Informatica

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 $l(f)$ la distanza del nodo x dalla radice dell'albero. Si dimostri che vale la relazione

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

   con l'uguaglianza che vale se 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 $A_1, \ldots, A_n, B_1, \ldots, B_m$ sono attributi di R ed $f_1, \ldots, f_m$ 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 $\gamma_{\{A_1,\ldots,A_n\},\{f_1(B_1),\ldots,f_m(B_m)\}}(R)$; a titolo di esempio ne forniamo una per la restrizione $\sigma_{cond}(R)$ e per la proiezione $\pi_{A_1,\ldots,A_n}(R)$; usiamo $t[A]$ per estrarre il campo A da t
   A. $\sigma_{cond}(R) = \{t \mid t \in R \; e \; cond(t) = True\}$
   B. $\pi_{A_1,\ldots,A_n}(R) = \{t[A_1], \ldots, t[A_n] \mid t \in R\}$
   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. $\pi_{A_1,\ldots,A_n,'f_1(C_1)',\ldots,'f_k(C_k)'}\gamma_{\{A_1,\ldots,A_n,B_1,\ldots,B_m\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R)$
   $\Rightarrow \gamma_{\{A_1,\ldots,A_n\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R)$
   B. $\sigma_{cond}(\gamma_{\{A_1,\ldots,A_n\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R))$
   $\Rightarrow \gamma_{\{A_1,\ldots,A_n\},\{f_1(C_1),\ldots,f_k(C_k)\}}(\sigma_{cond}(R))$
   C. $\gamma_{\{A_1,\ldots,A_n\},\{f_1('f_1(C_1)')\ldots,f_k('f_k(C_k)')\}}(\gamma_{\{A_1,\ldots,A_n,B_1,\ldots,B_m\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R))$
   $\Rightarrow \gamma_{\{A_1,\ldots,A_n\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R)$
   (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

1. The candidate is required to choose and solve two excercises out of the four below:

   (a) Consider a point-to-point link 2 Km in lenght. At what bandwidth would the propagation delay, at a speed of $2 \times 10^8 m/s$, equal the transmit delay for 100-byte packets?

   (b) Let A be a tree where every node has at most r sons and let F be the set of the leaves of A. Denote with $l(x)$ the distance between the node x and the root of the tree. Show that the following relation holds

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

   where the equality sign holds if and only if every non-leaf node has exactly r sons.

   (c) For a language L over an alphabet A, Brzozowski derivative of L over a character $c \in A$ is defined as:

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

      i. Consider the operators $\{\epsilon\}$, $L_1 \cup L_2$, $L_1 \cdot L_2$, L^*, which express, respectively, the language that only contain the empty string, language union, concatenation, and Kleene star, and express their derivative.
      Suggestion: to express the derivative of concatenation, you may use the following operator:

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

      notice that $nullable(L) \cdot M$ is equal to M when L contains ϵ, and is the empty set otherwise.

      ii. A language is 'regular' if it can be expressed using a regular expression, that is an expression built using the following operators: $\{\epsilon\}$, $\{a\}$ (the singleton that contains just $a \in A$), union, concatenation, Kleene star. Is it true that any Brzozowski derivative of a regular language is regular?

      iii. It is well known that equivalence of regular expressions is decidable.
      For two regular expressions E_1 and E_2 and a character $c \in A$, is it decidable whether $L(E_1) = D_c(L(E_2))$, where $L(E)$ is the language denoted by E?

   (d) The group-by operator

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

   often written as ${}_{A_1,\ldots,A_n}\gamma_{f_1(B_1),\ldots,f_m(B_m)}(R)$, extends the relational algebra with the ability to execute the operation that can be described by the following piece of SQL code:

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

where $A_1, \ldots, A_n, B_1, \ldots, B_m$ are attributes of R, and $f_1, \ldots, f_m$ are aggregation functions, which we will assume to be chosen among $count$, sum and $average$ only, operating on multisets (please ignore any issue related to null values).

i. Give a set-theoretic formal definition of the semantics of

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

as an example, we give here such a definition for restriction $\sigma_{cond}(R)$ and projection $\pi_{A_1,\ldots,A_n}(R)$; we use $t[A]$ to extract the A attribute from t.

A. $\sigma_{cond}(R) = \{t \mid t \in R.\ cond(t) = True\}$

B. $\pi_{A_1,\ldots,A_n}(R) = \{t[A_1], \ldots, t[A_n] \mid t \in R\}$

since f_i's operate on multisets, we suggest to exploit the notation $[T \mid t \in R.\ cond(t)]$ to indicate the multiset of R elements satisfying $cond$.

ii. Specify the attributes of the result and the good formation rules (those having to do with attributes) for γ, in the following style, here exemplified for σ:

A. $\sigma_{cond}(R)$: result attributes: the attributes of R; good formation: all the attributes of $cond$ must be attributes of R.

iii. Give some sufficient conditions (we do not require them to be necessary) such that the following rewritings do not modify the result of the expression, for any R; if needed, you may use functional dependencies

A. $\pi_{A_1,\ldots,A_n,'f_1(C_1)',\ldots,'f_k(C_k)'}\gamma_{\{A_1,\ldots,A_n,B_1,\ldots,B_m\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R)$
$\Rightarrow \gamma_{\{A_1,\ldots,A_n\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R)$

B. $\sigma_{cond}(\gamma_{\{A_1,\ldots,A_n\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R))$
$\Rightarrow \gamma_{\{A_1,\ldots,A_n\},\{f_1(C_1),\ldots,f_k(C_k)\}}(\sigma_{cond}(R))$

C. $\gamma_{\{A_1,\ldots,A_n\},\{f_1('f_1(C_1)')\ldots,f_k('f_k(C_k)')\}}(\gamma_{\{A_1,\ldots,A_n,B_1,\ldots,B_m\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R))$
$\Rightarrow \gamma_{\{A_1,\ldots,A_n\},\{f_1(C_1),\ldots,f_k(C_k)\}}(R)$
(please ignore the fact that the resulting attributes have different names, since $'f_i(C_i)'$ is a different string from $'f_i(f_i(C_i))'$.)

2. The candidate is required to choose one of the following areas and discuss one specific research problem inside that area:

(a) Data bases
(b) Computational complexity
(c) Programming languages
(d) Distributed systems