

Regex en R

Alin Castillo (alincastillo1995@gmail.com)

Contents

Bibliotecas	1
Anclajes	1
Ejercicio 1: Del vector x extraer todos los elementos que comiencen con la letra a.	1
Ejercicio 2: Del vector x extraer todos los elementos que terminen con la letra a.	1
Ejercicio 3: Del vector x extraer todos los elementos que tengan exactamente 4 letras.	1
Clases de caracteres y alternativas	2
Ejercicio 4: Del vector x extraer los elementos que contienen un punto (.), asterisco (*) y un espacio () en el centro	2
Ejercicio 5: Todas las palabras que comienzan con una vocal	2

Bibliotecas

```
library(stringr)
```

Anclajes

Suele ser útil anclar la expresión regular para que coincida desde el principio o el final de la cadena.

- \wedge para que coincida al comienzo de la cadena
- $\$$ para que coincida al final de la cadena

Ejercicio 1: Del vector x extraer todos los elementos que comiencen con la letra a.

```
x <- c("apple", "banana", "pear")
str_subset(string = x, pattern = "^a")
```

```
## [1] "apple"
```

Ejercicio 2: Del vector x extraer todos los elementos que terminen con la letra a.

```
str_subset(string = x, pattern = "a$")
```

```
## [1] "banana"
```

Ejercicio 3: Del vector x extraer todos los elementos que tengan exactamente 4 letras.

```
str_subset(string = x, pattern = "^....$")
```

```
## [1] "pear"
```

Clases de caracteres y alternativas

- `.`: coincide con cualquier caracter
- `\d`: coincide con cualquier dígito
- `\s`: coincide con cualquier espacio (espacio, tabulador, nueva línea)
- `[abc]`: coincide con a, b o c
- `[^abc]` coincide con cualquier cosa excepto con a, b o c

Ejercicio 4: Del vector x extraer los elementos que contienen un punto (`.`), asterisco (`*`) y un espacio () en el centro

```
x <- c("abc", "a.c", "a*c", "a c")
str_subset(string = x, pattern = "a[.]c")
```

```
## [1] "a.c"
```

```
str_subset(string = x, pattern = "a[*]c")
```

```
## [1] "a*c"
```

```
str_subset(string = x, pattern = "a[ ]c")
```

```
## [1] "a c"
```

Tener presente que usar `[]` funciona para gran parte de caracteres especiales, algunos de los no soportados son:

- `]`
- `\`
- `^`
- `-`

Se puede usar una alternativa para buscar entre una y otra expresión regular con `()`, así por ejemplo:

```
x <- c("grey", "gray")
str_subset(string = x, pattern = "gr(e|a)y")
```

```
## [1] "grey" "gray"
```

Para los siguientes ejemplos usaremos el corpus de palabras comunes `words`

```
palabras <- stringr::words
```

Ejercicio 5: Todas las palabras que comienzan con una vocal

```
str_subset(string = palabras, pattern = "^[aeiou]")
```

```
## [1] "a"      "able"   "about"  "absolute" "accept"
## [6] "account" "achieve" "across" "act"      "active"
## [11] "actual"  "add"     "address" "admit"    "advertise"
## [16] "affect"  "afford"  "after"   "afternoon" "again"
## [21] "against" "age"     "agent"   "ago"      "agree"
## [26] "air"     "all"     "allow"   "almost"   "along"
## [31] "already" "alright" "also"    "although" "always"
## [36] "america" "amount"  "and"     "another"  "answer"
## [41] "any"     "apart"   "apparent" "appear"   "apply"
## [46] "appoint" "approach" "appropriate" "area"     "argue"
## [51] "arm"     "around"  "arrange" "art"      "as"
## [56] "ask"     "associate" "assume"  "at"       "attend"
## [61] "authority" "available" "aware"   "away"     "awful"
```

## [66]	"each"	"early"	"east"	"easy"	"eat"
## [71]	"economy"	"educate"	"effect"	"egg"	"eight"
## [76]	"either"	"elect"	"electric"	"eleven"	"else"
## [81]	"employ"	"encourage"	"end"	"engine"	"english"
## [86]	"enjoy"	"enough"	"enter"	"environment"	"equal"
## [91]	"especial"	"europe"	"even"	"evening"	"ever"
## [96]	"every"	"evidence"	"exact"	"example"	"except"
## [101]	"excuse"	"exercise"	"exist"	"expect"	"expense"
## [106]	"experience"	"explain"	"express"	"extra"	"eye"
## [111]	"idea"	"identify"	"if"	"imagine"	"important"
## [116]	"improve"	"in"	"include"	"income"	"increase"
## [121]	"indeed"	"individual"	"industry"	"inform"	"inside"
## [126]	"instead"	"insure"	"interest"	"into"	"introduce"
## [131]	"invest"	"involve"	"issue"	"it"	"item"
## [136]	"obvious"	"occasion"	"odd"	"of"	"off"
## [141]	"offer"	"office"	"often"	"okay"	"old"
## [146]	"on"	"once"	"one"	"only"	"open"
## [151]	"operate"	"opportunity"	"oppose"	"or"	"order"
## [156]	"organize"	"original"	"other"	"otherwise"	"ought"
## [161]	"out"	"over"	"own"	"under"	"understand"
## [166]	"union"	"unit"	"unite"	"university"	"unless"
## [171]	"until"	"up"	"upon"	"use"	"usual"