TEOREMA (PUMPING LEMMA)

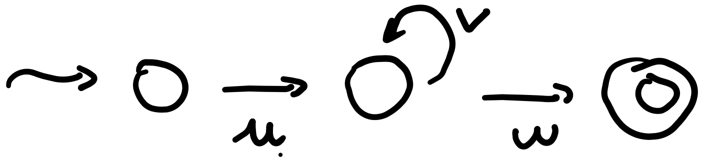
Per ogni linguaggio regolare L esiste una costante n tale che, se $z \in L$ e $|z| \geq n$, allora possiamo scrivere $z = uvw$ con $|uv| \leq n$, $|v| \geq 1$ e ottenere che $uv^i w \in L$ per ogni $i \geq 0$

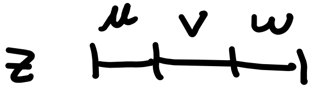
<u>OSSERVAZIONE</u>

QUESTO TEOREMA ci da una condizione necessaria ma non sufficiente per i linguaggi regolari.

Non posso utilizzarlo per dimostrare che un linguaggio L è regolare ma posso usarlo per dimostrare che un linguaggio L' non è regolare.

INFORMALMENTE: Se L è regolare esiste un certo numero n che ha queste caratteristiche, se ho una stringa z che appartiene a L lunga $\geq n$ la posso dividere in 3 parti lunghe u v w e la parte v la posso ripetere quante volte voglio ottenendo ancora stringhe del linguaggio.





$z \in \mathcal{L}(A) = L$

DIMOSTRAZIONE

DATO UN LINGUAGGIO L REGOLARE $\exists$ ASFD che lo riconosce.

Sia $\mathcal{A} = \langle \Sigma, Q, \delta, q_0, F \rangle$ un automa che riconosce L e sia $|Q| = n$

Sia $z \in L$ con $|z| = k \geq n \Rightarrow \bar{\delta}(q_0, z) \in F$

Supponiamo che $q_{i_0}, q_{i_1}, \dots, q_{i_k}$ sia la sequenza di stati attraversati da $\mathcal{A}$ durante la computazione z con $q_{i_0} = q_0$ e $q_{i_k} = \bar{\delta}(q_0, z) \in F$

In altre parole q_{i_1} è lo stato in cui arriva l'automa $\mathcal{A}$ dopo aver letto il primo simbolo di z, q_{i_2} lo stato in cui arriva l'automa $\mathcal{A}$ dopo aver letto il secondo simbolo di z ....

Indichiamo con z_n il prefisso di z di lunghezza n, allora avremo che q_{i_n} è lo stato in cui arriva l'automa $\mathcal{A}$ dopo aver letto z_n^{elementi} cioè $q_{i_n} = \bar{\delta}(q_0, z_n)$

DAL MOMENTO CHE $k \geq n$, deve esistere almeno uno stato in cui l'automa si porta almeno due volte durante la computazione su z, cioè esistono due stati nella sequenza $q_{q_i}, q_{1_i}, \dots, q_{i_k}$ che coincidono. In realtà questi due stati che coincidono si possono trovare già tra i primi $n+1$ elementi della sequenza $q_{q_i}, q_{1_i}, \dots, q_{i_k}$.

Possiamo affermare che

ESISTONO DUE INDICI r, s con $0 \leq r < s \leq n$ tali che $q_{i_r} = q_{i_s}$

cioè lo stato in cui arriva l'automa leggendo il prefisso di z di lunghezza r (leggendo z_r) è esattamente lo stesso stato in cui arriva l'automa leggendo il prefisso di z di lunghezza s (leggendo z_s) cioè ancora:

$$\bar{\delta}(q_0, z_r) = q_{i_r} = q_{i_s} = \bar{\delta}(q_0, z_s) \quad (A)$$

Poniamo $u = z_r$ $\quad uv = z_s$ $\quad uvw = z$

Chiaramente $|uv| = |z_s| = s \leq n$ e $|v| \geq 1$

perché $|u| = |z_r| = r < s = |z_s| = |uv|$ $\qquad |u| = |uv|$

$\cancel{|u|} < \cancel{|u|}|v| \qquad 0 < |v| \qquad \cancel{|u|} = \cancel{|u|}|v|$

$|v| > 0 \Rightarrow |v| \geq 1$

Inoltre dobbiamo dimostrare che $uv^iw \in L \quad \forall i \geq 0$

PROCEDIAMO PER INDUZIONE:

PASSO BASE: $i = 0$

$\bar{\delta}(q_0, uv^0w) = \bar{\delta}(q_0, u\varepsilon w) = \bar{\delta}(q_0, uw) = \bar{\delta}(\bar{\delta}(q_0, u), w) \overset{*}{=}$

$= \bar{\delta}(\bar{\delta}(q_0, z_s), w) = \bar{\delta}(\bar{\delta}(q_0, uv), w) = \bar{\delta}(q_0, uvw) = \bar{\delta}(q_0, z) \in F$

cioè $uv^0w = uw \in L$

PASSO INDUTTIVO $i > 0$

PER IPOTESI INDUTTIVA $uv^{i-1}w \in L$ cioè $\bar{\delta}(q_0, uv^{i-1}w) \in F$

Allore $\underline{\bar{\delta}(q_0, uv^{i}w)} = \bar{\delta}(q_0, uv^{1}\underbrace{v^{i-1}w}) = \bar{\delta}(\underbrace{\bar{\delta}(q_0, uv)}, v^{i-1}w) =$

$= \bar{\delta}(\bar{\delta}(q_0, zr), v^{i-1}w) = \bar{\delta}(\bar{\delta}(q_0, u), v^{i-1}w) = \bar{\delta}(q_0, uv^{i-1}w) \in F$

$\Rightarrow \bar{\delta}(q_0, uv^{i}w) \in F \quad \Rightarrow \quad \underline{\underline{uv^{i}w \in L}}$

$\square$ CVD

LINGUAGGI FORMALI => GRAMMATICHE

PROBLEMA DI ANDARE A CALCOLARE UNA FUNZIONE $f: \mathbb{N} \to \mathbb{N}$ può innescare un nuovo problema quello di calcolare e riconosce il linguaggio

$$L = \{ a^n b^{f(n)} \mid n \geq 0 \}$$

- POSSIAMO USARE LE ESPRESSIONI REGOLARI PER DEFINIRE UN LINGUAGGIO FORMALE
- UN ALTRO APPROCCIO È QUELLO generativo dove si usano strumenti formali che sono appunto le grammatiche formali, che consentono di costruire le stringhe di un linguaggio tramite un insieme prefissato di regole, dette regole di produzione
- APPROCCIO RICONOSCITIVO, che consiste nell'utilizzare macchine astratte dette automi riconoscitori che definiscono algoritmi di riconoscimento dei linguaggi stessi, vale a dire algoritmi che per un dato linguaggio $L \subseteq \Sigma^*$ stabiliscono se una stringa $x \in \Sigma^*$ appartiene a L o meno.

GRAMMATICHE DI CHOMSKY

LA GRAMMATICA È UN FORMALISMO CHE PERMETTE DI DEFINIRE UN INSIEME DI STRINGHE MEDIANTE L'IMPOSIZIONE DI UN PARTICOLARE METODO PER LA LORO COSTRUZIONE.

DERIVA DALLO STUDIO DI NOAM CHOMSKY PER LA COSTRUZIONE DI FRASI IN LINGUA INGLESE.

DEFINIZIONE UNA GRAMMATICA FORMALE $\mathcal{G}$ è una quartupla

$$\mathcal{G} = \langle V_T, V_N, P, S \rangle$$

V_T : è l'insieme finito non vuoto di simboli dell' ALFABETO TERMINALE i cui elementi sono detti CARATTERI TERMINALI, o TERMINALI

V_N : è l'insieme finito non vuoto di simboli dell'ALFABETO NON TERMINALE i cui elementi sono detti CARATTERI NON TERMINALI o NON TERMINALI o VARIABILI o CATEGORIE SINTATTICHE

· P è una relazione binaria di cardinalità finita su:

$$\underbrace{(V_T \cup V_N)^* \circ V_N \circ (V_T \cup V_N)^*}, \times (V_T \cup V_N)^*$$

P è detta insieme delle regole di produzione o delle regole sintattiche.

UNA COPPIA $\langle \alpha, \beta \rangle \in P$, indichiamo con $\alpha \to \beta$ la relazione tra α e β

· $S \in V_N$ è detto ASSIOMA ed è il simbolo non terminale di inizio ossia la categoria sintattica più generale.

NOTAZIONI

$V = V_T \cup V_N \Rightarrow [V^* \circ V_N \circ V^*] \times V^*$

$a, b, c \in V_T$

$A, B, X, Y \in V_N$

$x, y, z, w \in V_T^*$

$\alpha, \beta, \gamma \in (V_T \cup V_N)^* = V^*$

* STELLA DI KLEENE

V_T^* l'insieme di tutte le stringhe che possono essere formate usando i simboli presenti nell'alfabeto V_T

Se $V_T = \{a, b\}$ $V_T^* = \{\varepsilon, a, b, aa, ab, bb, \dots\}$

OSS. 1

LE PRODUZIONI DI UNA GRAMMATICA RAPPRESENTANO LE REGOLE MEDIANTE LE QUALI UNA STRINGA COMPOSTA TUTA DI CARATTERI TERMINALI PUÒ ESSERE TRASFORMATA (RISCRITTA) IN UN'ALTRA.

OSS 2.

Il linguaggio generato dalla grammatica è l'insieme delle stringhe costituite solo dai caratteri terminali ai quali si può arrivare partendo dall'assioma S e applicando una sequenza, arbitrariamente lunga, di passi di riscrittura.

UNA GRAMMATICA È DATA DA UN INSIEME DI REGOLE CHE APPLICATE GENERANO TUTTE LE STRINGHE DEL LINGUAGGIO (METODO ENUMERATIVO)