

Esame di ammissione al dottorato in Informatica anno 2010

1. Il candidato svolga due esercizi a scelta dei quattro seguenti:

- (a) Dato un canale punto-punto di 2 Km di lunghezza, su cui la luce si propaga con una velocità pari a $2 \times 10^8 m/s$, quale deve essere l'ampiezza di banda B del canale, espressa in Mbps, affinché il ritardo di propagazione P sia uguale al ritardo di trasmissione T per pacchetti di $100B$?
- (b) Sia A un albero in cui i nodi hanno al più r figli e sia F l'insieme delle sue foglie. Si indichi con $l(x)$ la distanza del nodo x dalla radice dell'albero. Si dimostri che vale la relazione

$$\sum_{f \in F} r^{-l(f)} \leq 1$$

con l'uguaglianza che vale se e solo se ogni nodo non foglia ha esattamente r figli.

- (c) Dato un linguaggio L su di un alfabeto A , la derivata di Brzozowski di L rispetto ad un carattere $c \in A$ è definita come:

$$D_c(L) = \{w \mid cw \in L\}$$

- i. Si considerino gli operatori standard $\{\epsilon\}$, $L_1 \cup L_2$, $L_1 \cdot L_2$, L^* , che esprimono rispettivamente il linguaggio che contiene solo la stringa vuota, l'unione, la concatenazione di linguaggi e la stella di Kleene, e se ne esprima la derivata. Suggerimento: per la derivata della concatenazione si consideri l'operatore

$$nullable(L) =_{def} L \cap \{\epsilon\}$$

osservando che $nullable(L) \cdot M$ è uguale ad M quando L contiene ϵ , ed è l'insieme vuoto altrimenti.

- ii. Un linguaggio è detto regolare se è possibile esprimerlo con un'espressione regolare, ovvero costruita a partire dagli operatori $\{\epsilon\}$, $\{a\}$ (il singoletto che contiene il solo carattere $a \in A$), unione, concatenazione, stella di Kleene. Possiamo affermare che la derivata di Brzozowski di un linguaggio regolare è regolare?
- iii. Ricordiamo che l'equivalenza di espressioni regolari è decidibile. Date due espressioni regolari E_1 ed E_2 ed un carattere $c \in A$, è decidibile se $L(E_1) = D_c(L(E_2))$, dove $L(E)$ è il linguaggio denotato da un'espressione regolare?
- (d) L'operatore di raggruppamento

$$\gamma_{\{A_1, \dots, A_n\}, \{f_1(B_1), \dots, f_m(B_m)\}}(R)$$

spesso indicato come $A_1, \dots, A_n \gamma_{f_1(B_1), \dots, f_m(B_m)}(R)$ estende l'algebra relazionale con la possibilità di eseguire un'operazione che ha lo stesso risultato di:

```

select    A1,...,An, f1(B1) as 'f1(B1)',..., fm(Bm) as 'fm(Bm)'
from      R
group by  A1,...,An

```

dove $A_1, \dots, A_n, B_1, \dots, B_m$ sono attributi di R ed f_1, \dots, f_m sono funzioni di aggregazione su multiinsiemi che supporremo scelte tra *count*, *sum* ed *average* (ignorando l'esistenza di valori nulli). Si chiede di:

- i. Dare una definizione insiemistica della semantica di $\gamma_{\{A_1, \dots, A_n\}, \{f_1(B_1), \dots, f_m(B_m)\}}(R)$; a titolo di esempio ne forniamo una per la restrizione $\sigma_{cond}(R)$ e per la proiezione $\pi_{A_1, \dots, A_n}(R)$; usiamo $t[A]$ per estrarre il campo A da t
 - A. $\sigma_{cond}(R) = \{t \mid t \in R \text{ e } cond(t) = True\}$
 - B. $\pi_{A_1, \dots, A_n}(R) = \{t[A_1], \dots, t[A_n] \mid t \in R\}$

dato che le f_i operano su multiinsiemi, suggeriamo di usare la notazione $[T \mid t \in R. cond(t)]$ per indicare il multiinsieme degli elementi di R che soddisfano *cond*.
- ii. Specificare gli attributi del risultato e le regole di buona formazione per l'operatore γ , sempre con riferimento agli attributi, nel seguente stile, esemplificato su σ :
 - A. $\sigma_{cond}(R)$: attributi del risultato: gli stessi di R ; buona formazione: tutti gli attributi che appaiono nell'espressione booleana *cond* devono essere attributi di R .
- iii. Dare alcune condizioni sufficienti (a scelta, non si richiede che siano anche necessarie) affinché le seguenti riscritture non modifichino il significato dell'espressione, per ogni R ; se utile, si faccia uso di dipendenze funzionali
 - A. $\pi_{A_1, \dots, A_n, 'f_1(C_1)', \dots, 'f_k(C_k)'} \gamma_{\{A_1, \dots, A_n, B_1, \dots, B_m\}, \{f_1(C_1), \dots, f_k(C_k)\}}(R)$
 $\Rightarrow \gamma_{\{A_1, \dots, A_n\}, \{f_1(C_1), \dots, f_k(C_k)\}}(R)$
 - B. $\sigma_{cond}(\gamma_{\{A_1, \dots, A_n\}, \{f_1(C_1), \dots, f_k(C_k)\}}(R))$
 $\Rightarrow \gamma_{\{A_1, \dots, A_n\}, \{f_1(C_1), \dots, f_k(C_k)\}}(\sigma_{cond}(R))$
 - C. $\gamma_{\{A_1, \dots, A_n\}, \{f_1('f_1(C_1)'), \dots, f_k('f_k(C_k)')\}}(\gamma_{\{A_1, \dots, A_n, B_1, \dots, B_m\}, \{f_1(C_1), \dots, f_k(C_k)\}}(R))$
 $\Rightarrow \gamma_{\{A_1, \dots, A_n\}, \{f_1(C_1), \dots, f_k(C_k)\}}(R)$
 (si ignori il fatto che gli attributi calcolati hanno nomi diversi, dato che $'f_1(C_1)'$ è una stringa diversa da $'f_1(f_1(C_1))'$).

2. Nell'ambito di una delle aree di ricerca sotto descritte, individuare un problema di ricerca e discuterlo:

- (a) basi di dati
- (b) complessità computazionale
- (c) linguaggi di programmazione
- (d) sistemi distribuiti