Bazy danych 2022

(częściowo) na podstawie slajdów Przemysławy Kanarek

4 kwietnia 2022

Dobry projekt bazy danych

NULL-e — krotki powinny pasować do schematu tabeli,

Redundancja — informacja nie powinna być zapisywana wielokrotnie,

Kontrola więzów — sprawdzanie własności klucza, unikalności i innych więzów powinno być łatwe,

Obliczanie złączeń — jest trudne, więc nie należy rozdrabniać zbytnio bazy.

Dobry projekt bazy danych - anomalie

Rozważmy tabelę: GR(idg, idk, prow_nazwisko, prow_biuro, lista_zapisanych_w_json)

złożone typy — lista_zapisanych_w_json → Zapisy(ids, idg, data_zapisu) **1NF**

Codd: *to permit data to be queried and manipulated using a "universal data sub-language"*

anomalia UPDATE — zmiana biura prowadzącego → zmiana wielu krotek (redundacja, niejednoznaczność)

anomalia INSERT — jak dodać nowego prowadzącego (bez grup)? Null?

anomalia DELETE — co po usunięciu ostatniej grupy prowadzącego?

Definicja i przykład

Definition (Zależność funkcyjna)

Dla relacji $R = A_1A_2 \ldots A_k$ oraz zbiorów jej atrybutów $\alpha, \beta \subseteq \{A_1A_2 \ldots A_k\}$ zachodzi zależność funkcyjna $\alpha \to \beta$, jeżeli dla każdego stanu r relacji R zachodzi:

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

W relacjach PR(ido,tytuł,nazwisko,adres), ST(ido,indeks,nazwisko,adres) i GR(idg,idk,idp,limit), TS(idg,termin,sala) zachodzą zależności

- w PR: ido $\to$ nazwisko,adres,tytuł,
- w ST: ido $\to$ nazwisko,adres,indeks oraz indeks$\to$ ido,nazwisko,adres;
- w GR: idg$\to$ idk,idp,limit
- w TS: termin,sala$\to$ idg

Spostrzeżenie: Jeśli K jest kluczem R, to $K \to R$.

Klucz relacji

Definition (Klucz relacji)

Kluczem relacji R nazywamy taki podzbiór K jej atrybutów, który:

- wyznacza funkcyjnie wszystkie atrybuty R, czyli $K \to R$ oraz
- jest minimalnym zbiorem o tej własności, czyli $(\forall L \subsetneq K)\neg(L \to R)$

Definition

Nadklucz relacji — dowolny zbiór atrybutów zawierający klucz relacji,

Klucz główny — jeden z kluczy relacji,

Klucz alternatywny — klucz relacji inny niż klucz główny,

Atrybut główny — atrybut (dowolnego) klucza relacji.

W relacji ST(ido,indeks,nazwisko,adres) w relacji TS(idg,termin,sala)

- Kluczem jest *indeks* i kluczem jest *ido*,
- Nadkluczami są: $\{indeks, nazwisko\}$ lub $\{ido, indeks, adres\}$,
- Jako klucz główny możemy wybrać *indeks*,
- Atrybuty główne to $\{ido, indeks\}$
- W relacji TS kluczem jest $\{termin, sala\}$

Przykład

- `Mafia(Miasto, Gang, Proceder)`
- ξ_1: `Gang` $\rightarrow$ `Miasto` oraz ξ_2: `Miasto, Proceder` $\rightarrow$ `Gang`.
- Co można wywnioskować o organizacji mafii wiedząc, że zachodzą ξ_1 i ξ_2?
- Jakie są klucze relacji `Mafia`?
- Anomalie?
- Podaj odwracalny rozkład relacji `Mafia` do postaci BCNF(!?).
- Rozkład?
- Odwracalny?
- BCNF?
- Czy rozkład ten zachowuje zależności?
- Jakie rodzi to problemy?

Reguły dla zależności funkcyjnych

Definition (Aksjomaty Armstronga i in.)

Dla relacji R i zbiorów jej atrybutów $\alpha, \beta, \gamma \subseteq R$ zachodzi:

- Zwrotność $(\beta \subseteq \alpha) \Rightarrow (\alpha \to \beta)$ (zależności trywialne)
- 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 $(\alpha \to \beta\gamma) \Rightarrow (\alpha \to \beta \wedge \alpha \to \gamma)$

Definition (Domknięcie zbioru zależności i zbioru atrybutów)

Dla relacji R i jej zbioru zależności funkcyjnych F:

- F^+ — **domknięciem zbioru zależności** F nazywamy zbiór wszystkich zależności wyprowadzalnych z F.
- $(\alpha)^+_F$ — **domknięciem zbioru atrybutów** $\alpha \subseteq R$ **względem** F nazywamy zbiór atrybutów, które można wyprowadzić z α za pomocą F.

Twierdzenie o znaczeniu Aksjomatów Armstronga

Definition (Aksjomaty Armstronga)

Dla relacji R i zbiorów jej atrybutów $\alpha, \beta, \gamma \subseteq R$ zachodzi:

Zwrotność $(\beta \subseteq \alpha) \Rightarrow (\alpha \to \beta)$ (zależności trywialne)

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

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

Theorem

Aksjomaty Armstronga stanowią zupełny, niesprzeczny i minimalny zbiór reguł pozwalający wyprowadzić ze zbioru zależności F każdą zależność funkcyjną prawdziwą w każdym stanie relacji, w którym spełnione są reguły F.

Wniosek

Tym samym zależności dowodliwe za pomocą aksjomatów Armstronga to zależności prawdziwe.

Uwagi ad. domknięcia zbiorów zależności i atrybutów

1. Zazwyczaj nie wyznaczamy F^+ — to zbiór duży i zawierający dużo nieciekawych informacji (np. zależności trywialne).
2. Efektywne wyznaczanie $(\alpha)^+_F$ jest potrzebne — pozwala zadecydować, co jest kluczem, czy w relacji jest redundancja itp.
3. Mamy algorytm, który pozwala wyznaczać $\chi = (\alpha)^+_F$:
   - $\chi \leftarrow \alpha$ (zwrotność)
   - dopóki χ zmienia się:
     - znajdź $\beta \in \chi$ taki, że istnieje $\beta \rightarrow \gamma \in F$ oraz $\gamma \setminus \chi \neq \emptyset$
     - $\chi \leftarrow \chi \cup \gamma$ (zastosuj rozszerzanie i przechodniość)
   - zwróć χ jako wynik
4. Mamy sposób, by porównywać zbiory zależności F i G dla tej samej relacji. Sprawdzamy, czy $F^+ = G^+$:
   - dla każdej zależności $\alpha \rightarrow \beta \in F$:
     - oblicz $\chi = (\alpha)^+_G$
     - jeśli $\beta \subseteq \chi$, to $\alpha \rightarrow \beta \in G^+$, w przeciwnym wypadku $F^+ \neq G^+$
   - powtórz to dla każdej zależności $\alpha \rightarrow \beta \in G$ i zbioru F
5. Zależności funkcyjne powinny być kontrolowane przez SZBD. Dlatego dobrze, by było ich mało. Zbiór F_{min} nazwiemy minimalnym pokryciem F jeśli jest równoważny F i nie zawiera zależności "nadmiarowych".

Postać normalna Boyce-Codda

Definition (Postać normalna Boyce-Codda, BCNF)

Relacja R ze zbiorem zależności funkcyjnych F jest w postaci normalnej Boyce-Codda, jeśli dla każdej nietrywialnej zależności $\alpha \to \beta$ $(\alpha \cap \beta = \emptyset)$ zbiór α jest nadkluczem.

Uwagi:

1. Relacja w BCNF ma tylko zależności trywialne i wynikające z nadklucza.
2. Kontrola zależności funkcyjnych w relacji w BCNF sprowadza się do kontroli własności klucza.

Przykłady:

1. `Mafia(Miasto, Gang, Proceder)`,
2. `Gang` $\to$ `Miasto`, `Miasto, Proceder` $\to$ `Gang`
3. `Lokalizacje(Miasto, Gang)`, `Procedery(Gang, Proceder)` są w BCNF

Rozkład relacji

- Rozkładem relacji R nazywamy zbiór relacji $\{R_1, \ldots, R_k\}$ taki, że $R = R_1 \cup \ldots \cup R_k$.
- Dla F — zbioru zależności R, rzutem F na R_i jest $F_i = \{\alpha \to \beta \in F^+ \mid \alpha, \beta \in R_i\}$.
- Dla r — stanu relacji R, stanem R_i jest $r_i = \pi_{R_i}(r)$.
- Złączenie naturalne jest operacją przeciwną do rozkładu.
- Rozkład R na $R_1, \ldots, R_k$ jest odwracalny, jeśli dla każdego poprawnego stanu r (spełniającego zależności F) zachodzi:

$$r \;=\; r_1 \bowtie r_2 \bowtie \cdots \bowtie r_k$$

- Rozkład R na $R_1, \ldots, R_k$ zachowuje zależności, jeśli:

$$F^+ \;=\; (F_1 \cup F_2 \cup \ldots \cup F_k)^+$$

- Rozkład relacji na składowe MUSI być odwracalny i POWINIEN zachowywać zależności.

Rozkład relacji do BCNF

Lemma

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

Lemma

*Każda relacja ma **odwracalny** rozkład na składowe w BCNF.*

Lemma

*Istnieją relacje, które nie mają **odwracalnego i zachowującego zależności** rozkładu na składowe w BCNF.*

Przykład

`Lokalizacje(Miasto, Gang)`, `Zajęcia(Gang, Proceder)` vs. `Miasto, Proceder` $\to$ `Gang`

Przykład rozkładu do BCNF

Dane:

- $R = KNGSUO$
- $F = \{K \to N, KU \to O, GS \to K, GU \to S, NG \to S\}$
- klucz R to GU

Rozkład:

R: R nie jest w BCNF. Rozkładamy wg $KU \to O$ na $R_1 = KUO$ i $R_2 = KNGSU$;

R_1 : $R_1 = \underline{KU}O$, $F_1 = \{KU \to O\}$ (BCNF)

R_2: $R_2 = KN\underline{G}S\underline{U}$, $F_2 = \{\underline{K \to N}, NG \to S, GS \to K, GU \to S\}$ (Nie-BCNF)

R_{21}: $R_{21} = K\underline{G}S\underline{U}$, $F_{21} = \{\underline{KG \to S}, GS \to K, GU \to S\}$. (Nie-BCNF)

R_{211}: $R_{211} = KGS$, $F_{211} = \{KG \to S, GS \to K\}$, klucze: KG i GS. (BCNF)

R_{212}: $R_{212} = K\underline{GU}$, $F_{212} = \{GU \to K\}$. (BCNF)

R_{22}: $R_{22} = \underline{K}N$, $F_{22} = \{K \to N\}$ (BCNF).

Wynik rozkładu: $R = KUO \cup KGS \cup KGU \cup KN$ i $\{KU \to O, KG \to S, GS \to K, GU \to K, K \to N\}$

Trzecia postać normalna

Definition (Trzecia postać normalna, 3NF)

Relacja R z zależnościami funkcyjnymi F jest w trzeciej postaci normalnej, jeśli każda zależność $\alpha \to B \in F$

- jest trywialna ($B \in \alpha$) albo
- wynika z nadklucza ($(\alpha)_F^+ = R$) albo
- ma po prawej stronie atrybut główny (B należy do jakiegoś klucza).

Lemma

Każda relacja ma odwracalny i zachowujący zależności rozkład na składowe w postaci 3NF.

Algorytm rozkładu do 3NF

Definition (F_{min})

Minimalnym pokryciem zbioru zależności funkcyjnych F nazwiemy równoważny F zbiór F_{min}, w którym:

- nie ma zależności trywialnych, np. $AB \rightarrow AC$
- nie ma zależności nadmiarowych, czyli wynikających z pozostałych zależności F_{min}, np. $A \rightarrow B, B \rightarrow C, A \rightarrow C$
- nie ma atrybutów lewostronnie nadmiarowych $AB \rightarrow C, A \rightarrow B$.

Algorytm rozkładu do 3NF

1. Wyznacz F_{min}.
2. Dla każdej zależności $\alpha \rightarrow \beta \in F_{min}$ utwórz składową $R_i = \alpha\beta$. Usuń składowe zawierające się w innych.
3. Jeśli żadna z utworzonych składowych nie zawiera klucza R, to dodaj do rozkładu składową K dla pewnego klucza K relacji R.

Przykład c.d.

- `Mafia(Miasto, Gang, Proceder, Szef)`
- ξ_1: `Gang` $\to$ `Miasto`, ξ_2: `Miasto, Proceder` $\to$ `Gang`, ξ_3: `Gang` $\to$ `Szef`
- Jakie są klucze relacji `Mafia`?
- Podaj odwracalny i zachowujący zależności rozkład relacji `Mafia` do postaci 3NF.
- `Miasta(Gang, Miasto)`, `Mafia(Miasto, Gang, Proceder)`, `Szefowie(Gang, Szef)`
- Dlaczego zachowuje zależności?
- Dlaczego jest odwracalny? Wskazówka: Zacznij od krotki z relacji z kluczem
- cf. BCNF: `Miasta(Miasto, Gang)`, `Procedery(Gang, Proceder)`, `Szefowie(Gang, Szef)`