Moshe Y. Vardi's First Love

Phokion G. Kolaitis

UC Santa Cruz & IBM Research

VardiFest 2022

2008 ACM SIGMOD E.F. Codd Innovations Award

- The ACM SIGMOD Edgar F. Codd Innovations Award is the top award in data management; it is "given for innovative and highly significant contributions of enduring value to the development, understanding, or use of database systems and databases."

- Moshe Y. Vardi won this award in 2008. The citation states that "Moshe Y. Vardi is the recipient of the 2008 SIGMOD Edgar F. Codd Innovations Award for fundamental contributions to the foundations of relational databases. He has made significant contributions to the foundations of relational databases by establishing deep connections between database theory, mathematical logic, complexity theory, and AI."

- In accepting the award, Moshe referred to database theory as his first love.

The Beginnings of the Love Story

Moshe Y. Vardi and Database Theory

- M.Sc. in Computer Science, The Weizmann Institute, 5/1980.
  Axiomatization of Functional and Join Dependencies in the Relational Model.
  Advisors: Catriel Beeri (Hebrew University) and Philip Rabinowitz

- Ph.D. in Computer Science, Hebrew University of Jerusalem, 9/1981
  The Implication Problem for Data Dependencies in the Relational Model.
  Advisor: Catriel Beeri

ICALP 1981- What Papers Looked Then

THE IMPLICATION PROBLEM FOR DATA DEPENDENCIES

Extended Abstract

C. Beeri and M.Y. Vardi*

Department of Computer Science
The Hebrew University of Jerusalem
Jerusalem 91904, Israel

ABSTRACT

In this paper we study the implication and the finite implication problems for data dependencies. When all dependencies are total the problems are equivalent and solvable but are NP-hard, i.e., probably computationally intractable. For non-total dependencies the implication problem is unsolvable, and the finite implication problem is not even partially solvable. Thus, there can be no formal system for finite implication. The meta decision problems of deciding for a given class of dependencies whether the implication problem is solvable or whether implication is equivalent to finite implication are also unsolvable.

1. INTRODUCTION

One of the important issues in the design of relational database schemas is the specification of the constraints that the data must satisfy to model correctly the part of the world under consideration. These constraints determine which databases are considered meaningful.

Of particular interest are the constraints called data dependencies. The first class of dependencies to be studied was the class of functional dependencies [Codd], which was followed by the class of multivalued dependencies [Fagl,Zan]. Recently, a number of generalizations of these dependencies have appeared; e.g., join dependencies [ABU,Riss], general dependencies [JP], and template dependencies [SU]. All these classes are subclasses of the class of tuple and equality generating dependencies of [BV2,Fag2,YP]. Intuitively, the meaning of a dependency is that if some tuples, fulfilling certain conditions, exist in the database, then either some other tuples must also exist therein, or some values in the given tuples must be equal.

A utilization of the above dependencies in the design of a relational database requires algorithms for determining whether a set of dependencies is redundant

*Research partially supported by Grant 1849/79 of the U.S.A.-Israel Binational Science Foundation.

The Fruits of the Love Story

The top two database theory conferences are

- The ACM Symposium on Principles of Database Systems (PODS)

and

- The International Conference on Database Theory (ICDT).

Over the years, Moshe has published

- 28 papers in PODS

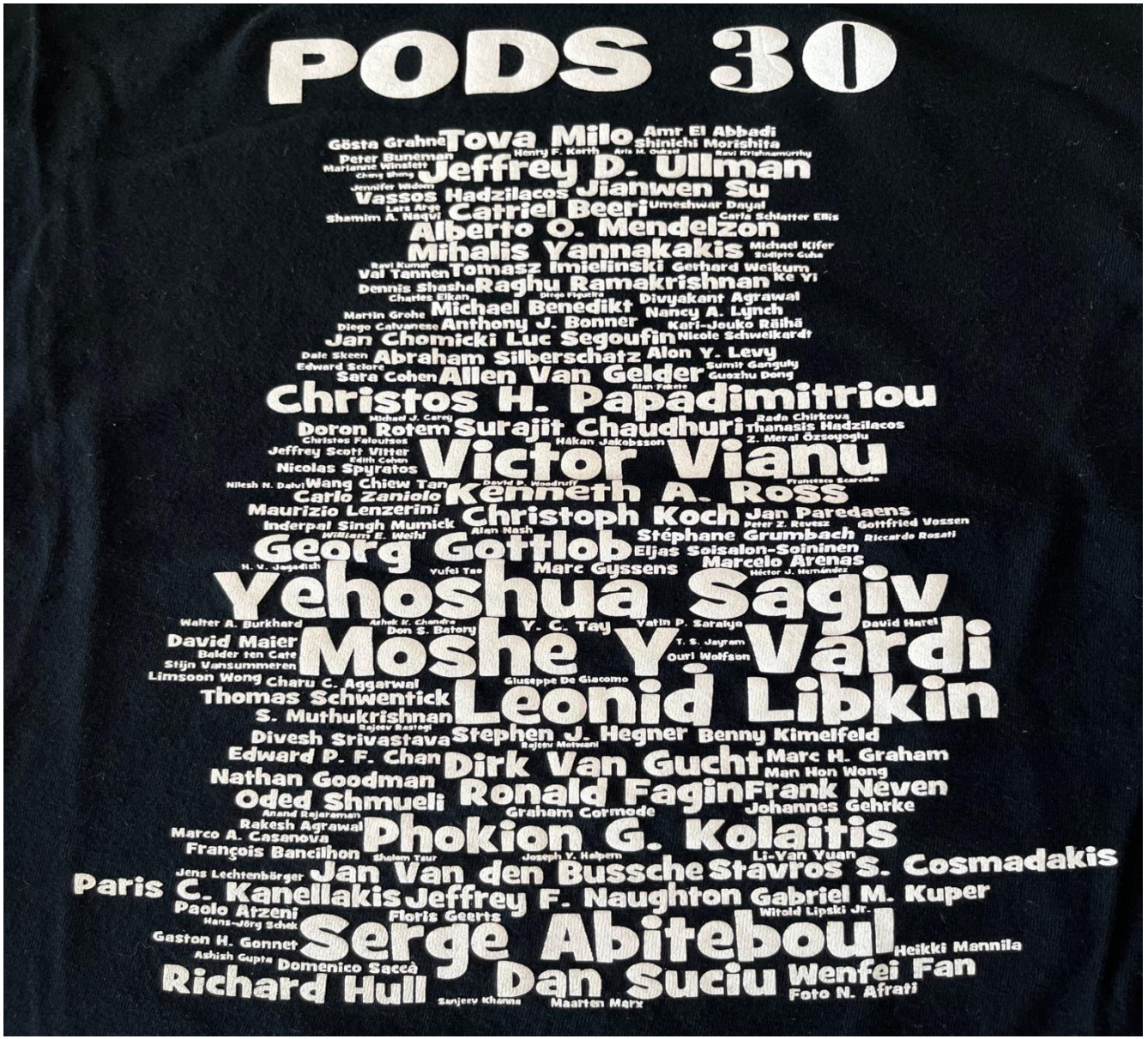
and

- 8 papers in ICDT.

The PODS Hall of Fame

# PAPERS	NAME	’82-’89	’90-’99	’00-’09	’10-’19	’20-
37	Leonid Libkin		8	14	13	2
31	Yehoshua Sagiv	12	10	8	1	
31	Georg Gottlob	1	4	14	12	
30	Victor Vianu	7	9	10	3	1
29	Benny Kimelfeld			4	22	3
28	**Moshe Y. Vardi**	**13**	**8**	**4**	**3**	
28	Serge Abiteboul	8	8	9	3	
28	Dan Suciu		6	9	10	3
28	Phokion G. Kolaitis	1	5	11	8	3
22	Tova Milo		9	7	6	
21	Ronald Fagin	2	2	11	6	

PODS 2011 – The 30th Anniversary



“Four Little Pieces”

I. A Proof Procedure for Data Dependencies

Catriel Beeri and Moshe Y. Vardi, J. ACM 31(4): 718-741 (1984).

II. The Complexity of Relational Query Languages

Moshe Y. Vardi, STOC 1982: 137-146.

III. Conjunctive-Query Containment and Constraint Satisfaction

Phokion G. Kolaitis and Moshe Y. Vardi,

J. Comput. Syst. Sci. 61(2): 302-332; earlier version in PODS 1998.

IV. Optimization of *Real* Conjunctive Queries

Surajit Chaudhuri and Moshe Y. Vardi, PODS 1993: 59-70.

Highlights of Piece I

- A Proof Procedure for Data Dependencies
  Catriel Beeri and Moshe Y. Vardi, J. ACM 31(4): 718-741 (1984).

- 446 citations to date; Moshe's first paper in the Journal of the ACM.

- Data dependencies are integrity constraints that the databases of interest ought to obey.

- The focus is on the implication problem for data dependencies: Given a set Σ of data dependencies and a data dependency ψ, does $\Sigma \models \psi$?

Highlights of Piece I

Conceptual Contributions: Introduction of

- Tuple-generating dependencies (tgds) $\forall \boldsymbol{x}\,(\varphi(\boldsymbol{x}) \rightarrow \exists \boldsymbol{y}\,\psi(\boldsymbol{x},\boldsymbol{y}))$
- Equality-generating dependencies (egds) $\forall \boldsymbol{x}\,\big(\varphi(\boldsymbol{x}) \rightarrow x_i = x_j\big)$,

where $\varphi(\boldsymbol{x})$ and $\psi(\boldsymbol{x},\boldsymbol{y})$ are conjunctions of atoms,

as unifying classes of data dependencies.

Technical Contributions:

- Systematic study of the chase procedure as a tool for reasoning about data dependencies.

  Chase procedure: "*obey the dependencies and do as they tell you*"
- Sufficient conditions for termination; analysis of non-termination.

Impact: Numerous uses in and applications to several areas, including data integration, data exchange, and knowledge representation.

Highlights of Piece II

- The Complexity of Relational Query Languages
  Moshe Y. Vardi, STOC 1982: 137-146.

- 1913 citations to date; Moshe's second most cited paper (after his paper with Pierre Wolper on "An Automata-Theoretic Approach to Automatic Program Verification" – 2249 citations)

- The focus is on the query evaluation problem for a language L:
  Given a database D and a query q in L, does $D \models q$?

Highlights of Piece II

Conceptual Contributions: Three versions of the query evaluation problem

Given a database D and a query q in L, does $D \models q$?

- Combined Complexity: Both D and q are part of the input
- Data Complexity: Fix the query; the input is just the database.
- Expression Complexity: Fix the database; the input is just the query.

Technical Contributions:

- Detailed study of these three versions for different query languages.
- "Empirical" finding: The expression complexity is typically exponentially higher than the data complexity.
- Immerman-Vardi Theorem: PTIME = FO + LFP on ordered structures.

Impact: A tremendously influential piece of work.

- A harbinger to parameterized complexity.

Highlights of Piece III

- Conjunctive-Query Containment and Constraint Satisfaction
  Phokion G. Kolaitis and Moshe Y. Vardi,
  J. Comput. Syst. Sci. 61(2): 302-332; earlier version in PODS 1998.

- 508 citations to date; this paper has a special meaning to Moshe and me because both of us were CS department chairs at that time.

- Conjunctive query: $q(\boldsymbol{x})$: $\exists \boldsymbol{y}\, \varphi(\boldsymbol{x}, \boldsymbol{y})$, where $\varphi(\boldsymbol{x}, \boldsymbol{y})$ is a conjunction of atoms (e.g., "there is a path of length 3 from x_1 to x_2").

- Conjunctive-query containment: $q(\boldsymbol{x}) \subseteq q'(\boldsymbol{x})$ means that for every database D, we have that q(D) $\subseteq$ q'(D).

Highlights of Piece III

Conceptual Contributions:

- The conjunctive-query containment problem (CQC) and the constraint satisfaction problem (CSP) are the same problem.
- Notions of uniform and non-uniform constraint satisfaction.

Technical Contributions:

- From tractable non-uniform CSP to tractable uniform CSP.
- Connections between Datalog and uniform CSP.
- Tight connection between structures of bounded treewidth and conjunctive queries definable by a bounded number of variables.

Impact: Winner of the

2008 ACM PODS Alberto O. Mendelzon Test-of-Time Award

Highlights of Piece IV

- Optimization of *Real* Conjunctive Queries
  Surajit Chaudhuri and Moshe Y. Vardi, PODS 1993: 59-70.

- 233 citations to date.

- The focus is on query processing under bag (multi-set) semantics

R

A	B
1	2
1	2
2	2

S

B	C
2	3
3	4

$R \bowtie_b S$

A	B	C
1	2	3
1	2	3
2	2	3

Highlights of Piece IV

Conceptual Contributions:

Gap between theory and practice:

Database theory is about set semantics, SQL is about bag semantics.

Technical Contributions:

- CQC under set semantics $\neq$ CQC under bag semantics.
- CQ Equivalence under bag semantics $\equiv$ Graph Isomorphism.

Problem:

Is conjunctive-query containment under bag semantics decidable?

(conjunctive-query containment under set semantics is NP-complete)

Impact: This problem remains open to date.

Conclusions and Future Directions

Conclusions:

- Moshe Y. Vardi is a towering figure in database theory.
- His contributions have been tremendously influential, his work in database theory remains relevant to date.

Future Directions:

- Moshe, it is never too late to go back to your first love.
- Database theory still needs you.