Esame di ammissione al dottorato in Informatica anno 2010

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

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

   often written as ${}_{A_1,\dots,A_n}\gamma_{f_1(B_1),\dots,f_m(B_m)}(R)$, 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. $\pi_{A_1,\ldots,A_n}(R) = \{t[A_1], ..., t[A_n] \mid t \in R\}$

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