ELEMENTI DI INFORMATICA TEORICA

Parte 2:
Linguaggi Formali e Automi

2.1 Linguaggi e grammatiche formali

Giovanni Amendola

Corso di laurea triennale in Informatica
Università della Calabria

26 marzo 2022

Anno Accademico 2021/2022

Grammatiche

- **Approccio generativo** per la descrizione dei linguaggi.
- Usate per formalizzare la **struttura** dei **linguaggi naturali** attraverso regole di trasformazione (**Gerarchia di Chomsky**, 1956).

Grammatiche

- **Approccio generativo** per la descrizione dei linguaggi.
- Usate per formalizzare la **struttura** dei **linguaggi naturali** attraverso regole di trasformazione (**Gerarchia di Chomsky**, 1956).

Disambiguare le proposizioni del linguaggio naturale

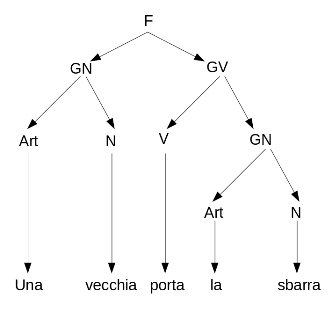
"Una vecchia porta la sbarra"

Grammatiche

- **Approccio generativo** per la descrizione dei linguaggi.
- Usate per formalizzare la **struttura** dei **linguaggi naturali** attraverso regole di trasformazione (**Gerarchia di Chomsky**, 1956).

Disambiguare le proposizioni del linguaggio naturale

"Una vecchia porta la sbarra"

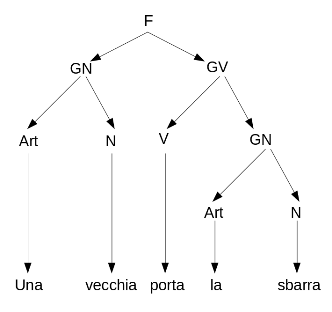


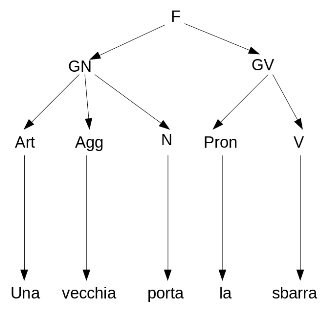
Grammatiche

- **Approccio generativo** per la descrizione dei linguaggi.
- Usate per formalizzare la **struttura** dei **linguaggi naturali** attraverso regole di trasformazione (**Gerarchia di Chomsky**, 1956).

Disambiguare le proposizioni del linguaggio naturale

"Una vecchia porta la sbarra"





Grammatiche

Esempio: "una vecchia porta la sbarra"

- **Alfabeto**:
  - **Simboli terminali**: vecchia, porta, sbarra, una, la
  - **Simboli non terminali**: F, GN, GV, N, V, Art, Agg, Pron

Grammatiche

Esempio: "una vecchia porta la sbarra"

- **Alfabeto**:
  - **Simboli terminali**: vecchia, porta, sbarra, una, la
  - **Simboli non terminali**: F, GN, GV, N, V, Art, Agg, Pron
- **Regole di produzione**:
  1. $F \rightarrow GN + GV$
  2. $GN \rightarrow Art + N \mid Art + Agg + N$
  3. $GV \rightarrow V + GN \mid Pron + V$
  4. $N \rightarrow vecchia \mid sbarra \mid porta$
  5. $V \rightarrow porta \mid sbarra$
  6. $Art \rightarrow una \mid la$
  7. $Pron \rightarrow la$
  8. $Agg \rightarrow vecchia$

Grammatiche

Esempio: "una vecchia porta la sbarra"

- **Alfabeto**:
  - **Simboli terminali**: vecchia, porta, sbarra, una, la
  - **Simboli non terminali**: F, GN, GV, N, V, Art, Agg, Pron
- **Regole di produzione**:
  1. $F \rightarrow GN + GV$
  2. $GN \rightarrow Art + N \mid Art + Agg + N$
  3. $GV \rightarrow V + GN \mid Pron + V$
  4. $N \rightarrow vecchia \mid sbarra \mid porta$
  5. $V \rightarrow porta \mid sbarra$
  6. $Art \rightarrow una \mid la$
  7. $Pron \rightarrow la$
  8. $Agg \rightarrow vecchia$

$$F(GN(Art(una) + N(vecchia)) + GV(V(porta) + GN(Art(la) + N(sbarra))))$$
$$F(GN(Art(una) + Agg(vecchia) + N(porta)) + GV(Pron(la) + V(sbarra)))$$

Grammatiche

Grammatiche

- Usate per lo sviluppo di specifiche per **linguaggi di programmazione**.
- Definire un linguaggio tramite un **alfabeto** e **regole di produzione**.

Grammatiche

- Usate per lo sviluppo di specifiche per **linguaggi di programmazione**.
- Definire un linguaggio tramite un **alfabeto** e **regole di produzione**.

Definizione: **Grammatica**

Una *grammatica formale* $\mathcal{G}$ è una quadrupla $\langle V_T, V_N, S, P\rangle$ dove V_T e V_N sono insiemi disgiunti, finiti e non vuoti. Inoltre

1. V_T è detto **alfabeto terminale** (cioè V_T è un alfabeto Σ).
2. V_N è detto **alfabeto non terminale** o *alfabeto delle variabili* o *alfabeto delle categorie sintattiche*.
3. S è il *simbolo iniziale* o **assioma**.
4. P è un insieme finito di **regole di produzione**. Una regola di produzione è una coppia $\langle v, w\rangle$ con $v \in (V_T \cup V_N)^+$ e contiene almeno un simbolo non terminale, mentre $w \in (V_T \cup V_N)^*$. Una produzione si rappresenta con $v \to w$.

Grammatiche

- Usate per lo sviluppo di specifiche per **linguaggi di programmazione**.
- Definire un linguaggio tramite un **alfabeto** e **regole di produzione**.

Definizione: Grammatica

Una *grammatica formale* $\mathcal{G}$ è una quadrupla $\langle V_T, V_N, S, P\rangle$ dove V_T e V_N sono insiemi disgiunti, finiti e non vuoti. Inoltre

1. V_T è detto **alfabeto terminale** (cioè V_T è un alfabeto Σ).
2. V_N è detto **alfabeto non terminale** o *alfabeto delle variabili* o *alfabeto delle categorie sintattiche*.
3. S è il *simbolo iniziale* o **assioma**.
4. P è un insieme finito di **regole di produzione**. Una regola di produzione è una coppia $\langle v, w\rangle$ con $v \in (V_T \cup V_N)^+$ e contiene almeno un simbolo non terminale, mentre $w \in (V_T \cup V_N)^*$. Una produzione si rappresenta con $v \to w$.

Il linguaggio generato da una grammatica è l'insieme delle **stringhe di terminali** ottenibili da una **sequenza finita di regole** di produzione (o riscrittura).

Grammatiche

Esempio: Grammatica

Definiamo la grammatica per generare il linguaggio delle parentesi bilanciate:

Grammatiche

Esempio: Grammatica

Definiamo la grammatica per generare il linguaggio delle parentesi bilanciate:

1. $V_T = \{(,)\}$.
2. $V_N = \{S\}$.
3. S.
4. $P = \{S \to (), S \to (S), S \to SS\}$
   Nota: invece di ripetere tre volte $S \to$ ..., si può scrivere $S \to () \mid (S) \mid SS$.

Quindi, $\mathcal{G} = \langle \{(,)\}, \{S\}, S, \{S \to () \mid (S) \mid SS\} \rangle$.

Grammatiche

Esempio: **Grammatica**

Consideriamo la seguente grammatica $\mathcal{G}$:

1. $V_T = \{a, b\}$.
2. $V_N = \{S\}$.
3. S.
4. $P = \{(1)\ S \to aSb, (2)\ S \to ab\}$.

- Ad esempio, se applichiamo due volte la prima regola e poi la seconda regola a partire dall'assioma S, abbiamo:

$$S \xrightarrow{(1)} aSb \xrightarrow{(1)} aaSbb \xrightarrow{(2)} aaabbb$$

- Possiamo generare la stringa a^5b^5? Come?

Derivazioni

Definizione: Derivazione diretta

Sia $\mathcal{G}$ una grammatica e siano v e w due stringhe formate da simboli in $(V_T \cup V_N)$. Diciamo che w si ottiene da v per **derivazione diretta** e scriviamo

$$v \Rightarrow w$$

se esistono due stringhe α e β (eventualmente vuote) con simboli in $(V_T \cup V_N)$ tali che

1. $v = \alpha l \beta$,
2. $w = \alpha r \beta$,
3. esiste in $\mathcal{G}$ la regola di produzione $l \to r$.

Ad esempio, se $v = ccSaba$, $w = ccaSaba$ e $S \to aS$ è una regola di produzione, allora $v \Rightarrow w$.

Derivazioni

Definizione: Derivazione

Sia $\mathcal{G}$ una grammatica e siano v e w due stringhe formate da simboli in $(V_T \cup V_N)$. Diciamo che w si ottiene da v per **derivazione** e scriviamo

$$v \stackrel{*}{\Rightarrow} w$$

se

1. $v = w$, oppure
2. esiste una sequenza di stringhe $s_1, \ldots, s_n$, tali che:
   - $s_1 = v$,
   - $s_n = w$,
   - $s_{i-1} \Rightarrow s_i$, per ogni $1 < i \leq n$.

Ad esempio, se $v = ccSaba$, $w = ccaaaSaba$ e $S \rightarrow aS$ è una regola di produzione, allora $v \stackrel{*}{\Rightarrow} w$.

Derivazioni

Esempio: Derivazione

Consideriamo la grammatica per generare il linguaggio delle parentesi bilanciate:

1. $V_T = \{(,)\}$.
2. $V_N = \{S\}$.
3. S.
4. $P = \{S \rightarrow (), S \rightarrow (S), S \rightarrow SS\}$.

Ad esempio, una derivazione da questa grammatica è la seguente:

$$S \Rightarrow (S) \Rightarrow (SS) \Rightarrow (()S) \Rightarrow (()())$$

ovvero

$$S \stackrel{*}{\Rightarrow} (()())$$

Derivazioni

Esempio: Derivazione

Consideriamo la grammatica $\mathcal{G} = \langle \{a, b, c\}, \{S, B, C\}, S, P \rangle$, dove P consiste in

(1) $S \to aS \mid B$

(2) $B \to bB \mid bC$

(3) $C \to cC \mid c$

Ad esempio, posso ottenere le seguenti derivazioni:

- $aaabB \overset{*}{\Rightarrow} aaabbC$
- $aS \overset{*}{\Rightarrow} aabbC$

Linguaggi generati da grammatiche

Definizione: Linguaggio generato da una grammatica

Sia $\mathcal{G} = \langle V_T, V_N, S, P\rangle$ una grammatica. Il **linguaggio** $\mathcal{L}(\mathcal{G})$ **generato** da $\mathcal{G}$ è l'insieme delle stringhe in $(V_T)^*$ derivabili dall'assioma S.

$$\mathcal{L}(\mathcal{G}) = \{s \mid s \in (V_T)^* \text{ e } S \overset{*}{\Rightarrow} s\}$$

Se $S \overset{*}{\Rightarrow} s$ e $s \in (V_T \cup V_N)^*$ diciamo che s è una stringa in *forma enunciativa*. Se s contiene solo simboli terminali diciamo che s è una stringa in *forma enunciativa terminale*.

Linguaggi generati da grammatiche

Definizione: Linguaggio generato da una grammatica

Sia $\mathcal{G} = \langle V_T, V_N, S, P \rangle$ una grammatica. Il **linguaggio** $\mathcal{L}(\mathcal{G})$ **generato** da $\mathcal{G}$ è l'insieme delle stringhe in $(V_T)^*$ derivabili dall'assioma S.

$$\mathcal{L}(\mathcal{G}) = \{s \mid s \in (V_T)^* \text{ e } S \overset{*}{\Rightarrow} s\}$$

Se $S \overset{*}{\Rightarrow} s$ e $s \in (V_T \cup V_N)^*$ diciamo che s è una stringa in *forma enunciativa*. Se s contiene solo simboli terminali diciamo che s è una stringa in *forma enunciativa terminale*.

Definizione: Stringhe generate da una grammatica

Sia $\mathcal{G}$ una grammatica. Se $S \overset{*}{\Rightarrow} s$ ed $s \in (V_T)^*$, allora diciamo che

$\mathcal{G}$ **genera** s.

Linguaggi generati da grammatiche

Esempio

Consideriamo la grammatica $\mathcal{G} = \langle \{a, b\}, \{S\}, S, P \rangle$, dove P consiste in

$$S \rightarrow aSb \mid ab$$

Linguaggi generati da grammatiche

Esempio

Consideriamo la grammatica $\mathcal{G} = \langle \{a, b\}, \{S\}, S, P \rangle$, dove P consiste in

$$S \to aSb \mid ab$$

In tal caso avremo:

$$\mathcal{L}(\mathcal{G}) = \{ab, aabb, aaa, bbb, ...\} = \{a^n b^n | n > 0\}$$

Linguaggi generati da grammatiche

Esempio

Consideriamo la seguente grammatica $\mathcal{G} = \langle \{a, b\}, \{S\}, S, P \rangle$, dove P consiste in

$$S \rightarrow aSa \mid bSb \mid a \mid b \mid \epsilon$$

Linguaggi generati da grammatiche

Esempio

Consideriamo la seguente grammatica $\mathcal{G} = \langle \{a, b\}, \{S\}, S, P \rangle$, dove P consiste in

$$S \rightarrow aSa \mid bSb \mid a \mid b \mid \epsilon$$

In tal caso avremo:

$$\mathcal{L}(\mathcal{G}) = \{\epsilon, a, b, aa, bb, aba, bab, aaa, bbb, ...\}$$

Si tratta del linguaggio delle stringhe palindrome sull'alfabeto $\{a, b\}$.

Linguaggi generati da grammatiche

Esempio

Consideriamo la seguente grammatica $\mathcal{G} = \langle\{a, b\}, \{S\}, S, P\rangle$, dove P consiste in

$$S \to aSa \mid bSb \mid a \mid b \mid \epsilon$$

In tal caso avremo:

$$\mathcal{L}(\mathcal{G}) = \{\epsilon, a, b, aa, bb, aba, bab, aaa, bbb, \ldots\}$$

Si tratta del linguaggio delle stringhe palindrome sull'alfabeto $\{a, b\}$.

Esempio

La grammatica $\mathcal{G} = \langle\{a, b\}, \{S, A\}, S, P\rangle$ con produzioni

- $S \to aA$
- $A \to Sb$

Linguaggi generati da grammatiche

Esempio

Consideriamo la seguente grammatica $\mathcal{G} = \langle \{a, b\}, \{S\}, S, P \rangle$, dove P consiste in

$$S \rightarrow aSa \mid bSb \mid a \mid b \mid \epsilon$$

In tal caso avremo:

$$\mathcal{L}(\mathcal{G}) = \{\epsilon, a, b, aa, bb, aba, bab, aaa, bbb, ...\}$$

Si tratta del linguaggio delle stringhe palindrome sull'alfabeto $\{a, b\}$.

Esempio

La grammatica $\mathcal{G} = \langle \{a, b\}, \{S, A\}, S, P \rangle$ con produzioni

- $S \rightarrow aA$
- $A \rightarrow Sb$

non genera alcuna stringa terminale, dunque $\mathcal{L}(\mathcal{G}) = \emptyset$.

Grammatiche equivalenti

Definizione: Grammatiche equivalenti

Due grammatiche $\mathcal{G}_1$ e $\mathcal{G}_2$ si dicono *equivalenti* se $\mathcal{L}(\mathcal{G}_1) = \mathcal{L}(\mathcal{G}_2)$.

Grammatiche equivalenti

Definizione: Grammatiche equivalenti

Due grammatiche $\mathcal{G}_1$ e $\mathcal{G}_2$ si dicono *equivalenti* se $\mathcal{L}(\mathcal{G}_1) = \mathcal{L}(\mathcal{G}_2)$.

Esempio

Consideriamo $\mathcal{G}_1 = \langle \{a, b\}, \{S, A\}, S, P \rangle$, dove P consiste in

$$S \to Ab \mid b$$
$$A \to aAa \mid aa$$

e $\mathcal{G}_2 = \langle \{a, b\}, \{S, A\}, S, P \rangle$, dove P consiste in

$$S \to Ab$$
$$A \to Aaa \mid \epsilon$$

Grammatiche equivalenti

Definizione: Grammatiche equivalenti

Due grammatiche $\mathcal{G}_1$ e $\mathcal{G}_2$ si dicono *equivalenti* se $\mathcal{L}(\mathcal{G}_1) = \mathcal{L}(\mathcal{G}_2)$.

Esempio

Consideriamo $\mathcal{G}_1 = \langle \{a, b\}, \{S, A\}, S, P \rangle$, dove P consiste in

$$S \to Ab \mid b$$
$$A \to aAa \mid aa$$

e $\mathcal{G}_2 = \langle \{a, b\}, \{S, A\}, S, P \rangle$, dove P consiste in

$$S \to Ab$$
$$A \to Aaa \mid \epsilon$$

Le due grammatiche sono equivalenti. Generano entrambe il linguaggio

$$\{a^{2n} b | n \geq 0\}.$$

Da linguaggio a grammatica

Esempio

Scrivere una grammatica per il linguaggio $\{a^n | n \geq 1\}$.

Da linguaggio a grammatica

Esempio

Scrivere una grammatica per il linguaggio $\{a^n | n \geq 1\}$.
Soluzione: $\mathcal{G} = \langle \{a\}, \{S\}, S, P \rangle$, con le seguenti produzioni:

1. $S \to aS$
2. $S \to a$

Da linguaggio a grammatica

Esempio

Scrivere una grammatica per il linguaggio $\{a^n | n \geq 1\}$.
Soluzione: $\mathcal{G} = \langle \{a\}, \{S\}, S, P \rangle$, con le seguenti produzioni:

1. $S \rightarrow aS$
2. $S \rightarrow a$

Esempio

Scrivere una grammatica per il linguaggio $\{a^n b^m | n \geq 1, m \geq 0\}$.

Da linguaggio a grammatica

Esempio

Scrivere una grammatica per il linguaggio $\{a^n | n \geq 1\}$.
Soluzione: $\mathcal{G} = \langle \{a\}, \{S\}, S, P \rangle$, con le seguenti produzioni:

1. $S \to aS$
2. $S \to a$

Esempio

Scrivere una grammatica per il linguaggio $\{a^n b^m | n \geq 1, m \geq 0\}$.
Soluzione: $\mathcal{G} = \langle \{a, b\}, \{A, B, S\}, S, P \rangle$, con le seguenti produzioni:

1. $S \to AB$
2. $A \to aA \mid a$
3. $B \to bB \mid \epsilon$

Da linguaggio a grammatica

Esempio

Scrivere una grammatica per il linguaggio $\{a^n b^n c^n | n \geq 1\}$.

Da linguaggio a grammatica

Esempio

Scrivere una grammatica per il linguaggio $\{a^n b^n c^n | n \geq 1\}$.

Soluzione

$\mathcal{G} = \langle \{a, b, c\}, \{S, A, B, C\}, S, P \rangle$, con le seguenti produzioni:

1. $S \rightarrow aSBC$
2. $S \rightarrow aBC$
3. $CB \rightarrow BC$
4. $aB \rightarrow ab$
5. $bB \rightarrow bb$
6. $bC \rightarrow bc$
7. $cC \rightarrow cc$