Package 'metaphonebr'

July 17, 2025

Title Custom 'MetaphoneBR' Phonetic Encoding for Brazilian Names

Version 0.0.4

Description Simplifies Brazilian names phonetically using a custom 'metaphoneBR' algorithm that preserves ending vowels.
Useful for name matching processing preserving gender information carried generally by ending vowels in Portuguese.
Mation (2025) <doi:10.6082/uchicago.15104>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

Imports lifecycle, stringi

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

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

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

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 20:30:07 UTC

Contents

---

metaphonebr   *Generates Phonetic Code (adapted Metaphone-BR) for Names in Por-*
                                   tuguese

Description

Applies a series of phonetic transformations to a person names vector to generate code that represents its approximate pronunciation in Brazilian Portuguese. The objective is to group similar sounding names, even though written in different forms.

Usage

```
metaphonebr(fullnames, verbose = FALSE)
```

Arguments

fullnames   A character vector for names to be processed.

verbose   Logical, if TRUE, print progress messages at each step. Default FALSE.

Details

The treatment process involves:

1. Preprocessing: Removal of accents, numbers and capitalize.
2. Removal of silent letters (initial H).
3. Simplification of common digraphs (LH, NH, CH, SC, QU, etc.).
4. Simplification of similar sounding consonants (C/K/S, G/J, Z/S, etc.).
5. Simplification of ending nasal sounds.
6. Removal of duplicated vowels.
7. Removal/trim of spaces and duplicated letters.

This is an adpation that does not follow strictly any published Metaphone algorithm, but was inspired by them considering brazilian portuguese context.

Value

A character vector with corresponding phonetic representation for each entry.

Examples

```
example_names <- c("Jo\u00e3o Silva", "Joao da Silva", "Maria", "Marya",
                   "Helena", "Elena", "Philippe", "Filipe", "Xavier", "Chavier")
phonetic_codes <- metaphonebr(example_names)
print(data.frame(Original = example_names, metaphonebr = phonetic_codes))

# With progress messages
phonetic_codes_verbose <- metaphonebr("Exemplo único", verbose = TRUE)
```

Index