

1. Unit: Exercises to XML

Information about the XML course can be found at
<http://www.stud.informatik.uni-goettingen.de/xml-lecture>

Exercise 1.1 (XML-Tree vs Directory-Tree)

Load `mondial-europe.xml` into `xmllint` and browse through the directory structure. First, change into the `country` element of Germany, then into the city of Göttingen. Then, change into the next city in the document.

Links to `xmllint` and the Mondial database can be found at
<http://www.stud.informatik.uni-goettingen.de/xml-lecture>.

Exercise 1.2 (Student-DTD)

- Write a DTD for XML documents with student data:
name, address and a student id, one or more subjects (computer science, law, chemistry, sociology etc)
- Write an XML document containing student data conforming to the DTD, and check it for validity using `xmllint`.

Exercise 1.3 (HTML-XHTML)

- Find a simple HTML document (e.g. your own personal student homepage, or any (simple) page from the Web) and convert it by hand from HTML to XHTML.
- Check the XHTML document for validity using the XHTML validator (<http://validator.w3.org/detailed.html>).

Hint: In your home directory in the CIP pool, there is a directory `public.html` which is your personal web directory. Files there are accessible via
<http://student.ifi.informatik.uni-goettingen.de/~<username>/<filename>>.

Exercise 1.4 (HTML-XHTML)

- Consider the following fragmentary XHTML DTD fragment :

```
<!ELEMENT html (head?,body)>
<!ELEMENT head (#PCDATA)>
<!ELEMENT body (p*)>
<!ELEMENT p (#PCDATA|table)*>
<!ELEMENT table (thead?,tbody)>
<!ELEMENT tbody (tr+)>
<!ELEMENT tr (td+)>
<!ELEMENT td (#PCDATA)>
```

Give a DFA that accepts the language of that DTD.

- Consider the following example of an HTML fragment (where nearly all closing tags are missing, and the table markup is far from correct):

```
<html>
<head><title>A very unprecise HTML page
<body>
  some text
  <p>
  <table>
    <tr> <td> eins.eins <td> eins.zwei
    <tr> <td> zwei.eins <td> zwei.zwei
  </table>
```

```
<p>  
  and some more text  
</html>
```

Extend your DFA such that it accepts this fragment.

Exercise 1.5 (DFAs and DTDs)

Consider the following DTD:

```
<!ELEMENT date (day,month,year?)>  
<!ELEMENT day (#PCDATA)>  
<!ELEMENT month (#PCDATA)>  
<!ELEMENT year (#PCDATA)>  
<!ELEMENT a (date*)>  
<!ELEMENT b (#PCDATA)>  
<!ELEMENT c (b+|(b?,a)*)>
```

Give a deterministic finite automaton for each element definition which accepts the corresponding *content model* and connect the automata to parse XML files according to this DTD.

Exercise 1.6 (Is XML a context-free language?) Consider XML as a formal language.

- (a) is the *language of all XML documents of a given document type*, specified by a DTD that does not contain any attributes context-free?
- (b) consider the case where the DTD contains attributes.
- (c) is the *language of all well-formed XML documents*, without known document type, or with no document type at all context-free?