Teoria dei linguaggi formali Introduzione

Prof. A, Morzenti

aa 2008-2009

ALFABETO: insieme ***finito*** di elementi

* cardinalità
* stringa, parola, frase

Stringa: insieme ordinato di elementi atomici eventualmente ripetuti

LINGUAGGIO: insieme di stringhe

La struttura insiemistica del linguaggio formale ha due livelli:

•insieme non ordinato di elementi non atomici che sono

•Insiemi ordinati di elementi atomici

Con *leggero abuso di notazione* a volte con Σ si indica l'alfabeto ma anche il linguaggio di stringhe lunghe 1

$$\sum = \{a_1, a_2, ..., a_k\}$$

$$|\Sigma| = k$$

$$\sum = \{a, b, c\}$$

$$abc, aabc, ac, bbb$$

$$\sum = \{a, b, c\}$$

$$L_1 = \{a\,b, a\,c\}$$

$$L_2 = \{bc, bbc\}$$

$$L_1 = \{a\,bc, aabbcc, abcabc, ...\}$$

$$|L_2| = |\{bc, bbc\}| = 2 \qquad |\varnothing| = 0 \qquad |bbc|_b = 2, |bbc|_a = 0$$

*LUNGHEZZA DI UNA STRINGA x: |x|
numero dei suoi elementi

$$|bbc| = 3$$
$$|abbc| = 4$$

* UGUAGLIANZA TRA STRINGHE: due stringhe sono uguali se
Hanno uguale lunghezza
I loro elementi, letti da sinistra a destra, coincidono

$$x = a_1 a_2 \dots a_h,\ y = b_1 b_2 \dots b_k$$
$$x = y \quad \text{se} \quad h = k$$
$$a_i = b_i \quad \text{per} \quad i = 1, \dots h;$$
$$bbc \neq bcb \neq bc$$

OPERAZIONI SULLE STRINGHE /1

CONCATENAMENTO (prodotto)

* operazione fondamentale
* è associativo
* lunghezza

$$x = a_1 a_2 \dots a_h,\ y = b_1 b_2 \dots b_k$$
$$x.y = a_1 a_2 \dots a_h b_1 b_2 \dots b_k = xy$$
$$(xy)z = x(yz)$$
$$|xyz| = |x| + |y| + |z|$$

STRINGA VUOTA

ε è elemento neutro rispetto al concatenamento: qualunque stringa concatenata a dx o a sn non muta

$$x\varepsilon = \varepsilon x = x$$
$$|\varepsilon| = 0$$

differenza tra ε e insieme vuoto

SOTTOSTRINGHE x=uyv

* sottostringhe: y
* prefissi : u
* suffissi: v
* sottrostr. proprie: y, con u $\neq$ε o v $\neq$ε

$$x = abccbc$$
$$prefissi: a, ab, abc, abcc, abccb, abccbc$$
$$suffissi: c, bc, cbc, ccbc, bccbc, abccbc$$
$$sottostringhe: ....., bc, cc, cb, ...$$

OPERAZIONI SULLE STRINGHE /2

RIFLESSIONE

$$x = atri \qquad x^R = irta$$
$$x = bon \qquad y = ton$$
$$xy = bonton$$
$$(xy)^R = y^R x^R = notnob$$

$$x = a_1 a_2 \ldots a_h$$
$$x^R = a_h a_{h-1} \ldots a_2 a_1$$
$$(x^R)^R = x$$
$$(xy)^R = y^R x^R$$
$$\varepsilon^R = \varepsilon$$

RIPETIZIONE

La potenza m-esima (con m maggiore o uguale a 1) è il concatenamento di una stringa con se stessa m-1 volte.

$$x^m = \underbrace{xxx \ldots x}_{1\,2\,3\,\ldots\;m}$$
$$x^m = x^{m-1}x, \quad m > 0$$
$$x^0 = \varepsilon$$

$$x = ab \quad x^0 = \varepsilon \quad x^1 = x = ab \quad x^2 = (ab)^2 = abab$$
$$y = a^3 = aaa \quad y^3 = a^3 a^3 a^3 = a^9$$
$$\varepsilon^0 = \varepsilon \quad \varepsilon^2 = \varepsilon$$

PRECEDENZA OPERATORI:

Elevamento a potenza/concatenamento

Riflessione/concatenamento

$$ab^2 = abb \qquad (ab)^2 = abab$$
$$ab^R = ab \qquad (ab)^R = ba$$

OPERAZIONI SUI LINGUAGGI /1

UNA OPERAZIONE DEFINITA SU UN LINGUAGGIO
È definita su tutte le sue stringhe

$$L^R = \{x \mid x = y^R \wedge y \in L\}$$

predicato caratteristico

$$\text{Prefissi(L)} = \{y \mid x = yz \wedge x \in L \wedge y, z \neq \varepsilon\}$$

prefissi propri

LINGUAGGIO PRIVO DI PREFISSI: nessuno dei prefissi propri delle sue frasi appartiene al linguaggio: Prefissi(L) ed L sono insiemi disgiunti.

$$L_1 = \{x \mid x = a^n b^n \wedge n \geq 1\} \quad a^2b^2 \in L_1 \;\; a^2b \notin L_1$$

L_1 è linguaggio privo di prefissi: ogni prefisso $a^n b^m$ con $n > m \geq 0$

$$L_2 = \{a^m b^n \mid m > n \geq 1\} \quad a^4b^3 \in L_2 \;\; a^4b^2 \in L_2$$

L_2 è linguaggio non privo di prefissi:

OPERAZIONI SUI LINGUAGGI / 2
Operazioni definite su due argomenti

CONCATENAMENTO

$$L'L'' = \{xy \mid x \in L' \wedge y \in L''\}$$

POTENZA m-esima

$$L^m = L^{m-1}L, m > 0$$

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

NB: $\{\varepsilon\} \neq \varnothing$

NB: conseguenze

$$\varnothing^0 = \{\varepsilon\} \quad L.\varnothing = \varnothing.L = \varnothing \quad L.\{\varepsilon\} = \{\varepsilon\}.L = L$$

OPERAZIONI SUI LINGUAGGI / 3

ESEMPI

$$L_1 = \left\{a^i \mid i \geq 0, pari\right\} = \left\{\varepsilon, a^2, a^4, a^6, \ldots\right\}$$

$$L_2 = \left\{b^j a \mid j \geq 1, dispari\right\} = \left\{ba, b^3a, b^5a, \ldots\right\}$$

$$L_1L_2 = \left\{a^i b^j a \mid \left(i \geq 0, pari\right) \wedge (j \geq 1, dispari)\right\}$$

$$= \left\{\varepsilon ba, a^2ba, a^4ba, \ldots \varepsilon b^3a, a^2b^3a, \ldots\right\}$$

$$(L_1)^2 = \left\{\varepsilon, a^2, a^4, a^6, \ldots\right\}\left\{\varepsilon, a^2, a^4, a^6, \ldots\right\} =$$

$$= \left\{\varepsilon, \varepsilon a^2, \varepsilon a^4, \ldots, a^2\varepsilon, a^4, \ldots, a^4\varepsilon, a^6 \ldots\right\} = L_1$$

per ogni coppia di numeri pari h e k, h+k e' pari a^{h+k} appartiene a L_1

ATTENZIONE:

$$\left\{x \mid x = y^m \wedge y \in L\right\} \subset L^m$$

$$m = 2 \quad L_1 = \left\{a, b\right\}$$

$$\left\{a^2, b^2\right\} \subset {L_1}^2 = \left\{a^2, ab, ba, b^2\right\}$$

OPERAZIONI SUI LINGUAGGI / 4

Stringhe di lunghezza finita:

L'operatore di potenza permette di definire in modo espressivo il linguaggio delle stringhe di lunghezza non superiore a un intero K finito

$$\boxed{\begin{array}{l} L = \{\varepsilon, a, b\}^3 \quad k = 3 \\ L = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, \ldots bbb\} \end{array}}$$

Si noti il ruolo di ε
Che consente di
Ottenere tutte le
Stringhe di lunghezza
0, 1, 2

$$\boxed{\begin{array}{c} \{\varepsilon, a, b\} \\ \{\varepsilon, a, b\} \\ \{\varepsilon, a, b\} \end{array}}$$

Se non si vuole nel risultato la stringa vuota:

$$\boxed{L = \{a, b\}\{\varepsilon, a, b\}^2}$$

OPERAZIONI SUI LINGUAGGI / 5

OPERAZIONI INSIEMISTICHE: sono definite le operazioni tradizionali: unione, intersezione, differenza, inclusione, inclusione stretta, uguaglianza

$\cup \quad \cap \quad \backslash \quad \subseteq \quad \subset \quad =$

LINGUAGGIO UNIVERSALE: insieme di tutte le stringhe di alfabeto Σ, di qualsiasi lunghezza anche 0.

$$\boxed{\begin{aligned} L_{universale} &= \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \ldots \\ L_{universale} &= \neg\varnothing \end{aligned}}$$

COMPLEMENTO di un linguaggio L di alfabeto Σ è definito come la differenza insiemistica rispetto al linguaggio universale (insieme delle stringhe di alfabeto Σ che non stanno in L)

$$\boxed{\neg L = L_{universale} \setminus L}$$

OPERAZIONI SUI LINGUAGGI / 6

ESEMPI

Il complemento di un linguaggio finito è sempre infinito

$$\neg\left(\{a,b\}^2\right) = \varepsilon \cup \{a,b\} \cup \{a,b\}^3 \cup \ldots$$

Non è detto che il complemento di un linguaggio infinito sia finito

$$L = \left\{a^{2n} \mid n \geq 0\right\} \quad \neg\mathrm{L} = \left\{a^{2n+1} \mid n \geq 0\right\}$$

Differenza

$$\Sigma = \{a,b,c\}$$

$$L_1 = \left\{x \mid |x|_a = |x|_b = |x|_c \geq 0\right\}$$

$$L_2 = \left\{x \mid |x|_a = |x|_b \wedge |x|_c = 1\right\}$$

$$L_1 \setminus L_2 = \varepsilon \cup \left\{x \mid |x|_a = |x|_b = |x|_c \geq 2\right\}$$

(stesso numero di a,b,c purché non 1)

$$L_2 \setminus L_1 = \left\{x \mid |x|_a = |x|_b \neq |x|_c = 1\right\}$$

(una sola c e ugual numero di a,b purchè diverso da 1)

Introduciamo operazione algebrica di uso freuente chiusura riflessiva e transitiva R* di una relazione R

Dati insieme A e relazione $R \subseteq A \times A$, elemento $(a_1,a_2) \in R$ indicato anche con $a_1 R a_2$

R* e` una ***relazione*** definita come

$$xR^*y \text{ sse } \exists x_o, x_1, \ldots x_n,\ n \geq 0 \text{ tali che } x=x_0,\ y=x_n, \text{ e } \forall i=1,\ldots n\ \ x_{i-1} R x_i$$

Similmente, ***chiusura*** (solo) *transitiva* R^+ idem c.s. con ***n ≥1***
potenza R^k con ***n=k***

Esempio: se su un grafo relazione R di ***adiacenza***
R* e` la relazione di ***raggiugibilita`***

Chiusura A* di un ***insieme*** A rispetto a un'operazione (funzione)
ottenuta da insieme A incrementato degli elementi ottenuti
applicando un numero qualsiasi di volte l'operazione

OPERAZIONI SUI LINGUAGGI / 7

OPERATORE STELLA: passaggio al limite dell'operatore di potenza (detto anche stella di Kleene o chiusura riflessiva e transitiva rispetto al concatenamento).
E' l'unione di tutte le potenze del linguaggio

$$L^* = \bigcup_{h=0\ldots\infty} L^h = L^0 \cup L^1 \cup L^2 \ldots = \varepsilon \cup L^1 \cup L^2 \ldots$$

$$L = \{ab, ba\} \quad L^* = \{\varepsilon, ab, ba, abab, abba, baab, baba, \ldots\}$$

(L è finito $\qquad$ L^* è infinito)

Ogni stringa del linguaggio stella può essere segmentata in sottostringhe appartenenti al linguaggio L

A volte il linguaggio stella coincide con quello di base

$$L = \{a^{2n} \mid n \geq 0\} \quad L^* = \{a^{2n} \mid n \geq 0\} \equiv L$$

OPERAZIONI SUI LINGUAGGI / 8

Se come linguaggio di base si prende Σ, Σ* contiene tutte le stringhe (è il linguaggio universale di alfabeto Σ). Si può anche dire che L è un linguaggio di alfabeto Σ, scrivendo $L \subseteq \Sigma^*$

PROPRIETA' DELLA STELLA

*monotonicità (con * l'insieme si allarga)	$L \subseteq L^*$
*chiusura rispetto al concatenamento	se $\left(x \in L^* \wedge y \in L^*\right)$ allora $xy \in L^*$
*idempotenza	$\left(L^*\right)^* = L^*$
*commutatività di stella e riflessione	$\left(L^*\right)^R = \left(L^R\right)^*$

Inoltre: $\boxed{\varnothing^* = \{\varepsilon\} \qquad \{\varepsilon\}^* = \{\varepsilon\}}$

Esempio: idempotenza $L_o=\{aa\}=\{a^2\}$, $L_1=L_o^*=\{a^{2n} \mid n\geq 0\}$, $L_1^*=L_1$

OPERAZIONI SUI LINGUAGGI / 9

OPERATORE STELLA
Esempio: identificatori I come stringhe di caratteri che iniziano con una lettera
e contengono un numero qualsiasi di lettere e di cifre; identificatori I_5 lunghi al ma. 5

$$\Sigma_A = \{A, B, \ldots, Z\} \quad \Sigma_N = \{0, 1, 2, \ldots, 9\}$$

$$I = \Sigma_A (\Sigma_A \cup \Sigma_N)^*$$

$$\text{se } \Sigma = \Sigma_A \cup \Sigma_N$$

$$I_5 = \Sigma_A (\Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \Sigma^4)$$

$$I_5 = \Sigma_A (\Sigma \cup \varepsilon)^4$$

OPERAZIONI SUI LINGUAGGI / 10

OPERATORE CROCE: chiusura transitiva (non riflessiva) rispetto al concatenamento. L'unitoria non contiene la potenza 0. Utile ma non indispensabile, derivato dalla stella.

$$L^+ = \bigcup_{h=1\ldots\infty} L^h = L^1 \cup L^2 \cup \ldots$$

$$\{ab, bb\}^+ = \{ab, bb, ab^3, b^2ab, abab, b^4, \ldots\}$$

$$\{\varepsilon, aa\}^+ = \{\varepsilon, a^2, a^4, \ldots\} = \{a^{2n} \mid n \geq 0\}$$

Uno stesso linguaggio può essere definito in più modi facendo un uso diverso degli operatori

Esempio: stringhe lunghe più di 4 caratteri:

$$\Sigma^4 \Sigma^*$$

$$(\Sigma^+)^4$$

OPERAZIONI SUI LINGUAGGI / 11

OPERATORE QUOZIENTE: accorcia una frase di un primo linguaggio, decurtandola di un suffisso, appartenente a un secondo linguaggio. NB: sbarra **in avanti**

$$\boxed{L = L'/L'' = \{ y \mid (x = yz \in L') \wedge z \in L'' \}}$$

Esempio

$$L' = \{ a^{2n} b^{2n} \mid n > 0 \}, \quad L'' = \{ b^{2n+1} \mid n \geq 0 \}$$

$$L'/L'' = \{ a^r b^s \mid (r \geq 2, pari) \wedge (1 \leq s < r, s\ dispari) \}$$

$$= \{ a^2 b, a^4 b, a^4 b^3, \ldots \}$$

$$L''/L' = \varnothing$$

nessuna stringa di L” ha una stringa di L’ come suffisso