Functional Dependencies on Nested Attributes: Algebraic, Logical and Topological Perspective

Sven Hartmann, Sebastian Link

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

Outline

1. **Background and Data Models**
2. **The Algebra of Nested Attributes**
3. **Algebraic Perspective**
4. **Logical Perspective**
5. **Topological Perspective**
6. **Conclusion and Future Work**

1.1 Functional Dependencies in the RDM

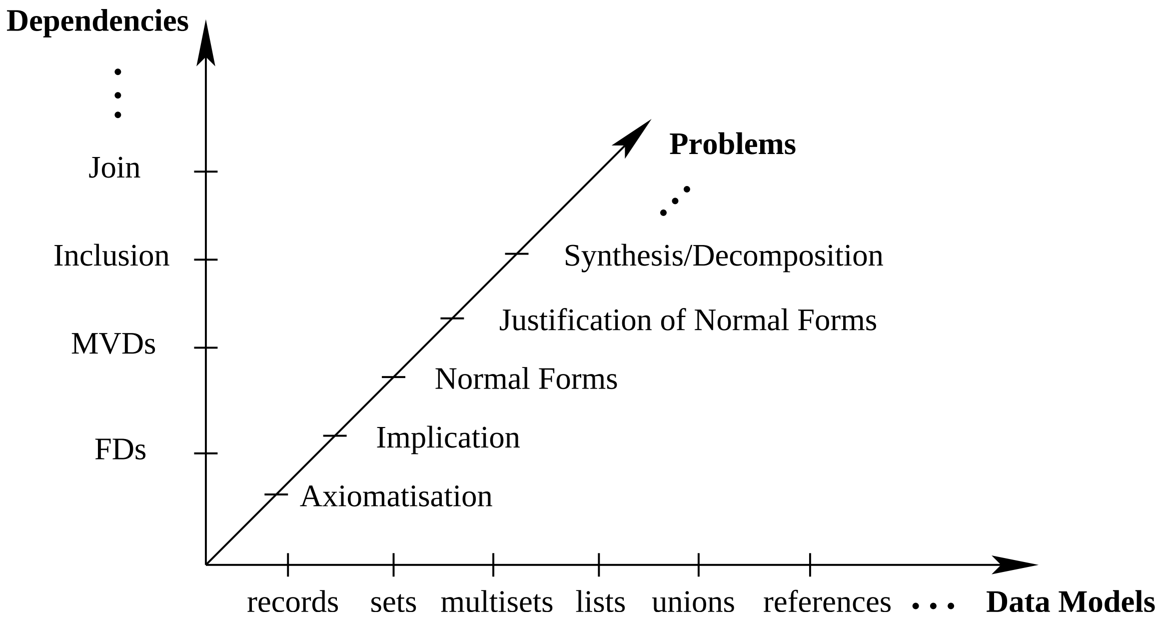
- FDs introduced in context of RDM by E.F. Codd in 1972
  - expression $X \rightarrow Y$ with $X, Y \subseteq R$
  - $\models_r X \rightarrow Y$ iff $t_1[Y] = t_2[Y]$, if $t_1[X] = t_2[X]$ for any $t_1, t_2 \in r$
- gain complete knowledge about consequences of semantics specified
  - **Boolean Algebra** $(\mathcal{P}(R), \subseteq, \cup, \cap, -, \emptyset, R)$ on R
  - **Armstrong Axioms** (1974)

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

- Fagin 1977: implication is equivalent to that of Horn clauses
- impact:
  - implication problem, equiv of sets of FDs, minimal covers
  - normal forms, redundancies, update anomalies, integrity checking

1.2 Advanced Data Models

- ER, UML, HERM, Nested RDM, Object-oriented/relational, XML
- extend achievements to complex objects in unified framework
- classify data models by the **types** they support



2.1 Database Schemata: Nested Attributes

- capture characteristics of objects in target database by attributes

$$N := A \mid \lambda \mid L(N_1, \dots, N_k) \mid L[N] \mid L\{N\} \mid L\langle N \rangle$$

- examples:
  - Shop(Customer,Trolley⟨Item(Article,Price)⟩,Discount)
  - Soccer{Match(Winner,Loser)}
  - Enrolments[Student(ID,Name,History[Course(No,Name,Grade)])]

2.2 Database Instances: Domain Assignment

- extend *dom* from flat to nested attributes ($dom(\lambda)=\{\text{ok}\}$)
- examples for nested tuples:
  - Shop(Customer,Trolley⟨Item(Article,Price)⟩,Discount):

    (Homer, ⟨(Donut,1.5),(Donut,1.5),(Chocolate,2),(Chocolate,2)⟩, 0)
    (Bart, ⟨(Donut,2),(Donut,2),(Chocolate,1.5),(Chocolate,1.5)⟩, 1)

  - Soccer{Match(Winner,Loser)}:

    {(Denmark, Sweden), (New Zealand, Australia)}
    {(Mexico, USA), (Brazil, Argentina), (Brazil, USA)}

- RDM: single application of record constructor
- Nested Relational Data Model: record and set constructor
- Object-oriented Data Models: record, set, multiset and list constructor

2.3 Subschemata: Subattributes

- recursively replacing attributes by λ gives different layers of info:

- some *subattributes* of
  Shop(Customer,Trolley$\langle$Item(Article,Price)$\rangle$,Discount):
  - Shop(λ,Trolley$\langle$Item(Article,Price)$\rangle$,Discount)
  - Shop(Customer,Trolley$\langle$Item(λ,λ)$\rangle$,Discount)
  - Shop(λ,Trolley$\langle$Item(Article,λ)$\rangle$,λ)
  - Shop(Customer,λ,Discount)

- formally:
  define subattribute relation $\leq$ on nested attributes (partial order)

2.4 Database Transformations: Projection Function

- subattributes represent at most as much info as their superattributes
- formally: for $M \leq N$ there is projection $\pi^N_M : dom(N) \rightarrow dom(M)$
- N = Shop(Customer,Trolley⟨Item(Article,Price)⟩,Discount) with

$$t\text{=(Bart,}\langle\text{(Donut,2),(Donut,2),(Chocolate,1.5),(Chocolate,1.5)}\rangle\text{,1)}$$

- M = Shop(Customer,Trolley⟨Item(λ,Price)⟩,Discount)

$$\pi^N_M(t)\text{=(Bart,}\langle\text{(ok,2),(ok,2),(ok,1.5),(ok,1.5)}\rangle\text{,1)}$$

- M = Shop(λ,Trolley⟨Item(λ,λ)⟩,Discount)

$$\pi^N_M(t)\text{=(ok,}\langle\text{(ok,ok),(ok,ok),(ok,ok),(ok,ok)}\rangle\text{,1)}$$

2.5 The Brouwerian Algebra of Subattributes

- subattribute order $\leq$ induces operations $\sqcup_N$, $\sqcap_N$, and $\dot{-}_N$
- $(Sub(N) = \{M \mid M \leq N\}, \leq, \sqcup_N, \sqcap_N, \dot{-}_N, N)$ is *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

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

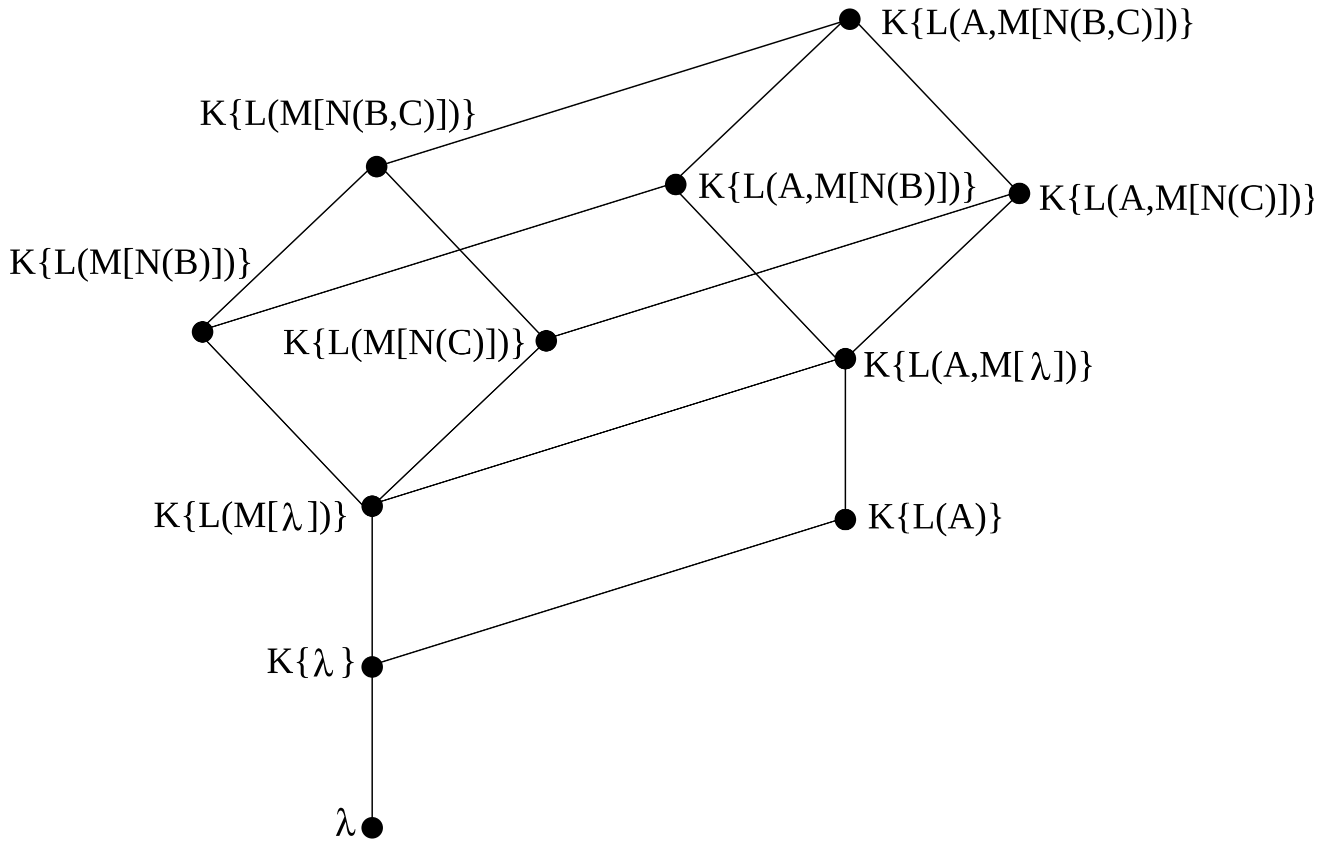
  for all $X \in Sub(N)$

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

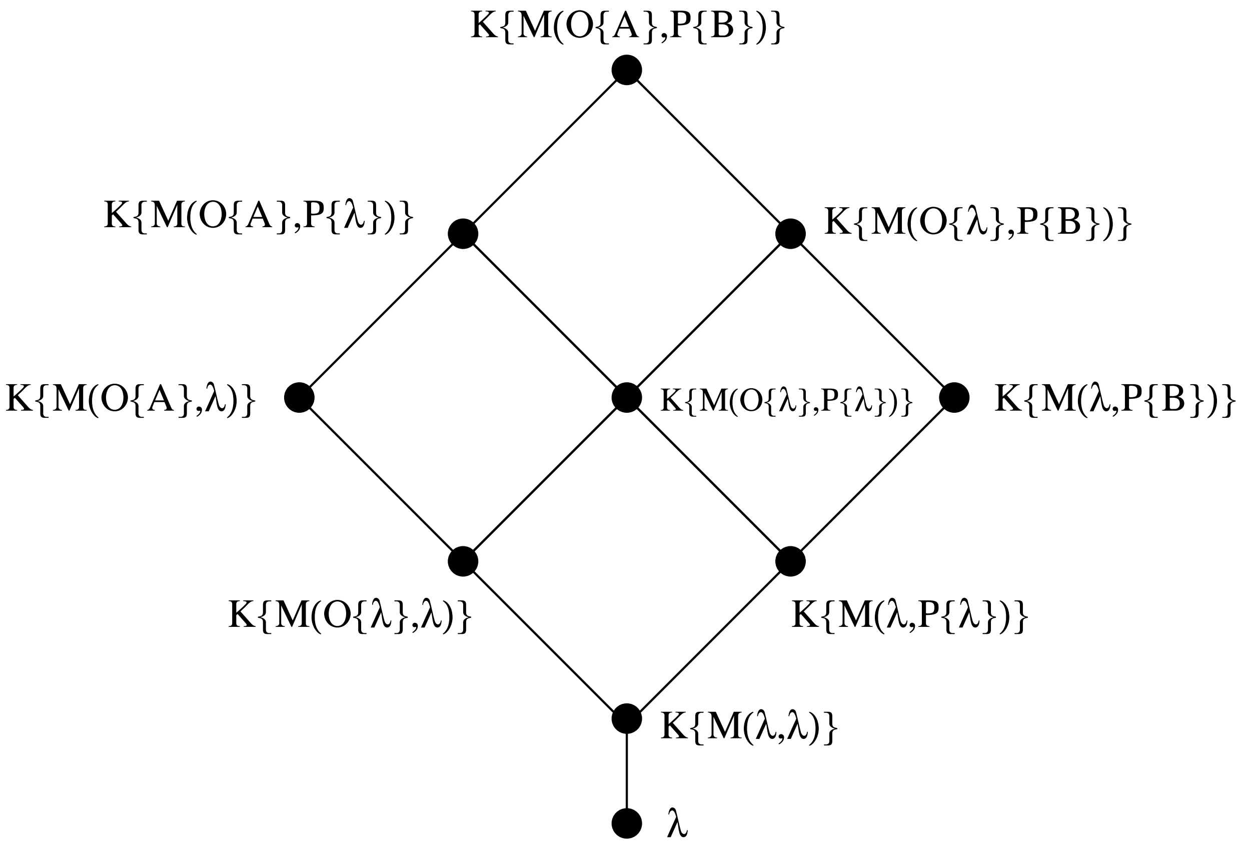
$$Y^{\mathcal{C}} \leq X \quad \text{if and only if} \quad X \sqcup Y = N$$

- $(Sub(N), \leq, \sqcup_N, \sqcap_N, (\cdot)_N^{\mathcal{C}}, \lambda_N, N)$ is not a **Boolean Algebra**

2.6 The Algebra of Nested Attributes: An Example



2.7 The Algebra of Nested Attributes: A further Example



3.1 Functional Dependencies

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

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

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

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

- Shop(λ,Trolley⟨Item(Article,Price)⟩,λ) $\rightarrow$ Shop(λ,λ,Discount)
- {Shop(λ,Trolley⟨Item(Article,λ)⟩,λ), Shop(λ,Trolley⟨Item(λ,Price)⟩,λ)} $\rightarrow$ Shop(λ,λ,Discount)
- implication: $\Sigma \models \tau$ iff $\models_r \tau$ if $\models_r \sigma$ for all $\sigma \in \Sigma$ and any (finite) r
- goal: find **sound** and **complete** $\mathfrak{R}$, i.e., $\Sigma^+_{\mathfrak{R}} \subseteq \Sigma^*$ and $\Sigma^* \subseteq \Sigma^+_{\mathfrak{R}}$

3.2 A fundamental Difference

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

$$t_1 = \{(\text{Denmark, Brazil}), (\text{Germany, Italy})\} \text{ and}$$
$$t_2 = \{(\text{Denmark, Italy}), (\text{Germany, Brazil})\}$$

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

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

3.4 Axiomatisation

Theorem 1. *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.,*

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

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

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

4.1 Join-Irreducibles

- one idea behind Fagin's Equivalence theorem: interpret attributes as propositional variables
- an element $a \in L$ of a lattice $(L, \sqsubseteq, \sqcup, \sqcap, 0)$ with bottom element 0 is called *join-irreducible* iff $a \neq 0$ and if $a = b \sqcup c$ holds for any $b, c \in L$, then $a = b$ or $a = c$
- let $\mathcal{B}(N)$ denote the join-irreducibles of $(Sub(N), \leq, \sqcup, \sqcap, \lambda_N)$
- for $N =$Soccer{Match(Winner,Loser)} we have Soccer{Match(λ,λ)}, Soccer{Match(Winner,λ)} and Soccer{Match(λ,Loser)} in $\mathcal{B}(N)$
- can't express Soccer{Match(Winner,λ)} $\rightarrow$ Soccer{Match(Winner,Loser)} since it is different from Soccer{Match(Winner,λ)} $\rightarrow$ {Soccer{Match(Winner,λ)},Soccer{Match(λ,Loser)}}

4.2 Extended Join-Irreducibles

- *extended join-irreducibles* form smallest $\mathcal{E}(N) \subseteq Sub(N)$ such that
  (i) $\mathcal{B}(N) \subseteq \mathcal{E}(N)$, and
  (ii) for all $X, Y \in \mathcal{E}(N)$ which are not reconcilable also $X \sqcup Y \in \mathcal{E}(N)$
- FDs are $\mathcal{X} \rightarrow \mathcal{Y}$ with $\leq$-antichains $\mathcal{X}, \mathcal{Y} \subseteq \mathcal{E}(N)$
- interpret extended join-irreducibles as variables via $\psi : \mathcal{E}(N) \rightarrow \mathcal{V}$
- $\sigma = \{X_1, \dots, X_n\} \rightarrow \{Y_1, \dots, Y_m\}$ gives set $\Phi(\sigma)$ of Horn clauses

$$\bigwedge_{i=1}^{k} \psi(X_i) \Rightarrow \psi(Y_1), \dots, \bigwedge_{i=1}^{k} \psi(X_i) \Rightarrow \psi(Y_m)$$

- Horn clauses can also encode the structure of N

$$\Pi_N = \{\psi(U) \Rightarrow \psi(V) \mid U, V \in \mathcal{E}(N), U \text{ covers } V\}$$

4.3 The Equivalence

Theorem 2. *Let N be a nested attribute, Σ a set of FDs and σ a single FD on N. Let Π_N denote the Horn clauses which encode the structure of N, and Π denote the corresponding set of Horn clauses for Σ. Then*

(i) Σ *implies* σ,
(ii) Σ *implies* σ *in the world of two-tuple instances, and*
(iii) $\Pi \cup \Pi_N$ *logically implies* π *for all* $\pi \in \Phi(\sigma)$

are equivalent. □

- this extends a well-known result by *Fagin et al.(1977)*, where
  - only single application of record constructor allowed,
  - join-irreducibles form anti-chain, and
  - join-irreducibles (attributes) suffice

4.4 A simple Example

- bijection ψ:

$$\begin{aligned} \text{Cup(Day},\lambda) &\leftrightarrow V_0, \\ \text{Cup}(\lambda,\text{Soccer}\{\text{Match}(\lambda,\lambda)\}) &\leftrightarrow V_1, \\ \text{Cup}(\lambda,\text{Soccer}\{\text{Match(Winner},\lambda)\}) &\leftrightarrow V_2, \\ \text{Cup}(\lambda,\text{Soccer}\{\text{Match}(\lambda,\text{Loser})\}) &\leftrightarrow V_3, \\ \text{Cup}(\lambda,\text{Soccer}\{\text{Match(Winner,Loser)}\}) &\leftrightarrow V_4 \end{aligned}$$

- $\text{Cup(Day},\lambda) \rightarrow \text{Cup}(\lambda,\text{Soccer}\{\text{Match(Winner,Loser)}\})$ implies $\text{Cup(Day},\lambda) \rightarrow \{\text{Cup}(\lambda,\text{Soccer}\{\text{Match(Winner},\lambda)\}, \text{Cup}(\lambda,\text{Soccer}\{\text{Match}(\lambda,\text{Loser})\})\}$

- equivalent to $\{V_0 \Rightarrow V_4, V_4 \Rightarrow V_3, V_4 \Rightarrow V_2, V_2 \Rightarrow V_1, V_3 \Rightarrow V_1\}$ implies $V_0 \Rightarrow V_2$ and $V_0 \Rightarrow V_3$

5.1 PO-Spaces

- remove structural rules, implementation details of implication problem
- *topological space* $\mathcal{T}$ is a structure $(S, \mathfrak{C})$ with set S and operation $\mathfrak{C}$ mapping subsets of S to subsets of S such that for all $A, B \subseteq S$:
  - $A \subseteq \mathfrak{C}A$,
  - $\mathfrak{C}A = \mathfrak{C}\mathfrak{C}A$,
  - $\mathfrak{C}(A \cup B) = \mathfrak{C}A \cup \mathfrak{C}B$, and
  - $\mathfrak{C}\emptyset = \emptyset$.
- subset A of S is *closed* iff $\mathfrak{C}A = A$
- poset $(S, \leq)$, for $A \subseteq S$: $\mathfrak{C}A = \{b \in S \mid b \leq a \text{ for some } a \in A\}$
- topological space $(S, \mathfrak{C})$ is called a *PO-space*

5.2 Units

- reduce notion of reconcilability to comparability wrt $\leq$
- $U \in Sub(N)$ is a *unit* of N iff U is $\leq$-maximal with

  $\forall X, Y \leq U$ if X and Y are reconcilable, then $X \leq Y$ or $Y \leq X$

- units of Cup(Day,Soccer{Match(Winner,Loser)}) are Cup(Day,λ) and Cup(λ,Soccer{Match(Winner,Loser)})
- $V, W \in Sub(N)$ reconcilable iff for all units U of N: $V \sqcap U$ and $W \sqcap U$ are $\leq$-comparable
- Cup(λ,Soccer{Match(Winner,λ)}) and Cup(λ,Soccer{Match(λ,Loser)}) are not comparable

5.3 Topological View on FDs

- $\mathcal{U}(N) = \{U_1, \dots, U_k\}$, FD on N is $\mathcal{X} \rightarrow \mathcal{Y}$ where
  - $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_k)$, $\mathcal{Y} = (\mathcal{Y}_1, \dots, \mathcal{Y}_k)$ and
  - $\mathcal{X}_i, \mathcal{Y}_i$ are closed sets of the PO-space on $(\mathcal{E}(U_i), \leq)$

- $\models_r \mathcal{X} \rightarrow \mathcal{Y}$ on N iff $\forall t_1, t_2 \in r$ we have $\forall i. \forall X \in \mathcal{X}_i . \pi_X^N(t_1) = \pi_X^N(t_2)$ implies $\forall i. \forall Y \in \mathcal{Y}_i . \pi_Y^N(t_1) = \pi_Y^N(t_2)$

- $\mathcal{Y} \subseteq \mathcal{X}$ iff $\forall i. \mathcal{Y}_i \subseteq \mathcal{X}_i$, and $\mathcal{X} \cup \mathcal{Y} = (\mathcal{X}_1 \cup \mathcal{Y}_1, \dots, \mathcal{X}_k \cup \mathcal{Y}_k)$

Theorem 3.

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

are minimal, sound and complete for FD-implication □

6 Conclusion and Future Work

- framework of nested attributes allows to capture data models by including corresponding type constructors
- theory of Brouwerian algebras can be used to extend many achievements from relational databases
- allows to study direct impact of type constructor on design problem without considering pecularities of specific data model
- study different classes of dependencies in different combinations of constructors
- increase expressiveness by studying embedded dependencies (allowing several Brouwerian algebras simultaneously)
- normal forms