KYUSHU UNIVERSITY 100th 2011
知の新世紀を拓く

形式言語理論

Formal Language Theory

3．正規表現

九州大学
KYUSHU UNIVERSITY

本日の内容

- 準備
- 正規表現
- ε-NFAと正規表現の等価性

準備

言語の連接

- アルファベット Σ 上の言語 L_1, L_2 に対し,

$$L_1 \cdot L_2 = \{ xy \mid x \in L_1, y \in L_2 \}$$

を L_1, L_2 の**連接**と呼ぶ.

- 演算子「・」を略して $L_1 \cdot L_2$ を $L_1 L_2$ と書くことが多い.

言語の冪乗

- アルファベット Σ 上の言語 L と非負整数 n に対し，言語 L^n を次で定める．

$$L^n = \begin{cases} \{\varepsilon\}, & \text{if } n = 0 \\ L^{n-1} \cdot L, & \text{if } n > 0 \end{cases}$$

- アルファベット Σ 上の言語 L に対し，言語 L^*, L^+ を次で定める．

$$L^* = \bigcup_{n=0}^{\infty} L^n$$

$$L^+ = \bigcup_{n=1}^{\infty} L^n$$

正規表現

正規表現の例

- 0,1から成る文字列のうち0で始まるもの.

  0(0+1)*　　　$\underline{0}$, $\underline{0}0$, $\underline{0}1$, $\underline{0}00$, $\underline{0}01$, …

- 0,1から成る文字列のうち00 で終わるもの.

  (0+1)* 00　　　$\underline{00}$, $0\underline{00}$, $1\underline{00}$, $00\underline{00}$, $01\underline{00}$, …

- 0,1から成る文字列のうち途中に011を含むもの.

  (0+1)* 011 (0+1)*　　　$\underline{011}$, $0\underline{011}$, $1\underline{011}$, $0\underline{011}0$, $0\underline{011}1$, …

正規表現の例

- 0,1から成る文字列のうち偶数個の0を含むもの.

$$1^*(01^*01^*)^*$$

1111, 11$\underline{0}$1$\underline{0}$1, 1$\underline{0}$111$\underline{0}$1$\underline{00}$11, …

正規表現の例

- 非負整数の十進表現.

$$(1+2+3+4+5+6+7+8+9)(0+1+2+3+4+5+6+7+8+9)^{*}+0$$

- ２の倍数である非負整数の十進表現.

$$(\ (1+2+3+4+5+6+7+8+9)(0+1+2+3+4+5+6+7+8+9)^{*}\ +\varepsilon\ )(0+2+4+6+8)$$

正規表現(regular expression)

【定義】Σをアルファベットとする．$A=\{), (, \varepsilon, \varnothing, \cdot, +, {}^{*}\}$は$\Sigma\cap A=\varnothing$を満たすとする．$\Sigma$上の正規表現とは，次の(1)-(5)から導かれる$\Sigma\cup A$上の文字列をいう．

(1) Σの要素は正規表現である．

(2) εは正規表現である．

(3) $\varnothing$は正規表現である．

(4) αとβが正規表現 ならば $(\alpha\cdot\beta)$ は正規表現である．

(5) αとβが正規表現 ならば $(\alpha+\beta)$ は正規表現である．

(6) α が正規表現ならば α^{*}は正規表現である．

$\Sigma=\{a,b\}$のとき

a は正規表現

b は正規表現

(a+b) は正規表現

(a+b)* は正規表現

(a・(a+b)*) は正規表現

正規表現の「意味」

【定義】 Σ 上の正規表現を Σ^* の部分集合に写す写像 $\|\cdot\|$ を次で定める.

(1) すべての $a\in\Sigma$ に対して $\|a\| = \{a\}$.

(2) $\|\varepsilon\| = \{\varepsilon\}$.

(3) $\|\varnothing\| = \varnothing$.

(4) 正規表現 α と β に対して $\|(\alpha\cdot\beta)\| = \|\alpha\|\cdot\|\beta\|$.

(5) 正規表現 α と β に対して $\|(\alpha+\beta)\| = \|\alpha\|\cup\|\beta\|$.

(6) 正規表現 α に対して $\|\alpha^*\| = \|\alpha\|^*$.

$$\|(a\cdot(a+b)^*)\| = \|a\|\cdot\|(a+b)^*\| = \{a\}\cdot\|(a+b)\|^* = \{a\}\cdot(\|a\|\cup\|b\|)^*$$

$$= \{a\}\cdot(\{a\}\cup\{b\})^* = \{a\}\cdot\{a,b\}^*$$

$$= \{ax \mid x\in\{a,b\}^*\}$$

正規表現の簡略化

- 正規表現 α に対し $(\alpha \cdot \alpha^*)$の省略形として α^+ を用いることがある.
- 演算の強さを

$$* > \cdot > +$$

と定めることにより, (,) の省略を許す.

$$((0(1^*))+0) \Rightarrow 01^*+0$$

ε-NFAと正規表現の等価性

ε-NFAと正規表現の等価性

【定理2.3】L を任意の言語とする．このとき，次の(1)(2)は等価である.

(1) ある正規表現 r が存在して $L = \|r\|$.

(2) ある ε-NFA M が存在して $L = L(M)$となる.

定理2.3の証明

- 定理2.1, 2.2より，次の2つの補題を示せばよい.

【補題2.1】すべての正規表現 r に対して $L(M) = \|r\|$となり次の(1)(2)を満たすε-NFA $M = (Q, \Sigma, \delta, q_0, F)$が存在する.

(1) M はただ一つの最終状態 f をもつ．すなわち $F=\{f\}$.

(2) すべての $a \in \Sigma \cup \{\varepsilon\}$に対して $\delta(f, a)=\varnothing$.

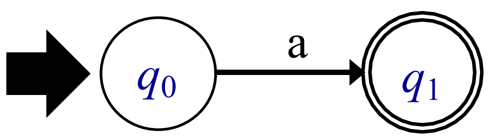
【補題2.2】すべてのDFA M に対して $L(M) = \|r\|$となる正規表現 r が存在する.

補題2.1の証明:アイデア(1／3)

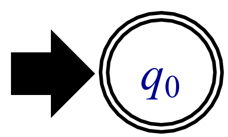
- 正規表現 r に含まれる演算子の個数 m に関する帰納法で証明する.
  - $m=0$ となるのは，次の3つの場合のいずれかである.
    - $r = a \in \Sigma$.
    - $r = \varepsilon$.
    - $r = \varnothing$.
  - $m>0$ となるのは，次の3つの場合のいずれかである.
    - $r = r_1 \cdot r_2$ (r_1, r_2は正規表現).
    - $r = r_1 + r_2$ (r_1, r_2は正規表現).
    - $r = r_1{}^*$ (r_1は正規表現).

補題2.1の証明：アイデア（2／3）

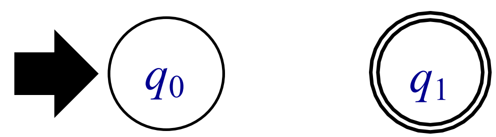
- m=0 のとき.



$r = a \in \Sigma$ に対する ε-NFA



$r = \varepsilon$ に対する ε-NFA



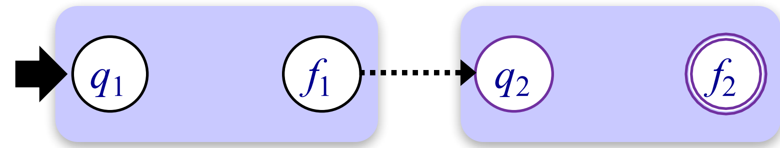
$r = \varnothing$ に対する ε-NFA

補題2.1の証明：アイデア（3／3）

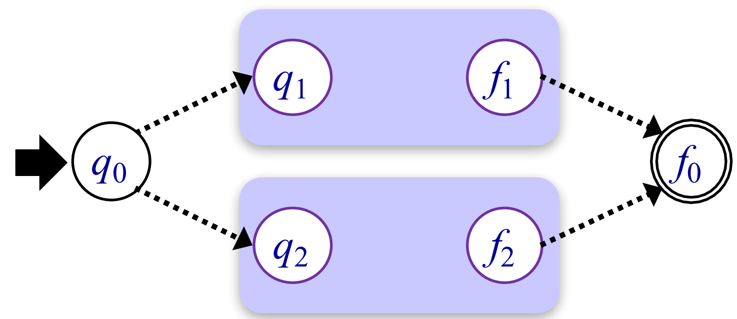
- $m>0$ のとき．

r_1 に対する ε-NFA

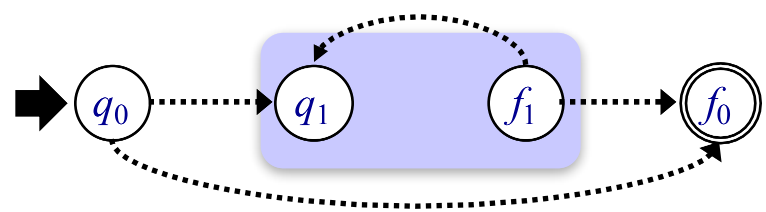
r_2 に対する ε-NFA



$r = r_1 \cdot r_2$ に対する ε-NFA

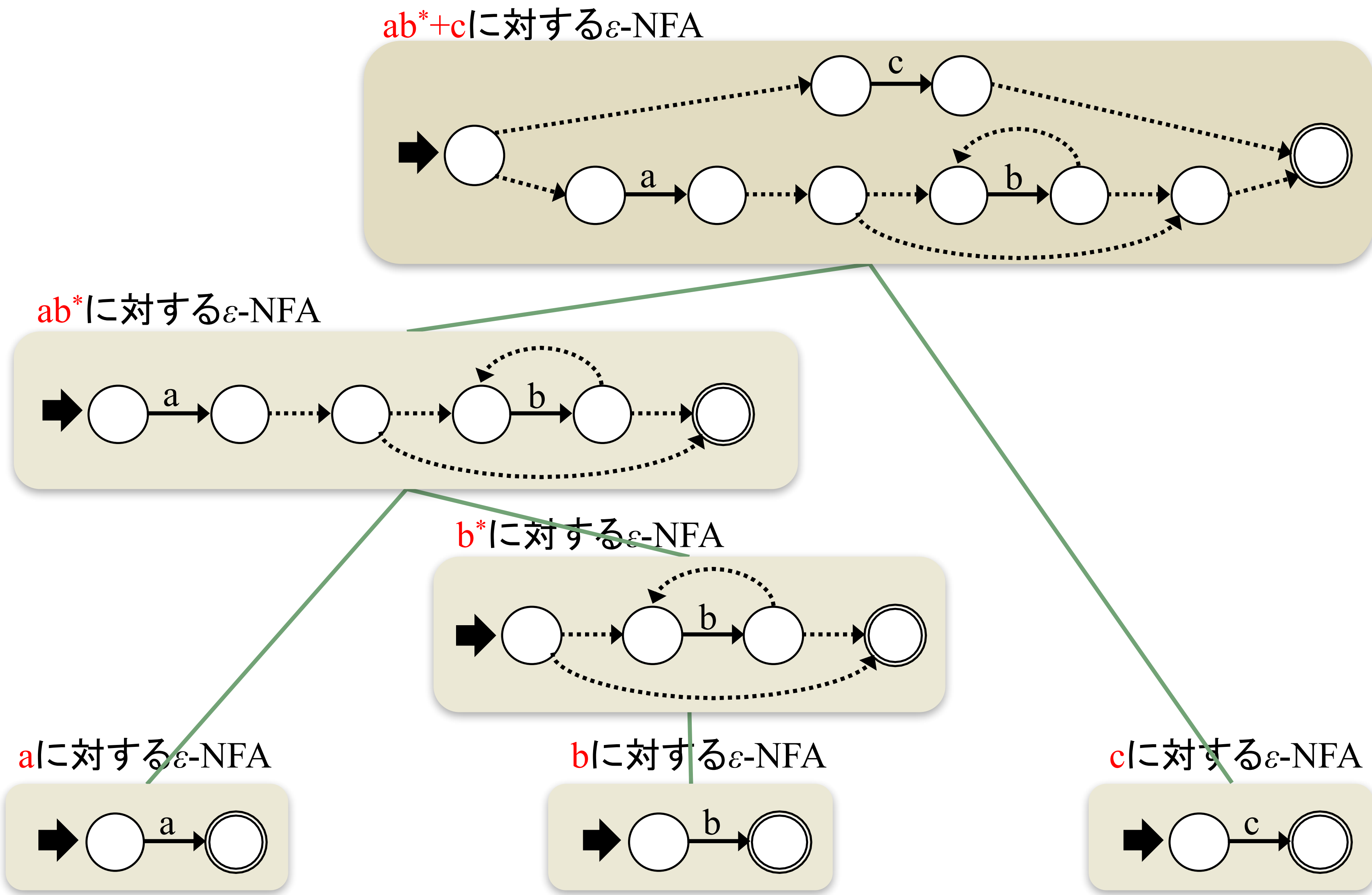


$r = r_1 + r_2$ に対する ε-NFA



$r = r_1^*$ に対する ε-NFA

例　正規表現ab^*+cに対するε-NFAの構成



補題2.1の証明（1／4）

- 正規表現 r に含まれる演算子の個数 m に関する帰納法で証明する.
- m=0のとき．補題の条件を満たす ε-NFA Mを次のように構成できる.

1. $r = a \in \Sigma$ ならば，$M=(\{q_0,q_1\}, \Sigma, \delta, q_0, \{q_1\})$ s.t.
$$\delta(q, c) = \begin{cases} \{q_1\}, & \text{if } (q, c) = (q_0, a) \\ \emptyset, & \text{otherwise.} \end{cases}$$
2. $r = \varepsilon$ ならば，$M=(\{q_0\}, \Sigma, \delta, q_0, \{q_0\})$ s.t.
$$\delta(q_0, c) = \emptyset \quad (\forall c \in \Sigma \cup \{\varepsilon\}).$$
3. $r = \emptyset$ ならば，$M=(\{q_0,q_1\}, \Sigma, \delta, q_0, \{q_1\})$ s.t.
$$\delta(q, c) = \emptyset \quad (\forall (q,c) \in \{q_0,q_1\} \times (\Sigma \cup \{\varepsilon\})).$$

補題2.1の証明（2／4）

- $m>0$のとき．補題の条件を満たす ε-NFAを次のように構成できる.
  1. $r = r_1 \cdot r_2$ (r_1, r_2は正規表現)ならば，帰納法の仮定より，r_1, r_2に対する ε-NFA M_1, M_2が存在する.
  $M_1=(Q_1, \Sigma, \delta_1, q_1, \{f_1\}), M_2=(Q_2, \Sigma, \delta_2, q_2, \{f_2\})$とする.
  一般性を失わず，$Q_1 \cap Q_2 = \emptyset$としてよい.
  このとき，$M=(Q_1 \cup Q_2, \Sigma, \delta, q_1, \{f_2\})$は補題の条件を満たす ε-NFAである．ここに，δ は次で定義される.

$$\delta(q,c) = \begin{cases} \delta_1(q,c), & \text{if } (q,c) \in Q_1 \times (\Sigma \cup \{\varepsilon\}) - \{(f_1,\varepsilon)\} \\ \delta_2(q,c), & \text{if } (q,c) \in Q_2 \times (\Sigma \cup \{\varepsilon\}) \\ \{q_2\}, & \text{if } (q,c) = (f_1,\varepsilon) \end{cases}$$

補題2.1の証明（３／４）

2. $r = r_1 + r_2$ (r_1, r_2は正規表現)ならば，帰納法の仮定より，r_1, r_2に対する ε-NFA M_1, M_2が存在する．
$M_1=(Q_1, \Sigma, \delta_1, q_1, \{f_1\})$, $M_2=(Q_2, \Sigma, \delta_2, q_2, \{f_2\})$とする．
一般性を失わず，$Q_1 \cap Q_2 = \emptyset$としてよい．
このとき，$M=(Q_1 \cup Q_2 \cup \{q_0, f_0\}, \Sigma, \delta, q_0, \{f_0\})$は補題の条件を満たす ε-NFAである．ここに，δ は次で定義される．

$$\delta(q,c) = \begin{cases} \{q_1, q_2\}, & \text{if } (q,c) = (q_0, \varepsilon) \\ \delta_1(q,c), & \text{if } (q,c) \in Q_1 \times (\Sigma \cup \{\varepsilon\}) - \{(f_1, \varepsilon)\} \\ \delta_2(q,c), & \text{if } (q,c) \in Q_2 \times (\Sigma \cup \{\varepsilon\}) - \{(f_2, \varepsilon)\} \\ \{f_0\}, & \text{if } (q,c) \in \{(f_1, \varepsilon), (f_2, \varepsilon)\} \\ \emptyset, & \text{otherwise} \end{cases}$$

補題2.1の証明(4／4)

3. $r = r_1^*$ (r_1 は正規表現)ならば，帰納法の仮定より，r_1に対する ε-NFAが存在する．
M_1=$(Q_1, \Sigma, \delta_1, q_1, \{f_1\})$ とする．
このとき，M=$(Q_1 \cup Q_2 \cup \{q_0, f_0\}, \Sigma, \delta, q_0, \{f_0\})$は補題の条件を満たす ε-NFAである．ここに，δ は次で定義される．

$$\delta(q,c) = \begin{cases} \{q_1\}, & \text{if } (q,c) = (q_0, \varepsilon) \\ \delta_1(q,c), & \text{if } (q,c) \in Q_1 \times (\Sigma \cup \{\varepsilon\}) - \{(f_1, \varepsilon)\} \\ \{q_1, f_0\}, & \text{if } (q,c) = (f_1, \varepsilon) \\ \emptyset, & \text{otherwise} \end{cases}$$

QED

補題2.2の証明（1／3）

- DFAを $M=(\{q_1,\ldots,q_n\}, \Sigma, \delta, q_1, F)$とする.
- 各$i, j \in \{1,\ldots,n\}$と各$k \in \{0,\ldots,n\}$に対して $R_{i,j}^{k}$ を次の①②を満たす文字列x全体の集合と定める.
  - ① $\delta(q_i, x) = q_j$.
  - ② xの任意の接頭辞 y $(0<|y|<|x|)$に対して$\delta(q_i, y) \in \{q_1,\ldots,q_k\}$.

状態q_iを出発して，途中$q_1,\ldots,q_k$の状態のみを経由しながら状態q_j へ至る状態遷移を引き起こす文字列の集合.

補題2.2の証明(2／3)

- すべての i, j, k について，$R^k_{i,j}$ は正規表現で表せることを帰納法で証明する.
- k=0のとき.
  - $R^0_{i,j}=\{a \mid a \in \Sigma, \delta(q_i, a)=q_j\} \cup \{\varepsilon \mid i=j\}$ が成り立つ.
  - $R^0_{i,j}$ は $\Sigma \cup \{\varepsilon\}$ の部分集合であるから有限集合であり，正規表現で表わせる.

補題2.2の証明（3／3）

- $k>0$のとき.
  - $R_{i,j}^{k}=R_{i,k}^{k-1}(R_{k,k}^{k-1})^{*}R_{k,j}^{k-1}\cup R_{i,j}^{k-1}\quad(k>0)$が成り立つ.
  - 帰納法の仮定より，$R_{i,k}^{k-1}$, $R_{k,k}^{k-1}$, $R_{k,j}^{k-1}$, $R_{i,j}^{k-1}$は正規表現で表すことができる．その正規表現を，順に，$r_{i,k}^{k-1}$, $r_{k,k}^{k-1}$, $r_{k,j}^{k-1}$, $r_{i,j}^{k-1}$とすると，$R_{i,j}^{k}$は以下の正規表現で表せる.

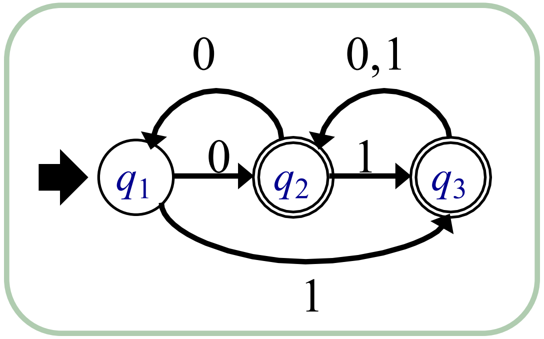
$$r_{i,k}^{k-1}(r_{k,k}^{k-1})^{*}r_{k,j}^{k-1}+r_{i,j}^{k-1}$$

  - $L(M)=\bigcup_{q_j\in F}R_{1,j}^{n}$ であるから，$F=\{j_1,\ldots,j_p\}$ とすると，$L(M)$は以下の正規表現で表せる.

$$r_{1,j_1}^{n}+\cdots+r_{p,j_p}^{n}$$

QED

例



	k=0	k=1	k=2	k=3
$r^k_{1,1}$	ε	ε	$(00)^*$	
$r^k_{1,2}$	0	0	$0(00)^*$	
$r^k_{1,3}$	1	1	0^*1	
$r^k_{2,1}$	0	0	$0(00)^*$	
$r^k_{2,2}$	ε	$\varepsilon + 00$	$(00)^*$	
$r^k_{2,3}$	1	$1+01$	0^*1	
$r^k_{3,1}$	$\varnothing$	$\varnothing$	$(0+1)(00)^*0$	
$r^k_{3,2}$	$0+1$	$0+1$	$(0+1)(00)^*$	
$r^k_{3,3}$	ε	ε	$\varepsilon + (0+1)0^*1$	

$$
\begin{aligned}
r^3_{1,2} = r^2_{1,3}(r^2_{3,3})^*r^2_{3,2}+r^2_{1,2} &= 0^*1(\varepsilon + (0+1)0^*1)^*(0+1)(00)^*+0(00)^* \\
&= 0^*1((0+1)0^*1)^*(0+1)(00)^*+0(00)^*
\end{aligned}
$$

$$
\begin{aligned}
r^3_{1,3} = r^2_{1,3}(r^2_{3,3})^*r^2_{3,3}+r^2_{1,3} &= 0^*1(\varepsilon + (0+1)0^*1)^*(\varepsilon + (0+1)0^*1) +0^*1 \\
&= 0^*1((0+1)0^*1)^*
\end{aligned}
$$

$$
r^3_{1,2} +r^3_{1,3} = 0^*1((0+1)0^*1)^*(\varepsilon + (0+1)(00)^*)+0(00)^*
$$