



## CS 280: Program #2

Spring 2015

Due March 6 (NOTE DATE CHANGE FROM SYLLABUS)

It's sometimes useful to make a little language for a simple problem. We are making a simple language to let us play with strings, and for this assignment we are building a lexical analyzer for this simple language.

1. The language has identifiers. An identifier starts with a letter and is followed by zero or more letters.
2. The language has integer constants. An integer constant is a sequence of one or more digits.
3. The language has string constants. A string constant is a sequence of characters, all on one line, enclosed in double quotes.
4. The language has 4 operators: + - \* /
5. The language has two keywords: "set" and "print"
6. White space is used to separate tokens and lines for readability
7. Statements in the language end in a semicolon
8. A comment begins with two slashes (//) and ends at a newline

For this assignment I have provided the header file `p2lex.h`, which is on the course website. You **MUST** use this header file for your assignment. You may not change it.

For this assignment you should produce and hand in these two files:

- a source file that `#includes` `p2lex.h` and implements the `getToken()` function
- a source file that implements a main program to test the lexical analyzer.

Your main program should meet the following specifications:

1. The program should take an optional command line argument containing the name of an input file.
2. The program should read from the standard input if no command line arguments are passed.
3. If the program is able to open the input file, it should read it by calling `getToken()` until it returns `DONE`
4. The program should keep a count of the number of times each token was seen, and print the result at the end of the program
5. The program should print out each unique ID that it finds after it is done printing out the counts