

## CS 288 2018S Section 006

### Homework 04

**Due:** At the beginning of class on Monday March 19<sup>th</sup>.

**Deliverables:** One text file, `regex.txt`, that contains the full `egrep` commands as typed at the Linux command prompt (12 commands in total).

1. The following questions refer to Edgar Allen Poe's poem "The Raven", which is available for download on Moodle.

Use **egrep** to find the answers to the following case-sensitive questions:

- a) The lines containing, but don't start with, the word **raven**.
- b) The blank lines.
- c) The lines with at least 40 characters.
- d) The nonempty lines with no punctuation characters.
- e) The lines with an uppercase letter ignoring the first character on each line.
- f) The lines that have at least two **p**'s (consecutive or nonconsecutive) followed by **ore**, not necessarily immediately after.
- g) The lines that have two consecutive words where the last letter in the first is the first letter in the second.
- h) The lines with 3 sets of double letters (the sets may or may not contain the same letters).
- i) The lines having 5 consecutive letters forming a palindrome.

*A word is defined as a series of letters with no space or punctuation characters in between.*

2. Give the three regular expressions required to produce the results:
  - a) Lines that contain **A** or **a** at some index  $m$ , and **Z** or **z** at some index  $n$ , where  $n > m$ .
  - b) Lines consisting solely of US postal code, allowing both the five-digit and nine-digit (called ZIP+4) formats.
  - c) Lines containing nothing but an integer between **0** and **211**, inclusive.