Zależności funkcyjne

Zależności funkcyjne

Zależności funkcyjne między atrybutami są rodzajem warunków integralności.

Definicja 3. Niech będzie dany zbiór atrybutów SCH oraz jego podzbiory X i Y. Mówimy, że Y jest funkcyjnie zależny od X, co zapisujemy $X \rightarrow Y$, wtedy i tylko wtedy, gdy dla każdej relacji R rozpiętej na schemacie *SCH* i dla każdych dwóch krotek t_1, $t_2 \in \mathrm{R}$ jest spełniony warunek:

$t_1(X) = t_2(X) \Rightarrow t_1(Y) = t_2(Y)$.

Zależności funkcyjne

Dla zależności funkcyjnych sformułowano zbiór reguł wnioskowania, które pozwalają na wyprowadzenie nowych zależności na podstawie istniejących. Nazywamy je aksjomatami Armstronga.

AKSJOMATY ARMSTRONGA

A1. $Y \subseteq X \Rightarrow X \rightarrow Y$ (zwrotność)

A2. $X \rightarrow Y \wedge Z \subseteq W \Rightarrow XW \rightarrow YZ$ (powiększenie)

A3. $X \rightarrow Y \wedge Y \rightarrow Z \Rightarrow X \rightarrow Z$ (przechodniość)

AKSJOMATY ARMSTRONGA

Dowód

A1.

$t_1\,(X) = t_2\,(X) \wedge Y \subseteq X \Rightarrow t_1\,(Y) = t_2\,(Y)$

Zależności wynikające z aksjomatu zwrotności są często nazywane trywialnymi.

AKSJOMATY ARMSTRONGA

A2.

Dowód przez zaprzeczenie.

Załóżmy, że : $t_1\ (XW) = t_2\ (XW) \wedge t_1\ (YZ) \neq t_2\ (YZ)$.

$t_1\ (XW) = t_2\ (XW) \wedge Z \subseteq W \Rightarrow t_1\ (XZ) = t_2\ (XZ) \Rightarrow$

$t_1\ (X) = t_2\ (X) \wedge t_1\ (Z) = t_2\ (Z)$

$t_1\ (Z) = t_2\ (Z) \wedge t_1\ (YZ) \neq t_2\ (YZ) \Rightarrow t_1\ (Y) \neq t_2\ (Y)$

Otrzymaliśmy: $t_1\ (X) = t_2\ (X) \wedge t_1\ (Y) \neq t_2\ (Y)$,

co jest sprzeczne z założeniem.

AKSJOMATY ARMSTRONGA

A3.

$$(t_1 (X) = t_2 (X) \Rightarrow t_1 (Y) = t_2 (Y)) \wedge$$
$$(t_1 (Y) = t_2 (Y) \Rightarrow t_1 (Z) = t_2 (Z)) \Rightarrow$$
$$(t_1 (X) = t_2 (X) \Rightarrow t_1 (Z) = t_2 (Z))$$

REGUŁY ARMSTRONGA

Z aksjomatów Armstronga wynikają następujące reguły:

D1. $X \rightarrow Y \wedge X \rightarrow Z \Rightarrow X \rightarrow YZ$ (suma)

D2. $X \rightarrow Y \wedge WY \rightarrow Z \Rightarrow XW \rightarrow Z$ (pseudoprzechodniość)

D3. $X \rightarrow Y \wedge Z \subseteq Y \Rightarrow X \rightarrow Z$ (rozkład)

REGUŁY ARMSTRONGA

Dowód

D1. $X \rightarrow Y \wedge X \rightarrow Z \Rightarrow X \rightarrow YZ$

$X \rightarrow Y \Rightarrow X \rightarrow YX$ (aksjomat A2)

$X \rightarrow Z \Rightarrow XY \rightarrow ZY$ (aksjomat A2)

$X \rightarrow YX \wedge XY \rightarrow ZY \Rightarrow X \rightarrow YZ$ (aksjomat A3)

REGUŁY ARMSTRONGA

Dowód

D2. $X \to Y \wedge WY \to Z \Rightarrow XW \to Z$

$X \to Y \Rightarrow XW \to YW$ (aksjomat A2)

$XW \to YW \wedge YW \to Z \Rightarrow XW \to Z$ (aksjomat A3)

REGUŁY ARMSTRONGA

Dowód

D3. $X \to Y \wedge Z \subseteq Y \Rightarrow X \to Z$

$Z \subseteq Y \Rightarrow Y \to Z$ (aksjomat A1)

$X \to Y \wedge Y \to Z \Rightarrow X \to Z$ (aksjomat A3)

REGUŁY ARMSTRONGA

Zbiór reguł wnioskowania jest zupełny (sound)
 i kompletny (complete). Oznacza to, że wszystkie wyprowadzone zależności są poprawne oraz że można wyprowadzić wszystkie zależności istniejące
w danym schemacie relacji.

Konsekwencja logiczna

Oznaczmy przez F zbiór zależności funkcyjnych między atrybutami schematu SCH.
Zależność funkcyjna f jest konsekwencją logiczną F, co zapisujemy $F \models f$, jeśli f jest spełnione dla wszystkich relacji o schemacie SCH.

Domknięcie zbioru zależności funkcyjnych F^+

Jest to zbiór zależności funkcyjnych będących konsekwencjami logicznymi F

Nasycenie atrybutu X^+

Zbiór F^+ zawiera zazwyczaj wiele elementów, nawet jeśli F nie jest zbiorem dużym. Za pomocą reguł wnioskowania można bowiem wyprowadzić wiele zależności. Wyznaczanie F^+ jest więc procesem czasochłonnym. Znacznie łatwiej można wyznaczyć nasycenie atrybutu X^+ .

Nasycenie atrybutu X^+

Jest to zbiór atrybutów prostych A takich, że zależność $X \rightarrow A$ można wyprowadzić zgodnie z regułami wnioskowania.

Twierdzenie 1

Zależność $X \rightarrow Y$ można otrzymać na podstawie reguł wnioskowania $\Leftrightarrow Y \subseteq X^+$

Twierdzenie 1

Dowód

Załóżmy, że $Y = \{A_1, A_2, \ldots, A_n\}$

1. $Y \subseteq X^+$.

Zgodnie z definicją X^+ jest zbiorem atrybutów A_i, takich, że prawdziwa jest zależność $X \rightarrow A_i$. Na podstawie reguły sumy $X \rightarrow X^+$.

$Y \subseteq X^+ \Rightarrow X^+ \rightarrow Y$ (aksjomat A1)

$X \rightarrow X^+ \wedge X^+ \rightarrow Y \Rightarrow X \rightarrow Y$ (aksjomat A3)

Twierdzenie 1

2. $X \to Y$

$X \to Y \Rightarrow X \to A_i$. (D3)

Oznacza to, że $A_i \in X^+ \Rightarrow Y \subseteq X^+$

WYZNACZANIE NASYCENIA ATRYBUTU

1. Przyjmujemy $X^0 = X$

2. W każdym następnym kroku powiększamy X_i ,

$X^{i+1} = X^i \cup S$, o atrybuty należące do następującego

zbioru S: $S = \{A: \exists\ Y \rightarrow Z \wedge Y \subseteq X^i \wedge A \subseteq Z\}$.

Ze względu na to, że $X^i \subseteq X^{i+1}{}_{\dots} \subseteq U$ wnioskujemy,

że metoda jest zbieżna.

Proces wyznaczania X^+ kończymy, gdy $X^i = X^{i+1}$.

REDUKT INFORMACYJNY

A – zbiór atrybutów

$B \subseteq A$ jest reduktem informacyjnym w A

wtedy i tylko wtedy gdy B identyfikuje A – B

i nie istnieje podzbiór właściwy $B' \subset B$,

taki że B' identyfikuje A – B'.

REDUKT ASOCJACYJNY

A – zbiór atrybutów

$B_l, B_r \subseteq A$, $B_l \cap B_r = \varnothing$

B_l jest reduktem asocjacyjnym w A

wtedy i tylko wtedy gdy B_l identyfikuje B_r

i nie istnieje podzbiór właściwy $B_l' \subset B_l$,

taki że B_l' identyfikuje $(B_l - B_l') \cup B_r$.

REDUKT ASOCJACYJNY

A – zbiór atrybutów

$B_l, B_r \subseteq A, B_l \cap B_r = \varnothing$

Redukt asocjacyjny B_l jest nierozszerzalny w A, wtedy i tylko wtedy nie istnieje zbiór B_r', taki że $B_r' \supset B_r$ i $B_l \cap B_r' = \varnothing$ i B_l identyfikuje B_r' .

REDUKT ASOCJACYJNY

A – zbiór atrybutów

B_l, $B_r \subseteq A$, $B_l \cap B_r = \varnothing$

Redukt asocjacyjny B_l jest nieredukowalny w A, wtedy i tylko wtedy nie istnieje zbiór B_l', taki że $B_l' \subset B_l$ i B_l' identyfikuje B_r .

Lp	Outlook	Temp	Humidity	Wind	Sport
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cold	Normal	Weak	Yes
6	Rain	Cold	Normal	Strong	No
7	Overcast	Cold	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cold	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Jedynym reduktem informacyjnym jest

{Outlook, Temp, Humidity, Wind} → {Sport}

Redukty asocjacyjne:

{Outlook, Temp, Wind} → {Sport}

{Outlook, Humidity, Wind} → {Sport}

Są to redukty nierozszerzalne i nieredukowalne.

POKRYCIA ZBIORÓW ZALEŻNOŚCI

Zbiory zależności F i G są równoważne, jeśli $F^+ = G^+$.

Mówimy, że F pokrywa G (i G pokrywa F).

Zbiory są równoważne $\Leftrightarrow$ każda zależność z F należy do G^+ i każda zależność z G należy do F^+ .

Twierdzenie 2

Każdy zbiór zależności funkcyjnych F jest pokryty zbiorem zależności G, w którym nie istnieje prawa strona o więcej niż jednym atrybucie.

POKRYCIA ZBIORÓW ZALEŻNOŚCI

Dowód:

Niech $X \rightarrow Y \in F$, $Y = \{A_1, A_2, \ldots, A_n\}$.

Niech G będzie zbiorem zależności postaci $X \rightarrow A_i$.

Atrybuty A_i odpowiadają zależnościom $X \rightarrow Y \in F$.

Na podstawie D3 $X \rightarrow Y \Rightarrow X \rightarrow A_i \Rightarrow G \subseteq F^+$.

Na podstawie D1 $X \rightarrow A_1 \wedge X \rightarrow A_2 \wedge \ldots X \rightarrow A_n \Rightarrow X \rightarrow Y \Rightarrow F \subseteq G^+$

ZBIÓR MINIMALNY

Wyznaczenie F^+ nie jest konieczne. Wystarczy wyznaczyć zbiór minimalny, czyli taki z którego wynikają wszystkie zależności należące do F^+ .

ZBIÓR MINIMALNY

Zbiór zależności F jest minimalny jeśli:

1. Prawa strona każdej zależności w F jest pojedyńczym atrybutem
2. Zbiór F – $\{X \rightarrow A\}$ nie jest równoważny F
3. Zbiór F – $\{X \rightarrow A\} \cup \{Z \rightarrow A\}$, gdzie $Z \subset X$

nie jest równoważny F.

ZBIÓR MINIMALNY

Warunek 2 oznacza, że zbiór F nie zawiera zależności redundantnych.

Warunek 3 oznacza, że zbiór F nie zawiera zależności z atrybutami nadmiarowymi po lewej stronie.

RÓWNOWAŻNOŚĆ ZBIORÓW

Sprawdzanie równoważności zbiorów F i G.

1. $G \subseteq F^+$ F pokrywa G (każdą zależność ze zbioru G można wywnioskować na podstawie zbioru F)
2. $F \subseteq G^+$ G pokrywa F (każdą zależność ze zbioru F można wywnioskować na podstawie zbioru G)

RÓWNOWAŻNOŚĆ ZBIORÓW

Przy sprawdzaniu równoważności można wykorzystać nasycenie atrybutu

1. Dla każdej zależności $X \rightarrow Y \in F$ wyznaczyć X^+ względem zbioru G.
2. Sprawdzić czy $X^+ \supseteq Y$.
3. Jeżeli warunek ten jest spełniony dla każdej zależności $X \rightarrow Y \in F$, to G pokrywa F, czyli $F \subseteq G^+$.

WYZNACZANIE KLUCZA

Twierdzenie 3

Niech R oznacza relację o schemacie SCH.

Niech F oznacza zbiór zależności funkcyjnych między atrybutami schematu SCH.

$X \rightarrow A \in F^{+} \Rightarrow SCH - \{A\} \rightarrow SCH \in F^{+} \Rightarrow$

Dowód:

$(SCH - \{A\}) \cup X \rightarrow (SCH - \{A\}) \cup A \in F^{+}$

(aksjomat A2)

WYZNACZANIE KLUCZA

Przy wyznaczaniu klucza wykorzystujemy twierdzenie 3. Jako pierwsze przybliżenie przyjmujemy zbiór wszystkich atrybutów: $K = SCH$.
Następnie usuwamy poszczególne atrybuty sprawdzając czy $K - \{A\} \rightarrow SCH \in F^+$.
Algorytm kończy się, gdy nie istnieje możliwość usunięcia żadnego atrybutu.
Otrzymany wynik zależy od kolejności w jakiej rozpatrujemy poszczególne atrybuty.

WYZNACZANIE KLUCZA – UWAGI DODATKOWE

Przy wyznaczaniu kluczy można wykorzystać następujące własności:

1. Każdy klucz kandydujący zawiera wszystkie atrybuty występujące tylko po lewej stronie zależności funkcyjnych
2. Nie istnieje klucz kandydujący zawierający atrybuty występujące tylko po prawej stronie zależności funkcyjnych
3. Jeżeli zbiór atrybutów występujących tylko po lewej stronie zależności funkcyjnych identyfikuje pozostałe atrybuty,
to tworzy on jedyny klucz relacji.

ROZKŁAD do 3NF

1. Wyznaczyć zbiór minimalny
2. Dla zależności postaci $X \rightarrow A_i$ utworzyć schemat $\{X, A_1, A_2, \ldots, A_n\}$
3. Jeżeli żaden ze schematów nie zawiera klucza, utworzyć schemat, do którego należą atrybuty kluczowe

ZWIĄZKI WIELOARGUMENTOWE

1. Pracownik może brać udział w realizacji różnych projektów. Wymogiem formalnym jest podpisanie kontraktu z odpowiednim wydziałem. Pracownik może zawierać kontrakty z wieloma wydziałami.
2. Między wydziałem i pracownikiem może istnieć tylko jeden kontrakt
3. Pracownik może podpisać kontrakty z wieloma wydziałami na udział w realizacji tego samego projektu.

1. Pracownik może brać udział w realizacji różnych projektów. Wymogiem formalnym jest podpisanie kontraktu z odpowiednim wydziałem. Pracownik może zawierać kontrakty z wieloma wydziałami.
2. Między wydziałem i pracownikiem może istnieć tylko jeden kontrakt
3. Pracownik może podpisać kontrakty z wieloma wydziałami na udział w realizacji tego samego projektu.

$K(W, E, P)$ $\qquad$ $Q(K) = M:N:1$ $\qquad$ $WE \rightarrow P$

Związki trójargumentowe

4. Każdy projekt jest realizowany na określonym wydziale i wydział może realizować wiele projektów

Związki trójargumentowe

4. Każdy projekt jest realizowany na określonym wydziale i wydział może realizować wiele projektów

$P \rightarrow W \Rightarrow EP \rightarrow W$

Związki trójargumentowe

5. Każdy wydział może uczestniczyć w realizacji tylko jednego projektu oraz projekt może być realizowany przez wiele wydziałów

Związki trójargumentowe

5. Każdy wydział może uczestniczyć w realizacji tylko jednego projektu oraz projekt może być realizowany przez wiele wydziałów

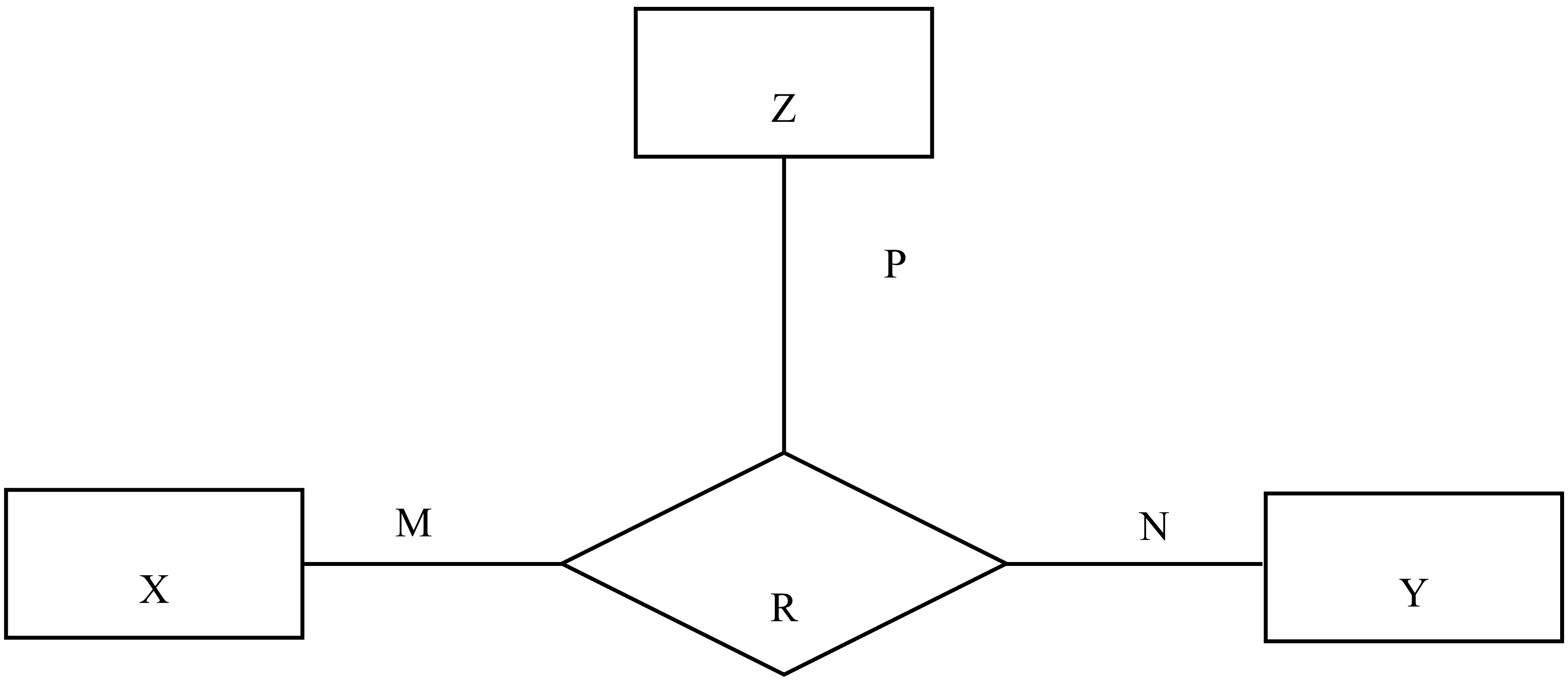
$W \rightarrow P \Rightarrow WE \rightarrow P$

Związki wieloargumentowe

$\mathbf{R}(X_1, X_2, \ldots , X_n)$

n – stopień związku, X_i – klucz i-tego zbioru

$\mathbf{Q}(X_1, X_2, \ldots , X_n) = M_1 : M_2 : \ldots : M_n$

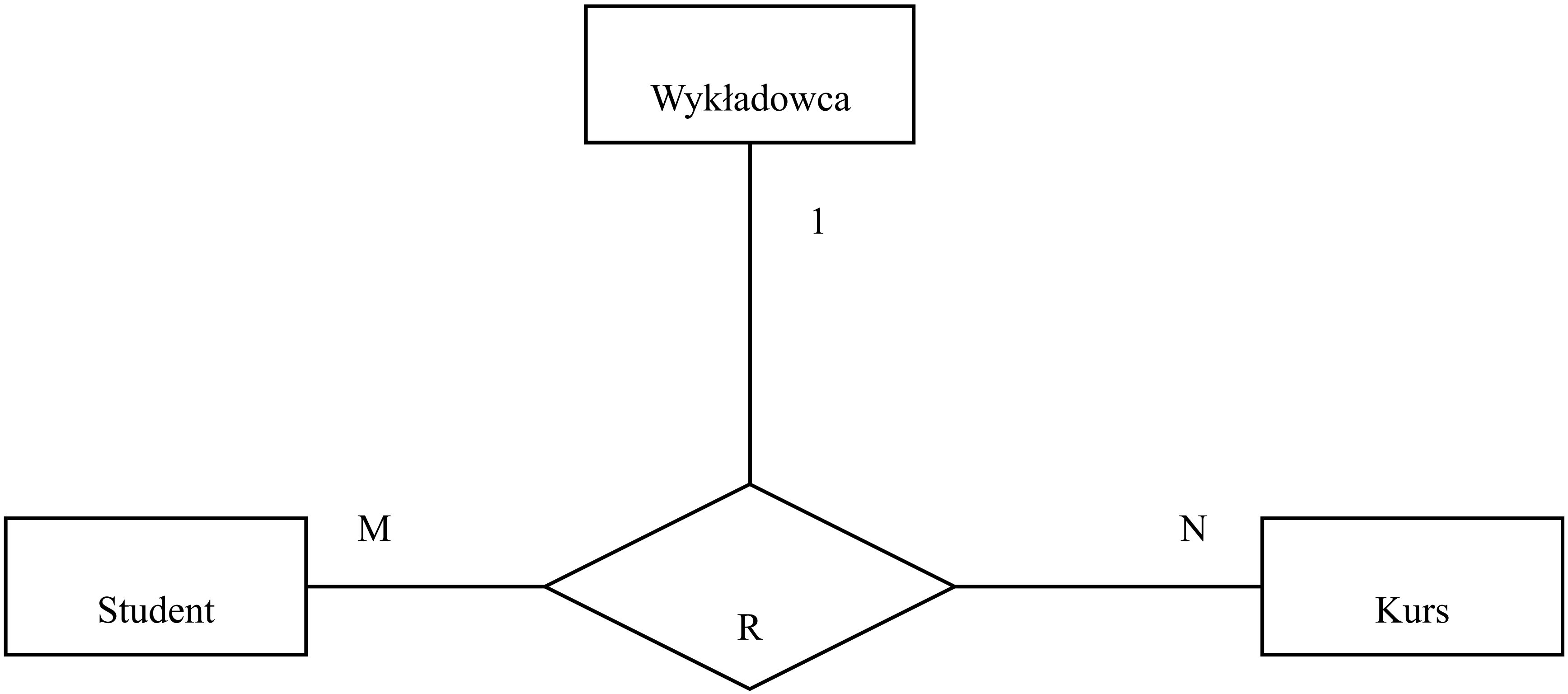

Z
P
M
X
R
N
Y

Związki wieloargumentowe

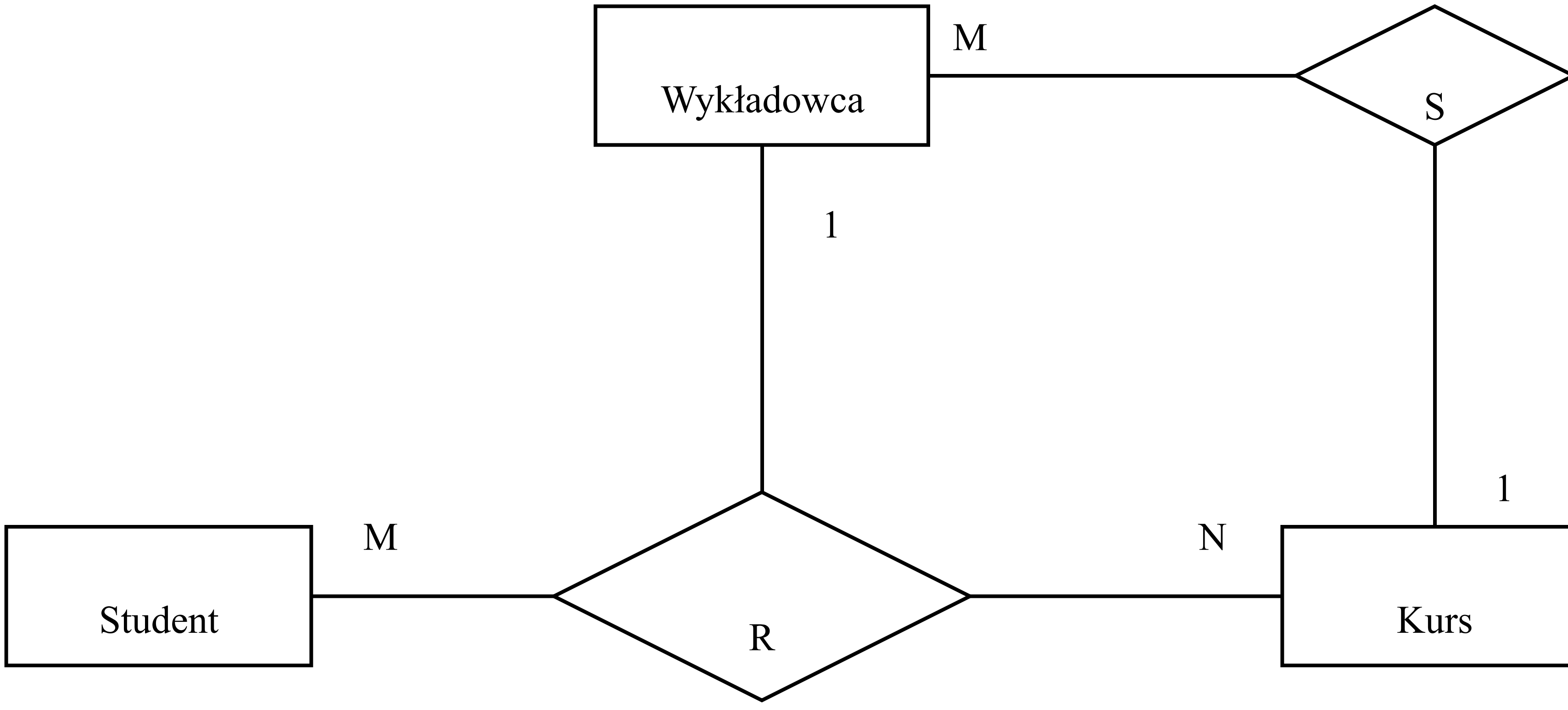
Związki trójargumentowe

1:1:1, a:1:1, a:b:1, a:b:c

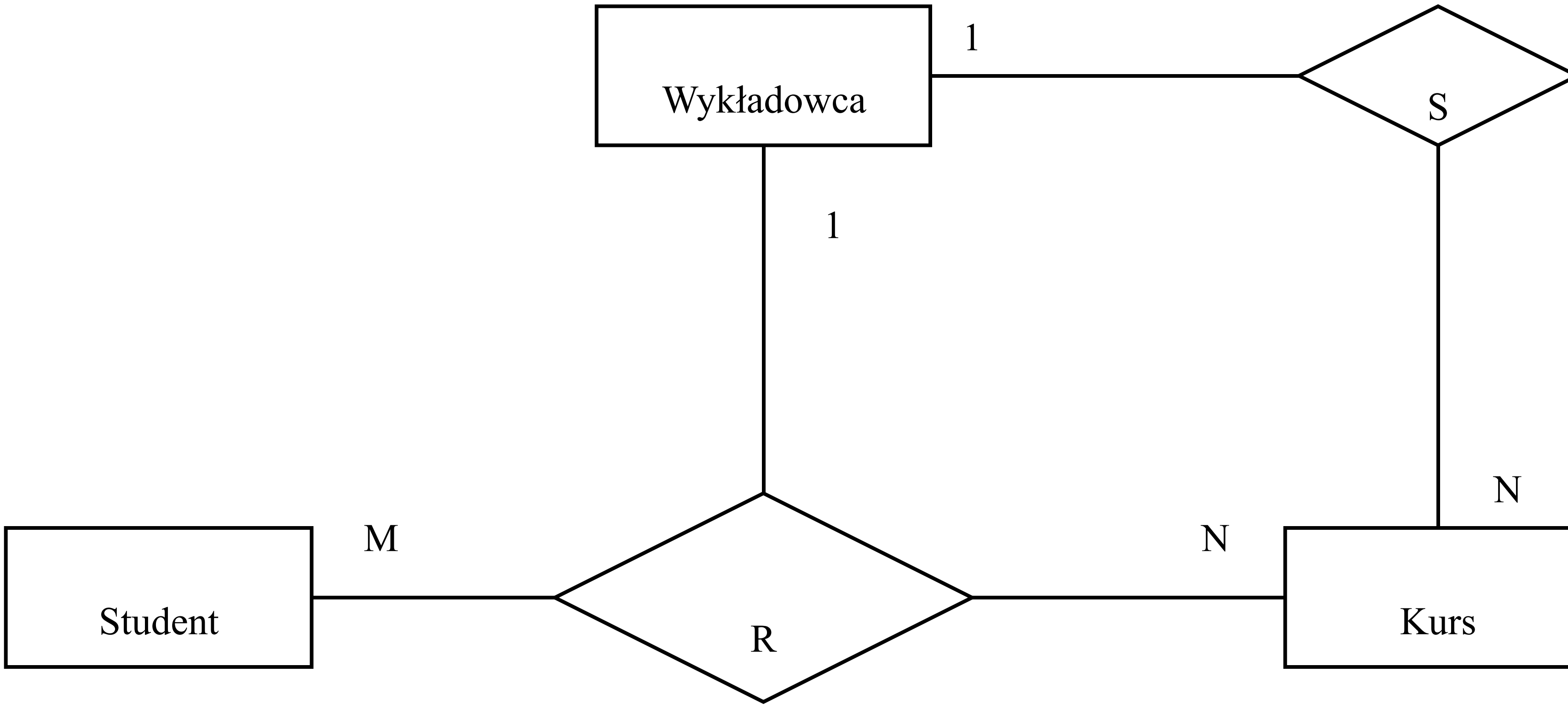
Kardynalności związków binarnych nie mogą mniejsze niż kardynalności związku trójargumentowego


Wykładowca
1
M
Student
R
N
Kurs

niedozwolone



dozwolone



Związki trójargumentowe

$XY \rightarrow Z$, $XZ \rightarrow Y$, $YZ \rightarrow X$ 1:1:1

(x1, y1, z1), (x2, y2, z2), (x1, y2, z3), (x3, y3, z3)

$XZ \rightarrow Y$, $YZ \rightarrow X$ 1:1:c

(x1, y1, z1) (x1, y1, z2), (x2, y1, z3), (x1, y2, z3)

$YZ \rightarrow X$ 1:b:c

(x1, y1, z1), (x1, y1, z2), (x1, y2, z1), (x2, y2, z2)

Brak a:b:c

(x1, y1, z1), (x1, y1, z2), (x1, y2, z1) , (x2, y1, z1)

Związki trójargumentowe

$XY \rightarrow Z, XZ \rightarrow Y, YZ \rightarrow X$ 1:1:1

$K1=XY, K2 = XZ, K3 = YZ$

$XZ \rightarrow Y, YZ \rightarrow X$ 1:1:c

$K1 = XZ, K2 = YZ$

$YZ \rightarrow X$ 1:b:c

$K=YZ$

Brak a:b:c

$K=XYZ$

Związki trójargumentowe

Kardynalność	Dopuszczalne zależności
1:1:1	każda
M:1:1	$X \rightarrow Y, X \rightarrow Z, Y \rightarrow Z, Z \rightarrow Y$
M:N:1	$X \rightarrow Z, Y \rightarrow Z$
M:N:P	brak

Związki wieloargumentowe

$\mathbf{R}(X_1, X_2, \ldots, X_n)$ $U = \{X_1, X_2, \ldots, X_n\}$

$U - \{X_i\} \rightarrow X_i$, $i = 1, 2, \ldots, n$ (2)

$U - \{X_i, X_j\} \rightarrow X_i$, $i \neq j, i, j = 1, 2, \ldots, n$ (3)

F – zbiór zależności (2)

L – zbiór atrybutów występujących tylko po lewej stronie zależności (2)

P – zbiór pozostałych atrybutów

Związki wieloargumentowe

Twierdzenie

W związku n-argumentowym $\mathbf{R}(X_1, X_2, \ldots , X_n)$
ze zbiorem F zależności funkcyjnych między
n atrybutami o postaci

$$U - \{X_i\} \rightarrow X_i \ , \ \ i = 1, 2, \ldots, n \qquad (2)$$

może istnieć zależność funkcyjna

$$U - \{X_i, X_j \} \rightarrow X_i , \ \ i \neq j, i, j = 1, 2, \ldots, n \quad (3)$$

jeżeli atrybut X_i nie należy do zbioru L atrybutów
występujących tylko po lewej stronie zależności (2).

Związki wieloargumentowe

Dowód

$U - \{X_i, X_j\} \rightarrow X_i \Rightarrow U - \{X_i\} \rightarrow X_i$

Jeśli $X_i \in P$, to $(U - \{X_i\} \rightarrow X_i) \in F$.

Jeśli $X_i \in L$, to $(U - \{X_i\} \rightarrow X_i) \notin F$.

ZWIĄZKI TRÓJARGUMENTOWE

$R(X, Y, Z)$

$XY \to Z,\ XZ \to Y,\ YZ \to X$

Narzucana zależność: $X \to Z$

$X \to Z \wedge XZ \to Y \Rightarrow X \to Y$

Nowy klucz X

Zbiór minimalny $X \to Z,\ X \to Y,\ YZ \to X$

Postać BCNF

ZWIĄZKI TRÓJARGUMENTOWE

R(X, Y, Z)
$XY \rightarrow Z$, $XZ \rightarrow Y$

Narzucana zależność: $X \rightarrow Z$
$X \rightarrow Z \wedge XZ \rightarrow Y \Rightarrow X \rightarrow Y$

Nowy klucz X
Zbiór minimalny: $X \rightarrow Z$, $X \rightarrow Y$
Postać BCNF

ZWIĄZKI TRÓJARGUMENTOWE

R(X, Y, Z)

$XY \rightarrow Z$, $XZ \rightarrow Y$

Narzucana zależność: $Y \rightarrow Z$

Nie ma nowego klucza.

Zbiór minimalny: $Y \rightarrow Z$, $XZ \rightarrow Y$

Postać 3NF