Quality-driven Query Answering and refinement

Ousmane Issa
Institut Supérieur d’Informatique, de Modélisation et de leurs Applications
Université Clermont Auvergne

Objectives
1. Investigate how data quality indicators can be used in order to improve the quality of query answers,
2. Ability to reason about quality-ranked query answers,
3. Design and implemente quality-aware query evaluation algorithm that enables the usage of quality indicators to compute context-dependant and quality-aware query answers,
4. Effective approach to automatically rewrite the query by taking into account the quality indicators of the involved data.

Introduction

Conclusions are done with extracted data from stored data. From these conclusions, decisions are made. If stored data contain erroneous data, extracted data from them can also contain error. So, wrong decision can be made from conclusions that are from these extracted data. To deal with this problem, some quality indicators are introduced. Given a query and a stored data called database, extracted data is answers of this query over database. Having a database that contains error and a query, our main goal is to compute answers of query over this database with a quantity that measures level of errors arise in answers. In literature, there are some related works but they are about error quantification in stored data. Using this quantity, when one takes decision with answers of query so one has idea about the risk of result.

Overview of the problem

Some interesting questions are:
1. How to identify these erroneous data from database ?
2. How to quantify errors of answers of queries ?
3. What means this error quantity for answers of queries ?
4. How to compute this quantity, once it is defined ?
5. depending on the context, what are the underlying filtering problems ?

First contribution, quantify inconsistency

First, we are interested in inconsistency,
Inconsistency arise when some integrity constraints defined over database are violated.
These constraints make it possible to give meaning to the data.
Example of integrity constraint is that two students can not have the same school number or two person can not have the same social security number.
An other example of integrity constraint is any disease is diagnosed before its surgical operation.

Related works

Measure of inconsistency in database [1, 2]. In case of [1], inconsistency is measured as the number of contradictions that can be arise in database when integrity constrains are applied over it. In [2], inconsistency measure is the number of cases that violate constraints.
Other related works are works about consistent query answering. For more details, reader can see [3, 4, 5, 6].

Formalization of the problem

Let I, IC and q respectively a database, a set of integrity constraints and a query.
1. Identify set of tuples in I that violate IC, it is denoted IncT(I, IC)
2. Compute answers of q over I, denoted q(I), and during this processing compute for each tuple t ∈ q(I) its set of set tuples in I from which it is derived, denoted prov(I, q, t)
3. All tuples in answers of query with their inconsistency degree, denoted qIC(t) is the following

\[ q^{inc}(I) = \langle t, m > : t \in q(I) \text{ and } m = \text{Min}_{E \in \text{prov}(I, q, t)}(|E \cap \text{IncT}(I, IC)|) \rangle \]

with || the cardinal function.

Running example

My thesis work is about quality of query answering over database that contains errors
My first contribution is definition of a new measure of inconsistency of query answers.

Conclusion

References

Measuring inconsistency in knowledgebases.
Modeling, measuring and monitoring the quality of information.
In ER’09 Proceedings of the ER 2009 Workshops (CoMoL, ETetheCoM, FP-UML, MOST-ONISW, QoIS, RIGiM, SeCoGIS) on Advances in Conceptual Modeling, 2009.
Database Repairing and Consistent Query Answering.
2011.
Consistent query answers in inconsistent databases.
A dichotomy in the complexity of counting database repairs.
[6] JEF WIJSEN.
Database repairing using updates.
ACM Transactions on Database Systems (TODS), 2005.
Provenance semirings.

Contact Information

Email: issaousmane7@gmail.com
Phone: +33767128651