P114
Rozložitelnost

Manipulace s HIT-atributy

10

Témata

- Podatributy
- Rozklad atributů
- Jak poznat rozložitelnost
- Věta o rozkladu
- Jádro datového modelu

Zopakování

- Atribut A je definovatelný nad množinou atributů $\{B_1, ..., B_n\}$ právě když
  $\exists f\ \forall wt\ ([A\ wt] = [f\ ([B_1\ wt], ..., [B_n\ wt])])$
  kde f je analytická funkce splňující základní předpoklad, w je možný svět, t je časový okamžik.
  Píšeme: $A \leftarrow (B_1, ..., B_n)$
- Množina atributů M je definovatelná nad množinou atributů N, je-li každý prvek z M definovatelný nad nějakou podmnožinou N.
  Píšeme: $M \leftarrow N$
- Atribut A je definovatelný nad atributem B, když je definovatelný nad množinou $\{B\}$.
  Píšeme: $A \leftarrow B$

Informační ekvivalence (zopakování)

- Nechť $M = \{A_1, ..., A_n\}$, $N = \{B_1, ..., B_m\}$ jsou množiny atributů a nechť $M \leftarrow N$ a zároveň $N \leftarrow M$. Potom říkáme, že M a N jsou informačně ekvivalentní. Píšeme: $M \approx N$
- Intuitivně je zřejmé, a lze dokázat, že dvě informačně ekvivalentní množiny atributů generují v každém stavu světa stejnou třídu propozic.
- Jiná intuice: z n-tice tabulek daných množinou atributů M dostaneme tutéž informaci jako z m-tice tabulek daných množinou atributů N
- VĚTA: Relace $\approx$ na množině atributů je ekvivalence.

Triviální odvození (zopakování)

- Řekneme, že atribut A vznikl z atributu B triviálním odvozením, jestliže A $\leftarrow$ B, tj. pro všechna w
  $A_w = [f\ B_w]$
  a funkce **f** obsahuje nejvýše identitu, připustně použitý singularizátor a/nebo existenční kvantifikátor.
- Triviální odvození jsou nejčastěji používána jako nástroj odvození odpovědi na dotazy nad databází. Pro formulování dotazů a transformací atributů je vhodnější využívat všech funkcí jednoduchých typů (přednáška 6) a neomezovat se pouze na H-typy.

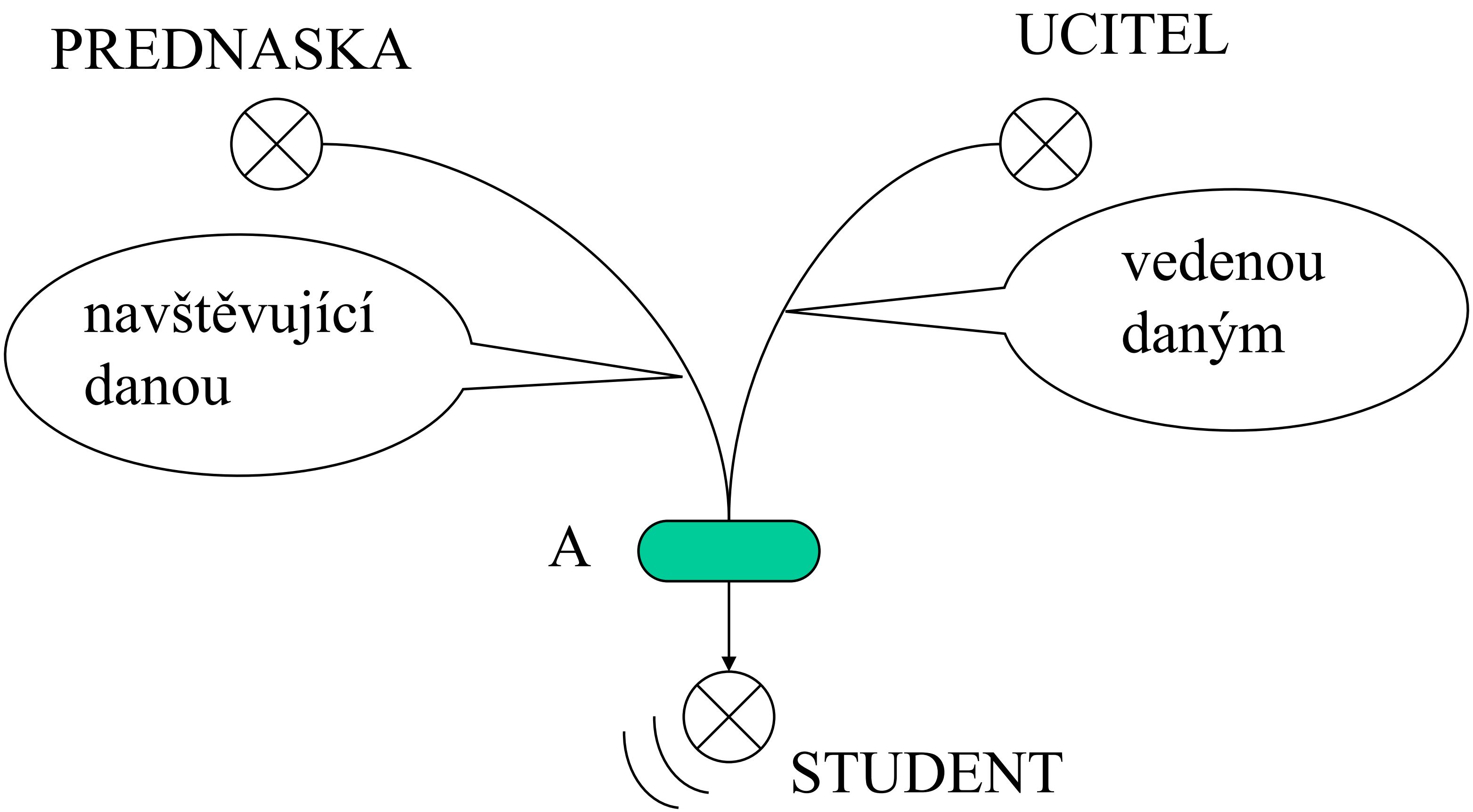
Podatribut

- Atribut A_1 vzniklý triviálním odvozením z atributu A budeme nazývat **podatributem** A.
  Každý podatribut, který není totožný s některou přípustnou rotací A, nazýváme **vlastní podatribut**.
- Zejména A je podatributem A.
- Složitost vlastního podatributu A_1 je vždy menší než složitost A.
- Všechny rotace a odvozené atributy v posledním příkladu přednášky 9 jsou podatributy.

Rozklad atributu

- Atribut A je **rozložitelný** na atributy $B_1, ..., B_n$, píšeme $A \lozenge (B_1, ..., B_n)$, když
  1) $B_1, ..., B_n$ jsou vlastní podatributy A, a
  2) $A \leftarrow (B_1, ..., B_n)$
- **DŮSLEDEK 1**:
  Jestliže $A \lozenge (B_1, ..., B_n)$, potom $A \approx \{B_1, ..., B_n\}$.
  DŮKAZ: plyne z definic.
- **DŮSLEDEK 2**:
  Jestliže $A \lozenge (B_1, ..., B_n)$, potom každá jeho přípustná rotace rotA je rovněž rozložitelná na $B_1, ..., B_n$.
  DŮKAZ: plyne z definice rotace, definice rozložitelnosti a věty 1 (viz přednáška 9).

Příklad: A/(W→ ((Prednaska, Ucitel)→ Student))



příklad - pokračování

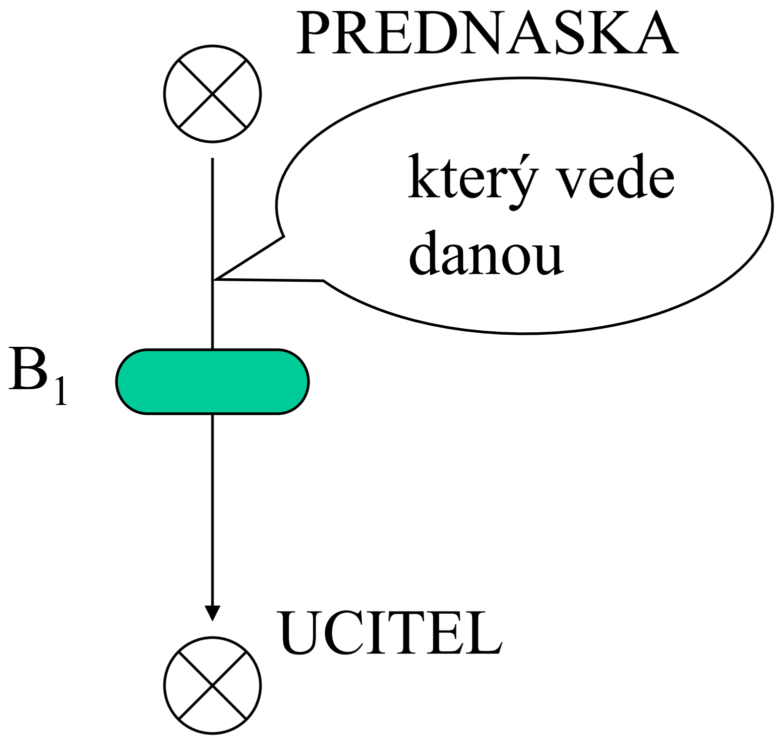
- uvažme atribut:
  B_1 = (#Ucitel)-s kteří vedou danou (#Prednaska)
  opravdu je jich víc?
  Nechť realita je, že danou přednášku vede vždy jenom jeden učitel. Potom bude správně:
  B_1 = (#Ucitel) který vede danou (#Prednaska)
- $B_1 \leftarrow A$ triviálně (pomocí =, $\exists$ a ι ve **f**)
- přidejme:
  B_2 = (#Student)-s kteří navštěvují danou (#Prednaska) /0,M:0,M
- $B_2 \leftarrow A$ triviálně (pomocí = a $\exists$ ve **f**)
- dokážeme napsáním algoritmu, že $A \leftarrow (B_1, B_2)$:

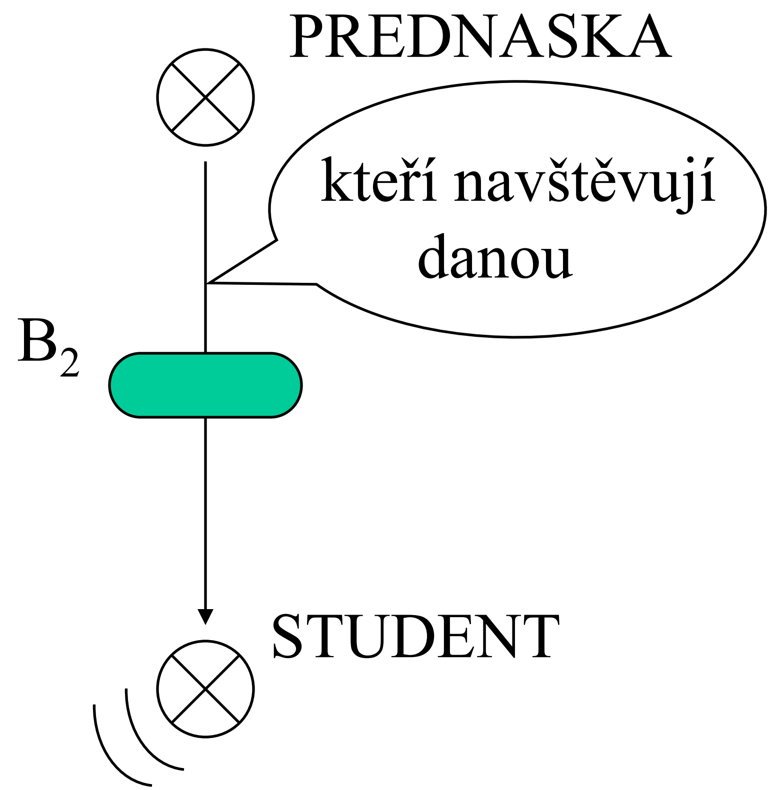
příklad - pokračování

- declare p:Prednaska, u:Ucitel
  declare s:Student
  for all p do
      for all u do
          for all s do
              if $[[B_{2w}\ p]\ s] \wedge [B_{1w}\ p] = u$
                  then write s
                  else endif
          endfor
      endfor
  endfor

- ve f stačila identita a konjunkce ...

$A \lozenge (B_1, B_2),\ A \approx \{B_1, B_2\}$





Poznámka o rozkladové konstrukci

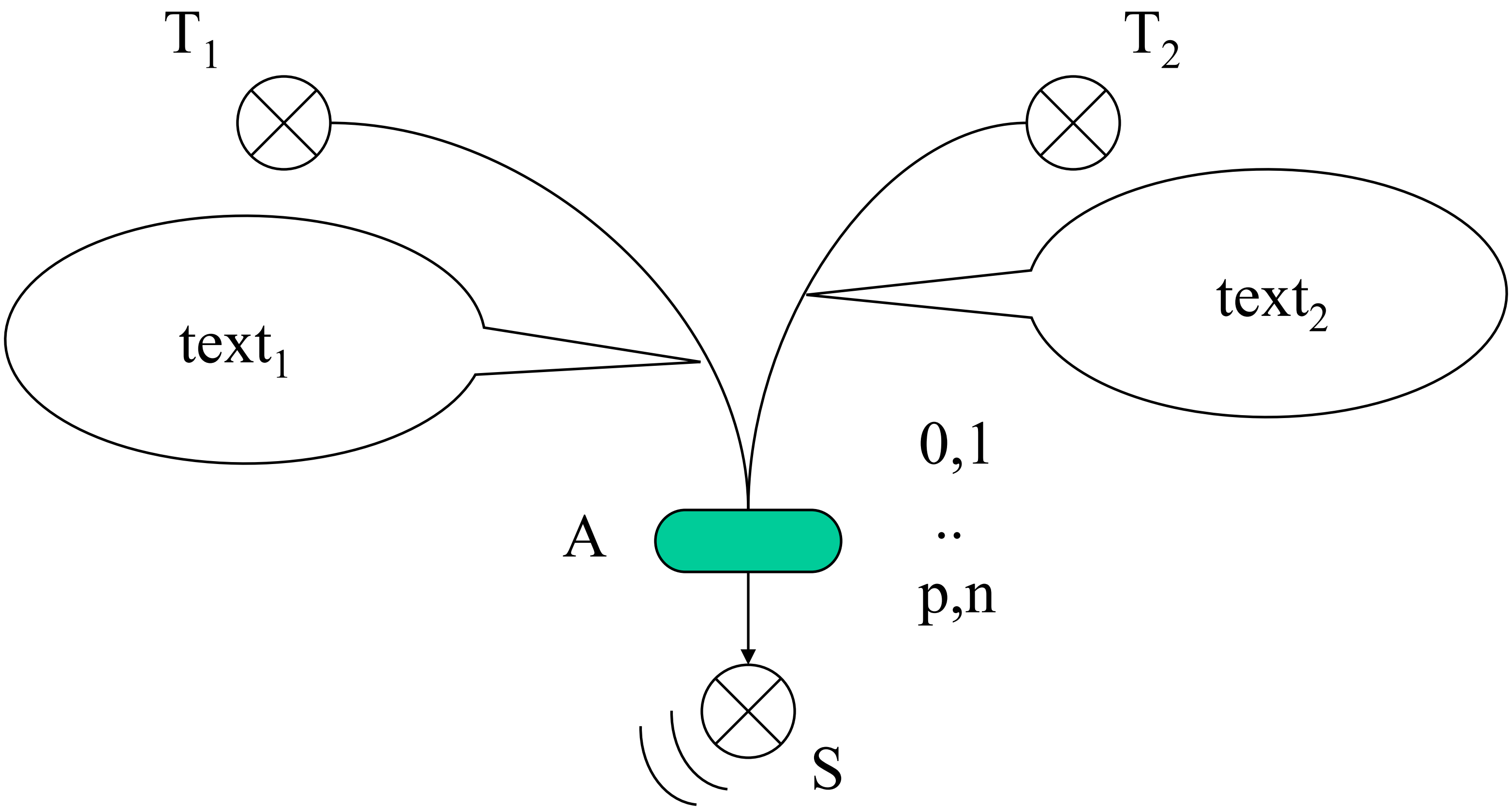
- Nechť A $\Diamond$ (B_1, ..., B_n), a nechť
  $A_w = [f\,(B_{1w}, ..., B_{nw})]$.

- Potom vždy f obsahuje konjunkce a identity
- V případě, že A je zadán v singulární rotaci, pak f obsahuje navíc singularizátor.

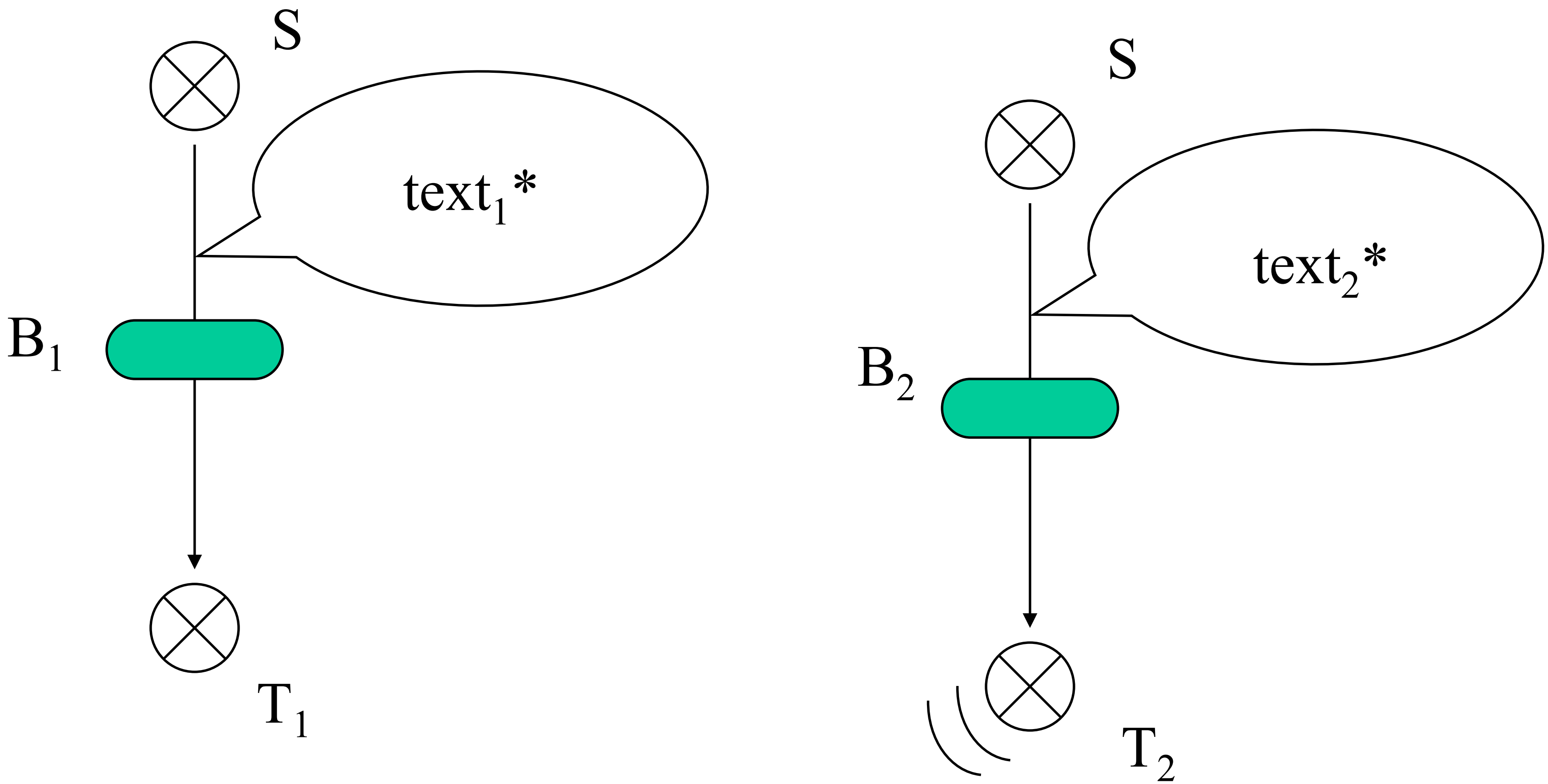
Jak poznat rozložitelnost

- (1) použitím definice:
  v diskusi s expertem se často k atributu složitosti > 2 objeví jeho podatribut; pak hledáme další podatribut původního a pokoušíme se zkonstruovat (algoritmem) původní atribut (viz min. příklad)
- (2) podle horního poměru
  je-li poměr atributu
  $A / (W \rightarrow ((T_1, ..., T_n) \rightarrow S))$ resp.
  $A / (W \rightarrow ((T_1, ..., T_n) \rightarrow (S \rightarrow Bool)))$
  tvaru **p,n:0,1**, tj. horní poměr obrácené funkce je 1, pak A je rozložitelný. (viz násl. obrázek)
- (3) použitím věty o rozkladu (viz dále)

Příklad: $A/(W \rightarrow ((T_1, T_2) \rightarrow (S \rightarrow Bool))$



$A \lozenge (B_1, B_2)$



Věta 2 (o rozkladu)

- Zobecnění obou příkladů
- Nechť A je atribut složitosti větší než 2. Nechť existuje jeho vlastní podatribut B_1 daný v přípustné singulární rotaci.
  Potom A ◊ (B_1, B_2), při čemž
  (a) definiční obor funkce B_{2w} je stejný jako definiční obor funkce B_{1w}
  (b) obor hodnot funkce B_{2w} je tvořen zbylými uzlovými typy atributu A, které nejsou v podatributu B_1.
- Netriviálnost přesné formulace věty (DM2) je dána parcialitou uvažovaných datových funkcí: je třeba se vypořádat s tím, že jedna funkce je na daném argumentu nedefinována (píšeme $\perp$), a druhá - skoro identická - vrací na tomtéž argumentu $\{\ \}$.

Jádro datového modelu

- Nechť **A** je libovolná množina HIT-atributů nad **BZT**. Jádrem množiny **A** je taková množina atributů **K**, pro kterou platí:
- (1) $\mathbf{K} \approx \mathbf{A}$
- (2) každý atribut v **K** je nerozložitelný
- (3) neexistuje atribut $A \in \mathbf{K}$ tak, že
  $A \leftarrow \mathbf{K} - \{A\}$

- Tedy jádro je minimální množina elementárních atributů, poskytující danou informační schopnost

Konceptuální model

- Konceptuální datový model je tvořen
  a) bází základních typů **BZT**
  b) jádrem **K** množiny **A** všech relevantních (datovým modelářem navržených) atributů
  c) množinou **C** integritních omezení (formulovaných datovým modelářem nad atributy z **K**)
- Použití: IDM, LDM, Object-Class Model, UML, ...

DM metodou HIT – minikurs
(jak pracovat s příručkou)

- Zopakování a přehled pojmů
- Příklady k jednotlivým jevům
- Metodická doporučení
  - Pojmenování
  - Zkoumání rozložitelnosti
  - Zkoumání odvoditelnosti
- Zadání semestrální práce