

Exercises on lexical analysis

February 22, 2010

1. Define regular expressions for the following languages:

(a) Floating point numbers. Examples:

3.1416 -3e4 +1.0e-5 .567e+8

In case of using a ' . ', a decimal digit is always requested.

- (b) All strings from the alphabet $\{a, b, c\}$ in which the first appearance of symbol b is always preceded by at least one appearance of symbol a .
- (c) All strings of lowercase letters that contain the five vowels in order (each vowel must appear only once). Example: `zfaehipojksuj`.
- (d) All strings of lowercase letters in which the letters are in ascending lexicographic order. Examples: `afhmnqsy`, `abcdz`, `dgky`, but the string `bdeaz` does not belong to the language.
- (e) Financial quantities in American notation. Examples:

`$*2,345.67` `$12,452,183.16` `$****12`

These have a leading dollar sign (\$), an optional string of asterisks (* – used on checks to discourage a fraud), a string of decimal digits, and an optional fractional part consisting of a decimal point (.) and two decimal digits. The string of digits to the left of the decimal point may consist of a single zero (0). Otherwise it must not start with a zero. If there are more than three digits to the left of the decimal point, groups of three (counting from the right) must be separated by commas (,).

- (f) Inexact constants in Scheme. Scheme allows real numbers to be explicitly inexact (imprecise). A programmer who wants to express all constants using the same number of characters can use sharp signs (#) in place of any lower-significance digits whose values are not known. A base-ten constant without exponent consists of one or more digits followed by zero or more sharp signs. An optional decimal point can be placed at the beginning, the end, or anywhere in between. Examples:

`35##` `35#.#` `356.3##` `.35##` `35.` `35#.`

2. Define finite automata (deterministic or nondeterministic) for the following languages:

$(ab|ac)^*$ $a(a|b)^*a$ $((\varepsilon|a)b^*)^*$ $(a|b)^*a(a|b)(a|b)$