

ALGORITMI PER INTERNET E WEB: INDICIZZAZIONE DI TESTI

Appello del 10 Gennaio 2008

COGNOME e NOME

N.MATRICOLA

Esercizio 1. (*10 punti*) Costruire l'automa di Aho-Corasick per i pattern $P_1 = \text{AGAGTAGTATAG}$, $P_2 = \text{TAG}$, $P_3 = \text{AGTA}$ e $P_4 = \text{GT}$, mostrando lo scheletro, la funzione φ e discutendo come sono stati individuati gli stati finali.

COGNOME e NOME

N.MATRICOLA

Esercizio 2. (*12 punti*) Per una stringa binaria S , il suo complemento \bar{S} ha l' i -esimo simbolo $\bar{S}[i]$ uguale al complemento di $S[i]$, ossia $\bar{S}[i] = 1$ se e solo se $S[i] = 0$. Per esempio, se $S = 0101110$ allora $\bar{S} = 1010001$. Introduciamo un simbolo speciale terminatore $\$$ che è più piccolo sia di 0 che di 1. Per una qualunque stringa binaria S , dimostrare che il suffix tree di $S\$$ e quello di $\bar{S}\$$ sono isomorfi.

COGNOME e NOME

N.MATRICOLA

Esercizio 3. (*12 punti*) Progettare un algoritmo che, preso un trie costruito per un insieme S di stringhe, permette di effettuare la *prefix search* di pattern P in S , con al più un errore (usando la distanza di edit). Discutere la complessità dell'algoritmo proposto.

[*Suggerimento: evitare la programmazione dinamica ma sfruttare il fatto che l'errore spezza P in due parti, un prefisso e un suffisso, che devono occorrere esattamente (ossia senza errori).*]