TD 2
Regular expressions

Version du September 20, 2020

Exercice 1 – Basic operators

We only consider in this exercise the following basic operators:

— the choice $(e_1 + e_2)$
— the concatenation $(e_1 e_2)$
— the repetition $(e^\star)$

You may omit superfluous parentheses as long as you respect the priority order of these operators (repetition has the priority over concatenation, which itself has priority over the choice).

Let $\Sigma = \{-, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, .\}$. Propose regular expressions recognizing the following sub-languages of $\Sigma^\star$.

1. Signed integers written in base 10. That is, with « − » in first position if needed, and no 0 as the first digit (except to represent 0).
2. Floating point numbers.
3. The decimal expansion of any real number, such as 3.141592, −318.29 or 42. Three additional constraints to spice things up:
   — a dot not followed by a any digit is not allowed,
   — again, the integer part cannot begin with 0, except for numbers comprised between −1 and 1,
   — −0 is not allowed.
4. Any natural number that is a multiple of 20.

Exercice 2 – Syntactic sugar

In addition to the previous basic operators, we now allow the following ones.

— For the regular expression e, $e^?$ is the abbreviation of $(\varepsilon + e)$.
— For the regular expression e, e^+ is the abbreviation of $ee^\star$.
— For some symbols $s_1, s_2, \ldots s_n$, $[s_1 s_2 \ldots s_n]$ stands for one of those symbols. This operator can easily be rewritten with the + operator. If $\Sigma = \{a, b, \ldots, z\}$ for instance, we have $[aeiou] = (a + e + i + o + u)$.
— If symbols of Σ are ordered (such as digits for instance, or our latin alphabet) $[s_1 - s_2]$ represents a symbol among those comprised between s_1 and s_2 (included). This operator can also be rewritten, for instance if $\Sigma = \{a, b, \ldots, z\}$, then we have $[a - e] = (a + b + c + d + e)$.

1. Simplify the regular expressions of the previous exercise using these new operators.
2. Let $\Sigma = \{-, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ., e\}$. Propose a regular expression recognizing a decimal number written in scientific notation. That is, a number such as $-1.234e56$, where
   — « − » is the sign. It can be absent.
   — « 1.234 » is the mantissa. It is comprised between 0 and 9.99999 . . ..
   — « e56 » is the exponent. It is optional, and should be interpreted as 10^{56}. The exponent is an integer number and can be signed, $2e{-}3$ represents 0.002 for instance.

   We allow superfluous 0s, as well as −0 (you should have understood that it is already quite a hassle to handle).

Exercice 3 – Simplification and equivalences

For each entry in the following list, decide whether the language denoted by the regular expression e is equal to, included in, contains, or is not comparable with the one denoted by the regular expression f for the alphabet $\Sigma = \{a, b, c\}$. Propose counter-examples when the two languages are different.

e	f
$a^\star b(ab)^\star$	$a^\star(bab)^\star$
$a(bb)^\star$	$ab^\star$
$a(a+b)^\star b$	$a^\star(a+b)^\star b^\star$
$abc + acb$	$a(b+c)(c+b)$
$a^\star bc + a^\star cb$	$a^\star(bc + a^\star cb)$
$(abc + acb)^\star$	$((abc)^\star(acb)^\star)^\star$
$(abc + acb)^+$	$((abc)^\star(acb)^\star)^+$
$(abc + acb)^\star$	$(abc(acb)^\star)^\star$
$(abc + acb)^\star$	$(a(bc)^\star(cb)^\star)^\star$

Exercice 4 – Intersection of languages

1. Is the intersection of two regular languages also a regular language?
2. Let L_1 and L_2 be the languages denoted by $ab+bc^+$ and $a^\star b^\star c^\star$, respectively. Propose a regular expression denoting the language $L_1L_2 \cap L_2L_1$.

Exercice 5

Determine which languages are equal to each other among the following ones.

$$(L \cup M)^\star \quad (LM)^\star L \quad L(LM)^\star \quad (L^\star \cup M)^\star \quad (M^\star \cup L)^\star \quad (L^\star M^\star)^\star \quad (M^\star L^\star)^\star \quad (L^\star \cup M^\star)^\star$$