Rae Jones

CSCE 240-H04

February 1, 2023

Project 1 – Extraction Documentation

I chose to use scabies for my final project. To do this, I copied the information on CDC's and webMD's websites into two separate text files. In order to format and gain information about these text files, I created code in Java. My program asks the user what disease they want to learn about and proceeds only if they answer "scabies". It then counts the lines, characters, and words in each original text file. All of this is written to a new output file for each source. The output file is also formatted with indentation.

CDC data source: https://wwwnc.cdc.gov/travel/diseases/scabies

WebMD data source:

https://www.webmd.com/skin-problems-and-treatments/ss/slideshow-scabies-overview

To help me understand PrintWrite, I used information from ProgramIz at

https://www.programiz.com/java-programming/printwriter.

To achieve the correct output files, I first added an IF statement that encapsulates the rest of my code. If the user states that they would like to learn about scabies, the code begins scanning the text files. If the user states otherwise, the code notifies the user and EXITS. When the code scans the ext files, it adds the data line-by-line to the output files. To indent correctly, it checks if the last char in the line is a "?" or a ":". If it is not, the line is indented by one tab. The

code also utilizes a FOR-LOOP that iterates through each char in the string to chars and find

spaces for word counting. The statistics for the original files are appended to the end of the

output files. The number of lines in the original data files and the new outputs do not match.