

On the Impact of Val on Pierre & ProvSQL

Pierre Senellart



institut
universitaire
de France



ValFest, 25 May 2024

Val and I

Val and I

did val tannen and pierre senellart work together?



Val Tannen and Pierre Senellart have indeed collaborated on research. Val Tannen, a computer science professor at the University of Pennsylvania, and Pierre Senellart, a computer scientist and professor at Télécom ParisTech, have worked together in the field of database theory and web data management. They have co-authored several research papers and have contributed significantly to the advancement of these fields through their joint work.

Val and I

Can you give me some examples of papers they wrote together?



Certainly! Val Tannen and Pierre Senellart have co-authored several research papers. Here are a few examples:

1. "**Provenance for MSO Queries on Trees**" - This paper discusses the provenance of monadic second-order (MSO) logic queries on trees, addressing the challenges of tracking the origins of data in tree-structured databases.
2. "**Provenance and XML: A General Approach and Its Efficiency**" - In this paper, they explore the application of provenance techniques to XML data, proposing a general approach and evaluating its efficiency.
3. "**Semiring-Annotated Data: Queries and Provenance Tractability**" - This work focuses on the use of semiring annotations in data to manage and query provenance information effectively, analyzing the tractability of such queries.
4. "**The Semantics and Provenance of Summaries in Data Integration**" - This paper delves into the semantics of data summaries in the context of data integration and discusses how provenance information can be managed and utilized.

These papers are representative of their collaborative efforts in advancing the understanding and management of data provenance, queries, and integration.

In Reality

- We've never worked together...

In Reality

- We've never worked together...
- But many interactions, and Val's research has had a deep impact on my development as a researcher

My First Database Conference



EDBT 2006
*International Conference on Extending
Database Technology*



My First Database Conference



EDBT 2006
International Conference on Extending
Database Technology



Models for Incomplete and Probabilistic Information

Todd J. Green and Val Tannen

University of Pennsylvania
{tjgreen, val}@cis.upenn.edu

Abstract. We discuss, compare and relate some old and some new models for incomplete and probabilistic databases. We characterize the expressive power of c -tables over infinite domains and we introduce a new kind of result, algebraic completion, for studying less expressive models. By viewing probabilistic models as incompleteness models with additional probability information, we define completeness and closure under query languages of general probabilistic database models and we introduce a new such model, probabilistic c -tables, that is shown to be

My First PODS



ACM SIGMOD/PODS 2007 Conference
2007 ACM SIGMOD International Conference on Management of Data
26th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems
11 - 14 June 2007
Beijing, China
<http://sigmod07.riit.tsinghua.edu.cn>



My First PODS



The banner features a photograph of a traditional Chinese pagoda on the left. The text on the right provides details about the conference, including the dates (11-14 June 2007), location (Beijing, China), and website (<http://sigmod07.riit.tsinghua.edu.cn>). Logos for SIGMOD and ACM SIGMOD are also present.

ACM SIGMOD/PODS 2007 Conference
2007 ACM SIGMOD International Conference on Management of Data
26th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems
11 - 14 June 2007
Beijing, China
<http://sigmod07.riit.tsinghua.edu.cn>



Provenance Semirings

Todd J. Green
tjgreen@cis.upenn.edu

Grigoris Karvounarakis
gkarvoun@cis.upenn.edu

Val Tannen
val@cis.upenn.edu

Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA

Summer 2016

- Frustrated about current state of Probabilistic DB systems, all unmaintained

Summer 2016

- Frustrated about current state of Probabilistic DB systems, all unmaintained
- Became the maintainer of Christoph Koch's and Dan Olteanu's MayBMS but increasingly harder to maintain, and even to compile

Summer 2016

- Frustrated about current state of Probabilistic DB systems, all unmaintained
- Became the maintainer of Christoph Koch's and Dan Olteanu's MayBMS but increasingly harder to maintain, and even to compile
- Decided to develop a new Probabilistic DB system, with focus on maintainability of generality: ProvSQL, in the form of a PostgreSQL extension

Key Inspirations behind ProvSQL

Key Inspirations behind ProvSQL

Definition 13. A **probabilistic c-table** (*pc-tables for short*) consists of a c-table T together with a finite probability space $dom(x)$ (whose outcomes are values in \mathbb{D}) for each variable x that occurs in T .

Key Inspirations behind ProVSQL

Definition 13. A **probabilistic c-table** (*pc-tables for short*) consists of a c-table T together with a finite probability space $\text{dom}(x)$ (whose outcomes are values in \mathbb{D}) for each variable x that occurs in T .

PROPOSITION 4.2. Let K be a commutative semiring and X a set of variables. For any valuation $v : X \rightarrow K$ there exists a unique homomorphism of semirings

$$\text{Eval}_v : \mathbb{N}[X] \rightarrow K$$

such that for the one-variable monomials we have $\text{Eval}_v(x) = v(x)$.

But How to Represent Provenance?

But How to Represent Provenance?

Circuits for Datalog Provenance

Daniel Deutch
Tel Aviv University
danielde@post.tau.ac.il

Sudeepa Roy
University of Washington
sudeepa@cs.washington.edu

Tova Milo
Tel Aviv University
milo@post.tau.ac.il

Val Tannen
University of Pennsylvania
val@cis.upenn.edu

Beyond the positive relational algebra

On the Limitations of Provenance for Queries With Difference *

Yael Amsterdamer

Tel Aviv University
and INRIA

Daniel Deutch

Ben Gurion University and INRIA

Val Tannen

University of Pennsylvania

Beyond the positive relational algebra

On the Limitations of Provenance for Queries With Difference *

Yael Amsterdamer

Tel Aviv University
and INRIA

Daniel Deutch

Ben Gurion University and INRIA

Val Tannen

University of Pennsylvania

Provenance for Aggregate Queries *

Yael Amsterdamer

Tel Aviv University
and University of Pennsylvania
yaelamst@post.tau.ac.il

Daniel Deutch

Ben Gurion University
and University of Pennsylvania
deutchd@cs.bgu.ac.il

Val Tannen

University of Pennsylvania
val@cis.upenn.edu

ProvSQL as of 2024

- **PostgreSQL extension**, works for all versions of PostgreSQL ≥ 9.6 , tested on Linux, Mac OS (x86/ARM), WSL

ProvSQL as of 2024

- **PostgreSQL extension**, works for all versions of PostgreSQL ≥ 9.6 , tested on Linux, Mac OS (x86/ARM), WSL
- Support for a **large subset** of SQL: multiset relational algebra + duplicate elimination + terminal aggregates

ProvSQL as of 2024

- **PostgreSQL extension**, works for all versions of PostgreSQL ≥ 9.6 , tested on Linux, Mac OS (x86/ARM), WSL
- Support for a **large subset** of SQL: multiset relational algebra + duplicate elimination + terminal aggregates
- Now **scalable** (currently experimenting with 10 GB TPC-H dataset)

ProvSQL as of 2024

- **PostgreSQL extension**, works for all versions of PostgreSQL ≥ 9.6 , tested on Linux, Mac OS (x86/ARM), WSL
- Support for a **large subset** of SQL: multiset relational algebra + duplicate elimination + terminal aggregates
- Now **scalable** (currently experimenting with 10 GB TPC-H dataset)
- Provenance computation in **arbitrary semirings** (through the universal semiring), in arbitrary **m-semirings** (through the free m-semiring), in arbitrary **semimodules** for aggregate provenance, **where-provenance**

ProvSQL as of 2024

- **PostgreSQL extension**, works for all versions of PostgreSQL ≥ 9.6 , tested on Linux, Mac OS (x86/ARM), WSL
- Support for a **large subset** of SQL: multiset relational algebra + duplicate elimination + terminal aggregates
- Now **scalable** (currently experimenting with 10 GB TPC-H dataset)
- Provenance computation in **arbitrary semirings** (through the universal semiring), in arbitrary **m-semirings** (through the free m-semiring), in arbitrary **semimodules** for aggregate provenance, **where-provenance**
- **Probabilistic query evaluation** through Boolean provenance, with a variety of techniques (esp., through knowledge compilation)

ProvSQL as of 2024

- **PostgreSQL extension**, works for all versions of PostgreSQL ≥ 9.6 , tested on Linux, Mac OS (x86/ARM), WSL
- Support for a **large subset** of SQL: multiset relational algebra + duplicate elimination + terminal aggregates
- Now **scalable** (currently experimenting with 10 GB TPC-H dataset)
- Provenance computation in **arbitrary semirings** (through the universal semiring), in arbitrary **m-semirings** (through the free m-semiring), in arbitrary **semimodules** for aggregate provenance, **where-provenance**
- **Probabilistic query evaluation** through Boolean provenance, with a variety of techniques (esp., through knowledge compilation)
- **Shapley value** computation

Thanks Val!

- for what you taught me through your papers

Thanks Val!

- for what you taught me through your papers
- for the idea of (algebraic?) elegance as a guiding principle for research

Thanks Val!

- for what you taught me through your papers
- for the idea of (algebraic?) elegance as a guiding principle for research
- for being part of both my PhD and Habilitation committees

Thanks Val!

- for what you taught me through your papers
- for the idea of (algebraic?) elegance as a guiding principle for research
- for being part of both my PhD and Habilitation committees
- for recommending me for jobs

Thanks Val!

- for what you taught me through your papers
- for the idea of (algebraic?) elegance as a guiding principle for research
- for being part of both my PhD and Habilitation committees
- for recommending me for jobs
- for your friendship