3rd year License Semi-structured data Academic year: 2024/2025

Lab₀₃

Exercise 01: Write a valid XML file against the following DTD

- <!ELEMENT directory (person)+>
- <!ELEMENT person (lastname, firstname, email+, tel*)>
- <!ATTLIST person profession (student | professor | assistant) "student">
- <!ELEMENT lastname (#PCDATA)>
- <!ELEMENT firstname (#PCDATA)>
- <!ELEMENT email (#PCDATA)>
- <!ELEMENT tel (#PCDATA)>
- <!ATTLIST tel type (fixed | portable) #REQUIRED>

Note:

1. You can link this DTD with your XML file using:

<!DOCTYPE Racine SYSTEM "file.dtd">

2. Verify that your file is well-formed and valid using the command:

xmllint --noout --valid file.xml

Or

xmllint --noout --dtdvalid file.dtd file.xml

(This command displays nothing if the file is well-formed and valid and displays errors otherwise).

Exercise 02:

Propose a BibTex.dtd file allowing the definition of XML documents representing Bibtex files (as described in Lab 02 exercise 03).

Exercise 03:

We want to represent the regulatory documents of construction. The regulatory documents are differentiated by the topicality (new or modified). They are grouped in the database of documents by the theme (e.g. accessibility for the disabled, acoustics, ventilation, asbestos, elevators, construction insurance, fire safety). Each document has a complex title which consists of a number, a date, a name, sometimes an acronym and is related to one or more specific areas. The date is required; it includes the date of the creation of a document and the date of its publication in an Official Journal

Exercise 04:

Suppose you need to define a structure for XML documents that list students in the Faculty of Computer Science. Consider the following requirements.

- A *studentlist* consists of one or more student entries.
- A student has a matriculation number by which they can be uniquely identified. A student entry consists of the student's name, the study program in which the student is enrolled, and the courses the student has taken so far. For each student, we need to record as well the year when the student enrolled.
- A student's name consists of the first name, the middle name, which is optional, and the last name.
- Study programs are "BSc in CS and Engineering", "BSc in Applied CS", "Master in CS", and "PhD in CS".

- Each course entry contains the name of the course. Courses are undergraduate (Bachelor) courses, postgraduate (Master) courses, or both. A course can also be mandatory or optional.
- A course entry contains the title of the course. We want to record as well whether the student has already received a course work mark for the course and, if so, the mark. Moreover, we want to record whether the student has already taken an exam for the course and, if so, what was the outcome of the exam.
 - 1. Write a DTD that defines the documents of type studentlist according to the requirements above.
 - 2. Give an instance of a studentlist document that is valid with respect to your DTD