

# 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*.

### **students.xml**

```
<?XML VERSION = "1.0" STANDALONE = "yes"?>
<students year="2019-2020">
  <student enrolmentDate = "2017">
    <name>John Smith</name>
    <course>
      <name>Software, Systems and Applications III</name>
      <submodule>Advanced Databases</submodule>
      <submodule>Web Technology</submodule>
    </course>
    <course>Computing Methodologies III</course>
  </student>

  <student enrolmentDate = "2016">
    <firstname>Anna</firstname>
    <lastname>Brown</lastname>
    <course>
      <name>Software, Systems and Applications IV</name>
      <submodule>Semantic Web</submodule>
    </course>
    <course>Computing Methodologies IV</course>
  </student>
</students>
```

### **teachers.xml**

```
<?XML VERSION = "1.0" STANDALONE = "yes"?>
<teachers>
  <teacher joiningDate = "2018" jobRole= "Professor">
    <name>Alexandra Cristea</name>
    <course>
      <name>Software, Systems and Applications III</name>
      <submodule>
        <name>Advanced Databases</name>
        <year>2018-2019</year>
        <year>2019-2020</year>
      </submodule>
    </course>
    <course>
      <name>Software, Systems and Applications IV</name>
      <submodule>
        <name>Semantic Web</name>
        <year>2018-2019</year>
        <year>2019-2020</year>
      </submodule>
    </course>
  </teacher>
</teachers>
```

The assignment is divided into two parts (Part A, Part B), as it is shown below. Details for marking are in the **Marking Scheme** below.

### Part A.

- a. Draw the directed tree structure of the *students.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.
- b. Write a DTD for the *teachers.xml* XML file above. Use a validator to ensure the correctness of your DTD and explain which one you used. *Hint:* You may need to search the web for some parts of the answer.
- c. Is it possible to write a DTD for the *students.xml* XML file above? Explain which construct would be problematic. How could this be done in XML for correct validation? *Hint:* You may need to search the web for some parts of the answer.

### Part B.

Assume that the above XML code is stored in their respective files, *students.xml* and *teachers.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 students who study 'Advanced Databases' this year.*
- b. *Find all teachers who teach 'Advanced Databases' this year.*

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

- c. *How many years has Professor Cristea been teaching 'Advanced Databases' (at Durham)?*
- d. *Find all students in year 3 currently taught by Alexandra.*
- e. *How many teachers and how many students are kept in the databases where the last name is not known?*

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	a.	All elements represented	5
		All attributes represented	5
		All values represented	5
		Whole structure correct	5
	b.	All elements represented	5
		All attributes represented	5
		Whole structure correct	5
	c.	Explanation	2
B	a.	Query	5
		(i) Interpretation	2
		(ii) Limitations	3
		(iii) Output	2
	b.	Query	5
		(i) Interpretation	2
		(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