

Introduzione ai Linguaggi Formali

Elementi dei linguaggi

- Alfabeto o vocabolario
 - Insieme **finito** di simboli di base
 - Esempi:
 - Alfabeto latino {a, b, ..., z}
 - Cifre {0, 1, ..., 9}
 - Alfabeto binario {0, 1}
- Stringa su un alfabeto A
 - Sequenza **finita** di simboli dell'alfabeto A
 - Sono consentite le ripetizioni

Esempi

- Alfabeto latino $A=\{a, b, \dots, z\}$
 - a è una stringa su A
 - aa è una stringa su A
 - aba, add, aza, ... sono stringhe su A
- Alfabeto delle cifre $D=\{0, 1, \dots, 9\}$
 - 0, 1, 2, ...,9 sono stringhe su D
 - 012, 999, 923456, ... sono stringhe su D

Lunghezza di una stringa

- La lunghezza di una stringa è il numero di simboli contenuti nella stringa
 - Indichiamo la lunghezza della stringa x con $|x|$
- Esempi:
 - $|a|=1$
 - $|991346|=6$
- La stringa vuota è una stringa che ha zero simboli
 - La indichiamo con ε
 - $|\varepsilon|=0$

Confronto di stringhe

- Due stringhe

- $x = x_1x_2 \dots x_n$

- $y = y_1y_2 \dots y_m$

sono uguali se e solo se

- $|x| = |y|$ ($n = m$)

- $x_i = y_i, \forall i$ ($1 \leq i \leq n$)

- Esempi

- ababs e baasb non sono uguali

- aabb e aabba non sono uguali

Concatenazione

- Date due stringhe x and y , la concatenazione (o prodotto) di x e y è una stringa xy (o $x \cdot y$), dove x è seguita da y
 - Esempio: stringhe su $A=\{a, b, c, d\}$
 - $x=abadd$
 - $y=dcc$
 - $xy=abadddcc$
 - $yx=dccabadd$
- Note
 - Una stringa x concatenata con ε è ancora x
 - Abbreviamo xx con x^2 , xxx con x^3 , ...
 - La concatenazione è associativa e non-commutativa

Sottostringhe

- Una stringa x è una sottostringa (o fattore) di una stringa s se esistono due stringhe y e z tali che $s=yxz$
 - y o z possono essere ε
 - Se $y = \varepsilon$, x è detta **prefisso**
 - Se $z = \varepsilon$, x è detta **suffisso**
 - Se sia y sia z sono ε , x è uguale a s
- Esempio: $s=aadabbc$
 - aad è un prefisso di s
 - $abbc$ è un suffisso s
 - ada è una sottostringa di s

Stella di Kleene

- La stella di Kleene è un operatore unario che si applica a un insieme di simboli o a un insieme di stringhe
 - È indicato con *
- Se A è un alfabeto, allora A^* è l'insieme di tutte le stringhe su simboli di A , inclusa la stringa vuota
- Esempi:
 - Se $A = \{a, b, c\}$ allora $A^* = \{\epsilon, a, b, c, aa, ab, ac, ba, bb, bc, ca, \dots\}$
 - Se $B = \{0, 1\}$ allora $B^* = \{\epsilon, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \dots\}$

Digressione formale (1)

- Un **semigrupp** è una coppia $\langle S, \circ \rangle$, ove
 - S è un insieme (**chiuso** rispetto a \circ)
 - \circ è un'operazione associativa su S
- Un **monoide** è un semigrupp tale che
$$\exists u \forall x (x \circ u = u \circ x = x)$$
(u è un'unità destra e sinistra)
- Un **grupp** è un monoide tale che
$$\forall x \exists x^{-1} (x \circ x^{-1} = x^{-1} \circ x = u)$$
(x^{-1} è l'inverso di x)

Digressione formale (2)

- Dato un semigruppo $\langle S, \circ \rangle$ e un sottoinsieme X di S :
 - X^+ denota il sottoinsieme di S generato da X , cioè tutte le sequenze della forma $x_1 \circ \dots \circ x_n$, con x_i in X e $n \geq 1$
- Per un monoide $\langle S, \circ \rangle$ con unità u ,
 $X^* = X^+ \cup \{u\}$
- X^* è detto il *monoide libero* generato da X

Linguaggi

- Un linguaggio è un insieme di stringhe su un alfabeto
- Linguaggi:
 - Italiano, inglese, francese
 - C, Java, Pascalma anche
 - Linguaggi grafici
 - Musica
 - Multimedia

Formalmente

- Un linguaggio L su un alfabeto A è un sottoinsieme di A^*
- Esempi
 - $A = \{a, b, c\}$
 - $A^* = \{\varepsilon, a, b, c, aa, ab, ac, ba, bb, bc, ca, \dots\}$
 - $L_1 = \{\varepsilon, a, b, c, bc, ca\}$
 - $L_2 = \{aa, ab, ac, ba, bb, bc, ca, cb, cc\}$

Operazioni

- Le operazioni sugli insiemi di applicano anche ai linguaggi
 - Un linguaggio è un insieme di stringhe
- Le operazioni sui linguaggi sono
 - Unione
 - Intersezione
 - Complemento
 - Differenza
 - Concatenazione
 - Potenza n-esima
 - Chiusura di Kleene

Operazioni insiemistiche (1)

- $L_1 \cup L_2$
 - Esempio:
 $L_1 = \{\varepsilon, a, b, c, bc, ca\}$
 $L_2 = \{ba, bb, bc, ca, cb, cc\}$
 $L_1 \cup L_2 = \{\varepsilon, a, b, c, ba, bb, bc, ca, cb, cc\}$
- $L_1 \cap L_2$
 - Esempio: $L_1 \cap L_2 = \{bc, ca\}$

Operazioni insiemistiche (2)

- $L_1 \setminus L_2$ (o $L_1 - L_2$)
 - Generalmente usato quando $L_2 \subseteq L_1$
 - Esempio:
 $L_1 = \{ba, bb, bc, ca, cb, cc\}$
 $L_2 = \{bc, ca\}$
 $L_1 \setminus L_2 = \{ba, bb, cb, cc\}$
- $L^c = A^* \setminus L$
 - A è l'alfabeto su cui L è definito
 - Esempio: $L_1^c =$ insieme di tutte le stringhe su $\{a,b,c\}^*$ tranne le stringhe di lunghezza 2 che iniziano con una 'b' o una 'c'

Concatenazione

- $L_1 \cdot L_2$ (o L_1L_2) = $\{x \cdot y \mid x \in L_1, y \in L_2\}$
 - Nota: ‘ \cdot ’ non è commutativo
 - $L_1 \cdot L_2 \neq L_2 \cdot L_1$

- Esempio

$$L_1 = \{\varepsilon, a, b, c, bc, ca\}$$

$$L_2 = \{ba, bb, bc, ca, cb, cc\}$$

$$L_1L_2 = \{ba, bb, bc, ca, cb, cc, aba, abb, abc, aca, acb, acc, bba, bbb, bbc, bca, bcb, bcc, cba, cbb, cbc, cca, ccb, ccc, bcba, bcbb, bcbc, bcca, bccb, bccc, caba, cabb, cabc, caca, cacb, cacc\}$$

Potenze

- L^n è ottenuto concatenando L con se stesso n volte
 - $L^0 = \{\varepsilon\}$
 - $L^i = L^{i-1} \cdot L$
- Esempi:
 - $L^2 = L \cdot L$
 - $L^3 = L \cdot L \cdot L$
 - $L^4 = L \cdot L \cdot L \cdot L$
 - ...
- Nota: ‘ \cdot ’ è associativo

Chiusura di Kleene

- $L^* = \bigcup_{n=0}^{\infty} L^n$

- $L^+ = \bigcup_{n=1}^{\infty} L^n$

quindi $L^* = L^+ \cup L^0 = L^+ \cup \{\varepsilon\}$

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

L^* e L^+ coincidono sse $\varepsilon \in L$

- Nota: $\{\varepsilon\} \neq \emptyset$

- $\{\varepsilon\} \cdot L = L$

- $\emptyset \cdot L = \emptyset$

Cosa rappresentano i linguaggi formali?

- Un linguaggio è un insieme di stringhe
 - $L_1 = \{bc, ca\}$
 - $L_2 = \{ba, bb, bc, ca, cb, cc\}$
 - $L_3 = \{x \in \{a,b\}^* \mid x = ay \wedge y \in \{a,b\}^*\}$
- Come si possono usare gli insiemi di stringhe nell'informatica?
 - I linguaggi formali non sono soltanto mere rappresentazioni matematiche

Linguaggi nell'informatica

- Un linguaggio è un modo di rappresentare o comunicare informazione
 - Non solo stringhe senza significato
- Ci sono molti tipi di linguaggi
 - Linguaggi naturali
 - Linguaggi di programmazione
 - Linguaggi logici
 - ...

Esempio (1)

- Consideriamo i seguenti linguaggi:
 - L_1 : insieme di documenti di tipo “Word/Mac”
 - L_2 : insieme di documenti di tipo “Word/PC”
- Operazioni:
 - L_1^c è l’insieme di documenti che non sono compatibili con “Word/Mac”
 - $L_1 \cup L_2$ è l’insieme di documenti che sono compatibili con Mac o PC
 - $L_1 \cap L_2$ è l’insieme di documenti che sono compatibili sia con Mac sia con PC

Esempio (2)

- Consideriamo i seguenti linguaggi:
 - L_1 : insieme di messaggi di posta elettronica
 - L_2 : insieme di messaggi di spam
- Operations:
 - $L_1 - L_2$ implementa un filtro
 - Le altre operazioni non hanno molto senso

Linguaggi in pratica

- Un linguaggio può rappresentare
 - Computazioni
 - Documenti
 - Parti di documenti
 - Programmi
 - Multimedia
- Le operazioni sui linguaggi creano nuove classi di linguaggi