



METODY INŻYNIERII WIEDZY

*ASOCJACYJNA REPREZENTACJA POWIĄZANYCH
TABEL I WNIOSKOWANIE*

IGOR CZAJKOWSKI

CELE PROJEKTU

- Transformacja dowolnej bazy danych w min. 3 postaci normalnej do postaci Asocjacyjnej Grafowej Struktury Danych
- Implementacja metod filtrowania, wyszukiwania, usuwania węzłów
- Porównanie wydajności i szybkości działania z relacyjną bazą danych SQL

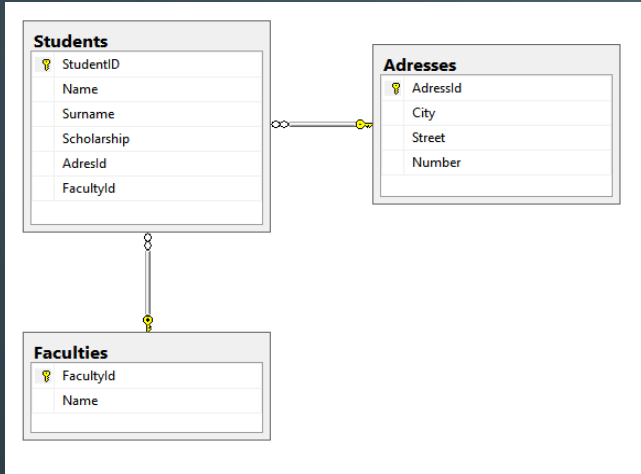
AGDS

- Umożliwiają horyzontalne i wertykalne powiązanie danych, bazy danych SQL jedynie powiązanie atrybutów poprzez relacje
- Brak konieczności wykonywania JOIN na tabelach
- Brak duplikatów
- Struktura grafowa

ZASTOSOWANE TECHNOLOGIE

- C#
- WPF
- GraphSharp
- MS SQL Server
- ANTRL v4

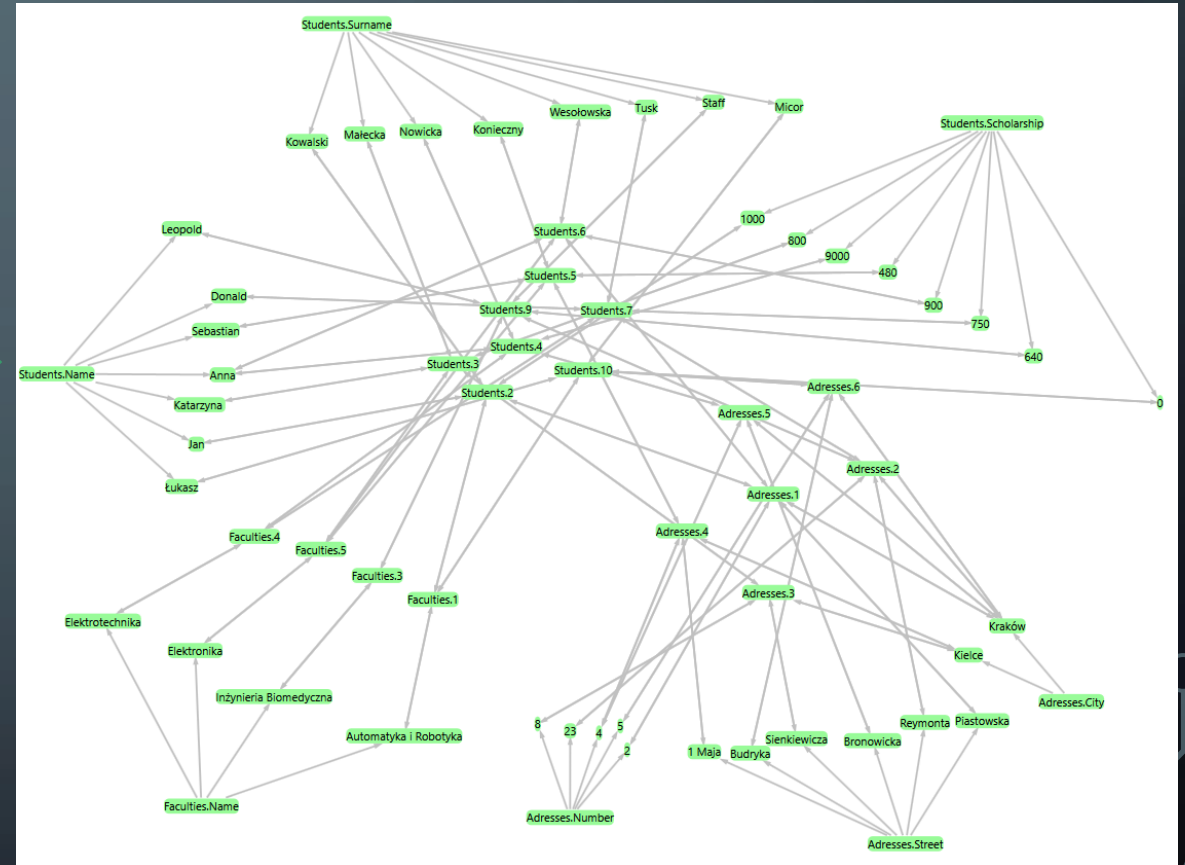
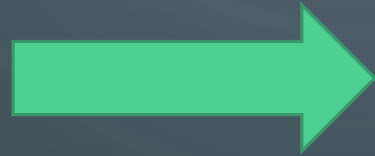
TRANSFORMACJA PRZYKŁADOWEJ BAZY



StudentID	Name	Surname	Scholarship	AdresId	FacultyId
1	2	Jan	Kowalski	1000	1
2	3	Katarzyna	Malecka	800	3
3	4	Anna	Nowicka	9000	5
4	5	Sebastian	Konieczny	480	4
5	6	Anna	Wesołowska	900	1
6	7	Donald	Tusk	750	2
7	9	Leopold	Staff	640	2
8	10	Łukasz	Micor	0	6

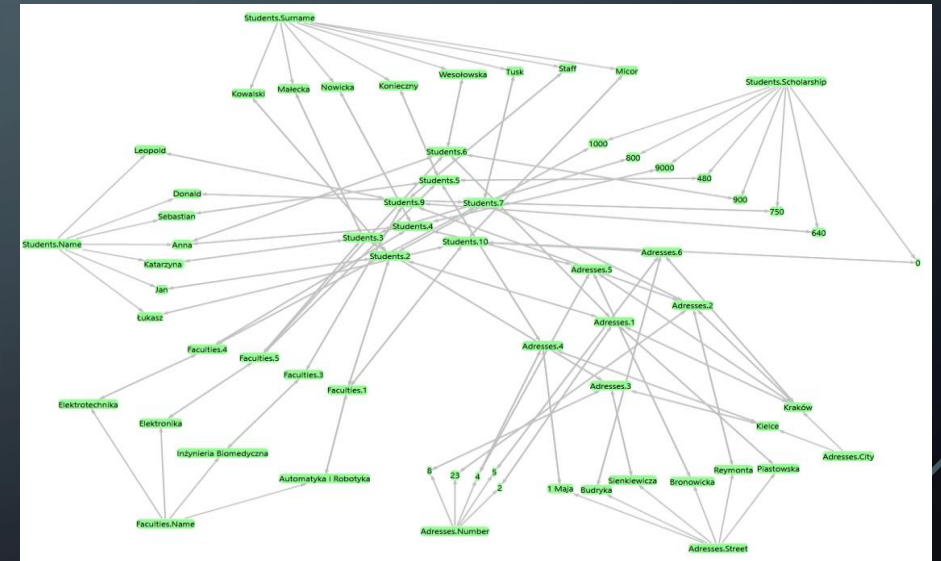
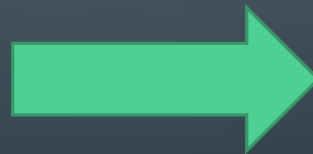
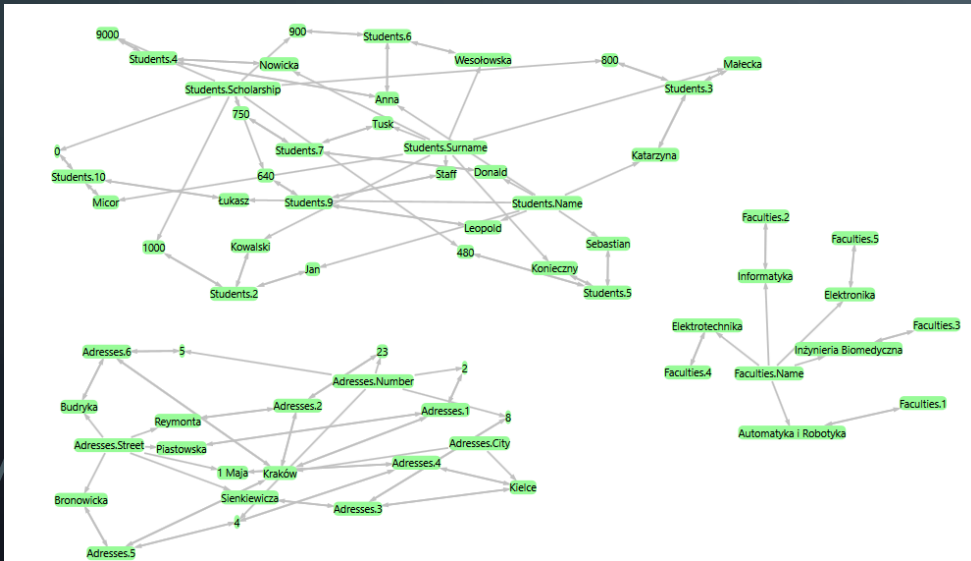
FacultyId	Name
1	Automatyka i Robotyka
2	Infomatyka
3	Inżynieria Biomedyczna
4	Elektrotechnika
5	Elektronika

AdresId	City	Street	Number
1	Kraków	Piastowska	2
2	Kraków	Reymonta	23
3	Kielce	Sienkiewicza	8
4	Kielce	1 Maja	4
5	Kraków	Bronowicka	4
6	Kraków	Budryka	5



BUDOWA GRAFU

1. Odczytanie metadanych z bazy SQL – nazwy tabel, relacje
2. Stworzenie podgrafów dla każdej tabeli
3. Połączenie ze sobą grafów każdej tabeli na podstawie relacji pomiędzy nimi



TRANSFORMACJA PRZYKŁADOWEJ BAZY

Rozkład przykładowej bazy NORTHWIND – prezentacja danych w postaci drzewa.

The screenshot shows a database tool window titled "AGDS". The interface includes a search bar, a "Zapytanie" field, and a "Limit odległości kluczy:" set to 3, with "Filtruj" and "Reset" buttons. The main area displays a tree view of the database schema. The tree is expanded to show the "Region" hierarchy, with "Northern" selected. The "Northern" region contains 17 territories, each with a unique ID and a description. The "Employees" table is also visible, showing a list of employees including "Andrew", "Anne", "Janet", and "Laura".

```
AGDS
Plik
Zapytanie
Limit odległości kluczy: 3 [Filtruj] [Reset]
Drzewo [Graf] [Graf Wyniki]
├── Orders.OrderDate
├── Orders.RequiredDate
├── Orders.ShippedDate
├── Orders.Freight
├── Orders.ShipName
├── Orders.ShipAddress
├── Orders.ShipCity
├── Orders.ShipRegion
├── Orders.ShipPostalCode
├── Orders.ShipCountry
├── Products.ProductName
├── Products.QuantityPerUnit
├── Products.UnitPrice
├── Products.UnitsInStock
├── Products.UnitsOnOrder
├── Products.ReorderLevel
├── Products.Discontinued
├── OrderDetails.UnitPrice
├── OrderDetails.Quantity
├── OrderDetails.Discount
├── Region.RegionDescription
│   ├── Eastern
│   │   ├── Territories.01581
│   │   ├── Territories.01730
│   │   ├── Territories.01833
│   │   ├── Territories.02116
│   │   ├── Territories.02139
│   │   ├── Territories.02184
│   │   ├── Territories.02903
│   │   ├── Territories.06897
│   │   ├── Territories.07960
│   │   ├── Territories.08837
│   │   ├── Territories.10019
│   │   ├── Territories.10038
│   │   ├── Territories.11747
│   │   ├── Territories.14450
│   │   ├── Territories.19713
│   │   ├── Territories.20852
│   │   ├── Territories.27403
│   │   ├── Territories.27511
│   │   └── Territories.40222
│   ├── Northern
│   ├── Southern
│   └── Western
├── Territories.TerritoryDescription
├── Employees.LastName
├── Employees.FirstName
│   ├── Andrew
│   ├── Anne
│   │   └── Employees.9
│   ├── Janet
│   └── Laura
```

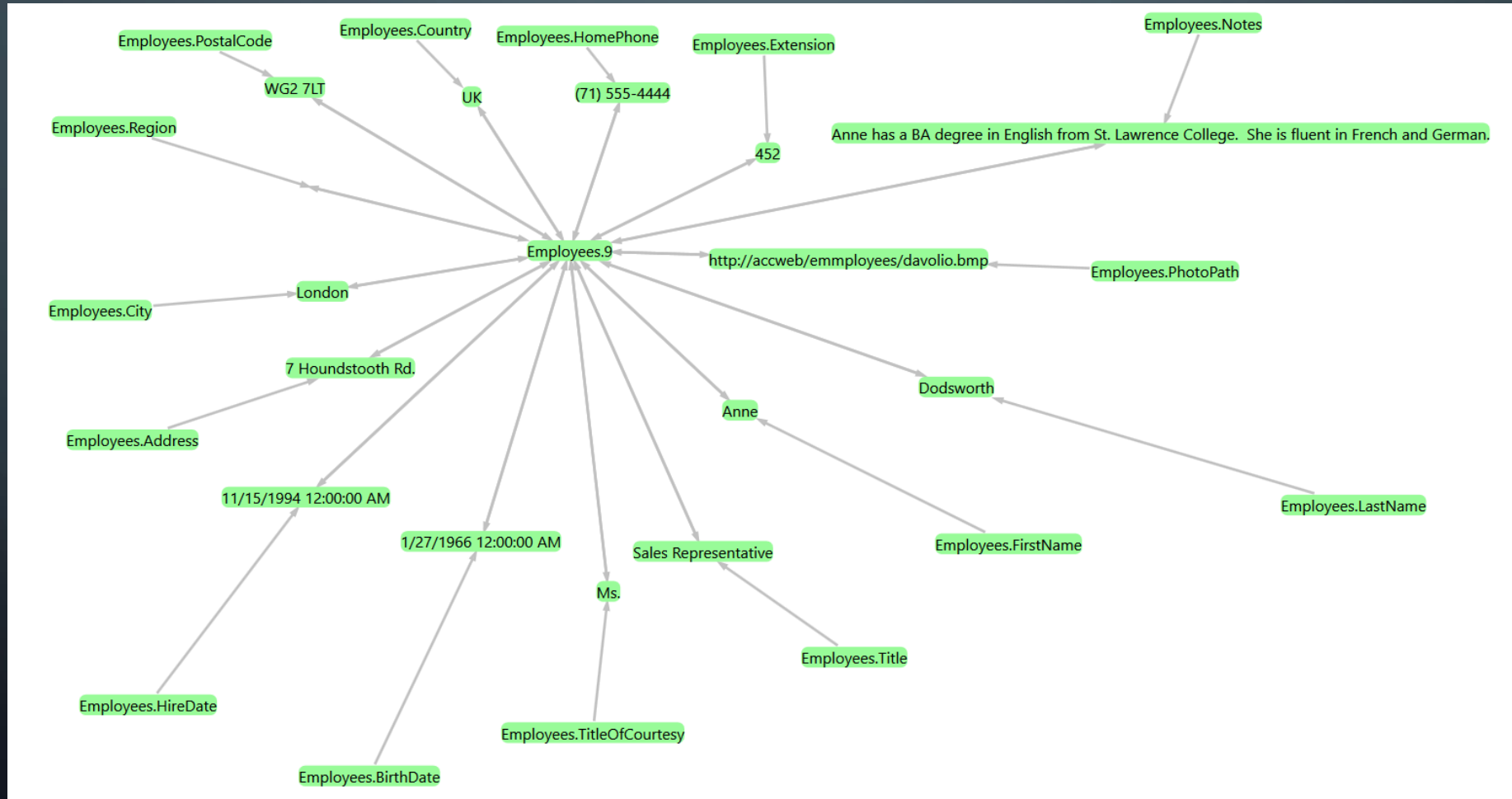
Wyszukano w 0 ms

WYSZUKIWANIE

- Wykorzystanie generatora parserów ANTLR do parsowania zapytań użytkownika
- Możliwość łączenia warunków za pomocą operatorów logicznych AND, OR
- Możliwość definiowania warunków: $=$, $<$, $>$, \leq , \geq
- Wykonywanie każdego zapytania osobno a następnie zwracanie odpowiednio iloczynu bądź sumy zbiorów w zależności od warunków.

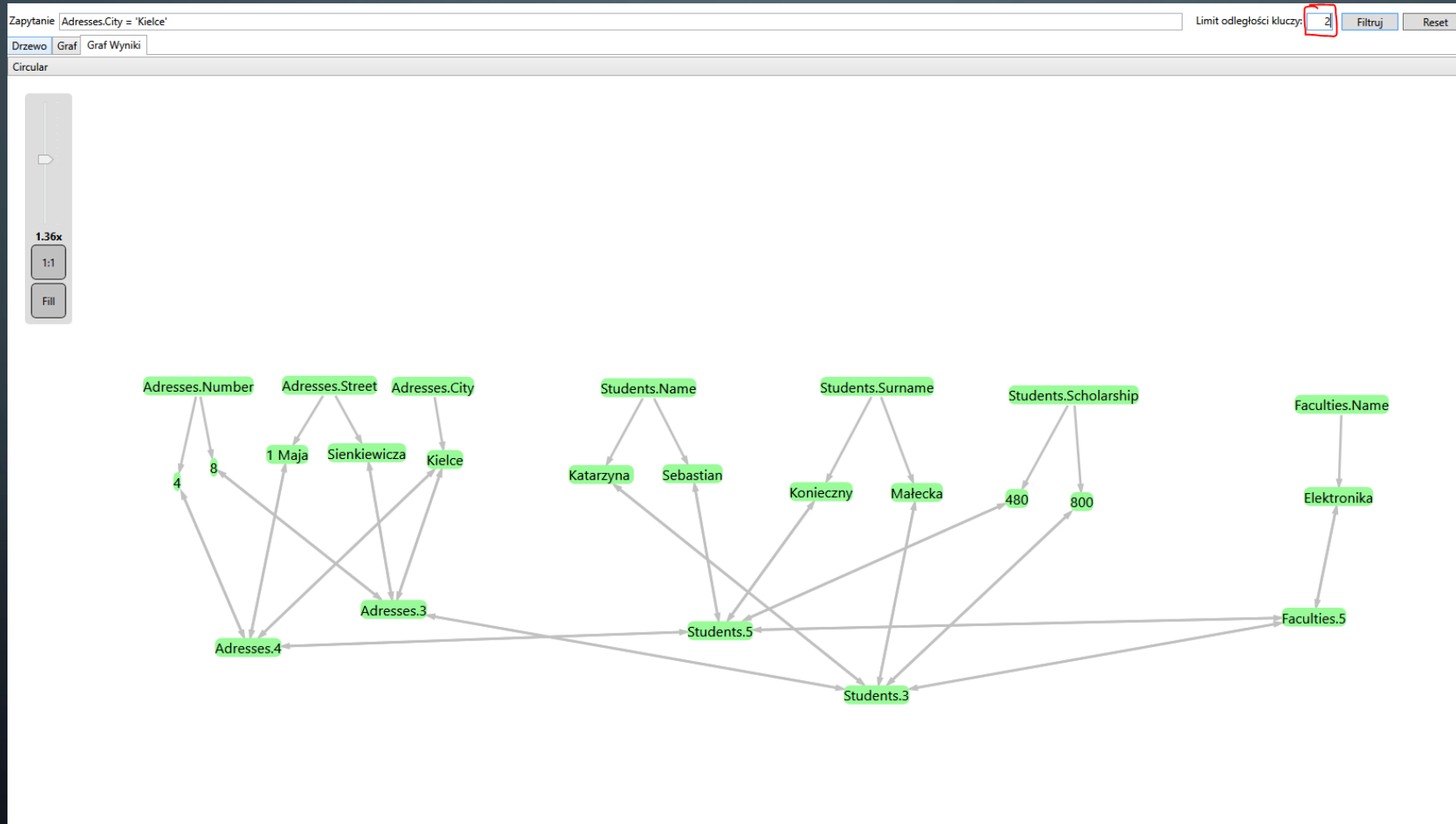
WYSZUKIWANIE

Employees.City = 'London' and Employees.FirstName = 'Anne'



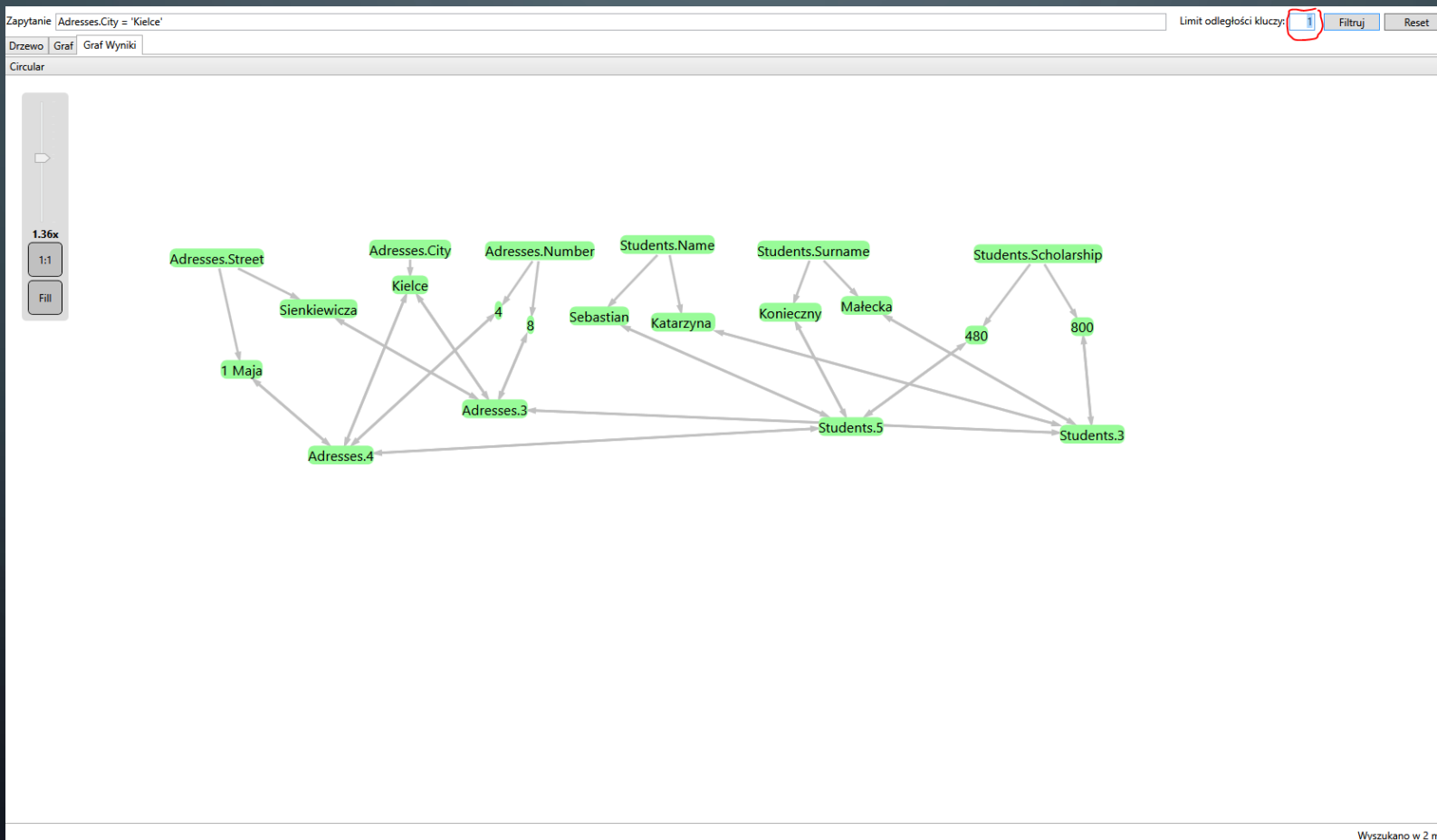
WYSZUKIWANIE

Możliwość przechodzenia do powiązanych tabel przez klucze obce:



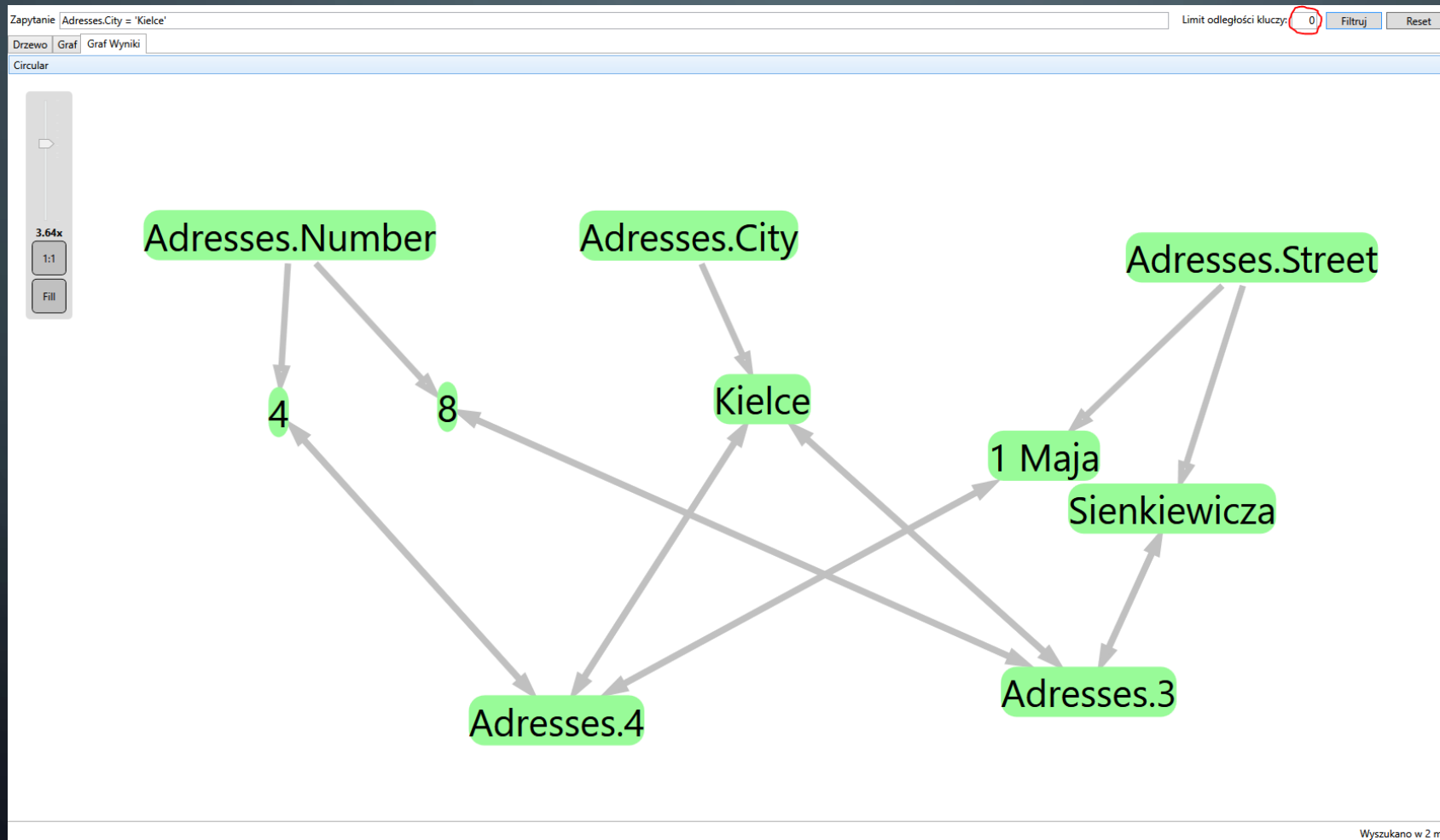
WYSZUKIWANIE

Możliwość przechodzenia do powiązanych tabel przez klucze obce:



WYSZUKIWANIE

Możliwość przechodzenia do powiązanych tabel przez klucze obce:



WYDAJNOŚĆ

Znajdź węzły pracowników i zamówień i powiązanych z nimi danych w bazie danych, którzy mieszkają w Londynie, mają na imię Anna, oraz wszystkie wysyłki powiązane z nimi mają adres dostarczenia w Argentynie: „Employees.City = 'London' and Employees.FirstName = 'Anne' or Orders.ShipCountry = 'Argentina' ”

SQL – 243MS

```
use NORTHWND

DECLARE @t1 DATETIME;
DECLARE @t2 DATETIME;

SET @t1 = GETDATE();
Select * from Orders
left join Customers on Orders.CustomerID = Customers.CustomerID
left join OrderDetails on Orders.OrderID = OrderDetails.OrderID
left join Employees on Orders.EmployeeID = Employees.EmployeeID
left join Shippers on Orders.ShipVia = Shippers.ShipperID
where Employees.City = 'London' and Employees.FirstName = 'Anne' or Orders.ShipCountry = 'Argentina';
SET @t2 = GETDATE();
SELECT DATEDIFF(millisecond,@t1,@t2) AS elapsed_ms;
```

	elapsed_ms
1	243

AGDS - 149 MS

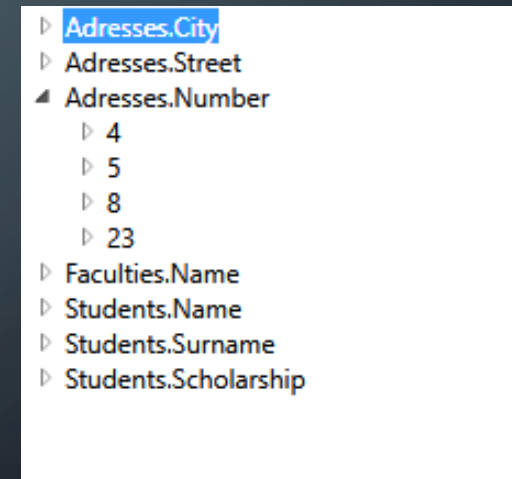
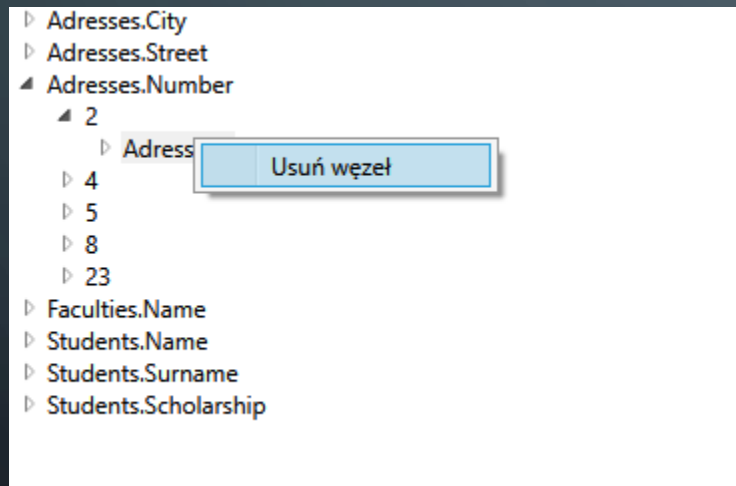
The screenshot shows the AGDS interface with a search query: `Employees.City = 'London' and Employees.FirstName = 'Anne' or Orders.ShipCountry = 'Argentina'`. The results are displayed as a tree view with columns expanded to show details for each record, including shipper information, order details, and employee information. A red box highlights the search execution time in the bottom right corner: "Wyszukano w 149 ms".

WYDAJNOŚĆ

- Dla małych ilości danych i bez używania JOIN w SQL wyniki SQL są nieznacznie lepsze.
- W przypadku dużych ilości danych AGDS jest szybszy, brak konieczności robienia JOIN.
- W przypadku AGDS duże zużycie pamięci – konieczność wczytania całej bazy danych do pamięci operacyjnej

USUWANIE WĘZŁÓW

- Możliwość usuwania kluczy głównych, w przypadku, gdy atrybut z którego usunięto klucz główny nie posiada innych dowiązanych kluczy również zostanie usunięty



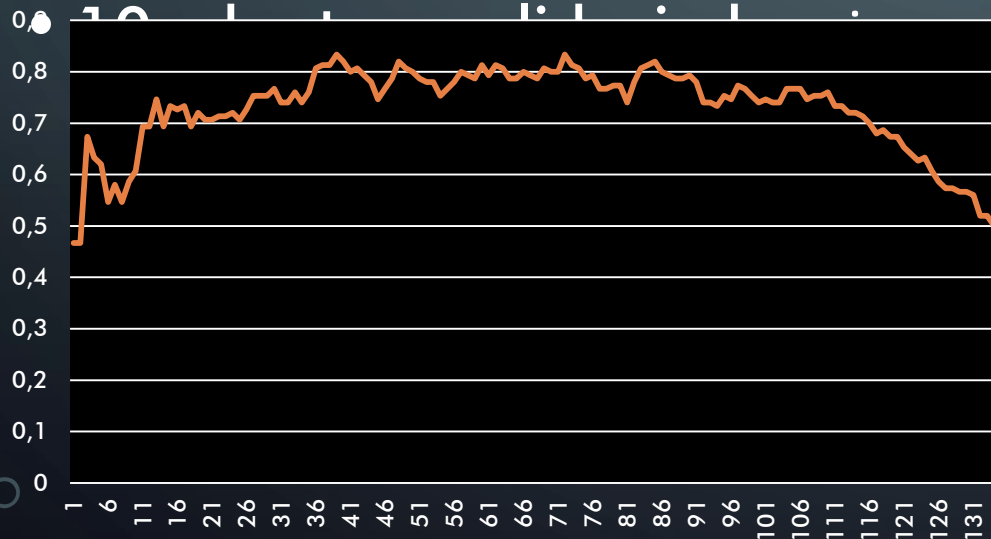
ROZWINIĘCIE PROJEKTU

- Przyspieszanie działania wyszukiwania
- Dodanie możliwości wyszukiwania bardziej złożonych elementów: daty, obrazy, blob.
- Zapis grafu do pliku – brak konieczności trzymania naraz całej bazy w pamięci i za każdym razem budowania grafu
- Reprezentacja wyników w postaci tabeli, analogicznie jak w SQL
- Możliwość zdefiniowania, które atrybuty chcemy otrzymać

POZOSTAŁE APLIKACJE

METODA K NAJBLIŻSZYCH SĄSIADÓW + WALIDACJA KRZYŻOWA

- Wyznaczanie optymalnego k na podstawie metody walidacji krzyżowej
- Ustalenie klasy wprowadzonego Irysu



```
file:///D:/szkola/4rok/MIW/ConsoleApplication1/bin/Debug/ConsoleApplication1.EXE
Best found k is: 39, effectiveness: 0,8333333333333333
LeafLength:
5,1
LeafWidth:
3,5
PetalLength:
1,4
PetalWidth:
0,2
Iris class is Setosa
```

ASOCJACYJNA GRAFOWA STRUKTURA DANYCH

- Transformacja danych z pliku *.txt do struktury AGDS
- Wyznaczanie wag w grafie dla wejściowego Irysa i znajdowanie Irysów podobnych

MainWindow
Zaladuj dane

	LeafLength	LeafWidth	PetalLength	PetalWidth	
0		0	3	0	OK
<					>
LeafLength 0					0
4.3	0.455696202531646				
4.4	0.443037974683544				
4.5	0.430379746835443				
4.6	0.417721518987342				
4.7	0.405063291139241				
4.8	0.392405063291139				
4.9	0.379746835443038				
5	0.367088607594937				
5.1	0.354430379746836				
5.2	0.341772151898734				
5.3	0.329113924050633				
5.4	0.316455696202532				
5.5	0.30379746835443				
5.6	0.291139240506329				
5.7	0.278481012658228				
5.8	0.265822784810127				
5.9	0.253164556962025				
6	0.240506329113924				
6.1	0.227848101265823				
6.2	0.215189873417722				
6.3	0.20253164556962				
LeafLength: 5.1; LeafWidth: 3.5; PetalLength: 1.4; PetalWidth 0.2; Type: Setosa	0.551947348403581				
LeafLength: 4.9; LeafWidth: 3; PetalLength: 1.4; PetalWidth 0.2; Type: Setosa	0.58668553236723				
LeafLength: 4.7; LeafWidth: 3.2; PetalLength: 1.3; PetalWidth 0.2; Type: Setosa	0.577413742661544				
LeafLength: 4.6; LeafWidth: 3.1; PetalLength: 1.5; PetalWidth 0.2; Type: Setosa	0.594734694076574				
LeafLength: 5; LeafWidth: 3.6; PetalLength: 1.4; PetalWidth 0.2; Type: Setosa	0.549430087183788				
LeafLength: 5.4; LeafWidth: 3.9; PetalLength: 1.7; PetalWidth 0.4; Type: Setosa	0.512438269197012				
LeafLength: 4.6; LeafWidth: 3.4; PetalLength: 1.4; PetalWidth 0.3; Type: Setosa	0.563451951395526				
LeafLength: 5; LeafWidth: 3.4; PetalLength: 1.5; PetalWidth 0.2; Type: Setosa	0.565031011683018				
LeafLength: 4.4; LeafWidth: 2.9; PetalLength: 1.4; PetalWidth 0.2; Type: Setosa	0.608190156228667				
LeafLength: 4.9; LeafWidth: 3.1; PetalLength: 1.5; PetalWidth 0.1; Type: Setosa	0.595241023190498				
LeafLength: 5.4; LeafWidth: 3.7; PetalLength: 1.5; PetalWidth 0.2; Type: Setosa	0.535327329280462				
LeafLength: 4.8; LeafWidth: 3.4; PetalLength: 1.6; PetalWidth 0.2; Type: Setosa	0.575597413742662				
LeafLength: 4.8; LeafWidth: 3; PetalLength: 1.4; PetalWidth 0.1; Type: Setosa	0.599850110198748				
LeafLength: 4.3; LeafWidth: 3; PetalLength: 1.1; PetalWidth 0.1; Type: Setosa	0.602961030602095				
LeafLength: 5.8; LeafWidth: 4; PetalLength: 1.2; PetalWidth 0.2; Type: Setosa	0.492911782489126				
LeafLength: 5.7; LeafWidth: 4.4; PetalLength: 1.5; PetalWidth 0.4; Type: Setosa	0.466060931130659				
LeafLength: 5.4; LeafWidth: 3.9; PetalLength: 1.3; PetalWidth 0.4; Type: Setosa	0.495489116654639				
LeafLength: 5.1; LeafWidth: 3.5; PetalLength: 1.4; PetalWidth 0.3; Type: Setosa	0.541947348403581				

EKSPLORACJA DANYCH

- Obliczanie wsparcia (support) s – częstotliwości lub ilość wystąpień wzorca lub zbioru elementów X w analizowanej encji lub transakcji.

- Dane z pliku Transaction.txt:

Przykład:

- ✓ CZĘSTE > 50%
- ✓ Cukier (80%)
- ✓ Kawa (60%)
- ✓ Jajka (60%)
- ✓ Mleko (40%)
- ✓ Orzeszki (40%)
- ✓ Masło (40%)
- ✓ Chleb (20%)
- ✓ Miód (20%)

ID TRANSAKCJI	ELEMENTY TRANSAKCJI
1	kawa, mleko, cukier, orzeszki
2	kawa, cukier, jajka
3	kawa, chleb, cukier, masło
4	orzeszki, cukier, miód, jajka
5	masło, mleko, jajka

```
file:///D:/szkoła/4rok/MIW/DataExploration/bin/Debug/DataExploration.EXE
kawa : 0,4444444444444444
mleko : 0,4444444444444444
cukier : 0,4444444444444444
orzeszki : 0,3333333333333333
jajka : 0,4444444444444444
chleb : 0,4444444444444444
mas?o : 0,3333333333333333
ser : 0,3333333333333333
```

NEURONOWE MAPY SAMOORGANIZUJĄCE SIĘ

- Budowa neuronowej mapy samoorganizującej się 2D
- Uczenie mapy
- Znajdywanie w mapie wzorców

```
file:///D:/szkoła/4rok/MIW/SOM/bin/Debug/SOM.EXE
SOM demo
Deserializing input
Enter x size of SOM:
4
Enter y size of SOM:
4
Enter narrowing:
1
Enter desired precision
0,001
Creating map
Iteration 0
Iteration 1
Iteration 2
Iteration 3
Iteration 4
Iteration 5
Iteration 6
Iteration 7
\\//////////////////////////////////////////////////////////////////
To show map node press 0
To show map node press 1
To enter element and find closest node press 2
To exit press q
0
[Setosa][Virginica][Virginica][Setosa]
[Versicolor][Virginica][Virginica][Setosa]
[Versicolor][Virginica][Virginica][Setosa]
[Versicolor][Virginica][Virginica][Setosa]
\\//////////////////////////////////////////////////////////////////
To show map node press 0
To show map node press 1
To enter element and find closest node press 2
To exit press q
2
Enter attribute number 0
5,1
Enter attribute number 1
3,5
Enter attribute number 2
1,4
Enter attribute number 3
0,2
5,10530472342111      3,37430493618839      1,71192436608388      0,339608964393142
Class is Setosa
```

DZIĘKUJĘ ZA UWAGĘ