Query XML

Balakrishna, D

Summer Full 2023

Loading the XML Library

```
# Load required libraries
library(XML)
```

Questions 1 to 3 are already included in the zip file

Question 4

```
# Load the XML document with DTD validation
xml_doc <- xmlParse(file = "CustomersAndOrders.xml", isURL = FALSE,
                    validate = TRUE)
```

Question 5

```
# Execute an XPath expression to get names of customers not living in "MA"
names_not_in_MA <- xpathSApply(xml_doc,
                               "//Customer[FullAddress/Region != 'MA']/CompanyName",
                               xmlValue)
cat("Names of customers that do not live in 'MA':\n", names_not_in_MA, "\n\n")
```

```
## Names of customers that do not live in 'MA':
##  Great Lakes Food Market Hungry Coyote Import Store Lazy K Kountry Store Let's Stop N Shop
```

Question 6

```
# Count how many customers are not living in "MA"
count_customers_not_in_MA <- length(names_not_in_MA)
cat("Number of customers that do not live in 'MA':", count_customers_not_in_MA,
    "\n\n")
```

```
## Number of customers that do not live in 'MA': 4
```

Question 7

```
# Execute an XPath expression to get the total amount paid for freight for
# customer with ID "GREAL"
total_freight_for_GREAL <- xpathSApply(xml_doc,
                                       "//Order[CustomerID='GREAL']/ShipInfo/Freight",
                                       xmlValue)
sum_total_freight_for_GREAL <- sum(as.numeric(total_freight_for_GREAL))
```

```r
cat("Total amount paid for freight by customer with ID 'GREAL':",
    sum_total_freight_for_GREAL, "\n\n")
```

```
## Total amount paid for freight by customer with ID 'GREAL': 1087.61
```

Question 8

```r
# Execute an XPath expression to get the average amount paid for freight for all
# orders shipped to the USA
total_freight_USA <- xpathSApply(xml_doc,
                                 "//Order[ShipInfo/ShipCountry='USA']/ShipInfo/Freight",
                                 xmlValue)
total_freight_USA <- sum(as.numeric(total_freight_USA))
average_freight_USA <- mean(total_freight_USA)
cat("Average amount paid for freight for all orders shipped to the USA:",
    average_freight_USA, "\n")
```

```
## Average amount paid for freight for all orders shipped to the USA: 1516.2
```