

Coursework 1 (Strand 1)

This coursework is worth 5% and is due on 20 October at 16:00. You are asked to implement a regular expression matcher and submit a document containing the answers for the questions below. You can do the implementation in any programming language you like, but you need to submit the source code with which you answered the questions, otherwise a mark of 0% will be awarded. You can submit your answers in a txt-file or pdf.

Disclaimer

It should be understood that the work you submit represents your own effort. You have not copied from anyone else. An exception is the Scala code I showed during the lectures or uploaded to KEATS, which you can use.

Tasks

The task is to implement a regular expression matcher based on derivatives. The implementation should be able to deal with the usual (basic) regular expressions

$$\emptyset, \epsilon, c, r_1 + r_2, r_1 \cdot r_2, r^*$$

but also with the following extended regular expressions:

$[c_1c_2 \dots c_n]$	a range of characters
r^+	one or more times r
$r^?$	optional r
$r^{\{n,m\}}$	at least n -times r but no more than m -times
$\sim r$	not-regular expression of r

In the case of $r^{\{n,m\}}$ we have the convention that $0 \leq n \leq m$. The meanings of the extended regular expressions are

$$\begin{aligned} L([c_1c_2 \dots c_n]) &\stackrel{\text{def}}{=} \{[c_1], [c_2], \dots, [c_n]\} \\ L(r^+) &\stackrel{\text{def}}{=} \bigcup_{1 \leq i} L(r)^i \\ L(r^?) &\stackrel{\text{def}}{=} L(r) \cup \{\emptyset\} \\ L(r^{\{n,m\}}) &\stackrel{\text{def}}{=} \bigcup_{n \leq i \leq m} L(r)^i \\ L(\sim r) &\stackrel{\text{def}}{=} \Sigma^* - L(r) \end{aligned}$$

whereby in the last clause the set Σ^* stands for the set of *all* strings over the alphabet Σ (in the implementation the alphabet can be just what is represented by, say, the type `char`). So $\sim r$ means 'all the strings that r cannot match'.

Be careful that your implementation of *nullable* and *der* satisfies for every r the following two properties (see also Question 2):

- $nullable(r)$ if and only if $[\] \in L(r)$
- $L(der\ c\ r) = Der\ c\ (L(r))$

Important! Your implementation should have explicit cases for the basic regular expressions, but also explicit cases for the extended regular expressions. That means do not treat the extended regular expressions by just translating them into the basic ones. See also Question 2, where you are asked to explicitly give the rules for $nullable$ and der for the extended regular expressions.

Question 1 (unmarked)

What is your King's email address (you will need it in Question 3)?

Question 2 (marked with 2%)

This question does not require any implementation. From the lectures you have seen the definitions for the functions $nullable$ and der for the basic regular expressions. Give the rules for the extended regular expressions:

$nullable([c_1c_2 \dots c_n])$	$\stackrel{\text{def}}{=} ?$
$nullable(r^+)$	$\stackrel{\text{def}}{=} ?$
$nullable(r^?)$	$\stackrel{\text{def}}{=} ?$
$nullable(r^{\{n,m\}})$	$\stackrel{\text{def}}{=} ?$
$nullable(\sim r)$	$\stackrel{\text{def}}{=} ?$
$der\ c\ ([c_1c_2 \dots c_n])$	$\stackrel{\text{def}}{=} ?$
$der\ c\ (r^+)$	$\stackrel{\text{def}}{=} ?$
$der\ c\ (r^?)$	$\stackrel{\text{def}}{=} ?$
$der\ c\ (r^{\{n,m\}})$	$\stackrel{\text{def}}{=} ?$
$der\ c\ (\sim r)$	$\stackrel{\text{def}}{=} ?$

Remember your definitions have to satisfy the two properties

- $nullable(r)$ if and only if $[\] \in L(r)$
- $L(der\ c\ r) = Der\ c\ (L(r))$

Question 3 (marked with 1%)

Implement the following regular expression for email addresses

$$([a-z0-9_-.]^\dagger) \cdot @ \cdot ([a-z0-9_-.]^\dagger) \cdot \cdot ([a-z.]^{\{2,6\}})$$

3. "aa
aa
aa"