Izvorni znanstveni članak

Dekompozicija relacijske sheme bez gubitka informacija

Vladimir Mateljan
Odsjek za informacijske znanosti, Filozofski fakultet
I. Lučića 3, Zagreb, Hrvatska
vmatelja@ffzg.hr

Goran Đambić
Croatia Airlines
Zagreb, Hrvatska
goran.djambic@croatiaairlines.hr

Sažetak

Postupkom normalizacije podataka relacijska shema dekomponira se na više manjih i pravilnijih podshema. Pomoću Rissanenovog kriterija nezavisnosti komponenata moguće je dokazati je li polazni skup informacija sačuvan, tj. da je dekompozicija početne relacijske sheme reverzibilna. Zahvaljujući tome Rissanenov test ima izvanrednu važnost u postupku oblikovanja baze podataka. Za Rissanenov test potrebno je imati komplet relacijskih shema koje se dobiju dekompozicijom, tj. potrebno je najprije izvršiti normalizaciju baze na relacijske sheme, a zatim izvršiti test. U radu se predlaže jedan novi test pomoću kojeg se može provjeriti reverzibilnost dekompozicije relacijske sheme bez da se prethodno izvršio postupak normalizacije.

Ključne riječi: funkcijska zavisnost, relacijska shema, normalizacija, Rissanenenov test, dekompozicija, reverzibilnost, gubitak informacija

1. Uvod

Normalizacija je metoda oblikovanja baze podataka. Osnovni ciljevi ove metode su eliminacija anomalija održavanja baze i svođenje redundancije u bazi na kontroliranu redundanciju. Normalizacija sintezom polazi od skupa funkcijskih zavisnosti zadanog na skupu atributa i direktno konstruira skup relacijskih shema u trećoj normalnoj formi. Konačan model baze podataka proizlazi iz skupa funkcijskih zavisnost koji je zadan na skupu atributa. Ako u takvim modelima, koji proizlaze iz skupa funkcijskih zavisnosti, vrijede Rissanenovi uvjeti reverzibilnosti, onda je dekompozicija početne relacijske sheme reverzibilna, odnosno tokom dekompozicije nije došlo do gubitka informacija. Za normalizaciju baze do treće normalne forme, u radu ćemo koristiti Bernsteinov algoritam za normalizaciju sintezom.

2. Bernsteinov algoritam za normalizaciju sintezom

Algoritam za vertikalnu normalizaciju sintezom definirao je Bernstein (1976) i glasi:

Ulaz : Skup funkcijskih zavisnosti (**FZ**) **F**.
Izlaz : Komplet relacijskih shema u trećoj normalnoj formi (**3NF**).

Postupak:

(1) Nađi atribut **Z** tako da vrijedi $\mathbf{Z}\in\mathbf{R}$.
(2) Konstruiraj funkcijsku zavisnost $\mathbf{R}\rightarrow\mathbf{Z}$ i dodaj je u **F**.
(3) Nađi reducirani neredundantni prstenasti pokrivač **G** za prošireni **F**.
(4) Za svaku sastavljenu funkcijsku zavisnost (**SFZ**) $(\mathbf{X_1},\mathbf{X_2},...,\mathbf{X_n})\rightarrow\mathbf{Y}$ u **G** konstruiraj relacijsku shemu $\mathbf{R_i}(\mathbf{X_1X_2...X_nY})$ s ključevima $\mathbf{X_1},\mathbf{X_2},...,\mathbf{X_n}$.
(5) Izbaci atribut **Z** iz relacijske sheme $\mathbf{R_i}$.
(6) Ispiši skup relacijskih shema u **3NF**.

3. Reverzibilnost dekompozicije relacijske sheme

Normalizacijom podataka, relacija se dekomponira na više projekcija. Pri tome ne smije doći do gubitka informacija, odnosno mora biti moguće operacijom prirodnog pridruživanja nad projekcijama uspostaviti polaznu relaciju. Za ovakvu dekompoziciju kažemo da je reverzibilna. Skup projekcija koje zadovoljavaju ovaj zahtjev Rissanen naziva nezavisnim komponentama relacije.

3.1. Rissanenov uvjet nezavisnosti komponenata

Prema Rissanenu (Rissanen, 1977), nezavisne komponente relacije moraju ispunjavati sljedeće uvjete reverzibilnosti, tj. moraju biti ispunjeni kriteriji:

(1) $\mathbf{R}=\mathbf{R_1}\cup\mathbf{R_2}$
(2) $(\mathbf{R_1}\cap\mathbf{R_2}\rightarrow\mathbf{R_1})\vee(\mathbf{R_1}\cap\mathbf{R_2}\rightarrow\mathbf{R_2})$

Riječima ove kriterije možemo izraziti na sljedeći način: Relacijske sheme $\mathbf{R_1}$ i $\mathbf{R_2}$, koje su dobivene dekompozicijom relacijske sheme **R,** su nezavisne ako je unija atributa u projekcijama relacije jednaka relacijskoj shemi te presjek $\mathbf{R_1}$ i $\mathbf{R_2}$ sadrži ključ bar jedne od njih.
Ako želimo ustanoviti jesu li zadovoljeni Rissannenovi uvjeti, moramo normalizirati bazu i za dobivene relacijske sheme u **3NF** provjeriti prethodna dva uvjeta.

3.2. Provjera reverzibilnosti normalizirane relacijske sheme

Dekompozicija je reverzibilna ako se skup relacijskih shema $\mathbf{R}$ sastoji od relacijskih shema $\mathbf{R_1},...,\mathbf{R_n}$ koje su dobivene postupkom vertikalne normalizacije sintezom i ako su ispunjeni Rissanenovi uvjeti reverzibilnosti $(\forall \mathbf{R_i} \in \mathbf{R})$ $(\exists \mathbf{R_j} \in \mathbf{R}, \mathbf{j} \neq \mathbf{i})$ $(\mathbf{R_i} \cap \mathbf{R_j}$ sadrži ključ od $\mathbf{R_i}$ ili od $\mathbf{R_j})$.

4. Testiranje reverzibilnosti na skupu funkcijskih zavisnosti

Pokazat ćemo da se provjera reverzibilnosti dekompozicije relacijske sheme može izvesti na skupu funkcijskih zavisnosti, bez da se prije baza normalizirala. Razmatrat ćemo skup funkcijskih zavisnosti $\mathbf{F}=\{\mathbf{X_1} \rightarrow \mathbf{Y_1}, \mathbf{X_2} \rightarrow \mathbf{Y_2},...,\mathbf{X_n} \rightarrow \mathbf{Y_n}\}$ u kojem niti jedna funkcijska zavisnost $\mathbf{X_i} \rightarrow \mathbf{Y_i}$ neće sadržavati isti atribut na svojoj lijevoj i desnoj strani. Naime, skup funkcijskih zavisnosti $\mathbf{F}$ uvijek se može zamijeniti skupom $\mathbf{F}=\{\mathbf{X_1} \rightarrow \mathbf{Y_1} \setminus \mathbf{X_1}, \mathbf{X_2} \rightarrow \mathbf{Y_2} \setminus \mathbf{X_2},...,\mathbf{X_n} \rightarrow \mathbf{Y_n} \setminus \mathbf{X_n}\}$.

Tvrdnja:

Neka je $\mathbf{F}=\{\mathbf{X_1} \rightarrow \mathbf{Y_1}, \mathbf{X_2} \rightarrow \mathbf{Y_2},...,\mathbf{X_n} \rightarrow \mathbf{Y_n}\}$ skup funkcijskih zavisnosti nad skupom atributa $\mathbf{R}$. Ako $\mathbf{F}$ sadrži **FZ** $\mathbf{X_r} \rightarrow \mathbf{Y_r}$ i $\mathbf{X_s} \rightarrow \mathbf{Y_s}$, $\mathbf{r} \neq \mathbf{s}$ za koje vrijedi:

(a) $\exists \mathbf{W}\ (\mathbf{W} \subset \mathbf{X_r} \cap \mathbf{Y_s})\ \&\ (\mathbf{W} \not\subset \bigcup_{i \neq r \vee i \neq s} \mathbf{X_i}\ \mathbf{Y_i})$

(b) $\exists\ \mathbf{X_{r0}}\ (\mathbf{X_{r0}} \subset \mathbf{X_r} \setminus \mathbf{W})\ \&\ (\mathbf{X_{r0}} \not\subset \bigcup_{i \neq r} \mathbf{X_i}\ \mathbf{Y_i})$

(c) $\exists \mathbf{Y_{r0}}\ (\mathbf{Y_{r0}} \subset \mathbf{Y_r})\ \&\ (\mathbf{Y_{r0}} \not\subset \bigcup_{i \neq r} \mathbf{X_i}\ \mathbf{Y_i})$

(d) $\exists\ \mathbf{X_{s0}}\ (\mathbf{X_{s0}} \subset \mathbf{X_s})\ \&\ (\mathbf{X_{s0}} \not\subset \bigcup_{i \neq s} \mathbf{X_i}\ \mathbf{Y_i})$

onda se Bernsteinovim algoritmom za normalizaciju sintezom dobiva skup relacijskih shema koji sadrži relacijsku shemu $\mathbf{R_r}$ koja nije povezana niti s jednom od preostalih relacijskih shema dobivenih u postupku normalizacije.

Drugim riječima:

Ako skup funkcijskih zavisnosti $\mathbf{F}$ nad skupom atributa $\mathbf{R}$ zadovoljava (a), (b),(c) i (d) onda se Bernsteinovim algoritmom za normalizaciju sintezom dobiva skup relacijskih shema koje ne zadovoljavaju Rissanenov test reverzibilnosti.

Dokaz:

Primijenimo li Bernsteinov algoritam za normalizaciju sintezom na skup funkcijskih zavisnosti $\mathbf{F}$ koji zadovoljava uvjete (a), (b),(c) i (d) tada imamo:

(1) Neka vrijedi $\mathbf{Z} \not\in \mathbf{R}$
(2) Zamijenimo skup **F** sa skupom $\mathbf{F}^{'}=\mathrm{F}\cup\{\mathbf{R}\rightarrow\mathbf{Z}\}$
(3) Reducirani neredundantni pokrivač:

Neredundantni pokrivač:

Funkcijska zavisnost $\mathbf{X_r}\rightarrow\mathbf{Y_r}$ u postupku traženja neredundantnog reduciranog pokrivača neće biti izbačena iz skupa funkcijskih zavisnosti. Naime, kako njezina desna strana sadrži atribut $\mathbf{Y_{r0}}$ koji, prema (b), nije niti u jednoj drugoj preostaloj funkcijskoj zavisnosti u $\mathbf{F}^{'}$, zaključujemo da ne može biti izvedena iz $\mathbf{F}^{'}\backslash\{\mathbf{X_r}\rightarrow\mathbf{Y_r}\}$.

Lijevo reduciranje:

Skup $\mathbf{X_{r0}}\subset\mathbf{X_r}\backslash\mathbf{W}$ neće biti izbačen u postupku lijevog reduciranja **FZ** $\mathbf{X_r}\rightarrow\mathbf{Y_r}$ jer, prema (b) i (c), vrijedi $\mathbf{Y_{r0}}\not\subset(\mathbf{X_r}\backslash\mathbf{A})^{+}\big|_{\mathbf{F}'}$ niti za jedan $\mathbf{A}\in\mathbf{X_{r0}}$.

Skup $\mathbf{X_{s0}}\subset\mathbf{X_s}$ neće biti izbačen u postupku lijevog reduciranja **FZ** $\mathbf{X_s}\rightarrow\mathbf{Y_s}$ jer, prema (c) i (d), vrijedi $\mathbf{Y_{r0}}\not\subset(\mathbf{X_r}\backslash\mathbf{A})^{+}\big|_{\mathbf{F}'}$ niti za jedan $\mathbf{A}\in\mathbf{X_{r0}}$.

Dakle, lijevim reduciranjem **FZ** $\mathbf{X_s}\rightarrow\mathbf{Y_s}$ preostat će **FZ** $\mathbf{X^{'}_s}\rightarrow\mathbf{Y_s}$ i vrijedit će $\mathbf{X_{s0}}\subset\mathbf{X^{'}_s}$.

Skup $\mathbf{W}\subset\mathbf{X_r}$ neće biti izbačen u postupku lijevog reduciranja **FZ** $\mathbf{X_r}\rightarrow\mathbf{Y_r}$ jer, prema (a), (b) i (c), vrijedi $\mathbf{Y_{r0}}\not\subset(\mathbf{X_r}\backslash\mathbf{B})^{+}\big|_{\mathbf{F}'}$ niti za jedan $\mathbf{B}\in\mathbf{W}$.

Niti jedan atribut iz skupa $\mathbf{Y_{s0}}$ neće biti izbačen iz lijeve strane **FZ** $\mathbf{R}\rightarrow\mathbf{Z}$ jer, prema (b), osim u $\mathbf{X_r}\rightarrow\mathbf{Y_r}$, skup $\mathbf{X_{r0}}$ se ne nalazi niti u jednoj drugoj **FZ** $\mathbf{X_i}\rightarrow\mathbf{Y_i}$ $(\mathbf{i}\neq\mathbf{r})$.

Pokažimo da u postupku lijevog reduciranja neće biti izbačen skup atributa $\mathbf{X_{s0}}$ sa lijeve strane **FZ** $\mathbf{R}\rightarrow\mathbf{Z}$.

Ako bi bio izbačen neki atribut $\mathbf{A}\in\mathbf{X_{s0}}$, značilo bi da postoji $\mathbf{X}\subset\mathbf{R}$ za koji bi vrijedila **FZ** $\mathbf{X}\rightarrow\mathbf{A}$, što je nemoguće, jer se, prema (d), atributi iz $\mathbf{X_{s0}}$ ne nalaze na desnoj strani niti jedne **FZ** iz skupa **F**.

Pokažimo da će u postupku lijevog reduciranja biti izbačen skup atributa **W** sa lijeve strane **FZ** $\mathbf{R}\rightarrow\mathbf{Z}$.

Funkcijsku zavisnost **FZ** $\mathbf{R \rightarrow Z}$ možemo pisati:

$\mathbf{X_1 Y_1 X_2 Y_2 ... (X_r \backslash W) Y_r ... X'_s\ Y_s ... X_n Y_n W \rightarrow Z}$

Prema (a) i (d), iz $\mathbf{X'_s \rightarrow Y_s}$ i $\mathbf{W \subset Y_s}$ slijedi $\mathbf{X'_s \rightarrow W}$. Prema tome, skup atributa **W** biti će u postupku lijevog reduciranja izbačen iz $\mathbf{R \rightarrow Z}$ jer je **W** određen skupom atributa $\mathbf{X'_s}$. Dakle, lijevim reduciranjem **FZ** $\mathbf{R \rightarrow Z}$ dobivamo **FZ** $\mathbf{R' \rightarrow Z}$, gdje skup atributa $\mathbf{R'}$ ne sadrži **W**.

Desno reduciranje:

U postupku desnog reduciranja, prema (b), skup $\mathbf{Y_{r0}}$ ne može biti izbačen iz skupa atributa sa desne strane **FZ** $\mathbf{X_r \rightarrow Y_r}$ jer se $\mathbf{Y_{r0}}$ nalazi samo u **FZ** $\mathbf{X_r \rightarrow Y_r}$ i niti u jednoj drugoj **FZ** $\mathbf{X_i \rightarrow Y_i}$ ($\mathbf{i \neq j}$).

U postupku desnog reduciranja niti jedan atribut $\mathbf{A \in W}$ ne može biti izbačen sa desne strane **FZ** $\mathbf{X_s \rightarrow Y_s}$ jer se **A** ne nalazi na desnoj strani niti jedne preostale funkcijske zavisnosti.

(4) i (5):

Zatvarač $\mathbf{X_r}^+ \big|_{F'}$ neće biti ekvivalentan niti jednom zatvaraču $\mathbf{X_i}^+ \big|_{F'}$ (i≠r) jer $\mathbf{X_r}$ sadrži skup $\mathbf{X_{r0}}$ koji nije u lijevoj strani niti jedne **FZ** $\mathbf{X_i \rightarrow Y_i}$ ($\mathbf{i \neq r}$). Iz toga slijedi **SFZ** $\mathbf{(X'_r) \rightarrow Y'_r}$ kojoj odgovara relacijska shema $\mathbf{R_r(X'_r Y'_r)}$ s ključem $\mathbf{X'_r}$. Nadalje, pokazat ćemo da $\mathbf{R_r}$ nije povezana niti s jednom od preostalih relacijskih shema koje su dobivene u postupku normalizacije.

Najprije ćemo pokazati da relacijska shema $\mathbf{R_z}$, koja je generirana funkcijskom zavisnošću $\mathbf{R \rightarrow Z}$, nije povezana s relacijskom shemom $\mathbf{R_r}$. Naime, relacijska shema $\mathbf{R_z}$ sastoji se od svih atributa koji su preostali nakon lijevog reduciranja funkcijske zavisnosti $\mathbf{R \rightarrow Z}$. Ti atributi sačinjavaju ključ relacijske sheme $\mathbf{R_z}$. Kako $\mathbf{R_z}$ sadrži skup $\mathbf{X_{s0}}$ koji nije u $\mathbf{R_r}$, slijedi da $\mathbf{R_r}$ ne može sadržavati strani ključ koji pokazuje na $\mathbf{R_z}$. S druge strane, ključ od $\mathbf{R_r}$ sadrži atribut **W** koji nije u $\mathbf{R_z}$. Prema tome, $\mathbf{R_z}$ ne pokazuje preko stranog ključa na $\mathbf{R_r}$. Dakle, relacijske sheme $\mathbf{R_r}$ i $\mathbf{R_Z}$ nisu međusobno povezane.

Sada ćemo pokazati da niti jedna druga od preostalih relacijskih shema dobivenih u postupku normalizacije nije povezana preko stranog ključa s relacijskom shemom $\mathbf{R_r}$. Naime, niti jedna relacijska shema, osim relacijske sheme $\mathbf{R_r}$, ne sadrži atribute $\mathbf{X_{r0}}$ koji su dio ključa od $\mathbf{R_r}$ pa prema tome niti jedna relacijska shema ne može pokazivati sa svojim stranim ključem na $\mathbf{R_r}$. Vrijedi i obratno, tj. $\mathbf{R_r}$ ne pokazuje sa svojim stranim ključem niti na jednu relacijsku shemu $\mathbf{R_i}$

(**i≠r**) koja je dobivena u postupku normalizacije. Naime, ako bi $\mathbf{R_r}$ pokazivala na neku relacijsku shemu $\mathbf{R_i}$ (**i≠r**), to bi značilo da $\mathbf{R_r}$ sadrži ključ od $\mathbf{R_i}$. To je nemoguće jer je jedini skup atributa koji je zajednički nekoj relacijskoj shemi $\mathbf{R_i}$ (**i≠r**), skup atributa **W** koji je desni dio **FZ** $\mathbf{X_s \rightarrow Y_s}$, a atributi desne strane **FZ** ne mogu biti ključni atributi.

Primjer:

Neka je zadan skup **FZ F**={**CE→DF**, **AFG→BH**,**EI→CD**,**D→C**,**CD→EI**} na relacijskoj shemi **R**=(**ABCDEFGHI**).

Uvedimo supstitucije:

$\mathbf{X_r}$=**AFG**, $\mathbf{Y_r}$=**BH**, $\mathbf{X_s}$=**CE**, $\mathbf{Y_s}$=**DF**, **W**=**F**, $\mathbf{X_{r0}}$=**AG**, $\mathbf{Y_{r0}}$=**BH**, $\mathbf{X_{s0}}$= **CE**

Za $\mathbf{X_r}$, $\mathbf{Y_r}$, $\mathbf{X_s}$, $\mathbf{Y_s}$, **W**, $\mathbf{X_{r0}}$, $\mathbf{Y_{r0}}$ i $\mathbf{X_{s0}}$ vrijede uvjeti iz prethodno dokazane tvrdnje. Prema tome zaključujemo da će se normalizacijom dobiti skup relacijskih shema koji će sadržavati relacijsku shemu $\mathbf{R_r}$(**X'Y'**) te da $\mathbf{R_r}$ neće biti povezana preko stranih ključeva niti s jednom od preostalih relacijskih shema koje će biti dobivene normalizacijom. Dakle, zaključujemo da će se normalizacijom relacijska shema **R** dekomponirati na skup relacijskih shema uz gubitak informacija.

Isti zaključak mogli smo dobiti normaliziranjem relacijske sheme **R** te provjerom Rissanenovih uvjeta reverzibilnosti. Normalizacijom se dobiva slijedeći skup relacijskih shema u **3NF**:

$\mathbf{R_1}$(**CDEFI**) s ključevima **CE**, **EI**, **D**
$\mathbf{R_2}$(**ABFGH**) s ključem **AFG**
$\mathbf{R_3}$(**AEGI**) s ključem **AEGI**

Međusobni presjeci relacijskih shema su:
1. $\mathbf{R_1 \cap R_2}$=**F** ne sadrži ključ niti iz $\mathbf{R_1}$ niti iz $\mathbf{R_2}$
2. $\mathbf{R_1 \cap R_3}$=**EI** sadrži ključ iz $\mathbf{R_1}$
3. $\mathbf{R_2 \cap R_3}$=**AG** ne sadrži ključ niti iz $\mathbf{R_2}$ niti iz $\mathbf{R_3}$

Iz **1.** i **3.** zaključujemo da relacijska shema $\mathbf{R_2}$ nije povezana niti s $\mathbf{R_1}$ niti s $\mathbf{R_3}$. Iz **2.** zaključujemo da je relacijska shema $\mathbf{R_3}$ povezana s $\mathbf{R_1}$. Prema Rissanenovim uvjetima slijedi da dekompozicija početne relacijske sheme **R** nije reverzibilna, odnosno da je tokom dekompozicije došlo do gubitka informacija. Može se primijetiti da relacija $\mathbf{R_2}$ u ovom primjeru odgovara relaciji $\mathbf{R_r}$ u tvrdnji. Prednost ispitivanja reverzibilnosti na skupu funkcijskih zavisnosti je u tome što se ne treba izvršiti postupak normalizacije relacijske sheme, tj. prije postupka normalizacije možemo zaključiti hoćemo li izgubiti informacije ako relacijsku shemu normaliziramo.

Zaključak

Pomoću Rissanenovog kriterija nezavisnosti komponenata moguće je pokazati je li pri normalizaciji relacijske sheme došlo do gubitka informacija. Da bi se izvršio Rissanenov test, potrebno je imati komplet relacijskih shema koje se dobiju normalizacijom početne relacijske sheme. U radu je pokazano da je moguće provjeriti je li dekompozicija relacijske sheme ostvarena bez gubitka informacija, a da se prethodno nije morala izvršiti normalizacija, čime se dobiva na efikasnosti testiranja reverzibilnosti jer postupak normalizacije može biti dugotrajan.

Literatura

Bernstein, P. A. Synthesizing third normal form relations from functional dependencies. //ACM Transitions on Database Systems, 1, 1976, pp. 277-298

Date, S. J. An introduction to Database Systems, Add.-Wesley Pub. Co, 1995, New York.

Maier, D. The theory of Relational databases, Comp. sciences Press, Rockville, MA.

Mateljan, V. The Possibility of Applaying the Calculus of Functional Dependences to a Knowledge Base. // Journal of information and organizational sciences. 4 (2000) , 1; 69-81

Rissanen, J. Independent Components of Relations // ACM Transitions on Database Syste ms, Vol. 2, No. 4, December 1977, pp. 317-325

Tkalac, S.; Mateljan, V. Algorithm for Nodes Arrangement in Graphic Reprezentation of Functional Dependencies Set. // Informatologija. 24 (1992), 3-4; 101-108.

Tkalac, S.; Mateljan, V. Proposition for modification of Bernstein's Algorithm for Vertical Normalization by Synthesis. // Informatologija. 24 (1992) , 1; 1-11.