Report date 9/11/2024

Paper report:

**Overview:**

The paper introduces a couple of scenarios in which data provenance analysis may be useful, namely:

* To define output trustworthiness based on the trustworthiness of the input it was generated by
* Minimizing the cost of computing the output in the cases where we have to pay for the input
* To compute clearance level to understand what permission are needed to watch the output
* For IVM and output maintenance scenarios
* Backtracking a faulty output to one or more faulty inputs

Out of these 6 points probably the most interesting and the most interesting for us are the 5th and the 6th.

Moving on the concept of which-provenance is mentioned, explaining that it consists in the tuples in the database that contributed to obtaining a particular output.Why-provenance and how-provenance have also been presented.

From how-provenance also started the analysis of the commutative semiring also used in “Provenance in databases: why, how and where” to define the how-provenance of tuples.

8 different semirings used in different scenarios are presented (not very clear why they are presented and many of them not later used).

Some properties of commutative semiring are presented and the analysis moves to the linear algebra spectrum which we are not very interested about.

The most interesting thing is the figure showing the hierarchy of different semiring that can be used for provenance going from the informative to the less informative based on the behavior of its semiring components.

A few other relevant papers and use cases are mentioned regarding the semirings.

Overall this paper is not useful for us since analyses semiring in a mathematical manner and not all the analysis and considerations are related to our study case. Also, I find that it does not add anything more relevant than what “Provenance in databases: why, how and where” implicitly or explicitly says.

**Authors:**

Todd J. Green, Val Tannen

**Year:**

2017