SI 413 Fall 2021: Homework 5

Due Monday, September 27

Your name:

Citations and collaborators:

Comments, suggestions, or questions for your instructor:

Fill out the first row of the table on a 0-5 scale before turning in.

This rubric is also available on the website under "Admin":
- **5**: Solution is completely correct, concisely presented, and neatly written.
- **4**: The solution is mostly correct, but one or two minor details were missed, or the presentation could be better.
- **3**: The main idea is correct, but there are some significant mistakes. The presentation is somewhat sloppy or confused.
- **2**: A complete effort was made, but the result is mostly incorrect.
- **1**: The beginning of an attempt was made, but the work is clearly incomplete.
- **0**: Not submitted.

Problem	1	2	3	4	Total
Self-assessment					
Final assessment					

1 Homographs and Synonyms

Pick a pair of two programming languages that you know, and come up with an example of each of the following in your two languages. As always, you can work together, but everyone must turn in unique examples.

a) A **homograph** is a code fragment that is the same *syntactically* between the two languages, but has different *semantics* in each.

b) A **synonym** is a code fragment that is the same *semantically* between the two languages, but has different *syntax*.

2 Scanner DFA

C++ and Java support a few different kinds of numerical constants, or "literals". The most basic are regular ints that you know and love like `15`, `256`, or `32`. There are also floating-point numbers like `3.7` or `.0684`.

For this problem, consider an `INT` token to be any sequence of 1 or more digits `[0-9]`, and a `FLOAT` token to be any sequence of 1 or more digits which contains exactly one decimal point `[.]`.

Draw the DFA for a scanner that accepts `FLOAT` and `INT` tokens. Be sure to label each accepting state with the type of token, and put characters or character ranges on each transition.

3 Bigger Scanner DFA

a) Modify your scanner DFA from the previous problem so that it also accepts an additional type of token, a `HEX` constant such as `0x3a5` or `0x7`.

For this problem, a `HEX` token contains the symbols `0x` followed by zero or more digits or letters in the range `a` through `f`.

b) Note that the previous definition allows for the string `0x` by itself to be considered a `HEX` token. What problem would there be if we disallowed this, so that `0x` is not a valid token but, for example, `0x3` is valid?

4 Ambiguous Grammar

Write a grammar that is ambiguous, and then show that it is ambiguous by coming up with a series of tokens that could be parsed in two different ways according to your grammar.