AN ABSTRACT OF THE DISSERTATION OF

Parisa Ataei for the degree of Doctor of Philosophy in Computer Science presented on June ?, 2021.

However, many other kinds of variation in databases arise in practice, and different kinds of variation often interact, but these scenarios are not well-supported by existing work.

Title: Theory and Implementation of a Variational Database Management System

Abstract approved: _____

Eric Walkingshaw

maybe better to rephrase slightly and use "data integration" since that's the more common name of the sub-field

In this thesis I present the variational database management system, a formal framework and its implementation for representing variation in relational databases and managing variational information needs. ~~The~~ A variational database is intended to support any kind of variation in a database. Specific kinds of variation in databases have already been studied and are well-supported, for example, schema evolution systems address the variation of ~~the schema of~~ a database 's schema ~~evolving~~ over time and integrated databases address ~~the~~ variation ~~of~~ caused by accessing data from multiple data sources simultaneously. ~~However, due to new business requirements new kinds of variation in databases raises and they are not addressed by existing work. Even though some of the new kinds of variation in databases are the result of combining well-studied kinds of variation, their solutions cannot address the new kind of variation.~~ For example, neither the schema evolution systems nor the database integration systems can address ~~the~~ variation that arises ~~in databases~~ when data sources combined in one database evolve over time.

This thesis collects a large amount of work: It defines the variational database framework and the syntax and [specific kind of] semantics of the variational relational algebra, a query language for variational databases. It also defines the properties of a generic variational database framework that must be met in order to satisfy the variety of information needs users when variation appears in databases and it [shows/proves] that the introduced framework satisfies all these needs. It illustrates two use cases of the variational database framework that represent two existing instances of variation in databases. It presents the variational database management system which is the implementation of

maybe rephrase to just "requirements of a variational database management system"
(i.e. don't mention users)

~~the~~ variational database**s** and variational relational algebra as an abstract layer written in Haskell on top of a traditional RDBMS. It also presents several theoretical results related to the framework and query language, such as syntax-based equivalence rules that preserve the semantics of a query, a type system for ensuring that a variational query is well formed with respect to the underlying variational schema, and a confluence property of the variational relational algebra type system and semantics with respect to the relational algebra type system and semantics.

Theory and Implementation of a Variational Database Management System

by

Parisa Ataei

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June ?, 2021
Commencement June 2021

Doctor of Philosophy dissertation of Parisa Ataei presented on June ?, 2021.

APPROVED:

Major Professor, representing Computer Science

Director of the School of Electrical Engineering and Computer Science

---

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Parisa Ataei, Author

ACKNOWLEDGEMENTS

[Eric. Committee. jeff. abu. parents. friends.]

TABLE OF CONTENTS

Page

1 Introduction 1

1.1 Motivation and Impact . . . 1
1.1.1 Motivating Example . . . 1

1.2 Contributions and Outline of this Thesis . . . 1

2 Background 2

2.1 The Relational Database . . . 2

2.2 Relational Algebra . . . 2

2.3 Variational Set . . . 2
2.3.1 Variational Set Configuration . . . 2

2.4 The Formula Choice Calculus . . . 2

3 The Variational Database Framework 3

3.1 Variational Needs in a Relational Database . . . 3

3.2 Variation Space in a Variational Database Framework . . . 3

3.3 Variational Schema . . . 3
3.3.1 Variational Schema Configuration . . . 3

3.4 Variational Table . . . 3
3.4.1 Variational Table Configuration . . . 3

3.5 Variational Database . . . 3
3.5.1 Variational Database Configuration . . . 4

3.6 Properties of a Variational Database Framework . . . 4

4 The Variational Query Language 5

4.1 Variational Relational Algebra . . . 5
4.1.1 VRA Configuration . . . 5
4.1.2 VRA Semantics . . . 5
4.1.3 VRA Type System . . . 5
4.1.4 VRA Variation-Minimization Rules . . . 5

4.2 Variational Query Language Properties . . . 5

5 Variational Database Management System (VDBMS) 6

5.1 Implemented Approaches . . . 6

TABLE OF CONTENTS (Continued)

Page

5.2 Experiments 6

6 Related Work 7

6.1 Instances of Variation in Databases 7

6.2 Instances of Database Variation Resulted from Software Development . . . 7

6.3 Variational Research 7

7 Conclusion 8

Bibliography 8

Appendices 9

A Variational Database Usecases 10

LIST OF ALGORITHMS

Algorithm Page

Chapter 1: Introduction

[points: - def variation in db and is everywhere - instances but context specific solution don't suffice - thus a generic framework that addresses problems]

[- dimension, behaviour - instances but context specific solution don't suffice - example of instances - well-studied: schema evolution - partial: SPL - new instance either out of combination of other or new - generic framework to instantiate for each instance and address all variational needs.]

1.1 Motivation and Impact

[motivation]

1.1.1 Motivating Example

[combination of instances, behaviours, and dimensions]

1.2 Contributions and Outline of this Thesis

[contribution]

Chapter 2: Background

[background]

2.1 The Relational Database

[relational database]

2.2 Relational Algebra

[relational algebra]

2.3 Variational Set

[vset]

2.3.1 Variational Set Configuration

[vset configuration.]

2.4 The Formula Choice Calculus

[formula choice calculus]

Chapter 3: The Variational Database Framework

[needs. must have configuration.]

3.1 Variational Needs in a Relational Database

[needs and examples of them.]

3.2 Variation Space in a Variational Database Framework

[fexp. evaluation.]

3.3 Variational Schema

[vsch]

3.3.1 Variational Schema Configuration

[vsch configuration.]

3.4 Variational Table

[vtab]

3.4.1 Variational Table Configuration

[vtab configuration]

3.5 Variational Database

[vdb]

3.5.1 Variational Database Configuration

[vdb configuration]

3.6 Properties of a Variational Database Framework

[well-formed vdb properties.context-specific properties.]
[show that they hold for vdb.]

Chapter 4: The Variational Query Language

[vql]

4.1 Variational Relational Algebra

[vra]

4.1.1 VRA Configuration

[vra configuration]

4.1.2 VRA Semantics

[vra semantics]

4.1.3 VRA Type System

[type sys]

4.1.4 VRA Variation-Minimization Rules

[rules]

4.2 Variational Query Language Properties

[prop. show for vra.]

Chapter 5: Variational Database Management System (VDBMS)

[vdbms]

5.1 Implemented Approaches

[apps]

5.2 Experiments

[exp.]

Chapter 6: Related Work

[related work! have to work on this!]

6.1 Instances of Variation in Databases

[schema evolution. database versioning. data integration. data provenance.]

6.2 Instances of Database Variation Resulted from Software Development

[SPL. data model. query.]

6.3 Variational Research

[blah]

Chapter 7: Conclusion

[conclusion]

APPENDICES

Appendix A: Variational Database Usecases

A.1 Variation in Space

[enron email usecase]

A.2 Variation in Time

[employee evolution usecase]