

rvest

Jan-Philipp Kolb

09 Mai 2017

Das Paket rvest

```
install.packages("rvest")
```

- rvest - Easily Harvest (Scrape) Web Pages

```
library(rvest)
```

```
ht <- read_html('https://www.google.co.in/search?q=guitar+repair+workshop')
links <- ht %>% html_nodes(xpath='//h3/a') %>% html_attr('href')
gsub('/url\\?q=', '', sapply(strsplit(links[as.vector(grep('url', links))], split='&'), '[', 1))
```

```
## [1] "http://theguitarrepairworkshop.com/"
## [2] "http://www.guitarservices.com/"
## [3] "http://www.guitarrepairbench.com/guitar-building-projects/guitar-workshop/guitar-workshop-proje"
## [4] "https://www.facebook.com/The-Guitar-Repair-Workshop-847517635259712/"
## [5] "https://www.taylorguitars.com/dealer/guitar-repair-workshop-ltd"
## [6] "http://www.laweekly.com/music/10-best-guitar-repair-shops-in-los-angeles-4647166"
## [7] "http://guitarworkshopglasgow.com/pages/repairs-1"
## [8] "https://www.justdial.com/Delhi-NCR/Guitar-Repair-Services/nct-10988623"
## [9] "https://www.justdial.com/Mumbai/Guitar-Repair-Services/nct-10988623"
```

Hin und weg

```
library(XML)
url4<-"http://api.openstreetmap.org/api/0.6/node/25439439"
obj4 <- xmlParse(url4)
xpathApply(obj4,"//tag[@k = 'railway:station_category']")[[1]]
```

```
## <tag k="railway:station_category" v="2"/>
```

- Wikipedia Artikel Bahnhofskategorien

Exkurs: Bahnhofskategorien

- rvest: Easily Harvest (Scrape) Web Pages

```
library(rvest)
bhfkat<-read_html(
  "https://de.wikipedia.org/wiki/Bahnhofskategorie")
df_html_bhfkat<-html_table(
  html_nodes(bhfkat, "table")[[1]], fill = TRUE)
```

Bahnhofskategorien Übersicht

Stufe	Bahnsteigkanten	Bahnsteiglänge	Reisende/Tag	Zughalte/Tag	Prozent	Kategorie
6	1	bis 90 m	bis 49	bis 10	> 90 %	1
5	2	> 90 bis 140 m	50 bis 299	11 bis 50	> 80 bis 90 %	2
4	3 bis 4	> 140 bis 170 m	300 bis 999	51 bis 100	> 60 bis 80 %	3
3	5 bis 9	> 170 bis 210 m	1000 bis 9999	101 bis 500	> 50 bis 60 %	4
2	10 bis 14	> 210 bis 280 m	10.000 bis 49.999	501 bis 1000	> 40 bis 50 %	5
1	ab 15	> 280 m	ab 50.000	ab 1001	> 25 bis 40 %	6
					bis 25 %	7

Figure 1:

Stufe	Bahnsteigkanten	Bahnsteiglänge	Reisende/Tag	Zughalte/Tag
6	01	> 000 bis 090 m	00000 bis 00049	000 bis 0010
5	02	> 090 bis 140 m	00050 bis 00299	011 bis 0050
4	03 bis 04	> 140 bis 170 m	00300 bis 00999	051 bis 0100
3	05 bis 09	> 170 bis 210 m	01000 bis 09999	101 bis 0500
2	10 bis 14	> 210 bis 280 m	10.000 bis 49.999	501 bis 1000
1	00i ab 15	> 280 m	00000i ab 50.000	000i ab 1001

Links

- Scraping CRAN with rvest