Package ‘unine’

Type     Package
Title    Unine Light Stemmer
Version  0.2.0
Maintainer Michaël Benesty <michael@benesty.fr>
Description Implementation of "light" stemmers for French, German, Italian, Spanish, Portuguuese, Finnish, Swedish. They are based on the same work as the "light" stemmers found in 'SolR' <https://lucene.apache.org/solr/> or 'ElasticSearch' <https://www.elastic.co/fr/products/elasticsearch>. A "light" stemmer consists in removing inflections only for noun and adjectives. Indexing verbs for these languages is not of primary importance compared to nouns and adjectives. The stemming procedure for French is described in (Savoy, 1999) <doi:10.1002/(SICI)1097-4571(1999)50:10%3C944::AID-ASI9%3E3.3.CO;2-H>.
URL     https://github.com/pommedeterresautee/unine,
        https://pommedeterresautee.github.io/unine/,
        http://members.unine.ch/jacques.savoy/clef/
BugReports https://github.com/pommedeterresautee/unine/issues
License MIT + file LICENSE
Encoding UTF-8
LazyData true
LinkingTo Rcpp
Imports Rcpp (>= 1.0.0), methods
SystemRequirements C++11
RoxygenNote 6.1.1
NeedsCompilation yes
Suggests testthat, covr, stringi
Author Michaël Benesty [aut, cre, cph], Jacques Savoy [cph]
Repository CRAN
Date/Publication 2019-04-13 22:24:09 UTC
**R topics documented:**

- finnish_stemmer .................................................. 2
- finnish_stoplist .................................................. 3
- french_stemmer .................................................. 3
- french_stoplist .................................................. 4
- german_stemmer .................................................. 4
- german_stoplist .................................................. 5
- italian_stemmer .................................................. 5
- italian_stoplist .................................................. 6
- portuguese_stemmer .............................................. 6
- portuguese_stoplist .............................................. 7
- spanish_stemmer .................................................. 7
- spanish_stoplist .................................................. 8
- swedish_stemmer .................................................. 8
- swedish_stoplist .................................................. 9
- unine ............................................................... 9

**Index**

<table>
<thead>
<tr>
<th>finnish_stemmer</th>
<th>Stem Finnish words</th>
</tr>
</thead>
</table>

**Description**

Stemmer for Finnish words

**Usage**

```
finnish_stemmer(words)
```

**Arguments**

- `words` a character containing the original words.

**Value**

character with stemmed words.

**Examples**

```
finnish_stemmer(c("taivas"))
```
finnish_stoplist  

Description
Finnish stop list

Usage
finnish_stoplist

Format
character containing the list of words

Source
Unine website

Examples

data("finnish_stoplist")

french_stemmer  

Description
Stemmer for French words

Usage
french_stemmer(words)

Arguments
words a character containing the original words.

Value
character with stemmed words.

Examples
french_stemmer(c("tester", "testament", "clients"))
french_stoplist  

Description
French stop list

Usage
french_stoplist

Format
character containing the list of words

Source
Unine website

Examples
data("french_stoplist")

german_stemmer  

Description
Stemmer for German words

Usage
german_stemmer(words)

Arguments
words a character containing the original words.

Value
character with stemmed words.

Examples
german_stemmer(c("kinder"))
german_stoplist

German stop list

Usage

german_stoplist

Format

character containing the list of words

Source

Unine website

Examples

data("german_stoplist")

---

italian_stemmer

Stem Italian words

Description

Stemmer for Italian words

Usage

italian_stemmer(words)

Arguments

words a character containing the original words.

Value

character with stemmed words.

Examples

italian_stemmer(c("arrivederci"))
### italian_stoplist  
**Italian stop list**

**Description**  
Italian stop list

**Usage**  
```r
italian_stoplist
```

**Format**  
- `character` containing the list of words

**Source**  
Unine website

**Examples**  
```r
data("italian_stoplist")
```

### portuguese_stemmer  
**Stem Portuguese words**

**Description**  
Stemmer for Portuguese words

**Usage**  
```r
portuguese_stemmer(words)
```

**Arguments**  
- `words`: a `character` containing the original words.

**Value**  
`character` with stemmed words.

**Examples**  
```r
portuguese_stemmer(c("adeus"))
```
**portuguese_stoplist**  

**Portuguese stop list**

**Description**
Portuguese stop list

**Usage**
portuguese_stoplist

**Format**
character containing the list of words

**Source**
Unine website

**Examples**
data("portuguese_stoplist")

---

**spanish_stemmer**  

**Stem Spanish words**

**Description**
Stemmer for Spanish words

**Usage**
spanish_stemmer(words)

**Arguments**
words  

**Value**
character with stemmed words.

**Examples**
spanish_stemmer(c("perros"))
**spanish_stoplist**  *Spanish stop list*

**Description**
Spanish stop list

**Usage**
spanish_stoplist

**Format**
- character containing the list of words

**Source**
Unine website

**Examples**
```
data("spanish_stoplist")
```

---

**swedish_stemmer**  *Stem Swedish words*

**Description**
Stemmer for Swedish words

**Usage**
swedish_stemmer(words)

**Arguments**
- `words`  a character containing the original words.

**Value**
- character with stemmed words.

**Examples**
```
swedish_stemmer(c("stiga"))
```
swedish_stoplist

---

**Description**

Swedish stop list

**Usage**

swedish_stoplist

**Format**

- character containing the list of words

**Source**

Unine website

**Examples**

data("swedish_stoplist")

---

unine

---

**unine: Unine Light Stemmer**

---

**Description**

Implementation of "light" stemmers for French, German, Italian, Spanish, Portuguese, Finnish, Swedish. They are based on the same work as the "light" stemmers found in 'SolR' [https://lucene.apache.org/solr/] or 'ElasticSearch' [https://www.elastic.co/fr/products/elasticsearch]. A "light" stemmer consists in removing inflections only for noun and adjectives. Indexing verbs for these languages is not of primary importance compared to nouns and adjectives. The stemming procedure for French is described in (Savoy, 1999) [doi:10.1002/(SICI)1097-4571(1999)50:10

**Author(s)**

- **Maintainer**: Michaël Benesty <michael@benesty.fr> [copyright holder]

Other contributors:

- Jacques Savoy <Jacques.Savoy@unine.ch> [copyright holder]
See Also

Useful links:

- https://github.com/pommedeterresautee/unine
- https://pommedeterresautee.github.io/unine/
- http://members.unine.ch/jacques.savoy/clef/
- Report bugs at https://github.com/pommedeterresautee/unine/issues
Index

* datasets
  finnish_stoplist, 3
  french_stoplist, 4
  german_stoplist, 5
  italian_stoplist, 6
  portuguese_stoplist, 7
  spanish_stoplist, 8
  swedish_stoplist, 9

character, 2–9

finnish_stemmer, 2
finnish_stoplist, 3
french_stemmer, 3
french_stoplist, 4

german_stemmer, 4
german_stoplist, 5

italian_stemmer, 5
italian_stoplist, 6

portuguese_stemmer, 6
portuguese_stoplist, 7

spanish_stemmer, 7
spanish_stoplist, 8
swedish_stemmer, 8
swedish_stoplist, 9

unine, 9
unine-package (unine), 9