

Assignment7_607

Alvaro Bueno

10/13/2017

```
html <- url('https://raw.githubusercontent.com/delagroove/dataScience/master/books.html')
webpage <- readLines(html)
htmltable <- readHTMLTable(webpage)
htmldata <- htmlParse(webpage, encoding="UTF-8")
result_href <- xpathSApply(htmldata, "//*[table/tr/td/a]", xmlGetAttr, 'href')
htmltable <- as.data.frame(htmltable)
htmltable$sourceUrl <- result_href

xData <- getURL('https://raw.githubusercontent.com/delagroove/dataScience/master/books.xml')
doc <- xmlParse(xData)

xmldata <- xmlToDataFrame(
  getNodeSet(doc, "//book")
)
result_href <- xpathSApply(doc, "//*[book/Source]", xmlGetAttr, 'url')
xmldata$sourceUrl <- result_href

jData <- getURL('https://raw.githubusercontent.com/delagroove/dataScience/master/books.json')
jsontable <- fromJSON(jData)
```

The tables are not exactly identical, but they can be if we rename the columns in the HTML table.

```
names(htmltable) <- c('Name', 'Author', 'Description', 'Published Date', 'Editorial', 'Format', 'NumPages')
kable(htmltable)
```

Name	Author
Last Night a DJ Saved My Life: The History of the Disc Jockey	Bill Brewster, Frank Broughton
Crime and Punishment	Fyodor Dostoyevsky, David McDuff (Translator)
One Hundred Years of Solitude	Gabriel García Márquez, Gregory Rabassa (Translator)

```
kable(xmldata)
```

Name	Author
One Hundred Years of Solitude	Gabriel García Márquez, Gregory Rabassa (Translator)
Last Night a DJ Saved My Life: The History of the Disc Jockey	Bill Brewster, Frank Broughton
Crime and Punishment	Fyodor Dostoyevsky, David McDuff (Translator)

```
kable(jsontable)
```

Name	Author
Crime and Punishment	Fyodor Dostoyevsky, David McDuff (Translator)
One Hundred Years of Solitude	Gabriel García Márquez, Gregory Rabassa (Translator)
Last Night a DJ Saved My Life: The History of the Disc Jockey	Bill Brewster, Frank Broughton