Package: formatBibtex (via r-universe)

September 19, 2024

Title Format BibTeX Entries and Files

Version 0.1.1.9000

Description Format BibTeX entries and files in an opinionated way.

Depends R (>= 4.0)

Imports utils

Suggests bibtex

License GPL (>= 3)

Encoding UTF-8

URL <https://wwenjie.org/formatbibtex>,
<https://github.com/wenjie2wang/formatBibtex>

BugReports <https://github.com/wenjie2wang/formatBibtex/issues>

RoxygenNote 7.2.3

Repository https://wenjie2wang.r-universe.dev

RemoteUrl https://github.com/wenjie2wang/formatbibtex

RemoteRef HEAD

RemoteSha 71ca9a2bfa9f645c0fd6617fe8772da57aef8a28

Contents

format_bibtex_entry   *Format BibTeX Entries in An Opinionated Way*

Description

This function tries to format the given BibTeX entries so that

- The "title" field is written in the title case with exceptions such as common prepositions with four or fewer letters. Special words such as "Bayesian" and "Markov" are protected by curly braces from case changing by BibTeX sytle.
- The "author" field follows a consistent fashion: "family name, given name". A period will be added to single letter acronym.
- The "journal" field (if any) is written in the title case.
- The "pages" field should use "–" instead of "-" between pages.

Usage

```
format_bibtex_entry(
  entry,
  fields = c("title", "author", "journal", "pages"),
  protected_words = getOption("formatBibtex.protected_words"),
  ...
)

format_bibtex_file(
  bibtex_file,
  output_file = bibtex_file,
  backup = (output_file == bibtex_file),
  dry_run = FALSE,
  ...
)
```

Arguments

entry	A bibentry object (created by utils::bibentry) representing BibTeX entries.
fields	A character vector specifying the fields to format. The available options are "title", "author", "journal", and "pages". Multiple choices can be specified.
protected_words	Optional words that needs protection by curly braces from cases changing by BibTeX style.
...	Other arguments passed to format_bibtex_entry.
bibtex_file	A character string presenting the BibTeX file that needs formatting.
output_file	A character string presenting the output BibTeX file. By default, the input BibTeX file will be overwritten with a backup file.

backup	A logical value. If TRUE, a backup file will be created to check and tweak the formatting options.
dry_run	A logical value. If TRUE, the formatted BibTeX entries will be returned without actually (over)writing a BibTeX file to the disk. The default value is FALSE.

Details

When emacs is available in the system, the function `format_bibtex_file()` will perform additional formatting with the help of the commands `bibtex-reformat` and `bibtex-sort-buffer`.

Value

A `bibtex` object.

Examples

```
library(formatBibtex)

## example BibTeX file that needs formatting
example_bib <- system.file("examples/example.bib", package = "formatBibtex")
print(readLines(example_bib), quote = FALSE)

## needs the package bibtex
has_bibtex <- requireNamespace("bibtex", quietly = TRUE)

## example of format_bibtex_entry()
if (has_bibtex) {
    bib <- bibtex::read.bib(example_bib)
    ## check the default words that need protection by curly braces
    (default_words <- getOption("formatBibtex.protected_words"))
    format_bibtex_entry(bib, protected_words = c(default_words, "SMEM"))
}

## example of format_bibtex_file()
if (has_bibtex) {
    output_file <- tempfile(fileext = ".bib")
    format_bibtex_file(example_bib,
                       output_file = output_file,
                       protected_words = c(default_words, "SMEM"))
    print(readLines(output_file), quote = FALSE)
}
```

---

format_string *Format a Character Vector*

Description

This function formats a character vector by following APA (American Psychological Assoication) style.

Usage

```
format_string(
  x,
  style = c("title", "sentence"),
  str2ws = "_",
  str4split = " |-",
  strict = TRUE,
  arguments = list(),
  protect_curly_braces = FALSE,
  lowercase_words = getOption("formatBibtex.lowercase_words"),
  protected_words = getOption("formatBibtex.protected_words"),
  punctuation = getOption("formatBibtex.punctuation"),
  ...
)
```

Arguments

x	A character vector that needs formatting.
style	A character string for specifying case style. "title" for title case or "sentence" for sentence case.
str2ws	Character expression that would be replaced with white space. By default, underscores will be replaced with white spaces.
str4split	Character expression that would be used to split the string into individual words.
strict	A logical value specifying whether to lower the cases for letters after the first letter. The default value is TRUE.
arguments	An optional argument list for specifying more details when replacing specified characters with whitespaces by using [gsub](#).
protect_curly_braces	
	A logical vector of length one indicating if curly braces represent a block that needs protection. This argument is mainly intended to format BibTeX entries.
lowercase_words	
	Some words that should be almost always be lowercase.
protected_words	
	Some words that should be kept in original case and should be not be convert to lowercase or uppercase.
punctuation	A character expression that should be considered as punctuation and should not be considered as a part of the protected words. The function would remove these punctuation before checking whether the words need protecting.
...	Other arguments that are not used now.

Details

The available style options are title case and sentence case. The first word of the string and the first word after colon will be always capitalized. For title case (default), the function would capitalize each word with a few exceptions, such as short conjunctions and prepositions. For sentence case,

the function would not capitalize a word unless it starts a string, come after a colon, or is a proper noun. We may specify some words that need protection from any kind of case conversion.

The function would also replace possible splitting characters, such as underscores with white spaces. Empty strings (`""` or `''`), NA, `character(0)`, and NULL are allowed and returned as they were.

Value

A character vector of the same length as the input.

Examples

```
library(formatBibtex)

## simple examples
foo <- c("iT IS_A_sIMPLe_ExamplE.", "let'S_do soMe_tesTs!")
format_string(foo, style = "title")
format_string(foo, style = "sentence")

## protect some words from formatting
## check out the built-in
(default_protected_words <- getOption("formatBibtex.protected_words"))

## e.g., protect ABCD from being converted to lowercase
bar <- c("on_the_convergence properties of the ABCD_algorithm",
         "teSt: tHe cluster is Running MCMC!")
format_string(bar, style = "sentence",
              protected_words = c(default_protected_words, "ABCD"))

## more tricky examples: protected words contain `str4split`
foo <- c("nineteenth- and twentieth-century writers",
         "well-differentiated cells with arXiv e-prints")
format_string(foo, str4split = "-| ",
              protected_words = c("arXiv", "e-prints"))

## trivial examples
format_string(NULL)
format_string(character(0))
format_string(character(3))
format_string(c(NA, "", "hello world!"))
```

Index