

Assignment / Query XML

Zongyu WU

Spring 2024

Question 4

In a code block, load the XML containing the DTD into R with validation.

```
library(XML)

xmlFile <- "orders.xml"

dom <- xmlParse(xmlFile, validate=T)
```

Question 5

In a new code block, execute an XPath expression that returns the names of all customers that are in the USA.

```
xpathEx <- "/Root/Customers/Customer[FullAddress/Country = 'USA']/ContactName"
usaCus <- xpathSApply(dom, xpathEx, xmlValue)
cat(usaCus)
```

```
## Howard Snyder Yoshi Latimer John Steel Jaime Yorres
```

Question 6

Using the result returned in (5) and any additional queries required, calculate the percentage of customers who do are not in the USA. Display the result as markup in your notebook.

```
xpathEx <- "count(/Root/Customers/Customer)"
numCus <- xpathSApply(dom, xpathEx, xmlValue)
```

20 percent of customers are not in the US.

Question 7

Using a combination of R and XPath, calculate the total amount paid for freight for all orders within the USA.

```
xpathEx <- "/Root/Orders/Order[ShipInfo/ShipCountry='USA']/ShipInfo/Freight"
freightValues <- xpathSApply(dom, xpathEx, xmlValue)

freightValues <- as.numeric(freightValues)

totalFreight <- sum(freightValues)
```

Total amount paid for freight for all orders within the USA is 1516.2.

Question 8

Using a combination of R and XPath, calculate the average amount paid for freight for all orders shipped to the USA.

Average amount paid for freight for all orders shipped to the USA is 68.92.