

Espressioni regolari in R

La sintassi accettata dalle funzioni `grep()` (`grepl`, `regexpr`, `gregexpr`, `sub` and `gsub`), e `strsplit()` è presentata nella pagina di aiuto [regex](#), accessibile da console o RStudio con il comando:

```
?base::regex
```

Vedi anche

- il [capitolo 14](#) del libro *R for Data Science* (H. Wickham e G. Grolemund, 2017, in aggiornamento online)
- il [capitolo 17](#) del libro *R Programming for Data Science* (R. D. Peng, 2020)
- voce [Ricerca e sostituzione di testo](#)

Uso nelle funzioni

Le espressioni regolari devono essere inserite fra virgolette (sono stringhe di testo):

```
# per il dataset  
library(dplyr)
```

```
# una parola  
starwars$skin_color[grep("blue", starwars$skin_color)]
```

```
[1] "white, blue"      "blue, grey"      "blue"  
[4] "blue, grey"      "white, blue"     "blue"  
[7] "grey, blue"       "red, blue, white"
```

```
# OR  
starwars$skin_color[grep("blue|white", starwars$skin_color)]
```

```
[1] "white, blue"      "white"           "white, red"      "blue,  
grey"  
[5] "blue"            "blue, grey"      "white, blue"     "white"  
[9] "blue"            "grey, blue"      "red, blue, white" "brown,  
white"
```

```
# nomi che cominciano per L  
starwars$name[grep("^L", starwars$name)]
```

```
[1] "Luke Skywalker"  "Leia Organa"      "Lando Calrissian" "Lobot"
```

```
[5] "Luminara Unduli" "Lama Su"
```

```
# nomi che contengono numeri
starwars$name[grep("\\d", starwars$name)]
```

```
[1] "C-3P0" "R2-D2" "R5-D4" "IG-88" "R4-P17" "BB8"
```

```
# set di caratteri e classi
# nomi che contengono numeri
starwars$name[grep("[0-9]", starwars$name)]
```

`[:digit:]`, `[:alpha:]`, ecc. vengono interpretati 'letteralmente', ovvero come set di caratteri (esempio: ":", d, i, g, i, t"), e vanno inseriti con doppia parentesi quadra:

```
# nomi che contengono numeri
starwars$name[grep("[:digit:]", starwars$name)]
```

Argomenti delle funzioni basate su grep

Tutte le funzioni sopra citate condividono i seguenti argomenti:

- `ignore.case` = FALSE: la ricerca sarà *case sensitive*
- `perl` = FALSE: vengono usate le *extended regexp* di default; indicare TRUE per usare lo stile Perl

Pagina di aiuto su [grep](#)

glob2rx

La funzione `glob2rx()` traduce le espressioni con *wildcard* (asterisco `"*"`) in regex:

```
glob2rx("L*")
```

```
"^L"
```

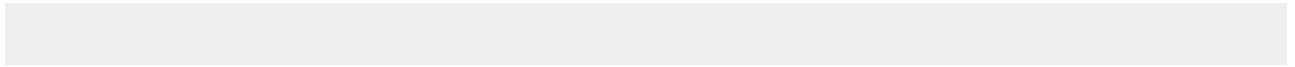
Quindi:

```
# nomi che cominciano per L
starwars$name[grep(glob2rx("L*"), starwars$name)]
```

```
[1] "Luke Skywalker" "Leia Organa" "Lando Calrissian" "Lobot"
[5] "Luminara Unduli" "Lama Su"
```

Script di esempio

nome_file.R



[Concetti di base](#), [Analisi testuale](#), [Gestione dei dati](#), [Coding](#)

From:

<https://www.agnesevardanega.eu/wiki/> - **Ricerca Sociale con R**

Permanent link:

https://www.agnesevardanega.eu/wiki/r/concetti_di_base/regex

Last update: **11/08/2025 14:37**

