

ALGORITMI PER INTERNET E WEB: INDICIZZAZIONE DI TESTI

Verifica 20 Dicembre 2007

COGNOME

NOME

Esercizio 1. (*10 punti*) Per la stringa $S = \text{AATATAATAAT}$ costruire:

1. l'automa deterministico per la ricerca del pattern S in un testo;
2. l'automa non-deterministico per la ricerca di S con al più $k = 3$ differenze in un testo;
3. l'albero dei suffissi per il testo $S\$$, dove $\$$ indica il terminatore di fine stringa.

ALGORITMI PER INTERNET E WEB: INDICIZZAZIONE DI TESTI

Verifica 20 Dicembre 2007

COGNOME

NOME

Esercizio 2. *(12 punti)*

1. Descrivere un algoritmo che, preso in ingresso le liste invertite per una collezione di documenti D , calcola la coppia di termini di distanza al più k che occorre più frequentemente nei documenti di D (ossia nel massimo numero di documenti dove entrambi i termini occorrono). La descrizione può essere fatta anche a parole, però deve illustrare l'idea principale ed essere sufficientemente precisa come nello pseudocodice.
2. Discutere la complessità della soluzione trovata.

ALGORITMI PER INTERNET E WEB: INDICIZZAZIONE DI TESTI

Verifica 20 Dicembre 2007

COGNOME

NOME

Esercizio 3. *(12 punti)*

1. Dato il suffix array per un testo T di n simboli, descrivere un algoritmo per trovare la più lunga sottostringa di T che occorre almeno q volte in T stesso. Come sopra, la descrizione può essere fatta anche a parole, però deve illustrare l'idea principale ed essere sufficientemente precisa come nello pseudocodice.
2. Discutere la complessità della soluzione trovata.