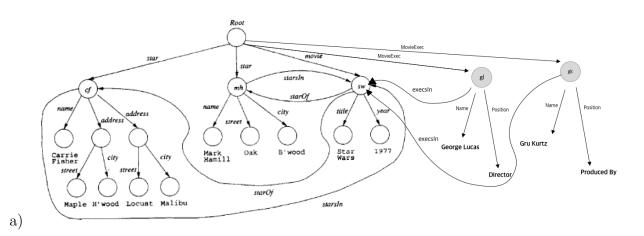
CSC343 Worksheet 9 Solution

June 28, 2020

1. Exercise 11.1.1:



Notes:

- Semistructured data
 - serves as a model suitable for **databases integration**, that is, for describing the data contained in two or more databases that contain similar data with different schemas
 - It serves as the underlying model for notations such as XML, to be taken up in Section 2, that are being used to share information on the web.
- Semistructured Data Representation
 - is a collection of nodes















4. The difference is that UML must fit data into its schema, where as the semi structured data allows whatever schema information that is appropriate to be attached to data

$\underline{\mathbf{Notees:}}$

- Semi-structured Data
 - Is schemaless

- Is motivated primarily by its flexibility
- One could enter data at will, and attach to the data whatever schema information you felt was appropriate for that data.
- Makes query processing harder

• Structured Data

- Is rigid framework into which data is placed.
- Data must fit into schema
- Fixed schema allows data to be organized with data structures that support efficient answering of queries
- e.g. UML, E/R, Relational, ODL

```
5. a)
         <? xml version = "1.0" encoding="utf-8" standalone = "yes">
         <StarMovieData>
              <Star starID="cf" starredIn="sw">
    3
                  <Name>Carrie Fisher</Name>
                  <Address>
                      <Street>123 Maple St.</Street>
    6
                      <City>Hollywood</City>
                  </Address>
                  <Address>
   9
                      <Street>5 Locust Ln.</Street>
                      <City>Malibu</City>
                  </Address>
              </Star>
              <Star starID="mh" starredIn="sw">
   14
                  <Name>Mark Hamill</Name>
                  <Street>456 Oak Rd.</Street>
                  <City>Brentwood</City>
              </Star>
   18
              <Movie movieID="sw" starsOf="cf", "mh">
   19
                  <Title>Star Wars</Title>
   20
                  <Year>1977</Year>
   21
              </Movie>
              <MovieExec movieExecID="gl" execsIn="sw">
   23
                  <Name>George Lucas</Name>
   24
                  <Position>Director</Position>
   25
              </MovieExec>
   26
              <MovieExec movieExecID="gk" execsIn="sw">
   27
                  <Name>Gru Kurtz</Name>
                  <Position>Produced By</Position>
   29
              </MovieExec>
   30
          </StarMovieData>
   31
```

• XML

- is called Extensible Markup Language
- is an example of semistructured data
- XML with and without a Schema

- has two different types
 - 1. Well-formed XML
 - * allows to invent your own tags
 - * corresponds very-similarly to semi-structured data

Example:

```
<? xml version = "1.0" encoding="utf-8" standalone = "yes</pre>
      <StarMovieData>
           <Star>
               <Name>Carrie Fisher</Name>
               <Address>
                   <Street>123 Maple St.</Street>
                   <City>Hollywood</City>
               </Address>
               <Address>
9
                   <Street>5 Locust Ln.</Street>
                   <City>Malibu</City>
11
               </Address>
12
          </Star>
13
           <Star>
14
               <Name>Mark Hamill</Name>
               <Street>456 Oak Rd.</Street>
               <City>Brentwood</City>
17
          </Star>
18
           <Movie>
               <Title>Star Wars</Title>
20
               <Year>1977</Year>
           </Movie>
      </StarMovieData>
```

2. Valid XML

- * Involes "Document Type Definition"
- * specifies allowable tags and gives a grammar for how they may be nested

Attributes

- is used to represent connections in a semistructured data graph

Example:

```
</Address>
               <Address>
9
                   <Street>5 Locust Ln.</Street>
11
                   <City>Malibu</City>
               </Address>
12
           </Star>
          <Star starID="mh" starredIn="sw">
14
               <Name>Mark Hamill</Name>
               <Street>456 Oak Rd.</Street>
               <City>Brentwood</City>
          </Star>
18
           <Movie starID="sw" starOf="cf", "mh">
19
               <Title>Star Wars</Title>
               <Year>1977</Year>
21
           </Movie>
22
      </StarMovieData>
23
```

Namespaces

- **Syntax:** xmlns:name:URI
- Is similar to import numpy as np in python
- Is used to distinguish tags coming from different sources, i.e. HTML

Example:

Retireving element *StarMovieData* from document **infolab.stanford.edu/movies**. Set md as the name of import

```
<md:StarMovieData xmlns:md="http://infolab.stanford.edu/
movies">
```

```
b)
       <? xml version = "1.0" encoding="utf-8" standalone = "yes">
       <StarMovieData>
 2
           <Star starID="cf" starredIn="sw">
               <Name>Carrie Fisher</Name>
 4
               <Address>
                    <Street>123 Maple St.</Street>
 6
                    <City>Hollywood</City>
               </Address>
 8
               <Address>
 9
                    <Street>5 Locust Ln.</Street>
                    <City>Malibu</City>
               </Address>
13
           </Star>
           <Star starID="mh" starredIn="sw">
14
               <Name>Mark Hamill</Name>
               <Street>456 Oak Rd.</Street>
16
               <City>Brentwood</City>
17
18
           </Star>
           <Movie movieID="sw" starsOf="cf", "mh">
19
               <Title>Star Wars</Title>
20
               <Year>1977</Year>
21
```

```
</Movie>
           <Movie movieID="esb" starOf="cf", "mh">
23
               <Title>Empire Strikes Back</Title>
24
               <Year>1980</Year>
25
           </Movie>
26
           <Movie movieID="roj" starOf="cf", "mh">
27
               <Title>Return of Jedi</Title>
28
               <Year>1983</Year>
29
           </Movie>
30
31
       </StarMovieData>
32
c)
       <? xml version = "1.0" encoding="utf-8" standalone = "yes">
       <StarMovieData>
 2
           <Star starID="cf" starredIn="sw">
 3
               <Name>Carrie Fisher</Name>
               <Address>
                    <Street>123 Maple St.</Street>
 6
                    <City>Hollywood</City>
               </Address>
 8
               <Address>
 9
                    <Street>5 Locust Ln.</Street>
                    <City>Malibu</City>
               </Address>
           </Star>
           <Star starID="mh" starredIn="sw">
14
               <Name>Mark Hamill</Name>
15
               <Street>456 Oak Rd.</Street>
16
               <City>Brentwood</City>
           </Star>
18
           <Movie movieID="sw" starsOf="cf", "mh" movieIn="fx">
19
               <Title>Star Wars</Title>
20
               <Year>1977</Year>
21
           </Movie>
22
           <Movie movieID="esb" starOf="cf", "mh" movieIn="fx">
23
               <Title>Empire Strikes Back</Title>
24
               <Year>1980</Year>
25
           </Movie>
26
           <Movie movieID="roj" starOf="cf", "mh" movieIn="fx">
27
               <Title>Return of Jedi</Title>
28
               <Year>1983</Year>
29
           </Movie>
30
           <Studio studioID="fx" movieOf="esb", "roj", "sw">
31
               <Name>Fox</Name>
               <Address>Hollywood</Address>
34
           </Studio>
       </StarMovieData>
35
```

6. Consider the following relation Classes:

class	type	country	numGuns	bore	displacement
Bismarck	bb	Germany	8	15	42000
Iowa	ъъ	USA	9	16	46000
Kongo	Ъc	Japan	8	14	32000
North Carolina	bb	USA	9	16	37000
Renown	bc	Gt. Britain	6	15	32000
Revenge	bb	Gt. Britain	8	15	29000
Tennessee	bь	USA	12	14	32000
Yamato	bb	Japan	9	18	650CO

(a) Sample data for relation Classes

```
<? xml version = "1.0" encoding="utf-8" standalone = "yes">
2
      <shipsData>
          <Classes>
              <Class>Bismarck</Class>
              <Type>bb</Type>
              <Country>Germany</Country>
              <NumGuns>8</NumGuns>
              <Bore>15</Bore>
              <Displacement>42000
          </Classes>
          <Classes>
              <Class>Iowa</Class>
              <Type>bb</Type>
13
              <Country>USA</Country>
14
              <NumGuns>9</NumGuns>
              <Bore>16</Bore>
16
              <Displacement>46000</Displacement>
17
          </Classes>
          <Classes>
19
              <Class>Kongo</Class>
20
              <Type>bc</Type>
21
              <Country>Japan</Country>
              <NumGuns>8</NumGuns>
23
              <Bore>14</Bore>
24
              <Displacement>32000
25
          </Classes>
26
          <Classes>
27
28
              <Class>North Carolina</Class>
              <Type>bb</Type>
29
              <Country>USA</Country>
30
              <NumGuns>9</NumGuns>
31
              <Bore>16</Bore>
32
              <Displacement>37000</Displacement>
33
          </Classes>
34
          <Classes>
35
              <Class>Renown</Class>
36
```

```
<Type>bc</Type>
37
               <Country>Gt. Britain</Country>
               <NumGuns>6</NumGuns>
39
               <Bore>15</Bore>
40
               <Displacement>32000
41
          </Classes>
42
          <Classes>
43
               <Class>Revenge</Class>
44
               <Type>bb</Type>
45
46
               <Country>Gt. Britain</Country>
               <NumGuns>8</NumGuns>
47
               <Bore>15</Bore>
48
               <Displacement>29000
49
          </Classes>
          <Classes>
51
               <Class>Tennessee</Class>
52
               <Type>bb</Type>
53
               <Country>USA</Country>
               <NumGuns>12</NumGuns>
55
               <Bore>14</Bore>
56
               <Displacement>32000</Displacement>
57
          </Classes>
58
          <Classes>
59
               <Class>Yamato</Class>
60
               <Type>bb</Type>
61
               <Country>Japan</Country>
62
               <NumGuns>9</NumGuns>
63
               <Bore>18</Bore>
64
               <Displacement>65000</Displacement>
65
          </Classes>
66
      </shipsData>
67
```

7. • DocRoot

SubElement

```
Subelement parentID="" childID="" position="" />
```

• ElementAttribute

```
<ElementAttribute elementID="" name="" value="" />
```

• ElementValue

```
<ElementValue elementID="" name="" value="" />
```

Notes:

- Empty element
 - is an element that is complete by itself
 - rather than being composed of a start tag, data and an end tag, the empty element is a combined start and end tag.

Example:

- * Phone is not an empty element
- * extension is an empty element
- 8. The following will be used as an example.



```
<ElementValue elementID="cf-name" name="name" value="Carrie</pre>
     Fisher"/>
               <SubElement parentID="cf" childID="cf-add_1" position="1">
6
                   <City elementID="cf-add_1-city" value="Hollywood">
                   <Street elementID="cf-add_1-street" value="Maple">
               </SubElement>
9
               <SubElement parentID="cf" childID="cf-add_2" position="2">
10
                   <ElementValue elementID="cf-add_2-city" value="Malibu">
                   <ElementValue elementID="cf-add_1-street" value="Locust"</pre>
     >
               </SubElement>
13
          </SubElement>
14
          <SubElement parentID="root" childID="mh" position="2">
               <ElementAttribute elementID="mh-starsIn" name="starsIn"</pre>
     value="sw"/>
               <ElementValue elementID="mh-name" value="Mark Hamill"/>
17
               <SubElement parentID="mh" childID="mh-add_2" position="1">
18
                   <ElementValue elementID="mh-add_2-city" value="Bollywood"</pre>
19
     ">
                   <ElementValue elementID="mh-add_2-street" value="Oak">
20
               </SubElement>
21
          </SubElement>
22
          <SubElement parentID="root" childID="sw" position="3">
23
               <ElementAttribute elementID="mh-starsOf_1" name="starsOf"</pre>
     value="cf"/>
              <ElementAttribute elementID="mh-starsOf_2" name="starsOf"
25
     value="mh"/>
               <ElementValue elementID="sw-title" value="Star Wars" />
26
               <ElementValue elementID="sw-year" value="1977" />
27
          </SubElement>
28
      </DocRoot>
29
```

```
9. a)
         <? xml version = "1.0" encoding="utf-8" standalone = "yes">
         <StarMovieData>
              <Star starID="cf" starredIn="sw">
                  <Name>Carrie Fisher</Name>
   4
                  <Address>
                      <Street>123 Maple St. <Street>
   6
                      <City>Hollywood</City>
                  </Address>
   8
                  <Address>
   9
                      <Street>5 Locust Ln.</Street>
                      <City>Malibu</City>
                  </Address>
   13
              </Star>
              <Star starID="mh" starredIn="sw">
   14
                  <Name>Mark Hamill</Name>
                  <Address>
   16
                      <Street>456 Oak Rd.</Street>
   17
                      <City>Brentwood</City>
   18
                  </Address>
   19
              </Star>
   20
              <Movie movieID="sw" starsOf="cf mh">
   21
```

```
<Title>Star Wars</Title>
               <Year>1977</Year>
23
          </Movie>
24
          <Movie movieID="esb" starsOf="cf mh">
25
               <Title>The Empire Strikes Back</Title>
26
               <Year>1980</Year>
27
          </Movie>
28
          <Movie movieID="roj" starsOf="cf mh">
29
               <Title>Return of the Jedi</Title>
30
31
               <Year>1983</Year>
          </Movie>
32
      </StarMovieData>
33
34
      <? xml version = "1.0" encoding="utf-8" standalone = "yes">
      <StarMovieData>
2
          <Star starID="cf" starredIn="sw">
3
               <Name>Carrie Fisher</Name>
               <Address>
                   <Street>123 Maple St. <Street>
6
                   <City>Hollywood</City>
               </Address>
               <Address>
9
                   <Street>5 Locust Ln.</Street>
                   <City>Malibu</City>
               </Address>
          </Star>
          <Star starID="hf" starredIn="sw fw">
14
               <Name>Harrison Ford</Name>
          </Star>
16
17
          <Star starID="mh" starredIn="sw">
18
               <Name>Mark Hamill</Name>
19
               <Address>
20
                   <Street>456 Oak Rd.</Street>
21
                   <City>Brentwood</City>
22
               </Address>
          </Star>
24
          <Movie movieID="sw" starsOf="cf mh hf">
25
               <Title>Star Wars</Title>
26
               <Year>1977</Year>
          </Movie>
28
          <Movie movieID="esb" starsOf="cf mh hf">
29
               <Title>The Empire Strikes Back</Title>
30
               <Year>1980</Year>
31
          </Movie>
          <Movie movieID="roj" starsOf="cf mh hf">
33
               <Title>Return of the Jedi</Title>
34
               <Year>1983</Year>
35
          </Movie>
36
          <Movie movieID="fw" starsOf="hf">
37
               <Title>Firewall</Title>
38
               <Year>2006</Year>
40
          </Movie>
```

```
41 </StarMovieData>
42
```

```
<? xml version = "1.0" encoding="utf-8" standalone = "yes">
b)
       <StarMovieData>
           <Star starID="cf" starredIn="sw">
 3
               <Name>Carrie Fisher</Name>
                <Address>
                    <Street>123 Maple St. <Street>
 6
                    <City>Hollywood</City>
               </Address>
 8
               <Address>
 9
10
                    <Street>5 Locust Ln.</Street>
                    <City>Malibu</City>
               </Address>
           </Star>
13
           <Star starID="hf" starredIn="sw fw">
14
               <Name>Harrison Ford</Name>
15
           </Star>
17
           <Star starID="mh" starredIn="sw">
18
               <Name>Mark Hamill</Name>
               <Address>
20
                    <Street>456 Oak Rd.</Street>
21
                    <City>Brentwood</City>
               </Address>
23
           </Star>
24
           <Movie movieID="sw" starsOf="cf mh hf">
25
               <Title>Star Wars</Title>
26
27
               <Year>1977</Year>
           </Movie>
28
           <Movie movieID="esb" starsOf="cf mh hf">
29
               <Title>The Empire Strikes Back</Title>
30
               <Year>1980</Year>
31
           </Movie>
32
           <Movie movieID="roj" starsOf="cf mh hf">
33
               <Title>Return of the Jedi</Title>
                <Year>1983</Year>
35
           </Movie>
36
           <Movie movieID="fw" starsOf="hf">
               <Title>Firewall</Title>
38
                <Year>2006</Year>
39
           </Movie>
40
           <Movie movieID="hhs" starsOf="cf">
41
               <Title>Hannah and Her Sisters</Title>
42
               <Year>1985</Year>
43
           </Movie>
44
       </StarMovieData>
45
46
```

```
d) <? xml version = "1.0" encoding="utf-8" standalone = "yes"> <StarMovieData>
```

```
<Star starID="cf" starredIn="sw">
               <Name>Carrie Fisher</Name>
               <Address>
                   <Street>123 Maple St.<Street>
                   <City>Hollywood</City>
               </Address>
8
               <Address>
Q
                   <Street>5 Locust Ln.</Street>
                   <City>Malibu</City>
               </Address>
          </Star>
          <Star starID="hf" starredIn="sw fw">
14
               <Name>Harrison Ford</Name>
           </Star>
          <Star starID="mh" starredIn="sw">
17
               <Name>Mark Hamill</Name>
18
               <Address>
19
                   <Street>456 Oak Rd.</Street>
20
                   <City>Brentwood</City>
21
               </Address>
          </Star>
          <Star starID="md" starredIn="bi">
24
               <Name>Matt Damon</Name>
25
          </Star>
26
           <Movie movieID="sw" starsOf="cf mh hf">
27
               <Title>Star Wars</Title>
28
               <Year>1977</Year>
29
          </Movie>
30
           <Movie movieID="esb" starsOf="cf mh hf">
31
               <Title>The Empire Strikes Back</Title>
               <Year>1980</Year>
          </Movie>
          <Movie movieID="roj" starsOf="cf mh hf">
35
               <Title>Return of the Jedi</Title>
36
               <Year>1983</Year>
          </Movie>
38
          <Movie movieID="fw" starsOf="hf">
39
               <Title>Firewall</Title>
40
               <Year>2006</Year>
41
          </Movie>
42
          <Movie movieID="hhs" starsOf="cf">
43
               <Title>Hannah and Her Sisters</Title>
44
               <Year>1985</Year>
45
          </Movie>
46
          <Movie movieID="bi" starsOf="md">
47
               <Title>The Bourne Identity</Title>
48
               <Year>2002</Year>
49
          </Movie>
50
      </StarMovieData>
51
52
```

```
10<sub>1</sub> <!DOCTYPE BankData [
2 <!ELEMENT BankData (Accounts*, Customers*)>
3 <!ELEMENT Accounts(acctNo, type, balance)>
```

```
<! ATTLIST Accounts
                   acctNo ID #REQUIRED
5
                   acctIn IDREFS #IMPLIED
6
          <!ELEMENT Customers(firsrName, lastName, idNo, account)>
               <! ATTLIST Accounts
9
                   idNo ID #REQUIRED
                   account IDREFS #IMPLIED
11
12
13
          <!ELEMENT acctNo (#PCDATA)>
          <!ELEMENT type (#PCDATA)>
14
          <!ELEMENT balance (#PCDATA)>
15
          <!ELEMENT firstName (#PCDATA)>
          <!ELEMENT lastName (#PCDATA)>
17
          <!ELEMENT idNo (#PCDATA)>
          <!ELEMENT account (#PCDATA)>
19
      ] >
20
21
```

```
Correct Solution:
      <!DOCTYPE BankData [</pre>
          <!ELEMENT BankData (Accounts*, Customers*)>
2
           <! ELEMENT Accounts(acctNo, type, balance)>
               <! ATTLIST Accounts
                   acctNo ID #REQUIRED
5
                   acctIn IDREFS #IMPLIED
          <!ELEMENT Customers(firsrName, lastName, idNo, account)>
               <! ATTLIST Accounts
9
                   idNo ID #REQUIRED
10
                   account IDREFS #IMPLIED
11
12
          <!ELEMENT acctNo (#PCDATA)>
13
          <!ELEMENT type (#PCDATA)>
14
          <!ELEMENT balance (#PCDATA)>
15
          <!ELEMENT firstName (#PCDATA)>
16
          <!ELEMENT lastName (#PCDATA)>
17
          <!ELEMENT idNo (#PCDATA)>
18
          <!ELEMENT account (#PCDATA)>
19
      ] >
20
21
```

```
<!ELEMENT LastName (#PCDATA)>
           <!ELEMENT Team (#PCDATA)>
           <! ELEMENT Teams(Name)>
11
12
               <! ATTLIST Teams
                    teamId ID #REQUIRED
13
                    playerOf IDREFS #IMPLIED
                    fanOf IDREFS #IMPLIED
           <!ELEMENT Name (#PCDATA)>
           <!ELEMENT Fans(FirstName, LastName)>
18
               <! ATTLIST Fans
19
                   faId ID #REQUIRED
20
                    fanIn IDREFS #IMPLIED
21
22
23
      ] >
24
```

Notes:

• There is no need to write same tags twice. (e.g. FirstName, LastName)

```
12_1
        <!DOCTYPE GenealogyData [</pre>
            <!ELEMENT TeamData (Players*, Teams*, Fans*)>
 2
            <!ELEMENT Players(FirstName, LastName, Team)>
 3
                 <!ATTLIST Players
                     playerId ID #REQUIRED
                     playerIn IDREFS #IMPLIED
 6
                 >
            <!ELEMENT FirstName (#PCDATA)>
 8
            <!ELEMENT LastName (#PCDATA)>
 9
            <!ELEMENT Team (#PCDATA)>
            <! ELEMENT Teams (Name) >
 11
                 <! ATTLIST Teams
 12
                     teamId ID #REQUIRED
 13
                     playerOf IDREFS #IMPLIED
 14
 15
                     fanOf IDREFS #IMPLIED
 16
            <!ELEMENT Name (#PCDATA)>
 17
            <!ELEMENT Fans(FirstName, LastName)>
 18
                 <! ATTLIST Fans
 19
                     faId ID #REQUIRED
 20
                     fanIn IDREFS #IMPLIED
 21
 22
       ] >
 23
```

Notes:

• PCDATA

- Is text that will be parsed by a parser.
- Tags inside the text will be treated as markup and entities will be expanded.

• CDATA

- Is text that will not be parsed by a parser.
- Tags inside the text will not be treated as markup and entities will not be expanded
- e.g. attributes

```
14<sub>1</sub> <!DOCTYPE ShipBattleData [

2
3 ]>
```

15. **Notes:**

• XML Schema

- is an alternative way to provide a schema for XML documents
- is more powerful than DTD
 - * allows to declare types such as integer or float
 - * gives ability to declare keys and foreign keys

- The Form of an XML Schema
 - Each XML-Schema document has the form

- Elements
 - Syntax:

```
<xs:element name = element name type = element type>
  constraints and / or structure information
</xs:element>
```

Example:

```
<xs:element name = "title" type = "xs:string" />
<xs:element name = "Year" type = "xs:integerg" />
3
4
```

- Complex Types
 - is an XML element that contains other elements and/or attributes. ^[1]
 - Syntax:

```
<xs:complexType name = type name>
  <xs:sequence>
    list of element definitions
  </xs:sequence>
</xs:complexType>
```

Example:

- Complex Types Attributes
 - Syntax:

<xs:attribute name = attribute name type = type name other information about the attribute />

- * other information includes
 - 1. default value
 - 2. required
 - 3. type

Example:

```
<? xml version = "1.0" encoding = "utf-8" ?>
1)
    <xs:schema xmlns:xs = "http://www.w3.org/2001/XMLSchema">
        <xs:complexType name = "movieType">
3)
4)
            <xs:attribute name = "title" type = "xs:string"</pre>
                use = "required" />
            <xs:attribute name = "year" type = "xs:integer"</pre>
5)
                use = "required" />
        </xs:complexType>
6)
        <xs:element name = "Movies">
7)
8)
            <xs:complexType>
                                                                                        attribute definition
9)
                   <xs:element name = "Movie" type = "movieType"</pre>
10)
                       minOccurs = "0" maxOccurs = "unbounded" />
               </xs:sequence>
11)
                                                                                          linkage of attribute definition
12)
            </xs:complexType>
         </xs:element>
13)
     </xs:schema>
                                                                                             element definition
```

Above allows us to have the tag <Movies title="value" year="1">

- Restricted Simple Types
 - Can be used as an attribute or element
 - 1. Restricting numerical values by setting lower bound and upperbound
 - 2. Restricting avlues to an enumerated type

References:

1. w3School: Complex Type, link