Theoretical Computer Science

Winter semester 2021/22

Prof. Dr. Georg Schied

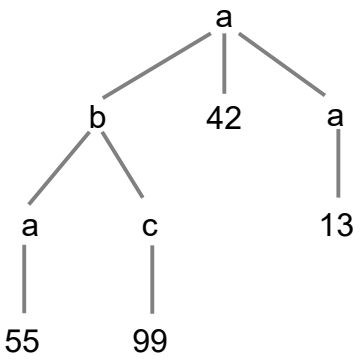
Assignment 3

Deadline: Wednesday, 27 October 2021

- 10 out of 20 points must be achieved in order to pass.

Exercise 3.1

Represent the following tree as an XML document:



Exercise 3.2 – obligatory (4 points)

a) Are the following XML documents *well-formed*? If not, indicate *all* errors.

b) Represent the structure of the well-formed XML documents as trees.

(1)

```
<a>
    <b> w </b>
    <d>
       <e> x <e>
       <f>
    </d>
</a>
<c>
   <d>
       <e> y </e>
   <f>
   </d>
      <e> z </e>
   </f>
</c>
```

(2)

```
<a><b>w</b><d><e>x</e></d><d><e>y</e></d><f>z</f></a>
```

(3)

```
<a>
   <b>xx</b>
   <a>
      <c>yy</c>
      zz
   </a>
   <a>
      xx
   </a>
</a>
```

Exercise 3.3

Specify regular expressions for the following languages:

a) L_a = {$w \in$ {0,1}* | w contains 01 as a substring}

b) L_b = {w $\in$ {a,b,c}* | w either starts with `a` and ends with `bc` or `cb`,
or starts with `b` and ends with `ac` }

c) L_c = {$w \in$ {a,b,c}* | $|w|_b$ is even} (0 is also considered as even)

Exercise 3.4 – obligatory (6 points)

Specify regular expressions for the following languages:

a) L_a = {$w \in$ {a,b}* | $|w|$ is odd}

b) L_b = {xwx | $x \in \Sigma$, $w \in \Sigma^*$ }, where Σ = {a, b, c},
i.e. all strings that start and end with the same symbol

c) L_c = {$w \in$ {a, b}* | w does contain neither aa nor bb as a sub-string}

Exercise 3.5

There exist many different number formats. One of them is as following, where groups of three digits are separated by apostrophes:

0
42
486
9’386
719’528
83’748’694’846

Indicate a regular expression that describes all possible numbers written this way.

Exercise 3.6 – obligatory (5 points)

Define the format of telephone numbers using regular expressions. Telephone numbers are allowed in the following format variants:

Format	Examples	Remark
Local number	123450	Does not start with 0
National number with area code	01234-123450 01234/123450	Starts with a single zero. Characters - or / are used as separator between area code and local number
International number with country code	(0049)1234-123450 (00500)800/9080	Starts with 00 and one to three digits as country indication. National area code is without leading 0.

Area code and local number may consist of one or more digits.

Exercise 3.7

Simplify the following regular expressions.

a) aa | a(bε | εa)

b) a(a | b) | (aa | ab)* ε

Exercise 3.8 – obligatory (5 points)

Simplify the following regular expressions. Show the intermediate steps of your transformations.

a) a((a | bb) a) | aaa

b) (ε | a(bd|cd) ((ac|ab)d)*)*

Exercise 3.9

Are the following equivalences valid for all regular expression *R* and *S*? Give a short justification.

(1) $R\,R^* = R^*\,R$

(2) $(RS)^* = R^*\,S^*$

(3) $(R \mid \varepsilon)^* = R^*$

(4) $(R \mid S)^* = R^* \mid S^*$

(5) $(R^* \mid S^*)^* = (R \mid S)^*$