

Package ‘Snowball’

April 18, 2012

Version 0.0-8

Date 2012-04-18

Title Snowball Stemmers

Author Kurt Hornik

Maintainer Kurt Hornik <Kurt.Hornik@R-project.org>

Description Snowball stemmers.

Imports RWeka (>= 0.3-19), rJava (>= 0.6-3)

License BSD

Repository CRAN

Date/Publication 2012-04-18 09:45:23

R topics documented:

SnowballStemmer	1
Index	3

SnowballStemmer	<i>R/Weka Snowball Stemmer</i>
-----------------	--------------------------------

Description

R interface to Weka’s Snowball stemmers.

Usage

```
SnowballStemmer(x, control = NULL)
```

Arguments

x	a character vector with words to be stemmed.
control	an object of class <code>Weka_control</code> , or a character vector of control options, or NULL (default). Available options can be obtained on-line using the Weka Option Wizard WOW , or the Weka documentation.

Details

The Snowball stemmers contain the Porter stemmer and several other stemmers for different languages. See <http://snowball.tartarus.org/> for more information.

SnowballStemmer is an interface to Weka's wrapper classes for the Java version of the Snowball stemmers. The corresponding jar cannot be included in package **RWeka** due to license restrictions, and hence is made available via the separate package **Snowball**.

The Omegahat package **Rstem** provides an R interface to a C version of Porter's word stemming algorithm.

Value

A character vector with the stemmed words.

Note

MacOS X users may need to perform `Sys.setenv(NOAWT = TRUE)` before loading **Snowball** (and any other package using **rJava**).

See Also

Other R interfaces to Weka stemmers ([RWeka_stemmers](#))

Examples

```
## Test the supplied vocabulary for the default stemmer ('porter'):
source <- readLines(system.file("words", "porter", "voc.txt",
                               package = "Snowball"))
result <- SnowballStemmer(source)
target <- readLines(system.file("words", "porter", "output.txt",
                               package = "Snowball"))
## Any differences?
any(result != target)
```

Index

*Topic **character**
SnowballStemmer, 1

RWeka_stemmers, 2

SnowballStemmer, 1

Weka_control, 2

WOW, 2