



SAPIENZA
UNIVERSITÀ DI ROMA

“SAPIENZA” UNIVERSITÀ DI ROMA
INGEGNERIA DELL'INFORMAZIONE,
INFORMATICA E STATISTICA
DIPARTIMENTO DI INFORMATICA

Automi: Calcolabilità e Complessità

Appunti integrati con il libro "Introduzione alla teoria della computazione",
Michael Sipser

Author
Alessio Bandiera

18 novembre 2023

Indice

Informazioni e Contatti	1
1 Linguaggi ed espressioni regolari	2
1.1 Stringhe e linguaggi	2
1.1.1 Stringhe	2
1.1.2 Linguaggi	3
1.1.3 Funzioni di Hamming	3
1.2 Determinismo	4
1.2.1 Definizioni	4
1.2.2 Linguaggi regolari	6
1.3 Non determinismo	7
1.3.1 Definizioni	7
1.3.2 Equivalenze	9
1.4 Operazioni regolari	10
1.4.1 Unione	10
1.4.2 Intersezione	12
1.4.3 Concatenazione	13
1.4.4 Elevamento a potenza	14
1.4.5 Star	15
1.4.6 Complemento	17
1.5 Espressioni regolari	18
1.5.1 Definizioni	18
1.6 Configurazioni	20
1.6.1 Configurazioni di DFA	20
1.6.2 Configurazioni di NFA	21
1.7 Non determinismo generalizzato	23
1.7.1 Definizioni	23
1.7.2 Equivalenze	24
1.8 Linguaggi non regolari	30
1.8.1 Pumping lemma	30
2 Linguaggi e grammatiche context-free	33
2.1 Grammatiche context-free	33
2.1.1 Definizioni	33
2.1.2 Ambiguità	35

2.1.3	Forma normale di Chomsky	37
2.2	Operazioni context-free	39
2.2.1	Unione	39
2.2.2	Concatenazione	39
2.3	Automi a pila	40
2.3.1	Definizioni	40
2.3.2	Equivalenze	44
2.4	Linguaggi non context-free	49
2.4.1	Pumping lemma	49

Informazioni e Contatti

Prerequisiti consigliati:

- TODO: DA DECIDERE

Segnalazione errori ed eventuali migliorie:

Per segnalare eventuali errori e/o migliorie possibili, si prega di utilizzare il **sistema di Issues fornito da GitHub** all'interno della pagina della repository stessa contenente questi ed altri appunti (link fornito al di sotto), utilizzando uno dei template già forniti compilando direttamente i campi richiesti.

Gli appunti sono in continuo aggiornamento, pertanto, previa segnalazione, si prega di controllare se l'errore sia ancora presente nella versione più recente.

Licenza di distribuzione:

These documents are distributed under the [GNU Free Documentation License](#), a form of copyleft intended to be used on manuals, textbooks or other types of document in order to assure everyone the effective freedom to copy and redistribute it, with or without modifications, either commercially or non-commercially.

Contatti dell'autore e ulteriori link:

- Github: <https://github.com/ph04>
- Email: alessio.bandiera02@gmail.com
- LinkedIn: [Alessio Bandiera](#)

1

Linguaggi ed espressioni regolari

1.1 Stringhe e linguaggi

1.1.1 Stringhe

Definizione 1.1.1.1: Alfabeto

Si definisce **alfabeto** un qualsiasi insieme finito, non vuoto; i suoi elementi sono detti **simboli** o **caratteri**.

Esempio 1.1.1.1 (Alfabeto). $\Sigma = \{0, 1, x, y, z\}$ è un alfabeto, composto da 5 simboli.

Definizione 1.1.1.2: Stringa

Sia Σ un alfabeto; una **stringa su Σ** è una sequenza finita di simboli di Σ ; la **stringa vuota** è denotata con ε .

- Data una stringa w di Σ , allora $|w|$ è la lunghezza di w .
- Se w ha lunghezza $n \in \mathbb{N}$, allora è possibile scrivere che $w = w_1 w_2 \cdots w_n$ con $w_i \in \Sigma$ e $i \in [1, n]$.

Esempio 1.1.1.2 (Stringa). Sia $\Sigma = \{0, 1, x, y, z\}$ un alfabeto; allora una sua possibile stringa è $w = x1y0z$.

Definizione 1.1.1.3: Stringa inversa

Sia Σ un alfabeto, e $w = w_1 w_2 \cdots w_n$ una sua stringa; allora si definisce l'**inversa** di w come segue:

$$w^{\mathcal{R}} := w_n w_{n-1} \cdots w_1$$

Definizione 1.1.1.4: Concatenazione

Sia Σ un alfabeto, e $x = x_1x_2 \cdots x_n, y = y_1y_2 \cdots y_n$ due sue stringhe; allora xy è la stringa ottenuta attraverso la **concatenazione** di x ed y .

Per indicare una stringa concatenata con se stessa k volte, si utilizza la notazione

$$x^k = \underbrace{xx \cdots x}_{k \text{ volte}}$$

Si noti che per ogni stringa x su Σ , si ha che $x^0 = \varepsilon$.

Definizione 1.1.1.5: Prefisso

Sia Σ un alfabeto, ed x, y due sue stringhe; allora x è detto essere un **prefisso** di y , se $\exists z \mid xz = y$, con z stringa in Σ .

Esempio 1.1.1.3 (Prefisso). Sia $\Sigma = \{a, b, c\}$ un alfabeto; allora la stringa $x = ab$ è prefisso della stringa $y = abc$, poiché esiste una stringa $z = c$ tale per cui $xz = y$.

1.1.2 Linguaggi**Definizione 1.1.2.1: Linguaggio**

Sia Σ un alfabeto; si definisce **linguaggio** un insieme di stringhe di Σ . Un linguaggio è detto **prefisso**, se nessun suo elemento è prefisso di un altro. Il linguaggio vuoto si indica con \emptyset .

Esempio 1.1.2.1 (Linguaggio binario). Il linguaggio binario, che verrà utilizzato estensivamente, è il seguente:

$$\Sigma = \{0, 1\}$$

1.1.3 Funzioni di Hamming**Definizione 1.1.3.1: Distanza di Hamming**

Sia Σ un alfabeto, e siano x, y due sue stringhe tali che $|x| = |y|$; si definisce **distanza di Hamming** tra x ed y il numero di caratteri per cui x ed y differiscono. In simboli, date due stringhe $x = x_1 \cdots x_n, y = y_1 \cdots y_n$ con $n \in \mathbb{N}$, si ha che

$$d_H(x, y) := |\{i \in [1, n] \mid x_i \neq y_i\}|$$

Esempio 1.1.3.1 (Distanza di Hamming). Siano $x = 1011101$ ed $y = 1001001$ due stringhe sull'alfabeto $\Sigma = \{0, 1\}$; poiché differiscono per 2 caratteri, si ha che $d_H(x, y) = 2$.

Definizione 1.1.3.2: Peso di Hamming

Sia $\Sigma = \{0, \dots, 9\}$ l'alfabeto composto dalle 10 cifre decimali, e sia x una sua stringa; si definisce **peso di Hamming** di x il numero di elementi di x diversi da 0. In simboli, data una stringa $x = x_1 \cdots x_n$, con $n \in \mathbb{N}$, si ha che

$$w_H(x) := |\{i \in [1, n] \mid x_i \neq 0\}|$$

Osservazione 1.1.3.1: Peso di Hamming di stringhe binarie

Sia $\Sigma = \{0, 1\}$ l'alfabeto binario; allora, il peso di Hamming di una sua stringa è il numero di 1 che la compongono.

1.2 Determinismo

1.2.1 Definizioni

Definizione 1.2.1.1: DFA

Un **DFA** (*Deterministic Finite Automaton*) è una quintupla $(Q, \Sigma, \delta, q_0, F)$, dove

- Q è l'**insieme degli stati** dell'automa, un insieme *finito*
- Σ è l'**alfabeto dell'automa**, un insieme *finito*
- $\delta : Q \times \Sigma \rightarrow Q$ è la **funzione di transizione**, che definisce la relazione tra gli stati
- $q_0 \in Q$ è lo **stato iniziale**
- $F \subseteq Q$ è l'**insieme degli stati accettanti**, sui quali le stringhe possono terminare

Esempio 1.2.1.1 (DFA). Un esempio di DFA è il seguente:

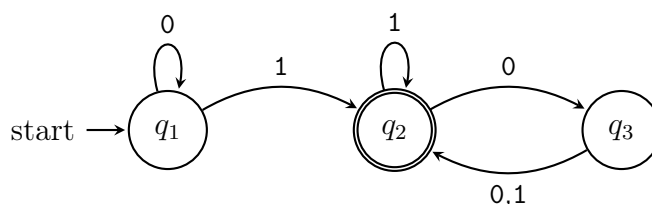


Figura 1.1: Un DFA.

esso può essere descritto secondo la quintupla $(Q, \Sigma, \delta, q_0, F)$ come segue:

- $Q = \{q_1, q_2, q_3\}$
- $\Sigma = \{0, 1\}$

- δ è la seguente:

	0	1
q_1	q_1	q_2
q_2	q_3	q_2
q_3	q_2	q_2

- q_1 è lo stato iniziale
- $F = \{q_2\} \subseteq Q$

Definizione 1.2.1.2: Stringhe accettate (DFA)

Sia $M = (Q, \Sigma, \delta, q_0, F)$ un DFA, e sia $w = w_1 \cdots w_n$ una stringa tale per cui $\forall i \in [1, n] \quad w_i \in \Sigma$; allora, M **accetta** w se esiste una sequenza di stati $r_0, \dots, r_n \in Q$ tali per cui

- $r_0 = q_0$
- $\forall i \in [0, n-1] \quad \delta(r_i, w_{i+1}) = r_{i+1}$
- $r_n \in F$

Definizione 1.2.1.3: Linguaggio di un automa

Sia M un automa; allora il **linguaggio di** M è un insieme $L(M)$ contenente tutte le stringhe accettate da M ; simmetricamente, si dice che M **riconosce** $L(M)$.

Esempio 1.2.1.2 (Linguaggio di un automa). Si consideri il seguente automa M_1 :



Figura 1.2: Un automa M_1 .

sapendo che $\Sigma = \{0, 1\}$, che q_1 è lo stato iniziale, e che $F = \{q_2\}$, è facilmente verificabile che

$$L(M_1) = \{w \mid w = w_1 \cdots w_{n-1}1, n \in \mathbb{N}\}$$

ovvero, M_1 accetta tutte e sole le stringhe che terminano per 1.

Definizione 1.2.1.4: Linguaggi riconosciuti da automi

Dato un linguaggio A , ed un automa M , si dice che M **riconosce** A se e solo se

$$A = \{w \mid M \text{ accetta } w\}$$

Definizione 1.2.1.5: Linguaggi di una classe di automi

Sia \mathcal{C} una classe di automi; allora, l'**insieme dei linguaggi** riconosciuti dagli automi della classe \mathcal{C} è denotato col seguente simbolismo:

$$L(\mathcal{C}) := \{L \mid \exists M \in \mathcal{C} : L(M) = L\}$$

dove L è un linguaggio, ed M è un automa della classe \mathcal{C} .

1.2.2 Linguaggi regolari**Definizione 1.2.2.1: Linguaggio regolare**

Un linguaggio è detto **regolare** se e solo se esiste un DFA che lo riconosce. La classe dei linguaggi regolari è denotata con REG. Allora, in simboli, si ha che

$$\text{REG} := \mathcal{L}(\text{DFA})$$

Esempio 1.2.2.1 (Linguaggi regolari). Sia $\Sigma = \{0, 1\}$ l'alfabeto binario, ed L il seguente linguaggio:

$$L := \{w \mid w = 0^n 1, n \in \mathbb{N} - \{0\}\}$$

Tale linguaggio è regolare, poiché esiste il seguente DFA che lo riconosce:



Figura 1.3: Un DFA che riconosce L .

1.3 Non determinismo

1.3.1 Definizioni

Definizione 1.3.1.1: NFA

Un **NFA** (*Nondeterministic Finite Automaton*) è un automa in cui possono esistere varie scelte per lo stato successivo in ogni punto. Durante la computazione, ogni volta che viene incontrata una scelta, la macchina si *divide*, e ognuno dei vari automi risultanti computa le varie scelte indipendentemente.

Formalmente, un NFA è una quintupla $(Q, \Sigma, \delta, q_0, F)$, dove

- Q è l'**insieme degli stati**, un insieme *finito*
- Σ è l'**alfabeto dell'automata**, un insieme *finito*
- $\delta : Q \times \Sigma_\epsilon \rightarrow \mathcal{P}(Q)$ è la **funzione di transizione**, che definisce la relazione tra gli stati
- $q_0 \in Q$ è lo **stato iniziale**
- $F \subseteq Q$ è l'**insieme degli stati accettanti**

dove $\Sigma_\epsilon := \Sigma \cup \{\epsilon\}$.

Se il simbolo di input successivo non compare su alcuno degli archi uscenti dallo stato occupato da una copia della macchina, quella copia cessa di proseguire; inoltre, se *una qualunque copia* della macchina è in uno stato accettante, l'NFA accetta la stringa di input. Si noti che questa divisione è descritta dall'insieme potenza $\mathcal{P}(Q)$, poiché da ogni stato si può arrivare ad un *insieme* di stati.

Si noti che il determinismo è un caso particolare di non determinismo, dunque un DFA è sempre anche un NFA; in simboli $\text{DFA} \subseteq \text{NFA}$.

Esempio 1.3.1.1 (NFA). Un esempio di NFA è il seguente:

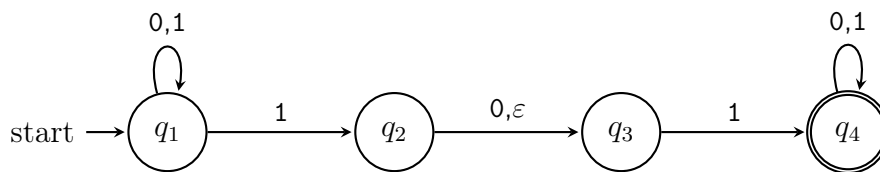


Figura 1.4: L'NFA N .

esso può essere descritto secondo la quintupla $N = (Q, \Sigma, \delta, q_0, F)$ come segue:

- $Q = \{q_1, q_2, q_3, q_4\}$
- $\Sigma = \{0, 1\}$
- δ è la seguente:

	0	1	ε
q_1	$\{q_1\}$	$\{q_1, q_2\}$	\emptyset
q_2	$\{q_3\}$	\emptyset	$\{q_3\}$
q_3	\emptyset	$\{q_4\}$	\emptyset
q_4	$\{q_4\}$	$\{q_4\}$	\emptyset

- q_1 è lo stato iniziale
- $F = \{q_4\} \subseteq Q$

Ad esempio, se N legge l'input 010110, la sua computazione è la seguente:



Nota: nel momento in cui vengono incontrati ε -archi, si giunge allo stato successivo della computazione nello stesso step dell'input appena elaborato, senza produrre un passo ulteriore, producendo inoltre un nuovo ramo di computazione.

Definizione 1.3.1.2: Stringhe accettate (NFA)

Sia $N = (Q, \Sigma, \delta, q_0, F)$ un NFA, e sia $w = w_1 \cdots w_n$ una stringa tale per cui $\forall i \in [1, n] \quad w_i \in \Sigma$; allora, M **accetta** w se esiste una sequenza di stati $r_0, \dots, r_n \in Q$ tali per cui

- $r_0 = q_0$
- $\forall i \in [0, n-1] \quad r_{i+1} \in \delta(r_i, w_{i+1})$
- $r_n \in F$

1.3.2 Equivalenze

Definizione 1.3.2.1: Equivalenza tra automi

Due automi si dicono **equivalenti** se e solo se riconoscono lo stesso linguaggio.

Teorema 1.3.2.1: DFA ed NFA

Le classi dei DFA e degli NFA sono equivalenti; in simboli

$$\mathcal{L}(\text{NFA}) = \mathcal{L}(\text{DFA})$$

Dimostrazione.

Prima implicazione. Si noti che

$$\mathcal{L}(\text{NFA}) \subseteq \mathcal{L}(\text{DFA}) \iff \forall A \in \mathcal{L}(\text{NFA}) \quad \exists M \in \text{DFA} \mid A = L(M)$$

Allora, sia $A \in \mathcal{L}(\text{NFA})$, e dunque esiste un NFA $N = (Q, \Sigma, \delta, q_0, F)$ in grado di riconoscerlo. Inoltre, si definisca

$$\forall k \geq 0 \quad E(R) := \bigcup_{r \in R} \delta^k(r, \varepsilon)$$

l'insieme degli stati raggiungibili da stati in R , applicando (anche ripetutamente) un numero arbitrario di ε -archi (si noti che per $k = 0$ si ha che $R \subseteq E(R)$).

Allora, sia $M = (Q', \Sigma, \delta', q'_0, F)$ il DFA definito come segue:

- $Q' := \mathcal{P}(Q)$, scelto tale da rappresentare ogni possibile stato di N ;
- $\forall R \in Q', a \in \Sigma \quad \delta'(R, a) := \bigcup_{r \in R} E(\delta(r, a))$, scelta tale in quanto, per un certo insieme di stati $R \in Q'$ di N , a $\delta'(R, a)$ viene assegnata l'unione degli stati che sarebbero stati raggiunti in N dagli $r \in R$ con a , calcolati dunque attraverso $\delta(r, a)$, aggiungendo infine i possibili ε -archi;
- $q'_0 := E(\{q_0\})$, scelto tale da far iniziare M esattamente dove aveva inizio N , comprendendo anche i possibili ε -archi iniziali;
- $F' := \{R \in Q' \mid \exists r \in R : r \in F\}$, che corrisponde all'insieme degli insiemi di stati di N contenenti almeno uno stato accettante in N .

Allora M è in grado di riconoscere A per costruzione, poiché il DFA costruito emula l'NFA di partenza, tenendo anche in considerazione gli ε -archi. Dunque, sia N che M riconoscono A , e per definizione sono di conseguenza equivalenti.

Seconda implicazione. Poiché il determinismo è un caso particolare del non determinismo, si ha che $\text{DFA} \subseteq \text{NFA} \implies \mathcal{L}(\text{DFA}) \subseteq \mathcal{L}(\text{NFA})$.

□

Corollario 1.3.2.1: Linguaggi regolari ed NFA

Un linguaggio è regolare se e solo se esiste un NFA che lo riconosce; in simboli

$$\text{REG} = \mathcal{L}(\text{NFA})$$

Dimostrazione. Per il [Teorema 1.3.2.1](#), si ha che

$$\text{REG} := \mathcal{L}(\text{DFA}) = \mathcal{L}(\text{NFA})$$

e dunque segue la tesi. □

1.4 Operazioni regolari

1.4.1 Unione

Definizione 1.4.1.1: Unione

Siano A e B due linguaggi su un alfabeto Σ ; allora, si definisce l'**unione** di A e B il seguente linguaggio:

$$A \cup B := \{x \mid x \in A \vee x \in B\}$$

Si noti che, per ogni linguaggio L , è vero che $\emptyset \cup L = L \cup \emptyset = L$.

Esempio 1.4.1.1 (Unione). Sia $\Sigma = \{\mathbf{a}, \dots, \mathbf{z}\}$ l'alfabeto composto da 26 lettere, e siano $A = \{\mathbf{uno}, \mathbf{due}\}$ e $B = \{\mathbf{tre}, \mathbf{quattro}\}$ due linguaggi su Σ . Allora, si ha che

$$A \cup B = \{\mathbf{uno}, \mathbf{due}, \mathbf{tre}, \mathbf{quattro}\}$$

Proposizione 1.4.1.1: Chiusura sull'unione

Siano A e B due linguaggi regolari su un alfabeto Σ ; allora $A \cup B$ è regolare.

Dimostrazione I. Per definizione, A e B sono linguaggi regolari, dunque esistono due DFA

$$M_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$$

$$M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$$

tali da riconoscere rispettivamente A e B . Allora, sia $M = (Q, \Sigma, \delta, q_0, F)$ il DFA definito come segue:

- $Q := Q_1 \times Q_2 = \{(r_1, r_2) \mid r_1 \in Q_1 \wedge r_2 \in Q_2\}$, scelto tale in quanto permette di avere tutte le possibili combinazioni di stati dei due automi di partenza;
- $\forall (r_1, r_2) \in Q, a \in \Sigma \quad \delta((r_1, r_2), a) := (\delta_1(r_1, a), \delta_2(r_2, a))$, scelta tale in quanto permette di simulare entrambi gli automi di partenza contemporaneamente, mandando ogni stato di M_1 ed M_2 dove sarebbe andato nei rispettivi automi di appartenenza;

- $q_0 := (q_1, q_2)$, scelto tale in quanto deve essere lo stato in cui entrambe gli automi in ipotesi iniziavano;
- $F := (F_1 \times Q_2) \cup (Q_1 \times F_2) = \{(r_1, r_2) \mid r_1 \in F_1 \vee r_2 \in F_2\}$, scelto tale in quanto permette di simulare gli stati accettanti di entrambi gli automi, e vanno prese tutte le coppie che vedono almeno uno dei due stati come accettanti poiché altrimenti non si accetterebbero delle stringhe accettate da M_1 ed M_2 in partenza.

Allora, poiché M è in grado di simulare M_1 ed M_2 contemporaneamente, per costruzione accetterà ogni stringa di A e di B , dunque riconoscendo $A \cup B$, e di conseguenza $A \cup B$ è regolare per definizione. \square

Dimostrazione II. Per definizione, A e B sono linguaggi regolari, dunque per il [Teorema 1.3.2.1](#) esistono due NFA

$$\begin{aligned} N_1 &= (Q_1, \Sigma, \delta_1, q_1, F_1) \\ N_2 &= (Q_2, \Sigma, \delta_2, q_2, F_2) \end{aligned}$$

tali da riconoscere rispettivamente A e B . Allora, sia $N = (Q, \Sigma, \delta, q_0, F)$ l'NFA costruito come segue:

- $Q := \{q_0\} \cup Q_1 \cup Q_2$, dove q_0 è un nuovo stato, e Q è scelto tale da includere sia N_1 che N_2 ;

$$\bullet \forall q \in Q, a \in \Sigma_\varepsilon \quad \delta(q, a) := \begin{cases} \delta_1(q, a) & q \in Q_1 \\ \delta_2(q, a) & q \in Q_2 \\ \{q_1, q_2\} & q = q_0 \wedge a = \varepsilon \\ \emptyset & q = 0 \wedge a \neq \varepsilon \end{cases}, \text{ scelta tale da poter eseguire}$$

contemporaneamente gli NFA N_1 ed N_2 , definendo per casi la funzione di transizione; infatti, si noti che si è posta $\delta(q_0, \varepsilon) := \{q_1, q_2\}$ in modo da collegare il nuovo stato q_0 a q_1 e q_2 , gli stati iniziali di N_1 ed N_2 rispettivamente;

- q_0 è il nuovo stato, che rappresenta lo stato iniziale di N ;
- $F := F_1 \cup F_2$, scelto tale da costruire N in modo che accetti una stringa se e solo se la accetterebbero N_1 o N_2 .

Figura 1.5: Rappresentazione dell'NFA N descritto.

Allora, l'NFA risultante N è in grado di computare contemporaneamente N_1 ed N_2 , ed è dunque in grado di riconoscere A e B contemporaneamente; di conseguenza, N riconosce $A \cup B$, che risulta dunque essere regolare per definizione. \square

1.4.2 Intersezione

Definizione 1.4.2.1: Intersezione

Siano A e B due linguaggi su un alfabeto Σ ; allora, si definisce l'**intersezione** di A e B il seguente linguaggio:

$$A \cap B := \{x \mid x \in A \wedge x \in B\}$$

Esempio 1.4.2.1 (Intersezione). Sia $\Sigma = \{a, \dots, z\}$ l'alfabeto composto da 26 lettere, e siano $A = \{\text{uno}, \text{due}\}$ e $B = \{\text{uno}, \text{tre}\}$ due linguaggi su Σ . Allora, si ha che

$$A \cap B = \{\text{uno}\}$$

Proposizione 1.4.2.1: Chiusura sull'intersezione

Siano A e B due linguaggi regolari su un alfabeto Σ ; allora $A \cap B$ è regolare.

Dimostrazione. Per definizione, A e B sono linguaggi regolari, dunque esistono due DFA

$$\begin{aligned} M_1 &= (Q_1, \Sigma, \delta_1, q_1, F_1) \\ M_2 &= (Q_2, \Sigma, \delta_2, q_2, F_2) \end{aligned}$$

tali da riconoscere rispettivamente A e B . Allora, sia $M = (Q, \Sigma, \delta, q_0, F)$ il DFA definito come segue:

- $Q := Q_1 \times Q_2 = \{(r_1, r_2) \mid r_1 \in Q_1 \wedge r_2 \in Q_2\}$, scelto tale in quanto permette di avere tutte le possibili combinazioni di stati dei due automi di partenza;
- $\forall (r_1, r_2) \in Q, a \in \Sigma \quad \delta((r_1, r_2), a) := (\delta_1(r_1, a), \delta_2(r_2, a))$, scelta tale in quanto permette di simulare entrambi gli automi di partenza contemporaneamente, mandando ogni stato di M_1 ed M_2 dove sarebbe andato nei rispettivi automi di appartenenza;
- $q_0 := (q_1, q_2)$, scelto tale in quanto deve essere lo stato in cui entrambe gli automi in ipotesi iniziavano;
- $F := F_1 \times F_2 = \{(r_1, r_2) \mid r_1 \in F_1 \wedge r_2 \in F_2\}$, scelto tale in quanto vanno prese tutte e sole le coppie composte da 2 stati entrambe accettanti negli automi di partenza.

Allora, poiché M è in grado di simulare M_1 ed M_2 contemporaneamente, ma accetta solo quando accettavano entrambe gli automi di partenza, per costruzione accetterà ogni stringa di $A \cap B$, che di conseguenza risulta essere regolare per definizione. \square

1.4.3 Concatenazione

Definizione 1.4.3.1: Concatenazione

Siano A e B due linguaggi su un alfabeto Σ ; allora, si definisce la **concatenazione** di A e B il seguente linguaggio:

$$A \circ B = \{xy \mid x \in A \wedge y \in B\}$$

Si noti che, per ogni linguaggio L , è vero che

- $\emptyset \circ L = L \circ \emptyset = \emptyset$
- $\{\varepsilon\} \circ L = L \circ \{\varepsilon\} = L$.

Inoltre, il simbolo \circ può essere talvolta omissso.

Esempio 1.4.3.1 (Concatenazione). Sia $\Sigma = \{a, \dots, z\}$ l'alfabeto composto da 26 lettere, e siano $A = \{\text{uno}, \text{due}\}$ e $B = \{\text{tre}, \text{quattro}\}$ due linguaggi su Σ . Allora, si ha che

$$A \circ B := \{\text{unotre}, \text{unoquattro}, \text{duetre}, \text{duequattro}\}$$

Proposizione 1.4.3.1: Chiusura sulla concatenazione

Siano A e B due linguaggi regolari su un alfabeto Σ ; allora $A \circ B$ è regolare.

Dimostrazione. Per definizione, A e B sono linguaggi regolari, dunque per il [Teorema 1.3.2.1](#) esistono due NFA

$$\begin{aligned} N_1 &= (Q_1, \Sigma, \delta_1, q_1, F_1) \\ N_2 &= (Q_2, \Sigma, \delta_2, q_2, F_2) \end{aligned}$$

tali da riconoscere rispettivamente A e B . Allora, sia $N = (Q, \Sigma, \delta, q_0, F)$ l'NFA costruito come segue:

- $Q := Q_1 \cup Q_2$, scelto tale da includere entrambe gli automi N_1 ed N_2 di partenza;
- $q_0 := q_1$, scelto tale da far iniziare l'esecuzione dell'automa su N_1 ;
- $F := F_2$, scelto tale da far terminare l'esecuzione dell'automa su N_2 ;
- $\forall q \in Q, a \in \Sigma_\varepsilon \quad \delta(q, a) := \begin{cases} \delta_1(q, a) & q \in Q_1 - F_1 \vee (q \in F_1 \wedge a \neq \varepsilon) \\ \delta_1(q, a) \cup \{q_2\} & q \in F_1 \wedge a = \varepsilon \\ \delta_2(q, a) & q \in Q_2 \end{cases},$

scelta tale da anteporre l'esecuzione di N_1 a quella di N_2 ; infatti, se $q \in Q_1 - F_1$ (è uno stato non accettante di N_1), oppure $q \in F_1$ ma $a \neq \varepsilon$, l'esecuzione di N_1 non viene alterata; diversamente, se invece $q \in F_1$ e $a = \varepsilon$, all'insieme di stati $\delta_1(q, \varepsilon)$ viene aggiunto q_2 , ovvero lo stato iniziale di N_2 , in modo da effettuare la concatenazione tra i due NFA non deterministicamente.



Figura 1.6: Rappresentazione dell'NFA N descritto.

Allora, l'NFA N costruito computa inizialmente N_1 , e se vengono raggiunti suoi stati accettanti, l'esecuzione prosegue attraverso N_2 , al fine di realizzare la concatenazione tra le stringhe. Di conseguenza, l'automa è in grado di riconoscere $A \circ B$ per costruzione, e dunque $A \circ B$ è regolare per definizione. \square

1.4.4 Elevamento a potenza

Definizione 1.4.4.1: Elevamento a potenza

Sia A un linguaggio; allora, si definisce **elevamento a potenza** di A il seguente linguaggio:

$$A^n := \underbrace{A \circ \dots \circ A}_{n \text{ volte}} = \begin{cases} A^0 := \{\varepsilon\} \\ A^n = A^{n-1} \circ A & n \geq 1 \end{cases}$$

Esempio 1.4.4.1 (Elevamento a potenza). Sia $\Sigma = \{\mathbf{a}, \dots, \mathbf{z}\}$ l'alfabeto composto da 26 lettere, e sia $A = \{\mathbf{uno}, \mathbf{due}\}$ un linguaggio su Σ . Allora, si ha che

$$A^2 = \{\varepsilon, \mathbf{uno}, \mathbf{due}, \mathbf{unouno}, \mathbf{unodue}, \mathbf{dueuno}, \mathbf{duedue}\}$$

Proposizione 1.4.4.1: Chiusura sull'elevamento a potenza

Sia A un linguaggio regolare su un alfabeto Σ , ed $n \in \mathbb{N}$; allora A^n è regolare.

Dimostrazione. La dimostrazione procede per induzione su $n \in \mathbb{N}$.

Caso base. Per $n = 0$, si ha che $A^0 = \{\varepsilon\}$, il quale è regolare poiché ad esempio il DFA

$$M_\varepsilon = (\{q\}, \Sigma, \delta, q, \{q\})$$

è in grado di riconoscerlo.

Ipotesi induttiva. Per $n \in \mathbb{N}$, si assume che A^n è regolare.

Passo induttivo. Per il caso $n + 1$, si ha che $A^{n+1} = A^n \circ A$ per definizione dell'elevamento a potenza; inoltre, per ipotesi induttiva A^n è regolare, A è regolare per ipotesi, e poiché REG è chiuso rispetto alla concatenazione per la [Proposizione 1.4.3.1](#), si ha che $A^{n+1} = A^n \circ A$ è regolare.

□

1.4.5 Star

Definizione 1.4.5.1: Star

Sia A un linguaggio; allora, si definisce l'operazione unaria **star** che definisce il seguente linguaggio:

$$A^* := \{x_1 \cdots x_k \mid k \geq 0 \wedge \forall i \in [1, k] \quad x_i \in A\} = \bigcup_{n \in \mathbb{N}} L^n = \{\varepsilon\} \cup L \cup L^2 \cup \dots$$

Si noti che $k = 0 \implies \varepsilon \in A^*$ per ogni linguaggio A .

Esempio 1.4.5.1 (Star). Sia $\Sigma = \{\mathbf{a}, \dots, \mathbf{z}\}$ l'alfabeto composto da 26 lettere, e sia $A = \{\mathbf{uno}, \mathbf{due}\}$ un linguaggio su Σ . Allora, si ha che

$$A^* := \{\varepsilon, \mathbf{uno}, \mathbf{due}, \mathbf{unouno}, \mathbf{unodue}, \mathbf{dueuno}, \mathbf{duedue}, \dots\}$$

Esempio 1.4.5.2 (Stringhe binarie). Si noti che nel caso di $\Sigma = \{0, 1\}$, si ha che Σ^* è l'insieme di ogni stringa binaria, di arbitraria lunghezza.

Proposizione 1.4.5.1: Chiusura sull'operazione star

Sia A un linguaggio regolare su un alfabeto Σ ; allora A^* è regolare.

Dimostrazione. Per definizione, A è un linguaggio regolare, dunque per il [Teorema 1.3.2.1](#) esiste un NFA

$$N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$$

tale da riconoscere A . Allora, sia $N = (Q, \Sigma, \delta, q_0, F)$ l'NFA costruito come segue:

- $Q := \{q_0\} \cup Q_1$, dove q_0 è un nuovo stato, posto prima di N_1 ;
- q_0 è il nuovo stato iniziale;
- $F := \{q_0\} \cup F_1$, poiché q_0 deve essere accettante, in modo tale da accettare ε (si noti che $\varepsilon \in A^*$ per ogni linguaggio A);
- $\forall q \in Q, a \in \Sigma_\varepsilon \quad \delta(q, a) := \begin{cases} \delta_1(q, a) & q \in Q_1 - F_1 \vee (q \in F_1 \wedge a \neq \varepsilon) \\ \delta_1(q, a) \cup \{q_1\} & q \in F_1 \wedge a = \varepsilon \\ \{q_1\} & q = q_0 \wedge a = \varepsilon \\ \emptyset & q = q_0 \wedge a \neq \varepsilon \end{cases},$

scelta tale da ricominciare l'esecuzione dell'automa ogni volta che viene raggiunto uno stato accettante in N_1 ; infatti, se $q \in Q_1 - F_1$ (è uno stato non accettante di N_1), oppure q è accettante e $a \neq \varepsilon$, l'esecuzione procede normalmente con $\delta_1(q, a)$; diversamente, se $q \in F_1$ ma $a = \varepsilon$, allora l'esecuzione deve ricominciare da capo per poter effettuare la concatenazione multipla delle stringhe in A che caratterizzano l'operazione star, e dunque a $\delta_1(q, a)$ viene aggiunto q_1 (lo stato iniziale di N_1); infine, ponendo $\delta(q_0, \varepsilon) := \{q_1\}$ si realizza l' ε -arco iniziale che collega q_0 (il nuovo stato) a q_1 , al fine di rendere N in grado di accettare ε .



Figura 1.7: Rappresentazione dell'NFA N descritto.

Allora, poichè l'NFA N è in grado di ricominciare l'esecuzione ogni volta che questa sarebbe terminata in N_1 , è in grado di accettare molteplici copie concatenate delle stringhe in A , in maniera non deterministica, e dunque per definizione N riconosce A^* , il quale risulta allora essere regolare per definizione. \square

1.4.6 Complemento

Definizione 1.4.6.1: Complemento

Sia A un linguaggio definito su un alfabeto Σ ; allora si definisce il **complemento** di A il seguente linguaggio:

$$\neg A := \{x \mid x \notin A\}$$

Esempio 1.4.6.1 (Complemento). Sia $\Sigma = \{0, 1\}$ l'alfabeto binario, e sia $A = \{00, 1\}$ un linguaggio definito su di esso. Allora il suo complemento è il seguente:

$$\neg A = \{\varepsilon, 0, 1, 01, 10, 000, \dots\}$$

Proposizione 1.4.6.1: Chiusura sul complemento

Sia A un linguaggio regolare su un alfabeto Σ ; allora $\neg A$ è regolare.

Dimostrazione. In ipotesi A è un linguaggio regolare, dunque per definizione esiste un DFA

$$M = (Q, \Sigma, \delta, q_0, F)$$

che lo riconosce; allora, è sufficiente considerare il DFA

$$M' := (Q, \Sigma, \delta, q_0, Q - F)$$

per ottenere un DFA tale da riconoscere $\neg A$; allora, $\neg A$ è regolare per definizione. \square

Definizione 1.4.6.2: Leggi di De Morgan

Siano A e B definiti su un alfabeto Σ ; allora, sono vere le seguenti, che prendono il nome di **leggi di De Morgan**:

$$i) \neg(A \cup B) = \neg A \cap \neg B$$

$$ii) \neg(A \cap B) = \neg A \cup \neg B$$

1.5 Espressioni regolari

1.5.1 Definizioni

Definizione 1.5.1.1: Espressione regolare

Sia Σ un alfabeto; allora, R si definisce **espressione regolare** se soddisfa una delle seguenti caratteristiche:

- $R = \emptyset$
- $R = \varepsilon$
- $R \in \Sigma$

Un'espressione regolare, dunque, è un modo compatto di definire un linguaggio. Si noti che le definizioni successive sono in grado di espandere la definizione appena fornita. Dato un alfabeto Σ , la classe delle espressioni regolari definite su di esso è denotata con $\text{re}(\Sigma)$.

Data un'espressione regolare R , con $L(R)$ si intende il linguaggio che R descrive, ovvero l'insieme di stringhe che R rappresenta. Dunque, sono vere le seguenti:

- $L(\emptyset) = \emptyset$
- $L(\varepsilon) = \{\varepsilon\}$
- $\forall a \in \Sigma \quad L(a) = \{a\}$

Definizione 1.5.1.2: Unione

Siano R_1 ed R_2 due espressioni regolari su un alfabeto Σ ; allora, si definisce l'**unione** di R_1 ed R_2 la seguente espressione regolare:

$$(R_1 \cup R_2)$$

e rappresenta uno qualsiasi dei caratteri di R_1 o di R_2 . Dunque, si ha che

$$\forall R \in \text{re}(\Sigma) \mid \exists R_1, R_2 \in \text{re}(\Sigma) : R = R_1 \cup R_2 \quad L(R) = L(R_1) \cup L(R_2)$$

Si noti che, per ogni espressione regolare R , è vero che $\emptyset \cup R = R \cup \emptyset = \emptyset$.

Esempio 1.5.1.1 (Unione). Sia $\Sigma = \{a, b, c\}$ un alfabeto; un esempio di espressione regolare di unione su Σ è il seguente:

$$R = (a \cup c)$$

ed il valore di questa espressione equivale da a oppure c , e dunque $L(R) = \{a, c\}$.

Esempio 1.5.1.2 (Espressioni regolari particolari). Sia $\Sigma = \{0, 1\}$ un alfabeto; l'espressione regolare $(0 \cup 1)$ rappresenta il linguaggio che consiste di tutte le stringhe di lunghez-

za 1 sull'alfabeto Σ , e dunque l'espressione regolare descritta si abbrevia talvolta con il simbolo Σ stesso.

Definizione 1.5.1.3: Concatenazione

Siano R_1 ed R_2 due espressioni regolari; allora, si definisce la **concatenazione** di R_1 ed R_2 la seguente espressione regolare:

$$(R_1 \circ R_2)$$

e rappresenta le stringhe che iniziano per R_1 e terminano con R_2 . Dunque, si ha che

$$\forall R \in \text{re}(\Sigma) \mid \exists R_1, R_2 \in \text{re}(\Sigma) : R = R_1 \circ R_2 \quad L(R) = L(R_1) \circ L(R_2)$$

Per indicare la concatenazione di R con sé stessa, si usa la seguente notazione

$$R^k := \underbrace{R \circ \dots \circ R}_{k \text{ volte}}$$

Si noti che, per ogni espressione regolare R , è vero che:

- $\emptyset \circ R = R \circ \emptyset = \emptyset$
- $\varepsilon \circ R = R \circ \varepsilon = R$

Inoltre, il simbolo \circ può essere talvolta omissso.

Esempio 1.5.1.3 (Concatenazione). Sia $\Sigma = \{a, b, c\}$ un alfabeto; un esempio di espressione regolare di concatenazione su Σ è il seguente:

$$R = (a \circ c)$$

che può essere scritto equivalentemente come ac , e dunque $L(R) = \{ac\}$.

Definizione 1.5.1.4: Star

Sia R un'espressione regolare; allora, si definisce l'operazione unaria **star** su R la seguente espressione regolare:

$$(R^*)$$

e tutte rappresenta le stringhe che si possono ottenere concatenando un qualsiasi numero di caratteri di R , e descrive dunque il linguaggio che consiste di tutte le stringhe dell'alfabeto di R . Dunque, si ha che

$$\forall R \in \text{re}(\Sigma) \mid \exists R_1 \in \text{re}(\Sigma) : R = R_1^* \quad L(R) = L(R_1)^*$$

Si noti che R^* comprende ε per qualsiasi espressione regolare R .

Spesso si usa la notazione $R^+ := RR^*$, ovvero le stringhe che si ottengono attraverso la concatenazione di 1 o più stringhe di R . Di conseguenza, si ha che $R^+ \cup \varepsilon = R^*$.

Esempio 1.5.1.4 (Star). Sia $\Sigma = \{a, b, c\}$ un alfabeto; un esempio di espressione regolare di star su Σ è il seguente:

$$R = (\Sigma^*)$$

che descrive il linguaggio che consiste di tutte le stringhe possibili sull'alfabeto, e dunque

$$L(R) = \{\varepsilon, a, b, c, aa, bb, cc, \dots\}$$

Esempio 1.5.1.5 (Espressioni regolari). Sia $\Sigma = \{0, 1\}$ un alfabeto; i seguenti sono esempi di espressioni regolari su Σ :

- $L(0^*10^*) = \{w \mid w \text{ contiene un solo } 1\}$;
- $L(\Sigma^*001\Sigma^*) = \{w \mid w \text{ contiene la stringa } 001 \text{ come sottostringa}\}$;
- $L(0\Sigma^*0 \cup 1\Sigma^*1 \cup 0 \cup 1) = \{w \mid w \text{ inizia e termina con lo stesso carattere}\}$, poiché si noti che l'ordine delle operazioni, a meno di parentesi, è (i) star, (ii) concatenazione, ed infine (iii) unione;
- $L((0 \cup \varepsilon)(1 \cup \varepsilon)) = \{\varepsilon, 0, 1, 01\}$;
- $L(\emptyset^*) = \{\varepsilon\}$, poiché \emptyset rappresenta il linguaggio vuoto, e dunque l'unica stringa che si può ottenere concatenando un qualsiasi numero di volte elementi del linguaggio vuoto, è ε .

Definizione 1.5.1.5: Linguaggi descritti da espressioni regolari

Dato un alfabeto Σ , si definisce **classe dei linguaggi di Σ descritti da espressioni regolari** il seguente insieme:

$$\mathcal{L}(\text{re}(\Sigma)) := \{L \subseteq \Sigma^* \mid \exists R \in \text{re}(\Sigma) : L = L(R)\}$$

1.6 Configurazioni

1.6.1 Configurazioni di DFA

Definizione 1.6.1.1: Estensione di δ (DFA)

Sia $M = (Q, \Sigma, \delta, q_0, F)$ un DFA; è possibile estendere la definizione della funzione di transizione δ , utilizzando la notazione dell'operazione star, mediante la seguente definizione ricorsiva:

$$\delta^* : Q \times \Sigma^* \rightarrow Q : (q, x) \mapsto \begin{cases} \delta^*(q, \varepsilon) = \delta(q, \varepsilon) \\ \delta^*(q, by) = \delta^*(\delta(q, b), y) \quad b \in \Sigma, y \in \Sigma^* \mid x = by \end{cases}$$

Si noti che tale funzione prende in input uno stato ed una stringa, e restituisce lo stato in cui il DFA si troverà al termine della lettura dell'intera stringa di input. La notazione δ^* è coerente con la definizione dell'operazione star, poiché viene calcolata la transizione di stati attraverso δ fintanto che l'input non è stato esaurito.

Definizione 1.6.1.2: Configurazione (DFA)

Sia $M = (Q, \Sigma, \delta, q_0, F)$ un DFA; una tupla $(q, x) \in Q \times \Sigma^*$ è detta **configurazione** di M se q è pari allo stato attuale della computazione di un certo input, ed x è la porzione di input rimanente da leggere.

Definizione 1.6.1.3: Relazione tra configurazioni (DFA)

Sia $M = (Q, \Sigma, \delta, q_0, F)$ un DFA, e siano $(p, x), (q, y) \in Q \times \Sigma^*$ due sue configurazioni durante la computazione di un certo input; allora, tali due configurazioni si dicono essere **in relazione** se e solo se dall'una è possibile passare all'altra. In simboli:

$$(p, x) \vdash_M (q, y) \iff \begin{cases} p, q \in Q \\ x, y \in \Sigma^* \\ \exists a \in \Sigma \mid x = ay \wedge \delta(p, a) = q \end{cases}$$

Osservazione 1.6.1.1: Chiusura transitiva di \vdash

Sia $M = (Q, \Sigma, \delta, q_0, F)$ un DFA; si noti che la chiusura transitiva della relazione tra configurazioni \vdash , ovvero \vdash^* , equivale a calcolare gli input attraverso la funzione di transizione estesa δ^* definita nella [Definizione 1.6.1.1](#).

Esempio 1.6.1.1 (Chiusura transitiva di \vdash). Sia $M = (Q, \sigma, \delta, q_0, F)$ un DFA, e sia $(p, x) \in Q \times \Sigma^*$ una sua configurazione durante la computazione di un certo input; inoltre, siano $a, b, c \in \Sigma$ tali che $x = abc$. Inoltre, siano vere le seguenti:

- $(p, abc) \vdash_M (p_1, bc)$
- $(p_1, bc) \vdash_M (p_2, c)$
- $(p_2, c) \vdash_M (q, \varepsilon)$

per certi $p, p_1, p_2, q \in Q$; allora, si ha che $(p, x) \vdash_M^* (q, \varepsilon)$, poiché è possibile raggiungere lo stato q , partendo da p , attraverso l'input $x = abc$.

Osservazione 1.6.1.2: Determinismo

Dato un DFA $M = (Q, \Sigma, \delta, q_0, F)$, è possibile fornire una definizione di *determinismo* alternativa, attraverso la relazione tra configurazioni definita nella [Definizione 1.6.1.3](#), come segue:

$$\forall q \in Q, a \in \Sigma, x \in \Sigma^* \quad \exists! p \in Q \mid (q, ax) \vdash_M (p, x)$$

1.6.2 Configurazioni di NFA

TODO

Osservazione 1.6.2.1: Linguaggio di un automa

Dato un automa M , la [Definizione 1.2.1.3](#) si può riscrivere in simboli come segue:

$$L(M) := \{x \in \Sigma^* \mid \delta^*(q_0, x) \in F \iff \exists q \in F \mid (q_0, x) \vdash_M^* (q, \varepsilon)\}$$

Si noti che, nella definizione, verrà presa in considerazione la funzione δ^* corrispondente alla classe dell'automata M in questione.

1.7 Non determinismo generalizzato

1.7.1 Definizioni

Definizione 1.7.1.1: GNFA

Un **GNFA** (*Generalized Nondeterministic Finite Automaton*) è una versione generalizzata di un NFA, in cui gli archi delle transizioni sono espressioni regolari sull'alfabeto dato.

Un GNFA legge blocchi di simboli dall'input, e si muove lungo gli archi che sono etichettati da espressioni regolari che possono descrivere il blocco di simboli letto.

Inoltre, essendo una versione generalizzata di un NFA, un GNFA può avere diversi modi di elaborare la stessa stringa di input, e accetta quest'ultima se la sua elaborazione può far sì che il GNFA si trovi in uno stato accettante al suo termine.

Nota: All'interno di questi appunti, a meno di specifica, si assume che ogni GNFA preso in considerazione abbia:

- un solo stato di inizio, privo di archi entranti, connesso con ogni altro stato, ma non con sé stesso;
- un solo stato accettante, privo di archi uscenti, non connesso con altri archi;
- ogni altro stato collegato con ogni altro stato, a meno di quello iniziale, anche con sé stessi.

Formalmente, dato un alfabeto Σ , un GNFA del tipo appena descritto è una quintupla $(Q, \Sigma, \delta, q_{\text{start}}, q_{\text{accept}})$ definita come segue:

- Q è l'**insieme degli stati** dell'automa, un insieme *finito*
- Σ è l'**alfabeto dell'automa**, un insieme *finito*
- $\delta : (Q - \{q_{\text{accept}}\}) \times (Q - \{q_{\text{start}}\}) \rightarrow \text{re}(\Sigma)$ è la **funzione di transizione**, che definisce la relazione tra gli stati; si noti che δ ha come dominio il prodotto cartesiano tra gli stati, e come codominio $\text{re}(\Sigma)$, poiché a differenza di un normale DFA o NFA, un GNFA prende come input 2 stati (che non possono essere né quello iniziale né quello accettante) e restituisce un'espressione regolare; inoltre, il dominio di δ è $(Q - \{q_{\text{accept}}\}) \times (Q - \{q_{\text{start}}\})$ per evitare di includere l'arco $(q_{\text{accept}}, q_{\text{start}})$
- $q_{\text{start}} \in Q$ è lo **stato iniziale**
- $q_{\text{accept}} \in Q$ è lo **stato accettante**

Esempio 1.7.1.1 (GNFA). Il seguente è il digramma di un GNFA sull'alfabeto $\Sigma = \{a, b\}$.

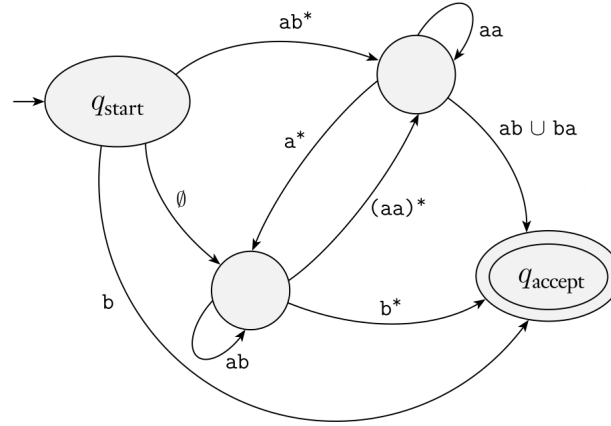


Figura 1.8: Un GNFA.

Definizione 1.7.1.2: Stringhe accettate (GNFA)

Sia $G = (Q, \Sigma, \delta, q_{\text{start}}, q_{\text{accept}})$ un GNFA, e sia $w = w_1 \cdots w_n$ una stringa tale per cui $\forall i \in [1, n] \quad w_i \in \Sigma^*$; allora, G **accetta** w se esiste una sequenza di stati $q_0, \dots, q_n \in Q$ tali per cui

- $q_0 = q_{\text{start}}$
- $\forall i \in [0, n-1] \quad w_i \in L(\delta(q_i, q_{i+1}))$, ovvero, w_i deve far parte del linguaggio rappresentato dall'espressione regolare sull'arco da q_i a q_{i+1}
- $q_n = q_{\text{accept}}$

1.7.2 Equivalenze**Metodo 1.7.2.1: GNFA di DFA**

Sia M un DFA; allora per costruire un GNFA ad esso equivalente, è sufficiente:

- aggiungere un nuovo stato iniziale, con un ε -arco entrante sul vecchio stato iniziale;
- aggiungere un nuovo stato accettante, con ε -archi entranti provenienti dai vecchi stati accettanti;
- sostituire gli archi con etichette multiple, con archi aventi come etichetta l'unione delle etichette;
- aggiungere archi etichettati con \emptyset tra gli stati non collegati (si noti che questa operazione non varia l'automa di partenza, poiché un arco etichettato con \emptyset non potrà mai essere utilizzato)

Esempio 1.7.2.1 (GNFA di DFA). Si consideri il seguente DFA

$$D = (\{b, c\}, \{0, 1, 2\}, \delta, b, \{c\})$$

rappresentato come segue:

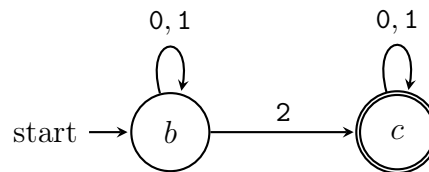


Figura 1.9: Il DFA D .

Allora, il suo GNFA equivalente è il seguente:



Figura 1.10: Il GNFA di D .

Algoritmo 1.7.2.1: Espressione regolare di un GNFA

Dato un GNFA $G = (Q, \Sigma, \delta, q_{\text{start}}, q_{\text{accept}})$, l'algoritmo restituisce un'espressione regolare equivalente a G .

```

1: function CONVERTGNFATOREGEX( $G$ )
2:   if  $|Q| == 2$  then
3:     return  $\delta(q_{\text{start}}, q_{\text{accept}})$ 
4:   else if  $|Q| > 2$  then
5:      $q \in Q - \{q_{\text{start}}, q_{\text{accept}}\}$ 
6:      $Q' := Q - \{q\}$ 
7:     for  $q_i \in Q' - \{q_{\text{accept}}\}$  do
8:       for  $q_j \in Q' - \{q_{\text{start}}\}$  do
9:          $\delta'(q_i, q_j) := \delta(q_i, q)\delta(q, q)^*\delta(q, q_j) \cup \delta(q_i, q_j)$ 
10:      end for
11:    end for
12:     $G' := (Q', \Sigma, \delta', q_{\text{start}}, q_{\text{accept}})$ 
13:    return convertGNFatoRegEx( $G'$ )
14:  end if
15: end function

```

Idea. L'algoritmo inizia prendendo in input un GNFA G , ed inizialmente viene controllato il numero di stati di G :

- se $|Q| = 2$, allora sicuramente $Q = \{q_{\text{start}}, q_{\text{accept}}\}$, e poiché si vuole restituire l'espressione regolare equivalente a G , di fatto quest'ultimo è costituito esclusivamente dall'espressione regolare posta sull'arco tra q_{start} e q_{accept} , dunque è sufficiente restituirla in output (riga 3);
- se $|Q| > 2$, allora viene costruito un GNFA G' , avente uno stato in meno, ovvero q (scelto alla riga 5), naturalmente diverso da q_{start} e da q_{accept} ; successivamente, per ogni coppia di stati (q_i, q_j) , viene definita $\delta'(q_i, q_j)$ in modo tale da accorpare tutte le possibili configurazioni di stati; ad esempio, prendendo in esame il seguente GNFA



dove R_1 , R_2 , R_3 ed R_4 sono espressioni regolari, è possibile vedere che il seguente GNFA è ad esso equivalente



poiché l'arco che q ha su sé stesso è stato descritto attraverso $(R_2)^*$, gli archi (q_i, q) e (q, q_j) sono stati inseriti per concatenazione, ed infine è stato unito l'altro possibile cammino verso q_j tramite unione;

- inoltre, si noti che tale espressione regolare — contenuta nella riga 9 — tiene in considerazione tutte le possibili configurazioni di archi tra stati di un GNFA, per come è stato definito il GNFA all'interno della [Definizione 1.7.1.1](#); infatti, per $|Q| > 2$, tra due stati q_i e q_j , oltre ad avere la garanzia che esista l'arco (q_i, q_j) , esiste sicuramente un terzo stato q intermedio tale per cui esistano archi (q_i, q) e (q, q_j) , ed esiste anche l'arco (q, q) ;
- infine, sia per la [Definizione 1.7.1.1](#), sia per come la riga 5 dell'algoritmo opera, per ogni GNFA si ha che $|Q| \geq 2$, dunque non è necessario gestire ulteriori casi.

Allora, l'algoritmo è in grado di restituire l'espressione regolare equivalente a G .

Dimostrazione. La dimostrazione procede per induzione su k , il numero di stati del GNFA.

Caso base. Se $k = 2$, ovvero se G ha solo 2 stati, necessariamente $Q = \{q_{\text{start}}, q_{\text{accept}}\}$, e dunque $\text{convertGNFatoRegEx}(G) = \delta(q_{\text{start}}, q_{\text{accept}})$, come descritto nella riga 3.

Ipotesi induttiva. Se G è un GNFA con $k - 1$ stati, allora $\text{convertGNFatoRegEx}(G)$ è un'espressione regolare equivalente a G .

Passo induttivo. È necessario dimostrare che, se G è un GNFA con k stati, allora $\text{convertGNFatoRegEx}(G)$ è un'espressione regolare equivalente a G . In primo luogo, è necessario dimostrare che G e G' sono equivalenti, e dunque sia w una stringa accettata da G ; allora, in un ramo accettante della computazione, G entra in una sequenza di stati

$$q_{\text{start}}, q_1, q_2, \dots, q_{\text{accept}}$$

dunque, se il ramo non contiene lo stato rimosso q , allora sicuramente G' accetta w ; viceversa, se q è contenuto all'interno di tale ramo, allora per quanto discusso all'interno dell'idea di dimostrazione, l'espressione regolare inserita al posto dello stato q rimosso è in grado di tenere in considerazione ogni possibile configurazione di archi, e dunque G' accetta sicuramente w ; infine, è possibile applicare la stessa osservazione per mostrare che una stringa accettata da G' deve essere necessariamente accettata anche da G . Allora, poiché G e G' sono equivalenti, e G ha k stati, allora necessariamente al termine del k -esimo passo dell'algoritmo, G' avrà $k - 1$ stati, e su di esso è possibile applicare l'ipotesi induttiva.

□

Esempio 1.7.2.2 (Espressioni regolari di GNFA). Sia G il GNFA dell'[Esempio 1.7.2.1](#); allora, la sua espressione regolare equivalente è ottenibile attraverso i seguenti passaggi:

- viene inizialmente rimosso lo stato b , ottenendo l'automa G' :


 Figura 1.11: G' , ovvero il GNFA G , dopo aver rimosso b .

- successivamente, viene rimosso lo stato c , ottenendo l'automa G'' :


 Figura 1.12: G'' , ovvero il GNFA G' , dopo aver rimosso c .

- allora, poiché G'' ha 2 stati, l'espressione regolare cercata — equivalente a G — è posta sull'arco (a, d) , ed è

$$(0 \cup 1)^*2(0 \cup 1)^*$$

Teorema 1.7.2.1: Linguaggi ed espressioni regolari

Un linguaggio è regolare se e solo se esiste un'espressione regolare che lo descrive; dunque, tutti e soli i linguaggi regolari sono descritti da espressioni regolari. In simboli, per un certo alfabeto Σ si ha che

$$\text{REG} = \mathcal{L}(\text{re}(\Sigma))$$

Dimostrazione.

Prima implicazione. Sia A un linguaggio regolare; allora, per definizione, esiste un DFA che lo riconosce, e sia questo M . Utilizzando il [Metodo 1.7.2.1](#), è possibile costruire un GNFA che sia equivalente ad M ; sia quest'ultimo G . Allora, è sufficiente applicare l'[Algoritmo 1.7.2.1](#) su G per ottenere l'espressione regolare ad esso equivalente; dunque, segue la tesi.

Seconda implicazione. Sia A un linguaggio su un alfabeto Σ , descritto da un'espressione regolare R . Allora, la dimostrazione procede costruendo degli NFA, per casi, come segue:

- $R = \emptyset \implies L(R) = \emptyset$; allora, il seguente NFA è in grado di riconoscere $L(R)$:

Figura 1.13: Un NFA in grado di riconoscere $L(\emptyset)$.

L'automa mostrato è descritto dalla quintupla $N = (\{q_0\}, \Sigma, \delta, q_0, \emptyset)$, dove $\forall a \in \Sigma \quad \delta(q_0, a) = \emptyset$.

- $R = \varepsilon \implies L(R) = \{\varepsilon\}$; allora, il seguente NFA è in grado di riconoscere $L(R)$:

Figura 1.14: Un NFA in grado di riconoscere $L(\varepsilon)$.

L'automa mostrato è descritto dalla quintupla $N = (\{q_0\}, \Sigma, \delta, q_0, \{q_0\})$, dove $\forall a \in \Sigma \quad \delta(q_0, a) = \emptyset$.

- $R \in \Sigma \implies L(R) = \{a\}$ per qualche $a \in \Sigma$; allora, il seguente NFA è in grado di riconoscere $L(R)$:

Figura 1.15: Un NFA in grado di riconoscere $L(a)$.

L'automa mostrato è descritto dalla quintupla $N = (\{q_1, q_2\}, \Sigma, \delta, q_1, \{q_2\})$, dove $\forall q \in \{q_1, q_2\}, x \in \Sigma \quad \delta(q, x) = \begin{cases} \{q_2\} & q = q_1 \wedge x = a \\ \emptyset & q \neq q_1 \vee x \neq a \end{cases}$.

- si noti che se esistono due espressioni regolari R_1 ed R_2 tali che $R = R_1 \cup R_2$, oppure $R = R_1 \circ R_2$, o ancora $R = R_1^*$, è sufficiente costruire gli NFA che sono stati costruiti nelle dimostrazioni della [Proposizione 1.4.1.1](#), della [Proposizione 1.4.3.1](#) e della [Proposizione 1.4.5.1](#).

Allora, per qualsiasi espressione regolare R , tale da descrivere un certo linguaggio A , è possibile costruire un NFA che riconosce il linguaggio che R descrive. Allora, per il [Corollario 1.3.2.1](#), A è regolare.

□

Osservazione 1.7.2.1: Equivalenze tra classi di automi

Si noti che, dal [Teorema 1.7.2.1](#), dall'[Algoritmo 1.7.2.1](#), e dal [Metodo 1.7.2.1](#), si ha che

$$\text{REG} = \mathcal{L}(\text{re}(\Sigma)) \supseteq \mathcal{L}(\text{GNFA}) \supseteq \text{REG}$$

dunque, segue l'importante risultato:

$$\text{REG} := \mathcal{L}(\text{DFA}) = \mathcal{L}(\text{NFA}) = \mathcal{L}(\text{GNFA}) = \mathcal{L}(\text{re}(\Sigma))$$

1.8 Linguaggi non regolari

1.8.1 Pumping lemma

Principio 1.8.1.1: Principio della piccionaia

Siano A e B due insiemi finiti, tali che $|B| < |A|$; allora, non esiste alcuna funzione iniettiva $f : A \rightarrow B$.

In altri termini, avendo una piccionaia con m caselle, non è possibile inserire più di m piccioni al suo interno: alcuni volatili dovranno necessariamente condividere la propria casella.

Definizione 1.8.1.1: Linguaggio non regolare

Un linguaggio A si definisce **non regolare** se non esiste un DFA in grado di riconoscerlo. In simboli

$$A \notin \text{REG}$$

Lemma 1.8.1.1: Pumping lemma (REG)

Sia A un linguaggio regolare; allora, esiste un $p \in \mathbb{N}$, detto **lunghezza del pumping**, tale che per ogni stringa $s \in A$ tale per cui $|s| \geq p$, esistono 3 stringhe $x, y, z \mid s = xyz$, soddisfacenti le seguenti condizioni:

- $\forall i \geq 0 \quad xy^iz \in A$
- $|y| > 0$ (o, equivalentemente, $y \neq \varepsilon$)
- $|xy| \leq p$

Dimostrazione. Poiché A è un linguaggio regolare in ipotesi, per definizione esiste un DFA $M = (Q, \Sigma, \delta, q_1, F)$ in grado di riconoscerlo. Allora, sia $p := |Q|$, e sia $s \in A \mid s = s_1s_2 \cdots s_n$ tale che $\forall i \in [1, n] \quad s_i \in \Sigma$, e $n \geq p$. Inoltre, sia

$$\forall i \in [1, n] \quad r_{i+1} := \delta(r_i, s_i)$$

la sequenza di stati attraversati da M mentre elabora s ; dunque, si ha che $r_{n+1} \in F$. Si noti che la sequenza di stati ha dunque cardinalità

$$|\{r_1, \dots, r_{n+1}\}| = n + 1$$

in quanto devono essere attraversati n archi, e dunque $n + 1$ stati. Inoltre, $n \geq p \iff n + 1 \geq p + 1$, e dunque per il [Principio 1.8.1.1](#), poiché $p := |Q|$, due dei primi $p + 1$ stati della sequenza devono necessariamente essere lo stesso stato; siano r_j il primo ed r_l il secondo, con $j \neq l$. Allora, sicuramente $l \leq p + 1$, poiché r_l è uno stato tra i primi $p + 1$.

Si pongano dunque

$$\begin{cases} x := s_1 \cdots s_{j-1} \\ y := s_j \cdots s_{l-1} \\ z := s_l \cdots s_n \end{cases}$$

Allora, si ha che:

- x porta M da r_1 ad r_j , y porta M da $r_j = r_l$ ad r_j — dunque y porta M da r_j ad r_j stesso — ed infine z porta M da r_j ad r_{n+1} , e poiché $r_{n+1} \in F$, M accetta sicuramente xy^iz per ogni $i \geq 0$;
- $j \neq l \implies |y| > 0$;
- $l \leq p + 1 \implies |xy| \leq p$, poiché r_l è lo stato che viene dopo aver letto $s_{l-1} \in y$, e dunque nel caso limite si ha che $l = p + 1 \implies |xy| = p$.

Allora, sono soddisfatte tutte le condizioni della tesi. \square

Esempio 1.8.1.1 (Dimostrazione del pumping lemma). La seguente rappresentazione raffigura un automa definito come segue:

$$M = (\{q_1, \dots, q_9, \dots, q_{13}\}, \Sigma, \delta, q_1, \{q_{13}\})$$

dove $|Q| = 13$, e $r_j = r_l = q_9$ è lo stato che si ripete all'interno dei primi $p + 1$ stati della sequenza presa in esame nella dimostrazione del [Lemma 1.8.1.1](#).

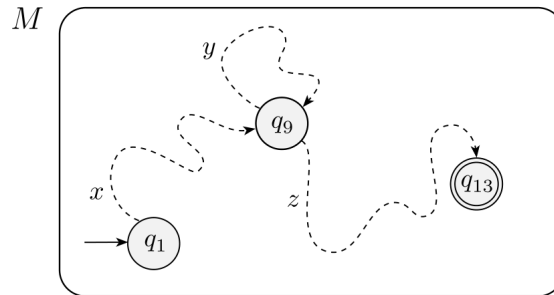


Figura 1.16: L'automata M descritto

Esempio 1.8.1.2 (Pumping lemma in REG). Si consideri il linguaggio

$$L := \{0^n 1^n \mid n \in \mathbb{N}\}$$

e per assurdo, sia $L \in \text{REG}$, e dunque per esso vale il [Lemma 1.8.1.1](#). Allora, sia $p \in \mathbb{N}$ la lunghezza del pumping di L , e si consideri la seguente stringa

$$s := 0^p 1^p \implies |s| = 2p > p$$

avente lunghezza sicuramente maggiore di p . Siano inoltre x, y, z tali da soddisfare il pumping lemma, ed in particolare $|xy| \leq p$, ma poiché $s := 0^p 1^p = xy^i z$ per ogni $i \in \mathbb{N}$, allora necessariamente la stringa xy è composta da soli 0, ed inoltre $|y| > 0$ implica che y ha almeno uno 0. Siano allora

- $k := |x| \implies x = 0^k$, poiché x è composta solo da 0
- $m := |y| > 0 \implies y = 0^m$, poiché y è composta solo da 0 (ed almeno uno)
- $|xy| \leq p \implies k + m \leq p \implies z = 0^{p-m-k} 1^p$ per gli 0 restanti

e, ponendo ad esempio $i = 2$, si ottiene che

$$xy^2 z = 0^k 0^{2m} 0^{p-m-k} 1^p$$

ma il numero di 1 di questa stringa è pari a p , mentre il numero di 0 è pari a

$$k + 2m + (p - m - k) = m + p$$

e poiché $m > 0$, si ha che $m + p > p$. Dunque $xy^2 z \notin L$ poiché non è una stringa della forma $0^n 1^n$. \nmid

Osservazione 1.8.1.1: Condizioni del pumping lemma (REG)

Si noti che la seconda condizione del [Lemma 1.8.1.1](#) stabilisce che $y \neq \varepsilon$, poiché altrimenti il teorema sarebbe trivialmente verificato; infatti, ammettendo $y = \varepsilon$, dato un linguaggio $A \in \text{REG}$, e presa una sua stringa $s \in A$, ponendo $p := |x|$ esistono sicuramente stringhe $x, z \mid s = xz$ tali da verificare le restanti condizioni del lemma, infatti:

- $\forall i \geq 0 \quad xy^i z = x\varepsilon^i z = xz =: s \in A$
- $|xy| = |x\varepsilon| = |x| \leq p := |x|$

e non sarebbe necessaria l'ipotesi per cui A debba essere regolare; dunque, non sarebbe possibile utilizzare tale lemma per determinare la classe di un dato linguaggio.

Linguaggi e grammatiche context-free

2.1 Grammatiche context-free

2.1.1 Definizioni

Definizione 2.1.1.1: Grammatica

Una grammatica è un insieme di regole di sostituzione di stringhe, in grado di produrre quest'ultime a partire da una variabile iniziale, mediante una sequenza di scambi. Le regole sono scritte nella forma

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

Definizione 2.1.1.2: Acontestualità

Si definisce **acontestualità** la condizione per cui il lato sinistro delle regole di una grammatica è composto sempre da un solo simbolo.

Definizione 2.1.1.3: CFG

Una **grammatica context-free** — detta anche **acontestuale** —, o CFG (*Context-Free Grammar*) è una quadrupla (V, Σ, R, S) , dove

- V è l'insieme delle **variabili**, un insieme *finito*
- Σ è l'insieme dei **terminali**, un insieme *finito*, dove $\Sigma \cap V = \emptyset$
- R è l'insieme delle **regole** o **produzioni**, un insieme *finito*
- $S \in V$ è la **variabile iniziale**, ed è generalmente il simbolo alla sinistra della prima regola della grammatica

Le CFG si scrivono nella forma

$$X \rightarrow Y$$

dove $X \in V$ e $X \rightarrow Y \in R$ è una regola della CFG. Due regole $X \rightarrow Y, X \rightarrow Z \in R$ possono essere accorpate con il simbolismo $X \rightarrow Y \mid Z$.

Siano u, v, w stringhe, e $A \rightarrow w \in R$; allora si dice che uAv **produce** uwv , denotato con

$$uAv \Rightarrow uwv$$

Date due stringhe u, v , se $u = v$, oppure esistono stringhe u_1, \dots, u_k con $k \geq 0$ tali che

$$u \Rightarrow u_1 \Rightarrow \dots \Rightarrow u_k \Rightarrow v$$

si dice che u **deriva** v , ed è denotato con $u \xRightarrow{*} v$.

Esempio 2.1.1.1 (CFG). Un esempio di CFG è il seguente:

$$\begin{aligned} A &\rightarrow 0A1 \\ A &\rightarrow B \\ B &\rightarrow \# \end{aligned}$$

In essa, si hanno:

$$\begin{aligned} V &:= \{A, B\} \\ \Sigma &:= \{0, 1, \#\} \\ S &:= A \in V \end{aligned}$$

Da essa, è possibile ottenere ad esempio la stringa 000#111 attraverso le seguenti sostituzioni:

$$A \Rightarrow 0A1 \Rightarrow 000A111 \Rightarrow 000B111 \Rightarrow 000\#111$$

Definizione 2.1.1.4: Linguaggio di una grammatica

Data una grammatica G , il suo **linguaggio** è l'insieme delle stringhe che la grammatica G è in grado di generare, ed è denotato con $L(G)$. In simboli, data una grammatica G , si ha che

$$L(G) = \{w \in \Sigma^* \mid S \xRightarrow{*} w\}$$

Definizione 2.1.1.5: CFL

Se G è una CFG, allora $L(G)$ è detto **linguaggio context-free**, o CFL (*Context-Free Language*).

Esempio 2.1.1.2 (CFL). Si prenda in esame la CFG dell'Esempio 2.1.1.1, e sia essa G . Allora, in tale grammatica il linguaggio risulta essere

$$L(G) = \{0^n \# 1^n \mid n \geq 0\}$$

Esempio 2.1.1.3 (Grammatiche e linguaggi). I seguenti sono esempi di linguaggi, e corrispondenti grammatiche che li descrivono:

- il linguaggio $L_1 := \{w \in \{0,1\}^* \mid w \text{ contiene almeno tre } 1\}$ è descritto dalla seguente grammatica:

$$\begin{aligned} S_1 &\rightarrow X1X1X \\ X &\rightarrow \varepsilon \mid 0X \mid 1X \end{aligned}$$

- il linguaggio $L_2 := \{w \in \{0,1\}^* \mid w = w^R \wedge |w| \equiv 0 \pmod{2}\}$ è descritto dalla seguente grammatica:

$$S \rightarrow 0S0 \mid 1S1 \mid \varepsilon$$

- il linguaggio $L_3 := \{a^i b^j c^{i+j} \mid i, j \geq 0\}$ è descritto dalla seguente grammatica:

$$\begin{aligned} S &\rightarrow aSc \mid X \\ X &\rightarrow bSc \mid \varepsilon \end{aligned}$$

Metodo 2.1.1.1: CFG di DFA

Dato un DFA $D = (Q, \Sigma, \delta, q_0, F)$, per ottenere una grammatica $G = (V, \Sigma, R, S)$ tale che $L(D) = L(G)$, è necessario:

- associare una variabile V_i , per ogni stato $q_i \in Q$, con $i \in [1, n]$ (assumendo $|Q| = n$);
- porre $S = V_0$;
- introdurre una regola $V_i \rightarrow aV_j$ se vale $\delta(q_i, a) = q_j$ per qualche coppia di stati $q_i, q_j \in Q$;
- aggiungere la regola $V_i \rightarrow \varepsilon$ se $q_i \in F$.

2.1.2 Ambiguità**Definizione 2.1.2.1: Derivazione a sinistra**

Sia G una grammatica; una stringa si dice essere **derivata a sinistra** se è stata ottenuta applicando regole di G sulle variabili più a sinistra disponibili.

Esempio 2.1.2.1 (Derivazione a sinistra). TODO

Definizione 2.1.2.2: Ambiguità

Sia G una grammatica; se esistono due stringhe u, v con $u \neq v$, tali che esiste una terza stringa z *derivata a sinistra* sia da u che da v attraverso le regole di G , si dice che G genera z **ambiguamente**.

Simmetricamente, G si dice essere **ambigua** se genera stringhe ambiguamente.

Esempio 2.1.2.2 (Stringhe generate ambiguamente). Si consideri la seguente grammatica:

$$E \rightarrow E+E \mid E*E \mid a$$

Attraverso essa, è possibile ottenere ad esempio la seguente stringa, applicando la derivazione a sinistra:

$$E \Rightarrow E*E \Rightarrow E+E*E \Rightarrow a+E*E \Rightarrow a+a*E \Rightarrow a+a*a$$

Si noti però che è possibile ottenere questa stringa applicando anche la seguente derivazione a sinistra:

$$E \Rightarrow E+E \Rightarrow a+E \Rightarrow a+E*E \Rightarrow a+a*E \Rightarrow a+a*a$$

Dunque, la grammatica descritta risulta essere ambigua, poiché esistono due derivazioni a sinistra per la stringa $a+a*a$.

Definizione 2.1.2.3: Linguaggi inerentemente ambigui

Un linguaggio si dice essere **inerentemente ambiguo** se non esistono grammatiche non ambigue che lo possano generare.

2.1.3 Forma normale di Chomsky

Definizione 2.1.3.1: Forma normale di Chomsky

Una CFG (V, Σ, R, S) è in **forma normale di Chomsky**, o CNF (*Chomsky Normal Form*), se ogni regola è della forma

$$\begin{aligned} A &\rightarrow BC \\ A &\rightarrow a \end{aligned}$$

dove $A, B, C \in V$, $a \in \Sigma$, $S := A$, e la regola $S \rightarrow \varepsilon \in R$ è sempre ammessa, la quale prende il nome di **ε -regola**.

Dunque, la CNF non permette di

- i) avere **regole unitarie**, ovvero regole della forma $X \rightarrow B \in R$, con $X, B \in V$;
- ii) avere regole della forma $X \rightarrow u \in R$, con $u \in (V \cup \Sigma)^* - \Sigma$;
- iii) avere regole della forma $X \rightarrow S \in R$, con $X \in V$.

Metodo 2.1.3.1: CNF di CFG

Sia $G = (V, \Sigma, R, S)$ una CFG; allora, è possibile rendere G in CNF attraverso i seguenti passaggi:

- vengono aggiunte la variabile $S_0 \in V$, e la regola $S_0 \rightarrow S \in R$, in modo da non avere S alla destra di nessuna regola in R ;
- ogni regola $A \rightarrow \varepsilon \in R$ — con $A \in V - \{S\}$ — viene rimossa da R , e successivamente per ogni regola della forma $X \rightarrow uAv \in R$ — per qualche $X \in V$ e $u, v \in (V \cup \Sigma)^*$ — viene aggiunta $X \rightarrow uv \in R$; inoltre, se $X \rightarrow A \in R$, allora viene aggiunta $X \rightarrow \varepsilon \in R$, solo se quest'ultima non era stata precedentemente rimossa;
- ogni regola unitaria $A \rightarrow B \in R$ viene rimossa, e per ogni regola $B \rightarrow u \in R$ — con $u \in (V \cup \Sigma)^*$ — viene aggiunta $A \rightarrow u \in R$, solo se questa non era una regola unitaria precedentemente rimossa;
- ogni regola restante $A \rightarrow u_1 \dots u_k$ con $k \geq 3$ — dove $u_1, \dots, u_k \in (V \cup \Sigma)^*$ — viene rimpiazzata con le regole

$$\begin{aligned} A &\rightarrow u_1 A_1 \\ A_1 &\rightarrow u_2 A_2 \\ A_2 &\rightarrow u_3 A_3 \\ &\vdots \\ A_{k-2} &\rightarrow u_{k-1} u_k \end{aligned}$$

dove $A_1, \dots, A_{k-2} \in V$, creando dunque una *catena di sostituzioni*, al fine di avere al più 2 variabili o terminali alla destra delle regole;

- infine, ogni $u_i \in \Sigma$ (dunque terminale) — per ogni $i \in [1, k]$ — viene rimpiazzato con una nuova variabile $U_i \in V$, e viene aggiunta la regola $U_i \rightarrow u_i \in R$.

Lemma 2.1.3.1: CFL generati da CFG in CNF

Ogni CFL è generato da una CFG in forma normale di Chomsky.

Dimostrazione. Sia L un CFL generato da una CFG G ; allora, attraverso il [Metodo 2.1.3.1](#), è possibile rendere G in forma normale di Chomsky, e dunque segue la tesi. \square

Esempio 2.1.3.1 (Convertire CFG in CNF). TODO

2.2 Operazioni context-free

2.2.1 Unione

Proposizione 2.2.1.1: Chiusura sull'unione

Siano L_1, \dots, L_n dei CFL; allora $\bigcup_{i=1}^n L_i$ è un CFL.

Dimostrazione. Poiché L_1, \dots, L_n sono CFL, allora esistono delle CFG G_1, \dots, G_n tali da generare rispettivamente ogni linguaggio in ipotesi; siano queste grammatiche definite come

$$\forall i \in [1, n] \quad G_i = (V_i, \Sigma_i, R_i, S_i)$$

Per dimostrare la tesi, è necessario dimostrare che esiste una CFG G tale che

$$L(G) = \bigcup_{i=1}^n L(G_i)$$

poiché $L(G) \in \text{CFL}$ per definizione; allora, si consideri la seguente CFG:

$$G = \left(\bigcup_{i=1}^n V_i \cup \{S\}, \bigcup_{i=1}^n \Sigma_i, \bigcup_{i=1}^n R_i \cup \{S \rightarrow S_1 \mid \dots \mid S_n\}, S \right)$$

e si noti che

$$w \in \bigcup_{i=1}^n L(G_i) \iff \exists j \in [1, n] \mid \begin{cases} w \in L(G_j) \iff S_j \xRightarrow{*} w \\ S \rightarrow S_j \in R \end{cases} \iff S \xRightarrow{*} w \iff w \in L(G)$$

Allora, segue la tesi. □

Esempio 2.2.1.1 (Unione di CFL). Siano L_1 ed L_2 due CFL generati rispettivamente dalle seguenti CFG G_1 e G_2 :

$$G_1 : S_1 \rightarrow 0S_11 \mid \varepsilon$$

$$G_2 : S_2 \rightarrow 1S_20 \mid \varepsilon$$

Allora, la seguente CFG G :

$$\begin{aligned} & S \rightarrow S_1 \mid S_2 \\ G : & S_1 \rightarrow 0S_11 \mid \varepsilon \\ & S_2 \rightarrow 1S_20 \mid \varepsilon \end{aligned}$$

è in grado di generare $L_1 \cup L_2$.

2.2.2 Concatenazione

Proposizione 2.2.2.1: Chiusura sulla concatenazione

TODO

Dimostrazione. Omessa. □

2.3 Automi a pila

2.3.1 Definizioni

Definizione 2.3.1.1: PDA

Un **PDA** (*Pushdown Automaton*) è una sestupla $(Q, \Sigma, \Gamma, \delta, q_0, F)$, dove

- Q è l'insieme degli stati, un insieme *finito*
- Σ è l'alfabeto dell'automa, un insieme *finito*
- Γ è l'alfabeto dello stack (o *pila*), un insieme *finito*
- $\delta : Q \times \Sigma_\varepsilon \times \Gamma_\varepsilon \rightarrow \mathcal{P}(Q \times \Gamma_\varepsilon)$ è la **funzione di transizione**, che definisce la relazione tra gli stati
- $q_0 \in Q$ è lo **stato iniziale**
- $F \subseteq Q$ è l'insieme degli stati accettanti

dove $\Gamma_\varepsilon := \Gamma \cup \{\varepsilon\}$.

Un PDA è un NFA dotato di uno **stack** illimitato, che gli consente di riconoscere alcuni linguaggi non regolari, poiché in esso è in grado di porre i simboli che legge dalla stringa di input, di fatto implementando un sistema di *memoria*.

Si noti che la macchina può usare differenti alfabeti per il suo input e la sua pila, infatti la definizione formale vede due alfabeti distinti, Σ e Γ rispettivamente. Inoltre, Γ_ε compare nel prodotto cartesiano del dominio di δ , poiché il simbolo in cima allo stack del PDA è in grado di determinare anch'esso la mossa seguente dell'automa; a tal proposito, $\varepsilon \in \Gamma$ permette di ignorare il primo elemento della pila. In aggiunta, Γ_ε compare anche all'interno dell'insieme potenza (si noti che un PDA è non deterministico) del codominio di δ , al fine di decidere se salvare il simbolo letto all'interno dello stack (tramite $\varepsilon \in \Gamma$ stesso). Dunque, un'operazione di un PDA sul suo stack può essere un *push*, un *pop*, o entrambe — ottenendo l'effetto di rimpiazzare l'elemento in cima allo stack.

Esempio 2.3.1.1 (PDA). Un esempio di PDA $M = (Q, \Sigma, \Gamma, \delta, q_1, F)$ è il seguente:

- $Q = \{q_1, q_2, q_3, q_4\}$
- $\Sigma = \{0, 1\}$
- $\Gamma = \{0, \$\}$
- $F = \{q_1, q_4\}$

e δ è data dalla seguente tabella di transizione:

Input:	0			1			ϵ		
Stack:	0	\$	ϵ	0	\$	ϵ	0	\$	ϵ
q_1									$\{(q_2, \$)\}$
q_2	$\{(q_2, 0)\}$			$\{(q_3, \epsilon)\}$					
q_3				$\{(q_3, \epsilon)\}$			$\{(q_4, \epsilon)\}$		
q_4									

Dunque, il suo diagramma di stato è il seguente:

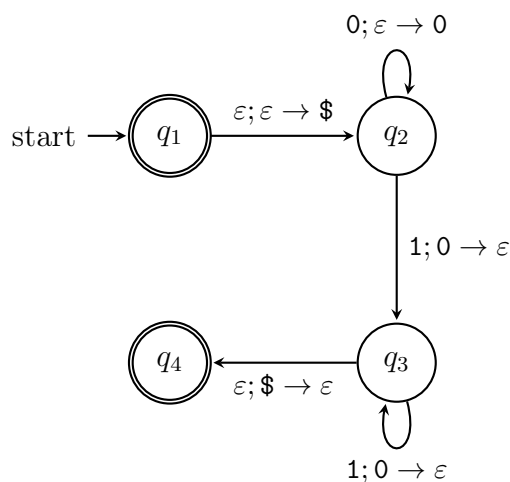


Figura 2.1: Il PDA M .

La notazione $a; b \rightarrow c$ presente sugli archi di questo diagramma sta ad indicare che se viene letto il simbolo a dall'input, M può sostituire b , se in cima al suo stack (attraverso un'operazione di *pop*) con c (mediante un'operazione di *push*). Si noti inoltre che ognuno dei simboli può essere ϵ , e dunque

- $\epsilon; b \rightarrow c$ indica che il ramo viene eseguito senza attendere alcun simbolo di input (si noti l'Esempio 1.3.1.1 per il non determinismo)
- $a; \epsilon \rightarrow c$ indica che, alla lettura di a , viene effettuato solamente il *push* di c nello stack
- $a; b \rightarrow \epsilon$ indica che, alla lettura di a , viene effettuato solamente il *pop* di b dallo stack

Definizione 2.3.1.2: Stringhe accettate (PDA)

Sia $M = (Q, \Sigma, \Gamma, \delta, q_0, F)$ un PDA, e sia $w = w_1 \cdots w_n$ una stringa tale per cui $\forall i \in [1, n] \quad w_i \in \Sigma_\varepsilon$; allora, M **accetta** w se esistono una sequenza di stati $r_0, \dots, r_n \in Q$ e una sequenza di stringhe $s_0, \dots, s_n \in \Gamma^*$ tali per cui

- $r_0 = q_0$
- $s_0 = \varepsilon$, ovvero, lo stack è inizialmente vuoto
- $\forall i \in [0, n-1] \quad \exists a, b \in \Gamma_\varepsilon, t \in \Gamma^* \mid (r_{i+1}, b) \in \delta(r_i, w_{i+1}, a) \wedge \begin{cases} s_i = at \\ s_{i+1} = bt \end{cases}$,
ovvero, M si muove correttamente in base allo stato, al simbolo nello stack, ed al prossimo simbolo di input; si noti che le stringhe s_0, \dots, s_n rappresentano di fatto il contenuto dello stack che M ha su un ramo accettante della computazione, infatti $s_i = at$ diventa $s_{i+1} = bt$ con l'iterazione successiva, dunque a è stato sostituito con b in cima alla pila
- $r_n \in F$

Osservazione 2.3.1.1: Stack vuoto

Un PDA non è in grado di controllare se il suo stack è vuoto, ma è possibile ottenere questo effetto come segue: si prenda in esame il PDA M dell'[Esempio 2.3.1.1](#); è importante notare che esso utilizza il simbolo $\$$ per capire se lo stack è vuoto o meno, poiché viene inserito sin dall'inizio (si osservi l'etichetta $\varepsilon; \varepsilon \rightarrow \$$ sull'arco (q_1, q_2)), e dunque se viene letto $\$$ in cima allo stack, M sa che quello è l'unico elemento contenuto al suo interno, e lo stack è di fatto vuoto.

Osservazione 2.3.1.2: Fine dell'input

Un PDA non è in grado di controllare se è stata raggiunta la fine della stringa di input, ma è possibile ottenere questo effetto come segue: si prenda in esame il PDA M dell'[Esempio 2.3.1.1](#); è importante notare che esso può sapere se è stata raggiunta la fine della stringa di input, poiché lo stato accettante q_4 può essere eventualmente raggiunto solamente alla lettura di ε e nel momento in cui è possibile rimuovere $\$$ dallo stack (si noti l'[Osservazione 2.3.1.1](#)).

Osservazione 2.3.1.3: Linguaggi non regolari e PDA

Si consideri il PDA dell'Esempio 2.3.1.1; esso opera come segue:

- viene posto \$ all'interno dello stack;
- fintanto che viene letto 0, viene *pushato* 0 nello stack;
- non appena viene letto 1, e fintanto che viene letto 1, viene rimosso 0 dallo stack;
- se all'interno dello stack è presente \$, allora l'automa accetta.

Dunque, di fatto, il PDA sta *contando* il numero di 0 e di 1 presenti all'interno della stringa, poiché la computazione avanza solamente se, per ognuno degli 1 letti, è possibile rimuovere uno 0 dallo stack. Infine, lo stato q_1 è accettante per poter riconoscere la stringa ε . Allora, l'automa riconosce il linguaggio

$$L = \{0^n 1^n \mid n \in \mathbb{N}\}$$

il quale non è regolare per l'Esempio 1.8.1.2. In simboli, si ha che

$$\text{REG} \subsetneq \mathcal{L}(\text{PDA})$$

Esempio 2.3.1.2 (PDA). Si consideri il seguente PDA:

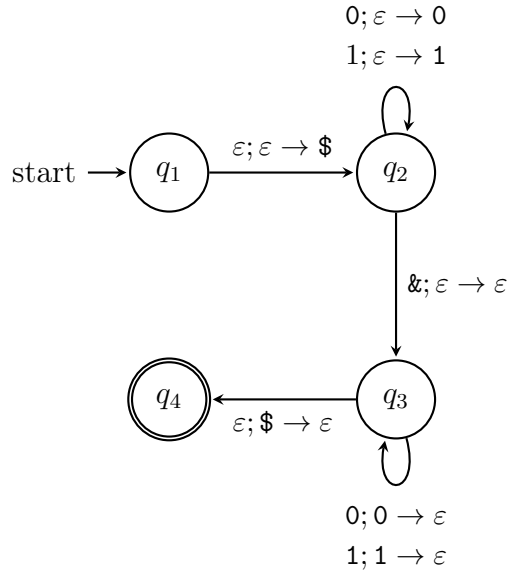


Figura 2.2: Un PDA.

Esso è in grado di riconoscere il linguaggio

$$L = \{w \in \{0, 1\}^* \mid w \& w^R\}$$

poiché:

- pone \$ all'interno dello stack, per sapere quando quest'ultimo diventa vuoto;
- pone all'interno dello stack qualsiasi simbolo diverso da &;
- una volta letto il simbolo &, l'automa cambia stato lasciando lo stack intatto, ed inizia a rimuovere da questo ogni simbolo che viene letto, *solo se coincide con il simbolo posto sulla sua cima*; di fatto, questa tecnica controlla che i simboli che vengono letti dopo & siano l'opposto di come sono stati letti inizialmente;
- se trova \$ nello stack — e dunque quest'ultimo è vuoto — accetta.

2.3.2 Equivalenze

Definizione 2.3.2.1: Inserimento di stringhe nello stack

Sia $P = (Q, \Sigma, \Gamma, \delta, q_0, F)$ un PDA; è possibile introdurre una notazione per effettuare l'**inserimento di una stringa nello stack** di P , come segue: dati due stati $p, q \in Q$, per inserire una stringa $u_1 \dots u_k \in \Gamma^*$ all'interno dello stack di P , devono esistere stati $r_1, \dots, r_{k-1} \in Q$ tali per cui

$$\begin{aligned}\delta(p, a, u_1 \dots u_k) &\ni (r_1, u_k) \\ \delta(r_1, \varepsilon, \varepsilon) &= \{(r_2, u_{k-1})\} \\ &\vdots \\ \delta(r_{k-1}, \varepsilon, \varepsilon) &= \{(q, u_1)\}\end{aligned}$$

In simboli, per indicare l'inserimento di una stringa all'interno dello stack di un PDA verrà utilizzato il seguente simbolismo

$$a; s \rightarrow xyz$$

per certi $a \in \Sigma_\varepsilon, s \in \Gamma, xyz \in \Gamma^*$.

Esempio 2.3.2.1 (Inserimenti di stringhe nello stack). Si consideri il seguente PDA:

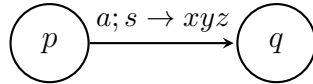


Figura 2.3: Un PDA che inserisce una stringa nello stack.

per certi $a \in \Sigma_\varepsilon, s \in \Gamma, xyz \in \Gamma^*$; dunque, con la notazione presente sull'arco (p, q) verrà sottointesa la seguente successione di stati — per certi stati $r_1, r_2 \in Q$:

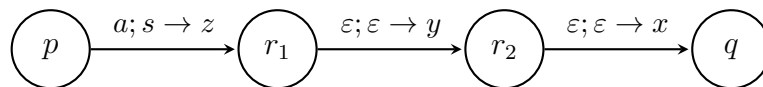


Figura 2.4: La formalizzazione del PDA precedente.

Metodo 2.3.2.1: PDA di CFG

Sia $G = (V, \Sigma, R, S)$ una CFG, e sia E l'insieme di stati tale da permettere di utilizzare la notazione descritta nella [Definizione 2.3.2.1](#); per costruire un PDA in grado di riconoscere $L(G)$, che avrà la forma

$$P := (\{q_0, q', q\} \cup E, \Sigma, V \cup \Sigma, \delta, q_0, \{q\})$$

si definiscono i seguenti:

- $\delta(q_0, \varepsilon, \varepsilon) := \{(q', S\$)\}$, ovvero, viene inserito $\$$ come marcatore nello stack di P (si noti l'[Osservazione 2.3.1.1](#)), e successivamente la stringa iniziale S di G ;
- $\forall A \in V \quad \delta(q', \varepsilon, A) := \{(q', w) \mid A \rightarrow w \in R, w \in (V \cup \Sigma)^*\}$ poiché, se viene incontrata una variabile A sulla cima dello stack di P , viene scelta una delle regole di G in grado di sostituire A (si noti che la sostituzione verrà effettuata non deterministicamente);
- $\forall a \in \Sigma \quad \delta(q', a, a) := \{(q', \varepsilon)\}$ per rimuovere i terminali dallo stack, evitando che possano essere ulteriormente rimpiazzati;
- $\delta(q', \varepsilon, \$) := \{(q, \varepsilon)\}$ per sfruttare il marcatore $\$$.

Il seguente è un diagramma che raffigura P (a meno degli stati in E):

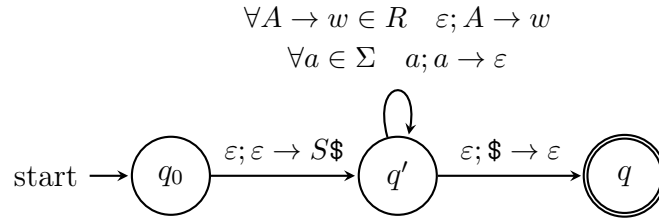
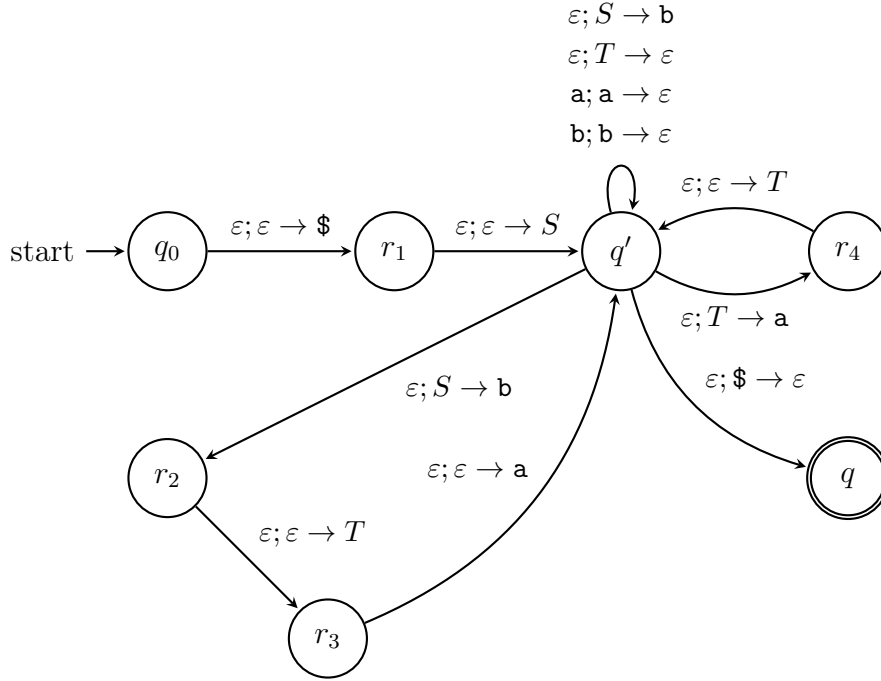


Figura 2.5: Il PDA P appena costruito.

Esempio 2.3.2.2 (PDA di CFG). Si consideri la seguente CFG:

$$G: \begin{array}{l} S \rightarrow aTb \mid b \\ T \rightarrow Ta \mid \varepsilon \end{array}$$

Attraverso il [Metodo 2.3.2.1](#), si ottiene il seguente PDA, in grado di riconoscere $L(G)$:

Figura 2.6: Un PDA in grado di riconoscere $L(G)$.**Teorema 2.3.2.1: CFL e PDA**

Un linguaggio è context-free se e solo se esiste un PDA che lo riconosce; in simboli

$$\text{CFL} = \mathcal{L}(\text{PDA})$$

Dimostrazione.

Prima implicazione. Sia L un CFL, e dunque per definizione esiste una CFG G tale che $L(G) = L$; allora, utilizzando il [Metodo 2.3.2.1](#), è possibile trasformare G in un PDA ad essa equivalente — ovvero, tale da riconoscere $L(G)$ — dunque segue la tesi.

Seconda implicazione. Sia L un linguaggio riconosciuto da un PDA definito come $P = (Q, \Sigma, \Gamma, \delta, q_0, F)$ — dunque $L = L(P)$ — e si consideri il seguente PDA

$$P' := (Q', \Sigma, \Gamma, \delta', q_0, F')$$

tale che:

- ogni transizione di stati possa effettuare solamente un'operazione di *push* o *pop*, *non simultaneamente*, dunque avente δ' definita come segue:

$$\forall p, q \in Q, a \in \Sigma, b, c \in \Gamma^* \quad (q, c) \in \delta(p, a, b) \implies \exists r \in Q' \mid \begin{cases} (r, \epsilon) \in \delta'(p, a, b) \\ (q, c) \in \delta'(r, \epsilon, \epsilon) \end{cases}$$

al fine di trasformare la seguente transizione



come segue:



e sia $Q_{\delta'}$ l'insieme di stati tali da permettere la non simultaneità appena descritta

- $Q' := Q \cup Q_{\delta'} \cup \{q_{\text{accept}}\}$
- $F' := \{q_{\text{accept}}\}$ sia costituito da un solo stato accettante, e dunque

$$\forall q \in F' \quad (q_{\text{accept}}, \varepsilon) \in \delta'(q, \varepsilon, \varepsilon)$$

in modo da far terminare la computazione di ogni stringa accettata in q_{accept}

- venga svuotato lo stack prima di accettare qualsiasi stringa, di conseguenza

$$\forall q \in F, a \in \Sigma \quad (q, \varepsilon) \in \delta'(q, \varepsilon, a)$$

Si noti che, per costruzione di P' , si ha che $L(P) = L(P')$, poiché P' è una versione di P in cui le transizioni possono effettuare solamente un'operazione per volta sullo stack, è presente un solo stato accettante e viene svuotato lo stack prima di accettare le stringhe.

Si consideri ora la CFG $G = (V, \Sigma, R, S)$ definita come segue:

- $V := \{A_{p,q} \mid p, q \in Q'\}$
- $S := A_{q_0, q_{\text{accept}}}$
- l'insieme di regole R è composto dall'unione dei seguenti:

$$\begin{aligned} R := & \{A_{p,p} \rightarrow \varepsilon \mid p \in Q'\} \cup \\ & \cup \{A_{p,q} \rightarrow aA_{r,s}b \mid p, q, r, s \in Q', a, b \in \Sigma_\varepsilon, \exists u \in \Gamma : (r, u) \in \delta'(p, a, \varepsilon), (q, \varepsilon) \in \delta(s, b, u)\} \cup \\ & \cup \{A_{p,q} \rightarrow A_{p,r}A_{r,q} \in R \mid p, q, r \in Q'\} \end{aligned}$$

Dunque, il linguaggio di G è il seguente:

$$L(G) = \{w \in \Sigma^* \mid A_{q_0, q_{\text{accept}}} \xRightarrow{*} w\}$$

La dimostrazione procederà ora provando che $L(G) = L(P')$. A tal scopo, sarebbe dunque sufficiente dimostrare che una stringa è accettata da P' se e solo se è derivabile, attraverso le regole di G , a partire da $A_{q_0, q_{\text{accept}}}$, ma è possibile generalizzare questa proposizione dimostrando che, per ogni coppia di stati $p, q \in Q'$, $A_{p,q}$ è in grado di derivare una stringa w se e solo se w porta P' dallo stato p — avendo lo stack vuoto — allo stato q — avendo ancora lo stack vuoto.

Prima implicazione. La dimostrazione procede per induzione sul numero di passaggi nella derivazione di w , a partire da $A_{p,q}$, attraverso le regole in R .

Caso base. Se la derivazione è composta solamente da 1 passaggio, allora è stata utilizzata una regola della forma $A_{p,q} \rightarrow u_1 \dots u_k$ dove $u_1, \dots, u_k \in \Sigma^*$, dunque esclusivamente terminali. Si noti però che le uniche regole della grammatica che vedono solamente terminali sulla loro destra sono le regole del primo insieme di definizione di R , ovvero forma $A_{p,p} \rightarrow \varepsilon$ dunque, se P' si trova nello stato p , avendo lo stack vuoto, sicuramente ε è in grado di portare P' in p stesso, mantenendo vuoto lo stack.

Ipotesi induttiva forte. Dati due stati $p, q \in Q'$, se $A_{p,q}$ è in grado di derivare una stringa w , applicando al più k passaggi — con $k \geq 1$ — allora w porta P' dallo stato p — avendo lo stack vuoto — allo stato q — avendo ancora lo stack vuoto.

Passo induttivo. È necessario dimostrare che l'ipotesi induttiva è ancora verificata per derivazioni costituite da $k + 1$ passaggi. Si consideri dunque una stringa w tale che $A_{p,q} \xRightarrow{*} w$ con $k + 1$ passaggi; per costruzione di R , il primo passaggio può essere $A_{p,q} \Rightarrow aA_{r,s}b \xRightarrow{*} x$ per certi stati $r, s \in Q'$ e terminali $a, b \in \Sigma_\varepsilon$, oppure $A_{p,q} \Rightarrow A_{p,r}A_{r,q} \xRightarrow{*} x$ per un qualche stato $r \in Q'$, e dunque:

- nel primo caso, sia y la sottostringa di w generata da $A_{r,s}$, dunque $A_{r,s} \xRightarrow{*} y \implies w = ayb$, e si noti che y è stata derivata con k passaggi; allora, per ipotesi induttiva forte, y porta P' dallo stato r — avendo lo stack vuoto — allo stato s — avendo ancora lo stack vuoto; inoltre, per costruzione di R , si ha che

$$A_{p,q} \rightarrow aA_{r,s}b \in R \iff \exists u \in \Gamma \mid \begin{cases} (r, u) \in \delta'(p, a, \varepsilon) \\ (q, \varepsilon) \in \delta'(s, b, u) \end{cases}$$

e dunque, assumendo che P' abbia lo stack vuoto:

- se si trova nello stato p , leggendo a , P' può andare nello stato r , inserendo u nello stack;
- se si trova nello stato r , leggendo y , P' può andare nello stato s , lasciando lo stack invariato — per ipotesi induttiva forte;
- se si trova nello stato s , leggendo b , P' può andare nello stato q , rimuovendo u dallo stack.

allora w è in grado di portare P' dallo stato p allo stato q , lasciando invariato lo stack.

- differentemente, nel secondo caso, siano y e z sottostringhe di w generate rispettivamente da $A_{p,r}$ ed $A_{r,q}$ dunque $\begin{cases} A_{p,r} \xRightarrow{*} y \\ A_{r,q} \xRightarrow{*} z \end{cases} \implies w = yz$, e si noti che y e z sono state derivate con $k + 1$ passaggi; allora, per ipotesi induttiva forte, y e z portano, rispettivamente, P' dagli stati p

ed r — avendo lo stack vuoto — agli stati r e q — avendo ancora lo stack vuoto; allora, w è in grado di portare P' dallo stato p allo stato q , lasciando invariato lo stack.

Seconda implicazione. La dimostrazione procede per induzione sul numero di passaggi — aventi lo stack vuoto — nella computazione di w , da parte di P' , tra gli stati p e q .

Caso base.

Ipoesi induttiva forte.

Passo induttivo.

In particolare, $A_{q_0, q_{\text{accept}}}$ è in grado di derivare una stringa w se e solo se w porta P' dallo stato q_0 — avendo lo stack vuoto — allo stato q_{accept} — avendo ancora lo stack vuoto — o equivalentemente, P' accetta w . Allora, segue che

$$A_{q_0, q_{\text{accept}}} \xRightarrow{*} w \iff w \in L(G) \iff w \in L(P')$$

e dunque, poiché $L(G) = L(P') = L(P) = L$, segue la tesi. □

2.4 Linguaggi non context-free

2.4.1 Pumping lemma

Lemma 2.4.1.1: Pumping lemma (CFL)

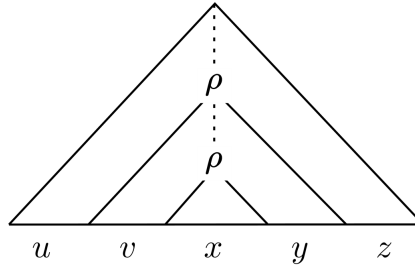
Sia A un CFL; allora, esiste un $p \in \mathbb{N}$, detto **lunghezza del pumping**, tale che per ogni stringa $s \in A$ tale per cui $|s| \geq p$, esistono 5 stringhe $u, v, x, y, z \mid s = uvxyz$ soddisfacenti le seguenti condizioni:

- $\forall i \geq 0 \quad uv^i xy^i z \in A$
- $|vy| > 0$ (o, equivalentemente, $v \neq \varepsilon \vee y \neq \varepsilon$)
- $|vxy| \leq p$

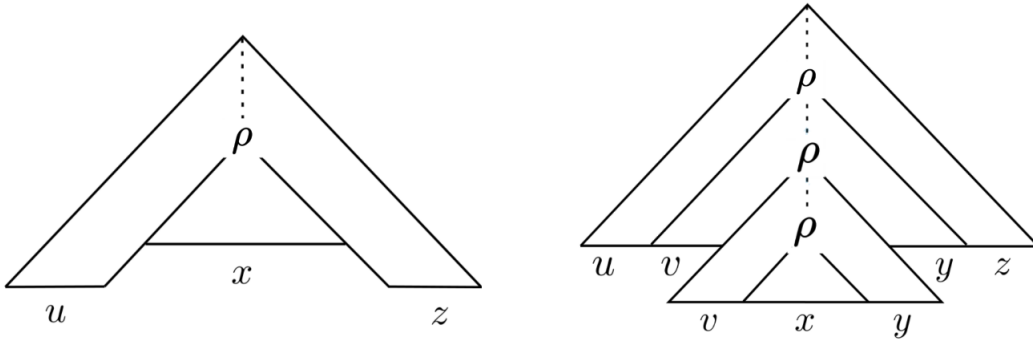
Dimostrazione. Poiché A è un CFL, per definizione esiste una CFG $G = (V, \Sigma, R, S)$ che lo genera; sia b il massimo numero di simboli nel lato destro di una regola in R , e dunque all'interno di ogni albero sintattico che è possibile costruire attraverso G non sarà possibile avere un nodo con più di b figli. Dunque, partendo da S , per la prima sostituzione effettuata, si avranno al più b foglie all'interno dell'albero; per la seconda, se ne avranno al più b^2 , ed in generale, un albero di altezza al più h avrà al più b^h foglie, dunque la stringa generata avrà lunghezza al più b^h . Di conseguenza, una stringa di lunghezza pari almeno a $b^h + 1$ deve essere stata generata da un albero avente altezza pari almeno ad $h + 1$.

Sia $p := b^{|V|+1}$, e sia $s \in A$ tale che $|s| \geq p$; si noti che $b^{|V|+1} \geq b^{|V|} + 1$, e poiché s ha lunghezza pari almeno a p , allora — per quanto detto precedentemente — s deve essere stata generata attraverso un albero sintattico avente altezza pari almeno a $|V| + 1$. Sia allora τ l'albero sintattico con minor numero di nodi tale da generare s ; poiché $h_\tau \geq |V| + 1$, il cammino \mathcal{C} più lungo dalla radice ad una foglia in τ ha lunghezza almeno $|V| + 1$, avente dunque almeno $|V| + 2$ nodi, dove il primo è la radice — ovvero S , una variabile — l'ultimo è la foglia — necessariamente un terminale — e tutti gli altri nodi nel mezzo sono variabili; dunque il cammino è composto da almeno $|V| + 1$ variabili. Si noti che $G = (V, \Sigma, R, S)$, dunque il numero di variabili di G è $|V|$; allora, per il [Principio 1.8.1.1](#), all'interno del cammino preso in considerazione, deve esistere TODO ALMENO? una variabile ripetuta più di 1 volta, e sia questa ρ . Si assuma che $\rho \neq S$. TODO SE $\rho = S$???

Siano u, v, x, y, z scelte come segue:



Poiché ρ è ripetuta, è possibile sostituire il suo sottoalbero nei punti in cui essa si ripete, ottenendo ad esempio i due seguenti alberi sintattici ancora validi:



Si noti che:

- sostituzione di sinistra genera la stringa uxz , che dunque è in A , mentre la sostituzione di destra genera la stringa uv^2xy^2z , che dunque è in A anch'essa, e si noti che sarebbe possibile continuare a sostituire il sottoalbero di ρ ripetutamente nella rappresentazione di destra; dunque, questo dimostra la prima condizione del lemma, ovvero $\forall i \geq 0 \quad uv^i xy^i z \in A$;
- se $v = y = \varepsilon$, allora TODO NON HO CAPITO

- l'occorrenza più in alto di ρ genera la stringa vxy , e ρ è stata scelta diversa da S , dunque ρ deve necessariamente trovarsi tra le $|V| + 1$ variabili più in basso di \mathcal{C} , il quale era stato scelto tale che fosse il più lungo cammino dalla radice ad una foglia; allora, il sottoalbero di ρ è alto al più $|V| + 1$ TODO PERCHÉ NON $|V|??$, dunque per quanto detto precedentemente, tale albero può generare una stringa di lunghezza al più $b^{|V|} + 1 =: p$, e dunque $|vxy| \leq p$.

Allora, sono soddisfatte tutte le condizioni della tesi. \square

Esempio 2.4.1.1 (Pumping lemma in CFL). Si consideri il linguaggio

$$L := \{0^n 1^n 2^n \mid n \in \mathbb{N}\}$$

e per assurdo, sia $L \in \text{CFL}$, e dunque per esso vale il [Lemma 2.4.1.1](#). Allora, sia $p \in \mathbb{N}$ la lunghezza del pumping di L , e si consideri la seguente stringa

$$w := 0^p 1^p 2^p \implies |w| = 3p > p$$

avente lunghezza sicuramente maggiore di p . Siano inoltre u, v, x, y, z tali da soddisfare il pumping lemma, ed in particolare $|vxy| \leq p$, ma poiché $w := 0^p 1^p 2^p = uv^i xy^i z$ per ogni $i \in \mathbb{N}$, allora si può verificare solo uno dei seguenti:

- vxy contiene solamente 0, dunque $v^0 xy^0$ contiene solamente 0
- vxy contiene solamente 1, dunque $v^0 xy^0$ contiene solamente 1
- vxy contiene solamente 2, dunque $v^0 xy^0$ contiene solamente 2
- vxy contiene soltanto 0 e 1, dunque $v^0 xy^0$ può contenere solo 0, solo 1 o sia 0 che 1
- vxy contiene soltanto 1 e 2, dunque $v^0 xy^0$ può contenere solo 1, solo 2, o sia 1 che 2

e dunque, indipendentemente dalla scelta di v, x, y TODO DA FINIRE IL FINALE NON HA SENSO

Corollario 2.4.1.1: CFL non è chiuso sull'intersezione

Siano L_1 ed L_2 due CFL; allora $L_1 \cap L_2$ non è necessariamente un CFL.

Dimostrazione. Siano G_1 e G_2 le due seguenti CFG:

$$\begin{array}{ll} S \rightarrow TU & S \rightarrow UT \\ G_1 : T \rightarrow 0T1 \mid \varepsilon & G_2 : T \rightarrow 1T2 \mid \varepsilon \\ U \rightarrow 2U \mid \varepsilon & U \rightarrow 0U \mid \varepsilon \end{array}$$

Si noti che la prima grammatica, partendo da S , compone stringhe costituite da TU , dove T è solo in grado di diventare una combinazione di $0T1$ o terminare — e dunque produce stringhe della forma $0^n 1^n$ — mentre U è solo in grado di diventare una combinazione di $2U$ o terminare — e dunque produce stringhe della forma 2^k ; allora, il suo linguaggio è

$$L(G_1) = \{0^n 1^n 2^k \mid n, k \in \mathbb{N}\}$$

Per ragionamento analogo, è possibile verificare che le stringhe che genera G_2 appartengono al linguaggio

$$L(G_2) = \{0^k 1^n 2^n \mid n, k \in \mathbb{N}\}$$

Infine, si noti che

$$L(G_1) \cap L(G_2) = \{0^n 1^n 2^n\}$$

poiché il primo linguaggio contiene le stringhe aventi stesso numero di 0 e di 1, ma arbitrario numero di 2, mentre il secondo contiene le stringhe aventi stesso numero di 1 e 2, ma arbitrario numero di 0, dunque la loro intersezione deve necessariamente essere il linguaggio composto da stringhe aventi stesso numero di 0, 1 e 2. Si noti però che, come dimostrato nell'Esempio 2.4.1.1, $\{0^n 1^n 2^n\} \notin \text{CFL}$. Allora, poiché $L(G_1), L(G_2) \in \text{CFL}$ per loro stessa definizione, segue la tesi. \square

Esempio 2.4.1.2 (Pumping lemma in CFL). Si consideri il linguaggio

$$L := \{ww \mid w \in \{0, 1\}^*\}$$

e per assurdo, sia $L \in \text{CFL}$, e dunque per esso vale il Lemma 2.4.1.1. Allora, sia $p \in \mathbb{N}$ la lunghezza del pumping di L , e si consideri la seguente stringa

$$w := 0^p 1^p 0^p 1^p \implies |w| = 4p > p$$

avente sicuramente lunghezza maggiore di p . Siano inoltre u, v, x, y, z tali da soddisfare il pumping lemma, ed in particolare $|vxy| \leq p$, ma poiché $w := 0^p 1^p 0^p 1^p = uv^i xy^i z$ per ogni $i \in \mathbb{N}$, allora si può verificare solo uno dei seguenti:

- vxy contiene solamente 0 o solamente 1, e dunque assumendo che ad esempio si trovi nella prima porzione di 0 di w , per ogni $i > 1$ si ha che la stringa diventa della forma

$$w := 0^{p+k} 101$$

per qualche k , e dunque non può essere della forma $w'w'$ con $w' \in \{0, 1\}^*$;

- vxy contiene sia 0 che 1, e dunque vxy si trova a cavallo tra due sequenze di 0 ed 1; allora, assumendo che ad esempio si trovi nella prima regione di w che comprende sia 0 che 1, per ogni $i > 1$, si ha che la stringa diventa della forma

$$w := 0^{p+j} 1^{p+h} 0^p 1^p$$

per qualche j ed h , e dunque non può essere della forma $w'w'$ con $w' \in \{0, 1\}^*$.

Allora, in nessuno dei casi possibili $w := uv^i xy^i z \in L \nmid$.

Corollario 2.4.1.2: CFL non è chiuso sul complemento

Sia L un CFL; allora $\neg L$ non è necessariamente un CFL.

Dimostrazione. Sia L il linguaggio del Esempio 2.4.1.2, e si consideri il suo complemento

$$L := \{ww \mid w \in \{0, 1\}^*\} \iff \neg L = \{0, 1\}^* - \{ww \mid w \in \{0, 1\}^*\}$$

e si consideri la seguente grammatica G :

$$\begin{aligned} S &\rightarrow A \mid B \mid AB \mid BA \\ G: \quad A &\rightarrow 0 \mid 0A0 \mid 0A1 \mid 1A0 \mid 1A1 \\ B &\rightarrow 1 \mid 0B0 \mid 0B1 \mid 1B0 \mid 1B1 \end{aligned}$$

e si noti che le stringhe derivate da A sono tutte le stringhe aventi lunghezza dispari ed uno 0 al centro, mentre le stringhe derivate da B sono tutte le stringhe aventi lunghezza dispari ed un 1 al centro.

Dunque, si ha che:

- presa una stringa $x \in \neg L$, e ponendo $n := |x|$, se n è dispari allora
 - se x ha 0 al centro, allora $A \xrightarrow{*} x$
 - se x ha un 1 al centro, allora $B \xrightarrow{*} x$

diversamente, se n è pari, allora sia i tale che $x_i \neq x_{\frac{n}{2}+i}$ — si noti che i deve necessariamente esistere, poiché se non esistesse allora vorrebbe dire che $\forall i \in [1, n] \quad x_i = x_{\frac{n}{2}+i} \iff \exists w \in \{0, 1\}^* \mid x = ww \iff x \in L$; siano inoltre

$$\begin{aligned} u &:= x_1 \dots x_{2i-1} \implies |u| = 2i - 1 - 1 + 1 \\ v &:= x_{2i} \dots x_n \implies |v| = n - 2i + 1 \end{aligned}$$

due sottostringhe di $x = uv$ di lunghezza dispari — si noti che $n - 2i + 1$ è dispari poiché n è pari in ipotesi; allora, i due caratteri centrali di u e v saranno rispettivamente

$$\begin{aligned} m(u) &= x_{\frac{2i-1+1}{2}} = x_i \\ m(v) &= x_{\frac{n+2i}{2}} = x_{\frac{n}{2}+1} \end{aligned}$$

allora, poiché $x_i \neq x_{\frac{n}{2}+1}$, e questi sono proprio i centri di due sottostringhe di x , aventi lunghezza dispari, si ha che x può essere generata attraverso le regole di G , poiché in essa sono presenti le regole $S \rightarrow AB \mid BA$; questo dimostra che $x \in \neg L \implies S \xrightarrow{*} x$;

- presa una stringa x tale che $S \xrightarrow{*} x$, e ponendo $n := |x|$, se n è dispari allora sicuramente $x \notin L \iff x \in \neg L$; diversamente, se n è pari, allora la stringa è stata generata attraverso una delle regole $S \rightarrow AB \mid BA$ necessariamente, e dunque — senza perdita di generalità — assumendo che x sia stata ottenuta attraverso la regola $S \rightarrow AB$, devono esistere $u, v \in \{0, 1\}^* \mid x = uv \wedge A \xrightarrow{*} u \wedge B \xrightarrow{*} v$, dove u e v sono due stringhe aventi lunghezza dispari; siano allora $l := |u| \implies |v| = n - l$, ottenendo che

$$m(u) = x_{\frac{l+1}{2}} = u_{\frac{l+1}{2}} = 0 \neq 1 = v_{\frac{n-l+1}{2}} = x_{\frac{n-l+1}{2}+l} = x_{\frac{n+l+1}{2}} = m(v)$$

allora, poiché $m(u)$ ed $m(v)$ devono necessariamente trovarsi l'uno nella prima metà di x , l'altro nella seconda metà (indipendentemente dalla scelta di u e v) si ha che la stringa x è necessariamente costituita da due stringhe $w, w' \in \{0, 1\}^* \mid w \neq w'$, poiché differiscono per almeno un carattere, ovvero $m(u) \neq m(v)$, di conseguenza $x \notin L \iff x \in \neg L$; questo dimostra che $S \xrightarrow{*} x \implies x \in \neg L$.

Dunque, poiché $S \xRightarrow{*} x \iff x \in \neg L$, si ha che

$$L(G) := \{w \in \{0, 1\}^* \mid S \xRightarrow{*} w\} = \neg L \iff \neg L \in \text{CFL}$$

e dunque segue la tesi. □

Osservazione 2.4.1.1: Condizioni del pumping lemma (CFL)

TODO