2

K-Relations and Beyond

Melita Hajdinjak and Andrej Bauer
University of Ljubljana
Slovenia

1. Introduction

Although the theory of relational databases is highly developed and proves its usefulness in practice every day Garcia-Molina et al. (2008), there are situations where the relational model fails to offer adequate formal support. For instance, when querying *approximate data* Hjaltason & Brooks (2003); Minker (1998) or data within a given range of distance or *similarity* Hjaltason & Brooks (2003); Patella & Ciaccia (2009). Examples of such similarity-search applications are databases storing images, fingerprints, audio clips or time sequences, text databases with typographical or spelling errors, and text databases where we look for documents that are similar to a given document. A core component of such *cooperative* systems is a treatment of imprecise data Hajdinjak & Mihelič (2006); Minker (1998).

At the heart of a cooperative database system is a database where the data domains come equipped with a *similarity relation*, to denote degrees of similarity rather than simply 'equal' and 'not equal'. This notion of similarity leads to an extension of the relational model where data can be annotated with, for instance, boolean formulas (as in incomplete databases) Calì et al. (2003); Van der Meyden (1998), membership degrees (as in fuzzy databases) Bordogna & Psaila (2006); Yazici & George (1999), event tables (as in probabilistic databases) Suciu (2008), timestamps (as in temporal databases) Jae & Elmasri (2001), sets of contributing tuples (as in the context of data warehouses and the computation of lineages or why-provenance) Cui et al. (2000); Green et al. (2007), or numbers representing the multiplicity of tuples (as in the context of bag semantics) Montagna & Sebastiani (2001). Querying such *annotated* or *tagged relations* involves the generalization of the classical relational algebra to perform corresponding operations on the annotations (tags).

There have been many attempts to define extensions of the relational model to deal with similarity querying. Most utilize fuzzy logic Zadeh (1965), and the annotations are typically modelled by a membership function to the unit interval, $[0, 1]$ Ma (2006); Penzo (2005); Rosado et al. (2006); Schmitt & Schulz (2004), although there are generalizations where the membership function instead maps to an algebraic structure of some kind (typically poset or lattice based) Belohlávek & V. Vychodil (2006); Peeva & Kyosev (2004); Shenoi & Melton (1989). Green et al. Green et al. (2007) proposed a general data model (referred to as the *$\mathcal{K}$-relation model*) for annotated relations. In this model tuples in a relation are annotated with a value taken from a *commutative semiring*, $\mathcal{K}$. The resulting positive relational algebra, $\mathrm{RA}^+_{\mathcal{K}}$, generalizes Codd's classic relational algebra Codd (1970), the bag algebra Montagna & Sebastiani (2001), the relational algebra on *c*-tables Imielinski & Lipski (1984), the probabilistic algebra on event tables Suciu (2008), and the provenance algebra Buneman et al. (2001); Cui et al. (2000). With relatively little work, the $\mathcal{K}$-relation model is also suitable as a basis for modelling data with similarities and simple, positive similarity queries Hajdinjak & Bierman (2011).

Geerts and Poggi Geerts & Poggi (2010) extended the positive relational algebra $\mathrm{RA}^+_{\mathcal{K}}$ with a difference operator, which required restricting the class of commutative semirings to commutative semirings with *monus* or *m-semirings*. Because the monus-based difference operator yielded the wrong answer for two semirings important for similarity querying, a different approach to modelling negative queries in the $\mathcal{K}$-relation model was proposed Hajdinjak & Bierman (2011). It required restricting the class of commutative semirings to commutative semirings with *negation* or *n-semirings*. In order to satisfy *all* of the classical relational identities (including the idempotence of union and self-join), Hajdinjak and Bierman Hajdinjak & Bierman (2011) made another restriction; for the annotation structure they chose *De Morgan frames*. In addition, since previous attempts to formalize similarity querying and the $\mathcal{K}$-relation model all suffered from an expressivity problem allowing only one annotation structure per relation (every tuple is annotated with a value), the *$\mathcal{D}$-relation* model was proposed in which every tuple is annotated with a tuple of values, one per attribute, rather than a single value.

Relying on the work on $\mathcal{K}$, $\mathcal{L}$-and $\mathcal{D}$-relations, we make some further steps towards a general model of annotated relations. We come to the conclusion that complete distributive lattices with finite meets distributing over arbitrary joins may be chosen as a general annotation structure. This choice covers the classical relations Codd (1970), relations on bag semantics Green et al. (2007); Montagna & Sebastiani (2001) Fuhr-Rölleke-Zimányi probabilistic relations Suciu (2008), provenance relations Cui et al. (2000); Green et al. (2007), Imielinksi-Lipski relations on *c*-tables Imielinski & Lipski (1984), and fuzzy relations Hajdinjak & Bierman (2011); Rosado et al. (2006). We also aim to define a general framework of $\mathcal{K}$, $\mathcal{L}$-and $\mathcal{D}$-relations in which all the previously considered kinds of annotated relations are modeled correctly. Our studies result in an attribute-annotated model of so called $\mathcal{C}$-relations, in which some freedom of choice when defining the relational operations is given.

This chapter is organized as follows. In §2 we recall the definitions of $\mathcal{K}$-relations and the positive relational algebra $\mathrm{RA}^+_{\mathcal{K}}$, along with $\mathrm{RA}^+_{\mathcal{K}}(\setminus)$, its extension to support negative queries. Section §3 recalls the definition of the tuple-annotated $\mathcal{L}$-relation model, the aim of which was to include similarity relations into the $\mathcal{K}$-relation framework of annotated relations. In §4 we present the attribute-annotated $\mathcal{D}$-relation model, where every attribute is associated with its own annotation domain, and we study the properties of the resulting calculus of relations. In section §5 we explore whether there is a common domain of annotations suitable for all forms of annotated relations, and we define a general $\mathcal{C}$-relation model. The final section §6 discusses the issue of ranking the annotated answers, and it gives some guidelines of future work.

2. The $\mathcal{K}$-relation model

In this section we recall the definitions of $\mathcal{K}$-relations and the positive relational algebra $\mathrm{RA}^+_{\mathcal{K}}$, along with $\mathrm{RA}^+_{\mathcal{K}}(\setminus)$, its extension to support negative queries. The aim of the $\mathcal{K}$-relation work was to provide a generalized framework capable of capturing various forms of annotated relations.

We first assume some base domains, or *types*, commonly written as τ, which are simply sets of ground values, such as integers and strings. Like the authors of previous work Geerts &

Poggi (2010); Green et al. (2007); Hajdinjak & Bierman (2011), we adopt the named-attribute approach, so a *schema*,

$$U = \{a_1\colon \tau_1, \ldots, a_n\colon \tau_n\}, \tag{1}$$

is a finite map from *attribute names* a_i to their types or domains

$$U(a_i) = \tau_i. \tag{2}$$

We represent an *U-tuple* as a map

$$t = \{a_1\colon v_1, \ldots, a_n\colon v_n\} \tag{3}$$

from attribute names a_i to values v_i of the corresponding domain, i.e.,

$$t(a_i) = v_i, \tag{4}$$

where $v_i \in \tau_i$ for $i = 1, \ldots, n$. We denote the set of all U-tuples by U-Tup.

2.1 Positive relational algebra $\mathrm{RA}^+_{\mathcal{K}}$

Consider generalized relations in which the tuples are annotated (tagged) with information of various kinds. A notationally convenient way of working with annotated relations is to model tagging by a function on all possible tuples. Green et al. Green et al. (2007) argue that the generalization of the positive relational algebra to annotated relations requires that the set of tags is a *commutative semiring*.

Recall that a *semiring*

$$\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1}) \tag{5}$$

is an algebraic structure with two binary operations (sum $\oplus$ and product $\odot$) and two distinguished elements ($\mathbf{0} \neq \mathbf{1}$) such that $(K, \oplus, \mathbf{0})$ is a commutative monoid[1] with identity element 0, $(K, \odot, \mathbf{1})$ is a monoid with identity element $\mathbf{1}$, products distribute over sums, and $\mathbf{0} \odot a = a \odot \mathbf{0} = \mathbf{0}$ for any $a \in K$ (i.e., $\mathbf{0}$ is an annihilating element). A semiring $\mathcal{K}$ is called commutative if monoid $(K, \odot, \mathbf{1})$ is commutative.

Definition 2.1 ($\mathcal{K}$-relation Green et al. (2007))**.** *Let $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$ be a commutative semiring. A $\mathcal{K}$-relation over a schema $U = \{a_1\colon \tau_1, \ldots, a_n\colon \tau_n\}$ is a function $A\colon U\text{-}Tup \to K$ such that its support,*

$$\mathrm{supp}(A) = \{t \mid A(t) \neq \mathbf{0}\}, \tag{6}$$

is finite.

Taking this extension of relations, Green et al. proposed a natural lifting of the classical relational operators over $\mathcal{K}$-relations. The tuples considered to be 'in' the relation are tagged with $\mathbf{1}$ and the tuples considered to be 'out of' the relation are tagged with $\mathbf{0}$. The binary operation $\oplus$ is used to deal with union and projection and therefore to combine different tags of the same tuple into one tag. The binary operation $\odot$ is used to deal with natural join and therefore to combine the tags of joinable tuples.

Definition 2.2 (Positive relational algebra on $\mathcal{K}$-relations Green et al. (2007))**.** *Suppose $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$ is a commutative semiring. The operations of the positive relational algebra on $\mathcal{K}$, denoted $RA^+_{\mathcal{K}}$, are defined as follows:*

[1] A monoid consists of a set equipped with a binary operation that is associative and has an identity element.

Empty relation: *For any set of attributes U, there is $\emptyset_U : U\text{-}Tup \to K$ such that*

$$\emptyset_U(t) \stackrel{\text{def}}{=} \mathbf{0} \tag{7}$$

for all U-tuples t.[2]

Union: *If $A, B \colon U\text{-}Tup \to K$, then $A \cup B \colon U\text{-}Tup \to K$ is defined by*

$$(A \cup B)(t) \stackrel{\text{def}}{=} A(t) \oplus B(t). \tag{8}$$

Projection: *If $A \colon U\text{-}Tup \to K$ and $V \subset U$, we write $f \downarrow V$ to be the restriction of the map f to the domain V. The projection $\pi_V A \colon V\text{-}Tup \to K$ is defined by*

$$(\pi_V A)(t) \stackrel{\text{def}}{=} \sum_{(t' \downarrow V) = t \text{ and } A(t') \neq \mathbf{0}} A(t'). \tag{9}$$

Selection: *If $A \colon U\text{-}Tup \to K$ and the selection predicate $\mathbf{P}$ maps each U-tuple to either $\mathbf{0}$ or $\mathbf{1}$, then $\sigma_{\mathbf{P}} A : U\text{-}Tup \to K$ is defined by*

$$(\sigma_{\mathbf{P}} A)(t) \stackrel{\text{def}}{=} A(t) \odot \mathbf{P}(t). \tag{10}$$

Join: *If $A \colon U_1\text{-}Tup \to K$ and $B \colon U_2\text{-}Tup \to K$, then $A \bowtie B$ is the $\mathcal{K}$-relation over $U_1 \cup U_2$ defined by*

$$(A \bowtie B)(t) \stackrel{\text{def}}{=} A(t \downarrow U_1) \odot B(t \downarrow U_2). \tag{11}$$

Renaming: *If $A \colon U\text{-}Tup \to K$ and $\beta \colon U \to U'$ is a bijection, then $\rho_\beta A \colon U'\text{-}Tup \to K$ is defined by*

$$(\rho_\beta A)(t) \stackrel{\text{def}}{=} A(t \circ \beta). \tag{12}$$

Note that in the case for projection, the sum is finite since A has finite support.

The power of this definition is that it generalizes a number of proposals for annotated relations and associated query algebras.

Lemma 2.1 (Example algebras on $\mathcal{K}$-relations Green et al. (2007))**.**

1. *The classical relational algebra with set semantics Codd (1970) is given by the $\mathcal{K}$-relational algebra on the boolean semiring $\mathcal{K}_{\mathbb{B}} = (\mathbb{B}, \vee, \wedge, \mathsf{false}, \mathsf{true})$.*
2. *The relational algebra with bag semantics Green et al. (2007); Montagna & Sebastiani (2001) is given by the $\mathcal{K}$-relational algebra on the semiring of counting numbers $\mathcal{K}_{\mathbb{N}} = (\mathbb{N}, +, \cdot, 0, 1)$.*
3. *The Fuhr-Rölleke-Zimányi probabilistic relational algebra on event tables Suciu (2008) is given by the $\mathcal{K}$-relational algebra on the semiring $\mathcal{K}_{prob} = (\mathcal{P}(\Omega), \cup, \cap, \emptyset, \Omega)$ where Ω is a finite set of events and $\mathcal{P}(\Omega)$ is the powerset of Ω.*
4. *The Imielinksi-Lipski algebra on c-tables Imielinski & Lipski (1984) is given by the $\mathcal{K}$-relational algebra on the semiring $\mathcal{K}_{c\text{-}table} = (\mathsf{PosBool}(X), \vee, \wedge, \mathsf{false}, \mathsf{true})$ where $\mathsf{PosBool}(X)$ is the set of all positive boolean expressions over a finite set of variables X in which any two equivalent expressions are identified.*

[2] As is standard, we drop the subscript on the empty relation where it can be inferred by context.

5. *The provenance algebra of polynomials with variables from X and coefficients from $\mathbb{N}$ Cui et al. (2000); Green et al. (2007) is given by the $\mathcal{K}$-relational algebra on the provenance semiring* $\mathcal{K}_{prov} = (\mathbb{N}[X], +, \cdot, 0, 1)$.

The positive relational algebra $\text{RA}^+_{\mathcal{K}}$ satisfies many of the familiar relational equalities Ullman (1988; 1989).

Proposition 2.1 (Identities of $\mathcal{K}$-relations Green et al. (2007); Hajdinjak & Bierman (2011))**.** *The following identities hold for the positive relational algebra on $\mathcal{K}$-relations:*

- *union is associative, commutative, and has identity $\emptyset$;*
- *selection distributes over union and product;*
- *join is associative, commutative and distributive over union;*
- *projection distributes over union and join;*
- *selections and projections commute with each other;*
- *selection with boolean predicates gives all or nothing, $\sigma_{\mathsf{false}}(A) = \emptyset$ and $\sigma_{\mathsf{true}}(A) = A$;*
- *join with an empty relation gives an empty relation, $A \bowtie \emptyset_U = \emptyset_U$ where A is a $\mathcal{K}$-relation over a schema U;*
- *projection of an empty relation gives an empty relation, $\pi_V(\emptyset) = \emptyset$.*

It is important to note that the properties of idempotence of union, $A \cup A = A$, and self-join, $A \bowtie A = A$, are missing from this list. These properties fail for the bag semantics and provenance, so they fail to hold for the more general model.

Green et al. only considered positive queries and left open the problem of supporting negative query operators.

2.2 Relational algebra $\text{RA}^+_{\mathcal{K}}(\setminus)$

Geerts and Poggi Geerts & Poggi (2010) recently proposed extending the $\mathcal{K}$-relation model by a difference operator following a standard approach for introducing a monus operator into an additive commutative monoid Amer (1984). First, they restricted the class of commutative semirings by requiring that every semiring additionally satisfy the following pair of conditions.

Definition 2.3 (GP-conditions Geerts & Poggi (2010))**.** *A commutative semiring* $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$ *is said to satisfy the* GP conditions *if the following two conditions hold.*

1. *The preorder $x \preceq y$ on K defined as*

$$x \preceq y \text{ iff there exists a } z \in K \text{ such that } x \oplus z = y \tag{13}$$

   is a partial order.[3]

2. *For each pair of elements $x, y \in K$, the set $\{z \in K; x \preceq y \oplus z\}$ has a smallest element. (As $\preceq$ defines a partial order, this smallest element must be unique, if it exists.)*

Definition 2.4 (m-semiring Geerts & Poggi (2010))**.** *Let* $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$ *be a commutative semiring that satisfies the GP conditions. For any $x, y \in K$, we define $x \ominus y$ to be the smallest element z such that $x \preceq y \oplus z$. A (commutative) semiring $\mathcal{K}$ that can be equipped with a* monus *operator $\ominus$ is called a* semiring with monus *or* m-semiring.

[3] While a *preorder* is a binary relation that is reflexive and transitive, a *partial order* is a binary relation that is refleksive, transitive, and antisymmetric.

Geerts and Poggi identified two equationally complete classes in the variety of m-semirings, namely

(1) m-semirings that are a boolean algebra (i.e., complemented distributive lattice with distinguished elements $\mathbf{0}$ and $\mathbf{1}$), for which the monus behaves like set difference, and
(2) m-semirings that are the positive cone of a lattice-ordered commutative ring, for which the monus behaves like the truncated minus of the natural numbers.

Recall that a *lattice-ordered ring* (or l-ring) is an algebraic structure $\mathcal{K} = (K, \vee, \wedge, \oplus, -, \mathbf{0}, \odot)$ such that $(K, \vee, \wedge)$ is a lattice, $(K, \oplus, -, \mathbf{0}, \odot)$ is a ring, operation $\oplus$ is order-preserving, and for $x, y \geq \mathbf{0}$ we have $x \odot y \geq \mathbf{0}$. An l-ring is commutative if the multiplication operation $\odot$ is commutative. The set of elements x for which $\mathbf{0} \leq x$ is called the *positive cone* of the l-ring.

Lemma 2.2 (Example m-semirings Geerts & Poggi (2010))**.**

1. *The boolean semiring,* $\mathcal{K}_{\mathbb{B}} = (\mathbb{B}, \vee, \wedge, \mathsf{false}, \mathsf{true})$*, is a boolean algebra. We have*

$$\mathsf{false} \ominus \mathsf{false} = \mathsf{false},\ \mathsf{false} \ominus \mathsf{true} = \mathsf{false},\ \mathsf{true} \ominus \mathsf{false} = \mathsf{true},\ \mathsf{true} \ominus \mathsf{true} = \mathsf{false}. \tag{14}$$

2. *The semiring of counting numbers,* $\mathcal{K}_{\mathbb{N}} = (\mathbb{N}, +, \cdot, 0, 1)$*, is the positive cone of the ring of integers,* $\mathbb{Z}$*. The monus corresponds to the truncated minus,*

$$x \ominus y = \max\{0, x - y\}. \tag{15}$$

3. *The probabilistic semiring,* $\mathcal{K}_{prob} = (\mathcal{P}(\Omega), \cup, \cap, \emptyset, \Omega)$*, is a boolean algebra. The monus corresponds to set difference,*

$$X \ominus Y = X \setminus Y. \tag{16}$$

4. *In the case of the semiring of c-tables,* $\mathcal{K}_{c\text{-}table} = (\mathsf{PosBool}(X), \vee, \wedge, \mathsf{false}, \mathsf{true})$*, the monus cannot be defined unless negated literals are added to the base set, in which case we get a boolean algebra. For any two expressions* $\phi_1, \phi_2 \in Bool(X)$ *we then have*

$$\phi_1 \ominus \phi_2 = \phi_1 \wedge \neg\phi_2, \tag{17}$$

where negation $\neg$ *over boolean expressions takes truth to falsity, and vice versa, and it interchanges the meet and the join operation.*

5. *The provenance semiring,* $\mathcal{K}_{prov} = (\mathbb{N}[X], +, \cdot, 0, 1)$*, is the positive cone of the ring of polynomials from* $\mathbb{Z}[X]$*. The monus of two polynomials* $f[X] = \sum_{\alpha \in I} f_\alpha x^\alpha$ *and* $g[X] = \sum_{\alpha \in I} g_\alpha x^\alpha$*, where* I *is a finite subset of* $\mathbb{N}^n$*, corresponds to*

$$f[X] \ominus g[X] = \sum_{\alpha \in I} (f_\alpha \dot{-} g_\alpha) x^\alpha, \tag{18}$$

where $\dot{-}$ *denotes the truncated minus on* $\mathbb{N}$*.*

Given an m-semiring, the positive relational algebra $\mathrm{RA}^+_{\mathcal{K}}$ can be extended with the missing difference operator as follows.

Definition 2.5 (Relational algebra on $\mathcal{K}$-relations Geerts & Poggi (2010))**.** *Let* $\mathcal{K}$ *be an m-semiring. The algebra* $RA^+_{\mathcal{K}}(\setminus)$ *is obtained by extending* $RA^+_{\mathcal{K}}$ *with the operator:*

Difference *If* $A, B : U\text{-}Tup \to K$*, then the difference* $A \setminus B : U\text{-}Tup \to K$ *is defined by*

$$(A \setminus B)(t) \stackrel{\text{def}}{=} A(t) \ominus B(t). \tag{19}$$

Geerts and Poggi show that their resulting algebra coincides with the classical relational algebra, the bag algebra with the monus operator, the probabilistic relational algebra on event tables, the relational algebra on c-tables, and the provenance algebra.

3. The $\mathcal{L}$-relation model

In this section we recall the definition of the $\mathcal{L}$-relation model, the aim of which was to include similarity relations into the general $\mathcal{K}$-relation framework of annotated relations.

3.1 Domain similarities

In a similarity context it is typically assumed that all data domains come equipped with a similarity relation or similarity measure.

Definition 3.1 (Similarity measures Hajdinjak & Bierman (2011))**.** *Given a type τ and a commutative semiring $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$, a* similarity measure *is a function $\rho\colon \tau \times \tau \to K$ such that ρ is reflexive, i.e. $\rho(x, x) = \mathbf{1}$.*

Following earlier work Shenoi & Melton (1989), only reflexivity of the similarity measure was required. Other properties don't hold in general Hajdinjak & Bauer (2009). For example, symmetry does not hold when similarity denotes driving distance between two points in a town because of one-way streets. Another property is transitivity, but there are a number of non-transitive similarity measures, e.g. when similarity denotes likeness between two colours.

Allowing only $\mathcal{K}$-valued similarity relations, Hajdinjak and Bierman Hajdinjak & Bierman (2011) modeled an answer to a query as a $\mathcal{K}$-relation in which each tuple is tagged by the similarity value between the tuple and the *ideal tuple*. (By an ideal tuple a tuple that perfectly fits the requirements of the similarity query is meant.) Prior to any querying, it is assumed that each U-tuple t has either desirability $A(t) = \mathbf{1}$ or $A(t) = \mathbf{0}$ whether it is in or out of A.

Example 3.1 (Common similarity measures)**.** *Three common examples of similarity measures are as follows.*

1. *An equality measure $\rho\colon \tau \times \tau \to \mathbb{B}$ where $\rho(x, y) \stackrel{\text{def}}{=} \mathsf{true}$ if x and y are equal and false otherwise. Here, $\mathbb{B} = \{\mathsf{false}, \mathsf{true}\}$ is the underlying set of the commutative semiring*

$$\mathcal{K}_{\mathbb{B}} = (\mathbb{B}, \vee, \wedge, \mathsf{false}, \mathsf{true}), \tag{20}$$

   called the boolean semiring.

2. *A fuzzy equality measure $\rho\colon \tau \times \tau \to [0, 1]$ where $\rho(x, y)$ expresses the degree of equality of x and y; the closer x and y are to each other, the closer $\rho(x, y)$ is to 1. Here, the unit interval $[0, 1]$ is the underlying set of the commutative semiring*

$$\mathcal{K}_{[0,1]} = ([0, 1], \max, \min, 0, 1), \tag{21}$$

   called the fuzzy semiring.

3. *A distance measure $\rho\colon \tau \times \tau \to [0, d_{max}]$ where $\rho(x, y)$ is the distance from x to y. Here, the closed interval $[0, d_{max}]$ is the underlying set of the commutative semiring*

$$\mathcal{K}_{[0,d_{max}]} = ([0, d_{max}], \min, \max, d_{max}, 0), \tag{22}$$

   called the distance semiring.

Because of their use the commutative semirings from this example were called similarity semirings.

A predefined environment of similarity measures that can be used for building queries is assumed—for every domain $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$ and every $\mathcal{K}$-relation over a schema $U = \{a_1 \colon \tau_1, ..., a_n \colon \tau_n\}$ there are similarity measures

$$\rho_{a_i} \colon \tau_i \times \tau_i \to K, 1 \leq i \leq n. \tag{23}$$

3.2 The selection predicate

In the original Green et al. model (Definition 2.2) the selection predicate maps U-tuples to either the zero or the unit element of the semiring. Since in a similarity context we expect the selection predicate to reflect the relevance or the degree of membership of a particular tuple in the answer relation, not just the two possibilities of full membership ($\mathbf{1}$) or non-membership ($\mathbf{0}$), the following generalization to the original definition was proposed Hajdinjak & Bierman (2011).

Selection: If $A \colon U\text{-Tup} \to K$ and the selection predicate

$$\mathbf{P} \colon U\text{-Tup} \to K \tag{24}$$

maps each U-tuple to an element of K (instead of mapping to either $\mathbf{0}$ or $\mathbf{1}$), then $\sigma_{\mathbf{P}} A \colon U\text{-Tup} \to K$ is (still) defined by

$$(\sigma_{\mathbf{P}} A)(t) = A(t) \odot \mathbf{P}(t). \tag{25}$$

Selection queries can now be classified on whether they are based on the attribute values (as is normal in non-similarity queries) or whether they use the similarity measures. Selection queries can also use constant values.

Definition 3.2 (Primitive predicate Hajdinjak & Bierman (2011))**.** *Suppose in a schema $U = \{a_1 \colon \tau_1, \ldots, a_n \colon \tau_n\}$ the types of attributes a_i and a_j coincide. Then given a commutative semiring $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$, for a given binary predicate θ, the* primitive predicate *$[a_i \, \theta \, a_j] \colon U\text{-}Tup \to K$ is defined as follows.*

$$[a_i \, \theta \, a_j](t) \stackrel{\text{def}}{=} \chi_{a_i \theta a_j}(t) = \begin{cases} \mathbf{1} & \text{if } t(a_i) \, \theta \, t(a_j), \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{26}$$

In words, $[a_i \, \theta \, a_j]$ behaves as the characteristic map of θ, where θ may be any arithmetic comparison operator among $=, \neq, <, >, \leq, \geq$.

Definition 3.3 (Similarity predicate Hajdinjak & Bierman (2011))**.** *Suppose in a schema $U = \{a_1 \colon \tau_1, \ldots, a_n \colon \tau_n\}$ the types of attributes a_i and a_j coincide. Given a commutative semiring $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$, the* similarity predicate *$[a_i \texttt{ like } a_j] \colon U\text{-}Tup \to K$ is defined as follows.*

$$[a_i \texttt{ like } a_j](t) \stackrel{\text{def}}{=} \rho_{a_i}(t(a_i), t(a_j)). \tag{27}$$

A symmetric version is as follows.

$$[a_i \sim a_j] \stackrel{\text{def}}{=} [a_i \texttt{ like } a_j] \cup [a_j \texttt{ like } a_i], \tag{28}$$

where union ($\cup$) of selection predicates is defined below.

Definition 3.4. *Given a commutative semiring $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$, union and intersection of two selection predicates $\mathbf{P}_1, \mathbf{P}_2 \colon$ U-Tup $\to K$ is defined as follows.*

$$(\mathbf{P}_1 \cup \mathbf{P}_2)(t) \stackrel{\text{def}}{=} \mathbf{P}_1(t) \oplus \mathbf{P}_2(t), \tag{29}$$

$$(\mathbf{P}_1 \cap \mathbf{P}_2)(t) \stackrel{\text{def}}{=} \mathbf{P}_1(t) \odot \mathbf{P}_2(t). \tag{30}$$

3.3 Relational difference

Whilst the similarity semirings support a monus operation in the sense of Geerts and Poggi Geerts & Poggi (2010), the induced difference operator in the relational algebra does not behave as desired.

- The fuzzy semiring, $\mathcal{K}_{[0,1]} = ([0,1], \max, \min, 0, 1)$, satisfies the GP conditions, and the monus operator is as follows.

$$x \ominus y = \min\{z \in [0,1]; x \leq \max\{y,z\}\} = \begin{cases} 0 & \text{if } x \leq y, \\ x & \text{if } x > y. \end{cases} \tag{31}$$

  This induces the following difference operator in the relational algebra.

$$(A \setminus B)(t) = \begin{cases} 0 & \text{if } A(t) \leq B(t), \\ A(t) & \text{if } A(t) > B(t). \end{cases} \tag{32}$$

  Hajdinjak and Bierman Hajdinjak & Bierman (2011) regret that this is not the expected definition. First, fuzzy set difference is universally defined as $\min\{A(t), 1 - B(t)\}$ Rosado et al. (2006). Secondly, in similarity settings only totally irrelevant tuples should be annotated with 0 and excluded as a possible answer Hajdinjak & Mihelič (2006). In the case of the fuzzy set difference $A \setminus B$, these are exclusively those tuples t where $A(t) = 0$ or $B(t) = 1$, and certainly not where $A(t) \leq B(t)$.
- The distance semiring, $\mathcal{K}_{[0,d_{\max}]} = ([0, d_{\max}], \min, \max, d_{\max}, 0)$, satisfies the GP-conditions, and the monus operator is as follows.

$$x \ominus y = \max\{z \in [0, d_{\max}]; x \geq \min\{y,z\}\} = \begin{cases} d_{\max} & \text{if } x \geq y, \\ x & \text{if } x < y. \end{cases} \tag{33}$$

  This induces the following difference operator in the relational algebra.

$$(A \setminus B)(t) = \begin{cases} d_{\max} & \text{if } A(t) \geq B(t), \\ A(t) & \text{if } A(t) < B(t). \end{cases} \tag{34}$$

  Again, in the distance setting, we would expect the difference operator to be defined as $\max\{A(t), d_{\max} - B(t)\}$. Moreover, this is a continuous function in contrast to the step function behaviour of the operator above resulting from the monus definition.

Rather than using a monus-like operator, Hajdinjak and Bierman Hajdinjak & Bierman (2011) proposed a different approach using *negation*.

Definition 3.5 (Negation)**.** *Given a set L equipped with a preorder, a* negation *is an operation $\neg :$ $L \to L$ that reverts order, $x \leq y \implies \neg y \leq \neg x$, and is involutive, $\neg\neg x = x$.*

Definition 3.6 (n-semiring Hajdinjak & Bierman (2011))**.** *A (commutative) n-semiring* $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1}, \neg)$ *is a (commutative) semiring* $(K, \oplus, \odot, \mathbf{0}, \mathbf{1})$ *equipped with negation,* $\neg\colon K \to K$ *(with respect to the preorder on K).*

Provided that $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1}, \neg)$ is a commutative n-semiring, the difference of $\mathcal{K}$-relations $A, B : U\text{-Tup} \to K$ may be defined by

$$(A \setminus B)(t) \stackrel{\text{def}}{=} A(t) \odot \neg B(t). \tag{35}$$

Each of the similarity semirings has a negation operation that, in contrast to the monus, gives the expected notion of relational difference.

Example 3.2 (Relational difference over common similarity measures)**.**

- *In the boolean semiring,* $\mathcal{K}_{\mathbb{B}} = (\mathbb{B}, \vee, \wedge, \mathsf{false}, \mathsf{true})$*, negation can be defined as complementation.*

$$\neg x \stackrel{\text{def}}{=} \begin{cases} \mathsf{true} & \text{if } x = \mathsf{false}, \\ \mathsf{false} & \text{if } x = \mathsf{true}. \end{cases} \tag{36}$$

  From the above we get exactly the monus-based difference of $\mathcal{K}_{\mathbb{B}}$*-relations.*

$$A(t) \odot \neg B(t) = A(t) \ominus B(t) = \begin{cases} \mathsf{false} & \text{if } B(t) = \mathsf{true}, \\ A(t) & \text{if } B(t) = \mathsf{false}. \end{cases} \tag{37}$$

- *In the fuzzy semiring,* $\mathcal{K}_{[0,1]} = ([0,1], \max, \min, 0, 1)$*, ordered by relation* $\leq$*, we can define a negation operator as*

$$\neg x \stackrel{\text{def}}{=} 1 - x. \tag{38}$$

  In the generalized fuzzy semiring $\mathcal{K}_{[a,b]} = ([a,b], \max, \min, a, b)$*, we can define* $\neg x \stackrel{\text{def}}{=} a + b - x$*. In the fuzzy semiring we thus get*

$$A(t) \odot \neg B(t) = \min\{A(t), 1 - B(t)\}, \tag{39}$$

  and in the generalized fuzzy semiring we get $A(t) \odot \neg B(t) = \min\{A(t), a + b - B(t)\}$*. These coincide with the fuzzy notions of difference on* $[0,1]$ *and* $[a,b]$*, respectively Rosado et al. (2006).*
- *In the distance semiring,* $\mathcal{K}_{[0,d_{max}]} = ([0, d_{max}], \min, \max, d_{max}, 0)$*, ordered by relation* $\geq$*, we can define a negation operator as*

$$\neg x \stackrel{\text{def}}{=} d_{max} - x. \tag{40}$$

  We again get the expected notion of difference.

$$A(t) \odot \neg B(t) = \max\{A(t), d_{max} - B(t)\}. \tag{41}$$

  This is a continuous function of $A(t)$ *and* $B(t)$*, and it calculates the greatest distance* d_{max} *only if* $A(t) = d_{max}$ *or* $B(t) = 0$*.*

Moreover, the negation operation gives the same result as the monus when $\mathcal{K}$ is the boolean semiring, $\mathcal{K}_{\mathbb{B}}$, the probabilistic semiring, $\mathcal{K}_{\text{prob}}$, or the semiring on c-tables, $\mathcal{K}_{c\text{-table}}$. Unfortunately, while the provenance semiring, $\mathcal{K}_{\text{prov}}$, and the semiring of counting numbers, $\mathcal{K}_{\mathbb{N}}$, both contain a monus, neither contains a negation operation. In general, not all m-semirings are n-semirings. The opposite also holds Hajdinjak & Bierman (2011).

3.4 Relational algebra on $\mathcal{L}$-relations

We have seen that the $\mathcal{K}$-relational algebra does not satisfy the properties of idempotence of union and self-join because, in general, the sum and product operators of a semiring are not idempotent. In order to satisfy *all* the classical relational identities (including idempotence of union and self-join) and to allow a comparison and ordering of tags, Hajdinjak and Bierman Hajdinjak & Bierman (2011) have restricted commutative n-semirings to De Morgan frames (with the lattice join defined as sum and the lattice meet as product). Recall that the lattice supremum $\vee$ and infimum $\wedge$ operators are always idempotent.

Definition 3.7 (De Morgan frame Salii (1983)). *A* De Morgan frame, $\mathcal{L} = (L, \bigvee, \wedge, \mathbf{0}, \mathbf{1}, \neg)$, *is a complete lattice* $(L, \bigvee, \wedge, \mathbf{0}, \mathbf{1})$ *where finite meets distribute over arbitrary joins, i.e.,*

$$x \wedge \bigvee_i y_i = \bigvee_i (x \wedge y_i), \tag{42}$$

and $\neg \colon L \to L$ *is a negation operation.*

Proposition 3.1 (De Morgan laws Salii (1983)). *Given a De Morgan frame* $\mathcal{L} = (L, \bigvee, \wedge, \mathbf{0}, \mathbf{1}, \neg)$, *the following laws hold.*

$$\neg \mathbf{0} = \mathbf{1} \tag{43}$$

$$\neg \mathbf{1} = \mathbf{0} \tag{44}$$

$$\neg(x \vee y) = \neg x \wedge \neg y \tag{45}$$

$$\neg(x \wedge y) = \neg x \vee \neg y \tag{46}$$

The similarity semirings from Example 3.1 are De Morgan frames, the same holds for the probabilistic semiring and the semiring on c-tables.

Definition 3.8 ($\mathcal{L}$-relation Hajdinjak & Bierman (2011)). *Let* $\mathcal{L} = (L, \bigvee, \wedge, \mathbf{0}, \mathbf{1}, \neg)$ *be a De Morgan frame. An* $\mathcal{L}$-relation *over a schema* $U = \{a_1 \colon \tau_1, \ldots, a_n \colon \tau_n\}$ *is a function* $A \colon U\text{-}Tup \to L$.

Definition 3.9 (Relational algebra on $\mathcal{L}$-relations Hajdinjak & Bierman (2011)). *Suppose* $\mathcal{L} = (L, \bigvee, \wedge, \mathbf{0}, \mathbf{1}, \neg)$ *is a De Morgan frame. The operations of the relational algebra on* $\mathcal{L}$, *denoted* $RA_{\mathcal{L}}$, *are defined as follows:*

Empty relation: *For any set of attributes* U *there is* $\emptyset_U \colon U\text{-}Tup \to L$ *such that*

$$\emptyset(t) \stackrel{\text{def}}{=} \mathbf{0} \tag{47}$$

for all U*-tuples* t.

Union: *If* $A, B \colon U\text{-}Tup \to L$ *then* $A \cup B : U\text{-}Tup \to L$ *is defined by*

$$(A \cup B)(t) \stackrel{\text{def}}{=} A(t) \vee B(t). \tag{48}$$

Projection: *If* $A \colon U\text{-}Tup \to L$ *and* $V \subset U$, *the projection of* A *on attributes* V *is defined by*

$$(\pi_V A)(t) \stackrel{\text{def}}{=} \bigvee_{(t' \downarrow V) = t \text{ and } A(t') \neq \mathbf{0}} A(t'). \tag{49}$$

Selection: *If* $A \colon U\text{-}Tup \to L$ *and the selection predicate* $\mathbf{P} \colon U\text{-}Tup \to L$ *maps each* U*-tuple to an element of* $\mathcal{L}$, *then* $\sigma_{\mathbf{P}} A \colon U\text{-}Tup \to L$ *is defined by*

$$(\sigma_{\mathbf{P}} A)(t) \stackrel{\text{def}}{=} A(t) \wedge \mathbf{P}(t). \tag{50}$$

Join: *If* $A\colon U_1$*-Tup* $\to L$ *and* $B\colon U_2$*-Tup* $\to L$*, then* $A \bowtie B$ *is the* $\mathcal{L}$*-relation over* $U_1 \cup U_2$ *defined by*

$$(A \bowtie B)(t) \stackrel{\text{def}}{=} A(t) \wedge B(t). \tag{51}$$

Difference: *If* $A, B\colon U$*-Tup* $\to L$*, then* $A \setminus B\colon U$*-Tup* $\to L$ *is defined by*

$$(A \setminus B)(t) \stackrel{\text{def}}{=} A(t) \wedge \neg B(t). \tag{52}$$

Renaming: *If* $A\colon U$*-Tup* $\to L$ *and* $\beta\colon U \to U'$ *is a bijection, then* $\rho_\beta A : U'$*-Tup* $\to L$ *is defined by*

$$(\rho_\beta A)(t) \stackrel{\text{def}}{=} A(t \circ \beta). \tag{53}$$

Unlike for $\mathcal{K}$-relations, we need not require that $\mathcal{L}$-relations have finite support, since De Morgan frames are complete lattices, which quarantees the existence of the join in the definition of projection.

It is important to note that since $\text{RA}_\mathcal{L}$ satisfies *all* the main positive relational algebra identities, in terms of query optimization, all algebraic rewrites familiar from the classical (positive) relational algebra apply to $\text{RA}_\mathcal{L}$ without restriction. Matters are a little different for the negative identities Hajdinjak & Bierman (2011). In fuzzy relations Rosado et al. (2006) many of the familiar laws concerning difference do not hold. For example, it is not the case that $A \setminus A = \emptyset$, and so it is not the case in general for the $\mathcal{L}$-relational algebra. Consequently, some (negative) identities from the classical relational algebra do not hold any more.

4. The $\mathcal{D}$-relation model

Notice that all tuples across all the $\mathcal{K}$-relations or the $\mathcal{L}$-relations in the database and intermediate relations in queries must be annotated with a value from the same commutative semiring $\mathcal{K}$ or De Morgan frame $\mathcal{L}$. To support simultaneously several different similarity measures (e.g., similarity of strings, driving distance between cities, likelihood of objects to be equal), and use these different measures in our queries (even within the same query), Hajdinjak and Bierman Hajdinjak & Bierman (2011) proposed to move from a tuple-annotated model to an attribute-annotated model. They associated every attribute with its own De Morgan frame. They generalized an $\mathcal{L}$-relation, which is a map from a tuple to an annotation value from a De Morgan frame, to a *$\mathcal{D}$-relation*, which is a map from a tuple to a corresponding tuple containing an annotation value for every element in the source tuple, referred to as a *De Morgan frame tuple*.

Definition 4.1 (De Morgan frame schema, De Morgan frame tuple, $\mathcal{D}$-relation Hajdinjak & Bierman (2011))**.**

- *A* De Morgan frame schema, $\mathcal{D} = \{a_1\colon \mathcal{L}_1, ..., a_n\colon \mathcal{L}_n\}$*, maps an attribute name,* a_i*, to a De Morgan frame,* $\mathcal{L}_i = (L_{a_i}, \bigvee_{a_i}, \wedge_{a_i}, \mathbf{0}_{a_i}, \mathbf{1}_{a_i}, \neg_{a_i})$.
- *A* De Morgan frame tuple, $s = \{a_1\colon l_1, ..., a_n\colon l_n\}$*, maps an attribute name,* a_i*, to a De Morgan frame element,* l_i.
- *Given a De Morgan frame schema,* $\mathcal{D}$*, a schema* U*, then a tuple* s *is said to be a* De Morgan frame tuple matching $\mathcal{D}$ over U *if* $dom(s) = dom(U) = dom(\mathcal{D})$*. The set of all De Morgan frame tuples matching* $\mathcal{D}$ *over* U *is denoted* $\mathcal{D}(U)$*-Tup.*
- *An* $\mathcal{D}$-relation over U *is a finite map from* U*-Tup to* $\mathcal{D}(U)$*-Tup.* Its support needs *not* be finite.

Definition 4.2 (Relational algebra with similarities Hajdinjak & Bierman (2011))**.** *The operations of the* relational algebra with similarities, *$RA_{\mathcal{D}}$, are defined as follows:*

Empty relation: *For any set of attributes U and corresponding De Morgan frame schema, $\mathcal{D}$, the empty $\mathcal{D}$-relation over U, $\emptyset_U$, is defined such that*

$$\emptyset_U(t)(a) \stackrel{\text{def}}{=} \mathbf{0}_a \tag{54}$$

where t is a U-tuple and $\mathcal{D}(a) = (L_a, \bigvee_a, \wedge_a, \mathbf{0}_a, \mathbf{1}_a, \neg_a)$.

Union: *If $A, B\colon U$-Tup $\to \mathcal{D}(U)$-Tup, then $A \cup B\colon U$-Tup $\to \mathcal{D}(U)$-Tup is defined by*

$$(A \cup B)(t)(a) \stackrel{\text{def}}{=} A(t)(a) \vee_a B(t)(a) \tag{55}$$

where $\mathcal{D}(a) = (L_a, \bigvee_a, \wedge_a, \mathbf{0}_a, \mathbf{1}_a, \neg_a)$.

Projection: *If $A\colon U$-Tup $\to \mathcal{D}(U)$-Tup and $V \subset U$, the projection of A on attributes V is defined by*

$$(\pi_V A)(t)(a) \stackrel{\text{def}}{=} \bigvee\nolimits_{(t' \downarrow V)=t \text{ and } A(t')(a) \neq \mathbf{0}_a} A(t')(a) \tag{56}$$

where $\mathcal{D}(a) = (L_a, \bigvee_a, \wedge_a, \mathbf{0}_a, \mathbf{1}_a, \neg_a)$.

Selection: *If $A\colon U$-Tup $\to \mathcal{D}(U)$-Tup and the selection predicate $\mathbf{P}\colon U$-Tup $\to \mathcal{D}(U)$-Tup maps each U-tuple to an element of $\mathcal{D}(U)$-Tup, then $\sigma_{\mathbf{P}} A\colon U$-Tup $\to \mathcal{D}(U)$-Tup is defined by*

$$(\sigma_{\mathbf{P}} A)(t)(a) \stackrel{\text{def}}{=} A(t)(a) \wedge_a \mathbf{P}(t)(a) \tag{57}$$

where $\mathcal{D}(a) = (L_a, \bigvee_a, \wedge_a, \mathbf{0}_a, \mathbf{1}_a, \neg_a)$.

Join: *Let $\mathcal{D}_1 = \{a_1\colon \mathcal{L}_1, ..., a_n\colon \mathcal{L}_n\}$ and $\mathcal{D}_2 = \{b_1\colon \mathcal{L}'_1, ..., b_m\colon \mathcal{L}'_m\}$ be De Morgan frame schemata. Let their union, $\mathcal{D}_1 \cup \mathcal{D}_2$, contain an attribute, $c_i\colon \mathcal{L}_i$, as soon as $c_i\colon \mathcal{L}_i$ is in $\mathcal{D}_1$ or $\mathcal{D}_2$ or both. (If there is an attribute with different corresponding De Morgan frames in $\mathcal{D}_1$ and $\mathcal{D}_2$, a renaming of attributes is needed.) If $A\colon U_1$-Tup $\to \mathcal{D}_1(U_1)$-Tup and $B\colon U_2$-Tup $\to \mathcal{D}_2(U_2)$-Tup, then $A \bowtie B$ is the $(\mathcal{D}_1 \cup \mathcal{D}_2)$-relation over $U_1 \cup U_2$ defined as follows.*

$$(A \bowtie B)(t)(a) \stackrel{\text{def}}{=} \begin{cases} A(t \downarrow U_1)(a) & \text{if } a \in U_1 - U_2 \\ B(t \downarrow U_2)(a) & \text{if } a \in U_2 - U_1 \ . \\ A(t \downarrow U_1)(a) \wedge_a B(t \downarrow U_2)(a) & \end{cases} \tag{58}$$

Difference: *If $A, B\colon U$-Tup $\to \mathcal{D}(U)$-Tup, then $A \setminus B\colon U$-Tup $\to \mathcal{D}(U)$-Tup is defined by*

$$(A \setminus B)(t)(a) \stackrel{\text{def}}{=} A(t)(a) \wedge_a (\neg_a B(t)(a)) \tag{59}$$

where $\mathcal{D}(a) = (L_a, \bigvee_a, \wedge_a, \mathbf{0}_a, \mathbf{1}_a, \neg_a)$.

Renaming: *If $A\colon U$-Tup $\to \mathcal{D}(U)$-Tup and $\beta\colon U \to U'$ is a bijection, then $\rho_\beta A\colon U'$-Tup $\to \mathcal{D}(U')$-Tup is defined by*

$$(\rho_\beta A)(t)(a) \stackrel{\text{def}}{=} A(t)(\beta(a)). \tag{60}$$

As in the case of $\mathcal{L}$-relations it is required that every tuple outside of a similarity database is ranked with the minimal De Morgan frame tuple, $\{a_1\colon \mathbf{0}_1, \ldots, a_n\colon \mathbf{0}_n\}$, and every other tuple is ranked either with the maximal De Morgan frame tuple, $\{a_1\colon \mathbf{1}_1, \ldots, a_n\colon \mathbf{1}_n\}$, or a smaller De Morgan frame tuple expressing a lower degree of containment of the tuple in the database.

Proposition 4.1 (Identities of $\mathcal{D}$-relations Hajdinjak & Bierman (2011)). *The following identities hold for the relational algebra on $\mathcal{D}$-relations:*

- *union is associative, commutative, idempotent, and has identity $\emptyset$;*
- *selection distributes over union and difference;*
- *join is associative and commutative, and distributes over union;*
- *projection distributes over union and join;*
- *selections and projections commute with each other;*
- *difference has identity $\emptyset$ and distributes over union and intersection;*
- *selection with boolean predicates gives all or nothing, $\sigma_{\mathsf{false}}(A) = \emptyset$ and $\sigma_{\mathsf{true}}(A) = A$, where $\mathsf{false(t)(a)} = \mathbf{0}_\mathsf{a}$ and $\mathsf{true(t)(a)} = \mathbf{1}_\mathsf{a}$ for $\mathcal{D}(a) = (L_a, \bigvee_a, \wedge_a, \mathbf{0}_a, \mathbf{1}_a, \neg_a)$;*
- *join with an empty relation gives an empty relation, $A \bowtie \emptyset_U = \emptyset_U$ where A is a $\mathcal{D}$-relation over a schema U;*
- *projection of an empty relation gives an empty relation, $\pi_V(\emptyset) = \emptyset$.*

Each of the similarity measures associated with the attributes maps to its own De Morgan frame. Again, a predefined environment of similarity measures that can be used for building queries is assumed—for every $\mathcal{D}$-relation over U, where $\mathcal{D} = \{a_1 \colon \mathcal{L}_1, ..., a_n \colon \mathcal{L}_n\}$ and $\mathcal{L}_i = (L_i, \bigvee_i, \wedge_i, \mathbf{0}_i, \mathbf{1}_i, \neg_i)$ and $U = \{a_1 \colon \tau_1, ..., a_n \colon \tau_n\}$ there is a similarity measure

$$\rho_{a_i} \colon \tau_i \times \tau_i \to L_i, 1 \leq i \leq n. \tag{61}$$

In the $\mathcal{D}$-relation model, primitive and similarity predicates need to be redefined.

Definition 4.3 (Primitive predicates Hajdinjak & Bierman (2011)). *Suppose in a schema $U = \{a_1 \colon \tau_1, \ldots, a_n \colon \tau_n\}$ the types of attributes a_i and a_j coincide. Then for a given binary predicate θ, the* primitive predicate

$$[a_i\,\theta\,a_j] \colon U\text{-}Tup \to \mathcal{D}(U)\text{-}Tup \tag{62}$$

is defined as follows.

$$[a_i\,\theta\,a_j](t)(a_k) \stackrel{\text{def}}{=} \begin{cases} \chi_{a_i\theta a_j}(t) & \text{if } k = i \text{ or } k = j, \\ \mathbf{1}_k & \text{otherwise.} \end{cases} \tag{63}$$

In words, $[a_i\,\theta\,a_j]$ has value $\mathbf{1}$ in every attribute except a_i and a_j, where it behaves as the characteristic map of θ defined as follows.

$$\chi_{a_i\theta a_j}(t) \stackrel{\text{def}}{=} \begin{cases} \mathbf{1}_k & \text{if } t(a_i)\,\theta\,t(a_j), \\ \mathbf{0}_k & \text{otherwise.} \end{cases} \tag{64}$$

Similarity predicates annotate tuples based on the similarity measures.

Definition 4.4 (Similarity predicates Hajdinjak & Bierman (2011)). *Suppose in a schema $U = \{a_1 \colon \tau_1, \ldots, a_n \colon \tau_n\}$ the types of attributes a_i and a_j coincide. The similarity predicate $[a_i \texttt{ like } a_j] \colon U\text{-}Tup \to \mathcal{D}(U)\text{-}Tup$ is defined as follows.*

$$[a_i \texttt{ like } a_j](t)(a_k) \stackrel{\text{def}}{=} \begin{cases} \rho_{a_i}(t(a_i), t(a_j)) & \text{if } a_k = a_i, \\ \rho_{a_j}(t(a_i), t(a_j)) & \text{if } a_k = a_j, \\ \mathbf{1}_k & \text{otherwise.} \end{cases} \tag{65}$$

In words, $[a_i \texttt{ like } a_j]$ measures similarity of attributes a_i and a_j, each with its own similarity measure. The symmetric version is defined as follows.

$$[a_i \sim a_j] \stackrel{\text{def}}{=} [a_i \texttt{ like } a_j] \cup [a_j \texttt{ like } a_i]. \tag{66}$$

Now union and intersection of selection predicates are computed component-wise.

Given the similarity measures associated with attributes, it is possible to define similarity-based variants of other familiar relational operators, such as similarity-based joins Hajdinjak & Bierman (2011). Such an operator joins two rows not only when their join-attributes have equal associated values, but when the values are similar.

5. A common framework

In this section we explore whether there is a common domain of annotations suitable for all kinds of annotated relations, and we define a general model of $\mathcal{K}$, $\mathcal{L}$-and $\mathcal{D}$-relations.

5.1 A common annotation domain

We have recalled two notions of difference on annotated relations: the monus-based difference proposed by Geerts and Poggi Geerts & Poggi (2010) and the negation-based difference proposed by Hajdinjak and Bierman Hajdinjak & Bierman (2011). We have seen in §3.3 that the monus-based difference does not have the qualities expected in a fuzzy context. The negation-based difference, on the other hand, does agree with the standard fuzzy difference, but it is not defined for bag semantics (and provenance). More precisely, the semiring of counting numbers, $\mathcal{K}_{\mathbb{N}} = (\mathbb{N}, +, \cdot, 0, 1)$, cannot be extended with a negation operation. (The same holds for the provenance semiring.)

We could try to modify the semiring of counting numbers in such a way that negation can be defined. For instance, if we replace $\mathbb{N}$ by $\mathbb{Z}$, we get the ring of integers, $(\mathbb{Z}, +, \cdot, 0, 1)$, where negation can be defined as $\neg x \stackrel{\text{def}}{=} -x$. This implies $(A \setminus B)(t) = -A(t) \cdot B(t)$, which is not equal to the standard difference of relations annotated with the tuples' multiplicities Montagna & Sebastiani (2001). Some other modifications would give the so called *tropical semirings* Aceto et al. (2001) whose underlying carrier set is some subset of the set of real numbers $\mathbb{R}$ equipped with binary operations of minimum or maximum as sum, and addition as product.

Let us now study the properties of the annotation structures of both approaches.

Proposition 5.1 (Identities in an *m*-semiring Bosbach (1965)). *The notion of an m-semiring is characterized by the properties of commutative semirings and the following identities involving $\ominus$.*

$$x \ominus x = \mathbf{0}, \tag{67}$$

$$\mathbf{0} \ominus x = \mathbf{0}, \tag{68}$$

$$x \oplus (y \ominus x) = y \oplus (x \ominus y), \tag{69}$$

$$x \ominus (y \oplus z) = (x \ominus y) \ominus z, \tag{70}$$

$$x \odot (y \ominus z) = (x \odot y) \ominus (x \odot z). \tag{71}$$

Notice that even in a De Morgan frame a difference-like operation may be defined,

$$x \div y \stackrel{\text{def}}{=} x \wedge \neg y. \tag{72}$$

Clearly, negation is then expressed as $\neg x = 1 \div x$.

Proposition 5.2 (Identities in a De Morgan frame)**.** *In a De Morgan frame the following identities involving ÷ hold.*

$$\mathbf{1} \div \mathbf{0} = \mathbf{1}, \tag{73}$$
$$\mathbf{1} \div \mathbf{1} = \mathbf{0}, \tag{74}$$
$$\mathbf{1} \div (x \vee y) = (\mathbf{1} \div x) \wedge (\mathbf{1} \div y), \tag{75}$$
$$\mathbf{1} \div (x \wedge y) = (\mathbf{1} \div x) \vee (\mathbf{1} \div y), \tag{76}$$
$$\mathbf{0} \div x = \mathbf{0}, \tag{77}$$
$$\mathbf{1} \div (\mathbf{1} \div x) = x, \tag{78}$$
$$x \div (\mathbf{1} \div y) = x \wedge y, \tag{79}$$
$$\mathbf{1} \div (\mathbf{1} \div (x \vee y)) = x \vee y, \tag{80}$$
$$\mathbf{1} \div (x \div y) = (\mathbf{1} \div x) \vee y, \tag{81}$$
$$(x \div y) \wedge y = x \wedge (y \div y), \tag{82}$$
$$(x \div y) \vee y = (x \vee y) \wedge (\mathbf{1} \div (y \div y)). \tag{83}$$

Proof. The first four identities are exactly the De Morgan laws from Proposition 3.1. The rest holds by simple expansion of definitions and/or is implied by the De Morgan laws. □

Notice the differences between the properties of the monus-based difference $\ominus$ in an m-semiring and the properties of the negation-based difference ÷ in a De Morgan frame. For instance, in a De Morgan frame we do *not* have $x \div x = \mathbf{0}$ in general.

However, since neither of the proposed notions of difference give the expected result for all kinds of annotated relations, an annotation structure different from m-semirings and De Morgan frames is needed. Observe that by its definition, a complete (even bounded) distributive lattice, $\mathcal{L} = (L, \vee, \wedge, \mathbf{0}, \mathbf{1})$, is a commutative semiring with the natural order $\preceq$ being the lattice order, $a \oplus b = a \vee b$ and $a \odot b = a \wedge b$ for every a, b in L. Because lattice completeness assures the existence of a smallest element in every set and hence the existence of the monus (see Definition 2.3 on GP-conditions), a complete distributive lattice is an m-semiring. On the other hand, if a commutative semiring, $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$, is partially ordered by $\preceq$ and any two elements from K have an infimum and a supremum, it is a lattice, not necessarily bounded Davey & Priestley (1990). The lattice meet and join are then determined by the partial order $\preceq$, and they are, in general, different from $\oplus$ and $\odot$. Since $\mathbf{0} \oplus a = a$, we have $\mathbf{0} \preceq a$ for any $a \in K$, and $\mathbf{0}$ is the least element of the lattice. In general, a similar observation does not hold for $\mathbf{1}$, which is hence not the greatest element of the lattice.

The underlying carrier sets of all the semirings considered are partially ordered sets, even distributive lattices. The unbounded lattices among them (i.e., $\mathcal{K}_{\mathbb{N}}$ and $\mathcal{K}_{\text{prov}}$) can be converted into bounded (even complete) lattices by adding a greatest element. To achieve this we just need to replace $\mathbb{N} \cup \{\infty\}$ for $\mathbb{N}$ and define appropriate calculation rules for ∞.

Lemma 5.1 (Making unbounded partially ordered semirings bounded)**.**

1. *The semiring of counting numbers,* $\mathcal{K}_{\mathbb{N}} = (\mathbb{N}, +, \cdot, 0, 1)$*, partially ordered by*

$$n \preceq m \iff n \leq m, \tag{84}$$

may be extended to the partially ordered commutative semiring $(\mathbb{N} \cup \{\infty\}, +, \cdot, 0, 1)$ *by defining* $\infty + n = \infty$ *and* $\infty \cdot n = \infty$ *except* $\infty \cdot 0 = 0$*. The partial order* $\preceq$ *now determines a complete lattice structure,* $(\mathbb{N} \cup \{\infty\}, \max, \min, 0, \infty)$*.*

2. *The provenance semiring,* $\mathcal{K}_{prov} = (\mathbb{N}[X], +, \cdot, 0, 1)$*, partially ordered by*

$$f[X] \preceq g[X] \iff f_\alpha \leq g_\alpha \text{ for all } \alpha \in I, \tag{85}$$

where $f[X] = \sum_{\alpha \in I} f_\alpha x^\alpha$ *and* $g[X] = \sum_{\alpha \in I} g_\alpha x^\alpha$*, may be extended to the commutative semiring* $((\mathbb{N} \cup \{\infty\})[X], +, \cdot, 0, 1)$ *by defining* $x^\infty \cdot x^n = x^\infty$ *as well as* $\infty + n = \infty$ *and* $\infty \cdot n = \infty$ *except* $\infty \cdot 0 = 0$ *as before. The partial order* $\preceq$ *now determines a complete lattice structure on* $(\mathbb{N} \cup \{\infty\})[X]$ *with*

$$f[X] \wedge g[X] = \sum_{\alpha \in I} \min\{f_\alpha, g_\alpha\} x^\alpha, \tag{86}$$

$$f[X] \vee g[X] = \sum_{\alpha \in I} \max\{f_\alpha, g_\alpha\} x^\alpha. \tag{87}$$

The least element of the lattice is the zero polynomial, 0*, and the greatest element is the polynomial with all coefficients equal to* ∞*.*

To summarize, a complete distributive lattice is an m-semiring. If the lattice even contains negation, we have two difference-like operations; monus $\ominus$ and $\div$, which is induced by negation. There is a class of annotated relations when only one of them ($\ominus$ for bag semantics and provenance, $\div$ for fuzzy semantics) gives the standard notion of relational difference, and there is a class of annotated relations when they both coincide (e.g., classical set semantics, probabilistic relations, and relations on c-tables).

Proposition 5.3 (General annotation structure)**.** *Complete distributive lattices with finite meets distributing over arbitrary joins are suitable codomains for all considered annotated relations.*

Proof. The boolean semiring, the probabilistic semiring, the semiring on c-tables, the similarity semirings as well as the semiring of counting numbers and the provenance semiring (see Lemma 5.1) can all be extended to a complete distributive lattice in which finite meets distribute over arbitrary joins. The later property allows to model infinite relations satisfying all the desired relational identities from Proposition 4.1, including commuting selections and projections. Relational difference may be modeled with the existing monus, $\ominus$, or $\div$ if the lattice is a De Morgan frame where a negation exists. The other (positive) relational operations are modeled using lattice meet, $\wedge$, and join, $\vee$, or semiring sum, $\oplus$, and product, $\odot$. □

5.2 A common model

Recall that Green et al. Green et al. (2007) defined a $\mathcal{K}$-relation over $U = \{a_1 \colon \tau_1, \ldots, a_n \colon \tau_n\}$ as a function $A \colon U\text{-Tup} \to K$ with finite support. The finite-support requirement was made to ensure the existence of the sum in the definition of relational projection. When the commutative semiring $\mathcal{K} = (K, \oplus, \odot, \mathbf{0}, \mathbf{1})$ was replaced by a De Morgan frame, $\mathcal{L} = (L, \bigvee, \wedge, \mathbf{0}, \mathbf{1}, \neg)$, the finite-support requirement became unnecessary; the existence of the join in the definition of projection was quaranteed by the completeness of the codomain.

To model similarity relations more efficiently, Hajdinjak and Bierman Hajdinjak & Bierman (2011) introduced a $\mathcal{D}$-relation over U as a function from U-Tup to $\mathcal{D}(U)$-Tup assigning every element of U-Tup (row of a table) a tuple of different annotation values. We adopt Definition 4.1 to the proposed general annotation structure, and show that a tuple-annotated model may be injectively mapped to an attribute-annotated model.

Definition 5.1 (Annotation schema,annotation tuple,$\mathcal{C}$-relation)**.**

- *An* annotation schema, $\mathcal{C} = \{a_1 \colon \mathcal{L}_1, ..., a_n \colon \mathcal{L}_n\}$, *over* $U = \{a_1 \colon \tau_1, ..., a_n \colon \tau_n\}$ *maps an attribute name,* a_i*, to a complete distributive lattice in which finite meets distribute over arbitrary joins,* $\mathcal{L}_i = (L_{a_i}, \bigvee_{a_i}, \wedge_{a_i}, \mathbf{0}_{a_i}, \mathbf{1}_{a_i})$.
- *An* annotation tuple, $s = \{a_1 \colon l_1, ..., a_n \colon l_n\}$, *maps an attribute name,* a_i*, to an element of a complete distributive lattice in which finite meets distribute over arbitrary joins,* l_i*. The set of all* annotation tuples *matching* $\mathcal{C}$ *over* U *is denoted* $\mathcal{C}(U)$*-Tup.*
- *An* $\mathcal{C}$-relation over U *is a finite map from* U*-Tup to* $\mathcal{C}(U)$*-Tup.*

Proposition 5.4 (Injection of a tuple-annotated model to an attribute-annotated model)**.** *Let* $\mathcal{A}$ *be the class of all functions* $A \colon U$*-Tup* $\to L$ *where* U *is any relational schema and* $\mathcal{L} = (L, \bigvee, \wedge, \mathbf{0}, \mathbf{1})$ *is any complete distributive lattice with finite meets distributing over arbitrary joins. Let* $\mathcal{B}$ *be the class of all* $\mathcal{C}$*-relations over* U*,* $B \colon U$*-Tup* $\to \mathcal{C}(U)$*-Tup, where* $\mathcal{C}$ *is an annotation schema. There is an injective function* $F \colon \mathcal{A} \to \mathcal{B}$ *defined by*

$$F(A)(t)(a_i) \stackrel{\text{def}}{=} A(t) \tag{88}$$

for all attributes a_i *in* U *and tuples* $t \in U$*-Tup.*

Proof. For $A_1, A_2 \in \mathcal{A}$ with $A_1(t) \neq A_2(t)$ we clearly have $F(A_1)(t)(a_i) \neq F(A_2)(t)(a_i)$. □

Proposition 5.4 says that moving from tuple-annotated relations to attribute-annotated relations does not prevent us from correctly modeling the examples covered by the $\mathcal{K}$-relation model in which each tuple is annotated with a single value from $\mathcal{K}$. The annotation value just appears several times. We thus propose a model of $\mathcal{C}$-relations, a common model of $\mathcal{K}$, $\mathcal{L}$-and $\mathcal{D}$-relations, that is attribute annotated. The definitions of union, projection, selection, and join of $\mathcal{C}$-relations may be based on the lattice join and meet operations (like in Definitions 3.9 and 4.2) or, if there exist semiring sum and product operations different from lattice join and meet, the positive relational operations may be defined using these additional semiring operations (like in Definition 2.2). The definition of relational difference may be based on the monus or, when dealing with De morgan frames where a negation exists, the derived $\div$ operation.

Definition 5.2 (Relational algebra on $\mathcal{C}$-relations)**.** *Consider* $\mathcal{C}$*-relations where all the lattices* $\mathcal{L}_i = (L_{a_i}, \bigvee_{a_i}, \wedge_{a_i}, \mathbf{0}_{a_i}, \mathbf{1}_{a_i})$ *from annotation schema* $\mathcal{C} = \{a_1 \colon \mathcal{L}_1, ..., a_n \colon \mathcal{L}_n\}$ *are complete distributive lattices in which finite meets distribute over arbitrary joins. Let* $\bigtriangledown_{a_i}$ *and* $\bigtriangleup_{a_i}$ *stand for either the lattice* $\vee_{a_i}$ *and* $\wedge_{a_i}$ *or some other semiring* $\oplus_{a_i}$ *and* $\odot_{a_i}$ *operations defined on the carrier set* L_{a_i} *of a* $\mathcal{L}_i$*, respectively. Let* $-_{a_i}$ *stand for either the monus* $\ominus_{a_i}$ *or a* $\div_{a_i}$ *operation defined on* L_{a_i}*. The operations of the relational algebra on* $\mathcal{C}$*, denoted* $RA_{\mathcal{C}}$*, are defined as follows.*

Empty relation: *For any set of attributes* U *and corresponding annotation schema,* $\mathcal{C}$*, the empty* $\mathcal{C}$*-relation over* U*,* $\emptyset_U$*, is defined by*

$$\emptyset_U(t)(a) \stackrel{\text{def}}{=} \mathbf{0}_a. \tag{89}$$

Union: *If $A, B\colon U$-Tup $\to \mathcal{C}(U)$-Tup, then $A \cup B\colon U$-Tup $\to \mathcal{C}(U)$-Tup is defined by*

$$(A \cup B)(t)(a) \stackrel{\text{def}}{=} A(t)(a) \triangledown_a B(t)(a). \tag{90}$$

Projection: *If $A\colon U$-Tup $\to \mathcal{C}(U)$-Tup and $V \subset U$, the projection of A on attributes V is defined by*

$$(\pi_V A)(t)(a) \stackrel{\text{def}}{=} \triangledown_{(t' \downarrow V)=t \text{ and } A(t')(a) \neq \mathbf{0}_a} A(t')(a). \tag{91}$$

Selection: *If $A\colon U$-Tup $\to \mathcal{C}(U)$-Tup and the selection predicate $\mathbf{P}\colon U$-Tup $\to \mathcal{C}(U)$-Tup maps each U-tuple to an element of $\mathcal{C}(U)$-Tup, then $\sigma_{\mathbf{P}} A\colon U$-Tup $\to \mathcal{C}(U)$-Tup is defined by*

$$(\sigma_{\mathbf{P}} A)(t)(a) \stackrel{\text{def}}{=} A(t)(a) \vartriangle_a \mathbf{P}(t)(a). \tag{92}$$

Join: *Let $\mathcal{C}_1 = \{a_1\colon \mathcal{L}_1, ..., a_n\colon \mathcal{L}_n\}$ and $\mathcal{C}_2 = \{b_1\colon \mathcal{L}'_1, ..., b_m\colon \mathcal{L}'_m\}$ be annotation schemata. If $A\colon U_1$-Tup $\to \mathcal{C}_1(U_1)$-Tup and $B\colon U_2$-Tup $\to \mathcal{C}_2(U_2)$-Tup, then $A \bowtie B$ is the $(\mathcal{C}_1 \cup \mathcal{C}_2)$-relation over $U_1 \cup U_2$ defined as follows.*

$$(A \bowtie B)(t)(a) \stackrel{\text{def}}{=} \begin{cases} A(t \downarrow U_1)(a) & \text{if } a \in U_1 - U_2 \\ B(t \downarrow U_2)(a) & \text{if } a \in U_2 - U_1 \ . \\ A(t \downarrow U_1)(a) \vartriangle_a B(t \downarrow U_2)(a) & \end{cases} \tag{93}$$

Difference: *If $A, B\colon U$-Tup $\to \mathcal{C}(U)$-Tup, then $A \setminus B\colon U$-Tup $\to \mathcal{C}(U)$-Tup is defined by*

$$(A \setminus B)(t)(a) \stackrel{\text{def}}{=} A(t)(a) -_a B(t)(a). \tag{94}$$

Renaming: *If $A\colon U$-Tup $\to \mathcal{C}(U)$-Tup and $\beta\colon U \to U'$ is a bijection, then $\rho_\beta A\colon U'$-Tup $\to \mathcal{C}(U')$-Tup is defined by*

$$(\rho_\beta A)(t)(a) \stackrel{\text{def}}{=} A(t)(\beta(a)). \tag{95}$$

Relational algebra $\text{RA}_{\mathcal{C}}$ still satisfies *all* the main positive relational algebra identities.

Proposition 5.5 (Identities of $\mathcal{C}$-relations)**.** *The following identities hold for the relational algebra on $\mathcal{C}$-relations:*

- *union is associative, commutative, idempotent, and has identity $\emptyset$;*
- *selection distributes over union and difference;*
- *join is associative and commutative, and distributes over union;*
- *projection distributes over union and join;*
- *selections and projections commute with each other;*
- *difference has identity $\emptyset$ and distributes over union and intersection;*
- *selection with boolean predicates gives all or nothing, $\sigma_{\mathsf{false}}(A) = \emptyset$ and $\sigma_{\mathsf{true}}(A) = A$, where $\mathsf{false}(\mathsf{t})(\mathsf{a}) = \mathbf{0}_\mathsf{a}$ and $\mathsf{true}(\mathsf{t})(\mathsf{a}) = \mathbf{1}_\mathsf{a}$ for $\mathcal{C}(a) = (L_a, \bigvee_a, \wedge_a, \mathbf{0}_a, \mathbf{1}_a)$;*
- *join with an empty relation gives an empty relation, $A \bowtie \emptyset_U = \emptyset_U$ where A is a $\mathcal{C}$-relation over a schema U;*
- *projection of an empty relation gives an empty relation, $\pi_V(\emptyset) = \emptyset$.*

Proof. If the lattice join and meet are chosen to model the positive relational operations, the above identities are implied by Proposition 4.1. On the other hand, if some other semiring sum and product operations are chosen, the identities are implied by Proposition 2.1. □

The properties of relational difference are implied by the identities involving $\ominus$ (see Proposition 5.1) and/or the identities involving $\div$ (see Proposition 5.2), depending on the selection we make.

6. Conclusion

Although the attribute-annotated approach has many advantages, it also has some disadvantages. First, it is clear that asking all attributes to be annotated requires more storage than simple tuple-level annotation. Another problem is that since the proposed general annotation structure, complete distributive lattices with finite meets distributing over arbitrary joins, may not be linearly ordered, an ordering of tuples with falling annotation values is not always possible. Even if each lattice used in an annotation schema is linearly ordered, it is not necessarily the case that there is a linear order on the annotation tuples. Hence, it may not be possible to list query answers (tuples) in a (decreasing) order of relevance. In fact, a suitable ordering of tuples may be established as soon as the lattice of annotation values, $\mathcal{L} = (L, \bigvee, \wedge, \mathbf{0}, \mathbf{1})$, is *graded* Stanley (1997). Recall that a graded or ranked poset is a partially ordered set equipped with a rank function $\rho : L \to \mathbb{Z}$ compatible with the ordering, $\rho(x) < \rho(y)$ whenever $x < y$, and such that whenever y covers x, then $\rho(y) = \rho(x) + 1$. Graded posets can be visualized by means of a Hasse diagram. Examples of graded posets are the natural numbers with the usual order, the Cartesian product of two or more sets of natural numbers with the product order being the sum of the coefficients, and the boolean lattice of finite subsets of a set with the number of elements in the subset. Notice, however, that the ranking problem simply reflects a fact about ordered structures and not a flaw in the model.

The work on attribute-annotated models is very new and has, as far as we know, not been implemented yet Hajdinjak & Bierman (2011). A prototype implementation by means of existing relational database management systems is thus expected to be performed in short term. Another guideline for future research is the study of standard issues from relational databases in the general setting, including data dependencies, redundancy, normalization, and design of databases, optimization issues.

7. References

Aceto, L.; Ésik, Z. & Ingólfsdóttir, A. (2001). *Equational Theories of Tropical Semirings*, BRICS Report Series RS-01-21, University of Aarhus, Denmark.

Amer, K. (1984). Equationally Complete Classes of Commutative Monoids with Monus. *Algebra Universalis*, Vol. 18, No. 1, Jan 1984, -129 – -131, ISSN 0002-5240.

Belohlávek, R. & Vychodil, V. (2006). Relational Model of Data Over Domains with Similarities: An Extension for Similarity Queries and Knowledge Extraction, *Proceedings of the 2006 IEEE International Conference on Information Reuse and Integration*, pp. 207-213, ISBN 0-7803-9788-6, Waikoloa, USA, Sept 2006, IEEE Press, Piscataway, USA.

Bordogna, G. & Psaila, G. (2006). *Flexible Databases Supporting Imprecision and Uncertainty*, Springer-Verlag, Berlin Heidelberg, Germany.

Bosbach, B. (1965). Komplementäre Halbgruppen: Ein Beitrag zur instruktiven Idealtheorie kommutativer Halbgruppen. *Mathematische Annalen*, Vol. 161, No. 4, Dec 1965, -279 – -295, ISSN 0025-5831.

Buneman, P.; Khanna, S. & Tan, W. C. (2001). Why and Where: A Characterization of Data Provenance, *Proceedings of the 8th International Conference on Database Theory*, pp. 316-330, ISBN 3-540-41456-8, London, UK, Jan 2001, Springer-Verlag, London, UK.

Calì, A.; Lembo, D. & Rosati, R. (2003). On the Decidability and Complexity of Query Answering over Inconsistent and Incomplete Databases, *Proceedings of the 22nd Symposium on Principles of Database Systems*, pp. 260–271, ISBN 1-58113-670-6, San Diego, USA, June 2003, ACM Press, New York, USA.

Codd, E. F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*, Vol. 13, No. 6, June 1970, -377 – -387, ISSN 0001-0782.

Cui, Y.; Widom, J. & Wiener, J. L. (2000). Tracing the Lineage of View Data in a Warehousing Environment. *ACM Transactions on Database Systems*, Vol. 25, No. 2, June 2000, -179 – -227, ISSN 0362-5915.

Davey, B. A. & Priestley, H. A. (1990). *Introduction to Lattices and Order*, Cambridge University Press, Cambridge, UK.

Garcia-Molina, H.; Ullman, J. & Widom, J. (2008). *Database Systems: The Complete Book*, 2nd edition, Prentice Hall, New York, USA.

Geerts, F. & Poggi, A. (2010). On Database Query Languages for *K*-Relations. *Journal of Applied Logic*, Vol. 8, No. 2, June 2010, -173 – -185, ISSN 1570-8683.

Green, T. J.; Karvounarakis, G. & Tannen, V. (2007). Provenance Semirings, *Proceedings of the 26th Symposium on Principles of Database Systems*, pp. 31-40, ISBN 978-1-59593-685-1, Beijing, China, June 2007, ACM Press, New York, USA.

Hajdinjak, M. & Mihelič, F. (2006). The PARADISE Evaluation Framework: Issues and Findings. *Computational Linguistics*, Vol. 32, No. 2, June 2006, -263 – -272, ISSN 0891-2017.

Hajdinjak, M. & Bauer, A. (2009). Similarity Measures for Relational Databases. *Informatica*, Vol. 33, No. 2, May 2009, -135 – -141, ISSN 0350-5596.

Hajdinjak, M. & Bierman, G. M. (2011). Extending Relational Algebra with Similarities. To appear in *Mathematical Structures in Computer Science*, ISSN 0960-1295.

Hjaltason, G. R. & Samet, H. (2003). Index-Driven Similarity Search in Metric Spaces. *ACM Transactions on Database Systems*, Vol. 28, No. 4, Dec 2003, -517 – -580, ISSN 0362-5915.

Hutton, B. (1975). Normality in Fuzzy Topological Spaces. *Journal of Mathematical Analysis and Applications*, Vol. 50, No. 1, April 1975, -74 – -79, ISSN 0022-247X.

Imielinski, T. & Lipski, W. (1984). Incomplete Information in Relational Databases. *Journal of the ACM*, Vol. 31, No. 4, Oct 1984, -761 – -791, ISSN 0004-5411.

Jae, Y. L. & Elmasri, R. A. (2001). A temporal Algebra for an ER-Based Temporal Data Model, *Proceedings of the the 17th International Conference on Data Engineering*, pp. 33-40, ISBN 0-7695-1001-9, Heidelberg, Germany, April 2001, IEEE Computer Society, Washington, USA.

Ma, Z. (2006). *Studies in Fuzziness and Soft Computing: Fuzzy Database Modeling of Imprecise and Uncertain Engineering Information*, Vol. 195, Springer-Verlag, Berlin Heidelberg, Germany.

Ma, Z. & Yan, L. (2008). A Literature Overview of Fuzzy Database Models. *Journal of Information Science and Engineering*, Vol. 24, No. 1, Jan 2008, -189 – -202, ISSN 1016-2364.

Minker, J. (1998). An Overview of Cooperative Answering in Databases, *Proceedings of the 3rd International Conference on Flexible Query Answering Systems*, pp. 282-285, ISBN 3-540-65082-2, Roskilde, Denmark, May 1998, Springer-Verlag, Berlin Heidelberg.

Montagna, F. & Sebastiani, V. (2001). Equational Fragments of Systems for Arithmetic. *Algebra Universalis*, Vol. 46, No. 3, Sept 2001, -417 – -441, ISSN 0002-5240.

Patella, M. & and Ciaccia, P. (2009). Approximate Similarity Search: A Multi-faceted Problem. *Journal of Discrete Algorithms*, Vol. 7, No. 1, March -2009, -36 – -48, ISSN 1468-0904.

Peeva, K. & Kyosev, Y. (2004). *Fuzzy Relational Calculus: Theory, Applications And Software (Advances in Fuzzy Systems)*, World Scientific Publishing Company, ISBN 9-812-56076-9.

Penzo, W. (2005). Rewriting Rules to Permeate Complex Similarity and Fuzzy Queries within a Relational Database System. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 2, Feb 2005, -255 – -270, ISSN 1041-4347.

Rosado, A.; Ribeiro, R. A.; Zadrozny, S. & Kacprzyk, J. (2006). Flexible Query Languages for Relational Databases: An Overview, In: *Flexible Databases Supporting Imprecision and Uncertainty*, Bordogna, G. & Psaila, G. (Eds.), pp. 3-53, Springer-Verlag, ISBN 978-3-540-33288-6, Berlin Heidelberg, Germany.

Salii, V. N. (1983). Quasi-Boolean Lattices and Associations. In: *Lectures in Universal Algebra: Proceedings of Colloquia Mathematica Societatis János Bolyai*, Vol. 43, Szabo, L. & Szendrei, A. (Eds.), pp. 429-454, North-Holland, ISBN xxx, Amsterdam, The Nerherlands.

Schmitt, I. & Schulz, N. (2004). Similarity Relational Calculus and its Reduction to a Similarity Algebra, In: *Lecture Notes in Computer Science: Foundations of Information and Knowledge Systems - FoIKS 2004*, Vol. 2942, Seipel, D. & Turull Torres, J. M. (Eds.), pp. 252-272, Springer-Verlag, ISBN 3-540-20965-4, Berlin Heidelberg, Germany.

Shenoi, S. & Melton, A. (1989). Proximity Relations in the Fuzzy and Relational Database Model. *Fuzzy Sets and Systems*, Vol. 31, No. 3, July 1989, -285 – -296, ISSN 0165-0114.

Stanley, R. (1997). *Cambridge Studies in Advanced Mathematics 49: Enumerative Combinatorics*, Vol. 1., Cambridge University Press, Cambridge, England.

Suciu, D. (2008). Probabilistic Databases. *SIGACT News*, Vol. 39, No. 2, June 2008, -111 – -124, ISSN 0163-5700.

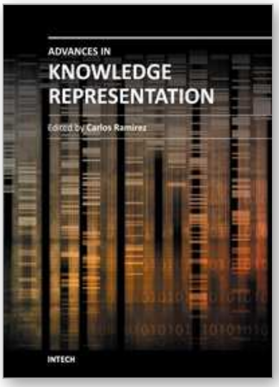
Ullman, J. D. (1988). *Principles of Database and Knowledge-Base Systems*, Vol. I, Computer Science Press, Inc., Rockville, USA.

Ullman, J. D. (1989). *Principles of Database and Knowledge-Base Systems: The New Technologies*, Vol. II, Computer Science Press, Inc., Rockville, USA.

Van der Meyden, R. (1998). Logical Approaches to Incomplete Information: A Survey, In: *Logics for Databases and Information Systems*, Chomicki, J. & Saake, G. (Eds.), page numbers (307-356), Kluwer Academic Publishers, ISBN 0-7923-8129-7, Norwell, USA.

Yazici, A. & George, R. (1999). *Fuzzy Database Modeling*, Physica-Verlag, Heidelberg, Germany.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, Vol. 8, No. 3, June 1965, -338 – -353, ISSN 0019-9958.



Advances in Knowledge Representation

Edited by Dr. Carlos Ramirez

ISBN 978-953-51-0597-8
Hard cover, 272 pages
Publisher InTech
Published online 09, May, 2012
Published in print edition May, 2012

Advances in Knowledge Representation offers a compilation of state of the art research works on topics such as concept theory, positive relational algebra and k-relations, structured, visual and ontological models of knowledge representation, as well as detailed descriptions of applications to various domains, such as semantic representation and extraction, intelligent information retrieval, program proof checking, complex planning, and data preparation for knowledge modelling, and a extensive bibliography. It is a valuable contribution to the advancement of the field. The expected readers are advanced students and researchers on the knowledge representation field and related areas; it may also help to computer oriented practitioners of diverse fields looking for ideas on how to develop a knowledge-based application.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821