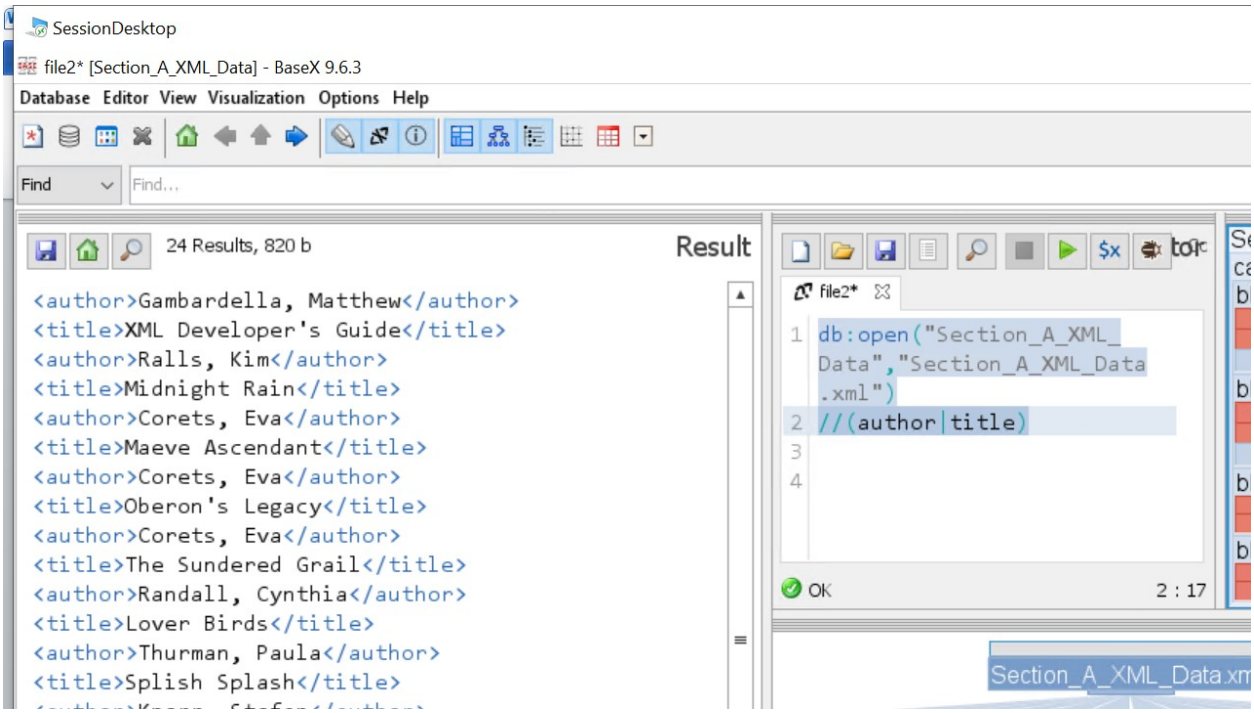
Assignment 3	Course : CIND110
Student : Samaraweera Wijesooriya	Date : August 8, 2022

Section A

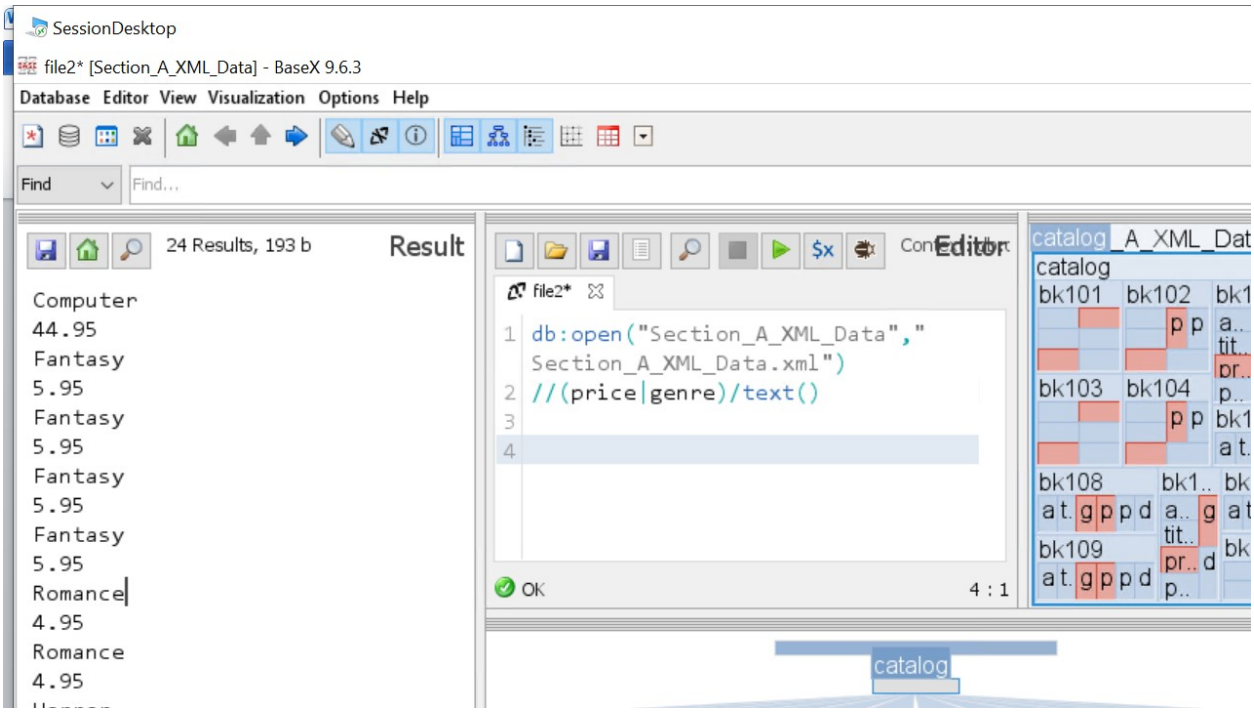
1. [4 Pts.] Write an XPath expression to find all authors along with their corresponding books.

   db:open("Section_A_XML_Data","Section_A_XML_Data.xml")
   //(author|title)



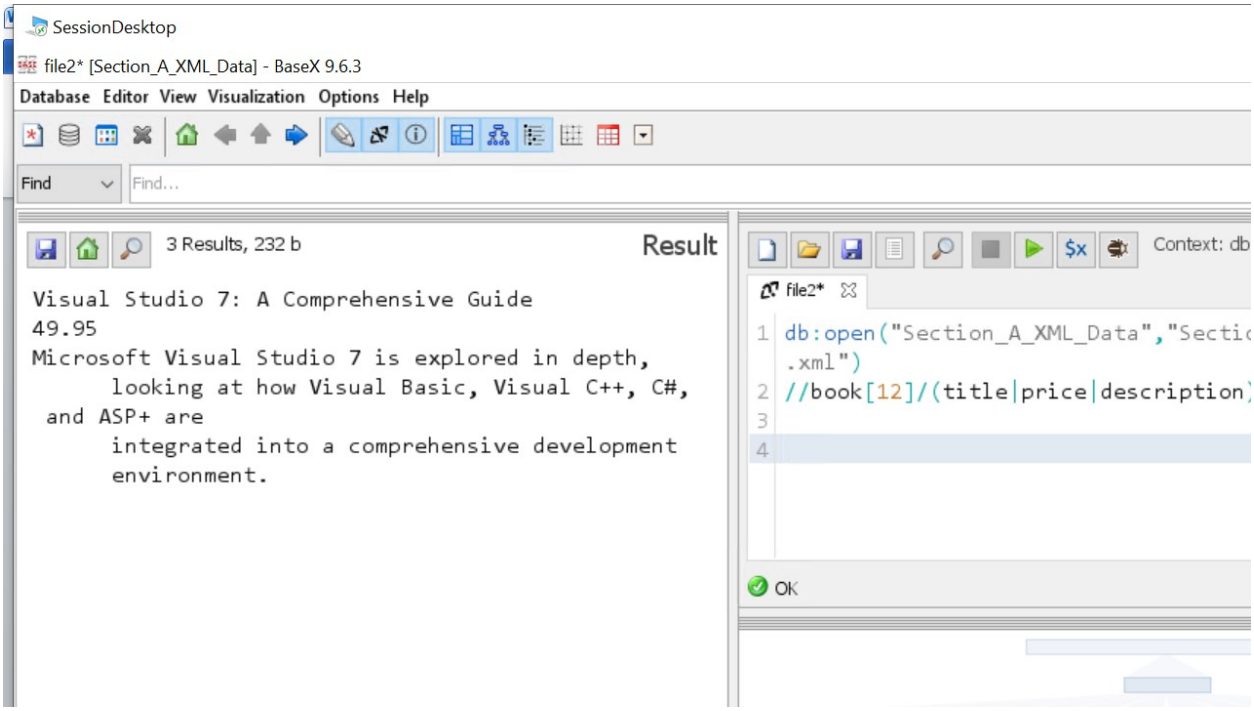
2. [4 Pts.] Write an XPath expression to find the prices of all books and their genre.

db:open("Section_A_XML_Data","Section_A_XML_Data.xml")
//(price|genre)/text()



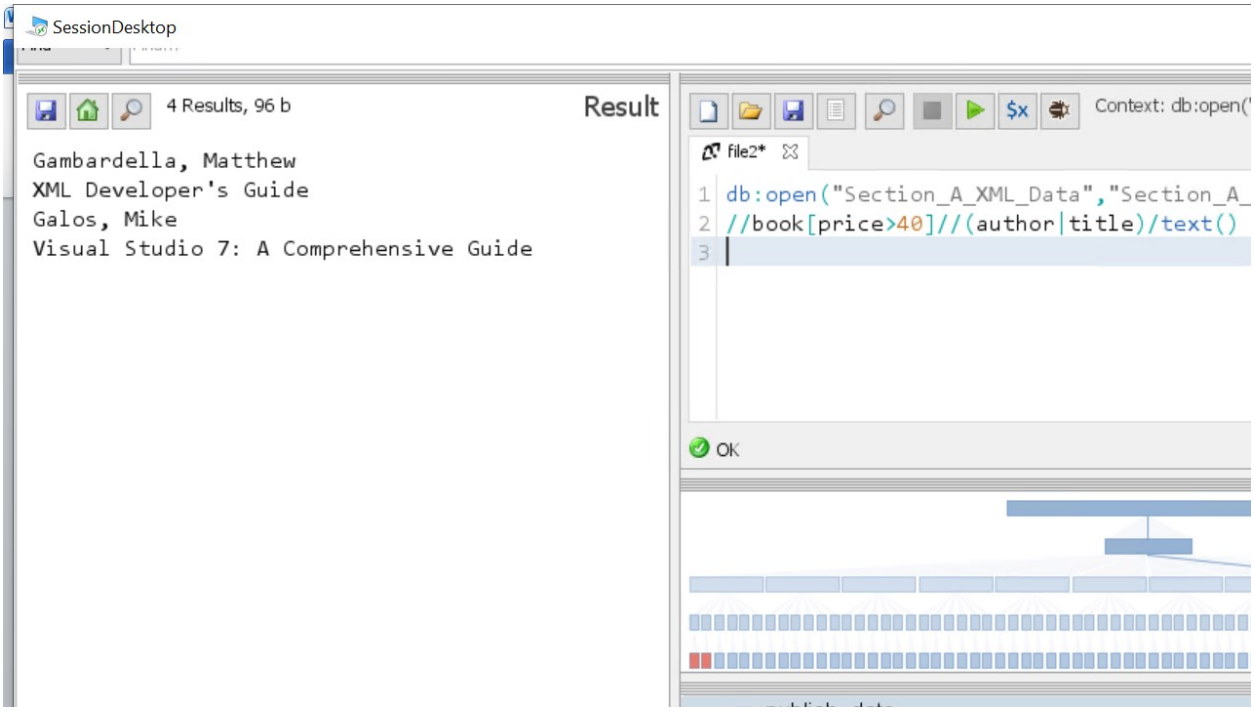
3. [4 Pts.] Write an XPath expression to find the title, price and the description in text of the last book in the catalog.

db:open("Section_A_XML_Data","Section_A_XML_Data.xml")
//book[12]/(title|price|description)/text()



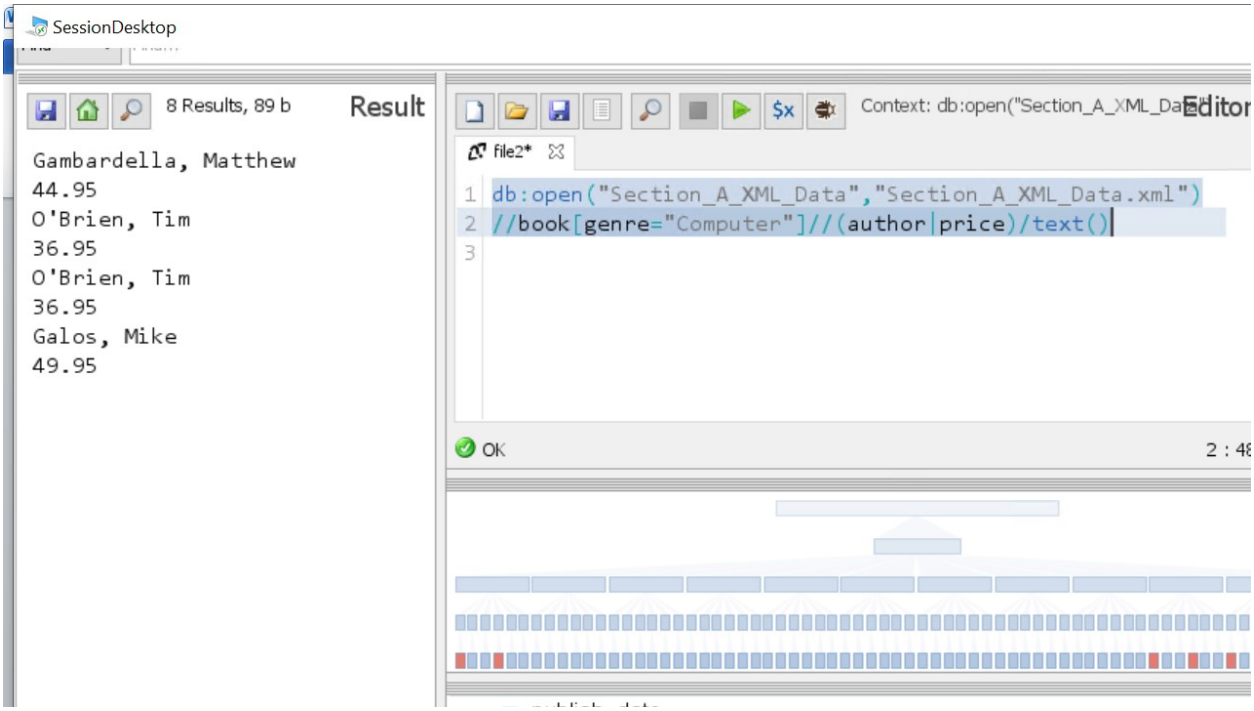
4. [4 Pts.] Write an XPath expression to find the authors and titles of the books which cost more than 40 dollars, along with the respective prices.

```
db:open("Section_A_XML_Data","Section_A_XML_Data.xml")
//book[price>40]//(author|title)/text()
```



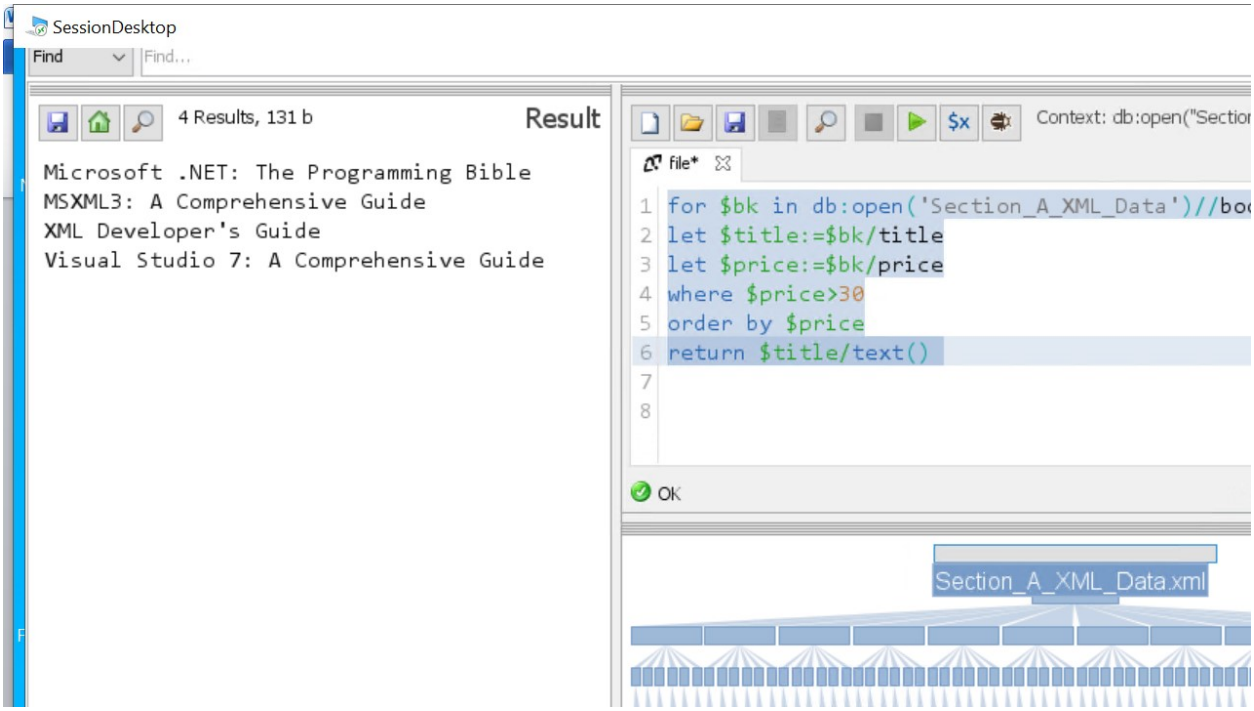
5. [4 Pts.] Write an XPath expression to find the authors and prices of the books belonging to Computer genre.

```
db:open("Section_A_XML_Data","Section_A_XML_Data.xml")
//book[genre="Computer"]//(author|price)/text()
```



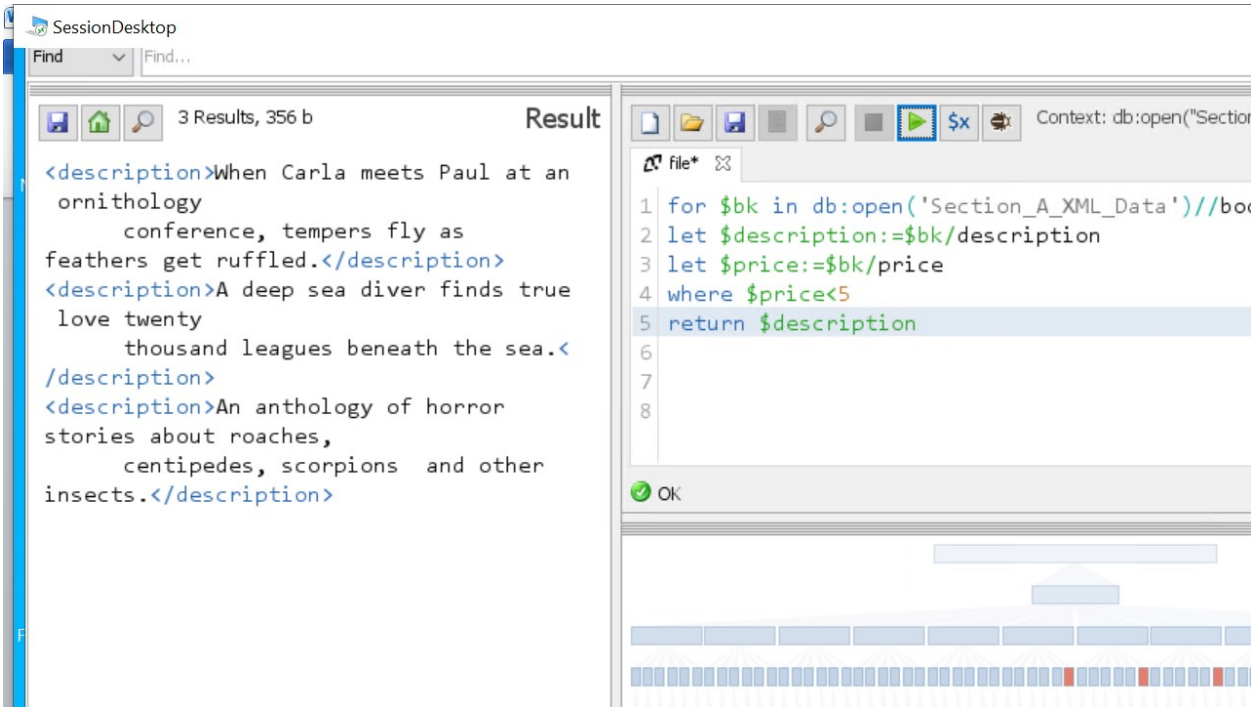
6. [6 Pts.] Write an XQuery (FLWOR) script to find the titles of the books arranged in ascending order of price, of which the price are more than 30 dollars.

```
for $bk in db:open('Section_A_XML_Data')//book
let $title:=$bk/title
let $price:=$bk/price
where $price>30
order by $price
return $title/text()
```



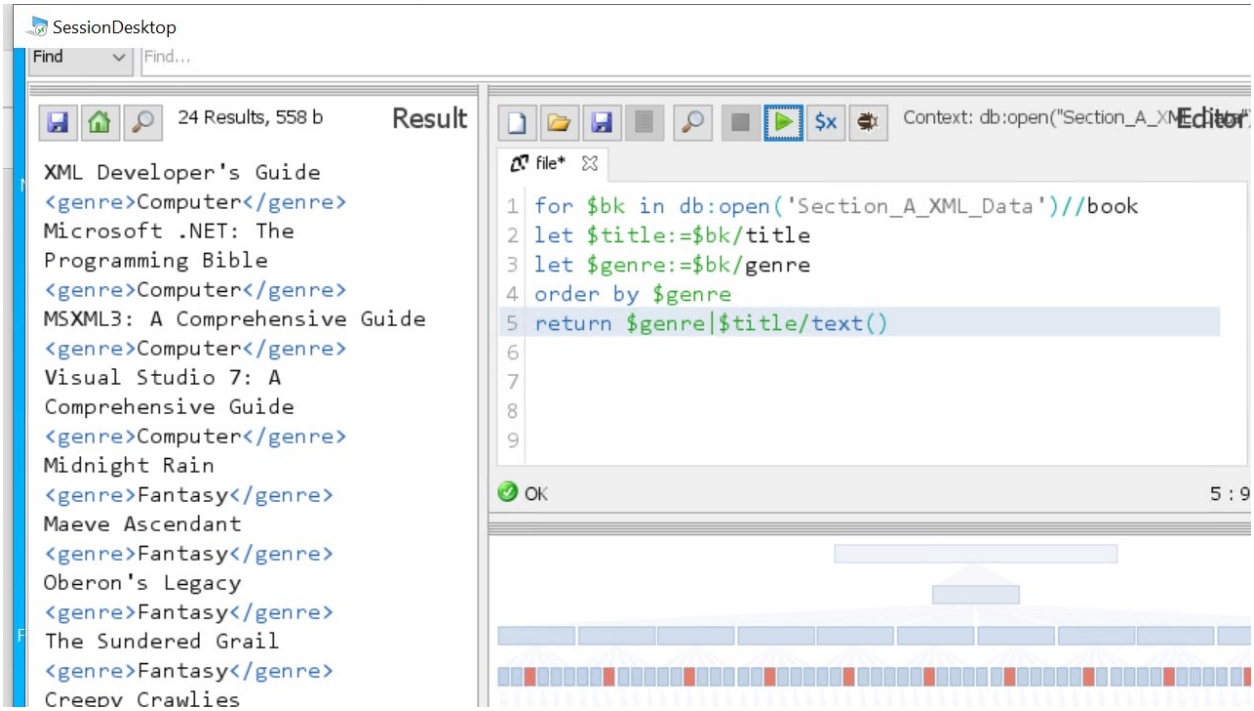
7. [6 Pts.] Write an XQuery (FLWOR) script to provide only the descriptions of the books, which cost less than 5 dollars.

```
for $bk in db:open('Section_A_XML_Data')//book
let $description:=$bk/description
let $price:=$bk/price
where $price<5
return $description
```



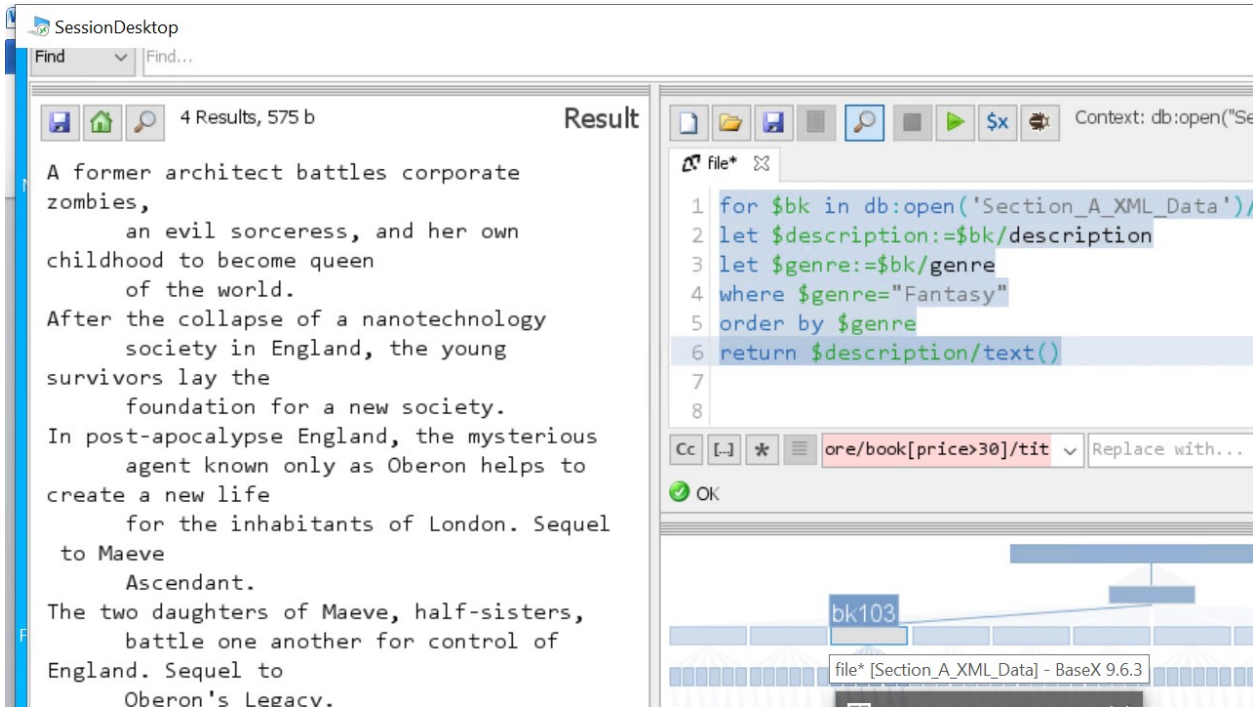
8. 8. [6 Pts.] Write an XQuery (FLWOR) script which gives the various genre along with the text of the title of the books in them.

```
for $bk in db:open('Section_A_XML_Data')//book
let $title:=$bk/title
let $genre:=$bk/genre
order by $genre
return $genre|$title/text()
```



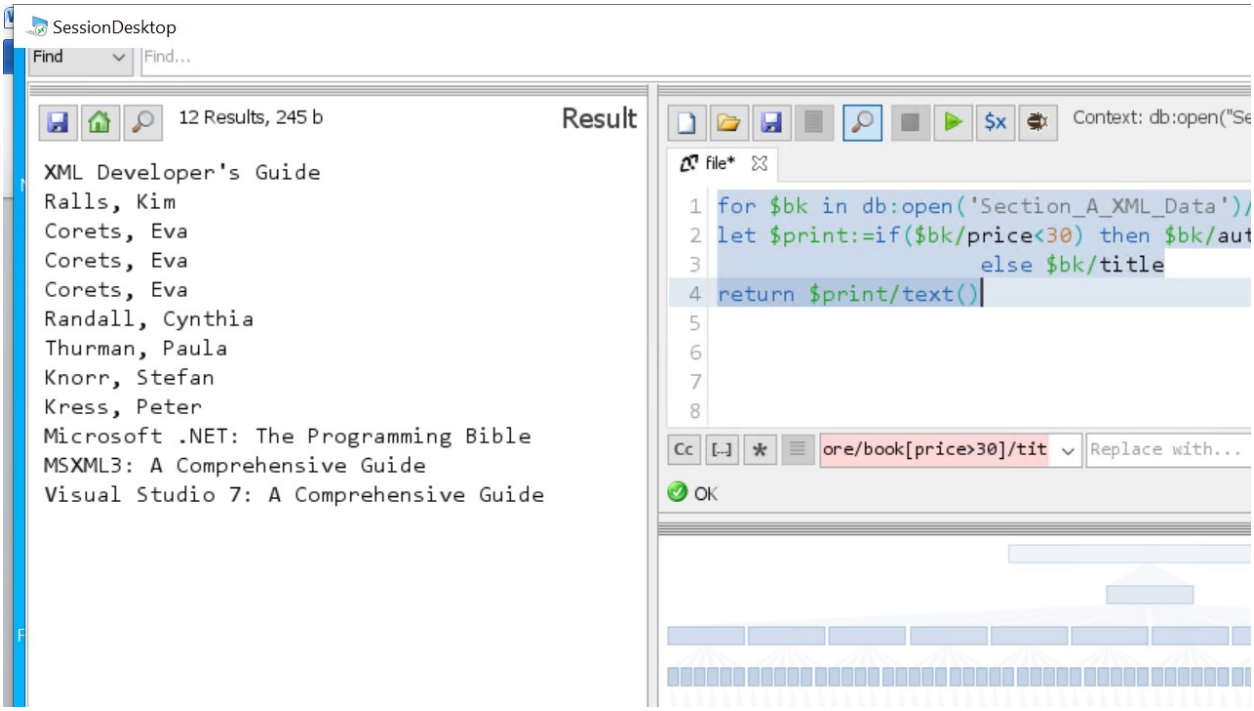
9. [6 Pts.] Write an XQuery (FLWOR) script which gives the description text showing that the book belongs to Fantasy genre.

```
for $bk in db:open('Section_A_XML_Data')//book
let $description:=$bk/description
let $genre:=$bk/genre
where $genre="Fantasy"
order by $genre
return $description/text()
```



10. [6 Pts.] Write an XQuery (FLWOR) script which gives the list of authors whose books cost less that 30 dollars and provides the titles of the books otherwise.

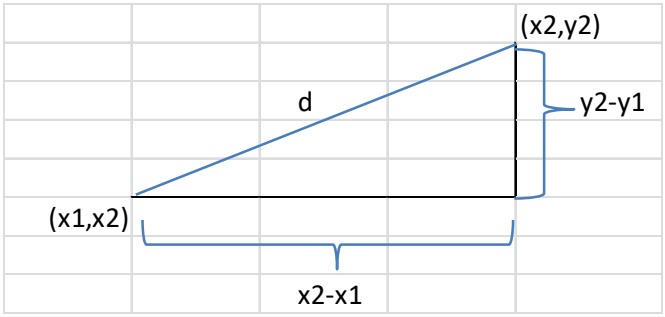
```
for $bk in db:open('Section_A_XML_Data')//book
let $print:=if($bk/price<30) then $bk/author
            else $bk/title
return $print/text()
```



Section B

1. [35 Pts.] Use the K-means algorithm to cluster this dataset. You can initiatiate the claculation by assuming K=2 and assume that the records with RIDs 103, and 104 are used as the initial cluster centroids.

Euclidean distance d=((x2-x1)^2 + (y2-y1)^2)^.5



1st iteration

K=2, c1 = RID 103, c2 = RID 104

RID	x_i		c_1		c_2		Distance 1 (|xi-c1|)	Distance 2 (|xi-c2|)	Nearest Cluster	New Centroid		
	age	yrs	age	yrs	age	yrs					age	yrs
101	30	5	50	15	25	5	22.4	5.0	c2	c1	51.7	21.7
102	50	25	50	15	25	5	10.0	32.0	c1	c2	28.3	6.7
103	50	15	50	15	25	5	0.0	26.9	c1			
104	25	5	50	15	25	5	26.9	0.0	c2			
105	30	10	50	15	25	5	20.6	7.1	c2			
106	55	25	50	15	25	5	11.2	36.1	c1			
							lowest distance					

c1 = mean of RID 102,103,106

(50+50+55)/3 = 51.7

New centroids c1 = (51.7,21.7) , c2=(28.3,6.7)

2nd iteration

c1 = (51.7,21.7) , c2=(28.3,6.7)

RID	x_i		c_1		c_2		Distance 1 (|xi-c1|)	Distance 2 (|xi-c2|)	Nearest Cluster	New Centroid		
	age	yrs	age	yrs	age	yrs					age	yrs
101	30	5	51.7	21.7	28.3	6.7	27.3	1.7	c2	c1	51.7	21.7
102	50	25	51.7	21.7	28.3	6.7	3.7	29.5	c1	c2	28.3	6.7
103	50	15	51.7	21.7	28.3	6.7	6.9	23.9	c1			
104	25	5	51.7	21.7	28.3	6.7	31.4	3.3	c2			
105	30	10	51.7	21.7	28.3	6.7	24.6	5.3	c2			
106	55	25	51.7	21.7	28.3	6.7	4.7	33.3	c1			

New centroids c1 = (51.7,21.7) , c2=(28.3,6.7)

2. [15 Pts.] Provide a brief description on the difference between describing discovered knowledge using clustering and describing it using classification?

There is no difference in distance using iteration 1 and 2. We identified 2 groups using clustering. So the 2 groups will be RID 102,103,106 and RID 101,104,105.

In Classification method, we classify sample in to known groups. We may use our knowledge about data in to groups . In this example we can classify using age range. Age between 25 and 49 , age between 50 and 55.

In classification we use our knowledge to put sample in to classes. In clustering we do it randomly without knowledge.