

# Package ‘malaytextr’

August 17, 2021

**Title** Text Mining for Bahasa Malaysia

**Version** 0.1.1

**Description** It is built to handle Bahasa Malaysia text. We provide functions and data sets that will help handling Bahasa Malaysia to be much easier. For word stemming in particular, we will find the Malay words in a dictionary and then proceed to remove “extra suffix” as explained in Khan, Rehman Ullah, Fitri Suraya Mohamad, Muh Inam UIHaq, Shahren Ahmad Zadi Adruce, Philip Nuli Anding, Sajjad Nawaz Khan, and Abdulrazak Yahya Saleh Al-Hababi (2017) <<https://ijrest.net/vol-4-issue-12.html>>. A dictionary of Malay words provided in this package can be used as a dictionary to perform word stemming.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.1

**URL** <https://github.com/zahiernasrudin/malaytextr>

**BugReports** <https://github.com/zahiernasrudin/malaytextr/issues>

**Imports** dplyr, magrittr, rlang, stringr

**Depends** R (>= 2.10)

**Suggests** rmarkdown, knitr, testthat (>= 3.0.0)

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**NeedsCompilation** no

**Author** Zahier Nasrudin [aut, cre] (<<https://orcid.org/0000-0002-7060-776X>>)

**Maintainer** Zahier Nasrudin <[zahiernasrudin@gmail.com](mailto:zahiernasrudin@gmail.com)>

**Repository** CRAN

**Date/Publication** 2021-08-17 16:30:21 UTC

## R topics documented:

|                          |   |
|--------------------------|---|
| malayrootwords . . . . . | 2 |
| remove_url . . . . .     | 2 |
| stem_malay . . . . .     | 3 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>5</b> |
|--------------|----------|

---

|                |                                 |
|----------------|---------------------------------|
| malayrootwords | <i>Data of Malay root words</i> |
|----------------|---------------------------------|

---

### Description

Data of Malay root words

### Usage

```
malayrootwords
```

### Format

A tibble with 4244 rows and 2 variables:

Col Word dbl Malay Word

Root Word dbl Malay Root Word

---

|            |                         |
|------------|-------------------------|
| remove_url | <i>Remove URL links</i> |
|------------|-------------------------|

---

### Description

Remove URL links

### Usage

```
remove_url(string)
```

### Arguments

string           String to change

### Details

remove\_url() is an approach to remove link(s) from a string

### Value

Returns a string with URL links removed

## Examples

```
x <- c("test https://t.co/fkQC2dXwnc", "another one https://www.google.com/ to try")
remove_url(x)
```

---

|            |                             |
|------------|-----------------------------|
| stem_malay | <i>Stemming Malay words</i> |
|------------|-----------------------------|

---

## Description

Malaytextr function to stem Malay words

## Usage

```
stem_malay(Word,
  dictionary,
  col_feature1,
  col_dict1,
  col_dict2)
```

## Arguments

|              |  |
|--------------|--|
| Word         | A data frame, or a character vector  |
| dictionary   | A data frame with a column of words to be stemmed and a column of root words |
| col_feature1 | Column that contains words to be stemmed from Word                           |
| col_dict1    | Column that will be used to match with col_feature1 from Word                |
| col_dict2    | Column that contains the root words from dictionary                          |

## Format

An object of class function of length 1.

## Details

stem\_malay() is an approach to find the Malay words in a dictionary and then proceed to remove "extra suffix" as explained by Khan et al. (2017), and then "prefix" and lastly, "suffix".

## Value

Returns a data frame with the following properties:

- Col Word: Renamed input from Word
- root\_word: An additional column which contains the word(s) after being stemmed.

## References

Khan, Rehman Ullah, Fitri Suraya Mohamad, Muh Inam UIHaq, Shahren Ahmad Zadi Adruce, Philip Nuli Anding, Sajjad Nawaz Khan, and Abdulrazak Yahya Saleh Al-Hababi. 2017. "Malay Language Stemmer."

**Examples**

```
#Specifying a character vector &
#use a dictionary from malaytextr package

stem_malay(Word = "banyaknya", dictionary = malayrootwords)

#A data frame,
#Use a dictionary from malaytextr package,
#With a dataframe, you will need to specify the column to be stemmed

x <- data.frame(text = c("banyaknya", "sangat", "terkedu", "pengetahuan"))

stem_malay(Word = x, dictionary = malayrootwords, col_feature1 = "text")
```

# Index

## \* **datasets**

malayrootwords, [2](#)

stem\_malay, [3](#)

malayrootwords, [2](#)

remove\_url, [2](#)

stem\_malay, [3](#)