

SETTEMBRE 2023

LESSICO 1

δ				
0	0	0	1	0
1	0	2	1	0
2	0	0	3	0
3	0	2	1	0
...				
...				

→ colonne = simboli accaduti

NUMERO DI OTTO = LUNGHEZZA PATTERN → 6

$$\delta(j, \sigma) = j+1 \quad \text{se } p[j+1] = \sigma \wedge j < m$$

$\downarrow \quad \downarrow \quad \downarrow$
stato stato $j+1$ pattern

↓
simbolo

$$\delta(j, \sigma) = k \quad \text{se } p[j+1] \neq \sigma \vee j = m$$

$$k = |\beta(p[1, j]\sigma)|$$

Si inizia dallo stato 0 dell'automa e si calcola $\delta(0, T[i])$

Dove $i=1$

Si continua con lo stato raggiunto e il prossimo simbolo del

testo $T[i]$ con $i=2$ finché $i=m$ compreso

Se $\delta(a, T[i]) = b$ e $b=m$ abbiamo trovato un'occorrenza
esatta del pattern in $T[i-m+1, i]$

E la scrittura continua con $T[i+1]$ (se $i+1 \leq m$) in uno stato
precedente allo stato b (cioè $(\delta(m, T[i]))$)

Esercizio 2 un vettore

La $B[i]$ è l'ultima colonna dell'ordinamento lessicografico delle rotazioni, quindi $B[i]$ è il simbolo nell'ultima posizione dell' i -esima rotazione nell'ordinamento lessicografico delle rotazioni.

$$T = c a b c b \$ \quad \$ < a < b < c$$

1	6	\$	c	a	b	c	b
2	2	a	b	c	b	\$	c
3	5	b	\$	c	a	b	c
4	3	b	c	b	\$	c	a
5	1	c	a	b	c	b	\$
6	4	c	b	\$	a	b	c
		S	F			B	

X La LAST_FIRST funzione è una funzione che dato $B[i]$ in input ritorna la posizione del simbolo nel vettore F .

La LAST_FIRST mappare è una procedura che dice che l' i -esima occorrenza del simbolo σ in B coincide con l' i -esima occorrenza del simbolo σ in F .

Esercizio 3

C'è un'altra cosa che una funzione $C()$ è una funzione $\text{occ}()$

La funzione C mappa quanti simboli σ ci sono nel vettore

	$C(\sigma)$
\$	0
a	1
b	2
c	4

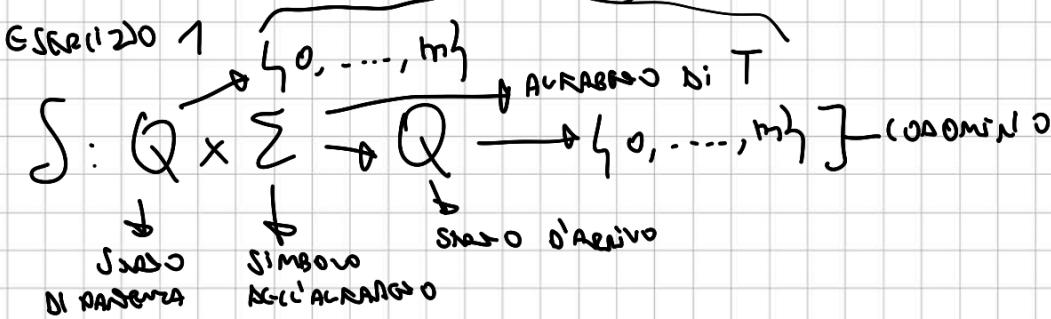
$$T = c a b c b \$$$

La funzione C dà la posizione massima di un simbolo σ nel vettore F .

La funzione $\text{occ}(i, \sigma)$ ci ritorna quanti simboli σ sono presenti in $B[1, i-1]$.

O_{cc}	b	a	b	c
1	0	0	0	0
2	0	0	1	0
3	0	0	1	1
4	0	0	1	2
5	0	1	1	2
6	1	1	1	2
7	1	1	2	2

27 GENNAIO 2023



$$\Sigma(j, \sigma) = j+1 \quad \text{se } j < m \wedge p[j+1] = \sigma$$

$$\Sigma(j, \sigma) = k \quad \text{se } j = m \vee p[j+1] \neq \sigma$$

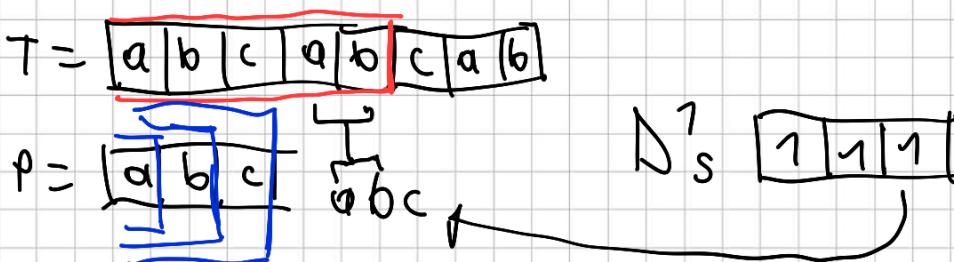
dove $m = |\rho| \in \mathbb{N}$ e $k = |\beta(p[1, j]\sigma)|$

ESEMPIO 2

La parola D_s^1 è una parola lunga come il pattern.
 Ciascuna posizione j della parola è 1 se il prefisso

del pattern $p[1, j]$ occorre come suffisso del pattern.

Nel caso con suffissi di lunghezza 1 su $T[1, 5]$



ESEMPIO 3

$T = abcabc\$ \quad \$ < a < b < c$

\$	a	b	c	a	b
a	b	\$	a	b	c
a	b	c	a	b	\$
b	\$	a	b	c	a
b	c	a	b	\$	a
c	a	b	\$	a	b
F	B				

OCC	\$	a	b	c
1	0	0	0	0
2	0	0	1	0
3	0	0	1	1
4	1	0	1	1
5	1	1	1	1
6	1	2	1	1
7	1	2	2	1

C
\$
a
b
c

$\text{occ}(i, \sigma) = \text{numero di simboli } \sigma \text{ in } B[1, i-1]$

La funzione OCC è una versione numero della BWT

Da cui si può calcolare la funzione C, di cui seguendo

il valore B, F ∈ stringa predizionale T

22 FEBBRAIO 2023

ESEMPIO 1

La parola D_i : è una parola lunga quanto il pattern

dove ogni posizione j nel vettore D_i è 1 se

il prefisso del pattern $P[1:j]$ occorre esattamente

in un suffisso del prefisso lungo i del testo ($\text{SUF}(T[1:j])$)

per calcolare la parola D_i si usa quella prefisso D_{i-1} e

la parola B_σ dove $\sigma = T[i]$

dove B_σ è una word lunga m (come il pattern) che in ogni

posizione j ha 1 se $P[j] = \sigma$

La parola D_i successiva si calcola così:

$$D_i = \text{SHIFT}_1(D_{i-1}) \text{ AND } B_\sigma$$

1) Un'occorrenza accade quando il bit meno significativo nella word D_i è 1

2) La posizione di inizio dell'occorrenza nel testo è $i - m + 1$

Allora l'occorrenza è $T[i - m + 1, i]$

ESEMPIO 2

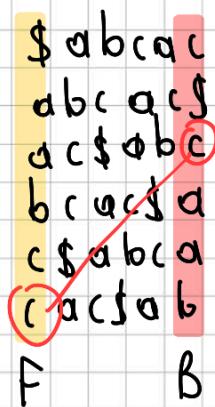
I \mathbb{Q} -intervalli sono intervalli $[a, b)$ di posizioni sul sa che contengono gli indici dei suffissi che condividono la stessa \mathbb{Q} come prefisso su $B^W T$

\nexists \mathbb{Q} -intervalli sono intervalli $[a, b)$ di posizioni di simboli preferibili i suffissi che iniziano con \mathbb{Q}

Esercizio 3

LA PROPOSTA (LAST - RIGHT) MAPPARE NICE THE LA τ -ORDINEZA DI σ
in B (ovvero il vettore B della BWJ) CORRISPOND ALIA
 τ OCCURENZE DI σ IN F

$$T = abc\alpha c\$ \quad f_c a < b < c$$



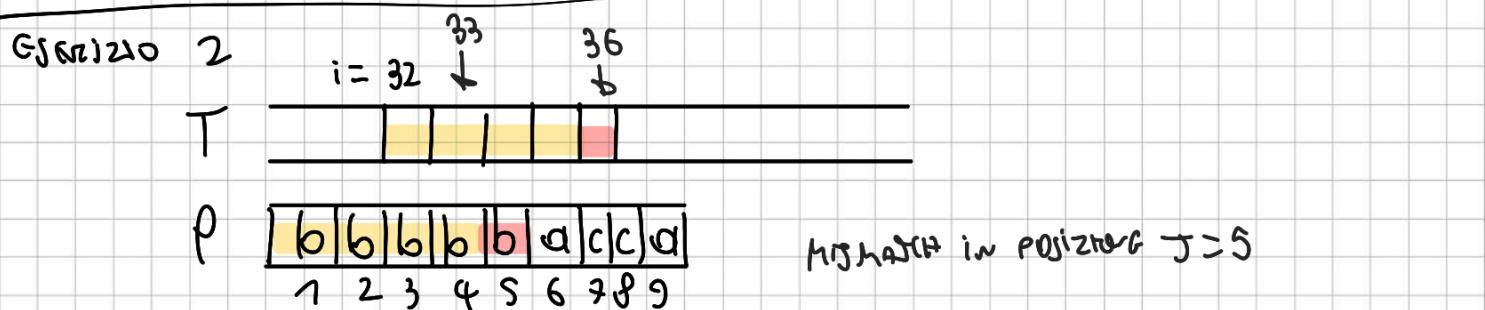
21 GIUGNO 2023

ESERCIZIO 1

Si inizia dallo stato 0, si legge il primo simbolo σ del testo $T[i]$, con la funzione di transizione $\delta(0, \sigma)$ si ricavano un nuovo stato j .

Si prosegue incrementando i finché $i \leq n$ con $n = |T|$

L'occurrenza corrente si visualizza tutte le volte che viene raggiunto lo stato $j=m$ con $m = |\rho|$



Posizione: $i + j - \ell(j-1) - 1 = 32 + 5 - 3 - 1 = 33$

$$\ell(4) = 3$$

Posizione: $i + j - 1 = 32 + 5 - 1 = 36$

Spiegazione: La finestra su T è in posizione 32, mentre il match su P è in posizione 4, quindi il mismatch appare in posizione $p=5$ allora $t=36$.

La nuova posizione della finestra è calcolata come la posizione della mismatch più 1 più il mismatch nel pattern.

Prima funzione nel testo

+ funzione ℓ (posizione di mismatch nel pattern - 1) - 1

La funzione ℓ risolve il bordo del pattern non posizionato.

Esercizio 3

LA LAST-MINJ FUNZIONE $j = LF(\cdot)$ RESTITUISCE LA POSIZIONE j NELL'SUFFIX ARRAY DEL SIMBOLO NELLA POSIZIONE i DELLA BUSTA ANGLO

$B[i]$

$$j = C(B[i]) + \text{occ}(i, B[i]) + 1$$

DONDE LA FUNZIONE C CONTIENE IL NUMERO DI SIMBOLI MINORI DEL SIMBOLO σ PARSO

MENNE LA FUNZIONE occ RESTITUISCE IL NUMERO DI OCCURRENZE DI σ IN $B[1, i-1]$ DONDE $\sigma = B[i]$

i È LA POSIZIONE DEL SIMBOLO $B[i]$ CHE PRECEDE LA ROTAZIONE E COINCIDE CON IL SIMBOLO $F[s]$ CHE È IL PRIMO SIMBOLO DEL VERSORE F DI UNA DATA ROTAZIONE NELL'SUFFIX ARRAY

25 LUGLIO 2023

Esercizio 1

NELLA DAGA DI PREPROCESSING VERRANNO COSTRUITE LE PAROLE B_σ

L'UNITE COME IL PASSO, DOVE CIASCUNA POSIZIONE J DEL VERSO E' 1 SSE

$$P[J] = \sigma$$

QUESTE POSSONO ESSERE CONGIUNTE IN SEQUENZA USANDO UNA MASCHINA

L'UNA COME IL PASSO DI ZERI CON LA PAROLA PIU' SIGNIFICATIVA A

PER CIASUNA POSIZIONE J DEL PASSO VENE RASO C'È RSHIFT AGLIA
SUCCESSIONE

MASCHINA.

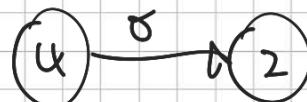
LA MASCHINA E' MESSA IN OR CON LA PAROLA B_σ (CHE STA SIA
INIZIALEZZATA CON TUTTI ZERI) CORRESPONDENTE AL SIMBOLO O LATTO
IN $P[J]$, LA MASCHINA ANDRA' QUINDI A' LA POSIZIONE J DENTRO $M[J]$

NEI RANGHI DI SCANNING SONO CALCOLATE LE WORD D_i L'UNA CHE
COME IL PASSO DONG CIASUNA POSIZIONE E' 1 SSE IL PREMESSO
DEL PASSO $P[1, J]$ OCCORRE COME SURRISIO DEL PREMESSO DEL
TESTO FINO ALLA POSIZIONE i , VENGONO SURF($T[1, i]$)

LE WORD D_i SONO CALCOLATE FACENDO UN RSHIFT AGLIA
WORD PRECEDENTE D_{i-1} IN AND CON LA WORD B_σ DOVE
 $\sigma = T[i]$

Esercizio 2

$$\rho = \begin{matrix} a & b & c & a & a & c \\ 1 & 2 & 3 & 4 & 5 & 6 \end{matrix}$$



Dopo che il nuovo simbolo e' muto del suo precedente

$$f(4, \sigma) = 2 \quad \text{QUINDI } T[i] = \sigma \quad \epsilon \quad J=2 \leq J_{i-1} = 4$$

$$\text{ACCORDE} \quad J_i = |B(P[1, J_{i-1}] \sigma)|$$

$$\text{Quindi } z = |\beta(a b c d \sigma)|$$

$$\text{Quindi } \sigma = b$$

$$\text{Quindi } T[i] = b$$

Esercizio 3

$$\begin{array}{c} F \\ \hline s \\ s \\ a \\ a \\ b \\ c \\ c \end{array} \quad \begin{array}{c} Q \\ \hline c \\ c \\ b \\ a \\ b \\ a \\ b \end{array}$$

$\leftarrow i=4$

$s < a < b < c$

Riconoscere il vettore F (che contiene la prima posizione dell'ordinamento lessicografico delle Q -esrazioni del testo)

$$j = LF(i)$$

j è la posizione nel vettore F della t -esima occorrenza del simbolo $F[j]$ (che coincide con la t -esima occorrenza per simbolo $\beta[i]$)

$$\text{Quindi } i \sim \text{questo caso} \quad LF(4) = 2$$