

P114

# Sémantika a její role

zaostření pozornosti

6

# Témata

- Sémantika
- informace
- logické vyplývání
- informační schopnost
- zaostření pozornosti
- sortalizace
- báze sort
- definice jednoduchých typů

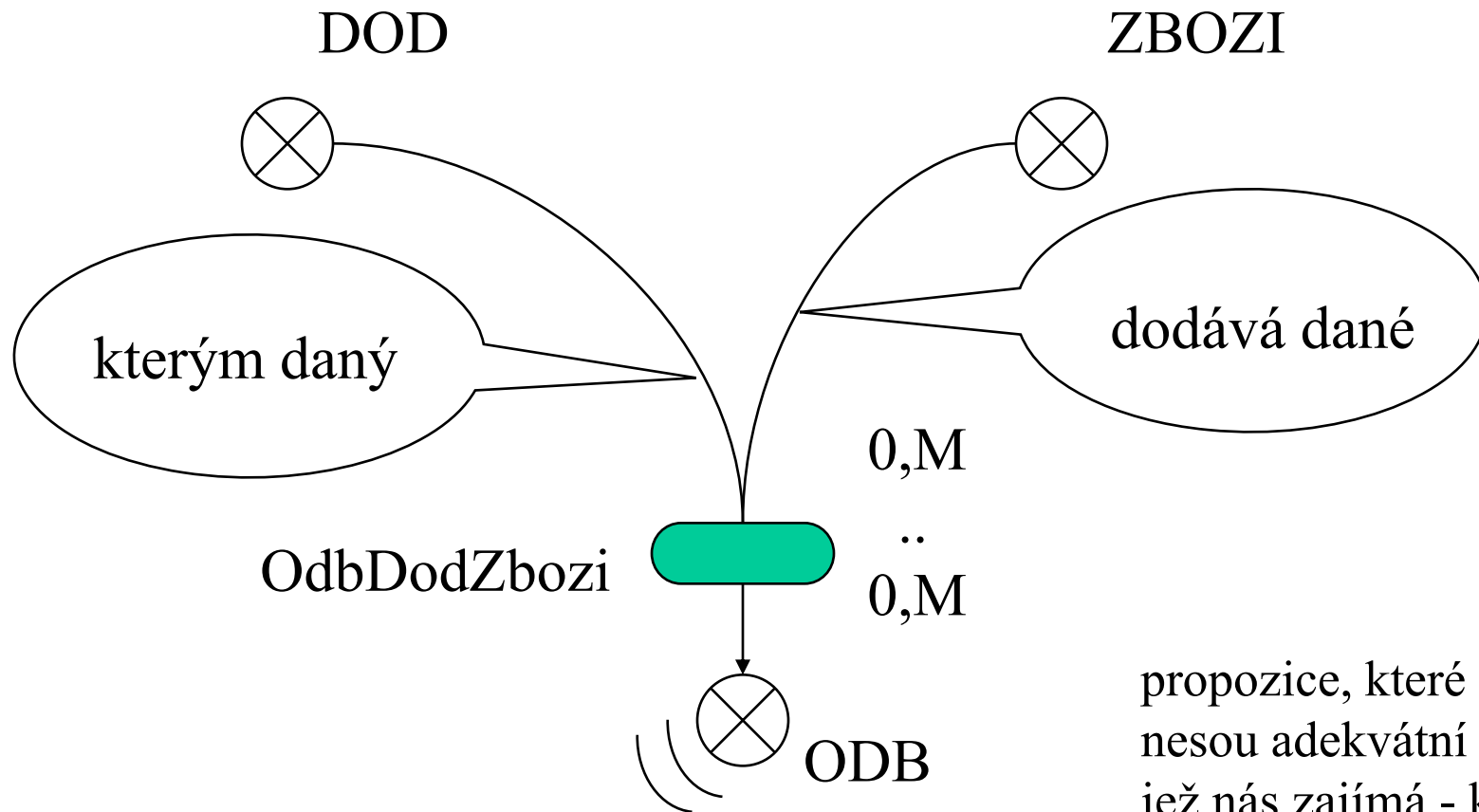
# sémantika

- vše o čem hovoří matematická logika i celá matematika, lze vybudovat pohodlně nad  $\mathbf{B} = \{\text{Bool}, \text{Univ}, \text{Tim}\} \dots$
- ... ale sémantika sdělení, používaných v přirozeném jazyce při popisu reálného světa, chybí
- sémantiku lze zahrnout do teorie podporující naše komunikace právě nad epistémickou bází  
 $\mathbf{EB} = \{\text{Bool}, \text{Univ}, \text{Tim}, \text{Wrd}\}$
- je to tzv. sémantika možných světů
- ta je základem pro náš přístup k DM ...

# sémantika a DM

- **běžný názor:** sémantiku nelze rozumně zachytit, proto se snažíme použít takové syntaktické prostředky, abychom na základě nich mohli alespoň vytušit sémantiku skrytou za našimi modely  
autoři UML, specialisti na DB, DWH, ...
- proti tomu stavíme Datové Modelování metodou HIT a *základní tezi:*  
„DM bez sémantiky je jako láska bez soulože ...“

# HIT-atribut: záznam sémantiky

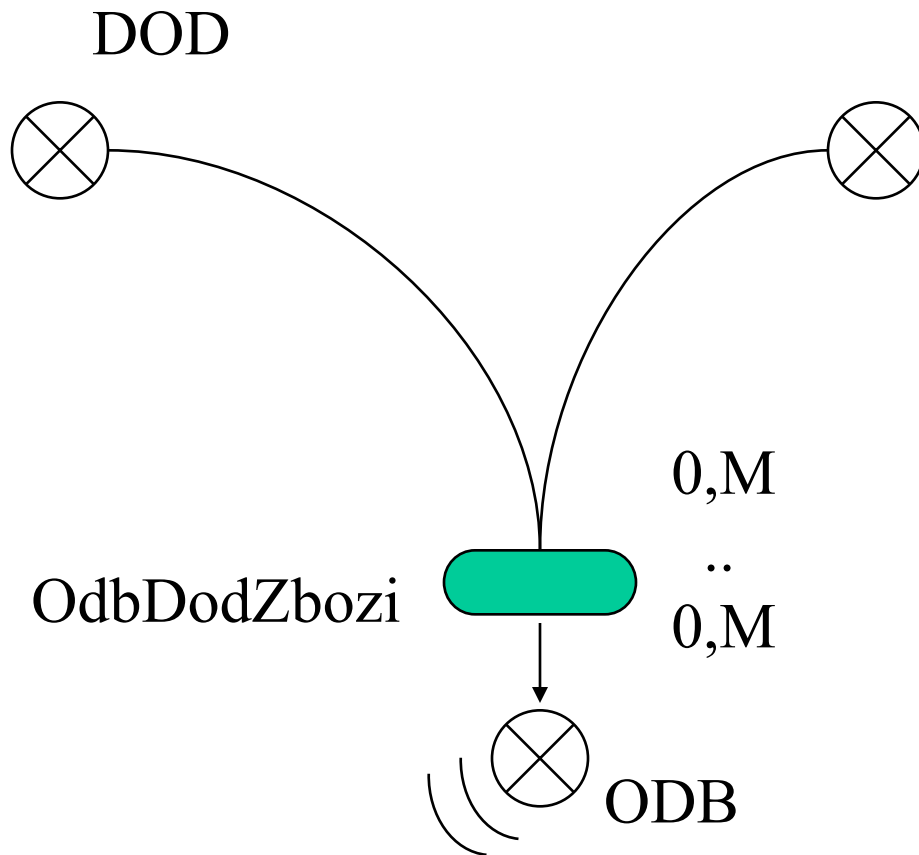


propozice, které generuje,  
nesou adekvátní informaci,  
jež nás zajímá - kvůli které  
vyrábíme IS

# Proč vyrábíme IS:

- chceme odpovědi na otázky typu:
  - kdo dodává jaké zboží do kterého obchodního řetězce?
  - komu co dodává daný dodavatel?
  - komu daný dodavatel dodává dané zboží?
  - ...
- Následující (1) a (2) jsou příklady propozic „generovaných“ atributem OdbDodZbozi:
  - (1) Kostelecké uzeniny dodávají jemné párky do prodejen Tesco a Makro.
  - (2) Dodavatel (MPK) dodává výrobek (Selský salám) do prodejen odběratele (Delvita).
- ... jsou to odpovědi na uvedené otázky

# diagram bez sémantiky:



generuje propozice tvaru:  
odběratelé (libovolným způsobem,  
resp. z libovolného důvodu)  
přiřazení ke dvojici dodavatel  
a zboží

to není dostatečný důvod  
pro tvorbu IS !!!

# sémantika a informace

Novák	3600
Horák	4200
Máloberu	2800

Plat ?

alimenty ?

Vitální kapacita  
plic ?



# data, propozice, sémantika

- propozice explikují význam dat:
- zaměstnanec Novák platí alimony 3 600 Kč
- zaměstnanec Horák má vitální kapacitu plic 4 200
- Tyto propozice dávají informaci, neboť snižují stupeň neurčitosti poznání reálného světa
- R. Carnap: informace obsažená v dané propozici je měřitelná počtem možných světů, které jsou touto propozicí vyloučeny (ve kterých nabývá pravdivostní hodnotu  $N$ )

# Co je to informace

## - základní přístupy

- Carnapovská informace: množství informace v propozici obsažené je dáno počtem možných světů pravdivostí dané propozice vyloučených
- Shanonovská informace: informace je míra snížení entropie (neurčitosti)
- čím více možných světů vyloučíme, tím více snížíme neurčitost

# Carnapovská informace

- Necht'  $P$  je množina propozic a  $R_1 \subseteq \text{Tim}$  je časový interval,  $q$  je nějaká propozice:

$$\begin{aligned} \text{Wrd}(P, R_1) &= \{w \in \text{Wrd} \mid [[qw]t] = \text{Pravda} \wedge q \in P \wedge t \in R_1\} \\ &= \{w \in \text{Wrd} \mid \forall q \in P \forall t \in R_1 ([[qw]t] = \text{Pravda})\} \end{aligned}$$

$\text{Wrd}(P, R_1)$  se nazývá **přípustný logický prostor** množiny propozic  $P$  vzhledem k časovému intervalu  $R_1$ .

- Čím je  $\text{Wrd}(P, R_1)$  menší, tím více informace je v Carnapově smyslu podáno.
- $\text{Wrd}(P, R_1) = \text{Wrd}$  ...  $P$  nedává žádnou informaci

# Carnapovská informace - pokračování

- Je-li  $R_1 = \text{Tim}$ , hovoříme o **přípustném logickém prostoru** **vzhledem k množině propozic  $P$**  a značíme  $\text{Wrd}(P)$ .
- Jestliže  $\text{Wrd}(P) \neq \text{Wrd}$ , pak  $\text{Wrd}(P)$  je vlastní podmnožina  $\text{Wrd}$ , a říkáme, že množina propozic  $P$  **dává (poskytuje) informaci**.
- Jestliže  $\text{Wrd}(P, R_1) \neq \text{Wrd}$ , pak  $\text{Wrd}(P, R_1)$  je vlastní podmnožina  $\text{Wrd}$ , a říkáme, že množina propozic  $P$  **dává (poskytuje) informaci v časovém intervalu  $R_1$** .

# logické vyplývání

- rozdíl Implikace ( $\Rightarrow$ ) vs. Logické vyplývání ( $\supset$ )
- implikace (pravdivostní funkce) (BoolBoolBool)-objekt nezávislý na stavu světa
- Propozice B logicky vyplývá z propozice A, značíme  $A \supset B$ , tehdy, když pro  $\forall w \in \text{Wrd}, \forall t \in \text{Tim}$  platí  $[[Aw]t] = \text{Pravda} \Rightarrow [[Bw]t] = \text{Pravda}$
- jestliže  $[[Aw]t] = \text{Nepravda}$  nebo je nedefinováno, pak o pravdivostní hodnotě propozice B nelze nic říci.

# uspořádání podle množství podávané informace

- (1) Necht'  $P, Q$  jsou množiny propozic takové, že pro každé  $q \in Q$  platí,  
bud'  $q \in P$  nebo  
 $\exists p \in P$  tak, že  $p \supset q$ .  
Potom pro přípustné logické prostory množin propozic  $P$  a  $Q$  platí  $\text{Wrd}(P, R_1) \subseteq \text{Wrd}(Q, R_1)$  pro libovolné  $R_1 \subseteq \text{Tim}$ . (Důkaz plyne z definice logického vyplývání).
- (2) Říkáme, že množina propozic  $Q$  **dává nejvýše tolik informace, jako množina propozic  $P$** , nebo že  $Q$  dává méně nebo stejně informace jako  $P$ .
- (3) Značíme  $Q \angle_i P$ . Relace  $\angle_i$  je částečné quasi-uspořádání na množině množin propozic. (Proveďte důkaz !)

# Funkce Cn - důsledek

- P je množina propozic

$((\text{Wrd}, \text{Tim}) \rightarrow \text{Bool}) \rightarrow \text{Bool}$  - objekt

označme  $((\text{Wrd}, \text{Tim}) \rightarrow \text{Bool}) = \text{Pr}$ , pak  $P / (\text{Pr} \rightarrow \text{Bool})$

- $Cn / (\text{Pr} \rightarrow \text{Bool}) \rightarrow (\text{Pr} \rightarrow \text{Bool})$

definovaná

$$Cn = \lambda p [\cup \lambda q (q \angle_i p)]$$

kde  $p, q :: (\text{Pr} \rightarrow \text{Bool})$ .

- Cn se nazývá funkce „důsledek“ a dává na každé množině propozic P množinu všech možných jejich logických důsledků

# Tvrzení o $C_n$

- **Tvrzení 1:**

Operace (funkce)  $C_n$  je idempotentní:

$$[C_n[C_n P]] = [C_n P]$$

- Proved'te důkaz!

- **Tvrzení 2:**

$Q \angle_i P$  právě když  $[C_n Q] \subseteq [C_n P]$

- Proved'te důkaz!

- **Důsledek:**

Množinu tříd propozic

$\{[C_n P] \mid P \text{ je libovolná množina propozic}\}$

lze částečně uspořádat.

- Proved'te důkaz!



## ... více o $C_n$

- $P_i$  třída propozic,  $i = 1, \dots, n$
- $P_i \subseteq [C_n P_i]$
- „obrácená trojúhelníková nerovnost“:  
$$\bigcup_{i=1}^n [C_n P_i] \subseteq [C_n (\bigcup_{i=1}^n P_i)]$$

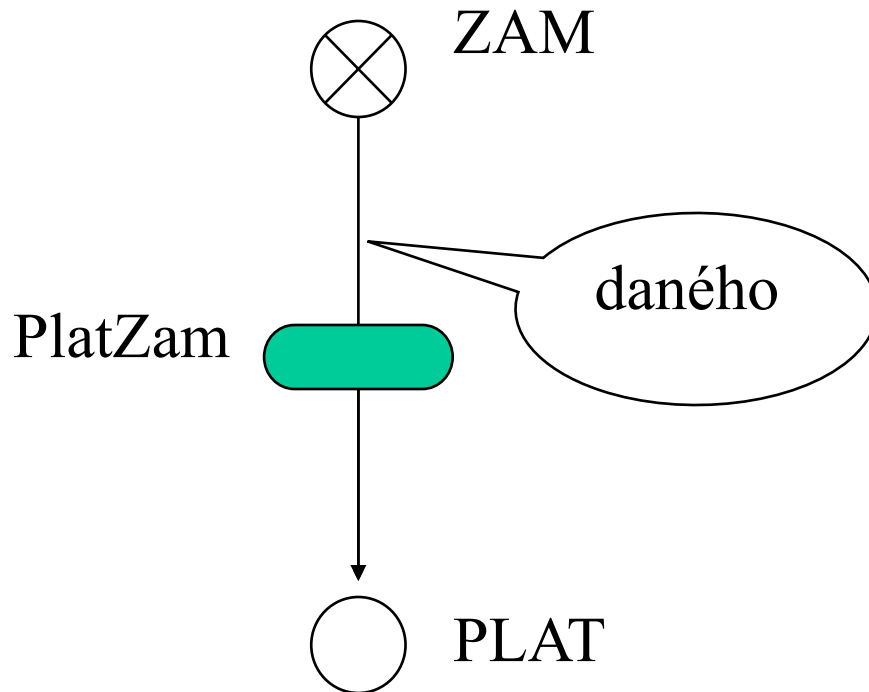
Proveďte důkaz!

- ... abychom věděli co to znamená, když říkáme „... to je přece logické, když ...“

# Informační schopnost IS

- **CO** se má udělat = specifikace provedení
  - jaké informace bude IS poskytovat = vymezení třídy dotazů nad daným IS zodpověditelných
- jak zadat třídu dotazů, které mají být daným IS zodpověditelné
- vytvořit seznam všech takových dotazů? **NE!**
- určit „bázi“ prostoru dotazů, které mají být zodpověditelné
- **Najít konstrukce, které generují propozice dávající právě požadované odpovědi**

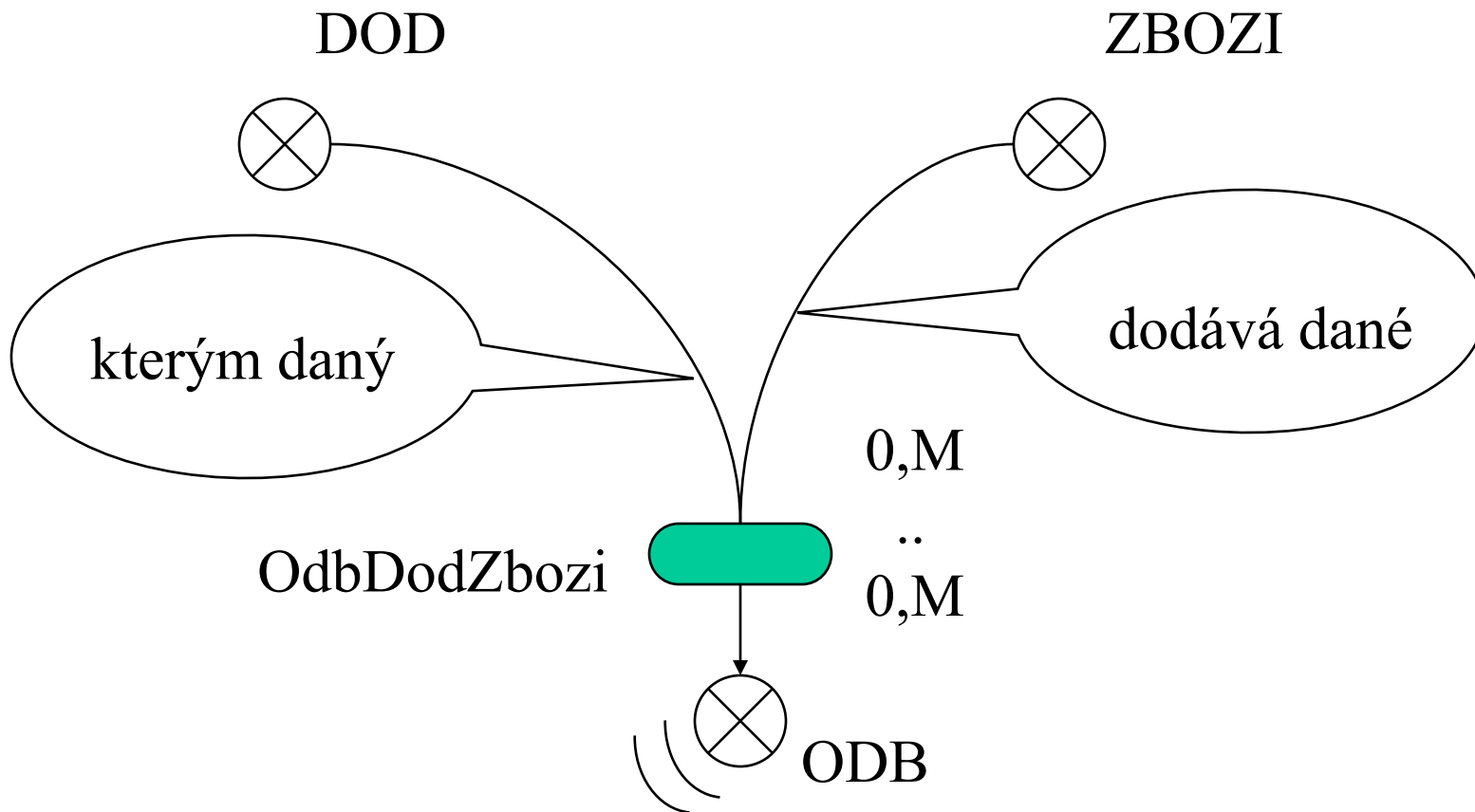
# konstrukce generující množiny propozic



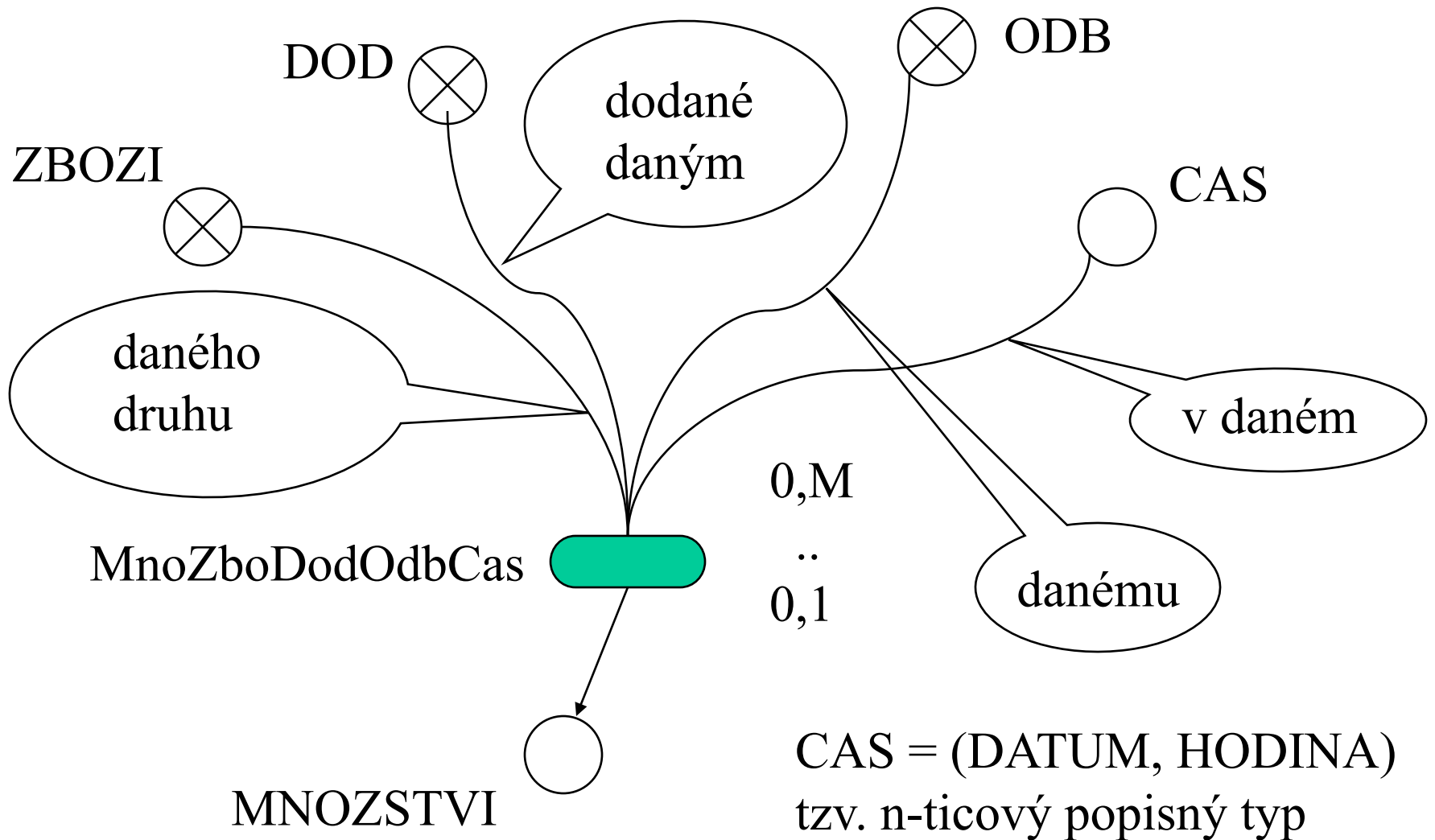
- typ hodnoty funkce
- typ argumentů
- typ samotné funkce
- role argumentů
- sémantika přiřazení

- $W::Wrd, t::Tim, z::ZAM, x::PLAT$
- $\lambda w \lambda t \lambda z \lambda x ([[[PlatZam(w)]t]z]=x)$   
(x je ale jediné - nutno použít singularizátor)
- $\lambda w \lambda t \lambda z \lambda x ([[[PlatZam(w)]t]z]=x) \quad (k)$
- Konstrukce (k) generuje všechny možné propozice tvaru :  
 „Zaměstnanec Novák má plat 15 000,-Kč“  
 „Zaměstnanec Mach má plat 17 500,-Kč“  
 ....  
 „Existuje zaměstnanec s platem nad 16 000,-Kč.“

? jaké propozice generuje atribut:



? jaké propozice generuje atribut:



# Bázové propozice generované atributem

- A je atribut,

$A / ((\text{Wrd}, \text{Tim}) \rightarrow (T \rightarrow S))$  nebo

$A / ((\text{Wrd}, \text{Tim}) \rightarrow (T \rightarrow (S \rightarrow \text{Bool})))$

T, S jsou entity nebo deskripce, resp. jejich n-tice

-- viz přednáška 3

- Množina bázových propozic generovaná atributem A ve stavu světa  $W / (\text{Wrd}, \text{Tim})$  je definována takto:

$$\text{BP}(A)^W = \lambda p (\exists x (\exists y ([A W] x] = y \wedge \wedge p = (\lambda w [[A w] x] = y))))),$$

kde  $p::\text{Pr}$ ,  $w::(\text{Wrd}, \text{Tim})$ ,  $x::T$ ,  $y::S$  nebo  $y::(S \rightarrow \text{Bool})$

# Informační kapacita atributu

- Informační kapacitou atributu  $A$  ve stavu světa  $W$  nazýváme množinu všech propozic generovaných atributem  $A$  ve  $W$ , tj. množinu  $P(A)^W$  všech logických důsledků básových propozic generovaných atributem  $A$  ve  $W$ :

$$P(A)^W = [Cn BP(A)^W]$$

- Informační kapacita množiny atributů  $\{A_1, \dots, A_n\}$  ve  $W$  je dána množinou všech logických důsledků propozic generovaných atributy  $A_i$ :

$$P(A_1, \dots, A_n)^W = [Cn \cup_{i=1}^n P(A_i)^W]$$



# Důsledky

- Informační kapacity (množin) atributů lze částečně uspořádat. -- proveďte důkaz!
- ... jsme schopni porovnávat informační schopnosti databází a IS

## ... zpět k propozicím generovaným PlatZam:

- Zaměstnanec Novák má plat 15 000,-Kč./ ((BoolTim)Wrd)  
(propozice)
- Zaměstnanec / (((BoolUniv)Tim)Wrd)  
(vlastnost individuí)
- Novák / Univ  
(jméno individua jako nálepka)
- má plat / (((TimUniv)Tim)Wrd)  
(fce z individuí do reálných čísel v závislosti na možném světě a čase)
- Propozice říká
  - 1) že Novák je zaměstnanec a
  - 2) že má plat a
  - 3) že ten plat je 15 000,-
- **... stavět konstrukce přímo nad epistémickou bází je pragmaticky neúnosné !**

# „Zaostření“ pozornosti

- proces modelování (zájmové části) světa se podobá zaostřování při fotografování: před našim objektivem je vše, my ale snímkem zobrazíme jasně pouze to, na co jsme zaostřili
- To znamená:
  - nad **EB** vybereme určité typy a na ty zaostříme pozornost
  - modely pak obsahují pouze konstrukce těchto „zaostřených“ typů
- Zaostřené typy jsou: funkce tzv. jednoduchých typů (viz přednáška 3) a funkce definující tzv. sorty (viz dále)

# příklady sort

Sorta ZAMESTNANEC:

je třída (extenze) individuí, kterou v určitém časovém intervalu  $R$  a v aktuálním světě  $w_a$  generuje konstrukce  $\lambda w \lambda t ([[Být\_zaměstnancem(w)]t])$ , kde  $Být\_zaměstnancem :: (((BoolUniv)Tim)Wrd)$

Sorta PLAT:

je třída (extenze) čísel konstruovaná konstrukcí  $\lambda r ([[Plat(r)]=Pravda)$ , kde  $Plat :: (BoolTim)$

# definice deskriptivní sorty

**Deskriptivní sortou** (popisnou sortou) je každá algoritmicky vyčíslitelná množina.

- Je to tedy rekursivní množina.
- Je to množina reprezentovatelná na počítači.

Je vždy dána nějakou definicí, která umožňuje rozhodnout, zda daný řetěz znaků resp. dané číslo patří do této sorty či nikoliv. Toto rozhodování je nezávislé na stavu světa (na Wrd a Tim).

Tím jsme mezi objekty našeho zkoumání zařadili i prvky jazyka, kterým se vyjadřujeme - zejména názvy konkrétních objektů jako „nálepky“.

# definice entitní sorty

Necht'  $R \subseteq \text{Tim}$  je rozumné časové okolí (bylo-je-bude) přítomnosti. Necht'  $r \in R$  je časový okamžik a  $w_a$  je aktuální svět.

Necht'  $T_1, \dots, T_m$  jsou ne nutně různé typy nad **EB**.

Necht'  $P_i / (((\text{Bool}T_i)\text{Tim})\text{Wrd})$  jsou konkrétní vlastnosti přisouditelné  $T_i$ -objektům.

Označme  $C(P_i, r, w_a)$  třídu  $T_i$ -objektů generovanou vlastností  $P_i$  v daném časovém okamžiku  $r$  a aktuálním světě  $w_a$ .

Potom:

$\bigcup_{r \in R} C(P_i, r, w_a)$  je **entitní sorta** definovaná vlastností  $P_i$ .

$\bigcup_{i=1..m} \bigcup_{r \in R} C(P_i, r, w_a)$  je **entitní sorta** definovaná disjunkcí vlastností  $P_i$ ,  $i=1, \dots, m$ .

$\bigcap_{i=1..m} \bigcup_{r \in R} C(P_i, r, w_a)$  je **entitní sorta** definovaná konjunkcí vlastností  $P_i$ ,  $i=1, \dots, m$ .

# sortalizace

- výběr vhodných entitních a deskriptivních sort pro popis zájmového výseku reality
- definice entitních sort
- popis hodnot deskriptivních sort
- sortalizace je to co jiní autoři nazývají klasifikací, tj. výběr základních tříd, nad nimiž budeme operovat : jak při modelování, tak při práci s databázovým systémem
  
- Uvědomme si, že entitní sorty jsou extenze!
- ...stejně jako prvky každé báze

# Jednoduché typy, HIT-atributy

- HIT-atributy (funkční závislosti) jsou objekty tzv. **jednoduchých typů**
  - (a)  $((((S_1, \dots, S_m)(T_1, \dots, T_n))\text{Tim})\text{Wrd})$
  - (b)  $((((\text{Bool}(S_1, \dots, S_m))(T_1, \dots, T_n))\text{Tim})\text{Wrd})$pokud platí, že alespoň jedna ze sort  $S_i$  nebo  $T_i$  je entitní sorta.
- číslo  $m+n$  se nazývá složitost HIT-atributu
- **HIT-atributy** konstruujeme nad tzv. bází sort **BS**. Sorty viz definice entitních a deskriptivních sort. **BS** je určena tak, že pragmaticky odpovídá našemu konkrétnímu zájmu.



# vztah sort k modelování

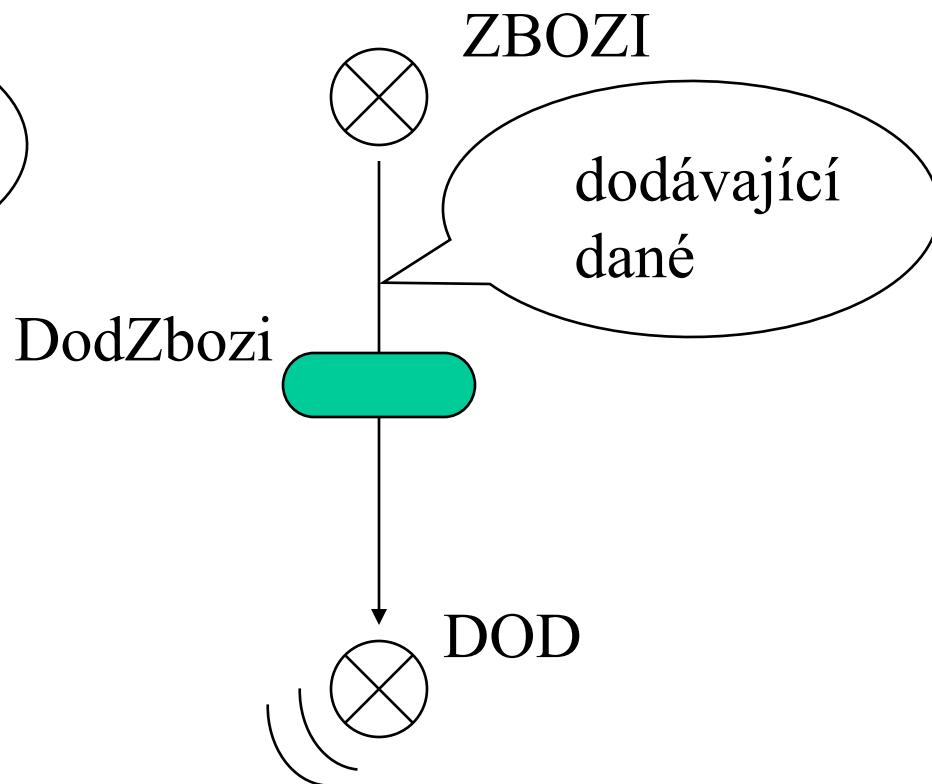
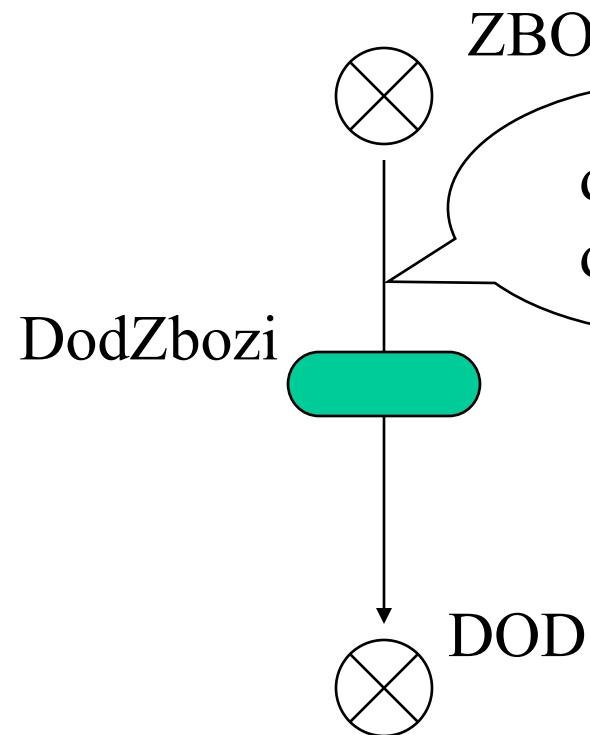
- **entitní a deskriptivní sorty** jsou objekty zájmu, o kterých hovoří uživatelé IS, a ze kterých bereme hodnoty funkcí a jejich argumentů při popisu výseku reálného světa
- jejich podmnožiny jsou definiční obory a obory hodnot funkcí uložených formou tabulek **v databázích**
- **v modelech** nás zajímají **konstrukce** těch **typů** (tj. funkcí nad **EB**), které „v rozumném časovém okolí přítomnosti“ a v aktuálním světě poskytují tyto sorty (pro případ entitních sort), resp. poskytují tyto sorty nezávisle na stavu světa (pro případ deskriptivních sort)

# Příklady k procvičení

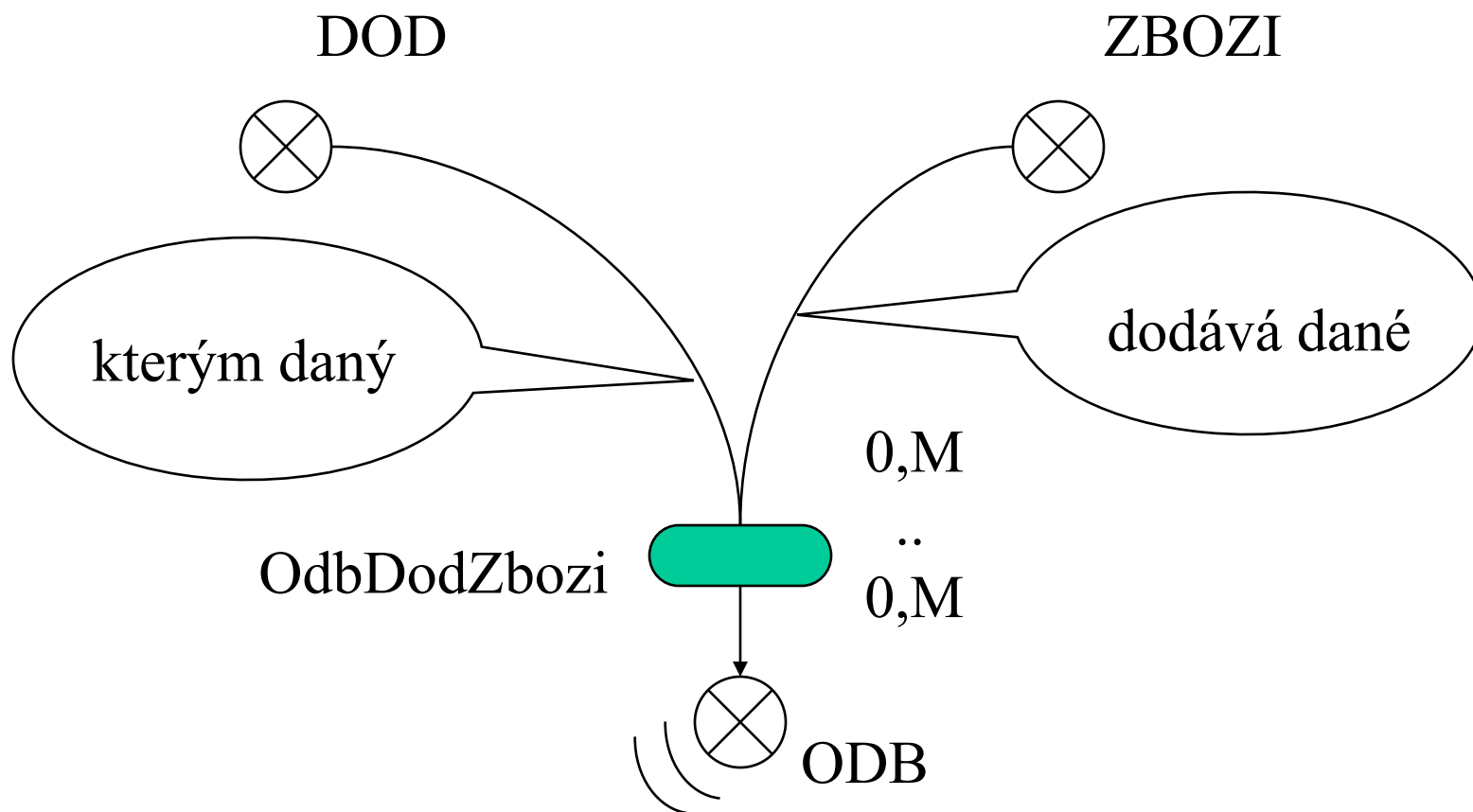
# Porovnejte informační kapacitu:

které bázové propozice generují?

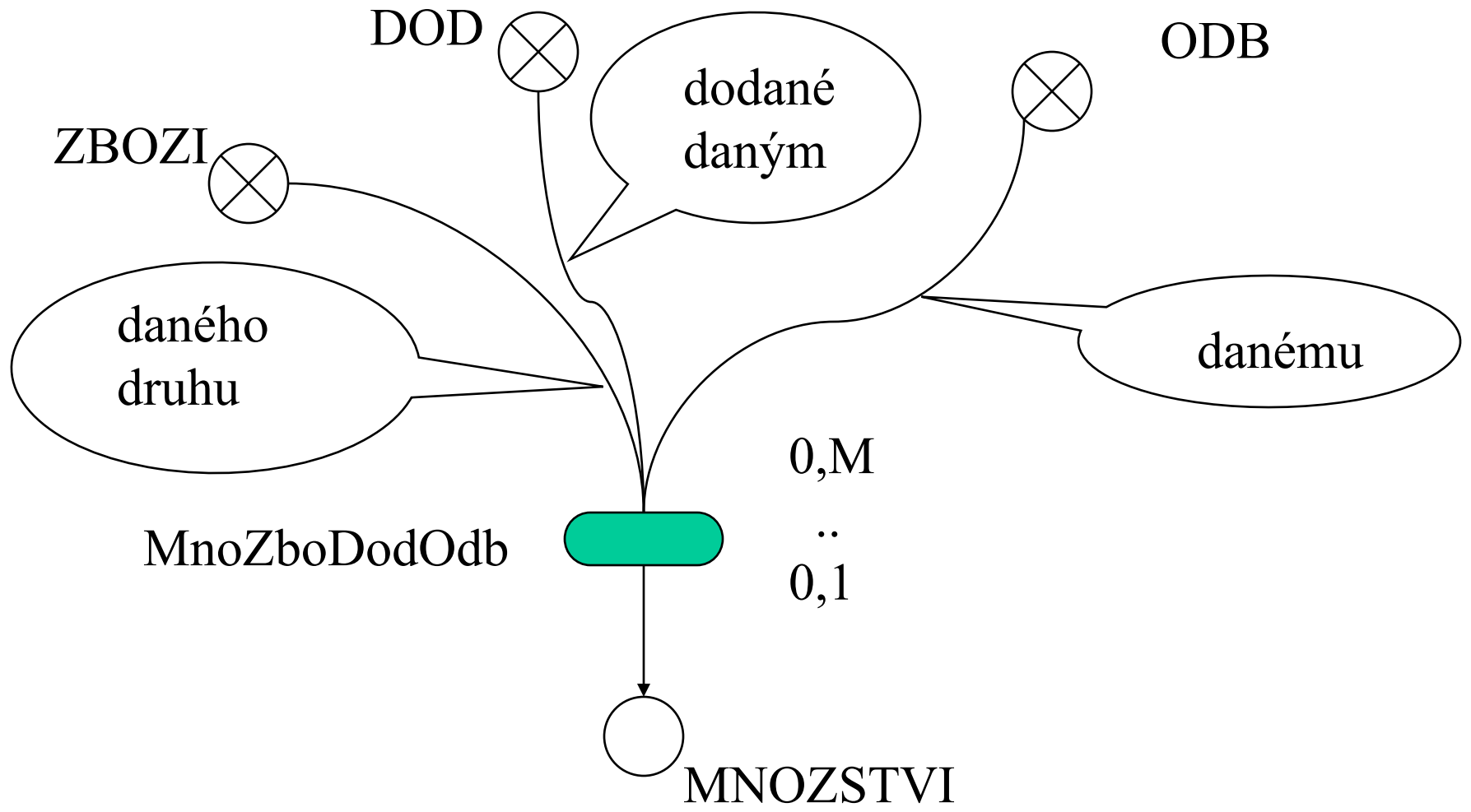
které propozice generují?



které bázové propozice generuje?  
které propozice generuje?



které bázové propozice generuje?  
které propozice generuje?



které bázové propozice generuje?  
které propozice generuje?

