

# XML Technologies: Exercise 6

## XQuery

This is a mandatory exercise and the result will be part of your final mark. The solution must be uploaded to OLAT by **May 6<sup>th</sup> at 15:59**. Late submissions will not be accepted.

Submit the following files in a zipped archive:

- rats.xq
- interesting.xq

Make sure the archive is named [lastname]\_[firstname]\_6.zip (for example *mueller-mathias\_6.zip*). Some parts of your submission may be automatically evaluated, so make sure to name your files *precisely* as prescribed, otherwise you might not get any points.

## 1 Introducing XQuery

XQuery is used to query XML documents. So far, you have learned that XSLT is for transforming XML documents, and XPath for pointing to nodes within XML files. We can now refine our paradigm of X-related technologies to include XQuery in the following way:

- XPath is for navigating XML documents
- XQuery is for querying XML documents
- XSLT is for transforming XML documents
- XML Schema is for validating documents
- XSL-FO is for defining how to display XML documents

### 1.1 Medical Markup

After contemplating X-wing construction and whiskies, we will now turn our attention towards other uses of XML. The US National Library of Medicine provides the general public with a huge database comprising bibliographical information. Unlike the previous XML documents, this is a real-world and up-to-date example of how XML is used.

The attached XML snippet `snippet.xml` is a small sample from a single MEDLINE citation. **Study it well before you continue.** The full citation, attached as `medsamp2016.xml`, covers information about the time and place of publication, the kind of publication, the publisher, the authors and their affiliation, the abstract, the topics, the chemical substances involved, any databases that were used etc.

## 1.2 Rats (2 points)

The rat is an important model organism in medicine, especially in genetics, since we share a fair amount of DNA (and a fair amount of mechanisms that regulate gene expression) with it.

**Write an XQuery that finds all MEDLINE citations in `medsamp2016.xml` where the Abstract text contains the word `rats` and returns their Article titles, as elements. The elements should be sorted by year of publication, showing first the article that was published most recently. (Don't worry about getting the sorting within the year right.) Save the query as `rats.xq`.**

## 2 Querying Several Documents (3 points)

As you know by now, XQuery is a full *superset* of XPath. Among the operations that are exclusive to XQuery (e.g. creating new nodes, reorder the results) are retrieving and linking XML data from several input files.

Consider the XML file `interesting-articles.xml` which contains the IDs of interesting MEDLINE articles. Naturally, we would like to know what those allegedly interesting articles are about.

**Write an XQuery that queries both the MEDLINE database file `medsamp2016.xml` and the `interesting-articles.xml` file, finds the interesting articles via their PMID, then for each of the four articles finds all keywords. For each interesting article, return all keywords, separated by commas, inside an article element:**

```
<!--Sample query result-->
<articles>
  <article PMID="12345678">
    Drosophila melanogaster ,
    Fruit flies , Balancer Chromosomes
  </article>
</articles>
```

Save the query as `interesting.xq`.