

Package ‘inlpubs’

January 13, 2022

Title USGS INL Project Office Publications

Version 1.0.2

Description Contains bibliographic information for the U.S. Geological Survey (USGS) Idaho National Laboratory (INL) Project Office.

Depends R (>= 3.6.0)

Imports checkmate, stats, tm

Suggests covr, DT, graphics, inlmisc, kableExtra, knitr, pkgdown, rmarkdown, RWeka, stringi, textutils, testthat, tools, usethis, utils

License CC0

URL <https://code.usgs.gov/inl/inlpubs>

BugReports <https://code.usgs.gov/inl/inlpubs/-/issues>

Copyright This software is in the public domain because it contains materials that originally came from the United States Geological Survey (USGS), an agency of the United States Department of Interior. For more information, see the official USGS copyright policy at <https://www.usgs.gov/information-policies-and-instructions/copyrights-and-credits>

Encoding UTF-8

VignetteBuilder knitr

SystemRequirements Package vignettes require pandoc (<https://pandoc.org/>) and phantomjs (<https://phantomjs.org/>). Text mining using n-grams requires Amazon Corretto (<https://aws.amazon.com/corretto/>).

BuildVignettes true

LazyData true

LazyDataCompression xz

RoxygenNote 7.1.2

NeedsCompilation no

Author Jason C. Fisher [aut, cre] (<<https://orcid.org/0000-0001-9032-8912>>)

Maintainer Jason C. Fisher <jfisher@usgs.gov>

Repository CRAN

Date/Publication 2022-01-13 18:52:46 UTC

R topics documented:

MineText	2
pubs	3
Index	5

MineText	<i>Mine Text Components in the INLPO Publications</i>
----------	---

Description

Performs a word frequency text analysis of **Idaho National Laboratory Project Office** (INLPO) publications.

Usage

```
MineText(
  pubs,
  components = c("title", "abstract"),
  ngmin = 1L,
  ngmax = ngmin,
  lowfreq = 1L
)
```

Arguments

pubs	pubs_data class. Bibliographic information, see pubs dataset for details.
components	character vector. One or more text components to analyze. Choices include the "title", "abstract", "annotation", and "citation" of the document.
ngmin, ngmax	integer number. Splits strings into <i>n-grams</i> with given minimal and maximal numbers of grams. An <i>n-gram</i> is an ordered sequence of <i>n</i> words taken from the body of a text. Requires that the RWeka package is available and that the environment variable JAVA_HOME points to where the Java software is located. Recommended for single text components only.
lowfreq	integer number. Lower frequency bound. Words that occur less than this bound are excluded from the returned frequency table.

Details

HTML entities are decoded when the **textutils** package is available.

Value

A word frequency table giving the number of times each word occurs in a publication's text component(s). A table column represents a single publication that is identified using its citation-key. And each row provides frequency counts for a particular word (also known as a 'term').

Author(s)

J.C. Fisher, U.S. Geological Survey, Idaho Water Science Center

See Also

`inlmisc::MakeWordCloud` function to create a word cloud.

Examples

```
m <- MineText(head(pubs, 3))
head(m)
## Not run:
d <- data.frame(word = rownames(m), freq = rowSums(m))
inlmisc::MakeWordCloud(d, display = TRUE)

## End(Not run)
```

pubs

Bibliographic Information of the INLPO

Description

Bibliographic information for reports, articles, maps, and theses related to scientific monitoring and research conducted by the U.S. Geological Survey (USGS), Idaho Water Science Center, **Idaho National Laboratory Project Office** (INLPO).

Usage

```
pubs
```

Format

An object of class 'pubs_data' that inherits behavior from the data frame class. Each record corresponds to a bibliographical item and contains the following variables:

key **BibTeX** key for the citation entry;
year year of publication;
citation bibliographic entry of class **bibentry**;
abstract abstract text string;
annotation Knobel and others (2005) annotation text string.

Row names are the BibTeX key for the citation entry.

Source

Many of these publications are available through the [USGS Publications Warehouse](#).

References

Knobel, L.L., Bartholomay, R.C., and Rousseau, J.P., 2005, Historical development of the U.S. Geological Survey hydrologic monitoring and investigative programs at the Idaho National Engineering and Environmental Laboratory, Idaho, 1949 to 2001: U.S. Geological Survey Open-File Report 2005–1223 (DOE/ID–22195), 93 p., doi: [10.3133/ofr20051223](#).

Examples

```
## Display table structure
str(pubs, max.level = 1, nchar.max = 50)

## Print the citation key for each entry in the bibliography:
rownames(pubs[1:10, ])

## Print citation, authors, and abstract for Fisher and others (2012):
key <- "FisherOthers2012"
ref <- pubs[key, "citation"]
print(ref, style = "citation", bibtex = TRUE)
format(ref$author, c("given", "family"))
txt <- pubs[key, "abstract"]
cat(strwrap(txt), sep = "\n")

## Print list of authors:
authors <- do.call("c", pubs$citation$author)
authors <- authors[!duplicated(authors)]
cat(format(authors[1:10]), sep = "\n")

## Not run:
## Export suggested citations from the bibliography:
txt <- sort(vapply(pubs$citation, function(x) {
  attr(unclass(x)[[1]], "textVersion")
}, character(1)))
txt <- head(c(rbind(txt, character(length(txt)))), -1)
txt <- strwrap(txt, width = 80, exdent = 2)
file <- tempfile(fileext = ".txt")
writeLines(txt, file)
file.show(file, title = "Suggested citations")

## End(Not run)
```

Index

- * **datasets**

- pubs, [3](#)

- * **utilities**

- MineText, [2](#)

bibentry, [3](#)

MakeWordCloud, [3](#)

MineText, [2](#)

pubs, [2, 3](#)