Intelligent-Information-Systems manuscript No.
(will be inserted by the editor)

Compact Representations for Efficient Storage of Semantic Sensor Data

Farah Karim · Maria-Esther Vidal · Sören Auer

Received: date / Accepted: date

Theorem 1 *Given G and G' such that G' is a factorized RDF graph of G. Let Q and Q' be SPARQL queries where Q' is a rewritten query of Q over G' generated by Algorithm 2. The problem of evaluating Q' against G' is in: (1) PTIME if query Q has only AND and FILTER operators; (2) NP-complete if query Q has expressions with AND, FILTER, and UNION operators; and (3) PSPACE-complete for OPTIONAL graph pattern expressions.*

Proof We proceed with a proof by contradiction. Assume that complexity of Q' is higher than Q. Then, UNION or OPTIONAL operators not included in Q are added to Q'. However, Algorithm 2 only changes triple patterns over G by triple patterns against G'. Additionally, Algorithm 2 includes new JOINs (AND operator). However, adding AND or FILTER operators does not affect the complexity of the problem of evaluating Q' over G', and contradicting the fact that the complexity of Q' is higher than Q.

Theorem 2 *The decomposition of the* `Observation` *universal table into factorized tables:* `Observation`, `Compact Observation Molecule`, *and* `Compact Measurement Mole- cule`, *is* loss-less join.

Proof Considering the following functional dependencies hold in the universal and factorized tables:

- `ObsMID` → `Type`, `Procedure`, `Property`, `MMID`
- `MMID` → `Value`, `Unit`
- `ObsID` → `SamplingTime`, `Timestamp`, `MID`, `ObsMID`

We can prove using the algorithm [2] that the factorized tables are a *loss-less join* decomposition of universal table T that includes all the attributes in the `Observation` universal plus `ObsMID` and `MMID`. The attributes of the `Observation` universal can be projected from G', thus, satisfying the *loss-less join* condition.

Theorem 3 *If G is an SSN RDF graph and G' is a factorized RDF graph of G, and T_1 is the factorized tabular representation of G', then T_1 is in third normal form with respect to the universal representation of G.*

Farah Karim
Leibniz University of Hannover, Welfengarten 1B, 30167 Hannover, Germany
Mirpur University of Science and Technology (MUST), Mirpur-10250 (AJK), Pakistan
E-mail: karim@l3s.de

Proof Recall [1], a table is in third normal form if for every $X \rightarrow Y$

- X is a super key, or
- $Y - X$ is a prime attribute

Considering that the following functional dependencies hold in both the universal, and factorized tables:

- `MMID` → `Value`, `Unit`
- `ObsMID` → `Type`, `Procedure`, `Property`, `MMID`
- `ObsID` → `SamplingTime`, `Timestamp`, `MID`, `ObsMID`

It can be demonstrated that all the tables created after factorization are in 3NF.

Theorem 4 *The decomposition of the* `Class Template (CT)` *based tables representing sensor data into the* `factorized CT` *based tables is* loss-less join.

Proof Consider the following functional dependencies hold in `CT` and `factorized CT` tables:

- `ObsMID` → `Procedure`, `Property`
- `MMID` → `Value`, `Unit`
- `ObsMID`, `MMID` → `ObsMID`, `MMID`
- `ObsID`, `ObsMID` → `ObsID`, `ObsMID`
- `ObsID`, `MID` → `ObsID`, `MID`
- `ObsID`, `SamplingTime` → `ObsID`, `SamplingTime`
- `SamplingTime` → `Timestamp`

We can prove using the algorithm[2] that the `factorized CT` based tables are a *loss-less join* decomposition of the `CT` based tables that includes all the attributes in the `CT` tables plus `ObsMID` and `MMID`. The attributes of the `CT` tables can be projected from G', thus, satisfying the *loss-less join* condition.

Theorem 5 *If G is an SSN RDF graph and G' is a factorized RDF graph of G, and T_2 is the* `Class Template (CT)` *based tabular representation of G', then T_2 is in third normal form with respect to the* `CT` *based tabular representation of G.*

Proof Recall [1], a table is in third normal form if for every $X \rightarrow Y$

- X is a super key, or
- $Y - X$ is a prime attribute

Considering the following functional dependencies hold in `CT` based tables:

- `ObsMID` → `Procedure`, `Property`
- `MMID` → `Value`, `Unit`
- `ObsMID`, `MMID` → `ObsMID`, `MMID`
- `ObsID`, `ObsMID` → `ObsID`, `ObsMID`
- `ObsID`, `MID` → `ObsID`, `MID`
- `ObsID`, `SamplingTime` → `ObsID`, `SamplingTime`
- `SamplingTime` → `Timestamp`

It can be demonstrated that all the factorized tables are in 3NF.

References

1. Codd, E.F.: Further normalization of the data base relational model. Data base systems pp. 33–64 (1972)
2. Jeffrey, D.U.: Principles of database and knowledge-base systems (1989)