Natural Language Processing
Lab 1

January 15, 2024

Dr Phil Smith

This lab sheet is to practice the concepts taught so far, which is just regular expressions. This is a good opportunity to have a look at the reading too as this contains some more formalisms for regular expressions that are useful to know.

1. Write regular expressions for the following languages.

    1. The set of all alphabetic strings.
    2. The set of all lowercase alphabetic strings ending in a b.
    3. The set of all strings from the alphabet a,b such that each a is immediately preceded and followed by a b.

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"). You may use \s to match a whitespace character to make things clear.
    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. ou may use \b to match the empty string, but only when it is not at the beginning or end of a word.
    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*).
    4. Write a pattern that places the first word of an English sentence in a register. Elegantly deal with punctuation.

3. Implement an ELIZA-like program, using substitutions such as those described on page 10. You might want to choose a different domain than a Rogerian psychologist, although keep in mind that you would need a domain in which your program can legitimately engage in a lot of simple repetition.

```
[a-z]* b
```

```
([A-Za-z]+)\s\|
([1-9]+[0-9]*)\b([A-Za-z]+)
(^.*)grotto(^.*)|(^.*)raven(^.*)
([A-Za-z]+)[\.\?!]
```

1

(1) [a-zA-z]+ (* if include ∅)

(2) ^[a-z]*[ab]$

bb ab bab bab

bb abab abb b

(3) ab

ab ab bb ab

ab bbb ab

ab

(? ab)

ab

[ab]

$[ab]^+$

$S = \{a, b\}$

ab

$(ab^+)^+$

$[S]^*$

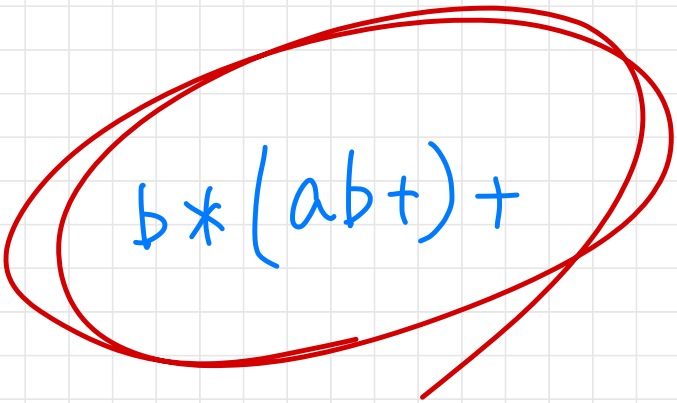
aa

$( )$ bb

ab (ab+)

abbb - - - abababab

((ab+)b*)+


b*(ab+)+

$ab\{1, n\}$

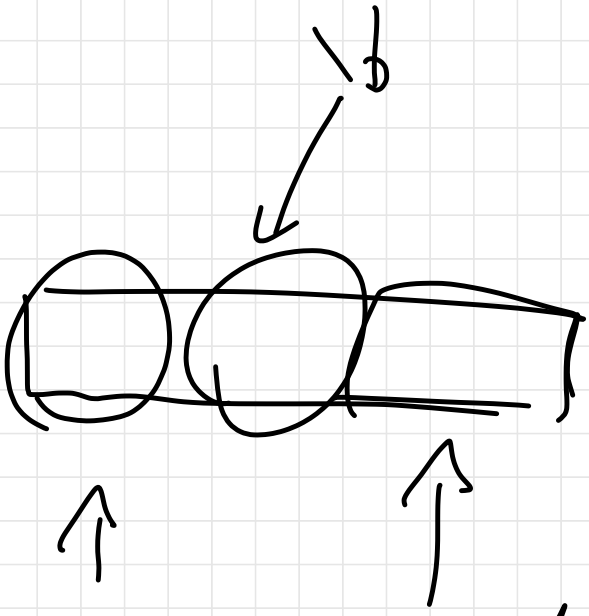
2

(1)

(.*)\s\1

(.*)(.*)\s\1\2

(2)



[

[+-][0-9]+ \b [a-zA-Z]+