# Package 'wikisourcer'

March 17, 2019

**Type** Package

**Title** Download Public Domain Works from Wikisource

**Version** 0.1.3

**Maintainer** Félix Luginbuhl <felix.luginbuhl@protonmail.ch>

**Description** Download public domain works from Wikisource <https://wikisource.org/>, a free library from the Wikimedia Foundation project.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**Depends** R (>= 2.10)

**Imports** tibble, magrittr, rvest, purrr, xml2, urltools

**RoxygenNote** 6.1.1

**URL** https://github.com/lgnbhl/wikisourcer

**BugReports** https://github.com/lgnbhl/wikisourcer/issues

**Suggests** dplyr, stringr, knitr, rmarkdown, ggplot2, tidyr, tidytext, widyr, SnowballC, ggraph, igraph

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Félix Luginbuhl [aut, cre]

**Repository** CRAN

**Date/Publication** 2019-03-17 08:20:02 UTC

## R topics documented:

`wikisource_book` *Download a book from Wikisource*

## Description

Download a book using the url of a Wikisource content page into a data frame. The Wikisource table of content page should link to all the Wikisource pages constituting the book. The text in the Wikisource pages is downloaded using the `wikisource_page()` function.

## Usage

```
wikisource_book(url, cleaned = TRUE)
```

## Arguments

| | |
|---|---|
| `url` | A url of a Wikisource content page listing the pages constituting the book. |
| `cleaned` | A boolean variable for cleaning Wikisource pages. |

## Details

The download could fail if the Wikisource paths listed into content page strongly differ from the url path of the content page.

## Value

A five column tbl_df (a type of data frame; see tibble or dplyr packages) with one row for each line of the text or texts, with columns.

**text** A character column

**title** A character column with the title of the Wikisource summary page

**page** Integer column with a number for the text from each Wikisource page downloaded

**language** A character column with a two letter string refering the language of the text

**url** A character column with the url of the Wikisource page of the text

## Examples

```
## Not run:

# download Voltaire's "Candide"
wikisource_book("https://en.wikisource.org/wiki/Candide")

# download "Candide" in French and Spanish
library(purrr)

fr <- "https://fr.wikisource.org/wiki/Candide,_ou_l%E2%80%99Optimisme/Garnier_1877"
es <- "https://es.wikisource.org/wiki/C%C3%A1ndido,_o_el_optimismo"
```

```
books <- map_df(c(fr, es), wikisource_book)

## End(Not run)
```

---

`wikisource_page` *Download a page from Wikisource*

---

## Description

Download the text of a Wikisource page into a data frame using its url.

## Usage

```
wikisource_page(wikiurl, page = NA, cleaned = TRUE)
```

## Arguments

| | |
|---|---|
| `wikiurl` | The url of a Wikisource page that will be downloaded. |
| `page` | A string naming the Wikisource page downloaded. |
| `cleaned` | A boolean variable for cleaning the Wikisource page. |

## Value

A four column tbl_df (a type of data frame; see tibble or dplyr packages) with one row for each line of the text or texts, with four columns.

**text** A character column

**page** A column naming the page downloaded

**language** A character column with a two letter string refering to the language of the text

**url** A character column with the url of the Wikisource page of the text

## Examples

```
## Not run:
# download Sonnet 18 of Shakespeare
wikisource_page("https://en.wikisource.org/wiki/Shakespeare%27s_Sonnets/Sonnet_18", "Sonnet 18")

# download Sonnets 116, 73 and 130 of Shakespeare
library(purrr)

urls <- paste0("https://en.wikisource.org/wiki/Shakespeare%27s_Sonnets/Sonnet_", c(116, 73, 130))
sonnets <- map2_df(urls, paste0("Sonnet ", c(116, 73, 130)), wikisource_page)

## End(Not run)
```

# Index