Teoria si practica bazelor de date. Proprietati ale formelor normale imerse

Prof.dr. Nicolaie GIURGITEANU
Catedra de Informatica Economica, Universitatea din Craiova

In this paper we present some properties of the nested normal forms such as reduced redundancy, design flexibility and generalizing the 4NF and BCNF normal forms. Also we prove some properties of these nested normal forms.

1 Introducere

Prin adaugarea unui comportament orientat obiect la o baza de date relationala se obtine o baza de dare relationala orientata obiect (*Object Relational Databases*). Deoarece aceasta extensie este aproape naturala exista deja o serie de produse software comerciale care suporta comportamentul orientat obiect.
Un comportament distinct al unui produs pentru baze de date relationale orientate obiect (ORDB) este acela ca o relatie poate fi scufundata într-o alta relatie, adica este o relatie imersa. Prin urmare proiectantii de baze de date trebuie sa fie capabili sa realizeze baze de date imerse cu proprietati impuse. Pentru bazele de date relationale obisnuite exista diverse forme normale care reprezinta adevarate standarde pentru proiectarea acestor tipuri de baze de date. În acelasi spirit s-au propus diverse forme normale si pentru bazele de date imerse si în acest sens citam formele normale imerse NNFmr, NNFmk, NNFed, NNFmd. Una din formele normale deosebit de importante este cea definita în [04], este vorba de forma normala prin partitii (*Partition Normal Form* - PNF). O relatie imersa satisface proprietatile unei forme normale prin partitii daca, atât ea cât si orice o alta relatie imersa inclusa în ea, nu contine tupluri cu valori identice pe atribute. Adica, oricare ar fi doua tupluri t_1 si t_2 apartinând relatiei imerse, întotdeauna vom avea $t_1[a]\neq t_2[a]$, oricare ar fi atributul a. Toate formele normale imerse enumerate mai sus sunt în acelasi timp si forme normale prin partitii, deoarece proprietatea PNF este una fundamentala pentru relatiile imerse.
Cunoasterea fortei si slabiciunilor formelor normale este una din conditiile esentiale ale proiectarii unei baze de date cu anumite proprietati impuse. Este cunoscut faptul ca formele 4NF sunt capabile sa elimine redundantele provocate de dependentele functionale, dar, în acelasi timp, ele nu pastreaza toate dependentele functionale. De asemenea se cunoaste faptul ca formele normale 3NF sunt conservative de jonctiune, pastreaza dependentele functionale dar nu elimina toate redundantele.
În cele ce urmeaza vom trata câteva din proprietatile formelor normale imerse si anume vom arata ca ele elimina redundantele si prezinta un mare grad de flexibilitate în proiectare si în acelasi timp sunt sub forma 4NF sau BCNF.
Dupa cum se stie, [02], schemele relationale obisnuite (ordinare) sunt scheme imerse si de aceea vom arata ca toate formele normale imerse enumerate mai sus implica forma normala 4NF, în raport cu o multime de multidependente si dependente functionale. Mai mult, daca pe schema relationala ordinara data sunt satisfacute doar dependente functionale, atunci formele normale imerse implica forma normala BCNF.

2. Generalizarea formelor 4NF si BCNF

2.1. Leme si definitii speciale

Pentru început avem nevoie de câteva definitii si leme.

Definitia 2.1. Fie U o multime de atribute, M si F, câte o multime de multidependente functionale si, respectiv, de dependente functionale peste U. O schema relationala R peste U ($R\subseteq U$) este în forma normala 4NF Fagin, în raport cu $M\cup F$ daca ori de câte ori are loc o multidependenta functionala netriviala $X\rightarrow\rightarrow Y$, în raport cu $M\cup F$, are loc si dependenta functionala $X\rightarrow Y$ în raport cu $M\cup F$.

Definitia 2.2. Fie U o multime de atribute si F o multime de dependente functionale peste U. O schema relationala R peste U ($R \subseteq U$) este în forma normala BCNF, în raport cu F, daca orice dependenta functionala netriviala este de forma $X \to R$.

De fapt, aceste doua definitii stabilesc conditiile ce definesc formele normale 4NF si BCNF [01]. Acum putem enunta si demonstra urmatoarele leme.

Lema 2.1. Fie M o multime de multidependente functionale peste schema relationala U. Daca $X \to\to W$ din M^+ este netriviala, atunci exista X', $X' \subseteq X$, astfel încât $X' \in LHS(M)$.

Demonstratie. Deoarece $X \to\to W$ nu este triviala deducem ca $W \not\subset X$ si, în plus XW?U. Fie X' o submultime a lui X. Sa demonstram ca X' este în LHS(M). Folosind regulile de derivare cu multidependentele functionale facem urmatorul rationament: $X' \to\to X'$ iar din MD5 deducem ca $X' \to\to X\text{-}X'$. Folosind regula de multireuniune deducem ca $X' \to\to X$. Deci X' este în LHS(M).

Lema 2.2. Fie M o multime de multidependente functionale si Z o cheie a sa [02]. Pentru orice $W \in DEP(Z)$, multidependenta functionala $Z \to\to W$ este netriviala si netransferabila.

Demonstratie. Faptul ca multidependenta este netriviala este evident. Sa aratam ca ea este si netransferabila. Presupunem prin absurd ca este transferabila, deci $\exists\ Z' \subset Z$ astfel încât $Z' \to\to (Z\text{-}Z')W$, prin urmare $Z' \to\to ZW$. De aici deducem ca pentru orice $V \in DEP(Z)$, cu $V \neq W$, $Z' \to\to V$ si prin urmare Z nu are dependente reduse. Acest lucru contrazice Lema 2.1.

Lema 2.3. Fie M o multime de multidependente functionale si Z o cheie a sa si fie $X \subset Z$. În aceste conditii exista $V \in DEP(X)$ astfel încât:

a) $Z \subset XV$, (adica X nu divide pe Z [02]);

b) pentru orice $W \in RDEP(X)$, $W \subset V$, unde RDEP(X) este multimea dependentelor fundamentale pe X, reduse [GN02];

c) multidependenta functionala $X \to\to V$ este redusa la stânga, iar daca X este si cheie atunci $V \in RDEP(X)$;

d) Z divide pe V;

e) X nu divide reuniunea tuturor dependentelor reduse ale lui Z.

Demonstratie. (a) Presupunem prin absurd ca X divide pe Z. În aceste conditii exista $V \in DEP(X)$ astfel încât $V_1 = V \cap Z \neq \Phi$ si $V_2 = C_v \cap Z \neq \Phi$, unde $C_v = U\text{-}XV$. Din multidependenta functionala $X \to\to V$, pe baza regulilor de derivare, deducem ca $Z \to\to V$ si prin urmare, pentru orice $W \in DEP(Z)$, fie $W \subseteq V$, fie $W \subseteq C_v$. Daca $W \subseteq V$, atunci, tinând seama de faptul ca $XV_1 \to\to XV_1C_v$ si ca $Z \subset XV_1C_v$, deducem ca $XV_1 \to\to W$, adica $Z \to\to W$ este redusa la stânga [02]. Daca $W \subseteq C_v$, deducem în mod asemanator ca mai sus ca $Z \to\to W$ este redusa la stânga. Prin urmare pentru orice $W \in DEP(Z)$, $Z \to\to W$ este reductibila la stânga. Dar aceasta este în contradictie cu ipoteza noastra ca Z este o cheie a lui M. Deci punctul a) este demonstrat.

(b) Deoarece $X \subset Z$, deducem ca X nu divide elemente ale lui DEP(Z). Adica, pentru orice $W \in DEP(Z)$ fie $W \cap V = \Phi$, fie $W \subseteq V$. Dar, daca $W \cap V = \Phi$, atunci $X \to\to W$ este reductibila la stânga si deci si $Z \to\to W$ este reductibila la stânga. Acest lucru vine în contradictie cu ipoteza noastra ca Z este o cheie a lui M.

(c) Daca $X \to\to V$ nu este redusa la stânga, atunci exista submultimi, X', nevide ale lui X, astfel încât $X' \to\to V$. Prin urmare X' divide pe Z ceea ce contrazice punctul (a).

(d) Deoarece $V \in DEP(X)$ si $X \subset Z$, deducem ca $Z \to\to V$. Daca Z nu divide pe V, atunci V-Z este în DEP(Z). Tinând acum seama de faptul ca $Z \to\to V$ si ca $V = (V\text{-}Z)\ (V \cap Z)$ deducem ca $Z \to\to (V\text{-}Z)$ este transferabila, ceea ce contrazice Lema 2.2.

(e) Se obtine direct din punctul (b).

Lema 2.4. Fie U o multime de atribute, M o multime de multidependente functionale peste U. Fie de asemenea Z o cheie a lui M, Z este în forma normala 4NF în raport cu M.

Demonstratie. Fie $X \subset Z$. Conform lemei 2.3 exista un $V \in DEP(X)$ astfel încât $Z \subset XV$. Deci nici o multidependenta $X \to\to W$, cu $W \in DEP(X)$ nu divide pe Z si prin urmare Z este în forma normala 4NF în raport cu M.

Lema 2.5. Fie U o multime de atribute si D multimea dependentelor si multidependentelor functionale pe U. Fie de asemenea E(D) înfasuratoarea lui D. Daca schema relationala R⊆U este în forma normala 4NF în raport cu E(D) atunci R este în forma 4NF si în raport cu D.

Demonstratie. Presupunem prin absurd ca R nu este în forma normala 4NF în raport cu D. Atunci exista o multidependenta functionala netriviala X→→Y∈D, care are proprietatea ca X? R. Din definitia lui E(D) deducem ca exista V∈DEP(X) cu X→→V, X? V. Prin urmare X? R si deci am ajuns la o contradictie cu faptul ca R este în forma normala 4NF în raport cu E(D).

2.2. Echivalenta formei normale imerse NNFMR cu forma normala 4NF si generalizarea formei normale BCNF

Teorema 2.1. Fie U o multime de atribute, M o multime de multidependente functionale peste U si F o multime de dependente functionale peste U. Fie de asemenea T un arbore, astfel încât *Aset(T)*⊆U si T sa fie format dintr-un singur nod (radacina). Arborele schema T este în forma normala 4NF în raport cu M ∪ F daca si numai daca el este în forma normala imersa NNFmr în raport cu M ∪ F.

Demonstratie. Fie D multimea dependentelor si multidependentelor functionale care au loc pe T în raport cu M ∪ F. Sa observam ca daca T este sub forma SNG [02], atunci pe el nu exista multidependente functionale, deci MVD(T) este vida.

Sa demonstram implicatia "⇐". Deoarece T este în forma normala 4NF în raport cu M ∪ F si M este vida deducem ca pe T nu pot avea loc decât dependente functionale, adica MVD(T) ∪ FD(T) se reduce la D si prin urmare el satisface primei conditii a formei NNFmr. Sa presupunem acum ca X → A este verificata pe T în raport cu M ∪ F. Deoarece T este în forma normala 4NF se verifica si dependenta X → T si, prin urmare se verifica si X → Parent(Na), unde Na este nodul care-l contine pe A.

În felul acesta am demonstrat ca T satisface si cea de a doua conditie a formei normale imerse NNFmr.

Sa demonstram acum implicatia directa "⇒". Deoarece T este în forma normala imersa NNFmr si MVD(T) este vida, dupa conditia 1, FD(T) se reduce la D. De aici deducem ca orice multidependenta functionala netriviala X →→ Y ce are loc pe T se obtine din dependenta netriviala X → Y. Prin urmare, din conditia 2 a formei normale imerse NNFmr, si tinând seama ca T este format numai dintr-un sigur nod, radacina, deducem ca X → T. Deci T este în forma normala 4NF.

Teorema 2.2. Fie U o multime de atribute si F o multime de dependente functionale pe U. Fie de asemenea T un arbore schema cu *Aset(T)* Í U. Arborele T este în forma normala BCNF daca si numai daca el este în forma normala imersa NNFmr.

Demonstratie. Deoarece avem doar dependente functionale, cele doua definitii de mai sus sunt echivalente si deci rezultatul se obtine imediat din Teorema 1.1.

2.3. Generalizarea formei normale 4NF de forma normala imersa NNFmk

Teorema 2.3. Fie U o multime de atribute si D o multime de multidependente si dependente functionale pe U. Fie de asemenea T un arbore schema de tip SNG cu *Aset(T)* Í U. Daca T este în forma normala imersa NNFmk în raport cu D, atunci el este si în forma normala 4NF în raport cu D.

Demonstratie. Deoarece T este un arbore SNG el consta dintr-un singur nod, radacina. Conform conditiei 4 a formei normale imerse NNFmk, radacina este o cheie a multimii multidependentelor functionale M = { X →→ Y ∈ D} ∪ {X →→ A | A ∈ Y}. Dupa lema 2.3 de mai sus, deducem ca radacina este în forma normal 4NF în raport cu M, iar din definitia lui M se deduce ca radacina este de asemenea în forma normal 4NF în raport cu D.

2.4. Generalizarea formei normale 4NF de forma normala imersa NNFmr

Teorema 2.4. Fie U o multime de atribute, D o multime de multidependente si dependente functionale peste *U*. Fie de asemenea *T* o schema formata dintr-un singur nod (schema SNG), astfel încât *Aset(T)* $\subseteq$ U. Daca *T* este în forma normala imersa NNFmr, în raport cu D atunci ele este si în forma normal 4NF, în raport cu D.

Demonstratie. Deoarece *T* este un arbore SNG el are un singur nod, anume radacina. Conform conditiei 4 a formei normale imerse NNFmr, radacina este o cheie a multimii M de multidependente functionale si deci, conform lemei 2.3, radacina lui *T* este în forma normal imersa NNFdf în raport cu M. Prin urmare, tinând seama si de definitia lui M, deducem ca radacina este în forma normala 4NF în raport cu D.

2.5. Generalizarea formei normale 4NF de forma normala imersa NNFed

Teorema 2.5. Fie *U* o multime de atribute si D o multime de multidependente si dependente functionale peste U. Fie *T* un arbore schema de tip SNG, astfel încât *Aset(T)* $\subseteq$ *U*. Daca *T* este NNFed în raport cu D atunci ea este si 4NF în raport cu D.

Demonstratie. Deoarece *T* este o schema SNG atunci ea este formata dintr-un singur nod, radacina.. Conform conditiei 4 din definitia formei normale imerse NNFed, radacina este cheie pentru E(D) si prin urmare, conform lemei 2, radacina este 4NF în raport cu E(D). Tinând seama acum si de lema 2.4, deducem ca radacina este în forma 4NF si în raport cu D.

2.6. Generalizarea formei normale 4NF de forma normala imersa NNFmd

Teorema 2.6. Fie *U* o multime de atribute si D o multime de multidependente si dependente functionale peste *U*. Fie *T* un arbore SNG astfel încât *Aset(T)* $\subseteq$ *U*. Raca *T* este în forma normala imersa NNFmd în raport cu D, atunci el este si în forma normala 4NF în raport cu D.

Demonstratie. Conform conditiei 4 a formei normale imerse NNFmd, radacina K este o cheie a lui $M^{/}$ si se gaseste în LHS($M^{/}$), unde multimea multidependentelor functionale $M^{/}$ este definita conform definitiei 10. K este înlocuita prin cheia candidata T. Deoarece K este o cheie pentru $M^{/}$, dupa lema 2, K este în forma 4NF în raport cu $M^{/}$. Vom arata ca T este în forma normala 4NF în raport cu D. Presupunem prin absurd ca T nu este în forma normal 4NF în raport cu D. Prin urmare, exista o multidependenta netriviala X $\to\to$ Y care are loc pe *T* în raport cu D astfel încât X ? *T*. Astfel, din X $\subseteq$ *T* si D, deducem ca exista o multidependenta functionala X $\to\to$ Z pe U cu Y = Z $\cap$ *T*. Vom demonstra acum ca X^{+} $\to\to$ W, unde W = Z $\cap$ K este o multidependenta functionala netriviala care are loc pe K în raport cu $M^{/}$ astfel încât X^{+} ? K. Prin urmare, K nu este 4NF în raport cu $M^{/}$, ceea ce este o contradictie. Sa aratam mai întâi ca X^{+} $\to\to$ W are loc pe K în raport cu $M^{/}$. Deoarece D implica multidependenta functionala X $\to\to$ Z, el va implica si multidependenta $X^{+}\to\to$ Z. Din definitia lui $M^{/}$, deducem ca X^{+} $\to\to$ Z este în $M^{/}$ si deoarece X$\subseteq$*T* si *T* $\subseteq$ K, deducem ca X $\subseteq$K. De aceea X^{+} $\subseteq$ K^{+}. Întru-cât K este în LHS($M^{/}$), deducem ca avem K = K^{+} si deci X^{+} $\subseteq$ K. Acum, din X^{+} $\subseteq$ K si X^{+} $\subseteq$ Z în $M^{/}$, deducem ca $X^{+}\to\to$ W are loc pe K în raport cu $M^{/}$. Sa aratam acum ca $X^{+}\to\to$ W este netriviala. Deoarece X $\to\to$ Y este netriviala si are loc pe *T*, fie Y, fie *T* - (XY), este nevida. Mai mult, X ? Y, altfel *T* nu este o cheie candidata. De aceea, exista un atribut A $\in$ Y astfel încât X ? A. Deoarece A $\in$ Y si Y = Z $\cap$ *T*, deducem ca A $\in$ Z si A $\in$ T. Deoarece T $\subseteq$K, mai deducem ca A $\in$ K si de aceea A $\in$ W. Acum, deoarece X ? A deducem ca A $\notin$ X^{+}. În mod similar se deduce ca exista un atribut B astfel încât B $\in$ T - (XY), X ? B si B $\in$ K - (X+W). Prin urmare X^{+} $\to\to$ W este netriviala pe K.

2.7. Concluzii asupra generalizarii formei normale 4NF prin forme normale imerse

Concluzia 2.1. Pentru teoremele 2.3, 2.4, 2.5 si 2.6 reciproca nu este adevarata, asa cum se poate constata din exemplul de mai jos.

Exemplu 2.1. Fie R = {A, B}, o schema relationala si fie D = {A $\to\to$ B} multimea dependentelor si multidependentelor functionale pe R. Se constata usor ca R este sub forma normala 4NF în raport cu D. Cu toate acestea, deoarece D contine doar multidependente triviale fiecare din multimile de multidependente functionale M, E(D) si $M^{/}$, asa cum sunt ele definite, contin doar dependente triviale. Prin urmare, conform definitiei 4.3 din [GN01], nici una din aceste multimi nu are o cheie. Deci R nu satisface conditiei 4 din definitia formelor normale NNFmk, NNFed, NNFmd.

Exemplul 2.2. Fie U=ABC, R={AB, AC} si fie M={ $A \rightarrow\rightarrow B$, $A \rightarrow\rightarrow C$}. De aici deducem ca M este libera de conflict iar R este descompunerea unica sub forma 4NF a lui U în raport cu M. Cu toate acestea nici AB si nici AC nu sunt sub forma normala NNFmk, sau NNFed, sau NNFmd. deoarece nici AB si nici AC nu sunt chei ale lui R în raport cu M (singura cheie fiind A). Pe de alta parte atât AB cât si AC nu sunt nici în forma 4NF si nici în forma NNFed. Sa observa însa ca schema relationala imersa A(B)*(C)*, care are doua drumuri AB si AC este în forma NNFmk NNFed, NNFmd. Cu toate acestea, aceasta schema relationala nu este în forma 4NF. Mai mult, conform Teoremei 3.1 de mai jos, A(B)*(C)* este si în forma NNFmr.

Concluzia 2.2. Toate formele normale imerse enumerate mai sus, implica forma normala BCNF, daca se dau numai dependente functionale. Acest lucru se obtine imediat din teoremele 2.3, 2.4, 2.5, 2.6.

Exemplul 2.1 pare a fi unul trivial. Cu toate acestea implicatiile lui sunt diverse si sunt foarte importante. În primul rând, D este echivalent cu multimea vida a multidependentelor functionale, care este libera de conflicte. De aceea, NNFmk, NNFed si NNFmd, cel putin în forma lor obisnuita, prezinta dificultati în generarea schemelor relationale imerse în raport cu multimile libere de conflict ale multidependentelor functionale. În al doilea rând, relatiile maimulte la mai multe între doua atribute, precum cele dintre A si B din exemplul 2.1, apar frecvent în practica. Pentru formele normale este important ca ele sa fie bine definite chiar pentru situatiile mai simple.

3. Reducerea redundantelor

3.1. Multimi de multidependente libere de conflict

Pentru început vom arata ca daca o un arbore schema *T* este în forma normal NNFmk în raport cu o multime de multidependente libera de conflict, atunci el este si în forma normala imersa NNFmr. Reciproca acestui rezultat nu este adevarata. Mai mult, daca un arbore *T* este consistent în raport cu o multime M de multidependente functionale libera de conflict, si daca fiecare drum al lui *T* este în forma normala 4NF în raport cu M, atunci structura imersa a lui *T* este fara valori redundante. Daca nu exista dependente functionale cele trei forme normale NNFmk, NNFed si NNFmd sunt echivalente. De aceea, în acest caz, rezultatele de mai sus se pastreaza si pentru formele normale imerse NNFed si NNFmd. Pentru moment vom considera ca nu avem dependente functionale.

Lema 3.1. Fie *U* o multime de atribute si fie D o multime de multidependente functionale, libera de conflict, peste *U*. Fie *T* un arbore schema astfel încât $Aset(T) \subseteq U$. Daca *T* este sub forma NNFmk în raport cu D, atunci orice drum din *T* este sub forma 4NF în raport cu D.

Demonstratie. Din faptul ca multimea dependentelor pe *U* este vida si din faptul ca T este în forma normala imersa NNFmk, deducem ca D este echivalenta cu multimea M din definitia formei normale imerse NNFmk. Dar, M implica MVD(T), tot din definitia lui NNFmk. Cum MVD(*T*) este formata numai din dependente triviale deducem ca si M este formata numai din dependente triviale. Deci, pe orice drum din *T* sunt valabile numai multidependentele triviale si ca atare, orice drum este sub forma 4NF, conform definitiei acestei forme [01].

Lema 3.2. Fie U o multime de atribute si fie de asemenea, *T* un arbore schema astfel încât $Aset(T) \subseteq$ U. Daca MVD(*T*) nu implica o multidependenta $X \rightarrow\rightarrow Y$ pe *Aset(T)*, atunci $X \rightarrow\rightarrow Y$ divide cel putin un drum din *T*.

Demonstratie. Vezi lema 5.1 [07].

Lema 3.3. Fie *U* o multime de atribute. Fie de asemenea M o multime de multidependente functionale, peste *U*, libera de conflict.

Fie *T* un arbore schema, consistent cu M, astfel încât *Aset(T)* ⊆ *U*. Fie D multimea multidependentelor implicate de M care au loc pe *T*. Daca toate drumurile din *T* sunt în forma normala 4NF în raport cu M, atunci MVD(*T*) determina D pe *Aset(T)*.

Demonstratie. Asa cum este mentionat si în [02], o multime de multidependente libera de conflict permite o descompunere unica sub forma 4NF [04]. Notând cu *Path(T)* multimea tuturor drumurilor din *T*, putem spune ca *Path(T)* este în forma normala 4NF în raport cu M, care este libera de conflict si *T* este consistent cu M si în plus, *Path(T)* este descompunerea unica sub forma 4NF a lui *Aset(T)* în raport cu M. Sa presupunem acum ca MVD(T) nu determina D. Prin urmare exista o multidependenta X →→ Y în D astfel încât ea nu se poate obtine din MVD(T). Coform lemei 10, X →→ Y divide cel putin un drum din *T* si deci, *Path(T)* nu este unica descompunere 4NF a lui *Aset(T)* în raport cu M. Deci am ajuns la o contradictie.

Sa vedem când un arbore schema este consistent cu o multime de restrictii. Fie D o multime de multidependente si dependente functionale peste o multime *U* de atribute si fie *T* un arbore schema astfel încât *Aset(T)* ⊆ *U*. Spunem despre arborele *T* ca este consistent cu D, daca pentru orice multidependenta functionala X →→ Y din MVD(T), D implica o multidepenta X →→ Z pe *U* care are proprietatea ca Y = Z ∩ *Aset(T)*.

Observatia 3.1. Un arbore schema este totdeauna consistent cu multime multidependentelor si dependentelor functionale data, pentru ca altfel schema sa implica o multidependenta care nu se obtine din dependentele si multidependentele date.

Teorema 3.1. Fie *U* o multime de atribute. Fie de asemenea M o multime de multidependente peste *U*, libera de conflict. Fie *T* un arbore schema astfel încât *Aset(T) U*. Daca *T* este în forma normala imersa NNFmk în raport cu M, atunci el este si în forma normala imersa NNFmr în raport cu M.

Demonstratie. Deoarece nu avem dependente functionale este suficient sa sa aratam ca *T* satisface prima conditie de a fi sub forma normala imersa NNFmr. Acum, deoarece *T* satisface prima conditie de a fi sub forma NNFmk, deducem ca T este consistent cu M. Prin urmare, multimea D definita în prima conditie a formei NNFmr implica MVD(T) pe *Aset(T)*. Conform lemei 9, Path(T) este sub forma 4NF în raport cu M si, deci, tinând seama si de lema 11, MVD(T) determina multimea D definita în conditia 1 a formei NNFmr pe *Aset(T)*.

Reciproca acestei teoreme nu este adevarata, asa cum o dovedeste exemplul 2.2 si Teorema 2.6.

Lema 3.3. Fie *U* o multime de atribute si M o multime de multidependente functionale peste *U*, libera de conflict. Fie de asemenea *T* un arbore schema astfel încât *Aset(T)* ⊆ *U*. Daca *T* este în forma NNFmr în raport cu M, atunci toate drumurile din *T* sunt sub forma 4NF în raport cu M.

Demonstratie. Presupunem prin absurd ca exista un drum P în *Path(T)* care nu este sub forma 4NF în raport cu M. Atunci, exista o multidependenta functionala netriviala X →→ Y pe P. Deoarece nu avem data si o multime de dependente functionale, deducem ca MVD(*T*) ∪ FD(T) nu implica X →→ Y, ceea ce vine în contradictie cu prima conditie a formei NNFmr.

Daca se da o multime de multidependente libera de conflict, atunci dupa lemele 3.1 si 3.2, daca *T* este un arbore schema care este sub forma NNFmk, atunci toate multidependentele de pe *T* sunt date de MVD(*T*). Astfel, structura imersa a lui *T* este capabila sa reduca redundantele. Tot din lemele 3.1 si 3.2 deducem ca si forma normala imersa NNFmr are un numar redus de redundante. Deoarece nu exista dependente functionale deducem ca NNFmk, NNFed si NNFmd sunt echivalente între ele si deci NNFed si NNFmd au si ele un numar redus de redundante.

Exemplul 3.1. Fie U = { Prof, Curs, Hobby, EchipamentHobby} si fie M = {Prof →→ Curs, Hobby →→EchipamentHobby} si fie relatia imersa din figura 3.1.

Prof	Curs	Hobby	EchipamentHobby
Barbu	CS209	Excursii	Sticla de apa
			Umbrela
Stere	CS415	Excursii	Sticla de apa
			Umbrela
	CS516	Dansuri	Costum
			Cizme

Fig. 3.1. Relatie imersa cu redundante

Sa observam ca M este libera de conflict. Sa consideram relatia din figura 3.1. Schema sa relationala violeaza conditiile de a fi sub una din formele normale imerse NNFmr, NNFmk, NNFed, NNFmd, deoarece unul din drumurile sale -Prof, Hobby, EchipamentHobby- nu este sub forma 4NF în raport cu M. Una din consecintele acestui lucru îl constituie prezenta redundantelor în relatia data. Asa cum se poate observa, valorile atributului EchipamentHobby, corespunzatoare valorii Excursii a atributului Hobby sunt memorate în doua locuri distincte. Dupa ce descompunem aceasta relatie imersa în cele doua subrelatii prezentate în figura 3.2, se constata ca a disparut orice redundanta. Sa mai observam ca orice drum din cele doua scheme relationale imerse din figura 3.2 este în forma normala 4NF si ambele scheme relationale imerse satisfac si conditiilor de a fi NNFmr, NNFmk, NNFed, NNFmd.

Prof	(Curs)*	(Hobby)*
Barbu	CS209	Excursii
Stere	CS415	Excursii
	CS516	Dansuri

Hobby	(EchipamentHobby)*
Excursii	Sticla de apa
	Umbrela
Dansuri	Costum
	Cizme

Fig. 3.2. O descompunere a relatiei imerse din figura 3.1

Încheiem prezentarile din aceasta sectiune concluzionând ca daca nu se dau dependente functionale, atunci toate formele normale imerse enumerate mai sus pot conduce la o eliminare dramatica a redundantelor în raport cu multimea libera de conflict a multidependentelor. Mai mult, daca nu se dau dependente functionale, atunci NNFmk implica NNFmr în raport cu multimea libera de conflict a multidependentelor. Aceleasi lucruri se pot spune si despre formele NNFed si NNFmd.

3.2. Reducerea redundantelor în cazul multimilor conflictuale

O multime de multidependente functionale care nu este libera de conflict vom spune ca este conflictuala. Pentru început vom enunta o proprietate importanta a formelor normale imerse NNFmr.

Teorema 3.2. Fie U o multime de atribute si fie D o multime de multidependente si dependente functionale peste U. Fie de asemenea un arbore schema T astfel încât $Aset(T) \subseteq U$. Daca T este consistent cu D, atunci T este în forma NNFmr în raport cu D daca si numai daca nici o relatie imersa R pe T, nu are redundante cauzate de multidependentele sau dependentele functionale implicate de D.

Demonstratie. Aceasta rezulta imediat din teoremele 5.1 si 5.2 din [07].

Asa cum se va observa din exemplul de mai jos formele normale imerse NNFmk, NNFed, NNFmd nu au aceasta proprietate.

Exemplul 3.2. Sa consideram o aplicatie de genealogie în care cineva este interesat de stramosii sai. Într-o astfel de aplicatie avem urmatoarele multidependente Persoana $\rightarrow\rightarrow$ Stramos si Persoana $\rightarrow\rightarrow$ Genealogist. Ultima multidependenta apare ca rezultat al faptului ca cineva, o persoana) este interesat sa afle informatii despre un anumit stramos al sau folosind mai multe servicii puse la dispozitie de câtiva genealogisti (adica Stramos si Genealogist sunt independente de Persoana). De asemenea pot sa apara si urmatoarele multidependente Stramos $\rightarrow\rightarrow$ Persoana si Stramos $\rightarrow\rightarrow$ Genealogist (Persoana si Genealogist sunt inde-

pendente de Stramos). Prin urmare multimea de multidependente data este {Persoana →→ Stramos, Persoana →→ Genealogist, Stramos →→ Persoana, Stramos →→ Genealogist }.
Sa consideram acum multimea total nescufundata din figura 3.3. Ea are o serie de redundante, de exemplu, ultima valoare Maria a atributului (Genealogist)* este redundanta deoarece ea se poate obtine prim multidependenta Stramos →→ Genealogist.

Persoana	(Stramos)*	(Genealogist)*
Stere	Ion Adam	Tudor Maria
Petre	Adam	Tudor
		Maria

Persoana	Stramos	Genealogist
Stere	Ion	Tudor
Stere	Adam	Tudor
Stere	Ion	Maria
Stere	Adam	Maria
Petre	Adam	Tudor
Petre	Adam	Maria

Fig. 3.3. O relatie imersa cu redundante cauzate de multidependente

Daca un arbore schema T nu îndeplineste conditiile 1 saui 2 ale formei normale NNFmr, atunci exista o relatie imersa pe T care are redundante [20]. Schema relationala din figura 3.3 nu îndeplineste conditiile 1 a formei normale NNFmr iar exemplul 3.3 a fost construit tocmai pentru a prezenta redundantele. Cu toate acestea, aceasta schema relationala imersa satisface conditiile formelor normale NNFmk, NNF ed si NNFmd. Tot din acest exemplu putem deduce ca formele normale NNFmk, NNFed si NNFmd admit relatii imerse cu redundante.
Relatia din figura 3.3 este nevoie sa fie descompusa în alte doua relatii imerse pentru a satisface conditiilor de a fi sub forma normala imersa NNFmr (figura 3.4).

Persoana	(Stramos)*
Stere	Ion
	Adam
Petre	Adam

Persoana	(Genealogist)*
Stere	Tudor
	Maria
Petre	Tudor
	Maria

Fig. 3.4. Descompunerea relatiei imerse din figura 3.3

Numarul de valori creste de la 9 la 11 si de aceea descompunerea relatiei imerse nu poate elimina valorile redundante. Pe de alta parte, toate formele normale imerse NNFmk, NNFed si NNFmd accepta schema relationala din figura 3.3 fara sa necesite ca relatia imersa sa fie descompusa. Sa mai notam si faptul ca multimea de multidependente este conflictuala deoarece ea nu îndeplineste Conditia 2 a definitiei multimilor libere de conflict.
Exemplul 3.3 se poate generaliza foarte usor. Sa presupunem ca M este o multime de multidependente netriviala libera de conflict. Fie *T* un arbore schema sub una din formele normale NNFmk, NNFed sau NNFmd în raport cu M. Conform teoremei 3.1, *T* este si în forma normala NNFmr în raport cu M. Deoarece M este o multime de multidependente netriviale, deducem ca arborele schema *T* are cel putin doua drumuri. Sa alegem doua muchii (V, W_1) si (V, W_2) în *T*, unde V este un parinte pentru vârfurile W_1 si W_2. Deoarece (V, W_2) este o muchie în *T* si acesta este în forma normala NNFmr în raport cu M, deducem ca M implica multidependenta Parent(V) →→ Z care are proprietatea ca, Child(W_2) =

proprietatea ca, Child(W_2) = *Aset(T)* $\cap$ Z. Acum vom adauga alte doua multidependente Child(W_1) $\to\to$ Parent(V) si Child(W_1) $\to\to$ Z. Dupa aceasta operatie M nu mai este libera de conflict deoarece Parent(V) $\to\to$ Z si Child(W_1) $\to\to$ Z încalca proprietatea intersectiei. Datorita noilor multidependente adaugate, *T* nu mai este în forma normala NNFmr. Cu toate acestea *T* ramâne în continuare sub forma normala NNFmk, NNFed sau NNFmd. Din aceste considerente formele normale imerse NNFmk, NNfed si NNFmd nu cer descompunerea arborelui schema *T* în timp ce forma normala imersa NNFmr necesita acest lucru, fara ca aceasta sa însemne si eliminarea redundantelor. provocate de cele doua multidependente incluse în plus.

În cele ce urmeaza vom argumenta dece descompunerea nu elimina redundantele atuci când mutimea de multidependente este conflictuala.

Daca o multime de multidependente este conflictuala atunci exista cel putin o descompunere sub forma normala 4NF. Prin urmare, chiar daca toate drumurile din *T* sunt sub forma normala 4NF, pot exista multidependente pe *T* care nu sunt obtinute pe baza multidependentelor din MVD(*T*). Deoarece structurile scufundate ale unei scheme relationale imerse nu pot decât sa puna în evidenta valorile redundante cauzate de MVD(*T*). multidependentele care provoaca redundantele sunt cele care nu sunt implicate de MVD(*T*). De aceea, chiar daca drumurile unui arbore schema sunt prelucrate pentru a satisface conditiile formei normale NNFmr, tot ramân valori redundante.

Urmatoarea teorema stabileste faptul ca daca multimea de multidependente date prezinta proprietatea intersectiei, atunci conditiile 1 si 2 ale formei normale NNFmk implica îndeplinirea conditiei 3 a acestei forme.

Teorema 3.3. Fie *U* o multime de atribute si fie M o multme de multidependente peste *U* care are si proprietatea intersectiei. Fie *T* un arbore schema astfel încât *Aset(T)* $\subseteq$ U. Daca sunt satisfacute conditiile 1 si 2 ale formei normale imerse NNFmk în raport cu M atunci *T* satisface de asemenea si conditia 3 a acestei forme normale în raport cu M.

Demonstratie. Presupunem ca *T* nu satisface conditia 3 a formei NNFmk. Presupunem deci ca exista o muchie (V, W) în *T* si ca exista o cheie X a lui M astfel încât exista Z $\in$ DEP_M(X) si Child(W) = Z $\cap$ *Aset(T)*. Presupunem de asemenea ca exista nodurile W_1, W_2, ... , W_n astfel încât Y = $\cup^n_{i=1}$ Child(W_i), X $\subseteq$ Parent(V)Y si, în plus, XY $\to\to$ Parent(V) nu are loc pe *T* în raport cu M. Parent(V) $\not\subset$ X pentru ca altfel XY $\to\to$ Parent(V) pe *Aset(T)*. Prin urmare X $\neq$ Parent(V). X $\not\subset$ Parent(V) altfel Parent(V)$\to\to$ Child(W) pe *Aset(T)* nu este redusa la stânga în raport cu M si deci *T* nu îndeplineste conditia 2 a formei normale NNFmk. De aceea X $\not\subset$ Parent(V). Deoarece *T* satisface conditia 1 a formei NNFmk, M va implica o multidependenta Parent(V) $\to\to Z'$ pe U astfel încât Child(W) = $Z' \cap$ *Aset(T)*. Deoarece nodurile lui *T* sunt disjuncte doua câte doua, X $\cap$ Z = Φ si Parent(V) $\cap$ Z' = Φ. Multidependenta X $\to\to$ Z implica X(Z - Z') $\to\to$ Z $\cap$ Z' si Parent(V) $\to\to$ Z' implica Parent(V)(Z' - Z) $\to\to$ Z $\cap$ Z'. Deoarece X $\cap$ Z = Φ si Parent(V) $\cap$ Z' = Φ deducem ca Z $\cap$ Z' este distinct fata de X(Z - Z') si fata de Parent(V)(Z' - Z). Conform proprietatii de intersectie a lui M avem ca X(Z - Z') $\cap$ Parent(V)(Z' - Z) $\to\to$ Z $\cap$ Z'. Totusi, X(Z - Z') $\cap$ Parent(V)(Z' - Z) = X $\cap$ Parent(V) si (Z $\cap$ Z') $\cap$ Aset(T) = Child(W). Deoarece Parent(V) $\not\subset$ X si X $\not\subset$ Parent(V), X $\cap$ Parent(V) este o submultime proprie a lui Parent(V). Prin urmare Parent(V) $\to\to$Child(W) pe *Aset(T)* nu este redusa la stânga în raport cu M, ceea ce este o contradictie.

4. Flexibilitate în proiectare

În cele ce urmeaza vom arata, prin câteva exemple, cât de flexibile sunt formele normale imerse.

Exemplul 4.1. Fie schema relationala R = Copil Jucarie si fie M = {Copil $\to\to$ Jucarie}. Sigur ca R este în forma normala 4NF în raport cu M. Asa cum am vazut în exemplul 2.1, deoarece M contine numai multidependente triviale, fiecare din urmatoarele submultimi de multidependente toarele submultimi de multidependente M, E(D) si M′, asa cum sunt ele statuate în definitiile din [GN02], contin numai multidependente triviale. Prin urmare, conform definitiei cheii ([GN02]), nici una din aceste multimi nu detine o cheie si deci R nu îndeplineste conditiile de a fi NNFmk, NNFed sau NNFmd. Cu toate acestea urmatoarele scheme relationale imerse R_1 = Copil Jucarie, R_2 = Copil (Jucarie)* si R_3 = Jucarie (Copil)* sunt în forma normala NNFmr. Pe care dintre ele o folosim depinde de aplicatia pe care vrem s-o realizam.

Exemplul 4.2. Fie U = {Prof, Hobby, HobbyEchipament} si fie multimea de multidependente functionale M = {Hobby →→ HobbyEchipament}. Sa observam ca M este libera de conflict. Singurele scheme relationale imerse permise de formele NNFmk, NNfed sau NNFmd sunt Hobby (Prof)*(HobbyEchipament)* deoarece singura cheie este Hobby. De notat este faptul ca aceasta schema relationala imersa este si sub forma NNFmr. Datele sunt memorate în aceasta schema relationala din punctul de vedere al lui Hobby. Presupunând ca am vrea sa memoram datele din punctul de vedere al lui Prof atunci schemele relationale imerse de care am avea nevoie ar fi Prof (Hobby)* si Hobby (HobbyEchipament)*. Ambele scheme relationale imerse satisfac conditiile de a fi sub forma NNFmr, dar, deoarece Prof nu este o cheie, Prof(Hobby)* nu satisface conditia a patra a formelor normale imerse NNFmk, NNFed si NNFmd.

Sa observam ca exemplele de mai sus sunt foarte naturale si apar frecvent în practica. Din aceste exemple putem deduce ca ultima conditie a formelor NNFed si NNFmd este foarte problematica si ca forma NNFmk mosteneste aceleasi probleme. Din aceste considerente putem deduce ca formele NNFmk, NNFed, NNFmd sunt mai restrictive si ca forma NNFmr prezinta cel mai flexibil grad în proiectare.

Bibliografie

[01] N. Giurgiteanu. Proiectarea bazelor de date relationale, Ed. SITECH, Craiova, 1997.

[02] N. Giurgiteanu. Teoria si practica bazelor de date. I. Scheme relationale, relatii si forme normale imerse (scufundate). Rev. Inormatica Economica, Vol VI, nr. 3/2002, pp. 48-55

[03] C. Beeri. On the Membership problem for functional and multvalued dependencies in relational databases. ACM Transactions on Database Systems, vol 5, nr. 3: 241-259, September 1980.

[04] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis. On the desirability of acyclic database schemes. Jonrnal of the ACM, 30(3):479-513, July 1983.

[05] R. Fagin. Multivalued dependencies and a new normal form for relational databases. ACM Transoctions on Database Systems, 2(3):262-278, September 1977.

[06] D. Maier. The Theory of Relationa! Databases. Computer Science Press, Rockville Maryland, 1983.

[07] W. Y. Mok, Y. K. Ng, and D. W. Embley. A normal form for precisely characterizing redundancy in neated relations. ACM Transactiona on Database Syatema, 21(1):77-106, March 1996.

[08] Z. M. Ozsoyoglu and L. Y. Yuan. A new normal form for nested relations. ACM Traraaactions an Database Syatema, 12(1):111-136, March 1987.

[09] Z. M. Ozeoyoglu and L. Y. Yuan. On the normalization in nested relational databases. Lecture Notes in Computer Science, 361, pages 243-271, 1989.

[10] M. A. Roth and H. F. Korth. The design of ¬1NF relational databases into nested normal form. In Proceedings of the 1987 ACM-SIGMOD Conference, pages 143-159 San Francisco California, May 1987.

[11] E. Sciore. Real-world MVD's. In Proceedsngs of the 198l ACM-SIGMOD Confenence, pages 121-132, Ann Arbor, Michigan, April 1981.