find_keywords_documentation

Mackenzie Ramalho

2024-04-26

Function 4: 'find_keywords()'

This function identifies keywords in the text based on their frequency or relevance.

Parameters:

- 'text': A character vector containing the text from which keywords are to be extracted.
- 'threshold': An integer specifying the minimum frequency threshold for a word to be considered a keyword. Defaults to 1.

Returns:

A character vector containing the keywords identified in the input text based on their frequency or relevance.

```r
calculate_word_frequency <- function(text) {
  tokens <- tokenize_text(text)
  word_freq <- table(tokens)
  return(word_freq)
}

tokenize_text <- function(text) {
  tokens <- strsplit(text, "\\s+")[[1]]
  return(tokens)
}
# Function to find keywords
find_keywords <- function(text, threshold = 1) {
  word_freq <- calculate_word_frequency(text)
  keywords <- names(word_freq[word_freq >= threshold])
  return(keywords)
}

#Example
keywords <- find_keywords("This is a sample sentence. Sample sentence contains sample words.")

keywords
```

```
## [1] "a"         "contains"  "is"        "sample"    "Sample"    "sentence"
## [7] "sentence." "This"      "words."
```

Use Case:

This function is beneficial for extracting important keywords from text data, which can provide insights into the main topics or themes present in the text. It is commonly used in text analysis tasks such as document summarization, keyword extraction, and content categorization.