

Klausur “Semistructured Data and XML”
Summer Term 2024
Prof. Dr. Wolfgang May
25. Juli 2024, 10:30–12:30
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).

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 (XML match ticker information)	15	
Aufgabe 4 (XPath (a))	3	
Aufgabe 5 (XPath (b))	4	
Aufgabe 6 (XPath/XQuery (c))	4	
Aufgabe 7 (XPath/XQuery (d))	6	
Aufgabe 8 (XPath/XQuery (e))	6	
Aufgabe 9 (XPath/XQuery (f))	6	
Aufgabe 10 (XQuery Update)	8	
Aufgabe 11 (XSLT)	16	
Aufgabe 12 (XML Markup Annotation Design)	5	
Aufgabe 13 (XPath (2b))	0	
Aufgabe 14 (XPath (3b))	0	
Aufgabe 15 (XPath/XQuery (4b))	0	
Aufgabe 16 (XPath/XQuery (4c))	0	
Aufgabe 17 (XPath/XQuery (4d))	0	
Summe	100	

Project: A Database for the European Football Championship (or a similar team sports event)

The scenario is about the 2024 European Football Championship: a database for a sports magazine like (in German) <https://www.kicker.de/europameisterschaft/tabelle>.

The database is intended to be maintained incrementally, from the early beginning, where only the participating countries are known, then updated when the first games are fixed, updated during the event (e.g. what are the matches in the next round), and (see further below) during the matches to provide data for a liveticker like <https://www.kicker.de/spanien-gegen-deutschland-2024-europameisterschaft-4772140/ticker>.

The assumed “snapshot” timepoint in the exam is the 119th minute (a regular match has 2x45 minutes, and in case of a draw, the match is extended by 2x15 minutes) of the first quarterfinal match, played between Spain and Germany on 5.7.2024 (to have an example for an update).

1. Several national teams, each from a different country, participate in the championship. Every team consists of several players, and has a coach. For every person, the name and the birthdate is stored. For every player, also the club team where he has a contract, and the country of this club are stored.

To reduce typing, the players elements are given in the exam.xml file (extend them as needed). The file can be found in StudIP under the “files” tab.

2. The coach of the team from *Germany* is *Julian Nagelsmann*, born 23.7.1987. The coach of the team from *Spain* is *Luis de la Fuente*, born 21.6.1961.
3. Furthermore, there are teams from Scotland, England (so, not only true countries), Hungary, Switzerland, Denmark and other countries.
4. In every match, two teams participate. The type of the match (group match, ..., quarterfinal, ... final), the date, the name of the city where it will take/takes/took place, and the result (care how you store this to make it queryable) are stored.
5. Sample match data (incomplete):
The german group matches: Germany vs. Scotland, 14.6.2024 in Munich, result 5:1.
Germany vs. Hungary, 19.6.2024 in Stuttgart, result 2:0.
Switzerland vs. Germany, 23.6.2024 in Frankfurt, result 1:1.
Round of 16: Germany vs. Denmark, 29.6.2023 in Dortmund, result 2:0.
Quarterfinal Germany vs. Spain, 5.7.2024 in Stuttgart, standing 1:1 at the “snapshot moment”. For the other quarterfinals, the teams are already known: Portugal vs. France, 5.7.2024 in Hamburg, ...
For the semifinals and the final, the date and place are already fixed, but obviously not, which teams will play them. The final takes place on 14.7.2024 in Berlin.
6. For every team participating in a match, it is stored which players were in the beginning formation.

7. Sample beginning formation data about the German team:
Neuer, Kimmich, Rüdiger, Kroos, Gündogan, Musiala and Havertz were in the beginning formation in every match of the German team.
Wirtz was also in the beginning formation in the (first three) matches against Scotland, Hungary, and Switzerland.
Sane was in the beginning formation in the matches against Denmark and in the “ongoing” match against Spain.

In the latter, Williams, Lopez and Yamal (and some others) were in the beginning formation for Spain.

8. The database should also contain the detailed ticker events during the matches as e.g. shown in <https://www.kicker.de/spanien-gegen-deutschland-2024-europameisterschaft-4772> ticker. Text examples for events are also given in the exam.xml file. These will be further detailed in Exercise 3.

Exercise 1 (XML [15 Points])

Design an XML structure for the above information. Use the frame given in the file exam.xml, which already contains the players' elements (to be completed to be useful! and to be moved around!) and the fulltext events that will be refined later.

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

to be extended here

### person elements to complete (if needed) and to use

<coach name="Julian Nagelsmann" bdate="1987-07-23"/>
<coach name="Luis de la Fuente" bdate="1961-06-21"/>

Sample Players of the German team:
<player name="Florian Wirtz" bdate="2003-05-04" club="Leverkusen" clubcountry="D"/>
<player name="Kai Havertz" bdate="1999-06-11" club="Arsenal" clubcountry="ENG"/>
<player name="Jamal Musiala" bdate="2003-02-26" club="Bayern" clubcountry="D"/>
<player name="Ilkay Gündogan" bdate="1990-10-24" club="Barcelona" clubcountry="E"/>
<player name="Antonio Rüdiger" bdate="1993-03-03" club="Real Madrid" clubcountry="E"/>
<player name="Leroy Sane" bdate="1996-01-11" club="Bayern" clubcountry="D"/>
<player name="Niclas Füllkrug" bdate="1993-02-09" club="Dortmund" clubcountry="D"/>
<player name="Emre Can" bdate="1994-01-12" club="Dortmund" clubcountry="D"/>
<player name="Toni Kroos" bdate="1990-01-04" club="Real Madrid" clubcountry="E"/>
<player name="Joshua Kimmich" bdate="1995-02-08" club="Bayern" clubcountry="D"/>
<player name="Thomas Müller" bdate="1989-09-13)" club="Bayern" clubcountry="D"/>
<player name="Manuel Neuer" bdate="1986-03-27)" club="Bayern" clubcountry="D"/>
<player name="Marc Andre ter Stegen" bdate="1992-04-30" club="Barcelona"
    clubcountry="E"/>
<player name="Deniz Undav" bdate="1996-07-19" club="Stuttgart" clubcountry="D"/>
<player name="Chris Führich" bdate="1998-01-09" club="Stuttgart" clubcountry="D"/>

Sample Players of the Spanish team:
<player name="Mikel Merino" bdate="1996-06-22" club="San Sebastian" clubcountry="E"/>
<player name="Mikel Oyarzabal" bdate="1997-04-21" club="San Sebastian"
    clubcountry="E"/>
<player name="Dani Olmo" bdate="1998-05-07" club="Leipzig" clubcountry="D"/>
<player name="Pedro Lopez" bdate="2002-11-25" club="Barcelona" clubcountry="E"/>
<player name="Nico Williams" bdate="2002-12-07" club="Bilbao" clubcountry="E"/>
<player name="Lamine Yamal" bdate="2007-07-13" club="Barcelona" clubcountry="E"/>
<player name="Marc Cucurella" bdate="1998-07-22" club="Chelsea" clubcountry="ENG"/>

Sample Player of the Scottish team:
<player name="Ryan Porteous" bdate="1999-03-25" club="Watford" clubcountry="ENG"/>

Sample Player of the Swiss team:
<player name="Dan Ndoye" bdate="2000-10-15" club="Bologna" clubcountry="I"/>

Sample Player of the Danish team:
<player name="Joachim Andersen" bdate="1996-05-31" club="Crystal Palace"
    clubcountry="ENG"/>
```

Sample fulltext events of the match Germany vs. Scotland:

<event> minute 10: Kroos passes to the right to Kimmich, pass to Wirtz, who scores the 1:0 for Germany.</event>
<event> minute 19: Goal by Musiala for Germany. The pass came from Havertz.</event>
<event> minute 42: Foul by Porteous in the penalty area against Gündogan.</event>
<event> minute 44: Decision to have a penalty for Germany. And red card against Porteous.</event>
<event> minute 44: Havertz executes the penalty and scores the 3:0.</event>
<event> minute 63: Havertz is substituted by Füllkrug.</event>
<event> minute 63: Wirtz is substituted by Sane.</event>
<event> minute 68: Pass by Musiala into the middle, Gündogan misses it, goal by Füllkrug for Germany.</event>
<event> minute 74: Musiala is substituted by Müller.</event>
<event> minute 76: Goal by Müller, but it was offside (so it does not count).</event>
<event> minute 80: Kroos is substituted by Can.</event>
<event> minute 87: Goal for Scotland: Own goal by Rüdiger who in a defending action pushes the ball with his head into the German goal to the 4:1</event>
<event> minute 90: Can stands free in the middle, pass from Müller, goal by Can for Germany.</event>
<event> minute 90: Final whistle. Germany wins 5:1 against Scotland.</event>

Sample events of the match Germany vs. Hungary:

<event> minute 22: Muddle in the Hungarian penalty zone, goal by Musiala for Germany.</event>
<event> minute 27: yellow card for Rüdiger for complaining too much.</event>
<event> minute 58: Havertz is substituted by Füllkrug.</event>
<event> minute 58: Wirtz is substituted by Sane.</event>
<event> minute 67: Goal by Gündogan for Germany.</event>
<event> minute 74: Musiala is substituted by Führich.</event>
<event> minute 84: Gündogan is substituted by Undav.</event>
<event> minute 90: Final whistle. Germany wins 2:0 against Hungary.</event>

Sample events of the match Switzerland vs. Germany:

<event> minute 28: Goal by Ndoye for Switzerland.</event>
<event> minute 76: Musiala is substituted by Füllkrug.</event>
<event> minute 76: Wirtz is substituted by Sane.</event>
<event> minute 90: Goal by Füllkrug for Germany.</event>
<event> minute 90: Final whistle. 1:1 draw between Switzerland and Germany.</event>

Sample events of the match Denmark vs. Germany:

<event> minute 52: Andersen played the ball with his arm in the German penalty area.</event>
<event> minute 53: Havertz executes the penalty and scores the 1:0.</event>
<event> minute 64: Gündogan is substituted by Füllkrug.</event>
<event> minute 68: Goal by Musiala for Germany.</event>
<event> minute 74: Musiala is substituted by Wirtz.</event>
<event> minute 90: Final whistle. Germany wins 2:0 against Denmark.</event>

Sample events of the (in the exam setting still ongoing) match Spain vs. Germany:

<event> minute 8: Lopez is substituted by Olmo.</event>
<event> minute 13: yellow card for Rüdiger after fouling Olmo.</event>

```
<event> minute 74: Musiala is substituted by Wirtz.</event>  
<event> minute 90: Final whistle. Germany wins 2:0 against Denmark.</event>
```

Sample events of the (in the exam setting still ongoing) match Spain vs Germany:

```
<event> minute 8: Lopez is substituted by Olmo.</event>  
<event> minute 13: yellow card for Rüdiger after fouling Olmo.</event>  
<event> minute 58: Sane is replaced by Wirtz.</event>  
<event> minute 51: Goal by Olmo for Spain.</event>  
<event> minute 57: Gündogan is substituted by Füllkrug.</event>  
<event> minute 67: yellow card for Kroos after fouling Olmo.</event>  
<event> minute 80: Williams is substituted by Merino.</event>  
<event> minute 89: Kimmich lays off by head to Wirtz, who takes the  
ball directly and hits it against the left inside post, from where  
it bounces into goal to the 1:1.</event>  
<event> minute 90: after a short break, there will be 2x 15 minutes  
overtime.</event>  
<event> minute 106: Füllkrug passes to Musiala, who shoots from about  
20 meters and hits Cucurella's hand in the Spanish penalty area.  
Referee Taylor lets the match continue. No penalty for Germany.</event>
```

The text for the event that has later to be inserted in the UPDATE
Exercise for the 119th minute:

```
<event> minute 119: Olmo serves the ball on Merino's head in the  
German penalty area, no German defender is there, and Merino heads  
the ball into the German goal. 2:1 for Spain!</event>
```

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 (XML match ticker information [15 Points])

The database should also contain the detailed ticker events during the matches as e.g. shown in

<https://www.kicker.de/spanien-gegen-deutschland-2024-europameisterschaft-4772140/ticker>. Sample events are already given in the file exam.xml as text-only events without any useful markup for easier handling via copy-replace.

Design a markup that allows to query this information, e.g., who scored the goals, which players are substituted in and out etc.:

- for every such event, the minute in which it happens should be stored in a way that allows for querying.
- Note that several events can happen in the same minute, and should be stored and listed in their “natural” order (e.g., first a foul, then a red card, and then a penalty kick).

- The information should allow for querying, e.g., who scored the goals, which players are substituted in and out etc.
- The events of the matches Germany-Scotland and Spain-Germany are mandatory for your XML; the others are optional.
- Do NOT store HTML, but choose a suitable XML markup of the entries.
- It is *not* required to extend the DTD!

Exercise 4 (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 or an XQuery query that returns the *names* of those German players who were at least once shown the yellow card.

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 5 (XPath (b) [4 Points])

Give an XPath query or an XQuery query that returns the *names* of those players who were not in the initial line-up in any match.

Write the XPath 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 6 (XPath/XQuery (c) [4 Points])

Give an XPath query or an XQuery query that returns the names of those players who play for the national team of the country where also their club is located.

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

Exercise 7 (XPath/XQuery (d) [6 Points])

Give an XQuery query that outputs the scorer list, i.e., for each player how many goals he scored for his team (ordered by the number of goals, list only players that scored at least one goal) in the form

```
<player name="..." goals="..."/>
```

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

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

Give an XQuery query that yields the names of those players who played in *every* match of their team (i.e. either in the initial line-up or substituted in during the match).

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

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

Give an XQuery query that outputs all pairs of players who play in the same club, but in different national teams, and there was/is at least one of the European Championship matches between their national teams where both of them participated (not necessarily at the same time, in case they were substituted in or out). `
` The output should be of the form

```
<result player1=".." country1=".." player2=".." country2=".." club=".."/>
```

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

Exercise 10 (XQuery Update [8 Points])

The liveticker events are maintained continuously. We assume to be in the 119th minute of the quarterfinal between Spain and Germany on 5.7.2024 when Merino scores the 2:1 for Spain. The text for the event is (as also given in the frame for exam.xml)

```
<event> minute 119: Olmo serves the ball on Merino's head in the
German penalty area, no German defender is there, and Merino heads
the ball into the German goal. 2:1 for Spain!</event>
```

Give the XQuery update statement(s) that extend the ticker appropriately (not necessary to execute, recall that updates can only be executed if you have the Saxon Enterprise Edition installed):

Exercise 11 (XSLT [16 Points])

Extend the given XSL stylesheet frame exam.xsl to an XSLT stylesheet that generates an HTML page for a match (use Germany vs. Scotland as example):

- A short heading part naming the participating teams, the result, and the players line-up list at the beginning, and
- –most important– the ticker contents as an HTML list or table (minute, text of the event), similar to <https://www.kicker.de/deutschland-gegen-schottland-2024-europameisterschaft-4772146/ticker> (the ticker of the D-SCO match). Every mentioning of a player should contain a link to the player's page on the Web site that is assumed to have the URL `<http://players/ID>` where *ID* is the id of the player.
- (5 points for: for every goal entry, output "GOAL!!! *x : y*" where *x : y* is the current new result after this goal)

Use the following call to execute it:

```
saxonXSL.bat -s:exam.xml exam.xsl
```

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

Exercise 12 (XML Markup Annotation Design [5 Points])

In Exercise 3, it was recommended not to store the ticker events in HTML. What would be the problems when storing the ticker contents directly in HTML?

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>
```

Additional Training Exercises

Exercise 13 (XPath (2b) [0 Points])

Give an XPath query or an XQuery query that returns the *names* of those players who did not participate in any match (i.e. neither in the initial line-up nor substituted in during the match).

Write the XPath query string in the file `query2b.xq` and call it with

```
saxonXQ.bat -s:exam.xml query2b.xq
```

Copy-and-paste the query from `query2b.xq` afterwards here:

Exercise 14 (XPath (3b) [0 Points])

Give an XPath query or an XQuery query that returns the *names* of those players who have scored a goal for their team, and have played during the whole match (i.e. were neither substituted in nor out during this match)

Write the XPath query string in the file `query3b.xq` and call it with

```
saxonXQ.bat -s:exam.xml query3b.xq
```

Copy-and-paste the query from `query3b.xq` afterwards here:

Exercise 15 (XPath/XQuery (4b) [0 Points])

Give an XQuery query that outputs the name of the youngest player of the EM. Give also his age at the date of the final (which he played until the 89th minute, and his team won the championship) in the following form:

```
<youngest name="..." age-at-final="..."/>
```

Copy-and-paste the query from `query4b.xq` afterwards here:

Exercise 16 (XPath/XQuery (4c) [0 Points])

Consider that the result of the game would not be stored xplicitly, but has to be computed from the ticker events.

Give an XQuery query that outputs for each match the result in the following form:

```
<match team1="..." team2="..." result="goals1:goals2"/>
```

(where `goals1` and `goals2` are the number of goals for the respective team)

Copy-and-paste the query from `query4c.xq` afterwards here:

Exercise 17 (XPath/XQuery (4d) [0 Points])

- a) Give an XQuery query that outputs for each player the total time (in minutes) that he played during the championship. The result should be in the in the form

```
<player name="..." minutes="..."/>
```

(In soccer, a player who is substituted out cannot be substituted in again.)

Copy-and-paste the query from `query4d.xq` afterwards here:

- b) Consider e.g. ice hockey, basketball or volleyball, where players are frequently substituted out and in again. Sketch how the total time playing can be computed in these cases.