

ALGORITMI PER INTERNET E WEB: INDICIZZAZIONE DI TESTI

Appello 1 Giugno 2004

COGNOME

NOME

Esercizio 1. (16 punti) Per un dato insieme S di stringhe definite su un alfabeto Σ , tale che nessuna stringa è prefisso di un'altra, si consideri il corrispondente trie compatto δ_S . Assumendo che $|\Sigma| = 2^k$, si codifichi quindi ciascun simbolo dell'alfabeto Σ usando k bit. Si operi una trasformazione delle stringhe in S in modo da ottenere un insieme S' di stringhe binarie. Presa $x \in S$ di lunghezza m , si produca una stringa binaria $x' \in S'$ di lunghezza mk , sostituendo i caratteri di Σ con la loro codifica a k bit. Sia $\delta_{S'}$ il trie compatto costruito sulle stringhe in S' .

1. Costruire un esempio per δ_S e $\delta_{S'}$.
2. Hanno δ_S e $\delta_{S'}$ la stessa forma di albero?
3. Preso un qualunque nodo $u \in \delta_S$, a cosa corrisponde in $\delta_{S'}$?
4. Quanto tempo richiede, al caso pessimo, la ricerca di una stringa y in δ_S e $\delta_{S'}$?

Motivare le risposte fornite.

ALGORITMI PER INTERNET E WEB: INDICIZZAZIONE DI TESTI

Appello 1 Giugno 2004

COGNOME

NOME

Esercizio 2. (*14 punti*) Si consideri la distanza di edit $d'(x, y)$ tra due stringhe x e y , in cui sono ammesse *solo* inserzioni e cancellazioni, mentre sono proibite le sostituzioni.

1. Fornire un esempio di due stringhe x e y per cui tale distanza è $d'(x, y) = 3$.
2. Descrivere un modo efficiente per calcolare $d'(x, y)$.