Relacja

Schemat relacji

Schemat relacji jest to zbiór $R = \{A_1, ..., A_n\}$, gdzie $A_1, ..., A_n$ są artybutami (nazwami kolumn) np. Loty = {Numer, Skąd, Dokąd, Odlot, Przylot}

Każdemu atrybutowi A przyporządkowana jest dziedzina *Dom(A)*, czyli zbiór dopuszczalnych wartości.

np. Dom(Numer) = NUMBER(3), DOM(Skąd) = CHAR (15), ...

Dziedziną relacji o schemacie $R = \{A_1, ..., A_n\}$ nazywamy sumę dziedzin wszystkich trybutów relaci: Dom(R) = Dom(A_1) $\cup$ Dom(A_2) $\cup$ ... $\cup$ Dom(A_n)
Relacja o schemacie $R = \{A_1, ..., A_n\}$ jest to skończony zbiór $r = \{ t_1,..., t_m\}$odwzorowań t_i: $R \rightarrow$ Dom(R) takich, że dla każdego j , $1 \le j \le n$, $t_i (A_j) \in$ Dom(A_j).

Krotka

Każde odwzorowanie t_i: $R \rightarrow$ Dom(R) nazywa się krotką (lub wierszem).

Krotkę można określić przez podanie wartości dla poszczególnych atrybutów:

t(Numer)=83, *t*(Skąd)='Warszawa', *t*(Dokąd)='Wrocław', *t*(Odlot)='6:50', *t*(Przylot)='8:00'

albo graficznie (wiersz tabeli)

Numer	Skąd	Dokąd	Odlot	Przylot
83	Warszawa	Wroclaw	6:50	8:00

Ograniczenie krotki *t* relacji *r* o schemacie *R* do zbioru atrybutów $X \subseteq R$ to odwzorowanie:

$t_{|X} : X \rightarrow$ Dom(X)

Na przykład: ograniczeniem krotki *t* (zdefiniowanej jak wyżej) do zbioru X={Skąd, Dokąd} jest krotka:

$t_{|X}$(Skąd) = 'Warszawa', $t_{|X}$ (Dokąd) = 'Wrocław'

co graficznie jest:

Skąd	Dokąd
Warszawa	Wroclaw

Zależność funkcyjna

Relacja *r* o schemacie $R = \{A_1, ..., A_n\}$spełnia zależność funkcyjną $X \rightarrow Y$ ($X,Y \subseteq R$) jeśli dla każdych dwóch krotek $t, u \in r$ zachodzi warunek:

jeśli $t_{|X} = u_{|X}$ to $t_{|Y} = u_{|Y}$

tzn. w ramach krotek relacji *r* wartości atrybutów zbioru *X* determinują jednoznacznie wartości atrybutów zbioru *Y* (dla każdej wartości w kolumnie Y istnieje dokładnie jedna związana z nią wartość w kolumnie X).

Zakłada się, że z każdym schematem relacji związany jest zbiór spełniających ją zależności funkcyjnych (zależnych od konkretnego zastosowania). Na przykład:
Numer $\rightarrow$ {Skąd, Dokąd, Odlot, Przylot}
{Skąd, Dokąd, Odlot} $\rightarrow$ {Numer, Przylot}
{Skąd, Dokąd, Przylot} $\rightarrow$ {Numer, Odlot }

Klucze

Nadkluczem relacji r o schemacie $R = \{A_1, ..., A_n\}$ nazywamy dowolny zbiór atrybutów $X \subseteq R$ taki, że zachodzi zależność funkcyjna $X \rightarrow R$. Tzn. wartość każdego atrybutu jest jednoznacznie zdeterminowana przez wartości atrybutów zbioru X . Jednym z nadkluczy jest zawsze zbiór wszystkich atrybutów R.

Kluczem relacji r o schemacie $R = \{A_1, ..., A_n\}$ nazywamy każdy minimalny nadklucz (nie zawierający żadnego nadklucza). A więc zbiór atrybutów X jest kluczem, jeśli wartość każdego atrybutu w R jest jednoznacznie zdeterminowana przez wartości atrybutów zbioru X i żaden podzbiór zbioru X nie ma już tej własności.

Zawsze istnieje co najmniej jeden klucz, a może być ich więcej, np.
{Numer}
{Skąd, Dokąd, Odlot}
{Skąd, Dokąd, Przylot}

Wyróżniony klucz nazywa się klucze głównym. Wchodzące w jego skład atrybuty są podkreślane.
Loty = {<u>Numer</u>, Skąd, Dokąd, Odlot, Przylot}

Operatory relacyjne

Operatory teorio-mnogościowe

$\cup$ (suma), $\cap$ (przecięcie), - (różnica)

Selekcja

dla atrybutu $A \in R$, oraz $a \in \mathrm{Dom}(A)$

$$\sigma_{A=a}(r) = \{t \in r \mid t(A) = a\}$$

(tzn. zbiór krotek relacji r, w których wartością atrybutu jest element a)
dla dowolnego warunku logicznego F

$$\sigma_F(r) = \{t \in r \mid t \text{ spelnia warunek } F\}$$

(tzn. zbiór krotek relacji r spełniający warunek F)

Pewne własności selekcji:

$$\sigma_F(r \cup s) = \sigma_F(r) \cup \sigma_F(s)$$

$$\sigma_{F \wedge G}(r) = \sigma_F(\sigma_G(r)) = \sigma_G(\sigma_F(r)) = \sigma_F(r) \cap \sigma_G(r)$$

$$\sigma_{F \vee G}(r) = \sigma_F(r) \cup \sigma_G(r)$$

Rzut

Rzut (projekcja) relacji na zbiór atrybutów $X \subseteq R$:

$$\pi_X(r) = \left\{ t_{|X} : t \in r \right\}$$

(tzn. ograniczenie wszystkich krotek relacji *r* do atrybutów zbioru *X*)

Przykładowe własności:

$\pi_X(\pi_Y(\mathrm{r})) = \pi_X(r)$ - przy założeniu $X \subseteq Y$

$\pi_X(r \cup s) = \pi_X(r) \cup \pi_X(s)$ - dla $\cap$ nie zachodzi

$\sigma_F\left(\pi_X(r)\right) = \pi_X\left(\sigma_F(r)\right)$ - przy założeniu $F(X)$

Równoważności:

SELECT $A_1, \dots, A_n$ FROM *r* WHERE *F*	$\pi_{\{A_1,\dots,A_n\}}\left(\sigma_F(r)\right)$

Iloczyn kartezjański

Dla dwóch relacji: *r* o schemacie $R = \{A_1, ..., A_n\}$ oraz *s* o schemacie $S = \{B_1, ..., B_m\}$, przy założeniu $R \cap S = \varnothing$, iloczynem kartezjańskim nazywamy relację $q = r \otimes s$ o schemacie $Q = \{A_1, ..., A_n, B_1, ..., B_m\}$, składającą się z krotek $t \in q$, dla których istnieją krotki $u \in r$ oraz $v \in s$, takie, że:

$t(A_i) = u(A_i)$ dla $1 \le i \le n$, (tzn. $u = t_{|R}$)

$t(B_i) = u(B_i)$ dla $1 \le i \le n$ (tzn. $v = t_{|S}$)

(tzn. iloczyn kartezjański relacji *r* i *s* jest zbiorem wszystkich możliwych kombinacji krotek tych relacji).

Równoważnie:

$t \in r \otimes s \Leftrightarrow t_{|R} \in r$ oraz $t_{|S} \in s$

W przypadku, gdy schematy relacji nie są rozłączne, najpierw zmieniamy nazwy atrybutów jednej z relacji (np. na $r.A_1, \dots, r.A_n$), a następnie stosujemy powyższą definicje.

Przykładowe własności:

$(r \cup s) \otimes q = (r \otimes q) \cup (s \otimes q)$

$\sigma_{F \wedge G}\left(r \otimes s\right) = \sigma_F(r) \otimes \sigma_G(s)$ jeśli $F(R)$, $F(S)$

$\pi_{X \cup Y}(r \otimes s) = \pi_X(r) \otimes \pi_Y(s)$ jeśli $X \subseteq R, Y \subseteq S$

Równoważności:

SELECT $C_1, \dots, C_k$ FROM *r,s* WHERE *F*	$\pi_{\{C_1,\dots,C_k\}}\left(\sigma_F(r \otimes s)\right)$

gdzie $\left\{C_1, \dots, C_k\right\} \subseteq \left\{A_1, \dots, A_n, B_1, \dots, B_m\right\}$

Złączenie naturalne

Dla dwóch relacji : *r* o schemacie $R = \{A_1, ..., A_n\}$ oraz *s* o schemacie $S = \{B_1, ..., B_m\}$ złączenie naturalne to relacja $q = r \oplus s$ o schemacie $Q = R \cup S$ taka, że:

$q = \{ t : \exists u \in r, v \in s \mid u_{|R} = t, v_{|S} = t \}$

(tzn. złączeniem naturalnym relacji *r* i *s* jest zbiór wszystkich możliwych połączeń krotek obu relacji, przy których ich wspólne atrybuty mają takie same wartości i nie są powtarzane.

Równoważna definicja:

$t \in r \oplus s \Leftrightarrow t_{|R} \in r$ oraz $t_{|S} \in s$

Uwagi:

Kolejność atrybutów w relacji jest nieistotna, podobnie jak kolejność krotek.

$R \cap S = \emptyset \Rightarrow r \oplus s = r \otimes s$

Złączenie naturalne da się również zdefiniować w następujący sposób:

$r \oplus s = \pi_{R \cup S}(\sigma_{r.D_1 = s.D_1 \wedge \ldots r.D_k = s.D_k}(r \otimes s))$ gdzie $R \cap S = \{D_1, \ldots, D_k\}$

Podstawowe własności złączenia naturalnego:

1. $q \oplus q = q$
2. $q \oplus r = r \oplus q$
3. $(q \oplus r) \oplus s = q \oplus (r \oplus s)$
4. $\pi_R(r \oplus s) \subseteq r$ oraz $\pi_S(r \oplus s) \subseteq s$
5. $q \subseteq \pi_R(q) \oplus \pi_S(q)$

Gdy zachodzi 4), złączanie nie traci informacji zawartych w *r* i *s*. Złączanie naturalne jednak nie wyklucza tracenia informacji (porównaj ze złączaniem zewnętrznym).

Gdy zachodzi 5), relacja q rozkłada się względem podziału na *R* i *S* z zachowaniem informacji. Tylko w takim przypadku można zastąpić w bazie danych relację *q* przez relacje $\pi_R(q)$ i $\pi_S(q)$.

Przykład:

Q={*A*,*B*,*C*}, *R*={*A*,*B*}, *S*={*B*,*C*}

$$q = \begin{array}{ccc} A & B & C \\ \hline a & b & c \\ a1 & b1 & c1 \end{array} \qquad \pi_R(q) = \begin{array}{cc} A & B \\ \hline a & b \\ a1 & b1 \end{array} \qquad \pi_S(q) = \begin{array}{cc} B & C \\ \hline b & c \\ b1 & c1 \end{array} \qquad \pi_R(q) \oplus \pi_S(q) = \begin{array}{ccc} A & B & C \\ \hline a & b & c \\ a & b & c1 \\ a1 & b & c \\ a1 & b1 & c1 \end{array} \neq q$$

Ocena poprawności relacji

Przykład:

relacja (zła): Dostawcy={Nazwa_dostawcy, Adres_dostawcy, Nazwa_towaru, Cena}

zależności funkcyjne: Nazwa_dostawcy $\rightarrow$ Adres_dostawcy

Nazwa_dostawcy, Nazwa_towaru $\rightarrow$ Cena

relacja (zła): Pracownicy={Id_pracownika, Nazwisko, Instytucja, Nazwa_instytucji, Adres_instytucji}
zależności funkcyjne: Id_pracownika $\rightarrow$ Nazwisko, Nazwa_instytucji
Nazwa_instytucji $\rightarrow$ Adres_instytucji

Ze schematem relacji wiąże się pewien zbiór zależności funkcyjnych F.
Zależność funkcyjna X $\rightarrow$Y wynika logicznie z zależności funkcyjnych F
(oznaczenie F|= X $\rightarrow$ Y), jeśli każda relacja r spełniająca wszystkie zależności w zbiorze F spełnia również zależność X $\rightarrow$ Y.

Jeśli
F= {Nazwa_dostawcy $\rightarrow$ Adres_dostawcy; Nazwa_dostawcy, Nazwa_towaru $\rightarrow$ Cena}
to
F|= Nazwa_dostawcy, Nazwa_towaru $\rightarrow$ Adres_dostawcy, Cena}

Domknięcie

Domknięcie zależności funkcyjnych F, oznaczane przez F^+ to zbiór wszystkich zależności funkcyjnych wynikających logicznie z zależności funkcyjnych F.
Przykładowo: A $\rightarrow$ C $\in$ {A $\rightarrow$ B; B $\rightarrow$ C}$^+$

Klucze – rozszerzona definicja

Nadkluczem relacji *r* o schemacie $R = \{A_1, ..., A_n\}$ i zbiorze zależności funkcyjnych F nazywamy dowolny zbiór atrybutów $X \in R$ taki, że: $X \rightarrow R \in F^+$

Kluczem relacji *r* o schemacie $R = \{A_1, ..., A_n\}$ i zbiorze zależności funkcyjnych F nazywamy dowolny minimalny nadklucz relacji r, tzn. dowolny zbiór atrybutów $X \in R$, taki, że:

1. $X \rightarrow R \in F^+$
2. dla żadnego $Y \subseteq X, X \neq Y$ nie zachodzi $Y \rightarrow R \in F^+$

Przykład:
R={Miasto, Ulica, Kod}, F={Miasto,Ulica $\rightarrow$ Kod; Kod $\rightarrow$ Miasto}
Kluczami tego schematu są: {Miasto, Ulica}, {Ulica, Kod}

$R=\{A,B,C,D\}$, $F=\{A \rightarrow C; B \rightarrow D\}$
Kluczem jest $\{A,B\}$

Zależność funkcyjna nie od klucza

Zależność funkcyjna $X \rightarrow Y$ jest zależnością od klucza, jeśli zbiór atrybutów X jest nadkluczem.
Zależność funkcyjna $X \rightarrow Y$ jest zależnością nie od klucza, jeśli:

1. jest nietrywialna (tzn. zbiór Y nie jest podzbiorem X)
2. nie jest zależnością od klucza

Zależności nie od klucza dla relacji Dostawcy: Nazwa_dostawcy $\rightarrow$ Adres_dostawcy

Zależności nie od klucza dla relacji Pracownicy: Nazwa_instytucji $\rightarrow$ Adres_instytucji

Algorytmy sprowadzania relacji do poprawnej postaci

Domknięcie tranzytywne

Dla zbioru $X \subseteq R$ jego domknięciem tranzytywnym względem zbioru zależności funkcyjnych F nazywamy następujący zbiór atrybutów: $X^+ = \{A \in R: X \rightarrow A \in F^+\}$
(tzn. zbiór wszystkich atrybutów, których wartości są w pełni zdeterminowane przez wartości atrybutów ze zbioru X)

Aby sprawdzić, czy dana zależność funkcyjna $X \rightarrow Y$ wynika logicznie ze zbioru zależności funkcyjnych F wystarczy wyznaczyć domknięcie tranzytywne zbioru X, tj. X^+. Zachodzi bowiem własność: $X \rightarrow Y \in F^+ \Leftrightarrow X \subseteq X^+$.

Algorytm wyznaczania domknięcia tranzytywnego:
Konstruujemy ciąg zbiorów:
1. $X_0 = X$
2. $X_{i+1} = X_i$ plus zbiór atrybutów A takich, że
   - istnieje pewna zależność funkcyjna $X \rightarrow Z \in F$
   - A należy do Z
   - $Y \subseteq X_i$

   dopóki nie zajdzie $X_i = X_{i+1}$.
3. Wtedy $X^+ = X_i$

Inaczej mówiąc: Jeśli w F istnieje zależność funkcyjna $Y \rightarrow Z$ oraz wszystkie atrybuty zbioru Y zostały już wygenerowane, wówczas mamy prawo wygenerować każdy z atrybutów zbioru Z.

Przykład:
$R=\{A,B,C,D,E,G\}$
$F=\{A,B \rightarrow C;\ D \rightarrow E,G;\ C \rightarrow A;\ B,E \rightarrow C;\ B,C \rightarrow D;\ C,G \rightarrow B,D;\ A,C,D \rightarrow B;\ C,E \rightarrow A,G\}$
$X=\{B,D\}$

Wówczas
$X0=\{B,D\}$
$X1=\{B,D,E,G\}$
$X2=\{B,C,D,E,G\}$
$X3=\{A,B,C,D,E,G\}=R$
Czyli $X^+=R$, co oznacza, że zbiór $\{B,D\}$ jest nadkluczem. Ponieważ $\{B\}^+=\{B\}$, $\{D\}^+=\{D,E,G\}$, więc $\{B,D\}$ jest kluczem.

Niech $Q=\{A_1, ..., A_n\}$ będzie schematem relacji, a F zbiorem zależności funkcyjnych. Usunięcie anomalii i redundancji relacji o schemacie Q odbywa się przez rozkład $Q = R \cup S$ na dwa schematy R i S.

Niech $\pi_Z(F) = \{ X \rightarrow Y \in F^+: X \cup Y \subseteq Z\}$ - rzut F na Z. Mówimy, że rozkład $Q = R \cup S$ zachowuje:

1. informacje, gdy dla każdej relacji q spełniającej zależności funkcyjne F zachodzi:

   $q = \pi_R(q) \oplus \pi_S(q)$
2. zależności funkcyjne, gdy

   $F^+ = (\pi_R(F) \cup \pi_S(F))^+$

   (czyli każda zależność funkcyjna $X \to Y \in F$ daje się wyprowadzić ze zbioru rzutów zależności $\pi_R(F) \cup \pi_S(F)$)

Własność: Rozkład $Q = R \cup S$ zachowuje informacje wtedy i tylko wtedy gdy

albo $(R \cap S) \to (R - S) \in F^+$ albo $(R \cap S) \to (S - R) \in F^+$

Przykład:
Relacja dostawcy i jej zależności funkcyjne
$Q=\{D,A,T,C\}$, $F=\{D \to A;\ D,T \to C\}$

Proponowany podział: $R = \{D,A\}$, $S = \{D,T,C\}$

Zauważmy, że: $R - S = \{A\}$, $R \cap S = \{D\}$, $S - R = \{T,C\}$.
A stąd:

ponieważ $D \to A \in F$, więc rozkład zachowuje informacje;

ponieważ $F = \pi_R(F) \cup \pi_S(F)$, więc rozkład zachowuje funkcyjne zależności.