CS4248 Assignment 1: Regexs and Language Models

By A0000000X (Please change this as appropriate)

This is a sample writeup.pdf file, to illustrate the expected format. You may choose to use this source file but need not to. Just follow the instructions required in the Assignment 1 instructions.

1. **Declaration of Original Work**.

By entering my Student ID below, I certify that I completed my assignment independently of all others (except where sanctioned during in-class sessions), obeying the class policy outlined in the introductory lecture. In particular, I am allowed to discuss the problems and solutions in this assignment, but have waited at least 30 minutes by doing other activities unrelated to class before attempting to complete or modify my answers as perthe Pokémon Go rule.

Signed, A000000X

2. **References**.

I give credit where credit is due. I acknowledge that I used the following websites or contacts to complete this assignment

- Regex Golf: https://alf.nu/RegexGolf for practicing regular expressions.

Part 1

Objective 1 — Tokenization, Zipf's law

A. Screenshot below.

   Answer here.

B. Plot here.

   Justification here.

C. Plot here.

Justification here.

D. Plot here.

   Justification here.

Objective 2 — How's the Weather?

Replace this with a high level description of my regular expression strategy to distinguish true requests for weather from other user inputs.

Objective 3 — Language Modeling

Here are some of the captured output conversations with respect to the generate_word, generate_text, an perplexity calls using the text corpus `test_corpus.txt` (replace as needed).

...

If you implemented any additional LM variants, you're welcomed to give high level documentation here.

Part 2

1. **Subtraction Regular Expressions**

Write your answer here.

2. **Language Models**

   a. True or False.

      Justify your answer here.

   b. True or False.

      Justify your answer here.

c. True or False.

   Justify your answer here.