Systemy Baz Danych

Algebra Relacyjna

Bartosz Zieliński

WYDZIAŁ FIZYKI
i INFORMATYKI STOSOWANEJ
Uniwersytet Łódzki

Algebra Relacyjna

- Definiuje operacje na relacjach dające w wyniku relacje.
- Zbiór wszystkich możliwych relacji wraz z tymi operacjami tworzy algebrę w sensie matematycznym.
- Operacje można składać w bardziej złożone wyrażenia.
- Pomiędzy operacjami występują nietrywialne związki pozwalające na wyrażenie tej samej relacji na różne sposoby.
- Wyrażenia algebry relacyjnej definiują zawsze zapytania niezależne od dziedziny.

Algebra Relacyjna
a Implementacja Wykonywania Zapytań w DBMS

- Wystarczy aby DBMS implementował operatory relacyjne.
- Aby wykonać dane zapytanie wystarczy teraz wykonać po kolei operacje z wyrażenia definiującego to zapytanie.

W rzeczywistych RDBMS-ach

- Interpreter SQL tłumaczy zapytanie SQL na wyrażenie **algebry relacyjnej**.
- Optymalizator może przekształcić to wyrażenie na **równoważne** (definiujące tą samą relację) ale wykonujące się szybciej, korzystając z reguł **algebry relacyjnej**.
- Dla niektórych operatorów istnieje kilka algorytmów je implementujących wybieranych przez optymalizator.

Operatory w Algebrze Relacyjnej

- **Przemianowanie** atrybutów (*rename*)
- **Projekcje** (*projections*)
- **Selekcje** (*selections*)
- Operatory teoriomnogościowe
  - **Unia**, **Przecięcie**, **Różnica**, **Iloczyn kartezjański**
- Operatory złączenia
  - **złączenie naturalne**, **złączenie równościowe** (*equijoin*), θ-złączenia, **złączenia zewnętrzne** (*outer joins*) itp.
- **Dzielenie** relacji
- **Obliczenia domenowe** (*domain computations*)
- **Agregacje**

Operacja Przemianowania Atrybutu $\rho_{B/A}$

Operacja $\rho_{B/A}(R)$ **przemianowania atrybutu** A na B zamienia w relacji R nazwę atrybutu A na B (aby wynik operacji był dobrze zdefiniowany wymagamy aby $A \in \mathrm{Attr}(R)$ i $B \notin \mathrm{Attr}(R)$).

Jobs

JobId	**Name**	**MinSalary**	**MaxSalary**
1	IT Specialist	8000	20000
2	Sales Specialist	5000	9000
3	Administration	7000	10000

$\rho_{\mathbf{MinimalnaPensja}/\mathbf{MinSalary}}(\mathbf{Jobs})$

JobId	**Name**	**MinimalnaPensja**	**MaxSalary**
1	IT Specialist	8000	20000
2	Sales Specialist	5000	9000
3	Administration	7000	10000

Unia, Przecięcie i Różnica

Operacje **unii** ($R \cup S$), **przecięcia** ($R \cap S$) i **różnicy** ($R \setminus S$) są zdefiniowane dla relacji R i S takich że $\mathrm{Attr}(R) = \mathrm{Attr}(S)$. Wtedy

$$\mathrm{Attr}(R \cap S) = \mathrm{Attr}(R \cup S) = \mathrm{Attr}(R \setminus S) := \mathrm{Attr}(R)$$

i

- $\mathrm{Rows}(R \cup S) := \mathrm{Rows}(R) \cup \mathrm{Rows}(S)$
  (krotki w $R \cup S$ to te i tylko te krotki które są w R lub S)
- $\mathrm{Rows}(R \cap S) := \mathrm{Rows}(R) \cap \mathrm{Rows}(S)$
  (krotki w $R \cup S$ to te i tylko te krotki które są w R i w S)
- $\mathrm{Rows}(R \setminus S) := \mathrm{Rows}(R) \setminus \mathrm{Rows}(S)$
  (krotki w $R \cup S$ to te i tylko te krotki które są w R i których nie ma w S)

Przykład Unii, Przecięcia i Różnicy

R

A	B
1	10
2	20
3	30
4	40

S

A	B
1	10
2	20
5	50

$R \cup S$

A	B
1	10
2	20
3	30
4	40
5	50

$R \cap S$

A	B
1	10
2	20

$R \setminus S$

A	B
3	30
4	40

Rzutowanie Krotek na Podzbiór Atrybutów

Definicja

Niech t będzie krotką i niech $X \subseteq \mathrm{Attr}(t)$. Wówczas **rzutowaniem krotki t na podzbiór atrybutów X** nazywamy krotkę $t|_X$ zdefiniowaną przez

$$\mathrm{Attr}(t|_X) := X, \quad (t|_X).A = t.A \text{ dla każdego } A \in X.$$

Przykład

$$t = \frac{\mathbf{A} \quad \mathbf{B} \quad \mathbf{C}}{1 \quad\; 2 \quad\; 3}, \quad t|_{\{\mathbf{A},\mathbf{C}\}} = \frac{\mathbf{A} \quad \mathbf{C}}{1 \quad\; 3}.$$

Złączenia Krotek

Definicja

Niech t_1 i t_2 będą krotkami takimi że $t_1|_X = t_2|_X$, gdzie $X := \mathrm{Attr}(t_1) \cap \mathrm{Attr}(t_2)$ (warunek ten jest zawsze spełniony gdy $X = \emptyset$). Wówczas **złączeniem t_1 i t_2** nazywamy krotkę $t_1 \bowtie t_2$ t.ż. $\mathrm{Attr}(t_1 \bowtie t_2) := \mathrm{Attr}(t_1) \cup \mathrm{Attr}(t_2)$ i

$$(t_1 \bowtie t_2).A := \begin{cases} t_1.A & \text{gdy } A \in \mathrm{Attr}(t_1) \\ t_2.A & \text{gdy } A \in \mathrm{Attr}(t_2) \end{cases}.$$

Przykład

$$\begin{array}{cc} A & \color{red}{B} \\ \hline 1 & \color{red}{2} \end{array} \bowtie \begin{array}{cc} \color{red}{B} & C \\ \hline \color{red}{2} & 3 \end{array} = \begin{array}{ccc} A & \color{red}{B} & C \\ \hline 1 & \color{red}{2} & 3 \end{array}, \qquad \begin{array}{cc} A & B \\ \hline 1 & 2 \end{array} \bowtie \begin{array}{c} C \\ \hline 3 \end{array} = \begin{array}{ccc} A & B & C \\ \hline 1 & 2 & 3 \end{array}$$

Iloczyn Kartezjański

Iloczyn kartezjański relacji jest określony dla par relacji o rozłącznych zbiorach atrybutów. Niech R i S będą relacjami takimi że $\mathrm{Attr}(R) \cap \mathrm{Attr}(S) = \emptyset$. Wówczas **iloczyn kartezjański** $R \times S$ jest relacją o atrybutach $\mathrm{Attr}(R \times S) := \mathrm{Attr}(R) \cup \mathrm{Attr}(S)$ i krotkach

$$\mathrm{Rows}(R \times S) := \{t_1 \bowtie t_2 \mid t_1 \in R, t_2 \in S\}$$

Zauważmy że $R \times S = S \times R$ (ponieważ $t_1 \bowtie t_2 = t_2 \bowtie t_1$).
Na przykład

R

A	B
1	10
2	20

S

C	D
3	30
4	40

$R \times S$

A	B	C	D
1	10	3	30
1	10	4	40
2	20	3	30
2	20	4	40

Projekcja na Podzbiór Atrybutów

Niech R będzie relacją i niech $X \subseteq \mathrm{Attr}(R)$. Wtedy $\pi_X(R)$ jest relacją o atrybutach X i krotkach

$$\mathrm{Rows}(\pi_X(R)) := \{t|_X \mid t \in R\}$$

Przykład projekcji na podzbiór atrybutów:

Employees

Id	FirstName	LastName	Salary	JobId
1	Toru	Takemitsu	10000.11	1
2	Philip	Glass	9000.00	3
3	Michael	Nyman	10000.50	1
4	Henryk	Górecki	11000.00	1
5	Thomas	Tallis	8000.80	2
6	Arvo	Pärt	15000.70	1
7	Arnold	Schönberg	6000.00	2
8	Anton	Webern	6500.12	2
9	Alban	Berg	6750.50	2
10	Olivier	Messiaen	9500.00	3

$\pi_{\{\mathbf{JobId}\}}(\mathbf{Employees})$

JobId
1
2
3

Operacja Selekcji

Przypuśćmy że R jest relacją i niech ϕ będzie warunkiem określonym na krotkach o atrybutach $\mathrm{Attr}(R)$. Wówczas $\sigma_\phi(R)$ jest relacją o atrybutach $\mathrm{Attr}(\sigma_\phi(R)) := \mathrm{Attr}(R)$ i krotkach

$$\mathrm{Rows}(\sigma_\phi(R)) := \{t \mid t \in R \text{ i } t \text{ spełnia } \phi\}$$

Zauważmy że

$$\begin{aligned}
\sigma_{\phi \text{ and } \psi}(R) &= \sigma_\phi(R) \cap \sigma_\psi(R) \\
\sigma_{\phi \text{ or } \psi}(R) &= \sigma_\phi(R) \cup \sigma_\psi(R) \\
\sigma_{\text{not } \phi}(R) &= R \setminus \sigma_\phi(R)
\end{aligned}$$

Operacje Selekcji cd.

Atomowymi operacjami selekcji

nazywamy operacje postaci $\sigma_{C \theta D}(R)$ i $\sigma_{C \theta v}(R)$ gdzie $C, D \in \mathrm{Attr}(R)$, $v \in \mathcal{U}$, a θ jest operatorem porównania takim jak $<$, $\leq$, $=$, itp.

Proste operacje selekcji

tworzone są z operacji atomowych przy pomocy operatorów boolowskich.

Złożone operacje selekcji

mogą korzystać także z kwantyfikatorów po innych relacjach.

Przykład Atomowej Operacji Selekcji

Employees

Id	FirstName	LastName	Salary	JobId
6	Arvo	Pärt	15000.70	1
7	Arnold	Schönberg	6000.00	2
8	Anton	Webern	6500.12	2
9	Alban	Berg	6750.50	2
10	Olivier	Messiaen	11000.00	3

$\sigma_{\textbf{Salary} \geq 10000}(\textbf{Employees})$

Id	FirstName	LastName	Salary	JobId
6	Arvo	Pärt	15000.70	1
10	Olivier	Messiaen	11000.00	3

Przykład Prostej Operacji Selekcji

Employees

Id	FirstName	LastName	Salary	JobId
6	Arvo	Pärt	15000.70	1
7	Arnold	Schönberg	6000.00	2
8	Anton	Webern	6500.12	2
9	Alban	Berg	6750.50	2
10	Olivier	Messiaen	11000.00	3

$\sigma_{\mathbf{JobId}=3 \vee \mathbf{Salary}>15000}(\mathbf{Employees})$

Id	FirstName	LastName	Salary	JobId
6	Arvo	Pärt	15000.70	1
10	Olivier	Messiaen	11000.00	3

Operacja Złączenia Naturalnego

Przypuśćmy że relacje R i S są takie że $\mathrm{Attr}(R) \cap \mathrm{Attr}(S) \neq \emptyset$.

Złączenie naturalne $R \bowtie S$

jest relacją o atrybutach $\mathrm{Attr}(R \bowtie S) := \mathrm{Attr}(R) \cup \mathrm{Attr}(S)$ i krotkach

$$\mathrm{Rows}(R \bowtie S)$$
$$= \left\{ t_1 \bowtie t_2 \mid t_1 \in R \wedge t_2 \in S \wedge t_1|_{\mathrm{Attr}_R \cap \mathrm{Attr}_S} = t_2|_{\mathrm{Attr}_R \cap \mathrm{Attr}_S} \right\}.$$

Przypomnijmy że

$$(t_1 \bowtie t_2)(A) := \begin{cases} t_1.A & \text{gdy } A \in \mathrm{Attr}(R) \\ t_2.A & \text{gdy } A \in \mathrm{Attr}(S) \end{cases} \qquad (= t_2 \bowtie t_1(A))$$

Zauważmy że $R \bowtie S = S \bowtie R$.

Przykład Operacji Złączenia Naturalnego

Employees

Id	FirstName	LastName	Salary	JId
1	Toru	Takemitsu	10000.11	1
2	Philip	Glass	9000.00	3
3	Michael	Nyman	10000.50	1
4	Henryk	Górecki	11000.00	1
5	Thomas	Tallis	8000.80	2
6	Arvo	Pärt	15000.70	1
7	Arnold	Schönberg	6000.00	2
8	Anton	Webern	6500.12	2
9	Alban	Berg	6750.50	2
10	Olivier	Messiaen	9500.00	3

Jobs

JId	Name	Min	Max
1	IT Specialist	8000	20000
2	Sales Specialist	5000	9000
3	Administration	7000	10000

$\mathbf{Employees} \bowtie \mathbf{Jobs}$

Id	FirstName	LastName	Salary	JId	Name	Min	Max
1	Toru	Takemitsu	10000.11	1	IT Specialist	8000	20000
2	Philip	Glass	9000.00	3	Administration	7000	10000
3	Michael	Nyman	10000.50	1	IT Specialist	8000	20000
4	Henryk	Górecki	11000.00	1	IT Specialist	8000	20000
5	Thomas	Tallis	8000.80	2	Sales Specialist	5000	9000
6	Arvo	Pärt	15000.70	1	IT Specialist	8000	20000
7	Arnold	Schönberg	6000.00	2	Sales Specialist	5000	9000
8	Anton	Webern	6500.12	2	Sales Specialist	5000	9000
9	Alban	Berg	6750.50	2	Sales Specialist	5000	9000
10	Olivier	Messiaen	9500.00	3	Administration	7000	10000

Złączenie Naturalne a Inne Operacje

Złączenie naturalne można zdefiniować korzystając z operacji **projekcji**, **selekcji**, **przemianowania** oraz **iloczynu kartezjańskiego**.

Niech $\mathrm{Attr}(R) \cap \mathrm{Attr}(S) = \{A_1, \ldots, A_n\}$ i niech $\{B_1, \ldots, B_n\} \subseteq \mathcal{A} \setminus \mathrm{Attr}(R) \cup \mathrm{Attr}(S)$. Oznaczmy $\sigma_{\vec{A}=\vec{B}} := \sigma_{A_1=B_1 \wedge \cdots \wedge A_n=B_n}$ oraz $\rho_{\vec{B}/\vec{A}} := \rho_{B_1/A_1} \circ \cdots \circ \rho_{B_n/A_n}$. Wówczas

$$R \bowtie S = \pi_{\mathrm{Attr}(R) \cup \mathrm{Attr}(S)}\left(\sigma_{\vec{A}=\vec{B}}\left(R \times \rho_{\vec{B}/\vec{A}}(S)\right)\right)$$

Skoro **złączenie naturalne** (inne złączenia zresztą też) może zostać zdefiniowane przy pomocy innych operacji po co traktować je jako odrębną operację komplikując algebrę relacyjną?

Złączenia, Iloczyny Kartezjańskie i Implementacja

Przypuśćmy że $\mathrm{Attr}(R) \cap \mathrm{Attr}(S) = \{A\}$, $B \notin \mathrm{Attr}(R) \cup \mathrm{Attr}(S)$ $|R| = N$ i $|S| = M$. Wówczas $|R \times \rho_{B/A}(S)| = MN$ ale $|R \bowtie S| \leq N$ jeśli wartość A jednoznacznie identyfikuje krotkę w S.

Istnieje wiele algorytmów implementacji złączenia niewymagających tworzenia danych pośrednich o rozmiarze MN.

- Przykładem jest naiwny algorytm który wymaga MN operacji ale korzysta jedynie z $|R| + |S| + |R \bowtie S|$ rekordów w pamięci:

  for all $t_1 \in R$
      for all $t_2 \in S$
          if$(t_1(A) = t_2(A))$ **yield** $t_1 \bowtie t_2$

- istnieją też znacznie bardziej efektywne algorytmy zarówno korzystające jak i nie korzystające z **indeksów**.

θ-Złączenia i Złączenia Warunkowe

Przypuśćmy że R i S są relacjami takimi że $\mathrm{Attr}(R) \cap \mathrm{Attr}(S) = \emptyset$. Niech $C \in \mathrm{Attr}_R$, $D \in \mathrm{Attr}_S$ i niech θ będzie jednym z binarnych operatorów porównania $<$, $>$, $\leq$, $\geq$, $=$, itp.

θ-złączenie relacji R i S (na C, D)

jest zdefiniowane jako $R \bowtie_{C\theta D} S := \sigma_{C\theta D}(R \times S)$

- W ogólności mamy dla pewnego warunku ϕ będącego kombinacją boolowską warunków postaci $A\theta B$, gdzie $A \in \mathrm{Attr}(R)$ a $B \in \mathrm{Attr}(S)$, **złączenie warunkowe** $R \bowtie_{\phi} S := \sigma_{\phi}(R \times S)$.
- Szczególny przypadek $R \bowtie_{A_1=B_1 \wedge \dots \wedge A_k=B_k} S$ nazywa się **złączeniem równościowym** (*equijoin*).

Dwa Oznaczenia Pomocnicze

Niech X będzie skończonym zbiorem atrybutów. Niech $\mathbf{null}_X$ będzie relacją zawierającą pojedynczą krotkę o atrybutach X i wartości $\mathbf{NULL}$ dla każdego z tych atrybutów, tzn., $\mathrm{Attr}(\mathbf{null}_X) = X$ i $\mathrm{Rows}(\mathbf{null}_X) = \{t\}$, gdzie $t.A = \mathbf{NULL}$ dla każdego $A \in X$.

Przypuśćmy że złączenie naturalne $R \bowtie S$ jest określone. Oznaczmy przez R_{-S} relację składającą się z krotek należących do R dla których nie istnieją odpowiadające (w sensie złączenia) krotki w relacji S:

$$R_{-S} := R \setminus \pi_{\mathrm{Attr}_R}(R \bowtie S)$$

Dodatkowo oznaczmy $R^{\bowtie}_{-S} := R_{-S} \times \mathbf{null}_{\mathrm{Attr}(S) \setminus \mathrm{Attr}(R)}$.

Złączenia Zewnętrzne

Złączenia zewnętrzne, zdefiniowane dla tych samych par relacji co złączenia naturalne, oprócz złączonych krotek zawierają także wszystkie krotki z jednej lub obu relacji biorących udział w złączeniu dla których nie istnieje odpowiadająca krotka z drugiej relacji

Wyróżniamy następujące rodzaje złączeń zewnętrznych:

- **Lewe złączenie zewnętrzne** (*left outer join*)

$$R^{(+)}\bowtie S := R \bowtie S \cup R^{\bowtie}_{-S}$$

- **Prawe złączenie zewnętrzne** (*right outer join*)

$$R \bowtie^{(+)} S := R \bowtie S \cup S^{\bowtie}_{-R}$$

- **Pełne złączenie zewnętrzne** (*full outer join*)

$$R^{(+)}\bowtie^{(+)} S := R \bowtie S \cup R^{\bowtie}_{-S} \cup S^{\bowtie}_{-R}$$

Przykłady Złączeń Zewnętrznych

R

X	Y
10	1
20	2
30	2
100	3

S

Y	Z
1	X1
2	X2
4	X4

$R^{(+)}\bowtie S$

X	Y	Z
10	1	X1
20	2	X2
30	2	X2
100	3	**NULL**

$R \bowtie^{(+)} S$

X	Y	Z
10	1	X1
20	2	X2
30	2	X2
NULL	4	X4

$R^{(+)}\bowtie^{(+)} S$

X	Y	Z
10	1	X1
20	2	X2
30	2	X2
100	3	**NULL**
NULL	4	X4

$R \bowtie S$

X	Y	Z
10	1	X1
20	2	X2
30	2	X2

Warunkowe Złączenia Zewnętrzne

Także złączenia warunkowe występują w (trzech) wersjach zewnętrznych

R

X
1
2
10

S

Y
0
9

$R^{(+)} \bowtie_{X<Y} S$

X	Y
1	9
2	9
10	**NULL**

$R \bowtie^{(+)}_{X<Y} S$

X	Y
NULL	0
1	9
2	9

$R^{(+)} \bowtie^{(+)}_{X<Y} S$

X	Y
NULL	0
1	9
2	9
10	**NULL**

$R \bowtie_{X<Y} S$

X	Y
1	9
2	9

Operator Dzielenia Relacji

Chcemy mieć operator $\div$ będący odwrotnością iloczynu kartezjańskiego w tym sensie że

$$(K \times S) \div S = K$$

Przypuśćmy że R is S są relacjami takimi że $\mathrm{Attr}(S) \subsetneq \mathrm{Attr}(R)$. Oznaczmy $X := \mathrm{Attr}(R) \setminus \mathrm{Attr}(S)$. Wówczas $R \div S$ jest relacją o atrybutach X zdefiniowaną jako największa relacja K taka że

$$K \subseteq \pi_X(R) \quad \text{i} \quad K \times S \subseteq R.$$

Przykład Użycia Operatora Dzielenia Relacji

- **Zadanie**: Podać identyfikatory programistów którzy znają wszystkie języki wymienione w **ImpLang**
- **Rozwiązanie**: $\mathbf{Programmers} \div \mathbf{ImpLang}$

Programmers

Id	Language
1	C++
1	Java
1	Haskell
2	Haskell
2	Java
3	C++
3	Java

ImpLang

Language
C++
Java

$\mathbf{Programmers} \div \mathbf{ImpLang}$

Id
1
3

Operator Dzielenia Relacji a Inne Operatory

Przypuśćmy że R is S są relacjami takimi że $\mathrm{Attr}(S) \subsetneq \mathrm{Attr}(R)$. Oznaczmy $X := \mathrm{Attr}(R) \setminus \mathrm{Attr}(S)$. Wówczas

$$R \div S = \pi_X(R) \setminus \pi_X\Big((\pi_X(R) \times S) \setminus R\Big)$$

gdzie $(\pi_X(R) \times S) \setminus R$ to zbiór elementów które powinny być w R gdyby był on iloczynem kartezjańskim rzutowania R na X z S, ale ich w R nie ma.

Przykłady Własności Operatorów Algebry Relacyjnej

Równości poniżej oznaczają że jeśli wyrażenie po jednej ze stron jest dobrze określone, to dobrze określone jest też wyrażenie po drugiej stronie i są one sobie równe.

- Naturalne złączenia i iloczyny kartezjańskie są **przemienne**, czyli $R \bowtie S = S \bowtie R$ i $R \times S = S \times R$.
- Naturalne złączenia i iloczyny kartezjańskie są **łączne**, czyli

  $$(R \bowtie S) \bowtie T = R \bowtie (S \bowtie T), \quad (R \times S) \times T = R \times (S \times T)$$

- Przypuśćmy że warunek ϕ zależy wyłącznie od atrybutów relacji R. Wówczas

  $$\sigma_\phi(R \bowtie S) = \sigma_\phi(R) \bowtie S.$$

  Podobna własność zachodzi dla iloczynów kartezjańskich.