

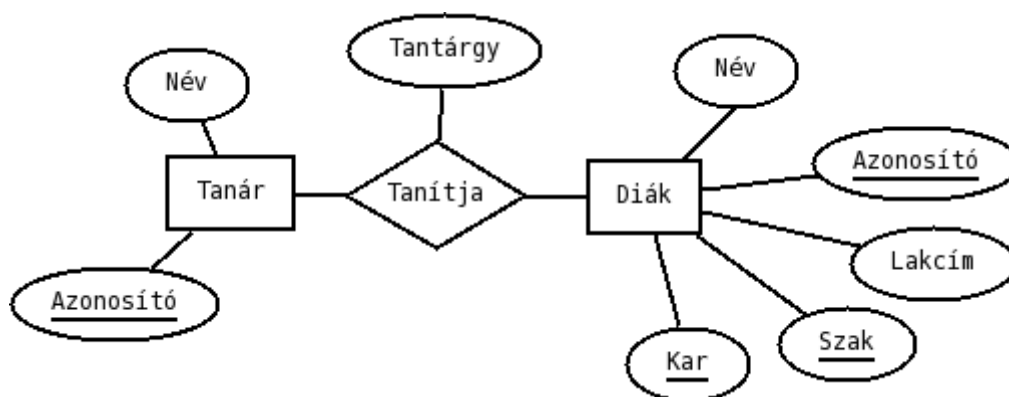
6. Gyakorlat

Relációs adatbázis normalizálása

Redundancia:

Az E-K diagramok felírásánál vagy az átalakításnál elképzelhető, hogy nem az optimális megoldást írjuk fel. Ekkor az adat redundáns lehet.

Példa:



Ebből a következő relációsémát kaphatjuk:

TANÁR(T.Azonosító, Név)

DIÁK(D.Azonosító, Név, Kar, Szak, Lakcím)

TANÍTJA(T.Azonosító, D.Azonosító, Kar, Szak, Tantárgy)

vagy egyszerűbb is lehet

TANÁR(T.Azonosító, Név)

DIÁK(D.Azonosító, Név, Kar, Szak, Lakcím, T.Azonosító, Tantárgy)

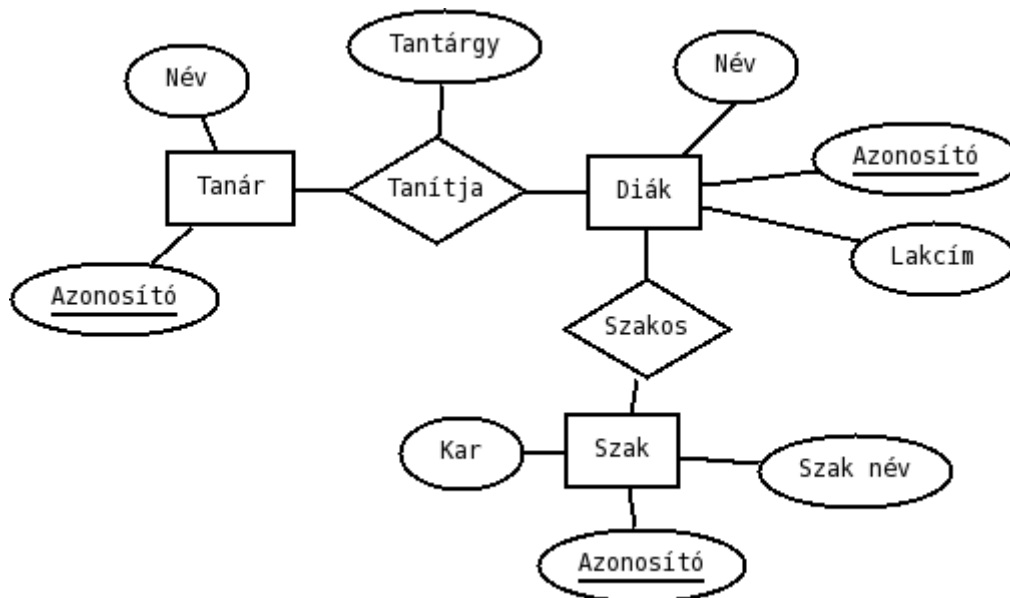
A Diák egyed esetében több attribútum alkotja a kulcsot, mivel elképzelhető, hogy egy diák több szakra is jár, valamint az sem kizárt, hogy különböző karoknak van azonos nevű szakja. Ilyen esetben a kar és szak redundáns adat lehet. Tegyük fel, hogy a tantárgyat mindig ugyanaz a tanár tanítja. Egy tanár több tantárgyat is taníthat.

A redundáns adat legfőképpen az adatok módosításánál okozhat problémát.

- Indul egy új szak, és még nincs hallgató, aki jelentkezett volna oda. Ha a fenti sémát követjük, akkor nem tudunk felvenni az adatbázisba új szakot, amíg nem jelentkezett rá senki.
- Megszűnik egy szak, akkor minden hallgatót törölni kell a szak alapján.
- Megváltozik egy szak neve, minden hallgatónál meg kell változtatni, persze figyelni kell arra, hogy melyik karon van az a szak, és oda jár- e a hallgató.

A redundanciát csökkenti az alábbi modell:

Vegyünk egy Szak egyedet, amely rendelkezik külön azonosítóval, attribútumként tárolja azt is, hogy melyik karhoz tartozik. A Diák egyednek így kevesebb attribútuma lesz, csak a személyére vonatkozó információk lesznek az attribútumok. Bármikor felvehető egy szak, vagy módosítható a neve.



A relációséma lehet:

TANÁR(T.Azonosító, Név)
 DIÁK(D.Azonosító, Név, Lakcím, SZ.Azonosító)
 SZAK(SZ.Azonosító, Kar, Szak név)
 TANÍTJA(T.Azonosító, D.Azonosító, Tantárgy)

Funkcionális függőség:

Legyen $R(A_1, \dots, A_n)$ egy relációséma, $P, Q \subseteq \{A_1, \dots, A_n\}$. P funkcionálisan függ Q -tól (jelölés $P \rightarrow Q$), ha bármely R feletti T tábla esetében valahányszor két sor megegyezik P -n akkor megegyezik Q -n is, vagyis minden $t_i \in T$ és $t_j \in T$ esetén

$$t_i(P) = t_j(P) \Rightarrow t_i(Q) = t_j(Q)$$

A $P \rightarrow Q$ funkcionális függőség triviális, ha $Q \subseteq P$.
 A $P \rightarrow Q$ függést teljesen nem triviálisnak nevezzük, ha $Q \cap P = \emptyset$.

Függőségek a fenti példában:

- $f_1: \{D.Azonosító\} \rightarrow \{D.Név, D.Lakcím\}$
- $f_2: \{SZ.Azonosító\} \rightarrow \{Szak név, Kar\}$
- de $\{D.Azonosító\} \rightarrow \{Szak\}$ nem igaz, mert egy diák több szakra is járhat.

Állítás: Egy $K \subseteq A$ attribútumhalmaz akkor és csak akkor superkulcs, ha $K \rightarrow A$.

Definíció: Relációsémának nevezzük egy $R(A, F)$ párt, ahol $A = \{A_1, \dots, A_n\}$, $F = \{f_1, \dots, f_m\}$ az A -n definiált funkcionális függések egy halmaza.

Armstrong-axiómák:

- A1: Reflexivitás: Ha $Y \subseteq X$, akkor $X \rightarrow Y$.
- A2: Bóvítés: Ha $X \rightarrow Y$, akkor $X \cup Z \rightarrow Y \cup Z$.
- A3: Transzitivitás: Ha $X \rightarrow Y$ és $Y \rightarrow Z$, akkor $X \rightarrow Z$.

Szétvágási szabály: ha $X \rightarrow \{B_1, \dots, B_k\}$ akkor $X \rightarrow B_1, \dots, X \rightarrow B_k$

Egyesítési szabály: ha $X \rightarrow B_1, \dots, X \rightarrow B_k$ akkor $X \rightarrow \{B_1, \dots, B_k\}$

Definíció: Az X attribútumhalmaz lezártja az F függőségi halmaz szerint $X^+ = \{A_i \mid X \rightarrow A_i\}$, vagyis az összes X -től függő attribútumból áll, amelyekre az $X \rightarrow A_i$ függőség F -ből levezethető.

Algoritmus X^+ kiszámítására:

Az $X = X^{(0)} \subset X^{(1)} \subset \dots \subset X^{(n)} = X^+$ halmzsorozatot képezzük. $X^{(i)}$ -ből $X^{(i+1)}$ előállítására: keressünk olyan F -beli $P \rightarrow Q$ függőséget, amelyre $P \subseteq X^{(i)}$, de Q már nem része $X^{(i)}$ -nek! Ha találunk ilyen, akkor $X^{(i+1)} := X^{(i)} \cup Q$, ha nem akkor $X^{(i)} = X^+$.

Példa:

$R = (Z, F)$, $Z = \{A, B, C, D, E\}$ és $F = \{ \{C\} \rightarrow \{A\}, \{B\} \rightarrow \{C, D\}, \{D, E\} \rightarrow \{C\} \}$

Határozzuk meg a B^+ halmazt!

$X^{(0)} = \{B\}$,

$X^{(1)} = \{B\} \cup \{C, D\} = \{B, C, D\}$

$X^{(2)} = \{B, C, D\} \cup \{A, C, D\} = \{A, B, C, D\}$

$X^{(3)} = X^{(2)}$

Állítás: Egy K attribútumhalmaz akkor és csak akkor szuperkulcs, ha $K^+ = A$.

Definíció: Az F függéshalmaz lezártja az összes F -ből levezethető függést tartalmazza. Jelölése: F^+ .

Definíció: Az F^+ egy részhalmazát bázisnak nevezzük, ha belőle F valamennyi függése levezethető.

Állítás: $F^+ = \{X \rightarrow Y \mid Y \subseteq X^+\}$, vagyis F^+ pontosan azokból az $X \rightarrow Y$ függésekből áll, amelyekre Y részhalmaza X^+ -nak.

Algoritmus F^+ meghatározására:

1. Vegyük az A attribútumhalmaz összes részhalmazát.
2. Minden X részhalmazhoz állítsuk elő X^+ -t.
3. Valamennyi $Y \subseteq X^+$ -ra az $X \rightarrow Y$ függőséget felvesszük F^+ -ba.

Felbontás (dekompozíció):

Legyen $R(A)$ egy relációséma, $X, Y \subset A$, $X \cup Y = A$. Ekkor az $R(A)$ séma felbontása X, Y szerint $R_1(X), R_2(Y)$. Táblákkal kifejezve: $T_1 = \pi_X(T)$ és $T_2 = \pi_Y(T)$.

Definíció: Egy felbontás hűségese, ha bármely R feletti T tábla esetén $T = T_1 * T_2$.

Tétel (Health tétele): Ha az $R(A)$ sémánál $A = B \cup C \cup D$, ahol B, C , és D diszjunktak és $C \rightarrow D$, akkor az $R_1(B \cup C), R_2(C \cup D)$ felbontás hűségese.

Normalizálás:

2NF:

Egy $R = (A, F)$ relációséma 2NF-ben van, ha minden másodlagos attribútum teljesen függ bármely kulcstól.

Következmény: Ha minden kulcs egyetlen attribútumból áll, akkor a séma 2NF-ben van.

Ha a sémában nincs másodlagos attribútum, akkor a séma 2NF-ben van.

2NF-re hozás:

Ha valamely K kulcsra $L \subset K$ és $L \rightarrow B$ (B az összes L -től függő attribútum halmaza), akkor a sémát felbontjuk $L \rightarrow B$ függőség szerint. Legyen $C = A - (L \cup B)$, ekkor az $R(A)$ sémát $R_1(C \cup L)$ és $R_2(L \cup B)$ sémákkal helyettesítjük.

Példa:

KURZUS(kurzuskód, kurzusnév, szak_kód, szak, kar)

KURZUS(kurzuskód, kurzusnév, szak_kód)

SZAK(szak_kód, szak, kar)

3NF:

Egy $R=(A,F)$ relációséma 3NF-ben van, ha minden másodlagos attribútuma közvetlenül függ bármely kulcstól.

3NF-re hozás:

Ha valamely K kulcsra $K \rightarrow Y \rightarrow B$ tranzitív függés fennáll, akkor a sémát felbontjuk az $Y \rightarrow B$ függőség szerint (B az összes Y -től függő másodlagos attribútumok halmaza). Legyen $C = A - (Y \cup B)$, ekkor az $R(A)$ sémát az $R_1(C \cup Y)$ és $R_2(Y \cup B)$ sémákkal helyettesítjük.

Boyce-Codd normálforma (BCNF):

Egy $R = (A,F)$ relációséma BCNF-ben van, ha bármely nemtriviális $L \rightarrow B$ függés esetén L szuperkulcs.

Állítás: Ha egy $R=(A,F)$ BCNF-ben van, akkor 3NF-ben is van.

BCNF-re hozás:

Ha $L \rightarrow B$ teljesen nemtriviális függés és L nem szuperkulcs, akkor a sémát felbontjuk az $L \rightarrow B$ függőség szerint. Legyen $C = A - (L \cup B)$, ekkor az $R(A)$ sémát $R_1(C \cup L)$ és $R_2(L \cup B)$ sémákkal helyettesítjük.

4NF:

Legyen $K, L \subseteq A$ és legyen $M = A - (K \cup L)$. Azt mondjuk, hogy K -től többértékűen függ L ($K \twoheadrightarrow L$), Ha bármely R feletti T tábla esetén ha t_i, t_j sorokra $t_i(K)=t_j(K)$, akkor van olyan sor, amelyre az alábbiak teljesülnek:

- $t(K) = t_i(K) = t_j(K)$
- $t(L) = t_i(L)$
- $t(M) = t_j(M)$

A $K \twoheadrightarrow L$ függés nemtriviális, ha $K \cap L = \emptyset$ és $K \cup L \neq A$.

Tétel (Fagin tétele): Az $R(A)$ relációsémánál legyen $A = B \cup C \cup D$, ahol, B, C, D diszjunktak. R felbontása az $R_1(B \cup C), R_2(C \cup D)$ sémákra akkor és csak akkor hűségese, ha $C \twoheadrightarrow D$ fennáll.

Definíció: Egy relációséma 4NF-ben van, ha minden nem triviális $K \twoheadrightarrow L$ függés esetén K szuperkulcs.

Ha egy relációséma 4NF-ben van, akkor BCNF-ben is van.

4NF-re hozás:

dekompozíció a függőség szerint: ha $K \twoheadrightarrow L$, nemtriviális függés, és K nem szuperkulcs, akkor az $R(A)$ sémát felbontjuk az $R_1(K \cup L)$ és $R_2(K \cup M)$ sémákra.

Feladatok:

1. Határozzuk meg a függőségeket az alábbi relációsémában!

HALLGATÓ(aha, név, város, irányítószám, utca, házszám, szak, kar)

2. Hozzuk 1NF, 2NF, 3NF alakra:

KÖLCSÖNZÉS(olvasójegy, név, lakcím, leltári szám, isbn, könyv címe, kölcsönzés dátuma, kölcsönzés sorszáma, visszahozta)