

Assignment 9

Robert Trieste
NYIT ID: 1289056

Overview

XML Lab 1 Objective:

1. Checking whether an XML is well-formed.
2. Parse using XML using DOM/SAX and extract information.
3. Validate against a DTD

Checking XML

Consider the following XML, called stkinfo.xml:

```
<?xml version="1.0" encoding="utf-8">
<contributors>
<contribution cname="DoorDash" cname="Food">
<text date="12-09-2021">The necessary threshold is <20 & >10
<status>Approved<status>
<totalassets>34567000</totalassets>
<totalliab>533400</totalliab>
<cash>763445000</cash>
</text>
<contribution cname="" />
<contribution cname="DoorDash" cname="DOOR">
<text date="14-09-2021">The necessary threshold is <20 & >10
<status>Approved<status>
<totalassets>112567000</totalassets>
<totalliab>5334000</totalliab>
<cash>763445000</cash>
</text>
<contribution cname="" />
</contributors>
```

Question 1.

Give all the reasons why the XML above is not well-formed.

```
Line 1: needs ? at end ; fix: <?xml version="1.0" encoding="utf-8"?>
Line 3: duplicate attribute; fix: <contribution cname="DoorDash" cname2="Food">
Line 4: cannot use characters like '<,>,&'; fix: <text date="12-09-2021">The
necessary threshold is 20_and_10
Line 5: need closing tag for 'status'; fix: </status>
```

Line 11: dupe attribute again; fix: <contribution cname="DoorDash" cname2="DOOR">
*closing tags when opening element tag uses an attribute/value pair like
cname="DoorDash" then the closing tag does not use any attributes; so close tag is
just </contribution>
Line 12: illegal characters again
Line 13: close status

Below is the valid version :

```
<?xml version="1.0" encoding="utf-8"?>

<contributors>
  <contribution cname="DoorDash" cname2="Food">
    <text date="12-09-2021">The necessary threshold is 20_and_10
      <status>Approved</status>
      <totalassets>34567000</totalassets>
      <totalliab>533400</totalliab>
      <cash>763445000</cash>
    </text>
  </contribution>
  <contribution cname="DoorDash" cname2="Food" >
    <text date="14-09-2021">The necessary threshold is 20_and_10
      <status>Approved</status>
      <totalassets>112567000</totalassets>
      <totalliab>5334000</totalliab>
      <cash>763445000</cash>
    </text>
  </contribution>
</contributors>
```

Now, check this via the xmllint Unix command. You can use your own LINUX instance or the Virtual Machine 10.22.12.220

```
xmllint -noout stkinfo.xml
```

If the XML is well-formed, the command will not output anything. If, on the other hand, the XML is not well-formed, an error will be output. Cut and paste the errors/output you received here.

****first installed necessary libraries on my AWS EC2 Ubuntu Linux box:
Command 'xmllint' not found, but can be installed with:**

```
sudo snap install libxml2      # version 2.9.9+pkg-0333, or  
sudo apt install libxml2-utils
```

See 'snap info libxml2' for additional versions.

```
ubuntu@ip-172-31-29-243:~/NYIT760/Assignment9$ sudo snap install libxml2  
libxml2 2.9.9+pkg-0333 (brlin) installed
```

Trieste Ubuntu Linux shell version of INCORRECT file Below:

```
ubuntu@ip-172-31-29-243:~/NYIT760/Assignment9$ xmllint -noout stkinfo_INCORRECT.xml  
warning: failed to load external entity "-noout"  
stkinfo_INCORRECT.xml:1: parser error : parsing XML declaration: '>' expected  
<?xml version="1.0" encoding="utf-8">  
      ^  
stkinfo_INCORRECT.xml:3: parser error : Attribute cname redefined  
<contribution cname="DoorDash" cname="Food">  
      ^  
stkinfo_INCORRECT.xml:4: parser error : StartTag: invalid element name  
<text date="12-09-2021">The necessary threshold is <20 & >10  
      ^  
stkinfo_INCORRECT.xml:4: parser error : xmlParseEntityRef: no name  
<text date="12-09-2021">The necessary threshold is <20 & >10  
      ^  
stkinfo_INCORRECT.xml:9: parser error : Opening and ending tag mismatch: status  
line 5 and text  
</text>  
      ^  
stkinfo_INCORRECT.xml:11: parser error : Attribute cname redefined  
<contribution cname="DoorDash" cname="DOOR">  
      ^  
stkinfo_INCORRECT.xml:12: parser error : StartTag: invalid element name  
<text date="14-09-2021">The necessary threshold is <20 & >10  
      ^  
stkinfo_INCORRECT.xml:12: parser error : xmlParseEntityRef: no name  
<text date="14-09-2021">The necessary threshold is <20 & >10  
      ^  
stkinfo_INCORRECT.xml:17: parser error : Opening and ending tag mismatch: status  
line 13 and text  
</text>  
      ^  
stkinfo_INCORRECT.xml:19: parser error : Opening and ending tag mismatch: status  
line 13 and contributors  
</contributors>  
      ^  
stkinfo_INCORRECT.xml:20: parser error : Premature end of data in tag text line 12
```

```
^
stkinfo_INCORRECT.xml:20: parser error : Premature end of data in tag contribution
line 11

^
stkinfo_INCORRECT.xml:20: parser error : Premature end of data in tag status line 5

^
stkinfo_INCORRECT.xml:20: parser error : Premature end of data in tag text line 4

^
stkinfo_INCORRECT.xml:20: parser error : Premature end of data in tag contribution
line 3

^
stkinfo_INCORRECT.xml:20: parser error : Premature end of data in tag contributors
line 2

^
ubuntu@ip-172-31-29-243:~/NYIT760/Assignment9$
```

Then modify the file to make it well-formed. Test it again and cut and paste the corrected xml document here:

Trieste Ubuntu Linux shell version of CORRECT file Below:

```
<?xml version="1.0" encoding="utf-8"?>

<contributors>
  <contribution cname="DoorDash" cname2="Food">
    <text date="12-09-2021">The necessary threshold is 20_and_10
      <status>Approved</status>
      <totalassets>34567000</totalassets>
      <totalliab>533400</totalliab>
      <cash>763445000</cash>
    </text>
  </contribution>
<contribution cname="DoorDash" cname2="Food" >
  <text date="14-09-2021">The necessary threshold is 20_and_10
    <status>Approved</status>
    <totalassets>112567000</totalassets>
    <totalliab>5334000</totalliab>
    <cash>763445000</cash>
  </text>
</contribution>
</contributors>
```

```
ubuntu@ip-172-31-29-243:~/NYIT760/Assignment9$ xmllint -noout Assign_9_Q1.xml
ubuntu@ip-172-31-29-243:~/NYIT760/Assignment9$
```

Parsing XML

In this section, we will use DOM or SAX to parse an input XML file and extract information from it.

JAVA Reference Guide on XML Parsing

For your reference: DOM at <https://docs.oracle.com/javase/tutorial/jaxp/dom/readingXML.html>
and about SAX : <https://docs.oracle.com/javase/tutorial/jaxp/sax/parsing.html>

Familiarize yourself with the XML format by looking over the file and consulting

Using the xml file you modified above, stkinfo.xml; Use a Java DOM or SAX parser to load and parse the map XML file. Your Java Program should print out the following:

- Total Assets
- Total Liability
- Cash

(PLEASE ADD COMMENTS TO YOUR CODE!)

Cut and paste your Java program in the text box below:

```
// java code here!
package com.trieste.xmlvalidator;

import java.io.File;
import javax.xml.parsers.DocumentBuilderFactory;
import javax.xml.parsers.DocumentBuilder;
import org.w3c.dom.Document;
import org.w3c.dom.NodeList;
import org.w3c.dom.Node;
import org.w3c.dom.Element;

public class XMLValidator
{
    public static void main(String[] args){
```

```
try {
    File inputFile = new File("datafiles/Assign_9_Q1.xml");
    DocumentBuilderFactory dbFactory = DocumentBuilderFactory.newInstance();
    /*
     * this is a public abstract class DocumentBuilderFactory extends Object.
     * this Defines a factory API that enables applications to obtain a parser
     * that produces DOM object trees from XML documents.
     *
     * DOM stands for "Document Object Model"
     * The DOM represents the document as nodes and objects; that way,
     * programming languages can interact with the page
     *
     * */
    DocumentBuilder dBuilder = dbFactory.newDocumentBuilder();
    Document doc = dBuilder.parse(inputFile);

    // scl - start at the root element ...
    doc.getDocumentElement().normalize();
    System.out.println("Root element : " + doc.getDocumentElement().getNodeName());

    NodeList nList = doc.getElementsByTagName("contribution");
    System.out.println("----- B E G I N -----");

    // lets iterate thru the array of nodes with a classic for loop
    // 3 components of for loop; initiation position; termination position;
    iterative step-amount (post-fix unary this case)
    for (int temp = 0; temp < nList.getLength(); temp++) {

        Node nNode = nList.item(temp);

        System.out.println("\nCompany : " + nNode.getNodeName());

        if (nNode.getNodeType() == Node.ELEMENT_NODE) {
            Element eElement = (Element) nNode;

            // lets create a series of print line statements then correspond to our
            XML elements and any inner attribute and value pairs
            System.out.println("Parent Company : "
                + eElement.getAttribute("cname"));

            System.out.println("Division : "
                + eElement.getAttribute("cname2"));

            System.out.println("Total Assets : "
                +
                eElement.getElementsByTagName("totalassets").item(0).getTextContent());

            System.out.println("Total Liabilities : "
                + eElement
                .getElementsByTagName("totalliab")
                .item(0)
                .getTextContent());

            System.out.println("Cash : "
                + eElement
```

```
                .getElementsByTagName("cash")
                .item(0)
                .getTextContent());
        }
    }
    System.out.println("----- E N D -----");
} catch (Exception e) {
    e.printStackTrace();
}
}
```

➔ CONSOLE OUTPUT from above:

"C:\Program Files\Java\jdk-12.0.1\bin\java.exe" "-javaagent:C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 2020.2.1\lib\idea_rt.jar=49849:C:\Program Files\JetBrains\IntelliJ IDEA Community Edition 2020.2.1\bin" -Dfile.encoding=UTF-8 -classpath D:\OneDrive\NYIT\CSCI-760_DatabaseSystems\JAVA\XMLValidator\out\production\XMLValidator
com.trieste.xmlvalidator.XMLValidator

Root element : contributors

----- B E G I N -----

Company :contribution

Parent Company : DoorDash

Division : Food

Total Assets : 34567000

Total Liabilities : 533400

Cash : 763445000

Company :contribution

Parent Company : DoorDash

Division : DOOR

Total Assets : 112567000

Total Liabilities : 5334000

Cash : 763445000

----- E N D -----

Process finished with exit code 0

Optional (Graduate Students Only)

Create a DTD file which validates your tpfixed.xml. Call it tp.dtd. Validate it using

xmllint:

```
xmllint -noout -dtd-valid tp.dtd tpfixed.xml
```

My Version:

one optional way to construct:

<https://docs.microsoft.com/en-us/troubleshoot/developer/visualstudio/visual-basic/language-compilers/validate-xml-document-by-dtd-xdr-xsd>

```
xmllint -noout -dtd-valid Assign_9_Q1.dtd Assign_9_Q1_WithDTD.xml
```

xml / dtd files can be separated but in same folder with references in XML to the DTD or they can be in combined style like below:

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE contributors
[
<!ELEMENT contributors (contribution,text,status,totalassets,totalliab,cash) >
<!ELEMENT contribution EMPTY>
<!ATTLIST contribution cname CDATA #REQUIRED >
<!ATTLIST contribution cname2 CDATA #REQUIRED >
<!ELEMENT text EMPTY>
<!ATTLIST text date CDATA #REQUIRED>
<!ELEMENT status CDATA #REQUIRED>
<!ELEMENT totalassets CDATA #REQUIRED>
<!ELEMENT totalliab CDATA #REQUIRED>
<!ELEMENT cash CDATA #REQUIRED>
<!ELEMENT contribution>
]>
```



```
<contributors>
  <contribution cname="DoorDash" cname2="Food">
    <text date="12-09-2021">The necessary threshold is 20_and_10
      <status>Approved</status>
      <totalassets>34567000</totalassets>
      <totalliab>533400</totalliab>
      <cash>763445000</cash>
    </text>
  </contribution>
<contribution cname="DoorDash" cname2="D00R" >
  <text date="14-09-2021">The necessary threshold is 20_and_10
    <status>Approved</status>
    <totalassets>112567000</totalassets>
    <totalliab>5334000</totalliab>
    <cash>763445000</cash>
  </text>
</contribution>
</contributors>

/*
*end work
*Robert Trieste
*/
```