1. [*10 points*] Consider the alphabet $\Sigma = \{a, b, c\}$. Let K, L, and M be the languages over M given by:

$$K = \{\varepsilon, a, b, c\} \qquad L = \{aa, ab\} \qquad M = \{c, cb, cbb, cbbb, cbbbb, \dots\}$$

Find the following languages. In each case, specify the elements of the language using set notatoin, or give a clear description in words of the set of strings that make up the language. You do not have to justify your answers, but an explanation might get you some partial credit for an incorrect answer.

a) $K \cup L$ **b)** L^2 **c)** K^2 **d)** M^2 **e)** M^R

f) KL **g)** LM **h)** K^* **i)** L^* **j)** M^*

Answer:

a) $K \cup L = \{\varepsilon, a, b, c, aa, ab\}$

b) $L^2 = LL = \{aaaa, aaab, abaa, abab\}$

c) $K^2 = KK = \{\varepsilon, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc\}$ — For example, $b \in KK$ because it can be written εb, and $\varepsilon \in KK$ because it can be written $\varepsilon\varepsilon$.

d) M^2 is the set of strings of b's and c's that start with c and contain exactly two c's.

e) $M^R = \{c, bc, bbc, bbbc, bbbbc, \dots\}$

f) $KL = \{aa, bb, aaa, aab, baa, bab, caa, cab\}$

g) LM is the set of strings that begin with aac or abc, followed by any number of b's.

h) K^* contains all strings made up of a's, b's, and c's.

i) L^* contains all strings made up of pairs of letters, where each pair is either aa or ab. (Jonathan had a better answer: Strings of a's and b's that have even length and a b can only occur in an even-numbered position.)

j) M^* contains all strings of b's and c's that begin with a c, plus the empty string. Since every string in M starts with a c, every non-empty string in M^* starts with a c. To see that every string of b's and c's that starts with c is in M^*, we can break the string into pieces that come from M. For example, $cbbcccbcbbbbb = (cbb)(c)(c)(cb)(cbbbb)$.

2. [*2 points*] True or False, if true, give a proof; if false, give a counterexample: Let L be a language over an alphabet Σ. If $L^R = L$, then every string in L is a palindrome. (A palindrome is a string, x, such that $x^R = x$.)

Answer:

This is false. As a counterexample, consider the language $L = \{ab, ba\}$. Neither of the strings in L is a palindrome, but $L^R = \{(ab)^R, (ba)^R\} = \{ba, ab\} = \{ab, ba\} = L$. (Note that as long as Σ has at least two symbols, then $L - \Sigma^*$ is also a counterexample.)

3. [*6 points*] Let $\Sigma = \{a, b\}$. Consider the languages that are generated by the following regular expressions over Σ. Give clear descriptions in words of the set of strings in each language. You need to make it clear how the languages differ. (Note that the following is **not** an acceptable sort of answer: Any number of a's, then a b, than any number of a's, then a b, then any number of a's, then a b, then any number of a's. Give simple characterizations of the strings that are generated.)

a) $a^*(bbb)a^*$ **b)** $(a|b)^*bbb(a|b)^*$

c) $a^*ba^*ba^*ba^*$ **d)** $a^*(b|\varepsilon)a^*(b|\varepsilon)a^*(b|\varepsilon)a^*$

Answer:

a) Strings of a's and b's that contain the substring bbb, and no other b's.

b) Strings of a's and b's that contain the substring bbb, but can also contain any number of additional b's, anywhere in the string.

c) Strings of a's and b's that contain exactly three b's. There can be any number of a's before, after, and in between the b's.

d) Strings of a's and b's that contain at most three b's, with any number of a's in any positions. (The subexpression $b|\varepsilon$ generates either a b or nothing at all, so the string can contain three b's, but each b is optional. Note that this language can also be generated without using ε: $a^*|a^*ba^*|a^*ba^*ba^*|a^*ba^*ba^*ba^*$.)

4. [*4 points*] Let $\Sigma = \{a, b\}$. Consider the languages that are generated by the following regular expressions over Σ. Give clear descriptions in words of the set of strings in each language.

a) $ab(a|b|c)^*ba \mid aba$ **b)** $(b|c|ab|ac)^*(a|\varepsilon)$

Answer:

[Note that there is an error in the statement of this problem. Σ has to be $\{a, b, c\}$, not just $\{a, b\}$ for the question to makes sense. The regular expressions in this problem are regular expressions over the alphabet $\{a, b, c\}$.]

a) All strings of a's and b's that begin with ab and end with ba. Note that aba satisfies this description but is not generated by $ab(a|b|c)^*ba$, so it has to be added as a special case.

b) All strings of a's and b's that do not contain the substring aa. The expression $(b|c|ab|ac)^*$ can generate b's and c's in any order, but the only way to add an a to the string is if the a is as part of ab or ac, so any a is immediately followed by a b or by a c. In particular, it's not possible to add two a's in a row to the string. Note that the string generated by $(b|c|ab|ac)^*$ is either empty or ends with b or c, so we don't get any strings that end with a, The $(a|\varepsilon)$ at the end of the regular expression makes it possible to add an a at the end of the string.

5. [*10 points*] For each of the following languages, give a regular expression that generates that language. Justify your answers by explaining why the regular expression generates the strings of the language. Be careful to note the alphabet in each case, and be careful to account for **all** of the strings that satisfy the given condition.

a) $L_1 = \{\, x \in \{a,b\}^* \mid \text{the first and last characters in } x \text{ are different} \,\}$

b) $L_2 = \{\, x \in \{a,b\}^* \mid x \text{ the number of } b\text{'s in } x \text{ is an even number} \,\}$

c) $L_3 = \{\, x \in \{a,b,c\}^* \mid x \text{ contains at least one } c\text{, and there are no } a\text{'s before that } c \,\}$

d) $L_4 = \{\, x \in \{a,b,c\}^* \mid x \text{ contains at least one of the substrings } aaa \text{ or } bbb \,\}$

e) $L_5 = \{\, x \in \{a,b,c\}^* \mid x \text{ contains } \textbf{both} \text{ of the substrings } aaa \text{ and } bbb \,\}$

Answer:

a) $(a(a|b)^*b) \mid (b(a|b)^*a)$ — We want strings that start with one letter and end with the other. A string generated by this regular expression matches either $(a(a|b)^*b)$ or $(b(a|b)^*a)$. The first part gives strings that begin with a and end with b, and the second gives strings that begin with b and end with a. Between the first and last character, $(a|b)^*$ can generate any sequence of a's and b's.

b) $(a^*ba^*b)^*a^*$ — The b's in the string are generated in pairs, since the pattern a^*ba^*b can't generate a b without generating a second b. The fact that a^*ba^*b can be repeated zero or more times means that the total number of b's can be any even number. Any number of a's can occur between the b's. The a^* at the end allows any number of a's to follow the last b, or, if there are no b's at all, it allows for a string that only contains a's.

c) $b^*c(a|b|c)^*$ — The first c in this expression forces any string that it can generate to contain at least one c. Only b's can occur before the first c. After the first c, the rest of the string can be any string of a's, b's, and c's.

d) $(a|b|c)^*(aaa|bbb)(a|b|c)^*$ — The subexpression $aaa|bbb$ forces any string that this can generate to contain either aaa or bbb as a substring. The substring can be preceded or followed by arbitrary strings of a's, b's, and c's.

e) $\big((a|b|c)^*aaa(a|b|c)^*bbb(a|b|c)^*\big) \mid \big((a|b|c)^*bbb(a|b|c)^*aaa(a|b|c)^*\big)$ — The key here is that the two substrings aaa and bbb can occur in either order, so the regular expression must account for that. The first half of the expression generates strings that contain aaa followed at some point later in the string by bbb; any string can occur before, between, and after those substrings. The second half of the expression is similar, except that the aaa substring occurs somewhere after the bbb substring. Another expression for the same language is $(a|b|c)^*\big(aaa(a|b|c)^*bbb \mid bbb(a|b|c)^*aaa\big)(a|b|c)^*$.