

Bazy danych i systemy zarządzania

Wykład IV

Operacje algebraiczne w relacyjnych bazach danych

Część II

Wyszukiwanie danych: operacja selekcji

Operacja **selekcji** jest podstawową operacją realizowaną w relacyjnych bazach danych. Jest ona stosowana do wyszukiwania danych spełniających zadane *kryteria* (warunki). Operację selekcji można więc traktować jako sposób *inetynsjonalnego* definiowania danych (interesującej nas relacji) w uniwersum zadawanym *ekstensjonalnie* przez relację wyjściową.

Niech R będzie relacją o schemacie $R(A_1, A_2, \dots, A_n)$. Zauważmy, że atrybuty A_1, A_2, \dots, A_n można traktować jak *zmiennne*, które dla poszczególnych rekordów przybierają określone wartości atomiczne. Niech Φ oznacza pewną formułę logiczną zbudowaną ze:

- stałych (w tym liczb i łańcuchów znakowych),
- atrybutów A_1, A_2, \dots, A_n pełniących rolę zmiennych; w formule Φ są one zmiennymi wolnymi (stosowana notacja: $[A_i]$),
- funkcji wbudowanych w systemie (określonych na stałych, zmiennych oraz wartościach funkcji),
- symboli relacji wbudowanych w systemie (np. $<$, $>=$, etc.),
- spójników logicznych dopuszczalnych w systemie, oraz
- nawiasów definiujących kolejność operacji.

Formuła Φ definiuje **kryterium selekcji**. Poprawnie zdefiniowana formuła Φ definiuje jednoznacznie relację R_Φ , taką, że:

$$R_\Phi = \{(d_1, d_2, \dots, d_n) \in R : \Phi\{A_1/d_1, A_2/d_2, \dots, A_n/d_n\}\}. \quad (1)$$

Zapis $\{A_1/d_1, A_2/d_2, \dots, A_n/d_n\}$ oznacza podstawienie stałych d_1, d_2, \dots, d_n za zmienne A_1, A_2, \dots, A_n . Operacja selekcji definiowana jest poprzez zadanie formuły Φ a jej wynik poprzez równość (1).

Operacja selekcji: definiowanie i własności

Z formułą Φ można jednoznacznie związać funkcję selekcji σ_Φ , określoną jak następuje:

$$\sigma_\Phi : R \longrightarrow R_\Phi.$$

Funkcja selekcji (wyboru) może być także interpretowana jako funkcja charakterystyczna poszukiwanej relacji, tzn.:

$$\sigma_\Phi(d_1, d_2, \dots, d_n) = \begin{cases} 1 & \text{jeżeli } (d_1, d_2, \dots, d_n) \in R_\Phi \\ 0 & \text{jeżeli } (d_1, d_2, \dots, d_n) \notin R_\Phi \end{cases}$$

Tak więc przy zadanej wyjściowej relacji R oraz warunku selekcji Φ , operacja selekcji może być zapisana w postaci:

$$\sigma_\Phi(R) = R_\Phi,$$

gdzie R_Φ jest zbiorem tych wszystkich n -krotek relacji R , które spełniają warunek Φ , $R_\Phi \subseteq R$.

W postaci logicznej operację selekcji można też zdefiniować jako:

$$\|\sigma_\Phi(R)\| = \|R\| \wedge \|\Phi\|.$$

Własności operacji selekcji:

- zostaje zachowany schemat relacji, oraz
- w odniesieniu do wyselekcjonowanych krotek nie jest tracona informacja (o ile jednocześnie nie jest zastosowana operacja projekcji, zostaje zachowany ich pełny opis),
- zazwyczaj znacznemu zmniejszeniu ulega liczba rekordów (w odniesieniu do relacji wyjściowej),
- wynikiem operacji selekcji może być zbiór zawierający 0, 1, lub wiele krotek relacji wyjściowej.

Przykład operacji selekcji

Rozważmy relację R zadaną tablicą Pracownik:

ID_prac	Nazwisko	Imię	Data ur	Stanowisko	Dział	Stawka
MT101	Abacki	Adam	61-01-01	robotnik	P10	550,00 zł
MT102	Abakowski	Alojzy	61-01-02	robotnik	P10	574, 00 zł
MT103	Adamski	Antoni	61-01-03	robotnik	P20	1275,00 zł
MT104	Adamski	Arnold	61-01-03	robotnik	P20	1280,00 zł
MT105	Adamski	Arnold	61-01-03	robotnik	P20	1295,00 zł
KT101	Aron	Antonina	61-01-03	robotnik	P10	575,00 zł
MU101	Batman	Bogusław	67-02-13	kierownik	P30	1224,00 zł
KU101	Celińska	Mirosława	69-03-08	analityk	F10	975,00 zł
MV101	Dioniziak	Dariusz	71-10-17	v-prezes	V	3000,00 zł

Niech $\Phi_1 = \text{Nazwisko} = \text{'Adamski'}$; wówczas $\sigma_1(R)$ jest tablicą postaci:

ID_prac	Nazwisko	Imię	Data ur	Stanowisko	Dział	Stawka
MT103	Adamski	Antoni	61-01-03	robotnik	P20	1275,00 zł
MT104	Adamski	Arnold	61-01-03	robotnik	P20	1280,00 zł
MT105	Adamski	Arnold	61-01-03	robotnik	P20	1295,00 zł

Podobnie, $\Phi_2 = (\text{Stanowisko} = \text{'robotnik'} \text{ AND } \text{Stawka} < 1000,00 \text{ zł}) \text{ OR } (\text{Stanowisko} = \text{'analityk'} \text{ AND } \text{Stawka} < 1000,00 \text{ zł})$ da wynik $\sigma_2(R)$:

ID_prac	Nazwisko	Imię	Data ur	Stanowisko	Dział	Stawka
MT101	Abacki	Adam	61-01-01	robotnik	P10	550,00 zł
MT102	Abakowski	Alojzy	61-01-02	robotnik	P10	574, 00 zł
KT101	Aron	Antonina	61-01-03	robotnik	P10	575,00 zł
KU101	Celińska	Mirosława	69-03-08	analityk	F10	975,00 zł

Definiowanie warunku selekcji

Z logicznego punktu widzenia, warunek selekcji Φ musi być *poprawnie zdefiniowaną formułą*, której wartość logiczna jest określona dla wszystkich rekordów tabeli wyjściowej. W praktyce, definiowanie kryterium selekcji odbywa się z wykorzystaniem:

- predefiniowanego szablonu użytkownika,
- właściwego dla systemu szablonu *Query by Example*,
- języka SQL (instrukcji wprowadzonej ręcznie, generowanej automatycznie lub z poziomu języka wyższego rzędu,
- specyficznego dla systemu języka zapytań (np. VTLIS, Datalog, Prolog).

Typowe elementy definiowania warunku selekcji obejmują:

- użycie stałych:
 - numerycznych, np. 13, +13, -13, 13,57,
 - tekstowych, np. 'Kowalski', "Kowalski",
 - logicznych, *True*, *False* (-1, 0),
 - datowych, np. #99-08-13#, '1999-08-13',
- nazw atrybutów i wyrażeń ścieżkowych, np. Nazwisko, [Data ur], Pracownik.Nazwisko,
- operatorów:
 - arytmetycznych, np. +, -, *, /,
 - tekstowych, np. +, &, %, _
 - logicznych, np. *and*, *or*, *not*, *xor*,
 - relacyjnych, np. =, <, >, <=, >=, <>, !=,
 - specjalnych (porównania), np. *Like*, *Between*,
- funkcji, np. *Abs(.)*, *Sqr(.)*, *Count(.)*,

Idea szablonu QBE

Definiowanie szablonu QBE

Aby zdefiniować szablon zapytania używając języka QBE należy określić:

- **Atrybuty** – zdefiniować, które pola mają być uwzględnione w definiowanym zapytaniu; podaje się też ich kolejność,
- **Źródło atrybutu** – tabela lub zapytanie z którego pochodzi dany atrybut (uwaga: w zapytaniu mogą występować atrybuty o takich samych nazwach, ale pochodzące z różnych źródeł),
- **Sortowanie** – czy i jak wyniki zapytania mają być posortowane,
- **Wyświetlanie** – czy dane pole ma być wyświetlane,
- **Kryterium selekcji** – formuły definiujące warunek selekcji, a więc *intensjonalną* definicję relacji docelowej.

Schemat szablonu QBE

<i>Pole :</i>	A_1	A_2	...	A_j	...	A_k
<i>Tabela :</i>	T_1	T_2	...	T_j	...	T_k
<i>Sortuj :</i>			...	$r/m/b$...	
<i>Pokaż :</i>			...	y/n	...	
<i>Kryteria :</i>	k_1	k_2	...	k_j	...	k_k
<i>Lub :</i>	k'_1	k'_2	...	k'_j	...	k'_k

(2)

W wyniku działania kwerendy otrzymuje się tabelę o schemacie $K(A_1, A_2, \dots, A_k)$; w zależności od tego, ile rekordów spełnia określone kryteria, w tabeli wynikowej może być 0, 1, lub wiele rekordów. Mogą one podlegać dalszemu przetwarzaniu.

System ACCESS – użycie wyrażeń

Definiowanie wyrażeń

Wyrażenia to definicje funkcji określające sposób wyliczania pewnych wartości; mogą one być konstruowane za pomocą *stałych*, atrybutów (pełniących rolę zmiennych), operatorów (arytmetycznych, logicznych, etc.) oraz predefiniowanych (w systemie) funkcji.

Stałe obejmują:

- **stałe numeryczne**, np. 13, +13, -13, +13,34, 13.35,
- **stałe tekstowe**, np. "Kowalski", "Kraków", "Jan Kowalski",
- **stałe logiczne** – PRAWDA/True oraz FAŁSZ/False; *liczbowymi odpowiednikami są -1 dla prawdy oraz 0 dla fałszu*,
- **stałe datowe** – *określają wartość daty i występują w ogranicznikach #, np. #07-22-98#*.

Operatory obejmują:

- **operatory arytmetyczne** – dodawanie (+), odejmowanie (-), mnożenie (*), dzielenie (/), potęgowanie (^),
- **operatory relacyjne** – =, >, <, <=, >=, <>, BETWEEN, IS NULL, IS NOT NULL, ich wynik jest wartością logiczną (PRAWDA lub FAŁSZ); mogą być stosowane do porównywania liczb (dat, walut) i tekstów,
- **operatory tekstowe** – +, &, są to operatory konkatencji tekstów,
- **operatory datowe** – odejmowanie dat (-) (wynik w dniach), modyfikacja daty (+, -), wynikiem jest data,
- **operatory logiczne** – koniunkcja (AND), alternatywa (OR) i negacja (NOT).

Operacja θ -złączenia (join)

Technologia relacyjnych baz danych umożliwia łączenie informacji zawartej w wielu tabelach; połączenia definiuje się pomiędzy rekordami dwóch wybranych tabel. Najbardziej ogólną operacją łączenia tabel jest operacja θ -złączenia, gdzie θ oznacza warunek dopuszczający połączenie rekordów.

Niech $R(A_1, A_2, \dots, A_n)$ oraz $S(B_1, B_2, \dots, B_m)$ będą dowolnymi relacjami, przy czym $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$. Niech θ będzie warunkiem logicznym określonym na atrybutach zbioru $\mathbf{A} \cup \mathbf{B}$ (niekoniecznie wszystkich); będziemy pisać $\theta(\mathbf{A}, \mathbf{B})$. Wynikiem operacji θ -złączenia relacji R i S jest relacja $T = R \bowtie_{\theta} S$ zdefiniowana w następujący sposób:

$$R \bowtie_{\theta(\mathbf{A}, \mathbf{B})} S = \pi_{\mathbf{A} \cup \mathbf{B}} \sigma_{\theta(\mathbf{A}, \mathbf{B})}(R \times S)$$

Tak więc, operacja θ -złączenia pozwala połączyć te rekordy tabeli R oraz S , które spełniają warunek θ dla wartości wybranych atrybutów (selekcja przy warunku $\theta(\mathbf{A}, \mathbf{B})$); ponadto, w tabeli wynikowej powtarzające się kolumny są pomijane (projekcja na atrybuty zbioru $\mathbf{A} \cup \mathbf{B}$). Jeżeli warunek θ jest zawsze prawdziwy (lub nie został określony), a zbiory \mathbf{A} i \mathbf{B} są rozłączne, to operacja θ -złączenia daje jako wynik iloczyn kartezjański relacji R oraz S .

Jeżeli warunek θ jest określony na pojedynczych atrybutach, np. A_i oraz B_j , to operację θ -złączenia relacji R oraz S można zapisać w postaci takiej jak $R \bowtie_{A_i \theta B_j} S$ lub $R \bowtie_{i \theta j} S$.

Niech B^1, B^2, \dots, B^k będą kolejnymi atrybutami zbioru \mathbf{B} nie należącymi do \mathbf{A} . Wówczas logiczna definicja operacji θ -złączenia ma postać:

$$\|(R \bowtie_{\theta(\mathbf{A}, \mathbf{B})} S)(d_1, \dots, d_n, d^1, \dots, d^k)\| = \|R(d_1, \dots, d_n)\|$$

$$\wedge \|S(d_{n+1}, \dots, d_{n+m})\| \wedge \theta(\mathbf{A}, \mathbf{B}),$$

przy czym $d^j = d_{n+i}$ dla $j = 1, 2, \dots, k$, $i \in \{1, 2, \dots, m\}$.

Przykład operacji θ -złączenia

Rozważmy relację R zadaną tablicą Pracownik:

Nazwisko	Imię	Data ur	Stawka
Abacki	Adam	61-01-01	550,00 zł
Abakowski	Alojzy	61-01-02	574, 00 zł
Adamski	Antoni	61-01-03	1275,00 zł
Adamski	Arnold	61-01-03	1280,00 zł
Adamski	Arnold	61-01-03	1295,00 zł
Aron	Antonina	61-01-03	575,00 zł
Batman	Bogusław	67-02-13	1224,00 zł
Celińska	Mirosława	69-03-08	975,00 zł
Dioniziak	Dariusz	71-10-17	3000,00 zł

oraz tablicę Nagroda

Data ur	Stawka1	Stawka2	Nagroda
61-01-03	500,- zł	1200,- zł	300,- zł
72-01-01	2999,- zł	3001,-zł	3000,- zł

przy czym θ niech będzie warunkiem postaci $\text{Pracownik.[Data ur]} \leq \text{Nagroda.[Data ur]}$ AND $\text{Pracownik.Stawka} > \text{Nagroda.Stawka1}$ AND $\text{Pracownik.Stawka} < \text{Nagroda.Stawka2}$. Wynikiem operacji θ -złączenia będzie tabela

Nazwisko	Imię	Data ur	Stawka	Stawka1	Stawka2	Nagroda
Abacki	Adam	61-01-01	550,00 zł	300,- zł	1300,- zł	300,- zł
Abakowski	Alojzy	61-01-02	574, 00 zł	300,- zł	1300,- zł	300,- zł
Aron	Antonina	61-01-03	575,00 zł	300,- zł	1300,- zł	300,- zł
Dioniziak	Dariusz	71-10-17	3000,00 zł	2999,- zł	3001,- zł	3000,- zł

Równozłączenie

Jednym z najczęstszych przypadków operacji θ -złączenia jest *operacja równozłączenia*. W przypadku równozłączenia warunek θ ma postać równości wybranych atrybutów obu tabel.

Niech $R(A_1, A_2, \dots, A_n)$ oraz $S(B_1, B_2, \dots, B_m)$ będą dowolnymi relacjami, przy czym $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$. Wynikiem operacji równozłączenia relacji R i S po atrybutach $A^1, A^2, \dots, A^k \in \mathbf{A}$ ze schematu relacji R oraz atrybutach $B^1, B^2, \dots, B^k \in \mathbf{B}$ ze schematu relacji S jest relacja

$$R \bowtie_{A^1=B^1 \wedge A^2=B^2 \wedge \dots \wedge A^k=B^k} S = \pi_{\mathbf{A} \cup \mathbf{B}} \sigma_{A^1=B^1 \wedge A^2=B^2 \wedge \dots \wedge A^k=B^k} (R \times S)$$

Tak więc, operacja równozłączenia pozwala połączyć te rekordy tabeli R oraz S dla których wartości wybranych atrybutów są równe. W tabeli wynikowej powtarzające się kolumny są pomijane (projekcja na atrybuty zbioru $\mathbf{A} \cup \mathbf{B}$). Zauważmy, że dla realizacji operacji równozłączenia nie musi istnieć wspólny atrybut dla obu relacji (może być $\mathbf{A} \cap \mathbf{B} = \emptyset$).

Jeżeli warunek θ jest określony na pojedynczych atrybutach, np. A_i oraz B_j , to operację θ -złączenia relacji R oraz S można zapisać w postaci takiej jak $R \bowtie_{A_i \theta B_j} S$ lub $R \bowtie_{i \theta j} S$.

Operacja równozłączenia jest szczególnym (ale najczęściej spotykanym) przypadkiem operacji θ -złączenia.

Maksymalna liczba elementów w relacji wynikowej dla operacji θ -złączenia wynosi $card(R) \cdot card(S)$. Jeżeli atrybuty definiujące połączenie z relacji R tworzą klucz (lub indeks jednoznaczny), to maksymalna liczba elementów relacji wynikowej wynosi $card(S)$; jeżeli oba zbiory atrybutów definiujących połączenie stanowią klucze, to maksymalna liczba elementów relacji wynikowej wynosi $\min(card(R), card(S))$.

Złączenie naturalne (natural join)

Operacja złączenia naturalnego stanowi z kolei szczególny przypadek operacji równozłączenia. Polega ona na połączeniu dwóch relacji według wartości wspólnych atrybutów.

Niech $R(A_1, A_2, \dots, A_n)$ oraz $S(B_1, B_2, \dots, B_m)$ będą dowolnymi relacjami, przy czym $\mathbf{A} = \{A_1, A_2, \dots, A_n\}$, $\mathbf{B} = \{B_1, B_2, \dots, B_m\}$. Istotnym założeniem pozwalającym na realizację złączenia naturalnego jest istnienie wspólnych atrybutów w łączonych relacjach, tzn.:

$$\mathbf{A} \cap \mathbf{B} \neq \emptyset.$$

Niech zatem $\mathbf{A} \cap \mathbf{B} = \{A^1, A^2, \dots, A^k\}$. Wynikiem operacji równozłączenia relacji R i S po atrybutach A^1, A^2, \dots, A^k jest relacja

$$R \bowtie_{R.A^1=S.A^1 \wedge \dots \wedge R.A^k=S.A^k} S = \pi_{\mathbf{A} \cup \mathbf{B}} \sigma_{R.A^1=S.A^1 \wedge \dots \wedge R.A^k=S.A^k} (R \times S)$$

Tak więc, operacja złączenia naturalnego pozwala połączyć te rekordy tabeli R oraz S dla których wartości *wszystkich wspólnych* atrybutów są równe. W tabeli wynikowej powtarzające się kolumny są pomijane (projekcja na atrybuty zbioru $\mathbf{A} \cup \mathbf{B}$). Zauważmy, że dla realizacji operacji złączenia naturalnego musi istnieć przynajmniej jeden wspólny atrybut dla obu relacji; ponadto, jeżeli znane są schematy obu relacji, to zbiór wszystkich wspólnych atrybutów jest zadany jednoznacznie i nie trzeba go podawać jawnie. Stosowane są wtedy zapisy:

$$R \bowtie S$$

$$R * S.$$

Operacja złączenia naturalnego jest szczególnym (ale najczęściej spotykanym) przypadkiem operacji równozłączenia. Maksymalna liczba elementów w relacji wynikowej jest ograniczana identycznie jak w przypadku równozłączenia.

Przykład operacji złączenia naturalnego

Rozważmy ponownie tablice dostawców i odbiorców materiałów budowlanych.

<i>Dostawca</i>	<i>Produkt</i>		<i>Produkt</i>	<i>Odbiorca</i>
<i>Budex</i>	<i>cegła</i>	oraz	<i>cegła</i>	<i>Wykbud</i>
<i>Budex</i>	<i>pustak</i>		<i>pustak</i>	<i>Wykbud</i>
<i>Budex</i>	<i>cement</i>		<i>cement</i>	<i>Rembud</i>
<i>Budex</i>	<i>piasek</i>		<i>cement</i>	<i>Wykbud</i>
<i>Matbud</i>	<i>pustak</i>		<i>piasek</i>	<i>Rembud</i>
<i>Matbud</i>	<i>cement</i>		<i>piasek</i>	<i>Wykbud</i>
<i>Matbud</i>	<i>gips</i>		<i>gips</i>	<i>Rembud</i>
<i>Matbud</i>	<i>piasek</i>			

W wyniku operacji złączenia naturalnego dostaniemy tabelę postaci:

<i>Dostawca</i>	<i>Produkt</i>	<i>Odbiorca</i>
<i>Budex</i>	<i>cegła</i>	<i>Wykbud</i>
<i>Budex</i>	<i>pustak</i>	<i>Wykbud</i>
<i>Budex</i>	<i>cement</i>	<i>Rembud</i>
<i>Budex</i>	<i>cement</i>	<i>Wykbud</i>
<i>Budex</i>	<i>piasek</i>	<i>Rembud</i>
<i>Budex</i>	<i>piasek</i>	<i>Wykbud</i>
<i>Matbud</i>	<i>pustak</i>	<i>Wykbud</i>
<i>Matbud</i>	<i>cement</i>	<i>Rembud</i>
<i>Matbud</i>	<i>cement</i>	<i>Wykbud</i>
<i>Matbud</i>	<i>gips</i>	<i>Rembud</i>
<i>Matbud</i>	<i>piasek</i>	<i>Rembud</i>
<i>Matbud</i>	<i>piasek</i>	<i>Wykbud</i>

Typy złączenia

W rezultacie złączenia (θ -złączenia, równozłączenia, złączenia naturalnego) relacji R i relacji S powstaje relacja wynikowa zawierająca tylko te konkatenacje rekordów obu tabel, dla których spełnione jest kryterium złączenia. Jest to tzw. *złączenie wewnętrzne* (inner join) i jest ono przyjmowane jako domyślne. W praktycznych przypadkach, może zaistnieć potrzeba rozszerzenia relacji wynikowej o rekordy jednej lub drugiej relacji nie pasujące do żadnego rekordu drugiej tabeli. Mamy wówczas możliwość utworzenia czterech innych typów złączeń:

- *złączenie lewostronne zewnętrzne* (*left outer join*), tzn. złączenie zawierające wszystkie rekordy tablicy R uzupełnione tymi rekordami z tablicy S , które spełniają warunek połączenia,
- *złączenie prawostronne zewnętrzne* (*right outer join*), tzn. złączenie zawierające wszystkie rekordy tablicy S uzupełnione tymi rekordami z tablicy R , które spełniają warunek połączenia,
- *złączenie zewnętrzne pełne* (*full outer join*), tzn. złączenie zawierające wszystkie rekordy tablic R oraz S uzupełnione wartościami typu $NULL$ w przypadku gdy do danego rekordu nie pasuje żaden rekord z drugiej tablicy; w istocie jest to kombinacja złączenia lewo- i prawostronnego,
- *złączenie zewnętrzne typu union*, tzn. złączenie zawierające wszystkie rekordy tablicy R nie pasujące do żadnego rekordu S uzupełnione tymi rekordami z tablicy S , które nie pasują do żadnego rekordu tablicy R ; ten typ złączenia stanowi przeciwieństwo złączenia wewnętrznego (dokładniej: uzupełnienie do iloczynu kartezjańskiego).

We wszystkich przypadkach pola puste wypełniane są wartościami $NULL$.

Przykłady złączeń

Rozważmy następujące tabele:

<i>Produkt</i>	<i>Dostawca</i>		<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	oraz	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Matbud</i>		<i>Budex</i>	222 – 222
<i>cement</i>	<i>Budex</i>		<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Matbud</i>		<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>		<i>Neobud</i>	666 – 666
<i>gips</i>	<i>Unibud</i>		<i>Probud</i>	777 – 777
	<i>Zerobud</i>		<i>Startbud</i>	

Wynikiem złączenia wewnętrznego (naturalnego) będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Budex</i>	222 – 222
<i>pustak</i>	<i>Matbud</i>	444 – 444
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666

Wynikiem złączenia typu UNION będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
	<i>Probud</i>	777 – 777
	<i>Startbud</i>	
<i>gips</i>	<i>Unibud</i>	
	<i>Zerobud</i>	

Przykłady złączeń

Wynikiem operacji lewostronnego złączenia zewnętrznego będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Budex</i>	222 – 222
<i>pustak</i>	<i>Matbud</i>	444 – 444
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666
<i>gips</i>	<i>Unibud</i>	
	<i>Zerobud</i>	

Wynikiem operacji prawostronnego złączenia zewnętrznego będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Budex</i>	222 – 222
<i>pustak</i>	<i>Matbud</i>	444 – 444
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666
	<i>Probud</i>	777 – 777
	<i>Startbud</i>	

Przykłady złączeń

Wynikiem operacji złączenia zewnętrznego pełnego będzie tabela postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>pustak</i>	<i>Budex</i>	111 – 111
<i>pustak</i>	<i>Budex</i>	222 – 222
<i>pustak</i>	<i>Matbud</i>	444 – 444
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666
<i>gips</i>	<i>Unibud</i>	
	<i>Zerobud</i>	
	<i>Probud</i>	777 – 777
	<i>Startbud</i>	

W wyniku poszukiwania wszystkich materiałów i ich dostawców nie posiadających telefonów otrzymamy tabelę:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>gips</i>	<i>Unibud</i>	

W wyniku poszukiwania wszystkich dostawców dostarczających określone materiały, np. *cement* oraz *piasek* posiadających telefony dostaniemy tabelkę postaci:

<i>Produkt</i>	<i>Dostawca</i>	<i>Telefon</i>
<i>cement</i>	<i>Budex</i>	111 – 111
<i>cement</i>	<i>Budex</i>	222 – 222
<i>piasek</i>	<i>Matbud</i>	444 – 444
<i>piasek</i>	<i>Neobud</i>	555 – 555
<i>piasek</i>	<i>Neobud</i>	666 – 666

Wymuszanie więzów integralności

Więzy integralności

W praktyce definiuje się jedynie dwa typy połączeń między tabelami, są to:

- połączenia typu *jeden-do-jeden*,
- połączenia typu *jeden-do-wielu*.

Połączenia typu *wiele-do-wielu* realizowane są poprzez zdefiniowanie tablicy pośredniej.

Istnieje możliwość wymuszenia więzów integralności (ACCESS) nakładanych na związek i dane w łączonych tabelach – realizowane jest wtedy sprawdzanie warunków zgodności danych. Warunki te są następujące:

- definiowany związek musi dotyczyć dwóch tabel w tej samej bazie danych,
- pole (pola) łączące tabeli głównej stanowią klucz lub indeks jednoznaczny,
- w odpowiadających im polach tabeli dołączanej (polach klucza obcego) nie mogą pojawiać się dane, nie występujące w tabeli głównej,
- odpowiadające sobie pola muszą być tego samego typu (Autonumer == Liczba całkowita długa).

Po wymuszeniu więzów integralności możliwe jest uaktywnienie następujących opcji:

- kaskadowa aktualizacja powiązanych pól,
- kaskadowe usuwanie powiązanych rekordów.

Opcje te pozwalają automatycznie zachowywać spójność danych.

Własności operacji algebry relacji

Operacje sumy uogólnionej $+$ (algebraicznej dla relacji zgodnych) i złączenia naturalnego $*$ (iloczynu kartezjańskiego dla relacji o rozłącznych zbiorach atrybutów) mają następujące własności:

1. Idempotentność: $R + R = R$, $R * R = R$,
2. Łączność: $R + (S + T) = (R + S) + T$, $R * (S * T)$,
3. Przemienność: $R + S = S + R$, $R * S = S * R$,
4. Prawa rozdzielności:
 - $((R + S) * T)(X, Y, Z) = ((R * T)(X, Z) + (S * T)(Y, Z))$,
 - $(R + (S * T))(X, Y, Z) = ((R + S)(X, Y) * (R + T)(X, Z))$
5. Prawa deMorgana:
 - $\overline{(R + S)}(X, Y) = (\overline{R} * \overline{S})(X, Y)$,
 - $\overline{(R * S)}(X, Y) = (\overline{R} + \overline{S})(X, Y)$,

Znane operacje algebraiczne w relacyjnych bazach danych obejmują 12 operacji: $R+S$, $R \cup S$, $R \setminus S$, $R \cap S$, $R * S$, $R \bowtie_{\theta} S$, $R \times S$, $R \div S$, $R[A^1, A^2, \dots, A^k]$, $R]B[$, \overline{R} , $\sigma_F(R)$. Można je wyrazić za pomocą pięciu operacji elementarnych:

- sumy,
- różnicy,
- projekcji,
- iloczynu kartezjańskiego,
- selekcji.

Np. dopełnienie $\overline{R}(A_1, A_2, \dots, A_n) = R[A_1] \times R[A_2] \times \dots \times R[A_n] \setminus R$.

Operację iloczynu można zdefiniować jako $R \cap S = \pi_R(\sigma_{=}(R \times S))$. Inna możliwość to $R \cap S = R \setminus (R \setminus S)$. Operację różnicy można zdefiniować jako z wykorzystaniem lewostronnego złączenia zewnętrznego lub za pomocą selekcji gdy do budowy kryterium można wykorzystać negację i warunek przynależności rekordu do zbioru.

Problemy realizacji zapytań i optymalizacji ich wykonania

Ogólne zasady optymalizacji zapytań

Dla optymalizacji zapytań należy stosować odpowiednie zasady konstrukcji:

- należy unikać mnożenia krotek (iloczynu kartezjańskiego, nieprecyzyjnego θ -złączenia, niepotrzebnych operacji złączenia),
- operacje selekcji i projekcji wykonywać na jak najniższym poziomie (jak najwcześniej),
- dla częstych operacji wyszukiwania lub zawierających ukryte wyszukiwanie posługiwać się indeksami,
- operację selekcji oraz operację złączenia i następujące po niej operacje projekcji należy wykonywać łącznie (jako jedną operację),
- unikać operacji wymagających (wielokrotnego) dostępu do dysku.

Wybrane zasady szczegółowe

- uszeregować kolejne warunki (kryteria) selekcji od najsilniejszych do najsłabszych (pierwsze operacje eliminują maksimum rekordów),
- dokonywać możliwie wcześnie projekcji na możliwie mały zestaw atrybutów (z opcją DISTINCT),
- operacje selekcji przesuwają "do wewnątrz" innych operacji (np. iloczynu kartezjańskiego, sumy, różnicy),
- operacje projekcji na wybrane atrybuty przesuwają "do wewnątrz" innych operacji (np. iloczynu kartezjańskiego, sumy).