The Implication Problem of Functional Dependencies in Complex-Value Databases

Sven Hartmann and Sebastian Link

Information Science Research Centre,
Massey University, Palmerston North,
New Zealand

1. **A brief Introduction to Databases**
2. **A Complex-value Data Model**
3. **Functional Dependencies**
4. **The Implication Problem**
5. **Future Work**

1.1 Databases: Syntax and Semantics

- *syntax*: database schema uses attributes (names) to define properties that all objects of target database are desribed by
  - purely relational FO-signature: Talk={Conference, Title, Author}
- *semantics*: database instance is any collection of objects each of which is given by values from domains of all attributes
  - (finite) structure by which logic of signature is interpreted
  -

Conference	Title	Author
WoLLIC	FDs in Complex-value DBs	S. Hartmann
WoLLIC	FDs in Complex-value DBs	S. Link

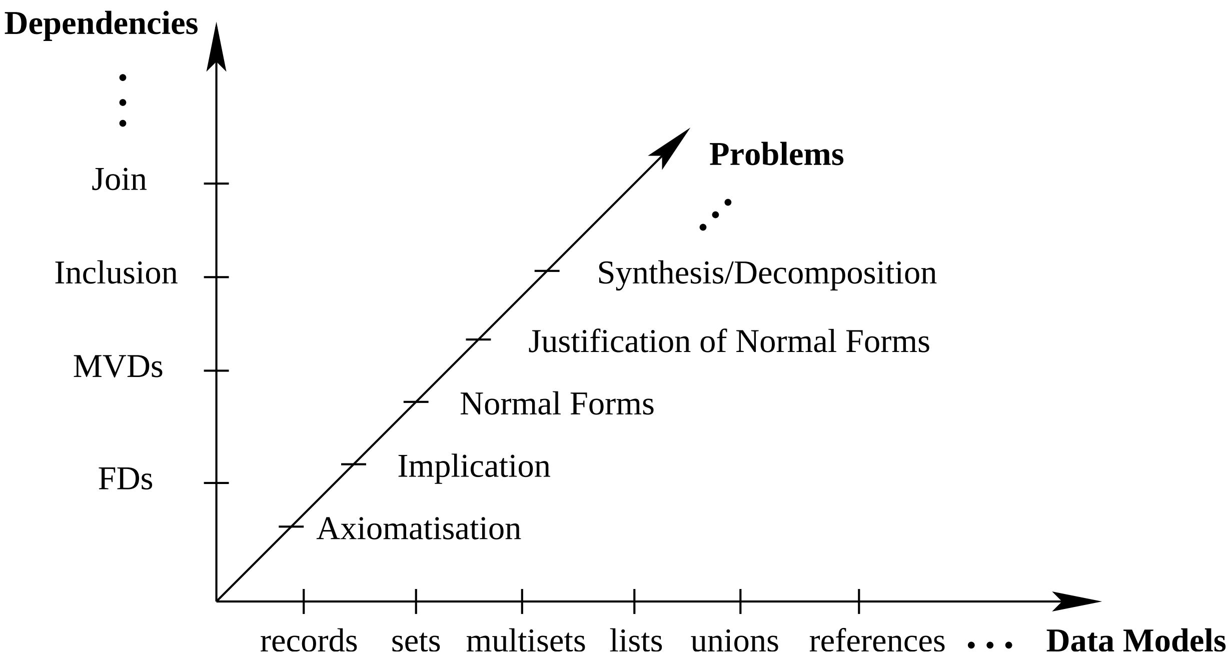
- *constraints*: conditions on the schema to restrict the database instances to those considered meaningful for the application in mind
  - certain classes of first-order formulae (at least 90 different)
  - Title, Author $\rightarrow$ Conference, Title $\rightarrow$ Author

1.2 Databases: The Implication Problems and its Implications

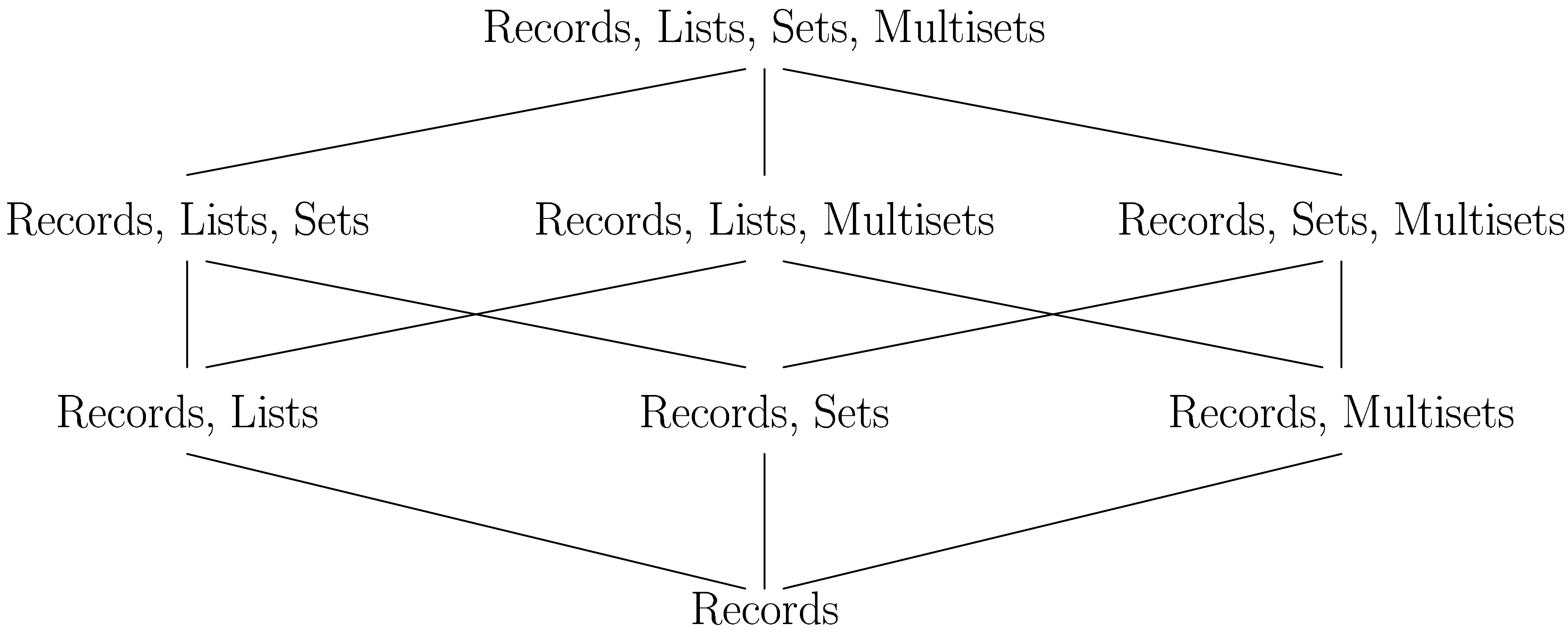
- given: schema S, set Σ of constraints in class $\mathcal{C}$, constraint σ in $\mathcal{C}$
- decide whether $\Sigma \models_{(\text{fin})} \sigma$ holds
- $\Sigma^*_{\mathcal{C}} = \{\sigma \in \mathcal{C} \mid \Sigma \models \sigma\}$ and $\Sigma^+_{\mathfrak{R}} = \{\sigma \mid \Sigma \vdash_{\mathfrak{R}} \sigma\}$
- $\mathfrak{R}$ sound ($\Sigma^+_{\mathfrak{R}} \subseteq \Sigma^*_{\mathcal{C}}$) and complete ($\Sigma^*_{\mathcal{C}} \subseteq \Sigma^+_{\mathfrak{R}}$) for implication of $\mathcal{C}$
- $\Sigma \models \sigma$ iff $\sigma \in \Sigma^*_{\mathcal{C}}$ iff $\sigma \in \Sigma^+_{\mathfrak{R}}$
- efficient solution to implication problem useful for
  - efficiently deciding equivalence of sets of constraints
  - efficiently computing minimal covers for sets of constraints
  - automatisation of database design tools

1.3 Databases: Data Models vs. Data Types

- RDM: single application of records
- Semantic Data Models: records (aggregation) and sets (grouping)
- HERM: records, sets and disjoint unions (cluster)
- OODBM: records, lists, sets (minimal) ...
- XML: records, lists, unions, multisets ...



1.4 The Framework of this Presentation



	Set	Multisets	Lists
Order	0	0	1
Multiplicity	0	1	1

2.1 Syntax: Nested Attributes

- capture characteristics of objects in target database by attributes
- finite set $\mathcal{U}$ of flat attributes and $dom(A)$ for all $A \in \mathcal{U}$
- use set $\mathcal{L}$ of labels with $\mathcal{U} \cap \mathcal{L} = \emptyset$ and $\lambda \notin \mathcal{U} \cup \mathcal{L}$
- **nested attributes** $\mathcal{N}\mathrm{A}(\mathcal{U}, \mathcal{L})$:
  - *flat attributes* $\mathcal{U} \subseteq \mathcal{N}\mathrm{A}$,
  - *null attribute* $\lambda \in \mathcal{N}\mathrm{A}$,
  - *record-valued attributes* $L(N_1, \dots, N_k) \in \mathcal{N}\mathrm{A}$, if $L \in \mathcal{L}$ and $N_1, \dots, N_k \in \mathcal{N}\mathrm{A}$ with $k \geq 1$
  - *set-valued attributes* $L\{N\} \in \mathcal{N}\mathrm{A}$, if $L \in \mathcal{L}$ and $N \in \mathcal{N}\mathrm{A}$
  - *list-valued attributes* $L[N] \in \mathcal{N}\mathrm{A}$, if $L \in \mathcal{L}$ and $N \in \mathcal{N}\mathrm{A}$
  - *multiset-valued attributes* $L\langle N\rangle \in \mathcal{N}\mathrm{A}$, if $L \in \mathcal{L}$ and $N \in \mathcal{N}\mathrm{A}$

2.2 Semantics: Domain Assignment

- extend mapping *dom* from flat attributes to nested attributes by:
  - $dom(\lambda) = \{ok\}$,
  - $dom(L(N_1, \dots, N_k)) = \{(v_1, \dots, v_k) \mid v_i \in dom(N_i)\}$,
  - $dom(L\{N\}) = \{\{v_1, \dots, v_n\} \mid v_i \in dom(N)\}$,
  - $dom(L\langle N\rangle) = \{\langle v_1, \dots, v_n\rangle \mid v_i \in dom(N)\}$,
  - $dom(L[N]) = \{[v_1, \dots, v_n] \mid v_i \in dom(N)\}$
- empty set, empty multiset, and empty list are $\emptyset, \langle\ \rangle, [\ ]$

2.3 Subattributes

- define $\leq \subseteq \mathcal{N}\mathrm{A} \times \mathcal{N}\mathrm{A}$ by:
  - $N \leq N$ for all nested attributes $N \in \mathcal{N}\mathrm{A}$,
  - $\lambda \leq A$ for all flat attributes $A \in \mathcal{U}$,
  - $\lambda \leq N$ for all set-, multiset- and list-valued attributes $N \in \mathcal{N}\mathrm{A}$,
  - $L(N_1, \ldots, N_k) \leq L(M_1, \ldots, M_k)$, if $N_i \leq M_i$ for all $i = 1, \ldots, k$,
  - $L\{N\} \leq L\{M\}$, if $N \leq M$,
  - $L\langle N \rangle \leq L\langle M \rangle$, if $N \leq M$,
  - $L[N] \leq L[M]$, if $N \leq M$
- subattribute relation $\leq$ on nested attributes is reflexive, anti-symmetric and transitive

2.4 Semantics on Subattributes: Projection Function

- for $M \leq N$ define $\pi^N_M : Dom(N) \rightarrow Dom(M)$ by:
  - $\pi^N_N : v \mapsto v$,
  - $\pi^N_\lambda : v \mapsto ok$,
  - $\pi^{L(N_1,\dots,N_k)}_{L(M_1,\dots,M_k)} : (v_1, \dots, v_k) \mapsto (\pi^{N_1}_{M_1}(v_1), \dots, \pi^{N_k}_{M_k}(v_k))$,
  - $\pi^{L\{N'\}}_{L\{M'\}} : S \mapsto \{\pi^{N'}_{M'}(s) \ : \ s \in S\}$,
  - $\pi^{L\langle N'\rangle}_{L\langle M'\rangle} : S \mapsto \langle\pi^{N'}_{M'}(s) \ : \ s \in S\rangle$,
  - $\pi^{L[N']}_{L[M']} : [v_1, \dots, v_n] \mapsto [\pi^{N'}_{M'}(v_1), \dots, \pi^{N'}_{M'}(v_n)]$
- $\emptyset, \langle\ \rangle, [\ ]$ mapped to themselves, except when projected on λ

2.5 Multiplicity

- consider nested attribute *Shopping(Person, Buys[Article])*
- with (Toni, [Shoes, Top, Shoes, Pants]) and (Seb, [ ])
- on *Shopping(Person, Buys[*λ*])*: (Toni, [ok, ok, ok, ok]) and (Seb, [ ])
- *Shopping(Person, Buys[*λ*])* still represents number of articles bought
- consider now *Shopping(Person, Buys*$\{$*Article*$\}$*)*
- with (Toni, $\{$Shoes, Top, Pants$\}$) and (Seb, $\emptyset$)
- projection on *Shopping(Person,Buys*$\{\lambda\}$*)*: (Toni, $\{$ok$\}$) and (Seb, $\emptyset$)
- *Shopping(Person, Buys*$\{\lambda\}$*)* still reveals that Toni bought something and Seb leaves empty-handed

2.6 Operations on Subattributes

- $Sub(N) = \{X \in \mathcal{N}A \mid X \leq N\}$: λ_N, $Y \sqcup_N Z$, $Y \sqcap_N Z$, $Y \dot{-}_N Z$:
  - $\lambda_N = L(\lambda_{N_1}, \dots, \lambda_{N_k})$, if $N = L(N_1, \dots, N_k)$, and $\lambda_N = \lambda$ else,
  - $X \leq Y$: $X \sqcup_N Y = Y$, $X \sqcap_N Y = X$, and $X \dot{-}_N Y = \lambda_N$,
  - $X \dot{-}_N \lambda_N = X$,
  - $N = L(N_1, \dots, N_k), X = L(X_1, \dots, X_k)$ and $Y = L(Y_1, \dots, Y_k)$:

$$X \circ_N Y = L(X_1 \circ_{N_1} Y_1, \dots, X_k \circ_{N_k} Y_k) \text{ for } \circ \in \{\sqcup, \sqcap, \dot{-}\}$$

  - $N = L\{M\}$, $X = L\{X'\}$, and $Y = L\{Y'\}$:

$$X \circ_N Y = L\{X' \circ_M Y'\} \text{ for } \circ \in \{\sqcup, \sqcap, \dot{-}\}$$

  - $N = L\langle M \rangle$, $X = L\langle X' \rangle$, and $Y = L\langle Y' \rangle$:

$$X \circ_N Y = L\langle X' \circ_M Y' \rangle \text{ for } \circ \in \{\sqcup, \sqcap, \dot{-}\}$$

  - $N = L[M]$, $X = L[X']$, and $Y = L[Y']$:

$$X \circ_N Y = L[X' \circ_M Y'] \text{ for } \circ \in \{\sqcup, \sqcap, \dot{-}\}$$

2.7 The Brouwerian Algebra of Subattributes

- $(Sub(N), \leq, \sqcup_N, \sqcap_N, \dot{-}_N, N)$ is a **Brouwerian Algebra**
  - $(Sub(N) = \{X \in \mathcal{N}A \mid X \leq N\}, \leq, \sqcup_N, \sqcap_N)$ is a lattice
  - N is top element
  - pseudo difference $Z \dot{-} Y$ of Z and Y in $Sub(N)$ satisfies

$$Z \dot{-} Y \leq X \quad \text{if and only if} \quad Z \leq Y \sqcup X$$

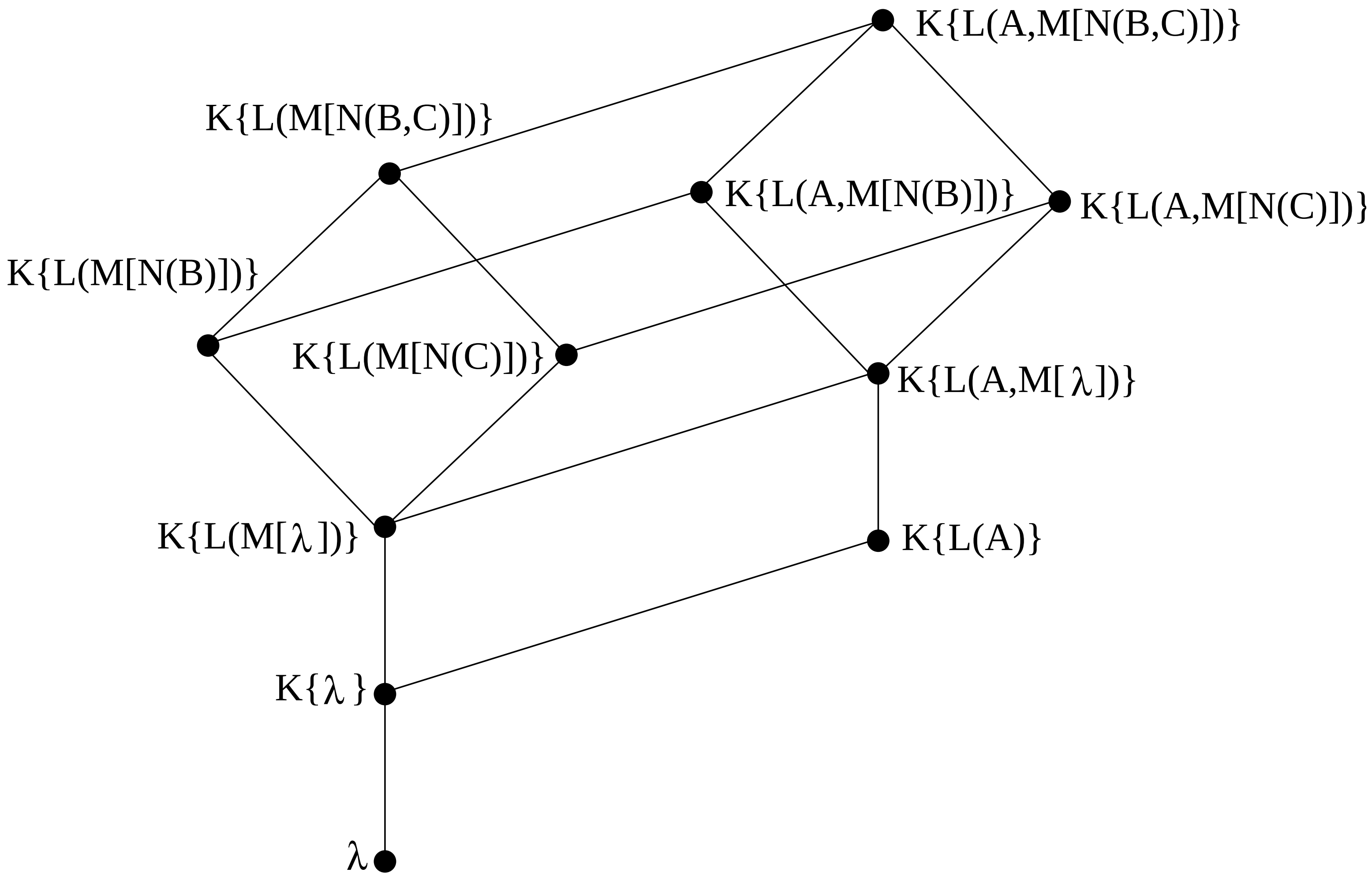
  for all $X \in Sub(N)$

- **Brouwerian Complement**: $Y_N^{\mathcal{C}} = N \dot{-}_N Y$ satisfies

$$Y^{\mathcal{C}} \leq X \quad \text{if and only if} \quad X \sqcup Y = N$$

- bottom element $\lambda_N = N \dot{-} N$
- $(Sub(N), \leq, \sqcup_N, \sqcap_N, (\cdot)_N^{\mathcal{C}}, \lambda_N, N)$ is not a **Boolean Algebra**
- every Brouwerian Algebra is distributive

2.8 The Algebra of Nested Attributes: An Example



3.1 Functional Dependencies

- **functional dependency** on nested attribute N is

$$\mathcal{X} \rightarrow \mathcal{Y} \qquad \text{with} \quad \mathcal{X}, \mathcal{Y} \subseteq Sub(N) \text{ non-empty}$$

- $r \subseteq Dom(N)$ **satisfies** $\mathcal{X} \rightarrow \mathcal{Y}$ on N ($\models_r \mathcal{X} \rightarrow \mathcal{Y}$) iff $\forall t_1, t_2 \in r$:

$$\pi_X^N(t_1) = \pi_X^N(t_2) \; \forall \, X \in \mathcal{X} \quad \text{implies} \quad \pi_Y^N(t_1) = \pi_Y^N(t_2) \; \forall \, Y \in \mathcal{Y}$$

- $N =$ Soccer{Match(Winner, Loser)} and $r = \{t_1, t_2\} \subseteq Dom(N)$:

t_1={(Greece, England), (France, Portugal)} and
t_2={(Greece, Portugal), (France, England)}

- $\models_r$ Soccer{Match(Winner)} $\rightarrow$ Soccer{Match(Loser)}
- $\not\models_r$ Soccer{Match(Winner)} $\rightarrow$ Soccer{Match(Winner, Loser)}
- values on subattributes X and Y do not determine values on $X \sqcup Y$
- shows: FDs cannot be simplified to $X \rightarrow Y$ with $X, Y \in Sub(N)$

3.2 Reconcilable Attributes

- $X, Y \in Sub(N)$ **reconcilable** iff one of the following holds:
  - $Y \leq X$ or $X \leq Y$,
  - $N = L(N_1, \dots, N_k), X = L(X_1, \dots, X_k), Y = L(Y_1, \dots, Y_k)$ where X_i and Y_i are reconcilable for all $i = 1, \dots, k$,
  - $N = L[N'], X = L[X'], Y = L[Y']$ where X' and Y' reconcilable
- Soccer{Match(Winner,λ)}, Soccer{Match(λ,Loser)} not reconcilable

3.3 Axiomatisation

- in presence of records, or records and lists:

$$\frac{}{X \to Y}\, Y \leq X, \quad \frac{X \to Y}{X \to X \sqcup_N Y}, \quad \frac{X \to Y,\ Y \to Z}{X \to Z}$$

form minimal axiomatisation

- $\mathcal{T}$ any non-empty subset of {lists, sets, multisets} apart from {lists}:

$$\frac{}{\mathcal{X} \to \mathcal{Y}}\, \mathcal{Y} \subseteq \mathcal{X}, \quad \frac{}{\{X\} \to \{Y\}}\, Y \leq X, \quad \frac{\mathcal{X} \to \mathcal{Y}}{\mathcal{X} \to \mathcal{X} \cup \mathcal{Y}},$$

$$\frac{}{\{X, Y\} \to \{X \sqcup_N Y\}}\, X, Y \text{ reconcilable}, \quad \frac{\mathcal{X} \to \mathcal{Y},\ \mathcal{Y} \to \mathcal{Z}}{\mathcal{X} \to \mathcal{Z}},$$

form minimal axiomatisation for records and $\mathcal{T}$

4.1 Closure and Units

- $\mathcal{X}^+ = \{Z \mid \mathcal{X} \to \{Z\} \in \Sigma^+\}$ for $\emptyset \neq \mathcal{X} \subseteq Sub(N)$
- $\mathcal{Y} \subseteq_{\text{gen}} \mathcal{X}$ iff $\forall Y \in \mathcal{Y}.\exists X \in \mathcal{X}.Y \leq X$
- $\mathcal{X} \to \mathcal{Y} \in \Sigma^+$ iff $\mathcal{Y} \subseteq \mathcal{X}^+$ iff $\mathcal{Y} \subseteq_{\text{gen}} \mathcal{X}^+_{\text{max}}$
- $N_i \in Sub(N)$ is **unit** of N ($N_i \in \mathcal{U}(N)$) iff N_i is $\leq$-maximal with property that all reconcilable $X, Y \leq N_i$ satisfy $X \leq Y$ or $Y \leq X$
- $N = L_1(L_2\langle L_3(A, B)\rangle, L_4[L_5(C, L_6\langle D\rangle)], L_7(E, L_8\{L_9(F, G, H)\}))$:
  - $N_1 = L_1(L_2\langle L_3(A, B)\rangle, \lambda, L_7(\lambda, \lambda))$,
  - $N_2 = L_1(\lambda, L_4[L_5(C, \lambda)], L_7(\lambda, \lambda))$,
  - $N_3 = L_1(\lambda, L_4[L_5(\lambda, L_6\langle D\rangle))], L_7(\lambda, \lambda))$,
  - $N_4 = L_1(\lambda, \lambda, L_7(E, \lambda))$,
  - $N_5 = L_1(\lambda, \lambda, L_7(\lambda, L_8\{L_9(F, G, H)\}))$
- N is the join of its mutually reconcilable units
- $\mathcal{X}_i = \{X \sqcap N_i \; : \; X \in \mathcal{X}\}$ for $N_i \in \mathcal{U}(N)$

4.2 Computing the Nested Attribute Closure

INPUT: $N \in NA$, $\mathcal{X} \subseteq Sub(N)$, set Σ of FDs on N; **OUTPUT**: $\mathcal{X}^{alg}_{max}$

VAR: $\mathcal{X}^{new}_i, \mathcal{X}^{old}_i, \mathcal{U}_i, \mathcal{V}_i \subseteq Sub(N)$, $N_1, \dots, N_k \in Sub(N)$;

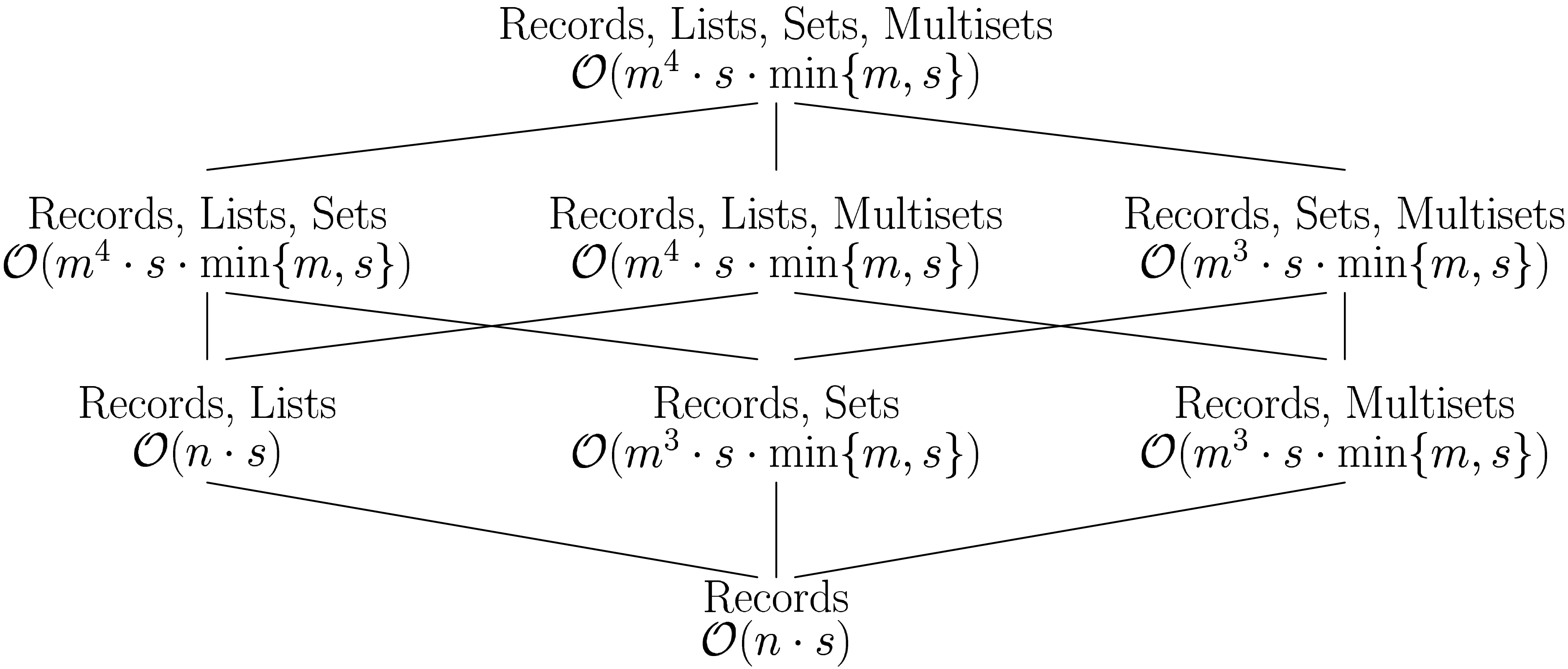
Compute $\mathcal{U}(N) = \{N_1, \dots, N_k\}$;
FOR $i = 1$ **TO** k **DO** $\mathcal{X}^{new}_i := \max(\{X \sqcap N_i \; : \; X \in \mathcal{X}\})$;
REPEAT
  FOR $i = 1$ **TO** k **DO** $\mathcal{X}^{old}_i := \mathcal{X}^{new}_i$;
  FOR each $\mathcal{U} \rightarrow \mathcal{V} \in \Sigma$ **DO**
    IF $\mathcal{U}_i \subseteq_{gen} \mathcal{X}^{new}_i$ for $i = 1, \dots, k$ **THEN**
      FOR $i = 1$ **TO** k **DO** $\mathcal{X}^{new}_i := \max(\mathcal{X}^{new}_i \cup \mathcal{V}_i)$;
    ENDIF;
  ENDDO;
UNTIL $\mathcal{X}^{new}_i = \mathcal{X}^{old}_i$ for $i = 1, \dots, k$;
$\mathcal{X}^{alg}_{max} := \{X_1 \sqcup \dots \sqcup X_k \; : \; X_i \in \mathcal{X}^{new}_i\}$;
RETURN$(\mathcal{X}^{alg}_{max})$;

4.3 Correctness

- $\mathcal{X}^{\text{alg}}_{\max} = \mathcal{X}^{+}_{\max}$
- both anti-chains
- $\mathcal{X}^{\text{alg}}_{\max} \subseteq_{\text{gen}} \mathcal{X}^{+}_{\max}$:
  - soudness of inference rules
- $\mathcal{X}^{+}_{\max} \subseteq_{\text{gen}} \mathcal{X}^{\text{alg}}_{\max}$:
  - $\Sigma = \Sigma_0 \subset \Sigma_1 \subset \cdots \subset \Sigma_s = \Sigma^+$
  - induction: $\mathcal{Y} \to \mathcal{Z} \in \Sigma_j$ with $\mathcal{Y} \subseteq_{\text{gen}} \mathcal{X}^{\text{alg}}_{\max}$ implies $\mathcal{Z} \subseteq_{\text{gen}} \mathcal{X}^{\text{alg}}_{\max}$
  - particularly: $\mathcal{Y} \to \mathcal{Z} \in \Sigma^+$ with $\mathcal{Y} \subseteq_{\text{gen}} \mathcal{X}^{\text{alg}}_{\max}$ implies $\mathcal{Z} \subseteq_{\text{gen}} \mathcal{X}^{\text{alg}}_{\max}$
  - $\mathcal{Y} = \mathcal{X}$ and $\mathcal{Z} = \mathcal{X}^{+}_{\max}$ shows $\mathcal{X}^{+}_{\max} \subseteq_{\text{gen}} \mathcal{X}^{\text{alg}}_{\max}$

4.4 Complexity - Upper Bounds

- $m = \sharp Sub(N)$, $s = \sharp\Sigma$, $n = \sharp\{X \in Sub(N) \mid X \text{ join-irreducible}\}$



5 Future Work

- study relationship between implication problems of constraints in complex-value databases and implication problems in certain logics
- Normal Forms:
  - syntactic conditions on nested attributes depending on given set of constraints
  - describe semantically well-designed database schemata (absence of redundancy etc.)
- more types: (disjoint) union, references
- other constraints: inclusion, join dependencies
- graph-theoretical approach in terms of homomorphisms between subtrees