Hand-in 4: XML Processing

This hand-in must be handed in as a single PDF file by each group using the LearnIT system no later than

Tuesday November 29, 23.59*.*

We will not be able to extend this deadline, as the exam office needs the list of students by the end of that week.

The fourth part of the project consists of extracting parts of the movie database, and documenting the output format by XML schemas. You must extract two files:

- `horror70s.xml` that contains information on movie references to or from horror movies made in the 1970s. Each movie should be represented by a single XML element <movie>...</movie>, inside of which should be information on the names of horror movies that either refer to, or are referred to by, the movie in question.
- `bacon-people.xml` that contains information on the collaborators of Kevin Bacon. This should be a list of actors and directors who have worked with him on a movie. For each collaborator there should be a name, other bibliographical information available in the database, and a list of the names and years of all joint movies.

Export from MySQL

The first step is to write an SQL query that returns sufficient information to create the document (but is not in the desired format). MySQL allows XML output in a standard format. For example, from the command line one can write:

```
mysql -uroot -e "SELECT * FROM movieRef" --xml > links.xml
```

to put the results of the given query, in am XML format, into the file `links.xml`. Other methods exist, such as XML export from MySQL Workbench (Query -> Export results). Choose any one you want; the XML format output may be slightly different depending on your method.

XML processing

The next step is to apply XPath and/or XQuery to transform the XML documents into the desired format. We suggest that you use the eXist system for this, though using any other XQuery interpreter is ok. To load the XML data into eXist use either the Admin Client on http://exist.itu.dk:8080 or your own installation of eXist. The latter is preferable, as `exist.itu.dk` has been observed to have long response times when there are many users.

XML Schema

Your last task is to make XML schemas, *MovieRefML* and *CollaboratorML,* that describe the data format produced by your conversion. Your schema should be as specific as possible, i.e., not allow significant deviations from the format and conventions used. Run a validator to ensure that your XML files are contained in the language described by the schema.

To be handed in

- Name of all group members who contributed to the hand-in.
- Your XML schemas, including a description of any nontrivial choices made.
- Your converted XML files.
- A description of how you converted the data, including the MySQL and XPath/XQuery statements used, and validator reports.

Course goal covered by this hand-in

After the course the students should be able to:
- write simple XML Schemas and simple XQuery.