

16 FEBBRAIO 2022

ESERCIZIO 1

La parola D^h_i è una parola lunga come il pattern dove in ogni posizione j c'è 1 o 0. C'è 1 se il prefisso del pattern lungo j ($P[1, j]$) compare come suffisso del prefisso lungo i del testo ($\text{SUF}_h(t[1, i])$) con distanza di edit h , ovvero h è il numero di operazioni di aggiunta, cancellazione o sostituzione che si possono fare per raggiungere un'occurrenza esatta del pattern.

Quindi dunque h è la distanza di edit mantenendo il indice i e la lunghezza del prefisso di T da confrontare con il prefisso corrispondente $T[1, i]$.

ESEMPIO

$$P = abcc \quad T = \begin{matrix} a & b & a & c & b & | & a & b \\ 1 & 2 & 3 & 4 & 5 \end{matrix}$$

$$D_5^2 = \boxed{\begin{array}{cccc} 1 & 1 & 1 & 1 \\ \hline 1 & 2 & 3 & 4 \end{array}}$$

$$D_5^1 = \boxed{\begin{array}{cccc} 1 & 1 & 0 & 0 \end{array}}$$

$$\overline{D_0^4} \quad \begin{matrix} T / \\ P \end{matrix} \quad \boxed{\begin{array}{cccccccc} & & & & & & & \end{array}}$$

$$\hookrightarrow \boxed{\begin{array}{cccccccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{array}}$$

I primi 4 sono 1 perché, parlando di un testo nullo, posso mantenere il prefisso del pattern lungo rispettivamente 1, 2, 3 e 4 facendo fino a 4 operazioni di aggiunta, dato che $h=4$ ovvero la distanza di edit dalla posizione s in poi invece non ci sono match appropriati perché posso aggiungere solo 4 simboli al testo nullo.

Esercizio 2



$$\text{posizant} = i + j - \varphi(j-1) - 1$$

finegent

$$\text{posizant} = i + j - 1$$

confusivo

Quindi possiamo calcolare la posizant da dove ricomincia il confusivo sul tesso, ma non possiamo calcolare la nuova posizant della finegent e la posizant da cui continua il confusivo sul pattern perché senza sapere il contenuto del pattern non possiamo calcolare il boldo, ovvero la prefix function φ

Esercizio 3

L'FM index è una rappresentazione numerica della BWT

E calcola 2 funzioni:

- $C(\sigma)$ consente il numero di simboli nel stesso che sono lessicograficamente minori di σ , per ogni σ
 - $occ(i, \sigma)$ consente quanti simboli σ sono contenuti nel vettore B della BWT fino alla posizione $i-1$ ($B[1, i-1]$) dove il vettore B della BWT contiene i simboli che preferiscono le σ s nel loro ordinamento lessicografico delle σ -rotazioni
- Quindi $C(\sigma)$ fa count down:0 $\sum_{\sigma' < \sigma} \sigma'$ e ritorna un numero da 0 a n con $n = |T|$
- Mentre occ ha come dominio $\{1, \dots, n\} \times \Sigma$ e codominio $\{0, \dots, n\}$

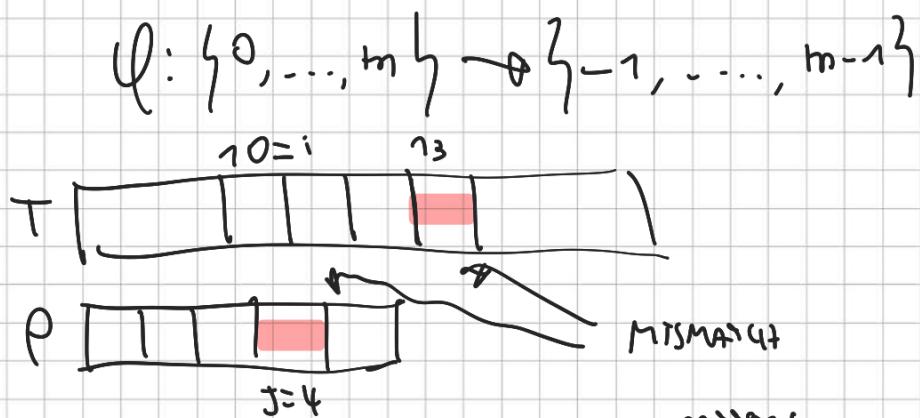
15 LUGLIO 2022

Esercizio 1

Quando durante l'esecuzione di KMP c'è un mismatch, viene usata la prefix function per trovare correttamente.

La prefix function calcola il lungo del prefisso del pattern lungo j , quindi:

- $j=0 \quad \ell(j) = -1$
- $m \geq j > 0 \quad \ell(j) = |\beta(p[1:j])|$



$$\text{posizione pattern} = i + j - \ell(j-1) - 1$$

$\downarrow \quad \downarrow$

posizione pattern PREFIX
FUNCTION

Esercizio 2

Δ_i^k : è una words column come il pattern che contiene i primi k caratteri della stringa i e il prefisso del pattern lungo j ($p[1:j]$)

Occorre che $\Sigma_{i=1}^k$ sia il prefisso del pattern lungo j del testo con distanza di edit $\leq k$ ($\text{SUF}_k(T[1:i])$)

ESERCIZIO 3

LA BWT B CONSIGLIA I SIMBOLI CHE PRECEDONO I SUCCESSI
LESSICOGRAFICAMENTE ORDINATI

OVVERO IL SIMBOLO $B[i]$ PREFERISCE IL SIMBOLO $F[i]$ CHE È
IL PRIMO SIMBOLO DEL SUCCESSO $S[i:]$

12 SETTEMBRE 2022

Esercizio 1

LA FUNZIONE DI TRANSIZIONE È UNA FUNZIONE δ CHE ASSOCIA
IL PASSATO A UNO STATO SUCCESSIVO DELL'AUTOMA.

$$\delta(j, \sigma) = j+1 \text{ SE } j < m \wedge P[j+1] = \sigma$$

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

DONDE σ È IL SIMBOLO DEL TESTO LUNGO, j È IL NUMERO ATTUALE
DELL'AUTOMA E P È IL PATTERNO

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

$$\delta: Q \times S \rightarrow Q$$

↓ ↓ ↓
STATO ALFABETO STATO

con $Q = \{0, 1, \dots, m\}$

ESEMPIO $P = aca$

$$T = bacaba$$

S: PARTE DALLO STATO 0 E SI LEGGE IL PRIMO SIMBOLO DEL TESTO $T[i]$
E SI RISOLVE UN NUOVO STATO, IN QUESTO CASO

$$\delta(0, b) = 0$$

POI SI INCARICA LA POSIZIONE i ; DEL QUALE CI SI RIPETE

$$\delta(0, T[i]) = 1 \quad \text{con } i=2$$

$$\delta(1, T[3]) = 2 \quad \delta(2, T[4]) = 3$$

QUANDO LO STATO j È UGUALE A $m = |P|$ ALLORA

C'È UN'OCCORRENZA ESTRAIA DEL PATTERNO NEL TESTO $T[i-j+1, j]$

Esercizio 2

Una parola D_i è una word lunga come il pattern lungo. Ciasun bit può essere 1 o 0.

il bit $j \in 1$ se il prefisso lungo j del pattern ($P[1:j]$)

ha un'occorrenza esatta in un suffisso del prefisso lungo

del resto ($\text{surf}(T[1:i])$)

altrimenti $D_i[j]$ sarà 0.

Si trova un'occorrenza esatta del pattern nel resto se il bit meno significativo è 1.

Esercizio 3

La backward extension estende un Q -intervallo $[b, e)$

aggiungendo un simbolo σ , trovando il σQ -intervallo $[b', e')$

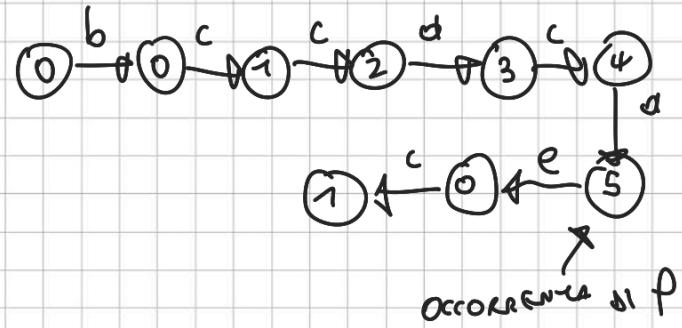
L'intervallo precedente $[b, e)$ contiene tutti i suffissi del resto che hanno come prefisso Q , il nuovo intervallo contiene i suffissi che hanno come prefisso σQ .

20 Giugno 2022

ESEMPIO 1 (SECONDA PARTE)

$$P = \underline{\underline{cc\quad ac\quad cd}} \quad T = \underline{b\quad c\quad c\quad a\quad c\quad d\quad e\quad c} \quad \Sigma = \{a, b, c, d, e\}$$

J	a	b	c	d	e	
0	0	0	1	0	0	
1	0	0	2	0	0	
2	3	0	2	0	0	
3	0	0	4	0	0	
4	5	0	2	0	0	
5	0	0	1	0	0	



$$T[i-m+1, i]$$

Dove i è la posizione del simbolo appena letto nel testo
 $\in m = |P|$

$$\text{Quindi } T[6-5+1, 6] = T[2, 6]$$

ESEMPIO 2

La parola D : È una parola lunga come il pattern

dove ogni posizione j della parola \bar{g} 1 se il prefisso

lungo j del pattern ($P[1, j]$) ha un'occorrenza esatta

in un successivo prefisso lungo i del testo ($SUF(T[1, i])$)

Il paragrafo Shift-and è chiamato così perché calcola

la parola D , successivamente facendo uno shift-1 della parola

precedente D_{i-1} mantenendo in AND con la parola B_0

dove σ è il simbolo del testo in posizione i ($T[i]$)

ESEMPIO 3

IL SURR^NX ARRAY È UN VETTORE VRBO QUANDO IL TEJO
(INCLUDENDO IL \$ FINALE) DOME OGNI POSIZIONE; CONSENTE L'INSERZIONE
DELL'i-ESIMO SURR^NX DEL TEJO NELL'ORDINAMENTO LEXICOGRANICO
NON SURPASSI DI T

QUINDI SE HO ;*i*; ALLORA NELL'ORDINAMENTO LEXICOGRANICO DEL SURR^NX;
HO CHE $T[S[i], h] < T[S[i^+], h]$ CON $h = |T|$