Report from Dagstuhl Seminar 25081

# Semirings in Databases, Automata, and Logic

## Guillermo Badia[*1], Manfred Droste[*2], Phokion G. Kolaitis[*3], Carles Noguera[*4], Sophie Brinke[†5], Lovro Mrkonjić[†6], and Gaia Petreni[†7]

1   **University of Queensland – Brisbane, AU.** `g.badia@uq.edu.au`
2   **Universität Leipzig, DE.** `droste@informatik.uni-leipzig.de`
3   **University of California – Santa Cruz, US & IBM Almaden Research Center – San Jose, US.** `kolaitis@ucsc.edu`
4   **University of Siena, IT.** `carles.noguera@unisi.it`
5   **RWTH Aachen, DE.** `brinke@logic.rwth-aachen.de`
6   **RWTH Aachen, DE.** `mrkonjiclovro@gmail.com`
7   **University of Siena, IT.** `gai.petreni@student.unisi.it`

── **Abstract** ──

This report documents the program and the outcomes of Dagstuhl Seminar 25081 "Semirings in Databases, Automata, and Logic", which was held from February 16 to 21, 2025. The seminar focused on semirings, a class of algebraic structures with many applications in computer science, particularly in databases and automata. Semirings are used in databases to annotate tuples in the input and output relations of queries (in particular, in the case of bag semantics, using the semiring of natural numbers) allowing to model several relevant aspects of databases. In automata theory, semirings allow to define weighted automata, which have applications in natural language processing, speech recognition, and image compression. Moreover, semirings are strongly related to the algebraic semantics of many-valued logics. The seminar brought together researchers from the communities mentioned above, and it developed a research agenda for studying semirings, guided by a collection of diverse applications. This led to several new collaborations between members from different communities, including joint work for publications.

---

\* Editor / Organizer
† Editorial Assistant / Collector

## 1    Executive Summary

*Guillermo Badia (University of Queensland – Brisbane, AU)*
*Manfred Droste (Universität Leipzig, DE)*
*Phokion G. Kolaitis (University of California – Santa Cruz, US & IBM Almaden Research*
*Center – San Jose, US)*
*Carles Noguera (University of Siena, IT)*

Semirings are fundamental algebraic structures that in recent times have found a number of applications to computer science, especially in the areas of databases and automata. On the side of databases, commercial query languages, such as SQL, use bag semantics, instead of set semantics, to evaluate relational database queries, which means that the semiring of the natural numbers is used to annotate tuples in the input and output relations. More generally, the annotations can be values in some fixed semiring; this gives a common generalization of both set semantics and bag semantics of database queries, and also makes it possible to model other situations in which one is interested, e.g., in the probability or the reliability of an answer. Furthermore, semirings of polynomials have been successfully used to carry out a rigorous study of provenance in databases. On the side of automata, semirings are used to define weighted automata, which are nondeterministic finite automata augmented with values from a semiring as weights on the transitions. These weights may model, e.g., the cost involved when executing a transition, the amount of resources or time needed for this, or the probability or reliability of its successful execution. Weighted automata have found numerous applications to natural language processing, speech recognition, and algorithms for digital image compression.

These applications have inspired numerous investigations in the logic-in-computer-science community. For instance, motivated by the work on semirings in databases, the evaluation of arbitrary first-order formulas in semirings has been recently explored in the literature, where, among other things, the relation between elementary equivalence (i.e., indistinguishability in first-order logic) and isomorphism on finite structures has been examined. In the same vein, various notions of locality in the semiring framework have been studied and zero-one laws and convergence laws have been obtained, whereas Ehrenfeucht–Fraïssé games have been used to characterize equivalence up to bounded quantifier rank under semiring semantics. As regards semirings and automata, Kleene's classical theorem on the characterization of languages recognized by automata with languages denoted by regular expressions was extended by Schützenberger to the behaviors of weighted automata and rational power series. By a fundamental theorem of Büchi, Elgot, and Trakhtenbrot, finite automata have the same expressive power on words as monadic second-order logic; therefore, various problems about this logic on words are decidable. In recent years, a suitable weighted logic was developed and shown to have the same expressive power on words, trees, and graphs as weighted automata. Consequently, this weighted logic has similar decidability properties on these structures as the unweighted monadic second-order logic. In turn, both of these approaches could be seen as part of the framework of many-valued logics, where algebraic structures expanding semirings with suitable operations have been studied for a long time.

This seminar brought together researchers from three communities that share an interest in semirings, namely databases, automata and multi-valued logic. The goal was to develop a research agenda, featuring tutorials, contributed talks, and discussions on open problems and future directions. Overall, the seminar was a success.

**Organization of the Seminar.** The seminar was held between February 16–21, 2025 from Monday to Friday. There were five invited tutorials with long 75-minute talks, two from the database community, two from the automata community, and one from the multi-valued logic community, that helped define a common language and a common set of problems, and 16 contributed 25-minute talks from experts in all these fields. The tutorials were given by

- Val Tannen on the semiring framework of data(base) provenance;
- Erik Paul on weighted automata and weighted logics;
- Erich Grädel on the model theory of semiring semantics;
- Sara Ugolini on the algebraic semantics of many-valued logics; and
- Peter Kostolányi on models of computation over semirings.

In addition, there was a session where five open problems (some of them rather major) were presented. Our collectors, Sophie Brinke, Lovro Mrkonjić, and Gaia Petreni, recorded all open problems, and later collected them for inclusion in this report. Finally, there was a closing session to formulate future research directions. The schedule was developed in a way that allowed sufficient time for informal discussions and exchange of ideas between participants.

**Outcomes of the Seminar.** There are several major outcomes:

- The presence of participants with very diverse backgrounds enabled us to exchange interesting ideas. For example, logicians and algebraists were strongly inspired by the problems in the database community, and database theoreticians learned relevant facts about semirings from the automata and algebra people. New collaborations emerged during the seminar, and there are some new papers with authors from different communities currently in the writing.
- We have assembled a list of open problems from all involved areas, which we included here. We hope that this list will help define the community interested in semirings, and will also inspire young researchers to contribute to this currently very timely area.
- A follow-up seminar is already being planned with a modified team of organizers (including two of the present ones, Guillermo Badia and Carles Noguera, together with Val Tannen and Thomas Eiter), adding the new community of AI.

## 2    Table of Contents

**Open problems**

## 3 Overview of Talks

### 3.1 The Semiring Framework for Data(base) Provenance

*Val Tannen (University of Pennsylvania – Philadelphia, US)*

Data from different sources mixes up in *data sharing* systems. We end up with complex relationships between participants' data. In well-disciplined systems, these relationships are specified by logical constraints that define *database transformations*. When participants receive data from the sharing system, they would like to apply some *trust* judgements to it. Their confidence may be based on external opinions about the other participants. To combine these opinions, they need to know how a data item they receive depends on the sources – other participants in the system. Knowing this means knowing the *provenance* of the data item.

This tutorial provides an overview of the *semiring framework* for databases and logics, which forms the foundation of the *semiring provenance* approach.

### 3.2 Weighted Automata and Weighted Logics

*Erik Paul (Universität Leipzig, DE)*

Weighted automata are a generalization of finite automata and allow the description of quantitative properties of languages. Finite automata themselves are known to be expressively equivalent to MSO logic due to the fundamental Büchi–Elgot–Trakhtenbrot Theorem. For a long time, a corresponding logical characterization of weighted languages recognizable by weighted automata had been missing.

In this tutorial, we provide an introduction to weighted automata and in particular to weighted logics, a generalization of classical MSO logic expressively equivalent to weighted automata. After giving some insights into the ideas behind weighted logics in their current form, we consider generalizations to data structures and automaton models beyond weighted automata over words as well as generalizations to weight structures beyond semirings. Finally, we present further evidence of the robustness of weighted logics by examining a generalization of the classical Feferman–Vaught Theorem to weighted logics over arbitrary finite structures and by showing how fragments of weighted logics capture languages recognizable by weighted automata restricted to a certain degree of ambiguity.

### 3.3 The Impact of State Mergingon Predictive Accuracy in Probabilistic Tree Automata: Dietze's Conjecture Revisited

*Johanna Björklund (University of Umeå, SE)*

Dietze's conjecture concerns the problem of equipping a tree automaton $M$ with weights to make it probabilistic, in such a way that the resulting automaton $N$ predicts a given corpus $C$ as accurately as possible. The conjecture states that the accuracy cannot increase if the states in $M$ are merged with respect to an equivalence relation $\sim$ so that the result is a smaller automaton $M^\sim$. Put differently, merging states can never improve predictions. This is under the assumption that both $M$ and $M^\sim$ are bottom-up deterministic and accept every tree in $C$. We prove that the conjecture holds, using a construction that turns any probabilistic version $N^\sim$ of $M^\sim$ into a probabilistic version $N$ of $M$, such that $N$ assigns at least as great a weight to each tree in $C$ as $N^\sim$ does.

### 3.4 Term Rewriting, Equality Saturation, the Chase, and Tree Automata

*Dan Suciu (University of Washington – Seattle, US)*

Equality saturation is an approach to the word problem that is currently popular in the compilers community. In this talk we describe the equality saturation framework, and relate it to two well-known concepts: the chase of TGDs and EGDs, and to tree automata. We also describe open problems, some of which require extensions of the chase, or of tree automata, to semirings.

### 3.5 The Transportation Problem for Positive Commutative Monoids

*Albert Atserias (UPC Barcelona Tech, ES)*

In the transportation problem as studied in the theory of resource allocation in economic theory, a finite set of suppliers seek to distribute their production among a finite set of consumers to completely fill their demands. As an optimization problem, the goal is to achieve so at the minimum total transportation cost. Varying the interpretation of what supply, demand, and cost is, the transportation problem can be used to model a variety of tasks from different areas, from the original resource allocation problem in economics, to the problem of image reconstruction in tomography, the relational consistency problem

in database theory, and the joint measurability problem of observables in quantum theory. Motivated by some of these applications, where the interpretations of supply and demand need not obey the arithmetic of the real numbers, we study the transportation problem in the abstract setting of positive semirings and, even more generally, in the setting of positive commutative monoids.

## 3.6     The Model Theory of Semiring Semantics

*Erich Grädel (RWTH Aachen, DE)*

Semiring semantics relies on the idea to evaluate logical statements not just by true or false, but by values in some commutative semiring. This approach has originally been motivated by the provenance analysis of database queries, but has meanwhile been extended to a systematic alternative semantics for many logical systems, in particular for first-order logic and fixed-point logics. Semiring semantics can provide very detailed information about a query or a logical statement, for instance concerning the combinations of atomic facts that imply its truth, or practical information about evaluation costs, confidence scores, access levels, or the number of successful evaluation strategies.

The development of semiring semantics raises the question to what extent classical techniques and results of logic extend to semiring semantics, and how this depends on the algebraic properties of the underlying semiring. This tutorial provides a survey on recent results in this research programme, covering aspects such as axiomatisability of finite semiring interpretations, 0-1 laws, locality, and Ehrenfeucht–Fraïssé games.

## 3.7     The algebraic semantics of many-valued logics

*Sara Ugolini (IIIA–CSIC – Bellaterra, ES)*

Many-valued logics are systems of nonclassical logics which go beyond the true/false dichotomy of classical logic. By considering more than two truth values, they provide a formal framework for reasoning with partial truth and truth degrees.

In this tutorial we give an overview of many valued logics from the perspective of their algebraic semantics. In particular, we focus on those logics whose models can be presented by means of residuated structures and their reducts; these indeed cover most of the well-known many-valued logics considered in the literature, e.g., Łukasiewicz logic, Gödel–Dummett logic, all logics arising from (left-)continuous t-norms, 3-valued Kleene logic.

In particular, we discuss how algebraic methods can be fruitfully used to study logical properties.

**References**

**1** N. Galatos, P. Jipsen, T. Kowalski, and H. Ono. *Residuated Lattices: An Algebraic Glimpse at Substructural Logics*, Studies in Logic and the Foundations of Mathematics, vol. 151, Elsevier, Amsterdam, The Netherlands, 2007.

**2** P. Hájek. *Metamathematics of Fuzzy Logics*, Kluwer Academic Publisher, Dordrecht, The Netherlands, 1998.

**3** G. Metcalfe, F. Paoli, and C. Tsinakis. *Residuated Structures in Algebra and Logic*, vol. 277. American Mathematical Society, 2023.

## 3.8 Compactness in Semiring Semantics

*Sophie Brinke (RWTH Aachen, DE)*

**Joint work of** Sophie Brinke, Anuj Dawar, Erich Grädel, Lovro Mrkonjić, Matthias Naaf

Semiring semantics for first-order logic evaluates formulae not just to true or false but to values from a commutative semiring. Depending on the underlying semiring, this allows us to track descriptions of the atomic facts that are responsible for the truth of a statement or practical information about the evaluation such as costs or confidence. Also classical semantics appears as a special case when the Boolean semiring is used, which raises the question to what extent model-theoretic results can be generalized beyond the Boolean semiring and how this relates to the algebraic properties of the underlying semiring. In this talk, we investigate the availability of compactness, which states that a set of sentences is satisfiable if every finite subset is. By defining satisfiability as the existence of non-zero valuations, this implication can be generalized to every absorptive semiring. For the, in Boolean semantics equivalent, formulation via entailment, the situation is different: Based on an order on the semiring, the entailment relation naturally extends to semiring semantics, but this yields a stronger variant of compactness, which fails for the tropical semiring, the Łukasiewicz semiring, and the semirings of generalized absorptive polynomials. However, it still holds for finite semirings and (possibly infinite) lattice semirings.

## 3.9 Codd's Theorem for Databases over Semirings

*Guillermo Badia (University of Queensland – Brisbane, AU)*

**Joint work of** Guillermo Badia, Phokion G. Kolaitis, Carles Noguera
**Main reference** Guillermo Badia, Phokion G. Kolaitis, Carles Noguera: "Codd's Theorem for Databases over Semirings", CoRR, Vol. abs/2501.16543, 2025.
**URL** https://doi.org/10.48550/ARXIV.2501.16543

Codd's Theorem, a fundamental result of database theory, asserts that relational algebra and relational calculus have the same expressive power on relational databases. We explore Codd's Theorem for databases over semirings and establish two different versions of this result for such databases: the first version involves the five basic operations of relational algebra, while in the second version the division operation is added to the five basic operations of relational algebra. In both versions, the difference operation of relations is given semantics using semirings with monus, while on the side of relational calculus a limited form of negation is

used. The reason for considering these two different versions of Codd's theorem is that, unlike the case of ordinary relational databases, the division operation need not be expressible in terms of the five basic operations of relational algebra for databases over an arbitrary positive semiring; in fact, we show that this inexpressibility result holds even for bag databases.

## 3.10    Semirings with Infinitary Operations

*Lovro Mrkonjić (RWTH Aachen, DE)*

Some applications of semirings require *infinitary semirings* with addition and multiplication operators over infinite collections of values. For example, in semiring semantics for first-order logic, existential and universal quantifiers over infinite domains are interpreted using infinitary addition and multiplication, respectively. In order to guarantee well-defined and informative semiring semantics in the infinitary setting, infinitary operations must satisfy core algebraic properties, such as distributivity and invariance under partition and re-ordering of the operands, which are analogous to associativity and commutativity from the finite setting.

We present an overview of infinitary semirings and precise definitions of their core algebraic properties, and we discuss further properties such as compactness. It turns out that some semirings do *not* admit infinitary operations, some semirings must be *extended* in order to allow infinitary operations, and other semirings, such as semirings induced by complete lattices, naturally admit infinitary operations. Finally, we outline the impact of infinitary operations on free polynomial semirings, which are used for provenance analysis in logic, databases and game theory. Our analysis shows that, with suitable definitions for infinitary semirings, large parts of the theory of semiring provenance can be successfully generalized to infinite domains.

## 3.11    Models of Computation over Semirings

*Peter Kostolányi (Comenius University – Bratislava, SK)*

The tutorial highlights a recent reinterpretation of certain parts of computational complexity theory from the viewpoint of a weighted automata theorist. It explains how the framework of weighted Turing machines over semirings can be used to capture various sorts of quantitative complexity classes – such as those of counting and optimisation problems – in a consistent manner as instances of weighted complexity classes over different semirings. Several basic results about the classical complexity classes of decision problems turn out to generalise to this weighted setting, while new appealing connections to weighted logics arise. The tutorial gives a brief overview of the most substantial observations of this kind.

## 3.12 Hoops in semiring semantics

*Tomasz Kowalski (Jagiellonian University – Kraków, PL)*

Many semirings used in semiring semantics are naturally ordered and admit an additional subtraction-like operation, often called *monus*. It turns out that multiplication-free reducts of such semirings are very well known in universal algebra and algebraic logic under the name of *dual hoops*. A dual hoop is an algebra $\mathbf{A} = (A; +, -, 0)$ such that $(A; +, 0)$ is a commutative monoid, and moreover the following identities hold:

- $x - x = 0$,
- $(x - y) + y = (y - x) + x$,
- $x - (y + z) = (x - y) - z$.

This implies that $x \leq y \iff x - y = 0$ defines an order, $-$ is a co-residual of $+$ in this order, and moreover the order is natural. In algebraic logic, these algebras are more often presented in the order-dual form, with $\cdot$, $\to$ and 1, so from now on I will always speak of hoops, relying on the reader to identify the correct order.

Hoops were first considered by Bosbach in [1, 2], then studied by Büchi and Owens in [3], and then investigated thoroughly by Blok and Ferreirim in [4]. Hoops have a number of pleasant algebraic properties, and a number of them follow from something more general than the hoop structure, namely, from the fact that they are *subtractive*. The property of subtractivity was isolated and thoroughly studied by Aglianò and Ursini in a series of articles starting from [5] (with a prequel by Ursini). The hoop structure adds natural (in the technical sense) residuated order to subtractivity, yelding algebras resembling ordered rings.

Furthermore, expanding a hoop to a semiring preserves the pleasant algebraic properties alluded to above. The theory of such expansions was studied, somewhat more generally, by Blok and Pigozzi in [6].

My talk will be a sightseeing tour, introducing such expanded hoops and presenting a few examples: in particular showing that some commonly used semirings such as $\mathbb{N}[X]$ have an implicit dual hoop structure.

### References

**1** B. Bosbach. *Komplementäre Halbgruppen. Axiomatik und Arithmetik*, Fundamenta Mathematicae 64, 257–287, 1969.
**2** B. Bosbach. *Komplementaere Halbgruppen. Kongruenzen und Quotienten*, Fundamenta Mathematicae 69, 1–14, 1970.
**3** J. R. Büchi and T. M. Owens. *Complemented monoids and hoops*, unpublished manuscript.
**4** W. J. Blok and I. M. A. Ferreirim. *Hoops and their implicational reducts*, Algebra Universalis 43, 233–257, 2000.
**5** P. Aglianò and A. Ursini. *On subtractive varieties II: General properties*, Algebra Universalis 36, issue 2, 222–259, 1996.
**6** W. J. Blok and D. Pigozzi. *On the structure of varieties with equationally definable principal congruences III*, Algebra Universalis 32, 545–608, 1994.

## 3.13   Valued Constraint Satisfaction Problem and Resilience in Database Theory

*Žaneta Semanišinová (TU Dresden, DE)*

A recent research topic in database theory is the computational complexity of resilience of queries. For a fixed conjunctive query, the problem is to compute the number of facts that need to be removed from a given database so that it does not satisfy the query. Mathematically, this problem can be viewed as removing tuples from relations of a relational structure so that it does not satisfy a fixed primitive positive sentence. In this talk, I will explain how resilience problems can be viewed as valued constraint satisfaction problems (VCSPs) of structures that are finite or at least finite-like (in the sense that they have an oligomorphic automorphism group). A VCSP is parameterized by a valued structure, which consists of a domain and cost functions, each defined on some finite power of the domain with values in the semiring of rational numbers extended with positive infinity. Building on the known results about VCSPs on finite domains, we explore how the setting can be generalized to infinite domains and give some general powerful hardness and tractability conditions for the resilience problem.

## 3.14   Polynomial-time Convergence of Datalogo over p-stable Semirings

*Hung Ngo (relationalAI – Berkeley, US)*

Datalogo is an extension of Datalog that allows for aggregation and recursion over an arbitrary commutative semiring. Unlike Datalog, the natural iterative evaluation of some Datalogo programs over some semirings may not converge. It is known that the commutative semirings for which the iterative evaluation of Datalogo programs is guaranteed to converge are exactly those semirings that are stable.

Previously, the best known upper bound on the number of iterations until convergence over $p$-stable semirings is exponential in data complexity. In this talk, we present a recent finding that the native iterative algorithm for Datalogo converges in polynomial time.

### 3.15 Not all semirings are strong

*Jacques Sakarovitch (IRIF, CNRS – Université Paris Cité, FR & LTCI, Télécom Paris, IPP, FR)*

Working with weighted automata naturally leads to the problem of the definition of the (Kleene) star operation on the semiring of (formal power) series. And proving that rational series are realised by weighted finite automata (one direction of the Kleene Theorem) may require to establish the identity: $(s_0 + s_p)^* = (s_0)^*(s_p(s_0)^*)^*$ where $s_0$ and $s_p$ are respectively the "constant term" and the "proper part" of a series $s$.

In my book "Elements of Automata Theory", I gave a proof of the above identity under the hypothesis that the weight semiring has the property that the product of two summable families is a summable family. I called such semirings "strong" and even though all semirings that I knew are strong, I stated the conjecture that there should exist some semirings which are not strong.

I have presented the problem but for want of time, I have not been able to describe the example of a semiring which is not strong that my colleague David Madore gave me and that will be the subject of a forthcoming publication. At least, I was able to announce the existence of such a semiring which justifies the title.

### 3.16 Semiring Circuits for Sum-Product Queries

*Paris Koutris (University of Wisconsin-Madison, US)*

In this talk, we ask the following question: given a Boolean Conjunctive Query, what is the smallest circuit that computes the provenance polynomial of the query over a given semiring? We answer this question by giving upper and lower bounds. We show a circuit construction that matches this bound when the semiring is idempotent. The techniques we use combine several central notions in database theory: provenance polynomials, tree decompositions, and disjunctive Datalog programs. We extend our results to lower and upper bounds for formulas. Finally, we briefly present open problems related to circuit constructions.

## 3.17 Rewriting Consistent Answers on Annotated Data and Semiring Circuits

*Jonni Virtema (University of Sheffield, GB)*

We study consistent answers of queries over databases annotated with values from a naturally ordered positive semiring. In this setting, the consistent answers of a query are defined as the minimum of the semiring values that the query takes over all repairs of an inconsistent database. The main focus is on self-join free conjunctive queries and key constraints, which is the most extensively studied case of consistent query answering over standard databases. We introduce a variant of first-order logic with a limited form of negation, define suitable semiring semantics, and establish that the consistent query answers of a self-join free conjunctive query under key constraints are rewritable in this logic if and only if the attack graph of the query contains no cycles. This result generalizes an analogous result of Koutris and Wijsen for ordinary databases, but also yields new results for a multitude of semirings, including the bag semiring, the tropical semiring, and the fuzzy semiring. Moreover, we relate our logic to particular arithmetic circuit complexity classes defined over semirings.

## 3.18 When do homomorphism counts help in query algorithms?

*Balder ten Cate (University of Amsterdam, NL)*

A query algorithm based on homomorphism counts is a procedure for determining whether a given instance satisfies a property by counting homomorphisms between the given instance and finitely many predetermined instances. In a left query algorithm, we count homomorphisms from the predetermined instances to the given instance, while in a right query algorithm we count homomorphisms from the given instance to the predetermined instances. Homomorphisms are usually counted over the semiring $\mathbb{N}$ of non-negative integers; it is also meaningful, however, to count homomorphisms over different semirings, including the Boolean semiring $\mathbb{B}$, where the homomorphism count indicates whether or not a homomorphism exists. I will present some results from joint work with Victor Dalmau, Phokion G. Kolaitis and Wei-Lin Wu (ICDT 2024), where we compare the relative expressive power of query algorithms over $\mathbb{N}$ and over $\mathbb{B}$. One of the main results asserts that if a property is closed under homomorphic equivalence, then that property admits a left query algorithm over $\mathbb{B}$ if and only if it admits a left query algorithm over $\mathbb{N}$.

## 3.19 How to Bake an Uncertainty Pie: Take Even Parts of Abstract Interpretation, K-relations, and Zonotopes and Mix Thoroughly

*Boris Glavic (University of Illinois – Chicago, US)*

Uncertainty arises naturally due to data quality issues such as missing values, constraint violations, and outliers. As the ground truth clean version of a dirty database is typically not recoverable, there is a need to reason about all possible outcomes of computations over all possible clean versions of the database and to determine which outcomes will be certainly returned by the computation irrespective of uncertainty about the ground truth clean dataset. These problems have been studied in reachability analysis in control theory and in the context of certain answers and consistent query answering in databases. As both problems are known to be computationally hard for many interesting classes of computation, there is a need for conservative approximations: under-approximate what is certainly known to be true and over-approximate what is possibly true. In this talk, I will introduce techniques for computing such approximations for relational queries, yielding a query semantics for uncertain data for full relational algebra queries with aggregation and sorting-based operations that has PTIME data complexity and an approach for training and inference with linear models where both the training data and test data may be dirty. On the side of over-approximating all possible outcomes both approaches are shown to be instances of abstract interpretation, a technique from programming languages for over-approximating a set of inputs as an element from a so-called abstract domain and for computations in the abstract domain that preserve this over-approximation. In the cases of relational queries, $K$-relations over a naturally-ordered semiring $K$ can be used as the abstract domain and standard $K$-relational query semantics is sufficient to preserve the over-approximation. To deal with aggregation and value uncertainty, the domain of these $K$-relations are intervals. In case of machine learning, the abstract domain of zonotopes, a restricted type of convex polytopes expressed as affine transformations of the unit hypercube, is used to encode uncertainty. This domain has the advantage of being able to encode correlations between values, but necessitates new methods for computing closed form solutions to models that will be briefly introduced in this talk.

### References

1   Su Feng, Boris Glavic, Oliver Kennedy: *Efficient Approximation of Certain and Possible Answers for Ranking and Window Queries over Uncertain Data*, Proceedings of the VLDB Endowment 16(6), 1346–1358, 2023.
2   Su Feng, Boris Glavic, Aaron Huber, Oliver Kennedy: *Efficient Uncertainty Tracking for Complex Queries with Attribute-Level Bounds*, Proceedings of the 46th International Conference on Management of Data, 528–540, 2021.
3   Su Feng, Aaron Huber, Boris Glavic, Oliver Kennedy: *Uncertainty Annotated Databases - a Lightweight Approach for Approximating Certain Answers*, Proceedings of the 44th International Conference on Management of Data, 1313–1330, 2019.

## 3.20 Complete Additively Idempotent Semirings and their Applications in Weighted Automata Theory, Social Network Analysis and Modal Logics

*Miroslav Ćirić (University of Niš – Serbia, RS)*

An additively-idempotent semiring is a semiring whose addition is an idempotent operation. In such a semiring, one can define a partial ordering, called the natural partial ordering, which is compatible both with addition and multiplication, and has the zero as the smallest element. Relative to the natural partial ordering, an additively-idempotent semiring is positively ordered, and besides, it forms an upper semilattice in which the binary supremum operation coincides with addition. In the case when that upper semilattice is complete, in the sense that every subset, whether finite or infinite, has a supremum, then it is convenient to treat infinite suprema as infinite sums, seeing that they satisfy all properties from the definition of infinite addition. If, in addition, infinite distributive laws hold, then we say that the considered semiring is a complete additively-idempotent semiring. The class of such semirings is quite broad and includes many important types of semirings, such as complete max-plus semirings, complete min-plus (tropical) semirings, as well as various semirings that are semiring reducts of algebras widely used in fuzzy set theory.

The key property of complete additively-idempotent semirings is the existence of residuation, that is, the existence of a pair of binary operations adjoined to multiplication, which enable certain types of inequations and equations to be solved in such a semiring. In addition, residuation is transferred to matrices and vectors over such a semiring, which gives the possibility to solve various inequalities and equations involving matrices and vectors. The purpose of this talk is to show how residuation can be applied in solving some very important problems in weighted automata theory (containment and equivalence, simulations and bisimulations, state reduction), in social network analysis (positional analysis, blockmodeling) and modal logics.

## 3.21 On the Boolean Closure of Deterministic Tree Automata

*Wolfgang Thomas (RWTH Aachen, DE)*

The tree languages recognizable by deterministic top-down tree automata form a proper subclass $D$ of the class of regular tree languages, and this proper containment also applies to its (strictly larger) Boolean closure $\mathrm{BC}(D)$. We present two results around the decades-long open problem whether for regular tree languages one can decide membership in $\mathrm{BC}(D)$. First, a characterization of $\mathrm{BC}(D)$ by a natural extension of deterministic top-down tree automata is presented, giving a convenient method to show that certain regular tree languages are outside $\mathrm{BC}(D)$. Secondly, it is shown that, for fixed $k$, it is decidable whether a regular tree language is a Boolean combination of $k$ tree languages of $D$.

## 4 Open problems

### 4.1 Circuit Size for Reachability

*Paris Koutris (University of Wisconsin-Madison, US)*

Let $K_n$ be the complete graph over $n$ vertices. Consider the provenance polynomial that sums over all simple paths from node 1 to node $n$ in $K_n$, and for every path takes the product of the weights $x_{ij}$ of every edge $\{i, j\}$ in the path. We interpret sum and product over an absorptive semiring (absorptive means that $1 + x = 1$ for every element $x$).

The question is: What is the size of the smallest circuit that represents this polynomial? We know that we can construct such a circuit of size $\mathcal{O}(n^3)$ for any absorptive semiring [1]. But we do not know whether this is optimal, or whether we can show a matching lower bound. The lower bound is open even for the case of the tropical semiring. The best lower bound is the trivial $\Omega(n^2)$ bound that says that we need to see all input edges.

**References**
**1** Stasys Jukna. *Lower Bounds for Tropical Circuits and Dynamic Programs*, Theory Comput. Syst. 57(1): 160–194, 2015. DOI: 10.1007/s00224-014-9574-4

### 4.2 Convergence rate of Datalogo over p-stable semirings

*Hung Ngo (relationalAI – Berkeley, US)*

In the paper "Polynomial Time Convergence of the Iterative Evaluation of Datalogo Programs" to be presented at PODS 2025, we proved that a Datalogo program converges after a polynomial number of steps. The convergence time is roughly $\mathcal{O}(pn^4)$, where $n$ is the output size. The best known lower bound is $\Omega((p+1)n)$.

The open problem is to close this gap. We think that an upper bound of $\mathcal{O}((p+1)n)$ is possible.

### 4.3 Subreducts of monus semirings

*Paolo Aglianò (University of Siena, IT)*

Many semirings have a natural structure of "semiring with monus", i.e. they possess a term operation which acts as a dual implication and it forms a residuated pair with + w.r.t. the natural ordering. However always asking such term to be present can cause unwanted problems with the current understanding of the theory. Of course it is possible to formulate the definitions in such a way that a semiring with monus is an algebra (in the universal algebraic sense). This suggests that a way out could be solving the following problem.

▶ **Problem.** Characterize those semirings that are subreducts of semirings with monus. Alternatively, characterize those semirings that are embeddable in a semiring with monus.

This could be useful because
1. from the general theory we know that this class is at least a quasivariety, so it has strong structural properties;
2. the embedding allows one to use the monus operation "as if it were present", but without having it interfere with the (quasi)equational properties of the class.


## 4.4 Weighted First-Order Definable Languages

*Manfred Droste (Universität Leipzig, DE)*

▶ **Theorem.** *For $L \subseteq \Sigma^*$, the following statements are equivalent.*
1. *$L$ is star-free.*
2. *$L$ is aperiodic.*
3. *$L$ is FO-definable.*
4. *$L$ is LTL-definable.*

The equivalences of statements 1 and 2, 2 and 3 respectively 3 and 4 are fundamental theorems of Schützenberger [1], McNaughton and Papert [2] resp. Kamp [3]. A proof of these and more equivalences can be found in the survey by Diekert and Gastin [4]. This raises the question whether analogous results hold for weighted languages $s \colon \Sigma^* \to \mathcal{S}$, where $\mathcal{S}$ is a semiring.

1′  It is still unclear how statement 1 shall be translated to the weighted setting. A star-free language in the classical sense is a language that can be constructed from finite languages using Boolean operators, including complementation, and concatenation, but no Kleene star. In the semiring setting, the problem arises what to do with the complement.
2′  For statement 2, an adequate translation is already known: $s$ is called aperiodic if $s = [\![\mathfrak{A}]\!]$ for some aperiodic, polynomially ambiguous NFA $\mathfrak{A}$. Polynomially ambiguous means that $\mathfrak{A}$ shall only admit polynomially many paths for any input.
3′  Statement (3) can be translated to: $s$ is definable in weighted FO.
4′  It is still open to define a suitable weighted version of LTL.

First results on formulations of statements 1′ and 2′ and their equivalence are contained in [5]. An equivalence of statements 2′ and 3′ was given more recently in [6].

There are natural formulations of a weighted LTL for which the implication 4′ implies 3′ is straightforward. The difficulty is to obtain an equivalence. An ideal result would be a formulation of a weighted LTL which, possibly through the implication of 4′ implies 2′, would lead to procedures for quantitative model checking, similar to the seminal developments for the unweighted case.

In summary, it is still open to find adequate formulations for statements 1′ and 4′ and to prove or disprove their equivalence to the remaining statements.

**References**

**1** M.-P. Schützenberger. *On Finite Monoids Having Only Trivial Subgroups*, Information and Control 8(2): 190–194, 1965. DOI: 10.1016/s0019-9958(65)90108-7

**2** R. McNaughton and S. Papert. *Counter-free Automata*, Research Monograph, vol. 65, MIT Press, 1971.

**3** J. A. W. Kamp. *Tense Logic and the Theory of Linear Order*, PhD Thesis, University of California, Los Angeles (UCLA), 1968.

**4** V. Diekert and P. Gastin. *First-order definable languages*, Logic and Automata: History and Perspectives (J. Flum, E. Grädel, Th. Wilke, eds.), Amsterdam University Press, 2008, pp. 261–306. ISBN: 978-90-5356-576-6.

**5** M. Droste and P. Gastin. *On aperiodic and star-free formal power series in partially commuting variables*, Formal Power Series and Algebraic Combinatorics (D. Krob, A. A. Mikhalev, A. V. Mikhalev, eds.), 12th Int. Conf., Moscow, Springer, 2000, pp. 158–169.

**6** M. Droste and P. Gastin. *Aperiodic weighted automata and weighted first-order logic*, Mathematical Foundations of Computer Science (MFCS 2019), LIPIcs, vol. 138, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019, pp. 76:1–76:15.

## 4.5 Conjunctive Query Containment under Bag Semantics

*Phokion G. Kolaitis (University of California – Santa Cruz, US & IBM Almaden Research Center – San Jose, US)*

The conjunctive query containment problem asks: given two conjunctive queries $q$ and $q'$ of the same arities, is it true that on every relational database $D$, it holds that $q(D) \subseteq q'(D)$?

In 1977, Chandra and Merlin showed that, for standard relational databases, this problem is NP-complete [1]. In 1993, Chaudhuri and Vardi raised the question of whether or not the conjunctive query containment problem is decidable for bag databases, i.e., when the databases considered are databases over the semiring $(\mathbb{N}, +, \times, 0, 1)$ of the natural numbers [2]. In spite of concerted efforts by several different groups of researchers, the conjunctive query containment problem for bag databases remains open to date. It is known, however, that it the containment problem for bag databases becomes undecidable if slightly larger classes of queries are considered, such as unions of conjunctive queries [3] or conjunctive queries that include negated equalities as atoms [4].

It should be noted that the conjunctive query containment problem is also meaningful for databases over naturally ordered semirings. Results concerning the decidability and complexity of this problem over such semirings have been obtained by T. J. Green in 2011 [5] and by Kostylev, Reutter, and Salamon in 2014 [6].

**References**

**1** A. K. Chandra and P. M. Merlin. *Optimal implementation of conjunctive queries in relational data bases*, STOC 1977: 77–90. DOI: 10.1145/800105.803397

**2** S. Chaudhuri and M. Y. Vardi. *Optimization of Real Conjunctive Queries*, PODS 1993: 59–70. DOI: 10.1145/153850.153856

**3** Y. E. Ioannidis and R. Ramakrishnan. *Containment of Conjunctive Queries: Beyond Relations as Sets*, ACM Trans. Database Syst. 20(3): 288–324, 1995. DOI: 10.1145/211414.211419

**4** T. S. Jayram, P. G. Kolaitis, E. Vee. *The Containment Problem for Real Conjunctive Queries with Inequalities*, PODS 2006: 80–89. DOI: 10.1145/1142351.1142363

**5**    T. J. Green. *Containment of Conjunctive Queries on Annotated Relations*, Theory Comput.
       Syst. 49(2): 429–459, 2011. DOI: 10.1007/s00224-011-9327-6

**6**    E. V. Kostylev, J. L. Reutter, A. Z. Salamon. *Classification of Annotation Semirings over
       Containment of Conjunctive Queries*, ACM Trans. Database Syst. 39(1): 1:1–1:39, 2014.
       DOI: 10.1145/2556524

## Participants

Samson Abramsky
University College London, GB

Paolo Aglianò
University of Siena, IT

Antoine Amarilli
Inria Lille – Nord Europe, FR

Albert Atserias
UPC Barcelona Tech, ES

Guillermo Badia
University of Queensland –
Brisbane, AU

Johanna Björklund
University of Umeå, SE

Camille Bourgaux
CNRS & ENS – Paris, FR

Sophie Brinke
RWTH Aachen, DE

Silvia Butti
University of Oxford, GB

Miroslav Ćirić
University of Niš – Serbia, RS

Víctor Dalmau
UPF – Barcelona, ES

Frank Drewes
University of Umeå, SE

Manfred Droste
Universität Leipzig, DE

Thomas Eiter
TU Wien, AT

Boris Glavic
University of Illinois –
Chicago, US

Erich Grädel
RWTH Aachen, DE

Marcel Jackson
La Trobe University –
Bundoora, AU

Phokion G. Kolaitis
University of California –
Santa Cruz, US & IBM Almaden
Research Center – San Jose, US

Peter Kostolányi
Comenius University –
Bratislava, SK

Egor Kostylev
University of Oslo, NO

Paris Koutris
University of Wisconsin-
Madison, US

Tomasz Kowalski
Jagiellonian University –
Kraków, PL

Dietrich Kuske
TU Ilmenau, DE

Filip Mazowiecki
University of Warsaw, PL

Lovro Mrkonjić
RWTH Aachen, DE

Hung Ngo
relationalAI – Berkeley, US

Carles Noguera
University of Siena, IT

Nina Pardal
University of Huddersfield, GB

Erik Paul
Universität Leipzig, DE

Vitaly Perevoshchikov
Herrischried, DE

Gaia Petreni
University of Siena, IT

Jacques Sakarovitch
IRIF, CNRS – Université Paris
Cité, FR & LTCI, Télécom Paris,
IPP, FR

Žaneta Semanišinová
TU Dresden, DE

Dan Suciu
University of Washington –
Seattle, US

Val Tannen
University of Pennsylvania –
Philadelphia, US

Balder ten Cate
University of Amsterdam, NL

Wolfgang Thomas
RWTH Aachen, DE

Sara Ugolini
IIIA–CSIC – Bellaterra, ES

Jonni Virtema
University of Sheffield, GB