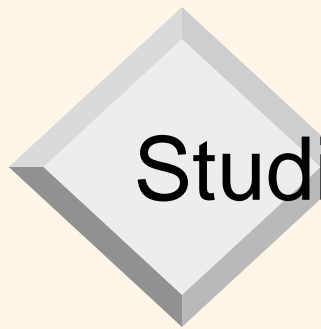


# Dotazovací jazyky I (NDBI001)

J. Pokorný



# Studium databází na informatice na MFF

Data science

Moderní DB systémy

Bioinformatické alg., DB, ...

Dotazovací jazyky II

Vyhledávání MM obsahu na webu

*Mgr.*

Dotazovací jazyky I

Transakce

Datové formáty

Administrace Oracle

Vyhledávání na Webu

DB systémy pro praxi

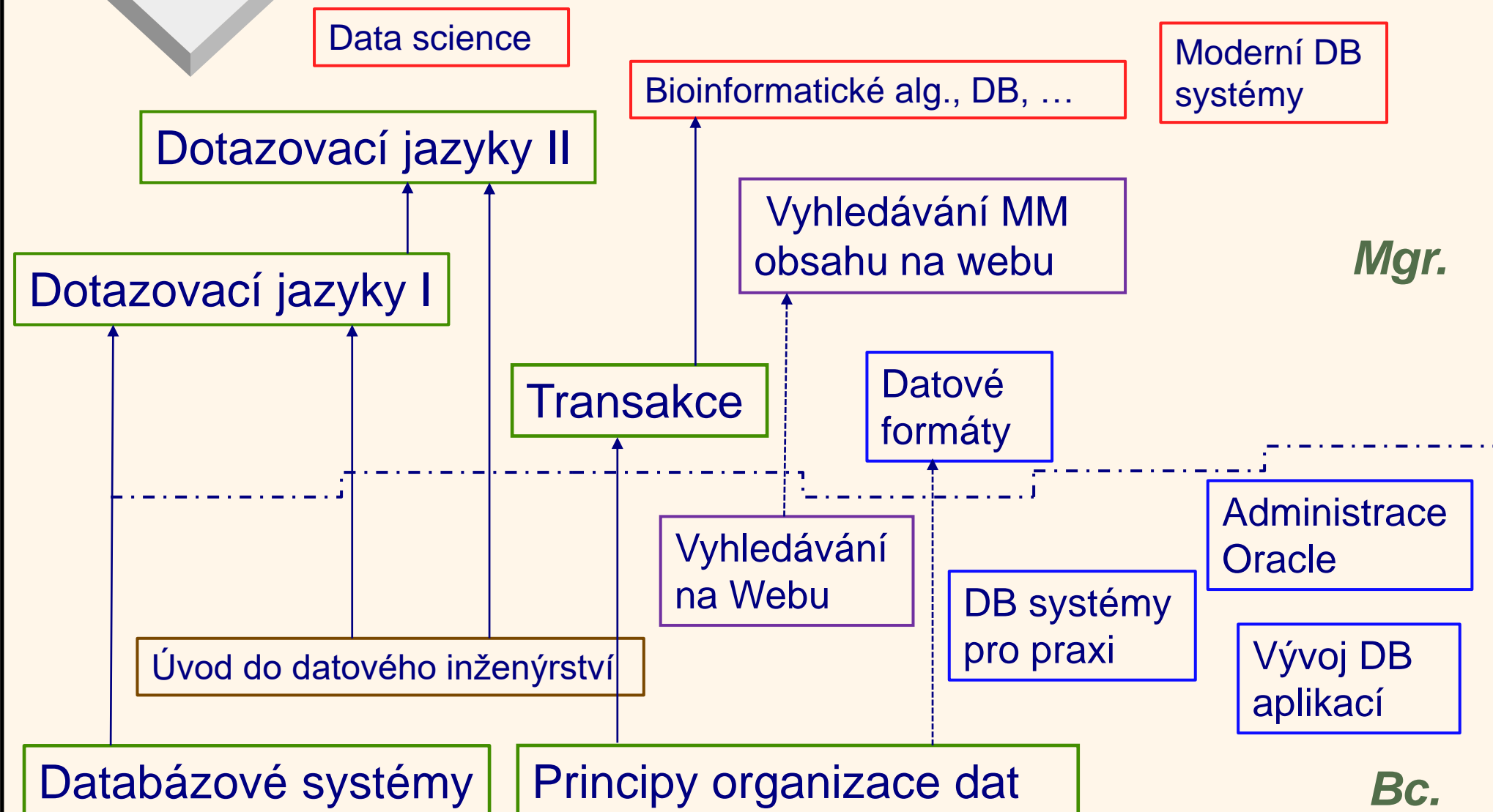
Vývoj DB aplikací

Úvod do datového inženýrství

*Bc.*

Databázové systémy

Principy organizace dat





# Obsah

1. Shrnutí pojmů RMD, relační kalkuly, algebry. Pojem databázového dotazu, dotazovacího jazyka a jeho vyjadřovací síly.
2. Detailní výklad jazyka SQL, standardizace jazyka, SQL92, SQL:1999. Tříhodnotová logika v SQL. Jednoduché a složité dotazy. Typy spojení v SQL.
3. Integritní omezení v SQL. Referenční integrita. Definice domén.
4. Pojem pohledu, systémového katalogu. Kritika SQL. Rozšíření SQL
5. Objektově-orientované a objektově-relační (OR) databáze. OR model v SQL:1999.
6. Operátor CUBE, SQL:2003
7. SQL-MM-Spatial
8. SQL:2003
9. Vyhodnocování dotazů.
10. Optimalizace a její role v SŘBD. Algebraická optimalizace.
11. Cenově orientovaná optimalizace, heuristické přístupy k optimalizaci.
12. Dotazování v XML.
13. SQL/XML.



# Další informace

## *Literatura:*


1. Pokorný, J.: Dotazovací jazyky, Science, 1994
2. Pokorný, J.: Dotazovací jazyky. Skripta UK, Vydavatelství Karolinum, 2002, 255 s.
3. Pokorný, J., Vařenta, M.: Databázové systémy. Česká technika – nakladatelství ČVUT, 2. přepr. vydání, 2020.
4. Pokorný, J.: Datová analýza a databáze pro prezenční a kombinovanou formu studia. Skripta, MB – ŠAVŠ, 2020  
<https://savs.tritius.cz/detail/241953>.

*Slajdy:* <http://www.ksi.mff.cuni.cz/~pokorny/vyuka.html#NDBI006>

*Zkouška:* písemná – příklady

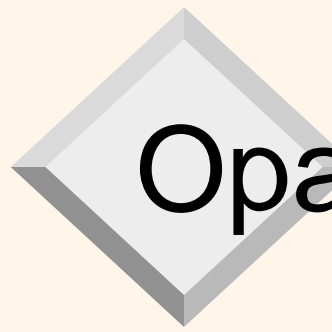
*Vztah ke státnicím:* Studijní plán Softwarové a datové inženýrství

*Doporučené pro zaměření:* Databázové systémy



# Dále o DJ1 - přednášky ZS 2023-2024

- 3.10. Opakování RA, RC, SQL1
- 10.10. SQL2
- 17.10. OR1
- 24.10. OR2
- 7.11. SQL:2003, Spatial
- 14.11. Vyhodnocení dotazů
- 21.11. Vyhodnocení dotazů
- 28.11. Optimalizace dotazů
- 5.12. Optimalizace dotazů
- 12.12. XML - XPath
- 19.12. XML - XQuery
- 26.12. volno
- 2.1. volno
- 9.1. XML - SQL/XML



# Opakování – základní pojmy

- Relační model dat (RMD)
- Relační algebra (RA)
- Doménový relační kalkulus (DRK)

# Relační model dat

Hrají	Název_k	Jméno_f	Čas
	Blaník	Top gun	29.03.94
	Blaník	Kmotr	08.03.94
	Mír	Nováček	10.03.94
	Mír	Top gun	09.03.94
	Mír	Kmotr	08.03.94

Schémata: Hrají(Název\_k, Jméno\_f, Čas)

Kina (Název\_k, Adresa, Jméno\_v)

Film (Jméno\_f, Herec, Režisér)

# Relační algebra - dotazovací jazyk

Předpoklady: Schéma DB  $R$ ;  $R(A)$ ,  $S(B) \in R$

- **projekce**  $R$  na množinu atributů  $C$ , kde  $C \subseteq A$

Značení:  $R[C]$

- **selekce**  $R$  podle podmínky  $\varphi$

Značení:  $R(\varphi)$

- **spojení**  $R$  a  $S$

Značení:  $R * S$

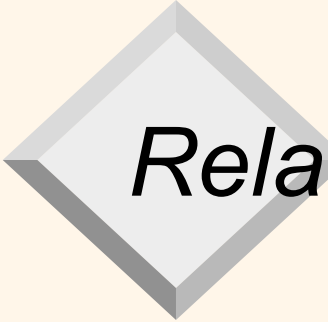
Př.:  $(\text{Hrají}(\text{Název}_k = \text{Mír})[\text{Jméno}_f, \text{Čas}] * \text{Film})[\text{Herec}]$

Další: **sjednocení**  $\cup$ , **průnik**  $\cap$ , **rozdíl**  $-$ ,

**kartézský součin**  $\times$

Stačí:  $\times$ ,  $\cup$ ,  $-$ , projekce, selekce





# *Relační algebra - dotazovací jazyk*

Další (odvozené) operace:

- spojení relací (přirozené,  $\theta$ -spojení, polospojení)
- dělení relací
- kompozice relací

**!** vnější spojení je mimo základní RMD

vyžaduje prázdné hodnoty

Pz.: vlastnosti relačních operací umožňují realizovat algebraickou optimalizaci dotazů



# *DRK - doménový relační kalkul*

Jde o podmnožinu predikátového kalkulu 1.řádu

- termy: proměnné, konstanty
- predikátové symboly: **R**, porovnávací ( $=, \neq, <, >, \geq, \leq$ )
- logické spojky ( $\neg, \wedge, \vee, \Rightarrow$ )
- Kvantifikátory ( $\exists, \forall$ )

Další pojmy: volné a vázané proměnné

ohodnocení proměnných, interpretace

predikátových symbolů, vyhodnocení formulí

*Dotaz v DRK* je výraz  $\{x_1, \dots, x_k \mid A(x_1, \dots, x_k)\}$



# *DRK - doménový relační kalkul*

*Př.:*

$\{x,y \mid \text{Kina}(x, \text{'Strašnická'}, y)\}$

$\{\text{herec}, \text{rež} \mid \text{Film}(\text{'Černí baroni'}, \text{herec}, \text{rež})\}$

$\{\text{herec} \mid \exists \text{rež} \text{ Film}(\text{'Černí baroni'}, \text{herec}, \text{rež})\}$

syntaktická zjednodušení:

- zavedeme atributy
- odstranění zbytečných  $\exists$

$\{a, jf \mid \exists k (\text{Hrají}(\text{Název}_k:k, \text{Jméno}_f:jf) \wedge \text{Kina}(\text{Název}_k:k, \text{Adresa}:a))\}$



# *DRK - doménový relační kalkul*

Složitější dotazy:

Najdi filmy, které dávají ve všech kinech, kde něco hrají.

$\{f \mid \forall k(\text{Hrají}(\text{Název}_k:k) \Rightarrow \text{Hrají}(\text{Název}_k:k, \text{Jméno}_f:f))\}$

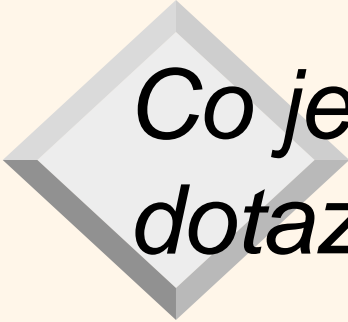
Problémy:

- jak kvantifikovat, když je doména nekonečná
- jak řešit některé dotazy s negací či disjunkcí

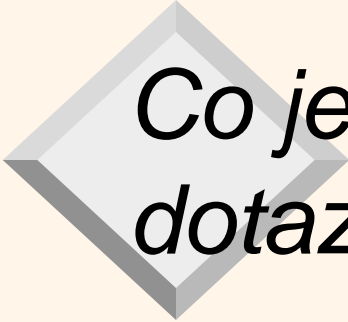
Př.:  $\{x \mid \neg R(x)\}$

$\{x,y \mid R('a',x) \vee S('b',y)\}$

Řešení: omezená interpretace, bezpečné výrazy



*Co je databázový dotaz, co je  
dotazovací jazyk?*



# *Co je databázový dotaz, co je dotazovací jazyk?*

Najdi filmy, které dávají ve všech kinech, kde něco hrají.

$$\{f \mid \forall k (\text{Hrají}(\text{Název}_k:k) \Rightarrow \text{Hrají}(\text{Název}_k:k, \text{Jméno}_f:f))\}$$
$$\{f \mid \neg \exists k (\text{Hrají}(\text{Název}_k:k) \wedge \neg \text{Hrají}(\text{Název}_k:k, \text{Jméno}_f:f))\}$$

# Co je databázový dotaz, co je dotazovací jazyk?

- (Databázový) dotaz typu  $(S \rightarrow T)$  je částečně rekurzivní funkce  $q$ , která pro každou databázi  $S^*$  poskytne odpověď  $q(S^*)$  typu  $T$ , nebo je na  $S^*$  nedefinovaná.

omezení:

- hodnoty v  $q(S^*)$  jsou z  $S^*$ ,
- odpověď nezávisí na reprezentaci dat v DB
- prvky DB se chápou jako neinterpretované objekty
- Dotazovací jazyk nad  $S$  je množina výrazů nad konečnou abecedou + významová funkce přiřazující každému výrazu dotaz.



# Vyjadřovací síla relačních jazyků

- **Vyjadřovací síla dotazovacího jazyka**  $J$  nad  $S$  je množina všech dotazů  $M(J)$ , které lze pomocí  $J$  vyjádřit.
  - $J_1 < J_2$  právě když  $M(J_1) \subset M(J_2)$
  - $J_1 \cong J_2$  právě když  $M(J_1) = M(J_2)$
- Dotazovací jazyk, kterým lze vyjádřit všechny databázové dotazy, se nazývá **úplný**.





# *Vyjadřovací síla relačních jazyků*

- Programování vs. relační algebra
  - relační algebra je jazyk velmi vysoké úrovně
- Dotazovací jazyk, který umožňuje realizovat relační algebru se nazývá **relačně úplný**.
- Komerční svět:
  - SQL,
  - jazyky formulářů,
  - obrázkové jazyky



# *Rozšíření relačních jazyků*

Problémy: dotazy na počet, průměr, hodnota v n-tici má být vypočítaná,

Dále: dotaz na všechny podřízené  
(tranzitivní uzávěr relace)

Otázka: lze vytvořit neprocedurální výpočetně úplný jazyk?

Dílčí řešení:

```
{k, počet | počet=COUNT(f | Hrají(Název_k:k, Jméno_f:f))}
```

zavedení nejmenšího pevného bodu

while, repeat, ...

Kompromis: SQL + uložené procedury