

ALGORITMI PER INTERNET E WEB: INDICIZZAZIONE DI TESTI

Appello 25 Gennaio 2002

COGNOME

NOME

Esercizio 1. (*20 punti*) Data una espressione regolare r e un insieme S di stringhe, si vogliono identificare le stringhe in S aventi un prefisso che soddisfa l'espressione regolare r . Per esempio, se $r = (a|b)^*c$ e $S = \{\text{abaco}, \text{abbacchio}, \text{abito}, \text{bacca}, \text{barca}\}$, le stringhe da fornire in uscita sono abaco, abbacchio e bacca. Tale ricerca è da effettuarsi assumendo di avere a disposizione l'automa nondeterministico δ_r per l'espressione regolare r e usando il trie T costruito sulle stringhe in S .

1. Scrivere lo pseudocodice di $\text{RICERCAREGEXPTRIE}(\delta_r, T)$, che realizza la ricerca descritta sopra.
2. Descrivere il comportamento di RICERCAREGEXPTRIE al caso pessimo.

ALGORITMI PER INTERNET E WEB: INDICIZZAZIONE DI TESTI

Appello 25 Gennaio 2002

COGNOME

NOME

Esercizio 2. (10 punti) Una sottosequenza di una stringa può essere vista come il risultato della cancellazione di zero o più caratteri dalla stringa di partenza. Una sottosequenza *comune* a due stringhe x e y è una stringa che risulta sottosequenza di entrambe x e y . Sia $s(x, y)$ la *sottosequenza comune più lunga* tra x e y .

Dimostrare la relazione $d'(x, y) = |x| + |y| - 2|s(x, y)|$ tra la lunghezza $|s(x, y)|$ e la distanza di edit $d'(x, y)$ in cui sono ammesse *solo* inserzioni e cancellazioni. In altre parole, il numero minimo di operazioni di edit può essere calcolato prendendo la sottosequenza comune più lunga $s(x, y)$.