



Výpočetní technika a lékařská informatika

Databáze

Mgr. Markéta Trnečková, Ph.D.

Data

■ data vs. informace

■ Data

- údaje získané pozorováním, nebo měřením
- představují fakta, text, obraz, zvuk, video, nejčastěji v kontextu sledovaného procesu nebo situace
- nezávislá na uživateli, většinou odráží současný stav reality

■ Informace

- informací se data a vztahy mezi nimi stávají vhodnou interpretací pro uživatele vytvořením struktur, které odhalují uspořádání, vzory, tendence a trendy
- strukturovaná, organizovaná, shrnutá a interpretovaná data, silně závislé na tom, kdo je požaduje, tedy závislé na uživateli (individuální hloubka znalostí, zkušeností ...)

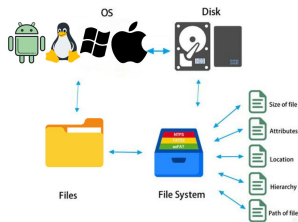
Kartotéka

- ukládání dat pomocí karet
- nevýhody:
 - složité vyhledávání
Potřebuji najít všechny pacienty s diabetem.
 - obtížná změna dat
Změnil se název pojišťovny.
 - nadbytečnost dat
U všech pacientů z Olomouce uložené stejné PSČ.
 - složité získávání informací
Potřebuji pacienty, kteří byli nemocní více jak 10×.



Souborový systém

- pro každou oblast samostatný soubor – pacienti, lékaři, léky
- jednodušší organizace dat a vyhledávání
- nevýhody:
 - nekonzistence = různé verze dat v různých souborech
 - datové anomálie
 - modifikace – změna je potřeba udělat u všech výskytu dat (i v různých souborech)
 - vkládání – pokud vkládáme např. novou fakturu, musíme k ní přiřadit klienta, který může být nový nebo již založený v souboru klientů = možnost způsobení nekonzistence
 - mazání – při smazání klienta mohou zůstat faktury bez klienta



Databázový systém

- databáze

- **Skládá se z**

- data – syrová fakta z oblasti zájmu uživatele
- metadata — popisují data a vztahy mezi daty v databázi
- programové vybavení – Systém řízení báze dat (SŘDB) = kolekce programů, která:
 - řídí databázovou strukturu
 - kontroluje přístup k datům uloženým v databázi
 - umožňuje přístup více uživatelům do databáze
 - pomáhá k efektivnější správě dat
- technické prostředky (hardware)
- uživatelé

Databázový systém

Příklad

Jaká data uchovává praktický lékař. Jakým způsobem je může organizovat pomocí kartotéky, souborového systému?

3

Příklad

Jaká bychom logicky tyto data mohli mezi sebou provázat? Jde to udělat pomocí databáze?

Databázový systém

Funkce SŘBD

- správa metadat
- správa dat
- transformace a prezentace dat
- správa zabezpečení
- víceuživatelský přístup
- zálohování a obnova
- správa integrity dat
- komunikační rozhraní s uživatelem

Databázový systém

Uživatelé

■ administrátor databáze, správce dat

- centrální kontrola
- definice paměťové struktury a přístupových metod
- přidělování práv
- ...

■ aplikační programátor

- vytváří aplikační programy

■ znalý uživatel

- využívají DBS ad hoc prostřednictvím databázového jazyka

■ naivní uživatel

- využívají DBS jen prostřednictvím aplikací

Databázový systém

Funkce databáze

■ správa dat

■ základní operace s daty:

- vkládání nových položek
- zrušení nepotřebných položek
- oprava, aktualizace vybraných položek
- vyhledávání informací v datech

■ instance databáze:

- zachycuje aktuální stav
- data uložená ke konkrétnímu časovému okamžiku

Databázový systém

Typy databází

- podle počtu uživatelů
 - jednouživatelská
 - víceuživatelská (stejná oprávnění / různá oprávnění)
- podle lokace
 - centralizovaná
 - distribuovaná
- podle způsobu použití
 - transakční nebo produkční – pro denní použití
 - datový sklad (warehouse) – pro získávání informací

Databázový systém

Vlastnosti dat

- **perzistentní** – trvale uložená, přetrvávají mezi operacemi
- **sdílená** – přístupná více uživatelům
- **integrovaná** – sjednocené informační systémy bez redundance
- **zabezpečená** – přístup a operace s daty mohou být omezeny systémem práv
- **konzistentní** – tytéž údaje na různých místech mají stejnou hodnotu
- **zachovávají** integritu – data odrážejí aktuální realitu
- **spolehlivá** – dají se rekonstruovat po chybě
- **rozsáhlá** – nestačí vnitřní paměť, používají se sofistikované algoritmy pro implementaci operací, ...

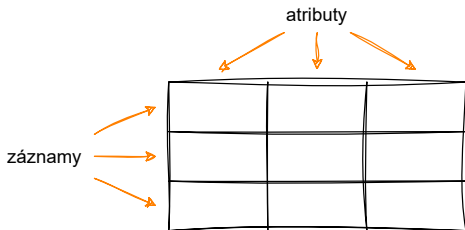
Databázový systém

System řízení bází dat

- Systém řízení bází dat (SŘBD) = software, který zprostředkovává komunikaci mezi koncovými uživateli, aplikacemi a samotnou databází s cílem uchovávat a analyzovat data
- Příklady:
 - MySQL – Vhodné pro webové projekty
 - MS SQL – Microsoft, menší projekty, využívající MS
 - Oracle – Velké databáze
- **SQL** – Simple Query Language

Relační databáze

- základní model – **tabulky**
- tabulky jsou propojeny prostřednictvím vztahů
- **atributy** – sloupce v tabulkách, charakteristiky objektů v tabulkách
- **záznamy** (n-tice, objekty) – řádky tabulek
- objekty jsou stejného druhu, ale každý je v databázi jen jednou (integrita entity)
- atributy \neq konkrétní hodnoty, pouze charakteristiky



Příklad v Excelu

- Budeme pracovat v MS Excel i když by se více hodilo používat MS Access
- **Snadná dostupnost:** téměř každý student má Excel nainstalovaný, Access méně běžně
- **Nízká bariéra učení:** věřím, že studenti již znají základní funkce (tabulky, filtry, vzorce)
- **Vizualizace a manipulace s daty:** snadné použití filtrů, kontingenčních tabulek a grafů pro rychlé demonstrace
- **Přechod na SQL a Access:** pochopení principů relačních databází v Excelu a snadnější přechod na Access nebo SQL v budoucnu



Atributy

- sloupce v tabulce, charakteristiky
- popisují instance v řádcích databáze
- **doména** – množina možných hodnot pro daný atribut
- **datový typ** – je odvozen z domény, např. celé číslo
- úkolem SŘBD je nedovolit zapsat do sloupce hodnotu, která neodpovídá datovému typu
- **atomickost atributu** – atribut je atomický (jednoduchý), pokud neobsahuje žádné smysluplné menší komponenty (jméno → jméno, příjmení)
- každý atribut obsahuje jednu hodnotu – pokud jich potřebujeme víc, musíme vytvořit novou tabulku
- příklad:
 - jméno, příjmení, telefonní číslo
 - více telefonních čísel na osobu
 - 2 tabulky: jméno, příjmení; telefonní číslo, osoba

Příklad v Excelu

Příklad

- 1 Vytvořte nový list a pojmenujte ho Pacienti a v něm tabulku obsahující sloupce Rodné číslo, Jméno, Příjmení, Rok_narození, Diagnóza
- 2 Do tabulky vložte cca 5 záznamů, vytvořte i pacienty, kteří mají stejnou diagnózu (např. Diabetes)

The screenshot shows the Microsoft Excel interface with a spreadsheet named 'Pacienti'. The spreadsheet contains the following data:

	A	B	C	D	E	F	G	H
1	RČ	Jméno	Příjmení	Rok_narození	Diagnóza			
2	805101/1234	Jana	Nováková	1980	Diabetes			
3	750615/2345	Petr	Svoboda	1975	Hypertenze			
4	905220/3456	Lucie	Dvořáková	1990	Diabetes			
5	851230/4567	Martin	Procházka	1985	Astma			
6	015305/5678	Eva	Králová	2001	Hypertenze			
7								
8								
9								

Datové typy

- **Integer** – číslo
- **Varchar** – krátký text (max. 255 znaků)
- **Text** – dlouhý text
- **Boolean** – logická hodnota (pravda/nepravda, . . .)
- **Date/DateTime** – datum, nebo datum a čas

Příklad

V excelu je možné nastavit sloupcům formát dat, ale také kontrolu vkládaných dat (Data → Ověření dat).

Přidejte vhodné omezení na hodnoty představující rok narození. Vyzkoušejte si vložit špatnou hodnotu.

Speciální atributy

Primární klíč

- n-tice nejsou uspořádané
- je potřeba identifikovat každou n-tici
- jeden atribut **unikátní** = primární klíč
- **!** každá n-tice má v tabulce unikátní hodnotu primárního klíče **!**
- každá tabulka **musí** obsahovat primární klíč

Příklad

Máme v tabulce Pacienti nějaký primární klíč?

Speciální atributy

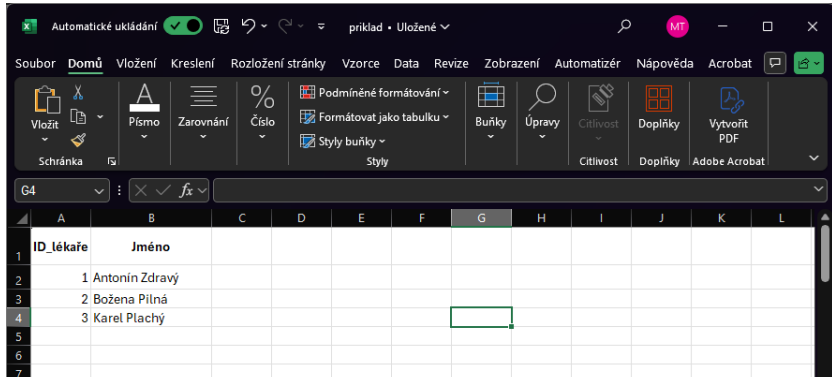
Cizí klíč

- primární klíče v databázi slouží k definování vztahů mezi tabulkami
- např. osoba: id-osoby, jméno, příjmení
- pokud se PK přeneso do jiné tabulky, stává se v ní cizím klíčem
- např. telefon: id-telefonu, telefonní číslo, id-osoby

Příklad v Excelu

Příklad

- 1 Vytvořte nový list a pojmenujte ho Lekari a v něm tabulku obsahující sloupce ID_Lékaře, Jméno
- 2 Do tabulky vložte několik záznamů
- 3 Do tabulky Pacienti přidejte sloupec Ošetřující_lékař (vhodně tabulky pomocí tohoto sloupce propojte)



The screenshot shows the Microsoft Excel interface. The ribbon is set to 'Domů' (Home). The formula bar shows 'G4'. The active cell is G4. The table below has the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID_lékaře	Jméno										
2		1 Antonín Zdravý										
3		2 Božena Pilná										
4		3 Karel Plachý										
5												
6												
7												

Vztahy

- logické propojení mezi různými tabulkami
- např. v databázi knihovny – tabulky autor a kniha
- význam vztahu (autor napsal knihu), násobnost

Násobnost vztahu

- kolik n-tic v první tabulce může být propojeno s jednou n-ticí ve druhé tabulce a naopak
- např. kolik knih může autor napsat, kolik autorů může napsat jednu knihu

vztah 1:n – one-to-many

- jednoduchý
- např. fotbalový tým (1 tým obsahuje více hráčů, každý hráč patří do 1 týmu)
- nový atribut cizího klíče v tabulce hráč, který identifikuje tým

vztah m:n – many-to-many

- záznam v obou tabulkách může souviset s libovolným počtem záznamů (nebo žádným) v druhé tabulce
- např. kniha, autor
- vztahy many-to-many vyžadují třetí tabulku, tzv. asociační nebo propojovací tabulku, protože relační systémy neumí vztah přímo reprezentovat
- např. tabulka kniha-autor

Násobnost vztahu

vztah 1:1 – one-to-one

- pouze jeden (nebo žádný) záznam na každé straně vztahu
- např. manželé – můžete být nebo nemusíte být ženatí/vdané, ale pokud ano, vy i váš partner máte jen jednoho manžela/manželku
- málo časté

Příklad

Jaký vztah je mezi Pacienty a Lékaři v ukázkovém příkladu?

Násobnost vztahu

Příklad

Abychom vztah definovali korektně, měli bychom přidat omezení dat ve sloupci Ošetřující_lékař, že jsou možné jen hodnoty z tabulky Lekari a sloupce ID_lékaře.

Data → Ověření dat → Povolit → Seznam → Vybereme sloupec odkud mohou hodnoty být.

Přidejte toto omezení.

Příklad v Excelu

Příklad

- 1 Vytvořte nový list a pojmenujte ho Léky a v něm tabulku obsahující sloupce ID_léku, Název, Účinná_látka
- 2 Do tabulky vložte cca 5 záznamů

The screenshot shows the Microsoft Excel interface with a spreadsheet titled 'příklad - Uložené'. The ribbon is set to 'Domů' (Home). The spreadsheet contains a table with the following data:

	A	B	C	D	E	F	G	H	I	J
1	ID_léku	Název	Účinná_látka							
2		101 Paralen	Paracetamol							
3		102 Metformin	Metformin							
4		103 Ventolin	Salbutamol							
5		104 Prestarium	Perindopril							
6										
7										

The bottom status bar shows the current sheet is 'Léky' and the zoom level is 100%.

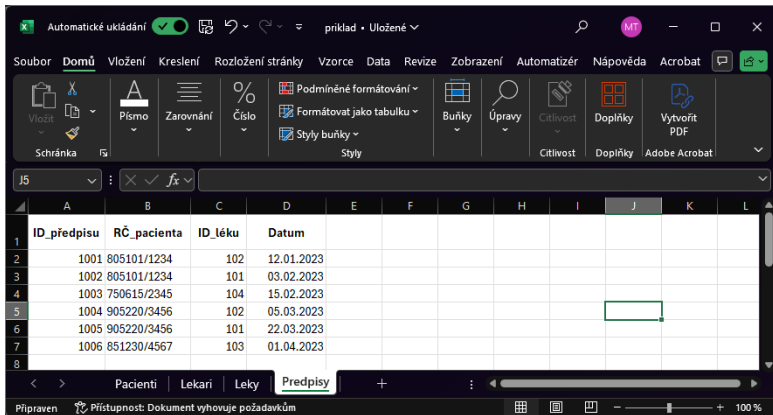
Příklad v Excelu

Příklad

1 Definujte vztah mezi Pacienty a Léky. Jak to uděláme?

Příklad v Excelu

- tabulka (nový list) Predpisy (vztah N:M mezi pacienty a léky)
- sloupce například: ID_předpisu, RČ_pacienta, ID_léku, Datum



The screenshot shows the Microsoft Excel interface with a table named 'Predpisy'. The table has four columns: ID_předpisu, RČ_pacienta, ID_léku, and Datum. The data is as follows:

	ID_předpisu	RČ_pacienta	ID_léku	Datum
1				
2	1001	805101/1234	102	12.01.2023
3	1002	805101/1234	101	03.02.2023
4	1003	750615/2345	104	15.02.2023
5	1004	905220/3456	102	05.03.2023
6	1005	905220/3456	101	22.03.2023
7	1006	851230/4567	103	01.04.2023
8				

Příklad

Nezapomeňte přidat omezení na hodnoty, které mohou být ve sloupcích RČ_pacienta a ID_léku.

Normalizace databáze

- proces strukturování relační databáze podle tzv. **normálních forem** s cílem snížit redundanci dat a zlepšit jejich integritu

První normální forma (1NF)

- doména každého atributu obsahuje pouze atomické hodnoty a hodnota atributu obsahuje vždy jen jednu hodnotu z dané domény
- **důvod**: ne-atomický atribut = omezená práce s tabulkou (např. hledání lidí ze stejného města)

Druhá normální forma (2NF)

- je v první normální formě (1NF), všechny atributy jsou **závislé na primárním klíči**
- špatný příklad: tabulka student (id, jméno, příjmení, univerzita, adresa univerzity)
- adresa univerzity není závislá na studentovi
- při změně adresy univerzity může vzniknout chyba (musíme ji měnit u všech studentů)
- **řešení**: nová tabulka s informacemi o univerzitě

Normalizace databáze

Třetí normální forma (3NF)

- je ve 2NF, všechny atributy jsou **závislé přímo na primárním klíči**
- příklad: zaměstnanec a plat
- pokud je plat odvozen z pracovní pozice, pak není závislý na primárním klíči, ale na pozici
- řešení: nová tabulka s pozicemi a platy

Příklad

Splňují naše tabulky normální formy? Pokud ne, jaké omezení zde může být?

Referenční integrita

- je vlastnost dat, která říká, že všechny jejich reference jsou platné
- dobře navržená databáze = SŘBD dokáže zajistit mezi propojenými tabulkami **referenční integritu**
- není možné odstranit autora, pokud existuje kniha, kterou napsal
- není možné změnit primární klíč, pokud je použit jako cizí klíč v jiné tabulce

Relační schéma

- atributy: A_1, A_2, \dots, A_n
- relační schéma: $R = (A_1, A_2, \dots, A_n)$
- např. kniha = (id, název, autor, počet-stran)
- relace nad relačním schématem: $r(R)$
- např. název(kniha)
- **instance**: konkrétní hodnoty – tabulka
- relační entita: n -tice (řádek v tabulce)
- např. 1 „Romeo a Julie“ „William Shakespeare“ 151

Základní operace

Selekce

- výběr řádků (n-tic)
- obvykle za použití podmínky → vyberou se řádky splňující podmínku
- `select * from book where number-of-pages > 100`
- podmínky lze skládat pomocí logických operátorů – or, and, ...

Příklad

- 1 Najděte pacienty, kteří jsou starší než 30 let.
- 2 Najděte všechny pacienty s diabetem.

Základní operace

Selekce

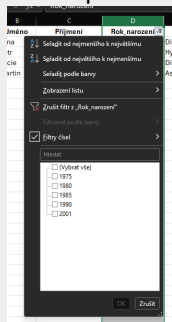
Příklad (Jak na to v excelu?)

1 Vyberte sloupec



2 Přidejte filtr

3 U sloupce se objeví šipka, kde je možné vytvořit filtr



Pozor! Sloupec musí mít správný datový typ (číslo), aby bylo možné roky porovnávat.

Základní operace

Projekce

- výběr sloupců, které chceme zobrazit
- `select author from book`

Příklad

Jak bychom projekci udělali v excelu?

Operace mezi tabulkami

Příklad

Přidejte nový list Leky2. Vytvořte zde kopii tabulky z listu Leky a přidejte nějaké další záznamy (můžete libovolně, ne jen na konec tabulky).

The screenshot shows the Microsoft Excel interface with the following data in the 'Leky' worksheet:

ID_léku	Název	Účinná látka
101	Paralen	Paracetamol
105	Aspirin	Acetylsalicylová kyselina
102	Metformin	Metformin
106	Ibuprofen	Ibuprofen
103	Ventolin	Salbutamol
107	Zyrtec	Cetirizin
104	Prestarium	Perindopril

The 'Leky2' worksheet is currently selected and empty in the bottom tab bar.

Operace mezi tabulkami

Rozdíl

- $r - s$
- obě relace musí být **kompatibilní** (stejný počet atributů, stejné domény)

Tabulka: r

A	B
α	1
α	2
β	1

Tabulka: s

A	B
α	2
β	3

Tabulka: $r - s$

A	B
α	1
β	1

Příklad

Pomocí podmíněného formátování můžeme zobrazit rozdíl tabulek Leky2 a Leky.

Domů → Podmíněné formátování → Nové pravidlo → Určit buňky k formátování pomocí vzorce.

=COUNTIFS(Leky!\$A:\$A;\$A2;Leky!\$B:\$B;\$B2;Leky!\$C:\$C;\$C2)=0

Nezapomeňte nastavit nějaký formát. Třeba pozadí buněk.

Operace mezi tabulkami

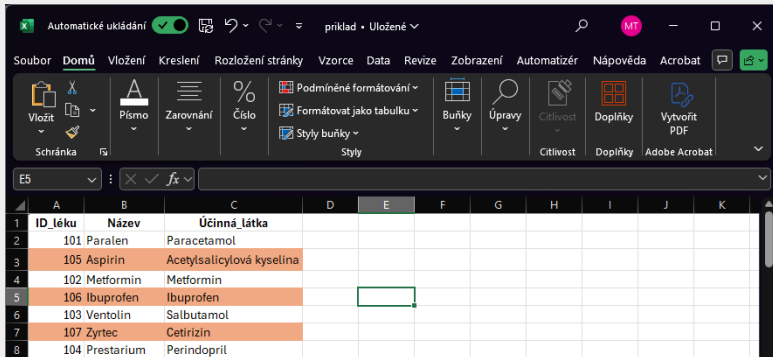
Rozdíl

Příklad

Pomocí podmíněného formátování můžeme zobrazit rozdíl tabulek Leky2 a Leky. Domů → Podmíněné formátování → Nové pravidlo → Určit buňky k formátování pomocí vzorce.

=COUNTIFS(Leky!\$A:\$A;\$A2;Leky!\$B:\$B;\$B2;Leky!\$C:\$C;\$C2)=0

Nezapomeňte nastavit nějaký formát. Třeba pozadí buněk.



The screenshot shows the Excel interface with the 'Podmíněné formátování' (Conditional Formatting) menu open. The 'Nové pravidlo' (New Rule) option is selected. The formula bar displays the formula: $=\text{COUNTIFS}(\text{Leky!}\$A:\$A;\$A2;\text{Leky!}\$B:\$B;\$B2;\text{Leky!}\$C:\$C;\$C2)=0$. The spreadsheet below shows a table with columns ID_léku, Název, and Účinná látka.

	A	B	C	D	E	F	G	H	I	J	K
1	ID_léku	Název	Účinná látka								
2	101	Paralen	Paracetamol								
3	105	Aspirin	Acetylsalicylová kyselina								
4	102	Metformin	Metformin								
5	106	Ibuprofen	Ibuprofen								
6	103	Ventolin	Salbutamol								
7	107	Zyrtec	Cetirizin								
8	104	Prestarium	Perindopril								

Operace mezi tabulkami

Sjednocení

- $r \cup s$
- **kompatibilní** relace

Tabulka: r

A	B
α	1
α	2
β	1

Tabulka: s

A	B
α	2
β	3

Tabulka: $r \cup s$

A	B
α	1
α	2
β	1
β	3

Příklad

Jak bychom mohli simulovat sjednocení tabulek?

Operace mezi tabulkami

Sjednocení

Příklad

Jak bychom mohli simulovat sjednocení tabulek?

Zkopírovat data z jedné tabulky do druhé a odstranit duplicitní záznamy.

Data → Odebrat duplicity

Operace mezi tabulkami

Kartézský součin

- $r \times s$
- relace mají různé atributy (pokud mají stejné jméno, jeden se musí přejmenovat)
- spojí všechny řádky r se všemi řádky s

Tabulka: r

A	B
α	1
β	2

Tabulka: s

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

Tabulka: $r \times s$

A	B	C	D	E
α	1	α	10	a
α	1	β	10	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b

Další operace

Přejmenování

- není klasická operace
- můžeme pojmenovat výsledek nějaké operace

Kompozice

- složitější operace = skládání základních operací
- např. kartézský součin $r \times s$ a výběr těch řádků, kde $A = C$

Tabulka: r

A	B
α	1
β	2

Tabulka: s

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

A	B	C	D	E
α	1	α	10	a
β	2	β	10	a
β	2	β	20	b

Další operace

- z praktických důvodů se zavádějí i další odvozené operace – zjednodušení

Průnik

- $r \cap s$
- kompatibilní relace
- výsledek: řádky obsažené v obou relacích

Tabulka: r

A	B
α	1
α	2
β	1

Tabulka: s

A	B
α	2
β	3

Tabulka: $r \cap s$

A	B
α	2

Příklad

Jak by se dal definovat průnik pomocí podmíněného formátování?

Další operace

Přirozené spojení (natural join)

- $r \bowtie s$
- relace r a s mají jeden nebo více párů shodně pojmenovaných sloupců
- sloupce musí mít stejný datový typ
- výsledek: spojení, kde shodně pojmenované sloupce mají stejné hodnoty

Tabulka: r

A	B
α	1
β	2

Tabulka: s

A	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

Tabulka: $r \bowtie s$

A	B	D	E
α	1	10	a
β	2	10	a
β	2	20	b

Příklad

Příklad

Ke každému pacientovi přidejte sloupec, ve kterém bude jméno ošetřujícího lékaře.

Příklad

Příklad

Ke každému pacientovi přidejte sloupec, ve kterém bude jméno ošetřujícího lékaře.

```
=SVYHLEDAT(F2;Lekari!$A:$C;2;NEPRAVDA)
```

Další operace

Dělení

- $r \div s$
- dotaz typu „pro všechna“
- vstupní relace: $r = (A_1, \dots, A_n, B_1, \dots, B_m)$, $s = (B_1, \dots, B_m)$
- výsledná relace: $r \div s = (A_1, \dots, A_n)$
- $r \div s = \{t \mid t \in \Pi_{R-S}(r) \wedge \forall u \in S : tu \in r\}$
- $\Pi_{R-S}(r)$ znamená výběr sloupců

Tabulka: r

A	B	C	D	E
α	a	α	a	1
α	a	γ	a	1
α	a	γ	b	1
β	a	γ	a	1
β	a	γ	b	3
γ	a	γ	a	1
γ	a	γ	b	1
γ	a	β	b	1

Tabulka: s

D	E
a	1
b	1

Tabulka: $r \bowtie s$

A	B	C
α	a	γ
γ	a	γ

Další operace

Tabulka: r

A	B	C	D	E
α	a	α	a	1
α	a	γ	a	1
α	a	γ	b	1
β	a	γ	a	1
β	a	γ	b	3
γ	a	γ	a	1
γ	a	γ	b	1
γ	a	β	b	1

0

Tabulka: s

D	E
a	1
b	1

Tabulka: $r \times s$

A	B	C
α	a	γ
γ	a	γ

$$D E = a 1$$

A	B	C
α	a	α
α	a	γ
β	a	γ
γ	a	γ

$$D E = b 1$$

A	B	C
α	a	γ
γ	a	γ
α	a	β