Reg. No: 2016/23/B/ST6/01438; Principal Investigator: prof. dr hab. Jerzy Marcinkowski

# One Step From the Relational Model. A Couple of Database Theory Puzzles.

a Research Proposal (popular science summary)

Jerzy Marcinkowski

December 2016

Our project is in the area of database theory. This theory, inspired by the practice of data bases, tries to: • identify, describe and study, in an abstract, mathematical way, the fundamental theoretical mechanisms behind the practical applications; • propose new concepts and mechanisms, which could be applicable in databases practice; • explain the failures of some practical ideas, by proving negative results.

In our proposal we describe, in some detail, four sub-areas/technical open problems in database theory we plan to work on. None of them can be honestly explained on one page of popular text. But let us at least try to tell something about one of them.

In a typical relational data base we collect some facts we know about. For example we may know facts concerning animals gossiping in the forest:

*Gossips-to(squirrel, bear), Gossips-to(squirrel, wolf), Gossips-to(wolf, fox)*

where the term *Gossips-to(squirrel,wolf)* means that **we know** that anything gossiped by squirrel will finally reach wolf. But be careful: it does not follow from our above example data base whether wolf's gossips reach squirrel – maybe they don't, but maybe they do and we just do not know that they do.

In the currently proposed extensions of the traditional relational model, apart from facts, we also can store in our data base our "*semantic*" knowledge about the relations described by the facts. This knowledge is stored as *rules*. For example all we know about gossiping can be formalized as two rules:

*Gossips-to(x,y)* $\Rightarrow \exists z$ *Gossips-to(y,z)*

*Gossips-to(x,y), Gossips-to(y,z)* $\Rightarrow$ *Gossips-to(x,z)*

First of the rules says that if $y$ is gossiped to („some $x$ gossips to $y$"), then there is some $z$ to whom $y$ herself gossips to. Second rule asserts that if gossips from $x$ reach $y$ and gossips from $y$ reach $z$ then gossips from $x$ reach $z$.

But data bases are not just about storing. They are about querying. For example we would like to know whether...

Exercise ...our knowledge about facts and rules implies that **we can be sure that**: (**a.**) *Gossips-to(squirrel, fox)*? (**b.**) *Gossips-to(fox, squirrel)*? (**c.**) $\exists x$ *Gossips-to(fox,x)* (which reads: "there is someone fox gossips to")? (**d.**) $\exists x$ *Gossips-to(x, squirrel)*? (Answers at the bottom of this page).

In recent years we learned a lot about algorithms evaluating such queries – data base queries in presence of rules. Depending on the particular format of the rules these algorithms can either be computationally easy, or hard, or may not exist at all. But there is one issue we do not understand well. Consider a query: $\exists x$ *Gossips-to(x,x)* - meaning "there is someone whose own gossips finally reach him back". Is this query, according to our knowledge contained in the (above) facts and rules, **surely** true?

This depends on whether we assume (additionally) that there are only finitely many animals in our forest. If we do, then the answer to the query is "yes". If we do not, then the answer is "no, we cannot be sure".

The theory we have so far, concerning algorithms evaluating data base queries in presence of rules, cannot really deal with this additional (natural) assumption that world is finite. One of the goals of our project is to fill this gap.

Answers to the exercise: a. yes; b. no; c. yes; d. no.