

WebTech-HTML,XML,JSON

Chunhui Zhu

October 11, 2017

Assignment - Working with XML and JSON in R

Pick three of your favorite books on one of your favorite subjects. At least one of the books should have more than one author. For each book, include the title, authors, and two or three other attributes that you find interesting.

Take the information that you've selected about these three books, and separately create three files which store the book's information in HTML (using an html table), XML, and JSON formats (e.g. "books.html", "books.xml", and "books.json").

To help you better understand the different file structures, I'd prefer that you create each of these files "by hand" unless you're already very comfortable with the file formats. Write R code, using your packages of choice, to load the information from each of the three sources into separate R data frames. Are the three data frames identical? Your deliverable is the three source files and the R code. If you can, package your assignment solution up into an .Rmd file and publish to rpubs.com. [This will also require finding a way to make your three text files accessible from the web].

R enviroment

1.HTML

```
setwd("C://Users/Ivy/Desktop/607/W7")
htmldata<-htmlTreeParse('Books.html', useInternalNodes = T)
htmldata
```

```
## <!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN" "http://www.w3.org/TR/REC-html40/loose
## <html>
## <head></head>
## <body><ul>
## <li>Book: My Data Science Book</li>
##       <li>Author: Vincent Granville</li>
##       <li>Attributes: Big Data, Statistic, Metrics</li>
##     </ul></body>
## </html>
##
```

```
htmlcontent<-getNodeSet(htmldata,"//ul//li")
htmlcontent<-sapply(htmlcontent,xmlValue)
htmlcontent
```

```
## [1] "Book: My Data Science Book"
## [2] "Author: Vincent Granville"
## [3] "Attributes: Big Data, Statistic, Metrics"
```

2.XML

```
library("methods")
```

```

xmldata<-xmlParse(file="Books.xml")
xmldata

## <?xml version="1.0" encoding="utf-8"?>
## <a1>
##   <h1>Book: R for Data Science</h1>
##   <h2>Author:Hadley Wickham, Garrett Grolemond </h2>
##   <h3>Atributs:Data Management, Tidy, Mutate </h3>
## </a1>
##

xmlcontent<-xmlRoot(xmldata)
xmlcontent1<-xmlToDataFrame(xmlcontent)
xmlcontent1

##                               text
## 1                Book: R for Data Science
## 2 Author:Hadley Wickham, Garrett Grolemond
## 3  Atributs:Data Management, Tidy, Mutate

```

3.JSON

```

jsondata<-fromJSON(file="books.json")
jsondata

## [[1]]
## [[1]]$Book
## [1] "R for Data Science"
##
##
## [[2]]
## [[2]]$Author
## [1] "Hadley Wickham, Garrett Grolemond"
##
##
## [[3]]
## [[3]]$Atributs
## [1] "Data Management, Tidy,Mutate"

```