

tokenize_text_documentation

Mackenzie Ramalho

2024-04-26

Function 2: 'tokenize_text()'

This function tokenizes the text into individual words or tokens.

Parameters:

- 'text': A character vector containing the text to be tokenized.

Returns:

A character vector containing individual words or tokens extracted from the input text.

```
# Function to tokenize text
tokenize_text <- function(text) {
  tokens <- strsplit(text, "\\s+")[1]
  return(tokens)
}

#Example
tokens <- tokenize_text("This is a sample sentence.")

tokens
```

```
## [1] "This"      "is"        "a"         "sample"    "sentence."
```

Use Case:

This function is useful for breaking down text data into smaller units, such as words or tokens, which can then be further analyzed or processed.