

Package ‘hyphenatr’

August 29, 2016

Encoding UTF-8

Title Tools to Hyphenate Strings Using the 'Hunspell' Hyphenation Library

Version 0.3.0

Description Identifying hyphenation points in strings can be useful for both text processing and display functions. The 'Hunspell' hyphenation library <<https://github.com/hunspell/hyphen>> provides tools to perform hyphenation using custom language rule dictionaries. Many hyphenation rules dictionaries are included. Words can be hyphenated directly or split into hyphenated component strings for further processing.

Depends R (>= 3.1.0)

NeedsCompilation yes

Copyright file inst/COPYRIGHTS

License GPL-3 + file LICENSE

LazyData true

Suggests testthat

Imports Rcpp, stringi

RoxygenNote 5.0.1

LinkingTo Rcpp

Author Bob Rudis [aut, cre],
Raph Levien [aut, cph] (original libhnpj),
Rene Engelhard [aut, cph] (libhnpj maintainer),
Chris Halls [aut, cph] (libhnpj maintainer),
Peter Novodvorsky [aut, cph] (libhnpj OOo integration),
László Németh [aut, cph] (libhnpj non-standard and compound hyphenation with Unicode support),
Nanning Buitenhuis [aut, cph] (libhnpj substrings.c),
Edwin de Jonge [ctb] (Language testing),
Gergely Daróczi [ctb] (Language testing)

Maintainer Bob Rudis <bob@rudis.net>

Repository CRAN

Date/Publication 2016-03-18 07:21:39

R topics documented:

curr_dict	2
hyphenate	2
hyphenatr	3
hyphendict_is_loaded	4
list_dicts	4
switch_dict	5

Index	6
--------------	----------

curr_dict	<i>Identify current hyphen rules language</i>
-----------	---

Description

Identify current hyphen rules language

Usage

```
curr_dict()
```

Examples

```
curr_dict()
```

hyphenate	<i>Hyphenate a character vector of words</i>
-----------	--

Description

Given a character vector (one word per element), this function will hyphenate the strings or return a list of separated hyphenated string components.

Usage

```
hyphenate(words, simplify = TRUE)
```

Arguments

words	character vector of words
simplify	if TRUE, will return words with = as the hyphen character. If FALSE, will return a list of separated, hyphenated word components. If a string (e.g. "-", "­" or '­') will use that character for the hyphen character.)

Value

a character vector or a list depending on the value of `simplify`.

Note

The `en_US` hyphenation rules dictionary is loaded by default

Examples

```
dat <- readLines(system.file("extdata/top10000en.txt", package="hyphenatr"))

out1 <- hyphenate(dat)
out2 <- hyphenate(dat, simplify=FALSE)
out3 <- hyphenate(dat, simplify="-")
out4 <- hyphenate(dat, simplify="&shy;")
```

hyphenatr

Tools to Hyphenate Strings Using the 'Hunspell' Hyphenation Library

Description

Identifying hyphenation points in strings can be useful for both text processing and display functions. The 'Hunspell' hyphenation library provides tools to perform hyphenation using custom language rule dictionaries. Words can be hyphenated directly or split into hyphenated component strings for further processing.

Details

Based on the [Hunspell hyphenation library](#) and includes hyphenation rules dictionaries from [Libre-Office](#).

Special thanks to Edwin de Jonge & Gergely Daróczi for language testing.

Note

The `en_US` hyphenation rules dictionary is loaded by default

Author(s)

Bob Rudis (@hrbrmstr)

hyphendict_is_loaded *Test to see if the hyphenation language rules dictionary is loaded*

Description

Test to see if the hyphenation language rules dictionary is loaded

Usage

```
hyphendict_is_loaded()
```

Note

The en_US hyphenation rules dictionary is loaded by default

list_dicts *List available hyphenation languages rules*

Description

The vast majority **OpenOffice Hunspell hyphenation rules dictionaries** are named in the format `location.code_COUNTRY.CODE` (e.g. en_US). This function returns the list of available dictionaries. Used the desired, supplied result in your call to [switch_dict](#).

Usage

```
list_dicts()
```

Note

The en_US hyphenation rules dictionary is loaded by default

Examples

```
list_dicts()
```

switch_dict	<i>Switch hyphen rules language</i>
-------------	-------------------------------------

Description

See References for the source of the hyphenation rules. For valid values, use `list_dicts()`.

Usage

```
switch_dict(lang)
```

Arguments

lang	hyphen rules language
------	-----------------------

Note

The en_US hyphenation rules dictionary is loaded by default

References

<https://github.com/LibreOffice/dictionaries>

Examples

```
switch_dict("hu_HU")
```

Index

`curr_dict`, 2

`hyphenate`, 2

`hyphenatr`, 3

`hyphenatr-package (hyphenatr)`, 3

`hyphendict_is_loaded`, 4

`list_dicts`, 4

`switch_dict`, 4, 5