

Fondamenti di informatica

UniVR - Dipartimento di Informatica

Fabio Irimie

1° Semestre 2025/2026

Indice

1	Introduzione	2
1.1	Cos'è l'informatica?	2
1.2	Origini dell'informatica	2
1.2.1	Calcolabilità	2
1.3	Nozioni di base	2
1.3.1	Basi di logica	2
1.3.2	Nozioni sugli insiemi	3
1.3.3	Nozioni sulle relazioni	4
1.3.4	Nozioni sulle funzioni	4
2	Funzioni calcolabili	5
2.1	Quante funzioni numerabili ci sono?	5
2.2	Funzioni vs Insiemi	7
3	Linguaggi regolari	8
3.1	Linguaggi formali	10
3.2	Linguaggi regolari (automi a stati finiti, DFA)	11
3.2.1	Come si dimostra che un linguaggio è regolare?	12
3.3	Automi a stati finiti non deterministici (NFA)	17
3.3.1	Linguaggio riconosciuto da un NFA	18
3.3.2	Conversione da NFA a DFA	20
3.4	Automi non deterministici con ϵ -transizioni (ϵ -NFA)	21
3.4.1	ϵ -closure	22
4	Espressioni regolari	23
4.1	Proprietà dei linguaggi regolari	25
4.1.1	Proprietà di chiusura	25
4.1.2	Proprietà di decidibilità	26
4.1.3	Esistenza dell'automa minimo	26
4.1.4	Condizione necessaria perchè un linguaggio sia regolare	31
4.1.5	Dimostrazione che un linguaggio non è regolare	33
5	Linguaggi context free	36
5.1	Grammatiche context free	36
5.1.1	Dimostrazione canonica che un linguaggio è context free	38
5.1.2	Dimostrazione alternativa	42
5.2	Alberi di derivazione	42
5.3	Ambiguità delle grammatiche	43
5.4	Forma normale	44
5.5	Minimizzazione delle grammatiche context free	44
5.5.1	Eliminazione dei simboli inutili	45
5.5.2	Eliminazione delle produzioni unitarie	46
5.5.3	Eliminazione delle ϵ -produzioni	47

1 Introduzione

1.1 Cos'è l'informatica?

È una scienza che studia la calcolabilità, cioè cerca di capire che problemi si possono risolvere con un programma. Nasce dall'unione di matematica, ingegneria e logica. Il computer è solo uno strumento, mentre la matematica è il linguaggio con cui si creano algoritmi che permettono di risolvere i problemi.

1.2 Origini dell'informatica

Hilbert, nel 1900, si pose l'obiettivo di formalizzare tutta la matematica con un insieme finito e non contraddittorio di assiomi. Nel 1931, invece, Gödel dimostrò che l'informatica non potrà mai rappresentare tutta la matematica, perché ci saranno sempre proposizioni vere ma non dimostrabili tramite il calcolo. Ci si iniziò a chiedere se esistessero modelli di calcolo meccanici in grado di risolvere tutti i problemi. Nel 1936, Turing propose la macchina di Turing, una **sol**a macchina programmabile in grado di risolvere tutti i problemi risolvibili:

$$\text{Int}(P, x) = \begin{cases} P(x) & \text{se } P(x) \text{ termina} \\ \uparrow & \text{se } P(x) \text{ non termina} \end{cases}$$

dove P è un programma e x è un input. La macchina di Turing è un modello teorico di calcolatore, che non esiste fisicamente, ma è in grado di simulare qualsiasi altro calcolatore. Da questo modello deriva la concezione di calcolabilità, cioè se un problema è intuitivamente calcolabile, allora esiste un programma in grado di risolverlo.

Altri modelli di calcolo che sono stati proposti sono:

- Lambda-calcolo
- Funzioni ricorsive
- Linguaggi di programmazione (Turing-completi)

Definizione utile 1.1. La Turing-completezza è la proprietà di un linguaggio di programmazione di essere in grado di simulare una macchina di Turing, cioè di poter risolvere qualsiasi problema risolvibile.

1.2.1 Calcolabilità

Un programma è calcolabile se termina, ma non è detto che termini in un tempo ragionevole. Non esistono algoritmi che possono dire se un programma termina o meno. Questo è un esempio di problema non calcolabile.

I problemi non calcolabili sono infinitamente più numerosi di quelli calcolabili

1.3 Nozioni di base

1.3.1 Basi di logica

Alcune nozioni di logica che ci serviranno in seguito:

- **Linguaggio del primo ordine:**

- Simboli relazionali (p, q, \dots)
- Simboli di funzione (f, g, \dots)
- Simboli di costante (c, d, \dots)

- **Simboli logici:**

- Parentesi ($()$) e virgola
- Insieme numerabile di variabili (v, x, \dots)
- Connettivi logici ($\neg, \wedge, \vee, \rightarrow, \leftrightarrow$)
- Quantificatori (\forall, \exists)

- **Termini:**

- Variabili
- Costanti
- f simbolo di funzione m -ario t_1, t_2, \dots, t_m termini, allora $f(t_1, t_2, \dots, t_m)$ è un termine.

- **Formula atomica:** p simbolo di relazione n -ario, t_1, t_2, \dots, t_n termini, allora $p(t_1, t_2, \dots, t_n)$ è una formula atomica.

- **Formula:**

- Formula atomica
- ϕ formula, allora $\neg\phi$ è una formula
- ϕ e ψ formule, allora $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\phi \rightarrow \psi)$, $(\phi \leftrightarrow \psi)$ sono formule.
- ϕ formula e v variabile, allora $\forall v.\phi$ e $\exists v.\phi$ sono formule.

1.3.2 Nozioni sugli insiemi

- $x \in A$ significa che x è un elemento dell'insieme A
- $\{x|P(x)\}$ si identifica insieme costituito dagli x che soddisfano la proprietà (o predicato) $P(x)$
- $A \subseteq B$ significa che A è un sottoinsieme di B se ogni elemento di A è anche in B
- $\mathcal{P}(S)$ denota l'insieme delle parti di S , ovvero l'insieme di tutti i sottoinsiemi di S ($\mathcal{P}(S) = \{X|X \subseteq S\}$)
- $A \setminus B = \{x|x \in A \wedge x \notin B\}$, $A \cup B = \{x|x \in A \vee x \in B\}$, $A \cap B = \{x|x \in A \wedge x \in B\}$
- $|A|$ denota la cardinalità di A , ovvero il numero di elementi in A .
- \bar{A} denota il complemento di A , ovvero $x \in \bar{A} \leftrightarrow x \notin A$

1.3.3 Nozioni sulle relazioni

- Prodotto cartesiano:

$$A_1 \times A_2 \times \cdots \times A_n = \{\langle a_1, a_2, \dots, a_n \rangle \mid a_1 \in A_1, \dots, a_n \in A_n\}$$

- Una **relazione** (binaria) è un sottoinsieme del prodotto cartesiano di (due) insiemi; dati A e B , $R \subseteq A \times B$ è una relazione su A e B
 - **Riflessiva**: $\forall a \in S$ si ha che aRa
 - **Simmetrica**: $\forall a, b \in S$ se aRb allora bRa
 - **Antisimmetrica**: $\forall a, b \in S$ se aRb e bRa allora $a = b$
 - **Transitiva**: $\forall a, b, c \in S$ se aRb e bRc allora aRc
- Per ogni relazione $R \subseteq S \times S$ la chiusura transitiva di R è il più piccolo insieme R^* tale che $\langle a, b \rangle \in R \wedge \langle b, c \rangle \in R \rightarrow \langle a, c \rangle \in R^*$
- Una relazione è detta **totale** su S se $\forall a, b \in S$ si ha che $aRb \vee bRa$
- Una relazione R di *di equivalenza* è una relazione binaria riflessiva, simmetrica e transitiva.
- Una relazione binaria $R \subseteq S \times S$ è un **pre-ordine** se è riflessiva e transitiva.
- R è un ordine parziale se è un pre-ordine antisimmetrico.
- $x \in S$ è **minimale** rispetto a R se $\forall y \in S. y \not R x$ (ovvero $\neg(yRx)$)
- $x \in S$ è **minimo** rispetto a R se $\forall y \in S. xRy$
- $x \in S$ è **massimale** rispetto a R se $\forall y \in S. x \not R y$ (ovvero $\neg(xRy)$)
- $x \in S$ è **massimo** rispetto a R se $\forall y \in S. yRx$

1.3.4 Nozioni sulle funzioni

- Una relazione f è una **funzione** se $\forall a \in A$ esiste uno ed un solo $b \in B$ tale che $(a, b) \in f$
- A dominio e B codominio di f . Il range di f è l'insieme di tutti i valori che f può assumere.
- f è **iniettiva** se $\forall a_1, a_2 \in A$ se $a_1 \neq a_2$ allora $f(a_1) \neq f(a_2)$
- Se $f : A \mapsto B$ è sia iniettiva che suriettiva allora è **biiettiva** e quindi esiste $f^{-1} : B \mapsto A$

2 Funzioni calcolabili

Un insieme è una proprietà ed è rappresentato da una funzione che indica se un elemento appartiene o meno all'insieme. I problemi da risolvere (in questo corso) hanno come soluzione una funzione sui naturali:

$$f : \mathbb{N} \rightarrow \mathbb{N}$$

Questo tipo di funzione è un **insieme** di associazioni input-output. Un esempio è la funzione quadrato:

$$f = \text{quadrato} = \{(0, 0), (1, 1), (2, 4), (3, 9), \dots\} = \{(n, n^2) | n \in \mathbb{N}\}$$

Quindi f è un insieme di coppie in $\mathbb{N} \subseteq \mathbb{N} \times \mathbb{N}$ la cui cardinalità è: $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$. Di conseguenza la funzione è un sottoinsieme di $\mathbb{N} \times \mathbb{N}$:

$$f \subseteq \mathbb{N} \times \mathbb{N} \quad f \in \mathcal{P}(\mathbb{N} \times \mathbb{N})$$

Dove $\mathcal{P}(\mathbb{N} \times \mathbb{N})$ è l'insieme delle parti di $\mathbb{N} \times \mathbb{N}$, cioè l'insieme di tutti i sottoinsiemi di $\mathbb{N} \times \mathbb{N}$.

Il numero di funzioni è:

$$f : \mathbb{N} \rightarrow \mathbb{N} = |\mathcal{P}(\mathbb{N} \times \mathbb{N})| = |\mathcal{P}(\mathbb{N})|$$

Esempio 2.1. Un esempio di insieme delle parti per l'insieme $A = \{1, 2, 3\}$ è:

$$\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

E le cardinalità sono:

$$\begin{array}{ccc} |A| = 3 & |\mathcal{P}(A)| = 8 = 2^3 & \\ & \downarrow & \\ |\mathbb{N}| = \omega & < \underbrace{|\mathcal{P}(\mathbb{N})| = 2^\omega}_{\text{Insieme delle funzioni, non numerabile}} & = |\mathbb{R}| \end{array}$$

Si ha quindi che **l'insieme delle funzioni non è numerabile**

Ci si chiede se queste funzioni sono tutte calcolabili:

Definizione 2.1. Una funzione **intuitivamente calcolabile** è una funzione descrivibile attraverso un algoritmo, cioè una sequenza finita di passi discreti elementari.

2.1 Quante funzioni numerabili ci sono?

Consideriamo Σ come un alfabeto finito, cioè una sequenza di simboli utilizzabili per scrivere un programma o algoritmo.

$$\Sigma = \{s_1, s_2, s_3, \dots, s_n\}$$

↓

Programma \subseteq sequenze di simboli in Σ

Con Σ^* si descrive l'insieme di tutte le sequenze finite di simboli in Σ , quindi l'insieme di tutti i possibili programmi è:

Programmi $\subseteq \Sigma^*$

Esempio 2.2. Se $\Sigma = \{a, b, c\}$ allora la sequenza di tutti i possibili simboli è:

$$\Sigma^* = \{\epsilon, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, \dots\}$$

dove ϵ è la stringa vuota.

La cardinalità di Σ^* è infinita numerabile (anche se Σ è finito).

$$|\Sigma^*| = |\mathbb{N}|$$

Si ha quindi che l'insieme dei programmi è numerabile:

$$|\text{Programmi in } \Sigma| \leq |\Sigma^*| = |\mathbb{N}|$$

e questo implica che l'insieme delle **funzioni calcolabili è numerabile**

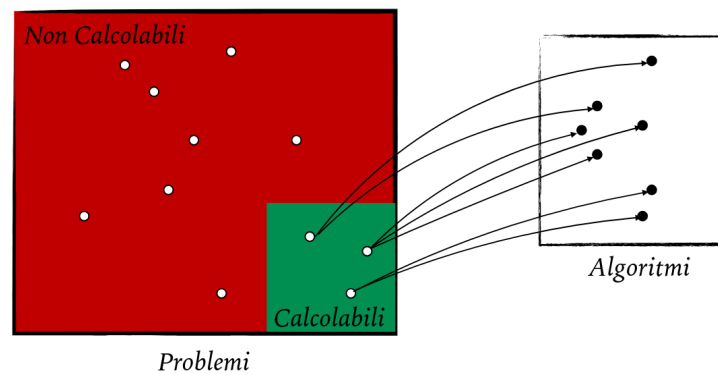


Figura 1: Cardinalità delle funzioni

Esempio 2.3. Prendiamo ad esempio la seguente funzione (serie di fibonacci):

$$f(n) \subseteq \mathbb{N} \times \mathbb{N}$$

Dove:

$$f(0) = 1, f(1) = 1, f(2) = 2$$

$$f(3) = 3, f(4) = 5, f(5) = 8$$

$$f(6) = 13, f(7) = 21, \dots$$

Definiamo un algoritmo ricorsivo:

$$\begin{cases} f(0) = 1 = f(1) \\ f(x+2) = f(x+1) + f(x) \end{cases}$$

Trovare un algoritmo non è possibile per tutte le funzioni, ma solo per quelle calcolabili.

Nell'insieme delle funzioni calcolabili ci sono:

- Funzioni totali, cioè definite per ogni input $n \in \mathbb{N}$ e terminano sempre
- Funzioni parziali, cioè non definite per ogni $n \in \mathbb{N}$

2.2 Funzioni vs Insiemi

Una funzione può essere vista come un linguaggio \mathcal{L}_f tale che:

$$f: \mathbb{N} \rightarrow \mathbb{N} \quad \leftrightarrow \quad \mathcal{L}_f = \{1^{f(x)} \mid x \in \mathbb{N}\} \quad \Sigma = \{1\}$$

Questo linguaggio permette di dire se un input appartiene o meno al linguaggio:

$$\sigma \in \Sigma^*$$

↓

$$\begin{cases} \sigma \in \mathcal{L}_f & \text{se appartiene al linguaggio} \\ \sigma \notin \mathcal{L}_f & \text{se non appartiene al linguaggio} \end{cases}$$

Parliamo di insiemi invece che di funzioni dove gli elementi dell'insieme dipendono dal calcolo della funzione.

Esempio 2.4. Prendiamo ad esempio le seguenti funzioni:

- Funzione costante (Finite)

$$f(x) = 2 \rightarrow \mathcal{L}_f \text{ è finito}$$

- Funzione lineare (Regolari)

$$f(x) = 2x \rightarrow \mathcal{L}_f \text{ è infinito numerabile}$$

C'è bisogno di una memoria finita per determinare se la stringa appartiene al linguaggio

- (Context free)

$$f(\sigma) = \sigma\sigma^{\text{reverse}}$$

$$\sigma = abc \quad \sigma^{\text{reverse}} = cba$$

Per calcolare questa funzione c'è bisogno di una memoria illimitata, cioè non si può sapere a priori quanta ce n'è bisogno, è sufficiente uno stack.

- Decidibile

$$f(x) = x^2$$

Per calcolare questa funzione c'è bisogno di una memoria illimitata

Le funzioni calcolabili sono divise in classi secondo la gerarchia di Chomsky:

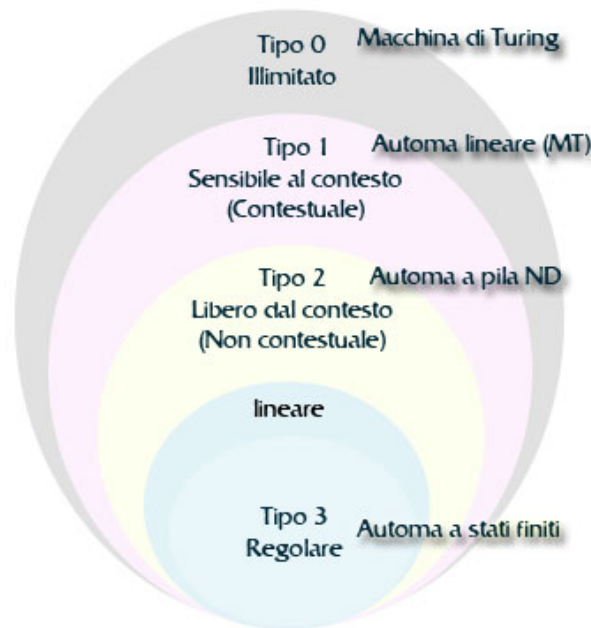


Figura 2: Gerarchia di Chomsky

3 Linguaggi regolari

Il principio di induzione è un meccanismo di definizione e dimostrazione che funziona **solo su insiemi infiniti**. Esistono due metodi di induzione:

- Induzione matematica
- Induzione strutturale

In questo corso tratteremo solo l'induzione matematica.

Un insieme A infinito con una relazione di ordine non riflessiva (senza l'uguale perchè l'elemento non è in relazione con sè stesso): $< : (A, <)$. $A = \mathbb{N}$ e $<$ è l'ordinamento stretto tra numeri naturali. La relazione di ordine deve essere **ben fondata**, quindi non devono esserci catene discendenti infinite, cioè una sequenza di elementi in ordine decrescente infinita:

$$a_0 > a_1 > a_2 > a_3 > \dots \rightarrow \text{non ben fondata}$$

Una relazione di ordine riflessiva non è ben fondata, perchè esistono catene infinite:

$$a_0 \geq a_1 \geq a_2 \geq a_3 \geq a_3 \geq a_3 \geq \dots \rightarrow \text{non ben fondata}$$

b minimale in A : $b \in A$ b è minimale se $\forall b' < b$. $b' \notin A$ Ad esempio: $\{1, 2, 3\}$ ha come minimali (di contenimento) $\{1, 2\}$ e $\{2, 3\}$.

Definizione 3.1 (Principio di induzione). Se A è un insieme ben fondato (con ordinamento $<$), e Π è una proprietà definita sugli elementi di A : $\Pi \subseteq A$, allora:

$$\forall a \in A . \underbrace{\Pi(a)}_{a \text{ soddisfa } \Pi} \iff \underbrace{\forall a \in A . [[\forall b < a . \Pi(b)] \Rightarrow \Pi(a)]}_{\text{Se dimostriamo } \Pi \text{ per ogni elemento più piccolo di } a, \text{ allora } \Pi \text{ vale anche per } a}$$

Consideriamo come caso base gli elementi minimali di A :

$$\text{Base}_A = \{a \in A \mid a \text{ minimale}\}$$

Se si dimostra che Π vale per tutti gli elementi minimali di A (la base):

$$\underbrace{\forall a \in \text{Base}_A . \Pi(a)}_{\substack{\text{Base} \\ \text{Dimostriamo } \Pi \text{ per ogni} \\ \text{elemento minimale di } A}} \wedge \underbrace{\forall a \in A \setminus \text{Base}_A . \forall b < a . \Pi(b)}_{\substack{\text{Passo induttivo} \\ \text{Ipotesi induttiva}}} \Rightarrow \underbrace{\Pi(a)}_{\text{Tesi da dimostrare}}$$

Esempio 3.1. Dimostriamo che:

$$\forall n \in \mathbb{N} \sum_{i=1}^n i = \frac{n(n+1)}{2}$$

L'insieme è:

$$A = \mathbb{N} \setminus \{0\} = \{1, 2, 3, \dots\}$$

- **Base**

$$\text{Base}_A = \{1\}$$

– dimostriamo la base

$$\sum_{i=1}^1 i = n(n+1)/2 = 1(1+1)/2 = 1$$

- **Passo induttivo:** Prendo $n \in \mathbb{N}$

– Ipotesi induttiva, cioè per ogni $m < n$ vale la proprietà:

$$\forall m < n . \sum_{i=1}^m i = \frac{m(m+1)}{2}$$

Dobbiamo dimostrare la proprietà per n :

$$\sum_{i=1}^n i = \sum_{i=1}^{n-1} i + n$$

$n - 1 < n$ quindi vale l'ipotesi induttiva

$$\sum_{i=1}^{n-1} i = \frac{(n-1)(n-1+1)}{2} = \frac{(n-1)n}{2}$$

↓

$$\begin{aligned}\sum_{i=1}^n i &= \sum_{i=1}^{n-1} i + n \\ &= \frac{(n-1)n}{2} + n \\ &= \frac{(n-1)n + 2n}{2} \\ &= \frac{n^2 - n + 2n}{2} \\ &= \frac{n(n+1)}{2} \quad \square\end{aligned}$$

È quindi dimostrato che:

$$\forall n \in \mathbb{N} \sum_{i=1}^n i = \frac{n(n+1)}{2}$$

3.1 Linguaggi formali

Definizione 3.2. Un linguaggio formale è un insieme di stringhe costruite su un alfabeto finito Σ .

Solitamente un linguaggio formale \mathcal{L} è un sottoinsieme di Σ^* , tipicamente infiniti, ma non necessariamente:

$$\mathcal{L} \subseteq \Sigma^*$$

I linguaggi sono divisi in:

- Linguaggi finiti
- Linguaggi regolari, il modello utilizzato è l'automa a stati finiti.

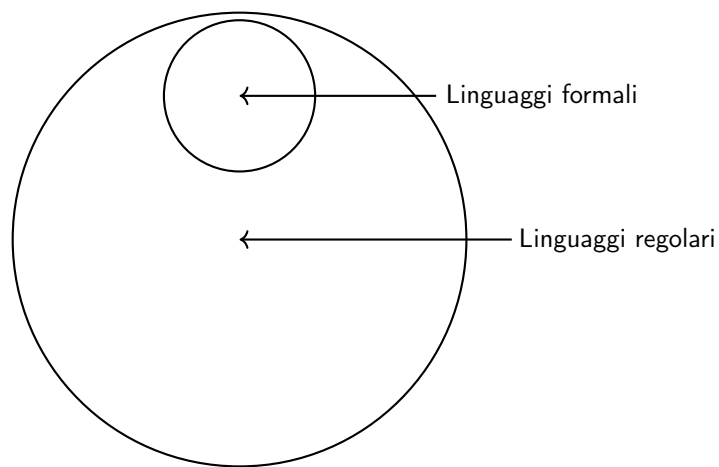


Figura 3: Linguaggi formali e linguaggi regolari

3.2 Linguaggi regolari (automi a stati finiti, DFA)

Il meccanismo più semplice per una memoria finita è l'automa a stati finiti

Consideriamo il linguaggio:

$$\mathcal{L}_f = \{1^{2n} \mid n \in \mathbb{N}\}$$

ha bisogno di due stati q_0 e q_1 . Lo stato q_0 rappresenta l'informazione di essere di lunghezza pari, mentre lo stato q_1 rappresenta l'informazione di essere di lunghezza dispari.

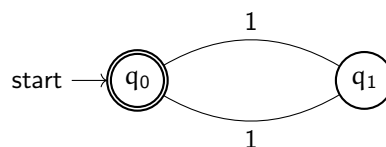


Figura 4: Automa a stati finiti per il linguaggio $\mathcal{L}_f = \{1^{2n} \mid n \in \mathbb{N}\}$

L'automa a stati finiti **deterministico** è definito come una quintupla:

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

dove:

- Q è un insieme **finito** di stati. Ogni stato rappresenta un'informazione
- Σ è un insieme **finito** di simboli (alfabeto). Ogni simbolo è un elemento atomico che posso leggere e che compone le stringhe da riconoscere
- $q_0 \in Q$ è uno stato e identifica lo stato iniziale. Lo stato finale viene indicato con un doppio cerchio
- $F \subseteq Q$ è l'insieme degli stati finali (di accettazione)

- $\delta : Q \times \Sigma \rightarrow Q$ È una **funzione di transizione** che dato uno stato e un simbolo, restituisce lo stato successivo ed è come se fosse una tabella che associa ad ogni coppia (stato, simbolo) uno stato:

$\Sigma \setminus Q$	q_0	q_1
1	q_1	q_0

La funzione deve essere **totale**, cioè deve essere definita per ogni coppia $(q, a) \in Q \times \Sigma$, quindi la tabella deve essere completa.

- $\hat{\delta} : Q \times \Sigma^* \rightarrow Q$ Descrive lo stato che raggiungono leggendo una sequenza di simboli

$$\begin{cases} \hat{\delta}(q, \epsilon) = q \\ \hat{\delta}(q, wa) = \delta(\hat{\delta}(q, w), a) \end{cases} \quad w \in \Sigma^*, \quad a \in \Sigma$$

È quindi la **chiusura transitiva** di δ

3.2.1 Come si dimostra che un linguaggio è regolare?

Esempio 3.2. Prendiamo in considerazione il seguente linguaggio:

$$L = \{\sigma \mid \sigma \text{ contiene almeno due } 1\}$$

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

Con le seguenti stringhe si ha:

- $\sigma = 011 \in \mathcal{L}$
- $\sigma = 1000100 \in \mathcal{L}$
- $\sigma = 00010 \notin \mathcal{L}$

L'informazione che codifica lo stato iniziale deve essere coerente con ϵ (stringa vuota)

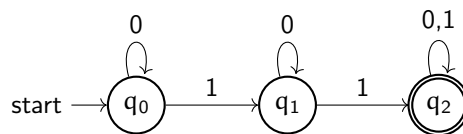


Figura 5: Automa a stati finiti per il linguaggio \mathcal{L}

Un linguaggio L è riconosciuto da $M = (Q, \Sigma, \delta, q_0, f)$ (DFA, Deterministic Finite Automaton) se: $L = L(M)$ dove $L(M)$ è il linguaggio di M definito come:

$$L(M) = \{\sigma \in \Sigma^* \mid \hat{\delta}(q_0, \sigma) \in F\}$$

Cioè sono tutte le stringhe che partendo da q_0 fanno raggiungere uno stato finale.

Definizione 3.3. Per dimostrare che L è regolare dobbiamo costruire M (almeno un M) e **dimostrare che** $L = L(M)$

$L = L(M)$ è un'uguaglianza insiemistica e si dimostra con due contenimenti:

$$L = L(M) \equiv L \subseteq L(M) \wedge L(M) \subseteq L$$

- Se un elemento si trova nel primo insieme, allora si trova anche nel secondo

$$L \subseteq L(M) \equiv \sigma \in L \Rightarrow \sigma \in L(M) \equiv \sigma \in L \Rightarrow \hat{\delta}(q_0, \sigma) \in F$$

- Se un elemento si trova nel secondo insieme, allora si trova anche nel primo

$$L(M) \subseteq L \equiv \sigma \in L(M) \Rightarrow \sigma \in L \equiv \hat{\delta}(q_0, \sigma) \in F \Rightarrow \sigma \in L$$

o per contrapposizione:

$$\sigma \notin L \Rightarrow \hat{\delta}(q_0, \sigma) \notin F$$

Questo dimostra che il linguaggio è regolare perchè è riconosciuto da un automa.

Esempio 3.3. Riprendendo l'esempio precedente:

$$L = \{\sigma \in \Sigma^* \mid \sigma \text{ contiene almeno due } 1\}$$

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

$M =$

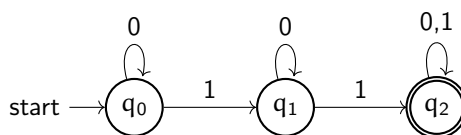


Figura 6: Automa a stati finiti per il linguaggio \mathcal{L}

Dimostriamo per induzione sulla lunghezza delle stringhe $\sigma \in \Sigma^*$ che se $x \in L$ allora $\hat{\delta}(q_0, x) \in F$ e se $x \notin L$ allora $\hat{\delta}(q_0, x) \notin F$.

$|\sigma| = 0$ non è **mai** sufficiente come base, ma è eventualmente la base **solo** per una delle due dimostrazioni. Bisogna quindi prendere la lunghezza più piccola che permette di avere sia $\sigma \in L$ che $\sigma \notin L$, in questo caso è $|\sigma| = 2$. Per ogni σ tale che $|\sigma| < 2$ $\sigma \notin L$ perchè non può contenere due 1 e non è riconosciuta da M dove il primo stato finale è raggiunto leggendo almeno due simboli.

$$\varepsilon \in L \quad \varepsilon \notin L$$

- **Base:** Controlliamo ogni stringa di lunghezza minima nel linguaggio

per provare il caso base. In questo caso la lunghezza minima è $|\sigma| = 2$

$$\begin{cases} \sigma = 11 \in L \text{ e } \hat{\delta}(q_0, 11) = q_2 \in F \\ \sigma = 10 \notin L \text{ e } \hat{\delta}(q_0, 10) = q_1 \notin F \\ \sigma = 01 \notin L \text{ e } \hat{\delta}(q_0, 01) = q_1 \notin F \\ \sigma = 00 \notin L \text{ e } \hat{\delta}(q_0, 00) = q_0 \notin F \end{cases}$$

- **Passo induttivo:** Assumiamo che valga l'**ipotesi induttiva**, cioè la tesi con un limite fissato:

$$\forall \sigma \in \Sigma^* . |\sigma| \leq n . \begin{cases} \sigma \in L \Rightarrow \hat{\delta}(q_0, \sigma) \in F \\ \sigma \notin L \Rightarrow \hat{\delta}(q_0, \sigma) \notin F \end{cases}$$

Vogliamo dimostrare che la tesi vale per $|\sigma| = n + 1$ (la successiva stringa che posso considerare).

$$\text{Tesi: } \begin{cases} \sigma \in L \Rightarrow \hat{\delta}(q_0, \sigma) = q_2 \text{ } \sigma \text{ contiene almeno due 1} \\ \sigma \notin L \Rightarrow \hat{\delta}(q_0, \sigma) = q_0 \text{ } \sigma \text{ non contiene 1} \\ \sigma \notin L \Rightarrow \hat{\delta}(q_0, \sigma) = q_1 \text{ } \sigma \text{ contiene esattamente un 1} \end{cases}$$

Ipotesi induttiva:

$$\forall \sigma \in \Sigma^* . |\sigma| \leq n . \text{ allora la tesi vale su } \sigma$$

Dimostrazione della tesi per σ tale che $|\sigma| = n + 1$. ($|\sigma'| = n$ quindi su σ' possiamo applicare l'ipotesi induttiva):

$$|\sigma| = n + 1 \rightarrow \sigma = \sigma'1 \vee \sigma = \sigma'0$$

– Supponiamo che σ appartenga al linguaggio e termini con 1:

$$\sigma \in L \wedge \sigma = \sigma'1$$

↓

* Se $\sigma' \in L$ applico l'ipotesi induttiva:

$$\hat{\delta}(q_0, \sigma') = q_2$$

$$\begin{aligned} \hat{\delta}(q_0, \sigma) &\stackrel{\sigma = \sigma'1}{=} \hat{\delta}(q_0, \sigma'1) \\ &= \delta(\hat{\delta}(q_0, \sigma'), 1) \\ &= \delta(q_2, 1) = q_2 \in F \end{aligned}$$

* Se $\sigma' \notin L$ allora σ' contiene esattamente un 1:

$$\hat{\delta}(q_0, \sigma') = q_1$$

$$\hat{\delta}(q_0, \sigma'1) = \delta(q_1, 1) = q_2$$

– Supponiamo che σ appartenga al linguaggio e termini con 0:

$$\sigma \in L \wedge \sigma = \sigma'0$$

Per definizione di L abbiamo che

$$\sigma \in L \wedge \sigma = \sigma'0 \Rightarrow \sigma' \in L$$

Dimostriamo l'ipotesi induttiva:

$$\hat{\delta}(q_0, \sigma') = q_2$$

allora

$$\begin{aligned}\hat{\delta}(q_0, \sigma) &= \hat{\delta}(q_0, \sigma'0) \\ &= \delta(\hat{\delta}(q_0, \sigma'), 0) \\ &= \delta(q_2, 0) = q_2 \in F\end{aligned}$$

– Supponiamo che σ non appartenga al linguaggio e contiene esattamente un 1:

* $\sigma = \sigma'0 \Rightarrow \sigma' \notin L$ e contiene esattamente un 1

Ipotesi induttiva:

$$\hat{\delta}(q_0, \sigma') = q_1$$

\Downarrow

$$\begin{aligned}\hat{\delta}(q_0, \sigma) &= \hat{\delta}(q_0, \sigma'0) \\ &= \delta(\hat{\delta}(q_0, \sigma'), 0) \\ &= \delta(q_1, 0) = q_1 \notin F\end{aligned}$$

* $\sigma = \sigma'1 \Rightarrow \sigma' \notin L$ e non contiene 1

Ipotesi induttiva:

$$\hat{\delta}(q_0, \sigma') = q_0$$

\Downarrow

$$\begin{aligned}\hat{\delta}(q_0, \sigma) &= \hat{\delta}(q_0, \sigma'1) \\ &= \delta(\hat{\delta}(q_0, \sigma'), 1) \\ &= \delta(q_0, 1) = q_1 \notin F\end{aligned}$$

– Supponiamo che σ non appartenga al linguaggio e non contiene 1

$$\sigma \notin L \wedge \sigma = \sigma'0$$

$\sigma = \sigma'1$ non è possibile per l'ipotesi che σ non contiene 1

Ipotesi induttiva:

$$\hat{\delta}(q_0, \sigma') = q_0$$

\Downarrow

$$\begin{aligned}\hat{\delta}(q_0, \sigma) &= \hat{\delta}(q_0, \sigma'0) \\ &= \delta(\hat{\delta}(q_0, \sigma'), 0) \\ &= \delta(q_0, 0) = q_0 \notin F\end{aligned}$$

Tutti i casi sono dimostrati, quindi abbiamo dimostrato che:

$$L = L(M) \Rightarrow L \text{ è regolare}$$

Esercizio 3.1. Consideriamo il seguente linguaggio:

$$L = \{\sigma \in \Sigma^* \mid \text{ogni sequenza di 0 è di lunghezza pari}\}$$

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

Si può accettare anche sequenze di lunghezza 0. Alcuni esempi sono:

$$101 \notin L$$

$$1111 \in L$$

$$10010000 \in L$$

$$00101 \notin L$$

L'automa a stati finiti M è il seguente:

- q_0 : Non sono stati letti 0
- q_1 : Sequenza di 0 consecutiva di lunghezza dispari
- q_2 : Sequenza di 0 consecutiva di lunghezza pari

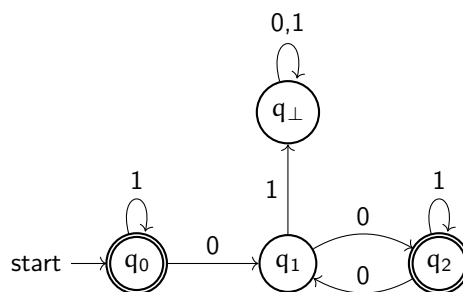


Figura 7: Automa a stati finiti per il linguaggio L

La tesi è:

- $\sigma \in L$ e non contiene 0: $\Rightarrow \hat{\delta}(q_0, \sigma) = q_0$
- $\sigma \in L$ e contiene una sequenza pari di 0: $\Rightarrow \hat{\delta}(q_0, \sigma) = q_2$
- $\sigma \notin L$ e contiene una sequenza **finale** dispari di 0: $\Rightarrow \hat{\delta}(q_0, \sigma) = q_1$
- $\sigma \notin L$ e contiene una sequenza dispari di 0 seguita da 1: $\Rightarrow \hat{\delta}(q_0, \sigma) = q_{\perp}$

Esercizio 3.2. Consideriamo il seguente linguaggio (ogni sequenza di 0 è di lunghezza almeno 2):

$$L = \{\sigma \in \Sigma^* \mid \exists n \geq 1. \sigma = 0^n \Rightarrow n \geq 2\}$$

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

L'automa a stati finiti M è il seguente:

- q_0 : Non contiene 0, oppure **tutte** le sequenze di 0 sono lunghe almeno 2
- q_1 : Esattamente uno 0
- q_2 : Almeno due 0

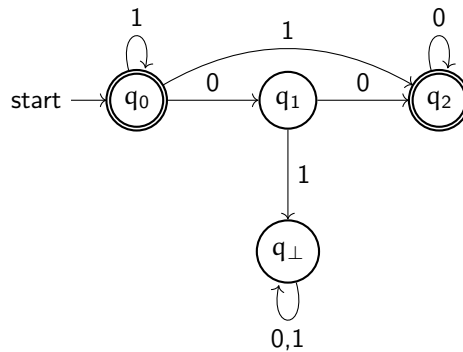


Figura 8: Automa a stati finiti per il linguaggio L

La tesi è:

- $\sigma \in L$ e $\sigma = \sigma'1 \Rightarrow \hat{\delta}(q_0, \sigma) = q_0$
- $\sigma \in L$ e $\sigma = \sigma'0 \Rightarrow \hat{\delta}(q_0, \sigma) = q_2$
- $\sigma \notin L$ e $\sigma = \sigma'0$ dove l'ultima sequenza di 0 è esattamente lunga 1:
 $\Rightarrow \hat{\delta}(q_0, \sigma) = q_1$
- $\sigma \notin L$ e σ contiene una sequenza lunga 1 di 0: $\Rightarrow \hat{\delta}(q_0, \sigma) = q_{\perp}$

3.3 Automi a stati finiti non deterministici (NFA)

Un automa a stati finiti non deterministico si crea quando ad un solo simbolo sono associate più transizioni. Quando questo succede gli stati vengono considerati in parallelo. Un NFA è definito come una quintupla:

$$N = \langle Q, \Sigma, \delta, q_0, F \rangle$$

- Q è un insieme finito di stati
- Σ è un insieme finito di simboli (alfabeto)

- $q_0 \in Q$ è uno stato e identifica lo stato iniziale
- $F \subseteq Q$ è l'insieme degli stati finali
- $\delta : Q \times \Sigma \rightarrow \mathcal{P}(Q)$ è una funzione di transizione che dato uno stato e un simbolo restituisce un insieme di stati **potenzialmente** raggiungibili. È possibile che esistano coppie associate all'insieme vuoto:

$$\emptyset \in \mathcal{P}(Q)$$

Inoltre non è obbligatorio avere un arco uscente per ogni simbolo di Σ .

- $\hat{\delta} : Q \times \Sigma^* \rightarrow \mathcal{P}(Q)$ Descrive gli stati che si possono raggiungere leggendo una sequenza di simboli:

$$\begin{cases} \hat{\delta}(q, \epsilon) = \{q\} \\ \hat{\delta}(q, wa) = \bigcup_{p \in \hat{\delta}(q, w)} \delta(p, a) \end{cases} \quad w \in \Sigma^*, \quad a \in \Sigma$$

È quindi la chiusura transitiva di δ

Esempio 3.4. Un esempio di NFA è il seguente:

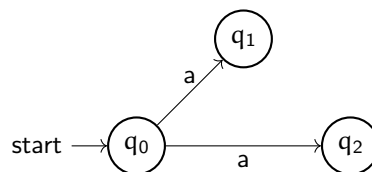


Figura 9: Esempio di NFA

$$\delta(q_0, a) = \{q_1, q_2\} \subseteq Q$$

3.3.1 Linguaggio riconosciuto da un NFA

Un linguaggio L è riconosciuto da un NFA N se:

$$L(N) = \{\sigma \in \Sigma^* \mid \hat{\delta}(q_0, \sigma) \cap F \neq \emptyset\}$$

Esempio 3.5. Consideriamo il seguente linguaggio:

$$L(N) = \{\sigma \in \Sigma^* \mid \sigma \text{ contiene almeno due } 1\}$$

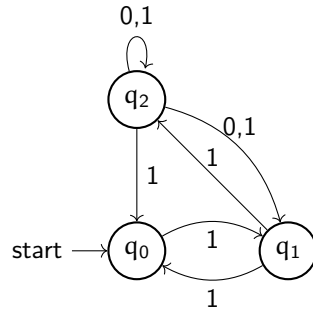


Figura 10: Esempio di NFA per il linguaggio L

Teorema 3.1 (Teorema di Rabin-Scott). Ogni linguaggio riconosciuto da un NFA è riconosciuto da un DFA.

$$\forall N = (Q, \Sigma, \delta, q_0, F) \exists M = (Q', \Sigma, \delta', q'_0, F') . L(N) = L(M)$$

Dimostrazione: Consideriamo un NFA $N = (Q, \Sigma, \delta, q_0, F)$ e costruiamo un DFA $M = (Q', \Sigma, \delta', q'_0, F')$. Gli insiemi degli stati sono:

$$Q' = \mathcal{P}(Q)$$

$$Q = \{q_0, q_1, q_2\}$$

\Downarrow

$$Q' = \left\{ \overset{q'_1}{\emptyset}, \overset{q'_0}{\{q_0\}}, \overset{q'_2}{\{q_2\}}, \overset{q'_3}{\{q_0, q_1\}}, \overset{q'_4}{\{q_0, q_2\}}, \overset{q'_5}{\{q_1, q_2\}}, \overset{q'_6}{\{q_0, q_1, q_2\}} \right\}$$

Lo stato iniziale rimane uguale per entrambi gli insiemi: $q'_0 = \{q_0\}$.

Gli insiemi degli stati finali sono:

$$F' = \{P \subseteq Q \mid P \cap F \neq \emptyset\} \quad P \in \mathcal{P}(Q)$$

Quindi:

$$F = \{q_2\}$$

$$F' = \left\{ \overset{q'_3}{\{q_2\}}, \overset{q'_6}{\{q_1, q_2\}}, \overset{q'_5}{\{q_0, q_2\}}, \overset{q'_7}{\{q_0, q_1, q_2\}} \right\}$$

La funzione di transizione è definita come:

$$\delta'(P, a) = \bigcup_{q \in P} \delta(q, a) \in \mathcal{P}(Q) \quad P \in Q' = \mathcal{P}(Q), a \in \Sigma$$

Quindi:

$$\begin{aligned}\underbrace{\delta'(q'_5, 1)}_{=\{q_1, q_2\}} &= \delta(q_1, 1) \cup \delta(q_2, 1) \\ &= \{q_0, q_2\} \cup \{q_0, q_1, q_2\} \\ &= \{q_0, q_1, q_2\} = q'_7\end{aligned}$$

Dimostriamo:

$$1. \hat{\delta}(q_0, \sigma) = \hat{\delta}'(q'_0, \sigma) = \{q_0\}$$

Dimostriamo per induzione su $|\sigma|$:

- Se $\sigma = \varepsilon$ allora:

$$\hat{\delta}'(q'_0, \varepsilon) = q'_0 = \{q_0\} = \hat{\delta}(q_0, \varepsilon)$$

per le definizioni

- Se $\sigma = \sigma'a$:

$$\begin{aligned}\hat{\delta}'(q'_0, \sigma'a) &= \delta'(\hat{\delta}'(q_0, \sigma'), a) \\ &= \delta'(\hat{\delta}(q_0, \sigma'), a) \\ &= \bigcup_{p \in \hat{\delta}(q_0, \sigma')} \delta(p, a) \\ &= \hat{\delta}(q_0, \sigma'a)\end{aligned}$$

Definizione di $\hat{\delta}'$ non deterministica

$$2. \sigma \in L(N) \iff \sigma \in L(M)$$

Dimostriamo la definizione di linguaggio riconosciuto in NFA:

$$\begin{aligned}\sigma \in L(N) &\iff \hat{\delta}(q_0, \sigma) \cap F \neq \emptyset \\ &\iff \hat{\delta}'(q'_0, \sigma) \cap F' \neq \emptyset \text{ ((1.))} \\ &\iff \hat{\delta}'(q'_0, \sigma) \in F' \\ &\iff \sigma \in L(M) \text{ (def. linguaggio accettato in DFA)}\end{aligned}$$

3.3.2 Conversione da NFA a DFA

Prendiamo in considerazione il seguente NFA:

$$L(N) = \{\sigma \in \Sigma^* \mid \sigma \text{ contiene almeno due } 1\}$$

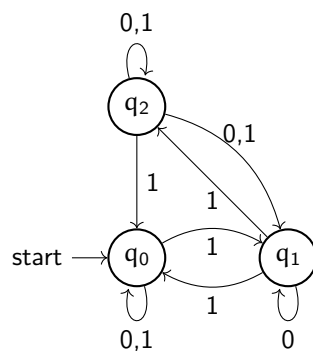


Figura 11: Esempio di NFA per il linguaggio L

Per trasformare un NFA in un DFA bisogna creare dei nuovi stati che raggruppano gli stati non deterministici. In questo caso:

- **Tabella degli stati della NFA:**

Stato	Input 0	Input 1
q ₀	{q ₀ }	{q ₀ , q ₁ }
q ₁	{q ₁ }	{q ₀ , q ₂ }
q ₂	{q ₁ , q ₂ }	{q ₀ , q ₁ , q ₂ }

Tabella 1: Tabella di transizione della NFA

- **Traduzione degli stati della NFA in stati del DFA:**

	0	1
	∅	∅

Tabella 2: Tabella di traduzione degli stati della NFA in stati del DFA

3.4 Automi non deterministici con ε -transizioni (ε -NFA)

Questo tipo di NFA permette di cambiare stato anche senza leggere simboli:

$$q_1 \xrightarrow{\varepsilon} q_2$$

Un ε -NFA è definito come un NFA con la differenza che la funzione di transizione è definita come:

$$\delta : Q \times (\Sigma \cup \{\varepsilon\}) \rightarrow \underbrace{\mathcal{P}(Q)}_{\text{Non determinismo}}$$

Esempio 3.6. Prendiamo ad esempio il seguente ε -NFA in cui leggendo solo a si può raggiungere sia q' che q'' :

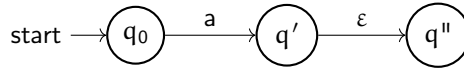


Figura 12: Esempio di ϵ -NFA

3.4.1 ϵ -closure

Per definire $\hat{\delta}$ bisogna prima definire la ϵ -closure.

Una ϵ -closure di uno stato q è l'insieme di tutti gli stati che si possono raggiungere da q seguendo archi etichettati con ϵ :

$$\epsilon\text{-closure} : Q \rightarrow \mathcal{P}(Q)$$

O definito per insiemi:

$$\epsilon\text{-closure} : \mathcal{P}(Q) \rightarrow \mathcal{P}(Q)$$

$$\epsilon\text{-closure}(P) = \bigcup_{p \in P} \epsilon\text{-closure}(p)$$

La funzione $\hat{\delta}$ è definita come:

$$\begin{cases} \hat{\delta}(q, \epsilon) &= \epsilon\text{-closure}(q) \\ \hat{\delta}(q, wa) &= \bigcup_{p \in \hat{\delta}(q, w)} \epsilon\text{-closure}(\delta(p, a)) \end{cases}$$

Esempio 3.7. La ϵ -closure dell'esempio precedente è:

$$\epsilon\text{-closure}(q') = \{q', q''\}$$

Il riconoscimento di un linguaggio è analogo a quello di un NFA:

$$L(N) = \{ \sigma \in \Sigma^* \mid \hat{\delta}(q_0, \sigma) \cap F \neq \emptyset \}$$

Teorema 3.2. Sia $M = \langle Q, \Sigma, \delta, q_0, F \rangle$ una ϵ -NFA, allora esiste una NFA M' tale che $L(M) = L(M')$.

Quindi l'insieme dei linguaggi riconosciuti da ϵ -NFA coincide con quello degli NFA, che a sua volta coincide con i linguaggi regolari.

$$M = \langle Q, \Sigma, \delta, q_0, F \rangle \text{ } \epsilon\text{-NFA}$$

Costruiamo una NFA

$$M' = \langle Q', \Sigma', \delta', q'_0, F' \rangle$$

Dove:

$$Q' = Q, \quad \Sigma' = \Sigma, \quad q'_0 = q_0$$

$$\delta'(q, a) = \hat{\delta}(q, a) \quad F' = \begin{cases} F \cup \{q_0\} & \text{se } \epsilon\text{-closure}(q_0) \cap F \neq \emptyset \\ F & \text{altrimenti} \end{cases}$$

in cui il numero degli stati rimane lo stesso, l'alfabeto non cambia e neanche lo stato iniziale.

4 Espressioni regolari

Le espressioni regolari sono operazioni algebriche sui linguaggi regolari. Le operazioni che si possono fare sono:

- **Unione:** Dati i linguaggi $L_1, L_2 \subseteq \Sigma^*$

$$L_1 \cup L_2 = \{\sigma \mid \sigma \in L_1 \vee \sigma \in L_2\}$$

- **Concatenazione:** Dati i linguaggi $L_1, L_2 \subseteq \Sigma^*$

$$L_1 \cdot L_2 = \{\sigma_1 \sigma_2 \mid \sigma_1 \in L_1 \wedge \sigma_2 \in L_2\} \equiv L_1 L_2$$

Ad esempio:

$$L_1 = \{0101, 010101\} \quad L_2 = \{000, 111\}$$

$$L_1 \cdot L_2 = \{0101000, 0101111, 010101000, 010101111\}$$

- **Stella di Kleene:** Dato un linguaggio $L \subseteq \Sigma^*$

$$L^* = \bigcup_{n \in \mathbb{N}} L^n$$

Cioè la concatenazione di L con se stesso n volte, con:

$$\begin{cases} L^0 = \{\varepsilon\} \\ L^{n+1} = L \cdot L^n \end{cases}$$

Ad esempio:

$$L = \{000, 111\}$$

$$L^0 = \{\varepsilon\}$$

$$L^1 = L \cdot L^0 = \{000, 111\}$$

$$L^2 = L \cdot L^1 = \{000000, 000111, 111000, 111111\}$$

$$L^3 = L \cdot L^2 = \left\{ \begin{array}{l} 000000000, 000000111, 000111000, 000111111, \\ 111000000, 111000111, 111111000, 111111111 \end{array} \right\}$$

\vdots

$$L^* = \{L^0, L^1, L^2, L^3, \dots\}$$

- L^* è l'insieme di tutte le possibili concatenazioni di stringhe appartenenti a L , compresa la stringa vuota ε
- $L^+ = \bigcup_{n \geq 0} L^n$ è definito come L^* senza la stringa vuota ε :

$$L^+ = L \cdot L^*$$

Definizione 4.1. Definiamo per induzione le espressioni regolari su un alfabeto Σ :

- **Caso base:**

- $\emptyset \subseteq \Sigma^*$ è un'espressione regolare che rappresenta il linguaggio vuoto
- ε è un'espressione regolare che rappresenta il linguaggio:

$$\{\varepsilon\} \subseteq \Sigma^*$$

- $a \in \Sigma$ è un'espressione regolare che rappresenta il linguaggio:

$$\{a\} \subseteq \Sigma^*$$

- **Passo induttivo:**

Siano r, s sono espressioni regolari che rappresentano il linguaggio

$$R \subseteq \Sigma^* \wedge S \subseteq \Sigma^*$$

- $r + s$ è un'espressione regolare che rappresenta il linguaggio $R \cup S$
- $r \cdot s$ è un'espressione regolare che rappresenta il linguaggio $R \cdot S$
- r^* è un'espressione regolare che rappresenta il linguaggio:

$$R^*$$

Esempio 4.1. Prendiamo ad esempio l'espressione regolare:

$$1^* + 0^* + (10)^*$$

che equivale a

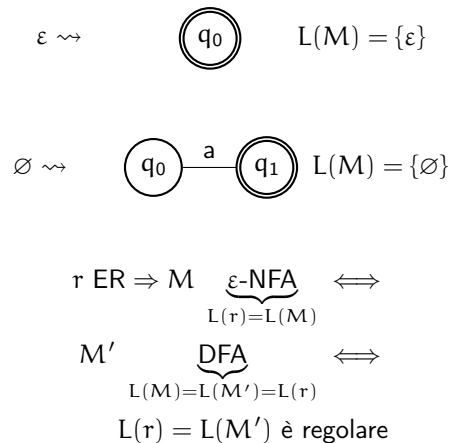
$$\{1^n \mid n \in \mathbb{N}\} \cup \{0^n \mid n \in \mathbb{N}\} \cup \{(10)^n \mid n \in \mathbb{N}\}$$

Teorema 4.1 (Teorema di equivalenza). Dato un DFA $M = \langle Q, \Sigma, \delta, q_0, F \rangle$ allora esiste un'espressione regolare r tale che $L(M) = L(r)$.

$$L \text{ Regolare} \xLeftrightarrow{\text{def}} \underbrace{\exists M \text{ DFA} . L}_{L(M)=L} \xRightarrow{\text{Th}} \exists r \in \underbrace{\text{ER}}_{\text{Espressioni Regolari}} . L(r) = L$$

Teorema 4.2. Data un'espressione regolare (ER) r esiste una ε -NFA M tale che: $L(r) = L(M)$

Quindi:



Esempio 4.2. Consideriamo il seguente linguaggio:

$$L = \{\sigma \in \Sigma^* \mid \sigma \text{ contiene almeno due } 1\}$$

Per dimostrare che è regolare si costruisce il DFA M e si dimostra $L = L(M)$. Però se si ha un'espressione regolare $r = 0^*10^*10^*$ **non** dimostra che L è regolare. Bisognerebbe dimostrare $L = L(r)$

4.1 Proprietà dei linguaggi regolari

4.1.1 Proprietà di chiusura

Indica se l'insieme dei linguaggi regolari è chiuso rispetto ad alcune operazioni, cioè se applicando queste operazioni a linguaggi regolari si ottengono sempre linguaggi regolari.

Operazioni:

- $*$
- \cup
- \cdot
- \cap ($L_1 \cap L_2 = \{\sigma \mid \sigma \in L_1 \wedge \sigma \in L_2\}$)
- $^-$ ($\bar{L} = \{\sigma \mid \sigma \text{ not in } L\}$)

Teorema 4.3. I linguaggi regolari sono chiusi rispetto alle operazioni di:

- Stella di Kleene
- Unione (finita)

- Concatenazione

Consideriamo i linguaggi regolari L_1, L_2 . Allora:

- L_1^* è regolare
- $L_1 \cup L_2$ è regolare
- $L_1 \cdot L_2$ è regolare

Teorema 4.4. I linguaggi regolari sono chiusi rispetto alla complementazione. Dato un automa a stati finiti, il linguaggio complementare è riconosciuto dall'automata complementare, cioè quello in cui gli stati finali diventano non finali e viceversa.

Per le leggi di De Morgan, i linguaggi regolari sono chiusi per intersezione finita:

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

4.1.2 Proprietà di decidibilità

Indica se esistono algoritmi che risolvono alcuni problemi sui linguaggi regolari. Consideriamo un insieme di stringhe accettate da un DFA (linguaggio regolare) M con n stati.

- $L(M) \neq \emptyset$ se e solo se accetta almeno una stringa di lunghezza $\leq n$
- $L(M)$ è infinito se e solo se accetta almeno una stringa di lunghezza l con $n \leq l < 2n$. Cioè se esiste un ciclo. Questo fornisce un estremo superiore per verificare se un linguaggio è infinito.
- $L(M_1) = L(M_2)$

4.1.3 Esistenza dell'automa minimo

Forniamo strategie per costruire un automa minimo.

Definizione 4.2 (Relazione di equivalenza e partizione). Data una relazione R su Σ . R è una relazione di:

- Equivalenza: $E \subset \Sigma \times \Sigma$
- Riflessiva: $\forall a. aRa$
- Simmetrica: $\forall a, b. aRb \Rightarrow bRa$
- Transitiva: $\forall a, b, c. aRb \wedge bRc \Rightarrow aRc$

La **relazione di equivalenza R induce una partizione** (unione di insiemi) di S :

$$R \subseteq S \times S$$

ovvero:

$$S = S_1 \cup S_2 \cup \dots \cup S_k$$

dove S_i sono una partizione di S . Inoltre:

- $\forall i, j . S_i \cap S_j = \emptyset$
- $\forall i . \forall a, b \in S_i . aRb$
- $\forall a \in S_i, b \in S_j, i \neq j . a \not R b$

Quindi R è come se dividesse S in insiemi disgiunti:

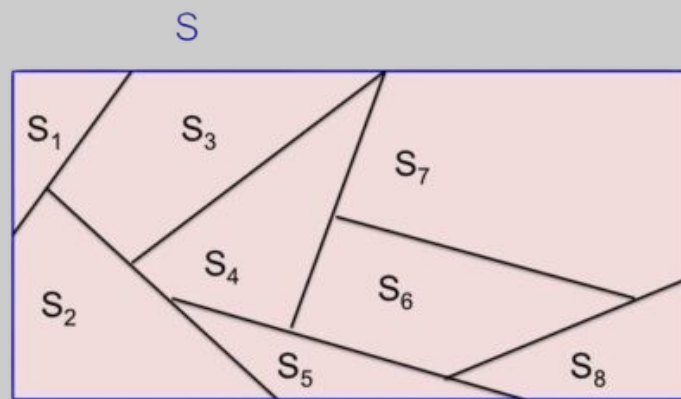


Figura 13: Esempio di partizione su S

S_i sono detti **classi di equivalenza** di R e si indica con:

$$a \in R \Rightarrow [a]_R = \{b \mid aRb\}$$

Ad esempio:

$$cRa \Rightarrow [a]_R \equiv [c]_R$$

Definiamo di seguito due relazioni di equivalenza:

Definizione 4.3 (Classe di equivalenza definita per tutti i linguaggi). Consideriamo un linguaggio $L \subseteq \Sigma^*$, possiamo definire una relazione R_L come:

$$R_L \subseteq \Sigma^* \times \Sigma^*$$

$$x, y \in \Sigma^* \quad xR_L y \iff \forall z \in \Sigma^* . xz \in L \iff yz \in L$$

Quindi o entrambi appartengono a L o entrambi non appartengono a L .

Definizione 4.4 (Classe di equivalenza definita per gli automi). Consideriamo un automa M DFA $\langle Q, \Sigma, \delta, q_0, F \rangle$. Possiamo definire una relazione R_M come:

$$R_M \subseteq \Sigma^* \times \Sigma^*$$

$$x, y \in \Sigma^* \quad x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$$

Due stringhe sono in relazione se raggiungono lo stesso stato.

Definizione 4.5 (Relazione invariante destra). Una relazione R su Σ^* : $R \subseteq \Sigma^* \times \Sigma^*$ è **invariante destra** se e solo se:

$$x, y \in \Sigma^* \quad x R y \implies \forall z \in \Sigma^* . xz R yz$$

Se due stringhe sono in relazione tra di loro, allora in qualunque modo vengano estese, rimarranno in relazione tra di loro. Quindi essere in relazione è invariante rispetto all'estensione della stringa **verso destra**

R_L e R_M sono relazioni invarianti destre.

Definizione 4.6 (Raffinamento). La relazione R_2 è **raffinamento** di R_1 se:

$$\begin{array}{l} R_2 \subseteq S \times S \\ R_1 \subseteq S \times S \end{array} \quad \text{di equivalenza}$$

e R_1 è più grossa di R_2 , cioè ogni classe di equivalenza di R_2 è contenuta in una classe di equivalenza di R_1 e quindi è come se fosse più dettagliata:

$$\forall x . [x]_{R_2} \subseteq [x]_{R_1}$$

Il numero di classi di equivalenza di R_2 è maggiore del numero di classi di equivalenza di R_1 .

Definizione 4.7. Dato un automa DFA $M = \langle Q, \Sigma, \delta, q_0, F \rangle$, ogni stato definisce un linguaggio L_q come:

$$q \in Q \quad : \quad L_q = \{ \sigma \in \Sigma^* \mid \hat{\delta}(q, \sigma) = q \}$$

Teorema 4.5 (Teorema di Myhill-Nerode). I seguenti enunciati sono equivalenti:

1. $L \subseteq \Sigma^*$ è un linguaggio regolare, ovvero esiste un DFA M tale che

$$L = L(M)$$

2. L è unione di classi di equivalenza (cioè è partizionato) indotte da una relazione di equivalenza R invariante destra e di indice finito, cioè se il numero di classi di equivalenza indotte è finito.
3. R_L è di indice finito

Dimostriamo che:

$$1. \implies 2. \xRightarrow[R \text{ raffina } R_L]{} 3. \xRightarrow[\text{Costruisce } M]{} 1.$$

- 1. \implies 2.

Ipotesi: L è un linguaggio riconosciuto da un DFA $M = \langle Q, \Sigma, \delta, q_0, F \rangle$
 $L = L(M)$.

Tesi: Esiste una relazione di equivalenza R invariante destra e di indice finito tale che L è unione di classi di equivalenza di R .

Prendiamo $R = R_M \quad x R_M y \iff \delta(q_0, x) = \delta(q_0, y)$.

1. Il numero delle classi di equivalenza di R_M è uguale al numero di stati di M : $|Q|$ è finito, quindi R_M è di indice finito
2. R_M è invariante destra

$$\begin{aligned} L = L(M) &= \{x \in \Sigma^* \mid \hat{\delta}(q_0, x) \in F\} \\ &= \{x \in \Sigma^* \mid \hat{\delta}(q_0, x) = q \wedge q \in F\} \\ &= \bigcup_{q \in F} \{x \in \Sigma^* \mid \hat{\delta}(q_0, x) = q\} \\ &= \bigcup_{q \in F} L_q \end{aligned}$$

Quindi L è unione di classi di equivalenza di R_M .

- 2. \implies 3.

Ipotesi: L è unione di classi di equivalenza di una relazione di equivalenza R invariante destra e di indice finito.

Tesi: R_L è di indice finito (numero di classi finito).

Dimostriamo che R è raffinamento di R_L perchè allora il numero di classi di equivalenza di R (finito per ipotesi) sarebbe maggiore del numero di classi di R_L (che quindi sarebbe finito).

Per dimostrare R raffinamento di R_L bisogna dimostrare che se due oggetti sono in relazione secondo la relazione più fine R , lo sono

anche secondo la relazione più grossa R_L :

$$\begin{aligned}\forall x, y . xRy &\implies xR_L y \\ &\equiv [y \in [x]_R \implies y \in [x]_{R_L}] \\ &\equiv [x]_R \subseteq [x]_{R_L} \text{ per raffinamento}\end{aligned}$$

Prendiamo xRy sapendo che R è invariante destra, cioè:

$$\forall z \in \Sigma^* . xRy \implies xzRyz$$

$$L = \cup \text{ classi di equivalenza}$$

Questo implica che o entrambe x e y appartengono a L o entrambe non appartengono a L :

$$xRy \implies x \in L \iff y \in L$$

$$[x]_R \subseteq L \quad \vee \quad [x]_R \text{ fuori da } L$$

$$\begin{aligned}xRy &\implies x \in L \iff y \in L \\ &\xRightarrow{\text{Invariante destra}} \forall z . xzRyz \xRightarrow{L = \cup \text{ classi}} xz \in L \iff yz \in L \\ &\xRightarrow{\text{Def. di } R_L} xR_L y\end{aligned}$$

Questo dimostra che dato un M generico R_M è un raffinamento di R_L .

- 3. \implies 1. **Ipotesi:** R_L è di indice finito.

Tesi: L è regolare, cioè è riconosciuto da un DFA M : $L = L(M)$.

Costruiamo M :

- $Q = \{[x]_{R_L} \mid x \in \Sigma^*\}$ è l'insieme delle classi di equivalenza di R_L (sono finite per ipotesi)
- Σ è l'alfabeto del linguaggio L
- $q_0 = [\epsilon]_{R_L}$ è la classe di equivalenza della stringa vuota
- $F = \{[x]_{R_L} \mid x \in L\}$ è l'insieme delle classi di equivalenza che contengono almeno una stringa appartenente a L
- $\delta(q, a) = \delta([x]_{R_L}, a) = [xa]_{R_L}$. Si potrebbe anche prendere un elemento qualsiasi $y \in [x]_{R_L}$ e definire:

$$yRx \implies [x]_{R_L} = [y]_{R_L} \implies [ya]_{R_L} = [xa]_{R_L}$$

e questo vale perchè R_L è invariante destra. Quindi la definizione di δ è una buona definizione perchè è indipendente dall'elemento che rappresenta la classe di equivalenza.

Bisogna dimostrare che $L = L(M)$:

– Dimostrazione per induzione che $\hat{\delta}([x], y) = [xy]$

$$\begin{aligned}\hat{\delta}(q_0, x) &= \hat{\delta}([ε]_{R_L}, x) = [x]_{R_L} \\ \implies x \in L(M) &\iff \hat{\delta}(q_0, x) \in F \\ &\iff \hat{\delta}([ε]_{R_L}, x) \in F \\ &\iff [x]_{R_L} \in F \\ &\iff x \in L \implies L(M) = L\end{aligned}$$

Dato L esiste un DFA M tale che $L = L(M)$ ed M ha il numero minimo di stati. (Quello costruito con le classi di equivalenza di R_L)

4.1.4 Condizione necessaria perchè un linguaggio sia regolare

Un linguaggio L è regolare se esiste un automa, quindi per dimostrare che un linguaggio non è regolare si può dimostrare che non esiste un automa. Per fare ciò si usa il Pumping lemma che fornisce una condizione Π_L **necessaria** alla regolarità di un linguaggio:

$$L \text{ regolare} \Rightarrow \Pi_L \quad \equiv \quad \neg \Pi_L \Rightarrow L \text{ non regolare}$$

Teorema 4.6 (Pumping lemma per linguaggi regolari). Consideriamo un linguaggio regolare L , allora esiste una costante $k \in \mathbb{N}$ tale che per ogni stringa di lunghezza maggiore di k nel linguaggio, esiste una suddivisione della stringa in tre parti u, v, w tale che:

$$\begin{aligned}\exists k \in \mathbb{N} . \forall z \in L : |z| \geq k \\ \Downarrow \\ \exists u, v, w \in \Sigma^* . z = uvw \wedge \begin{cases} |uv| \leq k \\ |v| > 0 \\ \forall i \in \mathbb{N} . uv^i w \in L \end{cases}\end{aligned}$$

cioè la parte v può essere "pompatà" (ripetuta i volte) e la stringa risultante appartiene ancora a L .

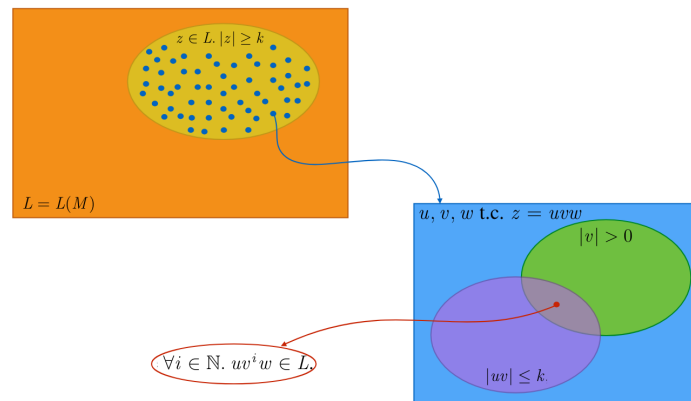


Figura 14: Rappresentazione grafica dell'insieme per il Pumping lemma

Dimostrazione:

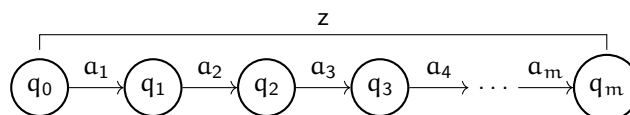
Consideriamo L regolare, allora esiste un automa DFA M tale che $L = L(M)$

$$M = \langle Q, \Sigma, \delta, q_0, F \rangle \quad |Q| = n \in \mathbb{N} \text{ stati finiti}$$

Vogliamo dimostrare che il lemma vale per $k = n$ (k è $|Q|$). Prendiamo una stringa nel linguaggio $z \in L$ tal che la sua lunghezza sia maggiore o uguale di k : $|z| \geq k$. Scriviamo z come insieme di caratteri:

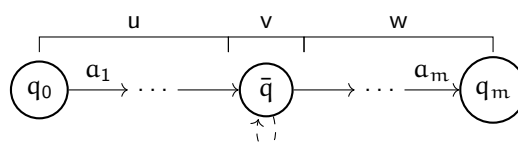
$$z = a_1 a_2 a_3 \dots a_m \quad m \geq k$$

Rappresentiamo l'elaborazione di z come una sequenza di stati:



Per leggere m simboli si attraversano $m + 1$ stati e quindi per **riconoscere** z si usano $m + 1$ stati, ma $m \geq n$, quindi si attraversano almeno $n + 1$ stati $\Rightarrow m + 1 \geq n + 1$. Questo implica che si attraversano più stati di quelli in Q , ovvero **almeno** uno stato è ripetuto nel riconoscimento di z .

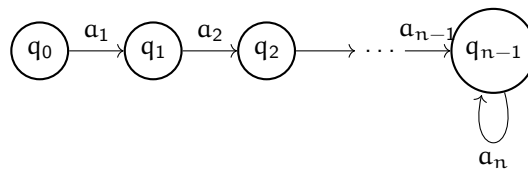
Supponiamo che $\bar{q} \in Q$ sia il primo stato (leggendo z) che viene **ripetuto**, in cui si torna per riconoscere z :



- u va da q_0 alla prima occorrenza di \bar{q} (il primo stato ripetuto)
- v va da \bar{q} a \bar{q} (percorso che porta a ritrovare lo stato ripetuto)
- w va da \bar{q} allo stato finale q_m

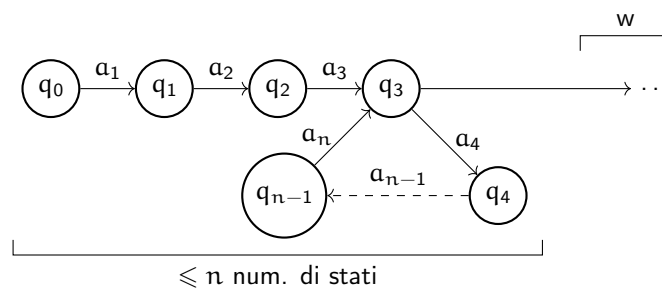
Questa suddivisione è accettabile per il teorema?

1. $|uv| \leq k = n$



Per arrivare allo stato finale q_n si attraversano $n - 1$ simboli

$$Q = \{q_0, q_1, \dots, q_{n-1}\} \rightarrow |Q| = n$$



2. $|v| > 0$ almeno un arco deve essere uscire dallo stato e ritornare nello stesso stato
3. uw arrivano in q_m , quindi:

$$\forall i. uv^i w \in L$$

e raggiunge lo stato finale q_m

4.1.5 Dimostrazione che un linguaggio non è regolare

Per dimostrare che un linguaggio non è regolare si può usare il pumping lemma:

$$L \text{ regolare} \Rightarrow \exists k. \forall z \in L : |z| \geq k$$

\Downarrow

$$\exists z = uvw. \begin{cases} |uv| \leq k \\ |v| > 0 \\ \forall i \in \mathbb{N}. uv^i w \in L \end{cases}$$

Per fare ciò si nega il pumping lemma:

$$A \rightsquigarrow B$$

$$\neg B \rightsquigarrow \neg A$$

$$\exists \rightsquigarrow \forall$$

Il pumping lemma si nega come segue:

- $\exists k \rightsquigarrow \forall K$: La dimostrazione **non** deve imporre vincoli su k
- $\forall z \rightsquigarrow \exists z \in L . |z| \geq k$: costruiamo noi la $z \in L$ di lunghezza $\geq k$
- $\exists uvw = z \rightsquigarrow \forall uvw = z: |v| > 0 \mid uv \leq k$
- $\forall i \in \mathbb{N} \rightsquigarrow \exists i \in \mathbb{N} . uv^i w \notin L$: troviamo un i che "rompe" la stringa

Quindi il pumping lemma negato diventa:

$$\forall k \in \mathbb{N} . \exists z \in L : |z| \geq k$$

\Downarrow

$$\forall uvw = z . \begin{cases} |uv| \leq k \\ |v| > 0 \\ \exists i \in \mathbb{N} . uv^i w \notin L \end{cases}$$

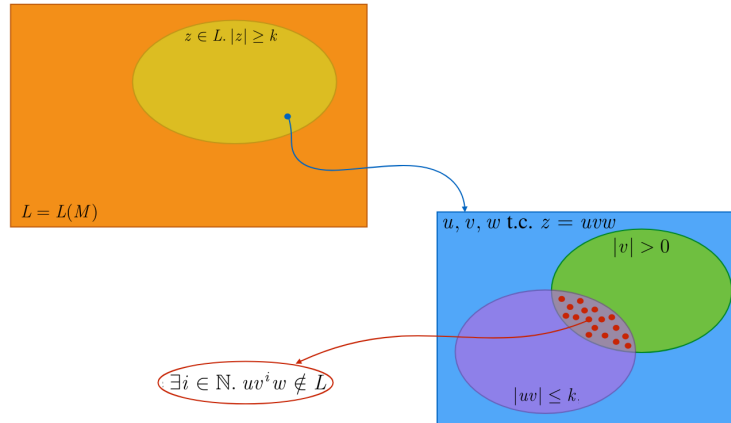


Figura 15: Rappresentazione grafica dell'insieme per il Pumping lemma negato

Esempio 4.3. Consideriamo il linguaggio:

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

Bisogna creare un automa che riconosca tutte le stringhe, ad esempio:

$$\varepsilon, 01, 0011, 000111, 00001111, \dots$$

Scriviamo le condizioni di appartenenza a L:

$$0^a 1^b \in L \iff a = b$$

(gli 0 devono essere uguali agli 1).

Ogni volta che si legge un 1 bisogna ricordarsi quanti 0 sono stati letti prima, però si possono leggere infiniti 0 e quindi servirebbero infiniti stati. Vogliamo quindi dimostrare che L non è regolare usando il pumping lemma negato:

$$\forall k \in \mathbb{N} . \exists z \in L : |z| \geq k$$

\Downarrow

$$\forall uvw = z . \begin{cases} |uv| \leq k \\ |v| > 0 \\ \exists i \in \mathbb{N} . uv^i w \notin L \end{cases}$$

Fissiamo $k \in \mathbb{N}$ (k non deve avere nessun vincolo). Qualunque sia k prendiamo la stringa:

$$z = 0^k 1^k \in L \quad |z| \geq k$$

Le uniche suddivisioni che vanno bene sono quelle in cui uv stanno nella parte degli 0 (perchè per ipotesi $|uv| \leq k$). Si può concludere che la sottostringa uv è composta da soli 0:

$$uv \in 0^k$$

Consideriamo la stringa:

$$z_i = uv^i w = 0^{k+(i-1)|v|} 1^k$$

(ripetere v i volte equivale ad aggiungere $(i-1)|v|$ volte v a quella già esistente). Ad esempio: Supponendo di avere 10 zeri un esempio di suddivisione è il seguente

$$\underbrace{000000}_u \underbrace{000}_v \underbrace{0111111111}_w$$

- Con $i = 0$ (togliere v)

$$0^{k+(0-1)|v|} 1^k = 0^{k-|v|} 1^k$$

- Con $i = 1$ (lasciare v una volta)

$$0^{k+(1-1)|v|} 1^k = 0^k 1^k$$

- Con $i = 3$ (ripetere v tre volte)

$$0^{k+(3-1)|v|} 1^k = 0^{k+2|v|} 1^k$$

Troviamo un i tale che $z_i \notin L$:

$$z_i = 0^{k+(i-1)|v|}1^k$$

Scegliamo ad esempio $i = 2$:

$$\begin{aligned} z_2 = 0^{k+|v|}1^k \in L &\iff \underbrace{k+|v|}_a = \underbrace{k}_b \\ &\iff |v| = 0 \text{ che è assurdo perchè } |v| > 0 \implies z_2 \notin L \end{aligned}$$

5 Linguaggi context free

I linguaggi context free sono più potenti dei linguaggi regolari, mantenendo comunque la decidibilità di molti problemi. Lo strumento che si utilizza per **definire** i linguaggi context free sono le **grammatiche**. Rispetto ai linguaggi regolari in cui l'automa era un riconoscitore, per i linguaggi context free la grammatica è un generatore. Un esempio di linguaggi context free sono i linguaggi di programmazione.

5.1 Grammatiche context free

Una grammatica context free è una quadrupla $E = \langle V, T, P, S \rangle$ dove:

- V è un insieme **finito** di simboli **non terminali** (variabili). Il loro ruolo è quello di rappresentare una categoria sintattica (ad esempio: espressione, istruzione, comando, ecc.)
- T è un insieme finito di simboli **terminali**. Sono i simboli effettivi che possono essere sostituiti a quelli non terminali. Ad esempio:

$$\underbrace{x}_{\text{terminale}} = \underbrace{\text{espressione}}_{\text{non terminale}}$$

- P è un insieme finito di **produzioni** (regole di derivazione). Indicano come sostituire i simboli non terminali con stringhe di simboli.
- $S \in V$ è il simbolo iniziale
- il **tipo di produzioni** (come possono essere fatte) determina il tipo di grammatica. In una **grammatica context free** le produzioni sono della forma:

$$A \rightarrow \alpha$$

dove $A \in V$ (non terminali) e $\alpha \in (V \cup T)^*$ (sequenza di simboli terminali e non terminali). Il vincolo principale è il fatto che a sinistra dell'implicazione ci sia esattamente **una sola variabile**. Questa implicazione significa che A può essere sostituito con α , indipendentemente dal contesto in cui A si trova, cioè dai simboli presenti prima e dopo A .

Esempio 5.1. Un esempio di grammatica sono le **espressioni booleane**. Consideriamo la grammatica: $G = \langle V, T, P, S \rangle$ dove:

$$V = \{E\}$$

$$T = \{0, 1, \text{and}, \text{or}, \text{not}, (,)\}$$

$$P = \begin{cases} E \rightarrow 0 \mid 1 \equiv E \rightarrow 0 E \rightarrow 1 \\ E \rightarrow (E \text{ and } E) \\ E \rightarrow (E \text{ or } E) \\ E \rightarrow \text{not } E \end{cases} = E \rightarrow 0 \mid 1 \mid (E \text{ and } E) \mid (E \text{ or } E) \mid (\text{not } E)$$

$$S = E$$

(\mid è l'operatore or). Una grammatica può essere scritta anche solamente con l'insieme delle sue produzioni.

Il **linguaggio generato** dalla grammatica G è una combinazione di produzioni.

Ad esempio:

$$A \rightarrow \beta \in P \quad \alpha, \gamma \in (V \cup T)^*$$

dove:

$$A \in V$$

$$\beta \in (V \cup T)^*$$

Si deriva la stringa a destra applicando la produzione $A \rightarrow \beta$

$$\alpha A \gamma \Rightarrow \alpha \beta \gamma$$

Si è generata una nuova stringa sostituendo A con β .

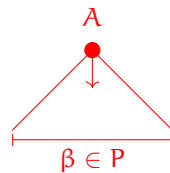


Figura 16: Rappresentazione grafica della produzione $A \rightarrow \beta$

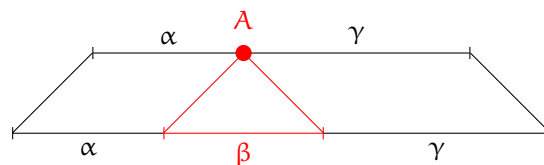


Figura 17: Rappresentazione grafica della derivazione $\alpha A \gamma \Rightarrow \alpha \beta \gamma$

La singola freccia (\rightarrow) indica una produzione, mentre la doppia freccia (\Rightarrow) indica una derivazione (applicazione di una produzione). Altri esempi di derivazioni:

- $E \rightarrow (E \text{ or } E) \equiv E \Rightarrow (E \text{ or } E)$
- $E \rightarrow (E \text{ and } E) \equiv \underbrace{(\underbrace{E}_{\alpha} \text{ or } E)}_{\gamma} \Rightarrow \underbrace{(\underbrace{(E \text{ and } E)}_{\beta} \text{ or } E)}_{\gamma}$

Definizione 5.1 (Derivazione). Partendo da una sequenza $\alpha \in V \cup T$ una derivazione consiste nell'applicazione di una produzione ad uno dei simboli non terminali presenti nella sequenza α . Considerando la stringa:

$$\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n \in (V \cup T)^*$$

e

$$\forall j \in [1, n-1] . \alpha_j \Rightarrow \alpha_{j+1}$$

allora si può scrivere l'applicazione di n produzioni come:

$$\alpha_1 \Rightarrow_n \alpha_n$$

Ad esempio:

$$\alpha_i \Rightarrow_0 \alpha_j \quad \text{Nessuna produzione applicata}$$

Si può scrivere la chiusura transitiva di \Rightarrow_n come:

$$\alpha \Rightarrow^* \beta \text{ se } \exists i . \alpha \Rightarrow_i \beta$$

Cioè esiste un numero di passi i tale che applicando i produzioni si passa da α a β . (Da α si deriva β)

Definizione 5.2 (Linguaggio generato). Il linguaggio L è generato dalla grammatica $G : L(G)$ con $G = \langle V, T, P, S \rangle$ è l'insieme delle sequenze di terminali generate da S , ovvero per le quali esiste una derivazione a partire da S :

$$L(G) = \left\{ \sigma \in T^* \mid S \xRightarrow{G}^* \sigma \right\}$$

Il linguaggio $L \subseteq T^*$ è **context free** se esiste una grammatica G context free che lo genera, ovvero tale che $L = L(G)$. Quindi per dimostrare che un linguaggio è context free basta costruire una grammatica che lo genera.

5.1.1 Dimostrazione canonica che un linguaggio è context free

Definizione 5.3 (Dimostrazione che un linguaggio è context free). Possiamo dimostrare che $L = L(G)$, quindi L è context free. Dimostriamo:

$$1. L \subseteq L(G)$$

Se σ appartiene al linguaggio, allora da S si può derivare σ :

$$L \subseteq L(G) \equiv \sigma \in L \implies \sigma \in L(G) \equiv S \Rightarrow^* \sigma$$

Dimostriamo per induzione sulla lunghezza di σ .

- **Caso base:** $|\sigma| =$ lunghezza minima in L
- **Passo induttivo:** Si suppone che per ogni σ di lunghezza n valga l'ipotesi induttiva, e si dimostra che vale anche per σ di lunghezza maggiore di n .

$\forall \sigma . |\sigma| = n$ vale l'ipotesi induttiva 1.

$$2. L(G) \subseteq L$$

Se σ appartiene al linguaggio generato dalla grammatica, allora σ appartiene al linguaggio:

$$L(G) \subseteq L \equiv \sigma \in L(G) \implies \sigma \in L \equiv S \Rightarrow^* \sigma$$

Dimostriamo per induzione sulla lunghezza della derivazione:

$$S \Rightarrow^* \sigma \equiv \exists i . S \Rightarrow_i \sigma$$

quindi l'induzione è su i .

- **Caso base:** La derivazione più corta possibile di simboli terminali
- **Passo induttivo:** Si suppone che per ogni $i \leq n$ se la S deriva in i passi in una stringa di soli terminali σ allora $\sigma \in L$:

$$\forall i \leq n . S \Rightarrow_i \sigma \text{ allora } \sigma \in L$$

Dimostriamo che $S \Rightarrow_{n+1} \sigma$ allora $\sigma \in L$.

Esempio 5.2. Consideriamo il seguente linguaggio:

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

Vogliamo dimostrare che L è context free. Per farlo bisogna costruire una grammatica G che genera L :

- S: Bisogna chiedersi se la stringa vuota appartiene a L.

1. $\varepsilon \in L$? Sì, per $n = 0$ $0^0 1^0 = \varepsilon$. ε va sempre generata dalla grammatica.

Quindi S diventa:

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

La grammatica di questo linguaggio quindi è:

$$G : S \rightarrow \varepsilon \mid 0S1 \text{ (grammatica context free)}$$

Dimostriamo: $L = L(G)$

1. $L \subseteq L(G)$ per induzione sulla lunghezza di $\sigma \in L$

- Tesi: $\sigma \in L$ allora $S \Rightarrow^* \sigma$
- **Base:** $n = 0$ $|\sigma| = 0$ è tale che $\sigma = \varepsilon$ e $\sigma \rightarrow \varepsilon$ ($S \Rightarrow \varepsilon$)
- **Passo induttivo:** L'ipotesi induttiva è la tesi, ma limitata a n:

$$\forall i \leq n. |\sigma| = i \quad \sigma \in L \text{ allora } S \Rightarrow^* \sigma$$

Dimostriamo che $|\sigma| = n$ e $\sigma \in L$ allora $S \Rightarrow^* \sigma$. Prendiamo $\sigma \in L$, ovvero:

$$\exists j. \sigma = 0^j 1^j \quad 2j = n$$

Bisogna trovare la derivazione. Possiamo osservare che:

$$\sigma = 0 \underbrace{0^{j-1} 1^{j-1}}_{2j-2 < n} 1$$

(abbiamo separato il primo 0 e l'ultimo 1 dalla sequenza centrale). σ può essere scritta come:

$$\sigma = 0 \underbrace{0^{j-1} 1^{j-1}}_{2j-2 < n} 1 = 0\sigma'1 \quad \text{con } \sigma' = 0^{j-1} 1^{j-1} \in L$$

Siccome $|\sigma'| < n$ e $\sigma' \in L$ allora vale l'ipotesi induttiva e quindi:

$$\Rightarrow \exists S \Rightarrow^* \sigma' \equiv 0^{j-1} 1^{j-1}$$

usiamo questa derivazione per costruire la derivazione di σ . Questa derivazione deve essere fatta in modo da avere come ultimo passo $S \Rightarrow \varepsilon$ per ottenere σ' :

$$S \Rightarrow^* 0^{j-1} 1^{j-1} \equiv S \Rightarrow 0^{j-1} S 1^{j-1} \Rightarrow 0^{j-1} \varepsilon 1^{j-1} = 0^{j-1} 1^{j-1}$$

Per arrivare a $\sigma = 0^j 1^j$ bisogna sostituire la S con $0S1$:

$$S \Rightarrow^* 0^j S 1^j \Rightarrow 0^{j-1} 0 S 1 1^{j-1} \Rightarrow 0^{j-1} 0 \epsilon 1 1^{j-1} = 0^j 1^j = \sigma \quad \square$$

Questo conclude la dimostrazione del primo punto, cioè trovare una derivazione per σ . Quindi abbiamo dimostrato che $L \subseteq L(G)$.

2. $L(G) \subseteq L$ Dimostriamo per induzione sulla lunghezza delle derivazioni

- **Tesi:** $S \Rightarrow^* \sigma$ allora $\sigma \in L$
- **Base:** $S \Rightarrow \epsilon$ ($S \rightarrow \epsilon$) e $\epsilon \in L$
- **Passo induttivo:** L'ipotesi induttiva è la tesi limitata a i :

$$\forall i \leq n. S \Rightarrow_i \sigma \text{ allora } \sigma \in L$$

Prendiamo una stringa generata in n passi:

$$S \Rightarrow_n \sigma \quad \equiv \quad S \Rightarrow_{n-1} \sigma_1 S \sigma_2 \Rightarrow \sigma_1 \sigma_2 = \sigma$$

Per come è fatta la grammatica σ_1 contiene solo 0 e σ_2 contiene solo 1. Ovvero:

$$\sigma_1 = 0^j \quad \sigma_2 = 1^k$$

Bisogna riuscire ad ottenere una derivazione più corta di n simboli terminali in modo tale da poter applicare l'ipotesi induttiva. Un solo passo indietro non basta perchè non si ha una derivazione di tutti simboli terminali, quindi si può andare ancora più indietro:

$$S \Rightarrow_{n-1} \sigma_1 S \sigma_2 \quad \equiv \quad S \Rightarrow_{n-2} \sigma'_1 \textcolor{red}{S} \sigma'_2 \Rightarrow \underbrace{\sigma'_1 0}_{\sigma_1} \textcolor{red}{S} \underbrace{1 \sigma'_2}_{\sigma_2}$$

Scorporiamo la derivazione di $n - 2$ passi in:

$$S \Rightarrow_{n-2} \sigma'_1 S \sigma'_2 \Rightarrow \sigma'_1 \sigma'_2$$

abbiamo una derivazione lunga $n - 1$ di simboli terminali dove:

$$\begin{aligned} \sigma_1 = \sigma'_1 0 &\implies \sigma'_1 = 0^{h-1} \\ \sigma_2 = 1 \sigma'_2 &\implies \sigma'_2 = 1^{k-1} \end{aligned}$$

Per ipotesi induttiva:

$$\xRightarrow{I.I.} \sigma'_1 \sigma'_2 \in L$$

Quindi:

$$\sigma'_1 \sigma'_2 = 0^{h-1} 1^{k-1} \in L \implies h - 1 = k - 1 \implies h = k$$

ma allora $\sigma = \sigma_1 \sigma_2 = 0^h 1^k$ è tale che $h = k$ e quindi $\sigma \in L$

Ricapitolando:

- Si parte da $S \Rightarrow_n 0^h 1^k$, ma non si può applicare l'ipotesi induttiva, quindi si va un passo indietro
- $S \Rightarrow_{n-1} 0^h S 1^k$, ma ancora non si può applicare l'ipotesi induttiva
- Si va ancora un passo indietro:

$$S \Rightarrow_{n-2} 0^{h-1} S 1^{k-1} \Rightarrow 0^{h-1} 0 S 1 1^{k-1} \Rightarrow 0^h 1^k$$

questo si può usare per generare $0^{h-1} 1^{k-1}$:

$$S \Rightarrow_{n-2} 0^{h-1} S 1^{k-1} \Rightarrow 0^{h-1} 1^{k-1}$$

che è lunga $n-1$ e quindi si può applicare l'ipotesi induttiva.

- Quindi $0^{h-1} 1^{k-1} \in L \implies h-1 = k-1 \implies h = k$ e quindi $0^h 1^k \in L \quad \square$

Siamo andati indietro tanto quanto serviva per generare una stringa di soli simboli terminali.

5.1.2 Dimostrazione alternativa

Esempio 5.3. Considerando l'esempio precedente, una dimostrazione alternativa che un linguaggio è context free consiste nel partire dall'inizio e fare passi avanti piuttosto che partire dalla fine e fare passi indietro. Questo non funziona sempre, ma in questo caso sì:

$$L(G) \subseteq L$$

$$\forall i < n . S \Rightarrow_i \sigma \text{ allora } \sigma \in L$$

• **Passo induttivo:**

$$S \Rightarrow_n \sigma \equiv S \Rightarrow 0 S 1 \Rightarrow_{n-1} \sigma \equiv 0 \sigma' 1$$

dove $S \Rightarrow_{n-1} \sigma'$. Ma allora si può applicare l'ipotesi induttiva e quindi $\sigma' \in L$ e questo significa che:

$$\exists j . \sigma' = 0^j 1^j$$

questo implica che:

$$\sigma = 0 \sigma' 1 = 0 0^j 1^j 1 = 0^{j+1} 1^{j+1} \in L$$

5.2 Alberi di derivazione

Data una grammatica context free $G = \langle V, T, P, S \rangle$, un albero di derivazione (parse tree) per G ha le seguenti caratteristiche:

1. Ogni nodo ha un'etichetta che è un simbolo di $V \cup T \cup \{\epsilon\}$
2. L'etichetta della radice è un simbolo in V

3. Ogni nodo interno (non foglia) ha etichetta in V
4. Se un nodo n è etichettato con $A \in V$ e i nodi figli n_1, n_2, \dots, n_k sono etichettati con $X_1, X_2, \dots, X_k \in V \cup T \cup \{\varepsilon\}$ allora esiste una produzione che genera un arco per ogni simbolo figlio:

$$A \rightarrow X_1 X_2 \dots X_k \in P$$

5. Se un nodo ha etichetta ε allora è una foglia ed è l'unico figlio del padre

Esempio 5.4. Consideriamo la grammatica

$$E \rightarrow 0 \mid 1 \mid (E \text{ and } E) \mid (E \text{ or } E) \mid (\text{not } E)$$

Consideriamo la derivazione:

$$E \Rightarrow^* ((0 \text{ or } 1) \text{ and } (\text{not } 0))$$

L'albero di derivazione è il seguente:

Teorema 5.1. Sia $G = \langle V, T, P, S \rangle$ una grammatica context free, allora:

$$S \Rightarrow^* \alpha \in T^*$$

se e solo se esiste un albero di derivazione con radice etichettata S e foglie etichettate (leggendo da sinistra verso destra) con i simboli di α .

5.3 Ambiguità delle grammatiche

Esempio 5.5. Consideriamo una grammatica per le espressioni aritmetiche senza parentesi:

$$E \rightarrow E + E \mid E * E \mid 0 \mid 1 \mid 2$$

Consideriamo la stringa:

$$2 * 0 + 1$$

Si hanno due casi:

- Si esegue prima la moltiplicazione:

$$2 * 0 + 1 = (2 * 0) + 1 = 1$$

- Si esegue prima l'addizione:

$$2 * 0 + 1 = 2 * (0 + 1) = 2$$

Notiamo che si ottengono due risultati diversi, quindi la grammatica è ambigua perchè la stessa stringa può essere generata da due alberi di derivazione distinti:

- Albero di derivazione con precedenza alla moltiplicazione
- Albero di derivazione con precedenza all'addizione

Definizione 5.4 (Ambiguità). Una grammatica è **ambigua** se esiste almeno una parola α con più di un albero di derivazione con radice S .

Il linguaggio generato da una grammatica ambigua è detto **inerentemente ambiguo**.

5.4 Forma normale

Le forme normali sono grammatiche le cui produzioni rispettano vincoli di "forma" specifici. Ci sono due forme normali importanti per le grammatiche context free:

- **Forma normale di Chomsky:** Tutte le produzioni hanno la forma:

$$A \rightarrow a \quad \vee \quad A \rightarrow BC$$

dove:

$$A, B, C \in V \quad a \in T$$

Questa forma genera sempre un albero binario (ogni nodo ha al massimo due figli).

- **Forma normale di Greibach:** Tutte le produzioni hanno la forma:

$$A \rightarrow a\alpha$$

dove:

$$A \in V \quad a \in T \quad \alpha \in V^*$$

Rappresenta l'automa a pila in cui α è il contenuto della pila e a è il simbolo da aggiungere.

5.5 Minimizzazione delle grammatiche context free

Ogni grammatica context free può essere riscritta in modo tale che:

1. **Eliminazione dei simboli inutili:** Ogni simbolo non terminale genera simboli terminali e ogni simbolo terminale è generato da almeno un simbolo non terminale
2. **Eliminazione delle produzioni unitarie:** Nessuna produzione è della forma:

$$A \rightarrow B \quad A, B \in V$$

perchè A è generabile da B e B è generabile da A .

3. **Eliminazione della ϵ -produzione:** Se $\epsilon \notin L$ allora **non** ci deve essere la produzione:

$$A \rightarrow \epsilon \quad A \in V$$

Definizione 5.5. Quando la grammatica è senza simboli inutili, senza produzioni unitarie e senza ϵ -produzioni, si può trasformare in forma normale di Chomsky.

5.5.1 Eliminazione dei simboli inutili

$x \in V \cup T$ è utile se esiste una derivazione

$$S \Rightarrow^* \underbrace{\alpha x \beta}_{1.a} \Rightarrow^* w \in T^*$$

Teorema 5.2. Per ogni grammatica context free $\forall G = \langle V, T, P, S \rangle$ esiste una grammatica equivalente $G' = \langle V', T, P', S \rangle$ senza simboli inutili:

$$\forall G . \exists G' . L(G) = L(G') \text{ e } G' \text{ non ha simboli inutili}$$

G' si ottiene applicando in sequenza le due trasformazioni viste (1, 2) (nell'ordine visto)

1. Eliminiamo i simboli che non generano terminali:

Esempio 5.6. Consideriamo le produzioni $S \rightarrow \alpha AB$ e $A \rightarrow A$. Il simbolo A è inutile perchè non genera mai simboli terminali.

Metodo algoritmico per eliminare i simboli che non generano terminali:

(a) Dopo la trasformazione

$$\forall A \in V . \exists w \in T^* . A \Rightarrow^* w$$

Allora $L(G) = \emptyset$ tutti i simboli sono inutili

(b) Prendiamo per ipotesi che $L(G) \neq \emptyset$, definiamo:

$$\Gamma(W) = \{A \in V \mid \exists \alpha \in (T \cup W)^* . A \rightarrow \alpha \in P\}$$

$$\begin{cases} \Gamma^0(W) = W \\ \Gamma^{i+1}(W) = \Gamma(\Gamma^i(W)) \end{cases}$$

Esempio 5.7. Consideriamo le seguenti produzioni:

$$\begin{aligned} A &\rightarrow A \\ S &\rightarrow aS \mid B \\ B &\rightarrow b \end{aligned}$$

Si parte con $W = \emptyset$:

- $\Gamma^0(\emptyset) = \{A \in V \mid \exists \alpha \in T^* . A \rightarrow \alpha \in P\} = \emptyset$
- $\Gamma^1(\emptyset) = \Gamma(\emptyset) = \{B\}$ perchè $B \rightarrow b \in P$

- $\Gamma^2(\emptyset) = \Gamma(\{B\}) = \{A \in V \mid \exists \alpha \in (T \cup \{B\})^* . A \rightarrow \alpha\}$
 $= \{B, S\}$
 - $\Gamma^3(\emptyset) = \Gamma(\{B, S\})$
 $= \{A \in V \mid \exists \alpha \in (T \cup \{B, S\})^* . A \rightarrow \alpha\}$
 $= \{B, S\}$
- Si è raggiunto il **punto fisso**, cioè non si ottengono più simboli nuovi.

Alla fine si ottiene l'insieme di tutti i simboli utili, cioè quelli che generano simboli terminali. Tutti gli altri (e le corrispondenti produzioni) vengono eliminati.

Quindi $S \rightarrow aS \mid B \quad B \rightarrow b$ è equivalente ma senza simboli inutili che non generano nulla

2. Eliminiamo i simboli non raggiungibili da S:

$$\Gamma(W) = \{X \in V \cup T \mid \exists A \in W . A \rightarrow \alpha X \beta \in P\} \cup S$$

Quindi ad esempio:

$$\Gamma(\emptyset) = \{S\}$$

Cioè si aggiunge ciò che possiamo raggiungere da S in un certo numero di passi.

Esempio 5.8. Consideriamo le seguenti produzioni:

$$S \rightarrow a \mid bC$$

$$C \rightarrow d \mid dE$$

$$E \rightarrow \varepsilon$$

$$F \rightarrow f$$

Si parte con $W = \emptyset$:

- $\Gamma(\emptyset) = \{S\}$
- $\Gamma(\{S\}) = \{X \in V \cup T \mid S \rightarrow \alpha X \beta \in P\} \cup \{S\} = \{S, a, b, C\}$
- $\Gamma(\{S, a, b, C\}) = \{X \in V \cup T \mid C \rightarrow \alpha X \beta \in P \vee S \rightarrow \alpha X \beta \in P\} = \{S, a, b, C, d, E\}$
- $\Gamma(\{S, a, b, C, d, E\}) = \{S, a, b, C, d, E, \varepsilon\}$

Quindi F è un simbolo non raggiungibile da S e quindi è inutile.

5.5.2 Eliminazione delle produzioni unitarie

L'idea è quella di sostituire le produzioni unitarie con le produzioni dello stesso simbolo non terminale a cui puntano. Ad esempio se abbiamo $S \rightarrow A \rightarrow 0A0$ si può compattare S come: $S \rightarrow 0A0$

Esempio 5.9. Consideriamo la seguente grammatica:

$$\begin{aligned} S &\rightarrow \varepsilon \mid A \mid B \\ A &\rightarrow 0 \mid 0A0 \mid B \mid 00 \\ B &\rightarrow 1 \mid 1B1 \mid A \mid 11 \end{aligned}$$

Sostituiamo in tutte le produzioni quello che è producibile dal non terminale a destra:

$$\begin{aligned} S &\rightarrow \varepsilon \mid \underbrace{0 \mid 0A0 \mid 00}_A \mid \underbrace{1 \mid 1B1 \mid 11}_B \\ A &\rightarrow 0 \mid 0A0 \mid 00 \mid \underbrace{1 \mid 1B1 \mid 11}_B \\ B &\rightarrow 1 \mid 1B1 \mid 11 \mid \underbrace{0 \mid 0A0 \mid 00}_A \end{aligned}$$

5.5.3 Eliminazione delle ε -produzioni

L'idea è quella di sostituire la transizione con tutto quello che può generare ε ad esempio:

Esempio 5.10. Consideriamo le produzioni:

$$\begin{aligned} S &\rightarrow \varepsilon \mid A \mid B \\ A &\rightarrow 0 \mid 0A0 \mid B \\ B &\rightarrow 1 \mid 1B1 \mid A \mid \varepsilon \end{aligned}$$

Siccome sia S che B vanno in ε si possono sostituire con:

$$B \rightarrow S$$

Anche A con un certo numero di derivazioni arriva in ε , quindi tutti i simboli non terminali che in un certo numero di passi generano ε sono:

$$\{S, A, B\}$$

Le nuove produzioni diventano:

$$\begin{aligned} S &\rightarrow \varepsilon \mid A \mid B \\ A &\rightarrow 0 \mid 0A0 \mid 00 \mid B \\ B &\rightarrow 1 \mid 1B1 \mid 11 \mid A \end{aligned}$$

Eliminando $B \rightarrow \varepsilon$ le stringhe 00 e 11 non sarebbero più generabili e quindi si aggiungono

In questo modo si sono eliminate tutte le ε -produzioni, tranne $S \rightarrow \varepsilon$ se $\varepsilon \in L$.