

ŚRODOWISKOWE BAZY DANYCH

cz. 2 - operacje

Adam Piórkowski

pioro@agh.edu.pl

<http://home.agh.edu.pl/~pioro/dyd/>

<http://home.agh.edu.pl/~pioro/dyd/SBD/SBD.html>

© Adam Piórkowski, wszelkie prawa zastrzeżone

Kraków, 2015

OPERACJE NA DANYCH

Proste działania teoriomnogościowe

- suma
- różnica
- iloczyn

Działania złożone

Wybieranie danych

- rzut (projekcja)
- selekcja

Złączenia tabel

- iloczyn kartezjański (CROSS JOIN)
- złączenie θ (teta)
- złączenie naturalne (NATURAL JOIN)
- równozłączenie / złączenie wewnętrzne (INNER JOIN)
- złączenia zewnętrzne (lewostronne, prawostronne, pełne – OUTER JOIN)
- autozłączenie (auto-join)
- złączenie częściowe (semi-join)
- iloraz

OPERACJE NA DANYCH

SUMA

Dane są relacje A i B:

$A(atr1, atr2), B(atr1, atr2)$

ATR1	ATR2
1	x
2	y
3	z

ATR1	ATR2
1	u
2	y

Suma

Sumą relacji A i B ($A \cup B$)

jest zbiór wszystkich krotek należących do relacji A lub B lub ich części wspólnej.

- duplikaty są odrzucane

$A \cup B$

ATR1	ATR2
1	x
2	y
3	z
1	u

OPERACJE NA DANYCH

RÓŻNICA

Dane są relacje A i B:

$A(atr1, atr2), B(atr1, atr2)$

ATR1	ATR2
1	x
2	y
3	z

ATR1	ATR2
1	u
2	y

Różnica

Różnicą relacji A i B ($A - B$) jest

zbiór wszystkich krotek należących do relacji A i nienależących do relacji B ani części wspólnej A i B.

$A - B$

atr1	atr2
1	x
3	z

OPERACJE NA DANYCH

ILOCZYN

Dane są relacje A i B:

$A(atr1, atr2)$, $B(atr1, atr2)$

A		
ATR1	ATR2	
1	x	
2	y	
3	z	

B		
ATR1	ATR2	
1	u	
2	y	

ILOCZYN

Iloczynem relacji A i B ($A \cap B$) jest

zbiór wszystkich krotek należących zarówno do relacji A jak i do relacji B (część wspólna).

Iloczyn można wyrazić przy pomocy różnicy: $A \cap B = A - (A - B)$

$A \cap B$

ATR1	ATR2
2	y

OPERACJE NA DANYCH

PROJEKCJA

Rzut (projekcja)

Operacja rzutu Π relacji A przekazuje dane relacji A do innej (nowej) relacji B, której zbiór atrybutów $\{b_1, \dots, b_n\}$ jest podzbiorem atrybutów A $\{a_1, \dots, a_n\}$

A

atr1	atr2	atr3	atr4
1	x	aa	dx
2	y	bb	dy

$\Pi_{atr1, atr2, atr4}(A)$

atr1	atr2	atr4
1	x	dx
2	y	dy

liczba krotek się nie zmienia, ale mogą powstać duplikaty

OPERACJE NA DANYCH

SELEKCJA

Selekcja

Operacja selekcji σ na relacji A tworzy podzbiór krotek, spełniających kryterium F, operujące na atrybutach A.

A

atr1	atr2	atr3	atr4
1	x	aa	123
2	y	bb	234
3	x	aa	576
4	y	bb	121
3	x	aa	260

$\sigma_{atr1 < 4, atr4 < 250}(A)$

atr1	atr2	atr3	atr4
1	x	aa	123
2	y	bb	234

OPERACJE NA DANYCH

Iloczyn kartezjański

A

atr1	atr2	atr3
11	x	aa
22	y	bb
33	x	bb

B

atr2	atr3	atr4	atr5
x	aa	123	11
y	bb	234	22

$A \bowtie B$

atr1	A.atr2	A.atr3	B.atr2	B.atr3	atr4	atr5
11	x	aa	x	aa	123	11
11	x	aa	y	bb	234	22
22	y	bb	x	aa	123	11
22	y	bb	y	bb	234	22
33	x	bb	x	aa	123	11
33	x	bb	y	bb	234	22

OPERACJE NA DANYCH

Złączenie θ (teta)

Złączenie θ dwóch relacji A i B ($A \bowtie_{\theta} B$) powstaje poprzez selekcję wyników $A \times B$ przy użyciu pewnego warunku θ . Warunek θ dotyczy najczęściej wybranych atrybutów relacji A i B.

A		
atr1	atr2	atr3
11	x	aa
22	y	bb
33	x	bb
44	y	bb

atr2	atr3	atr4	atr5
x	aa	123	11
y	bb	234	22

$A \bowtie_{\theta} B$

$\theta: \text{atr1} > \text{atr5}$

atr1	A.atr2	A.atr3	B.atr2	B.atr3	atr4	atr5
22	y	bb	x	aa	123	11
33	x	bb	x	aa	123	11
44	y	bb	x	aa	123	11
33	x	bb	y	bb	234	22
44	y	bb	y	bb	234	22

OPERACJE NA DANYCH

ILORAZ

Dane są relacje $A=\{a_1, a_2, \dots, a_m\}$ i $B=\{b_1, b_2, \dots, b_n\}$, $B \neq \emptyset$, $m > k$

Ilorazem relacji A i B ($A \div B$) jest nowa relacja C, która zawiera te krotki A, które zawierają wszystkie kombinacje krotek B.

Zbiór atrybutów relacji C zawiera jedynie atrybuty należące do relacji A i nienależące do relacji B.

A			
atr1	atr2	atr3	atr4
1	x	aa	12
2	y	aa	12
3	z	aa	12
1	x	bb	34
2	y	bb	34
3	z	bb	34
1	x	cc	56
2	y	cc	56

B	
atr3	atr4
aa	12
bb	34
cc	56

A ÷ B	
atr1	atr2
1	x
2	y

3	z
---	---

3	z	cc	56
---	---	----	----

OPERACJE NA DANYCH

ZŁĄCZENIE NATURALNE

$A(\text{atr1}, \text{atr2}, \text{atr3}, \dots)$, $B(\text{atr2}, \text{atr3}, \text{atrbis1}, \text{atrbis2}, \dots)$

Złączenie naturalne relacji A i B ($A \bowtie B$) kreuje nową relację, której zbiór atrybutów powstaje jako suma teoriiomnogościowa zbiorów atrybutów A i B. Operacja przebiega następująco:

1. wyznacza się $A \times B$,
2. z wyznaczonego iloczynu kartezjańskiego wybiera się tylko te krotki, które mają te same wartości dla atrybutów wspólnych dla A i B ($A.\text{atr2} <=> B.\text{atr2}$, $A.\text{atr3} <=> B.\text{atr3}$)
3. dokonuje się rzutu usuwającego zduplikowane kolumny ($A.\text{atr2}$, $B.\text{atr2} \rightarrow \text{atr2}$, ...)

A		
atr1	atr2	atr3
1	x	aa
2	y	bb
3	x	bb
4	y	bb

B			
atr2	atr3	atr4	atr5
x	aa	123	11
y	bb	234	22

OPERACJE NA DANYCH

ZŁĄCZENIE NATURALNE

$A \times B$						
atr1	A.atr2	A.atr3	B.atr2	B.atr3	atr4	atr5
1	x	aa	x	aa	123	11
2	y	bb	x	aa	123	11
3	x	bb	x	aa	123	11
4	y	bb	x	aa	123	11
1	x	aa	y	bb	234	22
2	y	bb	y	bb	234	22
3	x	bb	y	bb	234	22
4	y	bb	y	bb	234	22

atr1	A.atr2	A.atr3	B.atr2	B.atr3	atr4	atr5
1	x	aa	x	aa	123	11
2	y	bb	y	bb	234	22
4	y	bb	y	bb	234	22

$A \bowtie B$

atr1	atr2	atr3	atr4	atr5
1	x	aa	123	11
2	y	bb	234	22
4	y	bb	234	22

OPERACJE NA DANYCH

RÓWNOZŁĄCZENIE

Szczególny przypadek złączenia θ , w którym warunek θ łączy krotki o tej samej wartości dla wybranych atrybutów

$\theta : A.attr2 = B.attr3$

! Iloczyn kartezjański

A		B	
atr1	atr2	atr3	atr4
1	x	x	AA
2	y	y	BB
3	x	v	CC

$A \bowtie_{A.attr2 = B.attr3} B$

atr1	atr2	atr3	atr4
1	x	x	AA
2	y	y	BB
3	x	x	AA

OPERACJE NA DANYCH

Złączenie wewnętrzne (INNER JOIN)

Złączenie wewnętrzne to szczególny przypadek złączenia θ , w którym warunek θ łączy krotki dla wybranych atrybutów zadanym operatorem porównania ($=$, $<$, $>$, $<=$, $>=$)

$\theta : A.attr2 > B.attr3$

A		B	
atr1	atr2	atr3	atr4
1	x	x	AA
2	y	y	BB
3	x	v	CC

$A \bowtie_{A.attr2 > B.attr3} B$

atr1	atr2	atr3	atr4
1	x	v	CC
2	y	x	AA
2	y	v	CC
3	x	v	CC

OPERACJE NA DANYCH

Złączenie zewnętrzne lewostronne (LEFT OUTER JOIN)

Wynikiem złączenia zewnętrznego lewostronnego

A LEFT OUTER JOIN B ON (atrybutyzłączenia)

są wszystkie krotki relacji A, zawierające wartości atrybutów

z relacji B dla uzgodnionych warunkiem złączenia (atrybutyzłączenia) krotek z relacji B lub wartości puste.

A		B	
atr1	atr2	atr3	atr4
1	x	x	AA
2	x	y	BB
3	z	v	CC

$A \text{ LEFT OUTER JOIN } B \text{ ON } (A.attr2 = B.attr3)$

atr1	atr2	atr3	atr4
1	x	x	AA
2	x	x	AA
3	z		

OPERACJE NA DANYCH

Złączenie zewnętrzne prawostronne (RIGHT OUTER JOIN)

Wynikiem złączenia zewnętrznego prawostronnego

A LEFT OUTER JOIN B ON (atrybutyzłączenia)

są wszystkie krotki relacji B, zawierające wartości atrybutów

z relacji A dla uzgodnionych warunkiem złączenia (atrybutyzłączenia) krotek z relacji A lub wartości puste.

A		B	
atr1	atr2	atr3	atr4
1	x	x	AA
2	x	y	BB
3	z	u	CC

$A \text{ RIGHT OUTER JOIN } B \text{ ON } (A.attr2 = B.attr3)$

atr1	atr2	atr3	atr4
1	x	x	AA
2	x	x	AA
		y	BB
		u	CC

OPERACJE NA DANYCH

Złączenie zewnętrzne pełne (FULL OUTER JOIN)

Wynikiem złączenia zewnętrznego prawostronnego

A LEFT OUTER JOIN B ON (atrybuty_złączenia)

są wszystkie krotki relacji B, zawierające wartości atrybutów

z relacji A dla uzgodnionych warunkiem złączenia (atrybuty_złączenia) krotek z relacji A lub wartości puste.

A		B	
atr1	atr2	atr3	atr4
1	x	x	AA
2	x	y	BB
3	z	u	CC

A FULL OUTER JOIN B ON (A.atr2 = B.atr3)

atr1	atr2	atr3	atr4
1	x	x	AA
2	x	x	AA
		y	BB
		u	CC
3	z		

OPERACJE NA DANYCH

Autozłączenie (self-join)

Autozłączenie to operacja łączenia danych z tej samej tabeli.

Nr	Imię	Nazwisko	Nr szefa
1	Jerzy	Nowak	
2	Anna	Wójcik	1
3	Aneta	Mazur	2

→ s-j

Nr	Imię	Nazwisko	Sz.Imię	Sz.Nazw
1	Jerzy	Nowak		
2	Anna	Wójcik	Jerzy	Nowak
3	Aneta	Mazur	Anna	Wójcik

Złączenie częściowe (semi join)

Złączenie częściowe to takie złączenie dwóch tabel, w którym wyniki pochodzą tylko z jednej tabeli. Stosowane przy dekompozycji z zapytania zagnieżdżonego lub przy ograniczaniu rozwiązań.

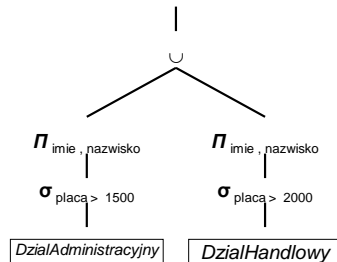
OPERACJE NA DANYCH

ZŁOŻENIA

Algebra w modelu relacyjnym dopuszcza tworzenie złożów operacji. Kolejność wykonywania można określić operatorem ().

$(\pi_{\text{IMIE, NAZWISKO}}(\sigma_{\text{PLACA} > 1500}(\text{DzialAdministracyjny}))) \cup (\pi_{\text{IMIE, NAZWISKO}}(\sigma_{\text{PLACA} > 2000}(\text{DzialHandlowy})))$

Powyższe złożenie można przedstawić przy pomocy drzewa operacji:



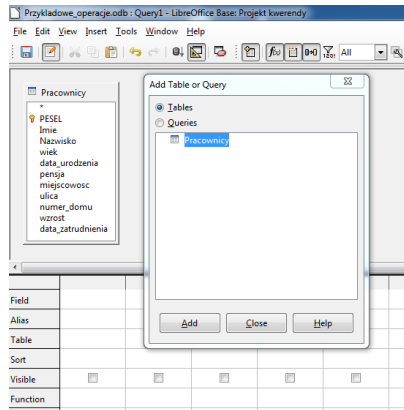
Przykładowe kwerendy w LibreOffice

Przykładowa tabela z danymi

Field Name	Field Type
PESEL	BigInt [BIGINT]
Imie	Text [VARCHAR]
Nazwisko	Text [VARCHAR]
wiek	Small Integer [SMALLINT]
data_urodzenia	Date [DATE]
penja	Decimal [DECIMAL]
miestowosc	Text [VARCHAR]
ulica	Text [VARCHAR]
numer_domu	Text [VARCHAR]
wzrost	Decimal [DECIMAL]
data_zatrudnienia	Date [DATE]

PESEL	Imie	Nazwisko	wiek	data_urodzenia	penja	miestowosc	ulica	numer_domu	wzrost	data_zatrudnienia
2222222	Ula	Nowak	29	1.05.1986	3500,00	Rzeszów	Cieplint	1	170,0	15.04.2013
3333333	Ela	Nowak	31	2.06.1984	5500,00	Kraków	Nowojki	1	180,0	15.04.2012
111111111	Ola	Nowak	32	1.03.1983	6000,00	Warszawa	Piekna	1	160,5	15.08.2013
123456567	Ala	Nowak	30	1.01.1985	5000,00	Warszawa	Śliczna	1	175,0	15.08.2012
987654321	Ewa	Nowak	28	1.02.1987	4000,00	Kraków	Wadowin	1	165,0	15.04.2011

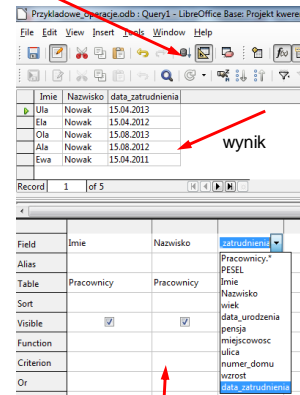
Przykładowe kwerendy w LibreOffice



Wybór tabel biorących udział w kwerendzie – na początek – jedna

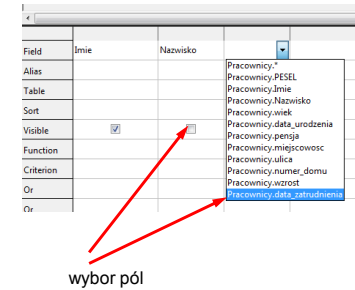
Przykładowe kwerendy w LibreOffice

uruchomienie zapytania



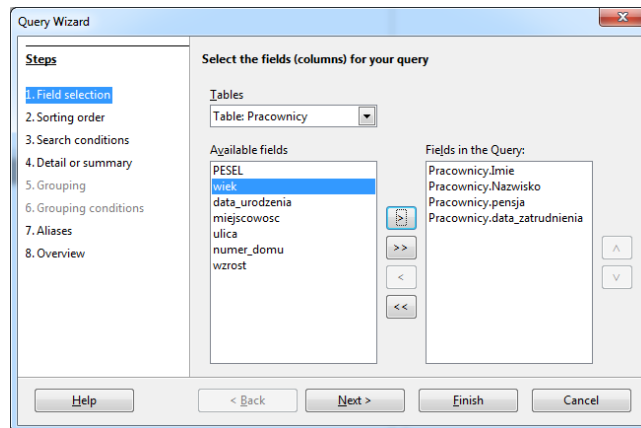
Projekt kwerendy

PROJEKCJA



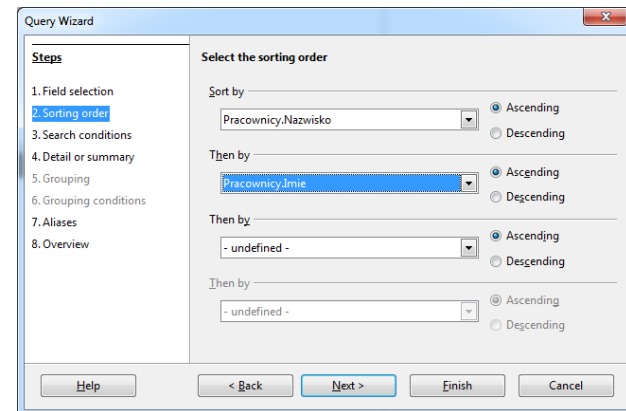
wybor pól

Przykładowe kwerendy w LibreOffice



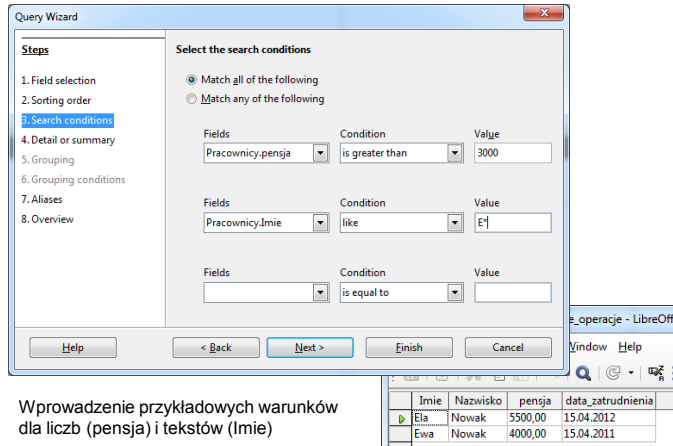
Inna metoda: użycie tzw. wizzarda

Przykładowe kwerendy w LibreOffice

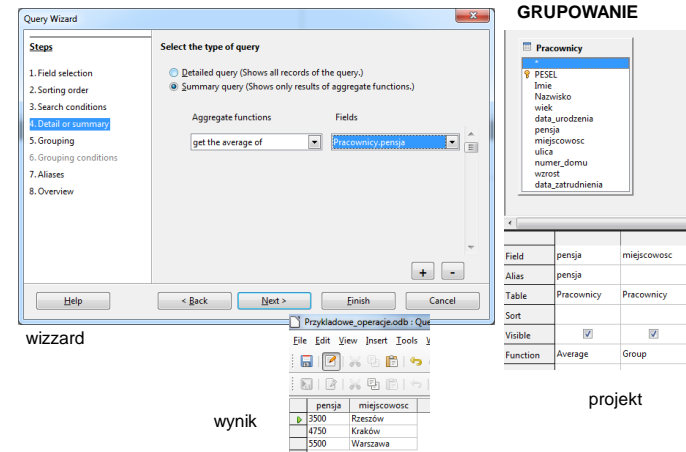


sortowanie danych

Przykładowe kwerendy w LibreOffice



Przykładowe kwerendy w LibreOffice



Przykładowe kwerendy w LibreOffice

GRUPOWANIE

