READING DATA FROM XML FILES

XML is a language that is useful to store and transport data. XML and JSON are industry standards to store data and distribute to other programmers.

- XML stands for eXtensible Markup Language
- XML is a markup language much like HTML
- XML was designed to store and transport data
- XML was designed to be self-descriptive
- XML uses user-defined tags

Example (cust-data.xml) :

```
<?xml version="1.0" encoding="utf-8" ?>
<customers>
  <customer type="business">
    <place>UK</place>
    <amount>15000</amount>
  </customer>
  <customer type="personal">
    <place>India</place>
    <amount>25000</amount>
  </customer>
  <customer type="personal">
    <place>Canada</place>
    <amount>20000</amount>
  </customer>

</customers>
```

Note: There should not be gap between <? and xml in the beginning word.

In the above xml document, the first line is not compulsory. It tells that this is an xml document. XML documents must contain one root element that is the parent of all other elements. In the above <customers> is the root element. The child tags under this root element are <customer> tags. The subchild tags are <place> and <amount>.

To load xml file, we use parse() function of ElementTree module in Python, as:

```
import xml.etree.ElementTree as ET
tree = ET.parse('F:/py/cust-data.xml')
```

To retrieve the root tag:

```
root = tree.getroot()
print(root.tag)
```

To retrieve the child tags:

```
for child in root:
    print('Child tag name= ', child.tag)
```

To retrieve each child tag information:

```
for c in root.findall('customer'):  # here <customer> is child tag
    a = c.attrib

    # retrieve subchild (other) tags
```

```
    other = c.find('place').text
    print('Place=', other)
```

See the complete program below:

readxml.py

```
# read data from xml file using ElementTree module
import xml.etree.ElementTree as ET
tree = ET.parse('F:/py/cust-data.xml')

# go to the root of the tree structure
root = tree.getroot()

# display root tag name
print('Root tag name= ', root.tag)

# display child tags names
for child in root:
    print('Child tag name= ', child.tag)

# display each child(attribute) info
for c in root.findall('customer'):
    a = c.attrib
    print(a)  #{'type': 'business"}
    t = a.get('type')
    print(t)  #business

    # retrieve subchild (other) tags
    other = c.find('place').text
    print('Place=', other)

    other1 = c.find('amount').text
    print('Amount=', other1)
```

Output:

```
F:\py>python x.py
Root tag name=  customers
Child tag name=  customer
Child tag name=  customer
Child tag name=  customer
{'type': 'business'}
business
Place= UK
Amount= 15000
{'type': 'personal'}
personal
Place= India
Amount= 25000
{'type': 'personal'}
personal
Place= Canada
Amount= 20000
```

Assignment:

As shown above, read all tags and their information from the file named gdp-data.xml.