How can I use grepl with multiple patterns in order to substitute the function of the -R flag?

Authored by

stats writer

June 24, 2024

Grepl is a function in R that allows for pattern matching within a string. It is commonly used with a single pattern to find and extract specific information from a string. However, by using multiple patterns with the grepl function, it can serve as a substitute for the -R flag in the command line interface. This means that instead of manually searching for and replacing multiple patterns in a file, grepl can perform the task efficiently and accurately. By providing a list of patterns to search for, grepl can quickly scan through a string and identify all matching patterns, making it a useful tool for data manipulation and analysis in R.

R: Use grepl with Multiple Patterns

You can use the following basic syntax with the grepl() function in R to filter for rows in a data frame that contain one of several string patterns in a specific column:

```
library(dplyr)

new_df <- filter(df, grepl(paste(my_patterns, collapse='|'), my_column))
```

This particular syntax filters the data frame for rows where the value in the column called my_column contains one of the string patterns in the vector called my_patterns.

The following example shows how to use this syntax in practice.

Example: How to Use grepl() with Multiple Patterns in R

Suppose we have the following data frame in R that contains information about various basketball teams:

```
#create data frame
df <- data.frame(team=c('Mavs', 'Hawks', 'Nets', 'Heat', 'Cavs'),
points=c(104, 115, 124, 120, 112),
status=c('Bad', 'Good', 'Excellent', 'Great', 'Bad'))

#view data frame
df
```

```
team points status
1 Mavs 104 Bad
2 Hawks 115 Good
3 Nets 124 Excellent
4 Heat 120 Great
5 Cavs 112 Bad
```

Suppose we would like to filter the data frame to only contain rows where the string in the status column contains one of the following string patterns:

'Good''Gre''Ex'

We can use the following syntax with the grepl() function to do so:

```
library(dplyr)

#define patterns to search for
my_patterns <- c('Good', 'Gre', 'Ex')

#filter for rows where status column contains one of several strings
new_df <- filter(df, grepl(paste(my_patterns, collapse='|'), status))

#view results
new_df

team points status
1 Hawks 115 Good
2 Nets 124 Excellent
3 Heat 120 Great
```

Notice that the data frame has been filtered to only contain the rows where the string in the status column contains one of the three patterns that we specified.

Note that by using the paste() function with the argument collapse='|' we actually searched for the string 'Good|Gre|Ex' in the status column.

Since the | symbol in R stands for "OR" we were able to search for rows that contained 'Good' or Gre' or 'Ex' in the status column.