

# Laboratorium z przedmiotu Data mining

Lab2 – reguły sekwencyjne

# Pojęcia podstawowe

- Transakcja – podzbiór elementów z pewnej dziedziny elementów z przypisanym pewnym znacznikiem czasowym.
- Sekwencja transakcji – zbiór transakcji dotyczący jednego obiektu
- Baza danych (BD) – zbiór sekwencji transakcji.

# Reguły sekwencyjne

Reguła sekwencyjna to wyrażenie w postaci:

$$A \Rightarrow B$$

gdzie A i B są sekwencjami

A – poprzednik reguły

B - następnik reguły

**Reguła sekwencyjna wyraża pewne  
następstwo czasowe zdarzeń.**

# Parametry reguły sekwencyjnych (1)

Parametry dostępne w pakiecie `arulesSequences`

- **Wsparcie** (ang. support)
- **Zaufanie** (ang. confidence)
- **Współczynnik podniesienia** (ang. lift)

# Parametry reguły sekwencyjnych (2)

Parametry dostępne w pakiecie `arulesSequences`:

- **maxsize** – maksymalna liczba elementów w składowej reguły, wartość domyślna 10
- **maxlen** – maksymalna liczba składowych reguły, wartość domyślna 10
- **mingap** – minimalna odległość w czasie pomiędzy dwoma kolejnymi składowymi reguły
- **maxgap** – maksymalna odległość w czasie pomiędzy dwoma kolejnymi składowymi reguły
- **maxwin** – *maksymalna odległość w czasie pomiędzy dwoma dowolnymi składowymi reguły (okno czasowe reguły)*

# Wsparcie

**Wsparcie reguły sekwencyjnej jest liczone jako liczba sekwencji transakcji, w których ta reguła występuje.**

Nie jest ważne, czy dana reguła występuje jeden, czy więcej razy w danej sekwencji transakcji.

wsparcie bezwzględne reguły  $(A \Rightarrow B) =$   
liczba sekwencji transakcji w BD zawierających  
 $A \cup B$  z narzuconymi warunkami czasowymi

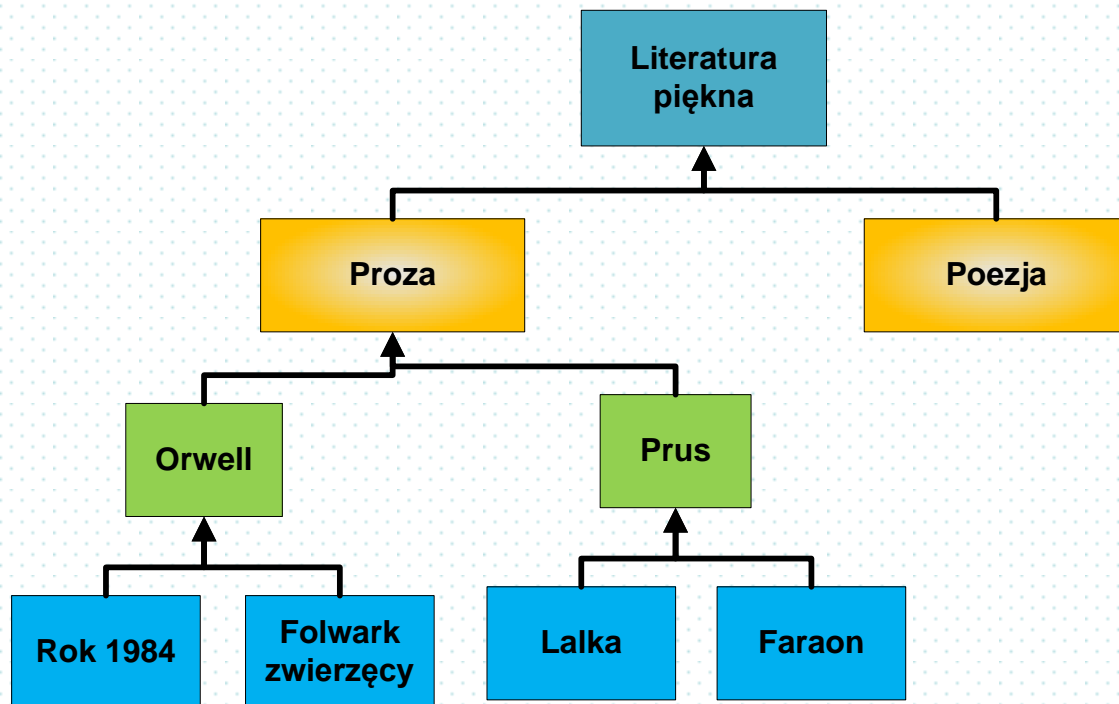
**Wsparcie względne** – częstość występowania reguły w bazie.

wsparcie względne  $(A \Rightarrow B) =$   
wsparcie bezwzględne  $(A \Rightarrow B)$  / liczba sekw. trans. w BD

# Wsparcie, zaufanie, współczynnik podniesienia

Interpretacja wsparcia i zaufania oraz współczynnika podniesienia dla reguł sekwencyjnych jest analogiczna do interpretacji tych wskaźników dla reguł asocjacyjnych.

# Hierarchia



**Zastosowanie hierarchii daje takie same rezultaty jak w przypadku reguł asocjacyjnych, czyli pozwala na odkrycie bardziej ogólnych reguł, odnoszących się do różnych poziomów.**



# Odkrywanie reguł przy użyciu oprogramowania R

Pakiet ***arulesSequence*** - wybrane metody:

- ***read\_baskets*** – wczytanie danych z sekwencjami
- ***itemFrequency*** – wyliczenie częstości występowania elementów
- ***ruleInduction*** – generowanie reguł sekwencyjnych ze zbioru sekwencji
- ***cspade*** – odkrywanie wzorców sekwencyjnych (częstych sekwencji) – algorytm SPADE,
- ***inspect*** – pokazanie reprezentacji tekstowej reguł/wzorców
- ***subset*** – selekcja reguł/wzorców sekwencyjnych