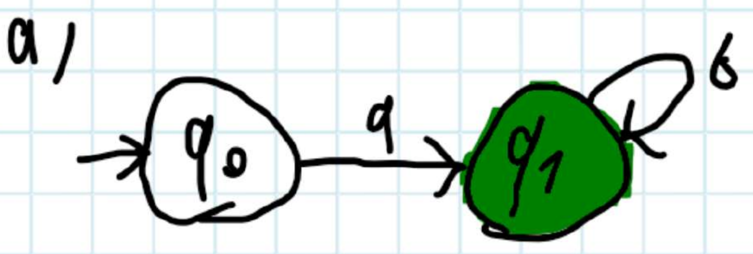
Карелин Константин М3104

Homework №6
Automata theory

1. For each given regular expression P, construct a DFA (Deterministic Finite Automaton), and find the number of accepted word of length at most 5, i.e. the size of the set $\mathcal{L}' = \{w \in \mathcal{L}(P) \mid |w| \le 5\}$. For "any" (`.`) and "negative" (`[^.]`) matches, assume that the alphabet is $\Sigma = \{a, b, c, d\}$.
   (a) P_1 = `ab*`
   (b) P_2 = `a+b?c`
   (c) P_3 = `[^cd]+c{3}`
   (d) P_4 = `[^a](.|ddd)?`
   (e) P_5 = `d(a|bc)*`
   (f) P_6 = `((a|ab)[cd]){2}`

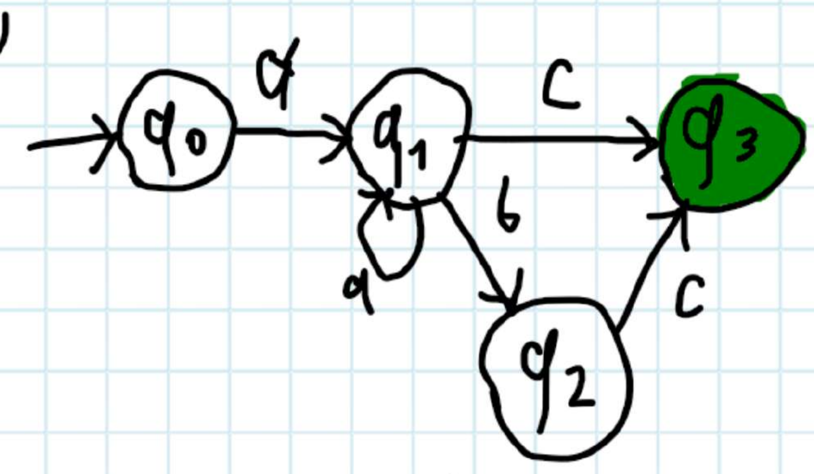
● — терминальная

a)



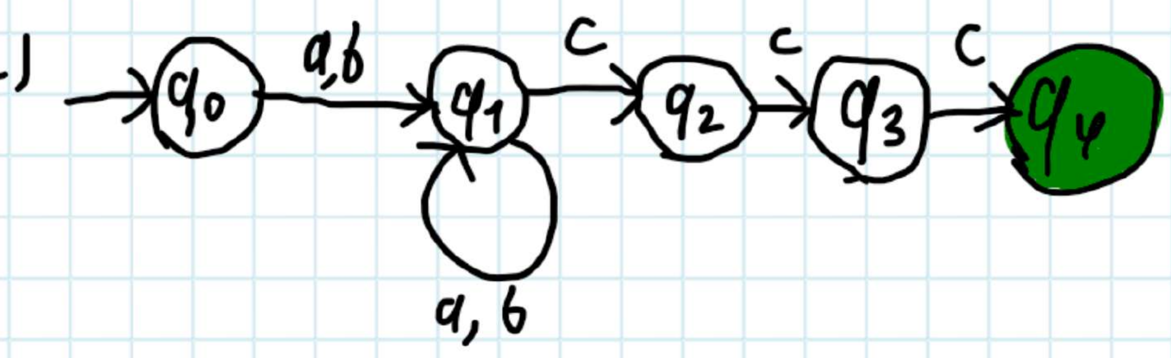
5 words

b)



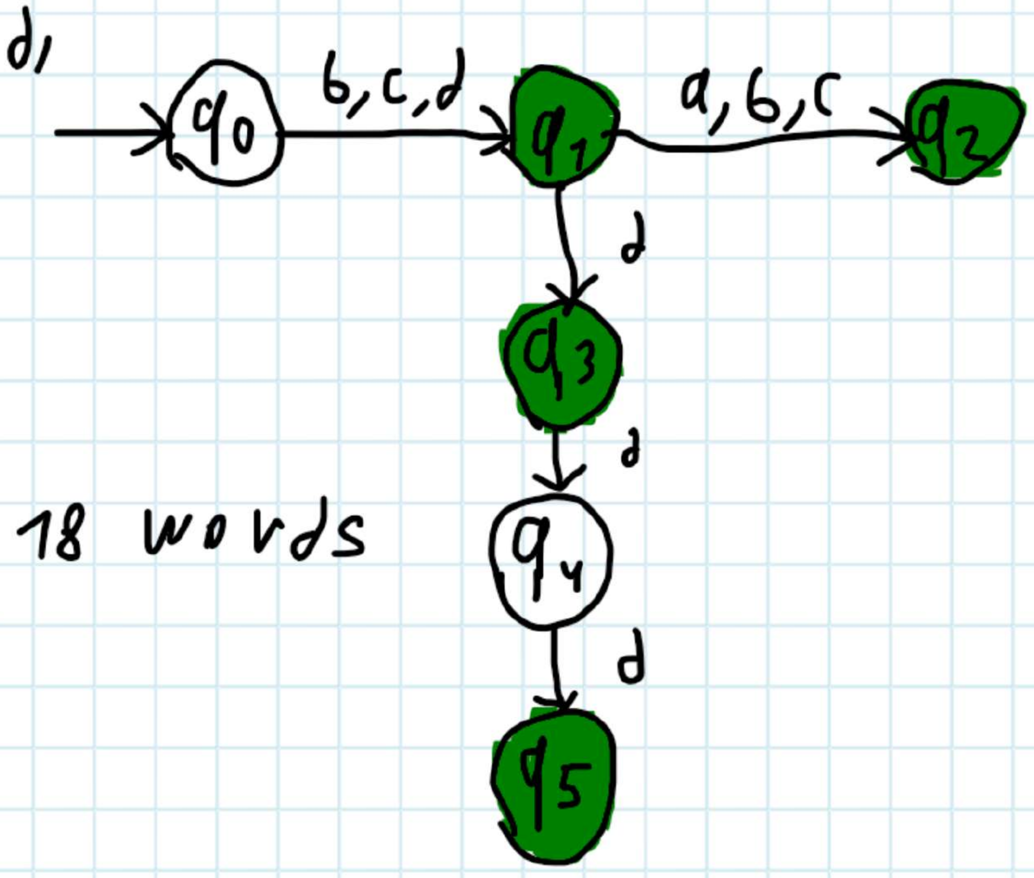
7 words

c)



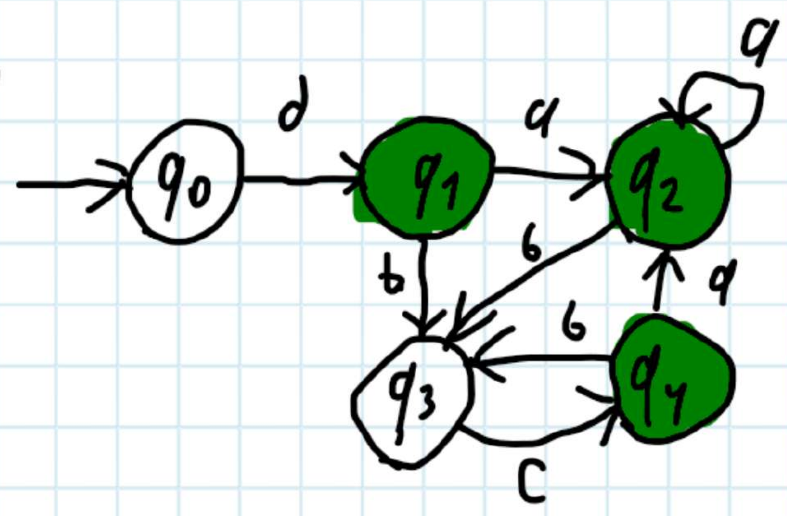
6 words

d)



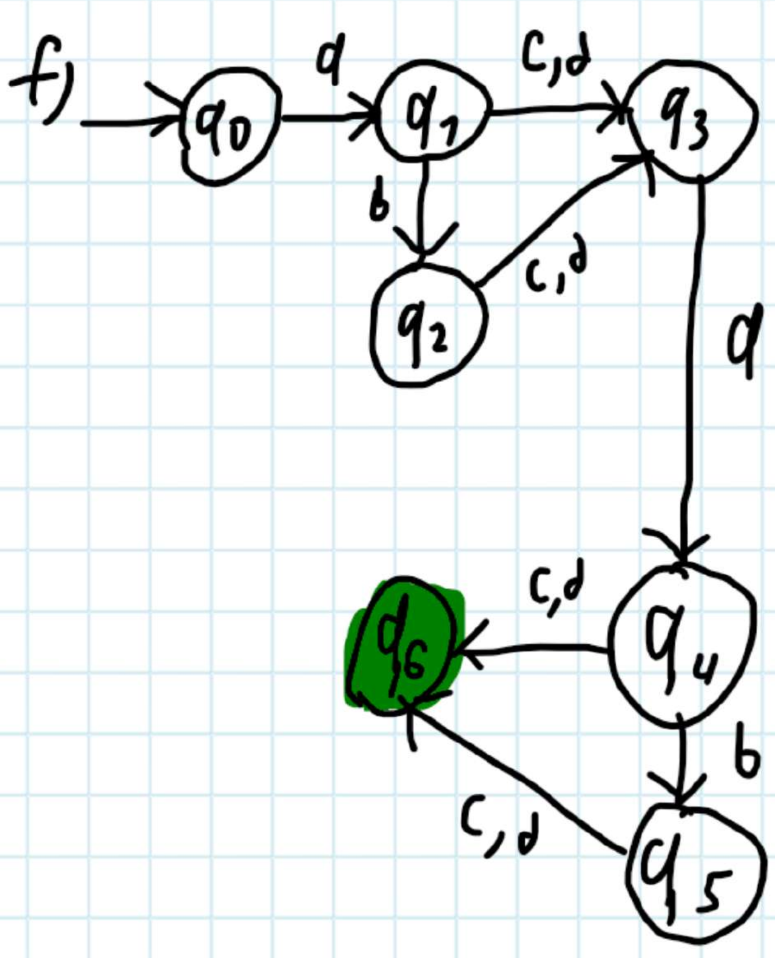
18 words

e)



11 word

f)



12 words

2. Describe the set of strings defined by each of these sets of productions in EBNF (extended Backus-Naur form).

(a) ⟨string⟩ ::= ⟨L⟩+ ⟨D⟩? ⟨L⟩+
⟨L⟩ ::= a | b | c
⟨D⟩ ::= 0 | 1

(b) ⟨string⟩ ::= ⟨sign⟩? ⟨N⟩
⟨sign⟩ ::= '+' | '−'
⟨N⟩ ::= ⟨D⟩ (⟨D⟩ | 0)*
⟨D⟩ ::= 1 | … | 9

(c) ⟨string⟩ ::= ⟨L⟩* (⟨D⟩+)? ⟨L⟩*
⟨L⟩ ::= x | y
⟨D⟩ ::= 0 | 1

(d) ⟨string⟩ ::= ⟨C⟩ ⟨R⟩*
⟨C⟩ ::= a | … | z | A | … | Z
⟨D⟩ ::= 0 | … | 9
⟨R⟩ ::= ⟨C⟩ | ⟨D⟩ | '_'

a) Какая-то комбинация из a, b, c, затем 0 или 1 или ничего, затем снова какая-то комбинация из a, b, c

b) Опциональный знак, а затем набор цифр не начинающийся с 0

c) Любая комбинация x, y (даже пустая), затем любая комбинация 0, 1, и снова любая комбинация x, y

d) Начинается с любой буквы английского алфавита (и строчные и заглавные) далее любые буквы английского алфавита или цифры или „_" (подчеркивани

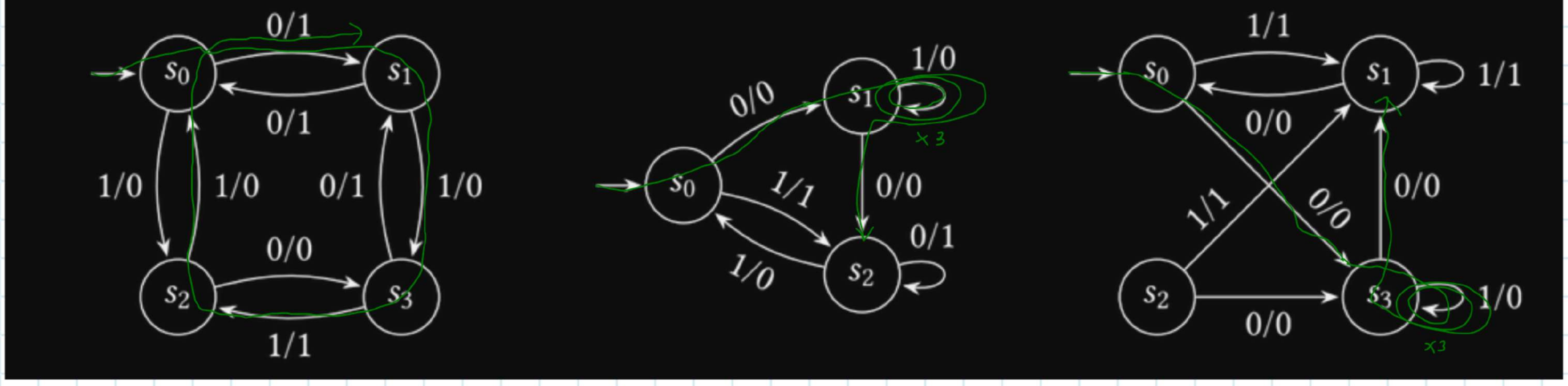
3. Let $\mathcal{G} = \langle V, T, S, P\rangle$ be the phrase-structure grammar with vocabulary $V = \{A, S\}$, terminal symbols $T = \{0, 1\}$, start symbol $S = S$, and set of productions P: $S \to 1S$, $S \to 00A$, $A \to 0A$, $A \to 0$.
(a) Show that 111000 belongs to the language generated by $\mathcal{G}$.
(b) Show that 11001 does not belong to the language generated by $\mathcal{G}$.
(c) What is the language generated by $\mathcal{G}$?

a) $S \to 1S \to 11S \to 111S \to 11100A \to 111000$

b) Т.к. в конце 11001 — единица, а закончить слово мы можем только A, которая преобразуется только в 0 => 11001 does not belong to the language generated by $\mathcal{G}$.

c) 1*|00|0*|0

4. Find the output generated from the input string 01110 for each of the following Mealy machines.



1 0 1 0 1   0 0 0 0 0   0 0 0 0 0

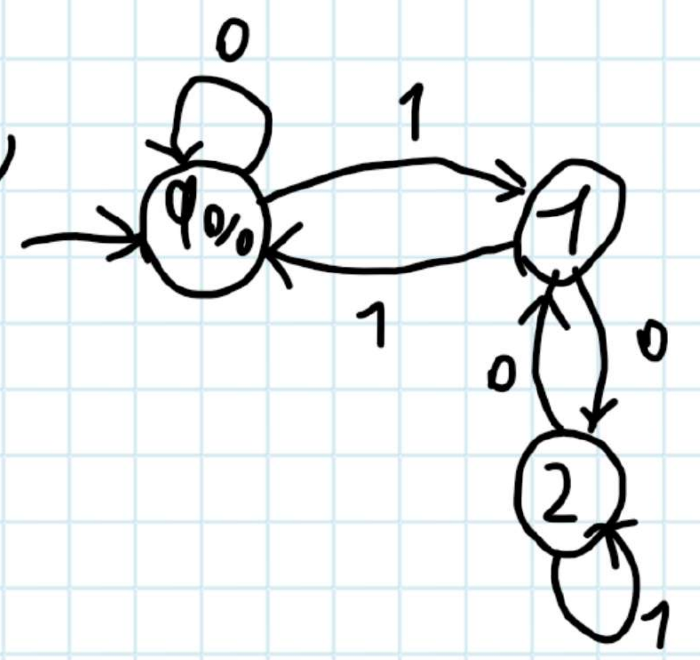
5. Construct a Moore machine for each of the following descriptions.

(a) Determine the residue modulo 3 of the input treated as a binary number. For example, for input ε (which corresponds to "value" 0) the residue is 0; 101 (5 in decimal) has residue 2; and 1010 (value 10) has residue 1.

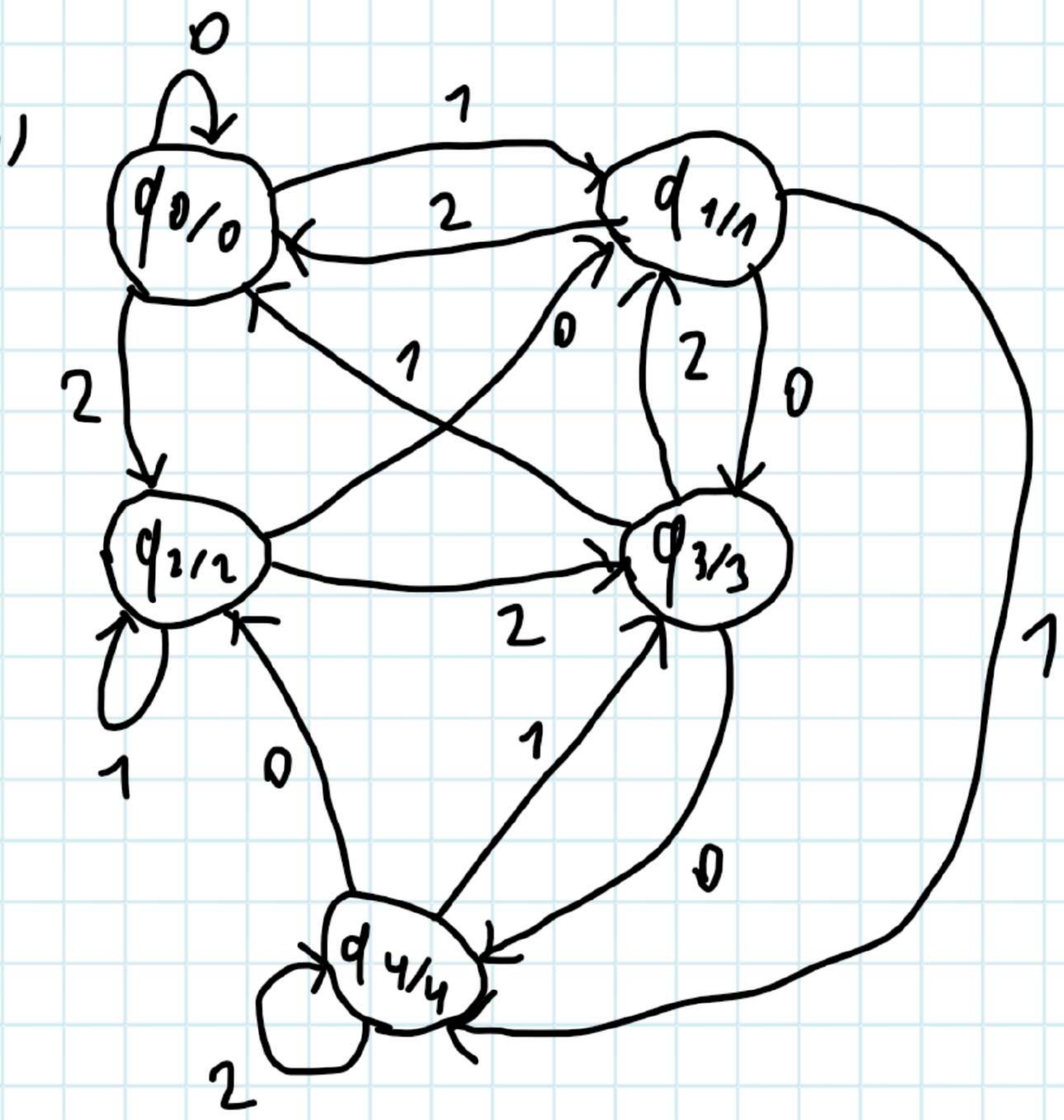
(b) Output the residue modulo 5 of the input from $\{0, 1, 2\}^*$ treated as a ternary (base 3) number.

(c) Output A if the binary input ends with 101; output B if it ends with 110; otherwise output C.

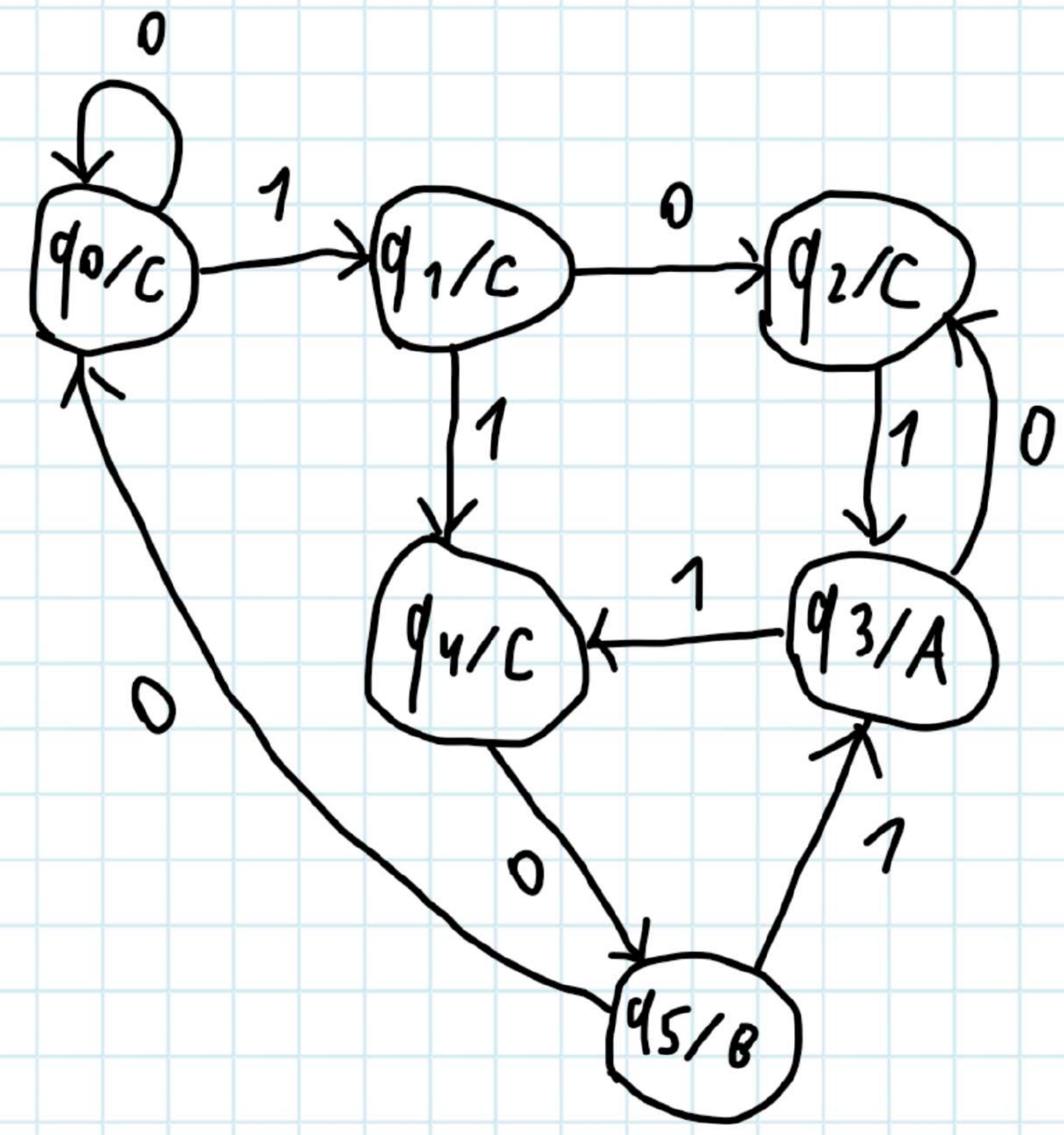
a)



b)



c)



6. Show that regular languages are *closed* under the following operations.

(a) Union, that is, if L_1 and L_2 are regular languages, then $L_1 \cup L_2$ is also regular.

(b) Concatenation, that is, if L_1 and L_2 are regular languages, then $L_1 \cdot L_2$ is also regular.

(c) Kleene star, that is, if L is a regular language, then L^* is also regular.

(d) Complement, that is, if L is a regular language, then $\overline{L} = \Sigma^* - L$ is also regular.

(e) Intersection, that is, if L_1 and L_2 are regular languages, then $L_1 \cap L_2$ is also regular.

a) Для L_1 и L_2 есть регулярные выражения, т.к. они регулярные языки. Тогда ~~для~~ $L_1 \cup L_2$ ~~РВ~~ можно описать объединением регулярных выражений для L_1 и L_2 — $R_{L_{12}} = R_1 | R_2$

b) Аналогично a) только $R_{L12} = R_1 \cdot R_2$

c) Таким же образом регулярное выражение для L^* это R^*

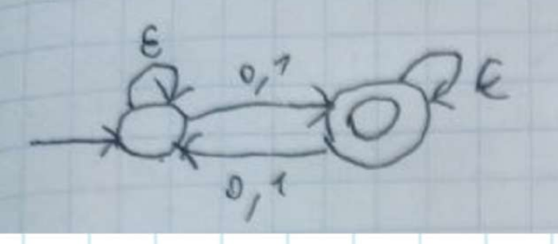
d) (По идее можно сделать так же, но для интереса) Если представить язык как автомат, то $\Sigma^* - L$ по сути просто инвертировать состояния (конечные $\leftrightarrow$ не конечные)

e) $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}} \rightarrow$ a)

7. Determine whether the following languages are regular or not. For non-regular languages, use Pumping lemma to prove that they are not regular. For each regular language, provide a regular expression and construct an ε-NFA.

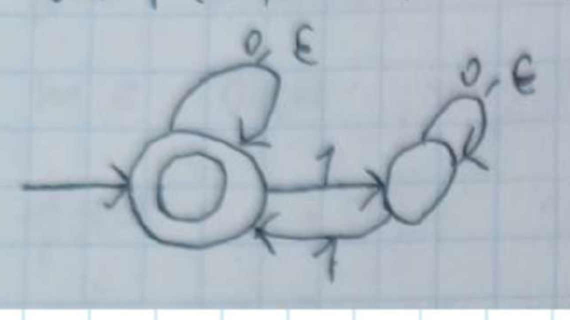
(a) $L_1 = \{w \in \{0,1\}^* \mid \text{length of } w \text{ is odd}\}$
(b) $L_2 = \{0^n 1^n \mid n \in \mathbb{N}\}$
(c) $L_3 = \{w \in \{0,1\}^* \mid w \text{ contains an even number of 1s}\}$
(d) $L_4 = \{1^{n^2} \mid n \in \mathbb{N}\}$

№7 regular
a) (0|1)(00|01|10|11)*



b) irregular
Возьмем $s = 0^p 1^p$ из L. S можно представить как $0^k, 0^l, 0^{p-k-l}$, где $k+l \le p$, $l > 0$
Проверим при $i = 2$:
$0^k 0^{2l} 0^{p-k-l} 1^p = 0^{p+l} 1^p$, что не $= 0^p 1^p$,
т.к. $p+l \ne p$, тогда L irregular

c) regular
0*|(0*10*10*)*



d) irregular
Возьмем $S = 1^{n^2}$ из L, по лемме о накачке представим к xyz, $xy \le p$, $y > 0$
Проверим $i = 2$:
xy^2z — это 1^{n^2+k}
что $\ne 1^{n^2}$, значит L-irregular

10. Solve the following regex crosswords. Fill each cell with a single ASCII character (an uppercase letter, a digit, a punctuation mark, or a space). Each row/column, when read left to right or top to bottom must match the regular expression(s) given for that row/column.

	[^SPEAK]*	EP|IP|EF
HE|LL|O+	H	E
[PLEASE]+	L	P

	(FI|A)*	(YE|OT)K	(.)[IF]*	[NODE]*	(FY|F|RG)*
(Y|F)(.)\2[DAF]\1	F	O	O	D	F
(U|O|I)*T[FRO]+	I	T	F	O	R
[KANE]*[GIN]*	A	K	I	N	G

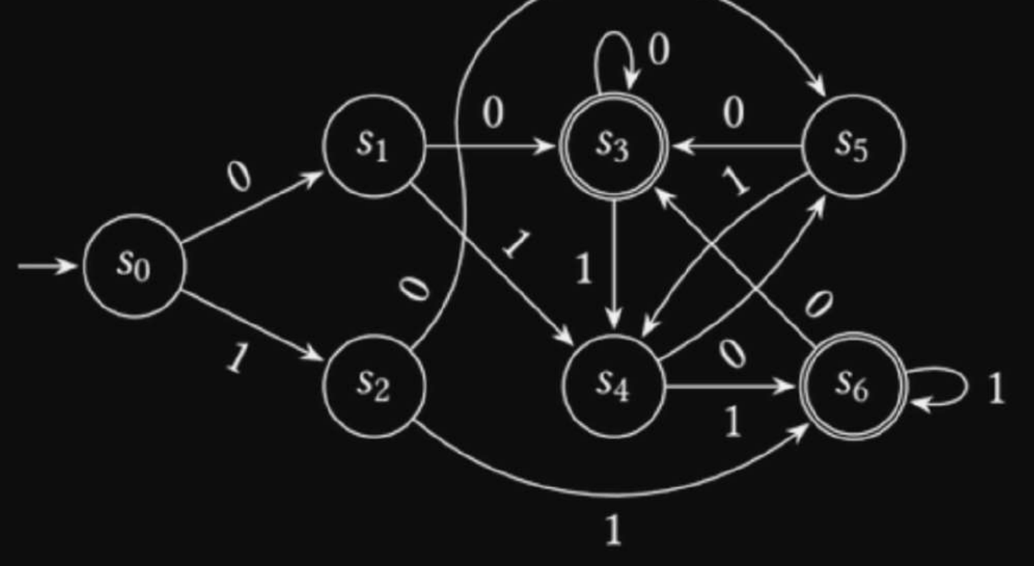
	[^NRU](NO|ON)	(D|FU|UF)*	(FO|A|R)*	(N|A)*
[RUNT]*	T	U	R	N
O.*[HAT]	O	F	F	A
(.)*DO\1	N	D	O	N

	[^MCI]*	.A	(TM|BF)	
(CAT|A-T)+	A	-	T	[^KI\sP]+
[MA\-\sE]+	E	A	M	(M|APS|EA)*
	[AI][E\s]	[A\-Z]*	[\sT\-M]*	

	(.)\1(.)\2	[C\sOU]*	[^PU\sH]*	
.[LUH]+	P	U	L	.*L+
(P|K)[^U]+	P	_	F	[PUF\s]*
.*C+[TIF]	I	C	T	[TIC]*
(NO|ONE|ION)*	I	O	N	[NOI\sE]+
	[PIF]*	.*[OWE]*	(TN|LF|TF)*	

	.*E.*	[^P]I(IT|ME)	(EM|FE)(IT|IP)	(TS|PE|KE)*
(EP|ST)*	S	T	E	P
T[A-Z]*	T	I	M	E
.M.T	E	M	I	T
.*P.[S-X]+	P	E	T	S

9. Consider the finite-state automaton $M = (\Sigma, Q, q_0, F, \delta)$ depicted below.



(a) Find the k-equivalence classes of M for $k = 0, 1, 2, 3$.

(b) Find the $*$-equivalence classes of M.

(c) Construct the quotient automaton $\overline{M}$ of M.

▹ The quotient automaton $\overline{M}$ of the deterministic finite-state automaton $M = (\Sigma, S, s_0, F, \delta)$ is the finite state automaton $\overline{M} = (\Sigma, \overline{S}, [s_0]_{R^*}, \overline{F}, \overline{\delta})$, where the set of states $\overline{S}$ is the set of R^*-equivalence classes of S; the transition function $\overline{\delta}$ is defined by $\overline{\delta}([s]_{R^*}, a) = [\delta(s, a)]_{R^*}$ for all states $[s]_{R^*}$ of $\overline{M}$ and input symbols $a \in \Sigma$; and $\overline{F}$ is the set consiting of R^*-equivalence classes of final states of M.

a) $k = 0$

$[S_0]_R = \{S_0, S_1, S_2, S_4, S_5\}$

$[S_3]_R = \{S_3, S_6\}$

$k = 1$

$[S_0]_R = \{S_0\}$

$[S_1]_R = \{S_1, S_3, S_5\}$

$k = 2$

$[S_0]_R = \{S_0, S_1, S_3, S_5\}$

$[S_2]_R = \{S_2, S_4, S_6\}$

$k = 3$

$[S_0]_R = \{S_0, S_1, S_2, S_3, S_4, S_5, S_6\}$

b) $[S_0]_{R*} = \{S_0\}$

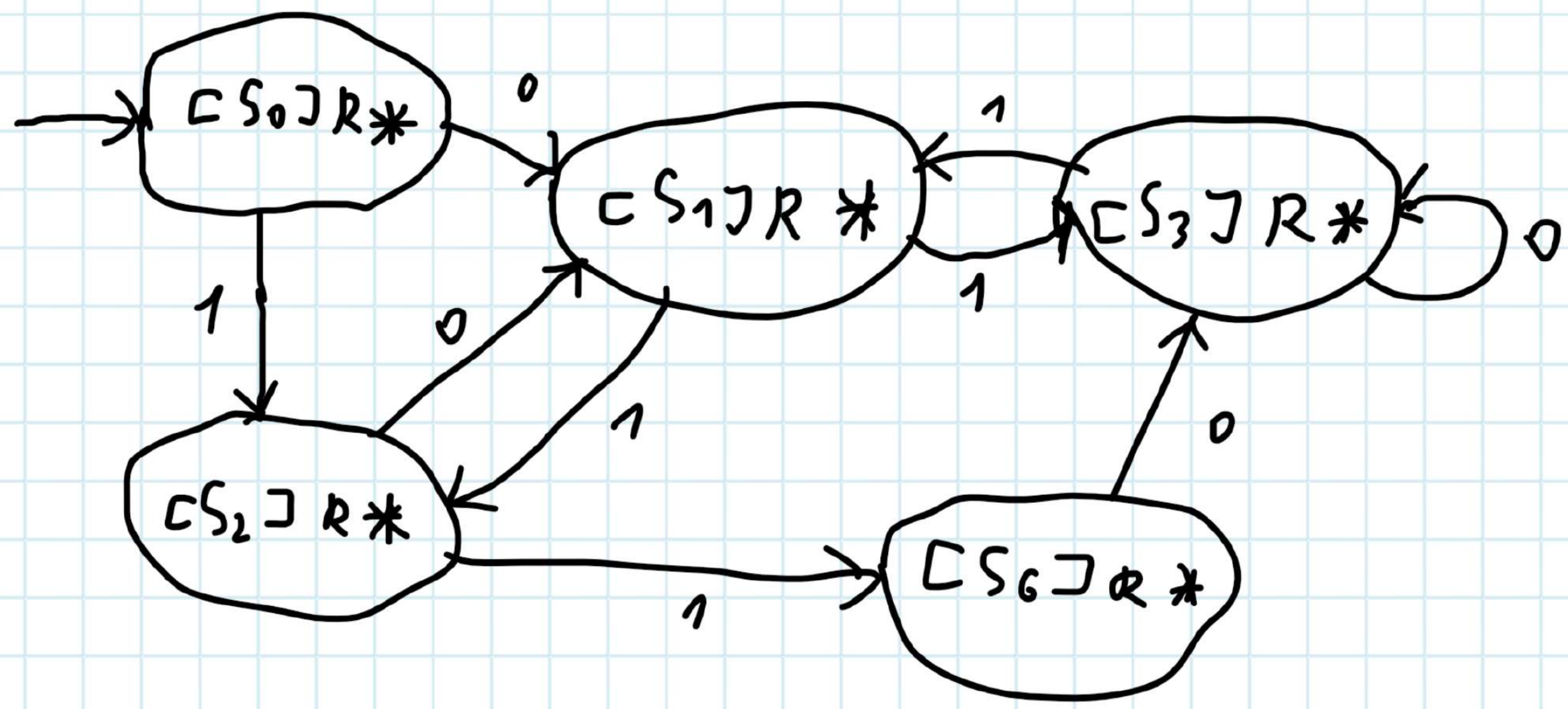
$[S_1]_{R*} = \{S_1, S_5\}$

$[S_2]_{R*} = \{S_2, S_4\}$

$[S_3]_{R*} = \{S_3\}$

$[S_6]_{R*} = \{S_6\}$

c)



8. Consider a finite-state automaton $M = (\Sigma, Q, q_0, F, \delta)$ and a non-negative integer k. Let R_k be the relation on the set of states of M such that $s\, R_k\, t$ if and only if for every input string $w \in \Sigma^*$ with $|w| \le k$, $\delta(s, w)$ and $\delta(t, w)$ are both final states or both not final states. Furthermore, let R^* be the relation on the set of states of M such that $s\, R^*\, t$ if and only if for every input string $w \in \Sigma^*$, regardless of length, $\delta(s, w)$ and $\delta(t, w)$ are both final states or both not final states.

(a) Show that for every nonnegative integer k, R_k is an equivalence relation on S. Two states s and t are called k-equivalent if $s\, R_k\, t$.

(b) Show that R^* is an equivalence relation on S. Two states s and t are called $*$-equivalent if $s\, R^*\, t$.

(c) Show that if two states s and t are k-equivalent ($k > 0$), then they are also $(k-1)$-equivalent.

(d) Show that the equivalence classes of R_k are a *refinement* of the equivalence classes of R_{k-1}.

(e) Show that if two states s and t are k-equivalent for every non-negative integer k, then they are $*$-equivalent.

(f) Show that all states in a given R^*-equivalence class are final or all are not final.

(g) Show that if two states s and t are $*$-equivalent, then $\delta(s, a)$ and $\delta(t, a)$ are also $*$-equivalent for all $a \in \Sigma$.

a) Reflexivity:

$\forall s \in S\ sR_k s$, т.к. $\forall w\ |w| \le k$ автомат останется в состоянии s

Symmetry:

$\forall w\ |w| \le k\ \delta(s,w)$ и $\delta(t,w)$ – конечные состояния, также и обратно

Transitivity:

$\forall w\ |w| \le k\ \delta(s,w)$ и $\delta(t,w)$ – конечные состояния $\Rightarrow$

$\delta(t,w)$ и $\delta(u,w)$ - конечные состояния (аналогично для не конечных)

$\Rightarrow$ equivalence relation

b) Т.к. конечность состояний однозначна (т.к. определяется по исходному состоянию и переходам по слову w)

можно сделать вывод, что a) верно для любой длины строк

c) Если $sR_k t$, то $\forall w\ |w| \le k\ \delta(s,w)$ и $\delta(t,w)$ - оба конечные или не конечные,

тогда $\forall w\ |w| \le k-1\ \delta(s,w)$ и $\delta(t,w)$ - оба конечные или не конечные, т.к. $|w| \le k-1$ подмножество $|w| \le k$

d) По c) если $sR_k t$, то $sR_{k-1} t$, поэтому R_k is refinement of R_{k-1}

e) Если s и t k-equivalen'тные $\forall k \in \mathbb{N}$, то $\forall w\ |w| \le k\ \delta(s,w)$ и $\delta(t,w)$ оба конечные или не конечные

f) Если sR^*t, то $\forall w\ |w| \le k\ \delta(s,w)$ и $\delta(t,w)$ оба конечные или не конечные

g) f)

Рассмотрим $\alpha \in \Sigma$:

Пусть $s' = \delta(s,\alpha)$ и $t' = (t,\alpha)$

$\forall w\ \delta(s',w) = \delta(\delta(s,\alpha),w)$ и $\delta(t',w) = \delta(\delta(t,\alpha),w)$

Т.к. sR^*t, следует, что $\delta(s,\alpha w)$ и $\delta(t,\alpha w)$ оба конечные или не конечные

Значит $\delta(s',w)$ и $\delta(t',w)$ тоже

$\Rightarrow s'R^*t' \Rightarrow \delta(s,\alpha)\, R^*\, \delta(t,\alpha)$