



Zależności funkcyjne

Niech D = (O, AT), gdzie O jest zbiorem obiektów, a AT – zbiorem ich atrybutów. Niech $A,B\subseteq AT$.

- $A \rightarrow B$ jest definiowana jako *zależność funkcyjna*, jeśli $\forall x \in \mathcal{O}, [x]_A \subseteq [x]_B$.
- Zależność funkcyjna A→B jest zwana minimalną, jeśli ∀C⊂A, C→B nie jest zależnością funkcyjną.
- Zależność A→B jest zwana trywialną, jeśli B⊆A.

3



Przykład: Zależności funkcyjne

D = (O, AT), gdzie O = {1,2,3,4,5,6}, AT = {abcd}.

	u	~	•	-
1	1	1	2	2
2	1	1	2	2
3	2	2	2	2
4	3	3	3	3
5	5	5	6	6

6 5 5 6 6

- {cd} → {ab} nie jest zależnością funkcyjną.
- {ab} → {cd} jest zależnością funkcyjną, ale nie jest minimalną.
- $\{a\} \rightarrow \{cd\}$ jest minimalną zależnością funkcyjną.
- $\{ab\} \rightarrow \{a\}$ jest zależnością trywialną.

4



Własności zależności funkcyjnych

Twierdzenie (Podstawowe reguły stosowane w algorytmie TANE). Niech $a \in X$ i $b \notin X$.

- r1: Jeśli X \ {a} → {a} jest zależnością funkcyjną,
 to ∀Y⊃X, Y \ {a} → {a} jest zależnością funkcyjną, ale nie jest minimalną.
- r2: Jeśli $X \setminus \{a\} \to \{a\}$ i $X \to \{b\}$ są zależnościami funkcyjnymi, to $\forall Y \supset X$, $Y \setminus \{b\} \to \{b\}$ jest zależnością funkcyjną, ale nie jest minimalną.

5



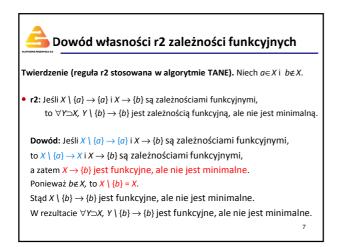
Dowód własności r1 zależności funkcyjnych...

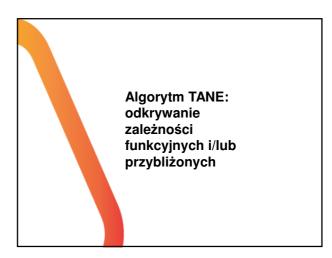
Twierdzenie (reguła r1 stosowana w algorytmie TANE). Niech $a \in X$.

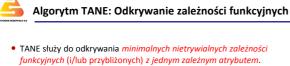
 r1: Jeśli X \ {a} → {a} jest zależnością funkcyjną, to ∀Y⊃X, Y \ {a} → {a} jest zależnością funkcyjną, ale nie jest minimalną.

Dowód: Ponieważ $a \in X$, to $\forall Y \supset X$, $Y \setminus \{a\} \supset X \setminus \{a\}$. Stąd $X \setminus \{a\} \rightarrow \{a\}$ jest zależnością funkcyjną. W rezultacie $\forall Y \supset X$, $Y \setminus \{a\} \rightarrow \{a\}$ jest zależnością funkcyjną, ale nie jest minimalną.

6

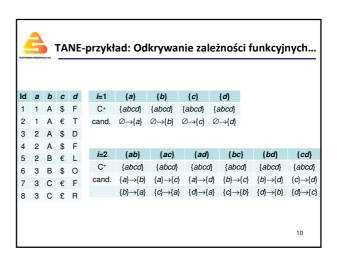


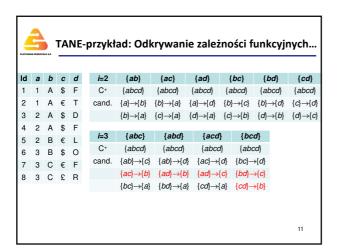


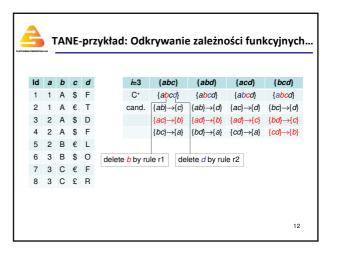


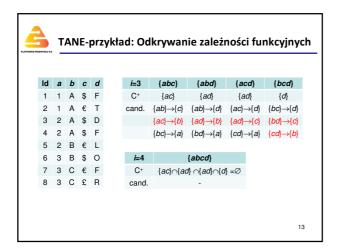
- funkcyjnych (i/lub przybliżonych) z jednym zależnym atrybutem.
- W i-tej iteracji algorytmu TANE, tworzone są zależności kandydujące $X \setminus \{a\} \to \{a\}$ ze zbiorów atrybutów X o liczności i.
- Nowe zbiory atrybutów o liczności i + 1 są tworzone ze zbiorów atrybutów o liczności i, tak jak w algorytmie Apriori.
- Podstawowe reguły r1 i r2 są używane w celu uniknięcia wytwarzania zbiorów atrybutów o liczności i+1, z których nie jest możliwe uzyskanie żadnych minimalnych nietrywialnych zależności funkcyjnych.

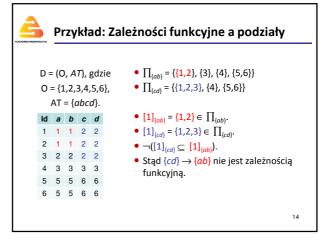
9

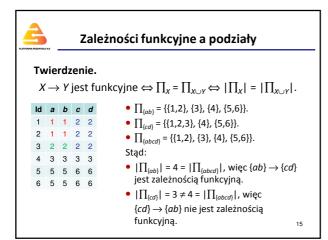


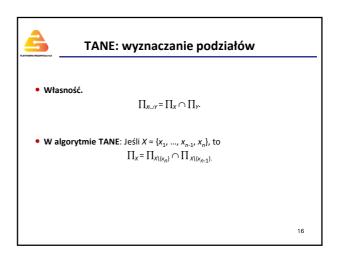


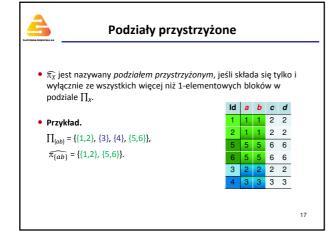


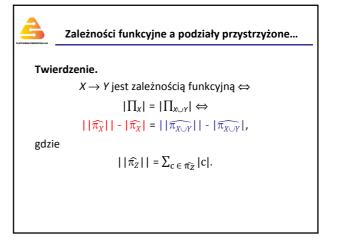


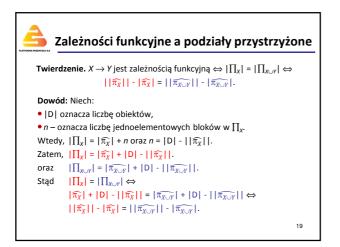


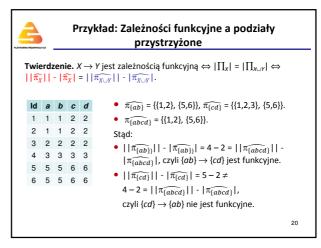




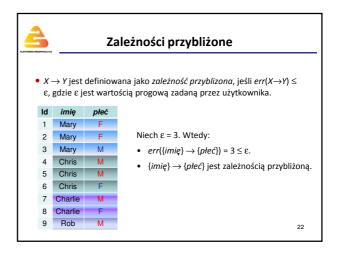


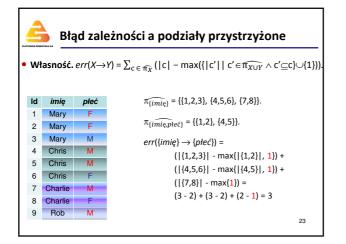


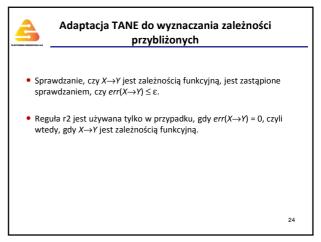














Literatura

 Ykä Huhtala, Juha Kärkkäinen, Pasi Porkka, Hannu Toivonen: TANE: An Efficient Algorithm for Discovering Functional and Approximate Dependencies. <u>Comput. J. 42</u>(2): 100-111 (1999)



Ćwiczenia...

- 1. Wyznacz podziały $\Pi_{(o)}$, $\Pi_{(b)}$ i $\Pi_{(ob)}$ na podstawie tabeli na slajdzie 10. Użyj tych podziałów do uzyskania odpowiedzi na następujące pytania:
 - Czy {a} → {b} jest zależnością funkcyjną?
 - Czy $\{b\} \rightarrow \{a\}$ jest zależnością funkcyjną?
 - Jaka jest wartość błędu zależności err({a} → {b})?
 - Jaka jest wartość błędu zależności $err(\{b\} \rightarrow \{a\})$?
 - Niech ε = 2. Czy $\{a\} \rightarrow \{b\}$ jest zależnością przybliżoną?
 - Niech $\varepsilon = 2$. Czy $\{b\} \rightarrow \{a\}$ jest zależnością przybliżoną?

25



Ćwiczenia

- 2. Wyznacz podziały przystrzyżone $\widehat{\pi_{\{a\}}}$, $\widehat{\pi_{\{b\}}}$ i $\widehat{\pi_{\{ab\}}}$ na podstawie tabeli na slajdzie 10. Użyj tych podziałów do uzyskania odpowiedzi na następujące pytania:
 - Czy $\{a\} \rightarrow \{b\}$ jest zależnością funkcyjną?
 - Czy $\{b\} \rightarrow \{a\}$ jest zależnością funkcyjną?
 - Jaka jest wartość błędu zależności err({a} → {b})?
 - Jaka jest wartość błędu zależności $err(\{b\} \rightarrow \{a\})$?
 - Niech ε = 1. Czy $\{a\} \rightarrow \{b\}$ jest zależnością przybliżoną?
 - Niech ε = 1. Czy $\{b\} \rightarrow \{a\}$ jest zależnością przybliżoną?