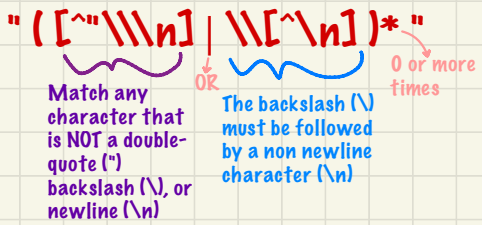
Homework 1

2.1) Write regular expressions to capture the following.

(a) Strings in C. These are delimited by double quotes ("), and may not contain newline characters. They may contain double-quote or back-slash characters if and only if those characters are "escaped" by a preceding backslash. You may find it helpful to introduce shorthand notation to represent any character that is not a member of a small specified set.

A regular expression for a string in C can be :



EBNF Syntax:

string —> quote valid_char* quote

valid_char —> (bkslsh bkslsh | bkslsh quote | not(invalid))

invalid —> bkslsh | quote | newline

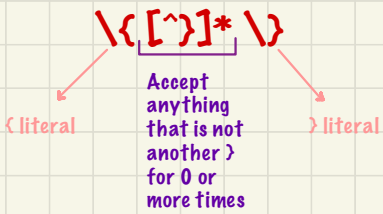
quote —> "

bkslsh —> \

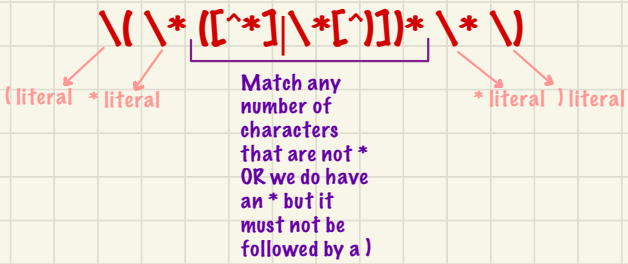
newline —> \n

(b) Comments in Pascal. These are delimited by (* and *) or by { and }. They are not permitted to nest.

A regular expression for comments of the form { } in Pascal can be:



A regular expression for comments of the form (* *) in Pascal can be:



EBNF Syntax:

comment —> open_b (not(close_b))* close_b
—> open_p star ((star not(close_p) | not(star))* star close_p

open_p —> (

close_p —>)

star —> *

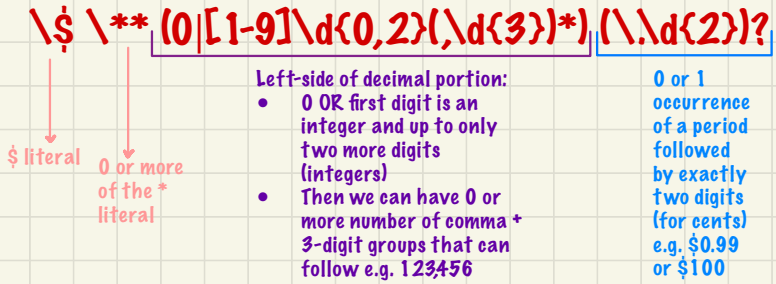
open_b —> {

close_b —> }

(f) Financial quantities in American notation. These have a leading dollar sign ($), an optional string of asterisks (*—used on checks to discourage fraud), a string of decimal digits, and an optional fractional part consisting of a decimal point (.) and two decimal digits. The string of digits to the left of the decimal point may consist of a single zero (0). Otherwise it must not start with a zero. If there are more than three digits to the left of the decimal point, groups of three (counting from the right) must be separated by commas (,). Example: $**2,345.67. (Feel free to use "productions" to define abbreviations, so long as the language remains regular.)

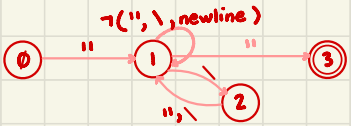
A regular expression for financials quantities in American notation can be:

(d is shorthand for integer)

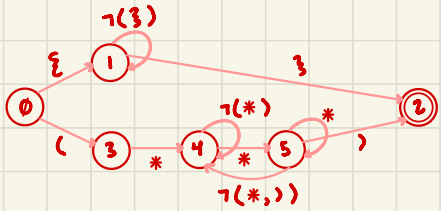


2.2) Show (as "circles-and-arrows" diagrams) the finite automata for Exercise 2.1.

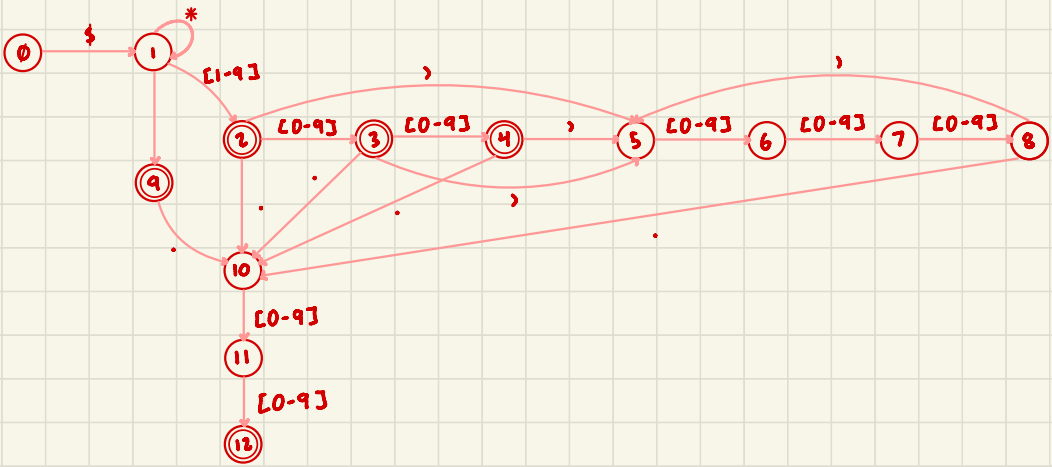
(a)



(b)



(c)



NOTE: A double circle indicates a final accepting state