Package 'tidystringdist'

April 3, 2019

Title String Distance Calculation with Tidy Data Principles

Version 0.1.4

Description Calculation of string distance following the tidy
data principles. Built on top of the 'stringdist' package.

License MIT + file LICENSE

Imports attempt, rlang, stringdist, tibble

Suggests dplyr, knitr, magrittr, rmarkdown, testthat

VignetteBuilder knitr

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

Collate 'globalVariables.R' 'tidycomb.R' 'utils.R' 'tidystringdist.R'

NeedsCompilation no

Author Colin Fay [aut, cre] (¡https://orcid.org/0000-0001-7343-1846¿),
Dmytro Perepolkin [ctb] (¡https://orcid.org/0000-0001-8558-6183¿)

Maintainer Colin Fay <contact@colinfay.me>

Repository CRAN

Date/Publication 2019-03-20 23:10:03 UTC

R topics documented:

---

tidy_comb   *Tidy combine*

Description

Get all combinations from a dataframe column or from a list

1

Usage

```
tidy_comb(data, base, ...)

## S3 method for class 'data.frame'
tidy_comb(data, base, ...)

## Default S3 method:
tidy_comb(data, base, ...)
```

Arguments

`data`	data object containing the list of words, either a list or a data.frame
`base`	the base word to compare with all the words
`...`	if data is a data.frame, the col where the words to combine are

Value

a tibble with all possible combination of elements from a list

Examples

```
tidy_comb(iris, "this", Species)
tidy_comb(state.name, "Paris")
```

`tidy_comb_all`	*Tidy combine all*

Description

Get all combinations from a dataframe column

Usage

```
tidy_comb_all(data, ...)

## S3 method for class 'data.frame'
tidy_comb_all(data, ...)

## Default S3 method:
tidy_comb_all(data, ...)
```

Arguments

`data`	a list or a data.frame with the elements to combine
`...`	if data is a data.frame, the col where the words to combine are

Value

a tibble with all possible combination of elements from a list

Examples

```
tidy_comb_all(iris, Species)
tidy_comb_all(state.name)
```

---

tidy_stringdist	*Tidy stringdist calculation*

Description

Tidy stringdist calculation

Usage

```
tidy_stringdist(df, v1 = V1, v2 = V2, method = c("osa", "lv", "dl",
  "hamming", "lcs", "qgram", "cosine", "jaccard", "jw", "soundex"), ...)
```

Arguments

df	a dataframe containing the strings to compare
v1	the name of the first columns
v2	the name of the second columns
method	one of the methods implemented in the stringdist package — "osa", "lv", "dl", "hamming", "lcs", "qgram", "cosine", "jaccard", "jw", "soundex". See `stringdist-metrics`
...	other parameters passed to `stringdist`

Value

a tibble with string distance

Examples

```
proust <- tidy_comb_all(c("Albertine", "Françoise", "Gilberte", "Odette", "Charles"))
tidy_stringdist(proust)
```