Exercise 1: Bookkeeping. Indicate in a sentence or two how much time you spent on this homework, how difficult you found it subjectively, and what you found to be the hardest part. Tell me something about yourself that I do not already know. Any non-empty answer will receive full credit.

Also, if your opinions have changed since the last assignment, indicate one thing you like about the class so far and one thing you would change about it.

Exercise 2: Induction. Prove by induction the following statement about the denotational semantics of IMP. For any BExp b and any initial state \( \sigma \) such that \( \sigma(x) \) is even, if

\[
[\textbf{while } b \textbf{ do } x := x + 2 ] \sigma \Rightarrow \sigma'
\]

and if \( \sigma'(x) \neq \bot \), then \( \sigma'(x) \) is even. Unlike in the previous assignment, this time you should use denotational semantics for the proof.

If you wish, you may consider the simplified language as in class in which case the problem can be rephrased as follows: for any BExp b and any initial state \( x \) such that \( x \) is even, if

\[
[\textbf{while } b \textbf{ do } x := x + 2 ] \Rightarrow x = x'
\]

and if \( x' \neq \bot \), then \( x' \) is even.

Hint: your proof should proceed by mathematical induction.
Exercise 3: Regular Expressions are commonly used as abstractions for string matching. Here is an abstract grammar for regular expressions:

\[
\begin{align*}
  r &::= "x" \quad \text{singleton – matches the character } x \\
  & \mid \text{empty} \quad \text{skip – matches the empty string} \\
  & \mid r_1 \ r_2 \quad \text{concatenation – matches } r_1 \text{ followed by } r_2 \\
  & \mid r_1 \ | \ r_2 \quad \text{or – matches } r_1 \text{ or } r_2 \\
  & \mid r^* \quad \text{Kleene star – matches 0 or more occurrences of } r \\
  & \mid \text{.} \quad \text{matches any single character} \\
  & \mid ["x" - "y"] \quad \text{matches any character between } x \text{ and } y \text{ inclusive} \\
  & \mid r^+ \quad \text{matches 1 or more occurrences of } r \\
  & \mid r^? \quad \text{matches 0 or 1 occurrence of } r
\end{align*}
\]

We will call the first five cases the primary forms of regular expressions. The last four cases can be defined in terms of the first five. We also give an abstract grammar for strings (modeled as lists of characters):

\[
\begin{align*}
  s &::= \text{nil} \quad \text{empty string} \\
  & \mid "x" :: s \quad \text{string with first character } x \text{ and other characters } s
\end{align*}
\]

We write “bye” as shorthand for “b :: “y” :: “e” :: nil. This exercise requires you to give big-step operational semantics rules of inference related to regular expressions matching strings. We introduce the following judgment:

\[\vdash r \text{ matches } s \text{ leaving } s'\]

The interpretation of the judgment is that the regular expression \( r \) matches some prefix of the string \( s \), leaving the suffix \( s' \) unmatched. If \( s' = \text{nil} \), then \( r \) matched \( s \) exactly. For example,

\[\vdash "h"("e"+) \text{ matches } \text{"hello" leaving } \text{"llo"\}]

Note that this operational semantics may be considered non-deterministic because we expect to be able to derive all three of the following:

\[\vdash ("h" \ | \ "e")^* \text{ matches } \text{"hello" leaving } \text{"ello" \}]}\]

\[\vdash ("h" \ | \ "e")^* \text{ matches } \text{"hello" leaving } \text{"hello" \}]}\]

\[\vdash ("h" \ | \ "e")^* \text{ matches } \text{"hello" leaving } \text{"llo" \}]}\]

Here are two rules of inference:

\[
\begin{align*}
  s &= "x" :: s' \\
  \vdash "x" \text{ matches } s \text{ leaving } s' \\
  \vdash \text{empty matches } s \text{ leaving } s
\end{align*}
\]

Give big-step operational semantics rules of inference for the other three primal regular expressions.
Exercise 4: We will use denotational semantics to model the fact that a regular expression can match a string leaving many possible suffixes. Let $S$ be the set of all strings, let $\mathcal{P}(S)$ be the powerset of $S$, and let $\text{RE}$ range over regular expressions. We introduce a semantic function:

$$\mathcal{R} : \text{RE} \rightarrow (S \rightarrow \mathcal{P}(S))$$

The interpretation is that $\mathcal{R}(r)$ is a function that takes in a string-to-be-matched and returns a set of suffixes. We might intuitively define $\mathcal{R}$ as follows:

$$\mathcal{R}[\left[ r \right]](s) = \{ s' \mid \vdash r \text{ matches } s \text{ leaving } s' \}$$

In general, however, one should not define the denotational semantics in terms of the operational semantics. Here are two correct semantic functions:

$$\mathcal{R}[\left[ \text{“x”} \right]](s) = \{ s' \mid s = \text{“x”} :: s' \}$$
$$\mathcal{R}[\left[ \text{empty} \right]](s) = \{ s \}$$

Give the denotational semantics functions for the other three primal regular expressions. Your semantics functions may not reference the operational semantics.

Exercise 5: We want to update our operational semantics for regular expressions to capture multiple suffixes. We want our new operational semantics to be deterministic—it should give the same the same answer as the denotational semantics above. We introduce a new judgment as follows:

$$\vdash r \text{ matches } s \text{ leaving } S$$

And use rules of inference like the following:

$$\vdash \text{“x” matches } s \text{ leaving } \{ s' \mid s = \text{“x”} :: s' \} \quad \vdash \text{empty matches } s \text{ leaving } \{ s \}$$
$$\vdash r_1 \text{ matches } s \text{ leaving } S \quad \vdash r_2 \text{ matches } s \text{ leaving } S'$$
$$\vdash r_1 \mid r_2 \text{ matches } s \text{ leaving } S \cup S'$$

You must do one of the following:

- either give operational semantics rules of inference for $r*$ and $r_1r_2$. Your operational semantics rules may not reference the denotational semantics. You may not place a derivation inside a set constructor, as in: $\{ x \mid \exists y. \vdash r \text{ matches } x \text{ leaving } y \}$. Each inference rule must have a finite and fixed set of hypotheses.

- or argue in one or two sentences that it cannot be done correctly in the given framework. Back up your argument by presenting two attempted but “wrong” rules of inference and show that each one is either unsound or incomplete with respect to our intuitive notion of regular expression matching.
Part of doing research in any area is getting stuck. When you get stuck, you must be able to recognize whether “you are just missing something” or “the problem is actually impossible”.

**Exercise 6:** Download the Homework 3 code pack from the course webpage. The README.txt describes the code pack, like in Homework 2. Write an interpreter for regular expressions based on the denotational semantics. In particular, you must write a function

\[
\text{matches : } \text{RE} \rightarrow S \rightarrow \mathcal{P}(S)
\]

in hw3.ml. Your interpreter must handle all of the regular expression forms, not just the primal ones. The Makefile includes a “make test” target that you should use (at least) to test your work. Modify the example.re file to include your best test case. Rename hw3.ml to your_last_name-hw3.ml and rename example.re to your_last_name-example.re for submission. Do not modify any other files.

Note that in the concrete syntax for RE, we use the " symbol to delimit strings and the ' symbol to delimit characters.