- 推荐书目：
    - 《形式语言与自动机导论》 Perter Linz
    - 《オートマトンと形式言語入門》 岡留剛
    - 《オートマトン・言語リトンの基礎》 米田政明

有限状态自动机

Alphabet 与文字列

記号的非空有限集合：Alphabet，用Σ表示.　　取有限文字排列（可重复取）：文字列（string），　空文字列：ε　　或"語"

Σ中全部文字列集合：Σ^*，Σ的閉包 closure.　除去空文字列：Σ^+　　即 $\Sigma^* = \{\varepsilon\} + \Sigma^+$

$x = a_1 \cdots a_n \ (n \geq 0;\ a_1 \cdots a_n \in \Sigma)$，$y = b_1 \cdots b_m$

連接 concat　$x \cdot y = xy = a_1 \cdots a_n b_1 \cdots b_m$.　$\forall x \in \Sigma^*$, s.t. $x \cdot \varepsilon = \varepsilon \cdot x = x$　　xy是文字列

長さ　$|x| = n$

言語　Σ^*的子集．对于语言L_1, L_2，積 $L_1 \cdot L_2 = \{xy \mid x \in L_1, y \in L_2\} = L_1 L_2$　　$L_1 L_2$是语言　　有序

$L^0 = \{\varepsilon\}$，$L^n = L^{n-1} \cdot L$，$L^* = \cup_{n=0}^{\infty} L^n$（Kleene closure），$L^+ = \cup_{n=1}^{\infty} L^n$

e.g. $L = \{ab, c\}$，　　言語族：语言的集合

$L^* = \{\varepsilon, ab, c, abc, cab, abab, cc, \cdots\}$

直積集合　笛卡尔积，Cartesian product　$A \times B = \{(x, y) \mid x \in A, y \in B\}$　有序

べき集合　幂集，Power set　$2^A = \{S \mid S \subseteq A\}$　集合的集合

有限状態オートマトン　有限状态自动机：有状态＋执行计算

DFA，決定性有限状态自动机　$M = (Q, \Sigma, \delta, q_0, F)$

Q：状態的非空有限集合.　Σ：记号（输入）的非空有限集合，Alphabet

$\delta: Q \times \Sigma \to Q$　状態遷移関数　　一张表

$q_0 \in Q$　初期状態.　$F \subseteq Q$　受理状態的集合　accepting state　　终态、可为空

決定性：　$Q \times \Sigma \to Q$ 是唯一且确定的．　∀状态 ＋ ∀记号 → 唯一确定的新状态

様相/時点表示：(q, ax)　　q当前状态，当前输入a（下一步），x剩余输入列

$\hat{\delta}: Q \times \Sigma^* \to Q$，s.t. $\forall q \in Q, \hat{\delta}(q, \varepsilon) = q$；　$\forall w \in \Sigma^*, \forall a \in \Sigma, \hat{\delta}(q, wa) = \delta(\hat{\delta}(q, w), a)$

$\hat{\delta}(q, x)$：　文字列x依次读完后的最终状态　　可用δ表示$\hat{\delta}$，都用δ

$(q_0, 1011) \vdash_M (q_1, 011) \vdash_M (q_0, 11) \vdash_M (q_1, 1) \vdash_M (q_2, \varepsilon)$

$(q_0, 1011) \vdash_M^* (q_2, \varepsilon)$

DFA的动作　$(q, ax) \vdash_M (q', w)$，$(q_0, w_0) \vdash_n^* (q_n, w_n)$．$\vdash_M^*$ 为 $\vdash_M$ 的闭包

若 $\delta(q_0, w) \in F$ / $(q_0, w) \vdash_M^* (q, \varepsilon)$，$q \in F$ $\implies$ w可被M受理

Mの受理する言語　$L(M) = \{w \in \Sigma^* \mid \delta(q_0, w) \in F\} = L$

非決定性有限状态自动机 NFA　　$\delta: Q \times \Sigma \to 2^Q$

DFA是特殊的NFA，s.t. $|\delta(q, a)| = a$

受理文字列 $w \in \Sigma^*$，s.t. $\delta(q_0, w) \cap F \neq \phi$

带空转移的有限状态自动机 ε-NFA　　不读取输入，状态发生变化．$\delta(q_1, \varepsilon) \ni q_2$

$\delta: Q \times \Sigma \cup \{\varepsilon\} \to 2^Q$

ε-閉包　e-NFA，$Q', Q'' \subseteq Q$，s.t. $\forall q \in Q', \delta(q, \varepsilon) \ni q'$，$q'$形成的集合构成$Q''$，称为$Q'$的ε闭包.

NFA、DFA等价性　　$\exists$ NFA M，$L = L(M) \iff \exists$ DFA M'，$L = L(M')$　　DFA, NFA, ε-NFA语言受理能力相同

NFA转化为DFA

子集法

从初期状态开始

对于NFA，假设有 $\delta(q_0, a) \vdash \delta(\{q_0, q_1\}, \varepsilon)$，则令 $\{q_0, q_1\}$ 为对应DFA的新状态

对于DFA新状态，只要含有至少1个原NFA中的受理状态，其就为DFA的受理状态

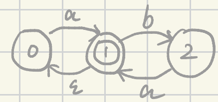
ϕ 称为死状态.

ε-NFA 转化为 NFA

初期状态：原初期状态 + ε达到的状态

当前输入a 能达到的状态 = "ε…εaε…ε" 能达到的状态.

e.g.



DFA:

	a	b
{0}	{0,1}	ϕ
{0,1}	{0,1}	{2}
ϕ	ϕ	ϕ
{2}	{0,1}	ϕ

DFA的最简化

分组法：对状态分组

$M_d \longrightarrow M_s$

① M_d 状态分为受理/非受理 2组

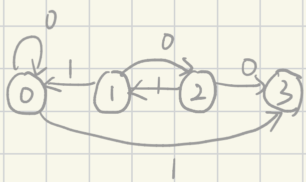
② 在每个组中，将不同输入迁移到同一组状态的状态，移入新组

③ 重复② 直到无新组

④ 把每一组作为 M_s 新状态

含有 M_d {受理、初始} 状态，即为 {受理、初始} 状态

			Md	0	1	表中值为Ms			
0	0	0	0	4	1	(0,0)	(0,0)	(0,3)	
3	0	0	1	2	4	(0,0)	(2,0)	(2,0)	
2	2	0	2	1	3	(0,1)	(0,1)	(3,1)	
1	1	1	3	5	2	(1,0)	(1,2)	(1,2)	
0	0	0	4	0	1	(0,0)	(0,0)	(0,3)	
1	1	1	5	3	2	(1,0)	(1,2)	(1,2)	4



正規表現 正则表达式

描述正则语言的方法

递归定义：$\Sigma \cap A = \phi$，$A = \{), (, \varepsilon, \phi, \cdot, +, * \}$ 即保留字符

Σ 的要素、ε, ϕ 都是正規表現 即还包括空文字列和空集

若 α, β 是正規表現，则 $\alpha+\beta$, $\alpha\cdot\beta$, α^*, (α) 都是正規表現

+：并集 ·：连接 *：闭包

匹配优先度（高→低）

运算

r为正規表現，L(r) 代表 r表示的言語。

ϕ 表示空集，ε 表示 $\{\varepsilon\}$，对于 $\forall a \in \Sigma$，a表示 $\{a\}$

$L(r_1+r_2) = L(r_1) \cup L(r_2)$ 并集

$L(r_1 \cdot r_2) = L(r_1) \cdot L(r_2)$ 连接

$L((r_1)) = L(r_1)$

$L(r_1^*) = (L(r_1))^*$

e.g.

$0(0+1)^*$: 0开始，由0,1组成

$(0+1)^*00$, 以00结尾，由0,1组成

$1^*(01^*01^*)^*$ 由0,1构成，含偶数个0

$ab+ba$: $\{ab, ba\}$

a^* : $\{a^n \mid n \geq 0\}$

$(a^2)^*+(b^3)^*$: $\{a^{2n} \mid n \geq 0\} \cup \{b^{3n} \mid n \geq 0\}$

$aa\cdot(a+b)^*\cdot bbb$: $\{aa\}\{a,b\}^*\{bbb\}$

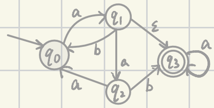
$a^4 \cdot a^* \cdot (\varepsilon+b+b^2+b^3)$

$\{a^n b^m \mid n \geq 4, m \leq 3\}$

正規言語 RL　DFA/NFA能受理的语言

语言L：　∃正規表現 r, s.t. $L=L(r)$ $\longleftrightarrow$ ∃DFA M, s.t. $L=L(M)$

DFA ⟶ r　状态消去法　NFA亦可



$(a(b+aa))^* a(\varepsilon+ab)a^*$

目标：消去除了q_0和受理状态以外的状态。

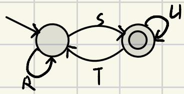
① 优先选出/入迁移之和少的状态 q_i

② 分组 In, Out，分别为迁移进q_i的状态，迁移出q_i，不包含q_i

③ 匹配 In→Out 中的状态，用正規表現写出消去q_i后的迁移动作

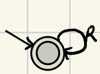
直到仅剩q_0和受理状态。此时：

i) $q_0 \notin F$



$r=(R+SU^*T)^*SU^*$

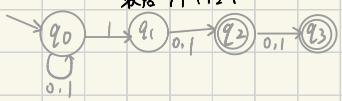
ii) $q_0 \in F$



$r=R^*$

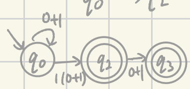
iii) $|F|>1$，考虑每一个受理状态视为唯一受理状态，得到r_i

最后 $r_1+r_2+\cdots$



1. 消q_1　In Out

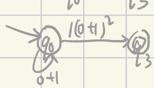
$q_0 \longrightarrow q_2$



2. 视q_3为唯一受理状态

消q_2 In Out

$q_0 \longrightarrow q_3$



$r_1=(0+1)^*\cdot 1\cdot(0+1)^2$

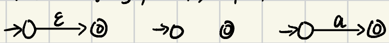
3. q_2为受……消q_3

$r_2=(0+1)^*\cdot 1\cdot(0+1)$

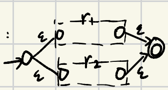
4. $r=r_1+r_2$ ✓

r → ε-NFA　递归转换

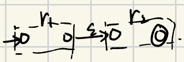
$L(\varepsilon)=\{\varepsilon\}$, $L(\phi)=\phi$, $L(a)=\{a\}$



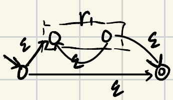
r_1+r_2:



$r_1\cdot r_2$:



r_1^*



形式文法　生成语言 / 自动机:受理语言)

$P_1: S_0 \longrightarrow aS_1$

$P_2: S_1 \rightarrow aS_2$

$P_3: S_2 \rightarrow aS_0$

$P_4: S_2 \longrightarrow a$

$S_0 \rightarrow aS_1 \rightarrow aS_2 \rightarrow aS_0$ / a　$\Rightarrow \{a^{3n} \mid n\geq 1\}$　導出

$G=\langle N,\Sigma,P,S_0\rangle$

N: 非終端記号的有限集合　状态（表示要继续生成）

Σ: 終端記号的有限集合　字母表

P: 導出規則的有限集合　$u\rightarrow v$, $u,v\in(N\cup\Sigma)^*$

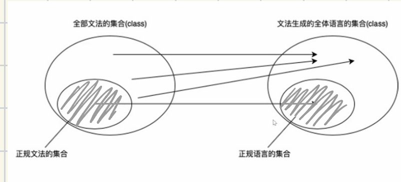
S_0　初期記号　$S_0\in N$

正规文法/正则文法(RG)生成==正规言語==(RL)

① $A \to aB$ / $A \to a$，$A, B \in N$，$a \in \Sigma$

② 允许 $S_0 \to \varepsilon$

$\{L(M) \mid M$是有限状态自动机$\} = \{L(G) \mid G$是正规文法$\}$



NFA $\leftrightarrow$ RG

$q \xrightarrow{a} p$ ：$A_q \to aA_p$；$q \xrightarrow{a} ⓟ$ $\begin{cases} A_q \to aA_q \\ A_q \to a \end{cases}$；$\to q$ ：$S_0 = S_{q_0}$

正規言語泵引理　ポンプの補題/反復補題

已知以下命题相互等价：①L是RL ②∃受理L的NFA和ε-NFA ③∃表示L的RG

e.g 证明 $L_1 = \{a^p b^p \mid p > 0\}$ 不是正规语言　即找不出受理L的DFA　引入泵引理，用以证明L不是正规语言.

假设有DFA M，$L(M) = L_1$.　令 $\delta(q_0, a^i) = q_i$，$i = 1, 2, \cdots$

$N = |Q|$，$\exists j > N$，s.t. $\exists i < j$，$\delta(q_0, a^j) = q_j = q_i$.

已知 $\delta(q_0, a^i b^i) \in F$，则有 $\delta(q_0, a^j b^i) \in F$，即 $a^j b^i \in L(M)$，受理语言不是L_1，矛盾.

即，不存在DFA M，其受理语言为L_1 $\longleftrightarrow$ L_1不是正规语言　uv^*w

☆引理：若L是正则语言，$\exists n \in Z^+$，s.t. $\forall z \in L, |z| \geq n$，有 $z = uvw$，$|uv| \leq n$，$|v| > 0$，$\forall i \geq 0$，$uv^iw \in L$

称n为泵长度。　n与|Q|相关，极端情况下 $n = |Q| + 1$

即，$\exists z \in L, |z| \geq n$，不满足上述，则L不是正则语言。

找一个z，$\forall$ $z = uvw$，$\exists i$ $uv^iw \notin L$

重新证明上述。

假设L_2是RL，泵长度为n，$z = a^n b^n = uvw$

$\because |uv| \leq n$　$\therefore uv = a^{|uv|}$.　$w = a^t b^n$，$t \geq 0$ ($t = n - |uv|$)

令 $i = 0$，$uv^0w = a^{n-|v|} b^n \notin L$ (去掉v)

《形式——》4.3节，P81

闭性 closure property

e.g $\forall a, b \in Z^+$，s.t. $a + b \in Z^+ \longrightarrow Z^+$对"+"运算封闭

Z^+对乘法封闭，对 −/÷ 不封闭。

μ是集合F上的一个n元运算。$A \subset F$ 对μ封闭：$\forall x_1, \cdots x_n \in A$，$\mu(x_1, \cdots x_n) \in A$.

正规语言运算闭性：①正规语言的class对连接，并，闭包 封闭。

证：L_1, L_2为正规语言。则$\exists r_1, r_2$，$L(r_1) = L_1$，$L(r_2) = L_2$

则 $r_1 + r_2$，$r_1 r_2$，r_1^*为正规表现.　因此 $L_1 \cup L_2$，$L_1 L_2$，L_1^*是正规语言

② 对补集封闭

$L \subseteq \Sigma^*$，$\bar{L} = \Sigma^* - L$

交换受理状态

证：$L(M) = L$，$M = \langle Q, \Sigma, \delta, q_0, F \rangle$　取 $M' = \langle Q, \Sigma, \delta, q_0, Q - F \rangle$，有 $L(M') = \Sigma^* - L$ □.

③ 交集封闭.

证：L_1, L_2为正规语言，$\bar{L_1}, \bar{L_2}$也是. 则 $L_1 \cap L_2 = \overline{\overline{L_1 \cap L_2}} = \overline{\bar{L_1} \cup \bar{L_2}}$，由上，□

泵引理的更多例子:

1. $L=\{ww^R \mid w\in\{0,1\}^*\}$ 　　$z=a^n b^n b^n a^n$
2. $L=\{a^n \mid n$是完全平方数$\}$ 　　$z=a^{n^2}$
3. $L=\{a^n \mid n\geq 2$ 质数$\}$

$p\geq n$ 最小质数，$z=a^p$ 　$z_i=uv^iw=a^{p+(i-1)|v|}$. 若 $i-1=p$，$p+p|v|=p(|v|+1)$ 为合数.

下推自动机 PDA (DPDA, APDA) 　Pushdown Automaton

PDA = ε-NFA + 栈 (LIFO)

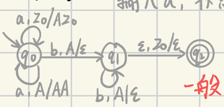
动作：状态变为 q，(q,α) 将栈顶替换为 α

由 (p,a,Z) 决定：(当前)状态 p，输入记号 a，栈顶记号 Z

当前栈顶 　目标栈顶

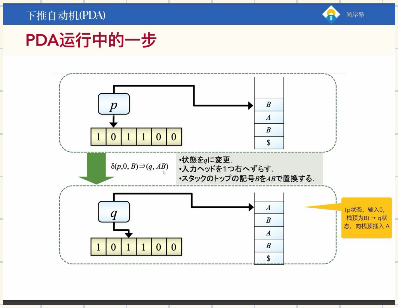
$\delta(p,a,Z)=(q,\alpha)$ 　　(Z_0表示栈底)

输入 a，状态 $p\to q$，栈顶 $Z\to\alpha$ 　　Z_0表示空 / ε表示空：2种规定



输入 aaabbb：

一般用空转移排出 Z_0 栈底



$(q_0, aaabbb, Z_0)\vdash(q_0, aabbb, AZ_0)\vdash(q_0, abbb, AAZ_0)$

$\vdash(q_0, bbb, AAAZ_0)\vdash(q_1, bb, AAZ_0)\vdash(q_1, b, AZ_0)\vdash(q_1,\varepsilon,Z_0)\vdash(q_2,\varepsilon,\varepsilon)$ 受理✓

$L=\{a^nb^n \mid n\geq 1\}$

确定性下推自动机 DPDA 　$M=(Q,\Sigma,\Gamma,\delta,q_0,Z_0,F)$

Q：状态非空集合 　Σ：字母表 　Γ：栈字母表 　$q_0\in Q$ 初期状态 　$Z_0\in\Gamma$ 栈底符号(初期記号) 　$F\subseteq Q$ 受理状态集合

δ：$(Q\times(\Sigma\cup\{\varepsilon\})\times\Gamma)$ 的子集 $\longrightarrow Q\times\Gamma^*$，状态迁移函数

存在空转移 　　并非所有组合都可迁移

确定性：

① 对 $\delta(q_i,\varepsilon,A)$，保证仅存在这唯一的迁移 (对于 (q_i,A))

② 对于未定义的组合，不受理

关于受理状态且空栈：$L(M)=\{w\in\Sigma^* \mid (q_0,w,Z_0)\vdash_M^*(q,\varepsilon,\varepsilon),\ q\in F\}$

关于受理状态：确定性上下文无关语言 (DCFL). RL是DCFL的子集。

关于空栈：　DPDA M 能受理 w，则 wy 不能受理。

证：对于 w 之后，$\delta(q,\varepsilon,Z_0)\neq\phi$ 　空转移排出 Z_0，这里排 Z_0 若不是空转移，则 w 输入后栈必为空

y 首个字母为 a，$\delta(q,a,X)\neq\phi$，不符合确定性 　　无法接受新输入

NPDA $\delta: (Q\times(F\cup\{\varepsilon\})\times\Gamma)$ の子集 $\longrightarrow 2^{Q\times F^*}$

PDA默认

受理语言：① 受理状态：$(q_0, w, Z_0) \vdash_M^* (q, \varepsilon, \alpha)$，$q\in F$，$\alpha\in F^*$ $L_N(M)$

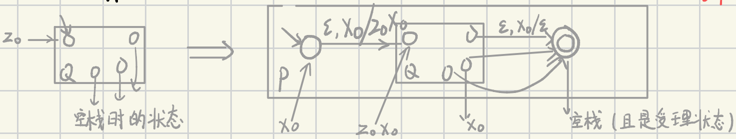
② 空栈 ---- .. $(q, \varepsilon, \varepsilon)$，$q\in Q$ $L_A(M)$

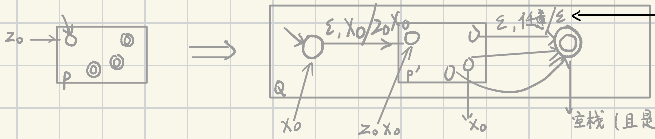
③ both ------ $(q, \varepsilon, \varepsilon)$，$q\in F$ $L(M)$

$L\subseteq\Sigma^*$：$\exists$ NPDA M $L=L(M) \Leftrightarrow \exists$ NPDA M'，$L=L_A(M') \Leftrightarrow \exists$ NPDA M，$L''=L_N(M'')$ 等价性

DPDA不存在此类等价性

证明：



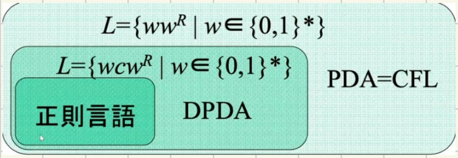


因此不符合DPDA确定性

$\{w\#w^R \mid w\in\{0,1\}^+\}$ 可找到NPDA，DPDA，$\{ww^R \mid w\in\{0,1\}^+\}$ 找不到DPDA.

即，NPDA对语言的受理能力更强。

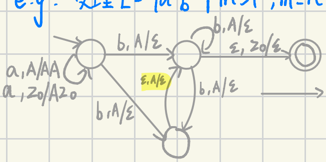
NPDA受理的语言：文脈自由言語（上下文无关语言，CFL） DPDA：确定上下文无关语言



PDA不能受理的语言. e.g. $L=\{a^nb^nc^n \mid n\geq 1\}$，$\{w\#w \mid w\in\{0,1\}^+\}$ 不是文脈自由言語

栈已排空，无法控制c的数 受限于栈结构

e.g. 受理 $L=\{a^nb^m \mid m\geq 1, m\leq n\leq 2m\}$ 的NPDA



$\longrightarrow$ 利用 ε, A/ε 一次就要走 b, A/ε 一次，因此 $m\geq\frac{1}{2}n$

文脈自由文法 CFG $G=\langle N,\Sigma,P,S\rangle$ 生成文脈自由語言

N：非終端記号 Σ：終端記号 $S\in N$：初期記号

$P\subseteq N\times(N\cup\Sigma)^*$ $N\to(N\cup\Sigma)^*$ 箭头左边仅能为1个非终端记号 $A\to\alpha$

A生成的字符串与A的上下文无关

e.g. $G=\langle N,\Sigma,P,S\rangle$，$N=\{S\}$，$\Sigma=\{a,b\}$，$P=\{S\to ab, S\to aSb\}$

$S\to ab \mid aSb$

$\longrightarrow L=\{a^nb^n \mid n\geq 1\}$

括号匹配 $N=\{S\}$，$\Sigma=\{(,)\}$，$P=\{S\to SS \mid (S) \mid ()\}$

CFG 化简 建议顺序：2→3→1，见黑书P109. PDA无所谓最小化，可见 CFG→PDA

1. 除去无效记号 去除所有包含無効記号的規則

① 死记号（无法最终生成终端记号的非终端记号），去除所有包含死记号的规则

$A\to B|\alpha$，A也为生记号

找生记号：某规则：$A\to\alpha$，α为终端记号，则A为生计号

α为终端记号+生计号……

} 重复

e.g. $S\to ABaC,\ A\to BC,\ B\to b|\varepsilon,\ C\to D|\varepsilon,\ D\to d$

② 不可到达记号

2. 除去 ε-生成规则

$\Rightarrow V_N=\{A,B,C\}$

$P:\ S\to ABaC|BaC|AaC|ABa|aC|Aa|Ba|a$

$A\to BC|B|C,\ B\to b,\ C\to D$

① 找形如 $A\to\varepsilon$，A加入V_N ② 重复：$A_1\cdots A_n\in V_N$，$B\to A_1\cdots A_n$ 则 B加入V_N ③ 对于剩下的规则，用ε代替任意个RHS的可空记号

3. 除去单位生成规则 e.g. $A\to B$ 即仅改名

$A\to a,\ B\to A|b \Rightarrow A\to a,\ B\to a|b$

① 找形如$A\to B$ ② 找 $A\xrightarrow{*}B$ ③ 对于 $A\xrightarrow{*}B$，若存在$B\to\alpha$，则添加$A\to\alpha$

CFG 标准型

CNF，GNF

注意RHS为单变量，即不允总以下替换：$A\to BC,\ B\to b,\ C\to c \longrightarrow A\to bc$

Chomsky CNF：$A\to BC,\ A\to a$. $[S\to\varepsilon]$

Greiback GNF：$A\to a\alpha\ (a\in\Sigma,\alpha\in N^*)$

对于任意不产生ε的CFL L，$\exists$ CNF G_1，GNF G_2，s.t. $L(G_1)=L(G_2)=L$

$A\longrightarrow X_1\ldots X_N\in P$，$X_i\in N\cup\Sigma$，$n\geq 2$ CFG → CNF

① $X_i=a\in\Sigma$，$\Rightarrow A\to X_1\cdots X_{i-1}X_i'X_{i+1}\cdots X_N$，$X_i'\to a$ 使RHS只有非终端记号

② 对于 $A\to X_1\cdots X_N$，$X_i\in N$，$n\geq 3$，$\Rightarrow A\to X_1Y_1,\ Y_1\to X_2Y_2,\cdots\ Y_{n-3}=X_{n-2}Y_{n-2},\ Y_{n-2}\to X_{n-1}X_n$ 拆

CFG ⟶ CNF ⟶ GNF

$A\xrightarrow{*}A\gamma,\ \gamma\in(N\cup\Sigma)^*$：A为左递归的. $A\to A\gamma\to A\gamma\gamma\to A\gamma\gamma\gamma\cdots$ 存在左递归的非终端记号→左递归的文法

GNF 不允许左递归

① 消 $A\to B\gamma$：$B\to\beta_1,\ B\to\beta_2\ldots\beta_i\in(N\cup\Sigma)^*\Rightarrow A\to\beta_1\gamma,\ A\to\beta_2\gamma\ldots$

② 消 $A\to A\gamma$：$A\to\beta_1,\ A\to\beta_2\ldots\ \Rightarrow A\to\beta_1Z,\ A\to\beta_2Z\ldots,\ Z\to\gamma,\ Z\to\gamma Z$ 改为右递归

重复① ②

e.g. $S\to AA,\ S\to 0,\ A\to SS,\ A\to 1$

$S\to AA$：$S\to SSA,\ S\to 1A$

$S\to SSA$：$S\to 0Z,\ S\to 1AZ,\ Z\to SA,\ Z\to SAZ$

$A\to SS$：$A\to 0S,\ A\to 1AS,\ A\to 0ZS,\ A\to 1AZS$

$Z\to SA$：$Z\to 0A|1AA|0ZA|1AZA$

$Z\to SAZ$：$Z\to 0AZ|1AAZ|0ZAZ|1AZAZ$ All done。

CFL泵引理

求长度

CFL L, $\exists n\in Z^+$, $\forall z\in L, |z|\ge n$, $z=uvwxy$, s.t. $|vwx|\le n$, $|vx|>0$, $\forall i\ge 0$, $uv^iwx^iy\in L$

$(A\to BC, A\to\alpha)$

对于 CFG CNF G，由于非终端记号个数有限，则随着导出文字列长度增加(至n)时，恰有某非终端记号(A)被导出2次。

(当长度大于 $|N|$ 时，必定存在这样的A —— 鸽巢定理)

再次出现

$S \xrightarrow[G]{*} uAy \xrightarrow[G]{*} uvAxy \longrightarrow uvwxy = z$　　若A再次出现k次，则导出 uv^kwx^ky　　计算理论导引 P77

$A\xrightarrow[G]{*} vAx$　　$A\xrightarrow[G]{*} w$

e.g. 证明 $L=\{a^nb^nc^n|n>0\}$ 不是 CFL　　CFL可以数2个变量，不能数3个，可见NPDA受理语言。

假设 L是CFL，泵长度为p　　令 $z=a^pb^pc^p=uvwxy$

$\because |vwx|\le p$ $\therefore$ vwx有 $\underbrace{a\cdots a}_{①}\underbrace{b\cdots b}_{②}\underbrace{c\cdots c}_{③}$ ④ ⑤ 分类讨论。　即改变i，只能改变abc中{1,2}个元素长度 □

CFL的闭性

$G_1=\langle N_1,\Sigma,P_1,S_1\rangle$, $G_2=\langle N_2,\Sigma,P_2,S_2\rangle$.　$L(G_1), L(G_2)\in CFL$

$G=\langle N,\Sigma,P,S\rangle$, $S\notin N_1\cup N_2$　　利用CFG证明

并　$N=N_1\cup N_2\cup\{S\}$, $P=P_1\cup P_2\cup\{S\to S_1|S_2\}$, $L(G)=L_1\cup L_2$

连　- - - - - - -　$P=P_1\cup P_2\cup\{S\to S_1S_2\}$, $L(G)=L_1L_2$

闭包　$N=N_1\cup\{S\}$, $P=P_1\cup\{S\to SS_1|\varepsilon\}$, $L(G)=L_1^*$

对交集不封闭　e.g. $L_1=a^nb^nc^m$, $L_2=a^nb^mc^m$, $L_1\cap L_2=a^nb^nc^n\notin CFL$

对补集不封闭　$L_1,L_2\in CFL$, $\overline{\overline{L_1}\cup\overline{L_2}}=L_1\cap L_2\notin CFL$ □

CFG→PDA

CFG $G=\langle N,\Sigma,P,S\rangle$, NPDA $M=\langle Q,\Sigma,\Gamma,\delta,q_0,Z_0,F\rangle$

$Q=\{q_0,q_1,q_2\}$, $\Gamma=\{Z_0\}\cup N\cup\Sigma$, $F=\{q_2\}$

δ: $(Q\times(\Sigma\cup\{\varepsilon\})\times\Gamma)$'子集 $\to 2^{Q\times\Gamma^*}$

$\delta(q_0,\varepsilon,Z_0)\to\{(q_1,SZ_0)\}$　　空转移压入S

★ $\delta(q_1,\varepsilon,A)\to\{(q_1,\alpha)|A\to\alpha\in P\}$　$A\to\alpha$　栈顶为非终端记号，利用空转移替换，作任意次ε，栈内使终为LZ_0

★ $\delta(q_1,a,a)\to\{(q_1,\varepsilon)\}$　栈顶与输入同时为终端记号a，弹出a　。仅当输入=L时，才能清空栈，即受理。

$\delta(q_1,\varepsilon,Z_0)\to\{(q_2,\varepsilon)\}$　　结束

PDA→CFG

$q_f \xrightarrow{\varepsilon,Z_0/\varepsilon} f$

既使M仅有一个受理状态，$|F|=1$

CFG $G=\langle N,\Sigma,P,S\rangle$, NPDA $M=\langle Q,\Sigma,\Gamma,\delta,q_0,Z_0,F\rangle$

构造 $M'=\langle Q',\Sigma,\Gamma,\delta',q_0,Z_0,F'\rangle$, $Q'=Q\cup\{f\}$, $F'=\{f\}$. $\forall q_f\in F$, 有 $\delta'(q_f,\varepsilon,Z_0)=\delta'(q_f,\varepsilon,Z_0)\cup\{(f,Z_0)\}$, 其它一致。

$N=\{L(q,Z,p)|q,p\in Q',Z\in\Gamma\}\cup\{S\}$　其中 $L(q,Z,p)$ 是一个非终端记号，也表示状态 $q\to p$，Z出栈。

对于 P

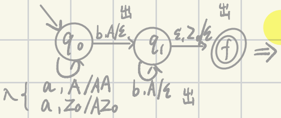
$p\xrightarrow{a,X/Z_1\cdots Z_k} r$　　$p\to r$　　$r\xrightarrow{弹出Z_1} s_1$　　$s_{k-1}\xrightarrow{弹出Z_k} q$

有 $(r,Z_1\cdots Z_k)\in\delta(p,a,X)$, s.t. $L(p,Z,q)\to aL(r,Z_1,s_1)L(s_1,Z_2,s_2)\cdots L(s_{k-1},Z_k,q)$　输入a，出X入$Z_1\cdots Z_k$，再全部弹出

$L(q,Z,p)$ 意味 $q\xrightarrow{Z/\varepsilon}p$，则 $L(q,Z,p)\to$ 栈顶为Z时所有 $q\to p$ 且Z出栈的路径。

有 $(q,\varepsilon)\in\delta(p,a,Z)$，则 $L(p,Z,q)\to a$　　baseline

$p\xrightarrow{a,Z/\varepsilon}q$，读入a，弹出Z，$p\to q$



$S \to L(q_0, Z_0, f)$ 惯例：添加初始状态 $L(q_1, Z_0, f) \to \varepsilon$ 惯例：空转移弹出栈底记号

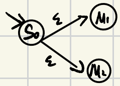
$L(q_0, Z_0, f) \to aL(q_0, A, q_1)L(q_1, Z_0, f)$

$L(q_0, A, q_1) \to aL(q_0, A, q_1)L(q_1, A, q_1)$

$L(q_0, A, q_1) \to b$, $L(q_1, A, q_1) \to b$

DFA的交并补

并：即M_3受理M_1，M_2各自的语言。

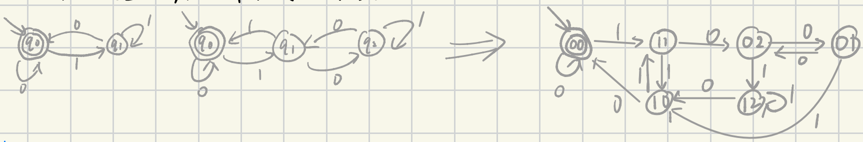


M_1的初期状态

M_2的……

补：M_1 M_2受理/不能受理语言相反：反转受理、非受理状态即可

交：M_3受理M_1，M_2都能受理的语言

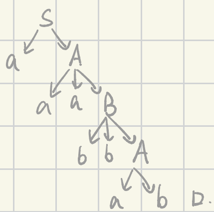


導出木 推导树

構文解析 parsing语法分析，判断文字列w是否属于G　　另方法 CYK.

推导树即为parsing结果

e.g. $P=\{S\to aA, S\to bB, A\to aaB, A\to ab, B\to bbA, B\to bA\}$ 判断aaabbab



【定义】CFG $G = <N, \Sigma, P, S>$ 的導出木（推导树）是满足下述条件的树

- 1. 以 $A \in N \cup \Sigma$ 为标签，且只有一个顶点 A 的树是推导树
- 2. 对于各个生成规则 $A \to \epsilon \in P$，以记号 ϵ 为标签的节点是顶点 A 的孩子的树是推导树
- 3. 若数个推导树的根节点 $B_1, \ldots, B_k \in N \cup \Sigma$ 的子节点分别是 $T_1, \ldots, T_k$，那么对于生成规则 $A \to B_1 \ldots B_k$，以 A 为根节点，将前述推导树的根节点 $B_1, \ldots, B_k$ 作为 A 的子节点的树是推导树

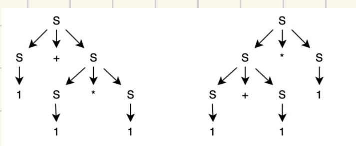
△ LL文法　左至右读入w，提前向右查看k个字符便能parsing的G：LL(k)文法

用LL分析文法LL(k)：自顶向下分析　　LR分析：自底向上。

曖昧さ 歧义性　导出w时存在多个推导数。

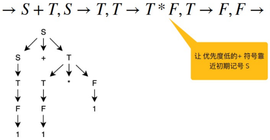
消歧义：一般越靠近S优先级越低

△ e.g. $P=\{S\to 1, S\to S+S, S\to S*S\}$, $w=1+1*1$



文法 G 的规则为 $P = \{S \to 1, S \to S + S, S \to S * S\}$，已知在 G 中 * 记号的优先度高于 + 记号（即优先匹配 * 记号的规则），请构造文法去除 G 中的歧义性

$P' = \{S \to S + T, S \to T, T \to T * F, T \to F, F \to 1\}$



最左/右导出：对最左/右的非终端记号进行替换

图灵机 (TM)

有限个状态。读写头 Head 向左、右移动，读/写。纸带长度无限。无数据，填入 Blank (B)。

① 读入当前指向记号 a.

② 当前状态为q. (q, a) 决定: 状态迁移/a位置写入替换/左右移动 Head　　$\delta(p,1) \ni (q,0,R)$　状态 $p \to q$，写入0，右移

③ $\delta(p,b) = \phi$，受理状态、停止。否则执行①②.　可以无法停止。　停机问题

$M = \langle Q, \Sigma, \Gamma, \delta, q_0, B, F \rangle$

Q: 状态，$\Sigma \subseteq \Gamma$: 输入记号，Γ: 纸带记号 (包括B,等)

δ: $(Q \times \Gamma)$'s 子集 $\longrightarrow Q \times \Gamma \times \{L, R\}$

$q_0 \in Q$: 初始状态，$B \in \Gamma - \Sigma$: 空白记号，$F \subseteq Q$ 受理状态集合

图灵完备: 一个计算模型可模拟任何图灵机。

句構造文法　PSG

$G = \langle N, \Sigma, P, S \rangle$

$S \in N$ 开始记号

P: $\alpha \longrightarrow \beta$ $(\alpha \in (N \cup \Sigma)^* N (N \cup \Sigma)^*, \beta \in (N \cup \Sigma)^*)$　左侧的α至少含有一个非终端记号

句構造言語　PSL

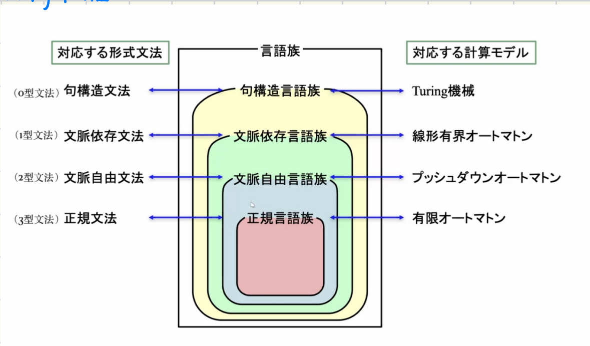
线性有界自动机 LBA

左右有¢、\$作为边界。

文脈依存文法　CSG　$G = \langle N, \Sigma, P, S \rangle$

$S \to \varepsilon$, $\alpha A \beta \longrightarrow \alpha \gamma \beta$ $(A \in N, \gamma \in (N \cup \Sigma)^+, \alpha, \beta \in (N \cup \Sigma)^*)$

Chomsky 階層



利用闭性证明语言非正则

$L = \{w \mid w$ 中 a, b 数量相同$\}$.　假设L是RL，则 $L \cap a^* b^* = a^n b^n \in RL$，但 $a^n b^n \notin RL$.

设计自动机

从 NFA 开始设计

$\Sigma = \{0, 1\}$，输入二进制，受理3的倍数的DFA