HW 1: Lexical Analysis

CSC 4351, Spring 2014

Due: 17 February 2014

1. Consider the following format for floating point literals. Assume a fraction part consisting of a string of one or more decimal digits with an optional decimal point (period) that can be added immediately before or after *any* of the digits. In addition, an optional exponent can be appended to the fraction part, consisting of an e or E, followed by an optional sign, followed by one or more decimal digits. A floating point literal must contain either a decimal point, or an exponent, or both.

   Examples: `.0, 0., 0.1, .01e-01, 01E00`

   (a) Give a regular expression for this construct. You may use ϵ.

   (b) Use Thompson's construction to translate the regular expression into an NFA.

   Make sure that you don't accept integers!

2. Consider scanning literals for representing carbohydrates in chemistry. Such literals consist of one or more *parts*, each consisting of C, H, or O, optionally followed by a decimal integer greater than 1 (it can greater than 9). You may use D_0 to mean decimal digits 0–9, D_1 to mean 1–9, and D_2 to mean 2–9.[1]

   (a) Give a regular expression for these carbohydrate literals. You may use the iteration operators * and + and ?.

   (b) Give a DFA that recognizes exactly these literals (no ϵ edges). Don't forget to mark the start state and all accepting states.

   Make sure that you don't accept the empty string!

---

[1]This is a computer science question, not a chemistry question. In particular, this specification allows bogus chemical formulas such as H_{42}.