DD1334 Databasteknik

Laboration 3: Semi-Structured Databases

Andreas Gustafsson, Hedvig Kjellström and Michael Minock and John Folkesson

---

The purpose of Laboration 3 is to get acquainted with a semi-structured database modeling technique, XML. You will learn 1) how to store data in XML format; 2) how to find data from XML files using XQuery.

The recommended reading for Laboration 3 is the material relating to XML.

Come prepared to the review session! *The review will take 10 minutes or less, so have all papers in order. Passing requires:*

Completed Task 1 with 9 of the 10 queries right.

See the Lab Grading page on canvas for the due dates for the labs and the grading of late assignments.

Laboration 3 is intended to take 18h to complete.

Video XML Database

XML (eXtensible Markup Language) is a powerful tool that is widely used for a variety of things. One of the primary uses of XML is to store data in a human readable format. Data is enclosed in nested tags that simplifies searching for patterns in a XML file.

In this assignment we will work with a video database in XML format. It can be found in the file videos.xml, found on the KTH Canvas website.

If you are using the CSC Ubuntu machines, you can use XQilla to evaluate the queries in Task 1. To evaluate a query, put it in a text file. Execute the query as:

```
xqilla <queryfile>
```

Task 1: XQuery

Formulate XML queries using XQuery that correspond to the following questions to the video database. Group the results into XML elements where applicable (see the answers below). If you got it right, you should get the same answers as noted below. Note however that your queries are not necessarily correct just because the output matches, you should consider special cases.

Notice that the book example 12.19 does not work on our system. One has to use the syntax 'order by' not just 'order'

Another tip is

```
let $nl := "&#10;"
```

…

```
return concat($nl,'   ',$t)}
```

will place a new line before it outputs whatever $t is.

another tip is how {} can allow you to insert expressions into text outputs:

```
return <movie director="{$d}"/>
```

another tip is how 'for … at… in' works. Here $count will tell us how far into $qes we have come and we can reference it in the for loop.

```
for $q at $count in $qes
```

another tip is the empty return () which is handy because you have to have an else

```
return if … then … else ()
```

Before the review: Write down your queries. Then pose all queries to the system and save the output, e.g. as a text file. Count % correct queries and make note of this.

At the review: Show the queries and answers (with % correct noted) to the assistant. Both group members should be prepared to answer questions about the queries.

1. Which movies have the genre "special"?
   Answer: *<title>Everest</title>*
   <title>National Geographic's Mysteries of Egypt</title>

2. Which director has directed at least two movies, and which movies has he directed?
   Answer: *<movie director="Tokko">*
   <title>Doquissåpan, avsnitt 1</title>
   <title>Doquissåpan, avsnitt 2</title>
   <title>Doquissåpan, avsnitt 4</title></movie>

3. Which are the top ten recommended movies?
   Answer: *<title>Enemy of the State</title>*
   <title>Star Wars</title>
   <title>Au Revoir Les Enfants</title>
   <title>Pinocchio</title>
   <title>Doquissåpan, avsnitt 1</title>
   <title>Doquissåpan, avsnitt 2</title>
   <title>Doquissåpan, avsnitt 4</title>
   <title>The Fugitive</title>
   <title>MUPPET TREASURE ISLAND</title>
   <title>Yellow Submarine</title>

4. Which actors have starred in the most movies?
   Answer: *actor="Ford, Harrison"*
                      actor="Doqumenteriet2011"

5. Which is one of the highest rating movie starring both Brad Pitt and Morgan Freeman?
   Answer: *<title>Seven</title>*

6. Which actors have starred in a PG-13 movie between 1997 and 2006 (including 1997 and 2006)?
   Answer: *Ford, Harrison*
                      Jones, Tommy Lee
                      Heche, Anne
                      Schwimmer, David
                      Doqumenteriet2011

7. Who have starred in the most distinct types of genre?
   Answer: *actor="Ford, Harrison"*

8. Which director have the highest sum of user ratings?
   Answer: *Tokko*

9. Which movie should you recommend to a customer if they want to see a horror movie and do not have a laserdisk?
   Answer: *<title>Seven</title>*

10. Group the movies by genre and sort them by user rating within each genre.
    Answer: *<genre genre="action">*
                        <title>Enemy of the State</title>
                        <title>Star Wars</title >
                        <title>The Fugitive</title></genre >
                       <genre genre="comedy">
                        <title>Clerks</title>
                        <title>Six Days Seven Nights</title></genre>
                       <genre genre="drama">
                        <title>Contact</title>
                        <title>2001: A Space Odyssey</title></genre>
                       <genre genre="family">
                        <title>Pinocchio</title>
                        <title>MUPPET TREASURE ISLAND</title></genre>
                       <genre genre="foreign">
                        <title>Au Revoir Les Enfants</title>
                        <title>La Femme Nikita</title></genre>
                       <genre genre="horror">
                        <title>Seven</title>
                        <title>The Fear</title></genre>
                       <genre genre="musical">
                        <title>Yellow Submarine</title>
                        <title>Carmen</title></genre>
                       <genre genre="special">
                        <title>Everest</title>

```
<title>National Geographic's Mysteries
        of Egypt</title></genre>
<genre genre="documentary">
  <title>Doquissåpan, avsnitt 1</title>
  <title>Doquissåpan, avsnitt 2</title>
  <title>Doquissåpan, avsnitt 4</title></genre>
```