Package: CSeqpat (via r-universe)

September 14, 2024

Type Package

Title Frequent Contiguous Sequential Pattern Mining of Text

Version 0.1.2

Author Anantha Janakiraman

Maintainer Anantha Janakiraman <anantharaman.j@gmail.com>

Description Mines contiguous sequential patterns in text.

Depends R (>= 3.1.0)

Imports NLP, tm, utils

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Date/Publication 2018-07-27 04:10:03 UTC

Repository https://hz6yc3.r-universe.dev

RemoteUrl https://github.com/cran/CSeqpat

RemoteRef HEAD

RemoteSha 6f3196ef0a8ee601d8b350f58393b418b5eaff13

Contents

CSeqpat *Mining Frequent Contiguous Sequential Patterns in a Text Corpus*

Description

Takes in the filepath and minimum support and performs pattern mining

Usage

```
CSeqpat(filepath, phraselenmin = 1, phraselenmax = 99999, minsupport = 1,
  docdelim, stopword = FALSE, stemword = FALSE, lower = FALSE,
  removepunc = FALSE)
```

Arguments

filepath	Path to the text file/text corpus
phraselenmin	Minimum number of words in a phrase
phraselenmax	Maximum number of words in a phrase
minsupport	Minimum absolute support for mining the patterns
docdelim	Document delimiter in the corpus
stopword	Remove stopwords from the document corpus (boolean)
stemword	Perform stemming on the document corpus (boolean)
lower	Lower case all words in document corpus (boolean)
removepunc	Remove punctuations from document corpus (boolean)

Value

A dataframe containing the frequent phrase patterns with their absolute support

Examples

```
test1 <- c("hoagie institution food year road ",
"place little dated opened weekend fresh food")
tf <- tempfile()
writeLines(test1, tf)
CSeqpat(tf,1,2,2,"\t",TRUE,FALSE,TRUE,FALSE)
```

Index