

От вектора к текстам в R

Konstantin Ivanin

25 сентября 2017 г

Расстояния и другое

Загрузим библиотеки

```
library(stringr)
library(tidyverse)
library(tidyr)
library(readr)
library(stringi)
library(stringr)
```

Чистка html тегов

В мире достаточно много готовых решений по чистке html и xml тегов, так что настало время изобрести велосипед. Напишите функцию `tagless()`, которая убирает из строки теги.

```
tagless <- function(str) {
  return(str_replace_all(str, '<[^>]*>', ''))
}

tagless('<greeting>Hello, world!</greeting>')
```

```
## [1] "Hello, world!"
```

Лишние пробелы

Напишите функцию `spaceless()`, которая убирает лишние пробелы.

```
spaceless <- function(str) {
  return(str_replace_all(str, '\\s+', ' '))
}

spaceless(c("two spaces", "five spaces"))
```

```
## [1] "two spaces" "five spaces"
```