

1 Exercise session 2

1. By the UNIX command `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)
 - `\b`: empty string in the edge of word `\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` with input file `esim.c` on homepage (<http://www.cs.joensuu.fi/pages/whamalai/tepe04/esim.c>)!

What kind of patterns do you find with the following command?
`egrep '(uu)|(sata)|(issa)|(oita)|(alla)' esim.c`

Do you invent a simpler command, which finds exactly the poem lines?

2. Construct an `egrep`-command to find the following lines from `esim.c`:
 - 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 decimal numbers.)
3. Let's consider the following languages of the alphabet $\Sigma = \{a, b\}$. For each of the languages give two strings that are members and two strings that are not members!

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

6. 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) $(a \cup ab)(a^* \cup ab)^*b$

7. Construct the regular expressions corresponding the following languages:

- a) $\{w \in \{a, b\}^* \mid \text{the third last character of } w \text{ is } a\}$
- b) $\{w \in \{a, b\}^* \mid w \text{ contains either substring } ab \text{ or } ba\}$
- c) $\{w \in \{a, b\}^* \mid w \text{ contains the string } aba \text{ but doesn't contain the string } bab\}$

8. Construct the regular expressions corresponding the following languages:

- a) $\{w \in \{a, b\}^* \mid w \text{ contains an even number of characters } a\}$
- b) $\{w \in \{a, b\}^* \mid \text{the length of } w \text{ is odd}\}$
- c) $\{w \in \{a, b\}^* \mid \text{number of characters } b \text{ is multiple of } 3\}$

9. Read the story about decision problems

<http://www.cs.joensuu.fi/pages/whamalai/tepe04/story.html> and complete it!