Package ‘spellcheckr’

October 28, 2016

Type Package
Title Correct the Spelling of a Given Word in the English Language
Version 0.1.2
Author Selva Prabhakaran
Maintainer Selva Prabhakaran <selva86@gmail.com>
Description Corrects the spelling of a given word in English using a modification of Peter Norvig’s spell correct algorithm (see <http://norvig.com/spell-correct.html>) which handles up to three edits. The algorithm tries to find the spelling with maximum probability of intended correction out of all possible candidate corrections from the original word.
License GPL-2 | GPL-3
Encoding UTF-8
LazyData true
Imports data.table, dplyr, stringr
RoxygenNote 5.0.1
NeedsCompilation no
Depends R (>= 2.10)
Repository CRAN
Date/Publication 2016-10-28 10:04:43

R topics documented:
correct ................................................................. 2
dict ................................................................. 3
Index 4
Description

Correct the spelling of a given word in the English language.

Usage

```
correct(wd = NULL, maxedit = 2, dictionary = NULL)
```

Arguments

- **wd**: Character. The word to be spell corrected.
- **maxedit**: Integer. The maximum number of edits allowed to reach the correct word. Max allowed is 3.
- **dictionary**: Character Vector of eligible words to be considered. Repeated words will get more weightage.

Details

This is based on Peter Norvig's spell correct algorithm [http://norvig.com/spell.py](http://norvig.com/spell.py). But this one is modified to handle up to three edits.

Value

The corrected word.

Author(s)

Selva Prabhakaran <selva86@gmail.com>

Examples

```
data(dict)
correct("scaret")
correct("beliebe")
```
### dict

<table>
<thead>
<tr>
<th>dict</th>
<th>dict</th>
</tr>
</thead>
</table>

**Description**

A dataset containing the valid English dictionary words originally obtained from a work of Jane Austen.

**Usage**

```
data(dict)
```

**Format**

A data frame with 245462 rows and 2 columns

**Details**

- . The actual word
- N The number of occurrences of the word in the document.
Index

*Topic datasets
dict, 3

correct, 2
dict, 3
splits (dict), 3