

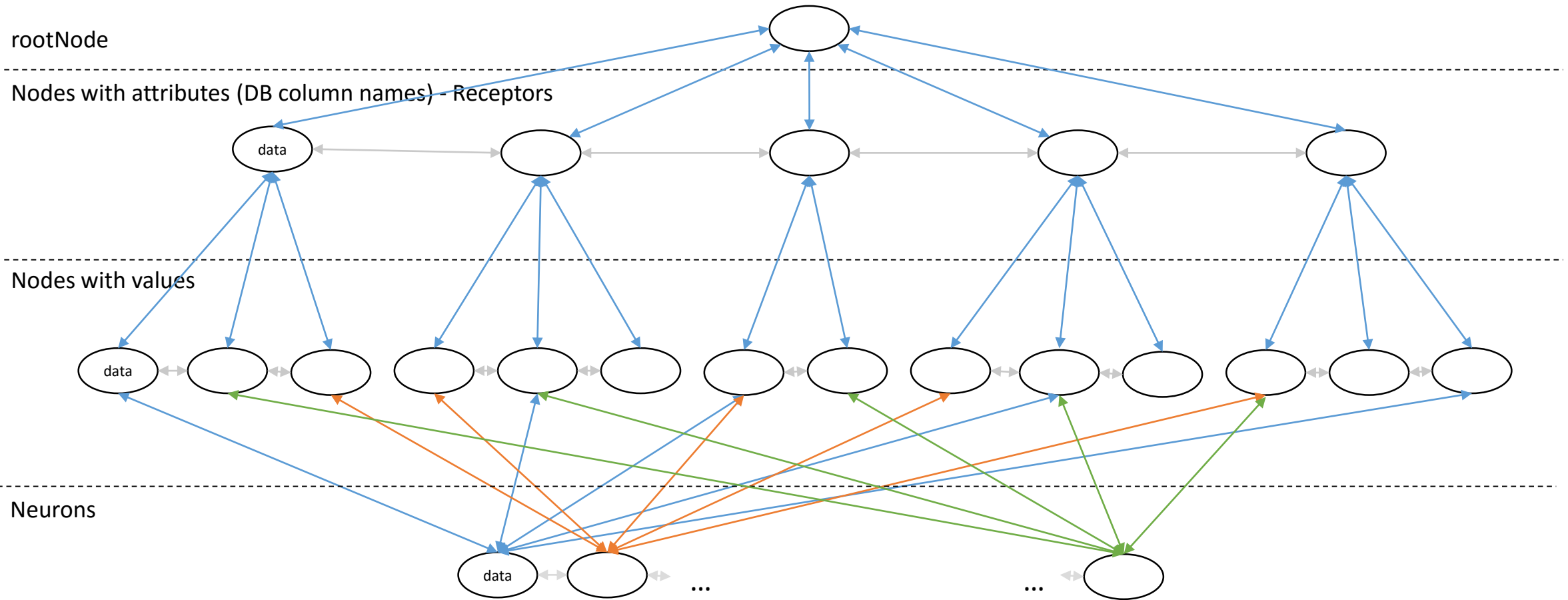
Projekt 2

Asocjacyjna reprezentacja powiązanych tabel i wnioskowanie

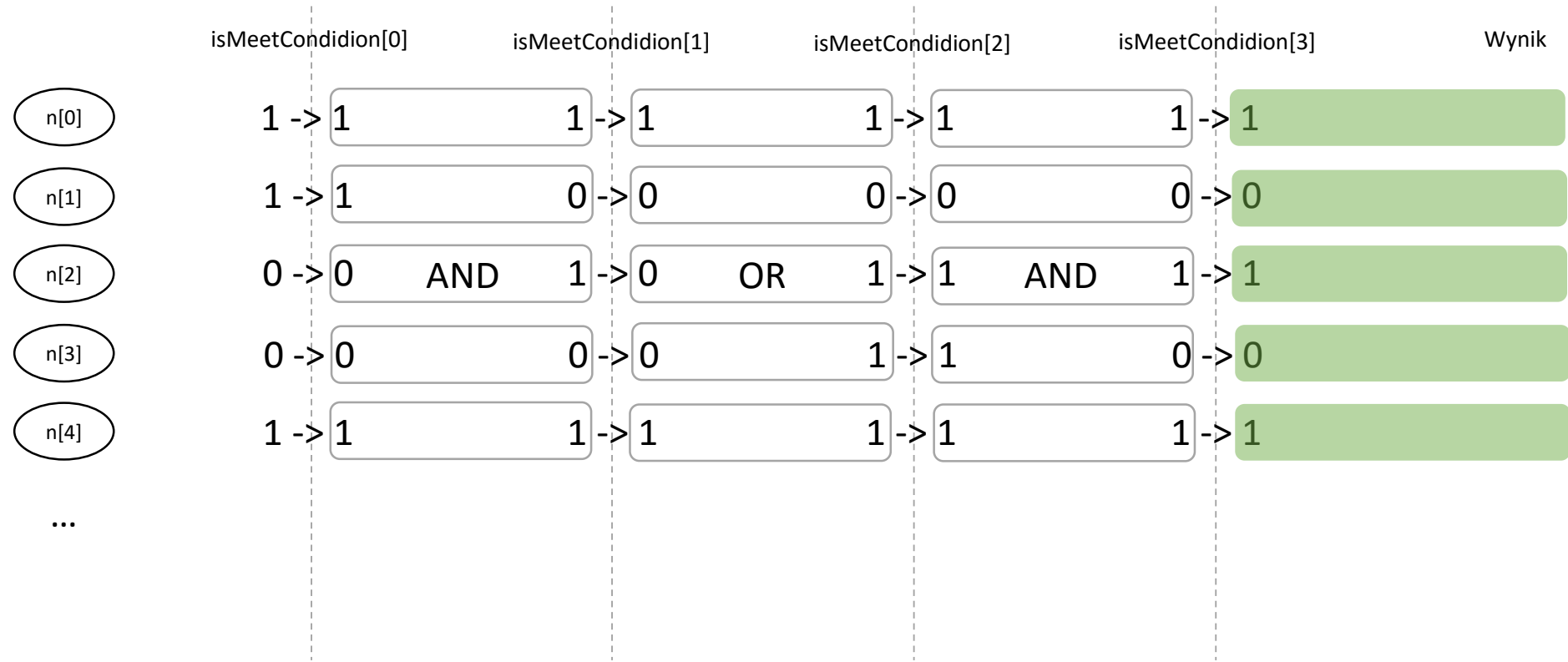
Program napisany w języku C# (Visual Studio 2015) z użyciem relacyjnej bazy danych SQL Server 2014

<https://github.com/MarcinKasprowicz/knowledge-engineering/>

Budowa grafu AGDS



Asocjacyjne wyrażenia logiczne



Program

https://github.com/MarcinKasprowicz/knowledge-engineering/tree/master/6-AGDS_BD

- Wywołanie

```
.\AANG.exe "Data Source=DESKTOP-VGMN9F4\SQLEXPRESS;Initial Catalog=master;Integrated Security=True" "SELECT * FROM country"
```

- \$1 – Pierwszy parametr - SQL Server connection string.
- \$2 – Drugi parametr - Zapytanie SQL SELECT na podstawie którego zostanie zbudowany graf.

```
PS E:\Studia\8. Data engineering\1. KNN algorithm\6-AGDS_BD\AANG\AANG\bin\Debug> .\AANG.exe "Data Source=DESKTOP-VGMN9F4\SQLEXPRESS;Initial Catalog=master;Integrated Security=True" "SELECT * FROM country"
Data set loaded in 00:00:00.0540347
Neural Network built in 00:00:00.4431200
Please select what you want to do:
[f]ilter      [b]enchmark  [q]uit
b
Write query using SQL WHERE Syntax with mathematical operators:
IndepYear > 1947 AND GovernmentForm = 'Republic'

SQL Result - 80 rows

Avarage execution time for DB method - 963,317717717718 mikrosekund

AGDS Result - 80 rows

Avarage execution time for AANG method - 129,811611611612 mikrosekund

Please select what you want to do:
[f]ilter      [b]enchmark  [q]uit
q
```

Porównanie czasów wykonania

Podział na ilość kryteriów filtrowania

- Dataset – studenci – 9 rekordów

| | 1 (μ s) | 2 | 3 | 4 | 5 |
|---------|--------------|---------|---------|---------|---------|
| DB | 1032,63 | 1027,01 | 1039,66 | 1018,28 | 1015,51 |
| AGDS | 14,60 | 19,88 | 25,84 | 28,25 | 29,68 |
| DB/AGDS | ~70,7 | ~51,6 | ~40,2 | ~36,0 | ~34,21 |

- Dataset – world – 239 rekordów

<http://dev.mysql.com/doc/index-other.html>

| | 1 (μ s) | 2 | 3 | 4 | 5 |
|---------|--------------|---------|---------|---------|---------|
| DB | 1298,13 | 1178,09 | 1501,95 | 1537,05 | 1546,67 |
| AGDS | 106,97 | 255,12 | 281,54 | 437,09 | 451,84 |
| DB/AGDS | ~12,1 | ~5,2 | ~5,3 | ~3,5 | ~3,4 |

Porównanie czasów wykonania

Podział na ilość kryteriów filtrowania

- Dataset – Bike sharing (day) – 731 rekordów

<https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>

| | 1 (μ s) | 2 | 3 | 4 | 5 |
|---------|--------------|---------|---------|---------|---------|
| DB | 2910,13 | 3475,28 | 1953,23 | 1407,96 | 3663,01 |
| AGDS | 186,30 | 243,08 | 264,29 | 650,31 | 1196,95 |
| DB/AGDS | ~15,6 | ~14,3 | ~7,4 | ~2,1 | ~3,0 |

- Dataset – AirQualityUCI – 9357 rekordów

budowanie grafu trwa około 10 minut

<https://archive.ics.uci.edu/ml/datasets/Air+Quality>

| | 1 (μ s) | 2 | 3 | 4 | 5 |
|---------|--------------|----------|----------|----------|----------|
| DB | 40927,79 | 93006,25 | 54546,36 | 39664,65 | 87362,20 |
| AGDS | 3263,25 | 6723,15 | 7365,40 | 8032,78 | 15830,01 |
| DB/AGDS | ~12,5 | ~13,8 | ~7,4 | ~4,9 | ~5,5 |

Wnioski

- Przy bazie danych składających się z 10 tyś rekordów budowa grafu trwa 10 minut. Przy 20 tyś budowa trwa więcej niż 30 minut.
- Wyszukiwanie rekordów w asocjacyjnym grafie jest o wiele szybsze niż w relacyjnej bazie danych!
- Przy lepszej implementacji wnioskowania wyrażenia logicznego AND wyniki prawdopodobnie byłyby lepsze.