Reasoning about Functional Dependencies in an Abstract Data Model

Sven Hartmann and Sebastian Link

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

1. **Motivation & Revision of the RDM**
2. **An Abstract Data Model**
3. **Algebraic Foundations**
4. **Axiomatisation**
5. **Extensions**

1 Motivation & Revision of RDM

- apart from RDM **no universal agreement** on data models exists
- abstract approach based on **type systems** allows general treatment of all data models which support the types investigated
- **functional dependencies** well-studied in RDM:
  - **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 sound and complete set of Inference Rules
  - **Finite Implication Problem** decidable in linear time
  - **Boyce-Codd Normal Form** guarantees absence of redundancies and update anomalies caused by FDs
- here: **values, records of values and finite sets of values**

2.1 An Abstract Data Model: Nested Attributes

- capture characteristics of objects in target database by attributes
- $(\mathcal{U}, \mathcal{D}, \mathbb{B}, \mathit{type} : \mathcal{U} \to \mathbb{B}, \mathit{dom} : \mathbb{B} \to \mathcal{D})$
- type system $\mathbf{t} := \mathbf{b} \mid (\mathbf{a_1} : \mathbf{t_1}, \dots, \mathbf{a_n} : \mathbf{t_n}) \mid \{\mathbf{t}\}$
- extend *dom* for $\mathbb{B}$ to *Dom* for $\mathbb{T}$
- **nested attributes** $\mathcal{N}A(\mathcal{U}, \mathcal{L})$:
  - *flat attributes* $\mathcal{U} \subseteq \mathcal{N}A$, *null attribute* $\lambda \in \mathcal{N}A$
  - *tuple-valued attributes* $L(N_1, \dots, N_m) \in \mathcal{N}A$ for $L \in \mathcal{L}$, $N_1, \dots, N_m \in \mathcal{N}A$
  - *set-valued attributes* $L\{N\} \in \mathcal{N}A$ whenever $L \in \mathcal{L}$ and $N \in \mathcal{N}A$
- extend type assignment *type* on $\mathcal{U}$ to *Type* on $\mathcal{N}A(\mathcal{U}, \mathcal{L})$

2.2 An Abstract Data Model: Subattributes

- $\equiv \subseteq \mathcal{N}A \times \mathcal{N}A$ defined as smallest equivalence relation with
  - $L(N_1, \dots, N_m) \equiv L(N_{\pi(1)}, \dots, N_{\pi(m)})$ for all $\pi \in \mathcal{S}_m$
  - $L(N_1, \dots, N_k) \equiv L(M_1, \dots, M_k)$ whenever $N_i \equiv M_i$
  - $L(N_1, \dots, N_m, \lambda) \equiv L(N_1, \dots, N_m)$,
  - $L(\lambda) \equiv \lambda$ and
  - $L\{N\} \equiv L\{M\}$ iff $N \equiv M$
- $\leq \subseteq \mathcal{N}A \times \mathcal{N}A$ defined as smallest partial order with
  - $\lambda \leq N$ for all $N \in \mathcal{N}A$,
  - $L(N_1, \dots, N_k) \leq L(M_1, \dots, M_k)$ whenever $N_i \leq M_i$, and
  - $L\{N\} \leq L\{M\}$ whenever $N \leq M$
- **projection function** $\pi^N_M : Dom(N) \to Dom(M)$ for $M \leq N$

2.3 An Abstract Data Model: Schemata & Instances

- **database schema**: $(N, \mathcal{K}, \Sigma)$ with
  - nested attribute $N \in \mathcal{N}\mathrm{A}$
  - **key** $\mathcal{K} \subseteq Sub(N)$
  - **constraint set** Σ on N
- **instance** of N: finite set $r \subseteq Dom(N)$ such that
  - π_K^N is 1-to-1 on r for all $K \in \mathcal{K}$
  - $\models_r \Sigma$

3.1 Algebraic Foundations: A Brouwerian Algebra

- $Sub(N) = \{X \in \mathcal{N}A \mid X \leq N\}$: $Y \sqcup_N Z$, $Y \sqcap_N Z$ and $Y \dot{-}_N Z$ are:
  - $Y \sqcup_N Z = Z$ iff $Y \leq Z$ iff $Y \sqcap_N Z = Y$
  - $Z \dot{-}_N \lambda = Z$, and $Z \leq Y$ iff $Z \dot{-}_N Y = \lambda$,
  - if $N = L\{B\}, Y = L\{C\}, Z = L\{D\}$ with $C, D \leq B$, then

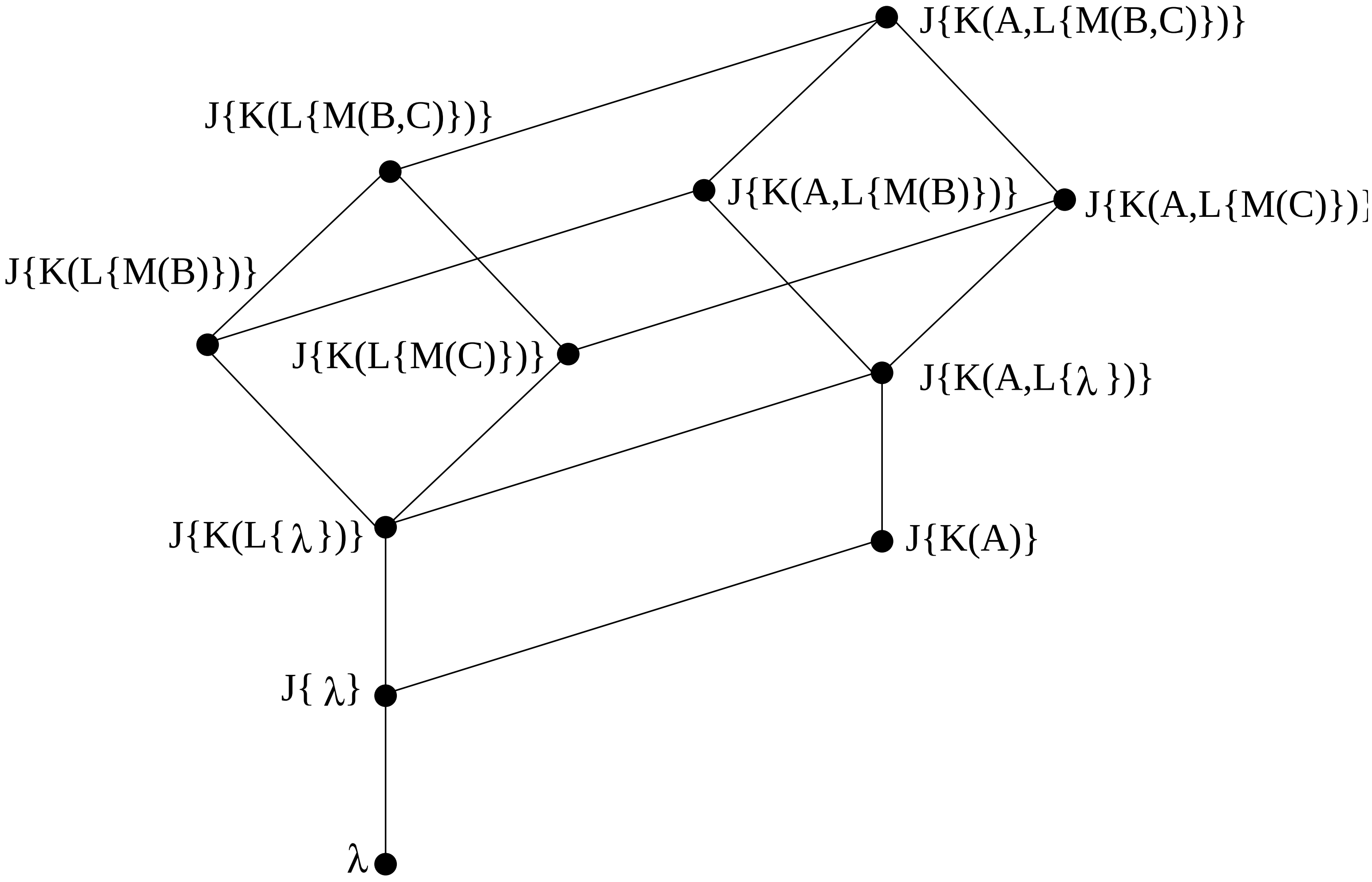
    $$Y \circ_X Z = L\{C \circ_B D\} \quad \text{for } \circ \in \{\sqcup, \sqcap\}$$

    and if $Z \not\leq Y$, then $Z \dot{-} Y = L\{D \dot{-}_B C\}$
  - $N = L(A_1, \ldots, A_n), Y = L(B_1, \ldots, B_n)$, $Z = L(C_1, \ldots, C_n)$:

    $$Y \circ_N Z = L(B_1 \circ_{A_1} C_1, \ldots, B_n \circ_{A_n} C_n) \quad \text{for } \circ \in \{\sqcup, \sqcap, \dot{-}\}$$

- $(Sub(N), \leq, \sqcup_N, \sqcap_N, \dot{-}_N, N)$ is a **Brouwerian Algebra**
- define **Brouwerian Complement** $Y_N^{\mathcal{C}} = N \dot{-}_N Y$
- $(Sub(N), \leq, \sqcup_N, \sqcap_N, (\cdot)_N^{\mathcal{C}}, \lambda, N)$ is not **boolean**

3.2 Algebraic Foundations: An Example



4.1 Axiomatisation: Functional Dependencies

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

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

- finite $r \subseteq Dom(N)$ **satisfies** $\mathcal{X} \to \mathcal{Y}$ on N ($\models_r \mathcal{X} \to \mathcal{Y}$) 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 $\models_r \tau$ whenever $\models_r \sigma$ for all $\sigma \in \Sigma$
- **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 $\Sigma^* \subseteq \Sigma^+$

4.2 Axiomatisation: Semi-Disjoint Attributes

- $N = \text{Soccer}\{\text{Match}(\text{Winner}, \text{Loser})\}$

- $r = \{t_1, t_2\} \subseteq Dom(N)$ with

$$t_1 = \{(\text{Germany, Chile}), (\text{Brasil, Argentina})\} \text{ and}$$
$$t_2 = \{(\text{Germany, Argentina}), (\text{Brasil, Chile})\}$$

- $\models_r \text{Soccer}\{\text{Match}(\text{Winner})\} \rightarrow \text{Soccer}\{\text{Match}(\text{Loser})\}$

- $\not\models_r \text{Soccer}\{\text{Match}(\text{Winner})\} \rightarrow \text{Soccer}\{\text{Match}(\text{Winner}, \text{Loser})\}$

- values on subattributes X and Y do not determine values on $X \sqcup Y$

- call $X, Y \in Sub(N)$ **semi-disjoint** iff there are subattributes $X' \leq X$ and $Y' \leq Y$ with $X' \sqcap Y' = \lambda$ and $X \sqcup Y = X' \sqcup Y'$

4.3 Axiomatisation: Completeness

- The **generalized Armstrong Axioms for FDs**

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

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

  form a sound and complete set of inference rules.

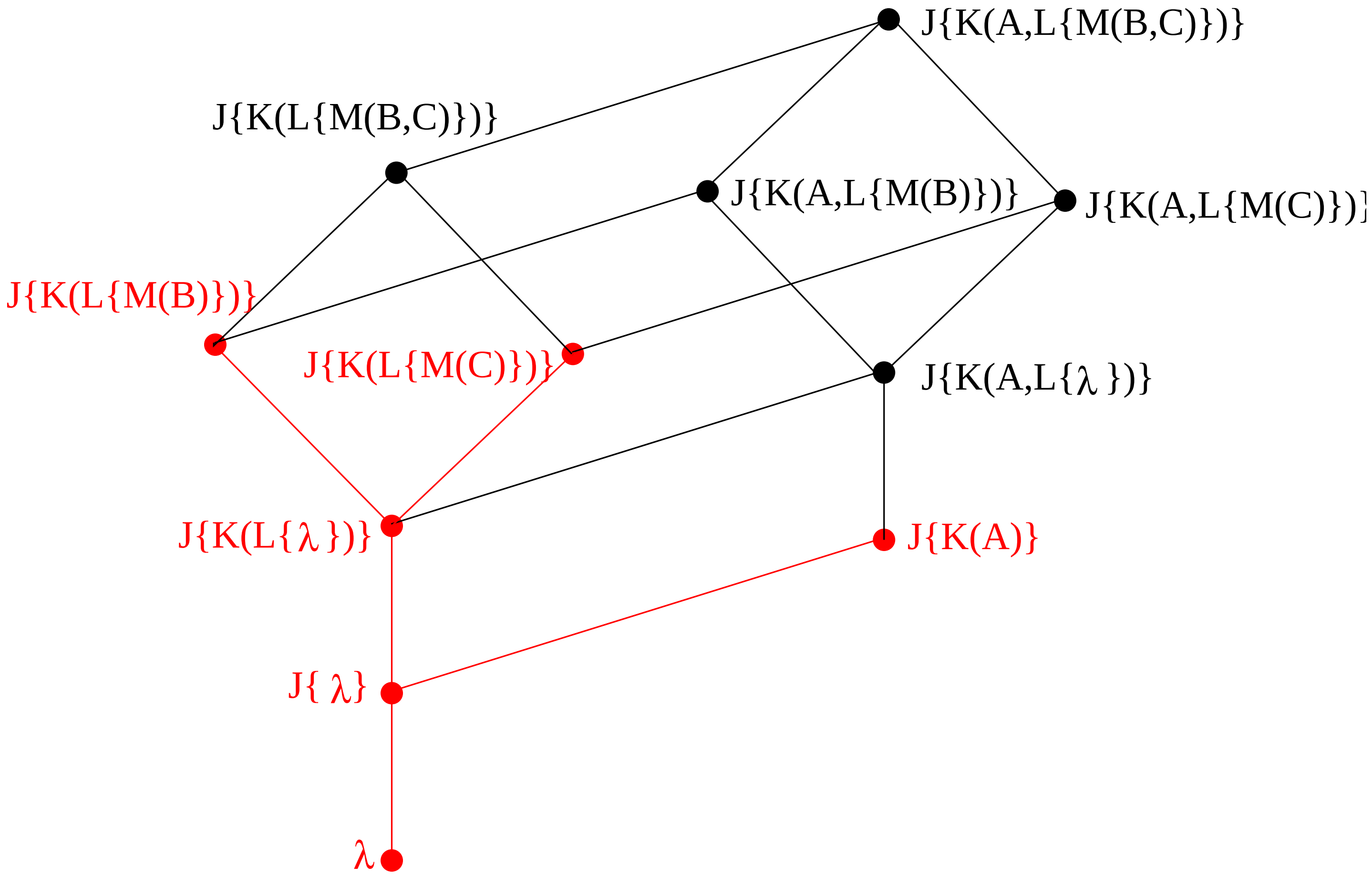
- Sketch of the Completeness Proof:
  - Σ set of fds on some N, take $\mathcal{X} \to \mathcal{Y} \notin \Sigma^+$
  - $\mathcal{X}^+ = \{Z \mid \mathcal{X} \to \{Z\} \in \Sigma^+\}$ ideal
  - $X, Y \in \mathcal{X}^+$ and $X \sqcap_N Y = \lambda$ implies $X \sqcup_N Y \in \mathcal{X}^+$
  - define $r = \{t_1, t_2\} \subseteq Dom(N)$ by: $\pi_W^N(t_1) = \pi_W^N(t_2)$ iff $W \in \mathcal{X}^+$
  - $\not\models_r \mathcal{X} \to \mathcal{Y}$, $\models_r \Sigma$ implies $\mathcal{X} \to \mathcal{Y} \notin \Sigma^*$

4.4 Axiomatisation: A combinatorial Lemma

- $N \in \mathcal{N}\mathrm{A}$, $\emptyset \neq \mathcal{I} \subseteq Sub(N)$ $\leq$-ideal, $X \sqcap_N Y = \lambda$ implies $X \sqcup_N Y \in \mathcal{I}$: $\exists t_N, t'_N \in Dom(N)$ with $\pi^N_X(t_N) = \pi^N_X(t'_N)$ iff $X \in \mathcal{I}$
- $\mathcal{I} = \{\lambda\}$, $\mathcal{I} = \{\lambda, A\}$ for $N \in \mathcal{U}$
- $N = L(N_1, \dots, N_m)$: $\mathcal{I}_j = \{X \sqcap_N L(N_j) : X \in \mathcal{I}\} \subseteq Sub(L(N_i))$
- $N = L\{M\}$: $\mathcal{I} = \{L\{X\} : X \in \mathcal{J}\} \cup \{\lambda\}$ for ideal $\mathcal{J} \subseteq Sub(M)$
- define $t_N = \{\tau_M(X) : X \leq M\}$ and $t'_N = \{\tau_M(X) : X \in \mathcal{J}\}$ by:
  - $\tau_\lambda(\lambda) = ok$,
  - $\tau_A(\lambda) = a'$, $\tau_A(A) = a$ with $a, a' \in Dom(A), a \neq a'$,
  - $\tau_{L(N_1,\dots,N_k)}(L(M_1, \dots, M_k)) = (\tau_{N_1}(M_1), \dots, \tau_{N_k}(M_k))$,
  - $\tau_{L\{N\}}(L\{M\}) = \begin{cases} \{\tau_N(M)\} & : M \neq \lambda \\ \emptyset & : M = \lambda \end{cases}$, $\tau_{L\{N\}}(\lambda) = \{\tau_N(\lambda)\}$.
- $\pi^N_\lambda(t_N) = ok = \pi^N_\lambda(t'_N)$
- $\{\pi^M_V(\tau_M(X)) : X \leq M\} = \{\pi^M_V(\tau_M(X)) : X \in \mathcal{J}\}$ iff $V \in \mathcal{J}$

4.5 Axiomatisation: An Example I

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



- $J\{K(A)\} \rightarrow J\{K(L\{M(B)\})\}$, $J\{K(A)\} \rightarrow J\{K(L\{M(C)\})\}$

4.6 Axiomatisation: An Example II

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

$$\{(a,\{b,c\}),(a,\emptyset),(a,\{(b',c)\}),(a,\{(b,c')\}),(a,\{(b',c')\}),(a',\{b,c\}),(a',\emptyset),(a',\{(b',c)\}),(a',\{(b,c')\}),(a',\{(b',c')\})\}.$$

- $t'_N = \{\tau_{K(A,L\{M(B,C)\})}(Y) : Y \in \mathcal{X}^+\}$ is

$$\{(a,\{b,c\}),(a,\emptyset),(a,\{(b',c)\}),(a,\{(b,c')\}),(a',\{b,c\})\}$$

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

W	$\pi^N_W(t_N)$	$\pi^N_W(t'_N)$
λ	ok	ok
$J\{\lambda\}$	$\{ok\}$	$\{ok\}$
$J\{K(L\{\lambda\})\}$	$\{\emptyset,\{ok\}\}$	$\{\emptyset,\{ok\}\}$
$J\{K(L\{M(B)\})\}$	$\{\{b\},\{b'\},\emptyset\}$	$\{\{b\},\{b'\},\emptyset\}$
$J\{K(L\{M(C)\})\}$	$\{\{c\},\{c'\},\emptyset\}$	$\{\{c\},\{c'\},\emptyset\}$
$J\{K(L\{M(B,C)\})\}$	$\{\{(b,c)\},\{(b',c)\},\{(b,c')\},\{(b',c')\},\emptyset\}$	$\{\{(b,c)\},\{(b',c)\},\{(b,c')\},\emptyset\}$
$J\{K(A)\}$	$\{a,a'\}$	$\{a,a'\}$
$J\{K(A,L\{\lambda\})\}$	$\{(a,\{ok\}),(a,\emptyset),(a',\{ok\}),(a',\emptyset)\}$	$\{(a,\{ok\}),(a,\emptyset),(a',\{ok\})\}$
$J\{K(A,L\{M(B)\})\}$	$\{(a,\{b\}),(a,\{b'\}),(a,\emptyset),(a',\{b\}),(a',\{b'\}),(a',\emptyset)\}$	$\{(a,\{b\}),(a,\{b'\}),(a,\emptyset),(a',\{b\})\}$
$J\{K(A,L\{M(C)\})\}$	$\{(a,\{c\}),(a,\{c'\}),(a,\emptyset),(a',\{c\}),(a',\{c'\}),(a',\emptyset)\}$	$\{(a,\{c\}),(a,\{c'\}),(a,\emptyset),(a',\{c\})\}$
N	t_N	t'_N

4.7 Axiomatisation: An Application

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

$$\frac{}{\{X, Y\} \to \{X \sqcup_N Y\}}$$

- sufficiency is not hard to see
- show X, Y not semi-disjoint implies existence of some $r \subseteq Dom(N)$ with $\not\models_r \{X, Y\} \to \{X \sqcup_N Y\}$
- by 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 ideal such that $S \sqcap_N T = \lambda$ implies $S \sqcup_N T \in \mathcal{I}$ for all $S, T \in \mathcal{I}$

5.1 Extensions: More Results

- **Finite Implication Problem**: $\Sigma \models \mathcal{X} \rightarrow \mathcal{Y}$ with FDs on N
- decidable in $\mathcal{O}(n^3 \cdot s \cdot \min\{s, n\})$ where $n = |\ Sub(N)\ |$ and $s = |\ \Sigma\ |$
- generalisation of BCNF results in sufficient conditions for nested attributes to be non-redundant and avoid abnormal update behaviour
- weaker normal form **HLNF** proposed which exactly captures non-redundance and absence of abnormal update behaviour
- **MVDs and FDs** based on record and list types have been studied regarding axiomatisation and finite implication problem
- **XML**: several classes of FDs, Axiomatisations

5.2 Extensions: Future Research

- richer type system:

$$b \mid (a_1 : t_1, \ldots, a_n : t_n) \mid \{t\} \mid [t] \mid \langle t \rangle \mid (a_1 : t_1) \oplus \cdots \oplus (a_n : t_n)$$

- rational tree structures:

$$t \ := \ \ell \mid b \mid (a_1 : t_1, \ldots, a_n : t_n) \mid \{t\} \mid \ell : t$$

- different classes of dependencies: **Join dependencies**

- **XML**: FIP and Normalisation