4 Teoria relacyjnych baz danych (cz.1)

4.1 Zależności funkcyjne

Def.1 Mówimy, że dla relacji $R = A_1A_2 \ldots A_k$ zachodzi zależność funkcyjna $\alpha \to \beta$, gdzie $\alpha, \beta \subseteq \{A_1, A_2, \ldots, A_k\}$, jeżeli dla każdego stanu r relacji R zachodzi:

$$\forall(t_1, t_2 \in r)((\forall A \in \alpha)(t_1.A = t_2.A)) \Rightarrow ((\forall B \in \beta)(t_1.B = t_2.B)).$$

$\Box_{Def.1}$

Przykład 1: Dla relacji

$$Wróżka = (pesel, nazwisko, imię, dataUr, miasto, ulica, dom, kodPocz, tel, zodiak)$$

zachodzą zależności funkcyjne[1]:

$$\begin{array}{rcl}
pesel & \to & (nazwisko, imię, dataUr), \\
dataUr & \to & zodiak, \\
(miasto, ulica, dom) & \to & kodPocz, \\
kodPocz & \to & miasto, \\
(kodPocz, ulica, dom) & \to & tel, \\
?(nazwisko, imię, dataUr, miasto, ulica, dom) & \to & pesel.
\end{array}$$

Natomiast *nie zachodzą* zależności:

$$\begin{array}{rcl}
(nazwisko, imię) & \to & (miasto, ulica, dom), \\
(miasto, ulica, dom) & \to & nazwisko, \\
kodPocz & \to & (ulica, miasto).
\end{array}$$

$\Box_{Prz.1}$

[1]Warto zauważyć, że zależności funkcyjne zależą od specyfiki zagadnienia rzeczywistego. Na przykład, wiele z podanych w przykładzie jest dyskusyjnych. Zależność pierwszą możemy przyjąć tylko wtedy, gdy zamierzamy wszystkie osoby identyfikować przez PESEL lub kod, który na pewno nie będzie pokrywał się z PESEL-em (np. numer paszportu może nie mieć tej własności). Założenie drugie może przyjąć tylko kiepska wróżka, która nie zważa na godzinę urodzenia i dokładny moment przejścia Słońca w kolejny znak Zodiaku. Założenie piąte można przyjąć, gdy wróżka planuje dla każdego adresu zapisać tylko jeden numer kontaktowy. Ostatnia zależność, to przykład założenia całkowicie zbędnego, mało ryzykownego, bo pewnie takie dwie osoby się nie trafią, ale... (Przyp. 2009)

Def.2

- *Kluczem relacji* R nazywamy taki podzbiór K atrybutów R, że:
  1. $K \to R$,
  2. $(\forall L, L \subset K)$ nie zachodzi $L \to R$.
- *Nadkluczem relacji* R nazywamy dowolny zbiór atrybutów R zawierający podzbiór właściwy będący kluczem R.
- Jeden z kluczy relacji nazywamy *kluczem głównym*.[2]
- Pozostałe klucze, to *klucze kandydujące*, inaczej *alternatywne*.
- *Atrubutem głównym* R nazywamy atrybut należący do dowolnego (!!!) klucza R.

$\Box_{Def.2}$

Przykład 2: W relacji z poprzedniego przykładu kluczami są:

- $K_1 = (pesel)$[3]
- $K_2 = (nazwisko, imię, dataUr, miasto, ulica, dom)$,
- $K_3 = (nazwisko, imię, dataUr, kodPocz, ulica, dom)$

przykładowe nadklucze, to:

- $(pesel, zodiak)$,
- $(nazwisko, imię, dataUr, miasto, ulica, dom, kodPocz)$.

Atrybuty główne, to: *pesel*, *nazwisko*, *imię*, *dataUr*, *miasto*, *ulica*, *dom*, *kodPocz*. $\Box_{Prz.2}$

[2]Wybór klucza ma zazwyczaj znaczenie dla rzeczywistej sprawności bazy danych. Zazwyczaj za klucz główny wybieramy zbiór atrybutów najczęściej używany w praktyce do identyfikacji krotek. Według klucza wiele SZBD organizuje sobie dane tabeli, stąd wyszukiwanie według klucza głównego bywa zazwyczaj szybsze. Klucz główny jest także używany do identyfikacji krotki w innych tabelach (jako klucz obcy), powinien więc to być ten najbardziej naturalny i najłatwiejszy w obsłudze klucz. W szczególności, gdy wszystkie klucze tabeli są złożone z wielu atrybutów, warto rozważyć wprowadzenie „sztucznego” *identyfikatora krotki.* (Przyp. 2009)

[3]Z przyczyn wyjaśnionych w poprzednim przypisie, to jest dobry kandydat na klucz główny. (Przyp. 2009)

4.2 Reguły wyprowadzania zależności funkcyjnych

Dla α, β, γ bedących podzbiorami atrybutów pewnej relacji R zachodzą zależności (trzy pierwsze nazywamy *Aksjomatami Armstronga*):

Zwrotność (zal.trywialne) $(\beta \subseteq \alpha) \Rightarrow (\alpha \to \beta)$;

Rozszerzanie $(\alpha \to \beta) \Rightarrow (\alpha\gamma \to \beta\gamma)$,

Przechodniość $(\alpha \to \beta \wedge \beta \to \gamma) \Rightarrow (\alpha \to \gamma)$,

Sumowanie $(\alpha \to \beta \wedge \alpha \to \gamma) \Rightarrow (\alpha \to \beta\gamma)$,

Rozkładanie $(\alpha \to \beta\gamma) \Rightarrow (\alpha \to \beta \wedge \alpha \to \gamma)$.

Def.3 Niech F będzie zbiorem zależności funkcyjnych relacji R i niech α będzie podzbiorem atrybutów R. *Domknięciem* F nazywamy zbiór F^+ wszystkich zależności funkcyjnych dotyczących atrybutów R i wyprowadzalnych z F.
Dwa zbiory zależności F i G nazywamy *równoważnymi*, jeżeli ich domknięcia są równe, tzn. $F^+ = G^+$.
Domknięciem α *względem* F (oznaczamy je $(\alpha)^+_F$) nazywamy maksymalny względem zawierania zbiór atrybutów $\beta \subseteq R$, taki że $\alpha \to \beta \in F^+$. $\Box_{Def.3}$

Uwaga 1: Aksjomaty Armstronga są *zupełnym*, *niesprzecznym* i *minimalnym* zbiorem reguł pozwalającym wyprowadzić ze zbioru F każdą zależność funkcyjną prawdziwą w każdym stanie relacji spełniającym F.

Uwaga 2: Pomimo, że aksjomaty Armstronga mogą stanowić podstawę algorytmicznej procedury wyznaczania F^+, to jednak takiej procedury nie opiszemy. Powód jest prosty. Nawet dla małych zborów F, zbiór F^+ może być bardzo duży i w większości nieciekawy. W szczególności wszystkich zależności trywialnych jest wkładniczo wiele, a nie niosą one interesujących informacji. W dalszej części wskażemy, jak obywać się bez wyznaczania F^+.

Uwaga 3: Aksjomaty Armstronga stanowią jednak podstawę efektywnego algorytmu wyznaczania domknięcia zbioru atrybutów względem zbioru zależności funkcyjnych, czyli $(\alpha)^+_F$. Zbiór χ to wyznaczany zbiór atrybutów wyprowadzalnych z α, czyli zachodzi niezmiennik $\alpha \to \chi$.

- $\chi \leftarrow \alpha$ (zwrotność);
- dopóki χ zmienia się:
  - znajdź zbiór $\beta \subseteq \chi$, taki że w F jest zależność $\beta \to \gamma$ i $\gamma \not\subseteq \chi$;
  - $\chi \leftarrow \chi \cup \gamma$, czyli zastosuj rozszerzanie do $\beta \to \gamma$ i przechodniość do $\alpha \to \chi$ i $\chi \to \gamma$;

Przykład 3: Dla $R = ABCD$ i $F = \{A \to B, BC \to D, D \to A\}$ kilka zależności z domknięcia F, to:

$$F^+ = F \cup \{AB \to BA, ABC \to A, ABCD \to A, AC \to AC, \ldots, AC \to D, BC \to A\}.$$

Następnie

$$(A)^+_F = AB \text{ oraz} (BC)^+_F = ABCD.$$

Zbiory zależności F i $G = \{A \to B, B \to D, D \to A\}$ nie są równoważne, ponieważ $(B)^+_F = B$ a $(B)^+_G = BDA$. $\Box_{Prz.3}$

Uwaga 4: Testowanie równoważności zbiorów zależności funkcyjnych F i G przeprowadza się poprzez sprawdzenie, czy dla każdej zależności $\alpha \to \beta \in F$ zachodzi $\beta \subseteq (\alpha)^+_G$. Taka własność dowodzi, że $F \subseteq G^+$. Sprawdzenie zawierania również w drugą stronę dowodzi, że $F^+ = G^+$.

Uwaga 5: Zależności funkcyjne dla bazy danych powinny być sprawdzane przez SZBD. Aby działał on wydajnie, zbiór tych zależności powinien być jak najmniej liczny. Zbiór ten redukujemy zazwyczaj do pewnego minimalnego pokrycia, tzn.podzbioru $F_{min} \subseteq F$, takiego że $(F)^+ = (F_{min})^+$ i F_{min} jest stosunkowo mały.

4.3 Postać normalna Boyce-Codda

Def.4 Rozważmy relację R ze zbiorem zależności funkcyjnych F. Przedstawmy F tak, by wszystkie zależności miały po prawej stronie po jednym atrybucie (reguła rozkładania). Relacja R jest w *postaci normalnej Boyce'a-Codda (BCNF)*, jeżeli:

$$(\forall \alpha \to B \in F) \ \ ((B \in \alpha) \ \vee \ ((\alpha)^+_F = R)),$$

tzn. każda zależność z F jest albo trywialna, albo wynika z nadklucza.[4] $\Box_{Def.4}$

Przykład 4:

- Relacja $R = ABCDE$ (A – ImNaz, B – DataUr, C – Adres, D – Zodiak, E – Tel) z zależnościami

  $$F = \{AB \to C, B \to D, C \to E, E \to C\}$$

  nie jest w BCNF. Kluczem R jest AB, więc wszystkie zależności, oprócz pierwszej, naruszają BCNF. Zależność $B \to D$ nazywamy częściową (wynika z części klucza). Zależność $C \to E$ nazywamy przechodnią (jej lewa strona nie należy do klucza).

- Relacja $R = ABC$ z $F = \{AB \to C\}$ jest w BCNF.

[4] Zauważmy, że gdy relacja jest w BCNF, SZBD ma bardzo ułatwione zadanie. Kontrola zależności funkcyjnych sprowadza się *tylko i wyłącznie* do kontroli własności klucza, a więc unikalności wartości w pewnej kolumnie czy kolumnach. (Przyp. 2009)

- Dla relacji $R = KMU$ z zależnościami

$$F = \{MU \to K, K \to M\},$$

klucze to $K_1 = MU$ i $K_2 = KU$, wszystkie atrybuty są więc główne. Zależność $K \to M$ narusza BCNF.

$\square_{Prz.4}$

4.4 Sprowadzenie (rozkład) relacji do postaci normalnej

Def.5

- *Rozkładem relacji* R nazywamy relacje R_1, R_2,..., R_k, takie że $R = R_1 \cup R_2 \cup \cdots \cup R_k$.
- Jeżeli F jest zbiorem zależności funkcyjnych R, to zbiorami zależności funkcyjnych dla $R_1, \ldots, R_k$ są odpowiednio $F_1, \ldots, F_k$, gdzie $F_i = \{\alpha \to \beta \in F^+ \mid \alpha, \beta \in R_i\}$ jest nazywane *rzutem* F na R_i, co oznaczamy $\pi_{R_i}(F)$.[5]
- Jeżeli r jest stanem R, to stanami $R_1, \ldots, R_k$ są odpowiednio $r_1, \ldots, r_k$, gdzie $r_i = \{t^{[R_i]} \mid (\exists s \in R)(s.R_i = t.R_i)\}$ jest nazywane *rzutem* (stanu) R na R_i.
- Operacją przeciwną do rozkładu jest *złączenie naturalne* oznaczane $\bowtie$. Dla relacji R_1 i R_2 ich złączeniem naturalnym jest relacja R o schemacie $R_1 \cup R_2$. Jeżeli r_1 i r_2 są odpowiednio stanami R_1 i R_2, to:
  $r_1 \bowtie r_2 = \{t^{[R_1 \cup R_2]} \mid ((\exists t_1 \in r_1)(t_1 = t.R_1)) \wedge ((\exists t_2 \in r_2)(t_2 = t.R_2))\}$.
- Rozkład R na $R_1, \ldots, R_k$ nazywamy *odwracalnym*, jeżeli dla każdego poprawnego (tzn. spełniającego zależności F) stanu r relacji R zachodzi

$$r = \pi_{R_1}(r) \bowtie \pi_{R_2}(r) \bowtie \cdots \bowtie \pi_{R_k}(r),$$

  tzn. po zrzutowaniu stanu relacji na składowe i następnie złączeniu naturalnym otrzymanych stanów składowych, dostajemy wyjściowy stan relacji.[6]
- Rozkład R na $R_1, \ldots, R_k$ *zachowuje zależności*, jeżeli $\pi_{R_1}(F) \cup \cdots \cup \pi_{R_k}(F)$ i F są sobie równoważne.[7] $\square_{Def.5}$

[5] W przedstawionej definicji rzutu bardzo ważny jest znaczek +. Rzut zbioru zależności, to wszystkie zależności z *domknięcia* F, które dotyczą atrybutów składowej R_i, a nie *tylko* zależności wybrane żywcem z F dotyczące atrybutów składowej R_i. Wyznaczanie F_i wymaga zazwyczaj sporo pracy, a nie tylko rzutu okiem na F i R_i. (Przyp. 2009)

[6] Nie wolno nam rozkładać relacji na składowe, jeśli nie mamy zagwarantowanej odwracalności. Wykonanie tej operacji, gdy nie jest odwracalna, powoduje, że nie będziemy potrafili odtworzyć stanu oryginalnej relacji, nawet poprzez złączenie naturalne — zazwyczaj dorzuci ono sporo krotek, których nie było w relacji oryginalnej przed rozkładem. (Przyp. 2009)

[7] Relację można rozłożyć na składowe bez zachowania zależności, ale wówczas należy utrzymywać dodatkowy zbiór zależności *globalnych*, które mogą dotyczyć atrybutów z różnych tabel. Kontrolowanie takich zależności jest trudne i czasochłonne, ale nie niemożliwe. Stąd przy rozkładzie relacji na składowe odwracalność jest warunkiem koniecznym, a zachowanie zależności — mocno zalecanym. (Przyp. 2009)

Lemat 1: Niech R będzie relacją, a F jej zbiorem zależności funkcyjnych. Jeżeli $\alpha \to \beta \in F^+$ jest nietrywialna (tzn. $\alpha \cap \beta = \emptyset$), to rozkład $R_1 = \alpha \cup \beta$, $R_2 = R \setminus \beta$ jest odwracalny.

Dowód: patrz tablica.

Lemat 2: Dla każdej relacji R istnieje odwracalny rozkład na składowe będące w BCNF.

Dowód: Dopóki R ma składową niebędącą w BCNF:

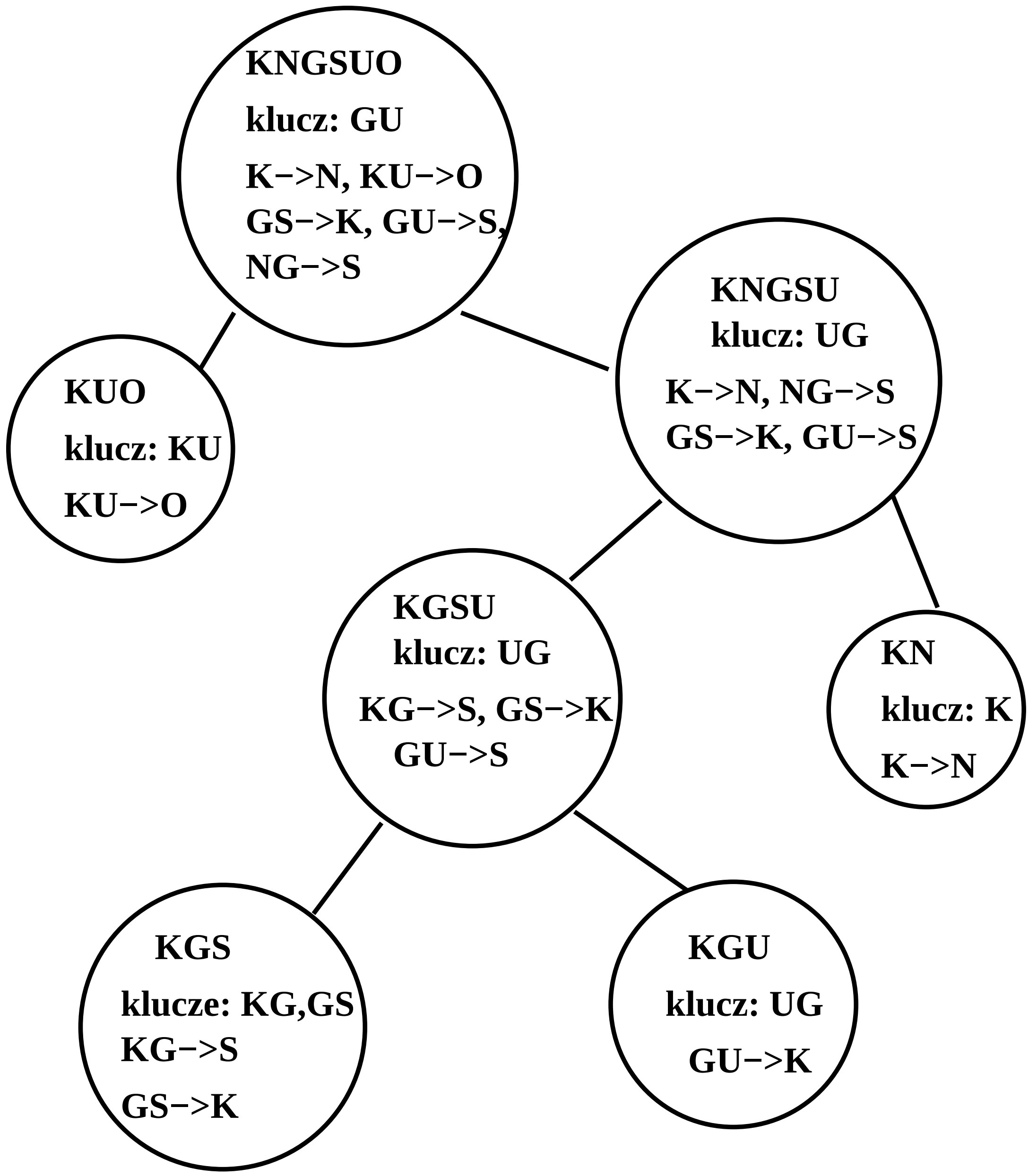
1. Wybierz złą składową R_i oraz zależność $\alpha \to \beta$ z $F_i = \pi_{R_i}(F)$ naruszającą BCNF.
2. Rozłóż R_i na składowe $R_i' = \alpha \cup \beta$ oraz $R_i'' = R_i \setminus \beta$.
3. Wyznacz F_i' oraz F_i'' będące rzutami F_i na R_i' oraz R_i''. $\square_{Lem.2}$

Lemat 3: Istnieją relacje, dla których nie istnieje odwracalny i zachowujący zależności rozkład do BCNF.

Dowód: Rozważmy $R = KMU$ i $F = \{K \to M, MU \to K\}$. R nie jest w BCNF. Gdy (dowolnie) rozłożymy relację R, to tracimy zależność $MU \to K$. $\square_{Lem.3}$

Przykład 5:

R=KNGSUO, F={K−>N, KU−>O, GS−>K, GU−>S, NG−>S}



Uwaga: Dla danej relacji algorytm z lematu 3 może dać raz rozkład zachowujący zależności, a raz nie zachowujący, w zależności od wyboru zależności wyznaczających rozkład relacji w kolejnych krokach.