COMPILER CONSTRUCTION #03

Particulars:

Name:	Asad Abbas
Reg No:	200901024
Date:	08-01-2023

Code:

```python
import xml.etree.ElementTree as ET  #Python parser
import openpyxl  #Use this for manipulating excel files

tree = ET.parse('compiler.xml')  #Parse XML file
root = tree.getroot() #fetch XML file

#Create new excel file
```

```python
workbook = openpyxl.Workbook()
worksheet = workbook.active

#Adding the column headers
worksheet.cell(row=1, column=1).value = 'Book ID'
worksheet.cell(row=1, column=2).value = 'Author Name'
worksheet.cell(row=1, column=3).value = 'Title'
worksheet.cell(row=1, column=4).value = 'Genre'
worksheet.cell(row=1, column=5).value = 'Price'
worksheet.cell(row=1, column=6).value = 'Publish Date'
worksheet.cell(row=1, column=7).value = 'Description'

row = 2 #adding data in 2nd row
for book in root.findall('book'):
  book_id = book.get('id')
  author_name = book.find('author').text
  title = book.find('title').text
  genre = book.find('genre').text
  price = book.find('price').text
  publish_date = book.find('publish_date').text
  description = book.find('description').text

  worksheet.cell(row=row, column=1).value = book_id
  worksheet.cell(row=row, column=2).value = author_name
  worksheet.cell(row=row, column=3).value = title
  worksheet.cell(row=row, column=4).value = genre
  worksheet.cell(row=row, column=5).value = price
  worksheet.cell(row=row, column=6).value = publish_date
  worksheet.cell(row=row, column=7).value = description

  row += 1

file_name = input("Enter the name for the Excel file: ")  #input to ask for
filename (it will be stored in the same folder as this file)

try:
  workbook.save(file_name + '.xlsx')   #Save the Excel file
  print("Excel file successfully created!")
except Exception as e:
  print("There was an error saving the Excel file:", e)
```

Output:

```
' 'c:\Users\Asad\Desktop\Semester 5\Compiler Construction\Assignment\:
Enter the name for the Excel file: 200901024_Assign_03
Excel file successfully created!
PS C:\Users\Asad\Desktop\Semester 5\Compiler Construction\Assignment>
```

Explanation:

In order to parse the given XML file into Excel using Python, we write a script for this purpose. We use the following two libraries:

- **xml.etree.ElementTree** for parsing the XML file
- **openpyxl** for manipulating Excel files.

Firstly, we parse the XML file using ElementTree and get the root element of the document. Then, a new Excel workbook is created and the active worksheet is obtained using openpyxl module. Column headers are added to the worksheet, and then the data for each book in the XML file is extracted and added to the worksheet. Finally, the user is prompted to enter a file name for the Excel file and the workbook is saved with that name.

Why use xml.etree.ElementTree

We could use DOM (Document Object Model,
it is a fully functional XML parser that supports all DOM capabilities, such as namespace support and validation. However, compared to certain other parsers, it could be slower and require more RAM. Similarly, SAX parsers are event-based, which can be more efficient than DOM parsers, but it requires more complex code to handle the events and extract the data you need. xml.parsers.expat also requires complex code. Therefore, xml.etree.ElementTree is the most suitable parser for the current requirements.

GitHub Link:

https://github.com/asadnaqviii/Semester-5--OS---CC