1 Exercises about regular expressions

1. By the UNIX comman egrep (extended grep) you can search for patterns in the text, which can be defined as regular expressions. The basic syntax of `egrep is following: egrep <expression> <file>`, in which the expression can be

   - list of characters in the brackets, e.g. $[abcd]$: any of the characters a, b, c, d
   - $(expression1)(expression2)$: concatenation of two expressions
   - $(expression1)|(expression2)$: either expression1 or expression2
   - $(expression)*$: 0 or more times the expression (closure of the expression)
   - $\backslash b$: empty string in the edge of word $\backslash B$: empty string in the middle of word

   Notice! The expression might need single quotation marks ('expression') More information by command man egrep.

   Test the command egrep! You may use as your inputfile any (C-)program, e.g. file esim.c in the homepage. What kind of patterns do the following commands find?

   egrep '[1]*'
   egrep '[1][0]'
   egrep '[1]|[0]'

2. Construct an egrep-command to find the following lines:
   a) Rows, which consist numbers.
   b) Rows, which consist either word *while* or *for*
   c) Rows, which consist number 10
   d) Rows, which consist integers. (Notice! Your command should not accept desimal numbers.)

3. Let's consider the following languages of the alphabet $\Sigma = \{a, b\}$. For each of the languages give two strings that are memebers and two strings that are not memebers!

a) a^*b^*
b) $a(ba)^*b$
c) $a^* \cup b^*$
d) $(aaa)^*$
e) $(\epsilon \cup a)b$
f) $\Sigma^* a \Sigma^* a \Sigma^* a \Sigma^*$

4. What strings belong to the language described by the following expression? $(c \cup h \cup m \cup r)at((c \cup t)a \cup (s \cup t)o)ught(m \cup l \cup tw \cup r)ice$

5. How would you define a regular expression to find information about snowstorms? Notice that the word "snowstorm" may be connected to something else like "snowstorming" or "snow- and thunderstorm warning".

6. Which strings belong to the language $L(\emptyset^*)$? What about $L(\epsilon^*)$?

7. Find a shortest string which belongs to the language described by the following expression!

   a) $a^*(b \cup abb)b^*b$
   b) $a^*b^*b(a \cup (ab)^*)^*b^*$
   c) $(a \cup ab)(a^* \cup ab)^*b$

8. Construct the regular expressions corresponding the following languages:
   a) $\{w \in \{a,b\}^* | \text{the third last character of } w \text{ is } a\}$
   b) $\{w \in \{a,b\}^* | w \text{ contains either substring } ab \text{ or } ba\}$
   c) $\{w \in \{a,b\}^* | w \text{ contains the string } aba \text{ but doesn't cointain the string } bab\}$

9. Construct the regular expressions corresponding the following languages:
   a) $\{w \in \{a,b\}^* | w \text{ contains an even number of characters } a\}$
   b) $\{w \in \{a,b\}^* | \text{the length of } w \text{ is odd}\}$
   c) $\{w \in \{a,b\}^* | \text{number of characters } b \text{ is multiple of } 3\}$

10. Give a simpler form for the following expressions! (such that they still describe the same language!)

a) $(0 \cup 1 \cup 01 \cup 11)^*$
b) $(0^* \cup 10^*)^*$
c) $1^*(011^*)^* \cup 1^*(011^*)^*0$