WebScrapingUsingR

Dalton
2020-12-04

Purpose of this algorithm, is to scrape the contact data of each MP

- Reading https://www.ourcommons.ca/Members/en/search?searchTypeId=11&parliamentId=43&sessionId=96&caucus=lib&province=ON&gender=B&page=1
- Reading the members URLs into a dataframe
- http://www.ourcommons.ca/Members/en/jean-claude-poissant(89246)/contact and the personal website of the MP if the previous link.
- Extracting the first and second name, the email address, the phone number, the twitter handle, the facebook handle, the personal website and the personal email address of each member.

Reading csv file

```
# Reading csv file
library(readr)
members_table <- read.csv("Members.csv")

head(members_table)
  ...Identity.Value Title First.Name Last.Name                 Constituency
1                 1   Mr.     Yvonne     Abraham                   Saint-Lambert
2                 2   Mrs.    Leslyn      Acton                 Terra Nova-The Peninsulas
3                 3   Mr.     Gerald      Adams                 Manitoba Interlake-Gimli
4                 4   Mrs.    Chrys       Adams                 Cloverdale-Langley City
5                 5   Mr.     Ali         Ehsassi               Willowdale
6                 6   Mr.     Rob         Ellis                 Kenora-Rainy River
  Province.Territory Gender Caucus Email.Address Phone.Number Twitter Facebook
1             Quebec      F    Lib            NA           NA      NA       NA
2 Newfoundland and...      F    Lib            NA           NA      NA       NA
3           Manitoba      M    Lib            NA           NA      NA       NA
4            British      F    Lib            NA           NA      NA       NA
5            Ontario      M    Lib            NA           NA      NA       NA
6            Ontario      M    Lib            NA           NA      NA       NA
  Personal.Website Personal.Email
1               NA            NA
2               NA            NA
3               NA            NA
4               NA            NA
5               NA            NA
6               NA            NA
```

Reading all the members URLs into a dataframe

```
# Reading all the members URLs into a dataframe
library(stringr)
Pages_Urls <- "https://www.ourcommons.ca/Members/en/search"

# Initialize the dataframe
Profile_Urls <- data.frame(Pages_Urls)

# Loop through the pages and add the profile URLs to the dataframe
for(i in 2:(nrow(members_table)/20+1)){
Pages_Urls <- paste0("https://www.ourcommons.ca/Members/en/search?page=",i)
Profile_Urls <- rbind(Profile_Urls, Pages_Urls)
}
tail(Profile_Urls, 10)
                                                                Pages_Urls
NA                                                                    <NA>
NA.1                                                                  <NA>
NA.2                                                                  <NA>
NA.3                                                                  <NA>
NA.4                                                                  <NA>
NA.5                                                                  <NA>
NA.6                                                                  <NA>
NA.7                                                                  <NA>
NA.8                                                                  <NA>
NA.9                                                                  <NA>

# Initialize the dataframe
Profile_Urls <- data.frame(Profile
```

```
http://www.chinaforbidcity.com/
http://www.gov.cn/
http://www.gov.cn/
http://www.gov.cn
```