Vivian Do
CSC 4330
Nov 08 2023

Homework 7

1.

In the text:

a. (5 points). How many lines (not sentences, but lines) begin with "Harry"?

b. (5 points). How many lines contain the word "Harry"

c. (10 points). How many instances of the word "Harry" are there?

Note that there are instances of "Harry's" --- please count those as well.

Note that the count for (a.) will be less than (or equal to) the count for (b.), which will be less than (or equal to) the count of (c.).

a)
236 lines

b)
1084 lines

c)
903 instances
// I'm not sure if my code is correct. My results are different from using "Ctrl + F", which gives 1327 instances.

2. (20 points).

In the text, how many times does the word "had" appear?

*Note: be careful about the following examples of tricky sentences in the text:

Had that been what the thieves were looking for? (1 time)

And so he had. (1 time)

Hedwig hadn't brought Harry anything so far. (0 times)

Then a small voice came out of the shadows. (0 times)

Mrs. Dursley had had a nice, normal day. (2 times)

*Note: that you may do this in several steps if you wish, e.g., perform a regex whose result can go to some intermediate file, and then perform another regex on this intermediate file, etc.

691 times
// I used a similar code as 1c

3. (20 points).

Remember the regular language of 2. of Homework 7, the language with an odd number of a's? Now we want to generalize it to larger alphabets. In the text, how many words have an odd number of A's/a's?

*Note: watch out for contractions such as "can't" or "shouldn'ta" (Hagrid has somewhat of a thick accent :) as well as compound words such as "last-minute" or "best-played". There is even an instances of "what's-her-name", which has both "'" and "-" (but has an even number of A's/a's)

*Note: again, that you may do this in several steps if you wish, using any programming language and its features (not just regexes), e.g., Python's "split()" function to split the text into some intermediate file, perform a regex on this file to produce a second intermediate file, etc. See the working examples in the slides "regexes-in-practice.txt" for some inspiration.

24612 words

4. (20 points).

Remember the regular languages of 1. and 2. of Homework 7? That is:

a. The language composed of all strings over the alphabet {a,b} of the form: zero or more a's followed by zero or more b's, i.e., {"", a, b, aa, ab, abb, aaa, aab, abb, bbb, ...}, where "" is the empty string.

b. The language composed of all strings over the alphabet {a,b} that contain an odd number of a's.

Show that (a.) and (b.) are both context-free languages (10 points each), by giving a corresponding context-free grammar. Note that stating just the production rules is sufficient to describe the context-free grammar.

a)
S -> emptystring
S -> aS
S -> bS

- this rule derives any string that conforms to the language

b)
S -> bS | aSa | emptystring

- this rule creates strings with odd numbers of a’s

5. (20 points).

Consider the language composed of all strings over the alphabet {a,b} that contain twice as many b's as a's, i.e., {"", abb, bab, bba, aabbbb, ababbb, abbabb, abbbab, ...} (note that the empty string is there for completeness, i.e., twice as many as zero is zero). Show that this language is context-free by giving a corresponding context-free grammar.

S -> emptystring
S -> aSbS
S -> bSaS
S -> bSbS

- this rule creates strings which for every “a” followed by two “b”s or for every “b” followed by two “a”s the pattern will repeat