COMP3381 Software, Systems and Applications III

Advanced Databases

XML and XQuery

Lecturer: Prof. Alexandra I. Cristea
Assessment: summative
Hand-in method: DUO
Document format: PDF

Consider the following two XML documents, *books.xml* and *media.xml*.

books.xml

```
<?XML VERSION = "1.0" STANDALONE = "yes"?>
<books>
	<book publicationDate = "1934">
		<title> Mary Poppins </title>
		<author> P. L. Travers </author>
		<illustrator> Mary Shepard </illustrator>
	</book>
	<book publicationDate = "1935">
		<title> Mary Poppins Comes Back </title>
		<author> P. L. Travers </author>
		<illustrator> Mary Shepard </illustrator>
	</book>
	<book publicationDate = "1943">
		<title> Mary Poppins Opens the Door </title>
		<author> P. L. Travers </author>
		<illustrator> Mary Shepard </illustrator>
	</book>
	<book publicationDate = "1952">
		<title> Mary Poppins in the Park </title>
		<author> P. L. Travers </author>
		<illustrator> Mary Shepard </illustrator>
	</book>
	<book publicationDate = "1962">
		<title> Mary Poppins From A to Z </title>
		<author> P. L. Travers </author>
		<illustrator> Mary Shepard </illustrator>
	</book>
	<book publicationDate = "1975">
		<title> Mary Poppins in the Kitchen </title>
		<author> P. L. Travers </author>
		<illustrator> Mary Shepard </illustrator>
	</book>
	<book publicationDate = "1982">
		<title> Mary Poppins in Cherry Tree Lane </title>
		<author> P. L. Travers </author>
		<illustrator> Mary Shepard </illustrator>
	</book>
	<book publicationDate = "1988">
		<title> Mary Poppins and the House Next Door </title>
		<author> P. L. Travers </author>
		<illustrator> Mary Shepard </illustrator>
	</book>
</books>
```

media.xml

```
<?XML VERSION = "1.0" STANDALONE = "yes"?>
<media>
	<movie year="1949" type="TV series" language="En">
		<title> Mary Poppins </title>
		<actress> Mary Wickes </actress>
		<actor> David Opatoshu </actor>
		<actor> Tommy Rettig </actor>
		<producer> CBS Television </producer>
	</movie>
	<movie year="1964" type="film" language="En">
		<title> Mary Poppins </title>
		<actress> Julie Andrews </actress>
		<actor> Dick Van Dyke </actor>
		<actor> David Tomlison </actor>
		<actress> Glynis Johns </actress>
		<producer> Walt Disney </producer>
		<director> Robert Stevenson </director>
		<academyNominations> 13 </academyNominations>
	</movie>
	<movie year="1983" type="film" language="Ru">
		<title> Mary Poppins, Goodbye </title>
		<actress> Natalya Andreychenko </actress>
		<actor> Albert Filozov </actor>
		<actor> Oleg Tabakov </actor>
	</movie>
	<theater year="2004" type="musical" language="En">
		<title> Mary Poppins (musical) </title>
		<producer> Cameron Mackintosh </producer>
	</theater>
	<broadcast year="2010" type="BBC Radio" language="En">
		<title> Mary Poppins </title>
		<actress> Juliet Stevenson </actress>
		<producer> BBC Radio 4 Extra </producer>
	</broadcast>
	<movie year="2013" type="film" language="En">
		<title> Saving Mr. Banks </title>
		<actress> Emma Thompson </actress>
		<actor> Tom Hanks </actor>
		<academyNominations> 1 </academyNominations>
	</movie>
	<movie year="2018" type="film" language="En">
		<title> Mary Poppins Returns </title>
		<actress> Emily Blunt </actress>
		<actor> Lin-Manuel Miranda </actor>
		<actress> Emily Mortimer </actress>
		<actor> Ben Whishaw </actor>
	</movie>

</media>
```

The assignment is divided into two parts (Part A, Part B), as it is shown below. The contribution of each of these two parts is shown below; details for marking are in the Marking Scheme below.

Part A. (35%)
Draw the directed tree structure of the *books.xml* XML file (i.e. the Hierarchical Tree Model). Please ensure that you include in your tree all the necessary labels.

Hint: If you have not enough space, you can write the names of the nodes *next* to the nodes (instead of *inside* the nodes), and use a star-like display.

Part B. (65%)
Assume that the above XML code is stored in their respective files, *books.xml* and *media.xml*. Based on these XML files, write the following queries as required below.

Write an **XPath** query for each of the following:

a. *Find all Mary Poppins books which were published before WWII.*
b. *Find all movies that have academy nominations*.

Write an **XQuery** query for each of the following:

c. *How many books have there been published about Mary Poppins?*
d. *List the books that have been published before the second Mary Poppins movie appeared.*
e. *How much later has the most recent Mary Poppins movie appeared, compared to the second book about Mary Poppins?*

Additionally, for each query a.-e. above you must:
(i) Explain how you interpreted the (natural language) query, and thus why your answer is appropriate.
(ii) Explain also limitations of your approach, if there are any.
(iii) Please provide also the output of these queries, when applying them to the two XML files above.

Furthermore, for each query c.-e. above, additionally to (i)-(iii) above, you must:
(iv) Comment if the XQuery query could be answered with an XPath query, and give the XPath query if that is the case. Otherwise, explain why the query can only be written as an XQuery query.

Hints: For maximum marks in Part B, consider writing the query in such a way, that if the databases would be extended, you would still get the correct answer; consider this for (ii) above. Also consider formatting your output, where possible. Check that your interpretation for (i) is not an over-simplification.

Marking Scheme

Part	Question No.	Detail	Marks
A		All elements represented	10
		All attributes represented	10
		All values represented	10
		Whole structure correct	5
B	a.	Query	5
		(i) Interpretation	3
		(ii) Limitations	3
		(iii) Output	2
	b.	Query	5
		(i) Interpretation	3
		(ii) Limitations	3
		(iii) Output	2
	c.	Query	5
		(i) Interpretation	2
		(ii) Limitations	1
		(iii) Output	1
		(iv) XPath+Explanation	4
	d.	Query	5
		(i) Interpretation	2
		(ii) Limitations	1
		(iii) Output	1
		(iv) XPath+Explanation	4
	e.	Query	5
		(i) Interpretation	2
		(ii) Limitations	1
		(iii) Output	1
		(iv) XPath+Explanation	4