Project 2017/18

Formal Languages and Compilers

MSc in Computer Science

University of Camerino

Prof. Luca Tesei

Description

Let $\Sigma = \{a-z, A-Z, 0-9\}$ be an alphabet. The objective is to use the ANTLR4 tool to derive an interpreter for regular expressions followed by a list of comma-separated words. Given the input $r, w_1, \ldots, w_n$ - where r is a regular expression on Σ and each w_i is a word on Σ - a sequence of n comma-separated "OK" or "KO" must be produced as output telling whether or not the corresponding word belongs to the language denoted by r.

The following table gives two examples.

Input	**result**
$a + bc$, a, b, bc	OK, KO, OK
$a * b$, aab, ba, b, abb	OK, KO, OK, KO

The part of grammar for the regular expressions must be designed with the usual precedence of operators ($*$ has precedence on concatenation that has precedence on $+$). The use of parentheses must be possible to override the precedence rules if desired. The regular expression ϵ should be syntactically represented by the underscore symbol '_'.

Hint: derive, during the parsing of r, an internal representation (as an object of an appropriately defined class) of the NFA corresponding to it obtained by the Thompson algorithm. Then, the automaton object can be used to decide whether or not each of the following words w_i belongs to the language denoted by r using the definition of acceptance for NFAs.

Submission

Prepare a written report describing your grammar, your code and some test results. You can use screenshots to show some of the results.

Send by email to the teacher all the files (as a jar) plus the report in pdf. The sending must occur before the starting of a written test (fixed for each exam session): a student that has not sent the project before the written test can not participate to the written test. The exam is passed when both the following conditions are satisfied:

1. the project has been sent and has been approved;
2. the written test has been passed;

In case a student does not pass the written test, he/she does not have to resend the (approved) project before the next attempt(s) to the written test. The final grade is a combination of the grades of the two tasks.