

Klausur “Semistructured Data and XML”
Summer Term 2025
Prof. Dr. Wolfgang May
6. August 2025, 13:30–16:30
Working Time: 120 Minutes
(carried out as a computer-based ILIAS exam)

Vorname:

Nachname:

Matrikelnummer:

Setting: The usage of saxon (with the aliases saxonValid, saxonXQ, and saxonXSL defined as in the course) and xmllint (validation error messages are better with xmllint) is recommended. Web access (e.g. for XPath/XQuery and XSLT documentation) is allowed. It was also recommended to have the slides, and a condensed self-prepared “cheat sheet” (preparation of a cheat sheet is a very effective way to work through the materials). Help by other persons or AI tools is NOT allowed.

Answers might be given in English or German (most answers are program code anyway).

In the text, the german translation is sometimes given in parentheses.

Give *all* answers via the ILIAS system.

Like in a “paper exam”, also solutions that do maybe not work (or do not work completely) can be delivered and will be graded with appropriate partial points.

For **passing** the exam, **50** points are sufficient.

	Max. Punkte	Erreichte Punkte
Aufgabe 1 (XML)	15	
Aufgabe 2 (DTD)	12	
Aufgabe 3 (XPath (a))	3	
Aufgabe 4 (XPath/XQuery (b))	4	
Aufgabe 5 (XQuery (c))	4	
Aufgabe 6 (XPath/XQuery (d))	6	
Aufgabe 7 (XQuery (e))	8	
Aufgabe 8 (XPath/XQuery (f))	6	
Aufgabe 9 (XPath/XQuery (g))	6	
Aufgabe 10 (XQuery Update)	6	
Aufgabe 11 (XSLT)	14	
Aufgabe 12 (XML Consistency)	8	
Aufgabe 13 (XQuery Update (abstract))	8	
Summe	100	

Note:

Project: Season Schedule for a Theater

The scenario is to manage the schedule for a theater for the upcoming season 2025/26 - the pieces, the performances for every day of the season, the stage actors and which roles they play.

1. The rooms where performances take place are the “Big Hall”, the “Small Hall”, the café and the foyer.

The Big Hall has 800 seats, and the Small Hall has 300 seats. In the café there are 100 seats, and in the foyer 200.

2. During the season, the theater performs a number of pieces, usually several times each. For every such piece, the title, the type (tragedy, comedy, opera, ...), the name of the composer(s) and writer(s), and the duration of the local staging is stored.

- The tragedy “Hamlet” by *William Shakespeare* (that was used in the lecture and exercises, see below) which has a duration of 3 hours,
- the opera “Don Carlos” by the composer *Giuseppe Verdi* (based on the play from 1787 by *Friedrich Schiller*) which has a duration of 4 hours,
- the tale “The Little Prince” by *Antoine de St. Exupéry* as a performance for children with a duration of 80 minutes,
- the opera “Orpheus in the Underworld” by *Hector Crémieux* and *Ludovic Halévy* (writers) and *Jacques Offenbach* (music) which has a duration of 90 minutes.

3. The text and stage directions of *some* plays is available in separate XML sources following the structure of the “Hamlet” play as known from the lecture, which is available at <https://www.stud.informatik.uni-goettingen.de/xml-lecture/Shakespeare/hamlet.xml>. (if you have it locally, you can also use `doc('hamlet.xml')` to access it.)

Design your main XML file in such a way that it allows to state queries that access the XML of the plays, in case that it is available.

4. Also the characters acting in the piece are listed, with their name (hint: care for compatibility with the above play XML source(s)), their gender (if applicable) and optionally a short comment:

- In “Hamlet”, the main characters are Claudius (King of Denmark), Hamlet (nephew of Claudius), Gertrude (Queen of Denmark, mother to Hamlet); the sample data also contains two minor characters: Reynaldo (servant to Polonius) and Lucianus (another nephew to Claudius).
- In “Don Carlos”, there are Philipp II (King of Spain), Elisabeth (Queen of Spain), Don Carlos (Philipp’s son and crown prince), Princess of Eboli (a Spanish aristocrat), and some more.

- In “the Little Prince”, the main characters are the pilot, the little prince, and the fox. The fox does not have a gender.
 - In “Orpheus in the Underworld”, the main characters are Orpheus (a musician) and Eurydice (Orpheus’ wife).
5. There are the stage actors working at the theater. For every actor, the name, the gender, and the type of voice (soprano, alto, tenor, bass) is stored.
- Alice is female and has a soprano voice.
 - Bob is male and has a tenor voice.
 - Carol is female and has an alto voice.
 - Dave is male and has a bass voice.
 - Eva is female and has also an alto voice.
 - Frank is male and has also a tenor voice.
 - Grace is female and has also a soprano voice.
 - Henry is male and has also a bass voice.
 - Michael and Nicolas are male, both are tenors.
 - and many more
6. The overall assignment for the whole season of stage actors to the roles is decided early (the actors need some time to learn their texts and to prepare) and stored. For each role, two persons are assigned:
- “Hamlet”: Bob and Frank are assigned to prepare for the Hamlet role, Dave and Henry are assigned to the role of King Claudius; Carol and Eva are assigned to Queen Gertrude, Michael and Nicolas are each assigned to both (minor characters) Reynaldo and Lucianus.
 - “Don Carlos”: Michael and Bob are assigned to Don Carlos; Frank and Nicolas are assigned to Philipp, Grace and Eva are assigned to Princess Eboli.
 - “The Little Prince”: Eva and Grace are assigned to the Little Prince role, Dave and Michael are assigned to the Pilot role, and Nicolas and Henry are assigned to the Fox role.
 - “Orpheus in the Underworld”: Henry and Dave are assigned to the Orpheus role, Grace and Alice are assigned to the Eurydice role.
7. The theater schedule is fixed before the beginning of the season and lists for all performances the date, time, and the hall where the performance takes place. (Note that not every day there are performances; usually Friday to Sunday and some performances on weekdays).
During the season, also the current number of the tickets sold for every performance is stored.
- Sample data:
- The season starts on Friday, 3.10.2025, 20:00 with “Hamlet” in the *Big Hall*. Currently, there are already 543 tickets sold.

- On Saturday, 4.10.2025 there is at 17:00 “The Little Prince” in the Big Hall for children. All 800 tickets are already booked by schools. A 20:30 “Orpheus in the Underworld” in the Big Hall. 400 tickets are sold.
 - On Sunday, 5.10.2025, there is at 20:00 “Don Carlos” in the Big Hall; 666 tickets are yet sold.
 - On Saturday, 11.10.2025, there is at 17:00 again a performance of “The Little Prince”, and at 20:00 again a performance of “Hamlet”, both in the Big Hall. 300 resp. 250 tickets are yet sold.
8. For every performance, the stage actors are assigned to the roles some time in advance and are stored in the XML:
- The assignment for the “Hamlet” performance on 3.10.2025 is as follows: Hamlet: Bob; Claudius: Dave; Gertrude: Carol; and Reynaldo and Lucianus are both played (they are on stage in different acts) by Michael.
 - The assignment for the “Little Prince” performance on 4.10.2025 is: Little Prince: Eva; Pilot: Dave; Fox: Michael.
 - The assignment for the “Orpheus in the Underworld” performance on 4.10.2025 is: Orpheus: Dave; Eurydice: Alice, ...
 - The assignment for “Don Carlos” on 5.10.2025 is Don Carlos: Michael; Philipp: Frank; Princess Eboli: Eva.
 - The assignment for the “Little Prince” performance on 11.10.2025 is again Little Prince: Eva; Pilot: Dave; Fox: Michael.
 - The assignment for “Hamlet” on 11.10.2025 will be done in Exercise 10.

Exercise 1 (XML [15 Points])

Design an XML structure for the above information. Use the frame given in the file `exam.xml` - it must contain significant enough data (at least 2 instances per element/attribute type).

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE put-name-here SYSTEM "exam.dtd">
```

Exercise 2 (DTD [12 Points])

Give the DTD for your document developed in Exercise 1, use the file `exam.dtd`.
Use one of the calls

```
xmllint -loadtdt -valid --noblanks -noout exam.xml
saxonValid.bat -s:exam.xml
```

for validating it (note that xmllint provides better error messages).

Upload or copy-and-paste the DTD from the file `exam.dtd` afterwards here:

Exercise 3 (XPath (a) [3 Points])

Use your `exam.xml` XML file as a basis for solving this and the following exercises.
None of the results should contain duplicates.

Give an XPath query that returns the *names* of those characters where at least one actor is staged for who has a bass voice.

Write the query string in the file `query1.xq` and call it with

```
saxonXQ.bat -s:exam.xml query1.xq
```

Copy-and-paste the query from `query1.xq` afterwards here:

Exercise 4 (XPath/XQuery (b) [4 Points])

Give an XPath query or an XQuery query that returns the *names* of those actors who are not staged in any performance on the weekend from 3.10.2025 until 5.10.2025.

Write the query string in the file `query2.xq` and call it with

```
saxonXQ.bat -s:exam.xml query2.xq
```

Copy-and-paste the query from `query2.xq` afterwards here:

Exercise 5 (XQuery (c) [4 Points])

Give an XQuery query that for each case, where in some performance an actor plays a role of a character that has a different gender, returns the date of the performance, the title of the piece, the name of the actor, and the role in some reasonable XML markup.

Copy-and-paste the query from `query3.xq` afterwards here:

Exercise 6 (XPath/XQuery (d) [6 Points])

Give an XQuery query that outputs for each piece the total number of tickets sold for it during this season, the number of performances that have been given, and the number of performances where all tickets have been sold.

Copy-and-paste the query from `query4.xq` afterwards here:

Exercise 7 (XQuery (e) [8 Points])

Give an XQuery query that checks that no collisions in the casting of a performance occur: a stage actor must not be casted for two roles that occur on the stage (here: have speeches) in the same scene (other multi-occurrences are allowed - the actor must then change his costume in-between). Concretely, output all actors and pairs of roles they are casted for where there would be a co-occurrence in some scene in the form

```
<conflict piece="..." date="..." actor="..." role1="..." role2="..."
act="..." scene="..." /> .
```

Obviously (and hopefully), the result is empty for the sample data. Describe how you can validate whether your query would correctly detect such cases (1P).

Copy-and-paste the query from query5.xq afterwards here:

Exercise 8 (XPath/XQuery (f) [6 Points])

Give an XQuery query that returns the names of those stage actors that are casted for every day in October where the theater shows some performance.

Copy-and-paste the query from query6.xq afterwards here:

Exercise 9 (XPath/XQuery (g) [6 Points])

Give an XQuery query that does the following: for a given date (e.g., the 2025-10-04, where two performances are given in the sample data) outputs an HTML printout that lists the performances of that day with the actual staging of the characters for this day. So the visitors can look up who are the actors that they see in the performance.

Copy-and-paste the query from query7.xq afterwards here:

Exercise 10 (XQuery Update [6 Points])

For the “Hamlet” performance on 11.10.2025, the staging of each role should be the actor who is listed to be prepared for this role (cf. item 6 in the project description), but did *not* play it in the “Hamlet” performance on 3.10.2025.

Give an XQuery Update statement that adds the appropriate XML contents to your file. (not necessary to execute, recall that updates can only be executed if you have the Saxon Enterprise Edition installed)

Exercise 11 (XSLT [14 Points])

Extend the given XSLT stylesheet frame `exam.xsl` to an XSLT stylesheet that generates HTML cheatsheets for the actors: Given a play (it should be usable for the general case - although we have the play text only for Hamlet), it should do the following for each character:

For each speech of the respective character, output (i) who is the last speaker before, and the last line of his speech, and (ii) then, the speech. If there is no other speech directly before, e.g., at the beginning of a scene, output “at begin of” and the SCENE’s title instead.

Use the following call to execute it:

```
saxonXSL.bat -s:exam.xml exam.xsl play=Hamlet
```

Copy-and-paste the XSLT stylesheet from exam.xsl afterwards here:

Exercise 12 (XML Consistency [8 Points])

Your XML designed in Exercise 1 probably does not guarantee that in the staging only actors can be casted for a role where they are also assigned to prepare for at the beginning of the season (see item 6 in the project description). (If your XML guarantees this already, you have already successfully solved this exercise).

Sketch (text, with XML and DTD fragments) how this could be guaranteed by a refined XML+DTD design.

Exercise 13 (XQuery Update (abstract) [8 Points])

Consider an XQuery Update statement with a rather simple for-clause like

```
for $e in some-xpath-expression
return (
  insert node <bla/> after $e/xxx,
  delete node $e/@yyy
)
```

where `some-xpath-expression` is formed only with `/`, `//`, and simple filters of the form `[@attr=value]`, and where every such *e*-element has at most one *xxx* subelement. Give an XSLT stylesheet pattern that executes the same updates on the XML tree.

The following frames can be used:

- Frame for XML file `exam.xml`:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE put-name-here SYSTEM "exam.dtd">
  to be extended here
```

- Frame for XML stylesheet `exam.xsl`:

```
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  version="2.0">
  to be extended here
</xsl:stylesheet>
```