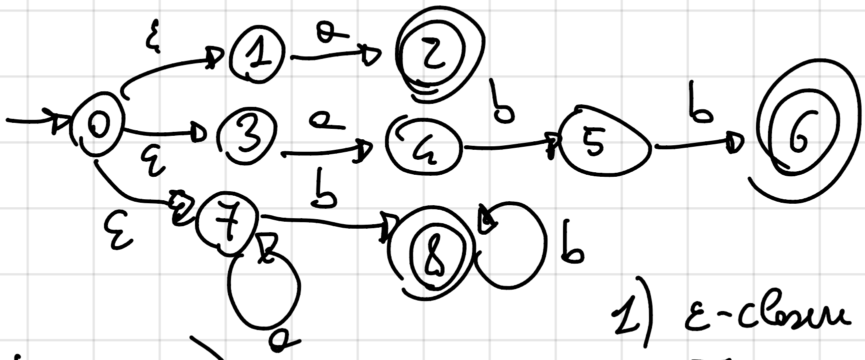
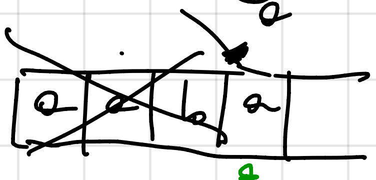




Input-pos

Last_Final = $\{2\}$

$$\bar{\delta} : \mathcal{P}(S) \times \Sigma \to \mathcal{P}(S)$$

$$\delta : S \times \Sigma \to \mathcal{P}(S)$$

$$\bar{\delta}(T, a) = \bigcup_{s \in T} \delta(s, a)$$

1) ε-closure$(\{0\}) = \{0, 1, 3, 7\}$

2) $\bar{\delta}(\{0,1,3,7\}, a) = \{2, 4, 7\}$

3) ε-closure$(\{2,4,7\}) = \{2, 4, 7\}$

4) $\bar{\delta}(\{2,4,7\}, a) = \{7\}$

5) ε-closure$(\{7\}) = \{7\}$

6) $\bar{\delta}(\{7\}, b) = \{8\}$

7) ε-closure$(\{8\}) = \{8\}$

8) $\bar{\delta}(\{8\}, a) =$ blocked

9) Output TOKEN 3 with lexeme aab

10) Reset and restart
(clean the input by eliminating the part already recognised)

11) ε-closure$(\{0\}) = \{0, 1, 3, 7\}$

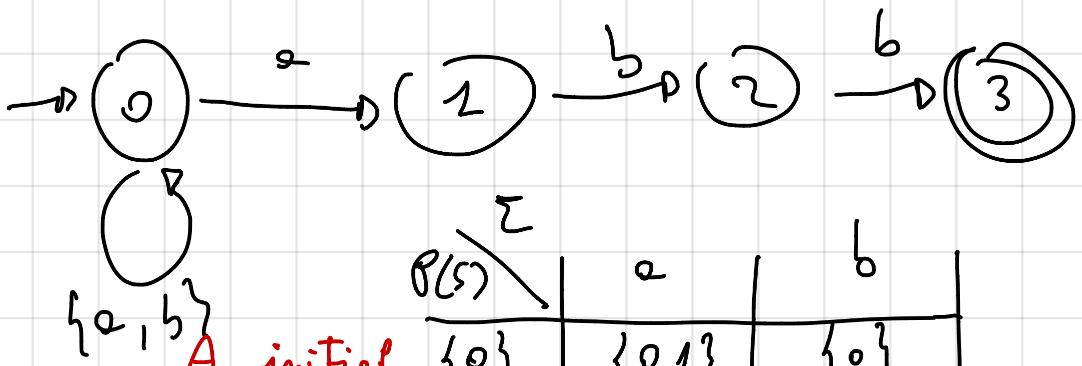
12) $\bar{\delta}(\{0, 1, 3, 7\}, a) = \{2, 4, 7\}$

13) ε-closure$(\{2, 4, 7\}) = \{2, 4, 7\}$

14) $\bar{\delta}(\{2, 4, 7\}, ?)$ blocked and terminate

15) output TOKEN 1 lexeme a

SUBSET CONSTRUCTION



Worklist

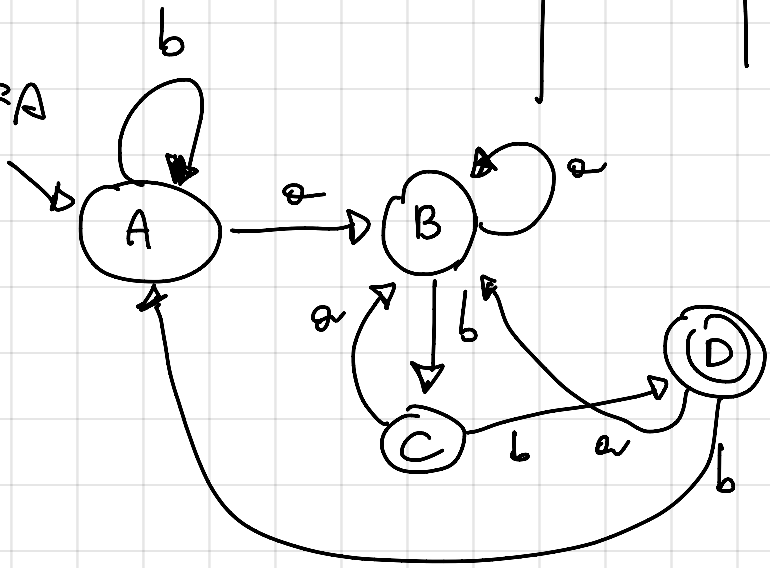
- $\cancel{\{0\}}$
- $\cancel{\{0,1\}}$
- $\cancel{\{0,2\}}$
- $\cancel{\{0,3\}}$

		$P(S)$ \ Σ	a	b
A	initial	{0}	{0,1}	{0}
B		{0,1}	{0,1}	{0,2}
C		{0,2}	{0,1}	{0,3}
D	final	{0,3}	{0,1}	{0}

$\bar{\delta}(\varepsilon\text{-closure}(\{0\}), a)$
$= \{0,1\}$

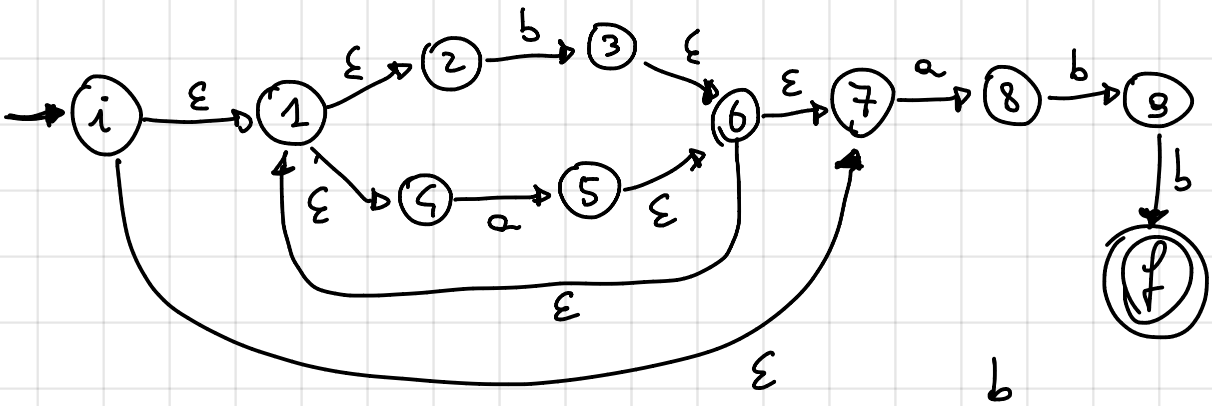
$\bar{\delta}(\varepsilon\text{-closure}(\{0\}), b) =$
$\{0\}$

DFA

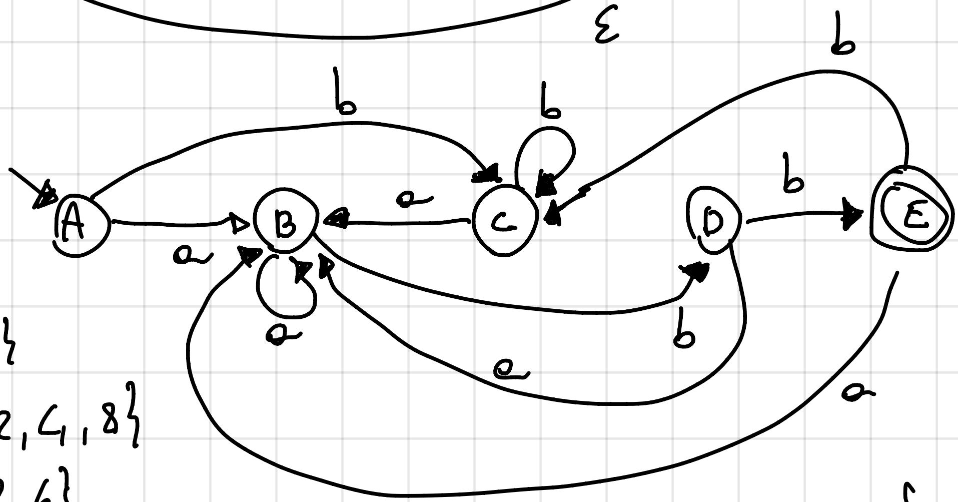


$(a|b)^*abb$ Thompson's

NFA



NFA → DFA



$A = \{i, 1, 2, 4, 7\}$

$B = \{5, 6, 7, 1, 2, 4, 8\}$

$C = \{3, 6, 7, 1, 2, 4\}$

$D = \{3, 6, 7, 1, 2, 4, 9\}$ $E = \{3, 6, 7, 1, 2, 4, f\}$

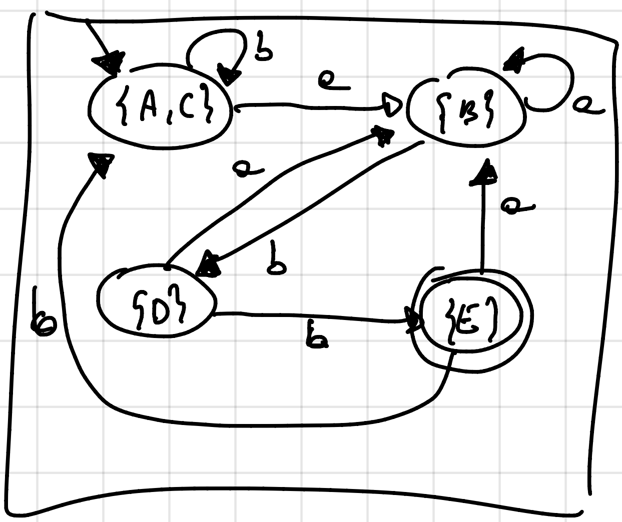
$\Pi_1 = (\{A, B, C, D\}, \{E\})$ cannot be refined

$A \xrightarrow{a} B$ $A \xrightarrow{b} C$

$B \xrightarrow{a} B$ $B \xrightarrow{b} D$

$C \xrightarrow{a} B$ $C \xrightarrow{b} C$

$D \xrightarrow{a} B$ $D \xrightarrow{b} E$



$\Pi_2 = (\{A, B, C\}, \{D\}, \{E\})$

$A \xrightarrow{b} C$

$B \xrightarrow{b} D$

$C \xrightarrow{b} C$

$A \xrightarrow{a} B$

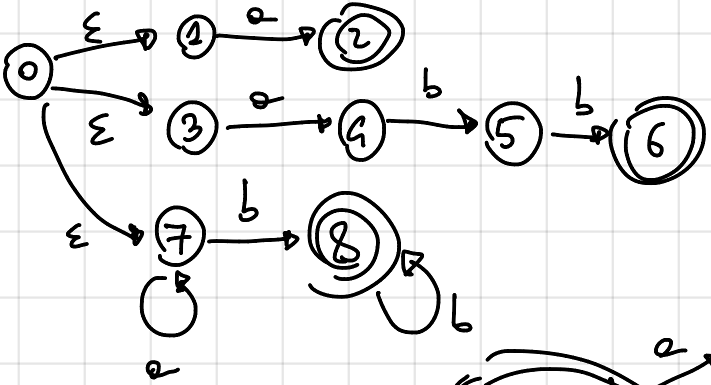
$C \xrightarrow{a} B$ NO DISTINCTION

$A \xrightarrow{b} C$

$C \xrightarrow{b} C$

$\Pi_3 = (\{A, C\}, \{B\}, \{D\}, \{E\})$

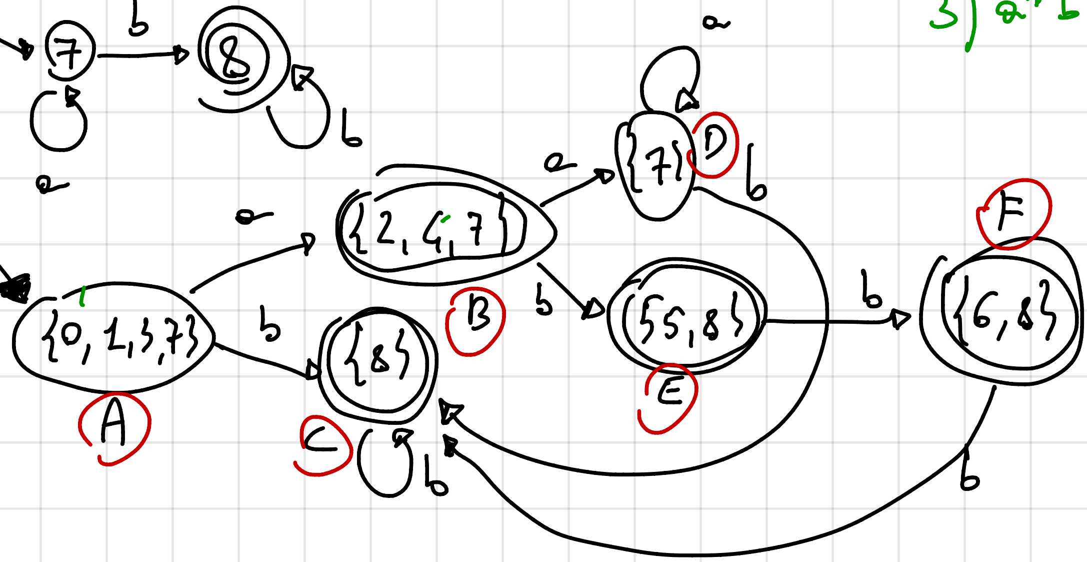
STOP



NFA

1) a T1
2) abb T2
3) a^*b^+ T3

DFA



$\pi_1 = (\{A,D\}, \{B,C,E,F\})$ ← This initial partition is not correct w.r.t. the behaviour of the LA

$A \xrightarrow{a} B$

$D \xrightarrow{a} D$

$\pi_2 = (\{A\}, \{D\}, \{B,C,E,F\})$

$B \xrightarrow{a} D$

$C \not\xrightarrow{a}$

$E \not\xrightarrow{a}$

$F \not\xrightarrow{a}$

$\pi_3 = (\{A\}, \{D\}, \{B\}, \{C,E,F\})$

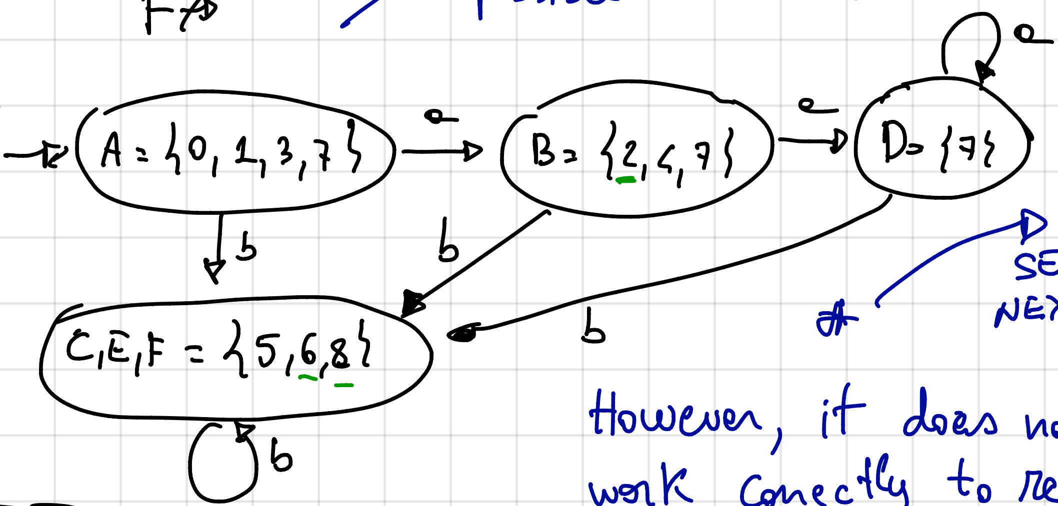
$C \xrightarrow{b} C$

$E \xrightarrow{b} F$

$F \xrightarrow{b} C$

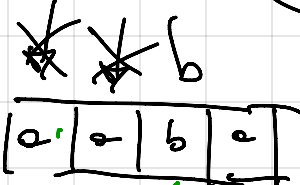
This automaton is CORRECTLY MINIMAL w.r.t. THE UNION OF possible LEXEMES

MINIMAL DFA



SEE NEXT PAGE

However, it does not work correctly to recognise the tokens



Last-final {6,8}

1) $A \xrightarrow{a} B$

2) $B \xrightarrow{a} D$

3) $D \xrightarrow{b} CEF$ → HERE The LA cannot decide which TOKEN is outputted, TOKEN 2 (state 6) OR TOKEN 3 (state 8)

THE INITIAL PARTITION TO CONSIDER MUST DISTINGUISH THE FINAL STATES WITH RESPECT TO THE SET OF TOKENS THAT CAN BE RECOGNISED.

In our example we have that final states are

$B = \{2,4,7\}$ i.e. TOKEN 1 (state 2)

$C = \{8\}$ i.e. TOKEN 3 (state 8)

$E = \{5,8\}$ i.e. TOKEN 3 (state 8)

$F = \{6,8\}$ i.e. TOKEN 2 AND TOKEN 3 (state 6 and state 8)

TO RETAIN THIS INFORMATION THESE FINAL STATES MUST BE DISTINGUISHED IF THE SET OF RECOGNISED TOKENS ARE DIFFERENT.

In our case they must be initially partitioned in three groups:

$\{B\} \rightarrow$ only TOKEN 1

$\{C, E\} \rightarrow$ only TOKEN 3

$\{F\} \rightarrow$ TOKEN 2 and TOKEN 3

Then the minimization can start as usual: $\rightarrow$

$\Pi_2 = (\underline{\{A, D\}}, \{B\}, \{C, E\}, \{F\})$

We try to partition $\{A, D\}$:

$A \xrightarrow{a} B$
$D \xrightarrow{a} D$ $\longrightarrow$ they can be distinguished

$\Pi_2 = (\{A\}, \{D\}, \{B\}, \underline{\{C, E\}}, \{F\})$

we try to partition $\{C, E\}$

$C \xrightarrow{b} C$
$E \xrightarrow{b} F$ $\longrightarrow$ they can be distinguished

$\Pi_3 = (\{A\}, \{D\}, \{B\}, \{C\}, \{E\}, \{F\})$

We stop because no group can be refined any more.

THUS, in this case the minimization, considering the information on tokens, does not eliminate states.

The LA works on the DFA obtained from the NFA

$\boxed{\cancel{a}}\boxed{\cancel{a}}\boxed{\cancel{b}}\boxed{\cancel{a}}$

4

Input_Pos - Last_final

Last_final = $\cancel{\{1\}}$ $\cancel{\{2\}}$ $\cancel{\{8\}}$ $\cancel{\{5\}}$ $\{2\}$

1) $A \xrightarrow{a} B$
2) Last_Final $\leftarrow \{2\}$, update Input_Pos..
3) $B \xrightarrow{a} D$
4) $D \xrightarrow{b} C$
5) Last_Final $\leftarrow \{8\}$, update Input_pos..
6) $C \not\xrightarrow{a}$ STOP
7) Ouput TOKEN 3 with lexeme aab
   Reset Last_final and input_pos
8) $A \xrightarrow{a} B$
9) Last_final $\leftarrow \{2\}$, upd....
10) $B \not\rightarrow$ end input
11) Output TOKEN 1 with lex. a

STOP