Package 'patentr'

July 23, 2025

Type Package

Title Access USPTO Bulk Data in Tidy Rectangular Format

Version 0.1.4

Description Converts TXT and XML data curated by the United States Patent and Trademark Office (USPTO). Allows conversion of bulk data after downloading directly from the USPTO bulk data website, eliminating need for users to wrangle multiple data formats to get large patent databases in tidy, rectangular format. Data details can be found on the USPTO website <https://bulkdata.uspto.gov/>. Currently, all 3 formats: 1. TXT data (1976-2001); 2. XML format 1 data (2002-2004); and 3. XML format 2 data (2005-current) can be converted to rectangular, CSV format.
Relevant literature that uses data from USPTO includes Wada (2020) <doi:10.1007/s11192-020-03674-4> and Plaza & Albert (2008) <doi:10.1007/s11192-007-1763-3>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.1

LinkingTo Rcpp

Imports Rcpp (>= 1.0.5), utils, lubridate (>= 1.7.9), magrittr (>= 2.0), dplyr (>= 1.0.2), rlang (>= 0.4.7), xml2 (>= 1.3.2), progress (>= 1.2.2)

URL https://JYProjs.github.io/patentr/

BugReports https://github.com/JYProjs/patentr/issues

Suggests testthat, covr, knitr, readr, rmarkdown, tibble

VignetteBuilder knitr

LazyData true

Depends R (>= 2.10)

NeedsCompilation yes

Author Raoul Wadhwa [aut, cre] (ORCID: <https://orcid.org/0000-0003-0503-9580>),

James Yu [aut],
Hayley Beltz [aut],
Milind Desai [aut],
Jacob Scott [aut],
Peter Erdi [aut]

Maintainer Raoul Wadhwa <raoulwadhwa@gmail.com>

Repository CRAN

Date/Publication 2021-09-12 15:20:02 UTC

Contents

get_bulk_patent_data 2
wku_to_pno 3
y1976w1 4

Index **5**

`get_bulk_patent_data` *Get Bulk Patent Data from USPTO*

Description

Download and convert bulk patent data to tidy format from the USPTO website <https://bulkdata.uspto.gov>. Data can be returned as a data frame or written to a file (see 'output_file' parameter). Since USPTO issues patents weekly, at minimum, all patents from a given week must be acquired at once.

Usage

```
get_bulk_patent_data(year, week, output_file)
```

Arguments

`year`	integer vector containing years from which patents should be collected
`week`	integer vector of weeks within the corresponding 'year' element from which patents should be collected
`output_file`	variable of class 'character'; will output to that file in CSV format

Value

either 'TRUE' (placeholder) or object of class 'data.frame' (see param 'output_file' for details)

Examples

```
## NOTE: none of the examples are run due to the download requirement
## Not run:
# download patents from the first week of 1976 and get data frame
patent_data <- get_bulk_patent_data(year = 1976, week = 1)

# download patents from the last 5 weeks of 1980 (and write to a file)
get_bulk_patent_data(year = rep(1980, 5), week = 48:52,
                     output_file = "patent-data.csv")

## End(Not run)
```

---

wku_to_pno *Get Patient Number from WKU*

Description

Convert WKU identifier provided in bulk patent files to patent number used in most sources. The References provided in bulk patent files are also in patent number format, not in WKU format.

Usage

```
wku_to_pno(wku)
```

Arguments

wku	character vector containing patent WKUs

Value

character vector containing patent numbers

Examples

```
# convert sample WKUs to patent number and print
sample_wku <- c("RE028671", "03930271")
print(wku_to_pno(sample_wku))
```

y1976w1 *Patents issued in week 1 of the year 1976.*

Description

A dataset containing information about patents issued by the United States Patent and Trademark Office (USPTO) <https://www.uspto.gov/> in the first week of the year 1976. This can be recreated by running the ‘get_bulk_patent_data‘ function in the ‘patentr‘ package and setting the ‘year‘ and ‘week‘ parameters to ‘1976‘ and ‘1‘, respectively.

Usage

```
y1976w1
```

Format

A data frame with 1379 rows and 9 variables:

WKU unique patient identifier

Title patent title

App_Date date on which patent application was submitted

Issue_Date date on which patent was issued by USPTO

Inventor patent inventor(s)

Assignee person(s)/corporation(s) to whom the patent was assigned

ICL_Class patent classification based on IPC system

References patents referenced by this patent

Claims free-text claims made about value of this patent

Source

https://www.uspto.gov/

Index

∗ **datasets**
y1976w1, 4

get_bulk_patent_data, 2

wku_to_pno, 3

y1976w1, 4