CS4300 Compiler Theory

First Project

Create a (f)lex program that will recognize numbers.

For the purposes of this exercise, define a number to be:

1. A finite string of decimal digits, with an optional leading "-", without leading zeros, containing at least one digit. If there is only one digit, that digit may be a zero.

   or

2. A finite string of decimal digits containing at least one digit, without leading zeros, with an optional leading "-", followed by a decimal point, followed by a finite string of decimal digits containing at least one digit. If there is only one digit preceding the decimal point, that digit may be a zero.

   or

3. A number of the form identified in 1. or 2., followed by an "e" or an "E", followed by a number of the form identified in 1.

   or

4. A finite string of decimal digits "0" through "7", with an optional leading "-", with exactly one leading zero, containing at least two digits. If there are exactly two digits, they may both be zeros.

   or

5. A finite string of hexadecimal (upper case only) digits containing at least one hexadecimal digit, with no leading zeros, preceded by the string "0x" or "0X", optionally preceded by a leading "-". If there is only one hexadecimal digit, that digit may be zero.

No numbers may contain white-space.

Assume that the input to your program will consist only of (ascii) digits, letters (upper or lower case), spaces, decimal points, "-", and new-lines.

The output of your program should be a report, telling whether "possible numbers" are actually numbers or not. A "possible number" is a string of non-white-space characters preceded by beginning-of-file, a space, or a new-line, and followed by a space, a new-line, or end-of-file. Your program should also report the number possible numbers, the number of numbers found, and the number of "possible numbers" that were not numbers.

Your group should submit two files through the cs homework submission system: a (f)lex program (named numbers.l), and a test data file (which you should name test-data.txt) containing appropriate test data to confirm that your program does the right thing.

Example:

If the input file contains

123 123a 0123 12.3 12.03 -123 0 12.3e-5 0x123 0xabc 0XA9B 0.12E5 -1.0023e4 -0xA98F 0XA98H 0.123e-18 0128 00123 goodbye

Your program output should contain the report:

123 number
123a not number
0123 number
12.3 number
12.03 number
-123 number
0 number
12.3e-5 number
0x123 number
0xabc not number
0XA9B number
0.12E5 number
-1.0023e4 number
-0xA98F number
0XA98H not number
0.123e-18 number
0128 not number
00123 not number
goodbye not number

Out of 19 possible numbers, there were 13 numbers, and 6 not numbers.

Note: my example does not cover all possible cases :-)

Note: your (f)lex program should contain appropriate comments, to at least identify who should receive credit ...