

# Package ‘CTM’

August 23, 2016

**Type** Package

**Title** A Text Mining Toolkit for Chinese Document

**Version** 0.1

**Date** 2016-08-23

**Author** Jim Liu

**Maintainer** Jim Liu <jimliu741523@gmail.com>

**Description** The CTM package is designed to solve problems of text mining and is specific for Chinese document.

**License** GPL-3

**LazyData** TRUE

**RoxygenNote** 5.0.1

**Imports** jiebaR

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2016-08-23 15:25:02

## R topics documented:

CDTM . . . . .	2
CTDM . . . . .	3
termCount . . . . .	4
<b>Index</b>	<b>5</b>

---

CDTM

*Document Term Matrix*

---

### **Description**

Constructs Document-Term Matrix from Chinese Text Documents.

### **Usage**

```
CDTM(doc, weighting, shortTermDeleted)
```

### **Arguments**

<code>doc</code>	The Chinese text document.
<code>weighting</code>	Available weighting function with matrix are <code>tf</code> , <code>tfidf</code> .
<code>shortTermDeleted</code>	Detect short word when <code>nchar &lt; 2</code> .

### **Details**

This function runs a Chinese word segmentation by `jiebaR` and builds a document-term matrix and there are two weighting functions with matrix, term frequency and term frequency inverse document frequency.

### **Author(s)**

Jim Liu

### **Examples**

```
library(CTM)
a <- "hello taiwan"
b <- "world of tank"
c <- "taiwan weather"
d <- "local weather"
text <- t(data.frame(a,b,c,d))
dtm <- CDTM(doc = text, weighting = "tfidf", shortTermDeleted = FALSE)
```

---

CTDM

*Term Document Matrix*

---

### **Description**

Constructs Term-Document Matrix from Chinese Text Documents.

### **Usage**

```
CTDM(doc, weighting, shortTermDeleted)
```

### **Arguments**

doc	The Chinese text document.
weighting	Available weighting function with matrix are tf, tfidf.
shortTermDeleted	Deltected short word when nchar <2.

### **Details**

This function run a Chinese word segmentation by jiebeR and build term-document matrix, and there is two weighting function with matrix, term frequency and term frequency inverse document frequency.

### **Author(s)**

Jim Liu

### **Examples**

```
library(CTM)
a <- "hello taiwan"
b <- "world of tank"
c <- "taiwan weather"
d <- "local weather"
text <- t(data.frame(a,b,c,d))
tdm <- CTDM(doc = text, weighting = "tfidf", shortTermDeleted = FALSE)
```

---

termCount	<i>Term Count</i>
-----------	-------------------

---

**Description**

Computing term count from text documents

**Usage**

```
termCount(doc, shortTermDeleted)
```

**Arguments**

doc                   The Chinese text document.  
shortTermDeleted       Deltected short word when nchar <2.

**Details**

This function run a Chinese word segmentation by jiebeR and compute term count from these text document.

**Author(s)**

Jim Liu

**Examples**

```
library(CTM)
a <- "hello taiwan"
b <- "world of tank"
c <- "taiwan weather"
d <- "local weather"
text <- t(data.frame(a,b,c,d))
count <- termCount(doc = text, shortTermDeleted = FALSE)
```

# Index

CDTM, [2](#)

CTDM, [3](#)

termCount, [4](#)