

Exercise Sheet 3: XPath

4.0 VU Semi-structured Data

Winter Semester 2024

The goal of this exercise is to gain practical experience in querying XML documents using XPath expressions.

Task 1

Consider the DTD file `takeout.dtd` that is provided with this exercise sheet. This DTD describes XML documents in the context of an example take-out restaurant. Specifically, an XML document that is valid according to this DTD contains the following content:

- offered dishes, where each dish has an identifier, a name, a description, as well as an indication whether the dish is vegetarian;
- staff members, where each person has an identifier, a name, and possibly a description of the transport that the person is using to deliver orders (e.g., bicycle, car, etc.);
- current orders of the restaurant, where each order contains a collection of items that need to be delivered to a particular address by a staff member, or to be picked up by a client at the restaurant. An order has an identifier and a total price. An item in an order has a reference to a dish identifier and stores the price of the item.

Subtask A

Create an example XML document `takeout.xml` that is valid with respect to `takeout.dtd`. It should contain some non-trivial content, so that you can train writing XPath expressions in a meaningful way (see below). If you wish, you may also modify `takeout.dtd` to support additional markup work (e.g., additional attributes, elements, etc.)

Subtask B

Write down XPath expressions for the following queries:

1. Retrieve the identifiers (the values of the `id` attribute) of all staff members working at the restaurant.
2. Retrieve the names (the values of the `name` attribute) of all vegetarian dishes.

3. Count the number of `item` elements in the 2nd order of the restaurant (i.e., in the second `order` element that appears as content of the `current-orders` element).
4. Retrieve client names (the values of the `client-name` attribute) of customers who have more than 2 items in their order.
5. Retrieve the names of staff members who are currently assigned to deliver some order.
6. Retrieve the identifiers of orders whose total price is more than 20. Specifically, the total price of an order is the sum of item prices over all items in the order.
7. Invent additional two queries of your choice. (Try to use XPath features that are challenging to you, not the ones that are easy.)

Document your queries and evaluation results in a file `takeout-xpath.txt`. For each query you should write down the corresponding XPath expression and explain the result of evaluating the expressions over the XML document in `takeout.xml`. For the queries you invented you should also write down an informal description to describe them (like in points 1-6 above).

We recommend using `http://xpather.com` for this exercise sheet, but any alternative tool for evaluating XPath expression can be used as well. As a command-line tool, you can use

```
xmllint --xpath "your_xpath_expression" document.xml
```

Task 2

Recall the scenario that was assigned to you¹. For this task, pick one (non-trivial) XML document in your scenario (e.g., from a previous exercise) and store it in a file `ex3.xml`. Your task is to write and explain 5 example XPath expressions for querying XML documents in your scenario. Your examples should demonstrate the key features of XPath that were discussed (different kinds of axes, node tests, predicates, functions, abbreviations, etc). Document your queries and evaluation results in a file `ex3-xpath.txt`. Specifically, for each example you should provide:

- a query description in natural language (like in Task 1),
- a corresponding XPath expression,
- a short description of the feature(s) the expression demonstrates,
- a description of the result of evaluating the XPath expression over `ex3.xml`.

When working on this exercise, focus on aspects that are challenging to you, not the ones that are easy. If you spend some significant time on some particular aspect, think how you can communicate your gained insights with fellow students during the session for work in groups.

¹Your scenario is based on the last digit of your student ID. Specifically, 0 or 1: Cinema; 2 or 3: Art Gallery; 4 or 5: University; 6 or 7: Music Festival; 8 or 9: Restaurant.

Guidance for work in groups

When presenting your solution in a student group, you should ideally focus on the more original content (i.e., the 2 invented queries from Subtask B as well as the whole Task 2).

- Initially, you should **briefly** discuss as a group and clarify if there were significant difficulties in using XPath to express the queries 1-6 from Subtask B. If there are remaining problems, please see if they can be quickly resolved during the discussion. If they cannot be quickly resolved, document it in the discussion log, and use the discussion forum and the extra week to resolve the issues.
- For the queries that you invented, please explain to your fellow students what they do, how they work, what features they use.

After listening to the presentation of your colleague, you may discuss the following:

- How their solutions can be improved?
- What were the main challenges in this exercise?

Time management: ~10 min presentation & discussion per student in a group + 10 minutes to fill in and scan the discussion log.

Individual Reflection

TL;DR: The instructions are basically the same as in the previous sheet; your individual reflection should be stored in `ex3-reflection.txt`

Note: this task should be performed after you had completed and had a group discussion for this exercise sheet. However, we recommend that you read and keep in mind the questions bellow during your work on exercise sheet.

Think carefully about your experience in working on the main task of this exercise, both in terms of individual work as well as the work in groups.

- What are 3 key takeaway technical concepts that you became familiar with when performing the above tasks?
- What are 3 key observations and experiences from your work in groups?
- Do you feel that your feedback to the fellow students was helpful for (i) improving their solutions, (ii) for improving your own understanding of the technical concepts, and (iii) for improving your own solutions?
- Did you improve your solution after the student discussion? If “yes”, then how? If “no”, then why not?
- Can you relate the topics covered here with your previous experience in other lectures or other topics in Computer Science that you are familiar with?
- What do you think can be done in order to help the learning process in the remaining part of the course?

Create a plain text file `ex3-reflection.txt` and **write an individual reflection** (no more than 300 words) on the questions asked above as well as other experiences that you had. This task should be done within 1 week after the session for work in groups.

Submission to TUWEL

After the session for work in groups you are welcome to improve your solutions by taking into account the feedback you received, for which you have 1 week. **Within 1 week after the work in groups, you should submit to TUWEL** a ZIP file `ex3.zip` that contains the following files:

1. The scan of the discussion log from the session for work in groups (in a PDF/JPEG/GIF formats, with clear visibility).
2. `takeout.xml`
3. `takeout.dtd` (in case you modified the originally supplied DTD)
4. `takeout-xpath.txt`
5. `ex3.xml`
6. `ex3-xpath.txt`
7. `ex3-reflection.txt`