REGULAR EXPRESSIONS & LANGUAGES

20/11/2008

E6.1

EXERCISE 1

Write regular expressions for the following languages:

a) The set of all strings consisting of zero or more a's followed by zero or more b's, followed by zero or more c's;

b) The set of all strings that consist of either 01 repeated one or more times or 010 repeated one or more times;

c) The set of strings that either begin or end (or both) with 01;

d) The set of strings over $\{x, y, z\}$ such that the number of y's is divisible by three;

e) The set of strings over $\{0, 1\}$ such that at least one of the last ten positions is a 1;

f) The set of strings over $\{0, 1, \ldots, 9\}$ such that the final digit has appeared before;

g) The set of strings over $\{0, 1, \ldots, 9\}$ such that the final digit has not appeared before.

EXERCISE 2

Give English descriptions of the languages over the alphabet $\{a, b, c\}$ defined by the following regular expressions:

a) $(a+b)(a+b)(a+b)$ b) $(\varepsilon+a)b(\varepsilon+c)$

c) $(cb)^* + b(cb)^* + (cb)^*c + b(cb)^*c$

EXERCISE 3

a) Show that for every regular language L we have $(L^*)^* = L^*$.

b) Show that for all regular languages L and M we have $(L^*M^*)^* = (L \cup M)^*$. [Note: $(L \cup M)^* = \mathcal{L}((L+M)^*)$]

SOLUTIONS (20/11/2008)

E6.2

1)a) $a^* b^* c^*$

1)b) $(01)(01)^* + (010)(010)^*$ or $(01)^+ + (010)^+$

1)c) $(01)(0+1)^* + (0+1)^*(01)$

Note: we assume that the strings are over $\{0,1\}$.

1)d) $((x+z)^* y (x+z)^* y (x+z)^* y (x+z)^*)^*$

1)e) Let $E_i = \underbrace{(0+1)\cdots(0+1)}_{i \text{ times}} 1 \underbrace{(0+1)\cdots(0+1)}_{(9-i) \text{ times}}$, $i \in \{0,1,\dots,9\}$.

Then $E = (0+1)^*(E_0 + E_1 + \cdots + E_9)$.

1)f) Let $E_d = 0+1+\cdots+9$. Then $E = E_d^* 0 E_d^* 0 + E_d^* 1 E_d^* 1 + \cdots + E_d^* 9 E_d^* 9$.

1)g) Let $E_0 = 1+2+\cdots+9$, $E_i = 0+\cdots+(i-1)+(i+1)+\cdots+9$ $(1 \le i \le 8)$,

$E_9 = 0+1+\cdots+8$, and $E_d = 0+1+\cdots+9$.

Then $E = E_d + E_0^+ 0 + E_1^+ 1 + \cdots + E_9^+ 9$.

(Also: $E = E_0^* 0 + E_1^* 1 + \cdots + E_9^* 9$.)

2)a) The set of all strings of length three that do not contain the symbol c: {aaa, aab, aba, abb, baa, bab, bba, bbb}.

2)b) The set of all strings with exactly one b, eventually preceded by an a and/or followed by a c: {b, ab, bc, abc}.

2)c) The set of all strings consisting of alternating b's and c's. Alternative regular expressions for the language are:

- $(\varepsilon + c)(bc)^*(\varepsilon + b)$
- $(bc)^* + (cb)^* + c(bc)^* + b(cb)^*$

3) a) We have to show that $L^* \subseteq (L^*)^*$ and $(L^*)^* \subseteq L^*$. E6.3

$\boxed{L^* \subseteq (L^*)^*}$ Trivial since $(L^*)^* =_{def} \{\varepsilon\} \cup L^* \cup L^*L^* \cup \ldots$

$\boxed{(L^*)^* \subseteq L^*}$ Given $w \in (L^*)^*$ we have to show that $w \in L^*$. If $w \in (L^*)^*$ then there exists $n \in \mathbb{N}$ such that $w = w_1 \cdots w_n$ where $w_i \in L^*$ $(1 \le i \le n)$. Since, for all $i \in \{1, 2, \ldots, n\}$, there exists $m_i \in \mathbb{N}$ such that $w_i = w_{i1} \cdots w_{im_i}$ where $w_{ij} \in L$ $(1 \le j \le m_i)$ we have that $w = (w_{11} \cdots w_{1m_1}) \cdots (w_{n1} \cdots w_{nm_n})$. Thus $w \in L^*$.

3) b) We have to show that $(L^*M^*) \subseteq (L \cup M)^*$ and $(L \cup M)^* \subseteq (L^*M^*)^*$.

$\boxed{(L^*M^*)^* \subseteq (L \cup M)^*}$ Given $w \in (L^*M^*)^*$ we have to show that $w \in (L \cup M)^*$. If $w \in (L^*M^*)^*$ then $w = w_1 \cdots w_n$ where $w_i \in L^*M^*$. Since, for all $i \in \{1, \ldots, n\}$, $w_i = u_{i1} \cdots u_{ik_i} v_{i1} \cdots v_{il_i}$ where $u_{ij} \in L$ and $v_{ij} \in M$ we have that:

$$w = (u_{11} \cdots u_{1k_1} v_{11} \cdots v_{1l_1}) \cdots (u_{n1} \cdots u_{nk_n} v_{n1} \cdots v_{nl_n}).$$

Thus $w \in (L \cup M)^*$.

$\boxed{(L \cup M)^* \subseteq (L^*M^*)^*}$ Given $w \in (L \cup M)^*$ we have to show that $w \in (L^*M^*)^*$. If $w \in (L \cup M)^*$ then $w = w_1 \cdots w_n$ where each w_i is in either L or M. If w_i is in L then w_i is also in L^* and, since ε is in M^*, $w_i = w_i\varepsilon$ is in L^*M^*. Similarly, if w_i is in M then w_i is in L^*M^*. Thus $w \in (L^*M^*)^*$.