Navrhovanie databáz

- **Formálne metódy**:
  identifikácia, či test objektov a optimalizácia návrhu databázy pre zvolený dátový model (relačný model)
- **Polo-formálne metódy**:
  analýza reálnej skutočnosti a komunikácia s koncovým užívateľom (ER-model, NIAM, ...)

Základné pojmy pre navrhovanie v relačnom modeli

Závislosti - $\forall\exists\wedge \Rightarrow$ (induktívne Hornové formuly)

- Funkčné závislosti $\mathbf{x} \rightarrow \mathbf{y}$
  $(\forall\ \mathbf{x y_1 z_1 y_2 z_2})(\ R(\mathbf{x y_1 z_1}) \wedge R(\mathbf{x y_2 z_2}) \Rightarrow \mathbf{y_1 = y_2}\ )$
- Multi-závislosti $\mathbf{x} \twoheadrightarrow \mathbf{y}$ (multivalued dependencies)
  $(\forall\ \mathbf{x y_1 z_1 y_2 z_2})(\ R(\mathbf{x y_1 z_1}) \wedge R(\mathbf{x y_2 z_2}) \Rightarrow R(\mathbf{x y_1 z_2})\ )$
- Uzáver množiny závislostí $\boldsymbol{F^*}$ je množina všetkých závislostí, ktoré vyplývajú z $\boldsymbol{F}$.
- Úplné pokrytie je $\boldsymbol{F^+}$ je priemet $\boldsymbol{F^*}$ na závislosti daného typu
- Pokrytie množiny závisloti $\boldsymbol{F}$ je ľubovolná množina závislostí $\boldsymbol{G}$ taká, že $\boldsymbol{F^+} = \boldsymbol{G^+}$.

Vlastnosti funkčných závislostí

(Armstrongové axiómy)

- (A1) $\mathbf{x} \subseteq \mathbf{y} \Rightarrow \mathbf{y} \rightarrow \mathbf{x}$ reflexívnosť
- (A2) $\forall \mathbf{z}\ \mathbf{x} \rightarrow \mathbf{y} \Rightarrow \mathbf{xz} \rightarrow \mathbf{yz}$ augmetation
- (A3) $(\mathbf{x} \rightarrow \mathbf{y}) \wedge (\mathbf{y} \rightarrow \mathbf{z}) \Rightarrow \mathbf{x} \rightarrow \mathbf{z}$ tranzitívnosť

Dôkaz dosadením do definície funkčnej závislosti:

(A1) $(\forall\ \mathbf{xyz}_1\mathbf{z}_2)(\mathbf{x} \subseteq \mathbf{y} \wedge R(\mathbf{yz}_1) \wedge R(\mathbf{yz}_2) \Rightarrow \mathbf{x} = \mathbf{x})$

(A2) $(\forall \mathbf{xzy}_1\mathbf{y}_2\mathbf{t}_1\mathbf{t}_2)(R(\mathbf{xzy}_1\mathbf{t}_1) \wedge R(\mathbf{xzy}_2\mathbf{t}_2) \Rightarrow \mathbf{y}_1 = \mathbf{y}_2)$
$(\forall \mathbf{xzy}_1\mathbf{y}_2\mathbf{t}_1\mathbf{t}_2)(R(\mathbf{xy}_1\mathbf{zt}_1) \wedge R(\mathbf{xy}_2\mathbf{zt}_2) \Rightarrow \mathbf{y}_1\mathbf{z} = \mathbf{y}_2\mathbf{z})$

(A3) $(\forall \mathbf{xy}_1\mathbf{z}_1\mathbf{y}_2\mathbf{z}_2\mathbf{t}_1\mathbf{t}_2)(R(\mathbf{xy}_1\mathbf{z}_1\mathbf{t}_1) \wedge R(\mathbf{xy}_2\mathbf{z}_2\mathbf{t}_2) \Rightarrow \mathbf{y}_1 = \mathbf{y}_2 = \mathbf{y})$
$(\forall \mathbf{xyz}_1\mathbf{z}_2\mathbf{t}_1\mathbf{t}_2)(R(\mathbf{xyz}_1\mathbf{t}_1) \wedge R(\mathbf{xyz}_2\mathbf{t}_2) \Rightarrow \mathbf{z}_1 = \mathbf{z}_2$

Ďalšie vlastnosti funkčných závislostí

- $(\mathbf{x} \rightarrow \mathbf{y}) \wedge (\mathbf{x} \rightarrow \mathbf{z}) \Rightarrow \mathbf{x} \rightarrow \mathbf{yz}$ (union rule)
- $(\mathbf{x} \rightarrow \mathbf{y}) \wedge (\mathbf{wy} \rightarrow \mathbf{z}) \Rightarrow \mathbf{wx} \rightarrow \mathbf{wz}$ (pseudotransivity)
- $(\mathbf{x} \rightarrow \mathbf{y}) \wedge (\mathbf{z} \subseteq \mathbf{y}) \Rightarrow \mathbf{x} \rightarrow \mathbf{z}$ (decomposition)

Dôkazová technika:

$$\left.\begin{array}{l} \mathbf{x} \rightarrow \mathbf{y} \Rightarrow \mathbf{x} \rightarrow \mathbf{xy} \quad \text{podľa (A2)} \\ \mathbf{x} \rightarrow \mathbf{z} \Rightarrow \mathbf{xy} \rightarrow \mathbf{yz} \quad \text{podľa (A2)} \end{array}\right\} \Rightarrow \quad \mathbf{x} \rightarrow \mathbf{yz} \quad \text{podľa (A3)}$$

Zvyšné dôkazy sa robia podobne podľa Armstrongových axióm. (Urobte ich ako cvičenie.)

Uzáver množiny atribútov

Nech $\mathbf{x}$ je množina atribútov a $\boldsymbol{F}$ je množina funkčných závislostí. Potom uzáverom $\mathbf{x}^+$ množiny $\mathbf{x}$ w.r.t. $\boldsymbol{F}$ rozumieme množinu $\mathbf{x}^+$ všetkých atribútov x takých, že $\mathbf{x} \rightarrow x$ pomocou závislostí v $\boldsymbol{F}$.

Lema: $\mathbf{x} \rightarrow \mathbf{y}$ sa dá odvodiť z $\boldsymbol{F}$ pomocou Armstrongových axióm práve vtedy keď $\mathbf{y} \subseteq \mathbf{x}^+$ w.r.t $\boldsymbol{F}$.

Pre každý atribút $a \in \mathbf{y} \subseteq \mathbf{x}^+$. Platí $\mathbf{x} \rightarrow a$ podľa definície uzáveru $\mathbf{x}^+$. Podľa union rule platí aj $\mathbf{x} \rightarrow \mathbf{y}$.

Naopak nech $\mathbf{x} \rightarrow \mathbf{y}$ sa dá odvodiť. Potom pre každe $a \in \mathbf{y}$ platí $\mathbf{x} \rightarrow a$ podľa decompsition rule a $a \in \mathbf{x}^+$.

Úplnosť „Armstrongových axióm"

Veta: Funkčná závislosť $\mathbf{x} \to \mathbf{y}$ sa dá odvodiť z ***F*** pomocou Armstrongových axióm práve vtedy, keď $\mathbf{x} \to \mathbf{y}$ je dôsledkom ***F***.

Dôkaz: Pretože Armstrongové axiómy sú dôsledkom definície funkčnej závislosti, dajú sa odvodiť len platné závislosti.

Opačne predpokladajme, že závislosť $\mathbf{x} \to \mathbf{y}$ platí ale nedá sa odvodiť pomocou Armstrongových axióm. Uvažujme reláciu **R**

	Atribúty x^+	ostatné atribúty
u:	1 1 … 1	1 1 … 1
v:	1 1 … 1	0 0 … 0

Všetky závislosti z F sú splnené v **R**

Nech $\mathbf{v} \to \mathbf{w}$ je dôsledkom ***F***, ale nie je splnené v **R**. Potom $\mathbf{v} \subseteq \mathbf{x}^+$. Predpoklad o **w** vedie vždy k sporu s predošlou lemou.

Charaktrizácia úplného pokrytia množiny funkčných závislostí

Množina $\boldsymbol{F}^+$ je príliš obsiahla, stačí však uvádzať maximálne závislosti. Závislosť je maximálna ak nemôžeme vynechať žiaden atribút na ľavej strane alebo pridať nejaký atribút na pravú stranu bez porušenia jej platnosti.

K charakterizácii stačia <u>nasýtené množiny= pravé strany maximálnych závislostí</u>. (Spätná rekonštrukcia maximálnych závislostí.)

Veta: Každá úplná množina funkčných závislostí má model v nejakej relácii nad doménou D= { 0, 1 }.

Vyplývanie medzi funkčnými závislosťami

Výpočet $\boldsymbol{F}^{+}$ a testovanie ekvivalencie $\boldsymbol{F}^{+} = \boldsymbol{G}^{+}$ je vo všeobecnosti náročná (exponenciálna) záležitosť. Našťastie stačí počítať uzávery množiny atribútov vzhľadom k $\boldsymbol{F}^{+}$.

Minimálne pokrytie množiny funkčných závislostí

Kánonické závislosti na pravej strane len jeden atribút.

Minimálne pokrytie je pokrytie kánonickými závislosťami z ktorých sa žiadna nedá vynechať bez toho, aby sa porušila vlastnosť byť pokrytím.

$AB \to C$ $D \to E$ $CG \to B$
$C \to A$ $D \to G$ $CG \to D$
$BC \to D$ $BE \to C$ $CE \to A$
$ACD \to B$ $CE \to G$

Minimálne pokrytia

$AB \to C$	$AB \to C$
$C \to A$	$C \to A$
$BC \to D$	$BC \to D$
$D \to E$	$D \to E$
$D \to G$	$D \to G$
$BE \to C$	$BE \to C$
$CE \to G$	$CE \to G$
$CD \to B$	$CG \to B$
$CG \to D$	

Nadkľúče a kľúče

Nech je daná relácia **R**(**U**). Potom množinu atribútov **K** takú, že **K** $\rightarrow$ **U** nazývame nadklúč. Minimálny nadklúč v zmysle množinovej inklúzie nazývame klúč.

Koľko kľúčov môže mať relácia o n atribútoch ?

Príklad:

R(A_1, ..., A_k, B_1, ..., B_k, C)

$\boldsymbol{F} = \{ A_i \leftrightarrow B_i \text{ pre } 1 \leqslant i \leqslant k \} \cup \{ A_1 \ldots A_k \rightarrow C \}$

Bezstrátové spojenia

$\rho=\{R_1, \ldots, R_k\}$, $R = R_1 \cup \ldots \cup R_k$

$$m_\rho(r) = \prod_{R_1}(r) \bowtie \ldots \bowtie \prod_{R_k}(r)$$ Join project mapping

Vlastnosti: $r \subseteq m_\rho(r)$

$m_\rho(r) = m_\rho(m_\rho(r))$

Hovoríme, že dekompozícia má bezstrátové spojenie ak $r = m_\rho(r)$.

Tabuľková metóda testovania.

R= SAIP

$S \to A$

$SI \to P$

S	A	I	P
a_1	a_2	b_{13}	b_{14}
a_1	b_{23}	a_3	a_4

Normálne formy (BCNF, 3NF)

BCNF: Relačná schéma R je v BCNF, keď pre každú v nej platnú funkčnú závislosť **x** $\rightarrow$ **y** platí **x** je nadkľúč.

3NF: Relačná schéma R je v 3NF, keď pre každú v nej platnú funkčnú závislosť **x** $\rightarrow$ **y** platí **x** je nadkľúč alebo **y** je prvok nejakého kľúča (primárny atribút) **R**.

Lema: a.) Každá binárna relácia je v BCNF.

b.) Ak R nie je v BCNF. Potom v nej existujú atribúty A a B také, že (**R** - AB) $\rightarrow$ A.

(Môže a nemusí platiť (**R** - AB) $\rightarrow$ B.)

3NF zachovávajúca závislosti

Príklad: R= MAP (Mesto, Adresa, PSČ)

Závislosti: MA $\rightarrow$ P, P $\rightarrow$ M

Hovoríme že dekompozícia $\mathbf{R} = \mathbf{R}_1 \cup \ldots \cup \mathbf{R}_k$ zachováva závislosti ***F*,** ak každá závislosť z ***F*** je v uzávere tých závislostí $\mathbf{x} \rightarrow \mathbf{y}$ z ***F***, že $\mathbf{xy} \subseteq \mathbf{R}_i$.

Algoritmus testovania zachovania závislosti $\mathbf{x} \rightarrow \mathbf{y}$

z := **x**; **while** sa **z** zmenilo **do**

for i := 1 **to** k **do**

$\mathbf{z} := \mathbf{z} \cup ((\mathbf{z} \cap \mathbf{R}_i)^+ \cap \mathbf{R}_i)$ {uzáver w.r.t. ***F***};

Algoritmus normalizácie do BCNF

Vstup: Relačná schéma **R** a množina funkčných závislostí ***F***.

Výstup: Množina relačných schém R_1 ... R_k v BCNF.

Metóda: Dekompozícia na dve schémy podľa predošlej lemy jednu **X**A zodpovedajúcu závislosti **X** $\rightarrow$ A, ktorá je v BCNF a druhú **R** - A, na ktorú použijeme algoritmus rekurzívne.

Algoritmus normalizácie podrobnejšie

```
Z :=  R;
repeat  bcnf:= decompose( Z, Y, A);
          Z :=  Z - A;
until bcnf;

function decompose( Z, Y, A): boolean ;
{ if Z neobsahuje atribúty A, B také, že A je v (Z - AB)+ then
        begin  Y:=  Z; bcnf= true end
   else  begin najdi A a B;
            Y:= Z - B;
            whille Y obsahuje A a B také, že (Y - AB)+ → A do
                    Y:= Y - B;
            bcnf := false;
          end;
   return(bcnf) }
```

Normalizácia do 3NF zachovávajúcej závislosti

Vstup: Relačná schéma **R** a minimálne pokrytie ***F*.**

Výstup: Relačné schémy bestrátovej dekompozície do 3NF.

Metóda: Ak ***F*** obsahuje závislosť, ktorá obsahuje všetky atribúty **R**, potom **R** je už v 3NF.
Inak každej funkčnej závislosti v ***F*** zodpovedá jedna relačná schéma. Treba pridať ešte relačnú schému pre atribúty **R**, ktoré sa nevyskytujú v žiadnej funkčnej závislosti ***F***. Tieto atribúty musia byť súčasťou každého kľúča, aby došlo k spojeniu treba ich doplniť na kľúč **R**.

Pravidlá pre multizávislosti

- (A1) $\mathbf{x} \subseteq \mathbf{y} \subseteq \mathbf{U} \Rightarrow \mathbf{y} \rightarrow \mathbf{x}$ reflexívnosť
- (A2) $\forall \mathbf{z}\ \mathbf{x} \rightarrow \mathbf{y} \Rightarrow \mathbf{xz} \rightarrow \mathbf{yz}$ augmentation
- (A3) $(\mathbf{x} \rightarrow \mathbf{y}) \wedge (\mathbf{y} \rightarrow \mathbf{z}) \Rightarrow \mathbf{x} \rightarrow \mathbf{z}$ tranzitívnosť
- (A4) $\mathbf{x} \twoheadrightarrow \mathbf{y} \Rightarrow \mathbf{x} \twoheadrightarrow \mathbf{U} - \mathbf{x} - \mathbf{y}$ complementation
- (A5) $(\mathbf{x} \twoheadrightarrow \mathbf{y}) \wedge (\mathbf{v} \subseteq \mathbf{w}) \Rightarrow \mathbf{wx} \twoheadrightarrow \mathbf{vy}$ augmentation
- (A6) $(\mathbf{x} \twoheadrightarrow \mathbf{y}) \wedge (\mathbf{y} \twoheadrightarrow \mathbf{z}) \Rightarrow \mathbf{x} \twoheadrightarrow (\mathbf{z} - \mathbf{y})$
- (A7) $\mathbf{x} \rightarrow \mathbf{y} \Rightarrow \mathbf{x} \twoheadrightarrow \mathbf{y}$
- (A8) $(\mathbf{x} \twoheadrightarrow \mathbf{y}) \wedge (\mathbf{z} \subseteq \mathbf{y}) \wedge (\mathbf{w} \cap \mathbf{y} = \emptyset) \wedge (\mathbf{w} \rightarrow \mathbf{z}) \Rightarrow \mathbf{x} \rightarrow \mathbf{z}$

4NF

Nech $\mathbf{D}^+$ je množina všetkýxh platných závislostí a multizávislostí v relačnej schéme **R**. Hovoríme, že relačná schéma **R** je v 4NF, ak pre každú multizávislosť $\mathbf{x} \twoheadrightarrow \mathbf{y}$ takú, že $\mathbf{R} = \mathbf{x} \cup \mathbf{y}$ a $\mathbf{x} \vee \mathbf{y}$, platí **x** je nadkľúč.

Veta: 4NF $\Rightarrow$ BCNF $\Rightarrow$ 3NF

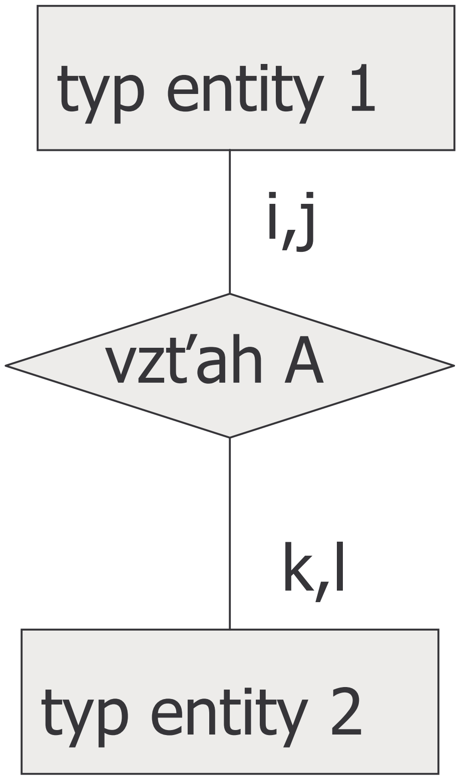
Poloformálne metódy - mapovanie reality

- Entitno-relačný model
- Binárny model
- NIAM
- Sémantický model
- O – O model
- Sieťový model - Bachmanové diagramy
- Automatické navrhovadlá (designer2000, access, …)
- HIT

Grafická reprezentácia

vizualizácia

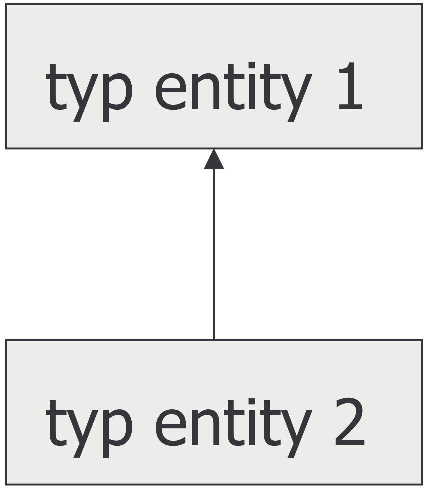
Entitno relačný (ERA) model



i, k = {
- 0 Entita daného typu sa nemusí vyskytovať vo vzťahu A
- 1 Každá entita daného typu sa musí vyskytnúť vo vzťahu A

j, l = {
- 1 Entita daného typu sa môže vyskytovať vo vzťahu A najviac raz
- n Bez ohraničení na počet výskytov entity daného typu vo vzťahu A

Vzťah generalizácie - is a



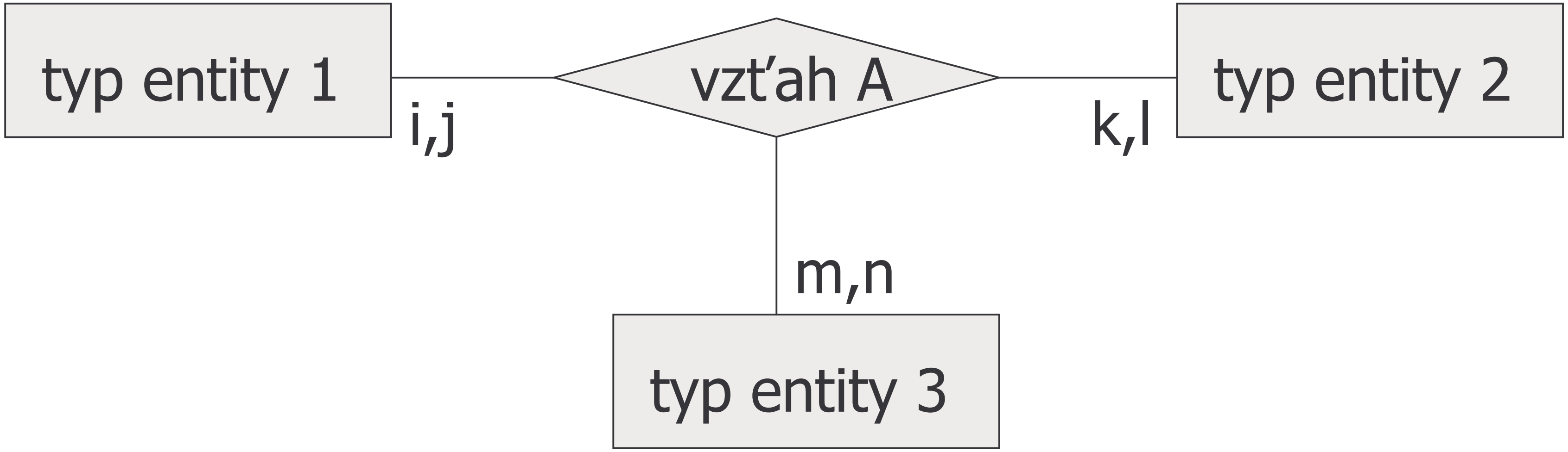
Typ entity 2 je špeciálnym prípadom typu entity 1.

- dedenie atribútov
- discriminated union
- nulové hodnoty

Atribúty - vpisujú sa do typov entít

označenie kľúčov (a cudzích kľúčov)

Ternárne a n-árne vzťahy



Problém ohraničení počtu výskytov

- objektifikácia binárneho vzťahu
- určenie funkčnej závislosti

Binárny model - NIAM

- slovný popis
- grafická reprezentácia

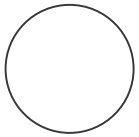
Pojmy:
- Typ
- Populácia
- Výskyt (occurrence)

Lexikálne (LOT) a nelexikálne typy objektov (NOLOT)

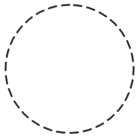
Typy vzťahov:

- *Idea* - medzi nelexikálnymi typmi objektov
- *Bridge* - medzi nelexikálným a lexikálnym objektom
- Phrase - medzi lexikálnymi objektami

Grafická notácia

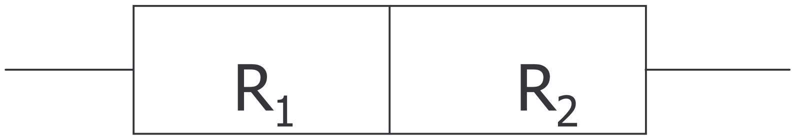


Nelexikálny typ objektu



Lexikálny typ objektu

Podtyp (is a)



Idea alebo bridge

Podmienky - constraints

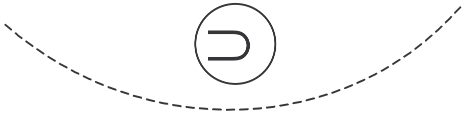
Nad menom role, znamená že táto rola jednoznačne určuje druhú rolu vo vťahu

surjekcia (totalita)

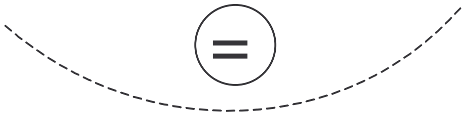
Podmienky - constraints

Disjunktnosť (vylúčenie) medzi podtypmi

Jednoznačné určenie výskytu (kombinácie)

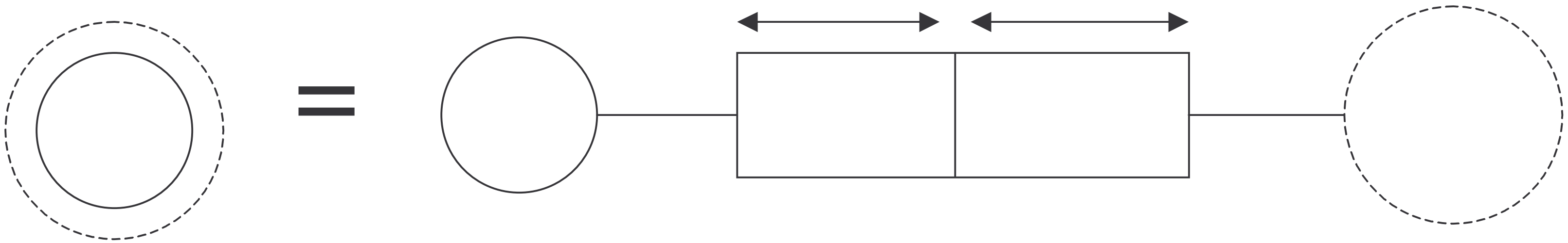


Inklúzia medzi populáciami rolí



Rovnosť populácii rolí

Makro



O – O model

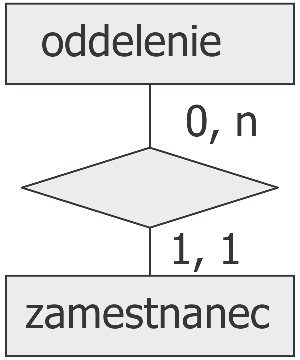
```
interface:  Študenti (
   attribute string meno;
   attribute integer rodné_číslo;
   attribute Struct(deň, mesiac, rok) Dátum_narodenia;
   relationship Set(Prednášky) zapísal_si
                     inverse Prednášky :: majú_zapísané )
```

V oblasti návrhu objektový model zodpovedá binárnemu modelu.

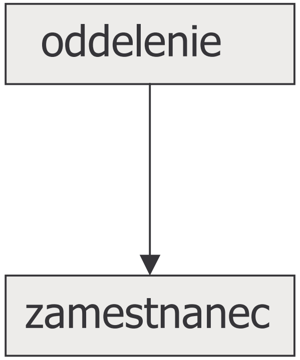
Navyše je detailnejší umožňuje podrobne popísať typy atribútov. Používa konštruktory typov (Set - množina, bag – multimnožina, struct – record, list –zoznam, array – pole, ...).

Základné konštrukcie I

ER-model



Sieťový

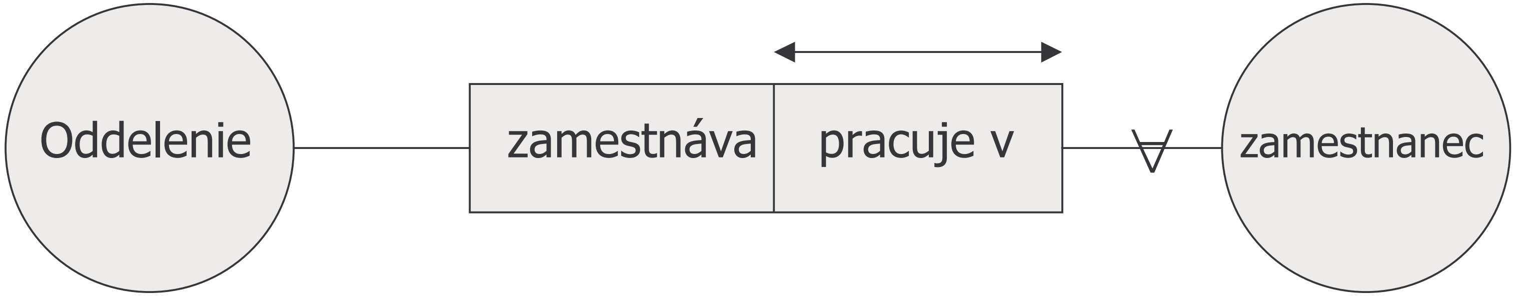


Relačný

Oddelenie(ČísOdd, …)

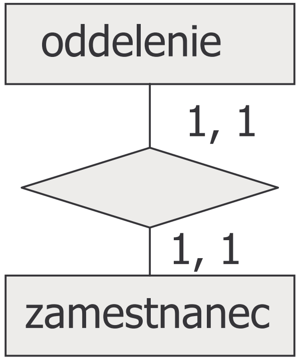
Zamestnanci(IdZam, ČísOdd, …)

Binárny model



Základné konštrukcie II

ER-model



Sieťový

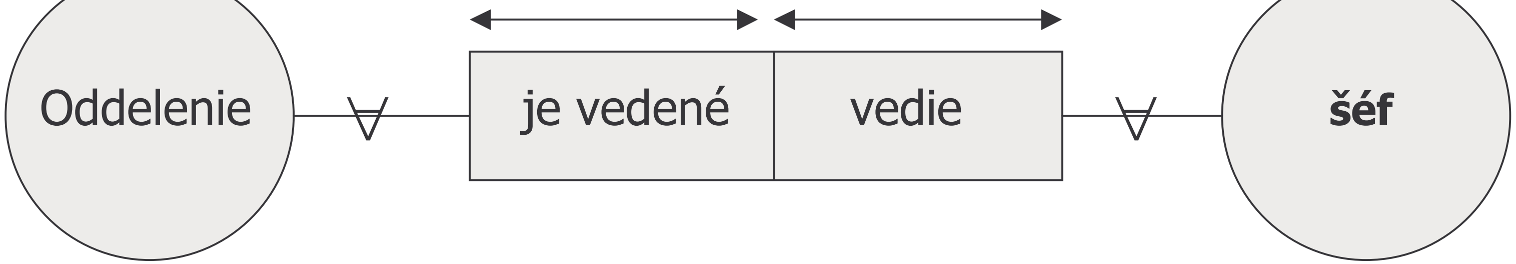
Record type
oddelenie(ČísOdd, IdŠéfa, ...)

Relačný

oddelenie(<u>ČísOdd</u>, <u>IdŠéfa</u>, ...)

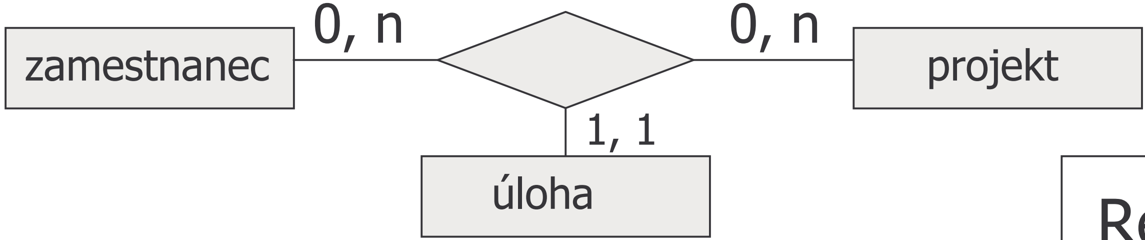
Existencia samostatných typov viet pre oddelenie a šéfa je možná, ale nie nutná.

Binárny model

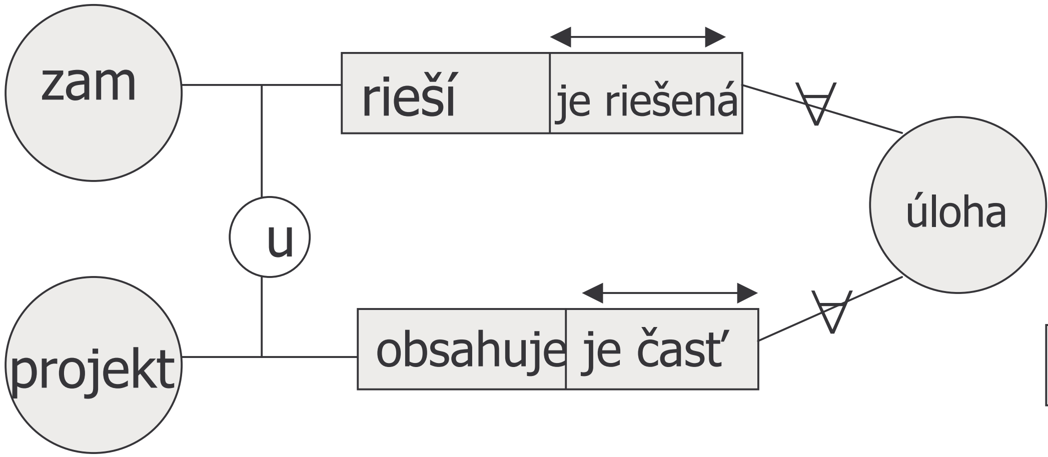


Ternárne vzťahy I

ER-model



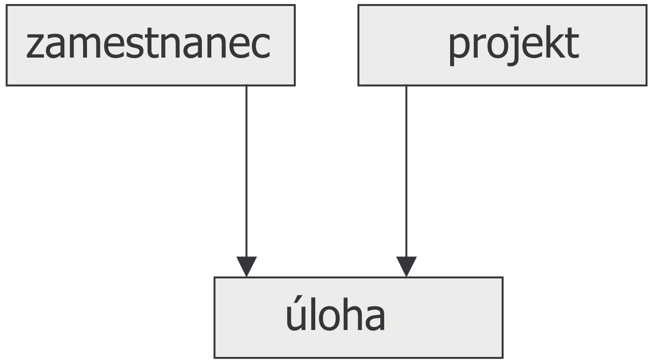
Binárny model



Relačný model

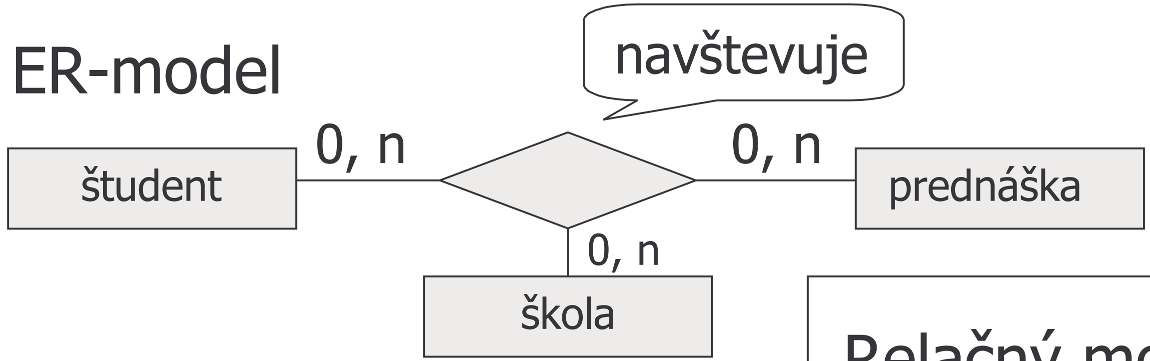
zamestnanci(IdZam, …)
projekty(ČísProj, …)
úlohy(IdZam, ČísProj, …)

Sieťový model

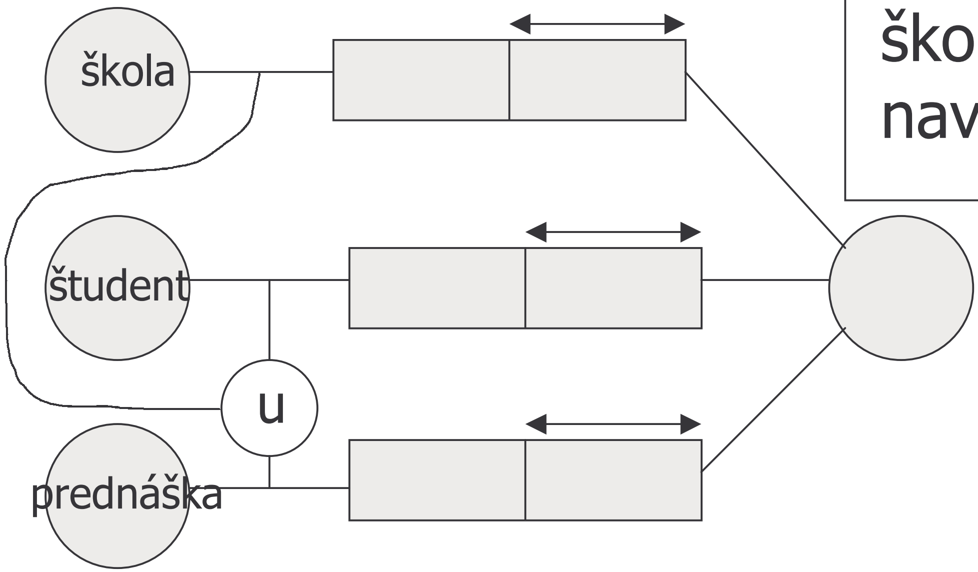


Ternárne vzťahy II

ER-model



Binárny model



Relačný model

študent(RodČís, …)
prednáška(názov, …)
škola(IČO, …)
navštevuje(názov, IČO, RodČís)