

CS440/540 Spring 2022

Assignment #1: Using Lex

Due date on blackboard

For this first assignment, you are going to use Lex to write a preprocessor to take input text and make it look presentable. **This is to be your own individual effort.** The output must adhere to the following rules:

1. The input text should be output as lines with length as close as possible to 60 characters **without exceeding this number**. Words should not be cut to meet this requirement.
 - Be careful to keep special symbols that can appear at the end of a word (period, question mark, exclamation point, semicolon, comma, colon, close parenthesis, close quote), symbols that can appear within words (apostrophe, @ and .), and symbols that can appear at the start of a word (open parenthesis, open quote) on the same line as the word.
 - If a word contains a hyphen, that hyphen should be used to break a word across lines IF the word is too long for the line and the part of the word up to and including the hyphen fits.
2. In the input text, paragraphs are separated by one or more blank lines. For the input file, a blank line is defined as zero or more blanks and tabs followed by an end of line marker (n). In the output text, there must be a single line separating paragraph. In addition, the first line of each paragraph should be indented 5 characters.
3. No line should begin with blanks unless it is the start of a paragraph.
4. Multiple blanks should be transformed to a single blank.
5. A line starting with the '*' character indicates a list item. Everything on the line is part of the given list item. A list item starts with the * character indented 5 spaces followed by a single space then the item. The item itself needs to be cut across lines adhering to the 60 character rule. The second and following lines will be indented 7 spaces shown in the online examples. There will be a blank line between list items (in the input) and between list items and paragraphs.
6. Any text following the '%' symbol to the end of the line should be removed.
7. Any other special characters that are not mentioned above should be removed.

Note – Be sure you are using Lex effectively. If you are writing a lot of C/Java code to manipulate the matched input, you are not making Lex do the work.

Submitting

Submit a single file named program1.l (if writing C code) or program1.jflex (if writing Java code) on blackboard. Your code must compile and run using one of the given makefiles.