract views and pdf downloads

**Value**

The output is an object of class `biorxiv_paper`. This object is a list where each item in the list is one of the papers found using `bx_search()`. Each of these items is itself a list with 2 items: `authors` and `paper`.

`author` This item is again a list with two items. The first item is called `'names'` and it is a vector of characters, where each element is the name and last name of one of the authors. The second item is called `'emails'` and it is a character vector, where each element of this vector is the email address of one of the authors.

`paper` This item is again a list with 5 items. The first item is called `'title'` and it is a character vector of one element. The second item is called `'abstract'` and it is a character vector of one element containing the whole abstract. The third item is called `'date'` and it is a character vector of one element. The fourth item is called `'DOI'` and it is a character vector of one element. The fifth item is called `'fulltext_url'` and it is a character vector of one element containing the full URL of the pdf.

**Examples**

```
bxEco <- bx_search("ecology", limit=1)
bxEcoData <- bx_extract(bxEco)
```

---

`bx_extract_single`      *Extract data from a single record*

---

**Description**

Generate an S3 object that represents a single paper

**Usage**

```
bx_extract_single(bxso_url)
```

**Arguments**

`bxso_url` The URL of a single biorxiv paper, so one element of the search results from `bx_search()`.

**Value**

The output is an object of class `biorxiv_paper`. This object is a list where each item in the list is one of the papers found using `bx_search()`. Each of these items is itself a list with 2 items: authors and paper.

`author` This item is again a list with two items. The first item is called 'names' and it is a vector of characters, where each element is the name and last name of one of the authors. The second item is called 'emails' and it is a character vector, where each element of this vector is the email address of one of the authors.

`paper` This item is again a list with 5 items. The first item is called 'title' and it is a character vector of one element. The second item is called 'abstract' and it is a character vector of one element containing the whole abstract. The third item is called 'date' and it is a character vector of one element. The fourth item is called 'DOI' and it is a character vector of one element. The fifth item is called 'fulltext\_url' and it is a character vector of one element containing the full URL of the pdf.

**Examples**

```
bxEco <- bx_search("ecology",limit=1)
bxEcoData <- bx_extract_single(bxEco[[1]])
```

---

bx_search	<i>Search biorxiv.org</i>
-----------	---------------------------

---

**Description**

return a list of URLs, IDS and number of results found from search of biorxiv.org

**Usage**

```
bx_search(query, limit = 10)
```

**Arguments**

<code>query</code>	The terms to search for in biorxiv.
<code>limit</code>	The number of results to return.

**Details**

This uses the generic search interface, therefore be aware that you'll have to do post download filtering if you want terms in a specific field

**Value**

A list with the following elements:

URL	The URLs of the biorxiv papers from the search terms.
ID	The IDs of the biorxiv papers from the search terms.
found	The total number of papers found for the query.
query	The query searched on biorxiv.
limit	The number of results to return as set by the user.

**Examples**

```
bxEco <- bx_search("ecology", limit=1)
```

# Index

`bx_download`, 2  
`bx_extract`, 2  
`bx_extract_single`, 3  
`bx_search`, 4