

# Package ‘patentr’

February 8, 2021

**Type** Package

**Title** Access USPTO Bulk Data in Tidy Rectangular Format

**Version** 0.1.0

**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, only TXT data (1976-2001) conversion is implemented; XML formats are in the process of being added. Relevant literature that uses data from USPTO includes Wada (2020) <[doi:10.1007/s11192-020-03674-4](https://doi.org/10.1007/s11192-020-03674-4)> and Plaza & Albert (2008) <[doi:10.1007/s11192-007-1763-3](https://doi.org/10.1007/s11192-007-1763-3)>.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**LinkingTo** Rcpp

**Imports** Rcpp (>= 1.0.5), data.table (>= 1.13), utils, lubridate (>= 1.7.9), magrittr (>= 2.0), dplyr (>= 1.0.2), rlang (>= 0.4.7)

**URL** <https://github.com/JYProjs/patentr>

**BugReports** <https://github.com/JYProjs/patentr/issues>

**Suggests** testthat

**NeedsCompilation** yes

**Author** Raoul Wadhwa [aut, cre],  
James Yu [aut],  
Peter Erdi [aut]

**Maintainer** Raoul Wadhwa <[raoulwadhwa@gmail.com](mailto:raoulwadhwa@gmail.com)>

**Repository** CRAN

**Date/Publication** 2021-02-08 11:10:02 UTC

**R topics documented:**

get\_bulk\_patent\_data . . . . . 2

**Index** . . . . . 3

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 = NULL)
```

**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	if ‘NULL’, returns a data frame; if a ‘character’ (single-element vector), 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)
```

# Index

`get_bulk_patent_data`, [2](#)