Functional Dependencies in Complex-Value Databases

Sebastian Link

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

Outline

1. **Motivation & Revision of the RDM**
2. **The Algebra of Nested Attributes**
3. **Axiomatisations**
4. **Minimality**
5. **Comparison**
6. **Extensions and Future Work**

1.1 Functional Dependencies in the RDM

- FDs introduced in context of RDM by E.F. Codd in 1972
  - expression $X \to Y$ with $X, Y \subseteq R$
  - $\models_r X \to Y$ iff $t_1[Y] = t_2[Y]$, if $t_1[X] = t_2[X]$ for any $t_1, t_2 \in r$
- gain complete knowledge about consequences of semantics specified
  - **Boolean Algebra** $(\mathcal{P}(R), \subseteq, \cup, \cap, -, \emptyset, R)$ on R
  - **Armstrong Axioms**

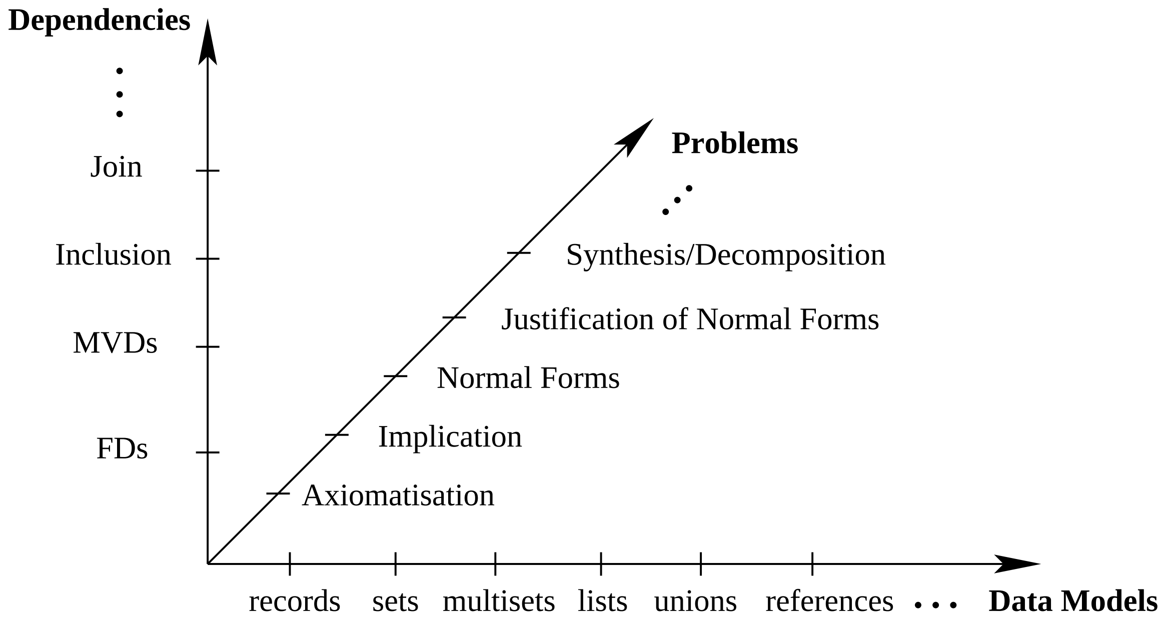
$$\frac{}{X \to Y} Y \subseteq X \qquad \frac{X \to Y}{X \to X \cup Y} \qquad \frac{X \to Y, Y \to Z}{X \to Z}$$

  form minimal, sound and complete set of Inference Rules

- important results established using axiomatisation:
  - **(Finite) Implication Problem** decidable in linear time
  - efficiently deciding equiv of sets of FDs, computing minimal covers
  - **Boyce-Codd Normal Form** and **Third Normal Form**
  - no redundancies and update anomalies, simple integrity checking

1.2 Challenges with Advanced Data Models

- apart from RDM **no universal agreement** on data models exists
- ER, UML, HERM, Nested RDM, Object-oriented/relational, XML
- Biskup (1995,1998):
  - find unifying framework
  - extend achievements to deal with complex object types
- classify data models by the **types** they support



[ht]

1.3 Complex Objects

- here: **values, records, lists, sets and multisets of values**
- ordered relations, time-series data, meteorological and astronomical data streams, runs of experimental data, multidimensional arrays, textual information, voices, sound, images, video, etc.
- subject to studies in deductive and temporal database community
- occur naturally in object-oriented databases (XML)
- bioinformatics: lists and sets occur naturally in genomic sequence databases
- fundamental data structures for number of computational frameworks: Gamma coordination language, Chemical Abstract Machine and P systems modeling membrane computing

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 $\emptyset, \langle\ \rangle, [\ ]$
- RDM: record-valued attributes only
- Nested Relational Data Model: record- and set-valued attributes only

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 $N \in \mathcal{N}\mathrm{A}$,
  - $L(N_1, \dots, N_k) \leq L(M_1, \dots, M_k)$, if $N_i \leq M_i$ for all $i = 1, \dots, k$,
  - $L\{N\} \leq L\{M\}$, 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 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 $Y = L(Y_1, \dots, Y_k)$:

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

  - $N = L\{M\}$, $X = L\{X'\}, Y = L\{Y'\}, \circ \in \{\sqcup, \sqcap\}$:

$$X \circ_N Y = L\{X' \circ_M Y'\}, \quad X \not\leq Y : X \dot{-}_N Y = L\{X' \dot{-}_M Y'\},$$

  - $N = L\langle M\rangle$, $X = L\langle X'\rangle, Y = L\langle Y'\rangle, \circ \in \{\sqcup, \sqcap\}$:

$$X \circ_N Y = L\langle X' \circ_M Y'\rangle, \quad X \not\leq Y : X \dot{-}_N Y = L\langle X' \dot{-}_M Y'\rangle,$$

  - $N = L[M]$, $X = L[X'], Y = L[Y'], \circ \in \{\sqcup, \sqcap\}$:

$$X \circ_N Y = L[X' \circ_M Y'], \quad X \not\leq Y : X \dot{-}_N Y = L[X' \dot{-}_M Y']$$

2.6 The Brouwerian Algebra of Subattributes

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

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

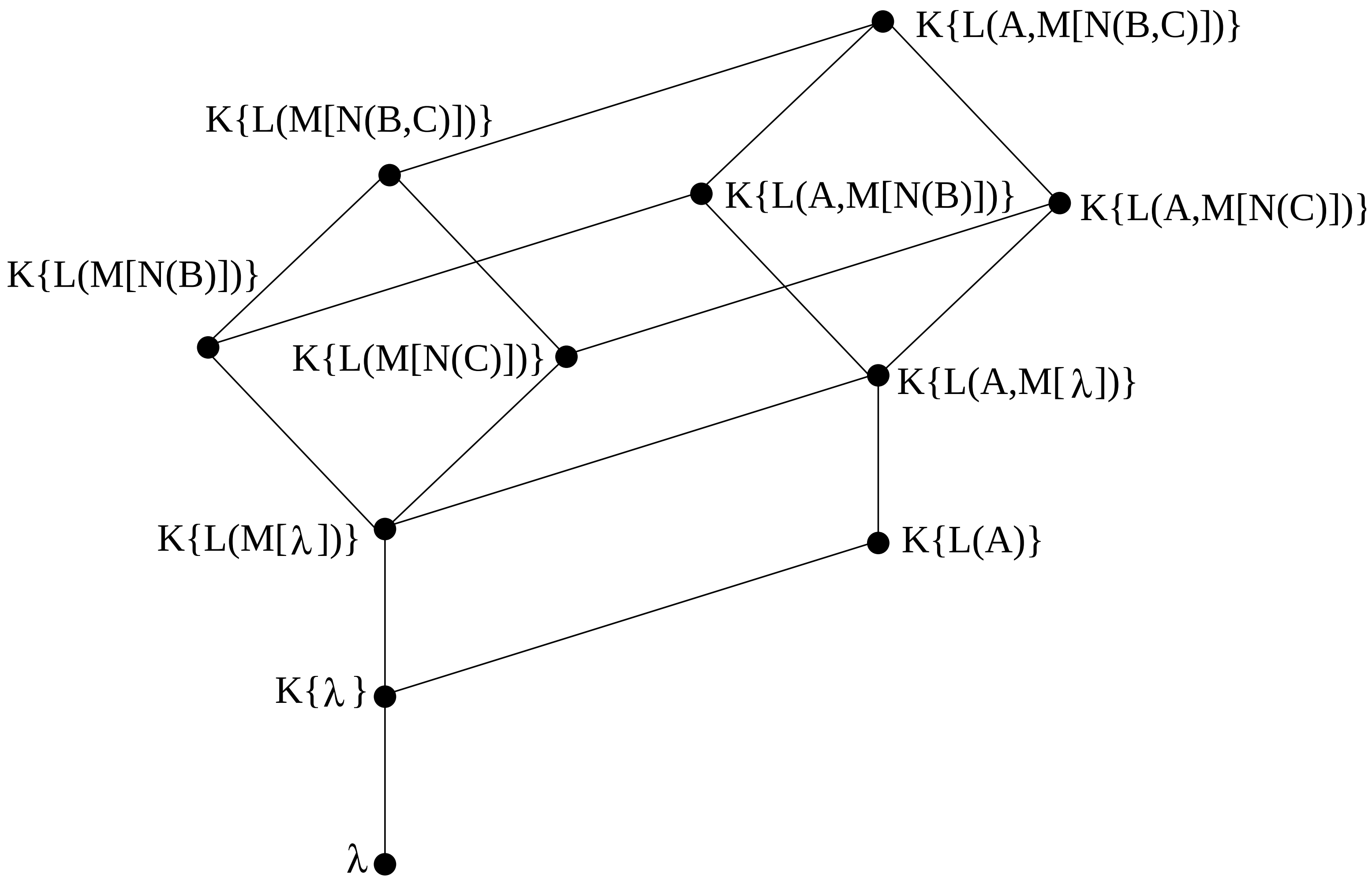
  for all $X \in Sub(N)$

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

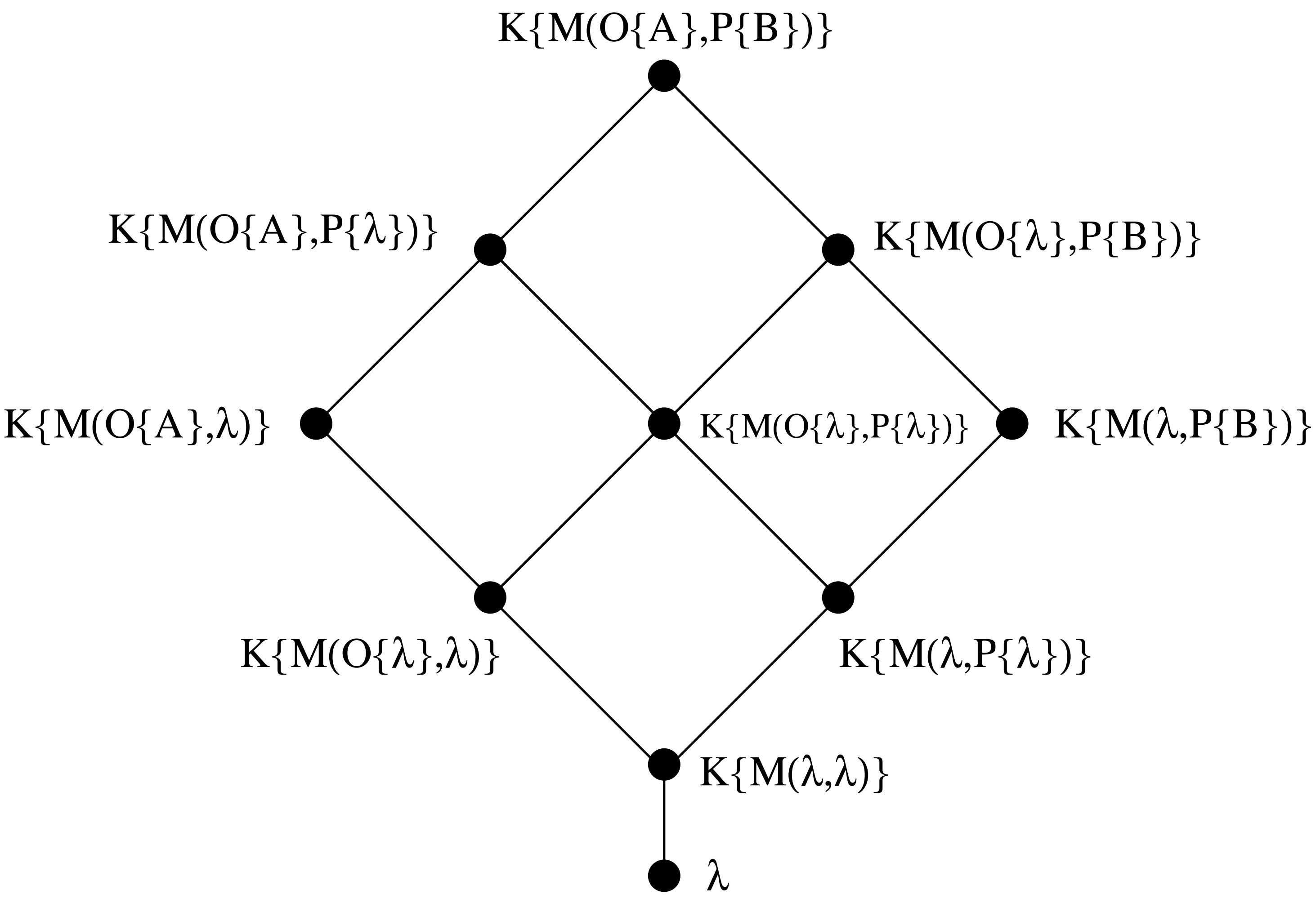
$$Y^{\mathcal{C}} \leq X \quad \text{if and only if} \quad X \sqcup Y = N$$

- 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

2.7 The Algebra of Nested Attributes: An Example



2.8 The Algebra of Nested Attributes: A further Example



3.1 Functional Dependencies

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

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

- finite $r \subseteq Dom(N)$ **satisfies** $\mathcal{X} \rightarrow \mathcal{Y}$ on N ($\models_r \mathcal{X} \rightarrow \mathcal{Y}$) iff

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

- implication: $\Sigma \models \tau$ iff $\models_r \tau$ if $\models_r \sigma$ for all $\sigma \in \Sigma$ and any (finite) r

- **semantic hull**: $\Sigma^* = \{\sigma \mid \Sigma \models \sigma\}$

- **syntactic hull**: $\Sigma^+ = \{\sigma \mid \Sigma \vdash_{\mathfrak{R}} \sigma\}$ for set $\mathfrak{R}$ of inference rules

- goal: find **sound** and **complete** $\mathfrak{R}$, i.e., $\Sigma^+ \subseteq \Sigma^*$ and $\Sigma^* \subseteq \Sigma^+$

3.2 A fundamental Difference

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

$$t_1{=}\{(\text{Borussia, Schalke}), (\text{Hansa, Bayern})\} \text{ and}$$
$$t_2{=}\{(\text{Hansa, Schalke}), (\text{Borussia, Bayern})\}$$

- $\models_r \text{Soccer}\{\text{Match}(\text{Winner})\} \rightarrow \text{Soccer}\{\text{Match}(\text{Loser})\}$
- $\not\models_r \text{Soccer}\{\text{Match}(\text{Winner})\} \rightarrow \text{Soccer}\{\text{Match}(\text{Winner}, \text{Loser})\}$
- values on subattributes X and Y do not determine values on $X \sqcup Y$
- the bad guys are: sets and multisets
- shows: FDs cannot be simplified to $X \rightarrow Y$ with $X, Y \in Sub(N)$
- FDs simpler in case of records and lists only

3.3 Semi-Disjoint Attributes

- $X, Y \in Sub(N)$ **semi-disjoint** if and only if one of the following holds:
  - $Y \leq X$ or $X \leq Y$,
  - $N = L(N_1, \ldots, N_k), X = L(X_1, \ldots, X_k), Y = L(Y_1, \ldots, Y_k)$ where X_i and Y_i are semi-disjoint for all $i = 1, \ldots, k$,
  - $N = L[N'], X = L[X'], Y = L[Y']$ where X' and Y' are semi-disjoint

- previously: Soccer{Match(Winner,λ)} and Soccer{Match(λ,Loser)} are not semi-disjoint

3.4 Major Result

- Let $N \in \mathcal{N}A$ and $X, Y, Z \in Sub(N)$. The Armstrong Axioms, i.e.,

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

form a minimal, sound and complete set of inference rules for the implication of FDs in the presence of records, and records and lists.

- Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq Sub(N)$ be non-empty, and $\mathcal{T}$ be any non-empty subset of {lists, sets, multisets} apart from {lists}. The generalised Armstrong Axioms, i.e.,

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

$$\frac{}{\{X, Y\} \to \{X \sqcup_N Y\}}\, X, Y \text{semi-disjoint}, \quad \frac{\mathcal{X} \to \mathcal{Y},\ \mathcal{Y} \to \mathcal{Z}}{\mathcal{X} \to \mathcal{Z}},$$

form a minimal, sound and complete set of inference rules for the implication of FDs in the presence of records and $\mathcal{T}$. □

3.5 Useful Inference Rules

- the following rules are derivable by the generalised Armstrong Axioms:
  - λ-axiom: $\dfrac{}{\mathcal{X} \to \{\lambda\}}$
  - union rule: $\dfrac{\mathcal{X} \to \mathcal{Y}, \mathcal{X} \to \mathcal{Z}}{\mathcal{X} \to \mathcal{Y} \cup \mathcal{Z}}$
  - subattribute rule: $\dfrac{\mathcal{X} \to \{Z\}}{\mathcal{X} \to \{Y\}}\ Y \leq Z$
  - subset rule: $\dfrac{\mathcal{X} \to \mathcal{Z}}{\mathcal{X} \to \mathcal{Y}}\ \mathcal{Y} \subseteq \mathcal{Z}$

3.6 The Completeness Proof: Outline

- Σ set of FDs on some N, take $\mathcal{X} \rightarrow \mathcal{Y} \notin \Sigma^+$ and show $\mathcal{X} \rightarrow \mathcal{Y} \notin \Sigma^*$
- $\mathcal{X}^+ = \{Z \mid \mathcal{X} \rightarrow \{Z\} \in \Sigma^+\}$ forms $\leq$-ideal (subattribute rule)
- $\lambda \in \mathcal{X}^+$ (λ axiom)
- semi-disjoint $X, Y \in \mathcal{X}^+$ imply $X \sqcup_N Y \in \mathcal{X}^+$ (restricted join axiom)
- define $r = \{t_1, t_2\} \subseteq Dom(N)$ by: $\pi_W^N(t_1) = \pi_W^N(t_2)$ iff $W \in \mathcal{X}^+$
- $\mathcal{Y} \not\subseteq \mathcal{X}^+$, $\mathcal{X} \subseteq \mathcal{X}^+$ (subset rule, union rule): $\not\models_r \mathcal{X} \rightarrow \mathcal{Y}$
- $\models_r \mathcal{U} \rightarrow \mathcal{V}$ for every $\mathcal{U} \rightarrow \mathcal{V} \in \Sigma$:
  - $\mathcal{U} \not\subseteq \mathcal{X}^+$: nothing to show
  - $\mathcal{U} \subseteq \mathcal{X}^+$: $\mathcal{X} \rightarrow \mathcal{X}^+ \in \Sigma^+$, $\mathcal{X} \rightarrow \mathcal{U} \in \Sigma^+$, $\mathcal{X} \rightarrow \mathcal{V} \in \Sigma^+$ imply $\mathcal{V} \subseteq \mathcal{X}^+$
- $\mathcal{X} \rightarrow \mathcal{Y} \notin \Sigma^*$

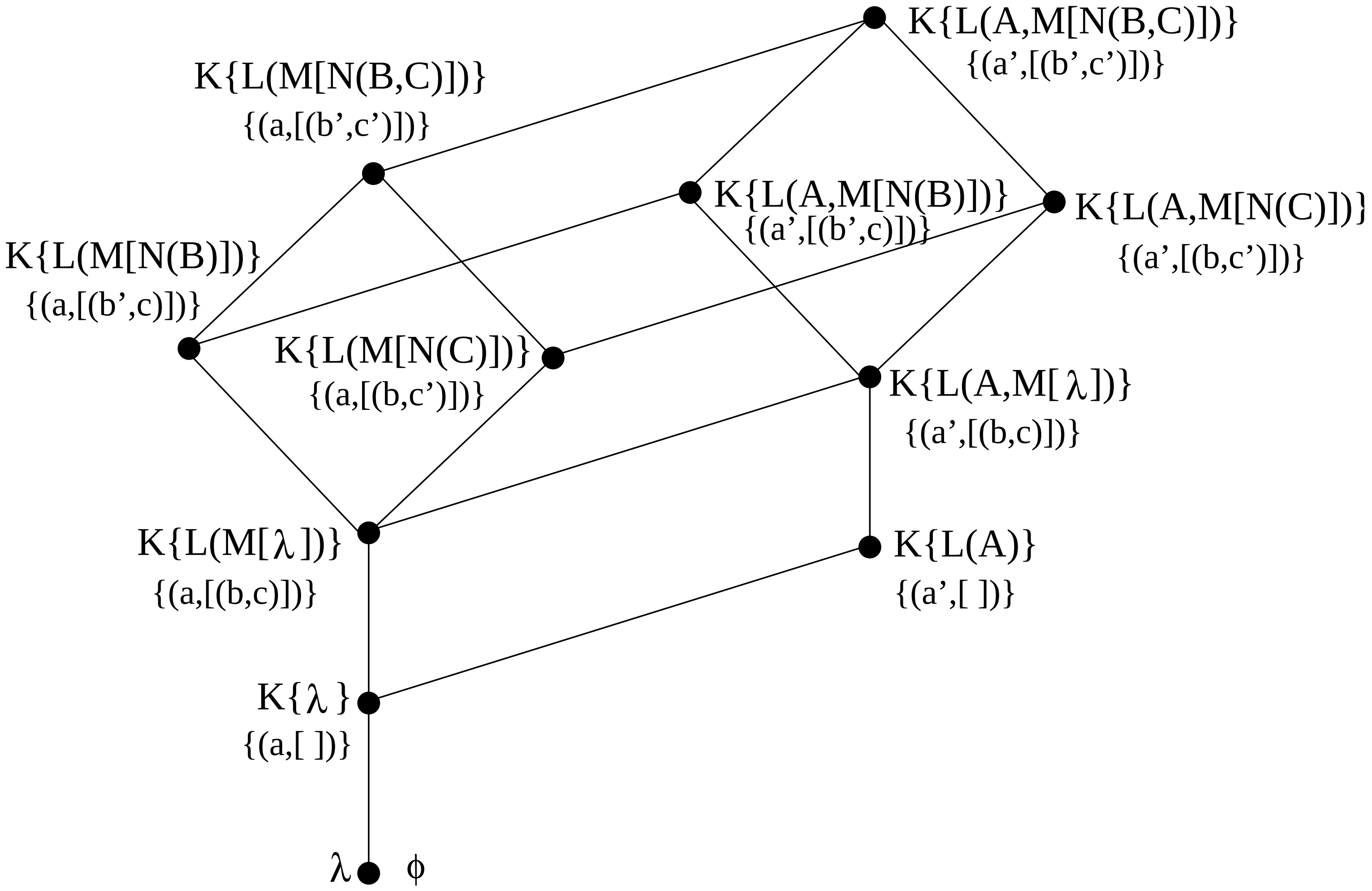
3.7 The Main Lemma

- $N \in \mathcal{N}$A, $\emptyset \neq \mathcal{I} \subseteq Sub(N)$ $\leq$-ideal closed under join of semi-disjoint attributes: $\exists t_N, t'_N \in Dom(N)$ with $\pi^N_X(t_N) = \pi^N_X(t'_N)$ iff $X \in \mathcal{I}$
- $\mathcal{I} = \{\lambda\}$, $\mathcal{I} = \{\lambda, A\}$ for $N \in \mathcal{U}$
- $N = L(N_1, \dots, N_k)$:
  - $\mathcal{I}_j = \{X \sqcap_N L(N_j) : X \in \mathcal{I}\} \subseteq Sub(L(N_j))$ non-empty $\leq$-ideals
  - closed under join of semi-disjoint attributes
  - define $t_N = (t_{L(N_1)}, \dots, t_{L(N_n)})$ and $t'_N = (t'_{L(N_1)}, \dots, t'_{L(N_n)})$
- $N = L[N']$:
  - $\mathcal{I} = \{L[M] : M \in \mathcal{J}\} \cup \{\lambda\}$ for ideal $\mathcal{J} \subseteq Sub(N')$
  - $\mathcal{J} = \emptyset$: define $t_N = [\,]$ and $t'_N = [n'] \in Dom(N)$, else:
  - $\mathcal{J}$ non-empty $\leq$-ideal closed under join of semi-disjoint attributes
  - define $t_N = [t_{N'}], t'_N = [t'_{N'}] \in Dom(N)$
- remains to consider set- and multiset-valued attributes

3.8 Identifying Terms

- identifying term $\tau_N(X)$ of $X \in Sub(N)$ defined as follows:
  - $\tau_\lambda(\lambda) = ok$,
  - $\tau_A(\lambda) = a, \tau_A(A) = a'$ with $a, a' \in Dom(A)$ and $a \neq a'$ for $A \in \mathcal{U}$,
  - $\tau_{L(N_1,\dots,N_k)}(L(M_1,\dots,M_k)) = (\tau_{N_1}(M_1),\dots,\tau_{N_k}(M_k))$,
  - $\tau_{L\{N\}}(L\{M\}) = \{\tau_N(M)\}$ and $\tau_{L\{N\}}(\lambda) = \emptyset$,
  - $\tau_{L\langle N\rangle}(L\langle M\rangle) = \langle\tau_N(M)\rangle$ and $\tau_{L\{N\}}(\lambda) = \langle\,\rangle$,
  - $\tau_{L[N]}(L[M]) = [\tau_N(M)]$ and $\tau_{L[N]}(\lambda) = [\,]$.
- $\pi^N_X(\tau_N(Y)) = \pi^N_X(\tau_N(X))$ implies $X \leq Y$
- $\pi^N_Y(\tau_N(X)) = \pi^N_Y(\tau_N(X \sqcap Y))$

3.9 Example: Identifying Terms

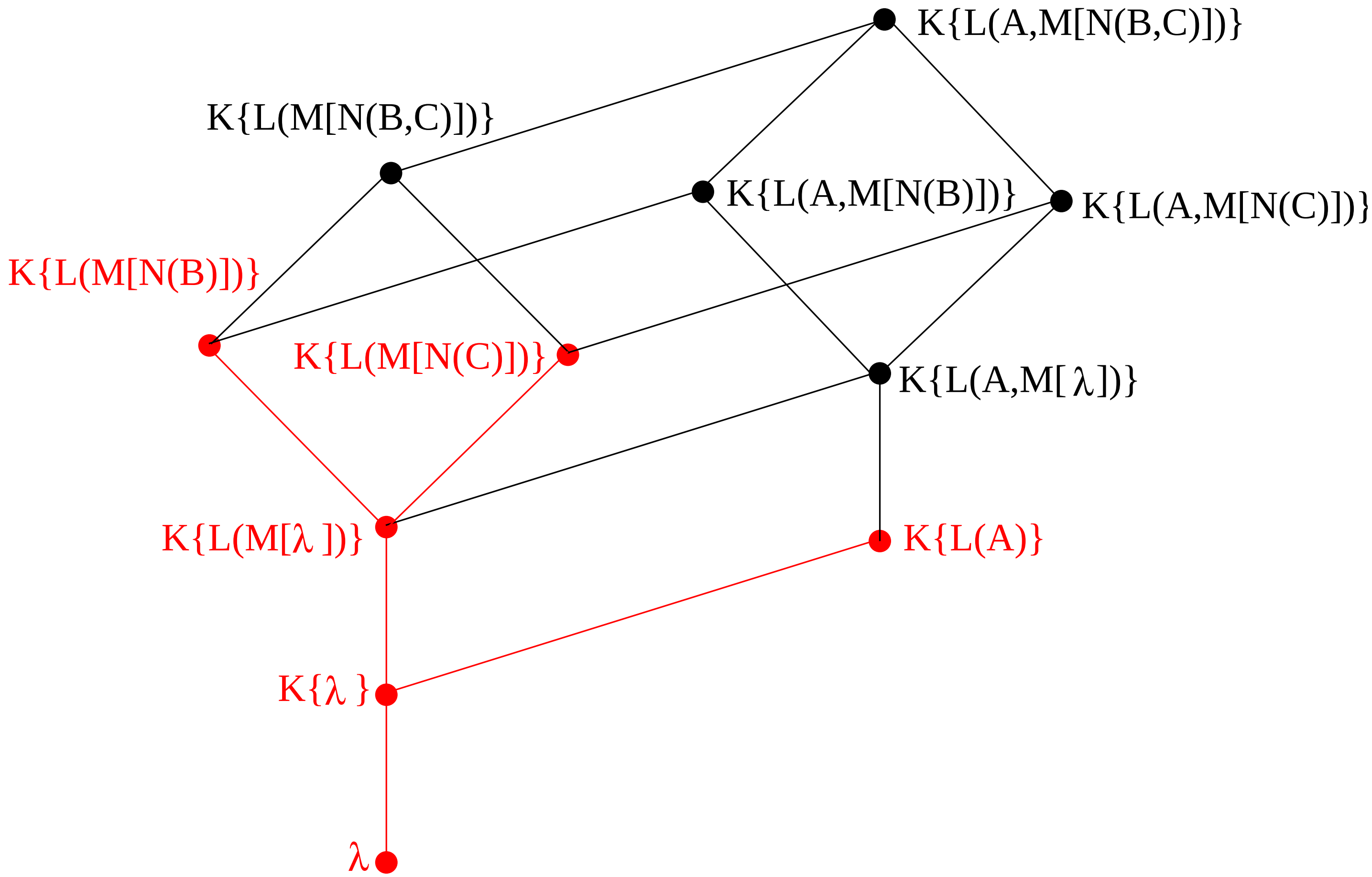


3.10 The Set Case

- $N = L\{P\} \in \mathcal{N}\mathrm{A}$, $\emptyset \neq \mathcal{X} \subseteq Sub(N)$ $\leq$-ideal: then there are $t_N, t'_N \in Dom(N)$ with $\pi^N_W(t_N) = \pi^N_W(t'_N)$ iff $W \in \mathcal{X}$
- $\mathcal{X} = \{L\{X\} : X \in \mathcal{Y}\} \cup \{\lambda\}$ for some $\mathcal{Y} \subseteq Sub(P)$
- $t_N = \{\tau_P(X) : X \leq P\}$ and $t'_N = \{\tau_P(X) : X \in \mathcal{Y}\}$
- $W = \lambda$: $\pi^N_\lambda(t_N) = ok = \pi^N_\lambda(t'_N)$ and $W = L\{V\}$: $\{\pi^P_V(\tau_P(X)) : X \leq P\} = \{\pi^P_V(\tau_P(X)) : X \in \mathcal{Y}\}$ iff $V \in \mathcal{Y}$
- $\{\pi^P_V(\tau_P(X)) : X \in \mathcal{Y}\} \subseteq \{\pi^P_V(\tau_P(X)) : X \leq P\}$ since $\mathcal{Y} \subseteq Sub(P)$
- $V \in \mathcal{Y}$ implies $\{\pi^P_V(\tau_P(X)) : X \leq P\} \subseteq \{\pi^P_V(\tau_P(X)) : X \in \mathcal{Y}\}$ since for all $X \leq P$ there is some $Y \in \mathcal{Y}$ with $\pi^P_V(\tau_P(X)) = \pi^P_V(\tau_P(Y))$ (for $X \notin \mathcal{Y}$ this is $Y = X \sqcap V$)
- $V \notin \mathcal{Y}$ implies $V \not\leq X$ for all $X \in \mathcal{Y}$ and $\pi^P_V(\tau_P(X)) \neq \pi^P_V(\tau_P(V))$ for all $X \in \mathcal{Y}$, i.e., $\pi^P_V(\tau_P(V)) \in \{\pi^P_V(\tau_P(X)) : X \leq P\}$, but $\pi^P_V(\tau_P(V)) \notin \{\pi^P_V(\tau_P(X)) : X \in \mathcal{Y}\}$

3.11 The Set Case: An Example I

- suppose $\mathcal{X}^+$ is:



- $K\{L(A)\} \to K\{L(M[N(B)])\}$, $K\{L(A)\} \to K\{L(M[N(C)])\}$

3.12 The Set Case: An Example II

- $t_N = \{\tau_{L(A,M[N(B,C)])}(X) : X \leq L(A,M[N(B,C)])\}$ is

$$\{(a',[(b',c')]);(a,[(b',c')]);(a',[(b',c)]);(a',[(b,c')]);(a,[(b',c)]);\\ (a,[(b,c')]);(a',[(b,c)]);(a,[(b,c)]);(a',[\,]);(a,[\,])\}$$

- $t'_N = \{\tau_{L(A,M[N(B,C)])}(Y) : Y \in \mathcal{Y}\}$ is

$$\{(a,[(b',c)]);(a,[(b,c')]);(a,[(b,c)]);(a',[\,]);(a,[\,])\}$$

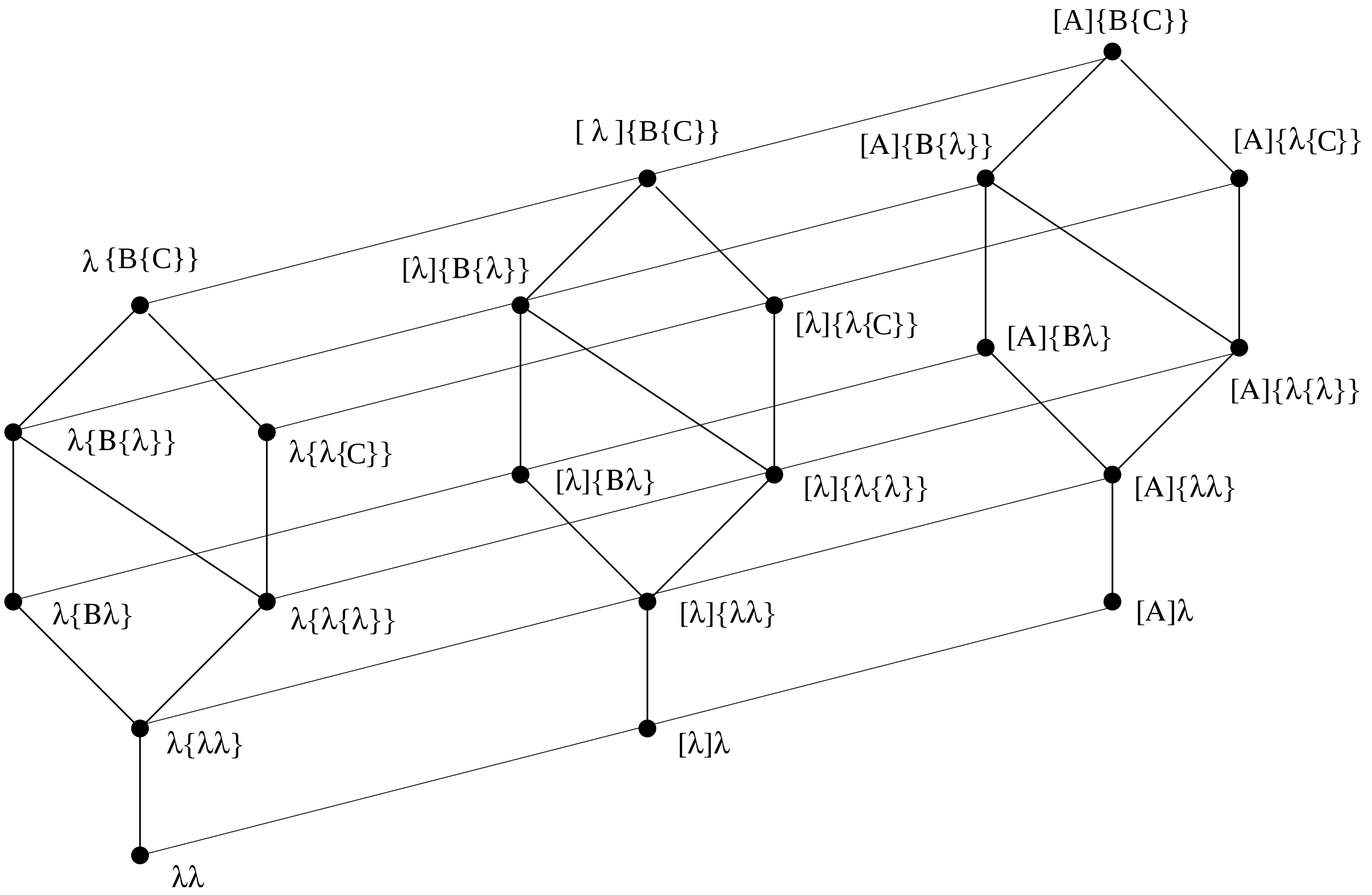
- projections $\pi^N_W(t_N)$ and $\pi^N_W(t'_N)$ for all $W \in Sub(N)$:

W	$\pi^N_W(t_N)$	$\pi^N_W(t'_N)$
$K\{L(M[N(B)])\}$	$\{(ok,[(b',ok)]);(ok,[(b,ok)]);(ok,[\,])\}$	
$K\{L(M[N(C)])\}$	$\{(ok,[(ok,c')]);(ok,[(ok,c')]);(ok,[\,])\}$	
$K\{L(A)\}$	$\{(a',ok);(a,ok)\}$	
$K\{L(M[N(B,C)])\}$	$\{(ok,[(b,c)]);(ok,[(b',c)]),(ok,[(b,c')]);$ $(\mathbf{ok},[(\mathbf{b}',\mathbf{c}')]);(ok,[\,])\}$	$\{(ok,[(b,c)]);(ok,[(b',c)]);$ $(ok,[(b,c')]);(ok,[\,])\}$
$K\{L(A,M[\lambda])\}$	$\{(a,[(ok,ok)]);(a,[\,]);(\mathbf{a}',[(\mathbf{ok},\mathbf{ok})]);(a',[\,])\}\}$	$\{(a,[(ok,ok)]);(a,[\,]);(a',[\,])\}$

3.13 The Multiset Case - Embedded Boolean Algebras

- strategy for set-valued attributes does not work for multiset-valued attributes since multiple occurences of projections do not vanish in a multiset
- for $X, Y \in Sub(N)$ with $X \leq Y$ define $[X, Y] = \{Z \in Sub(N) : X \leq Z \leq Y\}$
- $\{X_1, \ldots, X_k\}$ set of all $\leq$-maximal proper subattributes of X
- $([0_X, X], \leq, \sqcap, \sqcup, \overline{(\cdot)}, 0_X, X)$ forms Boolean Algebra with
  - $0_X = X \sqcap X_1 \sqcap \cdots \sqcap X_k$
  - $\overline{Y} = (X \dot{-} Y) \sqcup 0_X$ for all $Y \in [0_X, X]$

3.14 Embedded Boolean Algebras in $M = K(J[A], O\{P(B, Q\{C\})\})$



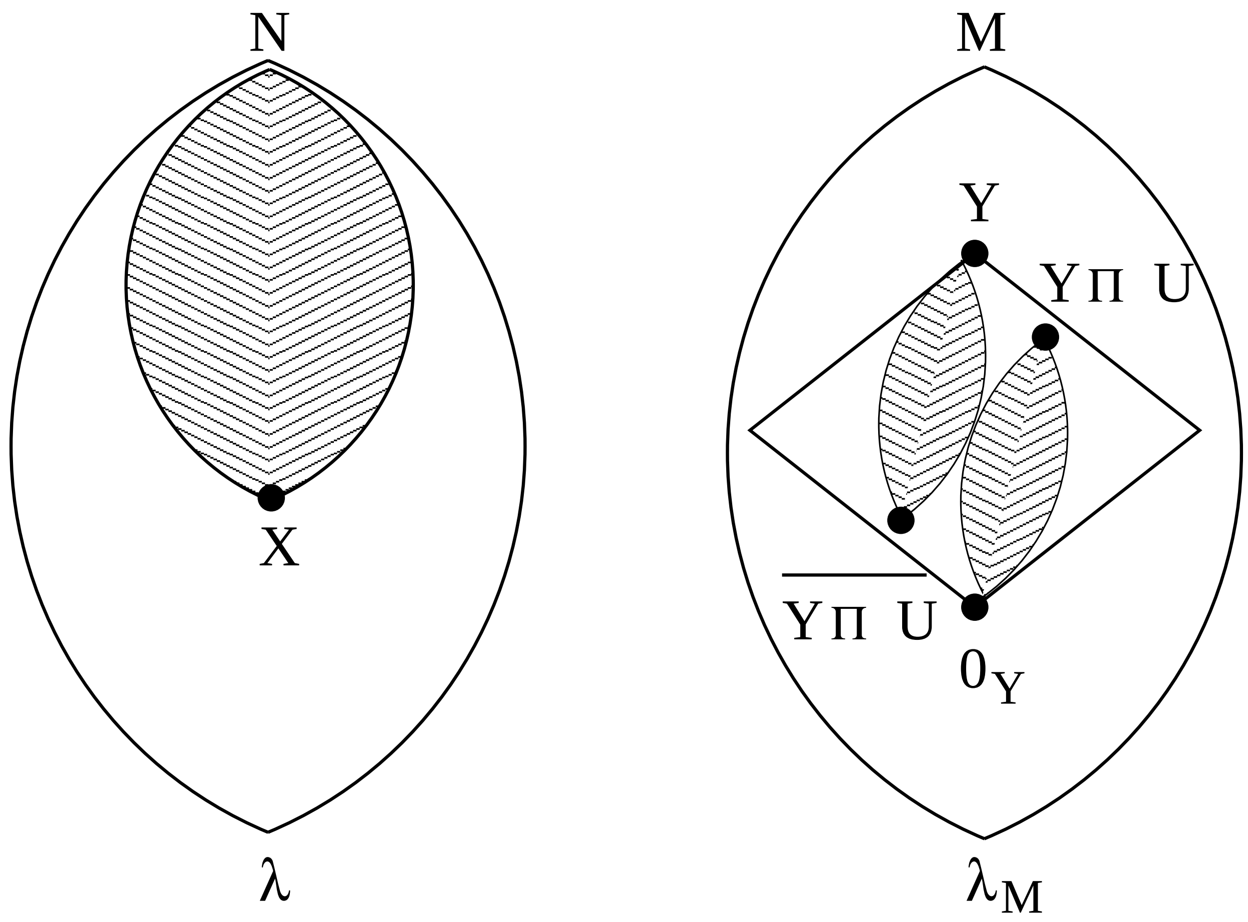
3.15 The Multiset Case - Preparations

- $N = L\langle M\rangle \in \mathcal{N}\mathrm{A}$ and $\lambda \neq X = L\langle Y\rangle \leq N$:

$$\exists t_1, t_2 \in Dom(N) \text{ with } \pi^N_W(t_1) \neq \pi^N_W(t_2) \text{ iff } X \leq W$$

- $\mathcal{L}_i$ denotes ith level of $([0_Y, Y], \leq)$
- $t_1 = \langle \tau_M(Z) \ : \ Z \in \mathcal{L}_i, i \text{ even}\rangle$ and $t_2 = \langle \tau_M(Z) \ : \ Z \in \mathcal{L}_i, i \text{ odd}\rangle$
- $\pi^M_Y(\tau_M(Y))$ element of either t_1 or t_2
- $\pi^N_X(t_1) \neq \pi^N_X(t_2)$, and $\pi^N_W(t_1) \neq \pi^N_W(t_2)$ whenever $X \leq W$
- $\pi^N_V(t_1) = \pi^N_V(t_2)$ for $\leq$-maximal $V \in Sub(N)$ with $X \not\leq V$
- $V = L\langle U\rangle$ where U is $\leq$-maximal $U \in Sub(M)$ with $Y \not\leq U$
- $Y \sqcap U$ is always a $\leq$-maximal proper subattribute of Y (co-atom)
- $Z \mapsto Z \sqcup \overline{Y \sqcap U}$ is bijection from $[0_Y, Y \sqcap U]$ to $[\overline{Y \sqcap U}, Y]$
- $\overline{Y \sqcap U}$ is atom: $\tau_M(Z \sqcup \overline{Y \sqcap U}) \in t_2$, if $\tau_M(Z) \in t_1$
- $\pi^M_U(\tau_M(Z)) = \pi^M_U(\tau_M(Z \sqcup \overline{Y \sqcap U}))$ for $Z \in [0_Y, Y \sqcap U]$

3.16 The Multiset Case - Illustration



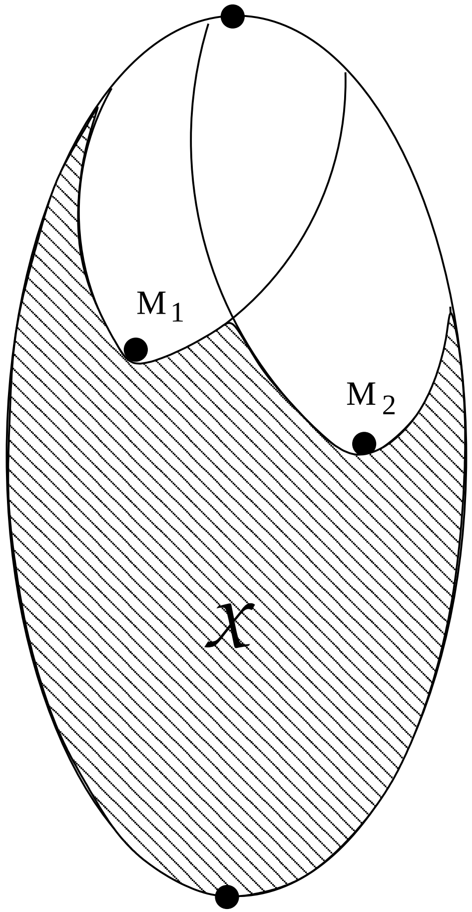
3.17 The Multiset Case

- $N = L\langle P\rangle \in \mathcal{N}\mathrm{A}$, $\emptyset \neq \mathcal{X} \subseteq Sub(N)$ $\leq$-ideal:

$$\exists t_N, t'_N \in Dom(N) \text{ with } \pi^N_W(t_N) = \pi^N_W(t'_N) \text{ iff } W \in \mathcal{X}$$

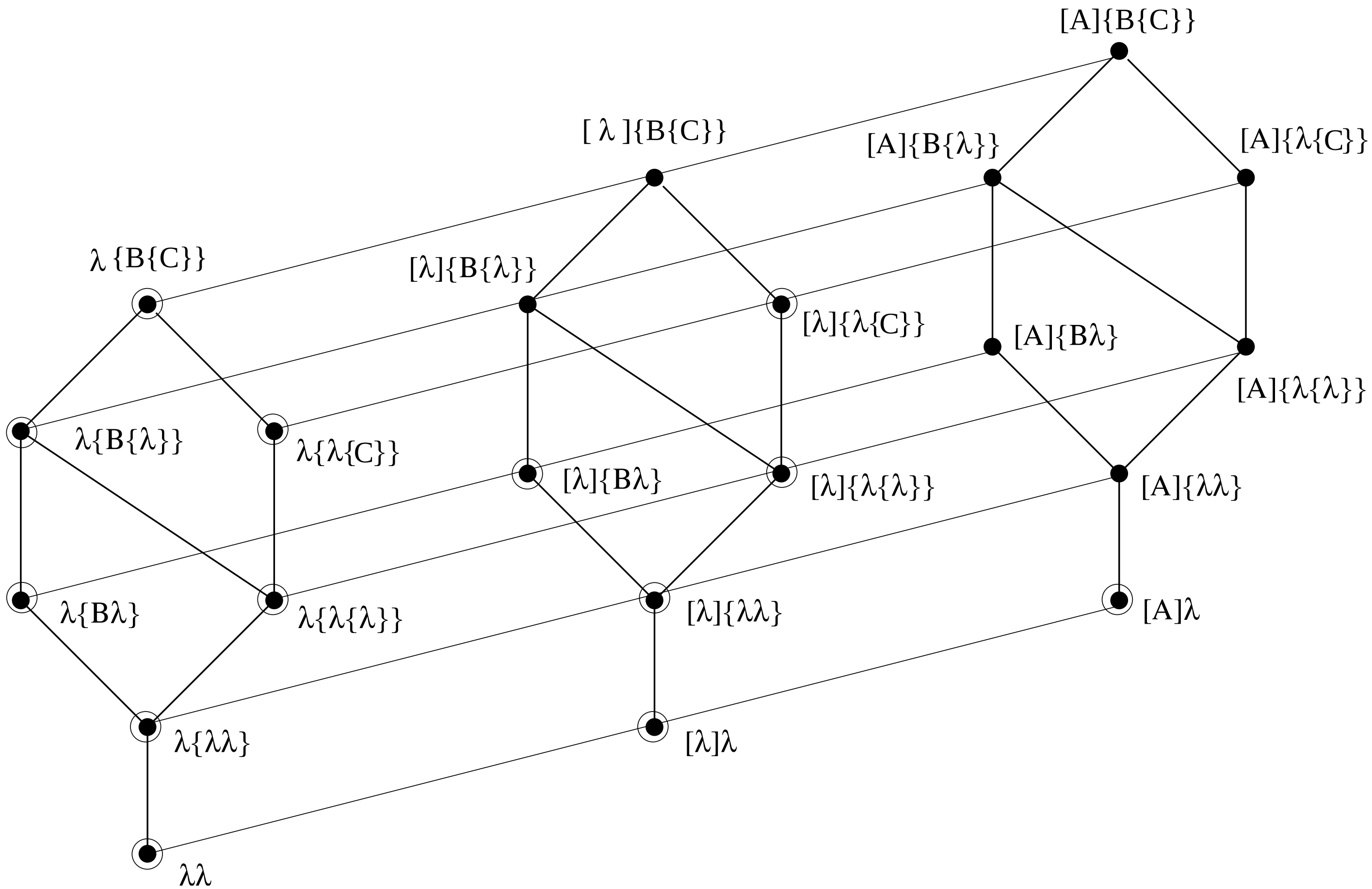
- $\{M_1, \dots, M_n\} \subseteq Sub(N)$ are all $\leq$-minimal with $M_i \notin \mathcal{X}$
- $\exists t_{M_i}, t'_{M_i} \in Dom(N)$ with $\pi^N_Z(t_{M_i}) \neq \pi^N_Z(t'_{M_i})$ iff $M_i \leq Z$
- $t_N = \bigcup_{i=1}^{n} t_{M_i}$ and $t'_N = \bigcup_{i=1}^{n} t'_{M_i}$ (multiset union)
- $W \in \mathcal{X}$:
  - $M_i \not\leq W$ for all $i = 1, \dots, n$
  - $\pi^N_W(t_{M_i}) = \pi^N_W(t'_{M_i})$ or all $i = 1, \dots, n$ and $\pi^N_W(t_N) = \pi^N_W(t'_N)$
- $W \notin \mathcal{X}$:
  - $\exists j$ with $1 \leq j \leq n$ such that $M_j \leq W$
  - $\pi^N_W(\tau_N(M_j))$ either in $\pi^N_W(t_N)$ or $\pi^N_W(t'_N)$, i.e. $\pi^N_W(t_N) \neq \pi^N_W(t'_N)$

3.18 Illustration



3.19 Example - Multiset Case

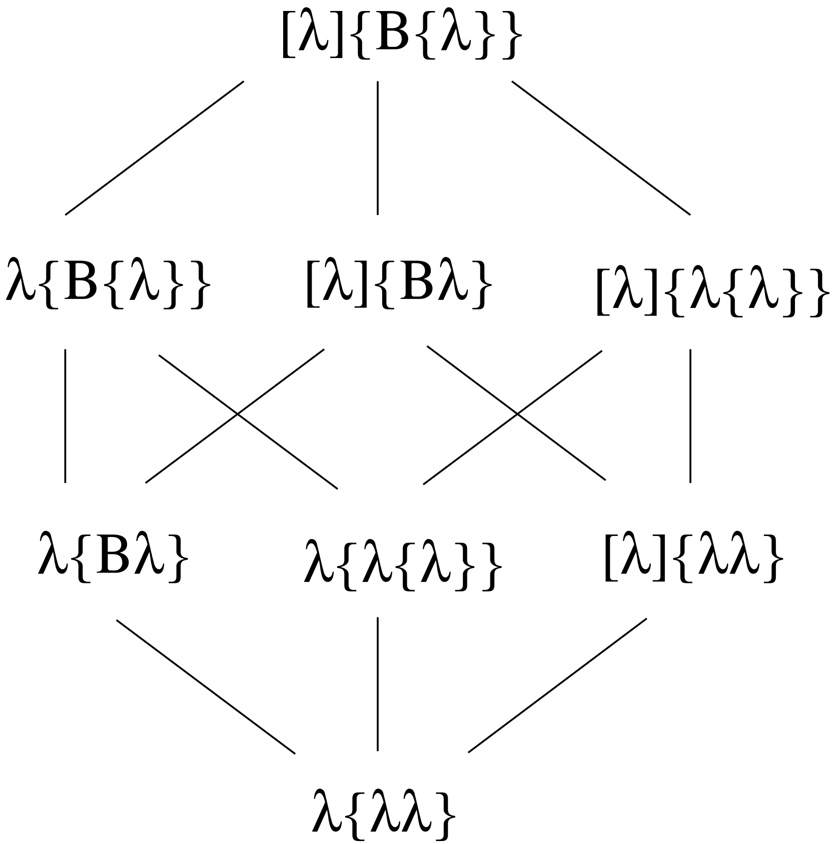
- $N = L\langle M \rangle$ with $M = K(J[A], O\{P(B, Q\{C\})\})$

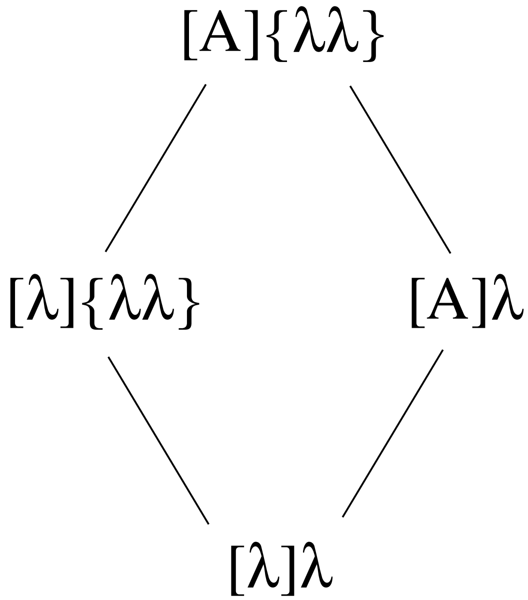


- $\mathcal{X} = \{L\langle X \rangle \ : \ X \in \mathcal{Y}\}$ and $\mathcal{Y}$ has all encircled attributes of M

3.20 Example continued - Multiset Case

- $\leq$-minimal $V \in Sub(N)$ with $V \notin \mathcal{X}$ are $V_1 = L\langle K(J[\lambda], O\{P(B, Q\{\lambda\})\})\rangle, V_2 = L\langle K(J[A], O\{P(\lambda, \lambda)\})\rangle$





3.21 Example continued - Multiset Case

- choose:

$$t'_1 = \langle([\,],\{(b,\emptyset)\});([\,],\{(b',\{c\})\});([a],\{(b',\emptyset)\});([a],\{(b,\{c\})\})\rangle$$
$$t'_2 = \langle([\,],\{(b',\emptyset)\});([\,],\{(b,\{c\})\});([a],\{(b,\emptyset)\});([a],\{(b',\{c\})\})\rangle$$
$$t''_1 = \langle([a],\emptyset);([a'],\{(b,\emptyset)\})\rangle$$
$$t''_2 = \langle([a],\{(b,\emptyset)\});([a'],\emptyset)\rangle$$

- multiset-union gives:

$$\begin{aligned} t_N = t'_1 \cup t''_1 &= \langle([\,],\{(b,\emptyset)\});([\,],\{(b',\{c\})\});([a],\{(b',\emptyset)\}); \\ &\quad ([a],\{(b,\{c\})\});([a],\emptyset);([a'],\{(b,\emptyset)\})\rangle \\ t'_N = t'_2 \cup t''_2 &= \langle([\,],\{(b',\emptyset)\});([\,],\{(b,\{c\})\});([a],\{(b,\emptyset)\}); \\ &\quad ([a],\{(b',\{c\})\});([a],\{(b,\emptyset)\});([a'],\emptyset)\rangle \end{aligned}$$

- $\pi^N_W(t_N) = \pi^N_W(t'_N)$ for all $\leq$-maximal $W \in \mathcal{X}$
- $\pi^N_{V_1}(t_N) \neq \pi^N_{V_1}(t'_N)$ and $\pi^N_{V_2}(t_N) \neq \pi^N_{V_2}(t'_N)$

3.22 A Remark on Semi-Disjointness

- show that semi-disjointness of $X, Y \in Sub(N)$ is an exact condition for soundness of

$$\overline{\{X, Y\} \rightarrow \{X \sqcup_N Y\}}$$

- sufficiency is not hard to see
- if X, Y not semi-disjoint, then $r \subseteq Dom(N)$ with $\not\models_r \{X, Y\} \rightarrow \{X \sqcup_N Y\}$ exists
- by Main Lemma: find suitable ideal $\mathcal{I}$ with $X, Y \in \mathcal{I}$ and $X \sqcup_N Y \notin \mathcal{I}$
- $\mathcal{I} = \{U \sqcup_N V : U \leq X, V \leq Y, U$ and V are semi-disjoint$\}$ is non-empty $\leq$-ideal that is closed under the join of semi-disjoint attributes

4.1 Minimality

- no proper subset of the generalised Armstrong Axioms is still complete
- all rules are independent of one another
- none of the rules can be infered from the other inference rules
- closure of set Σ of FDs on N under set $\mathfrak{R}$ of inference rules is $\Sigma^{\mathfrak{R}}$
- R is not implied by $\mathfrak{R}$, if there is some Σ and some σ on N with $\sigma \notin \Sigma^{\mathfrak{R}}$, but $\sigma \in \Sigma^{\mathfrak{R} \cup \{R\}}$

4.2 Independence of Reflexivity Axiom

- The reflexivity axiom is not implied by $\mathfrak{R}$ ={subattribute axiom, extension rule, restricted join axiom, transitivity rule}
- $N = L\{A\}, \Sigma = \emptyset$ and $\sigma = \{\lambda, L\{\lambda\}, L\{A\}\} \rightarrow \{\lambda\}$

	$\{\lambda\}$	$\{L\{\lambda\}\}$	$\{L\{A\}\}$	$\{\lambda, L\{\lambda\}\}$	$\{\lambda, L\{A\}\}$	$\{L\{\lambda\}, L\{A\}\}$	$\{\lambda, L\{\lambda\}, L\{A\}\}$
$\{\lambda\}$	×						
$\{L\{\lambda\}\}$	×	×		×			
$\{L\{A\}\}$	×	×	×	×	×	×	×
$\{\lambda, L\{\lambda\}\}$	×	×		×			
$\{\lambda, L\{A\}\}$	×	×	×	×	×	×	×
$\{L\{\lambda\}, L\{A\}\}$	×	×	×	×	×	×	×
$\{\lambda, L\{\lambda\}, L\{A\}\}$							

- $\sigma \notin \Sigma^{\mathfrak{R}}$
- $\{\lambda\} \subseteq \{\lambda, L\{\lambda\}, L\{A\}\}$: σ can be infered using the reflexivity axiom

4.3 Independence of Subattribute Axiom

- The subattribute axiom is not implied by $\mathfrak{R}$ ={reflexivity axiom, extension rule, restricted join axiom, transitivity rule}
- $N = L(A)$, $\Sigma = \emptyset$ and $\sigma = \{L(A)\} \to \{\lambda\}$

	$\{\lambda\}$	$\{L(A)\}$	$\{\lambda, L(A)\}$
$\{\lambda\}$	$\times$		
$\{L(A)\}$		$\times$	
$\{\lambda, L(A)\}$	$\times$	$\times$	$\times$

- $\sigma \notin \Sigma^{\mathfrak{R}}$
- $\lambda \leq L(A)$: σ can be infered using the subattribute axiom

4.4 Independence of Extension Rule

- The extension rule is not implied by $\mathfrak{R}$ ={reflexivity axiom, subattribute axiom, restricted join axiom, transitivity rule}

- $N = L(A)$, $\Sigma = \emptyset$ and $\sigma = \{L(A)\} \to \{\lambda, L(A)\}$

	$\{\lambda\}$	$\{L(A)\}$	$\{\lambda, L(A)\}$
$\{\lambda\}$	$\times$		
$\{L(A)\}$	$\times$	$\times$	
$\{\lambda, L(A)\}$	$\times$	$\times$	$\times$

- $\sigma \notin \Sigma^{\mathfrak{R}}$

- $\{L(A)\} \to \{\lambda\} \in \Sigma^{\mathfrak{R}}$: σ can be infered using extension rule and $\mathfrak{R}$

4.5 Independence of the Restricted Join Axiom

- The restricted join axiom is not implied by $\mathfrak{R}$ ={reflexivity axiom, subattribute axiom, extension rule, transitivity rule}
- $N = L(A, B), \Sigma = \emptyset$ and $\sigma = \{L(A), L(B)\} \rightarrow \{L(A, B)\}$
- $\sigma \notin \Sigma^{\mathfrak{R}}$
- $L(A)$ and $L(B)$ are semi-disjoint: σ can be infered using the restricted join axiom

4.6 Independence of the Transitivity Rule

- The transitivity rule is not implied by $\mathfrak{R}$ ={reflexivity axiom, subattribute axiom, extension rule, restricted join axiom}
- $N = L\{A\}, \Sigma = \emptyset$ and $\sigma = \{L\{\lambda\}, L\{A\}\} \to \{\lambda\}$

	$\{\lambda\}$	$\{L\{\lambda\}\}$	$\{L\{A\}\}$	$\{\lambda, L\{\lambda\}\}$	$\{\lambda, L\{A\}\}$	$\{L\{\lambda\}, L\{A\}\}$	$\{\lambda, L\{\lambda\}, L\{A\}\}$
$\{\lambda\}$	$\times$						
$\{L\{\lambda\}\}$	$\times$	$\times$		$\times$			
$\{L\{A\}\}$	$\times$	$\times$	$\times$		$\times$	$\times$	
$\{\lambda, L\{\lambda\}\}$	$\times$	$\times$		$\times$			
$\{\lambda, L\{A\}\}$	$\times$		$\times$		$\times$		
$\{L\{\lambda\}, L\{A\}\}$		$\times$	$\times$			$\times$	
$\{\lambda, L\{\lambda\}, L\{A\}\}$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

- $\sigma \notin \Sigma^{\mathfrak{R}}$
- $\{L\{\lambda\}, L\{A\}\} \to \{\lambda\}$ can be infered from $\{L\{\lambda\}, L\{A\}\} \to \{L\{\lambda\}\}, \{L\{\lambda\}\} \to \{\lambda\} \in \Sigma^{\mathfrak{R}}$ by the transitivity rule
- σ can be infered from Σ using the transitivity rule and $\mathfrak{R}$

5.1 Comparison - General Remarks

- generalises theory from RDM (records only)
- approach based on explicit subattributes deviates significantly from previous approaches in Nested RDM, object-oriented data models and XML (unnesting)
- yielding complementary expressiveness
- algebraic approach based on Brouwerian Algebra is original
- not aware on any work regarding lists or multisets
- comparison focuses on Nested RDM (records and sets)

5.2 Hara/Davidson: Reasoning about Nested Functional Dependencies (PODS99)

- well-defined path expressions using records/sets (no arbitrary nesting)
- axiomatisation provided, but RHS of FDs only single path
- case where RHS is union of paths particularly interesting for sets
- FDs $\{S\{L(A)\}, S\{L(B)\}\} \rightarrow S\{L(A, B)\}$ cannot be expressed
- consider Course(ID,Participants{Student(Number, Age, Grade)}) with Course(Participants{Student(Number)})→ Course(ID)
- different from Course:[Participant → ID]
- Course:[Participant:Number → Participant:Age] (age of every participant consistent over all courses) only expressible on nested attribute Student(Number, Age, Grade)
- λ attributes

5.3 Levene/Loizou: Semantics for Null-Extended Nested Relations

- null extended FDs admit null values
- study relationship between MVDs $X \twoheadrightarrow Y$ and FDs $X \rightarrow Y^*$ (Y refers to complete unnesting of the set-valued attribute Y^*)
- again defined on the basis of paths
- FDs from RDM cannot be expressed
- set-valued attributes can only occur on the RHS of null extended FDs
- $N = L(A, K\{M(B, S\{C\})\})$ expressed as $A(BC^*)^*$ (simplified)
- null extended FDs: $A \rightarrow (BC^*)^*$ or $AB \rightarrow C^*$
- last of these is not covered yet by our data model
- $L(A, K\{M(B)\}) \rightarrow L(K\{M(S\{C\})\})$ not expressible as null extended FD

5.4 Object-Oriented Data Models and XML

- Weddell: *Reasoning about Functional Dependencies Generalized for Semantic Data Models* (ToDS 1992)
- Bommel/Weddell: *Reasoning About Equations and Functional Dependencies on Complex Objects* (TKDE 1994)
- Tari/Stokes/Spaccapietra: *Object Normal Forms and Dependency Constraints for Object-Oriented Schemata* (ToDS 1997)
- Arenas/Libkin: *A Normal Form for XML Documents* (PODS 2002)
- Vincent/Liu: *Functional dependencies for XML* (APWeb 2003)
- approaches based on a relational representation of the data
- different expressiveness
- FDs of Arenas/Libkin are not axiomatisable at all

6.1 Extensions: More Results

- **Record and List Type:**
  - FD implication decidable in linear time
  - NLNF (weaker than BCNF) proposed and justified
  - MVDs: minimal axiomatisation and finite implication problem
  - FDs/MVDs: minimal axiomatisation and finite implication problem
- **Record, List, Set and Multiset Type:**
  - FD implication decidable in $\mathcal{O}(n^4 \cdot s \cdot \min\{s, n\})$
- **XML**: several classes of FDs, Axiomatisations

6.2 Extensions: Future Research

- CVNF
- decomposition/synthesis for NLNF, CVNF
- unions, references
- interactions
- different classes of dependencies: inclusion and join dependencies
- XML: FIP and Normalisation