Give regular expressions for each of the following languages over the alphabet $\{\mathtt{0}, \mathtt{1}\}$.

1. All strings containing the substring $\mathtt{000}$.

   Solution: $(\mathtt{0}+\mathtt{1})^*\mathtt{000}(\mathtt{0}+\mathtt{1})^*$ ■

2. All strings *not* containing the substring $\mathtt{000}$.

   Solution: $(\mathtt{1}+\mathtt{01}+\mathtt{001})^*(\varepsilon+\mathtt{0}+\mathtt{00})$ ■

   Solution: $(\varepsilon+\mathtt{0}+\mathtt{00})(\mathtt{1}(\varepsilon+\mathtt{0}+\mathtt{00}))^*$ ■

3. All strings in which every run of $\mathtt{0}$s has length at least 3.

   Solution: $(\mathtt{1}+\mathtt{0000}^*)^*$ ■

   Solution: $(\varepsilon+\mathtt{1})((\varepsilon+\mathtt{0000}^*)\mathtt{1})^*(\varepsilon+\mathtt{0000}^*)$ ■

4. All strings in which $\mathtt{1}$ does not appear after a substring $\mathtt{000}$.

   Solution: $(\mathtt{1}+\mathtt{01}+\mathtt{001})^*\mathtt{0}^*$ ■

5. All strings containing at least three $\mathtt{0}$s.

   Solution: $(\mathtt{0}+\mathtt{1})^*\mathtt{0}(\mathtt{0}+\mathtt{1})^*\mathtt{0}(\mathtt{0}+\mathtt{1})^*\mathtt{0}(\mathtt{0}+\mathtt{1})^*$ ■

   Solution (clever): $\mathtt{1}^*\mathtt{01}^*\mathtt{01}^*\mathtt{0}(\mathtt{0}+\mathtt{1})^*$ or $(\mathtt{0}+\mathtt{1})^*\mathtt{01}^*\mathtt{01}^*\mathtt{01}^*$ ■

6. Every string except $\mathtt{000}$. *[Hint: Don't try to be clever.]*

   Solution: Every string $w \neq \mathtt{000}$ satisfies one of three conditions: Either $|w| < 3$, or $|w| = 3$ and $w \neq \mathtt{000}$, or $|w| > 3$. The first two cases include only a finite number of strings, so we just list them explicitly. The last case includes *all* strings of length at least 4.

   $$\begin{aligned}&\varepsilon+\mathtt{0}+\mathtt{1}+\mathtt{00}+\mathtt{01}+\mathtt{10}+\mathtt{11}\\ &+\mathtt{001}+\mathtt{010}+\mathtt{011}+\mathtt{100}+\mathtt{101}+\mathtt{110}+\mathtt{111}\\ &+(\mathtt{1}+\mathtt{0})(\mathtt{1}+\mathtt{0})(\mathtt{1}+\mathtt{0})(\mathtt{1}+\mathtt{0})(\mathtt{1}+\mathtt{0})^*\end{aligned}$$

   ■

   Solution (clever): $\varepsilon+\mathtt{0}+\mathtt{00}+\big(\mathtt{1}+\mathtt{01}+\mathtt{001}+\mathtt{000}(\mathtt{1}+\mathtt{0})\big)(\mathtt{1}+\mathtt{0})^*$ ■

7. All strings w such that *in every prefix of* w, the number of 0s and 1s differ by at most 1.

Solution: Equivalently, strings that alternate between 0s and 1s: $(01+10)^*(\varepsilon+0+1)$ ■

*8. All strings containing at least two 0s and at least one 1.

Solution: There are three possibilities for how such a string can begin:

- Start with 00, then any number of 0s, then 1, then anything.
- Start with 01, then any number of 1s, then 0, then anything.
- Start with 1, then a substring with exactly two 0s, then anything.

All together: $000^*1(0+1)^* \;+\; 011^*0(0+1)^* \;+\; 11^*01^*0(0+1)^*$
Or equivalently: $\big(000^*1 \;+\; 011^*0 \;+\; 11^*01^*0\big)(0+1)^*$ ■

Solution: There are three possibilities for how the three required symbols are ordered:

- Contains a 1 before two 0s: $(0+1)^*\,1\,(0+1)^*\,0\,(0+1)^*\,0\,(0+1)^*$
- Contains a 1 between two 0s: $(0+1)^*\,0\,(0+1)^*\,1\,(0+1)^*\,0\,(0+1)^*$
- Contains a 1 after two 0s: $(0+1)^*\,0\,(0+1)^*\,0\,(0+1)^*\,1\,(0+1)^*$

So putting these cases together, we get the following:

$$\begin{aligned}&(0+1)^*\,1\,(0+1)^*\,0\,(0+1)^*\,0\,(0+1)^*\\ +\;&(0+1)^*\,0\,(0+1)^*\,1\,(0+1)^*\,0\,(0+1)^*\\ +\;&(0+1)^*\,0\,(0+1)^*\,0\,(0+1)^*\,1\,(0+1)^*\end{aligned}$$ ■

Solution (clever): $(0+1)^*\left(101^*0+010+01^*01\right)(0+1)^*$ ■

*9. All strings w such that *in every prefix of* w, the number of 0s and 1s differ by at most 2.

Solution: $\big(0(01)^*1+1(10)^*0\big)^*\cdot\big(\varepsilon+0(01)^*(0+\varepsilon)+1(10)^*(1+\varepsilon)\big)$ ■

★10. All strings in which the substring `000` appears an even number of times.
(For example, `0001000` and `0000` are in this language, but `00000` is not.)

Solution: Every string in $\{\mathtt{0},\mathtt{1}\}^*$ alternates between (possibly empty) blocks of `0`s and individual `1`s; that is, $\{\mathtt{0},\mathtt{1}\}^* = (\mathtt{0}^*\mathtt{1})^*\mathtt{0}^*$. Trivially, every `000` substring is contained in some block of `0`s. Our strategy is to consider which blocks of `0`s contain an even or odd number of `000` substrings.

Let X denote the set of all strings in $\mathtt{0}^*$ with an even number of `000` substrings. We easily observe that $X = \{\mathtt{0}^n \mid n = 1 \text{ or } n \text{ is even}\} = \mathtt{0} + (\mathtt{00})^*$.

Let Y denote the set of all strings in $\mathtt{0}^*$ with an *odd* number of `000` substrings. We easily observe that $Y = \{\mathtt{0}^n \mid n > 1 \text{ and } n \text{ is odd}\} = \mathtt{000}(\mathtt{00})^*$.

We immediately have $\mathtt{0}^* = X + Y$ and therefore $\{\mathtt{0},\mathtt{1}\}^* = ((X+Y)\mathtt{1})^*(X+Y)$.

Finally, let L denote the set of all strings in $\{\mathtt{0},\mathtt{1}\}^*$ with an even number of `000` substrings. A string $w \in \{\mathtt{0},\mathtt{1}\}^*$ is in L if and only if an even number of blocks of `0`s in w are in Y; the remaining blocks of `0`s are all in X. We consider two simple subcases to help us make sure we don't make mistakes.

- Exactly zero blocks of `0`s are in Y. This means all blocks of `0`s are in X and they alternate with blocks of `1`s. We capture this by $(X\mathtt{1})^*X$ where the last X is to allow ending in blocks of `0`s.
- Exactly two blocks of `0`s are in Y. The expression $(X\mathtt{1})^*Y\mathtt{1}\cdot(X\mathtt{1})^*Y(\varepsilon+\mathtt{1}(X\mathtt{1})^*X)$ captures this where we are paying attention in the expression to what comes after the second block of Y; either we end in that block or we need a `1` to separate a string that has zero blocks from Y.

Now we consider the general case with an even number of blocks in Y which can be zero, two or a multiple of two that can be simulated by repetitions. We can obtain this by considering the expression for exactly two blocks and allowing concatenation arbitrary number of times while being careful to ensure that we separate with a `1` as needed and allowing for the expression to end appropriately. The expression $\left((X\mathtt{1})^*Y\mathtt{1}\cdot(X\mathtt{1})^*Y\mathtt{1}\right)^*$ allows an arbitrary even number of blocks in Y with the caveat that it always ends with $Y\mathtt{1}$ if it has non-zero repetitions. We need to allow ending in Y or $(X\mathtt{1})^*X$ as in the second subcase above. Letting $Z = (X\mathtt{1})^*Y\mathtt{1}\cdot(X\mathtt{1})^*Y$ helps us simplify the final expression.

Putting the preceding observations together we obtain the following expression.

$$L = (Z\mathtt{1})^*(Z + (X\mathtt{1})^*X)$$

We will not plug in the expressions for X and Y because it will make the expression too long. There are likely to be shorter expressions based on better case analysis but debugging them is probably not worth it unless it is very important in some application.

Whew! ■