CIT 483/583 Lab06 – REXML

**Instructions**

Open and save this file in any MS word-compatible format as Lab06\_*Firstname*\_*Lastname*.<ext> and place your answers in that document. Do a **Save-As** and retain all of my content. Keep the document safe in case your submission fails, or you discover an error prior to the due date and wish to re-submit. Submit your document to the Lab06 dropbox in Canvas. The due date and any other pertinent information are noted in the Canvas item.

This lab should be completed on students.cs.nku.edu as a reference implementation and as a means of testing your answers. It can be done using IRB only, or you may create .rb files to test some or all of your code, but do not submit .rb files (programs).

**Place the answers in or immediately following each question and make sure your answers stand out from the questions by using a different font color.**

1. The following XML is not well-formed. Identify three (simple) changes that can be made to correct it. [The choice of single vs. double quotes is not an issue as long as they are matched by type.]

<receive>

<car make="Ford" model='Focus' year='2012">

<display>2012 Ford Focus</display>

</car>

<car make="Chevy" model='Impala' year="2009">

<display>2009 Chevy Impala<display>

</car>

</recieve>

a) year= "2012" (mismatched quotes)

b) <display> ending tag missing / - </display>

c) </recieve> should be </receive> (mismatch/typo)

2. Copy ~mccordt/master\_scoreboard.xml on students.cs.nku.edu to your CIT483 directory. Use irb to perform the following interactions and capture all output.

$ irb --simple-prompt

>> require 'rexml/document'

=> true

>> include REXML

=> Object

>> dom = Document.new(File.open 'master\_scoreboard.xml')

=> <UNDEFINED> ... </>

>> dom.class

=> REXML::Document

>> dom.root.class

=> REXML::Element

>> dom.root.elements.class

=> REXML::Elements

>> dom.root.elements.size

=> 15

>> dom.root.children.size

=> 31

>> dom.root

=> <games year='2015' month='07' day='19' modified\_date='2016-03-31T15:42:00Z' next\_day\_date='2015-07-20'> ... </>

>> dom.children

=> [<?xml ... ?>, #<REXML::Comment:0x000000011bfe80 @parent=<UNDEFINED> ... </>, @string="Copyright 2016 MLB Advanced Media, L.P. Use of any content on this page acknowledges agreement to the terms posted here http://gdx.mlb.com/components/copyright.txt">, "\n", <games year='2015' month='07' day='19' modified\_date='2016-03-31T15:42:00Z' next\_day\_date='2015-07-20'> ... </>]

>> dom.children.class

=> Array

>> dom.children[0]

=> <?xml ... ?>

>> dom.children[1]

=> #<REXML::Comment:0x000000011bfe80 @parent=<UNDEFINED> ... </>, @string="Copyright 2016 MLB Advanced Media, L.P. Use of any content on this page acknowledges agreement to the terms posted here <http://gdx.mlb.com/components/copyright.txt>">

>> dom.children[3]

=> <games year='2015' month='07' day='19' modified\_date='2015-07-26T07:31:29Z' next\_day\_date='2015-07-20'> ... </>

>> dom.children[3] == dom.root

=> true

>> dom.elements[1] == dom.root

=> true

>> dom.elements.to\_a('/games/game').size

=> 15

>> dom.elements.to\_a('/games/game/winning\_pitcher/').size

=> 13

>> dom.elements.to\_a('/games/game/winning\_pitcher/')[0]

=> <winning\_pitcher id='476454' last='Betances' first='Dellin' name\_display\_roster='Betances' number='68' era='1.47' wins='6' losses='2'/>

>> dom.elements.to\_a('/games/game')[3,2].each do |gg|

gg.elements.each {|el| p el.name }

end

[Abbreviate the output to the last 10-12 lines]

D' home\_name\_abbrev='WSH' away\_code='lan' away\_file\_code='la' away\_team\_id='119' away\_team\_city='LA Dodgers' away\_team\_name='Dodgers' away\_division='W' away\_league\_id='104' away\_sport\_code='mlb' home\_code='was' home\_file\_code='was' home\_team\_id='120' home\_team\_city='Washington' home\_team\_name='Nationals' home\_division='E' home\_league\_id='104' home\_sport\_code='mlb' day='SUN' gameday\_sw='P' double\_header\_sw='N' game\_nbr='1' tbd\_flag='N' away\_games\_back='-' home\_games\_back='-' away\_games\_back\_wildcard='' home\_games\_back\_wildcard='' venue\_w\_chan\_loc='USDC0001' location='Washington, DC' gameday='2015\_07\_19\_lanmlb\_wasmlb\_1' away\_win='53' away\_loss='40' home\_win='49' home\_loss='41' game\_data\_directory='/components/game/mlb/year\_2015/month\_07/day\_19/gid\_2015\_07\_19\_lanmlb\_wasmlb\_1' league='NN'> ... </>]

>> dom.elements.to\_a('/games/game/winning\_pitcher/').each do |p|

puts "Last name: #{p.attribute('last')} ERA: #{p.attribute('era')}"

end

[Abbreviate the output to the first 10-12 lines]

Betances 1.47

Last name: Betances ERA: 1.47

Last name: Estrada ERA: 3.22

Last name: Gonzalez ERA: 4.21

Last name: Rzepczynski ERA: 3.93

Last name: Greinke ERA: 1.30

Last name: Neris ERA: 3.52

Last name: Duffy ERA: 4.24

Last name: Jungmann ERA: 2.04

Last name: Keuchel ERA: 2.12

Last name: Torres ERA: 4.17

Last name: Chavez ERA: 3.21

Last name: Bumgarner ERA: 3.27

Last name: Arrieta ERA: 2.52

. . .

>> dom.elements.to\_a('//losing\_pitcher[@last="Villarreal"]')

=> [<losing\_pitcher id='543881' last='Villarreal' first='Pedro' name\_display\_roster='Villarreal' number='62' era='5.23' wins='1' losses='3'/>]

>> query = '//losing\_pitcher[@last="Villarreal"]/../winning\_pitcher'

=> "//losing\_pitcher[@last=\"Villarreal\"]/../winning\_pitcher">>

>> dom.elements.to\_a(query)[0].attribute("last")

=> last='Rzepczynski'

3. Use Ruby and ReXML to\_answer the following questions.

NOTE: You may continue your irb session from the previous question for testing; otherwise, you will have to load the required extensions and objects again, and initialize the DOM object again:

>> require 'rexml/document'

>> include REXML

>> dom = Document.new(File.open 'master\_scoreboard.xml')

The elements method allows you to reference items by name (not just by number). For example, to get the first game's winning\_pitcher attributes:

>> dom.elements.to\_a('//game')[0].elements['winning\_pitcher']

=> <winning\_pitcher id='476454' last='Betances' first='Dellin' name\_display\_roster='Betances' number='68' era='1.47' wins='6' losses='2'/>

To get selected winning \_pitcher elements, use [@attr = value] where = can be any relational operator:

dom.elements.to\_a('//winning\_pitcher[@losses > 10]')

(a) Who are the winning pitchers whose ERA is greater than 2?

dom.elements.to\_a('//winning\_pitcher[@era >2]').each { |p| puts p.attribute('last') }

Estrada

Gonzalez

Rzepczynski

Neris

Duffy

Jungmann

Keuchel

Torres

Chavez

Bumgarner

Arrieta

(b) Who are the winning pitchers whose ERA is less than 2?

dom.elements.to\_a('//winning\_pitcher[@era <2]').each { |p| puts p.attribute('last') }

Betances

Greinke

(c) Who are the winning pitchers who have fewer than 10 wins?

dom.elements.to\_a('//winning\_pitcher[@wins <10]').each { |p| puts p.attribute('last') }

Betances

Estrada

Gonzalez

Rzepczynski

Greinke

Neris

Duffy

Jungmann

Torres

Chavez

(d) Who are the losing pitchers who have more than 5 wins?

dom.elements.to\_a('//losing\_pitcher[@wins >5]').each { |p| puts p.attribute('last') }

Archer

Scherzer

Sale

Gallardo

Martinez

(e) What is the total sum of wins of all winning pitchers?

HINT: Use the starter code below that iterates over all winning\_pitcher elements and extracts the number of wins, which is a a REXML::Attribute. To treat it as a number, first get its String value, then call to\_i. Create a counter and, each time you get a number, add it to the counter.

total = 0

dom.elements.to\_a('/games/game/winning\_pitcher/').each do |p|

# process wins as described and add it to total

puts p.attribute('wins')

end

puts total

total = 0

dom.elements.to\_a('//winning\_pitcher/').each do |p|

total += p.attribute('wins').to\_s.to\_i

end

puts total

4. The following script contains a "listener" class that is based on the version in the lecture slides. It is written to isolate information on winning\_pitchers . Enter the code directly into IRB and answer the three questions that follow. You do not need to show IRB output.

require 'rexml/document'

require 'rexml/streamlistener'

include REXML

class MyListener

include StreamListener

def tag\_start (name, attr)

if name == "winning\_pitcher"

puts "Winning pitcher: #{name.inspect} −− #{attr.inspect}"

end

end

def tag\_end(name)

# puts "tag end: #{name.inspect}"

end

def text( val)

# puts "text : #{val.inspect}"

end

def comment(cmt)

# puts "comment: #{cmt.inspect}"

end

def entity (ent)

# puts "entity : #{ent.inspect}"

end

end # class

Load the class in IRB and then run:

file = File.open 'master\_scoreboard.xml'

Document.parse\_stream(file, MyListener.new)

NOTE: The way this code is written, the only way to find specific data or count things is to inspect the output manually. It is meant to demonstrate the difference between DOM and SAX parsing.

Also, if you need to reread the document, you have to re-open the file each time, as in:

file = File.open 'master\_scoreboard.xml'

Document.parse\_stream(file, MyListener.new)

a) Who are the winning pitchers whose *last* names begin with G"?

Gonzalez, Greinke

b) Who are the winning pitchers whose *last* names begin with A, B, C, or D"?

Betances, Duffy, Chavez, Bumgarner, Arrieta

c) Who has the highest number of losses among the winning pitchers?

Chavez