ASSIGNMENT 11.2: Query XML

Chandresh Lokesha

2024-03-27

```
library(XML)

# Load the XML file
xml_data <- xmlParse("orders.xml", useInternalNodes = TRUE, validate = TRUE)

# Run an XPath query to count the number of items in a specific list
total_customers <- xpathSApply(xml_data, "//Customers", xmlSize)

# Display the result
print(paste("Total number of 'Customers':", total_customers))
```

```
## [1] "Total number of 'Customers': 5"
```

```
# Run an XPath query to count the number of items in a specific list
total_orders <- xpathSApply(xml_data, "//Orders", xmlSize)

# Display the result
print(paste("Total number of 'Orders':", total_orders))
```

```
## [1] "Total number of 'Orders': 23"
```

Q5

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

```
customers_in_usa <- xpathSApply(
                    xml_data,
                    "//Customer[FullAddress/Country = 'USA']/CompanyName/text()",
                    xmlValue)
print(customers_in_usa)
```

```
## [1] "Great Lakes Food Market"    "Hungry Coyote Import Store"
## [3] "Lazy K Kountry Store"       "Let's Stop N Shop"
```

Q6

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.

```
# Percentage of customers not in the USA
percentage_not_in_usa <-
  ((total_customers - length(customers_in_usa)) / total_customers) * 100

# Display the result as markup
cat(sprintf("Percentage of customers not in USA: %.2f%%",
            percentage_not_in_usa))
```

```
## Percentage of customers not in USA: 20.00%
```

Q7

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

```
freight_amounts <- xpathApply(
                  xml_data,
                  "//Order[ShipInfo/ShipCountry = 'USA']/ShipInfo/Freight",
                  xmlValue)
# print(length(as.numeric(unlist(freight_amounts))))
# Convert freight amounts to numeric and sum them up
total_freight_usa <- sum(as.numeric(unlist(freight_amounts)))

# Display the total amount paid for freight for all orders within the USA
cat(sprintf("Total amount paid for freight for all orders
            within the USA: $%.2f", total_freight_usa))
```

```
## Total amount paid for freight for all orders
##             within the USA: $1516.20
```

Q8

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

```
num_orders_usa <- length(freight_amounts)

#  the average amount paid for freight for all orders shipped to the USA
average_freight_usa <- total_freight_usa / num_orders_usa
```

```r
# Display the average amount paid for freight for all orders shipped to the USA
cat(sprintf("Average amount paid for freight for all orders
            shipped to the USA: $%.2f", average_freight_usa))
```

```
## Average amount paid for freight for all orders
##             shipped to the USA: $68.92
```