

Package ‘rslp’

July 23, 2025

Type Package

Title A Stemming Algorithm for the Portuguese Language

Version 0.2.0

Maintainer Daniel Falbel <dfalbel@gmail.com>

Description Implements the ``Stemming Algorithm for the Portuguese Language" <[DOI:10.1109/SPIRE.2001.10024](https://doi.org/10.1109/SPIRE.2001.10024)>.

URL <https://github.com/dfalbel/rslp>

License MIT + file LICENSE

LazyData TRUE

Encoding UTF-8

RoxygenNote 7.0.2

Imports stringr, stringi, plyr, magrittr, tokenizers

Suggests dplyr, testthat, covr

NeedsCompilation no

Author Daniel Falbel [aut, cre]

Repository CRAN

Date/Publication 2020-05-11 14:40:03 UTC

Contents

apply_rules	2
extract_raw_rules	2
extract_replacement_rules	3
extract_rules	3
extract_rules_info	4
extract_rule_info	4
remove_accents	4
rslp	5
rslp_	5
rslp_doc	6
verify_sufix	7

apply_rules	<i>Apply rules</i>
-------------	--------------------

Description

Apply rules

Usage

apply_rules(word, name, steprules)

Arguments

- | | |
|-----------|---|
| word | word to which you want to apply the rules |
| name | the rule name, possible values are: 'Plural', 'Feminine', 'Adverb', 'Augmentative', 'Noun', 'Verb', 'Vowel' . |
| steprules | steprules as obtained from the function extract_rules. |

extract_raw_rules	<i>Extract raw rules</i>
-------------------	--------------------------

Description

Separate the seven kinds of rules

Usage

extract_raw_rules(raw_rules)

Arguments

- | | |
|-----------|--------------------------------|
| raw_rules | a charcter with the raw rules. |
|-----------|--------------------------------|

extract_replacement_rules	<i>Extract replacement rules</i>
---------------------------	----------------------------------

Description

Parses the the raw replacement rules.

Usage

```
extract_replacement_rules(raw_repl)
```

Arguments

`raw_repl` the part with replacement rules for each step rule.

extract_rules	<i>Extract Rules from file</i>
---------------	--------------------------------

Description

This function parse the rules that are disponible in the RLSP package disponible in the RSLP C source. This file has been downloaded and is installed with the package. It's path can be found using `system.file("steprules.txt", package = "rslp")` A parsed version is saved is also installed with the package and its path can be found using `system.file("steprules.rds", package = "rslp")`.

Usage

```
extract_rules(path = system.file("steprules.txt", package = "rslp"))
```

Arguments

`path` path to the raw steprules. Most of the times you don't have to change it.

extract_rules_info	<i>Extract Rules Info</i>
--------------------	---------------------------

Description

Extract all info from all rules

Usage

```
extract_rules_info(rules)
```

Arguments

rules	rules parsed before by extract_rule_info
-------	--

extract_rule_info	<i>Extract Rule Info</i>
-------------------	--------------------------

Description

Extract all info for one rule

Usage

```
extract_rule_info(rule)
```

Arguments

rule	the rule you want to extract infos
------	------------------------------------

remove_accents	<i>Remove Acccents</i>
----------------	------------------------

Description

A wrapppper for stringi package.

Usage

```
remove_accents(s)
```

Arguments

s	the string you want to remove accents
---	---------------------------------------

rslp

RSLP

Description

Apply the Stemming Algorithm for the Portuguese Language to vector of words.

Usage

```
rslp(
  words,
  steprules = readRDS(system.file("steprules.rds", package = "rslp"))
)
```

Arguments

words vector of words that you want to stem.

steprules as obtained from the function `extract_rules`. (only define if you are certain about it). The default is to get the parsed version of the rules installed with the package.

References

V. Orenco, C. Huyck, "A Stemming Algorithmm for the Portuguese Language", SPIRE, 2001, String Processing and Information Retrieval, International Symposium on, String Processing and Information Retrieval, International Symposium on 2001, pp. 0186, doi:10.1109/SPIRE.2001.10024

Examples

```
words <- c("gostou", "gosto", "gostaram")
rslp(words)
```

rslp_

RSLP_

Description

Apply the Stemming Algorithm for the Portuguese Language to a word.

Usage

```
rslp_(
  word,
  steprules = readRDS(system.file("steprules.rds", package = "rslp"))
)
```

Arguments

word	word to be stemmed.
steprules	as obtained from the function <code>extract_rules</code> .

rslp_doc

*RSLP Document***Description**

Apply the Stemming Algorithm for the Portuguese Language to vector of documents. It extracts words using the regex `"\b[:alpha:]\b"`

Usage

```
rslp_doc(
  docs,
  steprules = readRDS(system.file("steprules.rds", package = "rslp"))
)
```

Arguments

docs	chr vector of documents
steprules	as obtained from the function <code>extract_rules</code> . (only define if you are certain about it). The default is to get the parsed version of the rules installed with the package.

References

V. Orenge, C. Huyck, "A Stemming Algorithm for the Portuguese Language", SPIRE, 2001, String Processing and Information Retrieval, International Symposium on, String Processing and Information Retrieval, International Symposium on 2001, pp. 0186, doi:10.1109/SPIRE.2001.10024

Examples

```
docs <- c("coma frutas pois elas fazem bem para.")
rslp_doc(docs)
```

verify_sufix	<i>Verify</i>
--------------	---------------

Description

Given a list of suffixes, returns a vector of true or false indicating if the word has each one of the suffixes.

Usage

```
verify_sufix(word, rep_rules)
```

Arguments

word	word you which to verify replacement rules
rep_rules	data.frame of rules as specified in <code>steprules\$replacement_rule</code>

Index

`apply_rules`, [2](#)

`extract_raw_rules`, [2](#)

`extract_replacement_rules`, [3](#)

`extract_rule_info`, [4](#), [4](#)

`extract_rules`, [3](#)

`extract_rules_info`, [4](#)

`remove_accents`, [4](#)

`rslp`, [5](#)

`rslp_`, [5](#)

`rslp_doc`, [6](#)

`verify_suffix`, [7](#)