

AGH

*System szybkiego inteligentnego asocjacyjnego
wyszukiwania relacji pomiędzy danymi
wykorzystujący asocjacyjne grafowe struktury
danych AGDS.*

Autor: Paulina Sopata

Promotor: dr hab. Adrian Horzyk

Plan prezentacji



1. Cel i zakres pracy.

2. Wprowadzenie teoretyczne:

- asocjacyjne grafowe struktury danych AGDS,
- złożoność obliczeniowa.

3. Eksperymenty i omówienie wyników.

4. Podsumowanie.

Cel i zakres pracy

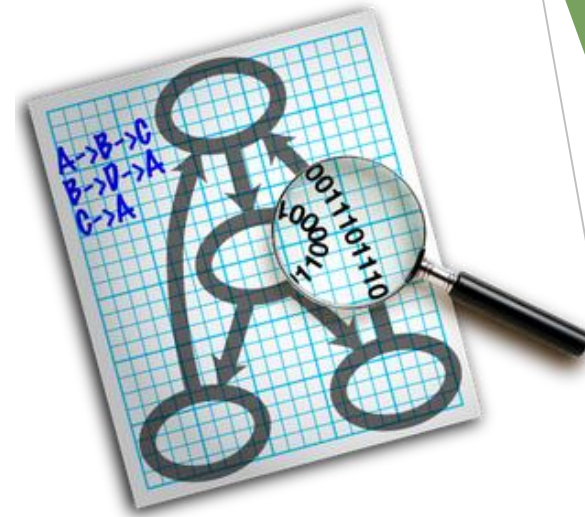


Cel pracy: Praca ma na celu opracowanie oraz implementację asocjacyjnego modelu danych, do którego transformowane są klasyczne tabele, pozwalając na szybkie (zwykle w czasie stałym $O(1)$) wnioskowanie na temat różnego rodzaju relacji automatycznie dostępnych w asocjacyjnych grafowych strukturach danych AGDS, bez konieczności żmudnego przeszukiwania danych w pętlach obliczeniowych.

Zakres pracy: Zakres pracy obejmuje również porównanie szybkości działania operacji na tych strukturach z operacjami wykonywanymi na klasycznych strukturach tabelarycznych, wraz z analizą oraz wyprowadzeniem wniosków na temat złożoności obliczeniowej.

Implementacja

- C++, QT Creator
- Graphviz, draw.io
- UCI Machine Learning Repository



draw.io

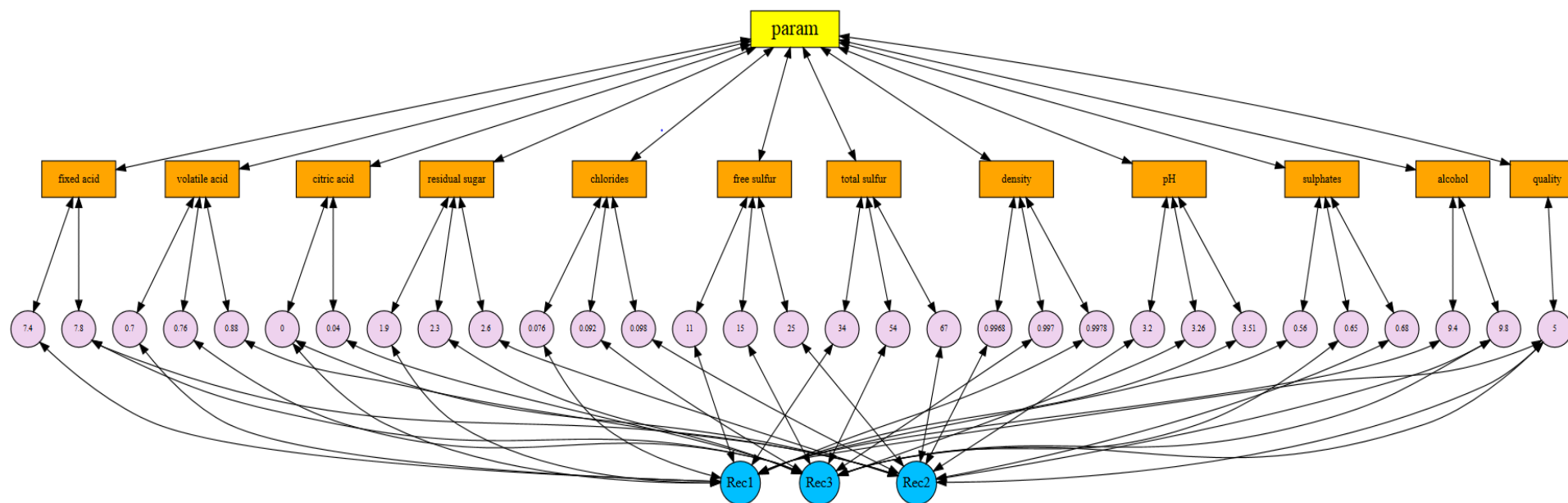


Asocjacyjne grafowe struktury danych AGDS



- Grafy AGDS mogą być wykorzystywane jako forma reprezentacji oraz przechowywania danych.
- Służą do znajdowania i przechowywania asocjacyjnych relacji pomiędzy danymi, obiektami i zdefiniowanymi na ich podstawie informacjami.
- Są wykorzystywane do łatwego oraz szybkiego znajdowania powiązanych danych oraz ich kombinacji, korelacji, podobieństw i różnic.
- Poprzez formę, w jakiej przetrzymywane są dane (brak duplikatów), bez strat kompresują i kontekstowo wiążą informacje, szczególnie w przypadku dużych zbiorów danych.

Asocjacyjne grafowe struktury danych AGDS



Złożoność obliczeniowa algorytmu

Powiązane zagadnienia

Złożoność czasowa - ilość czasu potrzebnego do wykonania zadania, wyrażona jako funkcja ilości danych.

Złożoność pesymistyczna $O()$ - ilość zasobów potrzebnych do wykonania algorytmu przy założeniu najbardziej "złośliwych", czyli najgorszych, danych.

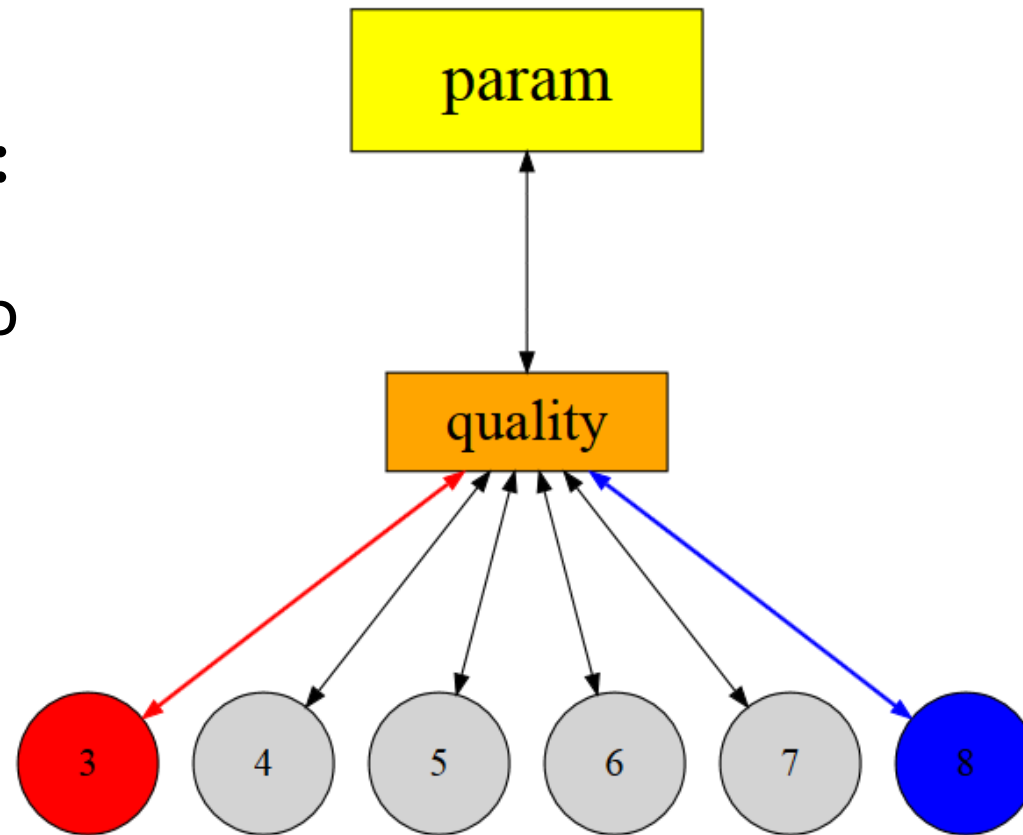
Złożoność obliczeniowa stała $O(1)$ - gdy czas wykonania algorytmu jest stały i niezależny od rozmiaru danych wejściowych.

Złożoność obliczeniowa liniowa $O(n)$ - czas działania jest liniowo proporcjonalny do rozmiaru danych wejściowych.

Eksperymenty

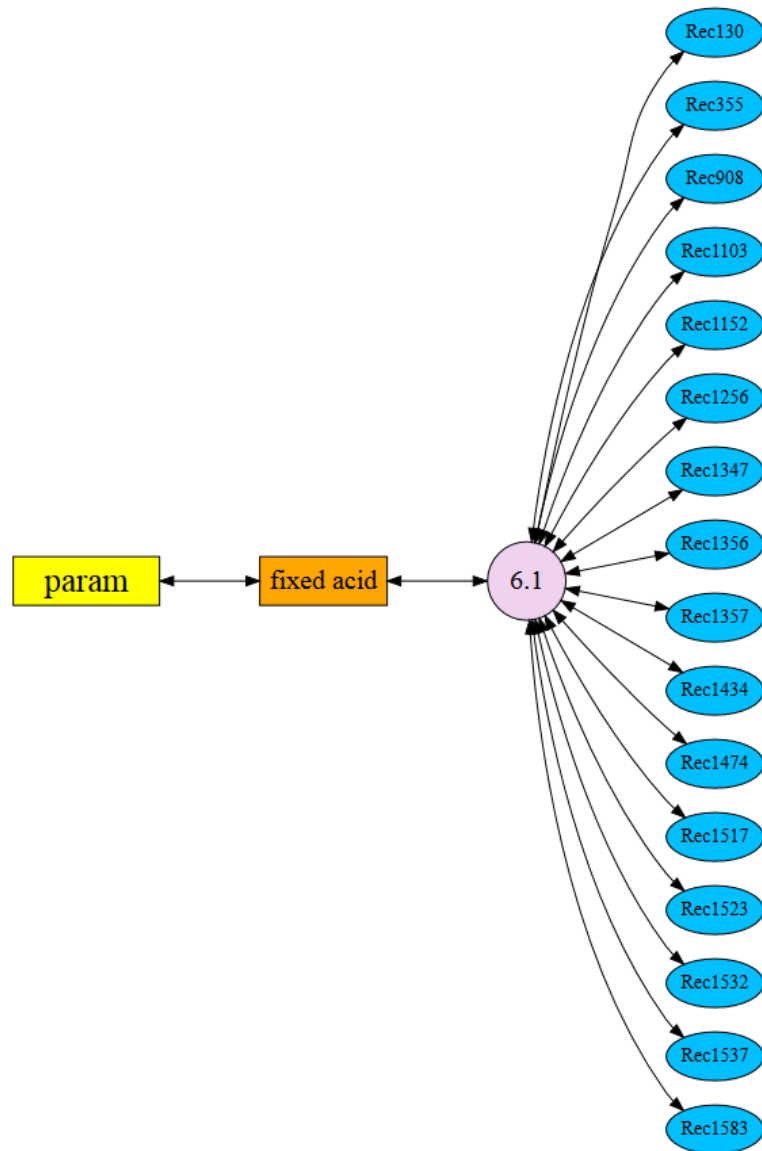
Minimum, maksimum, zakres

Złożoność obliczeniowa:
Graf: $O(1)$
Tabela: $O(n)$, $O(\log n)$ lub $O(1)$ w zależności od operacji i posortowania



Eksperymenty

Wyszukiwanie rekordów o określonej wartości



Złożoność obliczeniowa:

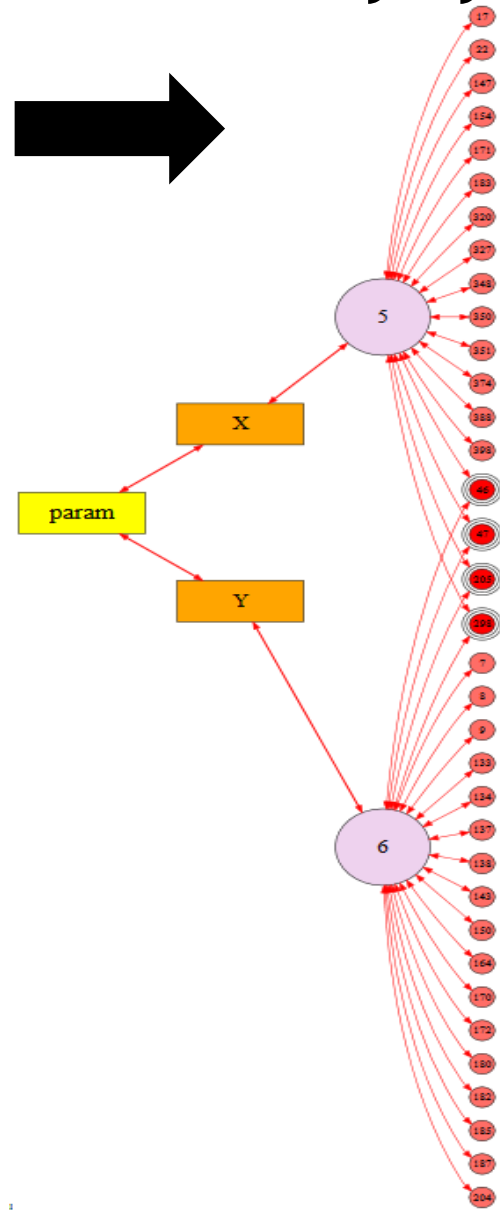
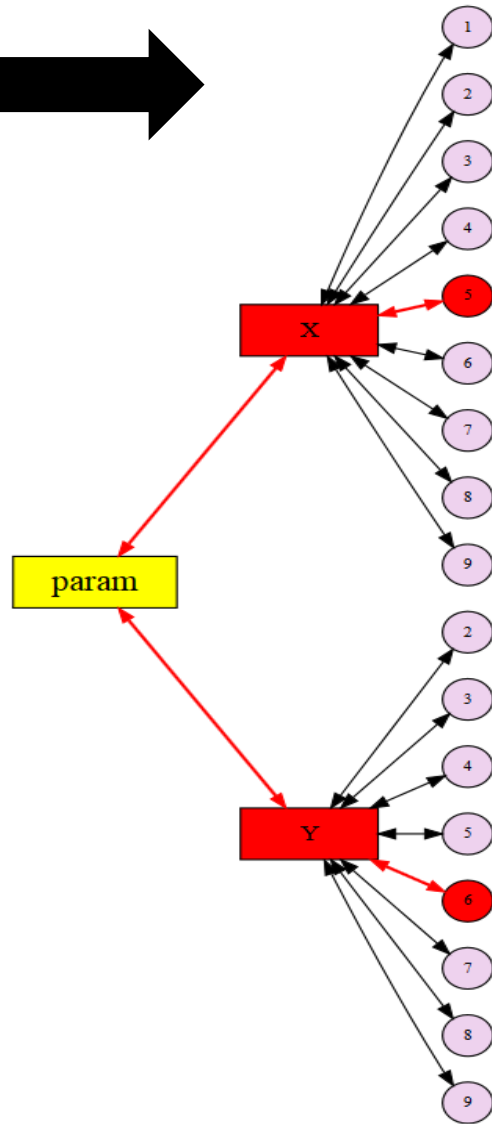
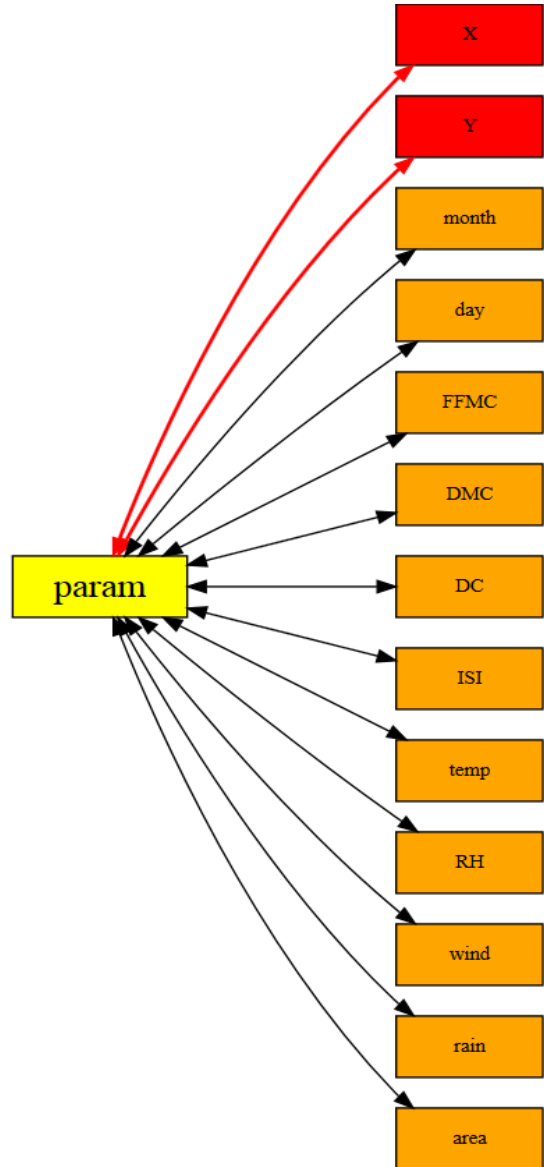
Graf: $O(1)$ lub $O(\log n)$

Tabela: $O(n)$ lub $O(\log n)$ w zależności od posortowania

Operacje były wykonywane na nieposortowanych tabelach. Jeżeli tablice posortujemy, a wyszukiwanie zostanie przeprowadzone np. wyszukiwaniem połówkowym, to złożoność obliczeniowa operacji na tabelach będzie wynosić w najlepszym przypadku $O(\log n)$. Jeżeli chodzi o operacje wykonywane na grafach złożoność obliczeniowa w najlepszym przypadku wynosi $O(1)$, a w najgorszym $O(\log n)$.

Eksperymenty

Wyszukiwanie relacji koniunkcji i alternatywy



Eksperymenty

Wyszukiwanie relacji koniunkcji i alternatywy



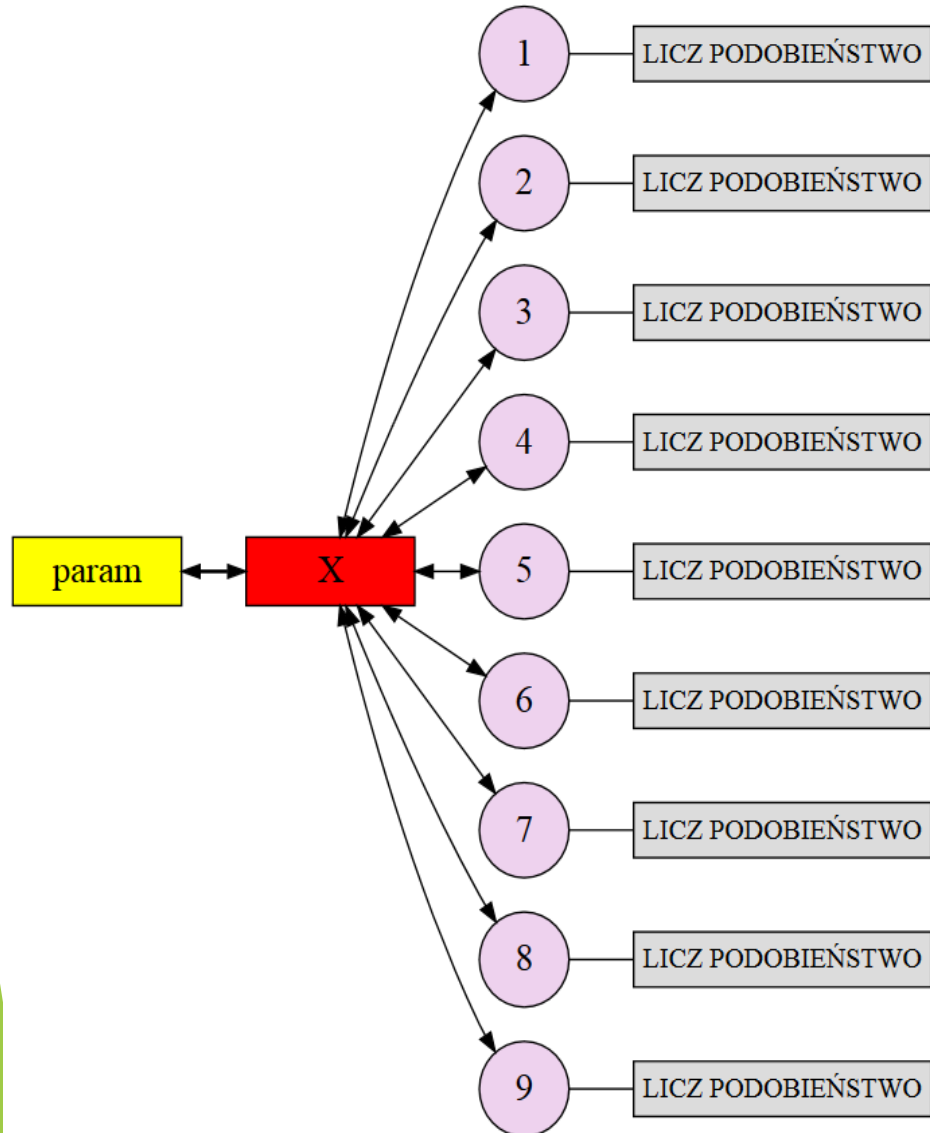
Złożoność obliczeniowa:

- Graf: $O(n)$
- Tabela: $O(n)$

Dlaczego pomimo tej samej złożoności obliczeniowej korzystniejsze jest wykonywanie operacji w grafie?

Eksperymenty

Obliczanie podobieństwa względem atrybutu



Wzór na podobieństwo:

$$P = 1 - \frac{|Val_A - Val_B|}{MAX - MIN}$$

Złożoność obliczeniowa:

Graf: $O(n)$

Tabela: $O(n)$

Podsumowanie



- Złożoność obliczeniowa algorytmów wyszukiwania relacji na grafowych strukturach danych jest korzystniejsza niż w przypadku operacji wykonywanych na klasycznych tabelach. W przedstawionych przypadkach jest stała lub liniowa .
- Czasy wykonywania operacji na grafowych strukturach danych są dużo mniejsze

Podsumowanie

Ile razy korzystniejszy jest czas wyszukiwania relacji w grafowej strukturze danych niż w tabeli?



	ForestFires	RedWine	OnlineNewsPopularity
Minimum, maksimum, zakres	5-krotnie korzystniejszy	15-krotnie korzystniejszy	1000-krotnie korzystniejszy
Wyszukiwanie rekordów o określonej wartości	2,5-krotnie korzystniejszy	2,5-krotnie korzystniejszy	700-krotnie korzystniejszy
Wyszukiwanie relacji koniunkcji i alternatywy	3,4-krotnie korzystniejszy	2,3 oraz 5-krotnie korzystniejszy	2,6-krotnie korzystniejszy
Obliczanie podobieństwa względem atrybutu	2,3 oraz 8,5-krotnie korzystniejszy	30-krotnie korzystniejszy	290 oraz 365-krotnie korzystniejszy



Dziękuję za uwagę.