David Schulz

CS 4980 001

3/14/2022

**Homework 1**

1. Write regular expressions for the following languages
   1. The set of all alphabetic strings
      1. [A-Za-z]\*
   2. The set of all lower-case alphabetic strings ending in a *b*
      1. [a-z]\*b
   3. The set of all strings from the alphabet *a, b* such that each *a* is immediately preceded by and immediately followed by a *b*
      1. (b(ab)\*)\*
2. Write regular expressions for the following languages. By “word”, we mean an alphabetic string separated from other words by whitespace, any relevant punctuation, line breaks, and so forth.
   1. The set of all strings with two consecutive repeated words (e.g., “Humbert Humbert” and “the the” but not “the bug” or “the big bug”)
      1. (\w+)\s+\1
   2. All strings that start at the beginning of the line with an integer and that end at the end of the line with a word
      1. ^\d.\*\b[A-Za-z]+$
   3. All strings that have both the word grotto and the word raven in them (but not, e.g., words like grottos that merely contain the word grotto)
      1. ^(?=.\*\bgrotto\b)(?=.\*\braven\b).\*$
   4. Write a pattern that places the first word of an English sentence in a register. Deal with punctuation.
      1. ^([A-Za-z'.-]\*)
3. Edit Distance
   1. Compute the edit distance (using insertion cost 1, deletion cost 1, substitution cost 2) of “leda” to “deal”. Show your work (using the edit distance grid).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | # | l | e | d | a |
| # | 0 | 1 | 2 | 3 | 4 |
| d | 1 | 2 | 3 | 2 | 3 |
| e | 2 | 3 | 2 | 3 | 4 |
| a | 3 | 4 | 3 | 4 | 3 |
| l | 4 | 3 | 4 | 5 | 4 |

Edit distance: 4

* 1. Figure out whether drive is closer to brief or to divers and what the edit distance is to each.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | # | d | r | i | v | e |
| # | 0 | 1 | 2 | 3 | 4 | 5 |
| b | 1 | 2 | 3 | 4 | 5 | 6 |
| r | 2 | 3 | 2 | 3 | 4 | 5 |
| i | 3 | 4 | 3 | 2 | 3 | 4 |
| e | 4 | 5 | 4 | 3 | 4 | 3 |
| f | 5 | 6 | 5 | 4 | 5 | 4 |

Drive -> Brief Distance: 4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | # | d | r | i | v | e |
| # | 0 | 1 | 2 | 3 | 4 | 5 |
| d | 1 | 0 | 1 | 2 | 3 | 4 |
| i | 2 | 1 | 2 | 1 | 2 | 3 |
| v | 3 | 2 | 3 | 2 | 1 | 2 |
| e | 4 | 3 | 4 | 3 | 2 | 1 |
| r | 5 | 4 | 3 | 4 | 3 | 2 |
| s | 6 | 5 | 4 | 5 | 4 | 3 |

Drive -> Divers Distance: 3

Divers is closer.

* 1. Now implement a minimum edit distance algorithm in Java and use your hand-computed results to check your code.