Prof. Dr. Stefan Brass
Institut für Informatik
MLU Halle-Wittenberg

December 15, 2017

XML and Databases
— Exercise Sheet 9 —

You only have to submit the parts marked as "Homework Exercises", i.e. Part d). But please think about the questions in Part a) before the meeting! Send your homework solutions to the instructor via EMail: `brass@informatik.uni-halle.de` (with "`xml17`" in the subject line). The official deadline is December 21, 10:00 (before the lecture time).

Repetition Questions

a) Answer the following questions about XPath:

- Name some differences between XPath 1.0 and XPath 2.0. I.e. what XPath 2.0 constructs cannot be used in a stylesheet that must run in a browser that supports only XPath 1.0?
- What is the meaning of the following XPath condition, i.e. for which attribute values this condition will be true?

  ```
  @weekday = ('Sat', 'Sun')
  ```

- Why might "`$x idiv 1`" be interesting? (Remember that the result of `idiv` is always an integer while the arguments of `idiv` can be of any numeric type.)
- What is the difference between the following XPath expressions?
  - `if @GUEST_STUDENT = true() then ... else ...`
  - `if @GUEST_STUDENT then ... else ...`
- What is the difference between the following XPath expressions?
  - `true`
  - `true()`
- What is the difference between the following XPath expressions?
  - `not /A`
  - `not(/A)`
- What is the meaning of "`not(*)`"?
- How can you check whether the value of an attribute `A` of the context node is the lexical representation of an integer (i.e. a sequence of digits, possibly with a sign)? How can you return the integer, if it is such an integer constant, and `-1` if it is not?

- What is the difference between the following XPath conditions?
  - `@A castable as xs:positiveInteger`
  - `@A instance of xs:positiveInteger`
- What is the difference between the following XPath expressions?
  - `@A cast as xs:positiveInteger`
  - `@A treat as xs:positiveInteger`
- What are the advantages and disadvantages of static type checking in XPath?
- Name ten important/common XPath functions. Also think about useful chapters for an XPath function manual. There actually is one, the “XPath and XQuery Functions and Operators 3.1”: [https://www.w3.org/TR/xpath-functions-31/].

In-Class Exercises

b) Consider the homework grades databases with data in elements:

```
<?xml version='1.0' encoding='ISO-8859-1'?>
<GRADES-DB>
    <STUDENTS>
        <STUDENT>
            <SID>101</SID>
            <FIRST>Ann</FIRST>
            <LAST>Smith</LAST>
        </STUDENT>
        ...
    </STUDENTS>
    <EXERCISES>
        <EXERCISE>
            <CAT>H</CAT>
            <ENO>1</ENO>
            <TOPIC>Relational Algebra</TOPIC>
            <MAXPT>10</MAXPT>
        <EXERCISE>
        ...
    </EXERCISES>
    <RESULTS>
        <RESULT>
            <SID>101</SID>
            <CAT>H</CAT>
            <ENO>1</ENO>
            <POINTS>10</POINTS>
        </RESULT>
        ...
    </RESULTS>
</GRADES-DB>
```

The file is available at: [http://users.informatik.uni-halle.de/˜brass/xml17/ex2.xml]

- Please write an XPath query to find the average number of points for Homework 1.
- What is the meaning of the following XPath query?

  ```
  for $p in max(//POINTS) return //RESULT[POINTS=$p]
  ```

- Does this query have the same result?

  ```
  //RESULT[not(number(POINTS) < //RESULT/POINTS)]
  ```

  This query uses only constructs of XPath 1.0.

- Why does the query not work without "`number`", i.e.

  ```
  //RESULT[not(POINTS < //RESULT/POINTS)]
  ```

  The actual data values are 5, 7, 8, 9, 10, 12. You observe that the query returns only `RESULT`-elements with 9 points. Why?

- Write an XPath query to print the last name of the student or students who got the maximal number of points for Homework 1. Write two versions: One with the maximum that was actually reached, and one with the `MAXPT` value stored in the exercise. (You can try it also with Homework 2, there no student got `MAXPT` points.)
- Print the last names of all students who have not submitted any homework.
- Print the last names of all students who have got at least 9 points for Homework 1 and for Homework 2.
- What is the average sum of homework points over all students who have submitted at least one homework?

Homework Exercises

d) Consider again the XML file for the classical music CDs:

- Data file:
  [http://www.informatik.uni-halle.de/~brass/xml17/cd.xml]
- XML Schema definition:
  [http://www.informatik.uni-halle.de/~brass/xml17/cd.xsd]
- DTD:
  [http://www.informatik.uni-halle.de/~brass/xml17/cd.dtd]

The document has the following elements:

- `CDDB`: `(composers, cds?, soloists?)`
- `composers`: `(composer*)`
- `composer`: `(pieceOfMusic*)`, attributes: `cno`, `firstName`, `name`, `born`, `died`.
- `pieceOfMusic`: `(recording*)`, attributes: `pno`, `title`, `key`, `opus`.
- `recording`: empty content, attributes: `rno`, `orchestra`, `conductor`.
- `cds`: `(cd*)`.
- `cd`: `(track*)`, attributes: `cdno`, `name`, `producer`, `numDiscs`, `totalTime`.
- `track`: empty content, attribute: `rno`.
- `soloists`: `(soloist*)`.
- `soloist`: `(performance*)`, attribute: `name`.
- `performance`: empty content, attributes: `rno`, `instrument`.

Please write the following queries in XPath and test it with an actual XPath implementation:

- Select the earlierst composer in the document, i.e. the composer with the minimum birth year (attribute `born`). Print the entire `composer` element.
- Print the last names of all composers that have written a symphony, i.e. the title of the music piece contains one of the substrings “`Sinfon`” or “`Symphon`”. Make sure that each composer appears only once, even if he has written more than one symphony (for different composers with the same name, it would be best to print the name twice, but you may also eliminate all duplicates). If you want, you can make the comparison case insensitive, but that is not required.
- Which CDs contain pieces of music from more than one composer? Print the name of the CD.
- What is the average number of tracks on a CD? For CD packs with more than one disc, divide the total number of tracks by the number of discs.