

clean_text_documentation

Mackenzie Ramalho

2024-04-26

Function 1: 'clean_text()'

This function cleans text data by removing special characters, numbers, and extra white spaces.

Parameters:

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

Returns:

A cleaned version of the input text with special characters, numbers, and extra white spaces removed, and all letters converted to lowercase.

```
# Function to clean text
clean_text <- function(text) {
  # Remove special characters
  cleaned_text <- gsub("[^a-zA-Z\\s]", "", text)
  # Remove extra white spaces
  cleaned_text <- gsub("\\s+", " ", cleaned_text)
  # Convert to lowercase
  cleaned_text <- tolower(cleaned_text)
  return(cleaned_text)
}

#Example
cleaned_text <- clean_text("Sample text with special characters: !@#%&^&*()")

cleaned_text
```

```
## [1] "sampletextwithspecialcharacters"
```

Use Case:

This function is particularly useful for pre processing text data before performing natural language processing tasks.