Automi e Linguaggi Formali
Esame del 12 Luglio 2017 – Parte 1

Tempo a disposizione: 1 h 30 min

Gli esercizi (1, 2, 3) e (4, 5, 6) vanno consegnati su due fogli differenti

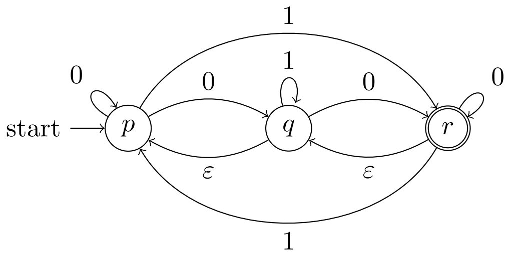
Linguaggi Regolari

1. Determinare il linguaggio definito dall'espressione regolare $\mathbf{0}^*\mathbf{1}(\mathbf{0}+\mathbf{1}\mathbf{0}^*\mathbf{1})^*$

   Soluzione. L'espressione regolare riconosce il linguaggio di tutte le stringhe sull'alfabeto $\{0,1\}$ che contengono un numero *dispari* di 1.

   L'espressione $(\mathbf{0}+\mathbf{1}\mathbf{0}^*\mathbf{1})^*$ genera tutte le stringhe con un numero pari di 1 (vedi Es. 1 dell'Esame del 28 Giugno), mentre $\mathbf{0}^*\mathbf{1}$ aggiunge un ulteriore 1 alla stringa, portando il numero di 1 ad essere dispari.

2. Dato il seguente ε-NFA

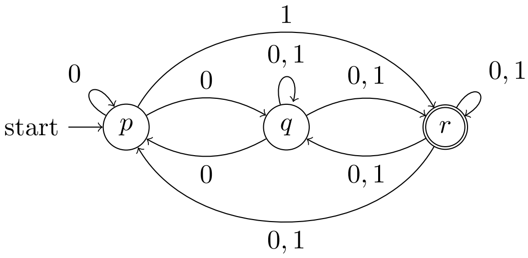


   (a) calcolare la ε-chiusura di ogni stato
   (b) costruire un DFA equivalente

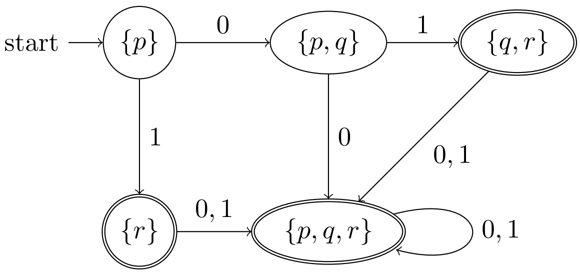
   Soluzione.

   (a) ECLOSE$(p) = \{p\}$ ECLOSE$(q) = \{p,q\}$ ECLOSE$(r) = \{p,q,r\}$

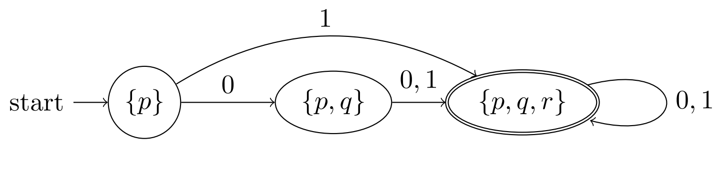
   (b) Per costruire un DFA equivalente si può procedere in due modi. Il primo modo utilizza due passaggi: prima si eliminano le ε-transizioni, ottenendo il seguente NFA (dove r rimane l'unico stato finale)



   e poi si usa la costruzione a sottoinsiemi per ottenere l'automa deterministico finale (dove gli stati non raggiungibili sono rimossi)



   Il secondo modo è quello di utilizzare la trasformazione diretta da ε-NFA a DFA descritta nel libro di testo (non vista a lezione), che porta al seguente DFA



3. Il linguaggio
$$L = \{w1w1w \mid w \in \{0, \ldots, 9\}^*\}$$
è regolare? Motivare la risposta.

Soluzione. Il linguaggio contiene tutte le parole formate da tre ripetizioni di una stringa w (di lunghezza arbitraria), separate dal simbolo 1. Intuitivamente, non può essere regolare perché per riconoscere le parole nel linguaggio occorre memorizzare tutti i simboli della prima occorrenza di w per poi controllare che si ripetano esattamente uguali nelle altre due ripetizioni.

Per dimostrarlo formalmente, assumiamo per assurdo che L sia regolare:

- sia $n > 0$ la lunghezza data dal Pumping Lemma;
- consideriamo la parola $v = 0^n10^n10^n$, che è di lunghezza maggiore di n ed è nella forma $w1w1w$ se poniamo $w = 0^n$. Quindi v appartiene ad L;
- sia $v = xyz$ una suddivisione arbitraria di v tale che $y \neq \varepsilon$ e $|xy| \leq n$;
- poiché $|xy| \leq n$, allora xy è completamente contenuta nella prima ripetizione di $w = 0^n$, e quindi sia x che y sono composte solo da 0. Inoltre, siccome $y \neq \varepsilon$, possiamo dire che $y = 0^p$ per qualche valore $p > 0$. Allora la parola xy^0z è nella forma $0^{n-p}10^n10^n$, e non può appartenere a L perché non esiste nessuna parola u che ci consenta di scrivere xy^0z come $u1u1u$.

Abbiamo trovato un assurdo quindi L non può essere regolare.

Note: L'esercizio è molto simile all'Es. 3 dell'Esame del 28 Giugno. Valgono le stesse considerazioni fatte nella soluzione pubblicata in precedenza.

Linguaggi Liberi dal Contesto

4. Data la seguente grammatica libera da contesto $G : B \to BB \mid (B) \mid \varepsilon$, rispondere alle seguenti due domande:

   (a) dimostrare, per induzione sulla lunghezza della derivazione, che L(G) consiste di stringhe in $\{(,)\}^*$ in cui le parentesi siano bilanciate, cioè tali che ogni parentesi aperta ha una corrispondente parentesi chiusa che la segue, e se la coppia di parentesi venisse eliminata, si otterrebbe di nuovo una stringa bilanciata.

   Soluzione: come richiesto esplicitamente dall'esercizio, usiamo un'induzione sulla lunghezza della derivazione. Quindi, consideriamo stringhe in $L(G)$ derivate da B in 1 passo, in $n > 1$ passi e in $n + 1$ passi.
   - **Base:** con 1 passo, la grammatica permette di derivare solo la stringa vuota ε che non avendo parentesi è una stringa bilanciata.
   - **Passo induttivo:** assumiamo come ipotesi induttiva che tutte le stringhe in $L(G)$ derivate in n passi (per qualsiasi $n \geq 1$) siano bilanciate. Da questo vogliamo dimostrare che anche le stringhe derivate in $n + 1$ passi sono bilanciate. Consideriamo quindi una qualsiasi derivazione di $n + 1$ passi: $B \Rightarrow^{(n+1)} w$. Il primo passo può essere l'applicazione della produzione $B \to (B)$ o $B \to BB$. Consideriamo i due casi separatamente:
     - $B \Rightarrow (B) \Rightarrow^n (w)$, dove $B \Rightarrow^n w$ è una derivazione di n passi e quindi ad essa si applica l'ipotesi induttiva che ci garantisce che w è bilanciata. Basta ora osservare che allora anche (w) è bilanciata, dove la seconda parentesi corrisponde alla prima e infatti eliminandole entrambe resteremmo con w che è bilanciata, come richiesto dalla definizione.
     - $B \Rightarrow BB \Rightarrow^n w'w''$, dove $B \Rightarrow^{n'} w'$ e $B \Rightarrow^{n''} w''$ con n' e $n'' > 0$ e $n' + n'' = n$, quindi, per ipotesi induttiva w' e w'' sono bilanciate, per cui la loro concatenazione, $w'w''$ è una stringa bilanciata.

   (b) Mostrare che la grammatica $G : B \to (B) \mid \varepsilon$ non genera tutte le stringhe in $\{(,)\}^*$ bilanciate.

Soluzione: basta osservare che la stringa bilanciata ()() non potrà in nessun caso essere generata senza la produzione $B \to BB$.

5. Dato l'automa a pila $P = (\{q, p\}, \{a, b\}, \{a, Z\}, \delta, q, Z, \{q\})$ dove δ è come segue:

$$\delta(q, a, Z) = \{(q, aZ)\}, \quad \delta(q, a, a) = \{(q, aa)\}, \quad \delta(q, b, a) = \{(q, \varepsilon)\}, \quad \delta(q, b, Z) = \{(p, \varepsilon)\},$$

rispondere alle seguenti due domande.

(a) Descrivere il linguaggio riconosciuto da P.

Soluzione: Si tratta del linguaggio costituito da tutte quelle stringhe in $\{a, b\}^*$ tali che in ogni prefisso il numero di a è almeno pari a quello dei b. Questo linguaggio compare negli ultimi tre esami di CFL.

(b) Trasformare P in un PDA P' che accetta per pila vuota lo stesso linguaggio accettato da P per stato finale.

Soluzione: Visto che P può vuotare la pila (per rifiutare un input in caso contenga più b di a) è necessario usare la costruzione standard per ottenere P' tale che $N(P') = L(P)$. Quindi avremo bisogno di uno stato iniziale q_1 con transizione: $\delta(q_1, \varepsilon, X) = \{(q, ZX)\}$, che inserisce sotto al simbolo Z di fondo stack di P, un proprio simbolo di fondo stack X che non è noto a P che quindi non ha mosse per X e perciò non potrà mai eliminare X dalla pila. Poi avremo bisogno di uno stato q_2 che ha il compito di svuotare la pila. Ricordate che q è l'unico stato finale di P, quindi avremo la mossa: $\delta(q, \varepsilon, ?) = \{(p_2, \varepsilon)\}$, e poi $\delta(q_2, \varepsilon, ?) = \{(q_2, \varepsilon)\}$, per svuotare la pila. Con ? si intende un qualsiasi simbolo della pila. Si ricorda che i simboli della pila di P' sono Z, X, e a.

6. Rispondere alle seguenti due domande che riguardano gli automi a pila deterministici.

(a) Come si dimostra che gli automi a pila deterministici riconoscono solo linguaggi non ambigui?

Soluzione: si osserva la costruzione generale di una grammatica CF che simula un PDA. La grammatica ha quei nonterminali $[q, X, p]$. Poiché la grammatica genera le stringhe che il PDA riconosce e la derivazione delle grammatica simula il calcolo del PDA, se il PDA è deterministico, esso ha un solo calcolo per ciascuna stringa accettata e anche la grammatica ha un'unica derivazione per ogni stringa generata. Insomma la grammatica è non ambigua.

(b) Qual è la differenza tra un linguaggio libero da contesto ambiguo ed una grammatica libera da contesto ambigua?

Soluzione: un linguaggio CF L è ambiguo (più precisamente, inerentemente ambiguo) quando ogni CFG che genera L è ambigua. Una grammatica invece è ambigua, se esistono due alberi di derivazione diversi della CFG che hanno uguale frontiera.