

Package ‘pullword’

July 13, 2021

Title R Interface to Pullword Service

Description R Interface to Pullword Service for natural language processing in Chinese. It enables users to extract valuable words from text by deep learning models. For more details please visit the official site (in Chinese) <<http://www.pullword.com/>>.

Version 0.3

Author Tong He <hetong007@gmail.com>

Maintainer Tong He <hetong007@gmail.com>

BugReports <https://github.com/hetong007/pullword/issues>

Depends R (>= 3.0.2),RCurl

License LGPL

NeedsCompilation yes

Repository CRAN

Date/Publication 2021-07-13 09:10:02 UTC

R topics documented:

isBIG5	2
isGB18030	2
isGB2312	3
isGBK	3
isUTF8	4
pullword	4
toUTF8	5

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

isBIG5

Indicate whether the encoding of input string is BIG5.

Description

Indicate whether the encoding of input string is BIG5.

Usage

```
isBIG5(string, combine = FALSE)
```

Arguments

string	A character vector.
combine	Whether to combine all the strings.

Value

Logical value.

Author(s)

Jian Li <<rweibo@sina.com>>

isGB18030

Indicate whether the encoding of input string is GB18030.

Description

Indicate whether the encoding of input string is GB18030.

Usage

```
isGB18030(string, combine = FALSE)
```

Arguments

string	A character vector.
combine	Whether to combine all the strings.

Value

Logical value.

Author(s)

Jian Li <<rweibo@sina.com>>

`isGB2312`*Indicate whether the encoding of input string is GB2312.*

Description

Indicate whether the encoding of input string is GB2312.

Usage

```
isGB2312(string, combine = FALSE)
```

Arguments

<code>string</code>	A character vector.
<code>combine</code>	Whether to combine all the strings.

Value

Logical value.

Author(s)

Jian Li <<rweibo@sina.com>>

`isGBK`*Indicate whether the encoding of input string is GBK.*

Description

Indicate whether the encoding of input string is GBK.

Usage

```
isGBK(string, combine = FALSE)
```

Arguments

<code>string</code>	A character vector.
<code>combine</code>	Whether to combine all the strings.

Value

Logical value.

Author(s)

Jian Li <<rweibo@sina.com>>

isUTF8	<i>Indicate whether the encoding of input string is UTF-8.</i>
--------	--

Description

Indicate whether the encoding of input string is UTF-8.

Usage

```
isUTF8(string, combine = FALSE)
```

Arguments

string	A character vector.
combine	Whether to combine all the strings.

Value

Logical value.

Author(s)

Jian Li <<rweibo@sina.com>>

pullword	<i>R Interface of PullWord service</i>
----------	--

Description

This function deals with communication with the server. The result will be parsed in a vector or a matrix, depends on show probability or not

pullword

Usage

```
pullword(input = NULL, file = NULL, threshold = 0, showProb = FALSE)
```

Arguments

input	The input text.
file	The input file.
threshold	The minimum probability for the words appearing in the result. Should be a real value between 0 and 1.
showProb	logical. The return value would be a data.frame if TRUE, or a vector otherwise.

Examples

```
require(pullword)
pullword('Replace this field with a Chinese sentence.', threshold=0, showProb=TRUE)
```

`toUTF8`*Convert encoding of Chinese string to UTF-8.*

Description

Convert encoding of Chinese string to UTF-8.

Usage

```
toUTF8(cnstring)
```

Arguments

`cnstring` A Chinese string vector.

Value

Converted vectors.

Author(s)

Jian Li <<rweibo@sina.com>>

Index

`isBIG5`, [2](#)

`isGB18030`, [2](#)

`isGB2312`, [3](#)

`isGBK`, [3](#)

`isUTF8`, [4](#)

`pullword`, [4](#)

`pullword-package (pullword)`, [4](#)

`toUTF8`, [5](#)