20181022a Selector gadget and simple scraping

cheungngo

22 October 2018

### Pre-requisites

library(rvest)

<https://cran.r-project.org/web/packages/rvest/vignettes/selectorgadget.html>  
Installing the selector gadget in chrome is also important to get the corresponding CSS selector from the desired fields in the webpage to be scraped.

### Specifying the url and decode by read\_html()

url = 'https://www.ncbi.nlm.nih.gov/pubmed/?term=schizophrenia+memantine'  
webpage = read\_html(url)  
webpage

## {xml\_document}  
## <html xmlns="http://www.w3.org/1999/xhtml" lang="en" xml:lang="en">  
## [1] <head xmlns:xi="http://www.w3.org/2001/XInclude">\n<meta http-equiv= ...  
## [2] <body>\n \t<div class="grid">\n\t\t\t<div class="col twelve\_col n ...

### Using the CSS selctor (by the selectorgadget) to scrap the data by html\_nodes() and html\_text()

# Suppose we are going to get the title   
# the CSS selector would be '.title' as shown by the selctor gadget  
title\_html = html\_nodes(webpage, '.title')  
head(title\_html)

## {xml\_nodeset (6)}  
## [1] <p xmlns:mml="http://www.w3.org/1998/Math/MathML" class="title"><a h ...  
## [2] <p xmlns:mml="http://www.w3.org/1998/Math/MathML" class="title"><a h ...  
## [3] <p xmlns:mml="http://www.w3.org/1998/Math/MathML" class="title"><a h ...  
## [4] <p xmlns:mml="http://www.w3.org/1998/Math/MathML" class="title"><a h ...  
## [5] <p xmlns:mml="http://www.w3.org/1998/Math/MathML" class="title"><a h ...  
## [6] <p xmlns:mml="http://www.w3.org/1998/Math/MathML" class="title"><a h ...

# convert the data to text  
title = html\_text(title\_html)  
head(title)

## [1] "Differential Effects of Extended Exercise and Memantine Treatment on Adult Neurogenesis in Male and Female Rats."   
## [2] "The use of memantine in neuropsychiatric disorders: An overview."   
## [3] "Study of the Effect of Memantine on Negative Sign in Patients with Schizophrenia and Schizoaffective Disorders."   
## [4] "Anti-dementia Drugs for Psychopathology and Cognitive Impairment in Schizophrenia: A Systematic Review and Meta-analysis."   
## [5] "Augmentation strategies for clozapine refractory schizophrenia: A systematic review and meta-analysis."   
## [6] "Medications Used for Cognitive Enhancement in Patients With Schizophrenia, Bipolar Disorder, Alzheimer's Disease, and Parkinson's Disease."

### Using gsub() to modify scraped data

# scraping the authors of the studies  
# '.desc' as detected by the selector  
author\_html = html\_nodes(webpage,'.desc')  
author = html\_text(author\_html)  
head(author)

## [1] "Cahill SP, Cole JD, Yu RQ, Clemans-Gibbon J, Snyder JS."   
## [2] "Lu S, Nasrallah HA."   
## [3] "Tavakoli-Ardakani M, Abbaspour H, Farhadi Nasab A, Mazaheri Meibodi A, Kheradmand A."  
## [4] "Kishi T, Ikuta T, Oya K, Matsunaga S, Matsuda Y, Iwata N."   
## [5] "Siskind DJ, Lee M, Ravindran A, Zhang Q, Ma E, Motamarri B, Kisely S."   
## [6] "Hsu WY, Lane HY, Lin CH."

# Suppose we only want the first author  
author = gsub(',.\*','',author)  
author = as.factor(author)  
head(author)

## [1] Cahill SP Lu S Tavakoli-Ardakani M  
## [4] Kishi T Siskind DJ Hsu WY   
## 17 Levels: Andrade C. Cahill SP Ding Z Hsu WY ... Zheng W

table(author)

## author  
## Andrade C. Cahill SP Ding Z   
## 1 1 1   
## Hsu WY Kantrowitz JT Kantrowitz JT.   
## 1 1 1   
## Kishi T Koola MM Koola MM.   
## 2 2 2   
## Light GA Lu S Omranifard V   
## 1 1 1   
## Siskind DJ Swerdlow NR Tavakoli-Ardakani M   
## 1 1 1   
## Veerman S Zheng W   
## 1 1

### combining the data into a dataframe

schiz\_memantine\_df = data.frame(Title = title, Author = author)  
head(schiz\_memantine\_df)

## Title  
## 1 Differential Effects of Extended Exercise and Memantine Treatment on Adult Neurogenesis in Male and Female Rats.  
## 2 The use of memantine in neuropsychiatric disorders: An overview.  
## 3 Study of the Effect of Memantine on Negative Sign in Patients with Schizophrenia and Schizoaffective Disorders.  
## 4 Anti-dementia Drugs for Psychopathology and Cognitive Impairment in Schizophrenia: A Systematic Review and Meta-analysis.  
## 5 Augmentation strategies for clozapine refractory schizophrenia: A systematic review and meta-analysis.  
## 6 Medications Used for Cognitive Enhancement in Patients With Schizophrenia, Bipolar Disorder, Alzheimer's Disease, and Parkinson's Disease.  
## Author  
## 1 Cahill SP  
## 2 Lu S  
## 3 Tavakoli-Ardakani M  
## 4 Kishi T  
## 5 Siskind DJ  
## 6 Hsu WY