Package 'nomesbr'

July 17, 2025

Type Package

Title Limpa e Simplifica Nomes de Pessoas (Name Cleaner and Simplifier)

Version 0.0.7

Description Limpa e simplifica nomes de pessoas para auxiliar no pareamento de banco de dados na ausência de chaves únicas não ambíguas. Detecta e corrige erros tipográficos mais comuns, simplifica opcionalmente termos sujeitos eventualmente a omissão em cadastros, e simplifica foneticamente suas palavras, aplicando variação própria do algoritmo metaphoneBR. (Cleans and simplifies person names to assist in database matching when unambiguous unique keys are unavailable. Detects and corrects common typos, optionally simplifies terms prone to omission in records, and applies phonetic simplification using a custom variation of the metaphoneBR algorithm.)
Mation (2025) <doi:10.6082/uchicago.15104>.

License MIT + file LICENSE

Encoding UTF-8

Language pt

RoxygenNote 7.3.2

Imports data.table, dplyr, httr2, stringr, tictoc

Suggests testthat (>= 3.0.0), mockery, knitr, rmarkdown

Depends R (>= 4.3.0)

Config/testthat/edition 3

URL https://github.com/ipeadata-lab/nomesbr, https://ipeadata-lab.github.io/nomesbr/

BugReports https://github.com/ipeadata-lab/nomesbr/issues

VignetteBuilder knitr

NeedsCompilation no

Author Rodrigo Borges [aut, cre] (ORCID:
<https://orcid.org/0000-0003-2076-1424>),
Pedro Cavalcanti G. Ferreira [aut] (ORCID:
<https://orcid.org/0000-0002-8540-1860>),
Lucas Mation [aut] (ORCID: <https://orcid.org/0000-0002-7461-932X>),
Ipea - Institue for Applied Economic Research [cph, fnd]

Maintainer Rodrigo Borges <rodrigoesborges@gmail.com>

Repository CRAN

Date/Publication 2025-07-17 20:30:02 UTC

Contents

`identificar_adicionar_nome_proprio`
Adiciona Nome Próprio Validado de 'nomes_proprios_compostos' .

Description

Adiciona Nome Próprio Validado de 'nomes_proprios_compostos' .

Usage

```
identificar_adicionar_nome_proprio(dt, s)

add_nome_proprio_to_word1_and_word2p(dt, s)
```

Arguments

`dt`	Um 'data.table'.
`s`	Nome da coluna (string) base para derivação das colunas de palavras (por exemplo, se 's = "nome_simpl"', espera 'nome_simpl1', 'nome_simpl2p').

Value

O 'data.table' 'dt' com colunas '_v2' adicionadas.

Examples

```
dt_nomes <- data.table::data.table(nome=c("MARIA DO SOCORRO SILVA",
"ANA PAULA DE OLIVEIRA","JOSE DAS FLORES"))
dt_nomes <- identificar_adicionar_nome_proprio(dt_nomes,"nome")
print(dt_nomes)
```

`limpar_nomes` *Limpa e Analisa Nomes em um data.table*

Description

Processa uma coluna de nomes em um ‘data.table‘, aplicando uma série de regras de limpeza para identificar e corrigir/marcar problemas comuns como menções a "FALECIDO", "CARTORIO", erros de digitação, espaços indevidos, etc.

Usage

```
limpar_nomes(d, s)

find_and_clean_NAnames_and_extra_spaces(d, s)
```

Arguments

`d`	Um objeto ‘data.table‘.
`s`	O nome da coluna (em string) dentro de ‘d‘ que contém os nomes a serem processados.

Details

A função executa os seguintes passos principais:

1. Cria uma cópia da coluna de nomes para limpeza.
2. Detecta e trata menções a "FALECIDO(A)".
3. Detecta e trata menções a "CARTORIO" e nomes de cidades comuns em registros.
4. Corrige espaçamento perto de caracteres especiais com ‘limpa_espaco_acento_til_apostrofe‘.
5. Identifica e trata nomes contendo termos problemáticos como "PAI", "MAE", "SEM", "NAO", exceto em contextos aceitáveis.
6. Identifica e trata casos de "NADA CONSTA" e variações.
7. Corrige E, DA, DE e variantes com caracter prévio indevido (ex: "EDAS" para "DAS" se aplicável).
8. Remove saudações como "SR.", "SRA.".
9. Remove termos como "IGNORADO", "DESCONHECIDO".
10. Remove repetições de partículas de ligação (ex: "DE DE").
11. Limpa letras repetidas no início ou meio de palavras.

Value

Um ‘data.table‘ modificado, contendo a coluna original, uma nova coluna com sufixo "_clean" com os nomes limpos, e colunas booleanas indicando a detecção de cada tipo de problema (ex: ‘falecido‘, ‘cartorio‘).

Examples

```
# Supondo que 'meu_DT' é um data.table com uma coluna 'nome_sujo'
DT_exemplo <- data.table::data.table(
id = 1:3,
nome_sujo = c("MARIA FALECIDA SSILVA", "CARTORIO DE PAZ", "JOAO D ARC")
)
DT_limpo <- limpar_nomes(DT_exemplo, "nome_sujo")
print(DT_limpo)
```

NA_strings *ADA CONSTA'*

Description

ADA CONSTA'

Usage

```
NA_strings
```

Format

An object of class `character` of length 1.

---

remove_PARTICULAS_AGNOMES *Remove Partículas, Agnomes e algumas Patentes de Nomes*

Description

Remove Partículas, Agnomes e algumas Patentes de Nomes

Usage

```
remove_PARTICULAS_AGNOMES(s)
```

Arguments

s	Vetor de caracteres contendo nomes.

Value

Vetor de caracteres com nomes simplificados.

Examples

```
vct_nomes <- c("JOAO DA SILVA FILHO","CORONEL JACINTO")
remove_PARTICULAS_AGNOMES(vct_nomes)
```

`segmentar_nomes` *Adiciona Colunas com Partes do Nome (w1, w2, w3, w2p, w12p)*

Description

Adiciona Colunas com Partes do Nome (w1, w2, w3, w2p, w12p)

Usage

```
segmentar_nomes(dt, s)

add_string_w1_w2_w3_and_w2p(dt, s)
```

Arguments

`dt`	Um ‘data.table‘.
`s`	Nome da coluna (string) em ‘dt‘ contendo os nomes.

Value

O ‘data.table‘ ‘dt‘ modificado por referência, com novas colunas.

Examples

```
dt_nomes <- data.table::data.table(nome=c("MARIA DO SOCORRO SILVA",
"ANA PAULA DE OLIVEIRA"))
dt_nomes <- segmentar_nomes(dt_nomes,"nome")
print(dt_nomes)
```

simplifica_PARTICULAS_AGNOMES_PATENTES

Cria coluna com agnomes, algumas patentes/cargos as remove, remove partículas

Description

Cria coluna com agnomes, algumas patentes/cargos as remove, remove partículas

Usage

```
simplifica_PARTICULAS_AGNOMES_PATENTES(d, s = "nome_clean")
```

Arguments

d	um objeto 'data.table'
s	string com nome da coluna de caracteres contendo nomes para simplificar. Por padrão, "nome_clean".

Value

data.table com novas colunas de nome simplificado e de marca agnomes_titulos

Examples

```
dt_nomes <- data.table::data.table(nome = c("JOAO DA SILVA FILHO",
"CORONEL JACINTO"))
dt_nomes <- simplifica_PARTICULAS_AGNOMES_PATENTES(d=dt_nomes,s="nome")
print(dt_nomes)
```

tabular_problemas_em_nomes

Tabula Problemas Detectados nos Nomes

Description

Cria uma tabela resumo contabilizando o número de ocorrências para cada tipo de problema detectado pela função 'marcar_problemas_e_limpar_nomes'.

Usage

```
tabular_problemas_em_nomes(d, s)

tabulate_name_poblems(d, s)
```

Arguments

d	O ‘data.table‘ retornado por ‘marcar_problemas_e_limpar_nomes‘.
s	O nome da coluna original (string) que foi processada.

Value

Um ‘data.table‘ com as colunas:

- ‘condition‘: O nome da condição/problema verificado.
- ‘N_detected‘: Número de vezes que a condição foi detectada.
- ‘N_made_NA‘: Número de detecções que resultaram na limpeza para ‘NA‘.
- ‘N_replaced‘: Número de detecções onde o nome foi alterado (não para ‘NA‘).

Examples

```
DT_limpo <- data.table::data.table(nome = c("JOSEE SILVA",
"RAIMUNDA DA DA SILVA"), nome_clean = c("JOSE SILVA",
"RAIMUNDA DA SILVA"),
falecido = NA, cartorio = NA,
espaco_TilAcentoApostrofe = NA,
nome_P_M_S_N = NA, nada_nao = NA,
nada_nao_consta2 = NA, final_missing = NA, Xartigo = NA, sr_sra = NA,
ignorado = NA, dededada = 1, letra_repetida = 1)
sumario <- tabular_problemas_em_nomes(DT_limpo, "nome")
print(sumario)
```

Index