LFC (Linguaggi Formali e Compilatori) - Note del Corso

Edoardo Lenzi

November 25, 2017

Contents

Chapter 1

Link

Vecchio sito
Note Drive

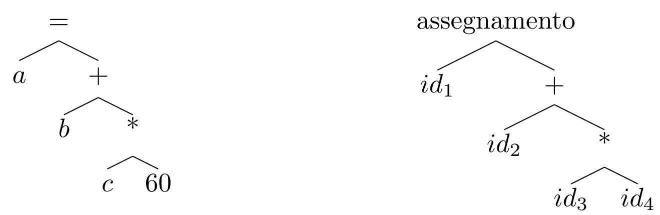
Chapter 2

Introduzione

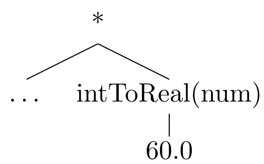
I linguaggi di **analisi lessicale** lavorano con simboli e caratteri; devo costruire una **tavola dei simboli** (specifica per un dato programma e compilatore). L'analisi restituisce dei **tokens** (puntatori) a record nella tavola dei simboli. La maggior parte delle implementazioni usano un numero come **identificatore**.

L'analisi sintattica invece studia la **grammatica del linguaggio**. Viene costruito un **abstract syntax tree**:

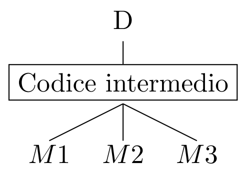
$$a = b + c \cdot 60$$



L'analisi semantica si occupa di vedere se c'é una corretta semantica (variabili dichiarate precedentemente). Se * necessita di un float allora 60 dev'essere convertito a float.



Generazione di **codice intermedio**: $\begin{array}{ll|l} temp_1 = & \text{intToReal}(60) & \text{VISITA DELL'ALBERO} \\ temp_2 = & id_3 * temp1 & \\ temp_3 = & id_2 + temp2 & \\ id_1 = & temp_3 & \end{array}$



2.1 Grammatiche

Una **grammatica** é una tupla G(V, T, S, P) con:

V	vocabolario
T	set simboli terminali
S	start symbol
P	set delle produzioni
V\T	simboli
ε	parola vuota, non puó essere un terminale!

Le produzioni sono della forma $\alpha \to \beta$ con α **stringa non vuota di simboli V con almeno un non terminale**, β stringa su V o ε.

Per convenzione i caratteri in maiuscolo denotano simboli non terminali mentre in minuscolo terminali. Quindi i simboli in T sono tutti lettere minuscole.

Considero X, Y variabili, generico simbolo in V e α β δ stringhe su V^*
$S \to aSb$
$S \to \varepsilon$ $\quad T = \{a, b\}$ (terminali), $(V \backslash T) = \{S, A\}$ (non terminali)
$S \to A$

2.1.1 Derivazioni

Date $\mu = \mu_1 \alpha \mu 2,\ \alpha \to \beta$, produzioni della grammatica G, $\gamma = \mu_1 \beta \mu_2$ é una derivazione in uno o piú passi partendo da μ.

Scrivendo $\mu \to^+ \gamma$ intendo che posso derivare γ da μ in uno o piú passi di derivazione.

2.1.2 Linguaggio generato da una grammatica

$$L(G) = \{w \ / \ w \in T^*, S \to_G^+ w\}$$

Dato il linguaggio L possono esistere piú grammatiche diverse tra loro che generano L. Pertanto dal linguaggio non posso risalire con certezza alla grammatica.

In generale dato un linguaggio generale L ed una grammatica G $\nexists$ un algoritmo per dimostrare che L = L(G).

Chapter 3

Linguaggi liberi

3.1 Grammatica libera dal contesto (context free)

Una grammatica generata é libera dal contesto (context free) se ogni sua produzione ha la forma:

$$A \to \beta$$

Con A non terminale. ($\alpha \to \beta$, α deve essere un solo simbolo non terminale, altrimenti é condizionata).
Con β qualsiasi (terminali, non terminali o ε).

Grammatiche libere si prestano in modo naturale a descrivere derivazioni in viste ad albero.

3.1.1 Esempi

$S \to aAb$
$aA \to aaAb$ Non é context free, genera $L(G) = \{a^n b^n \ / \ n > 0\}$.
$A \to \varepsilon$

$S \to aSb|\varepsilon$
$S \to aSb \to aaSbb \to aabb \implies \{a^n b^n \ / \ n \geq 0\}$ Context free, genera lo stesso L di quella sopra.
Serve per parentesizzare correttamente codice (genera $a^n b^n$).

$S \to AB$
$A \to Aa|a$ Tutto ció che deriva da A é indipendente da ció che deriva da B. $a^n b^m, \ n, m \geq 0$
$B \to Bb|b$

Se ho produzioni impossibili che non finiscono in terminali la grammatica genera un linguaggio vuoto ($(\{S, B, a\}, \ \{A\}, \ S, \ \{S \to aB\})$).

$(\{S\}, \ \{\}, \ S, \ \{S \to \varepsilon\})$ invece genera un linguaggio $\{\varepsilon\} \neq \emptyset$

$S \to 0B|1A$
$A \to 0|0S|1AA$ L(G) = {w / count(0,w)=count(1,w) }
$B \to 1|1S|0BB$

Definire una grammatica per $L = (a^k b^n \ / \ k, n > 0)$

$S \to aS|aB$
$B \to bB|b$

$S \to AB$
$A \to aA|a$
$B \to bB|b$

$S \to ab|aS|Sb$

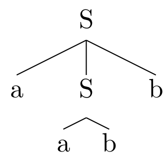
Definire G tale che $L(G) = \{a^k b^n c^{2n} \ / \ k, n > 0\}$
$S \to AB$
$A \to aA|a$
$B \to bBcc|bcc$

Definire G tale che $L(G) = \{a^k b^n d^{2k} \ / \ k, n > 0\}$
$S \to aSdd|B$
$B \to bB|b$

$S \to aSdd|aBdd$
$B \to bB|b$

3.2 Abstract syntax Tree

$$S \to aSb|ab$$



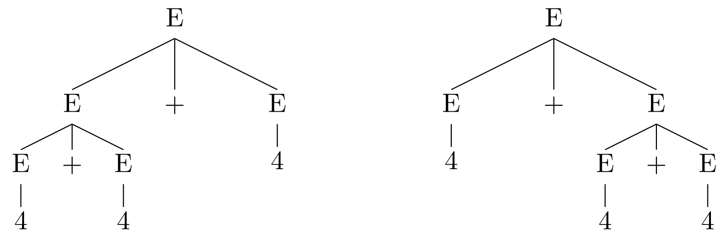
aabb, derivazione canonica $\mu \rightarrow \gamma$

3.3 Grammatiche ambigue

Nel caso di grammatiche libere si definiscono le **derivazioni canoniche destre e sinistre**, nel caso **rightmost** si richiede che ad ogni passo di derivazione $\mu \rightarrow \gamma$ venga rimpiazzato il terminale piú a destra di μ; nel caso **leftmost** quello piú a sinistra.

G é **ambigua** se esiste una parola del linguaggio generato da G, generabile con due derivazioni canoniche distinte entrambe destre o entrambe sinistre.

$E \rightarrow E + E | E * E | 4$ (il + associa a sinistra)



[Analogo con il *] Essendo derivazioni entrambe leftmost **G é ambigua**.

Occhio a non confondere la struttura dell'albero di derivazione con la sua sequenza di derivazioni.
La derivazione leftmost sostituisce prima i non terminali a sinistra e poi procede con i successivi.
Quello a sx prima spacca la E in E+E che poi viene spaccato in E+E dove poi vanno sostituiti alle E i non terminali 4.
Quello a dx invece spacca E in E+E, sostituisce alla prima E il 4 e poi passa alla sostituzione della seconda E con E+E.
In entrambi i casi la sostituzione dei non terminali avviene sempre prendendo il primo non terminale della stringa, cioé quello piú a sinistra.

Finché considero sostituzioni con un solo carattere a destra della produzione leftmost e rightmost sono del tutto equivalenti; la differenza arriva quando considero produzioni con sostituzioni su piú di un carattere perché mangi caratteri a possibili derivazioni future.

$S \rightarrow if\ b\ then\ S\ |\ if\ b\ then\ S\ else\ S\ |\ altro$
$if\ b\ then\ if\ b\ then\ altro\ else\ altro$
$if\ b\ then\ S\ else\ S$

3.4 Linguaggi liberi

Un linguaggio é libero se esiste una grammatica libera che lo genera.

Lemma 1: La classe dei linguaggi liberi é **chiusa all'unione** (se L_1 e L_2 sono liberi allora $L1 \cup L2$ é ancora libero)

$$\begin{aligned} L_1 \; libero &\implies \exists G_1 = (V_1, T_1, S_1, P_1) \; / \; L_1 = L(G_1) \\ L_2 \; libero &\implies \exists G_2 = (V_2, T_2, S_2, P_2) \; / \; L_2 = L(G_2) \\ (\{V_1 \cup V_2 \cup \{S\}, T_1 \cup T_2, &S, P_1 \cup P_2 \cup (S \to S_1 | S_2)\}) \end{aligned} \tag{3.1}$$

Devo rinominare i non terminali di G1 e G2 in modo da non avere omonimie (clash). Notare la produzione che agginge un nuovo start symbol per agganciarsi ai vecchi.

Lemma 2: La classe dei linguaggi liberi é **chiusa rispetto alla concatenazione** (se L1 e L2 sono liberi allora $\{w_1 w_2 \; / \; w_1 \in L_1, \; w_2 \in L_2\}$ é esso stesso un linguaggio libero).

$$G = (V_1 \cup V_2' \cup \{S\}, \; T_1 \cup T_2, \; S, \; P_1 \cup P_2' \cup \{S \to S_1 S_2'\})$$

$P_1 \cup P_2 \cup \{S \to S_1 S_2\}$ anche in questo caso bisogna stare attenti ad eliminare possibili clash dei simboli non terminali di G_2. Metto un apice per indicare una rinominazione dei non terminali ($V_2', \; S_2', \; P_2'$). In pratica concateno gli start symbols.

3.4.1 Esempio

Il linguaggio $\{a^n b^n \; / \; n > 0\}$ é libero perché $\exists$ una grammatica libera che lo genera (G_1):

$$G_1 \qquad S \to aSb/ab \; G_1 \text{ libera}$$
$$G_2 \qquad s \to aAb, \; A \to aaAb, \; A \to \varepsilon \; G_2 \text{ non libera} \; \square$$

3.5 Pumping Lemma per Linguaggi Liberi

Serve per dimostrare che un linguaggio non é libero. Ossia non esiste alcuna grammatica libera che lo genera.

Pumping lemma:
Sia L un linguaggio libero allora $\exists p \in \mathbb{N}^+ \; / \; \forall z \in L \; / \; |z| > p, \exists uvwxy$:

i) $z = uvwxy \; \wedge$

ii) $|vwx| \geq p \; \wedge$

iii) $|vx| > 0 \; \wedge$

$\forall i \in \mathbb{N} \; / \; uv^i wx^i y \in L$

3.5.1 Commento definizione

$\exists \; p \in \mathbb{N}^+$	esiste una costante $p > 0$
$\forall \; z \in L : \; \lvert z \rvert > p$	ogni parola con piú elementi di p
$\exists \; uvwxy \; / \; z = uvwxy$	esistono 5 sottostringhe che compongono z
$\lvert vwx \rvert \leq p$	la lunghezza delle 3 stringhe centrali é minore di p
$\lvert vx \rvert > 0$	la seconda e la quarta non sono mai entrambe nulle
$\forall \; i \in \mathbb{N} uv^i wx^i y \in L$	se ripeto i volte (i puó essere 0) la 2 e la 4 sono ancora in L

3.5.2 Dimostrazione Pumping Lemma

Osservazione 1: Data una grammatica G posso sempre creare una altra grammatica G' modificata dalla prima che genera lo stesso linguaggio.

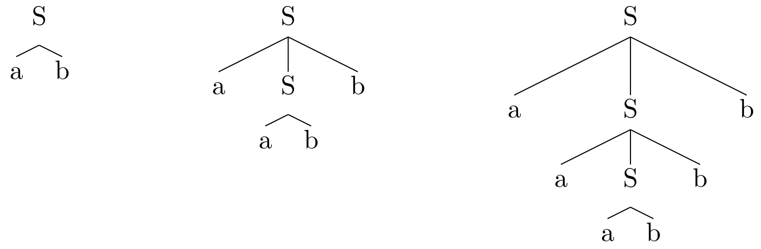
Osservazione 2: Una grammatica si puó portare in forma normale (di Chomsky) togliendo eventuali ridondanze o produzioni che terminano per forza con ε (a meno che non debba fare parte del linguaggio, ma a quel punto si scrive come $S \to \varepsilon$).

CNF

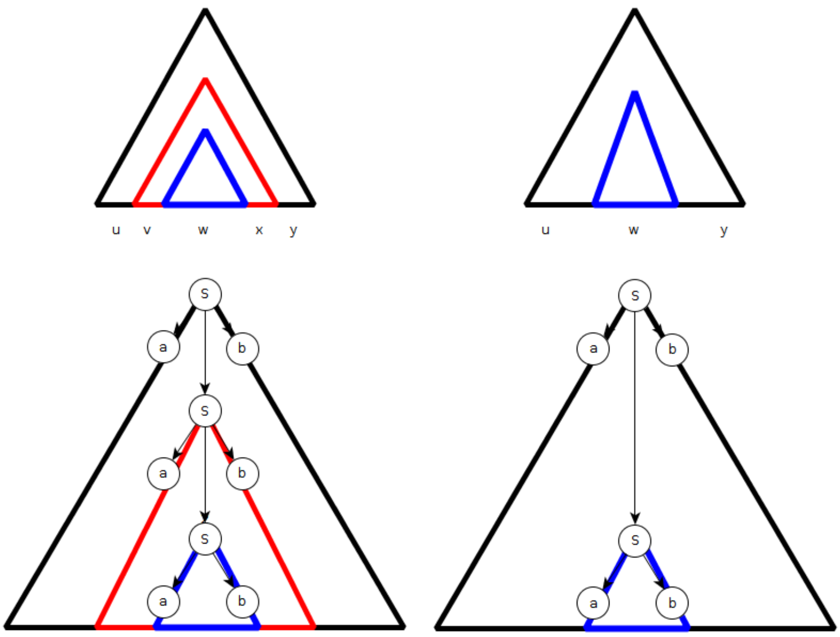
Chomsky Normal Form esige che una grammatica non abbia produzioni ridondanti:
ie) G_1 $S \to aSb|ab|B,\ B \to aBb|ab \leftarrow$ doppioni

Dimostrazione: L é un linguaggio libero $\implies \exists\, G$ in Chomsky Normal Form tale che L = L(G).

Definisco p come la lunghezza della parola piú lunga che puó essere derivata usando un albero di derivazione i cui cammini dalla radice sono lunghi al piú come il numero di simboli terminali della grammatica ($|V \backslash T|$).
$S \to aSb|ab$, ha due terminali, p = 2



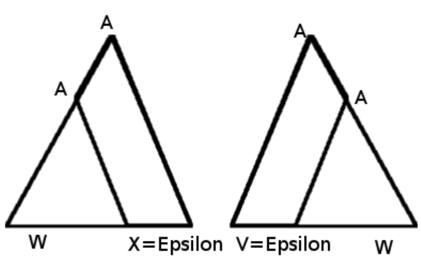
Guardo il cammino S $\to S_1 \to S_2 \to \ldots \to S_K$ di lunghezza K Se $z \in L\ \wedge\ |z| > p \implies$ nell'albero di derivazione di z esiste almeno un cammino la cui lunghezza é maggiore di $|V \backslash T| \implies$ allora esiste almeno un non-terminale che occorre almeno due volte lungo quel cammino (S, nell'esempio sotto).



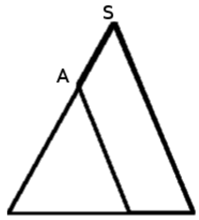
ie) p = 2, se prendo una qualunque parola piú lunga di 3 generata da G: aaaabbbb, la posso dividere in 5 con due pumpable: u=aa, v=a, w=abb, x=b, y=b

Se prendo $z \in L \wedge |z| > p \implies$ ho dovuto usare un albero di derivazione $/\ \exists$ al meno un cammino piú lungo di $|V \backslash T|$ per definizione di z. $\implies$ ho un non terminale ripetuto al meno due volte $\implies \exists$ al meno un non terminale che occorre al meno 2 volte lungo quel cammino

con **l'un-pumping** la parola sta sempre nel linguaggio (**taglio un pezzo di albero**) Dato che w e x non possono essere entrambi nulli al massimo avró $A \to aA,\ o\ A \to Aa$



ie) linguaggio libero $\{a, b\},\ S \to ab$



3.6 Pumping Lemma per assurdo

(tesi) = L libero $\implies \exists p \in \mathbb{N}^+ \ / \ \forall z \in L \ / \ |z| > p, \exists uvwxy$:

i) $z = uvwxy \wedge$

ii) $|vwx| \geq p \wedge$

iii) $|vx| > 0 \wedge$

$\forall i \in \mathbb{N} \ / \ uv^i wx^i y \in L$

$\neg(tesi) = \forall p \in \mathbb{N}^+ \ / \ \exists z \in L \ / \ |z| > p \ \forall \ uvwxy \ / \ \ [(z = uvwxy \wedge |vwx| \leq p \wedge |vx| > 0)$
$\implies \exists i \in N \ / \ uv^i wx^i y \notin L]$

Se si verifica la tesi negata il linguaggio non é libero.

Suppongo $L_1 = w_1 w_2 \ / \ w_1 = w_2, \ w_1, w_2 \in \{a, b\}^*$ **libero**;
Sia p un naturale qualunque sempre positivo;
Sia $z = a^p b^p a^p b^p$ $(|z| = 4p)$;
Allora $z \in L_1, \ |z| > p$
Siano $uvwxy \ / \ z = uvwxy, \ |vwx| \leq p \wedge |vx| > 0$ distinguiamo vari casi:

1) vwx é composto da 'a' che occorrono a sinistra (w_1)

2) vwx é a cavallo e contiene sia 'a' che 'b' in w_1

3) vwx contiene solo 'b' in w_1

4) 'b' in w_1 e 'a'$\in w_2$

5,6,7) ...speculare su w_2

- Nei casi 1, 3, 5, 7 considero le parole $z^1 = uv^0 wx^0 y$ (i=0);

- nel caso 1 sono certo di togliere alcune occorrenze di a quindi avró $z^1 = a^k b^p a^p b^p, \ k < p \implies z^1 \notin L$.

- Nel caso 3 $z^1 = a^p b^k a^p b^p, \ k < p \implies z^1 \notin L$.

- Nel caso 5 $z^1 = a^p b^p a^k b^p, \ k < p \implies z^1 \notin L$.

- Nel caso 7 $z^1 = a^p b^p a^p b^k, \ k < p \implies z^1 \notin L$.

- Nei casi 2, 4, 6 invece avró ancora $z^1 = uv^0wx^0y$ (i=0);

- Nel caso 2 $z^1 = a^kb^pa^pb^p,\ o\ a^pb^ka^pb^p,\ o\ a^jb^ka^pb^p,\ j,k < p \implies z^1 \notin L$.

- ...

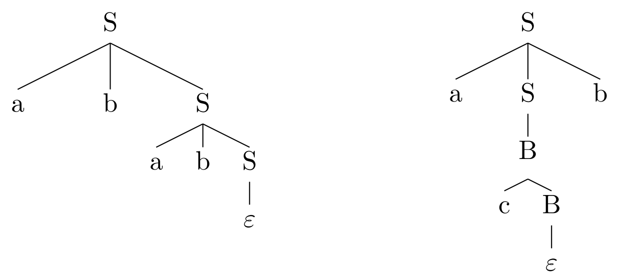
In pratica facendo l'unpumping in tutti i casi $z_1 \notin L$ quindi L non é libero.

3.6.1 How not to use Pumping Lemma

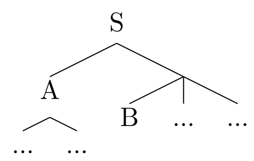
Se avessi preso $\{ww \ / \ w \in \{a,b\}^*\}$ sia $z = (ab)^p(ab)^p$, prendo p=4, $v \in \varepsilon, x = a, i = 0$ cosí non dimostro niente perché se voglio dimostrare con il pumping lemma la negazione della tesi devo dimostrare un asserto che vale $\forall p \in \mathbb{N}^+$. Pertanto non posso prendere un arbitrario p=4.

3.6.2 Esempi

ie) $\{a^nb^nc^j/n,j \geq 0\} = L_{17}$
$S \to aSb|B$
$B \to cB|\varepsilon$
$acb \in L_{17}$



$S \to abA|B$
$B \to cB|\varepsilon$
s



$S \to AB$
$A \to aAb|ab|\varepsilon$
$B \to Ab|c|\varepsilon$

3.6.3 Esempi Linguaggi Liberi

Essendo un linguaggio libero chiuso rispetto alla concatenazione, dati:

$L_1 = \{a^nb^nc^j \ / \ n,j \geq 0\}$ Libero perché concatenazione di $\{a^nb^n \ / \ n \geq 0\}$ e $\{c^j \ / \ j \geq 0\}$, entrambi liberi
$L_2 = \{a^jb^nc^n \ / \ n,j \geq 0\}$ Libero, inverso di L_1
$L_3 = \{a^nb^nc^n \ / \ n \geq 0\}$ Non é libero:
Suppongo L_3 libero, sia $p \in \mathbb{N}^+$, $z = a^pb^pc^p$ Allora $z \in L_3$, $|z| = 3p > p$
Spacco z in $A = a...a,\ B = b...b, C = c...c$
Siano $z = uvwxy \ \wedge \ |vwx| \leq p \ \wedge \ |vx| > 0$:

- vwx é composto da sole a in A
- vwx é composto da a in A e b in B
- vwx é composto da sole b in B
- vwx é composto da b in B e c in C
- vwx é composto da sole c in C

Considero la parola $z' = uv^0wx^0y$

1. $z' = a^k b^p c^p,\ k < p,\ z' \notin L_3$

3. $z' = a^p b^k c^p,\ k < p,\ z' \notin L_3$

5. $z' = a^p b^p c^k,\ k < p,\ z' \notin L_3$

2. $z' = a^k b^j c^p,\ k < p\ \vee\ j < p,\ z' \notin L_3$

4. $z' = a^p b^k c^j,\ k < p\ \vee\ j < p,\ z' \notin L_3$

Quindi visto che la parola non appartiene mai ad L_3 il linguaggio non é libero. □

Nota che $L1$ ed $L2$ risultano liberi anche facendo pumping lemma per assurdo perché nel caso in cui vwx cada nel terminale ripetuto j volte con l'unpumping la stringa appartiene comunque al linguaggio (quindi ho almeno un caso in cui appartiene al linguaggio e non posso applicare il pumping lemma per assurdo).

Quindi la classe di linguaggi liberi **non é chiusa rispetto all'intersezione**

$L_4 = \{a^n b^m c^{n+m}\ /\ n, m > 0\}$ Libero
$S \rightarrow aSc|aBc$
$B \rightarrow bBc|bc$

$L_5 = \{a^n b^m c^n d^m | n, m > 0\}$ Non libero
$L_6 = \{wcw^R\ /\ w \in \{a, b\}^+\}$ Libero
$S \rightarrow aSa|bSb|aca|bcb$

Chapter 4

Automi a stati finiti

Un NFA accetta/riconosce un certo linguaggio.

Sia N un NFA, allora il linguaggio riconosciuto/accettato da N é il set delle parole per le quali esiste almeno un cammino dallo stato iniziale di N ad uno stato finale di N.

notare che $\forall\ a \in A,\ a\varepsilon = \varepsilon a = a$.

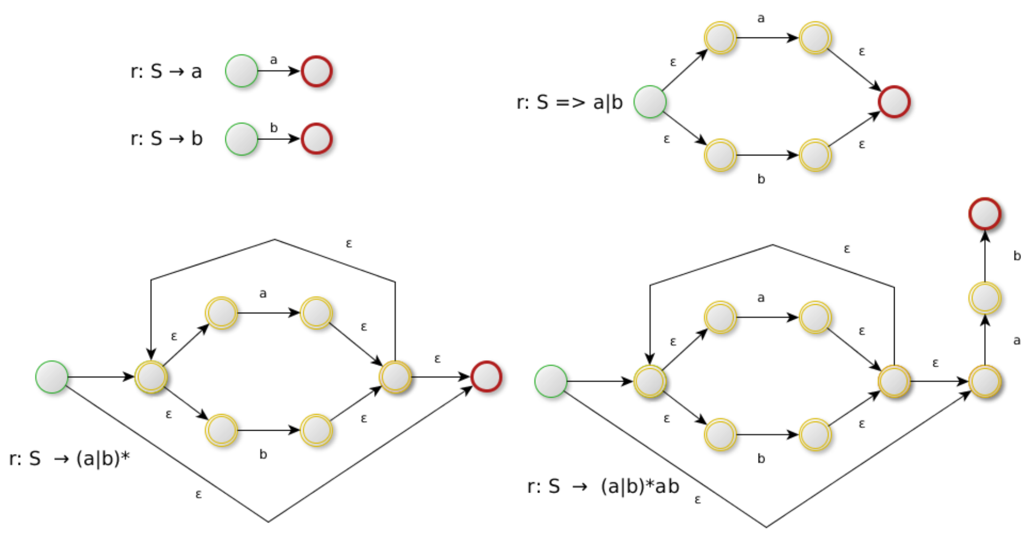
4.1 Thompson construction

input	regular expression r
output	NFA N $/ \ L(N) = L(r)$

Gli NFA usati nei passi della costruzione hanno:

- un solo stato finale
- non hanno archi entranti sul nodo iniziale
- non hanno archi uscenti dal nodo finale

Lemma: Lo NFA ottenuto dalle costruzini di Thompson ha al massimo 2k stati e 4k archi, con k lunghezza della re. r. **Osservazione**: Ogni passo della costruzione introduce al massimo 2 nodi e 4 archi.



Algoritmo a complessitá $O(|r|)$

4.2 Simulare un NFA

Il backtracking consiste nel seguire un percorso e se non va bene tornare in dietro e provarne un altro finché alla fine li provo tutti mal che vada.

$N = (S, A, move_n, S_0, F)$, S insieme stati, A i non terminali (label degli archi), S_0 stato iniziale, F set stati finali, $move_n$ funzione di transizione ($S \otimes A \to S$)
$t \in S, T \subset S$

4.2.1 $\varepsilon - closure$

> $\varepsilon - closure(\{t\})$ il set degli stati S raggiungibili tramite zero o piú $\varepsilon - transizioni$ da t (in pratica il nodo stesso e tutti i nodi raggiungibili con una $\varepsilon - transition$ [**applicato ricorsivamente**]).

Nota che $\forall t \in S,\ t \in \varepsilon - closure(t) \qquad \varepsilon - closure(T) = \cup_{t \in T} \varepsilon - closure(t)$

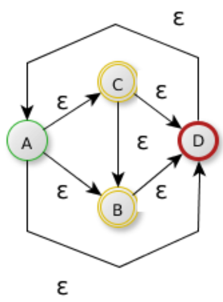
Questo algoritmo é piú performante del backtracking.

4.2.2 Algoritmo per la computazione $\varepsilon - closure$

Strutture dati:

- pila
- bool[] alreadyOn, dimensione $|S|$
- array[][] $move_n$

```
for(int i = 0; i < |S|; i++){
    alreadyOn[i] = false;
}
closure(t, stack){
    push t onto stack;
    alreadyOn[t] = true; //posso sempre arrivare a me stesso con una epsilon-transition
    foreach(i in move_n(t, epsilon)){
        if(!alreadyOn[i]){
            closure(i, stack);
        }
    }
}
```




```
alredyOn[F F F F];
closure(A, pila vuota)
    [A] [T F F F]
        //B non e' ancora nella pila
        closure(B, [A])
            [A, B] [T T F T]
            closure(D, [A, B])
                [A, B, D] [T T F T]
        closure(C, [A, B, D])
            [A, B, C, D] [T T T T]
```

4.2.3 Algoritmo per la simulazione di un NFA

input	NFA N, w$
output	yes se $w \in L(N)$, no altrimenti

```
N = (S, A, move_n, S_0, F)
states = epsilon-closure({S_0})
symbol = nextchar()
while(symbol != $){
    states = epsilon-closure(Unione_{t in states} di move_n(t, symbol));
    //l'insieme di tutti gli stati raggiungibili con la sottostringa letta fin ora
    symbol = nextchar();
}
if(states intersecato F != emptyset){
    return yes;
}
return no;
```

Algoritmo a complessitá $O(|w|(n+m))$

4.3 DFA

Automa a stati finiti, deterministico; una sottoclasse degli NFA che rispettano:

$$\text{DFA} \triangleq (S, A, move_d, s_0, F)$$
$$move_d \triangleq (S \otimes A) \rightarrow S$$

- non hanno $\varepsilon - transizioni$

- $\forall a \in A, s \in S, \ move_n(s,a)$ é un unico stato se ho una **funzione di transizione totale** (al piú uno stato se ho una **funzione di transizione parziale**)

Sink è il nodo pozzo dove **confluiscono tutte le transizioni non segnate** (per ogni stato se mi manca una transizione per un determinato terminale ne faccio una su sink); viene aggiunto per rendere la funzione di transizione una funzione di transizione totale

4.3.1 Linguaggio riconosciuto dal DFA

Dato il DFA D, L(D) é il linguaggio riconosciuto da D.
$L(D) = \{w = a_1, ..., a_k$ / $\exists$ cammino in D dallo stato iniziale al finale$\}$. $\varepsilon \in L(D) \iff s_0 \in F$.

4.3.2 Simulazione di un DFA con $move_d$ totale

input	w$, DFA $D = (S, A, move_d, F)$
output	yes se $w \in L(D)$, no altrimenti

```
state = s_0;
while(symbol != $ && state != bottom){
    //move_d(s, a) = bottom <=> move_d non e' definita su (s,a)
    //se sono in bottom sono in sink
    state = move_d(state, symbol);
    symbol = newxtchar();
}
if(state \in F)
    return yes;
return true;
```

Simulazione NFA costa $O(|w|(n+m))$ Simulazione DFA costa $O(|w|)$

4.4 Subset Construction

input $NFA(S^n, A, move_n, S_0^n, F^n)$
output $DFA(S^d, A, move_d, S_0^d, F^d)$

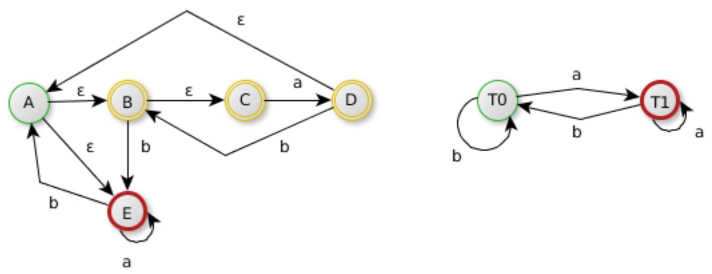
```
S_0^d = epsilon-closure({S_0^n});
 //raggruppo stati della epsilon closure in un unico stato S_0^d del DFA
states = {S_0^d};
tag S_0^d come non marcato;

while(exist T in states non marcato){
   marco T;
   foreach(a in A){ //guardo ogni arco
      T_1 = epsilon-closure(U_{t in T} di move_n(t,a));
         //tutti gli stati raggiungibili con una a-transition da uno stato in T
         //poi la loro epsilon closure
      if(T_1 != emptySet){
         move_d(T, a) = T_1;
         if(T_1 !in states){
            aggiungi T_1 a states come non marcato;
         }
      }
   }
}

foreach(T in states){
   if( (T intersecato F^n) != 0){
      metti T in F^d;
   }
}
```

Lo stato iniziale del DFA sará la $\varepsilon - closure$ dallo stato iniziale del NFA (quindi un set di stati). Considero lo stato iniziale del NFA e lo marco in grassetto poi espando T_0 con la $\varepsilon - closure$ dello stato iniziale.

Dallo stato T_0 guardo per ogni arco gli stati in cui arrivo e li marco in grassetto (T_1, T_2, ...); poi espando quelli in grassetto guardando le rispettive $\varepsilon - closure$.

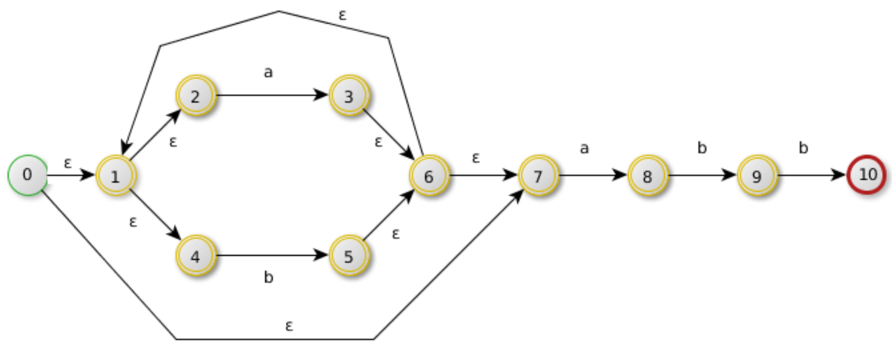
Alla fine guardo i set degli stati se due set coincidono mergio gli stati.

4.4.1 Esercizio

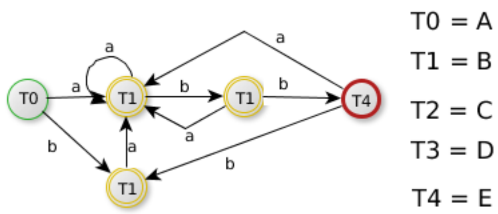


States	**a**	**b**
T0 = { **A** B C E }	T1	T2 = T0
T1 = { A B C **D** **E** }	T1	T0
T2 = { **A** B C **E** } = T0 (quindi T0 va in T0 tramite b)	come T0	come T0

4.4.2 Esercizio



States	a	b
S_0^d = { **0** 1 2 4 7 }	T1	T2
T1 = { 1 2 **3** 4 6 7 **8** }	T1	T3
T2 = { 1 2 4 **5** 6 7 }	T1	T2
T3 = { 1 2 4 **5** 6 7 9 }	T1	T4
T4 = { 1 2 4 **5** 6 7 **10** }	T1	T2



4.5 Partition Refinement

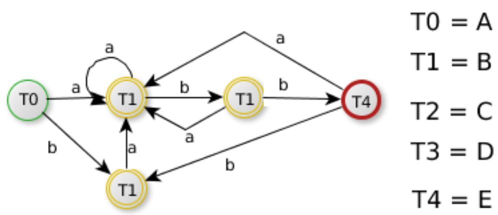
Guado gli archi, se tutta la partizione punta ad un nodo dell'altra transizione con lo stesso non terminale allora va bene; altrimenti spacco la partizione.

4.5.1 Algoritmo di Partition Refinement

Input	DFA D = $\{A, A, move_d, s_0, F\}$
Output	partizione di S in blocchi equidistanti

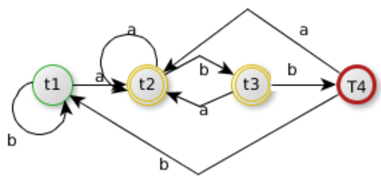
```
B_1 = F;
B_2 = S \ F;
P = {B_1, B_2};
while(exists B_i, B_j in P, exists a in A, B_i e'' partizionabile rispetto a (B_j, a)){
    sostituire B_i in P con split(B_i, (B_j, a));
}
```

4.5.2 Esempio



{ A B C D } { E }	Considero le partizioni dei terminali e non terminali
{ A B C D } { E }	Con a-transizione non esco dal primo set
{ A B C } { D } { E }	Con b-transizione vado da D in E (e A B C non vanno in E con b-transizioni)
{ A B C } { D } { E }	Con a-transizione non esco
{ A C } { B } { D } { E }	Con b-transizione vado da B in D e gli altri no quindi splitto
{ A C } { B } { D } { E }	vanno bene

Rinomino { A C } { B } { D } { E } in t_1, t_2, t_3, t_4



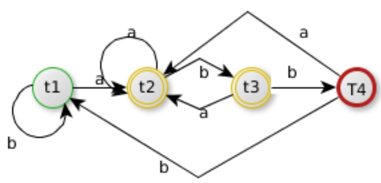
4.6 Algoritmo di minimizzazione di DFA

Input DFA $\mathbf{D} = \{S, A, move_d, s_0, F\}$ con $move_d$ totale
Output minimo DFA (**min(D)**) che riconosce lo stesso linguaggio del primo

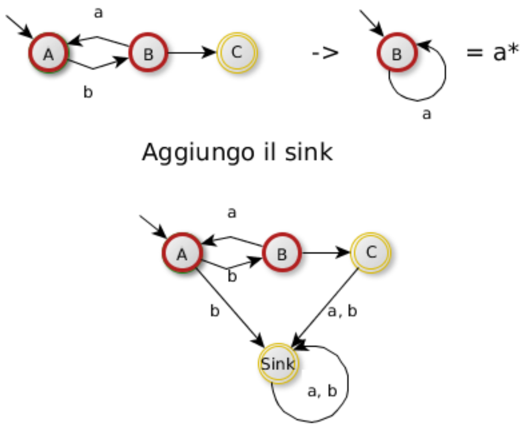
```
P = PartitionRefinement(DFA D);
// P = (B_1, ..., B_k);
foreach(B_i in P){
    var t_i;          //do un nome alla partizione
    if(s_o in B_i){
        t_i e'' iniziale per min(D); //setto lo stato iniziale di min(D)
    }
}

foreach(B_i in P, B_i in F){
    t_i e'' lo stato finale di min(D); //setto lo stato finale di min(D)
}

foreach( (B_i, a, B_j) tale che esiste s_i in B_i tali che move_d(s_i, a) = s_j){
    //per ogni tupla (stato, arco, stato) faccio la rispettiva transizione in min(D)
    setto una transizione temporanea in min(D) da t_i a t_j secondo il simbolo a;
}

foreach(dead state t_i){
    rimuovere t_i e tutte le transizioni da/verso t_i;
}
tutti i temporanei residui (sia stati che transizioni) sono gli stati e le transizioni
     di min(D);
```

Complessitá $O(nlgn)$.

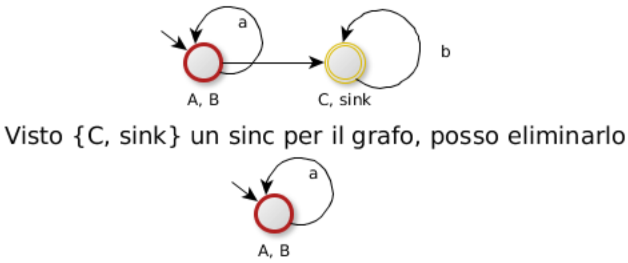
4.6.1 Esempio



Arrivato qua: rinominati { A C } { B } { D } { E } in t_1, t_2, t_3, t_4, applico la minimizzazione del DFA.

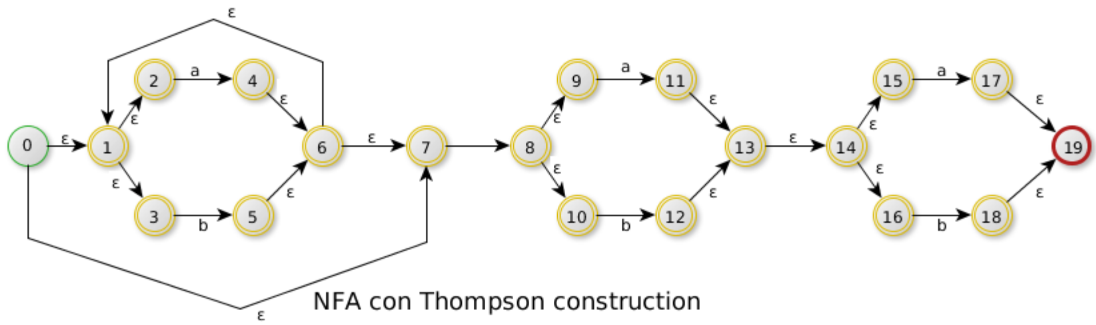


Ho le partizioni $\{A,\ B\}$, $\{C,\ sink\}$, applico partition refinement ma sono giá partizionati correttamente.

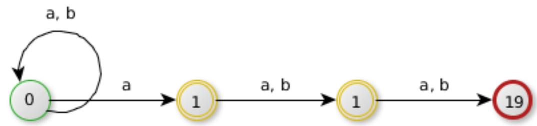


4.6.2 Esempio

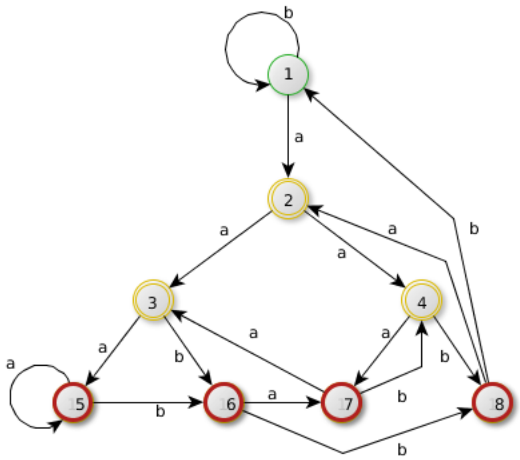
Sia $r = (a|b) * a(a|b)(a|b)$, per determinare il minimo DFA di riconoscimento di L(r) posso usare Thompson e spararmi in faccia o andare a naso.



NFA con Thompson construction



NFA a naso

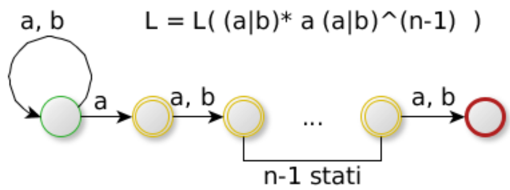


DFA Totale

4.6.3 Lemma

Lemma: $\forall n \in \mathbb{N}^+$, $\exists$ un NFA con (n+1) stati il cui minimo DFA equivalente ha almeno 2^n stati.

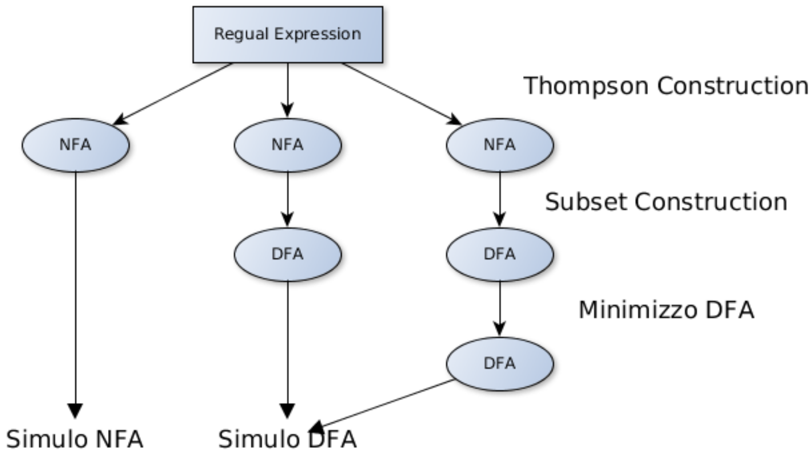
Dim:



Per assurdo suppongo esista un DFA minimo con meno di 2^n stati.
Osservo che esistono 2^n possibili parole di lunghezza n con simboli $\{a, b\}$.
$\Longrightarrow$ $\exists\, w_1 \neq w_2 \;/\; |w_1| = |w_2| = n$ e il loro riconoscimento conduce allo stesso stato del DFA.
$\Longrightarrow$ esiste almeno una posizione per cui w_1 e w_2 differiscono (considero quella piú a destra).

$w_1 = w_1'ax$, $w_2 = w_2'bx$ iniziano diversi ma finiscono con x entrambe. Considero $w_1'' = w_1'ab^{n-1}$ $w_2'' = w_2'bb^{n-1}$ raggiungo uno stato finale t; la seconda parola peró non appartiene al linguaggio, nonostante possa comunque raggiungere lo stato t. Pertanto contraddiciamo che t sia finale.

4.7 Ricapitolando



Algoritmo	Complessitá nello spazio	Complessitá nel tempo
Thompson Construction	$O(\|r\|)$	$O(\|r\|)$ || $O(n_n + m_n)$
Simulazione NFA	-	$O(\|w\|(n_n + m_n))$
Subset Construction	-	$O(n_d\|A\|(n_n + m_n))$
Minimazzazione DFA	-	$O(n_d lg(n_d))$
Simulazione DFA	-	$O(\|w\|)$

Chapter 5

Linguaggi Regolari o Lineari

5.1 Da DFA a Grammatica Regolare

Una grammatica é regolare se le produzioni sono della forma: $A \to \beta$, con β terminale non-terminale, viceversa o terminale e basta.

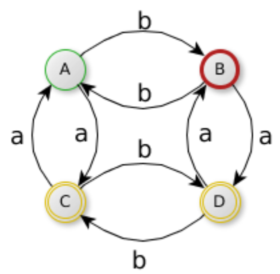
$$\begin{array}{lll} A \to aB & B \to b & \text{grammatica lineare destra} \\ B \to Ab & A \to a & \text{grammatica lineare sinistra} \end{array}$$

Dato un DFA D voglio trovare una grammatica regolare G tale che L(G) = L(D). Se ho una transizione $A \to B$ con una a-transizione diventerá $A \to aB$. Segno il nome del nodo che sto considerando prima della freccia e, dopo la freccia, il non terminale ed il nodo destinazione. Se ho un nodo foglia C avró $C \to \varepsilon$.

Se invece ho una grammatica regolare e voglio trovare un DFA D / $L(G) = L(D)$, faccio il procedimento inverso a prima; se ottengo un NFA basta fare Subset Construction.

5.1.1 Esempio

$L = \{w \ / \ w \in \{a,b\}^* \&\& |a| \ pari, \ |b| \ dispari\}$, L é regolare?



Sí é regolare.

5.1.2 Considerazioni

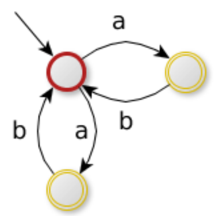
Regular expression, NFA e DFA hanno la stessa potenza espressiva, sono solo notazioni diverse.

Dal DFA posso sempre costruirmi una **grammatica regolare** equivalente.

Non devo fare l'errore di assumere che qualsiasi grammatica sia esprimibile attraverso un NFA.

5.1.3 Esempio

$L = \{w \ / \ w \in \{a,b\}^* \&\& |a| = |b|\}$, L é regolare?



Non potrá mai essere regolare, per il pumping lemma per i linguaggi regolari.

5.2 Pumping Lemma per Linguaggi Regolari

Sia L un linguaggio regolare $\implies \exists \ p \in \mathbb{N}^+ \ / \ \forall \ z \in L \ / \ |z| > p, \ \exists \ u, v, w \ / :$

i) $z = uvw \ \wedge$

i) $|uw| \leq p \ \wedge$

i) $|v| > 0, \ \forall i \in \mathbb{N}, \ uv^i w \in L$

5.2.1 Dimostrazione

L é regolare quindi puó essere riconosciuto da un automa a stati finiti.
Sia D il min DFA $\ / \ L(D) = L, \ p = |S|$, allora i cammini piú lunghi che non passano piú di una volta nel medesimo stato hanno al piú lunghezza (p-1).
Allora se $z \in L$ con $|z| > p$, z é riconosciuta tramite un cammino che attraversa almeno due volte uno stato.

5.2.2 Negazione testi Pumping Lemma per linguaggi regolari

$\forall p \in \mathbb{N}^+ \ / \ \exists \ z \in L \ / \ |z| > p. \ \forall \ uvw \ z = uvw \wedge |uw| \leq p \wedge |v| > 0) \implies \exists \ i \in \mathbb{N} \ / \ uv^i w \notin L)$

Lemma: $L = \{a^n b^n \ / \ n \geq 0\}$ non é regolare

Dim: Assumo per assurdo che L sia regolare, dato p un qualunque numero positivo e $z = a^p b^p$ allora $\forall \ uvw \ / \ z = uvw \wedge |uw| \leq p \wedge |v| > 0$ (la stringa v contiene solo (e almeno una) 'a').

allora uv^2w ha la forma $a^{p+k}b^p, \ k > 0$ allora $uv^2w \notin L$ il che contraddice il Pumping Lemma per linguaggi regolari.

5.2.3 Esercizio

$L_1 = \{w \ / \ w \in \{a,b\}^*$e contiene almeno una occorrenza di "aa "$\}$

L_1: $A \to aA|bA|aB$
$B \to aC$
$C \to aC|bC|\varepsilon$

$L_2 = \{ww \ / \quad \in \{a,b\}^*\}$
non é libero per il pumping lemma (giá dimostrato), quindi non é regolare.

$L_3 = \{ww^r \ / \ w \in \{a,b\}^*\}$
Non é regolare ma libero. $z = a^p b^p b^p a^p \in L_3$ visto che $uv < p$, uv é composta solo da a $uv^i w = a^p b^{2p} a^p \notin L_3$ quindi non puó essere regolare.

5.2.4 Esercizi di esame

Sia N_1 lo NFA con stato iniziale A e finale E con la seguente funzione di transizione:

	ε	a	b
A	$\{B, E\}$	∅	∅
B	$\{C\}$	∅	$\{E\}$
C	∅	$\{D\}$	∅
D	$\{E\}$	∅	$\{B\}$
E	∅	$\{E\}$	$\{A\}$

1) $aa \in L(N_1)$?

2) D é il DFA ottenuto da N_1, per subset construction, Q stato iniziale di D, $Q_{ab\_}$ lo stato di D che si raggiunge da Q tramite il cammino ab. Dire a quale sottoinsieme degli stati di N_1 corrisponde $Q_{ab\_}$.

1) Sí facendo $A \to B \to C \to D \to E \to E$ 2) Facendo la subset construction:

	a	b
$Q0 = \{A, B, C, D\}$	Q1	Q0
$Q1 = \{D, E\}$	Q2	Q0
$Q2 = \{E\}$	Q2	Q0

Chapter 6

Analisi Sintattica

$S \to cAd$
$A \to ab|a$

6.1 Parsing Top-down

Parto dal starting symbol ed espando le derivazioni dando prioritá alle derivazioni piú a sinistra. Cerco quindi di ricostruire una derivazione leftmost della stringa w data in input.

$w\$, \$ \notin V$
$w = cabd$

Per ricostruire la parola w parto dalla prima derivazione $S \to cAd$ derivo la A piú a sinistra (leftmost) e posso scegliere fra a ed ab; scelgo a e mi accorgo che ho sbagliato, torno in dietro e scelgo ab.

6.2 Parsing Top-down predittivo (o non ricorsivo)

Cambio la grammatica sopra in: $S \to cAd$ $A \to aB$ $B \to b|\varepsilon$ Cosí non sbaglio

6.2.1 Grammatica LL(1)

prima L: leggiamo la input string da sinistra
seconda L: ricostruiamo una leftmost derivazione
(1): decidiamo quale operazione effettuare guardando un solo simbolo in input
Le grammatiche LL(1) sono un subset delle grammatiche libere.

6.3 Algoritmi di Parsing

input buffer	w$
stack	bottom[$] top
parsing table	con tante righe quante non terminali, tante colonne quante terminali ($ incluso) in ogni cella metto un'eventuale trasformazione o "error "

6.3.1 Algoritmo di parsing non-ricorsivo

input	stringa w, tabella parsing non ricorsivo T, per G
output	derivazione leftmost di w se $w \in L(G)$, error() altrimenti

```
//init
buffer = {w$};
stack.push($S);

let b il primo simbolo di w
let x il top dello stack

while(x != $){
    if(x == b){
        pop(x);
```

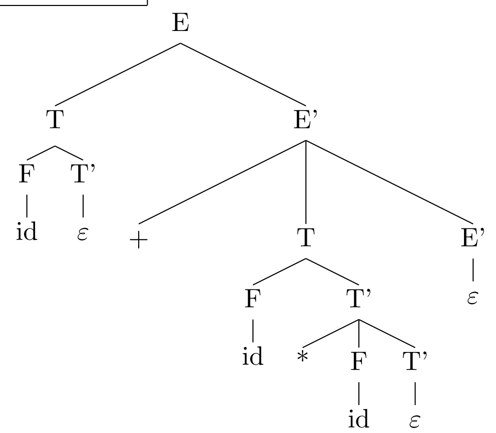
```
        let b il simbolo necessario di w;
    } else if(x e'' terminale){
        error();
    } else if(T[x,b] contiene X -> Y1...Yn){
        return X -> Y1...Yn;
        pop(x);
        push(Yn...Y1)
    }
    let x il top dello stack
}
```

6.3.2 Esempio

$E \rightarrow TE'$
$E' \rightarrow +TE'|\varepsilon$
$T \rightarrow FT'$
$T' \rightarrow *FT'|\varepsilon$
$F \rightarrow id$

	id	+	*	$
E	$E \rightarrow TE'$			
E'		$E' \rightarrow TE'$		$E' \rightarrow \varepsilon$
T	$T \rightarrow FT'$			
T'		$T' \rightarrow \varepsilon$	$T' \rightarrow *FT'$	$T' \rightarrow \varepsilon$
F	$F \rightarrow id$			

pila	input	output
$\$\underline{E}$	$\underline{id}$+id*id$	$E \rightarrow TE'$
$\$\ E\ \underline{T'}$		$E \rightarrow TE'$
$\$ET'\underline{F}$		$F \rightarrow id$
$\$\ ET'\ \underline{id}$		
$\$E'\underline{T''}$	$+\underline{id}$*id $	$T'' \rightarrow \varepsilon$
$\$\underline{E'}$		$E' \rightarrow TE'$
$\$\ E'T\ \underline{+}$	$\underline{id}$*id$	
$\$\ E'\underline{T}$		
	...Avanti così	



6.3.3 Esercizio

$S \rightarrow aA|bB$
$A \rightarrow c$
$B \rightarrow d$

$w = ac\$$

Parsing: $S \implies aA \implies ac$

Nella tabella metto le produzioni della grammatica:

	a	b	c	d	$
S	$S \to aA$	$S \to bB$			
A			$A \to c$		
B				$B \to d$	

6.4 Calcolo First

Data una generica $\alpha \in V^*$ per G=(V, T, S, P), first(α) é l'insieme dei simboli terminali b tali che $\alpha \implies bv$. Inoltre se $\alpha \implies \varepsilon$ allora $\varepsilon \in first(\alpha)$

6.4.1 Esercizio

$S \to A|B$
$A \to a|C$
$C \to \varepsilon$
Allora first(A) = $\{a, \varepsilon\}$ (ε perché posso fare $A \implies C \implies \varepsilon$).

6.4.2 Esercizio

$S \to A|B$
$A \to a|C$
$C \to bB$
Allora first(A) = $\{a, b\}$ (b perché posso fare $A \implies C \implies bB$, ma B non esiste).

6.4.3 Esercizio

$S \to A|B$
$A \to a|C$
$C \to bB$
$B \to c$
Allora first(A) = $\{a, b\}$ ($A \implies C \implies bB \implies bc$, ma tengo solo il primo simbolo (b))

6.4.4 Esercizio

$A \to A|C$
$C \to bB|\varepsilon$
$B \to c$
Allora first(A) = $\{a, b, \varepsilon\}$

G=(V,T,S,P) Sia $X \in V$. L'insieme first(X) viene calcolato come segue:

1) inizializzo first(X) vuoto $\forall\ X \in V$

2) se $X \in T$ allora first(X) = $\{X\}$

3) se $X \to \varepsilon \in P$ allora aggiungere ε ai first(X)

4) se $X \to Y_1...Y_n \in P$, con $n \geq 1$ allora uso la seguente procedura:

```
j = 1;
while(j <= n){
    aggiungere ai first(X) ogni b tale che b in first(Yj)
    if(epsilon in first(Yj)){
        j++;
    } else {
        break;
    }
}

if(j == n+1){
    aggiungere epsilon ai first(X);
}
```

Sia $\alpha = Y_1 ... Y_n,\ n \geq 1$, allora first($\alpha$) é calcolato sotituendo a X alpha:

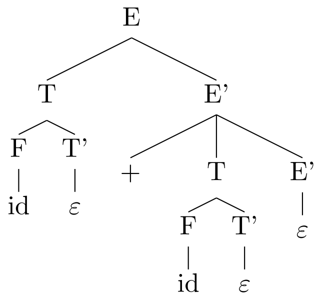
```
j = 1;
while(j <= n){
    aggiungere ai first(alpha) ogni b tale che b in first(Yj)
    if(epsilon in first(Yj)){
        j++;
    } else {
        break;
    }
}

if(j == n+1){
    aggiungere epsilon ai first(alpha);
}
```

6.4.5 Esercizio

$E \to TE'$
$E' \to +TE'|\varepsilon$
$T \to FT'$
$T' \to *FT'|\varepsilon$
$F \to (E)|id$

First: $E = \{id, (\}$ ovviamente ha gli stessi first di T per $E \to TE'$
$E' = \{+, \varepsilon\}$
$T = \{id, (\}$ ha gli stessi first di F per $T \to FT'$
$T' = \{*, \varepsilon\}$
$F = \{id, (\}$

Per generare id + id:



Mancano le parentesi fra i terminali nella tabella...

	id	+	*	\$
E	$E \to TE'$			
E		$E' \to TE'$		$E' \to \varepsilon$
T	$T \to FT'$			
T'		$T' \to \varepsilon$	$T' \to *FT'$	$T' \to \varepsilon$
F	$F \to id$			

6.5 Follow

$\forall\ A \in V \backslash T$, follow(A):

```
follow(A) = emptySet per ogni A in (V \ T);
follow(S).push($);

repeat{
    foreach(B -> alpha A beta in P){
        if(beta == epsilon){
            follow(A).push(follow(B));
        } else {
            follow(A).push(first(beta) \ epsilon);
            if(epsilon in first(beta)){
                follow(A).push(follow(B));
            }
        }
```

```
    }
} until (saturazione);
```

6.5.1 Esempio

$S \to aABb$
$A \to Ac|d$
$B \to CD$
$C \to e|\varepsilon$
$D \to f|\varepsilon$

	First	Follow
$S =$	$\{a\}$	$\{\$\}$
$A =$	$\{d\}$	$\{e, f, b$ (da $S \to aABb$), c (da $A \to Ac$)$\}$
$B =$	$\{e, f, \varepsilon\}$	$\{b$ (da $S \to aABb$)$\}$
$C =$	$\{a, \varepsilon\}$	$\{f$ (da $B \to CD$)$\}$
$D =$	$\{f, \varepsilon\}$	$\{\}$

6.5.2 Esempio

$S \to aA|aBc$
$A \to Bd|Cc$
$B \to e|\varepsilon$
$C \to f|\varepsilon$

	First	Follow
$S =$	$\{a, b\}$	$\{\$\}$
$A =$	$\{e, d, f, c\}$	$\{\$, (?)\}$
$B =$	$\{e, \varepsilon\}$	$\{c, d\}$
$C =$	$\{f, \varepsilon\}$	$\{c\}$

6.5.3 Esempio

$E \to TE'$
$E' \to +TE'|\varepsilon$
$T \to FT'$
$T' \to *FT'|\varepsilon$
$F \to (E)|id$

	First	Follow	
$E =$	$\{id, (\}$	$\{\$,)\}$	
$E' =$	$\{+, \varepsilon\}$	$\{\}$	ed eredita i follow di E
$T =$	$\{id, (\}$	$\{+\}$	ed eredita i follow di E,E’
$T' =$	$\{+, \varepsilon\}$	$\{\}$	ed eredita i follow di T
$F =$	$\{id, (\}$	$\{*\}$	ed eredita i follow di T, T’
		Quindi diventa:	
$E =$	$\{id, (\}$	$\{\$,)\}$	
$E' =$	$\{+, \varepsilon\}$	$\{\$,)\}$	
$T =$	$\{id, (\}$	$\{+, \$\}$	
$T' =$	$\{+, \varepsilon\}$	$\{+, \$\}$	
$F =$	$\{id, (\}$	$\{*, +, \$\}$	

Parsing di “$id + id * id\$$” $E \to TE' \to FT'E' \to idT'E' \to ...$

6.6 Costruzione delle tabelle di parsing predittivo top-down

input G=(V,T,S,P)
output Tabella T di parsing predittivo top-down se G é LL(1)

```
foreach((A -> alpha) in P){
    forall b in first(alpha), poniamo A -> alpha in T[A, b];
    if(epsilon in first(alpha)){
        forall x in follow(A) poniamo A -> alpha in T[A, x];
```

```
        }
    }

    poniamo error() in tutte le entry di T che sono rimaste vuote;

    if(la tabella non ha entry multiply-defined)
        G e'' LL(1);
```

6.6.1 Esempio

$E \to E + T|T$
$T \to T * F|T$
$F \to (E)|id$

	First	Follow	
$E =$	$\{(id\}$	$\{\$, +,)\}$	$\{\$, +,)\}$
$T =$	$\{(id\}$	$\{*\}$ ed eredita i follow di E	$\{\$, +, *,)\}$
$F =$	$\{(id\}$	$\{\}$ ed eredita i follow di T	$\{\$, +, *,)\}$

Guardo se é LL(1)

	id
E	$E \to E + T$ $E \to T$

Pur non sviluppando tutta la tabella si vede che ci sono entry multiple quindi non é LL(1).

Una grammatica G esibisce **left recursion** se per qualche $A \in V \backslash T$ e per qualche $\alpha \in V^*$, $A \to^* A\alpha$.

La left recursion é immediata se G ha almeno una produzione del tipo $A \to a\alpha$ (ovvero se succede nel passato). Nell'esempio di prima c'era left recursion immediata nei primi due casi ($E \to E\alpha \wedge T \to T\alpha$).

Proposizione: ogni grammaticha che esibisce left recursion non é LL(1).

$A \to B$
$B \to Aa$
Esempio di left recursion in piú passi.

6.7 Eliminazione Left Recursion immediata

6.7.1 Esempio

$A \to A\alpha|\beta$, con $\beta! = A \wedge \alpha! = \varepsilon$ diventa:
$A \to \beta A'$
$A' \to \alpha A'|\varepsilon$

Piú in generale $A \to A\alpha_1|...|A\alpha_n|\beta_1|...|\beta_n$, $\beta_1, ..., \beta_n! = A_y$, $\alpha_1, ..., \alpha_n! = \varepsilon$ diventa
$A \to \beta_1 A'|...|\beta_n A'$
$A' \to \alpha_1 A'|...|\alpha_m A'|\varepsilon$
Ho introdotto A' nuovo non terminale in G.

6.7.2 Esempio

Eliminare Left Recursion immediata da: $E \to E + T|T$
$T \to T * F|T$
$F \to (E)|id$

Diventa: $E \to TE'$
$E' \to +TE'|\varepsilon$
$T \to FT'$
$T' \to *FT'|\varepsilon$
$F \to (E)|id$

	First	Follow
$E =$	$\{id, (\}$	$\{\$,)\}$
$E' =$	$\{+, \varepsilon\}$	$\{\$,)\}$
$T =$	$\{id, (\}$	$\{+, \$,)\}$
$T' =$	$\{*, \varepsilon\}$	$\{+, \$,)\}$
$F =$	$\{id, (\}$	$\{*, +, \$,)\}$

Tabella di parsing

	id	+	*	(	)	$
E	$E \rightarrow TE'$			$E \rightarrow TE'$		$
E'		$E' \rightarrow +TE'$			$E' \rightarrow \varepsilon$	$E' \rightarrow \varepsilon$
T	$T \rightarrow FT'$			$T \rightarrow FT'$		$
T'		$T' \rightarrow \varepsilon$	$T' \rightarrow *FT'$		$T' \rightarrow \varepsilon$	$T' \rightarrow \varepsilon$
F	$F \rightarrow id$			$F \rightarrow (E)$	)	$

I campi vuoti sono error, non ci sono multiple entries quindi é LL(1).

6.7.3 Esempio

Eliminare Left Recursion immediata da: $E \rightarrow E + E|E * E|(E)|id$

Diventa: $E \rightarrow (E)E'|idE'$
$E' \rightarrow +EE'| * EE'|\varepsilon$

(Parte della tabella tanto non serve tutta)

	First	Follow	
$E =$	$\{id, (\}$	$\{\$,), +, *\}$ ed i follow di E'	$\{\$,), +, *\}$
$E' =$	$\{+, *, E\}$	$\{\}$ ed i follow di E	$\{\$,), +, *\}$

Visto che ho almeno una entry multipla la grammatica non é LL(1).

L'eliminazione della left recursion ci ha dato un grammatica che non é comunque LL(1). Nel nostro caso é anche ambigua.

> Lemma: L'eliminazione della left recursion NON elimina **l'ambiguitá**.

6.8 Left Factoring

$S \rightarrow aSb|ab$

	a	b	$
S	$S \rightarrow aSb$ $S \rightarrow ab$		

La grammatica non é LL(1). Possiamo peró fattorizzare le produzioni considerando una parte che é a sinistra ed é comune a piú produzioni, per ottenere una grammatica LL(1) che genera lo stesso linguaggio.

$S \rightarrow aA'$
$A' \rightarrow Sb|b$

DEF: Una grammatica G puó essere fattorizzata a sinistra quando esistono almeno due produzioni $A \rightarrow \alpha\beta_1$ e $A \rightarrow \alpha\beta_2$ per qualche $A \in V\backslash T$, $\alpha, \beta_1, \beta_2 \in V^* \wedge \alpha$ non comincia per A.

DEF: G puó essere fattorizzata a sinistra se: $A \rightarrow \alpha\beta_1$, $A \rightarrow \alpha\beta_2 \in P$ con $\alpha, \beta_1, \beta_2 \in V^*, \alpha$ non ha A come primo simbolo, $A \in V\backslash T$

Lemma: se G puó essere fattorizzata a sinistra allora G non é LL(1).

6.9 Algoritmo di fatturazione a sinistra

```
foreach(A in V\T){
    trovare il prefisso piu lungo comune a due o piu produzioni per A, chiamato alpha
    if(alpha != epsilon){
        sostituire A -> alpha beta_1|...|alpha beta_n|Y_1|...|Y_k
        con A -> alpha A' | Y_1 | Y_k
        con A' -> beta_1 | ... | beta_n e A' nuovo simbolo
    }
}
```

manca roba

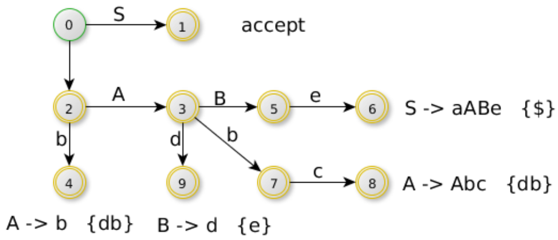
6.10 Bottom Up

Ricostruire, se $w \in L(G)$, una rightmost derivation al contrario

6.10.1 Esempio

$S \rightarrow aABe$
$A \rightarrow Abc|b$
$B \rightarrow d$

w = abbcde visto che é rightmost devo espandere B dato che é il non terminale piú a destra. $S \rightarrow aABe \rightarrow aAde \rightarrow aAbcde \rightarrow abbcde$



La sottolineatura significa che se arrivo in questo stato e sto leggendo come prossimo input una d o una b posso fare la riduzione della b usando A. Lo stesso vale per le altre, ovviamente con i loro simboli. La roba fra parentesi graffe si chiama look-ahead set.

Nel grafo faccio quindi i seguenti passi (i numeri sono i nodi):

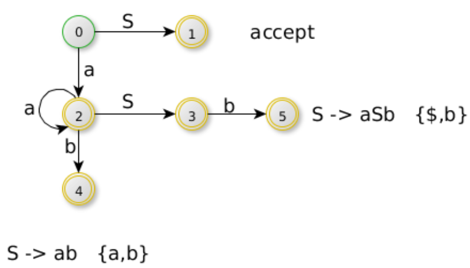
0	$abbcde\$$
$0 \rightarrow 2$ consumando 'a'	$a\|\|bbcde\$$
$2 \rightarrow 4$ consumando 'b'	$ab\|\|bcde\$$
4 riduco $A \rightarrow b$	$aA\|\|bcde$

A questo punto torno al nodo 2 ovvero il precedente. Vado quindi in 3, perché ho la A al posto della b che avevo prima.

torno a 2, vado in 3

$3 \rightarrow 7$ consumando 'b'	$aAb\|\|cde\$$
$7 \rightarrow 8$ consumando 'c'	$aAbc\|\|de\$$
8 riduco $A \rightarrow Abc$	$aA\|\|de$
torno a 7, torno in 3, vado in 9	
$3 \rightarrow 9$ consumando 'd'	$aAd\|\|e\$$
riduco $B \rightarrow d$	$aAB\|\|e\$$
torno a 3, vado in 5, vado in 6	
$5 \rightarrow 6$ consumando 'e'	$aABe\|\|\$$
6 riduco $S \rightarrow aABe$	$S\|\|\$$
torno a 0, vado in 1, ho finito	

Noi vogliamo avere grammatiche di tipo LALR(1). Grammatiche: $SLR(1) \subset LALR(1) \subset LR(1)$



$S \rightarrow aSb|ab$
$w = aaabbb\$$

0	$aaabbb\$$
$0 \to 2$	$a\|\|aabbb\$$
$2 \to 2$	$aa\|\|abbb\$$
$2 \to 2$	$aaa\|\|bbb\$$
$2 \to 4$	$aaab\|\|bb\$$
4 riduco $S \to ab$	$aaS\|\|bb\$$
torno a 2, vado in 3, vado in 5	
$3 \to 5$	$aaSb\|\|b\$$
5 riduco $S \to aSb$	$aS\|\|b\$$
torno a 3, vado in 5	
$3 \to 5$	$aSb\|\|\$$
5 riduco $S \to aSb$	$S\|\|\$$
torno a 0, vado in 1, ho finito	

Questa é una tabella:

	terminali ∪\$	$V \backslash T$
stati	shift-k: leggi un simbolo di input e vai allo stato reduce $A \to b$	goto-k: descrive le funzioni di tr identificate dai non terminali qu

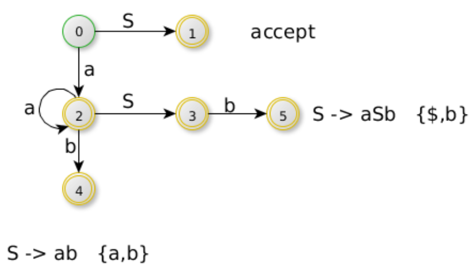
6.11 Algoritmo di shift/reduce

(comune a SLR(1), LR(1), LALR(1))

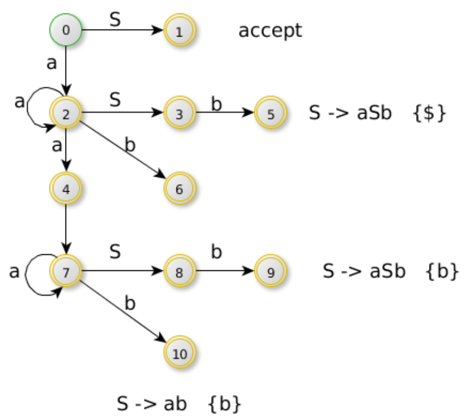
input w, tabella di parsing bottom-up di tipo $\diamond$, con $\diamond$ scelto fra $\{SLR(1), LALR(1), LR(1)\}$ G.
output derivazione rightmost di w se $w \in L(G)$, altrimenti error()

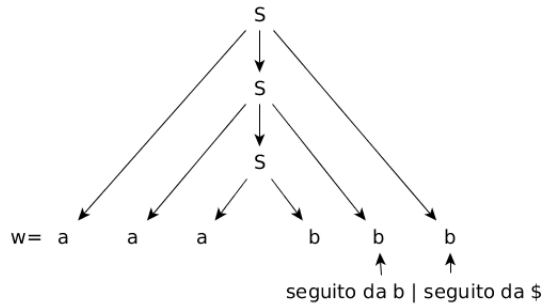
```
stack.push(s_0);
buffer = w$;
while(true){
    let s = stack.top();
    if(M[s,b] == shift-k){
        stack.push(b);
        stack.push(k);
        let b = buffer.readNext();
    } else if(M[s,b] == "reduce A -> beta"){
        stack.pop() 2|beta| simboli;
        let j tale che M[m, A] = gj;
        push(A);
        push(j);
        output "A -> beta";
    } else if(M[s,b] = accetta){
        break;
    } else {
        error();
    }
}
```

sketo $S \to aSb|ab$



Questo é uguale ma scritto diversamente per separare $\{\$, b\}$ in $\{4\}$ e $\{b\}$. Nel caso di $w = aaabbb\$$. Una caso rappresenta il rao piú in alto, mentre l'altro il secondo ramo (piú interno).





- Automa caratteristico
- Lookahead Function

Coppie diverse di questi due insiemi ci danno tipi di grammatiche diverse.

Gli automi che stiamo utilizzando devono essere in grado di ricordare abbastanza da essere in grado di tornare indietro fino al punto in cui abbiamo sostituito una certa sequenza di terminali/non terminali con un'altra.

$G = (V, T, S, P)$, aggiungo una produzione $S' \to S$, $G' = (V \cup \{S'\}, T, S', P \cup \{S' \to S\})$ $A \to \alpha.\beta$

All'inizio ho .S, ovvero non ho ancora letto nulla e devo leggere S.

All'inizio (il nodo iniziale), non ho ancora visto nulla. Visto che S puó iniziare con aSb o ab non sappiamo davanti a quale sviluppo ci troviamo. Il primo stato é quindi

$S' \to .S$
$S \to .aSb$
$S \to .ab$

Questo puó essere visto come un nodo. Da questo stato mi muovo verso un altro stato (con una a-transizione, perché vedo che iniziano quasi tutte con a). In questo stato avró: $S \to a.Sb$
$S \to a.b$

Adesso mi aspetto di vedere l'espansione di una S. Devo quindi aggiungere a questo nodo anche quelle produzioni, e diventa quindi: $S \to a.Sb$
$S \to a.b$
$S \to .aSb$
$S \to .ab$

Notare che le ultime due sono le stesse delle ultime due del nodo prima. Quella é la chiusura, mentre le due prima sono i generatori dello stato (kernel dello stato, **kernel items**).

Gli stati che sono terminali (ovvero che nol disegno prima avevano le transizioni scritte vicino), sono del tipo $S \to ab.$, ovvero che hanno incontrato di tutto e di cui si puó eseguire la riduzione. Questi si chiamano **reducing items**.

Dallo stato con 4 items che avevo prima, si puó fare una b-transizione che va in uno di quelli stati terminali, ovvero: $S \to ab$.
Questo perché la seconda produzione si aspetta b, che poi completa quello che viene generato da quella produzione. Sempre da quello stato con 4 produzioni partirá anche una a-transizione ed una S-transizione. Per vedere che transizioni devo avere, devo vedere la prima lettera dopo il punto per ogni item di quel nodo.

6.12 Items

$G = (V, T, S, P)$
$G' = (V \cup \{S'\}, T, S', P \cup \{S' \to S\})$, con $S' \notin V$.

Un LR(0)-item di G' é una produzione di G con un punto in qualche posizione del body, ovvero $A \to \alpha.\beta$. Alla produzione della forma $A \to \varepsilon$ corrisponde un solo LR(0)-item, ovvero $A \to .$

	iniziale	se $A = S' \wedge \alpha = \varepsilon \wedge \beta = S$, cioé se l'item é $S' \to .S$
	accepting	se $A = S' \wedge \alpha = S \wedge \beta = \varepsilon$, cioé se l'item é $S' \to S.$
L'item $A \to \alpha.\beta$ é detto:	kernel	se é un iniziale o tale che $\alpha! = \varepsilon$
	closure	se $\alpha = \varepsilon$ e non é iniziale
	reducing	se non é accepting e $\beta = \varepsilon$, cioé se il punto é in fondo $\wedge ! accepting$