# 50 Drugs

## Load libraries

```
library(rvest)
library(magrittr)
library(stringr)
library(stringi)
library(tidyverse)
library(knitr)
```

## Log-in and collect the links

```
fiftydrugs <- html_session("https://moodle.gla.ac.uk/login/index.php")

login <- fiftydrugs %>%
  html_node("form") %>%
  html_form() %>%
  set_values(username = "username", password = "password")
login$url <- "https://moodle.gla.ac.uk/login/index.php"

data <- fiftydrugs %>%
  submit_form(login, submit = '<unnamed>') %>%
  jump_to("https://moodle.gla.ac.uk/mod/data/view.php?d=424") %>%
  read_html() %>%
  html_nodes(".defaulttemplate p")
```

## Separate the links

```
links <- ""
for (i in 1:length(data)){
  links[i] <- data[[i]] %>%
    html_children() %>%
    .[2] %>%
    html_attrs() %>%
    as.character()
}

titles <- data %>%
  html_text()

class <- strsplit(titles,":") %>%
  map(1) %>%
  unlist

drug <- strsplit(titles,":") %>%
  map(2) %>%
  unlist
```

```
table <- data_frame(class, drug, links)
rm(titles, class, drug, i, links)
```

## Follow every link

```
linksdata <- ""

datacomplete <- fiftydrugs %>%
  submit_form(login, submit = '<unnamed>')

#need to fix
for (i in 1:length(table$links)){
linksdata[i] <- datacomplete %>%
    jump_to(table$links[i]) %>%
    read_html() %>%
    html_nodes("form li , #region-main h4")
}
for (i in 1:length(table$links)){
linksdata[[i]] <- datacomplete %>%
    jump_to(table$links[i]) %>%
    read_html() %>%
    html_nodes("form li , #region-main h4")
}
rm(i)
```

```
#function
splitAt <- function(x, pos) split(x, cumsum(seq_along(x) %in% pos))
```

```
listdataframe <- linksdata
#remove unnecessary objects
rm(data, datacomplete, fiftydrugs, login, linksdata)
```

## Rewrite linksdata into linksdataframe

```
#searches through the drug page, and selects the relevant parts
for (i in 1:length(table$links)){
subset <- listdataframe[[i]] %>%
rvest::html_attr("class") %>%
  is.na
listdataframe[[i]] <- listdataframe[[i]][subset]
#looks at the dom type (e.g li or h4)
examplelist <- listdataframe[[i]] %>%
rvest::html_name()
#see what's left
exampletext <- listdataframe[[i]] %>%
  html_text()
#seperate into sections by h4.
listdataframe[[i]] <- splitAt(exampletext, which(examplelist == "h4"))
names(listdataframe[[i]]) <- c("Example(s) of drugs:", "Mechanism of action:", "Indication(s):", "Side
#remove first item of each.
#should swap below for a match against the actual titles
```

```
listdataframe[[i]][[1]] <- list(listdataframe[[i]][[1]][-1])
listdataframe[[i]][[2]] <- list(listdataframe[[i]][[2]][-1])
listdataframe[[i]][[3]] <- list(listdataframe[[i]][[3]][-1])
listdataframe[[i]][[4]] <- list(listdataframe[[i]][[4]][-1])
listdataframe[[i]][[5]] <- list(listdataframe[[i]][[5]][-1])
listdataframe[[i]][[6]] <- list(listdataframe[[i]][[6]][-1])
listdataframe[[i]][[7]] <- list(listdataframe[[i]][[7]][-1])
}
rm(i, examplelist, exampletext, subset, splitAt)
```

## Split into dataframe

```
dataframe <- data_frame(listdataframe)

clearit <- function(x) {
unlist(x, recursive = FALSE) %>%
.[[1]] %>%
as_data_frame()
}

example <- clearit(dataframe[1,1])
for (i in 2:length(dataframe$listdataframe)){
  example <- rbind(example, clearit(dataframe[i,1]))
}

dataframe <- example
rm(example, i, clearit, listdataframe)
```

https://github.com/rstudio/webinars/blob/master/32-Web-Scraping/navigation-and-authentication.md

## Add moodle data to individual drug page data

```
totaldataframe <- cbind(table, dataframe)
head(totaldataframe)
```

```
## # A tibble: 6 x 10
##   class  drug    links  `Example(s) of ~ `Mechanism of a~ `Indication(s):`
##   <chr>  <chr>   <chr>  <list>           <list>           <list>
## 1 Haema~  Anti-~ https~ <chr [1]>        <chr [4]>        <chr [2]>
## 2 Haema~  Anti-~ https~ <chr [1]>        <chr [3]>        <chr [1]>
## 3 Haema~  Recom~ https~ <chr [2]>        <chr [3]>        <chr [3]>
## 4 Haema~  Hepar~ https~ <chr [1]>        <chr [4]>        <chr [3]>
## 5 Haema~  Hepar~ https~ <chr [1]>        <chr [4]>        <chr [3]>
## 6 Haema~  Vitam~ https~ <chr [1]>        <chr [3]>        <chr [2]>
## # ... with 4 more variables: `Side effects:` <list>, `Important
## #   pharmacokinetics / pharmacodynamics:` <list>, `Patient
## #   information:` <list>, `Other information:` <list>
```

### Export the data

```
Datestamp <- Sys.Date()

dataframeName <- str_c("archive/", Datestamp) %>%
str_replace_all(" ", "") %>%
str_c("drugs.Rda")
save(totaldataframe, file=dataframeName)

save(totaldataframe, file="drugs.Rda")

jsonName <- str_c("archive/", str_replace_all(Datestamp, " ", ""), "drugs.json")
totaldataframe %>%
  jsonlite::toJSON() %>%
  write(file = jsonName)

totaldataframe %>%
  jsonlite::toJSON() %>%
  write(file="drugs.json")
```