

# Package: internetarchive (via r-universe)

December 2, 2024

**Type** Package

**Title** An API Client for the Internet Archive

**Description** Search the Internet Archive (<<https://archive.org>>),  
retrieve metadata, and download files.

**Version** 0.1.6

**Date** 2016-12-08

**Maintainer** Ahmet Akkoc <themadprogramer@gmail.com>

**License** MIT + file LICENSE

**LazyData** TRUE

**Roxygen** list(wrap = FALSE)

**URL** <https://docs.ropensci.org/internetarchive>,  
<https://github.com/ropensci/internetarchive>

**BugReports** <https://github.com/ropensci/internetarchive/issues>

**Depends** R (>= 3.1.0)

**Imports** httr (>= 1.1.0), dplyr (>= 0.4.3)

**Suggests** knitr, testthat, rmarkdown

**VignetteBuilder** knitr

**RoxygenNote** 6.0.1

**Config/pak/sysreqs** libssl-dev

**Repository** <https://ropensci.r-universe.dev>

**RemoteUrl** <https://github.com/ropensci/internetarchive>

**RemoteRef** master

**RemoteSha** ce862d52c1541b670da5e991799b533b0c3b87f0

## Contents

ia_browse . . . . .	2
ia_download . . . . .	3
ia_files . . . . .	4
ia_get_items . . . . .	4
ia_item_id . . . . .	5
ia_keyword_search . . . . .	6
ia_list_fields . . . . .	6
ia_metadata . . . . .	7
ia_search . . . . .	7
internetarchive . . . . .	8

<b>Index</b>	<b>9</b>
--------------	----------

---

ia_browse	<i>Open an Internet Archive item in the browser</i>
-----------	---

---

### Description

Open an Internet Archive item in the browser

### Usage

```
ia_browse(item_id, type = c("details", "stream"))
```

### Arguments

item_id	The item identifier. If multiple item identifiers are passed in, only the first will be opened.
type	Which page to open: details is the metadata page, stream is the viewing page for items which are associated with a PDF.

### Value

Returns the item ID(s) passed to the function.

### Examples

```
# Distinguished Converts to Rome in America
ia_browse("distinguishedcon00scanuoft")
```

---

`ia_download`*Download files for Internet Archive items.*

---

## Description

Download files for Internet Archive items.

## Usage

```
ia_download(files, dir = ".", extended_name = TRUE, overwrite = FALSE,
            silence = FALSE)
```

## Arguments

<code>files</code>	A data frame of files returned by <code>ia_files</code> . You should filter this data frame to download only the files that you actually want.
<code>dir</code>	The directory in which to save the downloaded files.
<code>extended_name</code>	If this argument is FALSE, then the downloaded file will have a filename in the following format: <code>itemidentifier.extension</code> , e.g., <code>thedamnationofth00133gut.txt</code> . If there are multiple files of the same file type for an item, then the file names will not be unique. If this argument is TRUE, then the downloaded file will have a filename in the following format: <code>itemidentifier-original-filename.extension</code> , e.g., <code>thedamnationofth00133gut-133.txt</code> .
<code>overwrite</code>	If TRUE, this function will download all files and overwrite them on disk if they have already been downloaded. If FALSE, then if a file already exists on disk it will not be downloaded again but other downloads will proceed normally.
<code>silence</code>	If false, print the item IDs as they are downloaded.

## Value

A data frame including the file names of the downloaded files.

## Examples

```
## Not run:
if (require(dplyr)) {
  dir <- tempdir()
  ia_get_items("thedamnationofth00133gut") %>%
    ia_files() %>%
    filter(type == "txt") %>% # download only the files we want
    ia_download(dir = dir, extended_name = FALSE)
}

## End(Not run)
```

---

ia_files	<i>Access the list of files associated with an Internet Archive item</i>
----------	--

---

**Description**

Access the list of files associated with an Internet Archive item

**Usage**

```
ia_files(items)
```

**Arguments**

items            A list describing an Internet Archive items returned from the API.

**Value**

A list containing the files as a list of character vectors.

**Examples**

```
## Not run:
ats_query <- c("publisher" = "american tract society")
ids       <- ia_search(ats_query, num_results = 3)
items    <- ia_get_items(ids)
files    <- ia_files(items)
files

## End(Not run)
```

---

ia_get_items	<i>Get the metadata for Internet Archive items</i>
--------------	--

---

**Description**

Get the metadata for Internet Archive items

**Usage**

```
ia_get_items(item_id, silence = FALSE)
```

**Arguments**

item\_id            A character vector containing the ID for an Internet Archive item. This argument is vectorized, so you can retrieve multiple items at once.

silence            If false, print the item IDs as they are retrieved.

**Value**

A list containing the metadata returned by the API. List names correspond to the item IDs.

**Examples**

```
## Not run:
ia_get_items("thedamnationofth00133gut")

ats_query <- c("publisher" = "american tract society")
ids       <- ia_search(ats_query, num_results = 2)
ia_get_items(ids)

## End(Not run)
```

---

ia_item_id	<i>Access the item IDs from an Internet Archive items</i>
------------	---

---

**Description**

Access the item IDs from an Internet Archive items

**Usage**

```
ia_item_id(item)
```

**Arguments**

item	A list describing an Internet Archive items returned from the API. This argument is vectorized.
------	---

**Value**

A character vector containing the item IDs.

**Examples**

```
ats_query <- c("publisher" = "american tract society")
ids       <- ia_search(ats_query, num_results = 3)
items    <- ia_get_items(ids)
ia_item_id(items)
```

---

ia\_keyword\_search      *Perform an simple keyword search of the Internet Archive.*

---

**Description**

Perform an simple keyword search of the Internet Archive.

**Usage**

```
ia_keyword_search(keywords, num_results = 5, page = 1, print_total = TRUE)
```

**Arguments**

keywords	The keywords to search for.
num_results	The number of results to return per page.
page	When results are paged, which page of results to return.
print_total	Should the total number of results for this query be printed as a message?

**Value**

A character vector of Internet Archive item IDs.

**Examples**

```
ia_keyword_search("isaac hecker", num_results = 20)
```

---

ia\_list\_fields      *List accepted metadata fields*

---

**Description**

List accepted metadata fields

**Usage**

```
ia_list_fields()
```

**Value**

A list of the accepted metadata fields

**Examples**

```
ia_list_fields()
```

---

ia_metadata	<i>Access the item metadata from an Internet Archive item</i>
-------------	---

---

**Description**

Access the item metadata from an Internet Archive item

**Usage**

```
ia_metadata(items)
```

**Arguments**

`items` A list object describing an Internet Archive items returned from the API.

**Value**

A data frame containing the metadata, with columns `id` for the item identifier, `field` for the name of the metadata field, and `value` for the metadata values.

**Examples**

```
ats_query <- c("publisher" = "american tract society")
ids       <- ia_search(ats_query, num_results = 3)
items    <- ia_get_items(ids)
metadata <- ia_metadata(items)
metadata
```

---

ia_search	<i>Search the Internet Archive</i>
-----------	------------------------------------

---

**Description**

Perform an advanced search of the Internet Archive, specifying which metadata fields to search. Note that all searches are in the form of "contains," i.e., the title contains the search term.

**Usage**

```
ia_search(terms, num_results = 5, page = 1, print_url = FALSE,
          print_total = TRUE)
```

### Arguments

terms	A set of metadata fields and corresponding values to search. These should take the form of a named character vector.
num_results	The number of results to return per page.
page	When results are paged, which page of results to return.
print_url	Should the URL used for the query be printed as a message?
print_total	Should the total number of results for this query be printed as a message?

### Value

A character vector of Internet Archive item IDs.

### References

See the documentation on the Internet Archive's [advanced search page](#).

### Examples

```
query1 <- c("title" = "damnation of theron ware")
ia_search(query1)
query2 <- c("title" = "damnation of theron ware",
           "contributor" = "gutenberg")
ia_search(query2)
```

---

internetarchive

*Client for the Internet Archive API*

---

### Description

This client permits you to search ([ia\\_search](#)), retrieve item metadata ([ia\\_metadata](#)) and associated files ([ia\\_files](#)), and download files ([ia\\_files](#)) in a pipeable interface.



# Index

ia\_browse, 2  
ia\_download, 3  
ia\_files, 3, 4, 8  
ia\_get\_items, 4  
ia\_item\_id, 5  
ia\_keyword\_search, 6  
ia\_list\_fields, 6  
ia\_metadata, 7, 8  
ia\_search, 7, 8  
internetarchive, 8