

IS 606: Assignment 8

MUSA T. GANIYU

March 17, 2016

QUESTION:

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.

DIRECTION:

Write R code, using your packages of choice, to load the information from each of the three sources into separate R data frames.

XML Format

```
# Please kindly install these packages if you dont have it yet, and load it.
library(XML)
library(RCurl)
```

```
## Loading required package: bitops
```

```
library(knitr)
library(htmlltab)
```

```
## Warning: package 'htmlltab' was built under R version 3.2.4
```

```
library(jsonlite)
```

```
# Getting xml data loaded from URL
```

```
url = "https://raw.githubusercontent.com/mascotinme/MSDA-IS607/f896255f3b817cad1496bb767922bb7c72457422,
data = getURL(url, ssl.verifypeer=0L, followlocation=1L)
book1 <- xmlParse(data)
book <- xmlToDataFrame(book1)
kable(head(book))
```

name	isbn	author
Engineering mathematics: 7th Edition	978-1100005	Dexter Booth
American Sniper: The Autobiography of the Most Lethal Sniper in U.S. Military History	978-0062238863	Chris Kyle, S
the catcher in the rye	978-0316769488	j.d salinger

HTML FORMAT

Name	ISBN	Author
Engineering mathematics: 7th Edition	978-1100005	Dexter Boot
American Sniper: The Autobiography of the Most Lethal Sniper in U.S. Military History	978-0062238863	Chris Kyle,
the catcher in the rye	978-0316769488	J.D Salinger

```
# load HTML data from URL and ignore the warning statement(s).
options(warn = -1)

url = "https://raw.githubusercontent.com/mascotinme/MSDA-IS607/master/Books1.html"

books_html = htmltab(doc = url, which = 1)
books_html = data.frame(books_html, row.names = NULL)
kable(books_html)
```

Name	ISBN	Author
Engineering mathematics: 7th Edition	978-1100000000	Dexter Booth
American Sniper: The Autobiography of the Most Lethal Sniper in U.S. Military History	978-0062238863	Chris Kyle, S
The Catcher In The Rye	978-0316769488	J.D Salinger

JSON FORMAT

```
#loading json file from URL

url2 = "https://raw.githubusercontent.com/mascotinme/MSDA-IS607/master/Books.json"

books_json = getURL(url2)
books_json = fromJSON(books_json)
kable(books_json)
```

Question: Are the three data frames identical?

YES!, the three data frames are identical.

REFERENCES:

<http://www.rprimer.dk/index.php?page=showrule&index=2>

<https://cran.r-project.org/web/packages/jsonlite/vignettes/json-aaquickstart.html>

<http://www.w3schools.com/xml/>