

Recall from lecture that a *regular expression* is compact notation for a language (that is, a set of strings). Formally, a regular language is one of the following:

- The symbol $\emptyset$ (representing the empty set)
- Any string (representing the set containing only that string)
- $R + S$ for some regular expressions R and S (representing alternation / union)
- $R \cdot S$ or RS for some regular expressions R and S (representing concatenation)
- R^* for some regular expression R (representing Kleene closure / unbounded repetition)

In the absence of parentheses, Kleene closure has highest precedence, followed by concatenation. For example:

$$O + RE^* = O + R(E^*) = \{O, R, RE, REE, REEE, REEEE, REEEEEE, REEEEEEE, REEEEEEEE, \dots\}$$

$$O + (RE)^* = \{\epsilon, O, RE, RERE, RERERE, RERERERE, RERERERERE, RERERERERERE, \dots\}$$

$$(O + R)E^* = \{O, R, OE, RE, OEE, REE, OEEE, REEE, OEEEE, REEEE, OEEEEEE, \dots\}$$

$$(O + RE)^* = \{\epsilon, O, OO, RE, OOO, ORE, REO, OOOO, OORE, OREO, REOO, RERE, \dots\}$$

Give regular expressions for each of the following languages over the binary alphabet $\{0, 1\}$. (For extra practice, find multiple regular expressions for each language.)

- o. All strings.
1. All strings containing the substring 000 .
2. All strings *not* containing the substring 000 .
3. All strings in which every run of 0 s has length at least 3.
4. All strings in which the *last* 1 appears before the *first* substring 000 .
5. All strings containing at least three 0 s.
6. Every string except 000 . [Hint: Don't try to be clever.]

More difficult problems to work on later:

7. All strings w such that *in every prefix of w* , the number of 0 s and 1 s differ by at most 1.
8. All strings containing at least two 0 s and at least one 1 .
9. All strings in which every run has odd length. (For example, 0001 and 100000111 and the empty string ϵ are in this language, but 000000 and 001000 are not.)
- ★10. All strings w such that *in every prefix of w* , the number of 0 s and 1 s differ by at most 2.
- ★11. All strings in which the substring 000 appears an even number of times. (For example, 01100 and 000000 and the empty string ϵ are in this language, but 00000 and 001000 are not.)