

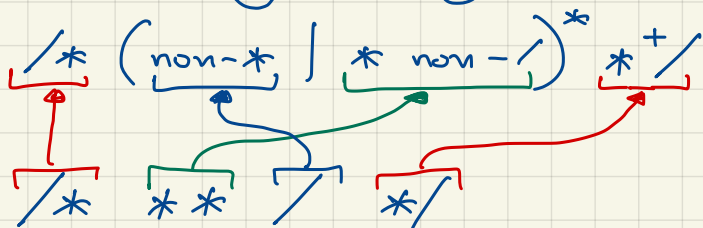
A1 sol



A1-sol

Q<sub>1</sub>

a) `/***/**/` is not a valid comment but it is accepted by the regular expression as shown below:



b) Correct expression:

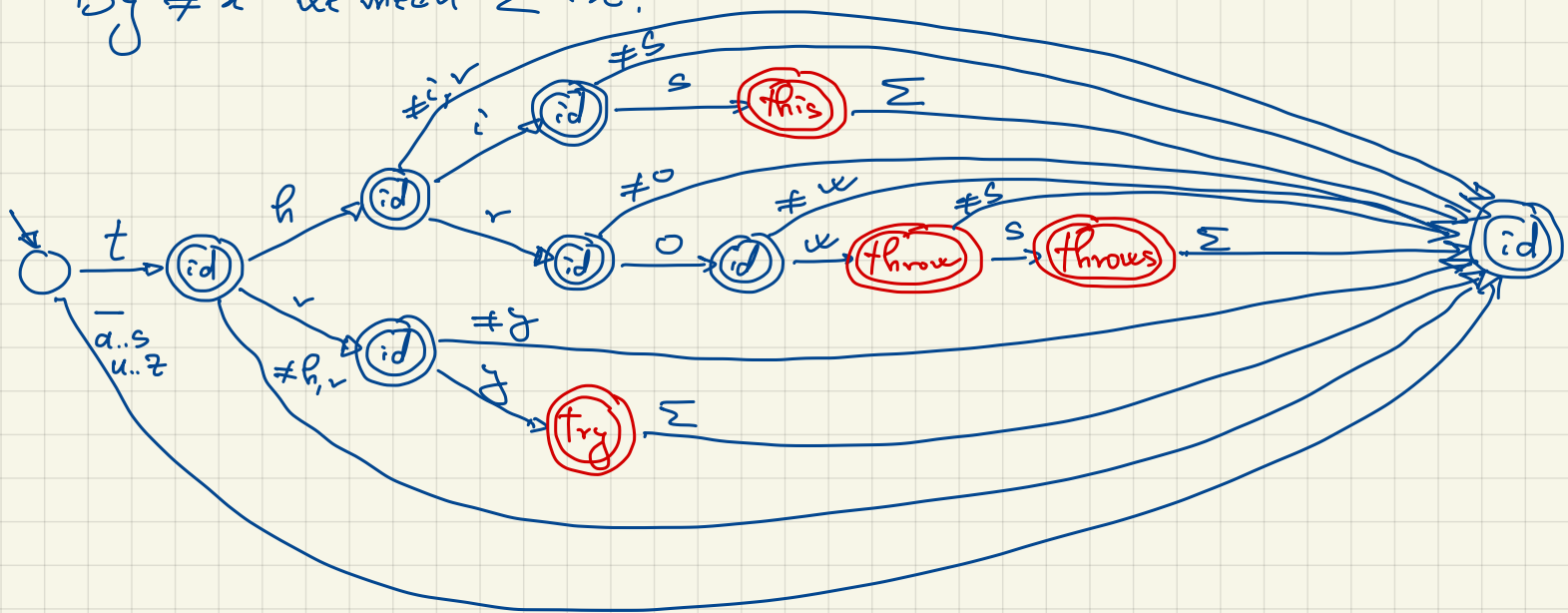
`/* (non-* | *+ (non-* /))* *+ /`

all `*`'s within the comment, except for those at the very end (`*...*/`), are matched here, followed by a `non-*/` character.

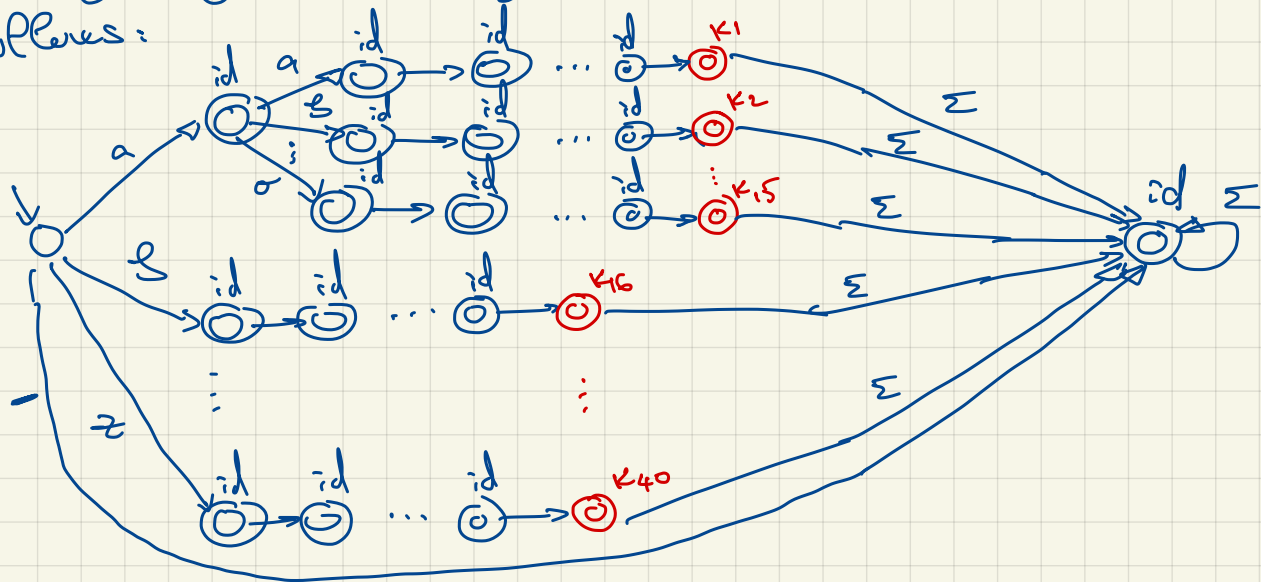
Q2 a)

Denote  $\Sigma = \{-, a, b, \dots, z, 0, 1, \dots, 9\}$ .

By  $\neq x$  we mean  $\Sigma \setminus x$ .



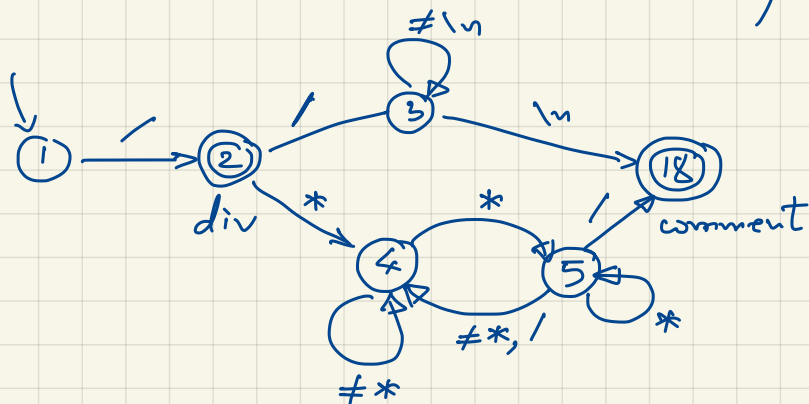
b) We need to maximize "branching". However, the number of letters, 26, imposes a limit on how much we can branch at the beginning. The largest number of states is achieved as follows:



For keywords  $K_i$ ,  $1 \leq i \leq 40$ , denoting  $|K_i|$  the number of characters in  $K_i$ , the number of states is  $\sum_{i=1}^{40} |K_i| - 14 + 2 = \sum_{i=1}^{40} |K_i| - 12$

Q<sub>3</sub> a)

The DFA is slightly different from the one in Fig. 2.6. The "comment" is a token with that part of DFA looking like this:

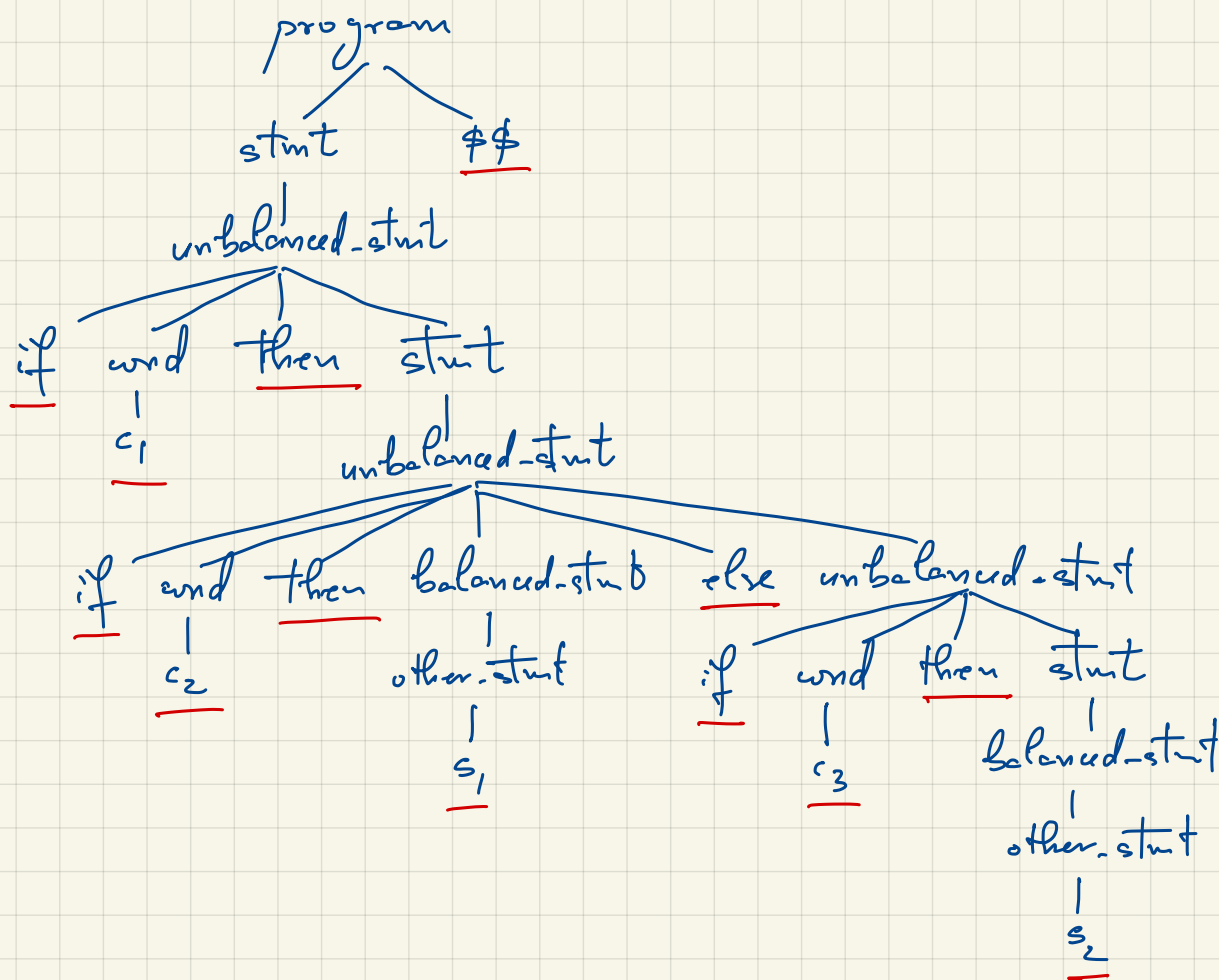


Therefore, we can choose

$t_1 = /$	- div
$w = /*$	- not a token
$t_2 = /**/$	- comment

b) Example: /\*\* characters  
unread

Q4 a



b

X	FIRST(X)	FOLLOW(X)
program	if(1,1) s <sub>i</sub> (1,1)	
stmt	if(1,2) s <sub>i</sub> (1,2)	
balanced-stmt	if(1,4) s <sub>i</sub> (1,5)	\$(2,1)
unbalanced-stmt	if(1,6)	else(2,4) \$(3,2)
other-stmt	s <sub>i</sub> (1,9)	\$(3,3)
and	c <sub>i</sub> (1,8)	else(3,5), \$(3,5)
		then(2,4)

c

PREDICT

1. if, s<sub>i</sub>
2. if, s<sub>i</sub>
3. if
4. if
5. s<sub>i</sub>

6. if
7. if
8. c<sub>i</sub>
9. s<sub>i</sub>

d

stmt → balanced-stmt  
stmt → unbalanced-stmt

}

same LHS, the PREDICT of both contains if

⑥ - there is no left-recursion

- there are common prefixes for productions 6 and 7

- new grammar has the following productions in place of former 6-7:

unbalanced\_start  $\rightarrow$  if and then unbal\_start\_tail

unbal\_start\_tail  $\rightarrow$  start

unbal\_start\_tail  $\rightarrow$  balanced\_start else unbalanced\_start

⑦ It is not LL(1). It has the same conflict from before:

start  $\rightarrow$  balanced\_start  
start  $\rightarrow$  unbalanced\_start

}

same LHS, the PREDICT of both contains if