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

1. All strings containing the substring 000.

   Solution: $(0+1)^*000(0+1)^*$ ■

2. All strings *not* containing the substring 000.

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

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

3. All strings in which every run of 0s has length at least 3.

   Solution: $(1+0000^*)^*$ ■

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

4. All strings in which 1 does not appear after a substring 000.

   Solution: $(1+01+001)^*0^*$ ■

5. All strings containing at least three 0s.

   Solution: $(0+1)^*0(0+1)^*0(0+1)^*0(0+1)^*$ ■

   Solution (clever): $1^*01^*01^*0(0+1)^*$ or $(0+1)^*01^*01^*01^*$ ■

6. Every string except 000. *[Hint: Don't try to be clever.]*

   Solution: Every string $w \neq 000$ satisfies one of three conditions: Either $|w| < 3$, or $|w| = 3$ and $w \neq 000$, 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.

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

■

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

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

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

*8. All strings containing at least two $\mathtt{0}$s and at least one $\mathtt{1}$.

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

- Start with $\mathtt{00}$, then any number of $\mathtt{0}$s, then $\mathtt{1}$, then anything.
- Start with $\mathtt{01}$, then any number of $\mathtt{1}$s, then $\mathtt{0}$, then anything.
- Start with $\mathtt{1}$, then a substring with exactly two $\mathtt{0}$s, then anything.

All together: $\mathtt{000}^*\mathtt{1}(\mathtt{0}+\mathtt{1})^* \;+\; \mathtt{011}^*\mathtt{0}(\mathtt{0}+\mathtt{1})^* \;+\; \mathtt{11}^*\mathtt{01}^*\mathtt{0}(\mathtt{0}+\mathtt{1})^*$

Or equivalently: $\big(\mathtt{000}^*\mathtt{1} \;+\; \mathtt{011}^*\mathtt{0} \;+\; \mathtt{11}^*\mathtt{01}^*\mathtt{0}\big)(\mathtt{0}+\mathtt{1})^*$ ■

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

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

So putting these cases together, we get the following:

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

Solution (clever): $(\mathtt{0}+\mathtt{1})^*\big(\mathtt{101}^*\mathtt{0}+\mathtt{010}+\mathtt{01}^*\mathtt{01}\big)(\mathtt{0}+\mathtt{1})^*$ ■

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

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

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

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

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

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

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

Finally, let L denote the set of all strings in $\{\texttt{0},\texttt{1}\}^*$ with an even number of $\texttt{000}$ substrings. A string $w \in \{\texttt{0},\texttt{1}\}^*$ is in L if and only if an odd number of blocks of $\texttt{0}$s in w are in Y; the remaining blocks of $\texttt{0}$s are all in X.

$$L = \big((X\texttt{1})^*Y\texttt{1}\cdot(X\texttt{1})^*Y\texttt{1}\big)^*(X\texttt{1})^*X$$

Plugging in the expressions for X and Y gives us the following regular expression for L:

$$\Big(((\texttt{0}+(\texttt{00})^*)\texttt{1})^*\cdot\texttt{000}(\texttt{00})^*\texttt{1}\cdot((\texttt{0}+(\texttt{00})^*)\texttt{1})^*\cdot\texttt{000}(\texttt{00})^*\texttt{1}\Big)^*\cdot((\texttt{0}+(\texttt{00})^*)\texttt{1})^*\cdot(\texttt{0}+(\texttt{00})^*)$$

Whew! ■