

String matching

Giovanni Righini

25 gennaio 2024



UNIVERSITÀ DEGLI STUDI DI MILANO

Allineamento di sequenze

Date due sequenze S_1 e S_2 di lunghezza n composte da caratteri 'A' e 'B', inserire eventuali caratteri bianchi per minimizzare il costo di allineamento.

Esempio:

- $S_1 = A A A B A B$
- $S_2 = B A B A B B$

Costo	A	B	blank
A	0	2	1
B	2	0	1
blank	1	1	-

Per esempio:

Soluzione 1:

- $S_1 = A A A B A B$
- $S_2 = B A B A B B$

Costo = 8

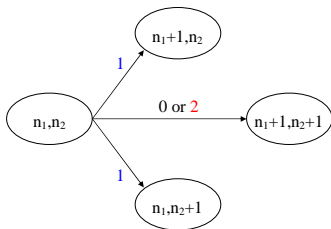
Soluzione 2:

- $S_1 = A A A B A B _$
- $S_2 = B A _ B A B B$

Costo = 4

Programmazione dinamica

- **Politica:** allineamento da sinistra a destra per entrambe le sequenze.
- **Stato:** numero di caratteri allineati su ciascuna sequenza (n_1, n_2) .
- **Estensione:** ogni stato è raggiungibile in al massimo tre modi.



Ci sono n^2 stati ed ogni stato ha 3 predecessori.

Complessità: $O(n^2)$ (polinomiale).

Il grafo degli stati

