COS 397 Spring 2013 Class Exercise $4 \quad 11$ questions; 3 pgs. 2013-01-24

This exercise is based on the paper ${ }^{1}$ describing Lex.

1. List the members of your group below. Underline your name.
2. What is the smallest Lex program?
3. Explain the significance of yytext and yyleng.
4. Write a Lex program that yields a lexer that removes all words of length 4, 8, and 13 . Repeat using sed instead of Lex.
5. Write a Lex program that yields a lexer that removes all leading and trailing whitespace from each line of the input. Repeat using sed instead of Lex.

[^0]6. Depict a finite-state automaton that may be used internally by the lexer generated by Lex for the program of Question 5.
7. Provide the regular expression described as follows in Section 4 of the paper:

Consider a language which defines a string as a set of characters between quotation (") marks, and provides that to include a " in a string it must be preceded by a $\backslash$. The regular expression which matches that is somewhat confusing [...]
8. Provide Lex code for a lexer that replaces all occurrences of the two consecutive words 'the Who' with 'The Who' (note uppercase T). Repeat using sed instead of Lex.
9. Provide Lex code for a lexer that replaces all occurrences of 'the Who' with 'The Who' (as in Question 8) and also replaces all occurrences of 'Who ever' with 'Whoever' and 'who ever' with 'whoever'; make your program as concise as possible. Repeat using sed instead of Lex.
10. Provide Lex code that produces a lexer that interprets its input as a regular expression and that outputs Lex code for a lexer that outputs the matches to that regular expression in its input.
11. Comment on the asymmetry in Lex's treatment of left and right contexts. (What is it? Why is it so? What are the alternatives?)


[^0]:    ${ }^{1}$ Michael E. Lesk and Eric Schmidt, "Lex—A Lexical Analyzer Generator," in Andrew G. Hulme and M. Douglas McIlroy (eds.), UNIX Vol. II: research system, 10th edition (Philadelphia, Pennsylvania: W. B. Saunders Company, 1990).

