Diskrétní struktury 1

učební text k přednáškám

Radim Bělohlávek

radim.belohlavek@acm.org

Katedra informatiky
Univerzita Palackého v Olomouci

2020

Text je určen studentům informatických oborů na Přírodovědecké fakultě Univerzity Palackého v Olomouci. Je zamýšlen jako studijní text k předmětu Diskrétní struktury 1. Text bude v některých částech ještě upraven. Zjištěné chyby a případné další připomínky mi prosím zasílejte na adresu radim.belohlavek@acm.org.

Text vychází z předchozích učebních textů, R. Bělohlávek, V. Vychodil, Diskrétní matematika pro informatiky, UP, Olomouc, 2004, a R. Bělohlávek, Úvod do informatiky, UP, Olomouc, 2008, které doporučuji jako další zdroje ke studiu (rozsah zejména prvního z nich je větší).

Obsah

1 Logika

Studijní cíle: Po prostudování této kapitoly by student měl rozumět základním pojmům logiky. Měl by také rozumět základům výrokové logiky.

Klíčová slova: logika, pravdivostní hodnota, logická spojka, formule, výroková logika.

1.1 Co a k čemu je logika

Logika je věda o správném usuzování.

Logika je vědou o správném usuzování. V logice jde o to, aby usuzování (neboli logické odvozování) mělo správnou *formu* bez ohledu na *obsah*. Uvažujme např. tvrzení „Prší.“ a „Jestliže prší, pak jsou silnice mokré“. Z nich lze odvodit tvrzení „Silnice jsou mokré.“ Uvažujme jinou dvojici tvrzení, např. „Petr má hlad.“ a „Jestliže má Petr hlad, pak se Petr snaží sehnat něco k jídlu.“ Z těchto tvrzení lze odvodit tvrzení „Petr se snaží sehnat něco k jídlu.“ V uvedených příkladech vyplývá z dané dvojice tvrzení další tvrzení. Uvedené dvojice tvrzení mají jiný obsah, neboť např. „Prší.“ znamená něco jiného než „Petr má hlad.“ Způsob, jakým jsme odvodili třetí tvrzení byl však v obou případech stejný. Říkáme, že obě odvození měla stejnou formu. Tuto formu je možné znázornit *symbolicky* takto: z tvrzení A, a tvrzení $A \to B$ (čteme „jestliže A, pak B“), plyne tvrzení B.

V logice jde o formu usuzování, ne o obsah usuzování.

Uvedené rysy jsou pro moderní logiku charakteristické, proto je zopakujme: logika studuje formy usuzování bez ohledu na obsah. Moderní logika má symbolický charakter, neboť jednotlivá tvrzení označujeme symboly (např. výše uvedenými symboly „A“ a „B“) a způsoby spojení tvrzení ve složitější tvrzení označujeme také symboly (např. výše uvedený „$\to$“). Pro uvedené rysy bývá moderní logika označována jako *logika formální*, popř. *symbolická*. Symbolický charakter umožňuje logice snadněji odhlédnout od nepodstatného (tedy od obsahu) a soustředit se na podstatné (např. na formy usuzování).

Logika má symbolický charakter.

Slovo „logika“ má v běžném životě i jiné významy. Tyto významy jsou od původního významu, který jsme si právě vysvětlili, odvozené, ale jsou nepřesné a zavádějící. Např. větou „Předložený návrh nemá žádnou logiku.“ chce autor patrně říct, že návrh je nesrozumitelný, popř. špatně zdůvodněný, popř. neracionální. Větou „Logika našeho podnikání spočívá v maximálním uspokojování potřeb zákazníka.“ chce autor nejspíš říct, že uspokojování potřeb zákazníků je hlavním rysem jeho podnikatelské strategie. „To je nelogické.“ znamená, že to je nesmyslné nebo nějak chybné. Podobných příkladů můžeme najít celou řadu.

Při studiu otázek, které v logice vznikají, se často používá matematických metod. Z tohto důvodu se někdy hovoří o *matematické logice*.

Budeme se zabývat klasickou logikou.

Dalším přívlastkem, který se v souvislosti s pojmem logika používá, je „klasická“, popř. „neklasická“. Zjednodušeně lze říci, že *klasickou logikou* se rozumí logika, která používá dvě pravdivostní hodnoty (pravda a nepravda) a tzv. klasické logické spojky. Mezi klasické logické spojky patří např. spojka „jestliže . . . , pak. . . “, se kterou jsme se setkali výše. Logika, která se zabývá i jinými spojkami než klasickými, popř. dalšími aspekty, kterými se klasická logika nezabývá, se nazývá *neklasická logika*. Příkladem neklasické spojky je spojka „je možné, že . . . “. Touto spojkou se zabývá modální logika. Dalším příkladem neklasické logiky je tzv. temporální logika (někdy nazývaná logikou času). Zabývá se tvrzeními, ve kterých hraje důležitou roli čas, např. „Po zelené naskočí na semaforu oranžová.“, „Každý den vychází Slunce.“ Dalším příkladem neklasické logiky je tzv. *fuzzy logika*. Ta se zabývá tvrzeními, které mohou mít kromě pravdivostních hodnot pravda a nepravda i jiné hodnoty. Např. tvrzení „Zákazník je spokojený.“ může mít pravdivostní hodnotu 1 (pravda), pokud je zákazník zcela spokojený, 0 (nepravda), pokud je nespokojený, ale i 0.8, pokud je víceméně spokojený, ale ne zcela.

Existují i neklasické logiky, např. modální logika, temporální logika, fuzzy logika.

Vztah logiky a informatiky je bohatý a různorodý. Se základy logiky by měl být obeznámen každý informatik. Znalost základů logiky nám umožňuje srozumitelně a jednoznačně se vyjadřovat a argumentovat. To je pochopitelně užitečné pro každého, nejen pro informatika. Pro informatika je to však důležité i proto, že svoje konstrukce a návrhy musí „sdělit počítači“, např. ve formě zdrojového kódu napsaného ve vhodném programovacím jazyku. Zdrojový kód obvykle obsahuje výrazy, které se vyhodnocují podle pravidel logiky (např. podmínky v příkazech větvení „if ... then ... else ... “). Logika nás těmto pravidlům učí. Zdrojový kód musí být přesný, jinak je program chybný. Chyby mohou mít dalekosáhlé následky (pomysleme na program pro výpočet mezd, program pro řízení elektrárny apod.). Zdrojový program musí být také srozumitelný, jinak mu nikdo jiný než jeho autor nebude rozumět (a po čase mu nebude rozumět ani jeho autor). Logika nás učí přesnosti i srozumitelnosti. To je další významný efekt studia logiky.

Vztah logiky a informatiky je bohatý a různorodý.

Logika nás učí přesnosti i srozumitelnosti.

Význam logiky pro informatiku spočívá také v metodě, kterou logika používá. Jak uvidíme v dalším výkladu, logika ve svém formálním přístupu používá místo přirozeného jazyka zjednodušený, zato přesný pojem jazyk logiky, zabývá se syntaktickými aspekty (jak se tvrzení nebo odvozování správně zapisují), sémantickými aspekty (jaký význam tvrzení a odvozování mají) apod. Podobně je to například v programování: programovací jazyk je jednoduchý, ale přesně definovaný, formální jazyk a při programování se potýkáme s různými syntaktickými aspekty (je zdrojový kód správně zapsán?) i sémantickými aspekty (dělá program to, co má?). V neposlední řadě spočívá význam logiky pro informatiku v tom, že řada oblastí informatiky využívá logiku jako nástroj. Pro příklad jmenujme logické programování, expertní systémy a obecněji umělou inteligenci, analýzu dat a verifikaci.

Logika má uplatnění v různých oblastech informatiky.

1.2 Základní pojmy logiky

1.2.1 Výroky, logické spojky, pravdivost výroků

Výrokem intuitivně rozumíme tvrzení (výpověď), u kterého má smysl uvažovat, zda je pravdivé. Výroky jsou např. následující tvrzení.

Výrok je tvrzení, které může být pravdivé, nebo nepravdivé.

Prší.

Byl jsem v obchodě a koupil jsem si knihu.

Když prší, jsou mokré silnice.

$2 + 2 = 4$ a $3 < 100$.

$2 + 2 = 6$.

Následující tvrzení ale nejsou výroky.

Kniha v obchodě.

$2 + 2$

Ať je pěkné počasí.

Z jednodušších výroků se vytvářejí složitější výroky pomocí tzv. *logických spojek*[1]. Logické spojky jsou speciální jazykové výrazy jako např.

Logické spojky jsou jazykové výrazy, kterými z jednodušších výroků vytváříme výroky složitější.

„... a ... “, „... nebo ... “, „jestliže ..., pak ... “,
„..., právě když ... “, „ne ... “ (tj. „není pravda, že ... “).

[1] Místo „logická spojka“ říkáme často jen „spojka“.

Například z výroku „2+2=4“ a výroku „Prší.“ vytvoříme pomocí spojky „a“ výrok „2+2=4 a prší.“ Z výroku „Prší.“ vytvoříme pomocí spojky „ne“ (tj. „není pravda, že . . . “) výrok „Neprší.“ (tj. „Není pravda, že prší.“). Z výroků „Prší.“ a „Silnice jsou mokré.“ vytvoříme pomocí spojky „jestliže . . . , pak . . . “ výrok „Jestliže prší, pak jsou silnice mokré.“

Průvodce studiem

Všimněme si, že pojem spojka zde používáme v širším významu než bývá běžné: spojkou chápeme jazykový výraz, jehož použitím na výroky dostaneme nový výrok. Např. výrok „Jestliže prší, pak jsou silnice mokré.“ vznikl použitím spojky „jestliže. . . , pak . . . “ na výroky „Prší.“ a „Silnice jsou mokré.“ Z hlediska českého jazyka je výraz „jestliže“ spojkou, jinou spojkou je výraz „pak“.

Všimněme si také, že při použití spojek na výroky měníme pořadí větných členů. Např. v právě uvedeném příkladu by přísně vzato výsledným výrokem měl být výrok „Jestliže prší, pak silnice jsou mokré.“. Tento výrok ale nezní pěkně. Podobné jazykové jevy budeme z hlediska logiky považovat za podružné a nebudeme se jimi zabývat. Ze stejného důvodu nebudeme rozlišovat např. tvrzení „Neprší.“ a „Není pravda, že prší.“

Klasická logika, které se věnujeme, se zabývá tzv. klasickými spojkami. Mezi ně patří výše uvedené spojky. s dalšími klasickými spojkami se seznámíme později.

Průvodce studiem

Kromě klasických spojek se v běžném životě používají i další spojky, např. „je možné, že . . . “, „je nutné, že . . . “, „ví se, že . . . “, „věří se, že . . . “. Např. z výroku „Vesmír je nekonečný.“ vytvoříme pomocí spojky „věří se, že . . . “ výrok „Věří se, že vesmír je nekonečný.“, pomocí spojky „je možné, že . . . “ pak výrok „Je možné, že vesmír je nekonečný.“ Spojkami „je možné, že . . . “, „je nutné, že . . . “ se zabývá modální logika, spojkami „ví se, že . . . “, „věří se, že . . . “ se zabývá epistemická logika. Tyto spojky jsou složitější než klasické spojky a my se jimi zabývat nebudeme.

Některé výroky jsou pravdivé (např. „2+2=4“), některé jsou nepravdivé (např. „2+2=6“). Jak ukazuje následující úvaha, existují i jiné výroky. Těmi se klasická logika nezabývá.

Průvodce studiem

U některých výroků má smysl uvažovat o jejich pravdivosti (jsou tedy výroky), přesto se zdráháme říci, že dané tvrzení je pravdivé, nebo že není pravdivé (je nepravdivé).

Uvažujme výrok „Člověk s výškou 180cm je vysoký“. Na otázku, zda je pravdivý, bychom přirozeně odpověděli, že do jisté míry ano, ale ne zcela. Intuitivně bychom mu tedy nepřiznali ani pravdivostní hodnotu 1 (pravdivý), ani 0 (nepravdivý), ale např. hodnotu 0.8, která vyjadřuje, že je skoro pravdivý. Studiem takových výroků se zabývá fuzzy logika. Fuzzy logika našla použití v mnoha oblastech, např. v automatickém řízení nebo v expertních systémech.

Pravdivost některých výroků, jako je např. „Za týden bude pršet.“, závisí na čase. Takovými výroky se zabývá temporální logika (logika času). My se takovými tvrzeními zabývat nebudeme.

Klasická logika se tedy zabývá jen speciálními výroky, a to těmi, které jsou buď pravdivé, nebo nepravdivé.

Pravdivostní hodnota výroku vyjadřuje, zda je výrok pravdivý, nebo nepravdivý.

Je-li výrok V pravdivý, říkáme také, že má pravdivostní hodnotu „pravda“. Místo „pravda“ používáme symbolické označení 1, a říkáme tedy, že V má pravdivostní hodnotu 1. Je-li výrok V nepravdivý, říkáme, že má pravdivostí hodnotu „nepravda“, popř. že má pravdivostní hodnotu 0. Pravdivostní hodnotu výroku V označujeme $||V||$ a, že je výrok V pravdivý, resp. nepravdivý, tedy zapisujeme zapisujeme

$$||V|| = 1, \qquad \text{resp. } ||V|| = 0.$$

Jak ale zjistíme pravdivostní hodnotu výroku? Podívejme se na následující výrok.

> Prší a venkovní teplota je menší než 15°C.

Tento výrok vznikl použitím spojky „a“ na výrok „Prší.“ a na výrok „Venkovní teplota je menší než 15°C.“ Výroky „Prší.“ a „Venkovní teplota je menší než 15°C.“ neobsahují žádné spojky. Takovým výrokům říkáme *atomické*. Pravdivostní hodnota výroku „Prší a venkovní teplota je menší než 15°C.“ závisí na pravdivostních hodnotách výroků „Prší.“ a „Venkovní teplota je menší než 15°C.“ Jsou-li oba výroky „Prší.“ i „Venkovní teplota je menší než 15°C.“ pravdivé, je i výrok „Prší a venkovní teplota je menší než 15°C.“ pravdivý. Jinak, tj. je-li některý z výroků „Prší.“ a „Venkovní teplota je menší než 15°C.“ nepravdivý, je výrok „Prší a venkovní teplota je menší než 15°C.“ nepravdivý.

Uvedený postup má obecnou platnost, a proto ho rozeberme podrobněji. Označme složený výrok „Prší a venkovní teplota je menší než 15°C.“ symbolem V. Atomické výroky „Prší.“ a „Venkovní teplota je menší než 15°C.“ označmě symboly V_1 a V_2. Označme dále spojku „a“ symbolem $\wedge$. Výrok V má tedy tvar (někdy říkáme formu)

$$V_1 \wedge V_2.$$

Pravdivostní hodnotu $||V_1 \wedge V_2||$ výroku $V_1 \wedge V_2$ vlastně „spočítáme“ z pravdivostních hodnot $||V_1||$ a $||V_2||$ výroků V_1 a V_2 pomocí významu spojky „a“. Význam spojky „a“ je dán následující tabulkou:

$\wedge^{\cdot}$	0	1
0	0	0
1	0	1

Tato tabulka popisuje, jak spojka „a“ přiřazuje dvojicím pravdivostních hodnot výsledné pravdivostní hodnoty. Dvojici 0 a 0 je tedy přiřazena hodnota 0, dvojici 0 a 1 hodnota 0, dvojici 1 a 0 hodnota 0, dvojici 1 a 1 hodnota 1, protože na průsečíku řádku označeného 0 a sloupce označeného 0 je hodnota 0 atd. Toto přiřazení (funkci) označujeme $\wedge^{\cdot}$ a nazýváme ho pravdivostní funkce spojky $\wedge$. Tabulka tedy říká $0 \wedge^{\cdot} 0 = 0$, $0 \wedge^{\cdot} 1 = 0$, $1 \wedge^{\cdot} 0 = 0$, $1 \wedge^{\cdot} 1 = 1$. Pravdivostní hodnota $||V_1 \wedge V_2||$ výroku $V_1 \wedge V_2$ je tedy dána vztahem

$$||V_1 \wedge V_2|| \;=\; ||V_1|| \wedge^{\cdot} ||V_2||.$$

Tomuto vztahu je třeba rozumět takto: Pravdivostní hodnotu $||V_1 \wedge V_2||$ výroku $V_1 \wedge V_2$ (levá strana rovnice) spočítáme použitím funkce $\wedge^{\cdot}$ na pravdivostní hodnoty $||V_1||$ a $||V_2||$ výroků V_1 a V_2. Tímto použitím získáme hodnotu $||V_1|| \wedge^{\cdot} ||V_2||$ funkce $\wedge^{\cdot}$ na pravdivostních hodnotách $||V_1||$ a $||V_2||$ (pravá strana rovnice). Hodnotu $||V_1|| \wedge^{\cdot} ||V_2||$ zjistíme z výše uvedené tabulky: je to hodnota na průsečíku řádku označeného $||V_1||$ a sloupce označeného $||V_2||$.

Otázkou zůstává, jak zjistíme pravdivostní hodnoty $||V_1||$ a $||V_2||$ výroků V_1 a V_2. Zde musíme rozlišit dva případy.

1. Je-li výrok V_i atomický, tj. neobsahuje logické spojky, pak jeho pravdivostní hodnota musí být dána „zvenčí“. Např. v našem případě jsou oba výroky V_1 („Prší.“) i V_2 („Venkovní teplota je menší než 15°C.“) atomické. Jejich pravdivostní hodnota je dána „zvenčí“ v tom smyslu, že nám ji někdo řekne, popř. ji sami zjistíme (podíváme se z okna, podíváme se na teploměr, najdeme na internetu apod.). Obecně budeme předpokládat, že existuje nějaký externí zdroj informací, označme ho e, pomocí kterého pravdivostní hodnotu $e(V_i)$ výroku V_i zjistíme.

2. Není-li výrok V_i atomický, tj. obsahuje logické spojky, pak je to složený výrok a jeho pravdivostní hodnotu spočítáme podobně jako jsme počítali pravdivostní hodnotu původního výroku V. Pokud je např. výrok V_1 složeným výrokem a má tvar $V_{11} \wedge V_{12}$, pak pravdivostní hodnotu $||V_1||$ výroku V_1, tj. hodnotu $||V_{11} \wedge V_{12}||$ výroku $V_{11} \wedge V_{12}$, spočítáme podle vztahu

$$||V_{11} \wedge V_{12}|| \;=\; ||V_{11}|| \wedge^{\cdot} ||V_{12}||.$$

   Výrok $V_1 \wedge V_2$ má tedy v tomto případě tvar $(V_{11} \wedge V_{12}) \wedge V_2$ (k vyjádření jeho struktury jsme použili závorky) a jeho pravdivostní hodnotu určíme následovně:

$$||V_1 \wedge V_2|| = ||(V_{11} \wedge V_{12}) \wedge V_2|| \;=\; ||V_{11} \wedge V_{12}|| \wedge^{\cdot} ||V_2|| = (||V_{11}|| \wedge^{\cdot} ||V_{12}||) \wedge^{\cdot} ||V_2||.$$

   Výroky V_{11} a V_{12} mohou být přitom opět složené nebo atomické a při určování jejich pravdivostních hodnot postupujeme obdobně.

Pravdivostní hodnota výroku V tedy závisí na významu logických spojek a na pravdivostních hodnotách atomických výroků, ze kterých se výrok V skládá. V našem případě závisí pravdivostní hodnota výroku V na pravdivostní funkci $\wedge^{\cdot}$ spojky „a“ (tj. na výše uvedené tabulce) a na pravdivostních hodnotách atomického výroku „Prší.“ a atomického výroku „Venkovní teplota je menší než 15°C.“ Tabulky pravdivostních funkcí logických spojek jsou určeny tím, jak spojky používáme v přirozeném jazyce. Jsou tedy dané a nelze je měnit (uvědomte si, že výše uvedená tabulka skutečně popisuje význam spojky „a“, jak ho známe z běžného jazyka). Jak jsme ale řekli, pravdivostní hodnoty atomických výroků jsou dány zvenčí. Někdo nám je musí sdělit nebo je musíme zjistit. Obecně předpokládáme, že pravdivostní hodnoty e(„Prší.“) a e(„Venkovní teplota je menší než 15°C.“) výroků „Prší.“ a „Venkovní teplota je menší než 15°C.“ zjistíme z nějakého externího zdroje informací e. Na tomto zdroji tedy pravdivostní hodnota výroku V závisí. Proto přesněji píšeme

$$||V||_e \qquad \text{místo} \qquad ||V||,$$

abychom explicitně zdůraznili, že jde o pravdivostní hodnotu výroku V při zadaném e. Na e se vlastně můžeme dívat jako na přiřazení, které atomickým výrokům přiřazuje pravdivostní hodnoty. Přiřazení e se proto v logice nazývá pravdivostní ohodnocení (ohodnocuje atomické výroky).

Stejně jako v případě spojky „a“ postupujeme i v případě ostatních logických spojek, např. „ne“, „nebo“, „jestliže ..., pak ...“ a „..., právě když ...“. Tyto spojky označujeme po řadě symboly $\neg$, $\vee$, $\rightarrow$ a $\leftrightarrow$. Jejich významy, tj. zobrazení popisující přiřazení pravdivostních hodnot, pak označujeme $\neg^{\cdot}$, $\vee^{\cdot}$, $\rightarrow^{\cdot}$, $\leftrightarrow^{\cdot}$. Přehled právě uvedených logických spojek podává tabulka 1.

Průvodce studiem

Každá logická spojka má své označení a svůj význam. Označením je symbol logické spojky. Významem je pravdivostní funkce logické spojky. Například označením spojky „a“ je symbol $\wedge$, významem spojky „a“ je pravdivostní funkce $\wedge^{\cdot}$. Symboly a pravdivostní funkce základních logických spojek ukazuje tabulka 1.

Pravdivostní hodnota výroku se počítá z pravdivostních

název	zápis v přirozeném jazyce	symbol	pravdivostní funkce	tabulka pravd. funkce
negace	„ne“	$\neg$	$\neg^{\cdot}$	$\begin{array}{c\|c} a & \neg^{\cdot} a \\ \hline 0 & 1 \\ 1 & 0 \end{array}$
konjunkce	„a“	$\wedge$	$\wedge^{\cdot}$	$\begin{array}{c\|cc} \wedge^{\cdot} & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$
disjunkce	„nebo“	$\vee$	$\vee^{\cdot}$	$\begin{array}{c\|cc} \vee^{\cdot} & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 1 \end{array}$
implikace	„jestliže . . . , pak . . . “	$\rightarrow$	$\rightarrow^{\cdot}$	$\begin{array}{c\|cc} \rightarrow^{\cdot} & 0 & 1 \\ \hline 0 & 1 & 1 \\ 1 & 0 & 1 \end{array}$
ekvivalence	„. . . , právě když . . . “	$\leftrightarrow$	$\leftrightarrow^{\cdot}$	$\begin{array}{c\|cc} \leftrightarrow^{\cdot} & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 1 \end{array}$

Tabulka 1: Základní logické spojky.

Příklad 1.1. Máme za úkol určit pravdivostní hodnotu výroku „$2 + 2 = 5$ nebo číslo 10 je dělitelné číslem 6.“ Jde o složený výrok, který vznikl použitím spojky disjunkce („nebo“) na atomické výroky „$2+2 = 5$“ a „Číslo 10 je dělitelné číslem 6.“ Označíme-li tyto atomické výroky symboly V_1 a V_2 a složený výrok symbolem V, můžeme výrok V zapsat symbolicky ve tvaru $V_1 \vee V_2$. Přitom víme, že „$2 + 2 = 5$“ je nepravdivý výrok a „Číslo 10 je dělitelné číslem 6.“ je také nepravdivý výrok, tj. $e(V_1) = 0$ a $e(V_2) = 0$ neboli $||V_1||_e = 0$ a $||V_2||_e = 0$. Pro pravdivostní hodnotu výroku „$2 + 2 = 5$ nebo číslo 10 je dělitelné číslem 6.“ potom dostaneme

$$\begin{aligned} & ||\text{„}2 + 2 = 5 \text{ nebo číslo 10 je dělitelné číslem 6.“}|| = \\ = \quad & ||V||_e = ||V_1 \vee V_2||_e = ||V_1||_e \vee^{\cdot} ||V_2||_e = 0 \vee^{\cdot} 0 = 0. \end{aligned}$$

Poznámka 1.2. Jak jsme již viděli, pro zápis složených výroků často používáme závorky, abychom jednoznačně vyznačili strukturu výroku. Např. kdybychom napsali „$2{\cdot}3 = 5$ a $2{+}2 = 5$ nebo $2{+}2 = 4$“, není jasné, jestli myslíme „$2{\cdot}3 = 5$ a ($2{+}2 = 5$ nebo $2 + 2 = 4$)“ nebo „$(2 \cdot 3 = 5$ a $2 + 2 = 5)$ nebo $2 + 2 = 4$“. Všimněme si, že při prvním uzávorkování dostaneme nepravdivý výrok, zatímco při druhém uzávorkování dostaneme výrok pravdivý. Závorky používáme i při symbolickém zápisu výroků. Píšeme např. $V_1 \wedge (V_2 \vee V_3)$, $(V_1 \wedge V_2) \vee V_3$.

Shrňme nyní, co víme o určování pravdivostní hodnoty výroku.

Průvodce studiem

Je-li dán výrok V v přirozeném jazyce a máme-li určit jeho pravdivostní hodnotu $||V||_e$ při daném e, postupujeme následovně.

1. Určíme atomické výroky $V_1, \ldots, V_n$, ze kterých se V skládá.
2. Určíme pravdivostní hodnoty $e(V_1), \ldots, e(V_n)$ atomických výroků $V_1, \ldots, V_n$. Hodnoty $e(V_i)$ jsou součástí zadání nebo je zjistíme nebo je známe.
3. Výrok V zapíšeme v symbolické podobě, dostaneme např. $V_1 \wedge (V_2 \to V_3)$.
4. Je-li V atomický výrok, pak $||V||_e = e(V)$.
5. Je-li V složený výrok, tj. má jeden z tvarů $\neg V_1$, $V_1 \wedge V_2$, $V_1 \vee V_2$, $V_1 \to V_2$, $V_1 \leftrightarrow V_2$, pak jeho pravdivostní hodnotu určíme podle pravidel
   - $||\neg V_1||_e = \neg^{\cdot} ||V_1||_e$,
   - $||V_1 \wedge V_2||_e = ||V_1||_e \wedge^{\cdot} ||V_2||_e$,
   - $||V_1 \vee V_2||_e = ||V_1||_e \vee^{\cdot} ||V_2||_e$,
   - $||V_1 \to V_2||_e = ||V_1||_e \to^{\cdot} ||V_2||_e$,
   - $||V_1 \leftrightarrow V_2||_e = ||V_1||_e \leftrightarrow^{\cdot} ||V_2||_e$.

Poznámka 1.3. Poznamenejme, že ohodnocení e (náš externí zdroj informací, který říká, které atomické výroky jsou pravdivé a které jsou nepravdivé) někdy není popsán úplně. U některých atomických výroků se totiž předpokládá, že je známo, zda jsou pravdivé či nikoli. Naše zadání potom je např.

> Určete pravdivostní hodnotu výroku „$2 + 2 = 4$ a $2 \cdot 3 = 5$“.

místo úplného

> Určete pravdivostní hodnotu výroku „$2 + 2 = 4$ a $2 \cdot 3 = 5$“, víte-li, že „$2 + 2 = 4$“ je pravdivý a „$2 \cdot 3 = 5$“ je nepravdivý výrok.

U zkráceného zadání se tedy předpokládalo, že víme $e(\text{„}2 + 2 = 4\text{“}) = 1$ a $e(\text{„}2 \cdot 3 = 5\text{“}) = 0$.

Určení pravdivostní hodnoty výroku v poznámce 1.3 je snadné. Výsledná hodnota je téměř hned vidět ze zadání. Ne vždy tomu tak ale je, jak ukazuje následující příklad.

Příklad 1.4. Určeme pravdivostní hodnotu výroku „($2 + 2 = 4$ a číslo 10 je dělitelné číslem 6), právě když není pravda, že Čína je nejlidnatější stát světa“, víme-li, že „Čína je nejlidnatější stát světa“ je pravdivý výrok.

Tento výrok se skládá ze tří atomických výroků, totiž z výroku „$2 + 2 = 4$“, výroku „číslo 10 je dělitelné číslem 6“ a výroku „Čína je nejlidnatější stát světa.“ Označíme-li tyto výroky V_1, V_2 a V_3, víme, že $e(V_1) = 1$, $e(V_2) = 0$, $e(V_3) = 1$. Symbolická podoba zadaného výroku je $(V_1 \wedge V_2) \leftrightarrow \neg V_3$. Pravdivostní hodnota $||(V_1 \wedge V_2) \leftrightarrow \neg V_3||$ je

$$||(V_1 \wedge V_2) \leftrightarrow \neg V_3|| = ||V_1 \wedge V_2|| \leftrightarrow^{\cdot} ||\neg V_3|| =$$
$$(||V_1|| \wedge^{\cdot} ||V_2||) \leftrightarrow^{\cdot} (\neg^{\cdot} ||V_3||) = (1 \wedge^{\cdot} 0) \leftrightarrow^{\cdot} (\neg^{\cdot} 1) = 0 \leftrightarrow^{\cdot} 0 = 1,$$

tedy výrok „($2 + 2 = 4$ a číslo 10 je dělitelné číslem 6), právě když není pravda, že Čína je nejlidnatější stát světa.“ je pravdivý.

1.2.2 Kvantifikátory

Některé výrazy přirozeného jazyka obsahují proměnné. Příkladem jsou věty:

Číslo x je větší nebo rovno 3.

$2 + y = 4$.

Jestliže je x dělitelné deseti, pak je x sudé.

$x + y \geq z$.

Tyto výrazy nejsou výroky. K tomu, aby se tyto výrazy staly výroky, bychom museli určit hodnotu proměnných, které se v těch výrazech vyskytují. Dosazením číselných hodnot za proměnné vzniknou z těchto výrazů výroky. V našem případě, pokud dosadíme 1 za x, 2 za y, 103 za z, dostaneme výroky

Číslo 1 je větší nebo rovno 3.

$2 + 2 = 4$.

Jestliže je 1 dělitelné deseti, pak je 1 sudé.

$1 + 2 \geq 103$.

První a čtvrtý výrok je nepravdivý, druhý a třetí je pravdivý.

Výrazy, obsahující proměnné, ze kterých se po dosazení hodnot za proměnné stanou výroky, se nazývají výrokové formy.

Výrazy obsahující proměnné, které se po dosazení hodnot za proměnné stanou výroky, se nazývají *výrokové formy*. Výroky jsme označovali písmeny, např. V. Výrokové formy bývá zvykem označovat písmenem, za kterým jsou v závorce uvedeny všechny proměnné, které forma obsahuje. Např. „Číslo x je větší nebo rovno 3.“ bychom označili $V(x)$, „$x + y \geq z$“ bychom označili $U(x, y, z)$ apod. Výraz, který vznikne z výrazu $U(x, y, z)$ dosazením hodnoty 6 za proměnnou y, označíme $U(x, 6, z)$ apod. Poznamenejme, že $U(x, 6, z)$ není výrok, neboť stále obsahuje proměnné. Výrokem bude např. výraz $U(1, 6, 100)$, tj. výraz „$1 + 6 \geq 100$“.

Z výrokových forem můžeme tedy tvořit výroky dosazením hodnot za proměnné. Pro každou proměnnou x, která se v dané výrokové formě vyskytuje, bychom ale měli zadat její *obor hodnot*, tj. množinu D_x všech hodnot, kterých může proměnná x nabývat. Obor hodnot D_x se někdy nezadává, zvláště je-li z nějakého důvodu zřejmý. To však může vést k nedorozumění, a proto bychom obor hodnot každé proměnné měli vždy zadat. Např. u výrazu „x je větší nebo rovno 3“ není jasné, co je oborem hodnot proměnné x. Tento obor D_x musíme zadat. Musíme např. říct, že x může nabývat hodnot z množiny všech celých čísel, tj. $D_x = \{0, 1, -1, 2, -2, 3, -3, \dots\}$.

Kvantifikátory jsou jazykové výrazy, kterými z výrokových forem vznikají výroky.

Další způsob, jak tvořit z výrokových forem výroky představují *kvantifikátory*. V klasické logice rozeznáváme dva kvantifikátory, obecný a existenční.

Obecný kvantifikátor s proměnnou x je výraz

„Pro každý x platí, že ...“,

popř. jen „Pro každý x ...“. Symbolicky se obecný kvantifikátor s proměnnou x označuje výrazem $(\forall x)$. Použitím obecného kvantifikátoru na výraz „x je větší nebo rovno 1“ dostaneme výraz

V klasické logice rozeznáváme obecný a existenční kvantifikátor. Obecný se značí symbolem $\forall$, existenční symbolem $\exists$.

„Pro každé x platí, že x je větší nebo rovno 1“,

což můžeme zapsat také symbolicky jako „$(\forall x)$ (x je větší nebo rovno 1)“, popř. „$(\forall x)$ $(x \geq 1)$“.

Existenční kvantifikátor s proměnnou x je výraz

„Existuje x tak, že platí ...“,

popř. jen „Existuje x tak, že ...“. Symbolicky se existenční kvantifikátor s proměnnou x označuje výrazem $(\exists x)$. Použitím existenčního kvantifikátoru na výraz „x je větší nebo rovno 1“ dostaneme výraz

„Existuje x tak, že x je větší nebo rovno 1.“,

což můžeme zapsat také symbolicky jako „$(\exists x)$ (x je větší nebo rovno 1).“, popř. „$(\exists x)$ ($x \geq 1$)“. Tento výraz je výrokem.

Průvodce studiem

Symboly $\forall$ a $\exists$ pro obecný a existenční kvantifikátor pocházejí z němčiny. Symbol $\forall$ vznikl otočením počátečního velkého „A“ ve slově „allgemein“, což je německý výraz pro „obecný“. Symbol $\exists$ vznikl otočením počátečního velkého „E“ ve slově „existentiell“, což je německý výraz pro „existenční“.

Použitím obecného nebo existenčního kvantifikátoru s proměnnou x na výrokovou formu, jejíž jedinou proměnnou je proměnná x, získáme výrok. To je snadno vidět. Např. ve výše uvedeném příkladu vznikl výrok „Existuje x tak, že x je větší nebo rovno 1.“ použitím existenčního kvantifikátoru na výraz „x je větší nebo rovno 1.“

Obecněji platí, že výrok vznikne použitím obecného nebo existenčního kvantifikátoru s proměnnou x na výraz, ve kterém je proměnná x jedinou proměnnou, která má v daném výrazu volný výskyt. Pojem volný výskyt proměnné zde nebudeme přesně definovat. Je to pojem intuitivně jasný, a proto zůstaneme v rovině intuice. Řekneme, že výskyt proměnné x v nějakém výrazu je volný, pokud se proměnná x v tomto výskytu nenachází v dosahu platnosti nějakého kvantifikátoru s proměnnou x. Uvažujme např. výraz

(Pro každé z je z větší než 0) nebo (existuje y tak, že x je menší než y).

Dosah platnosti kvantifikátoru je ta část výrazu, na kterou se kvantifikátor vztahuje. Např. dosah platnosti kvantifikátoru „Pro každé z“ je „z větší než 0“, dosah platnosti kvantifikátoru „existuje y“ je „x je menší než y“ apod. Proto výskyt proměnné z ve výrazu „z větší než 0“ není volným výskytem (z je totiž v dosahu platnosti kvantifikátoru „Pro každé z“). Výskyt proměnné y ve výrazu „x je menší než y“ není volným výskytem (y je totiž v dosahu platnosti kvantifikátoru „existuje y“), ale výskyt proměnné x ve výrazu „x je menší než y“ je volným výskytem, protože není v dosahu platnosti kvantifikátoru s proměnnou x.

Pravdivostní hodnoty výroků s kvantifikátory se určují podle jednoduchých pravidel.

Podívejme se nyní, jak se vyhodnocují pravdivostní hodnoty výroků, které obsahují kvantifikátory. Předpokládejme, že je dána výroková forma $V(x)$, kde x je proměnná s oborem hodnot D_x. Pravdivostní hodnotu $||(\forall x)V(x)||$ výroku $(\forall x)V(x)$, tj. výroku „Pro každé x platí $V(x)$“ definujeme pravidlem

$$||(\forall x)V(x)|| = \begin{cases} 1 & \text{pokud pro každé } m \in D_x \text{ je } ||V(m)|| = 1 \\ 0 & \text{jinak.} \end{cases}$$

Slovy: Výrok $(\forall x)V(x)$ je pravdivý, pokud pro každou hodnotu m z oboru D_x je výrok $V(m)$, který vznikne dosazením m do výrokové formy $V(x)$, pravdivý. Pravdivostní hodnotu $||(\exists x)V(x)||$ výroku $(\exists x)V(x)$, tj. výroku „Existuje x tak, že platí $V(x)$“ definujeme pravidlem

$$||(\exists x)V(x)|| = \begin{cases} 1 & \text{pokud aspoň pro jedno } m \in D_x \text{ je } ||V(m)|| = 1 \\ 0 & \text{jinak.} \end{cases}$$

Slovy: Výrok $(\exists x)V(x)$ je pravdivý, pokud pro alespoň jednu hodnotu m z oboru D_x je výrok $V(m)$, který vznikne dosazením m do výrokové formy $V(x)$, pravdivý.

Příklad 1.5. (1) Je dán výrok „Pro každé x platí, že jestliže x je dělitelné 6, pak x je dělitelné 3“. Oborem hodnot proměnné x je množina všech přirozených čísel, tj. $D_x = \{1, 2, 3, 4, \dots\}$. Určete pravdivostní hodnotu daného výroku.

Výrok můžeme symbolicky zapsat jako $(\forall x)(V(x))$, kde $V(x)$ je „Jestliže x je dělitelné 6, pak x je dělitelné 3.“ Podle výše uvedeného pravidla je $||(\forall x)(V(x))|| = 1$, právě když pro každé přirozené číslo m je $||V(m)|| = 1$. Přitom výrok $V(m)$ má tvar $V_1(m) \to V_2(m)$, kde $V_1(m)$ je „m je dělitelné 6“ a $V_2(m)$ je „m je dělitelné 3“. Je zřejmé, že $||V_1(m) \to V_2(m)|| = 1$, tj. že $||V(m)|| = 1$ (podrobněji: pro m dělitelná 6 je $||V_1(m) \to V_2(m)|| = ||V_1(m)|| \to^{\cdot} ||V_2(m)|| = 1 \to^{\cdot} 1 = 1$; pro m nedělitelná 6 je $||V_1(m) \to V_2(m)|| = ||V_1(m)|| \to^{\cdot} ||V_2(m)|| = 0 \to^{\cdot} ||V_2(m)|| = 1$). Proto je

$$||(\forall x)(V_1(x) \to V_2(x))|| = 1,$$

tj. výrok „Pro každé x platí, že jestliže x je dělitelné 6, pak x je dělitelné 3“ je pravdivý.

(2) Je dán výrok „Existuje x tak, že pro každé y platí, že $x \leq y$“. Oborem hodnot proměnných x i y je množina všech přirozených čísel, tj. $D_x = D_y = \{1, 2, 3, 4, \dots\}$. Určete pravdivostní hodnotu daného výroku.

Daný výrok můžeme symbolicky zapsat jako $(\exists x)(\forall y)(V(x, y))$, kde $V(x, y)$ je „$x \leq y$“. Přitom $(\forall y)(V(x, y))$ je výroková forma, kterou můžeme označit $U(x)$. Podle uvedených pravidel je $||(\exists x)(\forall y)(V(x, y))|| = 1$, tj. $||(\exists x)U(x)|| = 1$, právě když existuje přirozené číslo m tak, že $||U(m)|| = 1$. Zvolme za m číslo 1. $U(1)$ je výrok $(\forall y)(V(1, y))$. To je výrok tvaru $(\forall y)(W(y))$, kde $W(y)$ je výroková forma $V(1, y)$, tj. $W(y)$ je $1 \leq y$. Podle uvedených pravidel je $||(\forall y)(V(1, y))|| = 1$, právě když pro každé přirozené číslo m je $||V(1, m)|| = 1$, tj. když pro každé přirozené číslo m je $1 \leq m$. To je zřejmě pravda, a proto $||(\forall y)(V(1, y))|| = 1$, a tedy i $||(\exists x)(\forall y)(V(x, y))|| = 1$. Výrok „Existuje x tak, že pro každé y platí, že $x \leq y$“ je tedy pravdivý.

Bude-li ovšem oborem hodnot proměnných x i y je množina všech celých čísel, tj. $D_x = D_y = \{\dots, -2, -1, 0, 1, 2, \dots\}$, bude daný výrok nepravdivý (zdůvodněte).

Poznámka 1.6. Kvantifikátory se někdy objevují v následující podobě:

„Pro každé liché x platí, že $x^2 - 1$ je sudé.“

„Existuje sudé x tak, že x^2 je sudé.“

Obecný tvar těchto tvrzení je:

„Pro každé x splňující $P(x)$ platí $V(x)$.“

„Existuje x splňující $P(x)$ tak, že $V(x)$.“

Tato tvrzení je třeba chápat jako zkrácené zápisy tvrzení:

„Pro každé x platí, že jestliže $P(x)$, pak $V(x)$.“

„Existuje x tak, že $P(x)$ a $V(x)$.“

Taková tvrzení už jsou obvyklými tvrzeními s kvantifikátory. První dvě tvrzení v této poznámce můžeme tedy považovat za úspornější formy následujících tvrzení:

„Pro každé x platí, že jestliže x je liché, pak $x^2 - 1$ je sudé.“

„Existuje x tak, že x je sudé a x^2 je sudé.“

1.3 Základy výrokové logiky

Náš dosavadní výklad logiky ukázal několik základních pojmů a postupů, které jsou v logice důležité. Základní pojmy, se kterými jsme pracovali, tj. pojmy výrok a později výroková forma, jsem chápali jen intuitivně. Řekli jsme, že výrokem intuitivně rozumíme tvrzení, u kterého má smysl uvažovat o jeho pravdivosti. Tato intuitivní a tedy nepřesná definice v řadě případů stačí, má ale dvě zásadní nevýhody. Za prvé, je neurčitá a ponechává prostor pro spekulace o tom, co to vlastně výrok je. Proto byly i další naše definice přísně vzato nepřesné a stavěné spíše na intuici.[2]. Za druhé, je příliš široká a připouští tak i různá komplikovaná tvrzení, která mohou přinést zásadní problémy. Ukažme si to na příkladu tzv. paradoxu lháře.

Náš pojem výroku je neurčitý a příliš široký.

Průvodce studiem

Představme si člověka C, který říká „Lžu“. Podle našeho kritéria je to výrok. Je tento výrok pravdivý nebo ne?

Jsou dvě možnosti. Buď je to pravdivý výrok, nebo je to nepravdivý výrok.

Je-li výrok „Lžu.“ pravdivý, pak je pravda to, co C říká, tj. je pravda, že C lže. To, co C říká, je tedy nepravdivé, tedy i výrok „Lžu.“ je nepravdivý. Závěrem: Je-li výrok „Lžu.“ pravdivý, pak je tento výrok nepravdivý.

Je-li výrok „Lžu.“ nepravdivý, pak není pravda to, co C říká, tj. C nelže. To, co C říká, je tedy pravdivé, tedy i výrok „Lžu.“ je pravdivý. Závěrem: Je-li výrok „Lžu.“ nepravdivý, pak je tento výrok pravdivý.

Došli jsme k tomu, že výrok „Lžu.“ je pravdivý, právě když je nepravdivý. To je spor.

Přirozený jazyk je velmi bohatý. Obsahuje i výroky, které vedou k logicky sporným závěrům.

Paradox lháře ukazuje důležitou skutečnost. Přirozený jazyk je natolik bohatý, že obsahuje i výroky, jejichž analýza vede k logicky sporným závěrům. V případě paradoxu lháře je příčinou to, že výrok „Lžu.“ se odvolává sám na sebe (hovoří o své vlastní pravdivosti), podobně jako výrok „Tento výrok je nepravdivý.“

Má ale paradox lháře nějaké řešení? Jedním z možných řešení je vzdát se ambice pracovat se všemi možnými výroky přirozeného jazyka a místo toho pracovat jen s určitými výroky, které ke sporům nevedou. Tak postupuje moderní matematická logika.

Cílem této kapitoly je ukázat si základy nejjednodušší části matematické logiky, kterou je klasická výroková logika (přesněji bychom mohli říct formální systém klasické výrokové logiky). Výroková logika je prakticky důležitá a díky své jednoduchosti nám umožní ukázat, jak se v logice pracuje. Při výkladu se budeme znovu zabývat některými pojmy logiky zavedenými intuitivně v předchozí kapitole, tentokrát však přesněji a podrobněji.

1.3.1 Syntaxe: jazyk a formule výrokové logiky

Paradox lháře ukazuje, že pokud nijak neomezíme množinu výroků, se kterými pracujeme, můžeme odvodit sporný závěr. Výroky, se kterými se ve výrokové logice pracuje, jsou omezené. Ve výrokové logice navíc nepracujeme s výroky samotnými, ale pracujeme s formami (tvary) výroků. Formy výroků se nazývají formule a jsou to přesně definované, smysluplné řetězce symbolů. To ovšem není definice, definici pojmu formule uvedeme později. Příkladem formulí jsou řetězce $(p \wedge \neg q)$, $(p \to (q \wedge r))$, $(p \wedge r) \vee q$. Formule je volně řečeno to, co je společné výrokům se stejným tvarem. Např. formule

Formule výrokové logiky popisují tvar výroků. Konkrétní výroky dostaneme nahrazením výrokových symbolů atomickými výroky.

[2] Přesto jsme se se základními principy klasické logiky poměrně dobře seznámili. Mohli jsme sice postupovat zcela přesně už od začátku, ale bylo by to na úkor srozumitelnosti.

$(p \to (q \wedge r))$ popisuje tvar mnoha konkrétních výroků, např. výroků „Jestliže prší, pak jsou silnice mokré a hrozí nebezpečí smyku.“, „Jestliže inflace roste, pak lidé méně spoří a více utrácejí.“ Tyto konkrétní výroky můžeme z formule $(p \to (q \wedge r))$ dostat dosazením atomických výroků za symboly p, q, r, např. první výrok dostaneme dosazením „Prší.“ za p, „Silnice jsou mokré.“ za q a „Hrozí nebezpečí smyku.“ za r. Proto ze symbolům p, q, r, které se ve formulích výrokové logiky vyskytují, říká výrokové symboly. Tím, že ve výrokové logice pracujeme s formulemi, a ne s konkrétními výroky, můžeme lépe odhlédnout od obsahu a soustředit se na formu. To je záměrem logiky.

Začneme definicí jazyka výrokové logiky.

Definice 1.7. *Jazyk výrokové logiky* se skládá z

Jazyk výrokové logiky obsahuje symboly, ze kterých se skládají formule výrokové logiky.

- *výrokových symbolů* p, q, r, ..., popř. s indexy, p_1, p_2; předpokládáme, že máme neomezeně mnoho výrokových symbolů;[3]
- *symbolů výrokových spojek* $\neg$ (negace), $\wedge$ (konjunkce), $\vee$(disjunkce), $\to$ (implikace), $\leftrightarrow$ (ekvivalence);
- *pomocných symbolů* (,), [,], atd. (různé druhy závorek).

Ze symbolů jazyka sestávají formule výrokové logiky, které lze jednoduše definovat induktivně následující definicí.

Definice 1.8. Nechť je dán jazyk výrokové logiky. *Formule* daného jazyka výrokové logiky je definována následovně:

- každý výrokový symbol je formule (tzv. atomická formule);
- jsou-li φ a ψ formule, jsou i výrazy $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \to \psi)$, $(\varphi \leftrightarrow \psi)$ formule (tzv. složené formule).

Příklad 1.9. Formulemi jsou tedy jisté, smysluplně vytvořené konečné posloupnosti symbolů jazyka výrokové logiky.

Formulemi jsou např. posloupnosti p, q_1, $\neg p$, $(p \to q)$, $((p \wedge r) \vee p)$, $(\neg p \to (q \wedge \neg r))$. Posloupnost $(\neg p \to (q \wedge \neg r))$ je formule, což lze zdůvodnit následovně: r je formule (atomická), tedy i $\neg r$ je formule, což spolu s tím, že q je formule, dává, že $(q \wedge \neg r)$ je formule; dále jsou p, a tedy i $\neg p$ formule, a tedy konečně i $(\neg p \to (q \wedge \neg r))$ je formule.

Formulemi nejsou posloupnosti $\wedge p$, $p \wedge \vee p$, $pp \to (p\wedge))$, atd.

Všimněme si, že správně bychom měli říkat „formule daného jazyka výrokové logiky“. My však v případě, že jazyk je zřejmý z kontextu, popř. není důležitý, budeme říkat pouze „formule výrokové logiky“ nebo jen „formule“.

Poznámka 1.10 (konvence o vynechávání závorek)**.** Jak zná čtenář z aritmetiky, je pro zjednodušení zápisu a čtení užitečné vynechávat závorky tam, kde neutrpí jednoznačnost čtení. Podobně budeme postupovat i my. Např. místo $(p \to q)$ budeme psát jen $p \to q$. Dále se dohodneme na prioritách symbolů spojek: od největší po nejmenší je to $\neg$, $\wedge$, $\vee$, $\to$, $\leftrightarrow$. To nám umožní vynechávat závorky. Tak např. místo $(p \wedge (q \wedge r))$ můžeme psát jen $p \wedge (q \wedge r)$, místo $(p \to ((p \wedge q) \vee r))$ jen $p \to p \wedge q \vee r$ apod.

[3] Přesněji: předpokládáme, že máme nekonečně spočetně mnoho výrokových symbolů, což volněji řečeno znamená, je můžeme uspořádat do nekonečné posloupnosti.

1.3.2 Sémantika: pravdivost, tautologie, vyplývání

Zatím jsme se věnovali jen tzv. syntaxi výrokové logiky: víme, co je to jazyk výrokové logiky, a co jsou to formule. Zatím však nevíme, co to je pravdivá formule apod. Formule jsou jisté posloupnosti symbolů jazyka, samy o sobě však nemají žádný význam. V tomto smyslu můžeme říct, že formule je syntaktický objekt. Přiřazení významu syntaktickým objektům je záležitostí sémantiky. Právě sémantice výrokové logiky se v dalším budeme věnovat.

Definice 1.11. *(Pravdivostní) ohodnocení* je libovolné zobrazení e výrokových symbolů daného jazyka výrokové logiky do množiny $\{0, 1\}$, tj. ohodnocení e přiřazuje každému výrokovému symbolu p hodnotu 0 nebo 1.

Poznámka 1.12. (1) 0 a 1 reprezentují pravdivostní hodnoty nepravda a pravda. Hodnotu přiřazenou ohodnocením e symbolu p označujeme $e(p)$. Je tedy $e(p) = 0$ nebo $e(p) = 1$.

(2) Význam ohodnocení e můžeme chápat následovně. Jak jsme si řekli, výrokové symboly jsou pro nás symboly, které označují atomické výroky. Je-li $e(p) = 1$, znamená to pro nás, že atomický výrok označený symbolem p je pravdivý. Je-li $e(p) = 0$, znamená to, že atomický výrok označený symbolem p je nepravdivý.

Je-li dáno ohodnocení e, můžeme určit, jaká je pravdivostní hodnota dané formule. Pravdivostní hodnota každé formule je pravdivostním ohodnocením jednoznačně určena a je definována následovně.

Definice 1.13. Nechť je dáno ohodnocení e. *Pravdivostní hodnota formule* φ při ohodnocení e, označujeme ji $\|\varphi\|_e$, je definována následovně:

- Je-li φ výrokovým symbolem p, pak

  $$\|p\|_e = e(p).$$

- Je-li φ složená formule, tj. jednoho z tvarů $\neg\psi$, $\psi \wedge \theta$, $\psi \vee \theta$, $\psi \to \theta$, $\psi \leftrightarrow \theta$, pak

  $$\begin{aligned}
  \|\neg\psi\|_e &= \neg^{\cdot}\|\psi\|_e,\\
  \|\psi \wedge \theta\|_e &= \|\psi\|_e \wedge^{\cdot} \|\theta\|_e,\\
  \|\psi \vee \theta\|_e &= \|\psi\|_e \vee^{\cdot} \|\theta\|_e,\\
  \|\psi \to \theta\|_e &= \|\psi\|_e \to^{\cdot} \|\theta\|_e,\\
  \|\psi \leftrightarrow \theta\|_e &= \|\psi\|_e \leftrightarrow^{\cdot} \|\theta\|_e,
  \end{aligned}$$

  kde $\neg^{\cdot}$, $\wedge^{\cdot}$, $\vee^{\cdot}$, $\to^{\cdot}$, $\leftrightarrow^{\cdot}$ jsou pravdivostní funkce logických spojek z tabulky 1.

Poznámka 1.14. (1) Je-li $\|\varphi\|_e = 1$, resp. $\|\varphi\|_e = 0$, říkáme, že formule φ je při ohodnocení e pravdivá, resp. nepravdivá. Uvědomme si, že bez určení ohodnocení e nemá smysl říci „formule φ je pravdivá“ nebo „nepravdivá“. To přesně odpovídá situaci z kapitoly 1.2.1, kde jsme při určování pravdivostní hodnoty výroku museli znát pravdivostní hodnoty atomických výroků. Roli atomických výroků teď mají výrokové symboly.

(2) Část definice pravdivostní hodnoty formule, ve které se zavádí pravdivostní hodnota složené formule, můžeme alternativně popsat následovně:

$$\|\neg\psi\|_e = \begin{cases} 1 & \text{pokud } \|\psi\|_e = 0,\\ 0 & \text{jinak,}\end{cases}$$

$$\|\psi \wedge \theta\|_e = \begin{cases} 1 & \text{pokud } \|\psi\|_e = 1 \text{ a } \|\theta\|_e = 1,\\ 0 & \text{jinak,}\end{cases}$$

$$\|\psi \vee \theta\|_e = \begin{cases} 1 & \text{pokud } \|\psi\|_e = 1 \text{ nebo } \|\theta\|_e = 1, \\ 0 & \text{jinak,} \end{cases}$$

$$\|\psi \to \theta\|_e = \begin{cases} 1 & \text{pokud } \|\psi\|_e = 0 \text{ nebo } \|\theta\|_e = 1, \\ 0 & \text{jinak,} \end{cases}$$

$$\|\psi \leftrightarrow \theta\|_e = \begin{cases} 1 & \text{pokud } \|\psi\|_e = \|\theta\|_e, \\ 0 & \text{jinak.} \end{cases}$$

Snadno se vidí (ověřte si), že taková definice skutečně vede ke stejným pravdivostním hodnotám formulí.

Následující definice zavádí další užitečné pojmy.

Definice 1.15. Formule se nazývá

- *tautologie*, je-li při každém ohodnocení pravdivá,
- *kontradikce*, je-li při každém ohodnocení nepravdivá,
- *splnitelná*, je-li při aspoň jednom ohodnocení pravdivá.

Formule φ *sémanticky vyplývá* z množiny T formulí, jestliže φ je pravdivá při každém ohodnocení, při kterém jsou pravdivé všechny formule z T; to označujeme

$$T \models \varphi.$$

Poznámka 1.16. Splnitelné formule jsou tedy právě ty, které nejsou kontradikcemi. Že je formule φ tautologie, se někdy zapisuje $\models \varphi$.

Příklad 1.17. (1) Formule $p \vee \neg p$ i $p \to (p \vee q)$ jsou tautologie.

(2) Formule $p \wedge \neg p$ i $p \leftrightarrow \neg p$ jsou kontradikce.

(3) Formule $p \to \neg p$ je splnitelná, ale není to ani tautologie, ani kontradikce.

(4) Formule $p \to q$ sémanticky vyplývá z množiny formulí $T = \{p \to r, \neg q \to \neg r\}$.

Uvedené skutečnosti se snadno ověří tabulkovou metodou, kterou uvedeme níže. Je možné je také ověřit úvahou. Vezměme například formuli $p \to (p \vee q)$. Proč je tautologií? Kdyby existovalo ohodnocení e, při kterém by tato formule byla nepravdivá, pak z vlastností implikace plyne, že při tomto ohodnocení je p pravdivá a $p \vee q$ nepravdivá, což není možné, protože je-li p pravdivá, je i $p \vee q$ pravdivá.

Příklad 1.18. Dokažme, že pro libovolnou množinu T formulí a formule φ, ψ platí

$$T, \varphi \models \psi \text{ právě když } T \models \varphi \to \psi.$$

To je intuitivně poměrně jasné tvrzení. Přitom T, φ znamená $T \cup \{\varphi\}$, tj. T, φ označuje množinu T rozšířenou o formuli φ. Důkaz je velmi snadný, stačí si rozmyslet, co máme dokázat. Dokážeme, že z $T, \varphi \models \psi$ plyne $T \models \varphi \to \psi$ a že z $T \models \varphi \to \psi$ plyne $T, \varphi \models \psi$.

Předpokládejme nejdříve $T, \varphi \models \psi$ a dokažme $T \models \varphi \to \psi$. Máme dokázat, že je-li e ohodnocení, při kterém jsou pravdivé všechny formule z T, je při e pravdivá i formule $\varphi \to \psi$. Kdyby ale při e nebyla pravdivá formule $\varphi \to \psi$, musela by být při e φ pravdivá a ψ nepravdivá (z definice pravdivostní funkce spojky implikace). Je-li ale při e pravdivá φ i všechny formule z T, pak je dle předpokladu $T, \varphi \models \psi$ pravdivá i ψ, což je spor s tím, že ψ je nepravdivá.

Předpokládejme nyní $T \models \varphi \to \psi$ a dokažme $T, \varphi \models \psi$. Máme dokázat, že je-li e ohodnocení, při kterém je pravdivá φ i všechny formule z T, je při něm pravdivá i ψ. Dle předpokladu je ovšem při e pravdivá i $\varphi \to \psi$ a protože je při e pravdivá i φ, je při e pravdivá i ψ, což jsme měli dokázat.

p	q	r	$(p \to q) \wedge (p \to r)$
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

Tabulka 2: Tabulka pro formuli $(p \to q) \wedge (p \to r)$.

1.3.3 Tabulková metoda

Tabulková metoda představuje jednoduchý způsob, jak zjistit a přehledně zapsat pravdivostní hodnoty dané formule při všech možných ohodnoceních. Jsou-li $p_1, \ldots, p_n$ všechny výrokové symboly, které se vyskytují ve formuli φ, budeme místo φ psát také $\varphi(p_1, \ldots, p_n)$.

Podstata tabulkové metody je následující. Pro zadanou formuli $\varphi(p_1, \ldots, p_n)$ vytvoříme tzv. pravdivostní tabulku. Tabulka 2 je tabulkou pro formuli $(p \to q) \wedge (p \to r)$. Řádky tabulky odpovídají ohodnocením výrokových symbolů. Sloupce tabulky odpovídají symbolům $p_1, \ldots, p_n$ a formuli φ. Tabulka má tedy $n + 1$ sloupců, každý je v záhlaví označen příslušným výrazem, tj. postupně p_1, …, p_n, φ. Obsah řádku, který odpovídá ohodnocení e, je následující. Na místě odpovídajícímu sloupci tabulky s označením p_i je hodnota $e(p_i)$, tj. hodnota symbolu p_i při ohodnocení e. Na místě odpovídajícímu sloupci tabulky s označením φ je hodnota $||\varphi||_e$, tj. pravdivostní hodnota formule φ při ohodnocení e. Tabulka má tolik řádků, kolik je možností, jak ohodnotit symboly $p_1, \ldots, p_n$ hodnotami 0 a 1. Protože každému ze symbolů $p_1, \ldots, p_n$ můžeme přiřadit dvě možné hodnoty, máme celkem 2^n možností (2 možnosti pro p_1 krát 2 možnosti pro p_2 krát $\cdots$ krát 2 možnosti pro p_n, tj. $2 \times \cdots \times 2 = 2^n$ možností). Tabulka má tedy 2^n řádků.

To je potvrzeno v tabulce 2. Zde máme tři výrokové symboly p, q, r, tabulka má tedy $2^3 = 8$ řádků. V každém řádku uvedeme příslušné ohodnocení symbolů $p_1, \ldots, p_n$ a hodnotu formule φ při tomto ohodnocení. Např. ve třetím řádku tabulky 2 jsou uvedeny postupně hodnoty 0, 1, 0, 1, protože tento řádek odpovídá ohodnocení, které symbolům p, q a r přiřazuje postupně hodnoty 0, 1 a 0 a protože při tomto ohodnocení má formule $(p \to q) \wedge (p \to r)$ hodnotu 1. Je zvykem všechna možná ohodnocení symbolů uvádět přirozeně uspořádaná, tj. v prvních n sloupcích bude v prvním řádku $0 \ldots 0$ (n nul), ve druhém řádku pak $0 \ldots 01$ ($n - 1$ nul a jedna jednička), atd. až v posledním řádku bude $1 \ldots 1$ (n jedniček). Tak je to také v tabulce 2.

Tabulka dané formule popisuje tzv. *pravdivostní funkci* této formule. Například výše uvedená tabulka 2 popisuje pravdivostní funkci formule $(p \to q) \wedge (p \to r)$.

Poznámka 1.19. Řekli jsme, že v tabulce chceme zachytit hodnoty φ při všech možných ohodnoceních. Ohodnocení e je ale dáno tím, jaké hodnoty přiřazuje všem výrokovým symbolům, tedy nejen symbolům $p_1, \ldots, p_n$. Při popisu tabulkové metody jsme ale uvažovali jen hodnoty přiřazené symbolům $p_1, \ldots, p_n$. Dopustili jsme se tím jisté nepřesnosti, v zásadě nepodstatné. Přesně řečeno se má situace takto. Pravdivostní hodnota $||\varphi||_e$ závisí jen na hodnotách $e(p_1), \ldots, e(p_n)$, tj. na ohodnocení výrokových symbolů $p_1, \ldots, p_n$, a nezávisí na $e(p)$ pro $p \neq p_1, \ldots, p \neq p_n$, tj. ohodnocení výrokových

$\varphi \vee \neg\varphi$ (zákon vyloučeného třetího)

$\varphi \to \varphi$

Tabulka 3: Vybrané tautologie výrokové logiky. DOPLNIT

p	q	$\neg\neg p$	$(\neg q \to \neg p)$	q
0	0	0	1	0
0	1	0	1	1
1	0	1	0	0
1	1	1	1	1

Tabulka 4: Tabulka pro formule $\neg\neg p$, $(\neg q \to \neg p)$ a q.

symbolů jiných než $p_1, \ldots, p_n$. To je jasné proto, že $\varphi(p_1, \ldots, p_n)$ jiné výrokové symboly než $p_1, \ldots, p_n$ neobsahuje. Chceme-li v tabulce zachytit hodnoty φ při všech možných ohodnoceních, stačí tedy zachytit hodnoty φ pro všechna možná ohodnocení symbolů $p_1, \ldots, p_n$. Totiž, jak jsme řekli, hodnota $||\varphi||_e$ závísí jen na hodnotách $e(p_1), \ldots, e(p_n)$. Je-li e' jiné ohodnocení, které se s e shoduje v hodnotách přiřazených $p_1, \ldots, p_n$, tj. $e(p_1) = e'(p_1)$, $\ldots$, $e(p_n) = e'(p_n)$, pak je zřejmě $||\varphi||_e = ||\varphi||_{e'}$, tj. hodnoty formule φ při ohodnoceních e a e' jsou stejné. V daném řádku uvedená hodnota formule $||\varphi||_e$ je tedy hodnotou formule φ při každém ohodnocení e, které symbolům $p_1, \ldots, p_n$ přiřazuje hodnoty uvedené v prvních n sloupcích tohoto řádku. Například třetí řádek v tabulce 2 udává hodnotu formule $(p \to q) \wedge (p \to r)$ při ohodnocení e, kde $e(p) = 0$, $e(q) = 1$, $e(r) = 0$, $e(p_1) = 0$, $e(p_2) = 0$, $\ldots$ (p_1 a p_2 jsou výrokové symboly jazyka, které se nevyskytují ve formuli $(p \to q) \wedge (p \to r)$), ale i při ohodnocení e', $e'(p) = 0$, $e'(q) = 1$, $e'(r) = 0$, $e'(p_1) = 0$, $e'(p_2) = 1$, $\ldots$, a i při každém dalším ohodnocení e'', pro které je $e''(p) = 0$, $e''(q) = 1$, $e''(r) = 0$. Řádky tedy vlastně neodpovídají jednotlivým ohodnocením, ale celým třídám (skupinám) ohodnocení.

Tabulka vytvořená pro formuli φ umožňuje zjistit některé výše popsané vlastnosti formule φ: φ je tautologie, právě když ve sloupci odpovídajícím formuli φ jsou samé 1; φ je kontradikce, právě když ve sloupci odpovídajícím formuli φ jsou samé 0; φ je splnitelná, právě když ve sloupci odpovídajícím formuli φ je aspoň jedna 1. Vybrané tautologie ukazuje tabulka 3 (ověřte, že jde o tautologie).

Tabulkovou metodu můžeme jednoduše rozšířit pro více formulí. Předpokládejme, že máme formule $\varphi_1, \ldots, \varphi_m$, a že všechny proměnné vyskytující se v alespoň jedné z těchto formulí jsou právě $p_1, \ldots, p_n$. Odpovídající tabulka bude mít 2^n řádků a $n + m$ sloupců označených postupně $p_1, \ldots, p_n$ a $\varphi_1, \ldots, \varphi_m$. Do řádků píšeme ohodnocení e a hodnoty formulí $\varphi_1, \ldots, \varphi_m$ v těchto ohodnoceních: hodnoty $e(p_i)$ symbolů p_i v ohodnocení e píšeme do sloupců označených p_i. Příslušné hodnoty $||\varphi_j||_e$ formulí φ_j při ohodnocení e píšeme do sloupců označených φ_j. Příklad vidíme v tabulce 4, ve které je $n = 2$ (dva výrokové symboly: p a q) a $m = 3$ (tři formule: $\neg\neg p$, $(\neg q \to \neg p)$ a q).

Rozšířenou tabulkovou metodu můžeme použít ke zjištění, zda formule φ sémanticky plyne z formulí $\varphi_1, \ldots, \varphi_m$. Stačí vytvořit tabulku pro formule $\varphi_1, \ldots, \varphi_m$ a φ. Podle definice pak φ sémanticky plyne z $\varphi_1, \ldots, \varphi_m$, právě když v každém řádku, ve kterém mají formule $\varphi_1, \ldots, \varphi_m$ hodnotu 1, má také formule φ hodnotu 1. Z tabulky 4 např. vidíme, že formule q vyplývá z formulí $\neg\neg p$ a $(\neg q \to \neg p)$. Totiž, jediný řádek, ve kterém mají obě formule $\neg\neg p$ a $(\neg q \to \neg p)$ hodnotu 1, je čtvrtý řádek a v tomto řádku má formule q také hodnotu 1. Naopak, formule $\neg\neg p$ nevyplývá z formulí $(\neg q \to \neg p)$ a q, protože ve druhém řádku mají formule $(\neg q \to \neg p)$ a q hodnotu 1, ale formule $\neg\neg p$ tam má hodnotu 0.

φ	ψ	$\varphi \vee \psi$	$\neg(\neg\varphi \wedge \neg\psi)$	$\varphi \to \psi$	$\neg\varphi \vee \psi$
0	0	0	0	1	1
0	1	1	1	1	1
1	0	1	1	0	0
1	1	1	1	1	1

Tabulka 5: Tabulka k příkladu 1.21.

$\varphi \wedge \psi \equiv \psi \wedge \varphi$	$\psi \vee \varphi \equiv \varphi \vee \psi$	(komutativita)
$\varphi \wedge (\psi \wedge \theta) \equiv (\varphi \wedge \psi) \wedge \theta$	$\varphi \vee (\psi \vee \theta) \equiv (\varphi \vee \psi) \vee \theta$	(asociativita)
$\varphi \wedge (\psi \vee \theta) \equiv (\varphi \wedge \psi) \vee (\varphi \wedge \theta)$	$\varphi \vee (\psi \wedge \theta) \equiv (\varphi \vee \psi) \wedge (\varphi \vee \theta)$	(distributivita)
$\varphi \equiv \neg\neg\varphi$		(zákon dvojí negace)
$\neg(\varphi \wedge \psi) \equiv \neg\varphi \vee \neg\psi$	$\neg(\varphi \vee \psi) \equiv \neg\varphi \wedge \neg\psi$	(De Morganovy zákony)
$\varphi \to \psi \equiv \neg\varphi \vee \psi$	$\varphi \to \psi \equiv \neg(\varphi \wedge \neg\psi)$	
$\varphi \to \psi \equiv \neg\psi \to \neg\varphi$		(obměněná implikace)

Tabulka 6: Sémanticky ekvivalentní formule. DOPLNIT

Průvodce studiem

Tabulková metoda slouží k vypsání (tabelaci) hodnot zadaných formulí $\varphi_1, \ldots, \varphi_m$ v tabulce. Tabulka má 2^n řádků a $n+m$ sloupců, kde n je počet všech výrokových symbolů, které se vyskytují ve formulích $\varphi_1, \ldots, \varphi_m$. Do řádků píšeme všechna možná ohodnocení těchto symbolů a hodnoty formulí $\varphi_1, \ldots, \varphi_m$.

Pomocí tabulkové metody můžeme zjistit různé vlastnosti formulí a vztahy mezi formulemi, např. zda zadaná formule je tautologie, kontradikce, splnitelná, a také, zda zadaná formule sémanticky vyplývá z jiných zadaných formulí.

Tabulková metoda umožňuje snadno nahlédnout, že dvě formule jsou sémanticky ekvivalentní.

Definice 1.20. Formule φ a ψ se nazývají *(sémanticky) ekvivalentní*, což značíme

$$\varphi \equiv \psi,$$

právě když pro každé ohodnocení e je $\|\varphi\|_e = \|\psi\|_e$.

Snadno se vidí, že φ a ψ jsou ekvivalentní, právě když $\varphi \models \psi$ a $\psi \models \varphi$, tedy když ψ sémanticky plyne z φ a naopak.

Příklad 1.21. Ukažme, že (a) $\varphi \vee \psi \equiv \neg(\neg\varphi \wedge \neg\psi)$ a že (b) $\varphi \to \psi \equiv \neg\varphi \vee \psi$. Vztahy (a) i (b) jsou patrné z tabulky 5: (a) plyne z toho, že třetí a čtvrtý sloupec mají stejné hodnoty, (b) pak z toho, že stejné hodnoty mají i pátý a šestý sloupec.

Důležité dvojice navzájem ekvivalentních formulí ukazuje tabulka 6.

Pro formuli $\varphi \to \psi$ se často uvažují následující formule:

$\neg\psi \to \neg\varphi$ (obměněná implikace)

$\psi \to \varphi$ (obrácená implikace)

Jak ukazuje tabulka 7, $\neg\psi \to \neg\varphi$ je s $\varphi \to \psi$ ekvivalentní, ale $\psi \to \varphi$ ne. Ekvivalentnost implikace a k ní obměněné implikace se často využívá v důkazech. Máme-li dokázat tvrzení ve tvaru $\varphi \to \psi$, můžeme místo toho dokázat $\neg\psi \to \neg\varphi$, což může být snazší.

φ	ψ	$\varphi \to \psi$	$\neg\psi \to \neg\varphi$	$\psi \to \varphi$
0	0	1	1	1
0	1	1	1	0
1	0	0	0	1
1	1	1	1	1

Tabulka 7: Implikace, obměněná implikace a obrácená implikace

x_1	x_2	f
0	0	0
0	1	1
1	0	1
1	1	0

x_1	x_2	g
0	0	0
0	1	1
1	0	1
1	1	1

Tabulka 8: Tabulka booleovských funkcí dvou proměnných.

1.3.4 Booleovské funkce, normální formy, funkční úplnost

Definice 1.22. *Booleovská funkce* s n argumenty (někdy n-ární booleovská funkce) je libovolné zobrazení $f : \{0,1\}^n \to \{0,1\}$.

Booleovská funkce s n argumenty je tedy zobrazení, které každé uspořádaným n-ticím hodnot 0 a 1 přiřazuje hodnoty 0 nebo 1. Každou booleovskou funkci f s n argumenty lze zapsat tabulkou podobným způsobem jako u tabulkové metody. Argumenty funkce f označme proměnnými $x_1, \ldots, x_n$, odpovídající hodnotu funkce f pak $f(x_1, \ldots, x_n)$. Tabulka funkce f má 2^n řádků a $n+1$ sloupců. Sloupce označíme symboly $x_1, \ldots, x_n$ a f. Do každého řádku napíšeme hodnoty proměnných $x_1, \ldots, x_n$ a do posledního sloupce pak hodnotu $f(x_1, \ldots, x_n)$. Příkladem jsou booleovské funkce f a g dvou proměnných, které jsou zapsané v tabulce 8. Funkce f např. přiřazuje dvojici hodnot 0 a 1 hodnotu 1, tj. $f(0,1) = 1$, dále $f(0,0) = 0$, $f(1,0) = 1$, $f(1,1) = 1$. Pro funkci g je $g(0,0) = 0$, $g(0,1) = 1$, $g(1,0) = 1$, $g(1,1) = 1$.

Všimněme si, že výše uvedená funkce g je shodná s pravdivostní funkcí $\vee$˙ spojky disjunkce. Pravdivostní funkce každé ze spojek, se kterými jsme se setkali, jsou booleovské funkce. Pravdivostní funkce spojek $\wedge$, $\vee$, $\to$ a $\leftrightarrow$, jsou funkce 2 argumentů, pravdivostní funkce spojky $\neg$ je booleovská funkce jednoho argumentu. Funkce f z tabulky 8 ukazuje, že existují i jiné booleovské funkce než pravdivostní funkce $\wedge$˙, $\vee$˙, $\to$˙ a $\leftrightarrow$˙ logických spojek $\wedge$, $\vee$, $\to$ a $\leftrightarrow$.

Každou booleovskou funkci dvou proměnných můžeme totiž považovat za pravdivostní funkci logické spojky se dvěma argumenty. Z tohoto pohledu jsou tedy spojky $\wedge$, $\vee$, $\to$ a $\leftrightarrow$ jen některé z logických spojek se dvěma argumety. Skutečně, např. pravdivostní funkce spojky „buď ..., nebo ...“ je právě funkce f z tabulky 8 (výrok „Buď V_1, nebo V_2“ je pravdivý, právě když je pravdivý právě jeden z výroků V_1 a V_2).

Kolik booleovských funkcí s n argumenty existuje, tj. kolik existuje různých logických spojek s n argumenty?

Věta 1.23. *Existuje právě $2^{(2^n)}$ booleovských funkcí s n argumenty.*

Důkaz. Funkcí je tolik, kolika způsoby lze vyplnit příslušnou tabulku. Pro funkce s n argumenty má tabulka zřejmě 2^n řádků. V každém řádku je jedno volné místo pro

x_1	f_1
0	0
1	0

x_1	f_2
0	0
1	1

x_1	f_3
0	1
1	0

x_1	f_4
0	1
1	1

Tabulka 9: Všechny booleovské funkce jedné proměnné.

x_1	x_2	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Tabulka 10: Všechny booleovské funkce dvou proměnných.

hodnotu funkce, a tu můžeme vyplnit libovolným způsobem (napsat tam 0 nebo 1). Protože volných míst je 2^n, lze je hodnotami 0 nebo 1 vyplnit $2^{(2^n)}$ způsoby. □

Příklad 1.24. Tabulka 9 ukazuje všechny unární (1-ární) booleovské funkce. Dle věty 1.23 jsou skutečně čtyři, neboť $2^{(2^n)} = 2^{(2^1)} = 2^2 = 4$. Přitom f_3 je pravdivostní funkce spojky negace.

Příklad 1.25. Tabulka 10 ukazuje všechny binární (2-ární) booleovské funkce. Dle věty 1.23 jich je opravdu 16, neboť $2^{(2^n)} = 2^{(2^2)} = 2^4 = 16$. Všimněme si, že f_2, f_8, f_{10} a f_{14} jsou po řadě pravdivostní funkce spojek $\wedge$, $\vee$, $\leftrightarrow$ a $\rightarrow$. Funkce f_7 je pravdivostní funkce výše zmíněné spojky „buď ..., anebo ...“

Je zřejmé, že každá formule φ obsahující výrokové symboly $p_1, \ldots, p_n$ indukuje booleovskou funkci φ s n argumenty. Tuto funkci označme f_φ. Je to právě funkce, jejíž tabulku získáme vytvořením tabulky pro formuli φ. Například pro formuli $(p \rightarrow q) \wedge (q \rightarrow r)$ platí pro příslušnou funkci $f_{(p\rightarrow q)\wedge(q\rightarrow r)}$, že $f_{(p\rightarrow q)\wedge(q\rightarrow r)}(0,0,0) = 1$, ..., $f_{(p\rightarrow q)\wedge(q\rightarrow r)}(1,0,1) = 0$, ..., $f_{(p\rightarrow q)\wedge(q\rightarrow r)}(1,1,1) = 1$. Tato funkce znázorněna výše uvedenou tabulkou 2.

Normální formy

Zajímavé ale je, že platí také opačné tvrzení. Ke každé booleovské funkci f s n argumenty existuje formule φ_f, která indukuje právě funkci f. Platí dokonce, že formuli φ_f můžeme vzít takovou, že obsahuje pouze spojky $\neg$, $\wedge$ a $\vee$. Postup, jak takovou formuli získat, si nyní podrobně popíšeme. Zaveďme nejprve následující pojmy.

Definice 1.26. Nechť V je množina výrokových symbolů. Pak

- *literál* nad V je libovolný výrokový symbol z V nebo jeho negace;
- *úplná elementární konjunkce* nad V je libovolná konjunkce literálů, ve které se každý výrokový symbol z V vyskytuje právě v jednom literálu;
- *úplná elementární disjunkce* nad V je libovolná disjunkce literálů, ve které se každý výrokový symbol z V vyskytuje právě v jednom literálu;
- *úplná konjunktivní normální forma* nad V je konjunkce úplných elementárních disjunkcí nad V;

- *úplná disjunktivní normální forma* nad V je disjunkce úplných elementárních konjunkcí nad V.

Příklad 1.27. Uvažujme $V = \{p, q, r\}$. Pak

- literály jsou: $p, q, r, \neg p, \neg q, \neg r$; literál není např. $\neg\neg p$;
- ÚEK jsou: $p \wedge q \wedge r$, $\neg p \wedge q \wedge \neg r$; $p \wedge r$ není ÚEK;
- ÚED je např. $p \vee \neg q \vee r$;
- ÚKNF je např. $(p \vee q \vee \neg r) \wedge (p \vee \neg q \vee \neg r)$;
- ÚDNF je např. $(p \wedge q \wedge r) \vee (p \wedge q \wedge \neg r)$.

Předpokládejme nyní, že je tabulkou dána booleovská funkce $f(x_1, \ldots, x_n) : \{0,1\}^n \to \{0,1\}$. Uvažujme následující postup pro vytvoření úplné disjunktivní normální formy φ nad $V = \{p_1, \ldots, p_n\}$:

1. Pro každý řádek tabulky odpovídající ohodnocení e, při kterém má funkce f hodnotu 1 (tedy $f(e(p_1), \ldots, e(p_n)) = 1$) sestrojíme ÚEK

$$l_1 \wedge l_2 \wedge \cdots \wedge l_n,$$

kde

$$l_i = \begin{cases} p_i & \text{pro } e(p_i) = 1 \\ \neg p_i & \text{pro } e(p_i) = 0 \end{cases}$$

2. φ je disjunkcí ÚEK postupně sestrojených v bodu 1.

Příklad 1.28. Sestrojíme ÚDNF k booleovská funkci f dané následující tabulkou:

p	q	r	f
0	0	0	1
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

1. Projdeme řádky s 1 ve sloupci „f“ a vytvoříme příslušné ÚEK:

   ř. 1. : ÚEK je $\neg p \wedge \neg q \wedge \neg r$;

   ř. 2. : ÚEK je $\neg p \wedge \neg q \wedge r$;

   ř. 8. : ÚEK je $p \wedge q \wedge r$;

2. výsledná ÚDNF je disjunkcí ÚEK z kroku 1., tedy je to formule

$$(\neg p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge r) \vee (p \wedge q \wedge r)$$

Následující tvrzení říká, že výše popsaná konstrukce je správná.

Věta 1.29. *Nechť f je booleovská funkce, která nabývá aspoň pro jednu kombinaci argumentů hodnotu 1 (tedy nepředstavuje kontradikci). Pak ÚDNF φ sestrojená výše uvedeným postupem splňuje $f = f_\varphi$ (neboli φ indukuje funkci f; tabulky f a φ jsou stejné).*

Důkaz. Máme ukázat, že pro libovolné ohodnocení e platí, že $\|\varphi\|_e = 1$, právě když v tabulce funkce f je v řádku odpovídajícímu e hodnota 1. Uvědomme si, že φ má tvar $k_1 \vee \cdots \vee k_m$, kde každá ÚEK k_j má tvar $l_1 \wedge \cdots \wedge l_n$. Dokážeme oba požadované směry.

„$\Rightarrow$“: Nechť $\|\varphi\|_e = 1$. Z vlastností $\vee$ plyne, že pro nějakou $k_j = l_1 \wedge \cdots \wedge l_n$ musí být $\|k_j\|_e = 1$. Pak musí pro $l_i = p_i$ být $e(p_i) = 1$ a pro $l_i = \neg p_i$ pak $e(p_i) = 0$. Takové e je ale právě ohodnocení odpovídající řádku, díky kterému se k_j naší konstrukcí dostala do φ, tedy v tomto řádku musí být hodnota 1.

„$\Leftarrow$“: Je-li v nějakém řádku tabulky funkce f hodnota 1, uvažujme odpovídající ohodnocení e a odpovídající k_j. Z konstrukce plyne, že $\|k_j\|_e = 1$, a tedy $\|\varphi\|_e = \|k_1 \vee \cdots \vee k_m\|_e = 1$. □

Uvědomme si, že pokud má f ve všech řádcích 0, uvedený postup vrátí „prázdnou formuli“. Platí tedy:

Věta 1.30. *Ke každé booleovské funkci f, která nepředstavuje kontradikci, existuje ÚDNF φ tak, že $f = f_\varphi$.*

Důkaz. Požadovanou φ je například ÚDNF sestrojená k tabulce funkce f. □

Úlohu lze obměnit následovně. Místo funkce f je dána formule ψ a cílem je najít ÚDNF φ tak, aby φ a ψ byly sémanticky ekvivalentní (tedy měly stejné tabulky). Obměněnou úlohu lze vyřešit následovně. Nejdříve sestrojíme tabulku formule ψ, tedy tabulku booleovské funkce f_ψ indukované zadanou formulí ψ. Poté k této tabulce sestrojíme ÚDNF postupem uvedeným výše.

Příklad 1.31. Sestrojme ÚDNF formule $(p \to q) \wedge (p \to r)$. Vytvoříme tabulku dané formule a rovnou do ní přidáme sloupec, kam zapíšeme příslušné ÚEK:

p	q	r	$(p \to q) \wedge (p \to r)$	ÚEK
0	0	0	1	$\neg p \wedge \neg q \wedge \neg r$
0	0	1	1	$\neg p \wedge \neg q \wedge r$
0	1	0	1	$\neg p \wedge q \wedge \neg r$
0	1	1	1	$\neg p \wedge q \wedge r$
1	0	0	0	
1	0	1	0	
1	1	0	0	
1	1	1	1	$p \wedge q \wedge r$

Výsledná ÚDNF tedy je $(\neg p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge r) \vee (\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge q \wedge r) \vee (p \wedge q \wedge r)$.

Z definice je zřejmé, že pojem ÚKNF je duální k pojmu ÚDNF (z definice ÚKNF dostaneme definici ÚDNF záměnou konjunkcí za disjunkce). Je tedy přirozené, že k dané booleovské funkci f, popř. k dané formuli ψ, lze sestrojit ÚKNF způsobem, který je duální konstrukci ÚDNF. Postup je následující:

1. Pro každý řádek tabulky odpovídající ohodnocení e, při kterém má funkce f hodnotu 0 (tedy $f(e(p_1), \ldots, e(p_n)) = 0$) sestrojíme ÚED

   $$l_1 \vee l_2 \vee \cdots \vee l_n$$

   kde

   $$l_i = \begin{cases} p_i & \text{pokud } e(p_i) = 0 \\ \neg p_i & \text{pokud } e(p_i) = 1 \end{cases}$$

2. φ je konjunkcí ÚED postupně sestrojených v bodu 1.

Příklad 1.32. Sestrojme ÚKNF k formuli $(p \to q) \wedge (p \to r)$. Vytvoříme tabulku pravdivostní funkce formule $(p \to q) \wedge (p \to r)$ a přidáme k ní sloupec, do kterého zapíšeme příslušné ÚED.

p	q	r	$(p \to q) \wedge (p \to r)$	ÚED
0	0	0	1	
0	0	1	1	
0	1	0	1	
0	1	1	1	
1	0	0	0	$\neg p \vee q \vee r$
1	0	1	0	$\neg p \vee q \vee \neg r$
1	1	0	0	$\neg p \vee \neg q \vee r$
1	1	1	1	

Sestrojená ÚKNF je: $(\neg p \vee q \vee r) \wedge (\neg p \vee q \vee \neg r) \wedge (\neg p \vee \neg q \vee r)$.

Pro ÚKNF platí tvrzení analogická těm, které jsme dokázali pro ÚDNF:

Věta 1.33. *Nechť f je booleovská funkce, která nabývá aspoň pro jednu kombinaci argumentů hodnotu 0 (tedy nepředstavuje tautologii). Pak ÚKNF φ sestrojená výše uvedeným postupem splňuje $f = f_\varphi$ (neboli φ indukuje funkci f; tabulky f a φ jsou stejné).*

Věta 1.34. *Ke každé booleovské funkci f, která nepředstavuje tautologii, existuje ÚKNF φ tak, že $f = f_\varphi$.*

Z výše uvedeného je zřejmé, že platí také následující tvrzení.

Věta 1.35. *Ke každé formuli výrokové logiky, která není kontradikcí (tautologií) existuje s ní sémanticky ekvivalentní formule, která je ve tvaru ÚDNF (ÚKNF).*

Vyjadřování spojek jinými spojkami a funkční úplnost

Z výše uvedeného víme, že

$$\varphi \vee \psi \equiv \neg(\neg\varphi \wedge \neg\psi),$$

tedy že formule $\varphi \vee \psi$ a $\neg(\neg\varphi \wedge \neg\psi)$ jsou sémanticky ekvivalentní. Tyto dvě formule tedy při každém ohodnocení e nabývají stejných pravdivostních hodnot. To lze ekvivalentně vyjádřit následovně: pro libovolné pravdivostní hodnoty a a b platí

$$a \vee^{\cdot} b = \neg^{\cdot}(\neg^{\cdot} a \wedge^{\cdot} \neg^{\cdot} b).$$

Výše uvedené skutečnosti můžeme interpretovat tak, že spojku $\vee$ lze vyjádřit pomocí spojek $\neg$ a $\wedge$ nebo ekvivalentně tak, že booleovskou funkci $\vee^{\cdot}$ lze vyjádřit pomocí booleovských funkcí $\neg^{\cdot}$ a $\wedge^{\cdot}$ (přesněji: funkce $\vee^{\cdot}$ je složením funkcí $\neg^{\cdot}$ a $\wedge^{\cdot}$).

Následující vztahy ukazují další možnosti, jak jednu spojku vyjádřit pomocí jiných (snadno se ověří tabulkovou metodou):

$$\begin{aligned}
a \vee^{\cdot} b &= \neg^{\cdot} a \to^{\cdot} b \\
a \wedge^{\cdot} b &= \neg^{\cdot}(\neg^{\cdot} a \vee^{\cdot} \neg^{\cdot} b) \\
a \wedge^{\cdot} b &= \neg^{\cdot}(a \to^{\cdot} \neg^{\cdot} b) \\
a \to^{\cdot} b &= \neg^{\cdot}(a \wedge^{\cdot} \neg^{\cdot} b) \\
a \to^{\cdot} b &= \neg^{\cdot} a \vee^{\cdot} b
\end{aligned}$$

$\uparrow^{\cdot}$	0	1
0	1	1
1	1	0

$\downarrow^{\cdot}$	0	1
0	1	0
1	0	0

Tabulka 11: Pravdivostní funkce Shefferovy (↑) a Peirceovy (↓) spojky.

Z hlediska vyjadřování spojek jinými spojkami hrají důležitou roli tzv. Shefferova spojka, která se označuje ↑, popř. | nebo NAND, a Peirceova (také Nicodova) spojka, která se označuje ↓, popř. také NOR. Pravdivostní funkce těchto spojek ukazuje tabulka 11.

Snadno ověříme, že platí:

$$\begin{aligned} a \uparrow^{\cdot} b &= \neg^{\cdot}(a \wedge^{\cdot} b) \\ a \downarrow^{\cdot} b &= \neg^{\cdot} a \wedge^{\cdot} \neg^{\cdot} b \\ a \downarrow^{\cdot} b &= \neg^{\cdot}(a \vee^{\cdot} b) \end{aligned}$$

To je důvodem výše zmíněného označení NAND (negace konjunkce) a NOR (negace disjunkce). Navíc lze každou ze spojek $\neg$, $\wedge$ a $\vee$ vyjádřit pomocí $\uparrow$ i pomocí $\downarrow$ (ověřte tabulkovou metodou):

$$\begin{aligned} \neg^{\cdot} a &= a \uparrow^{\cdot} a \\ a \wedge^{\cdot} b &= (a \uparrow^{\cdot} b) \uparrow^{\cdot} (a \uparrow^{\cdot} b) \\ a \vee^{\cdot} b &= (a \uparrow^{\cdot} a) \uparrow^{\cdot} (b \uparrow^{\cdot} b) \end{aligned}$$

$$\begin{aligned} \neg^{\cdot} a &= a \downarrow^{\cdot} a \\ a \wedge^{\cdot} b &= (a \downarrow^{\cdot} a) \downarrow^{\cdot} (b \downarrow^{\cdot} b) \\ a \vee^{\cdot} b &= (a \downarrow^{\cdot} b) \downarrow^{\cdot} (a \downarrow^{\cdot} b) \end{aligned}$$

Vzniká přirozená otázka, zda existují spojky, kterými lze vyjádřit libovolnou booleovskou funkci. Takové spojky, resp. jejich pravdivostní funkce, tvoří tzv. funkčně úplný systém:

Definice 1.36. Množina $\{f_1, \ldots, f_k\}$ booleovských funkcí je *funkčně úplná*, pokud každou booleovskou funkci $f : \{0,1\}^n \to \{0,1\}$ lze vyjádřit jako složení některých funkcí z $\{f_1, \ldots, f_k\}$. Množina výrokových spojek je funkčně úplná, jestliže je funkčně úplná množina jejich pravdivostních funkcí.

Následující věta popisuje funkčně úplný systém.

Věta 1.37. $\{\neg^{\cdot}, \wedge^{\cdot}, \vee^{\cdot}\}$ *je funkčně úplná.*

Důkaz. Máme dokázat, že každou booleovskou funkci f lze získat složením $\neg^{\cdot}$, $\wedge^{\cdot}$ a $\vee^{\cdot}$. To plyne z věty 1.30 a z věty 1.34: K libovolné funkci f, který není kontradikcí, existuje odpovídající ÚDNF φ a k funkci f, která není tautologií, existuje odpovídající ÚKNF φ. Její pravdivostní funkce $\varphi^{\cdot}$ formule φ je složená z $\neg^{\cdot}$, $\wedge^{\cdot}$ a $\vee^{\cdot}$. □

Poznamenejme, že větu 1.37 lze ekvivalentně formulovat takto: množina spojek $\{\neg, \wedge, \vee\}$ je funkčně úplná.

Následující tvrzení je zřejmé.

Lemma 1.38. *Je-li možné každou funkci z množiny* $\{f_1, \ldots, f_k\}$ *vyjádřit jako složení některých funkcí z množiny* $\{g_1, \ldots, g_l\}$*, pak je-li* $\{f_1, \ldots, f_k\}$ *funkčně úplná, je i* $\{g_1, \ldots, g_l\}$ *funkčně úplná.*

Například každou funkci z $\{\neg^{\cdot}, \wedge^{\cdot}, \vee^{\cdot}\}$ lze vyjádřit složením funkcí z $\{\neg^{\cdot}, \to^{\cdot}\}$. Víme totiž, že $a \vee^{\cdot} b = \neg^{\cdot} a \to^{\cdot} b$ a $a \wedge^{\cdot} b = \neg^{\cdot}(a \to^{\cdot} \neg^{\cdot} b)$. Protože je $\{\neg^{\cdot}, \wedge^{\cdot}, \vee^{\cdot}\}$ úplná, je dle lemma 1.38 i $\{\neg^{\cdot}, \to^{\cdot}\}$ úplná.

Z funkční úplnosti množiny $\{\neg^{\cdot}, \wedge^{\cdot}, \vee^{\cdot}\}$, z uvedených vztahů o vzájemném vyjadřování spojek a z lemma 1.38 tedy plyne:

Věta 1.39. *Následující množiny logických funkcí jsou funkčně úplné:* $\{\neg^{\cdot}, \wedge^{\cdot}\}$, $\{\neg^{\cdot}, \vee^{\cdot}\}$, $\{\neg^{\cdot}, \to^{\cdot}\}$, $\{\uparrow^{\cdot}\}$, $\{\downarrow^{\cdot}\}$.

Žádná z následujících množin ale funkčně úplná není: $\{\neg^{\cdot}\}$, $\{\wedge^{\cdot}\}$, $\{\vee^{\cdot}\}$, $\{\wedge^{\cdot}, \vee^{\cdot}\}$, $\{\to^{\cdot}\}$, $\{\leftrightarrow^{\cdot}\}$, $\{\neg^{\cdot}, \leftrightarrow^{\cdot}\}$. Pokuste se zdůvodnit proč (poslední tři jsou obtížnější).

Shrnutí

Logika je věda o správném usuzování. V logice jde o formu usuzování, nikoli o obsah. Nejjednodušším formálním systémem logiky je klasický výroková logika. Zabývá se syntaktickými aspekty (jazyk, formule) i sémantickými aspekty (pravdivost, vyplývání).

Pojmy k zapamatování

- logika,
- logická spojka,
- výrok, pravdivostní hodnota výroku,
- symbol logické spojky a pravdivostní funkce logické spojky,
- kvantifikátor,
- jazyk a formule výrokové logiky,
- pravdivostní ohodnocení a pravdivostní hodnota formule, sémantické vyplývání,
- tabulková metoda,
- úplná disjunktivní normální forma, úplná konjunktivní normální forma,
- funkční úplnost.

Kontrolní otázky

1. *Jaké znáte logické spojky?*
2. *Co to je klasická a neklasická logika?*
3. *Jaký je vztah mezi obecným a existenčním kvantifikátorem?*
4. *Co to je formule výrokové logiky?*
5. *Vysvětlete, co to je tabulková metoda a k čemu slouží.*
6. *Vysvětlete pojmy sémantické vyplývání a sémanticky ekvivalentní formule.*
7. *Co to je booleovská funkce indukovaná danou formulí?*
8. *Zdůvodněte správnost konstrukce ÚDNF a ÚKNF.*
9. *Vyjádřete spojky* $\to$ *a* $\leftrightarrow$ *pomocí spojky* $\uparrow$ *a poté pomocí spojky* $\downarrow$.

Cvičení

1. Určete pravdivostní hodnotu výroku „Jestliže Čína je nejlidnatější stát světa, pak Petr je synem Marie.“. Přitom „Petr je synem Marie.“ je pravdivý výrok.

2. Je dán výrok „Pro každé x platí, že jestliže $2x+1$ je sudé, pak x je násobkem 5“. Přitom D_x je množina všech přirozených čísel. Je daný výrok pravdivý?
3. Jsou dány výroky „Pro každé x existuje y tak, že platí $x \leq y$“ a „Existuje y tak, že pro každé x platí $x \leq y$“. Oborem hodnot proměnných x i y je množina všech celých čísel, tj. $D_x = D_y = \{0, 1, -1, 2, -2, 3, -3, \dots\}$. Určete pravdivostní hodnoty daných výroků.
4. U každé z následujících formulí zjistěte, zda je tautologie, kontradikce, splnitelná.

$\varphi \vee \neg\varphi$	zákon vyloučeného třetího
$\neg(\varphi \wedge \neg\varphi)$	zákon sporu
$\neg(\varphi \wedge \psi) \leftrightarrow (\neg\varphi \vee \neg\psi)$	De Morganův zákon
$\neg(\varphi \vee \psi) \leftrightarrow (\neg\varphi \wedge \neg\psi)$	De Morganův zákon
$(\varphi \rightarrow \psi) \leftrightarrow (\neg\psi \rightarrow \neg\varphi)$	zákon kontrapozice

5. Zjistěte, zda formule ψ sémanticky plyne z φ: (a) φ je $(p \wedge q) \vee r$, ψ je $p \rightarrow (q \vee r)$; (b) φ je $(p \wedge q) \vee r$, ψ je $p \rightarrow (q \vee \neg r)$.
6. Přesvědčte se, že je-li $\psi \models \varphi$ a $\varphi \models \chi$, pak $\psi \models \chi$.

Úkoly k textu

1. Vypište všechny binární logické spojky.
2. Uveďte příklady výrokových forem $V(x, y)$ tak, že $(\forall x)(\exists y)(V(x, y))$ je pravdivý výrok a $(\exists y)(\forall x)(V(x, y))$ je nepravdivý výrok. Lze najít příklad formy $V(x, y)$ tak, aby $(\forall x)(\exists y)(V(x, y))$ byl nepravdivý výrok a aby $(\exists y)(\forall x)(V(x, y))$ byl pravdivý výrok?
3. Ukažte, že je-li $V(x)$ výroková forma, pak $||(\forall x)(V(x))|| = \min_{x \in D_x} ||V(m)||$ a $||(\exists x)(V(x))|| = \max_{x \in D_x} ||V(m)||$.
4. Ukažte, že je-li $V(x)$ výroková forma, pak $||(\forall x)(V(x))|| = ||\neg(\exists x)(\neg V(x))||$ a $||(\exists x)(V(x))|| = ||\neg(\forall x)(\neg V(x))||$.
5. Někdy se uvádí následující varianta paradoxu lháře: Kréťan říká „Všichni Kréťani lžou.“ Je to paradox, tj. vede tato situace ke sporu podobně jako vede ke sporu paradox lháře?
6. Zdůvodněte podle definice, že formule φ je tautologie, právě když φ sémanticky vyplývá z prázdné množiny formulí.

Řešení

1. Výrok je pravdivý.
2. Výrok je pravdivý. Nápověda: Pro každé přirozené číslo m je $2m + 1$ liché, tedy „Jestliže $2m + 1$ je sudé, pak m je násobkem 5“ je pravdivý výrok, viz tabulka pravdivostní funkce spojky implikace.
3. Výrok „Pro každé x existuje y tak, že platí $x \leq y$.“ je pravdivý. Výrok „Existuje y tak, že pro každé x platí $x \leq y$.“ je nepravdivý.
4. Všechny formule jsou tautologie.
5. (a) ano; (b) ne.

6. Jednoduchou úvahou přímo z definice: Nechť e je libovolné ohodnocení, při kterém je ψ pravdivá. Protože předpokládáme $\psi \models \varphi$, je při ohodnocení e pravdivá také φ, a tedy protože předpokládáme $\varphi \models \chi$, je při e pravdivá i χ, což jsme měli dokázat.

2 Množiny, relace, funkce

Studijní cíle: Po prostudování této kapitoly by student měl rozumět pojmům množina, relace a funkce. Měl by znát základní operace a vztahy definované pro množiny, základní operace definované pro relace, způsoby reprezentace relací a základní typy funkcí. Student by měl tyto pojmy znát aktivně, měl by tedy umět samostatně dokazovat jednoduchá tvrzení, hledat příklady a protipříklady.

Klíčová slova: množina, prvek, podmnožina, průnik, sjednocení, rozdíl, kartézský součin, relace, inverzní relace, skládání relací, funkce, injekce, surjekce, bijekce

2.1 Co a k čemu jsou množiny, relace a funkce

Množiny, relace a funkce jsou matematickými protějšky jevů, se kterými se setkáváme v každodenním životě. Množina je protějškem *souboru* (či *seskupení*). Relace je protějškem *vztahu*. Funkce je protějškem *přiřazení*. Pojmy množina, relace a funkce patří k základním stavebním prvkům diskrétní matematiky a matematiky vůbec. Umožňují přesné, srozumitelné a jednoduché vyjadřování. Používají se v matematice (bez jejich znalosti nemůžeme číst žádný matematický text) a v řadě aplikovaných oborů včetně informatiky (funkcionální programování, relační databáze, informační systémy, znalostní inženýrství a další).

Množina, relace a funkce jsou základní pojmy matematiky. V informatice se bez nich neobejdeme.

Průvodce studiem

S pojmy množina, relace a funkce se podrobně seznamte. Jsou to jednoduché pojmy. Byly zavedeny, abychom mohli přesně mluvit o souborech, seskupeních, systémech, vztazích, přiřazeních apod. Nenechte se svést tím, že víte, co je to seskupení nebo vztah. Když formalismus množin, relací a funkcí dobře zvládnete, ušetříte si spoustu práce v dalším studiu. Navíc budete umět praktické problémy dobře „uchopit“ a popsat. Když naopak formalismus množin, relací a funkcí nezvládnete, budete se s tímto nedostatkem v dalším studiu neustále potýkat.

2.2 Množiny

2.2.1 Pojem množiny

Pojem množina je matematickým protějškem běžně používaných pojmů *soubor*, *seskupení*, apod. *Množina* je objekt, který se skládá z jiných objektů, tzv. *prvků* té množiny. Tak například množina (označme ji S) všech sudých čísel, která jsou větší než 1 a menší než 9, se skládá z čísel 2, 4, 6, 8. Tato čísla jsou tedy prvky množiny S. Fakt, že S se skládá právě z prvků 2, 4, 6, 8 zapisujeme

Množina je objekt, který se skládá z jiných objektů, tzv. prvků množiny.

$$S = \{2, 4, 6, 8\}.$$

Množiny zpravidla označujeme velkými písmeny $(A, B, \ldots, Z)$, jejich prvky pak malými písmeny $(a, b, \ldots, z)$. Fakt, že x je prvkem množiny A označujeme

$$x \in A$$

a říkáme také, že x patří do A (popř. x je v A, A obsahuje x apod.). Není-li x prvkem A, píšeme $x \notin A$.

Daný objekt do dané množiny buď patří, nebo nepatří. Množina je jednoznačně dána

Množina je jednoznačně dána tím, jaké prvky obsahuje.

svými prvky, tj. tím, které prvky do ní patří a které ne. Nemá tedy smysl hovořit o pořadí prvků v množině (tj. pojmy „první prvek množiny“, „druhý prvek množiny“ atd. nemají smysl). Nemá také smysl uvažovat, kolikrát je daný prvek v dané množině (tj. říci „prvek x je v dané množině A třikrát“).

Speciální množinou je tzv. *prázdná množina*. Označuje se $\emptyset$. Tato množina neobsahuje žádný prvek, tj. pro každý prvek x platí $x \notin \emptyset$.

Příklad 2.1. Význačné množiny čísel mají svá speciální označení.

- $\mathbb{N}$ označuje množinu všech *přirozených čísel*. $\mathbb{N}$ tedy sestává z prvků $1, 2, 3, 4, 5, \ldots$.
- $\mathbb{Z}$ označuje množinu všech *celých čísel*. $\mathbb{Z}$ tedy sestává z prvků $0, 1, -1, 2, -2, 3, -3, 4, -4, \ldots$.
- $\mathbb{Q}$ označuje množinu všech *racionálních čísel*. $\mathbb{Q}$ tedy sestává z celočíselných zlomků, tj. z čísel $\frac{m}{n}$, kde $m \in \mathbb{Z}$, $n \in \mathbb{N}$.
- $\mathbb{R}$ označuje množinu všech *reálných čísel*. Ta obsahuje i iracionální čísla, např. $\sqrt{2}, \pi$ apod.

Prvky konečných množin lze očíslovat čísly 1, ..., n. Pokud to nelze, je množina nekonečná.

Množiny se dělí na konečné a nekonečné. Množina A se nazývá *konečná*, právě když existuje přirozené číslo n tak, že prvky této množiny lze očíslovat čísly $1, 2, \ldots n$ (tedy každému prvku přiřadíme některé z čísel $1, \ldots, n$ tak, že různým prvkům přiřadíme různá čísla a při očíslování každé z čísel $1, \ldots, n$ použijeme). Číslo n se přitom nazývá počet prvků množiny A a značíme ho $|A|$, tj. $|A| = n$. Říkáme také, že A má n prvků. Např. množina $\{2, 4, 6, 8\}$ je konečná. Zvolíme-li totiž $n = 4$, můžeme její prvky očíslovat např. následovně: prvku 2 přiřadíme číslo 1, prvku 4 číslo 2, prvku 6 číslo 3, prvku 8 číslo 4. Máme tedy $|\{2, 4, 6, 8\}| = 4$, tj. počet prvků množiny $\{2, 4, 6, 8\}$ je 4. Množina A se nazývá *nekonečná*, není-li konečná. Pak píšeme $|A| = \infty$ a říkáme, že A má nekonečně mnoho prvků. Např. množina $\mathbb{N}$ všech přirozených čísel je nekonečná. Více se o nekonečným množinách dozvíme v kapitole 7.

2.2.2 Zápisování množin

$\{a_1, \ldots, a_n\}$ je množina, která obsahuje právě prvky $a_1, \ldots, a_n$.

Množiny zapisujeme dvěma základními způsoby. Prvním je zápis tzv. *výčtem prvků*. Množina sestávající právě z prvků $a_1, \ldots, a_n$ se označuje $\{a_1, \ldots, a_n\}$. Příkladem je výše uvedený zápis $\{2, 4, 6, 8\}$. Zápis výčtem prvků můžeme použít u konečných množin. Druhým je zápis udáním tzv. *charakteristické vlastnosti*. Množina sestávající právě z prvků, které splňují vlastnost $\varphi(x)$, se označuje $\{x \mid \varphi(x)\}$.

$\{x \mid \varphi(x)\}$ je množina, která obsahuje právě prvky x splňující vlastnost $\varphi(x)$.

Vlastnost $\varphi(x)$ může být popsaná třeba i v přirozeném jazyce, ale musí mít jednoznačný smysl. Např. je-li $\varphi(x)$ vlastnost „x je sudé přirozené číslo větší než 1 a menší než 9“, můžeme uvažovat množinu označenou $\{x \mid \varphi(x)\}$. Ta je shodná s množinou označenou $\{2, 4, 6, 8\}$.

Místo „množina označená zápisem $\{\ldots\}$“ budeme říkat jen „množina $\{\ldots\}$“. Např. říkáme „množina $\{a, b, c\}$ má tři prvky“, „uvažujme množinu $\{x \mid x$ je sudé celé číslo$\}$“ apod.

Někdy se používá i pro zápis nekonečných množin způsob, který je podobný zápisu výčtem. Například množinu všech kladných sudých čísel zapíšeme $\{2, 4, 6, 8, \ldots\}$. Obecně tedy můžeme použít zápis $\{a_1, a_2, a_3, a_4, \ldots\}$, pokud je z prvků a_1, a_2, a_3, a_4 zřemá vlastnost charakterizující prvky popisované množiny. Poznamenejme také, že prázdná množina se někdy zapisuje $\{\}$.

Poznámka 2.2. Zápis výčtem prvků svádí k tomu mluvit o prvním prvku množiny, druhém prvku množiny, atd. Např. u množiny $\{2, 4, 6, 8\}$ máme tendenci říci, že 2 je prvním prvkem, 4 druhým prvkem atd. My však víme, že výrazy „první prvek množiny“, „druhý prvek množiny“ atd. nemají smysl. Množina je totiž dána jen tím, jaké prvky obsahuje, ne jejich pořadím. Při zápisu výčtem se ale pořadí objevuje. Správně bychom měli říci, že $\{2, 4, 6, 8\}$ označuje množinu, v jejímž zápise výčtem, který jsme použili, je prvek 2 na prvním místě, prvek 4 na druhém místě atd. Stejnou množinu můžeme zapsat výčtem a např. $\{4, 6, 2, 8\}$. V tomto zápise je prvek 2 na třetím místě.

Zápis výčtem svádí dále k tomu mluvit o tom, kolikrát se prvek v dané množině vyskytuje. Z technickým důvodů je výhodné připustit, aby se prvky v zápisu výčtem opakovaly. Můžeme např. napsat $\{2, 4, 6, 8, 2, 2, 4\}$. Takový zápis označuje stejnou množinu jako $\{2, 4, 6, 8\}$. Stejnou množinu označuje i $\{6, 6, 6, 2, 4, 8\}$. Záleží tedy jen na tom, které prvky se v zápise vyskytují, nezáleží na počtu jejich výskytu. Nelze tedy např. říci, že množina $\{2, 4, 6, 8, 2, 2, 4\}$ obsahuje tři prvky 2. Můžeme jen říci, že v zápise $\{2, 4, 6, 8, 2, 2, 4\}$ se prvek 2 vyskytuje třikrát.

Poznámka 2.3. (1) Zápis $\{x \in A \mid \varphi(x)\}$ označuje množinu $\{x \mid x \in A \text{ a } \varphi(x)\}$. Je to tedy zápis pomocí charakteristické vlastnosti. Označíme-li totiž $\psi(x)$ vlastnost, kterou prvek x splňuje, právě když patří do A a splňuje $\varphi(x)$, pak množina $\{x \in A \mid \varphi(x)\}$ je rovna množině $\{x \mid \psi(x)\}$. Např. množina $\{x \in \mathbb{Z} \mid x \leq 2\}$ je množina $\{x \mid x \in \mathbb{Z} \text{ a } x \leq 2\}$, tj. množina všech celých čísel, která jsou nejvýše rovna 2.

(2) Často se také používá zápis $\{a_i \mid i \in I\}$. Přitom I je nějaká množina (říká se jí *indexová*) a pro každý (index) $i \in I$ je a_i nějaký prvek. Pak $\{a_i \mid i \in I\}$ je množina

$$\{x \mid \text{existuje } i \in I \text{ tak, že } x = a_i\}.$$

$\{a_i \mid i \in I\}$ je tedy vlastně zápis pomocí charakteristické vlastnosti, neboť označuje množinu $\{x \mid \varphi(x)\}$, kde $\varphi(x)$ je „existuje $i \in I$, tak, že $x = a_i$“. Je-li každý prvek a_i množinou, nazývá se $\{a_i \mid i \in I\}$ indexovaný systém množin.

(3) Při zápise pomocí charakteristické vlastnosti se při popisu vlastnosti $\varphi(x)$ často používají obraty „pro každé y platí, že ...“ a „existuje y tak, že platí ...“. Jak je běžné, budeme tyto obraty zkráceně zapisovat (po řadě) pomocí „$\forall y \ldots$“ a „$\exists y \ldots$“ s případnými závorkami, které zajistí jednoznačný způsob čtení, popř. větší srozumitelnost. „$\forall y \in Y \ldots$“ a „$\exists y \in Y \ldots$“ znamenají „pro každé y z množiny Y platí, že ...“ a „existuje y z množiny Y tak, že platí ...“. Např. množina $\{x \mid \exists y \in \mathbb{N} : x = y^2\}$ je množina prvků x takových, že existuje přirozené číslo y tak, že $x = y^2$. Je to tedy množina všech druhých mocnin přirozených čísel. V Kapitole **??** se s kvantifikátory a jejich vlastnostmi seznámíme podrobněji.

Příklad 2.4. Podívejte se na následující množiny a jejich zápisy.

- $\{k \mid \exists n \in \mathbb{N} : k = 2^n\}$ označuje množinu všech kladných mocnin čísla 2. Stejnou množinu označuje $\{2, 4, 8, 16, \ldots\}$.
- $\{k \in \mathbb{N} \mid k \neq 1 \text{ a jestliže } \exists m, n \in \mathbb{N} : m \cdot n = k, \text{ pak } m = 1 \text{ nebo } n = 1\}$ označuje množinu všech prvočísel.
- $\{\{a, b\}, \{a\}, \{1, 2, 3, \{a, b\}\}\}$ je množina, která má tři prvky. Tyto prvky samy, tj. $\{a, b\}$, $\{a\}$, a $\{1, 2, 3, \{a, b\}\}$, jsou opět množiny. $\{a, b\}$ má dva prvky (a a b), $\{a\}$ má jeden prvek (a), $\{1, 2, 3, \{a, b\}\}$ má tři prvky (jsou to 1, 2, 3 a $\{a, b\}$). Vidíme tedy, že množina může obsahovat prvek, který je sám množinou. Tento prvek-množina sám může obsahovat prvky, které jsou množinami atd.

- $\{\emptyset\}$ je jednoprvková množina. Jejím jediným prvkem je $\emptyset$ (prázdná množina). Uvědomte si, že $\{\emptyset\}$ a $\emptyset$ jsou různé množiny ($\{\emptyset\}$ obsahuje jeden prvek, $\emptyset$ neobsahuje žádný). $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$ je čtyřprvková množina. Její prvky jsou $\emptyset$, $\{\emptyset\}$, $\{\{\emptyset\}\}$, $\{\emptyset, \{\emptyset\}\}$.
- Nechť $a_1 = p$, $a_2 = q$, $a_3 = r$, $a_4 = x$, $a_5 = y$, $a_6 = z$, $a_7 = 1$, $a_8 = r$, $I = \{1, 2, 3, 4\}$, $J = \{1, 2, 3, 7, 8\}$. Pak $\{a_i \mid i \in I\}$ je množina $\{p, q, r, x\}$, $\{a_i \mid i \in J\}$ je množina $\{p, q, r, 1\}$.
- $\{2^i \mid i \in \mathbb{N}\}$ je zápis typu $\{a_i \mid i \in I\}$, ve kterém $a_i = 2^i$ a $I = \mathbb{N}$. Je to množina všech kladných mocnin čísla 2.

Pouhý množinový zápis umožňuje přehledně vyjádřit strukturu, kterou chceme zachytit.

Množina je pojem, který intuitivně používáme v běžném životě, chceme-li označit několik objektů najednou („dát je do jednoho pytle“). Např. řekneme-li „ekonomické oddělení“, myslíme tím vlastně množinu zaměstnanců ekonomického odělení. Množinový zápis také umožňuje jednoduše vyjádřit hierarchickou strukturu. Předpokládejme pro jednoduchost, že v nemocnici pracuje ředitel (R), tři údržbáři (U_1, U_2, U_3), na oddělení chirurgie dva lékaři (C_1, C_2) a tři sestry (CS_1, CS_2, CS_3), na anesteziologicko-resuscitačním oddělení dva lékaři (A_1, A_2) a dvě sestry (AS_1, AS_2), na oddělení interním tří lékaři (I_1, I_2, I_3) a čtyři sestry (IS_1, IS_2, IS_3, IS_4). Pak strukturu zaměstnanců nemocnice popisuje jistým způsobem např. množina

$$\{\{R\}, \{U_1, U_2, U_3\}, \\ \{\{C_1, C_2, CS_1, CS_2, CS_3\}, \{A_1, A_2, AS_1, AS_2\}, \{I_1, I_2, I_3, IS_1, IS_2, IS_3, IS_4\}\}\}.$$

Při tomto pohledu se díváme takto: Zaměstnanci nemocnice jsou rozděleni do třech skupin, a to $\{R\}$ (vedení), $\{U_1, U_2, U_3\}$ (technický personál), $\{\{C_1, C_2, CS_1, CS_2, CS_3\}, \{A_1, A_2, AS_1, AS_2\}, \{I_1, I_2, I_3, IS_1, IS_2, IS_3, IS_4\}$ (zdravotnický personál). Zdravotnický personál se dále dělí na $\{C_1, C_2, CS_1, CS_2, CS_3\}$ (chirurgické oddělení), $\{A_1, A_2, AS_1, AS_2\}$ (anesteziologicko-resuscitační oddělení) a $\{I_1, I_2, I_3, IS_1, IS_2, IS_3, IS_4\}$ (interní oddělení). Jiným způsobem (vedení, technický personál, lékaři, sestry) popisuje strukturu zaměstnanců množina

$$\{\{R\}, \{U_1, U_2, U_3\}, \\ \{C_1, C_2, A_1, A_2, I_1, I_2, I_3\}, \{CS_1, CS_2, CS_3, AS_1, AS_2, IS_1, IS_2, IS_3, IS_4\}\}.$$

Pokud mluvíme o množině, jejíž prvky jsou opět množiny, říká se někdy místo „množina množin“ spíše „systém množin“ nebo „soubor množin“. Důvody k tomu jsou však jen estetické (zvukomalebné, „množina množin“ nezní dobře).

Je-li vlastnost $\varphi(x)$ popsaná slovně, nemusí být určitá, a pak $\{x \mid \varphi(x)\}$ nepopisuje množinu.

Poznámka 2.5. (1) Ne každou slovně popsanou vlastnost lze použít k zápisu množiny. Uvažujme např. zápis $\{x \mid x$ je číslo udávající ve stupních Celsia vysokou letní teplotu v Česku$\}$. Problém je v tom, že pojem „vysoká letní teplota v Česku“ není vymezen tak, že by každá teplota buď byla nebo nebyla vysoká. Např. by teploty 30 stupňů a více byly vysoké a teploty menší než 30 stupňů vysoké nebyly. Pojem „vysoká letní teplota v Česku“ je totiž vágní, určité teploty mu vyhovují lépe, určité hůře. Vágností se zabývá tzv. fuzzy logika a fuzzy množiny. Fuzzy množina se od (klasické, tj. „nefuzzy“) množiny liší v zásadě v tom, že objekt do fuzzy množiny může patřit v určitém stupni, např. 0 (vůbec nepatří), 0.2 (patří jen trochu), ..., 1 (úplně patří). Klasické množiny lze chápat jako hraniční případ fuzzy množin, kdy používáme pouze stupně 0 a 1.

(2) Přístup k množinám, který zde představujeme, je tzv. naivní (popř. intuitivní). Může však vést k zvláštním situacím, tzv. paradoxům. Začátkem 20. stol. na ně upozornil Bertrand Russel. Aby paradoxy odstranil, navrhl tzv. teorii typů a na ní vybudoval přístup k množinám, ve kterém se paradoxy neobjevují. Jiný, později mnohem

rozšířenější přístup k množinám, ve kterém se paradoxy nevyskytují, nabízí tzv. axiomatická teorie množin. Pro naše účely a i v řadě jiných situací však postačuje naivní přístup. Protože je také mnohem jednodušší, zůstaneme u něj.

Russellův paradox ukazuje meze našeho přístupu k množinám.

(3) Jedním z nejznámějších paradoxů naivního přístupu k množinám je tzv. Russellův paradox. Vypadá takto: Prvky množin mohou být opět množiny. Dále lze jistě uvažovat vlastnost „$x \notin x$“ a množiny objektů, které ji splňují. Označme ji N a nazvěme ji množinou všech normálních množin, tj. $x \in N$, právě když $x \notin x$. Poznamenejme na okraj, že všechny množiny, které jsme zatím viděli, byly normální. Protože N sama o sobě množina, můžeme se zeptat, zda platí $N \in N$, tj. zda N sama je normální. Je jasné, že musí být buď (a) $N \in N$, nebo (b) $N \notin N$. Zkusme ty možnosti rozebrat: (a) Když $N \in N$, pak N splňuje vlastnost prvků množiny N, tedy splňuje $x \notin x$, tedy $N \notin N$. Naopak, když (b) $N \notin N$, pak protože N splňuje vlastnost $x \notin x$, je dle definice normální a tedy patří do množiny všech normálních množin, tedy patří do N, tedy $N \in N$. Vidíme tedy, že z $N \in N$ plyne $N \notin N$ a z $N \notin N$ plyne $N \in N$, tedy $N \in N$ platí, právě když $N \notin N$. To je spor. Z přirozených předpokladů jsme přirozenými úvahami došli ke sporu, odtud název paradox. Russellův paradox má řadu populárních podob. Jednou z nich je tzv. paradox holiče: Ve městě je holič, který holí právě ty lidi, kteří neholí sami sebe. Otázka: Holí holič sám sebe?

2.2.3 Vztahy mezi množinami

Základní vztahy mezi množinami jsou rovnost a inkluze.

Základní vztahy mezi množinami jsou *rovnost* (označujeme ji symbolem $=$) a *inkluze* (označujeme ji symbolem $\subseteq$). Jsou-li A a B množiny, pak $A = B$ čteme „(množina) A se rovná (množině) B“ a $A \subseteq B$ čteme „(množina) A je podmnožinou (množiny) B“. Přitom

$$A = B \quad \text{znamená, že pro každý } x : x \in A, \text{ právě když } x \in B$$

a

$$A \subseteq B \quad \text{znamená, že pro každý } x : \text{ jestliže } x \in A, \text{ pak } x \in B.$$

Jinými slovy, $A = B$ znamená, že množiny A a B obsahují stejné prvky (neexistuje prvek, který by do jedné patřil ale do druhé ne). $A \subseteq B$ znamená, že všechny prvky množiny A jsou také prvky množiny B. $A \neq B$ znamená, že neplatí $A = B$. $A \not\subseteq B$ znamená, že neplatí $A \subseteq B$.

Všimněme si, že $A = B$ platí, právě když platí zároveň $A \subseteq B$ a $B \subseteq A$. Někdy je výhodné psát $A \subset B$, abychom označili, že $A \subseteq B$ a $A \neq B$. Dále si uvědomme, že pro všechny množiny A, B, C je $\emptyset \subseteq A$, $A \subseteq A$ a že jestliže $A \subseteq B$ a $B \subseteq C$, pak $A \subseteq C$.

Dokažme poslední tvrzení. Předpokládejme, že $A \subseteq B$ a $B \subseteq C$. Máme dokázat $A \subseteq C$, tedy že pro každý x platí, že když $x \in A$, pak $x \in C$. Zvolme tedy libovolný x a předpokládejme, že $x \in A$. Chceme ukázat $x \in C$. Uděláme to následovně. Z $x \in A$ a z předpokladu $A \subseteq B$ plyne, že $x \in B$. Dále z $x \in B$ a z předpokladu $B \subseteq C$ plyne $x \in C$. Důkaz je hotov.

Právě dokázané tvrzení je velmi jednoduché. Je tak jednoduché, že má člověk sklon říci „to je přece jasné, to není třeba dokazovat“. Tvrzení však mohou být složitější a složitější (viz dále) tak, že už nebudou „přece jasná“. Dokázat dané tvrzení, tj. vyjít z předpokladů a pomocí jednoduchých úvah (a popř. i pomocí známých tvrzení) dojít z předpokladů k závěru daného tvrzení, je pak jediným způsobem, jak se přesvědčit, že tvrzení platí. Ostatně uvědomme si, že i u velmi jednoduchých tvrzení je jediným korektním zdůvodněním důkaz. Říci „to je přece jasné“ nemá jako argument žádnou váhu. Za prvé, člověk se může splést (co, co se mu zdá jasné, tak ve skutečnosti nemusí být). Za druhé, a to je snad ještě důležitější, argumentujeme-li pomocí „to je jasné“, může se nám stát, že pojmy, o kterých mluvíme, vlastně pořádně nechápeme, že je chápeme jen povrchně, intuitivně. Umět dokázat i jednoduchá tvrzení (a tvrzení vůbec) je tedy i

dobrý test, jestli věci rozumíme (u složitějších tvrzení je dobré alespoň důkaz si přečíst a pochopit). Tedy naše doporučení: Čtěte důkazy a pokoušejte se je sami vymýšlet. To je užitečný zvyk nejen pro diskrétní matematiku. Naše zkušenost je následuící: Osvojit si důkazy (číst je, ty jednoduché i sami formulovat) vyžaduje počáteční časovou investici. Ta se ale vyplatí. Věcem lépe porozumíte, začnou se zdát jednoduché a začnete vidět souvislosti. Platí to nejen pro matematiku a informatiku, ale i pro každou oblast, ve které ze základních kamenů (pojmů, konstruktů, principů, ...) budujeme složitější systém.

Příklad 2.6. Platí např.

- $\{2\} = \{n \in \mathbb{N} \mid n \text{ je sudé prvočíslo}\}$,
- $\emptyset = \{k \in \mathbb{Z} \mid \exists n \in \mathbb{N} : 2k = 2n + 1\}$,
- $\{a, b, c, d\} = \{b, d, c, a\}$, $\{a, b, 1\} = \{1, a, a, b, b, b, 1\}$,
- $\{a, b\} \subseteq \{a, b, c, d\}$, $\{a, b\} \subseteq \{1, 2, a, b\}$,
- $\{a, \{a, b\}, \{\{a, 1\}, b\}\} \subseteq \{a, b, \{a, b\}, \{a, b, 1\}, \{\{a, 1\}, b\}\}$,
- $\{a, b\} \not\subseteq \{\{a, b, c\}\}$, $\{\{a, 1\}\} \not\subseteq \{a, b, 1, \{a\}, \{1\}\}$.

Proč platí $\{a, b\} \not\subseteq \{\{a, b, c\}\}$, tj. proč $\{a, b\}$ není podmnožinou $\{\{a, b, c\}\}$? Vždyť v $\{\{a, b, c\}\}$ jsou všechny prvky, které jsou v $\{a, b\}$ (pokušení které může plynout z povrchního chápání $\subseteq$). Zdůvodnění: Např. pro prvek a je $a \in \{a, b\}$, ale $a \notin \{\{a, b, c\}\}$.

Průvodce studiem

Podívejte se znovu na Příklad 2.6. Vztahy mezi množinami, které jsou v něm uvedené, a další podobné vztahy byste měli umět bez problémů zdůvodnit. Tak si ověříte, že těm úplně základním věcem rozumíte. Tady i na jiných místech v textu platí, že skoro nemá smysl číst text dál, dokud vám nebude jasné (tj. dokud nebudete umět pomocí definic zdůvodnit), proč např. platí $\{\{a\}\} \subseteq \{\{a\}, \{b\}\}$ a proč neplatí $\{\{a\}\} \subseteq \{\{a, b\}\}$. Než tyto věci začnete jasně chápat a vidět, může to chvíli trvat. Ten čas se vám ale vrátí. Zdůvodněte např., proč je $x \in A$, právě když $\{x\} \subseteq A$.

Množina, jejímiž prvky jsou právě všechny podmnožiny dané množiny X, se nazývá *potenční množina* množiny X a značí se 2^X. Tedy

Potenční množina množiny X je množina všech podmnožin množiny X.

$$2^X = \{A \mid A \subseteq X\}.$$

Vezměme např. $X = \{a, b\}$. X má čtyři podmnožiny. Jsou to $\emptyset$ (ta je podmnožinou každé množiny), $\{a\}$, $\{b\}$ a $\{a, b\}$ (množina je podmnožino sebe samé). Tedy $2^X = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$.

Příklad 2.7.

- Pro $X = \{a\}$ je $2^X = \{\emptyset, \{a\}\}$,
- pro $X = \{1, 2, 3\}$ je $2^X = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$,
- pro $X = \emptyset$ je $2^X = \{\emptyset\}$ (to si promyslete: jedinou podmnožinou množiny $\emptyset$ je $\emptyset$),
- pro $X = \{a, \{a\}\}$ je $2^X = \{\emptyset, \{a\}, \{\{a\}\}, \{a, \{a\}\}\}$.

Věta 2.8. *Je-li X konečná, pak $|2^X| = 2^{|X|}$. Tedy potenční množina n prvkové množiny má právě 2^n prvků.*

Důkaz. Nechť $X = \{x_1, \ldots, x_n\}$. Každou podmnožinu A množiny X můžeme zřejmě reprezentovat posloupností $b_1 \ldots b_n$, kde $b_i = 1$, když $x_i \in X$, a $b_i = 0$, když $x_i \notin X$. Například pro $X = \{x_1, x_2, x_3\}$ a $A = \{x_1, x_3\}$ je $b_1 b_2 b_3 = 101$. Podmnožin A je zřejmě právě tolik jako všech posloupností $b_1 \ldots b_n$ nul a jedniček. Těch je 2^n, protože b_1 může mít dvě hodnoty, nezávisle na tom b_2 dvě hodnoty, atd. To je celkem $2 \cdot \cdots \cdot 2$ (n krát), tedy 2^n, možností. □

2.2.4 Operace s množinami

Se skupinami objektů provádíme v běžném životě různé operace. Např. řekneme „samostatnost a logické uvažování jsou společné vlastnosti Jany a Aleny“. Z množinového pohledu tím myslíme následující. Jana i Alena mají nějaké vlastnosti. Množinu rysů Jany označme J, množinu rysů Aleny označme A. Množiny J a A jsou různé, např. označuje-li m vlastnost „je dobrá v matematice“, může být $m \in J$ (Jana je dobrá v matematice), ale $m \notin A$ (Alena není dobrá v matematice). Označme s a l vlastnosti „samostatnost“ a „logické uvažování“. Označme dále $J \cap A$ množinu vlastností, které patří do J i do A. Pak „samostatnost a logické uvažování jsou společné vlastnosti Jany a Aleny“ vlastně znamená $s \in J \cap A$ a $l \in J \cap A$.

Základní operace s množinami jsou průnik, sjednocení a rozdíl.

Mezi základní operace s množinami, se kterými se seznámíme, patří *průnik* (značí se $\cap$), *sjednocení* (značí se $\cup$) a *rozdíl* (značí se $-$). Jsou-li A a B množiny, definujeme množiny $A \cap B$, $A \cup B$ a $A - B$ předpisy

$$\begin{aligned} A \cap B &= \{x \mid x \in A \text{ a } x \in B\}, \\ A \cup B &= \{x \mid x \in A \text{ nebo } x \in B\}, \\ A - B &= \{x \mid x \in A \text{ a } x \notin B\}. \end{aligned}$$

Tedy x patří do $A \cap B$, právě když x patří do A i do B; x patří do $A \cup B$, právě když x patří do A nebo do B; x patří do $A - B$, právě když x patří do A, ale nepatří do B.

Příklad 2.9.

- Pro $A = \{a, b, e\}$, $B = \{b, c, d\}$ je $A \cap B = \{b\}$, $A \cup B = \{a, b, c, d, e\}$, $A - B = \{a, e\}$,
- pro $A = \{1, 2, a, b\}$, $B = \{1, a\}$ je $A \cap B = \{1, a\}$, $A \cup B = \{1, 2, a, b\}$, $A - B = \{2, b\}$, $B - A = \emptyset$,
- Pro $A = \{a\}$, $B = \{b, \{a\}\}$ je $A \cap B = \emptyset$, $A \cup B = \{a, b, \{a\}\}$,
- pro $A = \{\emptyset, a, \{a\}, \{a, b\}\}$, $B = \{b, \{a, \{b\}\}\}$ je $A \cap B = \emptyset$, $A \cup B = \{\emptyset, a, \{a\}, \{a, b\}, b, \{a, \{b\}\}\}$, $A - B = A$, $B - A = B$.

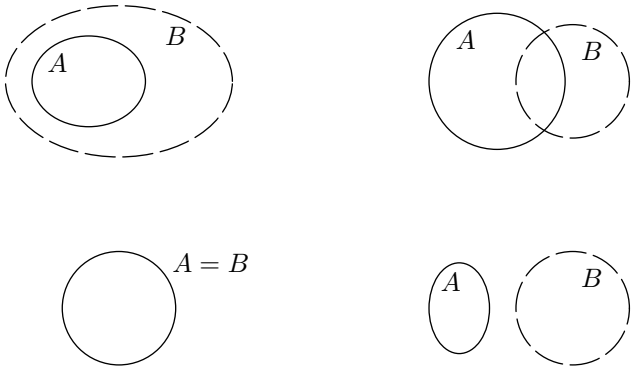
Množiny A a B se nazývají (navzájem) *disjunktní*, právě když $A \cap B = \emptyset$. Např. množiny $\{a, b, c, d\}$ a $\{1, 2, 3\}$ jsou disjunktní, množiny $\{a, b, 1\}$ a $\{1, 2, a\}$ disjunktní nejsou.

Často uvažujeme jednu množinu X, které říkáme *univerzum* (obor našich úvah) a pracujeme jen s množinami, které jsou podmnožinami X. Např. uvažujeme univerzum X všech občanů České republiky a potom pracujeme s jeho podmnožinami (např. množina dětí z X, množina zaměstnaných, množina důchodců apod.). Je-li dáno nějaké univerzum X a množina $A \subseteq X$, pak *doplněk* (někdy také *komplement*) množiny A je množina $X - A$ a značíme ji $\overline{A}$. Např. pro $X = \{a, b, c, d, e\}$ je $\overline{\{a, c\}} = \{b, d, e\}$.

Je-li $A = \{B_i \mid i \in I\}$ množina, jejíž prvky jsou opět množiny (tedy A je systém množin), definujeme sjednocení $\bigcup A$, a to vztahem

$$\bigcup A = \{x \mid \exists i \in I : x \in B_i\}.$$

Tedy $x \in \bigcup A$, právě když x patří do nějaké množiny, která je prvkem A. Například $\bigcup\{\{a, b, c\}, \{a, 1\}, \{1, 2\}\} = \{a, b, c, 1, 2\}$.



Obrázek 1: Vennovy diagramy.

Analogicky definujeme průnik systému množin, tedy

$$\bigcap A \;=\; \{x \mid \forall i \in I \,:\, x \in B_i\}.$$

Je například $\bigcap\{\{a,b,c\},\{a,1\},\{1,2\}\} = \{a\}$.

Vennovy diagramy slouží ke grafické ilustraci množin.

Průvodce studiem

Operace a základní vztahy mezi množinami můžeme ilustrovat pomocí tzv. *Vennových diagramů*[4]. Množiny se znázorňují (jsou reprezentovány) v rovině jako obrazce ohraničené uzavřenými křivkami (kružnice, ovály apod.). Přitom jedna množina může být reprezentována několika obrazci, které se neprotínají, popř. se jen dotýkají. Prvky množiny jsou ty body roviny, které se nacházejí uvnitř odpovídajícího obrazce (popř. se některé prvky v obrazci explicitně vyznačí křížkem). Ke každému obrazci se napíše symbol odpovídající množiny. Podívejte se na Obr. 1. Každá ze čtyř situací (vlevo dole, vpravo dole, vpravo nahoře, vlevo nahoře) znázorňuje dvě množiny, A a B. Pro situaci vlevo dole je $A = B$, vpravo dole jsou A a B disjunktní, vpravo nahoře A a B disjunktní nejsou, vlevo nahoře je $A \subseteq B$. Množina $A \cap B$ je reprezentována obrazcem, který je roven společné části obrazce reprezentujícího A a obrazce reprezentujícího B. Množina $A \cup B$ je reprezentována obrazcem, který je dán sloučením obrazce reprezentujícího A a obrazce reprezentujícího B. Fakt $A \subseteq B$ odpovídá situaci, kdy obrazec reprezentující A je obsažen v obrazci reprezentujícím B.

Vennovy diagramy umožňují názornou představu. Lze pomocí nich znázornit množiny, jejichž prvky můžeme chápat jako dvourozměrné. Některé množiny tak znázornit nemůžeme.

Podívejme se teď na některé základní vlastnosti.

Věta 2.10. *Pro množiny A, B, C platí*

$$\begin{aligned}
A \cap \emptyset = \emptyset, &\qquad A \cup \emptyset = A \\
A \cup A = A, &\qquad A \cap A = A \\
A \cup B = B \cup A, &\qquad A \cap B = B \cap A \\
(A \cup B) \cup C = A \cup (B \cup C), &\qquad (A \cap B) \cap C = A \cap (B \cap C) \\
A \cap (B \cup C) = (A \cap B) \cup (A \cap C), &\qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \\
A \cup (A \cap B) = A, &\qquad A \cap (A \cup B) = A
\end{aligned}$$

Před důkazem si uvědomme následující. Máme-li dokázat $A = B$, máme podle definice dokázat, že pro libovolný prvek x je $x \in A$, právě když $x \in B$. To lze dále rozložit na ověření toho, že z $x \in A$ plyne $x \in B$ a že z $x \in B$ plyne $x \in A$. Pojďme na důkaz Věty 2.10.

Důkaz. $A \cap \emptyset = \emptyset$: Zvolme libovolný x. Je $x \in A \cap \emptyset$, právě když (podle definice $\cap$) $x \in A$ a $x \in \emptyset$. Protože $x \in \emptyset$ je vždy nepravdivé, je vždy nepravdivý i výrok $x \in A$ a $x \in \emptyset$. Máme tedy dále $x \in A$ a $x \in \emptyset$, právě když $x \in \emptyset$. Celkem tedy máme $x \in A \cap \emptyset$, právě když $x \in \emptyset$, což dokazuje $A \cap \emptyset = \emptyset$.

$A \cup \emptyset = A$: $x \in A \cup \emptyset$, právě když (podle definice $\cup$) $x \in A$ nebo $x \in \emptyset$, právě když (protože $x \in \emptyset$ je vždy nepravdivé) $x \in A$.

$A \cup A = A$: $x \in A \cup A$, právě když $x \in A$ nebo $x \in A$, právě když $x \in A$.

$A \cap A = A$: Podobně jako předchozí, $x \in A \cap A$, právě když $x \in A$ a $x \in A$, právě když $x \in A$.

$A \cup B = B \cup A$: $x \in A \cup B$, právě když $x \in A$ nebo $x \in B$, právě když $x \in B$ nebo $x \in A$, právě když $x \in B \cup A$.

$A \cap B = B \cap A$: Podobně jako předchozí.

$(A \cup B) \cup C = A \cup (B \cup C)$: $x \in (A \cup B) \cup C$, právě když $x \in (A \cup B)$ nebo $x \in C$, právě když ($x \in A$ nebo $x \in B$) nebo $x \in C$, právě když (podle pravidel výrokové logiky) $x \in A$ nebo ($x \in B$ nebo $x \in C$), právě když $x \in A$ nebo $x \in B \cup C$, právě když $x \in A \cup (B \cup C)$.

$(A \cap B) \cap C = A \cap (B \cap C)$: Podobně jako předchozí.

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$: $x \in A \cap (B \cup C)$, právě když $x \in A$ a $x \in B \cup C$, právě když $x \in A$ a ($x \in B$ nebo $x \in C$), což je podle pravidel výrokové logiky právě když ($x \in A$ a $x \in B$) nebo ($x \in A$ a $x \in C$), právě když $x \in A \cap B$ nebo $x \in A \cap C$, právě když $x \in (A \cap B) \cup (A \cap C)$.

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$: Podobně jako předchozí.

$A \cup (A \cap B) = A$: $x \in A \cup (A \cap B)$, právě když $x \in A$ nebo ($x \in A$ a $x \in B$)), což je podle pravidel výrokové logiky právě když $x \in A$.

$A \cap (A \cup B) = A$: Podobně jako předchozí. □

Vidíme tedy, že řadu vlastností operací s množinami dostaneme jednoduše z odpovídajících pravidel výrokové logiky. Podívejme se ještě jednou na důkaz tvrzení $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. Z výrokové logiky víme, že formule $p \wedge (q \vee r)$ je ekvivalentní formuli $(p \wedge q) \vee (p \wedge r)$, tj. tyto formule mají při každém ohodnocení stejnou pravdivostní hodnotu. Vezmeme-li ohodnocení, které výrokovým symbolům p, q a r přiřazují po řadě pravdivostní hodnoty tvrzení $x \in A$, $x \in B$, $x \in C$, pak při tomto ohodnocení má formule $p \wedge (q \vee r)$ stejnou pravdivostní hodnotu jako tvrzení $x \in A \cap (B \cup C)$ (podívejte se do výše napsaného důkazu) a formule $(p \wedge q) \vee (p \wedge r)$ má stejnou pravdivostní hodnotu jako tvrzení $x \in (A \cap B) \cup (A \cap C)$. Proto je $x \in A \cap (B \cup C)$, právě když $x \in (A \cap B) \cup (A \cap C)$, tedy $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Průvodce studiem

Jedním z jednoduchých ale užitečných přínosů teorie množin je, že nám dává prostředky jednoduše a jednoznačně se vyjadřovat. Bez množinového formalismu bychom vše museli vyjadřovat opisem. K jednoduchosti: Zkuste např. opisem (tj. v přirozeném jazyku, bez množinového formalismu) popsat množinu ($A \cap (B \cup$

$(A \cap D))) \cup (B \cup E)$. K jednoznačnosti: Řekneme-li „seskupení sudých a lichých čísel“, máme nejspíš na mysli sjednocení množiny sudých a množiny lichých čísel. Řekneme-li ale „soubor malých a zelených mužíčků“, máme asi na mysli soubor těch mužíčků, kteří jsou zároveň malí a zelení (průnik množiny malých mužíčků a množiny zelených mužíčků), ale můžeme mít na mysli i soubor těch mužíčků, kteří jsou malí nebo zelení (sjednocení množiny malých mužíčků a množiny zelených mužíčků). Co přesně máme na mysli, vyplývá z kontextu nebo to musíme upřesnit. Množinový formalismus je naproti tomu jednoznačný.

Shrnutí

Množiny, relace a funkce patří k základním pojmům matematiky. Množina je matematický pojem, který je protějškem běžně používaného pojmu soubor, seskupení apod. Relace je protějškem pojmu vztah. Funkce je protějškem pojmu přiřazení.
Množina je dána tím, jaké prvky obsahuje. Speciální množinou je prázdná množina, ta neobsahuje žádný prvek. S množinami můžeme provádět různé operace. Mezi základní patří průnik, sjednocení a rozdíl. Základní vztah mezi množinami je vztah inkluze (být podmnožinou). Množiny zapisujeme nejčastěji výčtem prvků nebo udáním charakteristické vlastnosti.

Pojmy k zapamatování

- množina,
- inkluze,
- průnik, sjednocení, rozdíl,
- potenční množina.

Kontrolní otázky

1. *Může množina obsahovat daný prvek více než jedenkrát? Proč? Jsou množiny* $\{a, b\}$ *a* $\{b, a\}$ *různé? Proč?*
2. *Jaké znáte způsoby zápisu množin? Jsou množiny* $\{x \in \mathbb{R} \mid x^2 < 0\}$ *a* $\{x \in \mathbb{N} \mid x^4 < 0\}$ *stejné? Je některá z nich rovna* $\emptyset$*?*
3. *Platí, že když* $A \subseteq B$*, pak* $|A| = |B|$*? Co je to potenční množina dané množiny? Existuje množina, jejíž potenční množina je prázdná?*
4. *Jaké znáte množinové operace? Jaká je nutná a postačující podmínka pro to, aby* $A \cap X = A$*? Jaká pro* $A \cup X = A$*?*

Cvičení

1. Platí následující tvrzení?
   a) $\emptyset \subseteq \emptyset$
   b) $\emptyset \in \emptyset$
   c) $\{a\} \in \{a, b, c\}$
   d) $\{a\} \in \{\{a, b\}, c\}$
   e) $\{a, b\} \subseteq \{a, \{a, b\}\}$
   f) $\{a, b\} \subseteq \{a, b, \{a, b\}\}$

g) $A \in 2^X$

2. Nechť $A = \{a, 1, \{a, b\}\}$, $B = \{2, a, \{a\}\}$, $C = \{\emptyset, 2, 3, \{a, b\}\}$. Určete $A \cup B$, $A \cap C$, $C - A$, 2^B.
3. Nechť $A = \{a, \{b\}\}$, $B = \{a, b, \{a, b\}\}$. Určete $B \cap 2^A$, $(A \times B) \cap (B \times A)$.
4. Určete $2^\emptyset$, $2^{\{\emptyset\}}$, $2^{\{1\}}$, $2^{\{\{\emptyset\}\}}$, $2^{\{\emptyset,\{\emptyset\}\}}$.
5. Definujme operaci $\oplus$ vztahem

$$A \oplus B = (A \cup B) - (A \cap B).$$

Zjistěte, zda platí následující vztahy (vztahy dokažte nebo nalezněte protipříklady.)

a) $A \oplus A = A$

b) $A \oplus (B \cap C) = (A \oplus B) \cap (A \oplus C)$

c) $A \cap (B \oplus C) = (A \cap B) \oplus (A \cap C)$

d) $A \oplus (A \oplus A) = A$

e) $A \subseteq B \Rightarrow A \oplus C \subseteq B \oplus C$

6. Najděte nutnou a postačující podmínku pro to, aby a) $A \oplus B = A \cup B$, b) $A \oplus B = A$.
7. Najděte příklady množin A, B, C tak, aby platilo

a) $A \cap B = C \cap B$, ale $A \neq C \neq \emptyset$

b) $A \cap B \subset A \cap C$, ale $B \not\subseteq C$

c) $A \cup B = C \cup B$, ale $A \neq C$

d) $A \cup B \subset A \cup C$, ale $B \not\subseteq C$

8. Nechť pro množiny A, B, C platí $B \subset A \subset C$. Určete množinu X, pro kterou $A - X = B$ a $A \cup X = C$.

Úkoly k textu

1. Zdůvodněte (přesně podle definice), proč je prázdná množina podmnožinou každé množiny.
2. Může mít potenční množina množiny A méně prvků než množina A? Může jich mít stejně? Může jich mít více?
3. Jaké vztahy platí mezi množinou A a 2^A?

Řešení

1. a) ano, b) ne, c) ne, d) ne, e) ne, f) ano, g) ano.
2. $A \cup B = \{a, 1, 2, \{a\}, \{a, b\}\}$, $A \cap C = \{\{a, b\}\}$, $C - A = \{\emptyset, 2, 3\}$, $2^B = \{\emptyset, \{2\}, \{a\}, \{\{a\}\}, \{2, a\}, \{2, \{a\}\}, \{a, \{\}\}, \{2, a, \{a\}\}\}$.
3. $B \cap P(A) = \emptyset$, $(A \times B) \cap (B \times A) = \{\langle a, a\rangle\}$.
4. $2^\emptyset = \{\emptyset\}$, $2^{\{\emptyset\}} = \{\emptyset, \{\emptyset\}\}$, $2^{\{1\}} = \{\emptyset, \{1\}\}$, $2^{\{\{\emptyset\}\}} = \{\{\{\emptyset\}\}, \emptyset\}$, $2^{\{\emptyset,\{\emptyset\}\}} = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$.
5. a) ne, b) ano, c) ano, d) ano, e) ne
6. a) $A \cap B = \emptyset$, b) $B = \emptyset$.

7. a) $A = \{a, b, c\}$, $B = \{a, c, d\}$, $C = \{a, c\}$, b) $A = \{a, c\}$, $B = \{a, b\}$, $C = \{a, c\}$, c) $A = \{a\}$, $B = \{a, b, c\}$, $C = \{c\}$, d) $A = \{a, b\}$, $B = \{a, c\}$, $C = \{b, c, d\}$.
8. $X = C - B$ (jiné nejsou).

Studijní cíle: Po prostudování kapitol 2.3 a 2.4 by student měl rozumět pojmům relace a funkce. Měl by znát základní operace a vztahy definované nad těmito pojmy. Student by měl tyto pojmy znát aktivně, měl by umět samostatně dokázat jednoduchá tvrzení, hledat příklady a protipříklady.

Klíčová slova: kartézský součin, relace, reprezentace relací, inverzní relace, skládání relací, funkce, injekce, surjekce, bijekce

2.3 Relace

2.3.1 Pojem relace

Pojem relace je matematickým protějškem běžně používaného pojmu *vztah*. Různé objekty jsou nebo nejsou v různých vztazích. Např. číslo 3 je ve vztahu „být menší“ s číslem 5, ne však s číslem 2. Karel Čapek byl ve vztahu „být bratrem“ s Josefem Čapkem. Tři body v rovině mohou být ve vztahu „ležet na jedné přímce“.

Všimněme si, čím je vztah určen. Za prvé je to tzv. arita vztahu, tj. číslo udávající počet objektů, které do vztahu vstupují. Např. do vztahu „být bratrem“ vstupují dva objekty, ten vztah je binární, do vztahu „ležet na jedné přímce“ vstupují tři objekty, ten vztah je ternární. Za druhé jsou to množiny, jejichž prvky do vztahu vstupují. Např. do vztahu „být bratrem“ vstupují dva objekty, první je z množiny X_1 lidí, druhý je z množiny X_2 lidí. V tomto případě jsou X_1 a X_2 stejné, tj. $X_1 = X_2$. To tak ale nemusí být. Uvažujme např. vztah „mít“ mezi množinou X_1 nějakých objektů a množinou X_2 nějakých atributů. V tomto případě je obecně $X_1 \neq X_2$, např. $X_1 = \{$pes, kočka, běhat, stůl, rychlý, zelený, číst$\}$ a $X_2 = \{$„je podstatné jméno“, „je sloveso“$\}$. Je-li dána arita n a příslušné množiny $X_1, \ldots, X_n$, vztah je potom určen tím, které prvky x_1 z X_1, …, x_n z X_n v tom vztahu jsou a které ne. To nás přivádí k pojmu relace.

Základním pojmem je pojem uspořádané n-tice prvků. *Uspořádaná n-tice* objektů $x_1, \ldots, x_n$ (v tomto pořadí) se označuje $\langle x_1, \ldots, x_n \rangle$. Prvek x_i $(1 \leq i \leq n)$ se nazývá i-tá složka dané n-tice. Rovnost definujeme tak, že $\langle x_1, \ldots, x_n \rangle = \langle y_1, \ldots, y_m \rangle$, právě když $n = m$ a $x_1 = y_1$, …, $x_n = y_n$. n-tice a m-tice jsou si tedy rovny, právě když mají stejný počet složek a odpovídající si složky jsou stejné.

Definice 2.11. *Kartézský součin* množin $X_1, \ldots, X_n$ je množina $X_1 \times \cdots \times X_n$ definovaná předpisem

$$X_1 \times \cdots \times X_n = \{\langle x_1, \ldots, x_n \rangle \mid x_1 \in X_1, \ldots, x_n \in X_n\}.$$

Kartézský součin n množin je množina všech uspořádaných n-tic prvků z těchto množin.

Je-li $X_1 = \cdots = X_n = X$, pak $X_1 \times \cdots \times X_n$ značíme také X^n (n-tá kartézská mocnina množiny X). Uspořádanou 1-tici $\langle x \rangle$ obvykle ztotožňujeme s prvkem x (tj. $\langle x \rangle = x$). Potom X^1 je vlastně množina X.

Příklad 2.12.
- Pro $A = \{a, b, c\}$, $B = \{1, 2\}$ je $A \times B = \{\langle a, 1 \rangle, \langle a, 2 \rangle, \langle b, 1 \rangle, \langle b, 2 \rangle, \langle c, 1 \rangle, \langle c, 2 \rangle\}$, $B^2 = \{\langle 1, 1 \rangle, \langle 1, 2 \rangle, \langle 2, 1 \rangle, \langle 2, 2 \rangle\}$,
- pro $A = \{\{a\}, b\}$, $B = \{1\}$ je $A \times B = \{\langle \{a\}, 1 \rangle, \langle b, 1 \rangle\}$,
- pro $A = \{\emptyset, \{\emptyset\}\}$, $B = \{\emptyset\}$ je $A \times B = \{\langle \emptyset, \emptyset \rangle, \langle \{\emptyset\}, \emptyset \rangle\}$,
- pro $A = \{1, 2\}$, $B = \{b\}$ je $A \times B \times A = \{\langle 1, b, 1 \rangle, \langle 1, b, 2 \rangle, \langle 2, b, 1 \rangle, \langle 2, b, 2 \rangle\}$,

- pro $A = \emptyset$, $B = \{1, 2, 3\}$ je $A \times B = \emptyset$ (neexistuje totiž uspořádaná dvojice $\langle x, y\rangle$ tak, aby $x \in A$ a $y \in B$).

Můžeme přistoupit k definici pojmu relace.

Definice 2.13. Nechť $X_1, \ldots, X_n$ jsou množiny. *Relace* mezi $X_1, \ldots, X_n$ je libovolná podmnožina kartézského součinu $X_1 \times \cdots \times X_n$.

Relace mezi množinami $X_1, \ldots, X_n$ je podmnožina kartézského součinu $X_1 \times \cdots \times X_n$.

Poznámka 2.14. (1) Číslu n říkáme arita relace R, R se nazývá n-ární. Je-li $X_1 = \cdots = X_n = X$, nazývá se R také n-ární relace v množině X. Pro $n = 1, 2, 3, 4$ se místo n-ární používá také unární, binární, ternární, kvaternární. To, že R je unární relace v X, vlastně znamená, že $R \subseteq X$.

(2) O prvcích $x_1 \in X_1, \ldots, x_n \in X_n$ říkáme, že jsou (v tomto pořadí) v relaci R, pokud $\langle x_1, \ldots, x_n\rangle \in R$.

Relace je tedy množina sestávající z n-tic prvků příslušných množin. Obsahuje ty n-tice $\langle x_1, \ldots, x_n\rangle$, které mezi sebou mají zamýšlený vztah. Ty, které zamýšlený vztah nemají, neobsahuje. Běžně používaný, avšak jen intuitivně chápaný, pojem vztah je tedy pojmem relace matematizován. Pojem relace je přitom založen na pojmech množina a uspořádaná n-tice.

Příklad 2.15. (1) Pro $X = \{a, b, c\}$, $Y = \{1, 2, 3, 4\}$ jsou $\{\langle a, 2\rangle, \langle a, 3\rangle, \langle b, 1\rangle, \langle c, 1\rangle\}$, $\{\langle a, 2\rangle\}$, $\emptyset$, $X \times Y$ binární relace mezi X a Y. $\{\langle a, b, 2, 4, c\rangle, \langle a, a, 2, 2, a\rangle\}$ je relace mezi X, X, Y, Y, X. $\{\langle a, 1\rangle, \langle 2, c\rangle\}$ není binární relace mezi X a Y, protože dvojice $\langle 2, c\rangle$ nepatří do kartézského součinu $X \times Y$.

(2) Předpokládejme, že na rodinné oslavě jsou Adam (A), Bedřich (B), Cyril (C), Dominik (D), Egon (E), Marta (M), Naďa (N), Olga (O), Pavla (P) a Radka (R). Přitom Adam je synem Cyrila a Marty, Cyril je synem Egona a Olgy, Pavla je dcerou Dominka, Egon je synem Adama a Radky. Určete binární relaci R, která odpovídá vztahu „X je dítětem Y“, a ternární relaci S, která odpovídá vztahu „X je dítětem Y a Z“, kde Y je otec a Z je matka.

Jde o relace na množině {A, B, C, D, E, M, N, O, P, R}. R bude obsahovat všechny uspořádané dvojice $\langle x, y\rangle$ takové, že x je dítětem y. Tedy

$$R = \{\langle \mathrm{A}, \mathrm{C}\rangle, \langle \mathrm{A}, \mathrm{M}\rangle, \langle \mathrm{C}, \mathrm{E}\rangle, \langle \mathrm{C}, \mathrm{O}\rangle, \langle \mathrm{P}, \mathrm{D}\rangle, \langle \mathrm{E}, \mathrm{A}\rangle, \langle \mathrm{E}, \mathrm{R}\rangle\}$$

a

$$S = \{\langle \mathrm{A}, \mathrm{C}, \mathrm{M}\rangle, \langle \mathrm{C}, \mathrm{E}, \mathrm{O}\rangle, \langle \mathrm{E}, \mathrm{A}, \mathrm{R}\rangle\}.$$

(3) Platí-li navíc, že C je manželem M, E je manželem O a A je manželem R, můžeme uvažovat binární relaci T mezi množinou $X = \{$A, B, C, D, E$\}$ mužů a množinou $Y = \{$ M, N, O, P, R$\}$ žen, tj. $T \subseteq X \times Y$, která odpovídá vztahu „být manželem“. Pak bude

$$T = \{\langle \mathrm{C}, \mathrm{M}\rangle, \langle \mathrm{E}, \mathrm{O}\rangle, \langle \mathrm{A}, \mathrm{R}\rangle\}.$$

(4) Zapište jako binární relaci vztah dělitelnosti (tj. „x dělí y“ znamená, že existuje celé číslo k tak, že $x \cdot k = y$) na množině $X = \{2, \ldots, 10\}$.

Označme příslušnou relaci D. Je tedy $D \subseteq X \times X$, konkrétně

$$D = \{\langle 2, 2\rangle, \langle 2, 4\rangle, \langle 2, 6\rangle, \langle 2, 8\rangle, \langle 2, 10\rangle, \langle 3, 6\rangle, \langle 3, 9\rangle, \langle 4, 8\rangle, \langle 5, 10\rangle\}.$$

příjmení	jméno	narození	vzdělání	funkce
Adam	Jiří	1976	SŠ	prodejce
Kos	Jan	1961	VŠ	projektant
Malá	Magda	1955	SŠ	sekretářka
Rychlý	Karel	1967	VŠ	ředitel
⋮	⋮	⋮	⋮	⋮
Zahradník	Milan	1950	ZŠ	technik

Tabulka 12: Databáze z Příkladu 2.16.

Průvodce studiem

Zastavme se u pojmu relace. Podle definice 2.13 je relace podmnožina kartézského součinu. Dává to smysl? Relace má být matematickým protějškem pojmu vztah, který je přece každému jasný. Naproti tomu „podmnožina kartézského součinu“ zní nepřístupně a zbytečně komplikovaně. Pokud souhlasíte s předchozími dvěma větami, bude nejlepší, když si zkusíte sami navrhnout definici pojmu relace. Uvidíte, jestli přijdete na něco lepšího než je definice 2.13. Přitom ale dodržte „pravidla hry“: Vaše definice musí být jednoznačná (tj. musí být založana na jednoznačně definovaných pojmech) a musí být tak obecná, aby odpovídala pojmu vztah (tj. nemůžete se např. omezit jen na binární relace). Např. definice „Relace je dána tím, které prvky jsou v relaci se kterými.“ neobstojí. Je značně neurčitá a navíc je to definice kruhem (v definici pojmu relace se odkazujeme na pojme relace). Zkuste si představit, že podle této definice máte rozhodnout, zde něco je nebo není relace. Že je třeba, aby definice relace byla jednoznačná a jednoduchá vynikne nejlépe, když si uvědomíme, že relace můžeme chtít zpracovávat počítačem (a počítačových aplikací založených na relacích je celá řada). Předkládáme-li nejednoznačnou definici člověku, může nám to projít, ten člověk si definici třeba domyslí. U počítače nám to neprojde, počítač si nic nedomyslí. Kromě toho, jednoznačnost a jednoduchost definice patří k základům kultury vyjadřování nejen v matematice. Nejste-li tedy spokojeni s Definicí 2.13, zkuste teď sami navrhnout lepší a pak pokračujte ve čtení.

Porovnejte nyní váš návrh s Definicí 2.13 (nejlépe s kolegy nebo učitelem). Pokud jste lepší definici nevymysleli, vraťte se k Definici 2.13 a znovu ji posuďte.

Příklad 2.16. Pojem relace má ústřední roli v tzv. relačním databázovém modelu, který navrhl E. F. Codd.[5] Tzv. relační pohled na databáze spočívá v tom, že databázi, resp. databázovou tabulku, chápeme jako relaci. Např. databázi znázorněnou tab. 12, která obsahuje v řádcích informace o zaměstnancích, můžeme chápat jako 5-ární relaci R mezi množinami (těm se v databázích říká domény) $D_1 = \{\text{Adam, Kos, Malá, Rychlý, } \dots\}$, $D_2 = \{\text{Jiří, Jan, Magda, Karel, } \dots\}$, $D_3 = \{n \in \mathbb{N} \mid 1900 \leq n \leq 2004\}$, $D_4 = \{\text{ZŠ, SOU, SŠ, VŠ}\}$, $D_5 = \{\text{prodejce, projektant, sekretářka, ředitel, } \dots\}$, tedy $R \subseteq D_1 \times D_2 \times D_3 \times D_4 \times D_5$. Relace R je dána záznamy (řádky) v databázi, takže např. $\langle$ Adam, Jiří, 1976, SŠ, prodejce$\rangle \in R$. Je tedy:

$$\begin{aligned} R \;=\; \{&\langle\text{Adam, Jiří, 1976, SŠ, prodejce}\rangle, \langle\text{Kos, Jan, 1961, VŠ, projektant}\rangle,\\ &\langle\text{Malá, Magda, 1955, SŠ, sekretářka}\rangle, \langle\text{Rychlý, Karel, 1967, VŠ, ředitel}\rangle,\\ &\vdots\\ &\langle\text{Zahradník, Milan, 1950, ZŠ, technik}\rangle\} \end{aligned}$$

V relačních databázích jsou zavedeny i jiné operace než ty, které zavedeme my. Tyto

[5] Pěkně je o tom napsáno v knize C. J. Date: *The Database Relational Model: A Retrospective Analysis*. Addison Wesley, Reading, MA, 2001.

R	1	2	3	4
a	×	×		×
b		×		×
c	×			

Tabulka 13: Tabulka popisující binární relaci R mezi $X = \{a, b, c\}$ a $Y = \{1, 2, 3, 4\}$.

operace slouží k manipulaci a zpřístupňování dat v databázi a čtenář se s nimi může seznámit téměř v každé učebnici databázových systémů.

2.3.2 Vztahy a operace s relacemi

Relace jsou množiny (relace je podmnožina kartézského součinu). Proto s nimi lze provádět množinové operace ($\cap$, $\cup$, $-$) a lze na ně aplikovat vztah inkluze ($\subseteq$).

Relace jsou speciální množiny, a proto s nimi můžeme provádět všechny množinové operace.

Příklad 2.17. (1) Mějme $X = \{a, b, c\}$, $Y = \{1, 2, 3, 4\}$ a uvažujme binární relace $R = \{\langle a, 1\rangle, \langle a, 4\rangle, \langle c, 2\rangle, \langle c, 3\rangle, \langle c, 4\rangle\}$, $S = \{\langle a, 2\rangle, \langle a, 3\rangle, \langle a, 4\rangle, \langle b, 1\rangle, \langle b, 3\rangle\}$, $T = \{\langle a, 4\rangle, \langle c, 4\rangle\}$ mezi X a Y. Pak je např.

$$\begin{aligned} R \cap S &= \{\langle a, 4\rangle\}, \\ R \cup S &= \{\langle a, 1\rangle, \langle a, 2\rangle, \langle a, 3\rangle, \langle a, 4\rangle, \langle b, 1\rangle, \langle b, 3\rangle, \langle c, 2\rangle, \langle c, 3\rangle, \langle c, 4\rangle\}. \end{aligned}$$

Dále je $T \subseteq S$, $R \not\subseteq S$ apod.

(2) Nechť $\leq$ je relace uspořádání a $|$ relace dělitelnosti na množině $\mathbb{N}$ přirozených čísel. Tedy $\langle k, l\rangle \in \leq$, právě když k je menší nebo rovno l, a $\langle k, l\rangle \in |$, právě když l je dělitelné číslem k (v tomto případě, jako i u jiných případů binárních relací, běžně používáme tzv. infixovou notaci, tj. píšeme $k \leq l$ a $k|l$). Pak $| \subseteq \leq$, tj. relace $|$ je podmnožinou relace $\leq$. To vlastně znamená, že pro všechna přirozená čísla $k, l \in \mathbb{N}$ platí, že když $k|l$, pak $k \leq l$.

(3) Jsou-li R_1 a R_2 relace popisující nějaké databázové tabulky (viz Příklad 2.16), pak $R_1 \cup R_2$ je relace popisující databázovou tabulku, která vznikne sloučením (zřetězením) výchozích tabulek (přesně vzato, sloučením a vymazáním duplicitních výskytů databázových řádků). $R_1 \cap R_2$ je relace, která popisuje společné položky obou tabulek.

2.3.3 Operace s binárními relacemi

S relacemi však lze díky jejich speciální struktuře provádět i další operace. Zaměříme se na binární relace. Ty lze znázorňovat tabulkami. Např. relace $R = \{\langle a, 1\rangle, \langle a, 2\rangle, \langle a, 4\rangle, \langle b, 2\rangle, \langle b, 4\rangle, \langle c, 1\rangle\}$ mezi množinami $X = \{a, b, c\}$ a $Y = \{1, 2, 3, 4\}$ je znázorněna v Tab. 14. Tedy, je-li $\langle x, y\rangle \in R$, je v průsečíku řádku x a sloupce y symbol $\times$, jinak tam není nic.

Začneme tzv. inverzní relací. *Inverzní relací* k relaci $R \subseteq X \times Y$ je relace R^{-1} mezi Y a X definovaná předpisem

S binárními relacemi lze navíc provádět operace inverze a skládání.

$$R^{-1} = \{\langle y, x\rangle \mid \langle x, y\rangle \in R\}.$$

Příklad 2.18. Nechť relace R mezi $X = \{a, b, c\}$ a $Y = \{1, 2, 3\}$ je $R = \{\langle a, 1\rangle, \langle a, 2\rangle, \langle b, 2\rangle\}$. Pak inverzní relace k R je relace R^{-1} mezi Y a X daná $R^{-1} = \{\langle 1, a\rangle, \langle 2, a\rangle, \langle 2, b\rangle\}$.

Další operací je tzv. skládání. Je-li R relací mezi množinami X a Y a S relací mezi množinami Y a Z, pak *složením* relací R a S je relace $R \circ S$ mezi X a Z definovaná předpisem

$$R \circ S = \{\langle x, z\rangle \mid \text{existuje } y \in Y : \langle x, y\rangle \in R \text{ a } \langle y, z\rangle \in S\}.$$

R	b	h	k	o	r	s	v	z
1		×			×			
2		×						
3	×	×	×	×	×	×	×	×
4							×	
5		×			×		×	
6	×	×	×		×			
7	×	×				×		×

S	C	M	N	Sp	Za
b	×	×			
h	×	×		×	
k	×				
o					×
r	×			×	
s		×			
v			×	×	×
z		×			

Tabulka 14: K Příkladu 2.20: Tabulky popisující binární relaci R mezi pacienty a příznaky nemocí (vlevo) a relaci S příznaky nemocí a nemocemi (vpravo).

Tedy $\langle x, z\rangle$ patří do relace $R \circ S$, právě když existuje prvek $y \in Y$ tak, že $\langle x, y\rangle$ jsou v relaci R a $\langle y, z\rangle$ jsou v relaci S. Prvek y hraje roli prostředníka mezi x a z (y zprostředkuje vztah x a z).

Příklad 2.19. Nechť $X = \{a, b, c\}$, $Y = \{1, 2, 3, 4\}$ a $Z = \{\square, \triangle\}$. Uvažujme relace

$$R = \{\langle a, 1\rangle, \langle a, 2\rangle, \langle b, 2\rangle, \langle c, 4\rangle\} \text{ a } S = \{\langle 1, \square\rangle, \langle 1, \triangle\rangle, \langle 2, \square\rangle, \langle 3, \triangle\rangle, \langle 4, \square\rangle, \langle 4, \triangle\rangle\}.$$

Pak

$$R \circ S = \{\langle a, \square\rangle, \langle a, \triangle\rangle, \langle b, \square\rangle, \langle c, \square\rangle, \langle c, \triangle\rangle\}.$$

Například $\langle a, \square\rangle \in R \circ S$, protože $\langle a, 1\rangle \in R$ a $\langle 1, \square\rangle \in S$; prvek 1 je tedy prostředníkem. Jiným prostředníkem je prvek 2, neboť $\langle a, 2\rangle \in R$ i $\langle 2, \square\rangle \in S$. Dvojice $\langle b, \triangle\rangle$ ale do relace $R \circ S$ nepatří, protože neexistuje žádný prvek $y \in Y$, pro který $\langle y, \triangle\rangle \in S$.

Skládání relací je přirozená operace, kterou běžně používáme. Uvažujme následující příklad. Nechť X je množina pacientů, Y množina příznaků nemocí a Z množina nemocí. Nechť $R \subseteq X \times Y$ je relace „mít příznak“, tj. $\langle x, y\rangle \in R$ znamená, že pacient x má příznak y, a $S \subseteq Y \times Z$ je relace „být příznakem“, tj. $\langle y, z\rangle \in S$ znamená, že y je příznakem nemoci z (např. zvýšená teplota je příznakem chřipky). Pak pro pacienta $x \in X$ a nemoc $z \in Z$ znamená $\langle x, z\rangle \in R \circ S$, že existuje příznak $y \in Y$ tak, že pacient x má tento příznak a zároveň je tento příznak příznakem nemoci z. Tedy $\langle x, z\rangle \in R \circ S$ můžeme interpretovat jako „pacient x může mít nemoc z“.

Příklad 2.20. Nechť $X = \{1, 2, 3, 4, 5, 6, 7\}$ (X reprezentuje pacienty 1–7), $Y = \{b, h, k, o, r, s, v, z\}$ (b ... bolest hlavy, h ... horečka, k ... bolest končetin, o ... oteklé žlázy na krku, r ... rýma, s ... strnulý krk, v ... vyrážka, z ... zvracení), $Z = \{C, M, N, Sp, Za\}$ (C ... chřipka, M ... meningitida, N ... plané neštovice, Sp ... spalničky, Za ... zarděnky). Vztah „mít příznak“ mezi pacienty a příznaky je popsán relací $R \subseteq X \times Y$ znázorněnou v tab. 14 vlevo, vztah „být příznakem nemoci“ mezi příznaky a nemocemi je popsán relací $S \subseteq Y \times Z$ znázorněnou v tab. 14 vpravo. Složení relací R a S je relace $R \circ S \subseteq X \times Z$ znázorněná v tab. 15. Protože $\langle x, z\rangle \in R \circ S$ můžeme chápat tak, že pacient x může mít nemoc z, můžeme se na příklad dívat následovně. Ze vstupních informací R (dáno lékařským vyšetřením) a S (dáno znalostí lékaře) jsme odvodili nové informace reprezentované relací $R \circ S$. Ty říkají, že např. pacient 1 může mít chřipku, meningitidu nebo spalničky, že pacient 3 může mít libovolnou z uvažovaných nemocí (má všechny sledované příznaky), pacient 6 může mít libovolnou z uvažovaných nemocí (přestože nemá všechny sledované příznaky) atd.

Věta 2.21. *Pro relace $R \subseteq X \times Y$, $S \subseteq Y \times Z$, $T \subseteq Z \times U$ platí.*

$$\begin{aligned} R \circ (S \circ T) &= (R \circ S) \circ T \\ (R \circ S)^{-1} &= S^{-1} \circ R^{-1} \\ (R^{-1})^{-1} &= R \end{aligned}$$

$R \circ S$	C	M	N	Sp	Za
1	×	×		×	
2	×	×			
3	×	×	×	×	×
4			×	×	×
5	×	×	×	×	×
6	×	×		×	
7	×	×		×	

Tabulka 15: Tabulka popisující binární relaci $R \circ S$ mezi pacienty a nemocemi (viz Příklad 2.20).

Důkaz. $R \circ (S \circ T) = (R \circ S) \circ T$: Máme $\langle x, u\rangle \in R \circ (S \circ T)$, právě když existuje $y \in Y$ tak, že $\langle x, y\rangle \in R$ a $\langle y, u\rangle \in S \circ T$, právě když existuje $y \in Y$ tak, že $\langle x, y\rangle \in R$ a existuje $z \in Z$ tak, že $\langle y, z\rangle \in S$ a $\langle z, u\rangle \in T$, právě když existují $y \in Y$ a $z \in Z$ tak, že $\langle x, y\rangle \in R$, $\langle y, z\rangle \in S$, $\langle z, u\rangle \in T$, právě když existuje $z \in Z$ tak, že $\langle x, z\rangle \in R \circ S$ a $\langle z, u\rangle \in T$, právě když $\langle x, u\rangle \in (R \circ S) \circ T$.

$(R \circ S)^{-1} = S^{-1} \circ R^{-1}$: $\langle z, x\rangle \in (R \circ S)^{-1}$, právě když $\langle x, z\rangle \in (R \circ S)$, právě když existuje $y \in Y$ tak, že $\langle x, y\rangle \in R$ a $\langle y, z\rangle \in S$, právě když existuje $y \in Y$ tak, že $\langle z, y\rangle \in S^{-1}$ a $\langle y, x\rangle \in R^{-1}$, právě když $\langle z, x\rangle \in S^{-1} \circ R^{-1}$.

$(R^{-1})^{-1} = R$: $\langle x, y\rangle \in (R^{-1})^{-1}$, právě když $\langle y, x\rangle \in R^{-1}$, právě když $\langle x, y\rangle \in R$. □

Způsobů, jak skládat relace, existuje více.

Existují však i další přirozené způsoby, jak skládat relace. Předpokládejme opět, že $R \subseteq X \times Y$, $S \subseteq Y \times Z$. Pak $R \lhd S$, $R \rhd S$ a $R \,\square\, S$ jsou relace mezi X a Z definované předpisy

$$\begin{aligned} R \lhd S &= \{\langle x, z\rangle \mid \text{pro každé } y \in Y : \text{ pokud } \langle x, y\rangle \in R, \text{ pak } \langle y, z\rangle \in S\}, \\ R \rhd S &= \{\langle x, z\rangle \mid \text{pro každé } y \in Y : \text{ pokud } \langle y, z\rangle \in S, \text{ pak } \langle x, y\rangle \in R\}, \\ R \,\square\, S &= \{\langle x, z\rangle \mid \text{pro každé } y \in Y : \langle x, y\rangle \in R, \text{ právě když } \langle y, z\rangle \in S\}. \end{aligned}$$

Vraťme se k příkladu s pacienty, příznaky a nemocemi. $\langle x, z\rangle \in R \lhd S$ znamená, že všechny příznaky, které má pacient x, jsou příznaky nemoci z. $\langle x, z\rangle \in R \rhd S$ znamená, že pacient x má všechny příznaky nemoci z. $\langle x, z\rangle \in R \,\square\, S$ znamená, že pacient x má právě příznaky nemoci z. Uvědomme si, že relace R může vzniknout na základě lékařského vyšetření (lékař zjišťuje, jaké příznaky pacienti mají) a že relace S je „učebnicová znalost“ (lékařské knihy popisují příznaky jednotlivých nemocí). Obě R i S tedy mohou být dostupné např. v databázi. Všechna složení $R \circ S$, $R \lhd S$, $R \rhd S$ i $R \,\square\, S$ je pak možné z R a S jednoduše spočítat. Tyto relace poskytují netriviální informace o tom, kteří pacienti mohou mít které nemoci. Přitom pro daného pacienta x a danou nemoc y má každý z faktů $\langle x, z\rangle \in R \circ S$, $\langle x, z\rangle \in R \lhd S$, $\langle x, z\rangle \in R \rhd S$ i $\langle x, z\rangle \in R \,\square\, S$ přesně stanovený význam. Přitom nejslabší indikací toho, že pacient x má nemoc z je fakt $\langle x, z\rangle \in R \circ S$ (x má aspoň jeden příznak nemoci z), nejsilnější naopak fakt $\langle x, z\rangle \in R \,\square\, S$ (x má právě všechny příznaky nemoci z). Jak je vidět přímo z definice (rozmyslete si), relace $\lhd$ i $\rhd$ obsahují jako podmnožinu relaci $\square$; ta je jejich průnikem.

Příklad 2.22. Vraťme se k Příkladu 2.20. Relace $R \lhd S$, $R \rhd S$ a $R \,\square\, S$ jsou znázorněny v Tab. 16.

2.3.4 Binární relace a jejich reprezentace

$R \lhd S$	C	M	N	S	Za
1	×			×	
2	×	×		×	
3					
4			×	×	×
5				×	
6	×				
7		×	×	×	×

$R \rhd S$	C	M	N	S	Za
1					
2					
3	×	×	×	×	×
4			×		
5			×	×	
6	×				
7		×			

$R \square S$	C	M	N	S	Za
1					
2					
3					
4			×		
5				×	
6	×				
7		×			

Tabulka 16: Tabulka popisující binární relace $R \lhd S$, $R \rhd S$ a $R \square S$ mezi pacienty a nemocemi (viz Příklad 2.22).

Průvodce studiem

Chceme-li matematické pojmy zpracovávat v počítači, je třeba je vhodným způsobem v počítači reprezentovat. Musíme tedy navrhnout, jak by měl být matematický pojem (množina, relace apod.) v počítači (tj. v paměti počítače) uložen. Nejde ale jen o samotné uložení v paměti, nýbrž také o to, aby výpočty, které budou s danými pojmy prováděny, byly rychlé.

V této kapitole si ukážeme základní způsoby reprezentace binárních relací. Předpokládejme, že je dána binární relace R mezi konečnými množinami X a Y.

Matematické pojmy je třeba umět vhodně reprezentovat. Zvlášť důležitá je reprezentace v paměti počítače.

Reprezentace maticí (tabulkou)

Připomeňme, že matice typu $m \times n$ je obdélníkové schéma o m řádcích a n sloupcích, ve kterém se na každém místě odpovídajícím nějakému řádku a nějakému sloupci nachází nějaká (zpravidla číselná) hodnota. Označme takovou matici $\mathbf{M}$. Pro každé $i \in \{1, \ldots, m\}$ a $j \in \{1, \ldots, n\}$ označme m_{ij} prvek matice z průsečíku řádku i a sloupce j.

Průvodce studiem

Matice typu $m \times n$ je to samé co tabulka o m řádcích a n sloupcích. Rozdíl je jen v tom, že matice mají specifický způsob zápisu a že s maticemi jsou definovány různé standardní operace. Pojem matice používají matematici a inženýři, zvlášť když se s údaji zanesenými v matici budou provádět další operace. Pojem tabulka používá každý, kdo chce přehledným způsobem zapsat údaje o nějakých položkách (viz tabulkové procesory, nabídkové katalogy apod.).

Tabulky a matice představují základní způsob reprezentace binárních relací. Nechť R je relace mezi množinami $X = \{x_1, \ldots, x_m\}$ a $Y = \{y_1, \ldots, y_n\}$ a předpokládejme, že

R	1	2	3	4
a	×	×		×
b		×		×
c	×			

$$\mathbf{M}_R = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} .$$

Tabulka 17: Tabulka (vlevo) a matice (vpravo) popisující binární relaci R mezi $X = \{a, b, c\}$ a $Y = \{1, 2, 3, 4\}$.

prvky těchto množin jsou očíslovány, jak uvádějí indexy u x_i a y_j. Relaci R reprezentujeme tabulkou/maticí, ve které se na místě odpovídajícím řádku i a sloupci j nachází hodnota, která určuje, zda dvojice $\langle x_i, y_j \rangle$ je v relaci R. Obvykle se používá 1 (popř. ×) k označení $\langle x_i, y_j \rangle \in R$ a 0 (popř. prázdné místo) k označení $\langle x_i, y_j \rangle \notin R$. Matice $\mathbf{M}_R$ reprezentující relaci $R \subseteq \{x_1, \ldots, x_m\} \times \{y_1, \ldots, y_n\}$ je definována předpisem

Relaci lze reprezentovat maticí (tabulkou).

$$m_{ij} = \begin{cases} 1 & \text{je-li } \langle x_i, y_j \rangle \in R, \\ 0 & \text{je-li } \langle x_i, y_j \rangle \notin R\,. \end{cases} \tag{1}$$

$\mathbf{M}_R$ se nazývá *matice relace* R. Naopak také, každá binární matice $\mathbf{M}$ typu $m \times n$, tj. matice s hodnotami 0 a 1, reprezentuje relaci mezi $X = \{x_1, \ldots, x_m\}$ a $Y = \{y_1, \ldots, y_n\}$.

Příklad 2.23. V Tab. 17 vidíme tabulkovou a maticovou reprezentaci relace

$$R = \{\langle a, 1\rangle, \langle a, 2\rangle, \langle a, 4\rangle, \langle b, 2\rangle, \langle b, 4\rangle, \langle c, 1\rangle\}$$

mezi $X = \{a, b, c\}$ a $Y = \{1, 2, 3, 4\}$.

Maticová reprezentace je názorná. Její nevýhodou je velká paměťová náročnost.

Výhodou této reprezentace je přehlednost a to, že zjistit, zda $\langle x_i, y_j \rangle \in R$, lze rychle. Nevýhodou je paměťová náročnost. Např. pro reprezentaci relace na množině X s 1000 prvky zabírá je odpovídající matice rozměru 1000×1000 a má tedy 1000000 políček. V případě, že každý prvek z X je v relaci s (průměrně) 3 prvky z Y, obsahuje matice 3 tisíce jedniček a zbytek (997 tisíc) jsou nuly. Přitom uchovávat nuly je zbytečné, stačilo by uchovat informaci o tom, kde mají být jedničky. Pro takové případy se používají jiné reprezentace.

Pro binární matice můžeme zavést operace, které odpovídají operacím s relacemi. Mějme binární matice $\mathbf{M}, \mathbf{N}$ typu $m \times n$ a matici $\mathbf{K}$ typu $n \times k$. Definujme následující operace.

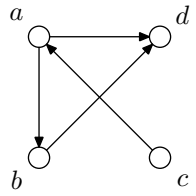
$$\begin{aligned}
&\mathbf{M} \vee \mathbf{N} = \mathbf{P}, && p_{ij} = \max\{m_{ij}, n_{ij}\} \\
&\mathbf{M} \wedge \mathbf{N} = \mathbf{P}, && p_{ij} = \min\{m_{ij}, n_{ij}\} \\
&\mathbf{M} - \mathbf{N} = \mathbf{P}, && p_{ij} = \max\{0, m_{ij} - n_{ij}\} \\
&\mathbf{M} \cdot \mathbf{K} = \mathbf{P}, && p_{ij} = \max\{m_{il} \cdot k_{lj};\, l = 1, \ldots, n\} \\
&\mathbf{M}^T, && m^T_{ij} = m_{ji}\ .
\end{aligned}$$

Například operace $\vee$ přiřazuje maticím $\mathbf{M}$ a $\mathbf{N}$ matici $\mathbf{P}$, jejíž každý prvek p_{ij} je roven minimu z hodnot m_{ij} a n_{ij}.

Operace s relacemi lze provádět pomocí vhodných operací s maticemi.

Věta 2.24. *Pro relace $R, S \subseteq X \times Y$, $U \subseteq Y \times Z$ je*

$$\begin{aligned}
\mathbf{M}_{R \cup S} &= \mathbf{M}_R \vee \mathbf{M}_S \\
\mathbf{M}_{R \cap S} &= \mathbf{M}_R \wedge \mathbf{M}_S \\
\mathbf{M}_{R - S} &= \mathbf{M}_R - \mathbf{M}_S \\
\mathbf{M}_{R \circ U} &= \mathbf{M}_R \cdot \mathbf{M}_U \\
\mathbf{M}_{R^{-1}} &= (\mathbf{M}_R)^T
\end{aligned}$$



Obrázek 2: Graf relace k Příkladu 2.26.

Důkaz. Důkaz je jednoduchý. Stačí porovnat definice operací s maticemi a definice operací s relacemi. □

Příklad 2.25. Na množině $X = \{a_1, a_2, a_3, a_4\}$ uvažujme relace $R = \mathrm{id}_X \cup \{\langle a_1, a_2\rangle, \langle a_1, a_3\rangle, \langle a_3, a_2\rangle\}$ a $S = \{\langle a_1, a_1\rangle, \langle a_2, a_4\rangle, \langle a_3, a_4\rangle, \langle a_4, a_1\rangle\}$. Přitom $\mathrm{id}_X = \{\langle a_1, a_1\rangle, \ldots, \langle a_6, a_6\rangle\}$. Matice těchto relací jsou

$$\mathbf{M}_R = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{a} \quad \mathbf{M}_S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} .$$

Matice relace $R \cup S$ je

$$\mathbf{M}_R \vee \mathbf{M}_S = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} .$$

Matice relace $R \cap S$ je

$$\mathbf{M}_R \wedge \mathbf{M}_S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

Matice relace $R \circ S$ je

$$\mathbf{M}_R \cdot \mathbf{M}_S = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} .$$

Matice relace R_{-1} je

$$(\mathbf{M}_R)^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

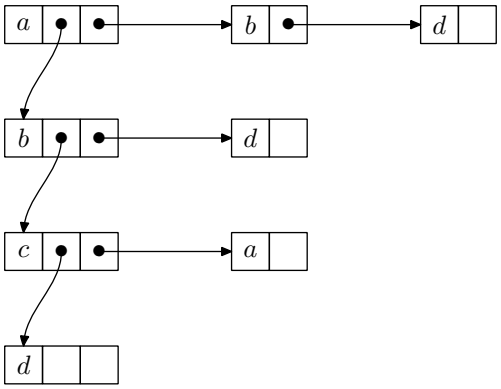
Reprezentace grafem

Relace na množině lze graficky znázornit pomocí tzv. grafů.

Grafy představují další názorný způsob reprezentace binárních relací. Graf binární relace R na množině X dostaneme tak, že každý prvek $x \in X$ znázorníme v rovině jako kroužek s označením daného prvku. Pokud $\langle x, y\rangle \in R$, nakreslíme z kroužku odpovídajícího x do kroužku odpovídajícího y orientovanou čáru s šipkou.

Příklad 2.26. Na Obr. 2 vidíme graf reprezentující binární relaci $R = \{\langle a, b\rangle, \langle a, d\rangle, \langle b, d\rangle, \langle c, a\rangle\}$ na množině $X = \{a, b, c, d\}$.

Upozorněme už teď, že graf je jedním ze základních pojmů diskrétní matematiky. Grafy se budeme zabývat v Kapitole 4. V tomto smyslu používáme v této kapitole pojem graf nepřesně. Podobně jak jsme ukázali, je možné reprezentovat i relace *R mezi X* a *Y*. Čtenář nechť si detaily rozmyslí sám.



Obrázek 3: Relace R z Příkladu 2.26 reprezentovaná seznamem seznamů.

Reprezentace seznamem seznamů

Tento způsob reprezentace je vhodný pro uložení binární relace R na množině X v paměti počítače. Na Obr. 3 je znázorněna reprezentace relace R z Příkladu 2.26 seznamem seznamů. Reprezentaci tvoří hlavní (spojový) seznam[6], ve kterém jsou uloženy všechny prvky množiny X. Na Obr. 3 je hlavní seznam znázorněn shora dolů spojenými čtverečky, které obsahují $a, \ldots, d$. Z každého prvku $x \in X$ hlavního seznamu vede seznam obsahující právě ty $y \in X$, pro které $\langle x, y\rangle \in R$. Na Obr. 3 jsou tyto seznamy znázorněny vodorovně. Např. z prvku a hlavního seznamu vede seznam obsahující b a d. To proto, že $\langle a, b\rangle \in R$ a $\langle a, d\rangle \in R$. Z prvku d nevede žádný seznam (tj. z d vede prázdný seznam), protože neexistuje $y \in X$ tak, že $\langle d, y\rangle \in R$.

Reprezentace seznamem seznamů je paměťově úsporná a je vhodná pro počítačové zpracování.

Vraťme se k relaci na množině X s 1000 prvky, kde každý prvek je v relaci s průměrně 3 prvky. Při reprezentaci seznamem seznamů budeme potřebovat 1000 políček pro prvky hlavního seznamu a pro každý z těchto prvků 3 další políčka pro prvky seznamu, který z tohoto prvky vede. To je celkem 4000 políček. Započítáme-li, že v každém políčku je třeba mít nejen označení prvku, ale i ukazatel na další políčko, je třeba zhruba 2×4000 paměťových buněk. Připomeňme, že maticová reprezentace takové relace vyžaduje 1000000 paměťových buněk[7].

2.4 Funkce (zobrazení)

2.4.1 Pojem funkce

Funkce je matematickým protějškem běžně používaného pojmu *přiřazení*. Objektům jsou často jednoznačným způsobem přiřazovány další objekty. Např. funkce sinus přiřazuje každému reálnému číslu x hodnotu $\sin(x)$, zaměstnancům jsou ve společnosti, kde pracují, přiřazována identifikační čísla apod. Takové přiřazení je možné chápat jako množinu dvojic $\langle x, y\rangle$, kde y je objekt přiřazený objektu x. Přiřazení je tedy možné chápat jako binární relaci mezi množinou X objektů, kterým jsou přiřazovány objekty, a množinou Y objektů, které jsou objektům z X přiřazovány. Taková relace R má díky jednoznačnosti přiřazení následující speciální vlastnost: je-li $\langle x, y_1\rangle \in R$ (objektu x je přiřazen objekt y_1) a $\langle x, y_2\rangle \in R$ (objektu x je přiřazen objekt y_2), pak $y_1 = y_2$ (jednoznačnost přiřazeného objektu, objektu x nemohou být přiřazeny dva různé objekty). To vede k následující definici.

[6] Spojový seznam je jednou ze základních datových struktur. Blíže viz jakoukoli učebnici algoritmů a datových struktur.

[7] Celá tato úvaha je zjednodušená, ale ilustruje podstatu věci.

Definice 2.27. Relace R mezi X a Y se nazývá *funkce* (někdy také *zobrazení*) množiny X do množiny Y, právě když pro každé $x \in X$ existuje $y \in Y$ tak, že

$$\langle x, y \rangle \in R,$$

a pro každé $x \in X$ a $y_1, y_2 \in Y$ platí, že

$$\langle x, y_1 \rangle \in R \text{ a } \langle x, y_2 \rangle \in R \text{ implikuje } y_1 = y_2.$$

Fakt, že R je funkce X do Y, označujeme $R : X \to Y$. Pro funkce používáme spíš $f, g, \ldots$ než $R, S, \ldots$. Je-li $f : X \to Y$ funkce a $x \in X$, pak ten $y \in Y$, pro který je $\langle x, y \rangle \in f$, označujeme $f(x)$, píšeme také $x \mapsto y$, popř. $x \mapsto f(x)$. V tom případě říkáme, že f zobrazuje prvek x na prvek y.

Příklad 2.28. Uvažujme množiny $X = \{a, b, c\}$, $Y = \{a, b, 1, 2\}$.

- Relace $R = \{\langle a, a \rangle, \langle b, b \rangle\}$ není funkce X do Y, protože k prvku $c \in X$ neexistuje prvek $y \in Y$ tak, že $\langle x, y \rangle \in R$.
- Relace $R = \{\langle a, a \rangle, \langle b, 2 \rangle, \langle c, a \rangle, \langle c, 2 \rangle\}$ není funkce X do Y, protože k prvku $c \in X$ existují dva různé prvky, které jsou s ním v relaci R. Máme totiž $\langle c, a \rangle \in R$, $\langle c, 2 \rangle \in R$, ale $a \neq 2$.
- Relace $R = \{\langle a, 2 \rangle, \langle b, b \rangle, \langle c, 2 \rangle\}$ je funkce X do Y.

Relace $R \subseteq X \times Y$, která splňuje, že když $\langle x, y_1 \rangle \in R$ a $\langle x, y_2 \rangle \in R$, pak $y_1 = y_2$, se někdy nazývá *parciální* (částečná) *funkce*. Přívlastek „parciální“ odkazuje k tomu, že může existovat $x \in X$, pro který neexistuje žádný $y \in Y$ tak, že $\langle x, y \rangle \in R$ (x se nezobrazí na žádný prvek).

Někdy se používá obrat „uvažujme funkci $y = f(x)$“, kde $f(x)$ je nějaký výraz, např. $y = x^2$ apod. Přitom se má za to, že je jasné, o jaké množiny X a Y se jedná (často je $X = Y = \mathbb{R}$, popř. $X \subseteq \mathbb{R}$). Pak jde vlastně o funkci $\{\langle x, y \rangle \mid x \in X, y \in Y, y = f(x)\}$.

2.4.2 Typy funkcí

Definice 2.29. Funkce $f : X \to Y$ se nazývá

- *prostá* (někdy také *injektivní*), právě když pro každé $x_1, x_2 \in X$, že z $x_1 \neq x_2$ plyne $f(x_1) \neq f(x_2)$,
- funkce množiny X *na* množinu Y (někdy také *surjektivní*), právě když pro každé $y \in Y$ existuje $x \in X$ tak, že $f(x) = y$,
- *vzájemně jednoznačná* (někdy také *bijektivní*), právě když je prostá a je to funkce na množinu Y (tj. injektivní a surjektivní).

Funkce je tedy prostá, právě když z $f(x_1) = f(x_2)$ plyne $x_1 = x_2$.

Příklad 2.30.

- Pro $X = \{a, b, c, d\}$ a $Y = \{1, 2, 3, 4\}$ je $f = \{\langle a, 1 \rangle, \langle b, 1 \rangle, \langle c, 4 \rangle, \langle d, 3 \rangle\}$ funkce X do Y. f není injektivní (protože $f(a) = f(b)$, ale $a \neq b$), ani surjektivní (neexistuje $x \in X$ tak, aby $f(x) = 2$), a tedy ani bijektivní.
- Pro $X = \{a, b\}$ a $Y = \{1, 2, 3\}$ je $f = \{\langle a, 1 \rangle, \langle b, 3 \rangle\}$ funkce X do Y, která je injektivní, ale není surjektivní (neexistuje $x \in X$ tak, aby $f(x) = 2$), a tedy ani bijektivní.

- Pro $X = \{a, b, c\}$ a $Y = \{1, 2\}$ je $f = \{\langle a, 2\rangle, \langle b, 2\rangle, \langle c, 1\rangle\}$ funkce X do Y, která není injektivní (protože $f(a) = f(b)$, ale $a \neq b$), ale je surjektivní, a tedy není bijektivní.
- Pro $X = \{a, b, c\}$ a $Y = \{1, 2, 3\}$ je $f = \{\langle a, 2\rangle, \langle b, 1\rangle, \langle c, 3\rangle\}$ funkce X do Y, která je injektivní i surjektivní, a i bijektivní.

Příklad 2.31. Podívejte se na následující funkce.

- $f = \{\langle x, y\rangle \in \mathbb{R} \times \mathbb{R} \mid y = x^2\}$ je funkce $\mathbb{R}$ do $\mathbb{R}$, která není injekce (např. $(-2)^2 = 2^2$) ani surjekce (např. neexistuje $x \in \mathbb{R}$ tak, že $x^2 = -1$). Uvažujeme-li ji však jako funkci množiny $\mathbb{R}$ do množiny $\{a \in \mathbb{R} \mid a \geq 0\}$ (nezáporná reálná čísla), je to surjekce.
- $f = \{\langle x, y\rangle \in \mathbb{R} \times \mathbb{R} \mid y = x^3\}$ je funkce $\mathbb{R}$ do $\mathbb{R}$, která je injekcí i surjekcí, tj. je bijekcí.
- $f = \{\langle x, y\rangle \in \mathbb{N} \times \mathbb{N} \mid y = x!\}$ je funkce $\mathbb{N}$ do $\mathbb{N}$ (faktoriál, tj. $x! = x \cdot (x-1) \cdots 2 \cdot 1$). Je to injekce, ale ne surjekce (např. číslo 3 není faktoriálem žádného čísla, tj. neexistuje $x \in \mathbb{N}$ tak, že $x! = 3$).

Podívejme se na některé vlastnosti funkcí.

Věta 2.32. *Pro funkce $f : X \to Y$, $g : Y \to Z$ platí*

a) $f \circ g$ je funkce.

b) Jsou-li f, g injekce, je $f \circ g$ injekce.

c) Jsou-li f, g surjekce, je $f \circ g$ surjekce.

Důkaz. Dokažme *a*). Nejprve musíme ukázat, že pro každé $x \in X$ existuje $z \in Z$ tak, že $\langle x, z\rangle \in f \circ g$. Protože je f funkce, existuje k $x \in X$ prvek $y \in Y$ tak, že $\langle x, y\rangle \in f$, a protože je g funkce, existuje k tomu y prvek $z \in Z$ tak, že $\langle y, z\rangle \in g$. Podle definice je tedy $\langle x, z\rangle \in f \circ g$. Nyní musíme ukázat, že když $\langle x, z_1\rangle \in f \circ g$ a $\langle x, z_2\rangle \in f \circ g$, pak $z_1 = z_2$. Když $\langle x, z_1\rangle \in f \circ g$ a $\langle x, z_2\rangle \in f \circ g$, pak podle definice pro nějaké $y_1, y_2 \in Y$ je $\langle x, y_1\rangle \in f$, $\langle y_1, z_1\rangle \in g$, a $\langle x, y_2\rangle \in f$, $\langle y_2, z_2\rangle \in g$. Protože f je funkce, musí být $y_1 = y_2$, a protože g je funkce, musí být $z_1 = z_2$. Tedy *a*) platí.

Dokažme *b*). Je-li $\langle x_1, z\rangle \in f \circ g$, $\langle x_2, z\rangle \in f \circ g$, existují $y_1, y_2 \in Y$ tak, že $\langle x_1, y_1\rangle \in f$, $\langle x_2, y_2\rangle \in f$, $\langle y_1, z\rangle \in g$, $\langle y_2, z\rangle \in g$. Protože g je injekce, platí $y_1 = y_2$. Platí tedy $\langle x_1, y_1\rangle \in f$, $\langle x_2, y_1\rangle \in f$, a protože f je injekce, je $x_1 = x_2$, tedy $f \circ g$ je injekce.

c) se dokáže podobně. □

Věta 2.33. *Funkce $f : X \to Y$ je bijekce, právě když existuje funkce $g : Y \to X$, pro kterou platí $f \circ g = \mathrm{id}_X$ a $g \circ f = \mathrm{id}_Y$; funkce g je přitom určena jednoznačně. Funkce g se pak nazývá inverzní k funkci f a značí se f^{-1}.*

Důkaz. Nechť f je bijekce. Ze surjektivity plyne, že pro každý prvek $y \in Y$ existuje $x \in X$ tak, že $f(x) = y$. Tento prvek je jediný, protože kdyby existoval $x' \neq x$, pro který $f(x') = y$, obdrželi bychom spor s předpokladem, že f je injektivní. Definujme tedy $g : Y \to X$ tak, že $g(y)$ je jediným prvkem $x \in X$, pro který $f(x) = y$. Je zřejmé, že $(f \circ g)(x) = g(f(x)) = x$ a $(g \circ f)(y) = f(g(y)) = y$. Pokud $g' : Y \to X$ splňuje $f \circ g = \mathrm{id}_X$ a $g \circ f = \mathrm{id}_Y$, je

$$g'(y) = g'((g \circ f)(y)) = [(g \circ f) \circ g'](y) = [g \circ (f \circ g')](y) = g((f \circ g')(y)) = g(y),$$

tedy $g' = g$ a g je tedy určena jednoznačně.

Nechť existuje g tak, že $f \circ g = \mathrm{id}_X$ a $g \circ f = \mathrm{id}_Y$. Pokud by f nebyla injektivní, existovaly by $x_1 \neq x_2$ tak, že $f(x_1) = f(x_2)$. Pak ale dle předpokladu $x_1 = (f \circ g)(x_1) = g(f(x_1)) = g(f(x_2)) = (f \circ g)(x_2) = x_2$, což je spor s předpokladem $x_1 \neq x_2$. f je tedy injektivní. f je i surjektivní, protože pro $y \in Y$ je dle předpokladu $g \circ f = \mathrm{id}_Y$ prvek $g(y) \in X$ prvkem, pro který $f(g(y)) = y$. □

Je-li f^{-1} inverzní funkce k funkci f, je dle definice f inverzní k f^{-1}.

Příklad 2.34.

- Nechť $X = Y = \mathbb{R}^+$ (množina kladných reálných čísel). Pak $f : x \mapsto x^2$ je bijekce a $f^{-1}(y) = \sqrt{y}$.
- Funkce $f : \mathbb{N} \to \mathbb{Z}$ definovaná předpisem

$$f(n) = \begin{cases} -\lfloor n/2 \rfloor & \text{pro } n \text{ liché,} \\ n/2 & \text{pro } n \text{ sudé,} \end{cases}$$

je bijekce, pro kterou

$$f^{-1}(n) = \begin{cases} 2n & \text{pro } n > 0, \\ -2n + 1 & \text{pro } n \leq 0. \end{cases}$$

Platí totiž $f(1) = 0$, $f(2) = 1$, $f(3) = -1$, $f(4) = 2$, $f(5) = -2$ atd.

Shrnutí

Kartézský součin množin $X_1, \ldots, X_n$ je množina všech uspořádaných n-tic prvků z těchto množin. Relace mezi množinami $X_1, \ldots, X_n$ je libovolná podmnožina kartézského součinu těchto množin. S relacemi lze provádět všechny množinové operace. S binárními relacemi lze provádět operace inverze a skládání. Binární relace se nejčastěji reprezentují tabulkou nebo grafem, v paměti počítače pak maticí nebo seznamem seznamů.
Funkce je zvláštní typ relace. Injekce, surjekce a bijekce jsou speciální typy funkcí.

Pojmy k zapamatování

- kartézský součin,
- relace, binární relace, inverzní relace, skládání binárních relací, reprezentace binárních relací,
- funkce, injekce, surjekce, bijekce.

Kontrolní otázky

1. *Je pravda, že každá neprázdná n-ární relace má aspoň n prvků? Proč?*
2. *Jaká je inverzní relace k relaci „být otcem“ na množině všech lidí (slovně ji popište)? Je-li R výše uvedená relace „být otcem“, co je relací $R \circ R$? Co jsou relace $R \lhd R$, $R \rhd R$?*
3. *Jaký je rozdíl mezi tabulkovou a maticovou reprezentací binární relace?*
4. *Nechť X a Y jsou množiny. Jaký vztah musí platit mezi $|X|$ a $|Y|$ pro to, aby existovala funkce $f : X \to Y$, která je injekcí, surjekcí, bijekcí?*
5. *Může být prázdná množina funkcí X do Y? Rozeberte v závislosti na množinách X a Y.*

Cvičení

1. Určete $A \times B$, kde $A = \{a, 1, \{a, b\}\}$ a $B = \{2, a, \{a\}\}$.
2. Dokažte následující vztahy.

$$\begin{aligned}
A \times B = \emptyset, \text{ právě když } A = \emptyset \text{ nebo } B = \emptyset \\
A \times B = B \times A, \text{ právě když } A \times B = \emptyset \text{ nebo } A = B \\
A \times (B \cup C) = (A \times B) \cup (A \times C) \\
(A \cup B) \times C = (A \times C) \cup (B \times C) \\
A \times (B \cap C) = (A \times B) \cap (A \times C) \\
(A \cap B) \times C = (A \times C) \cap (B \times C) \\
A \times (B - C) = (A \times B) - (A \times C) \\
(A - B) \times C = (A \times C) - (B \times C)
\end{aligned}$$

3. Najděte příklady relací, pro které platí (neplatí) $R \circ S = S \circ R$, $R^{-1} = R$.
4. Dokažte, že pro relace $R, R_1, R_2, U \subseteq X \times Y$, $S, S_1, S_2, V \subseteq Y \times Z$, $T \subseteq Z \times W$ platí

$$\begin{aligned}
(R^{-1})^{-1} = R \\
(R \circ S) \circ T = R \circ (S \circ T) \\
(R \circ S)^{-1} = S^{-1} \circ R^{-1} \\
\text{Je-li } R \subseteq U, S \subseteq V, \text{ pak } R \circ S \subseteq U \circ V \\
(R_1 \cup R_2)^{-1} = R_1^{-1} \cup R_2^{-1} \\
(R_1 \cap R_2)^{-1} = R_1^{-1} \cap R_2^{-1} \\
R \circ (S_1 \cup S_2) = R \circ S_1 \cup R \circ S_2 \\
R \circ (S_1 \cap S_2) = R \circ S_1 \cap R \circ S_2 \\
(R_1 \cup R_2) \circ S = R_1 \circ S \cup R_2 \circ S \\
(R_1 \cap R_2) \circ S = R_1 \circ S \cap R_2 \circ S
\end{aligned}$$

5. Které z následujících relací R jsou funkce X do Y?
   a) $X = Y = \mathbb{N}$, $R = \{\langle m, n\rangle \mid m \neq n\}$,
   b) $X = Y = \{a, b, c\}$, $R = \{\langle a, a\rangle, \langle a, b\rangle, \langle a, c\rangle\}$,
   c) $X = Y = \{a, b, c\}$, $R = \{\langle a, a\rangle, \langle b, a\rangle, \langle c, a\rangle\}$,
   d) X je množina všech českých slov, Y je množina všech písmen české abecedy $\{\langle w, l\rangle \mid w$ je české slovo s posledním písmenem $l\}$,
   e) $X = Y = \mathbb{R}$, $R = \{\langle x, y\rangle \mid x^2 + y^2 = 1\}$,
   f) $X = Y = \mathbb{R}$, $R = \{\langle x, y\rangle \mid x^2 = y\}$,
   g) $X = Y = \mathbb{R}$, $R = \{\langle x, y\rangle \mid x = y^2\}$.
6. Které z následujících funkcí jsou injektivní? Které jsou surjektivní?
   a) $f : N \to N$, $f(n) = n + 1$,
   b) $f : Z \to Z$, $f(i) = i + 1$,
   c) $f : K \to K$, kde $K = \{1, 2, \ldots, k\}$,

$$f(i) = \begin{cases} i + 1 & \text{pro } 1 \leq i < k \\ 1 & \text{pro } i = k \end{cases}$$

   d) $f : N \to \{0, 1, 2, 3\}$, kde

$$f(i) = \begin{cases} 0 & \text{jestliže } i \text{ je dělitelné 5, ale ne 11,} \\ 1 & \text{jestliže } i \text{ je dělitelné 11, ale ne 5,} \\ 2 & \text{jestliže } i \text{ je dělitelné 55,} \\ 3 & \text{v ostatních případech,} \end{cases}$$

e) $f : Q \to Q$, $f(i) = i^3$.

7. Najděte příklady funkcí f a g tak, aby
   a) g nebyla injekce, ale $f \circ g$ ano,
   b) f nebyla surjekce, ale $f \circ g$ ano.
8. Pro množinu U nechť je $\mathrm{id}_U = \{\langle u, u\rangle \mid u \in U\}$ relace identita. Ukažte, že pro relaci $f \subseteq X \times Y$ platí
   a) f splňuje, že z $\langle x, y_1\rangle \in f$ a $\langle x, y_2\rangle \in f$ plyne $y_1 = y_2$, právě když $f^{-1} \circ f \subseteq \mathrm{id}_Y$,
   b) je-li f funkce X do Y, pak je injektivní, právě když $f \circ f^{-1} = \mathrm{id}_X$.
   c) Je-li f funkce X do Y, pak je surjektivní, právě když $f^{-1} \circ f = \mathrm{id}_Y$.
9. Mějme $f : X \to Y$. Pro $A \subseteq X$ a $B \subseteq Y$ označme $f(A) = \{f(x) \mid x \in A\}$ a $f_{-1}(B) = \{x \in X \mid f(x) \in B\}$. Ukažte, že
   a) f je injektivní, právě když f^{-1} splňuje, že z $\langle x, y_1\rangle \in f$ a $\langle x, y_2\rangle \in f$ plyne $y_1 = y_2$,
   b) je-li f injektivní, pak pro každé $A, B \subseteq X$ platí $f(A \cap B) = f(A) \cap f(B)$, $f(A - B) = f(A) - f(B)$.
   c) pro každé $A, B \subseteq Y$ $f_{-1}(A \cap B) = f_{-1}(A) \cap f_{-1}(B)$, $f_{-1}(A - B) = f_{-1}(A) - f_{-1}(B)$.
10. Ukažte, že není-li f injektivní, neplatí bod b) z předchozího cvičení.
11. Dokažte, že pro funkce $f, f_1, f_2 : X \to Y$, $g, g_1, g_2 : Y \to Z$ platí, že
    a) je-li $f \circ g$ injekce, je f injekce,
    b) je-li $f \circ g$ surjekce, je g surjekce,
    c) je-li g injekce $f_1 \circ g = f_2 \circ g$, je $f_1 = f_2$,
    d) je-li f surjekce $f \circ g_1 = f \circ g_2$, je $g_1 = g_2$.

Úkoly k textu

1. Vraťme se k pojmu uspořádaná dvojice prvků. Tento pojem jsme chápali jako základní, tj. nedefinovaný. Je ho však možné definovat pomocí pojmu množina tak, že bude mít všechny požadované vlastnosti. Řekněme, že uspořádaná dvojice prvků a, b je množina $\langle a, b\rangle = \{a, \{a, b\}\}$. Ukažte, že $\langle a, b\rangle = \langle c, d\rangle$, právě když $a = c$ a $b = d$.
2. Dokažte zbývající části Věty 2.10
3. Dokažte Větu 2.24.
4. Ukažte, že pro konečné množiny X a Y existuje bijekce X do Y, právě když X a Y mají stejný počet prvků.
5. Dokažte bod c) z Věty 2.32.

Řešení

1. $A \times B = \{\langle a, 2\rangle, \langle a, a\rangle, \langle a, \{a\}\rangle, \langle 1, 2\rangle, \langle 1, a\rangle, \langle 1, \{a\}\rangle, \langle \{a, b\}, 2\rangle, \langle \{a, b\}, a\rangle, \langle \{a, b\}, \{a\}\rangle\}$.

2. Vztahy se dokážou jednoduše, rozepsáním přímo podle definice.
3. Mějme např. $X = Y = \{x, y, z\}$. $R \circ S = S \circ R$ platí pro $R = \{\langle x, y\rangle, \langle z, y\rangle\}$, $S = \{\langle y, x\rangle, \langle y, z\rangle\}$, neplatí pro $R = \{\langle x, y\rangle\}$, $S = \{\langle y, z\rangle\}$.
$R^{-1} = R$ platí např. pro $R = \{\langle x, y\rangle, \langle y, x\rangle\}$, neplatí např. pro $R = \{\langle x, z\rangle\}$.
4. Vztahy se dokážou jednoduše, rozepsáním přímo podle definice.
5. Funkcemi jsou relace R z c), d), f).
6. Injekce: a), b), c), e), surjekce: b), c), d).
7. Vezměme $X = \{x\}$, $Y = \{y_1, y_2\}$, $Z = \{z\}$. $f = \{\langle x, y_1\rangle\}$, $g = \{\langle y_1, z\rangle, \langle y_2, z\rangle\}$ splňují a) i b).
8. a) Nechť f splňuje, že z $\langle x, y_1\rangle \in f$ a $\langle x, y_2\rangle \in f$ plyne $y_1 = y_2$. Když $\langle y_1, y_2\rangle \in f^{-1} \circ f$, pak existuje $x \in X$ tak, že $\langle y_1, x\rangle \in f^{-1}$ a $\langle x, y_2\rangle \in f$, tj. $\langle x, y_1\rangle \in f$ a $\langle x, y_2\rangle \in f$. Z předpokladu plyne $y_1 = y_2$, tj. $f^{-1} \circ f \subseteq \mathrm{id}_Y$.
Naopak, nechť $f^{-1} \circ f \subseteq \mathrm{id}_Y$. Nechť $\langle x, y_1\rangle \in f$ a $\langle x, y_2\rangle \in f$. Pak $\langle y_1, y_2\rangle \in f^{-1} \circ f \subseteq \mathrm{id}_Y$. Protože $f^{-1} \circ f \subseteq \mathrm{id}_Y$, je $\langle y_1, y_2\rangle \in \mathrm{id}_Y$, tj. $y_1 = y_2$. Tedy z z $\langle x, y_1\rangle \in f$ a $\langle x, y_2\rangle \in f$ plyne $y_1 = y_2$.
b) a c) se dokážou podobnými úvahami.
9. a) přímo z definice.
b) $f(A \cap B) = f(A) \cap f(B)$: Protože $A \cap B \subseteq A$ i $A \cap B \subseteq B$, je dle definice $f(A \cap B) \subseteq f(A)$ i $f(A \cap B) \subseteq f(B)$. Z toho plyne $f(A \cap B) \subseteq f(A) \cap f(B)$. Naopak, pokud $y \in f(A) \cap f(B)$, pak existují $x_1 \in A$ a $x_2 \in B$ tak, že $f(x_1) = y$ a $f(x_2) = y$. Protože je f injekce, musí být $x_1 = x_2$. Tedy $x_1 \in A \cap B$, a proto $y \in f(A \cap B)$. Proto je $f(A) \cap f(B) \subseteq f(A \cap B)$.
$f(A - B) = f(A) - f(B)$: Nechť $y \in f(A - B)$, tj. existuje $x \in A - B$ tak, že $f(x) = y$. Proto je $y \in f(A)$. Kdyby $y \in f(B)$, pak existoval $x' \in B$ tak, že $f(x') = y$. Protože $x \in A - B$, je $x \neq x'$. To je ale spor s injektivitou f, protože máme $x \neq x'$ a $f(x) = y = f(x')$. Tedy je $y \in f(A) - f(B)$. Naopak, nechť $y \in f(A) - f(B)$. Pak $y = f(x)$ pro nějaký $x \in A$ a neexistuje $x' \in B$ tak, že $f(x') = y$. Proto je $x \in A - B$, a tedy $y \in f(A - B)$.
c) se dokáže podobnými úvahami.
10. Vezměme např. $X = \{x_1, x_2\}$, $Y = \{y\}$, funkci f danou předpisy $f(x_1) = y$, $f(x_2) = y$, množiny $A = \{x_1\}$, $B = \{x_2\}$. Pak $f(A \cap B) = \emptyset$ a $f(A) \cap f(B) = \{y\}$. Pro množiny $A = \{x_1, x_2\}$ a $B = \{x_2\}$ je $f(A - B) = f(\{x_1\}) = \{y\}$ a $f(A) = f(B) = \emptyset$.
11. Dokažme b). Kdyby g nebyla surjekce, pak by existoval $z \in Z$, kte kterému neexistuje $y \in Y$ tak, že $g(y) = z$. Proto nemůže existovat $x \in X$ tak, aby $f \circ g(x) = z$ (jinak by pro $y = f(x)$ bylo $g(y) = z$).
Dokažme d). Protože f je surjekce, existuje pro libovolný prvek $y \in Y$ prvek $x \in X$ tak, že $\langle x, y\rangle \in f$. Platí tedy $g_1(y) = g_1(f(x)) = f \circ g_1(x) = f \circ g_2(x) = g_2(f(x)) = g_2(y)$. Dokázali jsme, že pro libovolný prvek $y \in Y$ je $g_1(y) = g_2(y)$, tedy $g_1 = g_2$.
a) a c) se dokážou podobnými úvahami.

Kapitola 3 je převzata z textu R. Bělohlávek, V. Vychodil: Diskrétní matematika 1, 2, UP Olomouc, 2006. Bude upravena později.

3 Binární relace na množině

Studijní cíle: Po prostudování kapitoly by student měl znát běžné vlastnosti binárních relací na množině, jejich vzájemné vztahy a měl by mít představu o elementárních technikách, jak tyto vztahy rozpoznat. Dále by měl bých schopen k dané relaci najít její reflexivní, symetrický a tranzitivní uzávěr.

Klíčová slova: antisymetrie, asymetrie, irreflexivita, mocnina relace, reflexivita, relace, symetrie, tranzitivita, uzávěr (reflexivní, symetrický, tranzitivní), úplnost

3.1 Binární relace na množině

V kapitole 2.3 jsme zavedli pojem *relace* jakožto matematický protějšek běžně používaného pojmu *vztah.* Nyní se zaměříme na další vlastnosti a práci s relacemi, konkrétně s binárními relacemi na množině. Zopakujme, že binární relace R na množině $X \neq \emptyset$ je podmnožina kartézského součinu $X \times X$, to jest $R \subseteq X \times X$. Binární relace na množině jsou tedy matematickým protějškem vztahů mezi dvěma prvky množiny, například „x je menší než y“, „x má stejnou barvu jako y“, „x nezávisí na y“, ... Speciálními relacemi jsou *prázdná relace* $\emptyset$, *relace identity* $\omega_X = \{\langle x, x\rangle \,|\, x \in X\}$ (značí se také id_X), a *kartézský čtverec* $\iota_X = X \times X$.

Mnohé binární relace mají podobné vlastnosti a to i přesto, že jsou definovány na různých nosičích. Vezměme například množinu přirozených čísel $\mathbb{N}$ a definujme binární relaci R na $\mathbb{N}$:

$$R = \{\langle m, n\rangle \,|\, m \text{ má stejný počet cifer jako } n\}\,.$$

Dále uvažujme, že X označuje množinu všech lidí (z daného regionu) a definujme binární relaci R' na X následujícím předpisem

$$R' = \{\langle x, y\rangle \,|\, \text{rozdíl měsíčních příjmů } x \text{ a } y \text{ je menší než } 10\,000 \text{ Kč}\}\,.$$

Různé relace mohou mít analogické vlastnosti.

I když mají relace R, R' odlišené (námi přisouzené) interpretace, mají několik společných vlastností. Platí například, $\langle n, n\rangle \in R$ („n má stejný počet cifer jako n“) pro každé číslo $n \in \mathbb{N}$, analogicky $\langle x, x\rangle \in R'$ („rozdíl příjmů x a x je menší než 10 000 Kč“) pro každého člověka $x \in X$. Pro obě relace R, R' dále platí: pokud $\langle m, n\rangle \in R$, pak i $\langle n, m\rangle \in R$; pokud $\langle x, y\rangle \in R'$, pak i $\langle y, x\rangle \in R'$. V následující definici zavedeme vlastnosti binárních relací na množině.

Definice 3.1. Nechť R je binární relace na X. Řekneme, že R je

(i) *reflexivní,* pokud pro každé $x \in X$ platí $\langle x, x\rangle \in R$,
(ii) *irreflexivní,* pokud pro každé $x \in X$ platí $\langle x, x\rangle \notin R$,
(iii) *symetrická,* pokud pro každé $x, y \in X$ platí $\langle x, y\rangle \in R \to \langle y, x\rangle \in R$,
(iv) *asymetrická,* pokud pro každé $x, y \in X$ platí $\langle x, y\rangle \in R \to \langle y, x\rangle \notin R$,
(v) *antisymetrická,* pokud pro každé $x, y \in X$ platí $(\langle x, y\rangle \in R \wedge \langle y, x\rangle \in R) \to x = y$,
(vi) *úplná,* pokud pro každé $x, y \in X$ platí $\langle x, y\rangle \in R \vee \langle y, x\rangle \in R$,
(vii) *tranzitivní,* pokud pro každé $x, y, z \in X$ platí $(\langle x, z\rangle \in R \wedge \langle z, y\rangle \in R) \to \langle x, y\rangle \in R$.

Vlastnosti relací uvedené v definici 3.1 mají přirozenou interpretaci a u konečných binárních relací je lze vyčíst z jejich maticové a grafové reprezentace. Předpokládejme, že máme dánu binární relaci R na X.

- *Reflexivita* relace R vyjadřuje, že každý prvek $x \in X$ je v relaci R „sám ze sebou“. Relace R je reflexivní, právě když má binární matice $\mathbf{M}^R$ na diagonále samé jedničky, což je právě když je v orientovaném grafu $\langle X, R\rangle$ relace R u každého vrcholu „smyčka“.
- *Irreflexivita* relace R vyjadřuje, že *žádný* prvek $x \in X$ není v relaci R „sám se sebou“. Relace R je irreflexivní, právě když má binární matice $\mathbf{M}^R$ na diagonále samé nuly, což je právě když u žádného vrcholu orientovaného grafu $\langle X, R\rangle$ není „smyčka“. Relace nemůže být zároveň reflexivní a irreflexivní, nemusí mít ale ani jednu vlastnost z těchto dvou.
- *Symetrie* relace R vyjadřuje, že $\langle x, y\rangle \in R$, právě když $\langle y, x\rangle \in R$. To jest relace je symetrická pokud pro každé $x, y \in X$ máme buď současně $\langle x, y\rangle \in R$ a $\langle y, x\rangle \in R$, nebo současně $\langle x, y\rangle \notin R$ a $\langle y, x\rangle \notin R$. Relace R je symetrická, právě když její binární matice $\mathbf{M}^R$ je *symetrická podle diagonály,* to jest právě když je transponovaná matice $(\mathbf{M}^R)^T$ shodná s $\mathbf{M}^R$. V grafu relace se symetrie projevuje tak, že mezi vrcholy x, y buď není žádná hrana, nebo vede hrana z x do y i z y do x.
- *Asymetrie* relace R vyjadřuje, že do R nepadnou $\langle x, y\rangle$ a $\langle y, x\rangle$ současně. To jest relace je asymetrická pokud pro každé $x, y \in X$ máme buď současně $\langle x, y\rangle \notin R$ a $\langle y, x\rangle \notin R$, nebo do R padne právě jedna z dvojic $\langle x, y\rangle$ a $\langle y, x\rangle$. Z asymetrie přímo plyne irreflexivita, tím pádem asymetrie vylučuje reflexivitu. Pokud je relace R symetrická a asymetrická současně, pak $R = \emptyset$.
- *Antisymetrie* relace R vyjadřuje, že pro každé dva *různé* prvky $x, y \in X$ neplatí současně $\langle x, y\rangle \in R$ a $\langle y, x\rangle \in R$. R je antisymetrická, právě když každá dvě různá pole matice $\mathbf{M}^R$, která jsou souměrná podle diagonály, neobsahují dvě jedničky. V grafu relace se antisymetrie projevuje tak, že mezi dvěma různými vrcholy x, y je buď jedna hrana, nebo žádná. Z asymetrie plyne antisymetrie (obráceně obecně neplatí). Je-li R současně symetrická i antisymetrická, pak platí $R \subseteq \omega_X$.
- *Úplnost* relace R vyjadřuje, že pro každé dva $x, y \in X$ aspoň jedna z dvojic $\langle x, y\rangle, \langle y, x\rangle$ padne do R. Úplnost implikuje reflexivitu, to jest irreflexivita vylučuje úplnost, tím pádem i asymetrie vylučuje úplnost. R je úplná, právě když každá dvě pole matice $\mathbf{M}^R$, která jsou souměrná podle diagonály, obsahují aspoň jednu jedničku. V grafu $\langle X, R\rangle$ lze úplnost poznat tak, že mezi každými dvěma vrcholy vede aspoň jedna hrana.
- *Tranzitivita* relace R vyjadřuje, že pokud $\langle x, y\rangle \in R$ a pokud $\langle y, z\rangle \in R$, pak také $\langle x, z\rangle \in R$, to jest neformálně: pokud je x ve vztahu R s y (v grafu $\langle X, R\rangle$ vede hrana z x do y) a pokud je y ve vztahu R se z (v grafu $\langle X, R\rangle$ vede hrana z y do z), pak je i x ve vztahu R se z (v grafu $\langle X, R\rangle$ vede hrana z x do z). Řečeno ještě jinak, pokud v grafu $\langle X, R\rangle$ můžeme přejít z vrcholu x do vrcholu y po dvou hranách přes vrchol z, pak lze přejít x do y přímo (z x do y vede hrana).

Vlastnosti konečných relací je možné testovat zcela mechanicky prostě tím, že ověříme, zda-li platí definiční podmínky dané vlastnosti. Uvědomte si, že k prokázání, že daná vlastnost neplatí stačí najít jen jednu n-tici prvků, pro kterou definiční předpis neplatí – taková n-tice prvků nám slouží jako *protipříklad.* Například k tomu abychom zjistili, že relace R na X není symetrická stačí najít $x, y \in X$ tak, že $\langle x, y\rangle \in R$, ale $\langle y, x\rangle \notin R$. Pokud chceme ukázat, že vlastnost pro danou R platí, musíme provést test pro všechny prvky. Problém testování vlastností relací nastává v případě, kdy X je nekonečná množina – zde již mechanické testování „přes všechny prvky“ obecně nelze použít. V tomto případě lze obecně doporučit snažit se vysledovat vlastnosti relací z jejich popisu (definice) a poté je dokázat nebo vyvrátit protipříkladem. Někdy pomáhá představit si pouze konečnou podmnožinu X, vysledovat vlastnosti R zúžené na tuto konečnou podmnožinu a pak se je snažit dokázat obecně.

Příklad 3.2. (1) Nejprve si uvědomme vlastnosti speciálních relací $\emptyset$, ω_X a ι_X. $\emptyset$ je irreflexivní, symetrická, asymetrická, antisymetrická a tranzitivní, evidentně však není úplná ani reflexivní. ω_X je reflexivní, symetrická, antisymetrická a tranzitivní, není irreflexivní a není asymetrická. ω_X je úplná, právě když $|X| = 1$. ι_X je reflexivní, symetrická, tranzitivní a úplná, není irreflexivní, není asymetrická. ι_X je antisymetrická, právě když $|X| = 1$.

(2) Mějme dánu množinu $X = \{a, b, c, d\}$ a binární relaci R na X, kde

$$R = \{\langle a, a\rangle, \langle a, d\rangle, \langle b, b\rangle, \langle b, d\rangle, \langle c, a\rangle, \langle c, b\rangle, \langle c, c\rangle, \langle c, d\rangle, \langle d, d\rangle\}\,.$$

R je reflexivní, antisymetrická a tranzitivní (ostatní vlastnosti uvedené v definici 3.1 nemá).

(3) Vrátíme-li se nyní k motivačním příkladům z úvodu kapitoly, pak pro

$$R = \{\langle m, n\rangle |\, m \text{ má stejný počet cifer jako } n\}\,.$$

definovanou na množině celých čísel platí, že R je reflexivní, symetrická, tranzitivní (ostatní vlastnosti nemá). Relace

$$R' = \{\langle x, y\rangle |\, \text{rozdíl měsíčních příjmů } x \text{ a } y \text{ je menší než } 10\,000 \text{ Kč}\}\,.$$

je reflexivní a symetrická, ale obecně nemusí být tranzitivní (pokuste se vymyslet protipříklad).

(4) Mějme relaci R na množině celých čísel danou $R = \{\langle m, n\rangle \,|\, m - 2 \leq n\}$. R je reflexivní, protože $m - 2 \leq m$. R není tranzitivní, protože $4 - 2 \leq 2$, $2 - 2 \leq 0$, ale $4 - 2 \not\leq 0$, to jest $\langle 4, 2\rangle \in R$, $\langle 2, 0\rangle \in R$, ale $\langle 4, 0\rangle \notin R$. R je úplná, protože pro libovolná $m, n \in \mathbb{Z}$ máme buď $m \leq n$ (potom tím spíš $m - 2 \leq n$, tedy $\langle m, n\rangle \in R$), nebo $n \leq m$ (potom $\langle n, m\rangle \in R$). R není symetrická, protože třeba $\langle 5, 2\rangle \notin R$ a $\langle 2, 5\rangle \in R$. R není asymetrická, protože je reflexivní. R není antisymetrická, protože $\langle 1, 2\rangle \in R$ a $\langle 2, 1\rangle \in R$.

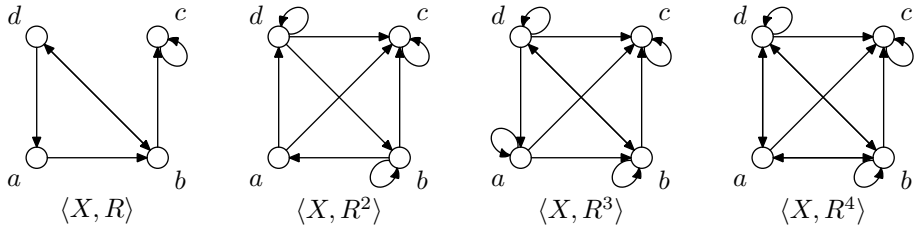
(5) Uvažujme libovolnou množinu U, dále zavedeme binární relaci R na 2^U následujícím předpisem: $R = \big\{\langle A, B\rangle \,|\, A, B \in 2^U \text{ a } A \text{ je podmnožina } B\big\}$. Relace R je reflexivní, antisymetrická a tranzitivní. Tuto relaci jsme si zavedli již v kapitole 2.2.3 na straně 36 jako množinovou inkluzi a fakt $\langle A, B\rangle \in R$ jsme zapisovali $A \subseteq B$. Analogicky bychom mohli chápat množinovou rovnost $A = B$ jako vyjádření příslušnosti $\langle A, B\rangle$ k průniku $R \cap R^{-1} = \omega_{2^U}$ (zdůvodněte proč).

Průvodce studiem

Asymetrii a antisymetrii nelze zaměňovat. Každá asymetrická relace je antisymetrická, ale obecně to neplatí obráceně. Například relace ω_X je antisymetrická, ale není asymetrická. Volně řečeno, antisymetrie vyjadřuje „téměř totéž co asymetrie“, až na prvky vyskytující se na diagonále. Platí, že antisymetrická relace je asymetrická, právě když je irreflexivní.

Věta 3.3. *Nechť R je binární relace na X. Pak*

(i) *R je reflexivní, právě když $\omega_X \subseteq R$,*
(ii) *R je irreflexivní, právě když $\omega_X \cap R = \emptyset$,*
(iii) *R je symetrická, právě když $R = R^{-1}$,*
(iv) *R je asymetrická, právě když $R \cap R^{-1} = \emptyset$,*
(v) *R je antisymetrická, právě když $R \cap R^{-1} \subseteq \omega_X$,*
(vi) *R je úplná, právě když $R \cup R^{-1} = \iota_X$,*
(vii) *R je tranzitivní, právě když $R \circ R \subseteq R$.*



Obrázek 4: n-tá mocnina relace

Důkaz. Tvrzení (i), (ii), (iii), (iv) a (vi) jsou zřejmá.
(v): Nechť R je antisymetrická a nechť $\langle x, y\rangle \in R \cap R^{-1}$. Pak $\langle x, y\rangle \in R$ a $\langle x, y\rangle \in R^{-1}$, tedy $\langle y, x\rangle \in R$. Odtud $x = y$, tedy $\langle x, y\rangle \in \omega_X$. Obráceně, předpokládejme $R \cap R^{-1} \subseteq \omega_X$. Pro $\langle x, y\rangle \in R$ a $\langle y, x\rangle \in R$ máme $\langle x, y\rangle \in R^{-1}$, tedy $\langle x, y\rangle \in R \cap R^{-1} \subseteq \omega_X$, z čehož $x = y$. Relace R je tedy antisymetrická.
(vii): Nechť R je tranzitivní a nechť $\langle x, y\rangle \in R \circ R$. Pak existuje $z \in X$ takové, že $\langle x, z\rangle \in R$ a $\langle z, y\rangle \in R$, tedy z tranzitivity $\langle x, y\rangle \in R$, to jest $R \circ R \subseteq R$. Obráceně, nechť platí $R \circ R \subseteq R$, pak pokud $\langle x, z\rangle \in R$ a $\langle z, y\rangle \in R$, pak $\langle x, y\rangle \in R \circ R \subseteq R$, tedy R je tranzitivní. □

Nyní zavedeme n-tou mocninu relace pomocí skládání relací.

Definice 3.4. Nechť R je binární relace na X. Pro každé $n \in \mathbb{N}$ definujeme binární relaci R^n na X:

n-tou mocninu relace zavádíme pomocí skládání relací.

$$R^n = \begin{cases} R & \text{pokud } n = 1, \\ R \circ R^{n-1} & \text{jinak.} \end{cases}$$

R^n se nazývá *n-tá mocnina R.*

Dle definice 3.4 máme $R^1 = R$, $R^2 = R \circ R^1 = R \circ R$, $R^3 = R \circ R^2 = R \circ (R \circ R) \ldots$ Podle věty 2.21 na straně 47 navíc platí, že $R^3 = R \circ (R \circ R) = (R \circ R) \circ R$, bez újmy tedy můžeme vynechávat závorky a psát pouze $R^3 = R \circ R \circ R$ a podobně. n-tou mocninu relace R na X lze vyjádřit následujícím způsobem: $\langle x, y\rangle \in R^n$ pokud existují $z_1, \ldots, z_{n-1} \in X$ tak, že

$$\langle x, z_1\rangle \in R, \langle z_1, z_2\rangle \in R, \ldots, \langle z_{n-2}, z_{n-1}\rangle \in R, \langle z_{n-1}, y\rangle \in R. \tag{2}$$

Tedy například pro $n = 2$ přejde (2) v

$$\langle x, z_1\rangle \in R, \langle z_1, y\rangle \in R,$$

pro $n = 3$:

$$\langle x, z_1\rangle \in R, \langle z_1, z_2\rangle \in R, \langle z_2, y\rangle \in R$$

a tak dále. Význam R^n si pro konečnou X nejlépe uvědomíme na orientovaných grafech příslušných R a R^n. Neformálně řečeno: v grafu R^n vede hrana z x do y (to jest $\langle x, y\rangle \in R^n$), právě když se v grafu R lze dostat z x do y po orientovaných hranách tak, že počet hran, přes které při tom přejdeme je roven n. Vše si demonstrujme na následujícím příkladu.

Příklad 3.5. Mějme binární relaci $R = \{\langle a, b\rangle, \langle b, c\rangle, \langle b, d\rangle, \langle c, c\rangle, \langle d, a\rangle, \langle d, b\rangle\}$ na čtyřprvkové množině $X = \{a, b, c, d\}$. Na obrázku 4 jsou zobrazeny orientované grafy odpovídající relacím $R = R^1$, R^2, R^3 a R^4, matice těchto relací vypadají následovně:

$$\mathbf{M}^R = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \ \mathbf{M}^{R^2} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}, \ \mathbf{M}^{R^3} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \ \mathbf{M}^{R^4} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Například $\langle a,b\rangle \in R$, $\langle b,d\rangle \in R$ a $\langle d,a\rangle \in R$, máme tedy $\langle a,a\rangle \in R^3$ – neformálně řečeno, z a jsme se dostali do a cestou přes tři hrany: $\langle a,b\rangle$, $\langle b,d\rangle$ a $\langle d,a\rangle$. Na druhou stranu ale třeba $\langle a,a\rangle \notin R^4$, protože neexistuje žádná „cesta z a do a“ přes čtyři hrany. Na tomto příkladu si dále všimněte, že obecně neplatí $R^n \subseteq R^{n+1}$, zde konkrétně $R^3 \not\subseteq R^4$.

V následujícím tvrzení si uvedeme některé další vlastnosti skládání a mocnění relací.

Věta 3.6. *Nechť R, S, T jsou binární relace na X, kde $S \subseteq T$. Pak*

(i) $S^{-1} \subseteq T^{-1}$,
(ii) $R \circ S \subseteq R \circ T$ *a* $S \circ R \subseteq T \circ R$,
(iii) *pokud je R tranzitivní, pak $R^n \subseteq R$ pro každé $n \in \mathbb{N}$,*
(iv) $R^m \circ R^n = R^{m+n} = R^n \circ R^m$ *pro každé $m, n \in \mathbb{N}$,*
(v) *pokud $|X| \leq n$ a $\langle x,y\rangle \in R^{n+1}$, pak $\langle x,y\rangle \in R^m$ pro nějaké $m \in \mathbb{N}$ splňující $m \leq n$.*

Důkaz. (i) je zřejmé.
(ii): Dokážeme $R \circ S \subseteq R \circ T$, druhý vztah se dokazuje analogicky. Nechť $\langle x,y\rangle \in R \circ S$, pak tedy existuje $z \in X$ tak, že $\langle x,z\rangle \in R$ a $\langle z,y\rangle \in S$. Jelikož ale $S \subseteq T$, dostáváme $\langle z,y\rangle \in T$, tedy $\langle x,y\rangle \in R \circ T$. Tím jsme dokázali $R \circ S \subseteq R \circ T$.
(iii): Triviálně $R^1 \subseteq R$. Indukcí ukážeme, že pokud tvrzení $R^n \subseteq R$ platí pro $n \in \mathbb{N}$, pak platí i pro $n+1$. Nechť tedy $R^n \subseteq R$. Jelikož je R tranzitivní relace, pak dle (ii) věty 3.6 a dle (vii) věty 3.3 dostáváme $R^{n+1} = R \circ R^n \subseteq R \circ R \subseteq R$.
(iv) plyne z definice a užitím věty 2.21 na straně 47.
(v): Mějme $|X| \leq n$, to jest X je nejvýš n-prvková množina a nechť $\langle x,y\rangle \in R^{n+1}$. Dle definice 3.4 tedy existují elementy $z_1, \ldots, z_n \in X$ takové, že

$$\langle x,z_1\rangle \in R, \langle z_1,z_2\rangle \in R, \ldots, \langle z_{n-1},z_n\rangle \in R, \langle z_n,y\rangle \in R.$$

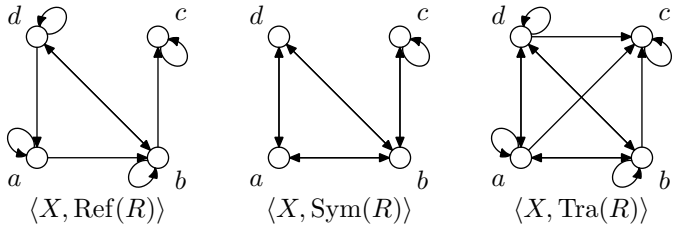
Označíme-li $z_0 = x$, pak má posloupnost $z_0, \ldots, z_n$ právě $n+1$ prvků, což vzhledem k počtu prvků množiny X (předpokládáme $|X| \leq n$) znamená, že musí existovat indexy $i, j \in \{0, \ldots, n\}$ takové, že $i < j$ a $z_i = z_j$. Vezmeme-li nyní posloupnost dvojic

$$\langle z_0,z_1\rangle \in R, \langle z_1,z_2\rangle \in R, \ldots, \langle z_{i-1},z_i\rangle \in R, \langle z_j,z_{j+1}\rangle \in R, \ldots, \langle z_{n-1},z_n\rangle \in R, \langle z_n,y\rangle \in R,$$

kterou jsme získali vyjmutím $(j-i)$ dvojic $\langle z_k,z_{k+1}\rangle$ $(i \leq k < j)$ z výchozí posloupnosti, pak zřejmě dostáváme $\langle x,y\rangle \in R^{n+1-(j-i)}$. Hledané $m \in \mathbb{N}$ je tedy například $m = n+1-(j-i)$. □

3.2 Uzávěry relací

Binární relace se často používají k popisu stavů systému a jejich návaznosti. Mějme množinu stavů X, ve kterých se může nacházet nějaký systém (například přístroj se může nacházet ve stavech „zapnut“, „v pohotovosti“, „v činnosti“, „vypnut“, …; tržní ekonomika se může nacházet ve stavu „konjunktury“, „inflace“, „deflace“, „stagflace“, „recese“, „deprese“, …). Předpokládejme, že R je binární relace na X s významem: $\langle x,y\rangle \in R$, právě když „systém může (v jednom kroku) přejít ze stavu x do stavu y“. Relace R (tak zvaná *přechodová relace*) nemusí být obecně tranzitivní: těžko si lze představit, že například ekonomika přejde přímo z „deprese“ do „konjunktury“. Kdybychom navíc na X definovali další binární relaci R' jakožto relaci *dosažitelnosti stavů* v obecně více krocích, pak by R' byla přirozeně *tranzitivní* a měla by navíc úzký vztah k přechodové relaci R: R by byla obsažena v R' a R' by byla nejmenší (ve smyslu množinové inkluze $\subseteq$) ze všech tranzitivních relací na X obsahujících R. Tento a podobné typy vztahů mezi relacemi si nyní zavedeme přesně.



Obrázek 5: Reflexivní, symetrický a tranzitivní uzávěr relace

Definice 3.7. Pro binární relaci R na X definujeme binární relace $\mathrm{Ref}(R)$, $\mathrm{Sym}(R)$, $\mathrm{Tra}(R)$ na X tak, že $\mathrm{Ref}(R)$ ($\mathrm{Sym}(R)$, případně $\mathrm{Tra}(R)$) je reflexivní (symetrická, případně tranzitivní) relace obsahující R a pro každou reflexivní (symetrickou, případně tranzitivní) relaci R' na X, kde $R \subseteq R'$, máme $\mathrm{Ref}(R) \subseteq R'$ ($\mathrm{Sym}(R) \subseteq R'$, případně $\mathrm{Tra}(R) \subseteq R'$). $\mathrm{Ref}(R)$ se nazývá *reflexivní uzávěr* R, $\mathrm{Sym}(R)$ se nazývá *symetrický uzávěr* R, $\mathrm{Tra}(R)$ se nazývá *tranzitivní uzávěr* R.

Předchozí definice je nekonstruktivní, neříká nic o tom, jak relace $\mathrm{Ref}(R)$, $\mathrm{Sym}(R)$ a $\mathrm{Tra}(R)$ vypadají a dokonce z ní ani přímo neplyne, zda takové relace existují pro každou R. V následující větě si ukážeme, že všechny výše uvedené uzávěry existují vždy ke každé binární relaci na libovolné množině a lze je konstruktivně popsat.

Věta 3.8. *Nechť R je binární relace na X. Pak*

$$\mathrm{Ref}(R) = R \cup \omega_X, \tag{3}$$

$$\mathrm{Sym}(R) = R \cup R^{-1}, \tag{4}$$

$$\mathrm{Tra}(R) = \textstyle\bigcup_{n=1}^{\infty} R^n. \tag{5}$$

Důkaz. Tvrzení pro $\mathrm{Ref}(R)$ je zřejmé.
Položme $S = R \cup R^{-1}$. Nyní prokážeme, že $\mathrm{Sym}(R) = S$. Platí $S^{-1} = (R \cup R^{-1})^{-1} = R^{-1} \cup (R^{-1})^{-1} = R^{-1} \cup R = S$. To jest S je symetrická dle bodu (iii) věty 3.3 a evidentně $R \subseteq S$. Mějme symetrickou relaci S' na X takovou, že $R \subseteq S'$. Nechť $\langle x, y\rangle \in S = R \cup R^{-1}$. Pokud $\langle x, y\rangle \in R$, pak triviálně $\langle x, y\rangle \in S'$. Pokud $\langle x, y\rangle \notin R$, pak $\langle x, y\rangle \in R^{-1}$, tedy $\langle y, x\rangle \in R$, to jest $\langle y, x\rangle \in S'$. Jelikož je S' symetrická, dostáváme odtud $\langle x, y\rangle \in S'$. Prokázali jsme tedy $S \subseteq S'$ a dohromady $\mathrm{Sym}(R) = R \cup R^{-1}$.
Položme $T = \bigcup_{n=1}^{\infty} R^n$, to jest $T = R^1 \cup R^2 \cup R^3 \cup \cdots$. Evidentně $R \subseteq T$. Ověříme tranzitivitu T: nechť $\langle x, z\rangle \in T$ a $\langle z, y\rangle \in T$. Z definice T plyne, že existují $m, n \in \mathbb{N}$ taková, že $\langle x, z\rangle \in R^m$ a $\langle z, y\rangle \in R^n$. Tedy $\langle x, y\rangle \in R^m \circ R^n = R^{m+n} \subseteq T$. Mějme tranzitivní relaci T' na X takovou, že $R \subseteq T'$. Pak dle bodů (ii) a (iii) věty 3.6 platí $R^n \subseteq (T')^n \subseteq T$ pro každé $n \in \mathbb{N}$. Odtud máme $T \subseteq T'$, tedy $\mathrm{Tra}(R) = \bigcup_{n=1}^{\infty} R^n$. □

Vztahu (5) je třeba rozumět tak, že výsledný tranzitivní uzávěr $\mathrm{Tra}(R)$ dostaneme tím, že sjednotíme nekonečně mnoho relací $R^1 \cup R^2 \cup R^3 \cup \cdots \cup R^n \cup R^{n+1} \cup \cdots$. Z pohledu informatika je toto vyjádření tranzitivního uzávěru pořád nekonstruktivní, protože pro vstupní relaci R nedává předpis pro algoritmus, který by vždy po *konečně mnoha krocích výpočtu* stanovil relaci $\mathrm{Tra}(R)$. Pokud je ale R definovaná na konečné množině X, pak jako důsledek bodu (v) věty 3.6 dostáváme, že $R^k \subseteq \bigcup_{i=1}^{n} R^i$ pro každé $k \in \mathbb{N}$, kde $k > n = |X|$. To jest máme

Důsledek 3.9. *Nechť R je binární relace na X, kde $|X| = n$. Pak* $\mathrm{Tra}(R) = \bigcup_{i=1}^{n} R^i$.
□

Příklad 3.10. (1) Vezměme relaci R z příkladu 3.5. Na obrázku 5 jsou zobrazeny orientované grafy odpovídající reflexivnímu, symetrickému a tranzitivnímu uzávěru R,

matice těchto relací vypadají následovně (pro přehlednost jsou nově přidané prvky zvýrazněny barevně):

$$\mathbf{M}^{\mathrm{Ref}(R)} = \begin{pmatrix} \mathbf{1} & 1 & 0 & 0 \\ 0 & \mathbf{1} & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & \mathbf{1} \end{pmatrix}, \quad \mathbf{M}^{\mathrm{Sym}(R)} = \begin{pmatrix} 0 & 1 & 0 & \mathbf{1} \\ \mathbf{1} & 0 & 1 & 1 \\ 0 & \mathbf{1} & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{M}^{\mathrm{Tra}(R)} = \begin{pmatrix} \mathbf{1} & 1 & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & \mathbf{1} & \mathbf{1} \end{pmatrix}.$$

(2) Mějme binární relaci R na množině přirozených čísel $\mathbb{N}$ definovánu $R = \{\langle m, n\rangle \mid m + 1 = n\}$. Relace R je irreflexivní, asymetrická, antisymetrická a není tranzitivní. Fakt $\langle m, n\rangle \in R$ lze intuitivně chápat jako „m je bezprostřední předchůdce n v množině přirozených čísel“. Pro $\mathrm{Tra}(R)$ máme $\langle m, n\rangle \in \mathrm{Tra}(R)$, právě když $m < n$. Na tomto příkladu je dobré uvědomit si, že v případě relací na nekonečných množinách můžeme mít $\bigcup_{i=1}^{n} R^i \subset \mathrm{Tra}(R)$ pro každé $n \in \mathbb{N}$. V tomto konkrétním případě: pro libovolné $n \in \mathbb{N}$ máme $\langle 1, n+2\rangle \in \mathrm{Tra}(R)$, ale $\langle 1, n+2\rangle \notin \bigcup_{i=1}^{n} R^i$. To jest důsledek 3.9 obecně nelze rozšířit pro relace na nekonečných množinách.

Průvodce studiem

Pokud binární relaci R na X interpretujeme jako přechodovou relaci, pak tranzitivní uzávěr $\mathrm{Tra}(R)$ relace R má přirozenou interpretaci jako relace dosažitelnosti. Tranzitivní uzávěry relací se často používají v informatice, například v teorii automatů: pokud R popisuje možnost změny stavu (elementární krok výpočtu) nějakého abstraktního výpočetního stroje, pak $\mathrm{Tra}(R)$ lze chápat jako relaci, která určuje „výpočet“, to jest $\langle x, y\rangle \in \mathrm{Tra}(R)$, pokud stroj během výpočtu začínajícího ve stavu x dojde do stavu y.

Shrnutí

Binární relace na dané množině lze chápat jako matematický protějšek vztahů mezi dvěma objekty dané množiny. Při zkoumání vlastností binárních relací na množině zavádíme abstraktní vlastnosti relací a zkoumáme jejich vzájemné vztahy. Mezi nejčastěji uvažované vlastnosti patří reflexivita, symetrie, antisymetrie, tranzitivita a úplnost. Ke každé relaci můžeme navíc stanovit její reflexivní, symetrický nebo tranzitivní uzávěr, to jest nejmenší reflexivní, symetrickou nebo tranzitivní relaci na dané množině, která obsahuje výchozí relaci.

Pojmy k zapamatování

- reflexivita, irreflexivita,
- symetrie, asymetrie, antisymetrie,
- úplnost,
- tranzitivita,
- mocnina relace,
- reflexivní uzávěr, symetrický uzávěr, tranzitivní uzávěr

Kontrolní otázky

1. *Existuje úplná a symetrická relace na X různá od ι_X?*
2. *Jaký je vztah mezi asymetrickými a antisymetrickými relacemi?*
3. *Platí, že relace je irreflexivní, právě když není reflexivní?*

4. Jaký význam má tranzitivní uzávěr relace a jak jej lze najít?

Cvičení

1. Zjistěte, jaké vlastnosti mají následující relace R na X.
   (a) $X = \mathbb{Z}$, $R = \{\langle m, n\rangle \mid m \text{ dělí } n \text{ beze zbytku}\}$.
   (b) $X = \{a, b, c, d\}$, $R = \{\langle a, b\rangle, \langle b, a\rangle, \langle b, b\rangle, \langle b, c\rangle\}$.
   (c) $X = \mathbb{Z} \times \mathbb{Z}$, $R = \{\langle\langle m, m'\rangle, \langle n, n'\rangle\rangle \mid m \leq n \text{ a } m' \leq n'\}$.
   (d) $X = \{0, 1, 2, 3, 4, 5, 6, 7\}$, $R = \{\langle m, n\rangle \mid m - n \leq 2\}$.
2. Najděte relaci R na $X = \{a, b, c, d\}$, tak aby
   (a) R byla reflexivní, tranzitivní, úplná a nebyla ani symetrická, ani antisymetrická,
   (b) R byla reflexivní, antisymetrická a nebyla tranzitivní,
   (c) R byla tranzitivní a asymetrická a zároveň $R \cup \omega_X$ byla úplná,
   (d) $\mathrm{Sym}(\mathrm{Tra}(R))$ nebyla tranzitivní.
3. Mějme relaci $R = \{\langle a, d\rangle, \langle c, d\rangle, \langle d, a\rangle\}$ na množině $X = \{a, b, c, d, e, f\}$. Stanovte $\mathrm{Ref}(R)$, $\mathrm{Sym}(R)$, $\mathrm{Tra}(R)$, $\mathrm{Sym}(\mathrm{Tra}(R))$, $\mathrm{Tra}(\mathrm{Sym}(R))$, $\mathrm{Tra}(\mathrm{Sym}(\mathrm{Ref}(R)))$.

Úkoly k textu

1. Mějme reflexivní relaci R na X, kde $|X| = n$. Dokažte, že $\mathrm{Tra}(R) = R^n$.

Řešení

1. Relace mají právě tyto vlastnosti
   (a) reflexivita, tranzitivita, (nesplňuje: irreflexivita, symetrie, asymetrie, antisymetrie, úplnost),
   (b) nesplňuje ani jednu vlastnost z definice 3.1,
   (c) reflexivita, tranzitivita, antisymetrie, (nesplňuje: irreflexivita, symetrie, asymetrie, úplnost),
   (d) reflexivita, úplnost, (nesplňuje: irreflexivita, symetrie, asymetrie, antisymetrie, tranzitivita).
2. Relace mají například následující tvar:
   (a) $R = \iota_X - \{\langle a, b\rangle, \langle a, c\rangle, \langle a, d\rangle\}$,
   (b) $R = \{\langle a, a\rangle, \langle a, b\rangle, \langle b, b\rangle, \langle b, c\rangle, \langle c, c\rangle, \langle d, d\rangle\}$,
   (c) $R = \{\langle a, b\rangle, \langle a, c\rangle, \langle a, d\rangle, \langle b, c\rangle, \langle b, d\rangle, \langle c, d\rangle\}$,
   (d) $R = \{\langle a, b\rangle\}$.
3. Relace mají následující tvar:

$$\begin{aligned}
\mathrm{Ref}(R) &= \{\langle a, a\rangle, \langle a, d\rangle, \langle b, b\rangle, \langle c, c\rangle, \langle c, d\rangle, \langle d, a\rangle, \langle d, d\rangle, \langle e, e\rangle, \langle f, f\rangle\}\,,\\
\mathrm{Sym}(R) &= \{\langle a, d\rangle, \langle c, d\rangle, \langle d, a\rangle, \langle d, c\rangle\}\,,\\
\mathrm{Tra}(R) &= \{\langle a, a\rangle, \langle a, d\rangle, \langle c, a\rangle, \langle c, d\rangle, \langle d, a\rangle, \langle d, d\rangle\}\,,\\
\mathrm{Sym}(\mathrm{Tra}(R)) &= \{\langle a, a\rangle, \langle a, c\rangle, \langle a, d\rangle, \langle c, a\rangle, \langle c, d\rangle, \langle d, a\rangle, \langle d, c\rangle, \langle d, d\rangle\}\,,\\
\mathrm{Tra}(\mathrm{Sym}(R)) &= \{\langle a, a\rangle, \langle a, c\rangle, \langle a, d\rangle, \langle c, a\rangle, \langle c, c\rangle, \langle c, d\rangle, \langle d, a\rangle, \langle d, c\rangle, \langle d, d\rangle\}\,,\\
\mathrm{Tra}(\mathrm{Sym}(\mathrm{Ref}(R))) &= \mathrm{Tra}(\mathrm{Sym}(R)) \cup \{\langle b, b\rangle, \langle e, e\rangle, \langle f, f\rangle\}\,.
\end{aligned}$$

Studijní cíle: Po prostudování kapitoly by student měl být seznámen s vlastnostmi nejčastěji používaných binárních relacích na množinách. Student by měl mít přehled o relaci ekvivalence a jejím vztahu k rozkladu na množině a k surjektivním obrazům. Dále by měl být seznámen se základními vlastnostmi a typy uspořádaných množin.

Klíčová slova: antiřetězec, ekvivalence (indukovaná zobrazením / příslušná rozkladu), ekvivalence, faktorová množina, Hasseův diagram, infimum, kužel (dolní, horní), kvaziuspořádání, lineární uspořádání, pokrytí, polosvaz (průsekový, spojový), princip duality, (minimální / nejmenší / maximální / největší) prvek, přirozené zobrazení, rozklad na množině, rozklad příslušný ekvivalenci, řetězec, supremum, svaz, třída ekvivalence / rozkladu, uspořádaná množina, uspořádání

Potřebný čas: 120 minut.

3.3 Ekvivalence

Ztotožněním nerozlišitelných prvků získáme „náhled“ na množinu.

V této podkapitole se budeme zabývat binárními relacemi, které lze interpretovat jako matematické protějšky *nerozlišitelnosti.* Motivace pro zkoumání tohoto fenoménu je v celku jasná. Pokud množina X reprezentuje velkou kolekci prvků, třeba nekonečnou, v některých případech můžeme chtít *ztotožnit* ty prvky z X, které jsou vzájemně nerozlišitelné některou svou vlastností a získat tak „zjednodušený náhled“ na X. Ztotožněním nerozlišitelných prvků můžeme získat množinu „reprezentantů“, která může být výrazně menší než výchozí množina X.

V úvodu kapitoly si nejprve definujeme speciální relaci, která je matematickým protějškem nerozlišitelnosti prvků. Jaké vlastnosti by tato relace měla mít? Určitě by měla být reflexivní, protože každé $x \in X$ je totožné s x, to jest „x nelze rozlišit od x“. Dále by relace měla být i symetrická: „pokud x nelze rozlišit od y, pak i y nelze rozlišit od x“. Další vlastností nerozlišitelnost je tranzitivita: „pokud x nelze rozlišit od y a y nelze rozlišit od z, pak x nelze rozlišit od z“. Uvědomte si, že kdybychom hledali matematický protějšek vlastnosti „být podobný“, pak by již automatické přijetí tranzitivity bylo přinejmenším diskutabilní. Nyní zavedeme relaci ekvivalence jako matematický protějšek nerozlišitelnosti.

Ekvivalence je matematický protějšek nerozlišitelnosti.

Definice 3.11. Reflexivní a symetrická binární relace na množině se nazývá *tolerance.* Tranzitivní tolerance se nazývá *ekvivalence.* Pro ekvivalenci E na množině X definujeme pro každý $x \in X$ množinu $[x]_E = \{y \in X \mid \langle x, y\rangle \in E\}$, kterou nazýváme *třída ekvivalence prvku x.*

Je-li E ekvivalence na X, pak vztah $\langle x, y\rangle \in E$ někdy čteme „x je E-ekvivalentní y“. Vzhledem k tomu, že E je symetrická, můžeme $\langle x, y\rangle \in E$ číst „x a y jsou E-ekvivalentní“. Třída ekvivalence $[x]_E$ je dle definice množina těch prvků $y \in X$, které jsou E-ekvivalentní x. Jinými slovy, $[x]_E$ obsahuje právě ty prvky z X, které nelze od x rozlišit ekvivalencí E.

Příklad 3.12. (1) ω_X a ι_X jsou ekvivalence na X, které navíc mají mezi všemi ekvivalencemi na X výsadní postavení. Relace ω_X musí být dle (i) věty 3.3 obsažena v každé ekvivalenci na X, to jest ω_X je nejmenší ekvivalence na X ve smyslu množinové inkluze $\subseteq$. Pro každý prvek $x \in X$ máme $[x]_{\omega_X} = \{x\}$. Naopak relace ι_X je evidentně největší ekvivalence na X. Pro každý prvek $x \in X$ máme $[x]_{\iota_X} = X$.

(2) Na $X = \{a, b, c\}$ existuje pět vzájemně různých ekvivalencí:

$$\begin{aligned}
\omega_X &= \{\langle a,a\rangle, \langle b,b\rangle, \langle c,c\rangle\},\\
E_1 &= \{\langle a,a\rangle, \langle a,b\rangle, \langle b,a\rangle, \langle b,b\rangle, \langle c,c\rangle\},\\
E_2 &= \{\langle a,a\rangle, \langle b,b\rangle, \langle b,c\rangle, \langle c,b\rangle, \langle c,c\rangle\},\\
E_3 &= \{\langle a,a\rangle, \langle a,c\rangle, \langle b,b\rangle, \langle c,a\rangle, \langle c,c\rangle\},\\
\iota_X &= \{\langle a,a\rangle, \langle a,b\rangle, \langle a,c\rangle, \langle b,a\rangle, \langle b,b\rangle, \langle b,c\rangle, \langle c,a\rangle, \langle c,b\rangle, \langle c,c\rangle\}.
\end{aligned}$$

(3) Uvažujme množinu celých čísel $\mathbb{Z}$ a $m \in \mathbb{N}$ a uvažujme binární relaci $\mathbb{Z}_m \subseteq \mathbb{Z} \times \mathbb{Z}$ definovanou

$$\mathbb{Z}_m = \{\langle a, b\rangle \mid a = b + t \cdot m \text{ pro nějaké } t \in \mathbb{Z}\}. \tag{6}$$

Snadno lze ukázat, že relace $\mathbb{Z}_m$ je ekvivalence na $\mathbb{Z}$. Podrobně, $\mathbb{Z}_m$ je zcela jistě reflexivní, platí $a = a + 0 \cdot m$, odtud plyne $\langle a, a\rangle \in \mathbb{Z}_m$. Pokud $\langle a, b\rangle \in \mathbb{Z}_m$, pak lze psát $a = b + t \cdot m$ pro nějaké $t \in \mathbb{Z}$. Tento výraz lze upravit na $a - t \cdot m = b$. Z vlastností součtu a součinu plyne $b = a + (-t) \cdot m$, to jest $\langle b, a\rangle \in \mathbb{Z}_m$ – relace je symetrická. Zbývá ověřit tranzitivitu: uvažujme $\langle a, b\rangle \in \mathbb{Z}_m$, $\langle b, c\rangle \in \mathbb{Z}_m$. Existují tedy $s, t \in \mathbb{Z}$, kde $a = b + s \cdot m$, $b = c + t \cdot m$. Dosazením za b dostáváme $a = c + t \cdot m + s \cdot m$, to jest $a = c + (t + s) \cdot m$. Relace $\mathbb{Z}_m$ se nazývá *ekvivalence modulo m*. O číslech $a, b \in \mathbb{Z}$ splňujících $\langle a, b\rangle \in \mathbb{Z}_m$ říkáme, že *a je ekvivalentní b modulo m*.

(4) Na $\mathbb{Q}$ můžeme uvažovat binární relaci $R = \{\langle x, y\rangle \mid x, y \in \mathbb{Q}, |x| = |y|\}$, to jest $\langle x, y\rangle \in R$, právě když mají x a y tutéž absolutní hodnotu. Zcela evidentně jde o ekvivalencí na $\mathbb{Q}$, přitom pro každé $x \in \mathbb{Q}$ máme $[x]_E = \{x, -x\}$. Speciálně máme $[0]_E = \{0\}$.

Další přirozený způsob zachycení informace o nerozlišitelnosti prvků množiny X je jejich „shlukování“ do podmnožin vzájemně nerozlišitelných prvků. Zřejmě každý prvek $x \in X$ je nerozlišitelný sám se sebou, takže ke každému $x \in X$ lze uvažovat neprázdnou podmnožinu $Y_x \subseteq X$ obsahující všechny prvky, které jsou od x nerozlišitelné. Zřejmě máme $X = \bigcup_{x \in X} Y_x$ a intuice také říká, že pokud lze x od x' rozlišit, pak neexistuje žádný prvek $z \in X$, který by byl zároveň nerozlišitelný od x i od x'. Následující definice shrnuje předchozí pozorování:

Rozklad na množině je matematický protějšek shluků nerozlišitelných prvků.

Definice 3.13. Nechť $X \neq \emptyset$ je množina. Systém množin $\Pi \subseteq 2^X$ splňující

(i) $Y \neq \emptyset$ pro každou $Y \in \Pi$,

(ii) pro každé $Y_1, Y_2 \in \Pi$ platí: pokud $Y_1 \cap Y_2 \neq \emptyset$, pak $Y_1 = Y_2$,

(iii) $\bigcup \Pi = X$,

se nazývá *rozklad na množině X*. Množiny $Y \in \Pi$ nazýváme *třídy rozkladu* Π. Pro prvek $x \in X$ označíme $[x]_\Pi$ tu třídu rozkladu Π, která obsahuje x.

Poznámka 3.14. Rozeberme nyní definici 3.13. Rozklad Π na X je systém neprázdných podmnožin X, přitom dle bodu (ii) požadujeme, aby každé dvě různé množiny $Y_1, Y_2 \in \Pi$, $Y_1 \neq Y_2$ byly disjunktní, to jest $Y_1 \cap Y_2 = \emptyset$, a konečně chceme, aby sjednocení všech množin z Π bylo rovno množině X – někdy proto říkáme, že rozklad na X je *disjunktní pokrytí množiny X*. Mějme například množinu $X = \{1, 2, 3, 4, 5, 6\}$. Například $\Pi_1 = \{\{1, 2\}, \{3, 4\}, \{5\}\}$ není rozklad na X, protože nesplňuje podmínku (iii) definice 3.13 ($6 \notin \bigcup \Pi_1$), stejně tak $\Pi_2 = \{\{1, 2, 3\}, \{3, 4\}, \{5, 6\}\}$ není rozklad na X, protože nesplňuje podmínku (ii) definice 3.13 ($\{1, 2, 3\} \cap \{3, 4\} \neq \emptyset$), na druhou stranu třeba $\{\{1, 5, 3\}, \{4\}, \{2, 6\}\}$ je rozklad na X.

Příklad 3.15. (1) Na množině X existují dva mezní rozklady. První z nich je rozklad Π, kde $[x]_\Pi = \{x\}$ pro každé $x \in X$, to jest všechny třídy rozkladu Π jsou jednoprvkové. Druhým mezním případem je rozklad $\Pi = \{X\}$, to jest Π obsahuje jedinou třídu, která je rovna celé X, tím pádem $[x]_\Pi = X$ pro každé $x \in X$.

(2) Mějme $X = \{a, b, c, d\}$. Na X existuje patnáct vzájemně různých rozkladů:

$\{\{a\}, \{b\}, \{c\}, \{d\}\}$, $\{\{a, b\}, \{c\}, \{d\}\}$, $\{\{b\}, \{a, c\}, \{d\}\}$, $\{\{b\}, \{c\}, \{a, d\}\}$,
$\{\{a\}, \{b, c\}, \{d\}\}$, $\{\{a, b, c\}, \{d\}\}$, $\{\{b, c\}, \{a, d\}\}$, $\{\{a\}, \{c\}, \{b, d\}\}$,
$\{\{a, c\}, \{b, d\}\}$, $\{\{c\}, \{a, b, d\}\}$, $\{\{a\}, \{b\}, \{c, d\}\}$, $\{\{a, b\}, \{c, d\}\}$,
$\{\{b\}, \{a, c, d\}\}$, $\{\{a\}, \{b, c, d\}\}$, $\{\{a, b, c, d\}\}$.

(3) Uvažujme rozklad $\Pi_7 = \{Z_0, Z_1, \ldots, Z_6\}$ na množině celých čísel $\mathbb{Z}$ takový, že každá třída Z_i rozkladu Π_7 obsahuje právě ta celá čísla, která jsou dělitelná číslem 7 se zbytkem i. To jest máme $Z_0 = \{0, 7, -7, 14, -14, \ldots\}$, $Z_1 = \{1, -1, 8, -8, 15, -15, \ldots\}$ a tak dále. Analogicky bychom mohli uvažovat rozklad Π_n pro každé $n \in \mathbb{N}$. Π_n se nazývá *systém zbytkových tříd modulo n.*

(4) Na množině racionálních čísel $\mathbb{Q}$ můžeme uvažovat rozklad Π takový, že pro každé $q \in \mathbb{Q}$ máme $[q]_\Pi = \{q, -q\}$. Jiným rozkladem na $\mathbb{Q}$ může být například systém $\Pi = \{Z, \{0\}, K\}$, kde $Z = \{x \in \mathbb{Q} \mid x < 0\}$, $K = \{x \in \mathbb{Q} \mid x > 0\}$.

Rozklady a ekvivalence vyjadřují de facto totéž.

Pozorný čtenář si jistě všiml, že ukázky rozkladů v předchozím příkladu byly analogické příkladům ekvivalencí z příkladu 3.12. Například je vidět, že v bodu (4) předchozího příkladu jsme uvedli ukázky rozkladů, které odpovídají třídám ekvivalence uvedených v bodu (4) příkladu 3.12. Vskutku, ekvivalence na množině popisují „totéž“ jako rozklady na množině o čemž se přesvědčíme ve zbytku kapitoly.

Věta 3.16. *Nechť Π je rozklad na X. Pak binární relace E_Π na X definovaná*

$$\langle x, y\rangle \in E_\Pi, \quad \textit{právě když} \quad [x]_\Pi = [y]_\Pi \tag{7}$$

je ekvivalence.

Důkaz. Pro každý $x \in X$ platí $[x]_\Pi = [x]_\Pi$ triviálně, to jest $\langle x, x\rangle \in E_\Pi$, čímž jsme prokázali reflexivitu E_Π. Nechť $\langle x, y\rangle \in E_\Pi$, tedy $[x]_\Pi = [y]_\Pi$, odtud zřejmě $[y]_\Pi = [x]_\Pi$, tedy $\langle y, x\rangle \in E_\Pi$. Nechť $\langle x, y\rangle \in E_\Pi$ a $\langle y, z\rangle \in E_\Pi$, pak $[x]_\Pi = [y]_\Pi = [z]_\Pi$, to jest $\langle x, z\rangle \in E_\Pi$. □

Definice 3.17. Ekvivalence E_Π definovaná (7) se nazývá *ekvivalence příslušná rozkladu* Π.

Nyní víme, že každému rozkladu přísluší ekvivalence. Dále prokážeme, že ke každé ekvivalenci přísluší rozklad a navíc, že rozklady a ekvivalence jsou ve vzájemně jednoznačné korespondenci. Nejprve si však dokážeme některé základní vlastnosti tříd ekvivalence.

Lemma 3.18. *Nechť E je ekvivalence na X. Pak platí*

(i) $x \in [x]_E$,
(ii) $y \in [x]_E$, *právě když* $\langle y, x\rangle \in E$,
(iii) $x \in [y]_E$, *právě když* $y \in [x]_E$,
(iv) $[x]_E = [y]_E$, *právě když* $y \in [x]_E$.

Důkaz. (i) plyne z reflexivity E, (ii) plyne ze symetrie E, (iii) je důsledek bodu (ii).
(iv): Pokud $[x]_E = [y]_E$, pak dle (i) máme $y \in [y]_E = [x]_E$. Předpokládejme tedy $y \in [x]_E$, to jest $\langle x, y\rangle \in E$. Pokud $z \in [x]_E$, pak $\langle z, x\rangle \in E$ dle (ii), z tranzitivity potom $\langle z, y\rangle \in E$, odtud $z \in [y]_E$ dle (ii). Prokázali jsme $[x]_E \subseteq [y]_E$. Opačná inkluze se prokazuje analogicky. Dohromady $[x]_E = [y]_E$. □

Věta 3.19. *Nechť E je ekvivalence na X. Pak systém množin $\Pi_E \subseteq 2^X$ definovaný*

$$\Pi_E = \{[x]_E \mid x \in X\} \tag{8}$$

je rozklad na množině X.

Důkaz. Postupně pro Π_E ověříme podmínky (i)–(iii) definice 3.13.
(i): Nechť $Y \in \Pi_E$, pak dle (8) existuje $x \in X$ tak, že $Y = [x]_E$. To jest $x \in [x]_E = Y \neq \emptyset$.
(ii): Předpokládejme, že $[x]_E \cap [y]_E \neq \emptyset$. Pak existuje $z \in X$, kde $z \in [x]_E$ a $z \in [y]_E$.

To jest máme $\langle x, z\rangle \in E$ a $\langle z, y\rangle \in E$, dále užitím tranzitivity $\langle x, y\rangle \in E$, což znamená $y \in [x]_E$, odtud $[x]_E = [y]_E$ dle bodu (iv) lemmy 3.18.
(iii): Jelikož $x \in [x]_E$ pro každý $x \in X$, máme $X \subseteq \bigcup_{x\in X}[x]_E = \bigcup \Pi_E$. $\bigcup \Pi_E \subseteq X$ platí triviálně, to jest $\bigcup \Pi_E = X$. □

Definice 3.20. Rozklad Π_E definovaný (8) se nazývá *rozklad příslušný ekvivalenci E.*

Rozklady a ekvivalence jsou ve vzájemně jednoznačné korespondenci.

Mějme ekvivalenci E a příslušný rozklad Π_E. Pokud uvažujeme ekvivalenci E_{Π_E} příslušnou Π_E, pak zřejmě $\langle x, y\rangle \in E$, právě když $[x]_{\Pi_E} = [y]_{\Pi_E}$, to je právě když $\langle x, y\rangle \in E_{\Pi_E}$. Dostáváme tak vztah $E = E_{\Pi_E}$. Analogické tvrzení lze ukázat pro rozklady, což dohromady shrnuje následující

Důsledek 3.21. *Pro ekvivalenci E na X a pro rozklad Π na X máme $E_{\Pi_E} = E$, $\Pi_{E_\Pi} = \Pi$.* □

Rozklad na množině X příslušný ekvivalenci E označujeme běžně X/E místo Π_E a někdy jej nazýváme *faktorová množina X podle E.* Jelikož $[x]_E = [x]_{\Pi_E}$, říkáme, že $[x]_E$ *je třída rozkladu množiny X podle E.* Uvažujeme-li o vztahu množiny X a rozkladu X/E, pak víme, že každému $x \in X$ přísluší třída rozkladu $[x]_E$, pro kterou $x \in [x]_E$. Pro ekvivalenci E na X tedy můžeme uvažovat zobrazení $f_E\colon X \to X/E$, kde

$$f_E(x) = [x]_E \tag{9}$$

Přirozené zobrazení je vždy surjektivní.

pro každý $x \in X$, a nazýváme jej *přirozené* (*kanonické*) *zobrazení.* Zcela očividně platí, že každé přirozené zobrazení f_E je surjektivní, protože každá třída $Y \in X/E$ je neprázdná, existuje tedy $y \in Y$, odtud $f_E(y) = [y]_E = Y$. Dále je vidět, že f_E je injektivní (a tím pádem bijekce), právě když $[x]_E = \{x\}$ pro každé $x \in X$, což je právě když $E = \omega_X$.

Průvodce studiem

Faktorová množina X/E představuje „zjednodušující pohled“ na výchozí množinu X, při kterém jsme ztotožnili ty prvky X, které od sebe nebyly rozlišitelné ekvivalencí E. Pro konečnou X a $E \neq \omega_X$ navíc platí, že faktorová množina X/E je ostře menší než výchozí množina X, to jest platí $|X/E| < |X|$. Faktorizace jako obecná metoda zmenšení výchozí množiny (třeba souboru dat) má aplikace v informatice například při analýze dat a shlukování.

Přirozené zobrazení lze chápat jako *zobrazení indukované ekvivalencí.* Nyní se budeme věnovat opačnému fenoménu, to jest budeme se snažit definovat ekvivalenci pro dané zobrazení. Nejprve si však řekněme, že kromě ekvivalencí a rozkladů existují další přirozené pohledy na to „jak zjednodušit nazírání“ na výchozí množinu. Jedním z nich je *surjektivní zobrazení.* Pokud je zobrazení $f\colon X \to Y$ surjektivní, pak lze obraz $f(x)$ prvku x chápat jako vyjádření: „prvek x nahradíme (zjednodušíme) prvkem $f(x)$“, neboli „$f(x)$ je zjednodušeným pohledem na x“. Surjektivita f zaručuje, že každý prvek z Y je „zjednodušeným pohledem“ na nějaký $x \in X$. Opět lze ukázat, že surjektivní zobrazení a ekvivalence mají zvláštní vztah. Pro zobrazení $f : X \to Y$ definujeme binární relaci E_f na X předpisem

$$\langle x, y\rangle \in E_f, \quad \text{právě když} \quad f(x) = f(y). \tag{10}$$

E_f je ekvivalence o čemž se můžete snadno přesvědčit. Ekvivalence E_f definovaná vztahem (10) se nazývá *ekvivalence indukovaná zobrazením f.* Nyní můžeme prokázat následující větu.

Věta 3.22 (o přirozeném zobrazení)**.** *Nechť $g : X \to Y$ je zobrazení. Pak existuje injektivní zobrazení $h\colon X/E_g \to Y$ takové, že $g = f_{E_g} \circ h$. Pokud je navíc g surjektivní, pak je h bijekce.*

Důkaz. Nejprve si uvědomme, že hledané h je zobrazení z faktorové množiny X/E_g, kde E_g je ekvivalence indukovaná zobrazením g, do množiny Y. Můžeme tedy položit $h([x]_{E_g}) = g(x)$ pro každé $x \in X$. Zobrazení h je zavedeno jednoznačně – hodnota $h([x]_{E_g})$ nezávisí na výběru prvku z třídy rozkladu $[x]_{E_g}$. Vskutku, pro $x' \in [x]_{E_g}$ máme $\langle x, x'\rangle \in E_g$, to jest $g(x) = g(x')$, odtud

$$h([x]_{E_g}) = g(x) = g(x') = h([x']_{E_g}).$$

Máme-li $g(x) = g(x')$, pak $\langle x, x'\rangle \in E_g$, to jest $[x]_{E_g} = [x']_{E_g}$, tedy zobrazení h je injektivní. Pro přirozené zobrazení $f_{E_g}\colon X \to X/E_g$ navíc platí $f_{E_g}(x) = [x]_{E_g}$, to jest

$$g(x) = h([x]_{E_g}) = h(f_{E_g}(x)) = (f_{E_g} \circ h)(x)$$

pro každé $x \in X$, což dokazuje $g = f_{E_g} \circ h$. Pokud je navíc g surjektivní, pak pro každé $y \in Y$ existuje $x \in X$ tak, že $y = g(x) = h([x]_{E_g})$, tedy i h je surjektivní, to jest h je bijekce. □

Poznámka 3.23. Podle předchozí věty tedy platí, že pokud je $g\colon X \to Y$ surjektivní zobrazení, to jest zobrazení nahrazující prvky z X jejich zjednodušenými protějšky z Y, pak Y je stejně mohutná množina jako faktorová množina X/E_g, kterou získáme rozkladem výchozí množiny X ekvivalencí indukovanou zobrazením g. V tomto smyslu lze tedy „zjednodušení“ dané surjektivním zobrazením g nahradit faktorovou množinou X/E_g. Opačně, pro faktorovou množinu X/E lze uvažovat přirozené (surjektivní) zobrazení $f_E\colon X \to X/E$. Ve smyslu „zjednodušení pohledu“ a „nerozlišitelnost“ jsou tedy ekvivalence a surjektivní zobrazení vzájemně nahraditelné.

Příklad 3.24. Vezměme surjektivní zobrazení $g : \mathbb{Z} \to \{-1, 0, 1\}$, které každému celému číslu $z \in \mathbb{Z}$ přiřazuje prvek z $\{-1, 0, 1\}$ předpisem

$$g(z) = \begin{cases} -1 & \text{pokud } z < 0, \\ 1 & \text{pokud } z > 0, \\ 0 & \text{jinak.} \end{cases}$$

Zobrazení g tedy reprezentuje funkci *signum.* Faktorová množina X/E_g se skládá ze tří tříd rozkladu: $\{x \in \mathbb{Z} \mid x < 0\}$, $\{0\}$ a $\{x \in \mathbb{Z} \mid x > 0\}$. Z intuitivního pohledu g i X/E_g reprezentují zjednodušení množiny celých čísel, které jsme získali „odhlédnutím od konkrétní číselné hodnoty a soustředěním se pouze na znaménko“. Poznamenejme, že i na úrovni konečné množiny X/E_g lze dělat řadu úvah o vlastnostech celé množiny $\mathbb{Z}$.

3.4 Uspořádání

Relace uspořádání je dalším typem speciální binární relace na množině. Uspořádání má mnoho interpretací. Na jednu stranu se na něj lze dívat jako na abstraktní relaci, jejíž speciální případy jsou relace „srovnávání čísel“, které dobře známe. Na druhou stranu ale uspořádání mohou mít rysy, které nemají přímou analogii při prostém porovnávání čísel. V některých případech je uspořádání interpretovatelné jako relace určující „hierarchii“, případně „závislosti“. Abychom hned na začátku předešli nedorozuměním, upozorněme na to, že uspořádání, kterému se budeme ve zbytku kapitoly věnovat, *nemá nic společného* se „škatulkováním“ – to jest kategorizací (prvků jisté množiny) podle nějakých jejich vlastností.

Uspořádání je v informatice pojem zcela zásadní ačkoliv si to někdy neuvědomujeme. Mezi základní výbavu každého informatika patří znalost *problému třídění* a typických *třídících algoritmů.* Problém třídění jako takový však de facto nemá smysl uvažovat pokud bychom na množině klíčů, podle kterých třídíme, nezavedli nějakou smysluplnou relaci uspořádání – obvykle ji však chápeme jako „určenou daným kontextem“ a explicitně ji nezdůrazňujeme. Uspořádání množin může výrazně zvýšit efektivitu některých algoritmů, například vyhledávání a podobně.

Definice 3.25. Reflexivní a tranzitivní binární relace R na X se nazývá *kvaziuspořádání.* Antisymetrické kvaziuspořádání se nazývá *uspořádání.* Úplné uspořádání se nazývá *lineární uspořádání* neboli *řetězec.* Pokud je R uspořádání na X, pak se $\langle X, R\rangle$ nazývá *uspořádaná množina.*

Relaci uspořádání na X obvykle značíme $\leq$ v souladu s intuitivním chápáním uspořádání a místo $\langle x, y\rangle \in \leq$ píšeme $x \leq y$. Zdůrazněme ale, že označení $\leq$ v tuto chvíli nemá (obecně) nic společného se srovnáváním čísel, na které jsme zvyklí. Pro uspořádání dále přijímáme značení $y \geq x$ jako zkratku za $x \leq y$. Pro vyjádření faktu $x \leq y$ a $x \neq y$ budeme používat stručný zápis $x < y$ a analogicky $x > y$.

Kvaziuspořádání obecně není antisymetrické, to jest je-li $\leq$ kvaziuspořádání na X, pak mohou existovat prvky $x, y \in X$, kde $x \leq y$, $y \leq x$ a $x \neq y$. Pokud je $\leq$ uspořádání, pak tato situace nastat nemůže, protože z $x \leq y$ a $y \leq x$ plyne $x = y$. Z předchozí definice je také patrný vztah mezi kvaziuspořádáním, uspořádáním a ekvivalencí:

- *symetrické kvaziuspořádání je ekvivalence,*
- *antisymetrické kvaziuspořádání je uspořádání.*

Uspořádání pořád ještě není formálním protějškem „uspořádání“ na které jsme zvyklí při porovnávání čísel. Je-li $\langle X, \leq\rangle$ uspořádaná množina, pak mohou existovat $x, y \in X$ (definice to nevylučuje), pro které neplatí ani $x \leq y$ ani $x \geq y$. V tomto případě říkáme, že prvky $x, y \in X$ jsou *nesrovnatelné,* což někdy značíme $x \parallel y$. V opačném případě (buď $x \leq y$ nebo $y \leq x$) řekneme, že prvky $x, y \in X$ jsou *srovnatelné.* Je-li $\leq$ lineární uspořádání na X, pak je $\leq$ úplná relace, což znamená, že každé dva prvky jsou srovnatelné. Lineární uspořádání tedy lze chápat jako matematický protějšek „tradičního srovnávání čísel“.

V řetězci jsou každé dva prvky srovnatelné.

Každá relace identity ω_X je uspořádání, které nazýváme *antiřetězec.* Je-li $\leq$ na X antiřetězec (jinými slovy: $\leq = \omega_X$), pak pro každé dva různé $x, y \in X$ máme $x \parallel y$. Antiřetězce jsou v jistém smyslu nejmenší uspořádání, protože každé uspořádání $\leq$ na X obsahuje ω_X.

Příklad 3.26. (1) Příkladem kvaziuspořádání, které není ani ekvivalence ani uspořádání je například relace $R = \{\langle a,a\rangle, \langle a,b\rangle, \langle b,a\rangle, \langle b,b\rangle, \langle c,a\rangle, \langle c,b\rangle, \langle c,c\rangle\}$ na množině $X = \{a, b, c\}$.

(2) Následující relace R jsou uspořádání na $X = \{x \mid \langle x, x\rangle \in R\}$:

(a) $R = \{\langle a,a\rangle, \langle b,b\rangle, \langle c,c\rangle\}$,
(b) $R = \{\langle a,a\rangle, \langle b,b\rangle, \langle c,a\rangle, \langle c,b\rangle, \langle c,c\rangle, \langle d,a\rangle, \langle d,b\rangle, \langle d,c\rangle, \langle d,d\rangle\}$,
(c) $R = \{\langle a,a\rangle, \langle b,a\rangle, \langle b,b\rangle, \langle c,a\rangle, \langle c,b\rangle, \langle c,c\rangle\}$,
(d) $R = \{\langle a,a\rangle, \langle b,a\rangle, \langle b,b\rangle, \langle c,a\rangle, \langle c,c\rangle, \langle d,a\rangle, \langle d,b\rangle, \langle d,d\rangle, \langle e,a\rangle, \langle e,b\rangle, \langle e,c\rangle, \langle e,d\rangle, \langle e,e\rangle\}$,
(e) $R = \{\langle a,a\rangle, \langle b,b\rangle, \langle c,a\rangle, \langle c,b\rangle, \langle c,c\rangle, \langle d,a\rangle, \langle d,b\rangle, \langle d,d\rangle\}$.

(3) Číselné množiny $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, ... jsme běžně zvyklí uspořádávat relací „menší rovno“, přitom tato relace je zjevně reflexivní, antisymetrická, tranzitivní i úplná – jedná se tedy o lineární uspořádání, kterému budeme dál říkat *přirozené uspořádání čísel* (přirozených, celých, racionálních, ...). Uvědomme si však, že přirozené uspořádání

Přirozené uspořádání množiny čísel není jediné

čísel není jediné možné uspořádání číselných množin! Vezměme si například množinu $\mathbb{N}$ a pro $x, y \in \mathbb{N}$ položme $x \leq y$, právě když x dělí y beze zbytku. Pak například $2 \leq 4$, ale $2 \not\leq 3$, protože 2 dělí 3 se zbytkem 1. Snadno nahlédneme, že takto zavedené $\leq$ je rovněž uspořádání na $\mathbb{N}$, které není lineární (například $2 \parallel 3$). Na $\mathbb{N}$ ale existují i lineární uspořádání různá od přirozeného uspořádání, dokonce je jich nekonečně mnoho. Označíme-li například $\leq$ přirozené uspořádání $\mathbb{N}$, pak $R = (\leq - \{\langle 1,2\rangle\}) \cup \{\langle 2,1\rangle\}$ je lineární uspořádání, ve kterém jsme oproti $\leq$ „zaměnili dvojku za jedničku“, to jest $2 < 1 < 3 < 4 < \cdots$.

(4) Uvažujme nyní množinu pravdivostních hodnot $X = \{0,1\}$. Pro $x, y \in X$ položíme $x \leq y$, právě když $x \to^{\cdot} y = 1$. Z vlastností logické operace $\to^{\cdot}$ plyne, že $\leq$ je lineární uspořádání na X, pro které platí $0 \leq 1$, to jest slovně: „nepravda je menší než pravda“.

(5) Relace R z bodu (5) příkladu 3.2 na straně 61 je uspořádání, které značíme $\subseteq$ a nazýváme jej množinová inkluze. Pro obecná U není $\subseteq$ lineární uspořádání. Analogicky bychom mohli zavést uspořádání $\supseteq$: pro $A, B \in 2^U$ klademe $A \supseteq B$, právě když B je podmnožina A. Obě dvě relaci mají zřejmý vztah, jedná se o vzájemně inverzní relace.

Pozorování z bodu (5) předchozího příkladu zobecňuje následující zřejmý princip.

Věta 3.27 (princip duality)**.**
Nechť $\leq$ je uspořádání na X. Pak $\leq^{-1}$ je uspořádání na X, které označujeme $\geq$. □

Nyní si řekněme něco o znázorňování konečných uspořádaných množin. Konečné uspořádání $\leq$ na X je relace, tím pádem ji můžeme reprezentovat binární maticí $\mathbf{M}^{\leq}$ nebo příslušným orientovaným grafem $\langle X, \leq\rangle$. Díky speciálním vlastnostem konečných uspořádání je však můžeme znázorňovat mnohem přehledněji pomocí speciálních diagramů. Ke každému uspořádání $\leq$ na X lze uvažovat odvozenou relaci $\prec$ definovanou předpisem

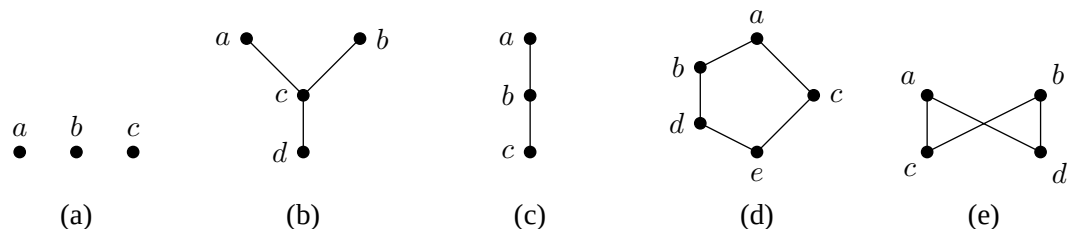
$$x \prec y, \text{ právě když } x < y \text{ a pro každé } z \in X \text{ platí: pokud } x \leq z \leq y \text{ pak } z \in \{x, y\}. \tag{11}$$

Relaci $\prec$ nazýváme *pokrytí* příslušné $\leq$, výraz $x \prec y$ čteme „x je pokryt y“ nebo „y pokrývá x“. Zřejmě máme $\prec \subseteq \leq$, to jest relace $\prec$ je obsažena v uspořádání $\leq$. Relace pokrytí je dle definice irreflexivní, asymetrická (plyne z vlastností $<$) a obecně není tranzitivní. Význam pokrytí příslušného uspořádané množině $\langle X, \leq\rangle$ je následující: $x \prec y$ znamená, že $x < y$ a zároveň neexistuje žádný prvek $z \in X$, který by se nacházel „mezi x a y“ vzhledem k uspořádání $\leq$. Relace $\prec$ tedy zachycuje informaci o prvcích, které se „bezprostředně pokrývají“. Pro uspořádání $\leq$ na konečné množině zřejmě máme $\leq = \mathrm{Tra}(\mathrm{Ref}(\prec))$, obecně to však neplatí. Uvažujme například množinu racionálních čísel $\mathbb{Q}$ a její přirozené uspořádání $\leq$. V tomto případě máme $\prec = \emptyset$, protože mezi každými dvěma různými racionálními čísly $x, y \in \mathbb{Q}$, $x < y$ existuje $z \in \mathbb{Q}$ tak, že $x < z < y$. Zřejmě tedy $\leq \neq \mathrm{Tra}(\mathrm{Ref}(\prec)) = \omega_{\mathbb{Q}}$.

Konečné uspořádané množiny znázorňujeme Hasseovými diagramy.

Na relaci pokrytí je založena jedna z metod jak znázornit konečnou uspořádanou množinu, tak zvané *Hasseovy diagramy uspořádaných množin.* Diagramy jsou složeny z uzlů reprezentujících prvky množiny X a hran, které vyznačují relaci pokrytí $\prec$ příslušnou danému uspořádání $\leq$ na X. Podrobněji, prvky množiny X znázorníme jako uzly (to jest „body“) v rovině tak, aby v případě, kdy $x < y$, ležel bod x níže než bod y. Dva body $x, y \in X$ spojíme v diagramu hranou (úsečkou), právě když $x \prec y$. Horizontální umístění bodů je vhodné volit tak, aby pokud možno nedocházelo ke křížení hran.

Příklad 3.28. Na obrázku 6 jsou zobrazeny Hasseovy diagramy relací uvedených v bodu (2) příkladu 3.26, v diagramech jsou pro přehlednost vyznačeny odpovídající



Obrázek 6: Hasseovy diagramy uspořádaných množin

uzly. Všimněte si, že obrázek (a) odpovídá tříprvkovému antiřetězci a obrázek (c) odpovídá tříprvkovému řetězci. Hasseův diagram dané uspořádané množiny má obecně mnoho možných nakreslení.

Nyní se budeme zabývat existencí speciálních prvků v uspořádaných množinách a jejich vzájemnému vztahu. Například vezmeme-li přirozené uspořádání $\leq$ na množině $\mathbb{N}$, pak o číslu 1 říkáme, že je „nejmenší“. Správně bychom ale měli říkat „nejmenší vzhledem k uspořádání $\leq$“, protože vlastnosti jako jsou „být nejmenší“, „být minimální“ a podobně, jsou těsně vázány k uvažovanému uspořádání. Nyní si tyto a analogické speciální prvky uspořádaných množin přesně zavedeme.

Definice 3.29. Nechť $\langle X, \leq \rangle$ je uspořádaná množina. Prvek $x \in X$ se nazývá

- *minimální,* jestliže pro každý $y \in X$ platí: pokud $y \leq x$, pak $x = y$,
- *nejmenší,* jestliže $x \leq y$ pro každý $y \in X$,
- *maximální,* jestliže pro každý $y \in X$ platí: pokud $y \geq x$, pak $x = y$,
- *největší,* jestliže $x \geq y$ pro každý $y \in X$.

Poznámka 3.30. V definici minimálního a nejmenšího prvku je podstatný rozdíl, i když to na první pohled možná není zřejmé. Prvek $x \in X$ je nejmenší, právě když je $x \in X$ menší (ve smyslu $\leq$) než všechny ostatní prvky z X. Minimalita prvku $x \in X$ znamená, že neexistuje žádný prvek, který by byl ostře menší (ve smyslu $<$) než x. Je téměř zřejmé, že pokud je x nejmenší, pak je také minimální, ale obráceně to již platit nemusí. Analogická situace nastává pro největší a maximální prvky. Nejprve si vztahy prvků ukážeme na příkladech.

Příklad 3.31. (1) Budeme-li uvažovat číselnou množinu $\mathbb{N}$ a její přirozené uspořádání $\leq$, pak číslo 1 je nejmenším a zároveň minimálním prvkem v $\langle \mathbb{N}, \leq \rangle$. Žádný největší ani maximální prvek v $\langle \mathbb{N}, \leq \rangle$ neexistuje. Pokud bychom uvažovali množinu $\mathbb{N} - \{1\}$, to jest množinu přirozených čísel bez jedničky, a pokud bychom na ní zavedli uspořádání $\leq$ předpisem: $x \leq y$, právě když x dělí y beze zbytku, pak by $\langle \mathbb{N} - \{1\}, \leq \rangle$ měla nekonečně mnoho minimálních prvků, kterými by byla právě všechna prvočísla. Na druhou stranu $\langle \mathbb{N} - \{1\}, \leq \rangle$ by neměla žádný nejmenší prvek, ani žádný největší či maximální prvek. Například $\mathbb{Q}$ uspořádaná přirozeným uspořádáním nemá žádný ze speciálních prvků uvedených v definici 3.29.

(2) Uvažujme uspořádané množiny dané diagramy na obrázku 6. Ad (a): prvky a, b, c jsou všechny zároveň maximální i minimální, žádný největší ani nejmenší prvek neexistuje. Ad (b): prvek d ne nejmenší a zároveň minimální, prvky a, b jsou maximální, žádný největší prvek neexistuje. Ad (c): a je největší a maximální, c je nejmenší a minimální. Ad (d): a je největší a maximální, e je nejmenší a minimální. Ad (e): a, b jsou maximální, c, d jsou minimální, žádné největší ani nejmenší prvky neexistují. Všimněte si, že v Hasseově diagramu jsou maximální prvky reprezentovány těmi uzly, které nemají žádné pokrytí – neexistuje žádný výše zakreslený uzel, se kterým by byl tento uzel spojen čarou. Největší prvek poznáme z Hasseova diagramu tak, že pod ním leží všechny ostatní prvky – je tedy zakreslen v diagramu nejvýš a do každého prvku se z něj lze dostat obecně přes několik hran. Analogicky pro minimální a nejmenší prvky.

(3) Potenční množina 2^U uspořádaná množinovou inkluzí $\subseteq$ má nejmenší a zároveň minimální prvek, kterým je $\emptyset$ a dále má i největší a zároveň maximální prvek, kterým je množina U.

V předchozím příkladu jsme mohli vypozorovat vztahy speciálních prvků z definice 3.29. Některé z nich si nyní prokážeme. V důkazu následující věty vhodně využijeme *principu duality* uvedeného ve větě 3.27. Přímo z definice 3.29 totiž pro každé $x \in X$ plyne, že x je největší (maximální / nejmenší / minimální) prvek v uspořádané množině $\langle X, \leq \rangle$, právě když je x nejmenší (minimální / největší / maximální) prvek v $\langle X, \geq \rangle$.

Věta 3.32. *Nechť $\langle X, \leq \rangle$ je uspořádaná množina. Pak platí*

(i) *v $\langle X, \leq \rangle$ existuje nejvýš jeden největší a jeden nejmenší prvek;*

(ii) *je-li $x \in X$ největší (nejmenší) prvek, pak je také maximální (minimální) a žádné další maximální (minimální) prvky se v X nevyskytují;*

(iii) *pokud je $\leq$ lineární uspořádání, pak je $x \in X$ největší (nejmenší) prvek, právě když je maximální (minimální).*

Důkaz. (i): Nechť $x \in X$ je největší prvek. Ukážeme, že pokud $y \in Y$ splňuje definiční podmínky největšího prvku, pak máme $x = y$. Pokud pro $y \in X$ platí: $z \leq y$ pro každé $z \in X$, pak máme i $x \leq y$. Jelikož je x největší prvek, máme $y \leq x$. To jest z antisymetrie dostáváme $x = y$. Tvrzení pro nejmenší prvek dostaneme užitím principu duality.
(ii): Nechť $x \in X$ je největší prvek a nechť $y \geq x$, pak máme rovněž $y \leq x$, odtud z antisymetrie $x = y$, to jest x je maximální. Pokud je $y \in X$ maximální prvek, pak $y \leq x$ implikuje $x = y$. Zbytek tvrzení plyne z principu duality.
(iii): Nechť $\leq$ je lineární uspořádání. Vzhledem k bodu (ii) stačí ukázat, že pokud je $x \in X$ maximální, pak je největší. Nechť je tedy x maximální a nechť $y \in X$. Jelikož je $\leq$ úplná relace, máme buď $x \leq y$ nebo $y \leq x$. Pokud $y \leq x$, jsme hotovi. Pokud $x \leq y$, pak z maximality plyne $y = x \leq x$. To jest x je největší prvek. Zbytek plyne z principu duality. □

Teď obrátíme naši pozornost k prvkům uspořádané množiny $\langle X \leq \rangle$, které mají speciální význam vzhledem k některým podmnožinám X. Uvažujeme-li například podmnožinu $Y \subseteq X$, můžeme se ptát, zda-li v X existuje nějaký prvek, který je větší (menší) než všechny prvky z Y. Pokud prvek těchto vlastností existuje, můžeme jej chápat jako „horní (dolní) mez“ množiny Y.

Definice 3.33. Nechť $\langle X, \leq \rangle$ je uspořádaná množina a nechť $Y \subseteq X$. Definujeme množiny,

$$\mathrm{L}(Y) = \{x \in X \mid x \leq y \text{ platí pro každé } y \in Y\}, \tag{12}$$

$$\mathrm{U}(Y) = \{x \in X \mid x \geq y \text{ platí pro každé } y \in Y\}. \tag{13}$$

$\mathrm{L}(Y)$ se nazývá *dolní kužel množiny Y v $\langle X, \leq \rangle$*. $\mathrm{U}(Y)$ se nazývá *horní kužel množiny Y v $\langle X, \leq \rangle$*.

Jinak řečeno, dolní kužel Y v $\langle X, \leq \rangle$ obsahuje právě ty prvky z X, které jsou menší nebo rovny všem prvkům obsaženým v Y, analogicky pro horní kužel. Definice nám okamžitě říká, jak dolní (horní) kužel nalézt, viz následující příklady.

Příklad 3.34. (1) Mějme uspořádanou množinu $\langle X, \leq \rangle$. Pro $Y = \emptyset$ máme $\mathrm{L}(\emptyset) = X = \mathrm{U}(\emptyset)$. Podrobněji, pro každý $x \in X$ je předpoklad $y \in \emptyset$ nepravdivý, to jest z vlastnosti implikace víme, že tvrzení „pokud $y \in \emptyset$, pak $x \leq y$ (případně $x \geq y$)“ platí triviálně pro každé $x \in X$.

(2) Vezmeme-li $\mathbb{Z}$ a její přirozené uspořádání, pak například máme $\mathrm{L}(\mathbb{N}) = \{x \in \mathbb{Z} \mid x \leq 1\}$, $\mathrm{U}(\{0\}) = \mathbb{N} \cup \{0\}$, $\mathrm{U}(\{x \in \mathbb{N} \mid x \text{ je sudé číslo}\}) = \emptyset$. Pro $\mathbb{Q}$ a její

přirozené uspořádání máme například $\mathrm{L}(\{\frac{1}{n} \mid n \in \mathbb{N}\}) = \mathrm{L}(\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\}) = \{q \in \mathbb{Q} \mid q \leq 0\}$.

(3) Uvažujme uspořádané množiny dané diagramy na obrázku 6. Uvedeme si některé zajímavé horní a dolní kužely. Ad (a): Máme $\mathrm{U}(\{x\}) = \{x\} = \mathrm{L}(\{x\})$ pro libovolný $x \in X$. Ad (b): $\mathrm{U}(\{a,b\}) = \emptyset$, $\mathrm{L}(\{a,b\}) = \{c,d\}$. Ad (c): $\mathrm{U}(\{a,b,c\}) = \{a\}$, $\mathrm{L}(\{a,b,c\}) = \{c\}$. Ad (d): $\mathrm{U}(\{b,c\}) = \mathrm{U}(\{c,d\}) = \{a\}$, $\mathrm{L}(\{b,c\}) = \mathrm{L}(\{c,d\}) = \{e\}$. Ad (e): $\mathrm{U}(\{a,b\}) = \emptyset$, $\mathrm{L}(\{a,b\}) = \{c,d\}$, $\mathrm{U}(\{c,d\}) = \{a,b\}$, $\mathrm{L}(\{c,d\}) = \emptyset$.

Řekli jsme, že horní a dolní kužel $Y \subseteq X$ je možné interpretovat jako množiny, které danou podmnožinu Y omezují shora a zdola. Platí navíc, že $\langle \mathrm{U}(Y), \leq_{\mathrm{U}(Y)} \rangle$, kde $\leq_{\mathrm{U}(Y)}$ je relace vzniklá ze $\leq$ zúžením na množinu $\mathrm{U}(Y)$, je opět uspořádaná množina. Duálně $\langle \mathrm{L}(Y), \leq_{\mathrm{L}(Y)} \rangle$ je uspořádaná množina. Můžeme tedy bez újmy zkoumat kužely coby uspořádané množiny. Obzvláště zajímavé pro nás je, pokud horní kužel $\mathrm{U}(Y)$ obsahuje nejmenší prvek, protože tento prvek lze interpretovat jako nejmenší horní mez podmnožiny Y v $\langle X, \leq \rangle$. Duálně k tomu, největší prvek dolního kuželu $\mathrm{L}(Y)$ lze interpretovat (pokud existuje) jako největší dolní mez podmnožiny Y v $\langle X, \leq \rangle$. Tyto speciální prvky nyní zavedeme jako supremum a infimum:

Definice 3.35. Nechť $\langle X, \leq \rangle$ je uspořádaná množina a nechť $Y \subseteq X$. Pokud má $\mathrm{L}(Y)$ největší prvek, pak se nazývá *infimum* Y a označuje se $\inf(Y)$. Pokud má $\mathrm{U}(Y)$ nejmenší prvek, pak se nazývá *supremum* Y a označuje se $\sup(Y)$.

Speciálně pro $\{x, y\} \subseteq X$ píšeme $\inf(x, y)$ místo $\inf(\{x, y\})$, stejně tak pro supremum. Supremum a infimum dané množiny obecně nemusí existovat. Ukažme si nejprve některé příklady.

Příklad 3.36. (1) Vezměme $\mathbb{N}$ uspořádanou přirozeným uspořádáním $\leq$. Pak zřejmě pro každé $x, y \in \mathbb{N}$ máme $\inf(x, y) = \min(x, y)$ a $\sup(x, y) = \max(x, y)$. V tomto případě tedy infimum i supremum existuje pro každé dva prvky. Dále zřejmě máme $\mathrm{L}(\mathbb{N}) = \{1\}$, tedy $\inf(\mathbb{N}) = 1$. Například ale $\mathrm{U}(\mathbb{N}) = \emptyset$, to jest $\sup(\mathbb{N})$ neexistuje. Zde vidíme, že i když supremum existuje ke každým dvěma prvkům z $\mathbb{N}$, nemusí nutně existovat k libovolné podmnožině $\mathbb{N}$.

(2) Vezměme $\mathbb{N} - \{1\}$ a položme $x \leq y$, právě když x dělí y beze zbytku. Pak například pro čísla 8 a 12 máme $\mathrm{L}(8, 12) = \{2, 4\}$, to jest $\inf(8, 12) = 4$. Analogicky $\mathrm{U}(8, 12) = \{24, 48, 72, 96, \dots\}$, tedy $\sup(8, 12) = 24$. Například ale $\inf(2, 3)$ neexistuje, protože $\mathrm{L}(2, 3) = \emptyset$. Supremum existuje ke každým dvěma prvkům a $\sup(x, y)$ je nejmenším společným násobkem x a y. Infimum existuje ke každým dvěma soudělným číslům a je rovno jejich největšímu společnému děliteli.

(3) Uvažujme uspořádané množiny dané diagramy na obrázku 6, uvedeme si infima a suprema některých podmnožin. Ad (a): $\inf(x, y)$ a $\sup(x, y)$ existuje, právě když $x = y$. Ad (b): $\inf(a, b) = c$, $\sup(a, b)$ neexistuje, $\inf(x, d) = d$ pro každé $x \in \{a, b, c, d\}$. Ad (c): Infimum a supremum existuje k libovolné podmnožině. Ad (d): $\sup(b, c) = \sup(c, d) = a$, $\inf(b, c) = \inf(c, d) = e$. Ad (e): $\sup(a, b)$ neexistuje (zde je $\mathrm{U}(\{a, b\}) = \emptyset$), $\inf(a, b)$ neexistuje (zde je sice $\mathrm{L}(\{a, b\}) = \{c, d\} \neq \emptyset$, ale $c \parallel d$, to jest $\mathrm{L}(\{a, b\})$ nemá největší prvek), analogicky $\sup(c, d)$ neexistuje ani $\inf(c, d)$ neexistuje.

Na základě existence infima či suprema ke každým dvěma prvkům definujeme speciální uspořádané množiny zvané polosvazy a svazy.

Definice 3.37. Nechť $\langle X, \leq \rangle$ je uspořádaná množina. Pokud pro každé $x, y \in X$ existuje $\inf(x, y)$, pak $\langle X, \leq \rangle$ nazveme *průsekový polosvaz.* Pokud pro každé $x, y \in X$ existuje $\sup(x, y)$, pak $\langle X, \leq \rangle$ nazveme *spojový polosvaz.* Je-li $\langle X, \leq \rangle$ průsekový i spojový polosvaz, pak $\langle X, \leq \rangle$ nazveme *svaz.*

Svaz je tedy uspořádaná množina, kde ke každým dvěma prvkům existuje jejich infimum i supremum. Nosiče svazu značíme obvykle L místo X, to jest svaz značíme $\langle L, \leq \rangle$. Dalším zřejmým pozorováním, které plyne z principu duality je fakt, že $\langle X, \leq \rangle$ je průsekový (spojový) polosvaz, právě když je $\langle X, \geq \rangle$ spojový (průsekový) polosvaz.

Ve svazu existuje ke každým dvěma prvků infimum a supremum.

Příklad 3.38. (1) Každá lineárně uspořádaná množina $\langle X, \leq \rangle$ je svaz, protože pro každé dva prvky $x, y \in X$ máme buď $x \leq y$, v tom případě $\inf(x, y) = x$ a $\sup(x, y) = y$, nebo platí $y \leq x$, v tom případě $\inf(x, y) = y$ a $\sup(x, y) = x$. Speciálně tedy číselné množiny $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ uspořádané přirozeným uspořádáním jsou svazy, ve kterých infima a suprema odpovídají minimu a maximu z obou prvků. Obecně v nich ale neexistují infima a suprema nekonečných podmnožin.

(2) $\mathbb{N} - \{1\}$ z bodu (2) příkladu 3.36 je spojový polosvaz, ale nejedná se o průsekový polosvaz. Kdybychom však dělitelností uspořádali celou množinu $\mathbb{N}$, pak by se jednalo i o průsekový polosvaz a tudíž o svaz.

(3) Uvažujme uspořádané množiny dané diagramy na obrázku 6, pak (a) není polosvaz (ani průsekový, ani spojový), (b) je průsekový polosvaz, ale nejedná se o spojový polosvaz, (c) a (d) jsou svazy, (e) není polosvaz (ani průsekový, ani spojový).

Na závěr podotkněme, že indukcí se lze snadno přesvědčit, že pokud je $\langle L, \leq \rangle$ svaz, pak v něm existuje supremum i infimum pro libovolnou konečnou a neprázdnou podmnožinu L o čemž se lze snadno přesvědčit matematickou indukcí: pro $x \in L$ máme triviálně $\inf(\{x\}) = x$; nechť infimum existuje pro každou $n-1$ prvkovou podmnožinu L, pak zřejmě

$$\inf(\{x_1, \ldots, x_n\}) = \inf(\{x_1, \ldots, x_{n-1}\}, x_n).$$

Tvrzení lze prokázat duálně pro suprema. Toto pozorování má následující důsledek. Pokud je $\langle L, \leq \rangle$ konečný svaz, to jest L je konečná množina, pak existuje $\inf(L)$ a $\sup(L)$, což dle definice infima a suprema znamená, že v $\langle L, \leq \rangle$ je $\inf(L)$ nejmenší prvek a $\sup(L)$ je největší prvek. Jinými slovy, každý konečný svaz má nejmenší prvek (značený 0) a největší prvek (značený 1). U nekonečných svazů to obecně neplatí, viz příklad 3.38.

Každý konečný svaz má největší a nejmenší prvek.

Shrnutí

Relace ekvivalence reprezentují matematický protějšek nerozlišitelnosti, jedná se o reflexivní, symetrické a tranzitivní relace. Ekvivalence jsou ve vzájemně jednoznačné korespondenci s rozklady na množině, to jest s disjunktními pokrytími množiny. Ekvivalence mají rovněž vztah k surjektivním zobrazením.
Kvaziuspořádání je reflexivní a tranzitivní relace. Relace uspořádání je reflexivní, antisymetrická a tranzitivní relace. Lineární uspořádání je reflexivní, antisymetrická, tranzitivní a úplná relace. V uspořádaných množinách se mohou nacházet speciální prvky (největší, maximální, nejmenší, minimální). Dalšími speciálními prvky uspořádaných množin vzhledem k jistým podmnožinám jsou infimum a supremum. Průsekový polosvaz je uspořádaná množina, ve které ke každým dvěma prvkům existuje jejich infimum, spojový polosvaz je uspořádaná množina, ve které ke každým dvěma prvkům existuje jejich supremum. Svaz je uspořádaná množina, ve které ke každým dvěma prvkům existuje jejich infimum i supremum.

Pojmy k zapamatování

- ekvivalence, třída ekvivalence, rozklad na množině, třída rozkladu,
- ekvivalence příslušná rozkladu, rozklad příslušný ekvivalenci, faktorová množina,
- přirozené zobrazení, ekvivalence indukovaná zobrazením,

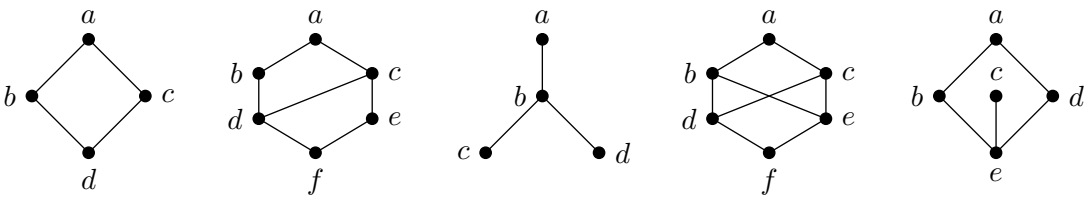
- kvaziuspořádání, uspořádání, uspořádaná množina, lineární uspořádání, řetězec, antiřetězec,
- princip duality, pokrytí, Hasseův diagram,
- minimální prvek, nejmenší prvek, maximální prvek, největší prvek,
- dolní kužel, horní kužel, infimum, supremum,
- spojový polosvaz, průsekový polosvaz, svaz

Kontrolní otázky

1. *Co víte o intuitivním významu ekvivalence?*
2. *Jaký je vztah rozkladu k ekvivalenci?*
3. *Proč je každá ekvivalenční třída neprázdná?*
4. *Jak vypadá nejmenší a největší ekvivalence na dané množině?*
5. *Jaký je rozdíl mezi kvaziuspořádáním, uspořádáním a ekvivalencí?*
6. *Co říká princip duality a jaký má intuitivní význam?*
7. *Která ekvivalence je zároveň uspořádání?*
8. *Mohou existovat v lineárně uspořádané množině dva různé maximální prvky?*

Cvičení

1. Rozhodněte, které z následujících relací R jsou ekvivalence. Proveďte diskusi.
   (a) $X = \{0,1,2,3\}$, $R = \{\langle 0,0\rangle, \langle 1,1\rangle, \langle 1,2\rangle, \langle 1,3\rangle, \langle 2,1\rangle, \langle 2,2\rangle, \langle 2,3\rangle, \langle 3,2\rangle, \langle 3,3\rangle\}$,
   (b) $X = 2^U$, $R = \{\langle A,B\rangle \mid A,B \in X,\ |A| = |B|\}$,
   (c) $X = 2^U$, $R = \{\langle A,B\rangle \mid A,B \in X,\ A \cap B \neq \emptyset\}$,
   (d) $R = S \cap S^{-1}$, kde S je reflexivní a tranzitivní relace na X.
2. K následujícím rozkladům Π nalezněte ekvivalence E_Π.
   (a) $\Pi = \{\{a,b,c\},\{d\},\{e,f\},\{g\}\}$,
   (b) $X = 2^{\{a,b\}}$, $\Pi = \{\{\{a\},\{a,b\}\},\{\emptyset,\{b\}\}\}$,
   (c) $X = \mathbb{N}$, $\Pi = \{\{0,2,4,\dots\},\{1\},\{3\},\{5\},\dots\}$.
3. K následujícím ekvivalencím E na X nalezněte rozklady Π_E.
   (a) $X = \{0,\dots,4\}$, $E = \{\langle 0,2\rangle, \langle 0,4\rangle, \langle 2,0\rangle, \langle 2,4\rangle, \langle 4,0\rangle, \langle 4,2\rangle\} \cup \omega_X$,
   (b) $X = \{0,1,2,3,4,5,6,7,8,9\}$, $E = \{\langle x,y\rangle \mid$ rozdíl $x - y$ je dělitelné třemi beze zbytku$\}$,
   (c) $X = \mathbb{N} \times \mathbb{N}$, $E = \{\langle\langle m,n\rangle,\langle p,q\rangle\rangle \mid m,n,p,q \in \mathbb{N},\ m+n = p+q\}$.
4. Určete, které z následujících relací R na X jsou kvaziuspořádání, uspořádání, případně řetězce.
   (a) $X = \{1,2,3,4\}$, $R = \{\langle 1,1\rangle, \langle 1,3\rangle, \langle 1,4\rangle, \langle 2,2\rangle, \langle 3,4\rangle, \langle 4,4\rangle\}$,
   (b) $X = \mathbb{Z}$, $\langle x,y\rangle \in R$, právě když buď $|x| = |y|$ a $x \geq y$, nebo $|x| < |y|$,
   (c) $X = \mathbb{Q}$, $R = \{\langle x,y\rangle \mid x,y \in \mathbb{Q},\ x^2 \leq y^2\}$,
   (d) $X = \mathbb{N} \times \mathbb{N}$, $R = \{\langle\langle m,n\rangle,\langle p,q\rangle\rangle \mid m,n,p,q \in \mathbb{N},\ m \leq p$ a $n \leq q\}$.
5. Uspořádané množiny zobrazené na následujících diagramech zapište binárními maticemi.



6. Vraťte se k předchozímu příkladu a určete nejmenší, minimální, největší a maximální prvky.

7. Vraťte se k předchozímu příkladu a rozhodněte, které z diagramů reprezentují průsekové případně spojové polosvazy či svazy.

Úkoly k textu

1. Nechť R je binární relace na X. Dokažte, že relace $E = \mathrm{Tra}(\mathrm{Sym}(\mathrm{Ref}(R)))$ je ekvivalence na X obsahující R a pro každou ekvivalenci E' na X obsahující R platí $E \subseteq E'$.

2. Pro každou množinu X označme $E(X)$ systém všech ekvivalencí na X. To jest například pro $X = \{a, b, c\}$ máme $E(X) = \{\omega_X, E_1, E_2, E_3, \iota_X\}$, kde E_1, E_2 a E_3 jsou ekvivalence z bodu (2) příkladu 3.12 na straně 67. Dokažte následující tvrzení.

   (a) Pro každé $E_1, E_2 \in E(X)$ platí $E_1 \cap E_2 \in E(X)$.
   (b) $\langle E(X), \subseteq \rangle$ je průsekový polosvaz, kde $\inf(E_1, E_2) = E_1 \cap E_2$.
   (c) Existuje X a $E_1, E_2 \in E(X)$ tak, že $E_1 \cup E_2$ není tranzitivní.
   (d) Pro každé $E_1, E_2 \in E(X)$ platí $\mathrm{Tra}(E_1, E_2) \in E(X)$.
   (e) $\langle E(X), \subseteq \rangle$ je spojový polosvaz, kde $\sup(E_1, E_2) = \mathrm{Tra}(E_1 \cup E_2)$.
   (f) $\langle E(X), \subseteq \rangle$ je svaz s nejmenším prvkem ω_X a největším prvkem ι_X.

   Předchozí tvrzení dokazujte postupně a vhodně využijte již prokázaných tvrzení.

Řešení

1. (a) není ekvivalence, protože není symetrická; (b) je ekvivalence; (c) pro $|X| = 1$ je ekvivalence, pro $|X| \geq 2$ není tranzitivní; (d) je ekvivalence: reflexivita a tranzitivita plyne z reflexivity a tranzitivity S a S^{-1}, symetrie plyne ze vztahu S a S^{-1}.
2. Ekvivalence mají následující tvar:
   (a) $\{\langle a,b\rangle, \langle a,c\rangle, \langle b,a\rangle, \langle b,c\rangle, \langle c,a\rangle, \langle c,b\rangle, \langle e,f\rangle, \langle f,e\rangle\} \cup \omega_{\{a,b,c,d,e,f,g\}}$,
   (b) $\{\langle\emptyset,\emptyset\rangle, \langle\emptyset,\{b\}\rangle, \langle\{b\},\emptyset\rangle, \langle\{b\},\{b\}\rangle, \langle\{a\},\{a\}\rangle, \langle\{a\},\{a,b\}\rangle, \langle\{a,b\},\{a\}\rangle, \langle\{a,b\},\{a,b\}\rangle\}$,
   (c) $\{\langle x,y\rangle \mid x, y \in \mathbb{N},\ x, y$ jsou obě sudá, nebo $x = y\}$.
3. Rozklady mají následující tvar:
   (a) $\{\{0,2,4\},\{1\},\{3\}\}$,
   (b) $\{\{0,3,6,9\},\{1,4,7\},\{2,5,8\}\}$,
   (c) $\{\{\langle 1,1\rangle\}, \{\langle 1,2\rangle, \langle 2,1\rangle\}, \{\langle 1,3\rangle, \langle 2,2\rangle, \langle 3,1\rangle\}, \{\langle 1,4\rangle, \langle 2,3\rangle, \langle 3,2\rangle, \langle 4,1\rangle\}, \dots\}$.
4. (a) není reflexivní, (b) řetězec, (c) kvaziuspořádání (není antisymetrické), (d) uspořádání.
5. Matice jsou uvedeny zleva doprava, pořadí sloupců a řádků dle abecedy.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

6. Zleva doprava: a je největší a maximální, d je nejmenší a minimální; a je největší a maximální, f je nejmenší a minimální; a je největší a maximální, c, d jsou minimální; a je největší a maximální, f je nejmenší a minimální; a, c jsou maximální, e je nejmenší a minimální.
7. Zleva doprava: svaz, svaz, spojový polosvaz, —, průsekový polosvaz.

4 Grafy a stromy

Studijní cíle: Po prostudování kapitoly 4 by student měl rozumět základním pojmům teorie grafů. Měl by dále znát základní tvrzení, která pro probírané pojmy platí. K vybraným úlohám teorie grafů by student měl znát algoritmy pro jejich řešení.

Klíčová slova: orientovaný graf, neorientovaný graf, izomorfismus grafů, podgraf, sled, tah, cesta, kružnice, vzdálenost vrcholů, ohodnocený graf, souvislost, komponenta, hledání cest, stupeň vrcholu, skóre, eulerovský tah, strom, kostra, hledání minimální kostry, kořenový strom

4.1 Co a k čemu jsou grafy

V životě se často setkáváme se situacemi, ve kterých existují určité objekty a spojení mezi nimi. Mezi některými objekty spojení existují, mezi jinými ne. Objekty mohou být křižovatky ve městě, spojeními pak ulice mezi jednotlivými křižovatkami. Někdy přitom na orientaci spojení nezáleží (je jedno, jestli vede spojení z objektu A do objektu B nebo vede-li spojení z B do A), někdy na orientaci záleží (může se stát, že vede spojení z A do B ale neexistuje spojení z B do A). V situaci s křižovatkami a ulicemi pro chodce na orientaci spojení nezáleží. Jsou-li totiž křižovatky A a B spojeny ulicí, která je lemována chodníkem, chodec o orientaci spojení neuvažuje, protože může jít z A do B i z B do A. Pro chodce je tedy důležité, že A a B jsou spojena, ale orientace spojení nehraje roli. Pro řidiče ale na orientaci záleží. Křizovatky A a B mohou být totiž spojeny jednosměrkou a pak je důležité, jestli spojení vede z A do B nebo z B do A. Jiným příkladem podobné situace je každý náčrtek, kde jsou body (objekty) spojené čarami (schema elektrického obvodu, vývojový diagram, schema hierarchické struktury ve firmě apod.).

Grafy představují místa a spojení mezi nimi.

Uvedenými situacemi se zabývá teorie grafů. Objekty se nazývají vrcholy, spojení pak hrany. Graf je dán množinou vrcholů a množinou hran mezi nimi. Nezáleží-li na orientaci hran, nazývá se graf neorientovaný, v opačném případě se nazývá orientovaný.

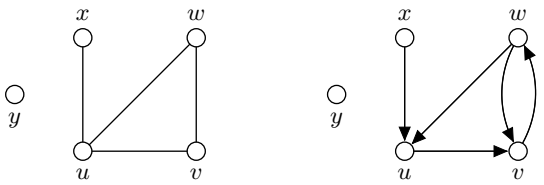
Grafy mají řadu různorodých aplikací.

Grafy mají řadu zajímavých aplikací. Ve městě může být pekařská firma, která má každé ráno do prodejen pečiva rozvézt zboží. Majiteli firmy přitom záleží na tom, aby se zbytečně neplýtvalo pohonnými hmotami, tj. aby byl při ranním rozvozu počet ujetých kilometrů co nejmenší. Z pohledu teorie grafů jde o úlohu najít cestu, která vychází z pekařství, prochází v libovolném pořadí všemi prodejnami pečiva a končí opět v pekařství. Přitom hledáme cestu, která je ze všech takových nejkratší. Jiný příklad představuje cestující, který chce najít co nejrychlejší vlakové spojení z jedné stanice do druhé. Může ji najít vyhledávacím programem na internetu. Samotný program vlastně hledá nejkratší cestu v grafu, který představuje železniční síť.

4.2 Neorientované a orientované grafy: základní pojmy

Příklad neorientovaného grafu vidíme na obr. 7 vlevo. Má vrcholy u, v, w, x a y. Vrcholy jsou znázorněny kroužky. Úsečky na obrázku znázorňují hrany. Například úsečka spojující kroužky označené x a u představuje hranu mezi vrcholy x a u. Mezi vrcholy x a y ale hrana není. Protože v neorientovaném grafu nemají hrany orientaci, můžeme hranu mezi vrcholy u a v reprezentovat neuspořádanou dvojicí, tedy dvouprvkovou množinou $\{u, v\}$. Hovoříme pak o hraně $\{u, v\}$. Neorientovaný graf tedy sestává z množiny vrcholů a množiny hran, tj. neuspořádaných dvojic vrcholů.

Příklad orientovaného grafu je na obr. 7 vpravo. Má opět vrcholy u, v, w, x a y. Hrany jsou ale orientované a jsou znázorněny šipkami. Reprezentujeme je tedy uspořádanými



Obrázek 7: Neorientovaný (vlevo) a orientovaný (vpravo) graf.

dvojicemi. Například hranu z u do v reprezentujeme uspořádanou dvojicí $\langle u, v\rangle$.[8] To nás vede k následující definici.

Definice 4.1 (neorientovaný a orientovaný graf)**.** *Neorientovaný graf* je dvojice $G = \langle V, E\rangle$, kde V je neprázdná množina tzv. *vrcholů* (někdy také *uzlů*) a

$$E \subseteq \{\{u, v\} \mid u, v \in V,\, u \neq v\}$$

je množina dvouprvkových množin vrcholů, tzv. *(neorientovaných) hran*.

Orientovaný graf je dvojice $G = \langle V, E\rangle$, kde V je neprázdná množina tzv. *vrcholů* (*uzlů*) a

$$E \subseteq V \times V$$

je množina uspořádaných dvojic vrcholů, tzv. *(orientovaných) hran*.

V neorientovaném grafu na směru spojení nezáleží, v orientovaném na něm záleží.

Hrana se tedy u neorientovaných grafů chápe jako dvouprvková množina $\{u, v\}$ vrcholů $u, v \in V$. Pak říkáme, že hrana $\{u, v\}$ vede mezi u a v, popř. spojuje u a v. To odpovídá našemu záměru: Graf je neorientovaný, tj. na pořadí vrcholů u hrany nezáleží. U orientovaných grafů se hrana chápe jako uspořádaná dvojice $\langle u, v\rangle$ vrcholů u a v. Pak říkáme, že hrana vede z u do v. I to odpovídá záměru: Graf je orientovaný, tj. na pořadí vrcholů u hrany záleží. Hrana $\langle u, v\rangle$ je tedy něco jiného než hrana $\langle v, u\rangle$. *Koncové vrcholy* hrany jsou u neorientované hrany $\{u, v\}$ vrcholy u a v, u orientované hrany $\langle u, v\rangle$ také vrcholy u a v.

Pokud neuvedeme jinak, budeme předpokládat, že grafy jsou konečné, tedy že mají konečnou množinu vrcholů, a tudíž i hran. Někdy se uvažují i nekonečné grafy. V neorientovaném grafu se někdy připouští hrany $\{v, v\}$, tedy hrany spojující vrchol v se sebou samým (smyčky). Naše definice takové hrany nepřipouští, podle potřeby ji ale lze rozšířit. Orientované smyčky, tj. hrany $\langle v, v\rangle$ ale naše definice připouští.

Příklad 4.2. Uvažujme množinu vrcholů $V = \{u, v, w, x, y\}$. Uvažujme množinu neorientovaných hran $E_1 = \{\{u, v\}, \{u, w\}, \{u, x\}, \{v, w\}\}$. $G_1 = \langle V, E_1\rangle$ je neorientovaný graf a vidíme ho znázorněný na obr. 7 vlevo. Uvažujme teď množinu orientovaných hran $E_2 = \{\langle u, v\rangle, \langle v, w\rangle, \langle w, u\rangle, \langle w, v\rangle, \langle x, u\rangle\}$. Pak $G_2 = \langle V, E_2\rangle$ je orientovaný graf a vidíme ho znázorněný na obr. 7 vpravo.

Příklad 4.3. (a) V je množina křižovatek v daném městě, $\langle k_1, k_2\rangle \in E$, právě když existuje ulice vedoucí z k_1 do k_2.

(b) V je množina všech lidí, $\{c, d\} \in E$, právě když c a d se znají. Tento neorientovaný „graf známostí“ představuje velkou sociální síť. Má zhruba 7,8 mld. vrcholů (tolik žije na Zemi lidí). Počet hran není známý. Poznamenejme v této souvislosti, že každý člověk je schopen udržovat aktivní známost se nejvýše přibližně 150 lidmi (Dunbarovo číslo).

[8] Připomeňme, že uspořádaná dvojice $\langle u, v\rangle$ je jiná než $\langle v, u\rangle$. Na pořadí prvků záleží.

(c) V je množina existujících webových stránek, $\langle p_1, p_2\rangle \in E$, právě když z p_1 vede odkaz (hyperlink) na p_2. Tento graf, tzv. webgraph, reprezentuje web. Odhady jeho velikosti se liší. Podle údajů InternetLiveStats z roku 2020 má asi 1,74 mld. vrcholů, z nichž je asi 25% aktivních; podle WorldWideWebSize.com z roku 2020 má asi 5,49 mld. vrcholů; podle Common Crawl Corpus měl v roce 2014 přibližně 1,72 mld. vrcholů a 64 mld. hran.

(d) V je množina všech neuronů v mozku, $\langle n_1, n_2\rangle \in E$, právě když z neuronu n_1 vede vlákno do n_2. Odhaduje se, že existuje asi 10 mld. neuronů a že každý z nich je spojen s přibližně deseti tisíci dalšími neurony.

Pro (neorientovaný nebo orientovaný) graf $G = \langle V, E\rangle$ se V a E nazývají množina vrcholů a množina hran grafu G a značí se $V(G)$ a $E(G)$. Bude-li z kontextu jasné, jde-li o graf orientovaný nebo neorientovaný, budeme psát jen "graf". Graf můžeme zadat přímo jeho obrázkem. Např. řekneme-li "uvažujme graf z obr. 7 vlevo", lze z tohoto obrázku určit jak množinu V vrcholů, tak množinu E hran. Znázornění grafu obrázkem je přitom přehlednější než jeho popis jakožto struktury $G = \langle V, E\rangle$.

K orientovanému grafu je třeba někdy uvažovat graf, který vznikne zanedbáním orientace hran. Říká se mu symetrizace orientovaného grafu.

Definice 4.4 (symetrizace)**.** *Symetrizace* orientovaného grafu $G = \langle V, E\rangle$ je neorientovaný graf $G' = \langle V, E'\rangle$, kde

$$\{u, v\} \in E' \quad \text{právě když} \quad \langle u, v\rangle \in E \text{ nebo } \langle v, u\rangle \in E.$$

Například graf na obr. 7 vlevo je symetrizací grafu vpravo.

Obrázek daného grafu není určen jednoznačně. Dva různé obrázky přitom mohou popisovat v zásadě stejné grafy, byť to na první pohled není patrné. Říkáme-li „v zásadě stejné", myslíme tím, že mají stejnou strukturu. V případě, že je graf dán obrázkem, mohou se obrázky dvou v zásadě stejných grafů lišit rozmístěním vrcholů, zakreslením hran, popř. také označením vrcholů. Grafy, které mají stejnou strukturu, se nazývají izomorfní.

Definice 4.5 (izomorfní grafy)**.** Nechť $G_1 = \langle V_1, E_1\rangle$ a $G_2 = \langle V_2, E_2\rangle$ jsou neorientované grafy. Bijekce $h : V_1 \to V_2$ se nazývá *izomorfismus* G_1 do G_2, pokud pro každé vrcholy $u, v \in V_1$ je

$$\{u, v\} \in E_1 \quad \text{právě když} \quad \{h(u), h(v)\} \in E_2.$$

Nechť $G_1 = \langle V_1, E_1\rangle$ a $G_2 = \langle V_2, E_2\rangle$ jsou orientované grafy. Bijekce $h : V_1 \to V_2$ se nazývá *izomorfismus* G_1 do G_2, pokud pro každé vrcholy $u, v \in V_1$ je

$$\langle u, v\rangle \in E_1 \quad \text{právě když} \quad \langle h(u), h(v)\rangle \in E_2.$$

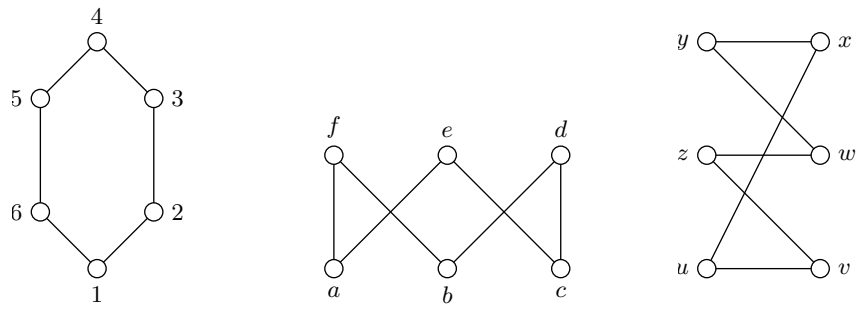
Pokud izomorfismus G_1 do G_2 existuje, nazývají se G_1 a G_2 *izomorfní*. V takovém případě píšeme $G_1 \cong G_2$.

Definice vlastně říká, že izomorfní grafy se liší jen „přejmenováním vrcholů". Toto přejmenování zabezpečuje bijektivní zobrazení h. Snadno se ověří, že relace být izomorfní je ekvivalence na třídě všech grafů (ověřte). Díky symetrii této relace tedy můžeme říkat, že G_1 a G_2 jsou izomorfní (pořadí grafů není podstatné; je-li $h : V_1 \to V_2$ izomorfismus G_1 do G_2, je $h^{-1} : G_2 \to G_1$ izomorfismus G_2 do G_1).

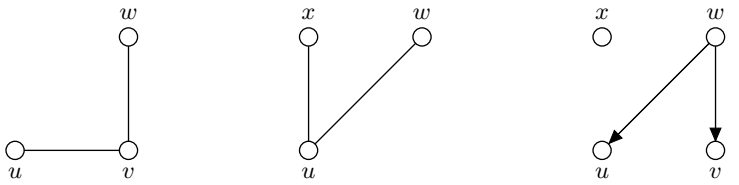
Příklad 4.6. Všechny grafy z obr. 8 jsou po dvou izomorfní. Např. bijekce h, pro kterou

$$h(1) = a,\ h(2) = e,\ h(3) = c,\ h(4) = d,\ h(5) = b,\ h(6) = f,$$

je izomorfismus prvního a druhého grafu.



Obrázek 8: Izomorfní neorientované grafy.



Obrázek 9: Podgrafy.

Části grafů se nazývají podgrafy.

Definice 4.7 (podgrafy)**.** (Orientovaný nebo neorientovaný) graf $\langle V_1, E_1 \rangle$ je *podgrafem* grafu $\langle V_2, E_2 \rangle$, právě když $V_1 \subseteq V_2$ a $E_1 \subseteq E_2$. Podgraf $\langle V_1, E_1 \rangle$ grafu $\langle V_2, E_2 \rangle$ se nazývá *indukovaný*, právě když E_1 obsahuje každou hranu z E_2, jejíž oba koncové vrcholy patří do V_1.

První dva grafy na obr. 9 jsou podgrafy grafu z obr. 7 vlevo, přitom první z nich není indukovaný (není v něm hrana $\{u, w\}$), druhý ano. Třetí graf na obr. 9 je podgrafem grafu z obr. 7 vpravo.

Důležitou oblastí je tzv. cestování v grafech. Představíme-li si vrcholy grafu jako místa a hrany jako spojnice, po kterých lze z míst do míst přecházet, lze se ptát, zda se z jednoho místa lze dostat do jiného, jaká je nejkratší cesta z jednoho do druhého a podobně. Úloh o cestování existuje celá řada. Pro jejich řešení byly vyvinuty efektivní algoritmy. Tyto úlohy nacházejí uplatnění i jinde, než tam, kde jde o skutečné cestování a kde třeba řidič auta potřebuje najít nejrychlejší cestu z výchozího do cílového města. Graf může například reprezentovat počítačovou síť a cestujícím může být blok informací (paket) v této síti. Představme si základní pojmy.

Definice 4.8 (cestování)**.** *Sled* v (neorientovaném nebo orientovaném) grafu[9] $G = \langle V, E \rangle$ je posloupnost

$$v_0, e_1, v_1, e_2, v_2, \ldots, e_n, v_n,$$

kde $v_i \in V$ jsou vrcholy, $e_j \in E$ jsou hrany a platí, že

- $e_i = \{v_{i-1}, v_i\}$ pro $i = 1, \ldots, n$, je-li G neorientovaný,
- $e_i = \langle v_{i-1}, v_i \rangle$ pro $i = 1, \ldots, n$, je-li G orientovaný.

Číslo n se nazývá *délka sledu*. Sled $v_0, e_1, v_1, e_2, v_2, \ldots, e_n, v_n$, se nazývá

[9]Názvosloví je nejednotné. Tedy to, co my budeme nazývat sled, tah a cesta, se v jiné literatuře může nazývat jinak.

- *uzavřený*, je-li $v_0 = v_n$,
- *tah*, neopakuje-li se v něm žádná hrana (tj. pro $i \neq j$ je $e_i \neq e_j$),
- *cesta*, neopakuje-li se v něm žádný vrchol (tj. pro $i \neq j$ je $v_i \neq v_j$),
- *kružnice*, je-li uzavřený a s výjimkou vrcholů v_0 a v_n jsou každé jeho dva vrcholy různé.

Vzdálenost z vrcholu u do vrcholu v je délka cesty z u do v, která má ze všech cest z u do v nejmenší délku.

Říkáme také, že sled $v_0, e_1, \ldots, v_n$ vede z v_0 do v_n. Z definice máme, že každý tah je sledem. Každá cesta je tahem, neboť neopakují-li se ve sledu vrcholy, nemohou se opakovat ani hrany. Kružnice nemůže být cestou, protože se v ní opakují vrcholy (první a poslední).

V grafu na obr. 7 vlevo je posloupnost $u, \{u,w\}, w, \{w,v\}, v, \{v,u\}, u, \{u,w\}, w$ sled, který není tahem (a tedy ani cestou), protože se v něm opakuje hrana $\{u,w\}$. Sled $u, \{u,w\}, w, \{w,v\}, v, \{v,u\}, u, \{u,x\}, x$ je tah, který není cestou, protože se v něm opakuje vrchol u. Sled $x, \{x,u\}, u, \{u,w\}, w, \{w,v\}, v, \{v,u\}, u, \{u,x\}, x$ je sice uzavřený, ale není to kružnice, protože se v něm opakuje vrchol u. Sled $u, \{u,w\}, w, \{w,v\}, v, \{v,u\}, u$, je kružnice.

Existují tedy sledy, které nejsou cestami. Následující věta ukazuje, že pokud nám jde o dosažitelnost z vrcholu do vrcholu, vystačíme s cestami.

Věta 4.9. *Existuje-li v grafu sled z vrcholu u do vrcholu v, existuje také cesta z u do v.*

Důkaz. Důkaz je jednoduchý. Opakuje-li se ve sledu $u, \ldots, v$ nějaký vrchol w, tj. má-li sled tvar $u, \ldots, w, \ldots, w, \ldots, v$, vynecháme posloupnost $w, \ldots,$. Dostaneme $u, \ldots, w, \ldots, v$, což je také sled z u do v. Pokud je už cestou, jsme hotovi. Pokud ne, opět vynecháme posloupnost mezi opakujícími se vrcholy. Protože je sled konečný, po konečném počtu kroků takto skončíme u cesty z u do v. □

Jaký význam mají pojmy z definice 4.8? Sled odpovídá putování bez omezení: Vyjdeme z nějakého místa, po libovolné hraně z něho přejdeme do jiného místa a tak dále, až dojdeme do koncového místa. Najít vhodný tah se bude snažit např. poštovní doručovatelka. Kdyby šla po hraně (tj. po ulici) vícekrát, zbytečně se nachodí. Hledáním vhodných cest se zabývají např. ve spedičních firmách při rozvozu zboží: Projet místem, kde se vyloží část zboží (tedy vrcholem příslušného grafu), vícekrát je příznakem nehospodárného naplánování rozvozu.

Následující pojmy zavedeme pro neorientované grafy. Pro orientované najdete návod v úkolech k textu.

Definice 4.10 (souvislost a komponenty)**.** Neorientovaný graf $G = \langle V, E \rangle$ se nazývá *souvislý*, právě když pro každé dva vrcholy $u, v \in V$ existuje sled z u do v. *Komponenta* neorientovaného grafu je každý jeho maximální souvislý podgraf.

Komponenta grafu $G = \langle V, E \rangle$ je tedy jeho podgraf $G' = \langle V', E' \rangle$, který je souvislý a pro který platí, že je-li $G'' = \langle V'', E'' \rangle$ souvislý podgraf grafu G, pro který $V' \subseteq V''$ a $E' \subseteq E''$, pak $V' = V''$ a $E' = E''$. Snadno se vidí, že komponenta grafu G je jeho podgraf $G' = \langle V', E' \rangle$, který je indukovaný množinou vrcholů $V' \subseteq V$ takovou, že každé dva vrcholy z V' lze spojit cestou a že k V' není možné přidat další vrchol z V, aby to stále platilo. Např. graf na obr. 7 vlevo není souvislý (vrcholy x a y nejsou spojeny

sledem). Jeho podgraf indukovaný vrcholy u, v a w je souvislý, ale není to komponenta, protože není maximální souvislý. Komponenty v tomto grafu jsou dvě. První je podgraf indukovaný vrcholy u, v, w, x, druhá je podgraf indukovaný vrcholem y. I obecně platí, že komponenty tvoří „rozklad grafu“.

Komponenty tvoří rozklad grafu.

Věta 4.11. *Nechť $G_1 = \langle V_1, E_1\rangle$, …, $G_n = \langle V_n, E_n\rangle$ jsou všechny komponenty grafu $G = \langle V, E\rangle$. Pak každý vrchol $v \in V$ patří právě do jedné V_i a každá hrana $e \in E$ patří právě do jedné E_i.*

Důkaz. Vezměme vrchol $v \in V$. Podgraf indukovaný $\{v\}$ je zřejmě souvislý. Proto je podgrafem nějakého maximálního souvislého podgrafu grafu G, tj. komponenty G_i. Proto $v \in V_i$, tj. v patří aspoň do jedné z množin $V_1, \ldots, V_n$. Ukažme, že v nemůže patřit do dvou různých $V_i \neq V_j$. Kdyby $v \in V_i \cap V_j$, pak uvažujme nějaký $v_i \in V_i - V_j$ (takový existuje, protože G_i a G_j jsou komponenty, a tedy $V_i \not\subseteq V_j$ a $V_j \not\subseteq V_i$). Protože G_i je souvislý, existuje sled $v_i, e_1, u_1, \ldots, v$. Uvažujme množinu vrcholů V', která vznikne z V_j přidáním všech vrcholů sledu $v_i, e_1, u_1, \ldots, v$. Pak $V_j \subseteq V'$ a podgraf indukovaný množinou V' je souvislý (vezmeme-li $v_1, v_2 \in V'$, pak existuje sled z v_1 do v i sled z v do v_2 a složením těchto sledů dostaneme sled z v_1 do v_2). Tedy G_j by nebyl maximální souvislý podgraf, tj. nebyl by komponentou, což je spor s předpokladem.

Že každá hrana $e \in E$ patří právě do jedné E_i dostaneme podobnou úvahou, když si uvědomíme, že pro $e = \{u, v\}$ je podgraf indukovaný množinou vrcholů $\{u, v\}$ je souvislý. □

Ulice, kterou v grafu reprezentujeme hranou, má ve skutečnosti nějakou délku, popř. propustnost. Stejně tak sklad, reprezentovaný v grafu pomocí vrcholu, může mít určitou kapacitu. V grafech se takovým doplňujícím informacím říká ohodnocení.

Definice 4.12 (ohodnocení)**.** *Hranové ohodnocení* grafu $\langle V, E\rangle$ s množinou hodnot D je funkce $w : E \to D$. *Vrcholové ohodnocení* grafu $\langle V, E\rangle$ s množinou hodnot D je funkce $w : V \to D$.

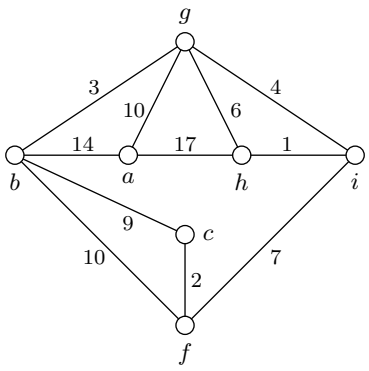
Je-li jasné, o jaké ohodnocení jde, říkáme jen ohodnocení. Grafu spolu s ohodnocením říkáme také hranově, popř. vrcholově ohodnocený graf, popř. jen ohodnocený graf. Množinou hodnot D je většinou nějaká množina čísel (to budeme automaticky předpokládat). Hodnota $w(e) \in D$ přiřazená funkcí w hraně $e \in E$ představuje např. délku hrany (vzdálenost mezi místy), její kapacitu (propustnost informačního nebo dopravního spojení) apod. Hodnota $w(u) \in D$ přiřazená funkcí w vrcholu $u \in V$ představuje např. propustnost uzlu, do kterého něco přichází a něco odchází, apod. Množina hodnot D může obsahovat libovolné prvky, např. názvy ulic nebo datové struktury obsahující strukturovanou informaci o dané hraně či vrcholu.

Příklad 4.13. Na obr. 10 je hranově ohodnocený graf. Hranové ohodnocení w je dáno předpisem $w(\{a, b\}) = 14, w(\{a, g\}) = 10, \ldots, w(\{h, i\}) = 1$.

Představuje-li hranové ohodnocení délky jednotlivých hran, je přirozené zavést pojem délky sledu, který zohledňuje toto ohodnocení. *Délka sledu $v_0, e_1, \ldots, e_n, v_n$ v hranově ohodnoceném grafu* je číslo

$$w(e_1) + \cdots + w(e_n),$$

je to tedy součet délek všech hran, které se ve sledu vyskytují. Např. délka sledu $b, \{b, f\}, f, \{f, i\}, i, \{i, h\}, h$ v grafu na obr. 10 je 18. Podobně jako v neohodnoceném grafu definujeme vzdálenost z u do v jako délku nejkratší cesty (délka se uvažuje vzhledem k ohodnocení). Uvědomte si, že přiřazuje-li hranové ohodnocení w každé hraně číslo 1, je délka sledu v takto ohodnoceném grafu rovna délce sledu v neohodnoceném grafu (viz definice 4.8).



Obrázek 10: Ohodnocený graf.

4.3 Hledání cest

Předpokládejme síť měst se známými vzdálenostmi mezi sousedními městy (tj. některé dvojice měst jsou spojeny silnicemi a my známe délky těchto spojujících silnic). Zajímá nás, jak se dostat z města A do města B nejkratší cestou (tj. tak, abychom ujeli co nejméně kilometrů). Jak to zjistit?

Z hlediska teorie grafů jde o problém hledání nejkratší cesty. Představme si totiž následující neorientovaný graf. Ke každému městu bude v grafu existovat vrchol (pro různá města různé vrcholy). Jsou-li města spojena silnicí, bude mezi jim odpovídajícími vrcholy v grafu hrana. Tento graf hranově ohodnotíme tak, že hodnota hrany bude rovna délce jí odpovídající silnice. Zjistit nejkratší cestu z A do B pak skutečně znamená najít nejkratší cestu (ve smyslu teorie grafů) která vede z uzlu odpovídajícího městu A do uzlu odpovídajícího B. V následujícím uvedeme jeden z nejznáměnších algoritmů hledání nejkratší cesty, tzv. Dijkstrův algoritmus.[10]

Algoritmus pracuje následovně. Na vstupu je neorientovaný graf $G = \langle V, E\rangle$, jeho hranové ohodnocení $w : E \to \mathbb{R}^+$ (každé hraně je přiřazeno kladné reálné číslo), a vrchol $s \in V$. Výstupem algoritmu je pro každý vrchol $v \in V$ číslo $d(v)$, které je vzdáleností z s do v. Algoritmus používá proměnné A, N, d, m, přitom A a N označují množiny vrcholů, d označuje funkci přiřazující vrcholům kladná reálná čísla a m označuje nezáporné reálné číslo. V každém kroku algoritmu je pro vrchol $v \in V$ hodnota $d(v)$ rovna délce nejkratší zatím nalezené cesty z s do v. Na začátku se nastaví $d(s) = 0$ a $d(v) = \infty$ pro ostatní vrcholy $v \neq s$ (krok 1. níže uvedeného algoritmu). Hodnota $d(v) = \infty$ znamená, že žádná cesta do v zatím nebyla nalezena. Tato hodnota se pak u vrcholů, do kterých existuje cesta z s, v průběhu výpočtu zmenšuje, v každém kroku obsahuje délku nejkratší zatím nalezené cesty z s do v, a na konci délku nejkratší cesty z s do v. U vrcholů v, do kterých cesta z s nevede, zůstává $d(v) = \infty$.

Množina A se na začátku nastaví na $A = V$ (krok 1). Během výpočtu obsahuje A ty vrcholy v, pro něž zatím nebyla stanovena konečná hodnota $d(v)$ (tj. $d(v)$ byla stanovena, ale v dalším výpočtu se ještě může změnit).

Algoritmus opakuje následující krok: zjistí nejmenší hodnotu $d(v)$ vrcholů z A (hodnota m v kroku 3). Množinu vrcholů v z A s touto nejmenší hodnotou označí N (krok 3). Z množiny A vyjme všechny vrcholy v, pro které je $d(v)$ nejmenší (krok 3). Vrchol v se tedy odstraní z A a vloží do N, právě když

$$d(v) = \min\{d(u) \mid u \in A\}.$$

[10] Edsger W. Dijkstra [dajkstra] —(1930–2002); jeden z nejvýznamnějších informatiků.

Každý takový vrchol v je pak považován za vrchol, pro nějž byla nalezena nejkratší cesta z s do v, délkou této cesty je $d(v)$, a kratší cesta do v se v dalších krocích algoritmu už nehledá (to je zajištěno tím, že se vrchol odstranil z A).

Vložení vrcholu v do N znamená, že v je považován za vrchol, přes který může do zbývajících vrcholů u množiny A vést kratší cesta, než je dosud nalezená nejkratší cesta do u. V tomto smyslu je tedy každý vrchol v z N kandidátem na zlepšení hodnoty $d(u)$. Algoritmus toto možné zlepšení prověří pro vrcholy u z A, do kterých vede z v hrana.

Algoritmus tedy (krok 4) pro každý $v \in N$ a pro každý $u \in A$, pro který existuje hrana $\{v, u\} \in E$, porovná hodnotu $d(v) + w(\{v, u\})$ (délka možné cesty z s do u, která vede přes v) s hodnotou $d(u)$ (délka dosud nejkratší nalezené cesty z s do u). Je-li $d(v) + w(\{v, u\}) < d(u)$ (tj. cesta přes v je kratší), změní se hodnota $d(u)$ na $d(u) = d(v) + w(\{v, u\})$.

Algoritmus se pak vrátí ke kroku 2. V něm se ověří, zda je v A ještě nějaký vrchol v s hodnotou $d(v) < \infty$ (uvědomte si, že při prvním průchodu krokem 2 je tato podmínka automaticky splněna díky $d(s) = 0$, proto jsme ji zatím nezmiňovali). Pokud ano, znamená to, že v A se nacházejí kandidáti na zlepšení hodnot $d(u)$ a pokračuje se znovu kroky 3 a 4 jako výše, tedy stanovením nové množiny N, odebráním vrcholů z A a tak dále. Pokud ale v A žádný vrchol v s hodnotou $d(v) < \infty$ není, algoritmus skončí.

Na konci je množina A buď prázdná, a to tehdy, když ke všem vrcholům z s cesta existuje, nebo je neprázdná, a obsahuje jen vrcholy v s hodnotou $d(v) = \infty$. Vrcholy s hodnotou $d(v) = \infty$ jsou právě ty, ke kterým z s neexistuje cesta.

Poznamenejme, že pro praktickou implementaci volíme místo ∞ nějakou velkou hodnotu, které nemůže být jinak dosaženo, např. součet délek všech hran zvětšený o 1. Následuje stručný popis algoritmu ($x := y$ znamená „do x přiřaď y“).

Algoritmus 4.14 (nalezení nejkratších cest z daného vrcholu)**.**

Vstup:	graf $G = \langle V, E\rangle$, vrchol $s \in V$,
	hranové ohodnocení $w : E \to \mathbb{R}^+$,
Výstup:	hodnota $d(v)$ pro každý $v \in V$, $d(v)$ je délka
	nejkratší cesty z s do v
Proměnné:	funkce $d : V \to \mathbb{R}^+$, číslo $m \in \mathbb{R}^+$,
	množiny $A, N \subseteq V$

1. $d(s) := 0$; pro každý $v \in V - \{s\}$: $d(v) := \infty$; $A := V$;
2. pokud neexistuje $v \in A$ takový, že $d(v) \neq \infty$, skonči;
3. $m := \min\{d(v) \mid v \in A\}$; $N := \{v \in A \mid d(v) = m\}$; $A := A - N$;
4. pro všechny $v \in N$, $u \in A$ takové, že $\{v, u\} \in E$: jestliže $d(v) + e(\{v, u\}) < d(u)$, pak $d(u) := d(v) + w(\{v, u\})$; pokračuj krokem 2.

Ukažme si činnost algoritmu na jednoduchém příkladě. Uvažujme graf na obr. 10. Budeme hledat nejkratší cesty z h do ostatních vrcholů. Nastavíme tedy $s = h$.

Krok 1: Nastaví se $A := \{a, b, c, f, g, h, i\}$, $d(a) = \infty$, $d(b) = \infty$, $d(c) = \infty$, $d(f) = \infty$, $d(g) = \infty$, $d(h) = 0$, $d(i) = \infty$.

Krok 2: Pokračuje se dál (protože podmínka ukončení není splněna).

Krok 3: Nastaví se $m := 0$, $N := \{h\}$, $A := \{a, b, c, f, g, i\}$.

Krok 4:. Upraví se $d(a) = d(h) + w(\{h, a\}) = 0 + 17 = 17$, $d(g) = d(h) + w(\{h, g\}) = 0 + 6 = 6$, $d(i) = d(h) + w(\{h, i\}) = 0 + 1 = 1$, tedy pro $v \in A$ je $d(a) = 17$, $d(b) = \infty$, $d(c) = \infty$, $d(f) = \infty$, $d(g) = 6$, $d(i) = 1$.

Krok 2: Pokračuje se dál.

Krok 3: Nastaví se $m := 1$, $N := \{i\}$, $A := \{a, b, c, f, g\}$.

Krok 4: Upraví se $d(f) = d(i)+w(\{i, f\}) = 1+7 = 8$, $d(g) = d(i)+w(\{i, g\}) = 1+4 = 5$, tedy pro $v \in A$ je $d(a) = 17$, $d(b) = \infty$, $d(c) = \infty$, $d(f) = 8$, $d(g) = 5$.

Krok 2: Pokračuje se dál.

Krok 3: Nastaví se $m := 5$, $N := \{g\}$, $A := \{a, b, c, f\}$.

Krok 4: Upraví se $d(a) = d(g) + w(\{g, a\}) = 5 + 10 = 15$, $d(b) = d(g) + w(\{g, b\}) = 5 + 3 = 8$, tedy pro $v \in A$ je $d(a) = 15$, $d(b) = 8$, $d(c) = \infty$, $d(f) = 8$.

Krok 2: Pokračuje se dál.

Krok 3: Nastaví se $m := 8$, $N := \{b, f\}$, $A := \{a, c\}$.

Krok 4: Upraví se $d(c) = d(f)+w(\{f, c\}) = 10$, tedy pro $v \in A$ je $d(a) = 15$, $d(c) = 10$.

Krok 2: Pokračuje se dál.

Krok 3: Nastaví se $m := 10$, $N := \{c\}$, $A := \{a\}$.

Krok 4: d se neupravuje, tedy pro $v \in A$ je $d(a) = 15$.

Krok 2: Pokračuje se dál.

Krok 3: Nastaví se $m := 15$, $N := \{a\}$, $A := \emptyset$.

Krok 4: d se neupravuje.

Krok 2: Výpočet se ukončí. Vzdálenosti $d(v)$ z h do v jsou tedy $d(a) = 15$, $d(b) = 8$, $d(c) = 10$, $d(f) = 8$, $d(g) = 5$, $d(h) = 0$, $d(i) = 1$.

K navrženému algoritmu musíme provést důkaz jeho správnosti.

Ručním ověřením zjistíme, že vypočítané vzdálenosti jsou správné. Musí tomu tak být vždy? Tedy, je Dijkstrův algoritmus správný v tom smyslu, že pro každý ohodnocený graf a jeho vrchol s budou po skončení výpočtu $d(v)$ délky nejkratších cest z s do v?

Průvodce studiem

Ke každému navrženému algoritmu je třeba provést důkaz jeho správnosti. Nestačí zjistit, že algoritmus pracuje správně na několika příkladech. Na jiných by mohl dávat nesprávné výsledky. Důkaz správnosti je ověření, že pro jakékoli přípustné hodnoty vstupů algoritmus vypočítá správné výstupy.

Například algoritmus 4.14 vychází z intuice, že hledáme-li nejkratší cestu lokálně, tj. z navštíveného vrcholu se snažíme jít do nejbližšího vrcholu, najdeme cestu, která je nejlepší i globálně. To ovšem není zřejmé a je třeba to ověřit. K tomu slouží důkaz správnosti.

Důkaz správnosti Dijkstrova algoritmu Pro $u, v \in V$ označme $\delta(u, v)$ délku nejkratší cesty z u do v. Předpokládejme, že $d(v)$ jsou hodnoty vypočítané algoritmem pro daný graf $\langle V, E\rangle$, ohodnocení w a vrchol s. Máme dokázat, že pro každý vrchol $v \in V$ je $d(v) = \delta(s, v)$. Dokážeme $d(v) \leq \delta(s, v)$ a $\delta(s, v) \leq d(v)$. Při tom budeme dokazovat indukcí podle počtu průchodů „cyklu“ 2.–4., tedy cyklu sestávajícího z kroků 2., 3., 4. Tím se rozumí následující: výpočet probíhá tak, že je proveden krok 1., pak se buď skončí, nebo se provedou kroky 2., 3., 4. (první průchod cyklem), pak se buď skončí, nebo se provedou kroky 2., 3., 4. (druhý průchod cyklem), atd. Provede-li se cyklus celkem n-krát, můžeme mluvit o 1., 2., ..., n-tém průchodu a o hodnotách proměnných v těchto průchodech. Hodnoty proměnných po i-tém průchodu cyklu 2.–4. budeme značit d_i, m_i, A_i a N_i (tj. např. $A_i = N_i - A_{i-1}$ apod.).

Nejdříve dokážeme

$$\delta(s,v) \leq d(v). \tag{14}$$

Protože $d(v) = d_j(v)$ pro nějaké j, stačí dokázat $\delta(s,v) \leq d_i(v)$. To dokážeme indukcí podle i. Pro $i = 0$ (tj. před prvním průchodem) je to zřejmé. Je totiž $\delta(s,s) = 0 = d_0(s)$ a pro ostatní v je jistě $\delta(s,v) \leq d_0(v) = \infty$. Předpokládejme, že platí $\delta(s,v) \leq d_i(v)$ a dokažme $\delta(s,v) \leq d_{i+1}(v)$. Pro vrchol v jsou dvě možnosti. Buď při $(i+1)$-tém průchodu cyklem v kroku 4 nedojde ke změně, tj. $d_{i+1}(v) = d_i(v)$, a pak je $\delta(s,v) \leq d_{i+1}(v)$ podle indukčního předpokladu. Nebo ke změně dojde, tj. $d_{i+1}(v) = d_i(u) + w(\{u,v\})$ pro nějaký u. Pak ale z indukčního předpokladu máme $\delta(s,u) + w(\{u,v\}) \leq d_i(u) + w(\{u,v\})$, a protože jistě platí $\delta(s,v) \leq \delta(s,u) + w(\{u,v\})$, máme $\delta(s,v) \leq d_i(u) + w(\{u,v\}) = d_{i+1}(v)$. (14) je dokázáno.

Označme nyní $D_1, \ldots, D_k$ všechny od sebe různé vzdálenosti vrcholů grafu od vrcholu s tak, že $0 = D_1 < D_2 < \cdots < D_k$ (tj. $D_1 = 0$ je vzdálenost s od s, D_k je vzdálenost nejvzdálenějšího vrcholu od s). Označme dále $V_i = \{v \in V \mid d(s,v) = D_i\}$ pro $i = 1, \ldots, k$, tj. V_i obsahuje právě vrcholy se vzdáleností D_i od s. Indukcí podle i dokážeme, že $D_i = m_i$ a $V_i = N_i$ pro každý provedený průchod i cyklem 2.–4. Z tohoto tvrzení už plyne požadovaná rovnost $d(v) = \delta(s,v)$: za prvé, každý $v \in V$, do kterého existuje cesta z s, patří do nějaké $V_i = N_i$ (pro ostatní je $d_i(v) = \infty$); za druhé, pro vrcholy v z N_i je $d_i(v) = m_i$ a výsledná vypočtená hodnota $d(v)$ je $d(v) = d_i(v)$ (v se odstraní z A a dál se s nimi nepracuje). Tedy pro každý vrchol v, do kterého existuje z s cesta, je $d(v) = d_i(v) = m_i = d(s,v)$.

Dokažme tedy $D_i = m_i$ a $V_i = N_i$. Pro $i = 1$ je to zřejmé: $D_1 = 0 = m_1$, $N_1 = \{s\} = V_1$. Předpokládejme, že tvrzení platí pro všechna $j < i$ a dokažme ho pro i: Vezměme libovolný $v \in V_i$. Pak podle definice V_i má nejkratší cesta $s, \ldots, u, e, v$ délku D_i. Cesta $s, \ldots, u$ je pak nejkratší cestou z s do u (jinak by $s, \ldots, u, e, v$ nebyla nejkratší z s do v, rozmyslete). Její délka je tedy některou z $D_j < D_i = D_j + w(\{u,v\})$. Podle indukčního předpokladu je $u \in V_j = N_j$, a proto $d_j(v) = D_i$ (podrobněji: kdyby $d_j(v) < D_i$, pak z $d(v) \leq d_j(v)$ je $d(v) < D_j = \delta(s,v)$, spor s (14); na druhou stranu se hodnota D_i do $d_j(v)$ dostane v j-tém cyklu přiřazením $d_j(v) := d_j(u) + w(\{u,v\})$ nebo už bylo $d_{j-1}(v) = D_i$). Proto i $d_i(v) = D_i$ (pro $j < k \leq i$ nemůže být $d_k(v) < d_j(v)$, pak by opět $d(v) < \delta(s,v)$). Tedy pro všechny $v \in V_i$ je $d_i(v) = D_i$.

Ukážeme teď $D_i = m_i$, tj. $D_i = \min\{d_i(u) \mid u \in A_{i-1}\}$. Kdyby existoval $u \in A_{i-1}$ tak, že $d_i(u) < D_i$, pak $u \notin V_i$ (neboť jsme ukázali, že pro $u \in V_i$ je $d_i(u) = D_i$). Tedy buď existuje $j < i$ a $u \in V_j = N_j$, což nelze (protože $N_j \cap A_{i-1} = \emptyset$), nebo existuje $j > i$ a $u \in V_j$, což také nelze, protože pak by $d(s,u) = D_j > D_i > d_i(u) \geq d(u)$, a to je spor s (14). Máme tedy $D_i = m_i$. Podle definice N_i je tedy $V_i \subseteq N_i$ (protože pro $u \in V_i$ je $d_i(u) = D_i = m_i$). Ale žádný jiný u z A_{i-1} do N_i nepatří. Pak by totiž musel $u \in V_j$ pro $j > i$, tedy by $d(s,u) = D_j > D_i = d_i(u) \geq d(u)$, což je spor s (14). Tedy je $V_i = N_i$. Důkaz správnosti Dijkstrova algoritmu je hotov. □

Poznámka 4.15. Dijkstrův algoritmus zjistí vzdálenosti vrcholů u od daného vrcholu s, ale nesdělí nám, kudy nejkratší cesta z s do u vede. Tuto informaci zjistíme následující úpravou uvedeného algoritmu. Ke každému vrcholu u udržujeme množinu vrcholů $\mathrm{pred}(u)$, které se na některé z nejkratších cest z s do u nachází těsně před u (nejkratších cest z s do u může obecně existovat více). Na začátku (např. po provedení kroku 1) nastavíme pro každý $u \in V$ hodnotu $\mathrm{pred}(u) := \emptyset$. Krok 4 rozšíříme následovně. Určí-li se v kroku 4 nová hodnota hodnota $d(u)$, provedeme

$$\mathrm{pred}(u) := \{v \in N \mid d(u) = d(v) + w(\{v,u\})\}.$$

Po skončení algoritmu pak pro každý uzel u, do kterého vede cesta z s (tj. $d(u) \neq \infty$), obsahuje výše popsané vrcholy.

Použijeme-li tento postup ve výše uvedeném příkladu pro $s = h$, pak například

$$\text{pred}(c) = \{f\},\ \text{pred}(f) = \{i\},\ \text{pred}(i) = \{h\},\ \text{pred}(h) = \emptyset,$$

tedy (v tomto případě jediná) nejkratší cesta z h do c je

$$h, \{h, i\}, i, \{i, f\}, f, \{f, c\}, c.$$

Poznámka 4.16. Lze ukázat, že pro $n = |V|$ a $m = |E|$ pro časovou složitost $T(n, m)$ Dijkstrova algoritmu v nejhorším případě platí

$$T(n, m) = O(n \log n + m).$$

4.4 Stupně vrcholů

Jednou ze základních a snadno zjistitelných informací o grafu je, kolik hran vchází a vychází do jednotlivých vrcholů. Je to informace, kterou dobře vnímáme i pohledem na obrázek grafu. V této kapitole ukážeme několik základních úvah založených na počtech hran jednotlivých vrcholů.

Definice 4.17 (stupeň vrcholu)**.** *Stupeň vrcholu* $v \in V$ grafu $\langle V, E\rangle$ je počet hran, pro které je v koncovým vrcholem, a značí se $\deg(v)$.

U orientovaných grafů se někdy zavádí vstupní a výstupní stupeň vrcholu jako počet hran, které do přicházejí, a počet hran, které z něj vycházejí. Stupeň vrcholu je pak součet vstupního a výstupního stupně. Pro graf na obr. 7 vlevo je $\deg(u) = 3$, $\deg(v) = 2$, $\deg(w) = 2$, $\deg(x) = 1$, $\deg(y) = 0$. Pro graf vpravo je $\deg(u) = 3$, $\deg(v) = 3$, $\deg(w) = 3$, $\deg(x) = 1$, $\deg(y) = 0$. V této kapitole budeme předpokládat, že grafy, kterými se zabýváme, jsou neorientované.

Věta 4.18. *V grafu* $G = \langle V, E\rangle$ *je* $\sum_{v\in V} \deg(v) = 2|E|$.

Důkaz. Máme dokázat, že součet stupňů všech vrcholů grafu je roven dvojnásobku počtu hran. Tvrzení je téměř zřejmé, uvědomíme-li si následující. Každá hrana $e \in E$ má dva vrcholy, u a v. Hrana e přispívá jedničkou do $\deg(u)$ (je jednou z hran, jejichž počet je roven $\deg(u)$), jedničkou do $\deg(v)$ a do stupně žádného jiného vrcholu nepřispívá. Hrana e tedy přispívá právě počtem 2 do $\sum_{v\in V} \deg(v)$. To platí pro každou hranu. Proto $\sum_{v\in V} \deg(v) = 2|E|$. □

Důsledek 4.19. *Počet vrcholů lichého stupně je v libovolném grafu sudý.*

Důkaz. Označme S a L množiny vrcholů, které mají sudý a lichý stupeň. Protože každý vrchol patří buď do S, nebo do L, je $\sum_{v\in V} \deg(v) = \sum_{v\in S} \deg(v) + \sum_{v\in L} \deg(v)$. Je jasné, že $\sum_{v\in S} \deg(v)$ je sudé číslo. Podle Věty 4.18 je $\sum_{v\in V} \deg(v) = 2|E|$, tedy $\sum_{v\in V} \deg(v)$ je sudé číslo. Proto i $\sum_{v\in L} \deg(v)$ musí být sudé číslo. Kdyby byl počet vrcholů s lichým stupněm lichý, byl by $\sum_{v\in L} \deg(v)$ součet lichého počtu lichých čísel, a tedy by $\sum_{v\in L} \deg(v)$ bylo liché číslo, což není možné. Počet vrcholů s lichým stupněm je tedy sudý. □

Uvedená tvrzení představují základní podmínky, které stupně každého grafu splňují. Představme si, že o grafu s vrcholy $v_1, \ldots, v_n$ nevíme nic víc než stupně jeho vrcholů, tj. známe posloupnost $\deg(v_1), \ldots, \deg(v_n)$ stupňů jeho vrcholů. Tato posloupnost se nazývá *skóre grafu* (někdy grafová posloupnost). Přitom dvě skóre považujeme za stejná, liší-li se jen permutací (seřazením) členů. Určuje skóre graf jednoznačným způsobem (až na izomorfismus)? Vezměme např. posloupnost $1, 1, 1, 1, 1, 1$. Jednoduchou úvahou dojdeme k tomu, že každé dva grafy, jejichž skóre je $1, 1, 1, 1, 1, 1$

jsou izomorfní. Jsou to grafy izomorfní s grafem $\langle V, E\rangle$, kde $V = \{a, b, c, d, e, f\}$ a $E = \{\{a, b\}, \{c, d\}, \{e, f\}\}$. Skóre $2, 2, 2, 2, 2, 2$ však graf jednoznačným způsobem neurčuje. Na množině vrcholů $V = \{a, b, c, d, e, f\}$ totiž můžeme mít dva grafy, které nejsou izomorfní, a přesto je $2, 2, 2, 2, 2, 2$ skóre každého z nich. První je dán množinou hran $\{\{a, b\}, \{b, c\}, \{a, c\}, \{d, e\}, \{e, f\}, \{d, f\}\}$ (dva trojúhelníky), druhý množinou $\{\{a, b\}, \{b, c\}, \{c, d\}, \{d, e\}, \{e, f\}, \{a, f\}\}$ (šestiúhelník). Grafy si nakreslete a zdůvodněte si to.

Jak víme, ne každá posloupnost čísel je skóre nějakého grafu. Např. $3, 4, 2, 0$ není skóre grafu (zkuste takový graf nakreslit a uvidíte, že to nejde). Můžeme to ale zjistit i bez kreslení. Stačí použít důsledek 4.19: Graf, jehož skóre je $3, 4, 2, 0$ by měl právě jeden vrchol lichého stupně, což není možné. Posloupnost $6, 2, 2, 0$ ale podmínce z Důsledku 4.19 vyhovuje, ale graf se skóre $6, 2, 2, 0$ také neexistuje (zkuste nakreslit). Vidíme, že podmínka z důsledku 4.19 je sice nutná, ale není postačující. Otázkou je, jestli existuje jednoduchá podmínka, kterou posloupnost nezáporných celých čísel splňuje, právě když je to skóre nějakého grafu. Ukážeme, že ano a že to, zda platí, dokonce lze ověřit jednoduchým algoritmem.

Věta 4.20. *Nechť $d_1 \geq d_2 \geq d_3 \geq \cdots \geq d_n$ jsou nezáporná celá čísla a $1 \leq d_1 \leq n-1$. Pak*

$$d_1, d_2, d_3, \ldots, d_n$$

je skóre nějakého grafu, právě když

$$d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$$

je skóre nějakého grafu.

Důkaz. Uvědomme si nejdříve následující věc. Posloupnost $d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$ má $n-1$ prvků. Přitom její část $d_2-1, d_3-1, \ldots, d_{d_1+1}-1$ má d_1 prvků a dostaneme ji odečtením jedničky z prvních d_1 prvků posloupnosti $d_2, d_3, \ldots, d_n$. Část $d_{d_1+2}, \ldots, d_n$ má $n-d_1-1$ prvků a je to posledních $n-d_1-1$ prvků prvků posloupnosti $d_2, d_3, \ldots, d_n$.

Ukážeme nejdříve, že když $d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$ je skóre, pak i $d_1, d_2, d_3, \ldots, d_n$ je skóre. Když je $d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$ skóre grafu $G = \langle V, E\rangle$, má G $n-1$ vrcholů (označme je $v_2, \ldots, v_n$), které mají po řadě stupně $d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, \ldots, d_{d_1+2}, \ldots, d_n$. Vytvořme z G nový graf $G' = \langle V', E'\rangle$ přidáním vrcholu v_1, tj.

$$V' = \{v_1, v_2, \ldots, v_n\},$$

a ke každému z vrcholů $v_2, \ldots, v_{d_1+1}$ přidejme hranu k vrcholu v_1, tj.

$$E' = E \cup \{\{v_1, v_j\} \mid j = 2, \ldots, d_1 + 1\}.$$

V grafu G' vede z vrcholu v_1 právě d_1 hran (tolik jsme jich přidali) a stupeň každého z vrcholů $v_2, \ldots, v_{d_1+1}$ se o 1 zvýšil. Stupně ostatních vrcholů se nezměnily (hrany jsme k nim nepřidávali). Skóre grafu G' je tedy $\deg(v_1), (d_2-1)+1, (d_3-1)+1, \ldots, (d_{d_1+1}-1)+1, d_{d_1+2}, \ldots, d_n$, což je právě $d_1, d_2, d_3, \ldots, d_n$. Dokázali jsme, že $d_1, d_2, d_3, \ldots, d_n$ je skóre grafu.

Ukažme teď naopak, že když $d_1, d_2, d_3, \ldots, d_n$ je skóre, je i $d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$ skóre. Je-li $d_1, d_2, d_3, \ldots, d_n$, uvažujme příslušný graf G s vrcholy $v_1, \ldots, v_n$ tak, že $\deg(v_i) = d_i$. Rozlišíme dva případy.

První případ: Vrchol v_1 stupně d_1 je hranou spojen s každým z vrcholů $v_2, \ldots, v_{d_1+1}$. Pak graf, který vznikne z G odstraněním vrcholu v_1 a hran, které z něj vycházejí (to jsou právě hrany $\{v_1, v_2\}, \ldots, \{v_1, v_{d_1+1}\}$), má právě skóre $d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} -$

$1, d_{d_1+2}, \ldots, d_n$, tedy posloupnost $d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$ je skóre grafu.

Druhý případ: Vrchol v_1 stupně d_1 spojený hranami s vrcholy $v_2, \ldots, v_{d_1+1}$ neexistuje. Pak tedy existuje vrchol v_i $(2 \leq i \leq d_1 + 1)$, který není spojen s v_1, a vrchol v_j $(d_1 + 1 < j \leq n)$, který je spojen s v_1.

Tvrdíme, že existuje $v_k \neq v_j$, který je spojen s v_i, ale není spojen s v_j: v_i je totiž spojen s d_i vrcholy, mezi kterými není v_1; v_j je spojen s $d_j \leq d_i$ vrcholy a jedním z nich je v_1; lze tedy zvolit vrchol v_k, který je spojen s v_i, ale ne s v_j; pokud $v_k \neq v_j$, jsme hotovi; pokud $v_k = v_j$, pak v_i je jedním z d_j vrcholů spojených s v_j; vynecháme-li tedy ze sousedů vrcholu v_i vrchol v_j a ze sousedů vrcholu v_j vrcholy v_1 a v_i, zbyde $d_i - 1$ sousedů vrcholu v_i, mezi kterými není v_1, ani v_i, a $d_j - 2$ sousedů vrcholu v_j; protože $d_i - 1 > d_j - 2$, existuje v_k spojený s v_i, který není spojen s v_j.

Vytvořme graf G', který vznikne z G odstraněním hran $\{v_1, v_j\}$ a $\{v_i, v_k\}$ a přidáním hran $\{v_1, v_i\}$ a $\{v_j, v_k\}$. Skóre grafu G' je opět $d_1, d_2, d_3, \ldots, d_n$. Záměnou jsme dosáhli toho, že z posloupnosti $v_2, \ldots, v_{d_1+1}$ je hranou spojeno s vrcholem v_1 více vrcholů v novém grafu G' než v původním G. Na graf G' je teď buď možné použít první případ, nebo ho lze postupným opakováním právě provedené úpravy převést na graf, na který první případ už použít lze. □

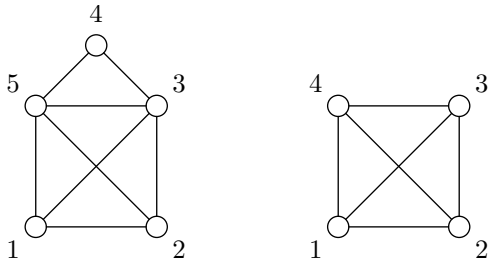
Věta 4.20 je základem pro následující algoritmus.

Algoritmus 4.21 (test skóre)**.**
Vstup: $n \in \mathbb{N}$, $\langle d_1, \ldots, d_n \rangle$, kde $d_1 \geq \cdots \geq d_n \geq 0$ jsou celá čísla;
Výstup: ANO, pokud $\langle d_1, \ldots, d_n \rangle$ je skóre, NE v opačném případě;

1. je-li $n = 1$ a $d_1 = 0$, odpověz ANO a skonči;
2. je-li $d_1 > n - 1$, odpověz NE a skonči;
3. vypočítej novou posloupnost $\langle d'_1, \ldots, d'_{n-1} \rangle = \langle d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n \rangle$;
4. je-li některé d'_i záporné, odpověz NE a skonči;
5. uspořádej $d'_1, \ldots, d'_{n-1}$ sestupně, přiřaď takto uspořádané hodnoty do $d_1, \ldots, d_{n-1}$, sniž n o 1 (tj. $n := n - 1$) a pokračuj bodem 1.

Ukažme, že algoritmus pracuje správně. Skončí-li algoritmus v bodě 1, je to správně, protože jednoprvková posloupnost 0 je skóre grafu (o jednom vrcholu a žádné hraně). Skončí-li algoritmus v bodě 2, je to správně, neboť stupeň žádného vrcholu nemůže být větší než počet vrcholů minus jedna. Pokud má aktuální posloupnost více než 1 člen, v bodě 3 se vypočítá nová posloupnost, která je dle věty 4.20 skóre grafu, právě když je původní posloupnost skóre grafu. Na základě věty 4.20 se tedy v tomto kroku testování redukuje na testování posloupnosti, která je o 1 člen kratší. Pokud se při výpočtu nové posloupnosti objeví záporné číslo, algoritmus skončí s odpovědí NE (stupeň vrcholu nemůže být záporný). Jinak algoritmus novou posloupnost sestupně setřídí a pokračuje znovu bodem 1. Je zřejmé, že algoritmus vždy skončí, neboť posloupnosti se vždy o 1 zkracují. Pro vstupní posloupnost délky n algoritmus skončí nejvýše po $n-1$ výpočtech nové posloupnosti. Algoritmus bychom mohli urychlit o krok, ve kterém by se testovaná posloupnost zkrátila o koncové nuly (zdůvodněte to).

Příklad 4.22. Algoritmus 4.21 použijeme ke zjištění, jestli jsou posloupnosti $6, 6, 5, 4, 3, 3, 3, 3, 2, 1, 0$ a $4, 3, 2, 1, 1$. Test posloupnosti $6, 6, 5, 4, 3, 3, 3, 3, 2, 1, 0$ vede



Obrázek 11: Nakreslete obrázky jedním tahem.

postupně na testy posloupností $5,4,3,3,2,2,2,2,1,0$, pak $3,2,2,2,2,1,1,1,0$, pak $2,1,1,1,1,1,1,0$, pak $1,1,1,1,0,0,0$, pak $1,1,0,0,0,0$, pak $0,0,0,0,0$, $0,0,0,0$, pak $0,0,0$, pak $0,0$, pak 0 a skončí odpovědí ANO v bodě 1. Test $4,3,2,1,1$ vede postupně na $2,1,0,0$, pak se v bodě 4. vypočítá posloupnost $0,-1,0$ a algoritmus skončí s odpovědí NE. To, že $4,3,2,1,1$ není skóre, můžeme poznat také přímo podle důsledku 4.19, protože $4,3,2,1,1$ má lichý počet lichých stupňů.

Se stupni vrcholů souvisí známá úloha nakreslit jedním tahem zadaný obrázek.

Průvodce studiem

Úloha o kreslení jedním tahem: Na obr. 11 jsou dva obrázky. Úkolem je nakreslit obrázek jedním tahem s tím, že žádnou hranu nesmíme nakreslit dvakrát a nesmíme zvednout tužku z papíru. U levého obrázku to jde (např. postupným spojováním vrcholů $1,2,3,5,1,3,4,5,2$), u pravého ne (zkuste to zdůvodnit).

S touto úlohou se asi každý setkal na základní škole. Ukážeme si, že rozhodnout, zda obrázek lze nakreslit jedním tahem, popř. zda je i možné přitom začít i skončit v jednom místě, jde jednoduše podle stupňu vrcholů. Na kreslení jednotažek navíc existuje algoritmus.

Definice 4.23 (eulerovský tah)**.** *Eulerovský tah*[11] je tah, který obsahuje všechny vrcholy grafu a ve kterém se každá hrana vyskytuje právě jednou. Je-li navíc uzavřený, nazývá se *uzavřený eulerovský tah*.

Eulerovský tah představuje kreslení „jedním tahem“. Chceme-li navíc při kreslení vyjít i skončit v jednom místě, musíme najít uzavřený eulerovský tah. Následující věta ukazuje, jak jednoduše poznat, zda eulerovský tah vůbec existuje.

Jestli graf má (uzavřený) eulerovský tah, lze jednoduše poznat ze stupňů jeho vrcholů.

Věta 4.24.

- *V neorientovaném grafu existuje uzavřený eulerovský tah, právě když je souvislý a každý vrchol má sudý stupeň.*
- *V neorientovaném grafu existuje neuzavřený eulerovský tah, právě když je souvislý a má právě dva vrcholy lichého stupně.*

Důkaz. Dokážeme nejdříve tvrzení pro uzavřené eulerovské tahy. Mějme graf $G = \langle V, E\rangle$.

Předpokládejme nejdřív, že v G existuje uzavřený eulerovský tah $v, e, \ldots, v$. Je jasné, že G je souvislý. Uvažujme libovolný vrchol $u \in V$ a množinu E_u všech hran, jichž je u koncovým vrcholem. Jejich počet je stupeň u, tj. $\deg(u) = |E_u|$. Pro $u \neq v$ je libovolný výskyt u v tahu $v, e, \ldots, v$ tvaru $\ldots, e, u, e', \ldots$, kde $e, e' \in E_u$, tj. každý výskyt u je doprovázen výskytem dvou hran z E_u (jednou z nich se do u vstoupí,

[11] Leonhard Euler (1707–1783), jeden z nejvýznamnějších matematiků.

druhou se vystoupí). Protože se v tahu každá hrana vyskytuje právě jednou, je jasné, že hran z E_u je sudý počet. Pro $u = v$ to platí s výjimkou prvního (tam hrana z v pouze vychází) a posledního výskytu (tam hrana do v pouze vchází). Každý z nich je doprovázen výskytem jedné hrany z E_u a proto je počet hran v E_u opět sudé číslo.

Předpokládejme teď, že G je souvislý a že každý jeho vrchol má sudý stupeň. Uvažujme tah $v_0, e_1, \ldots, e_n, v_n$ (označme ho t) v G, který má největší možnou délku (to můžeme: tah nemůže být delší než počet všech hran grafu G, tahů s největší délkou ale může být více). Všimněme si nejdřív, že musí být $v_0 = v_n$. Jinak by vrchol v_0 byl koncovým vrcholem lichého počtu hran (hrana z něj vychází a pak vždy vchází a vychází). Protože má v_0 sudý stupeň, existuje hrana $e = \{v, v_0\}$, která není obsažena v tahu t. Pak je ale $v, e, v_0, e_1, \ldots, e_n, v_n$ tah, který je delší než t, a to je spor s tím, že t má největší možnou délku. Tedy musí být $v_0 = v_n$. Dokážeme nyní sporem, že tah t obsahuje všechny hrany grafu G. Předpokládejme, že existuje $e = \{u, v\} \in E - \{e_1, \ldots e_n\}$. Protože G je souvislý, je v spojen nějakou cestou s každým vrcholem vyskytujícím se v tahu t. Nechť v_i je vrchol, pro který je taková cesta $v, \ldots, v_i$ nejkratší. Pak se žádná hrana cesty $v, \ldots, v_i$ neleží v tahu t (jinak by tato cesta nebyla nejkratší). Proto je

$$u, e, v, \ldots, v_i, e_{i+1}, v_{i+1}, \ldots, v_n = v_0, e_0, v_1, \ldots, v_i$$

tah, který je delší než t, což je spor s předpokladem. Tah t tedy obsahuje všechny hrany z E a je eulerovským tahem.

Část tvrzení, která se týká neuzavřených eulerovských tahů, se dokáže podobně (viz úlohy k textu). □

Průvodce studiem

Úloze rozhodnout, zda v grafu existuje uzavřený eulerovský tah, je podobná úloha tzv. hamiltonovské kružnice.[12] Hamiltonovská kružnice je kružnice, která obsahuje všechny vrcholy grafu. Připomeňme, že uzavřený eulerovský tah obsahuje všechny hrany grafu. Zatímco zjistit, zda v grafu existuje uzavřený eulerovský tah, je velmi snadné (podle věty 4.24 stačí ověřit, že každý vrchol má sudý stupeň), není znám rychlý algoritmus, který by zjistil, zda graf má hamiltonovskou kružnici. Navíc je pravděpodobné, že takový algoritmus ani neexistuje (zjistit existenci hamiltonovské kružnice je totiž tzv. NP-úplný problém).

Shrnutí

Graf je tvořen množinou vrcholů a hran spojujících některé vrcholy. Graf může být orientovaný nebo neorientovaný, podle toho, jestli rozlišujeme, zda orientace hran hraje roli. Posloupnost vrcholů a hran, která odpovídá možnému průchodu grafem, se nazývá sled. Rolišujeme několik typů sledů. Mezi důležité úlohy patří různé úlohy o cestování v grafech.

Pojmy k zapamatování

- orientovaný graf, neorientovaný graf, vrchol, hrana,
- izomorfismus grafů, podgraf,
- sled, délka sledu, uzavřený sled, tah, cesta, kružnice, vzdálenost vrcholů,
- ohodnocený graf,
- souvislost, komponenta, hledání cest,
- stupeň vrcholu, skóre, eulerovský tah.

Kontrolní otázky

1. *Vysvětlete rozdíl mezi pojmy orientovaný graf a neorientovaný graf.*
2. *Je-li graf G izomorfní s grafem G', je jeho podgrafem?*
3. *Jaký je rozdíl mezi pojmy sled, tah, cesta?*
4. *Může mít souvislý graf po odstranění jedné hrany tři komponenty?*
5. *Existuje graf, který má skóre* $7, 3, 1$*?*

Cvičení

1. Pro libovolné neorientované grafy G_1, G_2 a G_3 platí: $G_1 \cong G_1$; pokud $G_1 \cong G_2$, pak $G_2 \cong G_1$; pokud $G_1 \cong G_2$ a $G_2 \cong G_3$, pak $G_1 \cong G_3$. To samé platí pro orientované grafy. Dokažte.
2. Je-li graf G izomorfní s grafem G', je každý podgraf grafu G izomorfní nějakému podgrafu grafu G'. Dokažte.
3. Z definice plyne, že jsou-li konečné grafy $G_1 = \langle V_1, E_1 \rangle$ a $G_2 = \langle V_2, E_2 \rangle$ izomorfní, mají stejný počet vrcholů (izomorfismus je totiž bijekce $h : V_1 \to V_2$). Dokažte, že izomorfní grafy mají i stejný počet hran.
4. Nechť jsou dány grafy G s vrcholy $v_1, \ldots, v_n$ a G' s vrcholy $v'_1, \ldots, v'_n$ takové, že $\deg(v_i) = \deg(v'_i)$. Musí být G a G' izomorfní?
5. Jaký je největší možný součet stupňů vrcholů neorientovaného grafu s n vrcholy?
6. Určete (např. pomocí Algoritmu 4.21), zda jsou skóre posloupnosti
   (a) $6, 6, 6, 6, 5, 4, 3$,
   (b) $5, 5, 5, 5, 5, 4, 3$,
   (c) $4, 4, 4, 4, 3, 3, 2$.

   Nakreslete příslušné grafy.

Úkoly k textu

1. Dokončete důkaz věty 4.11, tj. ukažte, že každá hrana grafu je hranou právě jedné jeho komponenty.
2. Dokončete důkaz věty 4.24, tj. dokažte, že v neorientovaném grafu existuje neuzavřený eulerovský tah, právě když je souvislý a má právě dva vrcholy lichého stupně. Návod: Postupujte podobně jako pro uzavřené eulerovské tahy. Sečtením hran v eulerovském tahu dojdete k tomu, že když neuzavřený eulerovský tah existuje, mají právě dva vrcholy lichý stupeň. Naopak, když je graf souvislý a právě dva vrcholy, u a v, mají lichý stupeň, uvažujte opět nejdelší tah. O něm nejdříve dokažte, že jeho krajní vrcholy jsou u a v. Pak postupujte podobně jako u důkazu pro uzavřený eulerovský tah, tj. ukažte, že obsahuje všechny vrcholy i všechny hrany grafu.
3. Navrhněte algoritmus pro hledání uzavřeného eulerovského tahu a algoritmus pro hledání eulerovského tahu. Proveďte důkazy správnosti těchto algoritmů.
4. Upravte algoritmus 4.14 tak, aby fungoval i pro orientované grafy. Tj. vstupem bude ohodnocený orientovaný graf a jeho vrchol s. Výstupem budou čísla $d(v)$, kde $d(v)$ je délka nejkratší cesty z s do v. Proveďte důkaz správnosti.

Řešení

1. Snadno se totiž ukáže, že id_{V_1} je izomorfismus G_1 a G_2; je-li h izomorfismus G_1 a G_2, je h^{-1} izomorfismus G_2 a G_1; je-li g izomorfismus G_1 a G_2 a h izomorfismus G_2 a G_3, je $g \circ h$ izomorfismus G_1 a G_3.
2. Snadné.
3. Protože izomorfismus je vzájemně jednoznačné zobrazení vrcholů grafu G_2 vrcholům grafu G_1, platí, že pro libovolné různé vrcholy u a v grafu G_1 jsou $h(u)$ a $h(v)$ různé vrcholy grafu G_2, a dále, že každé dva různé vrcholy grafu G_2 jsou tvaru $h(u)$ a $h(v)$ pro nějaké vrcholy u a v grafu G_1.
   Definice izomorfismu tedy s přihlédnutím k právě uvedenému říká, že mezi dvěma vrcholy v jednom z těchto grafů vede hrana, právě když hrana vede i mezi dvěma vrcholy, které jim odpovídají v druhém grafu. Z toho je patrné, že oba grafy mají stejný počet hran.
   Dokázat to lze také ověřením toho, že zobrazení $\overline{h} : E_1 \to E_2$ definované pro $e = \{u, v\}$ předpisem
   $$\overline{h}(\{u, v\}) = \{h(u), h(v)\}$$
   je bijekce.
4. Ne. Uvažujme $n = 6$ a $E = \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4\}, \{v_4, v_5\}, \{v_5, v_6\}, \{v_6, v_1\}\}$ a $E = \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_1\}, \{v_4, v_5\}, \{v_5, v_6\}, \{v_6, v_4\}\}$.
5. $n \cdot (n - 1)$.
6. 1. ne, 2. ano, 3. ano.

Studijní cíle: Po prostudování kapitoly 4.5 by student měl rozumět speciálním grafům nazývaným stromy. Měl by znát základní vlastnosti stromů a algoritmy pro jejich procházení.

Klíčová slova: strom, list, kostra, kořenový strom, houbka vrcholu, výška stromu, m-ární strom, uspořádaný kořenový strom, preorder, postorder, inorder

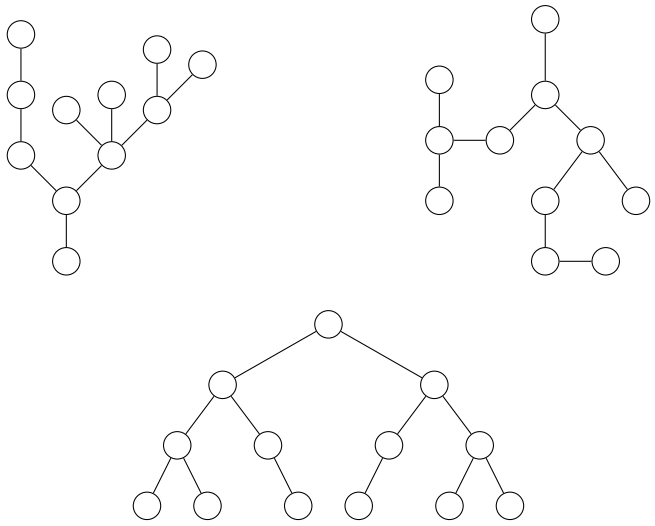
4.5 Stromy

Průvodce studiem

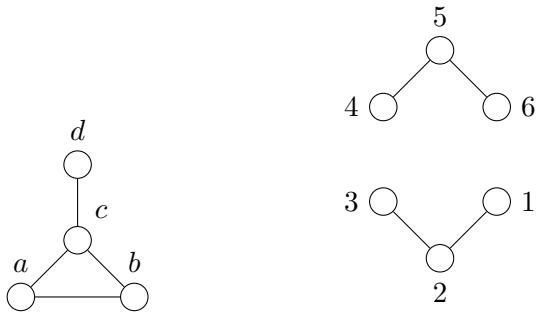
Stromy jsou speciální grafy, které dostaly název podle toho, že vypadají podobně jako stromy, popř. keře v přírodě. Typický strom-graf vypadá jako strom v přírodě. Má svůj kořen (speciální vrchol), ve kterém se větví (vedou z něj hrany) do míst (vrcholů), ve kterých se opět větví atd. Stromy jsou grafy, které mají nejčastější použití. Setkáváme se s nimi v běžném životě (různá členění, např. členění knihy na kapitoly, podkapitoly atd., mají stromovou strukturu), jako uživatelé počítačů (stromová struktura adresářů) i jako informatici (rozhodovací stromy, vyhledávací stromy).

4.5.1 Definice a základní vlastnosti

Stromy jsou speciální grafy, které dostaly název podle toho, že vypadají podobně jako stromy, popř. keře v přírodě. Zvolíme-li jeden vrchol stromu, můžeme ho považovat za kořen, ze kterého strom vyrůstá, větví se a roste do míst, ve kterých se opět větví. Tři stromy vidíme na obr. 12. Se stromy se setkáváme v běžném životě. Podobu stromů mají například různé hierarchické struktury. Kniha se člení na kapitoly a podkapitoly. Adresáře (složky) v počítači se člení na podadresáře, ty opět na podadresáře apod. Stromy mají rozsáhlé použití v informatice, zejména ve zpracování dat.



Obrázek 12: Stromy.



Obrázek 13: Grafy, které nejsou stromy.

Stromy budeme chápat jako neorientované grafy. Pojem strom lze zavést několika způsoby. My vyjdeme z následující definice.

Definice 4.25. *Strom* je neorientovaný souvislý graf bez kružnic.

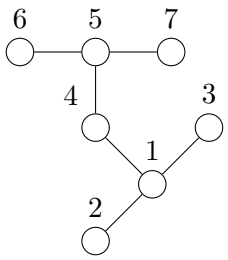
Že je graf bez kružnic neboli že neobsahuje kružnice, znamená, že žádný jeho podgraf není kružnicí. Všimněme si, že každý z grafů na obr. 12 je souvislý, neobsahuje kružnice, a je tedy stromem. Ani jeden z grafů na obr. 13 ale stromem není. Graf vlevo totiž obsahuje kružnici $a, \{a, b\}, b, \{b, c\}, c, \{a, c\}, a$. Graf vpravo není souvislý, protože v něm například neexistuje sled z vrcholu 1 do vrcholu 6.

Vrchol grafu se stupněm 1 se nazývá *koncový* (za koncový vrchol se někdy považuje i vrchol se stupněm 0). Koncový vrchol stromu se nazývá *list*.

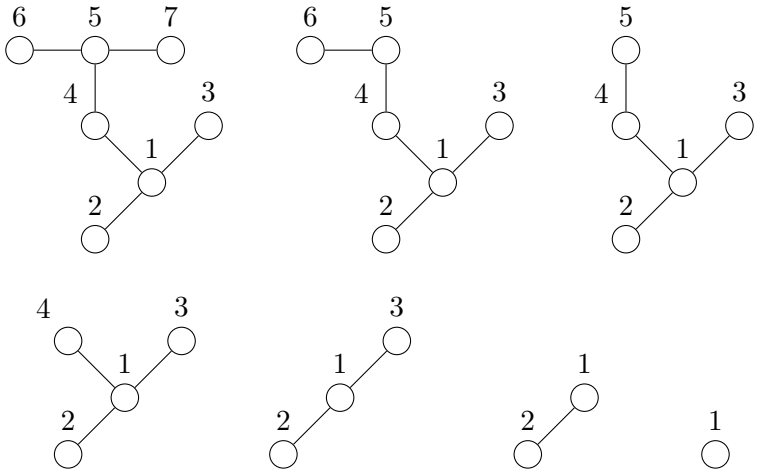
Uvažujme strom na obr. 14. Listy tohoto stromu jsou vrcholy 2, 3, 6 a 7. Jak se můžeme přesvědčit, odebereme-li z tohoto stromu nějaký list a hranu, která do něj vede, vznikne opět strom. Z něho můžeme opět odebrat nějaký list s hranou atd., až z tohoto stromu zbyde jediný uzel. Popsaný proces je vidět na obr. 15, ve kterém byly postupně odebrány vrcholy 7, 6, 5, 4, 3 a 2.

Můžeme uvažovat i obrácený proces, tedy vyjít z jednoho vrcholu a postupně k němu připojovat vrcholy, které se připojením stanou listy. Snadno se přesvědčíme, že ať zvolíme v uvažovaném stromu kterýkoli z jeho vrcholů 1–7, můžeme z něj takovým postupným přidáváním listů vytvořit celý strom.

Zmiňovanou vlastnost mají všechny stromy, tedy nejen uvažovaný strom z obr. 14. Tato vlastnost tedy nabízí jiný, konstruktivní pohled, podle kterého jsou stromy právě



Obrázek 14: Strom.



Obrázek 15: Odebírání (přidávání) vrcholů stromu.

grafy, které lze z jediného uzlu sestrojit připojováním nových vrcholů tak, že připojený vrchol se stane koncovým vrcholem zvětšeného grafu. Protože je tato vlastnost důležitá, zformulujeme ji a dokážeme obecně. Vyjdeme z následujícího pomocného tvrzení.

Věta 4.26. *V každém grafu bez kružnic (a tedy i v každém stromu) s aspoň dvěma vrcholy existují aspoň dva listy.*

Důkaz. Uvažujme cestu $v_0, e_1, \ldots, e_n, v_n$, která má ze všech cest největší délku (uvažujeme graf s konečnou množinou vrcholů, v něm má každá cesta délku menší, než je počet vrcholů, a tedy existuje cesta s největší délkou). Tvrdíme, že v_0 i v_n jsou listy. Dokážeme to sporem. Kdyby např. v_0 nebyl list, pak by existovala hrana $e = \{v, v_0\}$, která je různá od e_1 (jednou hranou vycházející z v_0 je e_1; není-li v_0 list, musí z něj vycházet ještě jiná hrana). Vrchol v musí být různý od každého v_i ($i = 1, \ldots, n$). Kdyby totiž $v = v_i$ pro nějaké $i = 1, \ldots, n$, pak by $v, e, v_0, \ldots, v_i$ byla kružnice, což není možné, protože uvažovaný graf kružnice neobsahuje. Pak je ale posloupnost $v, e, v_0, e_1, \ldots, v_n$ cesta, která je delší než $v_0, e_1, \ldots, v_n$, což je spor s předpokladem, že $v_0, e_1, \ldots, v_n$ je ze všech cest nejdelší. Podobně se ukáže, že v_n je list. □

Základem výše uvedené vlastnosti o konstrukci stromů připojováním vrcholů je následující tvrzení. Poznamenejme, že $G - v$ označuje graf, který vznikne z grafu G odstraněním vrcholu v a hran, které z něj vycházejí.

Věta 4.27. *Pro graf G a jeho koncový vrchol v jsou následující tvrzení ekvivalentní.*

1. *G je strom.*
2. *$G - v$ je strom.*

Důkaz. Předpokládejme, že G je strom. Máme ukázat, že $G - v$ je strom, tedy že je souvislý a neobsahuje kružnici. Souvislost: Vezměme libovolné dva vrcholy $u \neq w$ grafu $G - v$. Protože G je souvislý, existuje v G cesta $u, e, \ldots, w$. Protože v je koncový vrchol grafu G, musí být každý vrchol této cesty různý od v. Totiž, u i w jsou různé od v, protože to jsou vrcholy grafu $G - v$; žádný jiný vrchol cesty $u, e, \ldots, w$ nemůže být vrcholem v, protože v je koncový vrchol grafu G, a má tedy stupeň 1 (každý vnitřní vrchol cesty má stupeň aspoň 2). Cesta $u, e, \ldots, w$ je tedy cestou v grafu $G - v$. Proto je G souvislý. Kružnice: Kdyby $G - v$ obsahoval kružnici, byla by to zřejmě i kružnice v grafu G, což není možné, protože G je strom.

Předpokládejme naopak, že $G - v$ je strom. Ukážeme, že G je strom. Souvislost: Vezměme vrcholy $u \neq w$ v G. Když $u \neq v \neq w$, pak protože $G - v$ je strom, existuje v $G - v$ cesta z u do w. Ta je zřejmě i cestou v G. Když je $u = v$ (pro $w = v$ je to podobné), uvažujme vrchol x, ke kterému byl v v grafu G připojen (tj. $G - v$ vznikl odebráním vrcholu v a hrany $\{v, x\}$). Ze souvislosti $G - v$ plyne, že v něm existuje cesta $x, \ldots, w$. Je zřejmé, že pak je $u = v, \{v, x\}, x \ldots, w$ cesta v G, která spojuje u a w. G je tedy souvislý. Kružnice: Pokud by $v_0, \ldots, v_n$ byla kružnice v G, pak by neobsahovala v, protože v je koncový vrchol (vrcholy kružnice mají stupeň aspoň 2). Tato kružnice by tedy byla i kružnicí v $G - v$, což není možné, protože předpokládáme, že $G - v$ je strom. □

Uvědomme si, že věty 4.26 a 4.27 skutečně zdůvodňují správnost výše popsané konstrukce stromů. Je-li G strom s aspoň dvěma vrcholy, má dle věty 4.26 aspoň jeden list (tedy koncový vrchol) v. Podle věty 4.27 vznikne z G po odebrání v strom $G - v$. Ten má buď jeden vrchol, nebo má aspoň dva vrcholy a opět mu můžeme odebrat list. Tak nakonec skončíme s grafem o jednom vrcholu. Naopak, vyjdeme-li z libovolného stromu S (sestávajícího třeba z jediného vrcholu) a připojíme-li k jeho libovolnému vrcholu x hranou nový vrchol v, dostaneme graf G, pro který zřejmě platí, že S je právě grafem $G - v$. Proto je dle věty 4.27 graf G stromem.

Existují však i další užitečné charakterizace stromů. Následující věta uvádí některé z nich.

Věta 4.28. *Pro neorientovaný graf $G = \langle V, E \rangle$ jsou následující tvrzení ekvivalentní.*

1. *G je strom.*
2. *Mezi každými dvěma vrcholy grafu G existuje právě jedna cesta.*
3. *G je souvislý, ale vynecháním libovolné hrany vznikne nesouvislý graf.*
4. *G neobsahuje kružnice, ale přidáním jakékoli hrany vznikne graf s kružnicí.*
5. *G neobsahuje kružnice a $|V| = |E| + 1$.*
6. *G je souvislý a $|V| = |E| + 1$.*

Důkaz. „1. ⇒ 2.“: Předpokládejme, že G je strom. Protože G je podle definice souvislý, existuje mezi každými dvěma vrcholy cesta. Kdyby mezi nějakými vrcholy u a v existovaly dvě různé cesty, znamenalo by to, že v G je kružnice. To nyní ukážeme. Jsou-li $u, e_1, v_1, \ldots, e_n, v$ a $u, e'_1, v'_1, \ldots, e'_m, v$ dvě cesty, pak jejich spojením je uzavřený sled $s = u, e_1, v_1, \ldots, e_n, v, e'_m, \ldots, v'_1, e'_1, u$. Pokud ten ještě není kružnicí, opakuje se v něm nějaký vrchol $w \neq u$, tj. exsituje v něm úsek $w, \ldots, w$. Nahrazením tohoto úseku jen uzlem w toto opakování odstraníme. Pokud se ve zbylém uzavřeném úseku s' už žádný vrchol neopakuje, je s' hledanou kružnicí. Pokud ano, můžeme v něm opět nahradit nějakou část $w', \ldots, w'$ uzlem w'. Tak postupně dostaneme kružnici. To je ale spor s tím, že G je strom.

„2. ⇒ 3.“: Pokud v G mezi každými vrcholy existuje právě jedna cesta, je G zřejmě souvislý. Vynechme hranu $e = \{u, v\} \in E$. Kdyby byl graf G', který vynecháním vznikne, souvislý, existovala by v něm cesta $u, e_1, \ldots, v$ mezi u a v. To by ale znamenalo, že v G existují dvě cesty z u do v: jednou je $u, e_1, \ldots, v$, druhou je u, e, v. To je spor s tím, že mezi každými dvěma vrcholy je v G právě jedna cesta.

„3. ⇒ 4.“: Kdyby G obsahoval kružnici, pak odstraněním její libovolné hrany dostaneme zřejmě opět souvislý graf, což je spor s předpokladem 3. Kdyby po přidání hrany $e = \{u, v\}$ nevznikla kružnice, v G by neexistovala cesta mezi u a v (kdyby ano, přidáním e k této cestě dostaneme kružnici), a tedy G by nebyl souvislý, což je spor s předpokladem 3.

„4. ⇒ 5.“: Důkaz provedeme matematickou indukcí podle počtu $|V|$ vrcholů. Dokazujeme tedy tvrzení, že graf s n vrcholy, který splňuje podmínku 4, splňuje i podmínku 5. Pro graf s $n = 1$ vrcholem tvrzení zřejmě platí (pak je totiž $|E| = 0$, a tedy $|V| = |E|+1$). Předpokládejme, že tvrzení platí pro každý graf s n vrcholy. Nechť G má $n+1$ vrcholů a splňuje podmínku 4. Protože neobsahuje kružnice, má podle věty 4.26 list v. Odstraněním v a hrany, která do něj vede, získáme graf $\langle V', E' \rangle$, který je zřejmě také bez kružnic a ve kterém přidání libovolné hrany vytvoří kružnici. $\langle V', E' \rangle$ má ale n vrcholů, podle indukčního předpokladu proto splňuje $|V'| = |E'|+1$. Protože však $|V| = |V'|+1$ a $|E| = |E'| + 1$, platí i $|V| = |E| + 1$.

„5. ⇒ 6.“: Nechť $G_1 = \langle V_1, E_1 \rangle, \ldots, G_k = \langle V_k, E_k \rangle$ jsou všechny komponenty grafu G. Každý G_i je stromem, protože neobsahuje kružnice (je totiž podgrafem grafu G, který neobsahuje kružnice) a je souvislý (neboť je komponentou). Nyní stačí ukázat $k = 1$ (pak $G = G_1$, tedy G je souvislý, protože je sám komponentou). Z výše dokázaného plyne, že z podmínky 1 plyne 5, tedy každý strom G_i splňuje $|V_i| = |E_i| + 1$. Je zřejmé, že $|E| = |E_1| + \cdots + |E_k|$ a

$$|V| = |V_1| + \cdots + |V_k| = (|E_1| + 1) + \cdots + (|E_k| + 1) = |E_1| + \cdots + |E_k| + k = |E| + k.$$

Protože podle předpokladu je $|V| = |E| + 1$, platí $k = 1$.

„6. ⇒ 1.“: Dokážeme to indukcí podle počtu vrcholů. Má-li G jeden vrchol, je tvrzení zřejmé. Předpokládejme, že tvrzení platí pro každý graf s n vrcholy a že G má $n + 1$ vrcholů, je souvislý a splňuje $|V| = |E|+1$. Je $2|E| = 2(|V|-1) = 2|V|-2$ a dle věty 4.18 je součet stupňů vrcholů grafu G roven $2|V| - 2$. Ze souvislosti plyne, že každý vrchol má stupeň aspoň 1. Kdyby měl každý vrchol stupeň aspoň 2, byl by součet stupňů všech vrcholů aspoň $2|V|$, ale ten součet je $2|V| - 2 < 2|V|$. Tedy musí existovat vrchol v stupně právě 1, tj. list. Jeho odstraněním dostaneme graf $G' = \langle V', E' \rangle = G - v$, který je zřejmě souvislý, má n vrcholů a platí pro něj $|V'| = |V| - 1$, $|E'| = |E| - 1$. G' tedy splňuje $|V'| = |E'| + 1$ a z indukčního předpokladu plyne, že je to strom. Proto je G dle věty 4.27 strom.

□

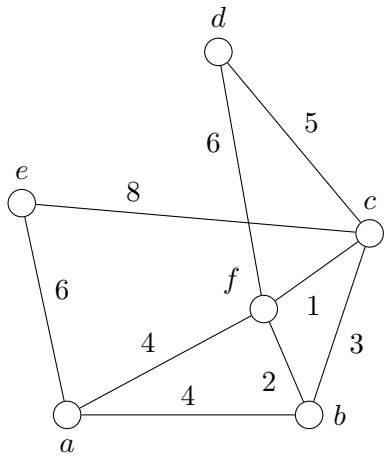
Zastavme se u některých charakterizací stromů uvedených ve větě 4.28. Vlastnosti 2., 3. a 4. jsou patrné z příkladů stromů, které jsme viděli. Vlastnosti 5. a 6. mimo jiné říkají, že počet vrcholů stromu je vždy o 1 větší než počet jeho hran. Například první graf na obr. 12 má 11 vrcholů a 10 hran, druhý graf také, třetí má 13 vrcholů a 12 hran. Je-li tedy dán graf, který má například 8 vrcholů a 6 hran, víme dle uvedených vlastností, že takový graf není stromem, aniž bychom znali další podrobnosti (například jeho obrázek). Neplatí ale, že je-li počet vrcholů grafu o 1 větší než počet hran, je tento graf stromem (ověřte).

4.5.2 Minimální kostra grafu

Představme si, že máme za úkol propojit města $v_1, \ldots, v_n$ elektrickým vedením, a to tak, aby výstavba vedení byla co nejlevnější. Musíme tedy rozhodnout, mezi kterými městy

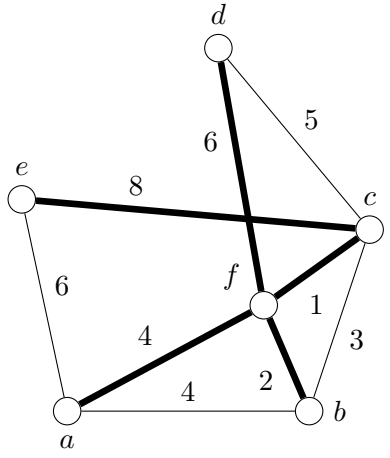
máme natáhnout elektrické dráty tak, aby se elektřina mohla dostat z každého města do každého jiného města (ne nutně přímým spojením, může téct přes ostatní města). Přitom víme, že mezi některými dvojicemi měst přímé propojení postavit nelze (mezi městy jsou hory, přehrady apod.). Pokud města v_i a v_j propojit lze, známe náklady na výstavbu vedení mezi v_i a v_j.

Příklad výše popsaného zadání ukazuje graf na obr. 16. Hrany vedou mezi těmi dvo-



Obrázek 16: Síť měst.

jicemi měst, mezi kterými lze vést přímé elektrické vedení, a jejich číselné hodnoty představují náklady na výstavbu takového vedení (v mil. Kč). Například náklady na propojení města a a města b jsou 4 mil. Kč, zatímco města e a d přímo propojit nelze. Pokud lze města u a v propojit, náklady propojení označíme w_{uv}. Je tedy $w_{ab} = 4$ apod. Možné propojení daných měst ukazuje obr. 17, ve kterém jsou tučně vyznačená spo-



Obrázek 17: Možné propojení měst elektrickým vedením.

jení, která se vybudují. Celkové náklady na vybudování tohoto propojení v miliónech Kč jsou

$$w_{af} + w_{bf} + w_{ce} + w_{cf} + w_{df} = 4 + 2 + 8 + 1 + 6 = 21.$$

Nevýhodou tohoto propojení – jak uvidíme později – je, že je zbytečně nákladné. Jak najít řešení, tedy nejlevnější propojení všech měst, si nyní ukážeme.

Popišme výše uvedený problém v řeči teorie grafů. Zadání problému, tedy města, jejich možná propojení a náklady na výstavbu propojení mezi městy, lze chápat jako

neorientovaný graf $G = \langle V, E\rangle$ s hranovým ohodnocením $w : E \to \mathbb{R}^+$. Mezi vrcholy u a v vede hrana, tedy $\{u, v\} \in E$, právě když u a v lze přímo propojit. Náklady propojení u a v jsou v takovém případě dány ohodnocením $w(\{u, v\})$ hrany $\{u, v\}$, což je kladné reálné číslo. Hledané propojení musí obsahovat všechny vrcholy grafu. Protože chceme, aby vybrané hrany (tedy vybudovaná propojení) propojovaly všechna města, chceme vlastně, aby podgraf grafu $G = \langle V, E\rangle$ daný množinou V všech vrcholů a množinou E' vybraných hran, byl souvislý. Vybraný podgraf $G' = \langle V, E'\rangle$ navíc nemá obsahovat kružnice, protože náklady by byly zbytečně vysoké. Po odstraněním libovolné hrany $\{u, v\}$ v případné kružnici totiž vybraný graf zůstane souvislým a my ušetříme $w(\{u, v\})$. Vybraný podgraf $G' = \langle V, E'\rangle$ má tedy obsahovat všechny vrcholy původního grafu, má být souvislý a nemá obsahovat kružnice. $G' = \langle V, E'\rangle$ má tedy být stromem. Takový podgraf $G' = \langle V, E'\rangle$ se nazývá kostra grafu $G = \langle V, E\rangle$. Je-li navíc ze všech koster nejlevnější, nazývá se tato kostra minimální.

Definice 4.29. *Kostra* neorientovaného grafu G je jeho podgraf G', který je stromem a obsahuje všechny vrcholy grafu G.

Je-li $w : E \to \mathbb{R}^+$ hranové ohodnocení grafu $G = \langle V, E\rangle$, nazývá se kostra $G' = \langle V, E'\rangle$ *minimální kostra*, pokud má ze všech koster grafu G nejmenší hodnotu $w(G')$, kde

$$w(G') = \sum_{\{u,v\}\in E'} w(\{u, v\}).$$

Tučně vyznačený podgraf na obr. 17 je tedy kostrou. Tato kostra ale není minimální (najděte kostru s menší hodnotou). Snadno se nahlédne, že graf má kostru, právě když je souvislý. Je také pochopitelné, že minimální kostra obecně není určena jednoznačně (v grafu může existovat více koster s nejmenší hodnotou).

Uvedeme nyní algoritmus pro hledání minimální kostry.

Algoritmus 4.30 (minimální kostra)**.**
Vstup: souvislý neorientovaný graf $G = \langle V, E\rangle$
s n vrcholy a m hranami,
ohodnocení $w : E \to \mathbb{R}^+$
Výstup: množina hran $E' \subseteq E$ taková, že $G' = \langle V, E'\rangle$ je minimální kostra grafu G

1. setřiď hrany vzestupně podle ohodnocení, tj. utvoř posloupnost $e_1, \ldots, e_m$ všech hran z E tak, že
$$w(e_1) \le \cdots \le w(e_m);$$
2. $E_0 = \emptyset$; $i := 0$;
3. dokud E_i neobsahuje $n - 1$ hran, prováděj:
$$i := i + 1;$$
$$E_i := \begin{cases} E_{i-1} \cup \{e_i\} & \text{pokud } \langle V, E_{i-1} \cup \{e_i\}\rangle \\ & \text{neobsahuje kružnici,} \\ E_{i-1} & \text{v opačném případě;} \end{cases}$$
4. $E' := E_i$.

Uvedený algoritmus tedy nejprve vzestupně setřídí hrany podle jejich hodnot. Pak hrany v tomto pořadí prochází a vytváří postupně množiny hran E_0, E_1, E_2, ..., až je

vytvořena množina E_i s právě $n-1$ hranami (níže uvidíme, že algoritmus vždy skončí s hodnotou $i \leq m$). Počáteční množina E_0 je prázdná. Pokud přidáním hrany e_i do množiny E_{i-1} dosud přidaných hran nevznikne kružnice, hranu přidáme a vytvoříme tak množinu E_i; pokud by kružnice vznikla, hranu nepřidáme a vezmeme $E_i := E_{i-1}$.

Ukažme, jak tento algoritmus hledá minimální kostru grafu z obr. 16.

Krok 1: Algoritmus setřídí hrany podle jejich hodnot. Protože v grafu existují hrany se stejnými hodnotami, není pořadí hran jednoznačné (závisí na konkrétním třídicím algoritmu). Předpokládejme, že setříděním vznikne posloupnost

$$e_1, e_2, \ldots, e_9 = \\ \{c,f\}, \{b,f\}, \{b,c\}, \{a,b\}, \{a,f\}, \{c,d\}, \{a,e\}, \{d,f\}, \{c,e\}.$$

Tato posloupnost je setříděná správně, protože odpovídající posloupnost hodnot $w(\{u,v\})$ hran této posloupnosti je

$$1, 2, 3, 4, 4, 5, 6, 6, 8.$$

Krok 2: Algoritmus nastaví $E_0 = \emptyset$ a $i = 0$.

Krok 3: Protože E_i, tedy E_0, neobsahuje $n-1 = 5$ hran, algoritmus zvýší hodnotu i, tedy provede $i := 1$, a pokusí se do E_0 přidat hranu $e_1 = \{c,f\}$. Protože tímto přidáním nevznikne kružnice, hranu přidá a vytvoří

$$E_1 = \{\{c,f\}\}.$$

Krok 3: E_i stále neobsahuje 5 hran, proto zvýší i na $i = 2$. Přidáním hrany $e_2 = \{b,f\}$ k E_1 kružnice nevznikne, proto se hrana přidá a vytvoří se

$$E_2 = \{\{c,f\}, \{b,f\}\}.$$

Krok 3: E_i neobsahuje 5 hran, proto se nastaví $i := 3$. Přidáním hrany $e_3 = \{b,c\}$ k E_2 by vznikla kružnice (spojovala by vrcholy b, c a f). Proto se hrana e_3 nepřidá a je

$$E_3 = \{\{c,f\}, \{b,f\}\}.$$

Krok 3: E_i stále 5 hran neobsahuje, proto se hodnota i zvýší na $i = 4$. Přidáním hrany $e_4 = \{a,b\}$ k E_3 kružnice nevznikne, a proto se hrana e_4 přidá. Máme tedy

$$E_4 = \{\{c,f\}, \{b,f\}, \{a,b\}\}.$$

Krok 3: E_i má méně než 5 hran, proto se provede $i := 5$. Přidáním hrany $e_5 = \{a,f\}$ by vznikla kružnice, proto se hrana e_5 nepřidá a je

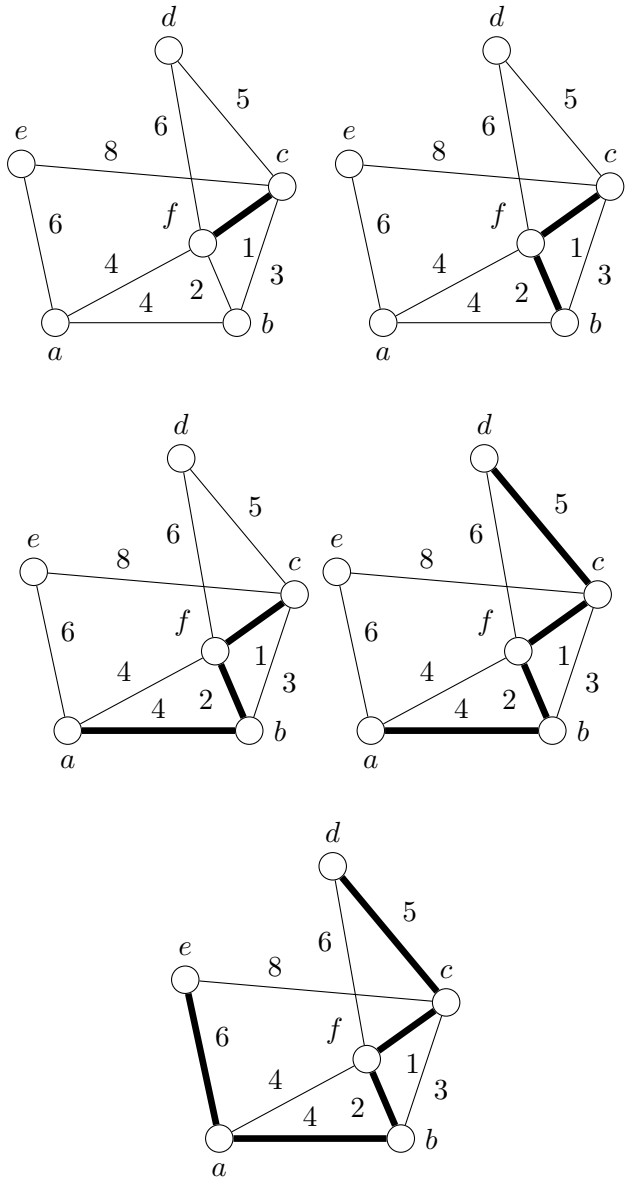
$$E_5 = \{\{c,f\}, \{b,f\}, \{a,b\}\}.$$

Krok 3: E_i má stále méně než 5 hran, proto provedeme $i := 6$. Přidáním hrany $e_6 = \{c,d\}$ kružnice nevznikne, proto hranu e_6 přidáme a získáme

$$E_6 = \{\{c,f\}, \{b,f\}, \{a,b\}, \{c,d\}\}.$$

Krok 3: E_i stále 5 hran neobsahuje, proto provedeme $i := 7$. Přidáním hrany $e_7 = \{a,e\}$ kružnice nevznikne, proto hranu e_7 přidáme. Získáme tedy

$$E_7 = \{\{c,f\}, \{b,f\}, \{a,b\}, \{c,d\}, \{a,e\}\}.$$



Obrázek 18: Hledání minimální kostry.

Krok 3: Protože E_i nyní obsahuje 5 hran, přejdeme na krok 4.

Krok 4: Výsledná množina hran, tedy minimální kostra grafu G, je

$$E' = E_7.$$

Uvedený postup ilustruje obr. 18. Hodnota získané kostry $G' = \langle V, E' \rangle$ je

$$\begin{aligned} w(G') &= w(\{c,f\}) + w(\{b,f\}) + w(\{a,b\}) \\ &\quad + w(\{c,d\}) + w(\{a,e\}) \\ &= 1 + 2 + 4 + 5 + 6 = 18. \end{aligned}$$

Vidíme, že tato kostra je levnější než výše uvedená kostra, která měla hodnotu 21. Snadno se přesvědčíme, že tato kostra je minimální.

Uvedený algoritmus je v literatuře známý jako Kruskalův algoritmus. Americký matematik a informatik Joseph Kruskal (1928–2010) ho publikoval v roce 1956.[13] Mnohem

[13] J. B. Kruskal, On the shortest spanning subtree of a graph and the traveling salesman problem. Proceedings of the American Mathematical Society. 7 (1)(1956): 48–50.

dříve, v roce 1926, však brněnský matematik Otakar Borůvka (1899-1995) publikoval jiný algoritmus pro hledání minimální kostry, dnes známý jako Borůvkův algoritmus.[14] Třetím známým je algoritmus původně objevený českým matematikem Vojtěchem Jarníkem (1897–1970)[15], který v roce 1957 znovuobjevil Robert Prim (1921–2009) a v roce 1959 potom Edsger Dijkstra (1930–2002).[16]

Pokud bychom měli Kruskalův algoritmus naprogramovat, aby ho mohl provádět počítač, musíme mimo jiné rozhodnout, jak testovat, zda graf $\langle V, E_{i-1} \cup \{e_i\}\rangle$ obsahuje kružnici, tedy zda přidáním hrany e_i ke grafu $\langle V, E_{i-1}\rangle$ vznikne kružnice. Tento test lze provést efektivně na základě následujícího pozorování. Graf $\langle V, E_{i-1}\rangle$ se rozpadá na komponenty, tj. na své maximální souvislé podgrafy. Pokud je takových komponent k, tvoří množiny vrcholů $V_1, \ldots, V_k$ těchto podgrafů rozklad množiny V (například graf vpravo na obr. 13 obsahuje dvě komponenty a příslušné množiny jsou $V_1 = \{1, 2, 3\}$ a $V_2 = \{4, 5, 6\}$). Protože graf $\langle V, E_{i-1}\rangle$ kružnice neobsahuje, neobsahuje kružnice ani žádná z jeho komponent. Každá jeho komponenta je tedy stromem. Dle věty 4.28 tedy přidáním hrany e_i vznikne kružnice, právě když oba vrcholy hrany e_i leží v jedné z množin V_l (kružnice nevznikne, když vrcholy hrany e_i leží v různých množinách). Takový test je možné provést poměrně rychle pomocí datových struktur pro uchování systému disjunktních množin.[17]

Zatím jsme se nezabývali zásadní otázkou: Je Kruskalův algoritmus správný? Tedy, lze dokázat, že pro každý souvislý neorientovaný graf $G = \langle V, E\rangle$ s ohodnocením $w : E \to \mathbb{R}^+$ algoritmus skončí a že $G' = \langle V, E'\rangle$ je minimální kostra daného grafu? Algoritmus je totiž poměrně jednoduchý a z jistého pohledu poměrně naivní. Z hran, které zbývají, přidává, pokud to jde, vždy hranu, která má nejmenší ohodnocení. Takovému postupu se říká žravý (angl. greedy) postup. Žravé postupy se často používají pro hledání suboptimálních řešení (tedy ne nutně optimálních, ale blízkých optimu) různých problémů. V našem případě však žravý postup funguje: Kruskalův algoritmus je správný.

Důkaz správnosti Kruskalova algoritmu (a) Nejdříve dokážeme, že algoritmus vždy skončí s hodnotou $i \leq m$ a vrátí množinu E', která obsahuje $n - 1$ hran.

Pokud algoritmus skončí s hodnotou $i < m$ (tedy neprošel všechny hrany z posloupnosti $e_1, \ldots, e_m$), pak E', tedy poslední vytvořená množina E_i, má zřejmě $n-1$ hran (jinak by nebyla splněna podmínka pro ukončení cyklu v kroku 3 a došlo by ke zvýšení hodnoty i).

Pokud algoritmus dojde k hodnotě $i = m$ (tedy tato hodnota se nastaví prvním příkazem kroku 3), pak se v druhé části kroku 3 zpracuje hrana e_m, tedy poslední hrana grafu G. Ukážeme sporem, že po provedení tohoto kroku má množina E_i, tedy množina E_m, $n-1$ hran, a tedy při dalším testu podmínky v kroku 3 dojde k přechodu na krok 4 a k ukončení algoritmu s množinou E' s $n - 1$ hranami.

Předpokládejme tedy, že E_m má méně než $n - 1$ hran. Protože vytvořený graf $\langle V, E_m\rangle$ má n vrcholů a méně než $n - 1$ hran a protože je sestrojen tak, že neobsahuje kružnice, nemůže být dle věty 4.28 souvislý (kdyby byl souvislý, byl by to dle definice strom s n vrcholy, ale strom s n vrcholy má $n - 1$ hran). Musí tedy existovat vrcholy u a v, které v sestrojeném grafu $\langle V, E_m\rangle$ nejsou propojeny cestou. Ve výchozím grafu G ale cestou

[14] O. Borůvka, O jistém problému minimálním. Práce Mor. přírodověd. spol. v Brně 3(1926): 37–58. O. Borůvka, Příspěvek k řešení otázky ekonomické stavby elektrovodních sítí. Elektrotechnický obzor 15(1926): 153–154.

[15] V. Jarník, O jistém problému minimálním. Práce Mor. přírodověd. spol. v Brně 6 (1930): 57–63.

[16] R. C. Prim, Shortest connection networks and some generalizations. Bell System Technical Journal 36 (6)(1957): 1389–1401. E. W. Dijkstra, A note on two problems in connexion with graphs. Numerische Mathematik 1 (1)(1959): 269–271.

[17] Lze ukázat, že časová složitost Kruskalova algoritmu v nejhorším případě je $T(n, m) = O(m \log n)$, kde $n = |V|$ a $m = |E|$. Lze také ukázat, že $T(n, m) = O(m \log m)$.

propojeny jsou, protože G je souvislý. Označme tuto cestu c. Alespoň jedna hrana e takové cesty nepatří do E_m (jinak by to byla cesta z u do v v grafu $\langle V, E_m\rangle$, ale ta neexistuje). Uvažujme nyní libovolnou takovou hranu $e = \{x, y\}$. Tato hrana je jednou z hran, které algoritmus prošel, tedy $e = e_i$ pro nějaké i, ale nepřidal. To znamená, že přidáním hrany e do E_{i-1} by vznikla kružnice v $\langle V, E_{i-1}\rangle$ obsahující $e = \{x, y\}$. Protože tato kružnice sestává z hran z množiny E_{i-1}, vyjmutím hrany e z této kružnice vznikne cesta $c_{x,y}$ z x do y v $\langle V, E_{i-1}\rangle$. To je i cesta v grafu $\langle V, E_m\rangle$. Nahradíme-li nyní každou hranu $e = \{x, y\}$ cesty c, která nepatří do E_m, cestou $c_{x,y}$, vznikne z c cesta z u do v obsahující jen hrany z E_m. To je spor s předpokladem, že v grafu $\langle V, E_m\rangle$ cesta z u do v neexistuje.

Dokázali jsme, že algoritmus vždy skončí s hodnotou $i \leq m$ a s množinou E' obsahující $n - 1$ hran.

(b) Dokažme nyní, že $G' = \langle V, E'\rangle$ je minimální kostrou. Z konstrukce množin E_i je zřejmé, že $G' = \langle V, E'\rangle$ neobsahuje kružnice. Protože G' má podle výše dokázané vlastnosti (a) $n - 1$ hran a protože má n vrcholů, je dle věty 4.28 stromem, a tedy kostrou grafu G.

Musíme tedy ukázat, že je-li $H = \langle V, F\rangle$ jiná kostra grafu G, má hodnotu aspoň takovou jako nalezená kostra G', tedy $w(G') \leq w(H)$. Víme, že E' obsahuje $n-1$ hran. Označme je $e'_1, \ldots, e'_{n-1}$, a to tak, že očíslování je v souladu s hodnotami těchto hran, tedy

$$w(e'_1) \leq \cdots \leq w(e'_{n-1}).$$

Graf $H = \langle V, F\rangle$ má $n-1$ hran (je to kostra grafu G, tedy strom s n vrcholy). Označme je $f_1, \ldots, f_{n-1}$, opět tak, že očíslování je v souladu s hodnotami, tj.

$$w(f_1) \leq \cdots \leq w(f_{n-1}).$$

Požadovanou nerovnost

$$w(G') = \sum_{i=1}^{n-1} w(e'_i) \leq \sum_{i=1}^{n-1} w(f_i) = w(H) \tag{15}$$

dokážeme následovně. Ukážeme, že

$$\text{pro každé } i = 1, \ldots, n-1 \text{ platí } w(e'_i) \leq w(f_i). \tag{16}$$

Dokážeme to sporem. Pokud (16) neplatí, označme j nejmenší index pro který $w(e'_j) > w(f_j)$. Protože $e'_1 = e_1$ (hrana e_1 je algoritmem vždy vybrána) je hrana s nejmenší hodnotou w, je jistě $j > 1$. Uvažujme množiny

$$\begin{aligned} E^* &= \{e'_1, \ldots, e'_{j-1}\} \text{ a} \\ F^* &= \{f_1, \ldots, f_j\}. \end{aligned}$$

K prokázání sporu stačí ukázat, že existuje hrana f_i z F^*, která není v E^* a pro kterou graf $\langle V, E^* \cup \{f_i\}\rangle$ neobsahuje kružnici. To se totiž vzhledem k tomu, jak algoritmus vybírá hrany, nemůže stát: Před přidáním hrany e'_j měl totiž přidat hranu f_i. To proto, že $w(f_i) \leq w(f_j) < w(e'_j)$, tudíž algoritmus hranu f_i navštívil před navštívením hrany e'_j. Měl ji přidat, protože by nevznikla kružnice, ale nepřidal ji (protože f_i není v E^*). To je požadovaný spor.

Ukažme tedy, že taková hrana f_i existuje. Graf $\langle V, E^*\rangle$ má n vrcholů, $j - 1$ hran, a proto vzhledem k $j - 1 < n - 1$ a k tomu, že nemá kružnice, nemůže být dle věty 4.28 stromem. Není tedy souvislý. Graf $\langle V, E^*\rangle$ se zřejmě rozpadá na $k \geq 2$ komponent $G_1, \ldots, G_k$, tedy na k maximálních souvislých podgrafů grafu $\langle V, E^*\rangle$. Označíme-li

$V_1, \ldots, V_k$ množiny vrcholů těchto komponent, obsahuje každý G_l právě ty hrany z E^*, které mají oba vrcholy ve V_l. Každý podgraf G_l je bez kružnic (protože ani $\langle V, E^* \rangle$ nemá kružnice), a je to tedy strom. Počet hran grafu G_l je tedy $|V_l| - 1$. Každá hrana z E^* je obsažena v jediném grafu G_l. Počet hran v E^* je tudíž roven součtu počtů $|V_l| - 1$ hran jednotlivých grafů G_l, neboli

$$|E^*| = \sum_{l=1}^{k} (|V_l| - 1) = \sum_{l=1}^{k} |V_l| - k = n - k. \tag{17}$$

Ukažme následující pomocné tvrzení (T): Počet hran z F^*, jejichž oba vrcholy se nacházejí v nějaké množině V_l, je nejvýše $n - k$. Protože $\langle V, F^* \rangle$ neobsahuje kružnice, neobsahuje kružnice ani jeho podgraf $\langle V_l, F_l^* \rangle$, kde F_l^* je množina hran z F^*, jejichž oba vrcholy jsou v V_l. Proto je $|F_l^*| \leq |V_l| - 1$, a tedy pro počet p hran z F^*, jejichž oba vrcholy se nacházejí v nějaké množině V_l skutečně platí

$$p = \sum_{l=1}^{k} |F_l^*| \leq \sum_{l=1}^{k} (|V_l| - 1) = n - k.$$

Existence požadované hrany f_i nyní plyne z tvrzení (T): Protože $n - k$ je dle (17) počet hran v E^* a protože F^* má o jednu hranu více než E^*, je jasné, že existuje hrana f_i množiny F^*, jejíž vrcholy leží v různých množinách V_l a $V_{l'}$. Hrana f_i nemůže být hranou z E^*, protože každá hrana z E^* má oba vrcholy v jedné množině V_l. Hrana f_i je tedy požadovanou hranou z F^*, která není v E^*. □

4.5.3 Kořenové stromy

Definice 4.31. *Kořenový strom* je dvojice $\langle G, r \rangle$, kde $G = \langle V, E \rangle$ je strom a $r \in V$ je vrchol, tzv. *kořen*.

Kořenový strom je tedy strom, ve kterém je vybrán jeden vrchol (kořen). Může to být kterýkoliv vrchol. Bývá to ale vrchol, který je v nějakém smysly na vrcholu hierarchie objektů, která je stromem reprezentována.

To, že je v kořenovém stromu jeden vrchol pevně zvolený a že ve stromu existuje mezi vrcholy jediná cesta, umožňuje ve kořenovém stromu zavádět uspořádání vrcholů. Na základě tohoto uspořádání se stromy kreslí. Základem je následující definice.

Definice 4.32. Nechť $\langle G, r \rangle$ je kořenový strom.

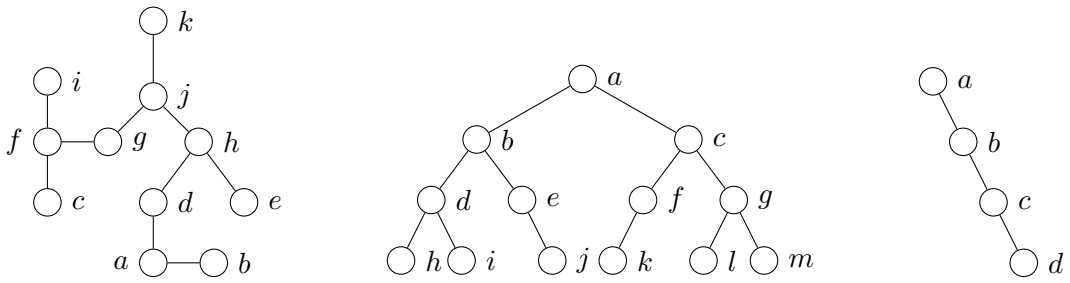
- Vrchol v se nazývá *potomek* vrcholu u (u se nazývá *předek* vrcholu v), právě když cesta z kořene r do v má tvar $r, \ldots, u, \ldots, v$.
- Vrchol v se nazývá *následník* (někdy *přímý potomek*) vrcholu u (u se nazývá *předchůdce* (někdy *rodič*) vrcholu v), právě když cesta z kořene r do v má tvar $r, \ldots, u, e, v$.
- Vrchol v se nazývá *list* kořenového stromu, právě když nemá žádného následovníka; jinak se nazývá *vnitřní vrchol*.
- *Hloubka* (někdy *úroveň*) vrcholu v je délka cesty od kořene r do v.
- *Výška* vrcholu v je délka nejdelší cesty neobsahující předky v, která vede z v do některého z listů.
- *Výška* (někdy *hloubka*) stromu je výška jeho kořene (ekvivalentně: je největší z hloubek jeho listů).

Kořenový strom se nazývá *m-ární*, právě když každý jeho vrchol má nejvýše m následovníků. 2-ární strom se nazývá *binární*; 3-ární strom se nazývá *ternární*.

Poznámka 4.33. (a) Protože ve stromu existuje mezi každými dvěma vrcholy právě jedna cesta, je každý vrchol, který není kořenem, následníkem právě jednoho vrcholu (jeho předchůdce je tedy určen jednoznačně). Vrchol ale může mít více následníků. Viz následující příklad.

(b) List v kořenového stromu $\langle G, r\rangle$ je také listem původního stromu G (dle definice nemá v následovníka, a protože má právě jednoho předchůdce, má stupeň 1, je tedy listem stromu G). Pokud má kořen r právě jednoho následovníka, není listem $\langle G, r\rangle$, ale vzhledem k tomu, že má stupeň 1, je listem listem G. Pokud kořen r nemá žádného následovníka, je listem (ale ne vnitřním vrcholem) v $\langle G, r\rangle$ i listem v G (pokud za list považujeme i vrchol stupně 0). Pokud má kořen více následovníků, je vnitřním vrcholem $\langle G, r\rangle$, a tedy není ani listem $\langle G, v\rangle$, ani listem G.

Příklad 4.34. (a) Uvažujme strom $G = \langle V, E\rangle$ z obr. 19 vlevo. Volbou libovolného vrcholu $r \in V$ získáme ze stromu G kořenový strom $\langle G, r\rangle$. Takto získané kořenové stromy jsou navzájem různé, protože mají různé kořeny, a liší se v nich i vztahy být potomkem a být následníkem, výšky a hloubky vrcholů i výška celého stromu.



Obrázek 19: Stromy z příkladu 4.34.

Zvolme například $r = i$ a uvažujme kořenový strom $\langle G, i\rangle$.

- Vezměme cestu $i, \{i, f\}, f, \{f, g\}, g, \{g, j\}, j, \{j, k\}, k$ (vynecháme-li hrany, můžeme takovou cestu stručně zapisovat i, f, g, j, k). Z její struktury je patrné, že k je potomkem vrcholů j, g, f a i (tyto uzly jsou tedy předchůdci k). Přitom k je přímý potomek, tedy následník, j (j je předchůdce k) a že k není následníkem žádného jiného vrcholu. Protože i je kořen a délka této cesty je 4, je hloubka vrcholu k rovna 4.

- Z cesty i, f, g, j, h, d, a, b je patrné, že hloubky vrcholů i, f, g, j, h, d, a, b jsou po řadě $0, 1, 2, 3, 4, 5, 6, 7$.

- Listy jsou vrcholy b, c, e a k. Délky cest z vrcholu i k těmto listům jsou po řadě $7, 2, 5$ a 4. Výška vrcholu i je tedy 7. Protože i, je kořenem, je jeho výška také výškou celého stromu.

- Výšky a hloubky všech vrcholů jsou následující:

vrchol	a	b	c	d	e	f	g	h	i	j	k
hloubka	6	7	2	5	5	1	2	4	0	3	4
výška	1	0	0	2	0	6	5	3	7	4	0

  Strom $\langle G, i\rangle$ je binární, protože počet následovníků každého vrcholu je nejvýše 2 (je to tedy i m-ární strom pro každé $m \geq 2$).

Zvolme nyní kořen $r = j$. Následovníci vrcholu j jsou v tomto případě vrcholy g, h a k a snadno se vidí, že $\langle G, j\rangle$ je ternární. Vidíme tedy, že změnou kořene se mohou změnit následovníci a předchůdci vrcholů. Změnit se mohou i výšky a hloubky vrcholů a listy. Například listy $\langle G, j\rangle$ jsou vrcholy b, c, e, i a k, výška kořene j, a tedy i výška $\langle G, j\rangle$ je 4.

(b) Uvažujme strom na obr. 19 uprostřed a zvolme $r = a$. Strom na obrázku je nakreslen tak, jak se kořenové strom kreslí: nejvýše (v úrovni 0) je kořen; pod ním (v úrovni 1) jsou jeho následníci, tedy vrcholy s hloubkou 1; pod nimi (v úrovní 2) jejich následníci, tedy vrcholy s hloubkou 2 atd. Strom $\langle G, a\rangle$ je binární.

(c) Zvolíme-li ve stromu na obr. 19 uprostřed $r = a$, získáme jiný kořenový strom, který má stejnou výšku jako strom z bodu (b).

Kořenové stromy z příkladu 4.34 (b) a (c) mají stejnou výšku, ale výrazně jiné počty vrcholů. Strom vlevo je však mnohem „košatější“ než ten vpravo, který je svým způsobem degenerovaný (je to seznam, každý vrchol kromě listu má právě jednoho následníka). Košatost (zaplněnost úrovní stromu vrcholy) znamená, že strom dané hloubky pojme mnoho vrcholů, což je důležité při použití stromů jako struktur pro ukládání a vyhledávání dat.

Definice 4.35. m-ární kořenový strom $\langle G, r\rangle$ výšky h se nazývá *zaplněný* pokud splňuje následující podmínky:

- každý vrchol s výjimkou vrcholů s hloubkou $h - 1$ má 0 nebo m následníků;
- každý list má hloubku h nebo $h - 1$.

Poznámka 4.36. Snadno se lze přesvědčit, že m-ární kořenový strom $\langle G, r\rangle$ výšky h je zaplněný, právě když vznikne postupným zaplňováním po vrstvách. Tedy má kořen, v hloubce 1 jeho m následníků, v hloubce 2 jejich m^2 následníků atd. až po hloubku $h - 1$ (ve které je všech možných m^{h-1} vrcholů); v hloubce h je pak k vrcholů, kde $1 \leq k \leq m^h$, které jsou libovolným způsobem připojeny v vrcholům hloubky $h - 1$.

Je-li strom zaplněný, pak z prvních dvou podmínek plyne, že každý vrchol hloubky menší než $h - 1$ má právě m následníků (podle první podmínky nemá buď žádného, nebo jich má m; kdyby neměl žádného, byl by to list hloubky menší než $h - 1$, což dle druhé podmínky není možné). V hloubce h pak je aspoň jeden vrchol, protože strom má výšku h.

Vznikne-li strom zaplňováním po vrstvách, je zřejmé, že obě podmínky z definice zaplněného stromu splňuje.

Poznámka 4.37. Pro m-ární stromy se používá různá terminologie související s „plností“ stromu. m-ární strom se například nazývá úplný, pokud každý jeho vrchol je list, nebo má právě m následníků (to je přísnější podmínka než naše podmínka zaplněnosti). Má-li navíc každý list stejnou hloubku, nazývá se takový strom perfektní (každý perfektní strom je zaplněný, ale ne naopak). Terminologie ale není jednotná.

Následujícím příkladem nahlédneme do problematiky tzv. vyhledávacích stromů.

Příklad 4.38 (binární vyhledávací stromy)**.** Binární vyhledávací stromy slouží k ukládání a vyhledávání dat (data jsou zpravidla číselné hodnoty). Jsou to binární kořenové stromy s dodatečnou informací, které splňuje jistá omezení usnadňující vyhledávání.

Dodatečná informace: Vrcholy binárního vyhledávacího stromu jsou ohodnoceny číselnými hodnotami; budeme předpokládat, že ohodnocení je funkce $w : V \to \mathbb{Z}$ (místo

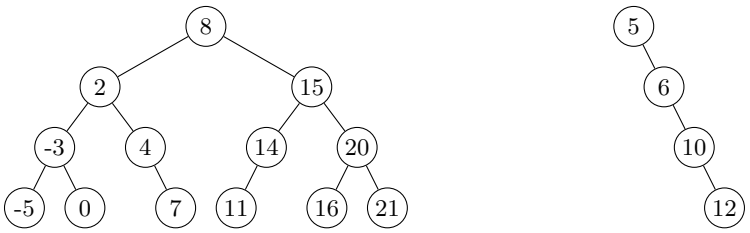
$\mathbb{Z}$ je možné uvažovat jinou číselnou množinu nebo obecněji jinou lineárně uspořádanou množinu). Číslo $w(v) \in \mathbb{Z}$ se interpretuje jako hodnota uložená ve vrcholu v. Ke každému následníku daného vrcholu je určeno, zda je to tav. levý následník, nebo pravý následník, přičemž pokud jsou následníci dva, jeden je levý a jeden pravý (levého následníka v kreslíme pod v vlevo, pravého vpravo).

Omezení:

- Je-li u levým následníkem vrcholu v nebo potomkem tohoto následníka, pak $w(u) \leq w(v)$;
- je-li u pravým následníkem vrcholu v nebo potomkem tohoto následníka, pak $w(u) \geq w(v)$.

Volně řečeno, vrcholy vlevo pod v mají hodnotu menší nebo rovnu hodnotě v; vrcholy vpravo pod v mají hodnotu větší nebo rovnu hodnotě v.

Dva binární vyhledávací stromy vidíme na obr. 20 (hodnoty vrcholů jsou kružnic znázorňujících vrcholy vepsány). Uvedená omezení umožňují v binárních vyhledávacích



Obrázek 20: Binární vyhledávací stromy.

stromech snadno vyhledávat. Předpokládejme, že máme za úkol zjistit, zda se ve stromu vyskytuje hodnota $x \in \mathbb{Z}$. Postup je následující:

1. $v :=$ kořen stromu;
2. je-li $w(v) = x$ odpověz Ano a skonči;
3. je-li $x < w(v)$ pak má-li v levého následníka u, proveď $v := u$ a jdi na krok 2, jinak odpověz Ne a skonči;
4. je-li $x > w(v)$ pak má-li v pravého následníka u, proveď $v := u$ a jdi na krok 2, jinak odpověz Ne a skonči.

Vyhledáváme-li ve stromu na obr. 20 vlevo hodnotu 15, algoritmus skončí s odpovědí Ano v hloubce $l = 1$, tedy po dvou porovnáních (hledanou hodnotu porováme postupně s hodnotami 8 a 15). V nejhorším případě dojdeme až do hloubky $l = 3$ (vyhledáváme-li např. hodnotu 0, která ve stromu je uložena, nebo hodnotu 1, která uložena není).

Poznámka 4.39. (a) Binární vyhledávací stromy jsou speciálním případem pozičních stromů. Kořenový strom se nazývá *poziční*, jestliže následníkům každého vrcholu jsou přiřazena přirozená čísla (pozice), přičemž různým následníkům daného vrcholu jsou přiřazena různá čísla. Mají-li všechna přiřazená čísla hodnotu nejvýše m, nazývá se poziční strom m-ární (pak je to zřejmě m-ární kořenový strom podle definice 4.32). Binární vyhledávací strom je tedy poziční binární kořenový strom s ohodnocením vrcholů, pokud je např. přiřazené číslo 1, jde-li o levého následníka, a 2, jde-li o pravého následníka.

(b) Pokud je podstatné pouze seřazení následníků každého vrcholu, nazývá se kořenový strom *uspořádaný* (budeme se jimi zabývat níže, viz definice 4.43). Má-li tedy vrchol kořenového stromu k následníků, musí být určeno, který z nich je prvním, druhým, ... až k-tým následníkem. Poziční strom lze považovat za uspořádaný (uspořádání je dáno přiřazenými pozicemi), naopak to neplatí (např. není jasné, zda jediný následník vrcholu v binárním stromu je levým nebo pravým následníkem).

Z příkladu 4.38 je zřejmé, má-li binární vyhledávací strom n vrcholů, tedy obsahuje-li n uložených hodnot, je při vyhledávání hodnoty v nejhorším případě třeba dostat se až do hloubky h rovné výšce stromu (a provést při tom $h+1$ porovnání s hledanou hodnotou). Čím je strom plnější, a tedy čím menší je jeho výška, tím rychlejší je vyhledávání.

Průvodce studiem

Vyhledávání v dobře zaplněném (dobře vyváženém) stromu přináší značnou výpočetní úsporu. Z příkladu 4.38 víme, že pro vyhledání hodnoty ve stromu s výškou h je v nejhorším případě potřeba hledanou hodnotu porovnat s $h + 1$ uloženými hodnotami. K vyhledávání hodnoty v seznamu (degenerovaném stromu, jako je jen na obr. 20 vpravo) je tedy třeba provést $n - 1$ porovnání. Pro 1500 hodnot je to tedy 1501 porovnání. n-hodnot lze ale umístit do zaplněného stromu a ten má dle věty 4.41 výšku $\lfloor \log_2 n \rfloor$. Pak stačí $\lfloor \log_2 n \rfloor + 1$ porovnání. Pro 1500 hodnot je to jen $\lfloor \log_2 1500 \rfloor + 1 = 11$ porovnání.

Poznámka 4.40. Binární vyhledávací stromy, které jsme si představili v příkladu 4.38, jsou základem mnoha datových struktur, kterým se říká vyhledávací stromy. Jde např. o tzv. AVL-stromy, červeno-černé stromy, 2-3 stromy, B-stromy a jiné. U takových datových struktur je třeba kromě metody vyhledávání navrhnout další metody, zejména metody pro vložení hodnoty a odstranění hodnoty. Důležité je, aby tyto metody – stejně jako výše uvedená metoda pro vyhledávání – pracovaly v čase (čas měříme počtem provedených instrukcí) úměrném $\log_2 n$ (popř. $\log_m n$ u m-árních stromů), kde n je počet vrcholů vyhledávacího stromu.Toho lze dosáhnout. Důležité a netriviální však je, aby při manipulaci se stromem (např. při přidávání a odstraňování hodnot) zajistit, aby výška stromu zůstala malá. To zajišťují metody pro tzv. vyvažování stromů, které je třeba při manipulaci stromem provádět.

Přirozeně se tedy objevuje otázka, jaká je výška binárního stromu s n vrcholy. Tato otázka je významná nejen v kontextu vyhledávacích stromů, ale kvůli řadě jiných situací, ve kterých se m-ární stromy vyskytují. Základní vztahy mezi výškou, počtem vrcholů, popř. počtem listů udávají následující tvrzení. Připomeňme, že pro reálné číslo x je $\lceil x \rceil$ nejmenší celé číslo, které je větší nebo rovno x, tj.

$$\lceil x \rceil = \min\{m \in \mathbb{Z} \mid x \leq m\} \quad \text{a} \quad \lfloor x \rfloor = \max\{m \in \mathbb{Z} \mid x \leq m\},$$

tedy $\lceil x \rceil$ vznikne zaokrouhlením x „nahoru“: $\lceil 1.2 \rceil = 2$, $\lceil 5.8 \rceil = 6$, $\lceil 3 \rceil = 3$; podobně $\lfloor x \rfloor$ vznikne zaokrouhlením „dolů“.

Věta 4.41. *V zaplněném binárním kořenovém stromu s n vrcholy a l listy, který má výšku h, platí*

(a) $2^h \leq n \leq 2^{h+1} - 1$;

(b) $h = \lfloor \log_2 n \rfloor$;

(c) $2^{h-1} \leq l \leq 2^h$;

(d) $\lceil \log_2 l \rceil \leq h \leq \lfloor \log_2 l \rfloor + 1$.

Důkaz. (a): Má-li strom výšku h, má dle poznámky 4.36 v hloubkách $l = 0, 1, \ldots, h-1$ postupně $2^0, 2^1, \ldots, 2^{h-1}$ vrcholů, celkem tedy $\sum_{l=0}^{h-1} 2^l$ vrcholů. Platí

$$\sum_{l=0}^{h-1} 2^l = 2^h - 1,$$

což lze nahlédnout vyjádřením součtu ve dvojkové soustavě: Protože zápis čísla 2^l sestává z 1 následované l nulami, je součet $\sum_{l=0}^{h-1} 2^l$ číslo, jehož zápisem je h cifer 1 (pro $h = 4$ je to v dvojkové soustavě $1 + 10 + 100 + 1000 = 1111$). Přičteme-li k tomuto součtu jedničku, získáme zřejmě číslo, jehož zápisem je 1 následovaná h nulami, což je číslo 2^h. Uvedený součet je tedy roven $2^h - 1$.

K těmto $2^h - 1$ vrcholům je v hloubce h navíc aspoň 1 vrchol, tedy $n \geq (2^h - 1) + 1 = 2^h$, nejvýše však 2^h vrcholů, tedy $n \leq (2^h - 1) + 2^h = 2 \cdot 2^h - 1 = 2^{h+1} - 1$. Odvodili jsme tedy $2^h \leq n \leq 2^{h+1} - 1$.

(b): Protože $\log_2$ je rostoucí funkce, dostaneme z odvozené nerovnosti $2^h \leq n \leq 2^{h+1} - 1$ nerovnost

$$\log_2 2^h \leq \log_2 n \leq \log_2(2^{h+1} - 1).$$

Protože z $x \leq y$ plyne $\lfloor x \rfloor \leq \lfloor y \rfloor$, dostaneme z této nerovnosti

$$\lfloor \log_2 2^h \rfloor \leq \lfloor \log_2 n \rfloor \leq \lfloor \log_2(2^{h+1} - 1) \rfloor. \tag{18}$$

Uvědomme si,že $\log_2 2^h = h$, a tedy protože h je celé číslo, je $\lfloor \log_2 2^h \rfloor = \lfloor h \rfloor = h$. Vzhledem k tomu, že $\log_2 2^h = h$, je $\log_2(2^{h+1} - 1) < h + 1$, zároveň ale $\log_2(2^{h+1} - 1) > h$, a tedy $\lfloor \log_2(2^{h+1} - 1) \rfloor = h$. Z nerovnosti (18) tedy plyne

$$h \leq \lfloor \log_2 n \rfloor \leq h,$$

musí tedy být $h = \lfloor \log_2 n \rfloor$.

(c): Vzhledem k tomu, co jsme uvedli na začátku bodu (a) tohoto důkazu, má strom s výškou h s nejmenším počtem listů za listy všechny vrcholy hloubky $h - 1$ až na jeden z nich, který má právě jednoho následníka. Nejmenší možný počet l listů je tedy roven počtu vrcholů hloubky $h - 1$, a těch je 2^{h-1}, což dokazuje první nerovnost. Naopak strom výšky h s největším počtem listů má u každého z 2^{h-1} dva následníky, které jsou právě listy; listů je tedy 2^h. V obecném případě je tedy $2^{h-1} \leq l \leq 2^h$.

(d): Zlogaritmováním nerovnosti z (c) dostaneme

$$h - 1 \leq \log_2 l \leq h.$$

Z druhé nerovnosti dostaneme $\lceil \log_2 l \rceil \leq \lceil h \rceil = h$; z první pak $h \leq \log_2 l + 1$, odkud dostaneme $h = \lfloor h \rfloor = \lfloor \log_2 l + 1 \rfloor = \lfloor \log_2 l \rfloor + 1$.

□

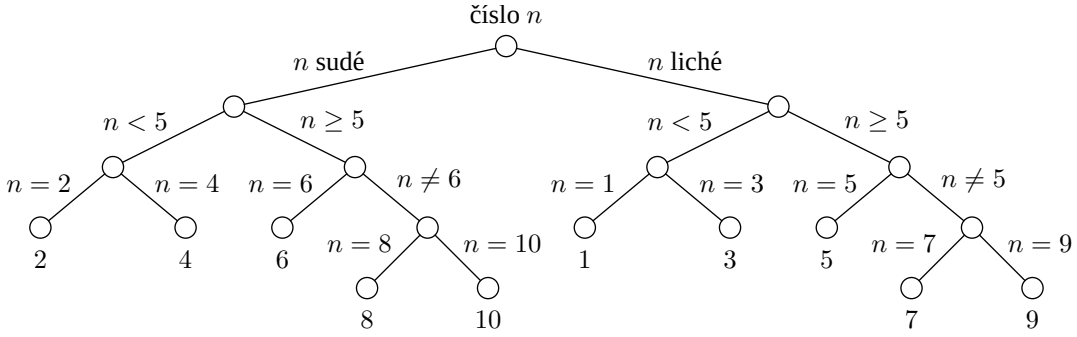
Věta 4.42. *V libovolném binárním kořenovém stromu s n vrcholy a l listy, který má výšku h, platí*

(a) $h + 1 \leq n \leq 2^{h+1} - 1$;

(b) $\lfloor \log_2 n \rfloor \leq h \leq n - 1$;

(c) $1 \leq l \leq 2^h$;

(d) $\lceil \log_2 l \rceil \leq h$.



Obrázek 21: Strom pro hádání čísla z $1, \ldots, 10$.

Důkaz. (a): První nerovnost plyne ze skutečnosti, že strom výšky h s nejmenším počtem vrcholů je řetězec (každý vrchol kromě listu má právě jednoho následníka) a ten má $h + 1$ vrcholů. Druhá plyne z druhé nerovnosti věty 4.41 (a) a toho, že strom výšky h s největším počtem vrcholů je zaplněný.

(b): Druhá nerovnost plyne z první nerovnosti v (a). První plyne z věty 4.41 (b) a z toho, že strom výšky h musí mít výšku rovnou nebo větší, než je výška stromu s n vrcholy, který je zaplněný (čtenář nechť tuto skutečnost zdůvodní sám).

(c): První nerovnost je triviální (jeden list má řetězec, tedy $l = 1$ může nastat). Druhá nerovnost plyne z druhé nerovnosti ve větě 4.41 (c) a z toho, že strom výšky h s největším počtem listů je ten, který je zaplněný a v hloubce h má všech možných 2^h listů.

(d): Nerovnost dostaneme zlogaritmováním druhé nerovnosti z bodu (c) a z toho, že $\lceil h \rceil = h$. □

Průvodce studiem

Vztahy mezi výškou, počtem vrcholů a počtem listů mají časté použití. Předpokládejme, že máme pomocí otázek typu Ano/Ne uhodnout kartu z balíčku 32 karet. Pro jednoduchost očíslujme karty čísly $1, \ldots, 32$, a předpokládejme tedy, že máme uhodnout číslo mezi 1 a 32. Jedna možnost (optimální) je první otázkou rozdělit čísla na dvě stejné části (např. otázka „Je číslo menší než 17?“, tj. části po 16 kartách. Další položená otázka, např. „Je číslo sudé?“ zbylých 16 karet opět rozdělí na dvě poloviny. Atd. až dojdeme k jednomu číslu, a to je to, které hádáme. Např. pro číslo 7 může být posloupnost dotazů kladených hadačem a posloupnost odpovědí následující. Dotaz: „Je číslo menší než 17?“ Odpověď: „Ano.“ Zbývají $1, \ldots, 16$. D: „Je číslo sudé?“ O: „Ne.“ Zbývají 1, 3, 5, 7, 9, 11, 13, 15. D: „Je číslo menší než 9?“ O: „Ano.“ Zbývají 1, 3, 5, 7. D: „Je číslo menší než 5?“ O: „Ne.“ Zbývají 5, 7. D: „Je to číslo 5?“ O: „Ne.“ Zbývá 7 a to je číslo, které hádáme. Strategie volit otázky tak, abychom vždy zredukovali počet možností (pokud možno) na polovinu se nazývá *metoda půlení*. Důležité je, že existuje-li l možností, metodou půlení se dobereme správné možnosti nejpozději v $\lceil \log_2 l \rceil$ krocích. Strom, který odpovídá naší situaci, je totiž vyvážený úplný binární strom o l listech. Výška stromu je právě počet otázek, které musíme v nejhorším případě položit. Na obr. 21 je strom, odpovídající situaci, kdy hádáme čísla z $1, \ldots, 10$.

Vraťme se k uspořádaným stromům.

Definice 4.43. Kořenový strom se nazývá *uspořádaný*, je-li ke každému vrcholu, který není listem, zadáno lineární uspořádání jeho potomků.

Formálněji, uspořádaný kořenový strom je struktura $\langle\langle V, E\rangle, r, \{\leq_v | \ v \in V\}\rangle$, kde $\langle\langle V, E\rangle, r\rangle$ je kořenový strom a pro vrchol $v \in V$ je $\leq_v$ lineární uspořádání na množině $P_v = \{v_1, \ldots, v_n\}$ všech potomků vrcholu v, pokud v není list, a $\leq_v = \emptyset$, pokud je v list. Jsou-li potomkové vrcholu v uspořádáni $v_1 \leq_v \cdots \leq_v v_n$, říkáme, že v_1 je první potomek v atd. V tomto pořadí je také kreslíme po vrchol v.

Častým úkolem spojeným se stromy je projít všechny vrcholy stromu a v každém provést nějakou akci. Ve vrcholech stromů mohou být například uloženy nějaké informace (například výše uvažované číselné hodnoty). Náš úkol může úkol může být vypsat všechny tyto informace (tj. projít všechny vrcholy a pro každý vrchol vypsat informaci, která je v něm uložena). Máme tedy za úkol pro každý vrchol provést operaci `zpracuj(v)`. Přitom `zpracuj(v)` může znamenat „vypiš informaci uloženou ve v“ apod.

Ukážeme si teď dva způsoby procházení kořenového stromu, tzv. *preorder* a *postorder*. Popíšeme je jako procedury `preorder(v)` a `postorder(v)`, které pracují následovně. `v` je vstupní parametr, za který můžeme dosadit libovolný vrchol stromu. Je-li pak `u` konkrétní vrchol stromu, znamená `preorder(u)` „vyvolání“ procedury `preorder` pro vrchol `u`. Stejně je to pro `postorder(u)`. Procedury mají tvar

```
preorder(v)
{
 zpracuj(v);
 jsou-li v1,...,vn potomci v v jejich usporadani, proved
   zpracuj(v1),...,zpracuj(vn).
}
```

a

```
postorder(v)
{
 jsou-li v1,...,vn potomci v v jejich usporadani, proved
   zpracuj(v1),...,zpracuj(vn);
 zpracuj(v).
}
```

Pro uspořádaný kořenový strom $\langle R, r\rangle$ způsobí `preorder(r)` průchod a zpracování stromu metodou preorder, `postorder(r)` způsobí průchod a zpracování metodou postorder. Tedy např. u metody preorder se nejprve zpracuje kořen r (`zpracuj(v)`) a má-li r potomky $v_1, \ldots, v_n$ (takto uspořádané), vyvolá se průchod metodou preorder ve vrcholu v_1 (`preorder(v1)`), po dokončení tohoto průchodu se se vyvolá průchod ve vrcholu v_2 (`preorder(v2)`) atd. až po průchod ve vrcholu v_n (`preorder(vn)`). Přitom průchod `preorder(v1)` ve v_1 probíhá tak, že se zpracuje v_1, tj. proběhne `zpracuj(v1)`, a pak dojde k vyvolání průchodů v případných potomcích vrcholu v_1.

Uvažujme uspořádaný kořenový strom na obr. 20 vlevo a předpokládejme, že uspořádání následníků vrcholů je dáno tím, jak jsou nakresleny (zleva doprava). Pokud `zpracuj(v)` vypíše číselnou hodnotu uloženou ve vrcholu v a je-li r kořen, pak průchodem preorder budou čísla vypsána v pořadí

$$8, 2, -3, -5, 0, 4, 7, 15, 14, 11, 20, 16, 21.$$

Při průchodu postorder budou vypsána v pořadí

$$-5, 0, -3, 7, 4, 2, 11, 14, 16, 21, 20, 15, 8.$$

U binárních uspořádaných stromů se někdy používá průchod metodou *inorder*.

```
inorder(v)
{
 je-li v1 prvni potomek vrcholu v, zpracuj(v1);
 zpracuj(v);
 je-li v1 prvni potomek vrcholu v, zpracuj(v2).
}
```

Při průchodu výše uvažovaného stromu inorder budou čísla vypsána v pořadí

$$-5, -3, 0, 2, 4, 7, 8, 11, 14, 15, 16, 20, 21.$$

Shrnutí

Stromy jsou speciální grafy, které lze definovat několika ekvivalentními způsoby, např. jako grafy bez kružnic. Stromy mají četné aplikace v informatice. Speciálními případy stromů jsou m-ární stromy, kořenové stromy, uspořádané kořenové stromy. Pro stromy jsou odvozeny užitečné vztahy mezi jejich charakteristikami.

Pojmy k zapamatování

- strom, list, kostra,
- kořenový strom, úroveň vrcholu, hloubka stromu, vyvážený strom,
- m-ární strom, uspořádaný kořenový strom, preorder, postorder, inorder.

Kontrolní otázky

1. *Co je to strom? Uveďte několik definic. Je strom totéž co kostra?*
2. *Proč se pojmy úroveň vrcholu a hloubka stromu zavádějí až pro kořenové stromy. Jaká je největší možná hloubka kořenového stromu s n vrcholy?*
3. *Je každý m-ární strom vyvážený?*

Cvičení

1. Nechť $\langle \langle V, E \rangle, r \rangle$ je kořenový strom a $v \in V$. Ukažte, že podgraf G_v indukovaný množinou $V_v = \{u \in V \mid \text{cesta z } u \text{ do } r \text{ prochází vrcholem } v\}$ je strom (tzv. podstrom indukovaný vrcholem v). Ukažte také, že $u \in V_v$, právě když úroveň vrcholu u je větší nebo rovna úrovni vrcholu v a existuje cesta z u do v, která neprochází kořenem r.
2. Ukažte, že v úplném n-árním stromu, který má n vrcholů, l listů a i vnitřních vrcholů (ty, které nejsou listy) platí (a) $n = mi + 1$, (b) $l = (m-1)i + 1$, (c) $i = (l-1)/(m-1)$.
3. Jaký je nejmenší počet listů úplného m-árního stromu výšky h? Jaký je nejmenší počet vrcholů úplného m-árního stromu výšky h?

Úkoly k textu

1. Dokažte podrobně Větu 4.27.
2. Dokažte, že v kořenovém stromu má každý vrchol právě jednoho rodiče.

Řešení

1. G_v neobsahuje kružnici, protože je to podgraf stromu, a ten neobsahuje kružnici. Zbývá ukázat, že G_v je souvislý. Když $u, w \in V_v$, pak dle definice cesta ta je ve stromu jediná) z u do r i cesta z w do r prochází vrcholem v. Vezmeme-li úseky z u do v (z první cesty) a z v do w (z druhé cesty), jejich spojením dostaneme cestu z u do w. Tedy G_v je souvislý, a tedy je to strom. Vezměme $u \in V_v$. Pak existuje cesta z u do r, která prochází v, tedy dle definice je úroveň u je větší nebo rovna úrovni v a existuje cesta z u do v, která neprochází kořenem r. Když je úroveň u je větší nebo rovna úrovni v a existuje cesta z u do v, která neprochází kořenem r, uvažujme cestu z u do r. Ta musí obsahovat v. Jinak by by hloubka u byla menší než hloubka v nebo by cesta vzniklá spojením cesty z u do r a cesty z r do v byla cestou z u do v, která obsahuje r (podrobně rozeberte).
2. Snadné, plyne téměř přímo z definic a základních vztahů.
3. Představte si, jak vypadá takový strom s nejméně vrcholy. Obsahuje kořen a v každé z následujících h úrovní má právě m vrcholů. Má tedy $1 + h \cdot m$ vrcholů (1 kořen plus m vrcholů v každé z h úrovní) a $(m-1)(h-1) + m$ listů (v 1. až $(h-1)$. úrovni po $m-1$ listech, v poslední úrovni m listů).

5 Kombinatorika

Studijní cíle: Po prostudování kapitol 5.1, 5.2 a 5.3 by student měl být znát základy kombinatorického počítání. Měl by znát pravidla součtu a součinu, pojmy permutace, variace a kombinace. Student by měl umět v základních úlohách samostatně provést správnou kombinatorickou úvahu. Měl by být schopen použít pravidla součtu a součinu k rozložení složitější úlohy na jednodušší.

Klíčová slova: kombinatorika, pravidlo součtu, pravidlo součinu, permutace, permutace s opakováním, variace, variace s opakováním, kombinace, kombinace s opakováním

5.1 Co a k čemu je kombinatorika

Kombinatorika je jednou z nejužitečnějších oblastí diskrétní matematiky. Zabývá se určováním počtu možností (konfigurací), které existují za určitých předepsaných podmínek. Může nás například zajímat, kolika způsoby je možné vyjádřit přirozené číslo n ve tvaru součtu $n_1 + \cdots + n_k$ přirozených čísel $n_1, \ldots, n_k$ přičemž nezáleží na pořadí čísel v součtu. Zde se jednou možností rozumí čísla $n_1, \ldots, n_k$. Předepsané podmínky v tomto případě říkají, že musí platit $n_1 + \cdots + n_k = n$ a dále že možnosti $n_1, \ldots, n_k$ a $n'_1, \ldots, n'_k$ se považují za shodné (počítají se jako jedna možnost), pokud se liší jen pořadím čísel (např. možnosti $1, 1, 2$ a $1, 2, 1$ se považují za shodné, $1, 1, 2$ a $1, 2, 2$ nikoli). Tak například pro číslo 3 existují 3 možnosti (ty možnosti jsou $1 + 1 + 1$, $1 + 2$, 3), pro číslo 4 existuje 5 možností ($1 + 1 + 1 + 1$, $1 + 1 + 2$, $1 + 3$, $2 + 2$, 4) atd.

Kombinatorika se zabývá určováním počtu možností, které mohou nastat za předepsaných podmínek.

Průvodce studiem

V časopise BYTE Magazine kdysi vyšla následující zpráva. "According to ... WEB Technologies' vice president of sales and marketing, the compression algorithm used by DataFiles/16 is not subject to the laws of information theory" (BYTE Magazine 17(6):45, June 1992). Představitelé WEB Technologies tvrdili, že jejich kompresní program DataFiles/16 komprimuje všechny typy souborů na přibližně jednu šestnáctinu jejich původní velikosti a že pro soubory velikosti aspoň 64KB je tato komprese bezeztrátová. Jednoduchá kombinatorická úvaha však ukazuje, že to není možné.

Uvažujme např. délku souboru $16n$ bitů. Existuje celkem 2^{16n} různých souborů délky $16n$ bitů. Každý takový soubor by podle WEB Technologies mělo být možné zkomprimovat na výsledný soubor délky nejvýše n bitů. Přitom existuje právě 2^k různých souborů délky k bitů. Tedy navzájem různých souborů délky nejvýše n bitů existuje $2^0 + 2^1 + 2^2 + \cdots + 2^n = 2^{n+1} - 1$. Protože ale $2^{n+1} - 1 < 2^{16n}$, není taková komprese možná. Kompresí totiž vyrobíme z daného souboru délky $16n$ bitů (těch je 2^{16n}) některý ze souborů délky nejvýše rovné n (těch je $2^{n+1} - 1$). Proto musí existovat různé soubory délky $16n$, které se kompresí převedou na stejný soubor délky nejvýše rovné n. Taková komprese tedy není bezeztrátová.

Kombinatorika má použití v mnoha praktických oblastech našeho života.

Příklad 5.1. Předpokládejme, že heslo pro přístup do databáze je posloupnost sestávající z právě 5 povolených znaků. Mezi povolené znaky patří písmena `a`,..., `z`, `A`, ..., `Z`, číslice 0, 1, ..., 9. Platí přitom, že heslo musí začínat písmenem. Kolik existuje různých hesel?

Protože písmen je 52 (26 malých a 26 velkých) a číslic je 10, použitím pravidla součinu (viz dále) zjistíme, že hesel je $52 \cdot 62^4 = 768\,369\,472$. Neznáme-li heslo, musíme tedy to spávné „uhodnout" z cca 768 miliónů možných hesel. Úvahy tohoto typu musí umět provádět každý, kdo se zabývá bezpečností počítačových systémů.

Příklad 5.2. Uvažujme následující variantu hazardní hry. Z osudí obsahujícího míčky s čísly $1,\ldots,20$ jsou vylosovány 3 míčky. Hra spočívá v tom, že si před losováním můžeme vsadit na námi vybraná 3 čísla. Za to zaplatíme 10 Kč. V případě, že uhodneme všechna 3 později vylosovaná čísla, vyhrajeme 20 000 Kč, jinak nevyhrajeme nic. Vyplatí se vsadit si?

Vybrat 3 míčky z 20 je možné 1 140 způsoby (je to počet kombinací 3 z 20, viz dále). My si vsadíme na 1 takový výběr. Pravděpodobnost, že trefíme ten správný, je tedy $\frac{1}{1140}$. Z dlouhodobého hlediska tedy vyhrajeme v 1 z 1 140 případů. V takových 1 140 případech tedy vyhrajeme $1 \times 20\,000 = 20\,000$ Kč, přitom za vsázení utratíme $1\,140 \times 10 = 11\,400$ Kč. Vsadit si se tedy vyplatí.

Příklad 5.3. Předpokládejme, že kódujeme elementární zprávy (zpráva může být znak nebo nějaká posloupnost znaků) tak, že každou zprávu zakódujeme jako posloupnost n symbolů 0 a 1, tzv. kódové slovo. Takovému kódu se říká binární kód délky n. Binární kód délky n tedy můžeme považovat za nějakou množinu posloupností délky n, které sestávají z 0 a 1. Např. $\{100, 010, 001\}$ je binární kód délky 3. Ten může být použit např. pro kódování výsledků nějakého procesu, kde výsledek je jeden z tří možných typů (prohra, remíza, výhra; rychlost ≤ 50 km/h, rychlost > 50, ale < 70 km/h, rychlost ≥ 70 km/h), tak, že např. prohra je kódována posloupností 100, remíza posloupností 010, výhra posloupností 001.

První otázka: Chceme-li kódovat k znaků binárním kódem délky n, jaké nejmenší n musíme zvolit, abychom zaručili možnost jednoznačného dekódování? Aby byl kód jednoznačně dekódovatelný, musí obsahovat aspoň k posloupností. Přitom posloupností z 0 a 1, které mají délku n, je právě 2^n (podle pravidla součinu, viz dále). Délka n musí tedy splňovat $k \leq 2^n$, tj. $\log_2 k \leq n$.

Druhá otázka: Předpokládejme, že na kódující posloupnosti 0 a 1 působí rušivé vlivy a že se proto může 0 změnit na 1 a 1 změnit na 0. Takové chyby jsou ale málo časté. Přijmeme-li posloupnost v délky n, může být zatížena chybami, a nemusí to tedy nutně být nějaké kódové slovo. Protože předpokládáme, že chyby jsou málo časté, je intuitivně přirozené chápat v jako chybou změněné kódové slovo w, a to takové w, které se ze všech kódových slov od v nejméně liší. Ve výše uvedeném případě např. 110 není kódovým slovem a jemu nejbližší kódová slova jsou 100 a 010. Vzdáleností posloupností přitom rozumíme počet pozic, na kterých se liší, tj. vzdálenost posloupností $a_1 \cdots a_n$ a $b_1 \cdots b_n$ je počet prvků množiny $\{i \mid a_i \neq b_i\}$. Vzdálenost 110 a 100 je tedy rovna 1 (liší se právě jednou pozicí). Pokud je kód dobře navržený, může dekódování probíhat tak, že přijatá posloupnost w délky n se opraví a výsledkem bude nejbližší kódové slovo. Jak jsme viděli, výše uvedený kód není dobře navržený, protože ke slovu 110 existují dvě kódová slova (100 a 010) se stejnou vzdáleností od 110. Jaký je největší počet k kódových posloupností binárního kódu délky n, který umožňuje opravu jednoduchých chyb? Přitom jednoduchá chyba je ta, která vznikne změnou právě jednoho symbolu posloupnosti (jednoduchou chybou vznikne z 001 např. 101, ale už ne 110). Uvažujme takto: Nechť takový kód obsahuje právě k kódových slov. Vezměme libovolné z nich a označme ho v. Slovem v bude interpretována nejen posloupnost v, ale i každá posloupnost, jejíž vzdálenost od v je 1 (tyto posloupnosti budou opraveny na v). Posloupností, které mají od v vzdálenost 1, je právě n (chyba může být na libovolné z n pozic). Slov, které připadají na kódové slovo v v tom smyslu, že budou po případné opravě jedné chyby převedeny na v, je tedy celkem $1 + n$. Protože na každé z k kódových slov takto připadá $n + 1$ navzájem různých posloupností délky n a protože počet všech posloupností nul a jedniček délky n je 2^n, musí být $k \cdot (n + 1) \leq 2^n$, tedy $k \leq \frac{2^n}{n+1}$. Největší počet kódových posloupností binárního kódu délky n, který umožňuje opravu jednoduchých chyb, je tedy $\frac{2^n}{n+1}$. Pro $n = 3$ je tedy největší počet $\frac{2^3}{3+1} = 2$. Kód s 3 kódovými slovy opravující jednoduché chyby tedy musí mít délku n aspoň 4 (protože pro délku $n = 3$ je největší počet kódových slov 2). Výše uvedený příklad tedy nelze

spravit tím, že vybereme jiná kódová slova délky 3.

Uvedené příklady představují typické problémy, kterými se kombinatorika zabývá. Přesněji řečeno, kombinatorika se, jako každá oblast matematiky, zabývá obecnými principy, které je potom možné na konkrétní situace z praktického života použít. Tak například předpokládejme, že víme, kolika způsoby je možné vybrat dvouprvkovou podmnožinu $\{x, y\}$ z daných n prvků. Označme počet těchto způsobů $D(n)$. Víme tedy, že $D(n) = \frac{n(n-1)}{2}$ (vyzkoušejte nebo to přímo odvoďte). Pak je snadné spočítat, že z 30 studentů je možné vybrat dvojici studentů 435 způsoby (neboť $435 = \frac{30 \cdot 29}{2}$), že existuje právě 499 500 způsobů jak vybrat dva míčky z tisíce ($499\,500 = \frac{1000 \cdot 999}{2}$) atd.

Upozorníme nyní na důležitou věc. I v kombinatorice se setkáme s tím, že pro různé situace odvodíme různé vzorce (jako výše uvedený vzorec $D(n) = \frac{n(n-1)}{2}$). Včas však varujme: Střežme se mechanického používání vzorců! V kombinatorice snad více než kde jinde platí, že k tomu, abychom byli vůbec schopni vybrat pro danou situaci „správný vzorec“, musíme situaci rozebrat a dokonale jí porozumět. Přitom toto porozumění je často netriviální záležitost (tomu tak není např. u derivování funkcí: máme-li například spočítat derivaci funkce $x^2 \cdot \sin(x)$, stačí znát vzorce pro derivování x^2, $\sin(x)$ a vzorec pro derivování součinu funkcí; použití vzorce pro úlohu spočítání derivace dané funkce je tedy téměř triviální). Řešení kombinatorické úlohy se spíše podobá řešení slovní úlohy: neexistuje obecný předpis pro řešení. Situaci musíme nejdříve dobře porozumět, pokud možno rozložit ji na jednodušší situace a ty potom vyřešit pomocí základních kombinatorických pravidel. Tato základní pravidla mohou mít podobu vzorců. Nesrovnatelně důležitější než naučit se vzorce, je ale naučit se používat základní kombinatorická pravidla (tato pravidla jsou pravým smyslem kombinatorických vzorců) a provádět kombinatorické úvahy. Bez porozumění kombinatorickým pravidlům nejsme schopni řešit jiné než triviální kombinatorické úlohy.

5.2 Pravidla součtu a součinu

Pravidlo součtu a pravidlo součinu jsou dvě základní kombinatorická pravidla. Mnoho dalších pravidel vzniká jejich kombinováním.

Základní kombinatorická pravidla jsou pravidlo součtu a pravidlo součinu.

Průvodce studiem

Pravidlo součtu: Lze-li úkol A provést m způsoby a lze-li úkol B provést n způsoby, přičemž žádný z m způsobů provedení úkolu A není totožný s žádným z n způsobů provedení úkolu B, pak provést úkol A nebo úkol B lze provést $m + n$ způsoby.

Pravidlo součtu je zjevné: množina způsobů, jak provést A nebo B, je sjednocením m-prvkové množiny způsobů, jak provést A, a n-prvkové množiny způsobů, jak provést B, a má tedy vzhledem k tomu, že výchozí dvě množiny způsobů nemají společný prvek, $m + n$ prvků. Ukážeme, jak lze pravidlo součtu použít.

Příklad 5.4. V knihovně je 5 knih, jejichž autorem je A. C. Doyle, a 10 knih, jejichž autorkou je A. Christie. Čtenář si tedy může vybrat 15 způsoby knihu, kterou napsali A. C. Doyle nebo A. Christie. Je-li A úkol „vybrat knihu, jejíž autorem je A. C. Doyle“ a B úkol „vybrat knihu, jejíž autorem je A. Christie“, pak je $m = 5$ a $n = 10$. Přitom přitom provést úkol A nebo úkol B znamená vybrat knihu, kterou napsali A. C. Doyle nebo A. Christie. Podle pravidla součtu to lze právě $m + n = 15$ způsoby. Použití pravidla součtu je oprávněné, protože žádná kniha, kterou napsal A. C. Doyle, není totožná s žádnou knihou, kterou napsala A. Christie.

Příklad 5.5. Množiny M a N jsou disjunktní (tj. nemají společné prvky) a platí $|M| = m$ a $|N| = n$. Kolika způsoby lze vybrat prvek, který patří do M nebo do N? Jsou-li A a B po řadě úkoly „vybrat prvek z množiny M“ a „vybrat prvek z množiny N“, pak předpoklady pravidla součtu jsou splněny (M a N nemají společné prvky), a proto existuje $m + n$ způsobů, jak vybrat prvek z M nebo N. Jinými slovy, jsou-li M a N disjunktní množiny, je $|M \cup N| = |M| + |N|$.

Poznamenejme, že předpoklad pravidla součtu, která říká, že žádný z m způsobů provedení úkolu A není totožný s žádným z n způsobů provedení úkolu B, je podstatná. Uvažujme příklad 5.5, ale vezměme množiny, které nejsou disjunktní, např. $M = \{a, b, c\}$, $N = \{b, c, d, e\}$. Jak snadno vidíme, existuje 5 způsobů, jak vybrat prvek z M nebo N, přitom $5 \neq 3 + 4 = m + n$.

Pravidlo součtu lze zobecnit na konečný počet úkolů: Pokud úkol A_1 lze provést m_1 způsoby, úkol A_2 lze provést m_2 způsoby, ..., úkol A_k lze provést m_k způsoby, přičemž po každou dvojici A_i a A_j ($i \neq j$) žádný z m_i způsobů provedení úkolu A_i není totožný s žádným z m_j způsobů provedení úkolu A_j, pak provést úkol A_1 nebo úkol A_2 nebo úkol A_k lze provést $m_1 + m_2 + \cdots + m_k$ způsoby.

Příklad 5.6. Nechť $M_1, \ldots, M_k$ jsou konečné po dvou disjunktní množiny. Kolik prvků má sjednocení $M_1 \cup \cdots \cup M_k$? Pomocí zobecněného pravidla součtu můžeme podobně jako v Příkladě 5.5 ukázat, že $|M_1 \cup \cdots \cup M_k| = |M_1| + \cdots + |M_k|$.

Průvodce studiem

Pravidlo součinu: Lze-li úkol C rozložit na po sobě následující úkoly A a B (tj. provést C znamená provést nejdřív A a potom B) a lze-li úkol A provést m způsoby a úkol B lze provést n způsoby, pak lze úkol C provést $m \cdot n$ způsoby.

Také pravidlo součinu je zjevné: každý způsob, jak řešit C vznikne volbou jednoho způsobu řešení úkolu A a jednoho způsobu řešení úkolu B. Vzhledem k tomu, že ke každému z m způsobů řešení A lze zvolit libovolný z n způsobů řešení B, máme celkem $m \cdot n$ způsobů. Princip součinu si ukážeme na příkladech.

Příklad 5.7. Kolik prvků má kartézský součin $M \times N$ dvou konečných množin M a N? Připomeňme, že $M \times N = \{\langle x, y\rangle \mid x \in M, y \in N\}$. Určit libovolný prvek $\langle x, y\rangle \in M \times N$ znamená totéž, co splnit úkol „zvol x a zvol y“. Tento úkol lze rozložit na úkol „zvol x“ a úkol „zvol y“. Prvek x lze přitom zvolit $|M|$ způsoby, prvek y lze zvolit $|N|$ způsoby. Podle pravidla součinu lze tedy úkol „zvol x a zvol y“ provést $|M| \cdot |N|$ způsoby. Proto $|M \times N| = |M| \cdot |N|$.

Podobně jako pravidlo součtu lze i pravidlo součinu zobecnit na na konečný počet úkolů: Lze-li úkol C rozložit na po sobě následující úkoly $A_1, \ldots, A_k$ a lze-li úkol A_i provést m_i způsoby (pro každé $i = 1, \ldots, k$), pak lze úkol C provést $m_1 + \cdots + m_k$ způsoby.

Příklad 5.8. Registrační značka vozidla má tvar PKC CCCC, kde P, K, a C jsou symboly a přitom P je některá z číslic 1–9, K je písmeno, určující příslušnost ke kraji (např. „T“ označuje Moravskoslezský kraj, „H“ označuje Královéhradecký apod.) a C je některá z číslic 0–9. Kolik lze v rámci jednoho kraje přidělit registračních značek?

První symbol lze zvolit 9 způsoby, druhý symbol nelze volit, protože je v rámci kraje pevně daný, třetí symbol lze zvolit 10 způsoby, stejně tak lze 10 způsoby zvolit čtvrtý, pátý, šestý i sedmý symbol. Podle zobecněného pravidla součinu tedy existuje v rámci jednoho kraje $9 \cdot 10 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 9 \cdot 10^5$ (devět set tisíc) možných různých registračních značek.

Pravidla součtu a součinu se často v jedné úloze kombinují. Ukažme jednoduchý příklad.

Příklad 5.9. Nechť A, B, C jsou konečné množiny, přičemž A a B jsou disjunktní. Kolik prvků má množina $(A \cup B) \times C$?

Úkol vybrat libovolně prvek z $(A \cup B) \times C$ lze rozložit na dva následující úkoly „vyber prvek z $A \cup B$“ a „vyber prvek z C“ (viz příklad 5.7). Přitom úkol „vyber prvek z $A \cup B$“ znamená „vyber prvek z A nebo vyber prvek z B“ a lze ho podle pravidla součtu provést $|A|+|B|$ způsoby (viz příklad 5.5). Proto lze dále podle pravidla součinu prvek z $(A \cup B) \times C$ vybrat $(|A|+|B|) \cdot |C|$ způsoby, tedy $|(A \cup B) \times C| = (|A|+|B|) \cdot |C|$.

5.3 Permutace, variace, kombinace

Kolika způsoby lze seřadit určitý počet objektů? Kolika způsoby lze vybrat určitý počet objektů z daných objektů, když na pořadí výběru záleží? Co když na pořadí výběru nezáleží? Co když se prvky ve výběru nemohou opakovat? Co když se opakovat mohou? Tyto a podobné otázky se často objevují v různých kombinatorických úlohách. Odpovědi na ně lze nalézt použitím pravidel součtu a součinu. Protože se však tyto otázky objevují opravdu často, odvodíme si vzorce, které na některé tyto otázky odpovídají. Vzorce, které odvodíme, patří k základům kombinatorického počítání. Nejprve však ještě jednou varování.

Průvodce studiem

Při používání kombinatorických vzorců, které uvedeme, je důležité vzorci dobře rozumět, „vidět do něj“, umět ho kdykoli odvodit. Důležitější než vzorce samotné jsou totiž úvahy, které k nim vedou. Vzorec je jen symbolickým vyjádřením závěru kombinatorické úvahy. Osvojíme-li si odpovídající úvahy, potřebné vzorce si nakonec můžeme odvodit sami (nebo je někde najdeme). Když si odpovídající úvahy neosvojíme, budou nám nejspíš vzorce k ničemu, neboť je u jen trochu složitějších úloh nebudeme umět používat. Čtenáři proto následující doporučení: Nesnažte se učit vzorce. Snažte se pochopit a naučte se sami provádět úvahy. Uvidíte, že věci jsou ve skutečnosti jednoduché.

5.3.1 Permutace

Student si u zkoušky vybere tři otázky. Může si vybrat, v jakém pořadí na ně bude odpovídat. Kolik má možností? Označme otázky A, B a C. Možná pořadí odpovídání jsou ABC, ACB, BAC, BCA, CAB, CBA. Je jich tedy šest. Tak přicházíme k pojmu permutace.

Permutace nějakých prvků je jejich seřazení.

Definice 5.10 (permutace)**.** *Permutace* n (navzájem různých) objektů je libovolné seřazení těchto objektů, tj. seřazení od prvního k n-tému. Počet permutací n objektů budeme značit $P(n)$.

Věta 5.11. $P(n) = n!$.

Důkaz. Jedno, ale libovolné, seřazení dostaneme tak, že vybereme 1. prvek (to lze provést n způsoby), poté vybereme 2. prvek (to lze provést $n-1$ způsobem, protože jeden prvek jsme již vybrali), poté vybereme 3. prvek (to lze provést $n-2$ způsoby), ..., nakonec vybereme n-tý prvek (to lze provést jedním způsobem, $n-1$ prvků totiž již bylo vybráno a zbývá poslední prvek). Podle pravidla součinu lze takový výběr provést $n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 1 = n!$ způsoby. Tedy $P(n) = n!$. □

Seřazujeme-li objekty, z nichž některé jsou stejné, provádíme tzv. permutace s opakováním.

U permutací s opakováním mohou být některé seřazované prvky stejné.

Definice 5.12 (permutace s opakováním)**.** Je dáno n objektů rozdělených do r skupin, které mají po řadě $n_1, \ldots, n_r$ objektů, tj. $n_1 + \cdots + n_r = n$. Objekty v každé ze skupin jsou navzájem nerozlišitelné. Každé seřazení těchto n objektů se nazývá *permutace s opakováním* (daným parametry $(n_1, \ldots, n_r)$). Počet takových permutací značíme $P(n_1, \ldots, n_r)$.

Věta 5.13. *Pro $n_1 + \cdots + n_r = n$ je $P(n_1, \ldots, n_r) = \frac{n!}{n_1! \cdot \cdots \cdot n_r!}$.*

Důkaz. Uvažujme libovolnou permutaci s opakováním. Očíslujme objekty v rámci každé z r skupin tak, aby se staly rozlišitelnými. Pak dané permutaci s opakováním odpovídá několik permutací očíslovaných objektů v tom smyslu, že na pozici i $(1 \leq i \leq n)$ je v permutaci s opakováním objekt z j-té skupiny $(1 \leq i \leq n)$, právě když je na pozici i v permutaci očíslovaných objektů objekt, který vznikl očíslováním z některého objektu z j-té skupiny.

Pro příklad: Mějme $n = 5$, $r = 2$, $n_1 = 3$, $n_2 = 2$, objekty první skupiny značíme A, objekty druhé skupiny značíme B. Po očíslování budeme mít objekty A_1, A_2, A_3, B_1, B_2. Permutaci s opakováním ABAAB odpovídá např. permutace $A_3B_1A_2A_1B_2$, ne však permutace $A_3A_2A_1B_1B_2$.

Kolik permutací očíslovaných objektů odpovídá každé permutaci s opakováním? Objekty první skupiny můžeme na jejich pozicích (ty jsou pro danou permutaci s opakováním dány pevně a je jich n_1) seřadit $n_1!$ způsoby (tolik je permutací n_1 prvků), ..., objekty r-té skupiny můžeme na jejich pozicích seřadit $n_r!$ způsoby. Protože seřazování objektů každé skupiny provádíme nezávisle na seřazování objektů libovolné jiné skupiny. Proto je celkový počet permutací očíslovaných objektů, které odpovídají libovolné permutaci s opakováním, roven $n_1! \cdot \cdots \cdot n_r!$.

To znamená, že počet všech permutací očíslovaných objektů, tedy $P(n)$, je $n_1! \cdot \cdots \cdot n_r!$-krát větší než celkový počet permutací s opakováním, tedy

$$P(n) = (n_1! \cdot \cdots \cdot n_r!) \cdot P(n_1, \ldots, n_r),$$

odkud plyne $P(n_1, \ldots, n_r) = \frac{n!}{n_1! \cdot \cdots \cdot n_r!}$. □

Příklad 5.14. Kolik slov (i nesmyslných) lze sestavit přerovnáním písmen ve slově POSTOLOPRTY? Počet slov je roven počtu seřazení písmen slova POSTOLOPRTY. Jde o permutace objektů s opakováním. Máme $n = 11$ objektů (písmen), které jsou rozděleny do $r = 7$ skupin odpovídajících jednotlivým písmenům P, O, S, T, L, R, Y. Počty objektů v jednotlivých skupinách jsou $n_P = 2$, $n_O = 3$, $n_S = 1$, $n_T = 2$, $n_L = 1$, $n_R = 1$, $n_Y = 1$. Počet slov je tedy $P(2, 3, 1, 2, 1, 1, 1) = \frac{11!}{2!3!2!}$.

5.3.2 Variace

Na lodi jsou čtyři důstojníci. Z nich je třeba jmenovat kapitána a jeho zástupce. Kolika způsoby to lze provést? Označme důstojníky písmeny A, B, C, D. Pak existuje těchto 12 způsobů: AB (A je kapitán, B jeho zástupce), AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC.

Variace je výběr, u kterého záleží na pořadí vybíraných prvků.

Definice 5.15 (variace)**.** Je dáno n (navzájem různých) objektů a číslo $r \leq n$. *Variace* r (objektů) z n (objektů) je libovolný výběr r objektů z daných n objektů, ve kterém záleží na pořadí vybíraných objektů. Počet takových variací značíme $V(n, r)$.

Ve výše uvedeném příkladu je $n = 4$ (máme 4 objekty) a $r = 2$ (vybíráme dva objekty). Variace BA je výběr, ve kterém je jako první vybrán objekt B a jako druhý objekt A. Variace BA a AB jsou různé (záleží na pořadí). Celkem existuje 12 takových variací, tj. $V(4, 2) = 12$.

Věta 5.16. $V(n, r) = n \cdot (n-1) \cdot \cdots \cdot (n-r+1)$.

Důkaz. Každá variace je dána tím, jaké objekty jsou na 1., 2., ..., r-tém místě. Objekt na 1. místě lze zvolit n způsoby (vybíráme z n objektů), objekt na 2. místě pak $n-1$ způsoby (vybíráme z $n-1$ objektů, protože jeden objekt je už na 1. místě), ..., objekt na r-tém místě lze vybrat $n-r+1$ způsoby (tolik objektů ještě k výběru zbývá). Podle pravidla součinu je tedy celkový počet takto provedených výběrů, tj. počet všech variací, $n \cdot (n-1) \cdot \cdots \cdot (n-r+1)$. □

Poznámka 5.17. Všimněme si, že $V(n, r) = \frac{n!}{(n-r)!}$. Skutečně,

$$\frac{n!}{(n-r)!} = \frac{n \cdot (n-1) \cdots (n-r+1) \cdot (n-r) \cdots 1}{(n-r) \cdots 1} = n \cdot (n-1) \cdots (n-r+1) = V(n, r)$$

Všimněte si také, že $V(n, n) = n! = P(n)$, tj. počet variací n a n je stejný jako počet permutací n objektů. To není náhoda. Variace n a n je vlastně výběr n prvků z n prvků, ve kterém záleží na pořadí. Je to tedy uspořádání, tj. permutace, n prvků (první vybraný prvek je v daném uspořádání na prvním místě, ..., n-tý vybraný prvek je v daném uspořádání na n-tém místě).

Výběry, ve kterých se prvky mohou opakovat, nazýváme variace s opakováním.

U variací s opakováním může být každý prvek vybrán několikrát.

Definice 5.18 (variace s opakováním)**.** Jsou dány objekty n různých typů. Objektů každého typu je neomezeně mnoho a jsou navzájem nerozlišitelné. *Variace r* (objektů) z n (objektů) *s opakováním* je libovolný výběr r objektů z daných objektů n typů, ve kterém záleží na pořadí vybíraných objektů. Počet takových variací značíme $\overline{V}(n, r)$.

Protože jsou prvky jednotlivých typů nerozlišitelné, jsou dvě variace s opakováním stejné, právě když mají na odpovídajících si místech (prvním až r-tém) objekty stejných typů.

Věta 5.19. $\overline{V}(n, r) = n^r$.

Důkaz. První prvek můžeme vybrat n způsoby, druhý prvek můžeme vybrat n způsoby, ..., r-tý můžeme vybrat n způsoby. Podle pravidla součinu lze tedy výběr provést $n \cdot \cdots \cdot n = n^r$ způsoby. □

Příklad 5.20. Zámek na kolo s kódem má pro nastavení kódu tři otáčecí kolečka. Na každém z nich lze nastavit číslice 0, 1, ..., 9. Předpokládejme, že nastavení a zkouška jedné číselné kombinace trvá 2 sekundy. Jak dlouho trvá v průměrném případě otevření zámku, neznáme-li správnou číselnou kombinaci (průměrný případ definujeme jako aritmetický průměr nejlepšího a nejhoršího případu)?

Číselné kombinace jsou 000 až 999. Jsou to tedy variace 3 z 10 s opakováním (3 pozice, 10 číslic). Těch je $10^3 = 1000$. V nejlepším případě nastavíme správnou kombinaci už v 1. pokusu (to trvá 2 sekundy), v nejhorším až v 1000. pokusu (to trvá 2000 sekund). V průměrném případě je to tedy $\frac{(2+2000)}{2} = 1001$ sekund (což je 16 minut a 41 sekund).

Poznámka 5.21. Variace s opakováním bychom mohli definovat jinak, a to následovně: Je dáno n (navzájem různých) objektů a číslo r. Variace r (objektů) z n (objektů) s opakováním (definovaná alternativně) je libovolný výběr r objektů z daných n objektů, ve kterém záleží na pořadí vybíraných objektů a ve kterém se prvky po výběru vracejí mezi prvky, ze kterých se vybírá. Uvědomme si, že způsob výběru zde je jiný, než v Definici 5.18. Důležité však je, že počet variací s opakováním je i v tomto případě $\overline{V}(n,r)$ (ověřte).

Příklad 5.22. Z místa A je třeba předávat na místo B zprávu o tom, jak dopadla akce konaná v místě A. Přitom existuje celkem 20 000 možných výsledků té akce. Předpokládejme, že pro zakódování výsledku se použije posloupnost $k = 2$ různých symbolů (např. 0 a 1), která má délku d. Jaká je nejmenší délka takové posloupnosti? Jak to bude při jiných počtech symbolů $k = 3, 4, 5$?

Jde o to, najít nejmenší délku d tak, aby posloupností k symbolů bylo aspoň 20000, tj. aby každý výsledek mohl být nějakou posloupností zakódován. Vybíráme-li z k symbolů posloupnost délky d, vybíráme vlastně variaci d z k s opakováním. Takových posloupností je tedy $\overline{V}(k,d) = k^d$. Chceme tedy najít nejmenší d tak, aby $k^d \geq 20000$. Pro $k = 2$ je $d = 15$, pro $k = 3$ je $d = 10$, pro $k = 4$ je $d = 8$, pro $k = 5$ je $d = 7$.

5.3.3 Kombinace

V táboře jsou 4 muži (označme je A, B, C, D). Kolika způsoby z nich lze vybrat dvoučlennou hlídku? Výběr hlídky je dán výběrem dvou z nich, tedy dvouprvkovou podmnožinou množiny $\{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$. Hlídky tedy mohou být $\{\mathrm{A}, \mathrm{B}\}$, $\{\mathrm{A}, \mathrm{C}\}$, $\{\mathrm{A}, \mathrm{D}\}$, $\{\mathrm{B}, \mathrm{C}\}$, $\{\mathrm{B}, \mathrm{D}\}$, $\{\mathrm{C}, \mathrm{D}\}$, je jich tedy 6.

Kombinace je výběr, u kterého nezáleží na pořadí vybíraných prvků.

Definice 5.23 (kombinace)**.** Je dáno n (navzájem různých) objektů a číslo $r \leq n$. *Kombinace* r (objektů) z n (objektů) je libovolný výběr r objektů z daných n objektů, ve kterém nezáleží na pořadí vybíraných objektů. Počet takových kombinací značíme $\binom{n}{r}$.

Čísla $\binom{n}{r}$ se nazývají *kombinační čísla* a označují se také $C(n,r)$ (čte se "en nad er").

Ve výše uvedeném příkladu je $n = 4$ (máme 4 objekty) a $r = 2$ (vybíráme dva objekty). Kombinace $\{\mathrm{A}, \mathrm{C}\}$ je výběr, ve kterém jsou vybrány A a C. Celkem existuje 6 takových kombinací, tj. $\binom{4}{2} = 6$.

Věta 5.24. $\binom{n}{r} = \frac{n!}{(n-r)!r!}$.

Důkaz. Víme, že $V(n,r) = \frac{n!}{(n-r)!}$. Uvědomme si, že každé kombinaci r z n odpovídá tolik variací r z n, kolika způsoby lze uspořádat r vybraných objektů (u kombinace záleží jen na vybraných objektech, ne na jejich uspořádání, kdežto u variace záleží i na jejich uspořádání). Např. kombinaci $\{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$ odpovídají variace ABC, ACB, BAC, BCA, CAB, CBA. Existuje $r!$ způsobů, jak uspořádat r objektů. Je tedy

$$\text{počet kombinací } r \text{ z } n \text{ krát počet uspořádání } r \text{ objektů} = \text{počet variací } r \text{ z } n,$$

tj.

$$\binom{n}{r} \cdot r! = V(n,r).$$

Odtud $\binom{n}{r} = \frac{V(n,r)}{r!} = \frac{n!}{(n-r)!r!}$. □

Přímo z odvozeného vzorce plyne

$$\binom{n}{r} = \binom{n}{n-r}.$$

Skutečně, $\binom{n}{n-r} = \frac{n!}{(n-(n-r))!(n-r)!} = \frac{n!}{(n-r)!r!} = \binom{n}{r}$. Dále platí, že

$$\binom{n}{n} = 1 \quad \text{a} \quad \binom{n}{0} = 1.$$

Poznámka 5.25. Vzorec pro $\binom{n}{r}$ lze odvodit také takto. Očíslujme n objektů, ze kterých vybíráme, čísly 1 až n. Kombinaci r z n můžeme vyjádřit jako řetězec n nul a jedniček, který obsahuje právě r jedniček, přičemž na i-tém místě toho řetězce je 1, právě když se v dané kombinaci nachází i-tý prvek. Např. jsou-li a, b, c, d první až čtvrtý prvek a, pak řetězci 0110 odpovídá kombinace $\{b, c\}$. Takový řetězců existuje právě tolik, kolik existuje permutací n prvků s opakováním, které jsou rozděleny do dvou skupin obsahujících r prvků (jedničky) a $n-r$ prvků (nuly). Takových permutací je podle věty 5.13 $\frac{n!}{(n-r)!r!}$.

Příklad 5.26. Kolika způsoby lze vybrat 4 předměty z nabídky 10 volitelných předmětů? Výběr předmětů je kombinace 4 z 10. Výběr lze tedy provést $\binom{10}{4} = \frac{10!}{6!\cdot 4!} = \frac{10\cdot 9\cdot 8\cdot 7}{4\cdot 3\cdot 2} = 210$ způsoby.

Řekneme si teď dva užitečné vztahy. Prvním z nich je, že pro $k < n$ platí

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}. \tag{19}$$

Odvoďme to: $\binom{n-1}{k-1} + \binom{n-1}{k} = \frac{(n-1)!}{(k-1)!(n-k)!} + \frac{(n-1)!}{k!(n-k-1)!} = \frac{k\cdot(n-1)!+(n-k)\cdot(n-1)!}{k!(n-k)!} = \frac{(k+(n-k))\cdot(n-1)!}{k!(n-k)!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$.

Druhým je tzv. *binomická věta*.

Věta 5.27. *Pro libovolná $a, b \in \mathbb{R}$ a nezáporné celé číslo n platí*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k. \tag{20}$$

Důkaz. Matematickou indukcí. Místo důkazu této věty uvedeme důkaz jejího důsledku, viz níže. □

Věta 5.28. *Pro reálné číslo x a nezáporné celé n je*

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k. \tag{21}$$

Důkaz. Věta je důsledkem binomické věty. Uvedeme přímý důkaz.

Dokážeme to indukcí přes n.

Indukční předpoklad: Pro $n = 0$ je tvrzení zřejmé. Např. pro $n = 0$ je $(1+x)^0 = 1$ a $\sum_{k=0}^{0} \binom{n}{k} x^k = \binom{0}{0} x^0 = 1$.

Indukční krok: Předpokládejme, že tvrzení platí pro $n-1$, tj. $(1+x)^{n-1} = \sum_{k=0}^{n-1}\binom{n-1}{k}x^k$, a dokažme ho pro n. Máme

$$\begin{aligned}
(1+x)^n &= (1+x)(1+x)^{n-1} = (1+x)\sum_{k=0}^{n-1}\binom{n-1}{k}x^k = \\
&= \sum_{k=0}^{n-1}\binom{n-1}{k}x^k + \sum_{k=0}^{n-1}\binom{n-1}{k}x^{k+1} = \sum_{k=0}^{n-1}\binom{n-1}{k}x^k + \sum_{k=1}^{n}\binom{n-1}{k-1}x^k = \\
&= \binom{n-1}{0}x^0 + \sum_{k=1}^{n-1}(\binom{n-1}{k-1}x^k + \binom{n-1}{k-1}x^k) + \binom{n-1}{n-1}x^n = \\
&= x^0 + \sum_{k=1}^{n-1}(\binom{n-1}{k-1} + \binom{n-1}{k-1})x^k + x^n = \\
&= \binom{n}{0}x^0 + \sum_{k=1}^{n-1}\binom{n}{k}x^k + \binom{n}{n}x^n = \sum_{k=0}^{n}\binom{n}{k}x^k.
\end{aligned}$$

Přitom jsme použili vzorec (19). □

Podobně lze dokázat následující, obecnější verzi: Pro libovolná $a, b \in \mathbb{R}$ a nezáporné celé číslo n platí

$$(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k. \tag{22}$$

Binomická věta má řadu použití.

Příklad 5.29. Určeme $(x^2+2y)^5$. Dle obecnější podoby binomické věty je

$$\begin{aligned}
(x^3+2y)^5 &= \binom{5}{0}(x^3)^5(2y)^0 + \binom{5}{1}(x^3)^4(2y)^1 + \binom{5}{2}(x^3)^3(2y)^2 + \\
&\quad \binom{5}{3}(x^3)^2(2y)^3 + \binom{5}{4}(x^3)^1(2y)^4 + \binom{5}{5}(x^3)^0(2y)^5 = \\
&= \binom{5}{0}x^{15} + \binom{5}{1}x^{12}2y + \binom{5}{2}x^9 4y^2 + \\
&\quad \binom{5}{3}x^6 8y^3 + \binom{5}{4}x^3 16y^4 + \binom{5}{5}32y^5 = \\
&= x^{15} + 10x^{12}y + 40x^9y^2 + 80x^6y^3 + 80x^3y^4 + 32y^5.
\end{aligned}$$

Příklad 5.30 (počet podmnožin n-prvkové množiny)**.** Dosazením $x = 1$ dostáváme

$$2^n = \sum_{k=0}^{n}\binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n}.$$

Protože $\binom{n}{k}$ je počet všech k-prvkových podmnožin n-prvkové množiny, udává součet vpravo počet 0-prvkových plus počet 1-prvkových plus ... plus počet n-prvkových, tj. počet všech podmnožin n-prvkové množiny. Pomocí binomické věty tedy vidíme, že je roven 2^n.

K tomu lze ale dojít i takto: Seřaďme prvky dané n-prvkové množiny za sebe. Představme si n pozic, které odpovídají 1., 2.,..., n. prvku. Do pozic budeme umísťovat 0 a 1. Podmnožiny jednoznačně odpovídají umístěním 0 a 1 do těchto pozic: je-li na i-té pozici 1, pak i-tý prvek patří do dané podmnožiny, je-li tam 0, pak do ní nepatří. Podmnožin n-prvkové množiny je tedy právě tolik, kolika způsoby lze do n pozic umístit nuly a jedničky. Tento počet je roven počtu variací n ze 2 (vybíráme z $\{0,1\}$), tedy je to $\overline{V}(n,2) = 2^n$.

Způsobů, jak vyřešit kombinatorický problém, bývá několik.

Průvodce studiem

Počet všech podmnožin n-prvkové množiny je 2^n. Lze k tomu dojít několika způsoby. Dva z nich jsme ukázali v Příkladu 5.30. Taková situace, kdy k jednomu výsledku můžeme dojít několika způsoby, je pro kombinatoriku typická. Různé způsoby odpovídají různým pohledům na věc. Například u počtu všech podmnožin byl první způsob "sečti počty všech 0-prvkových, 1-prvkových, ..., n-prvkových podmnožin", druhý způsob byl "představ si podmnožiny jako posloupnosti nul a jedniček a urči počet těchto posloupností". Obecný návod, jak si problém vhodně představit, není. Záleží jen na naší představivosti.

Výběr, ve kterém nezáleží na pořadí prvků a ve kterém se prvky mohou opakovat, se nazývá kombinace s opakováním. Vede k tomu následující úloha. V obchodě mají 4 typy zákusků (věnečky, řezy, špičky a trubičky). Máme koupit 6 zákusků. Kolika způsoby to lze provést? Jeden možný způsob je koupit 6 věnečků, další je koupit 6 větrníků, další je koupit 2 větrníky a 4 řezy, další je koupit věneček, řez, špičku a 3 větrníky atd. Důležité je, zaprvé, že pořadí zákusků v nákupu je nepodstatné, a zadruhé, že v nákupu mohou být zákusky stejného typu (zákusky se mohou opakovat).

U kombinací s opakováním může být každý prvek vybrán několikrát.

Definice 5.31 (kombinace s opakováním)**.** Jsou dány objekty n různých typů. Objektů každého typu je neomezeně mnoho a jsou navzájem nerozlišitelné. *Kombinace* r (objektů) z n (objektů) *s opakováním* je libovolný výběr r objektů z daných objektů n typů, ve kterém nezáleží na pořadí vybíraných objektů. Počet takových kombinací značíme $\overline{C}(n,r)$.

Že jsou objekty jednotlivých typů nerozlišitelné, znamená, že dvě kombinace s opakováním považujeme za stejné, právě když pro každý z n typů obsahují stejné počty objektů toho typu. U příkladu se zákusky to např. znamená, že každé dva nákupy obsahující dva větrníky a čtyři špičky, považujeme za stejné (byť v jednom nákupu mohou být jiné dva věnečky než ve druhém).

Věta 5.32. $\overline{C}(n,r) = \binom{n+r-1}{n-1}$.

Důkaz. Podívejme se na výběr takhle. Máme n přihrádek, které odpovídají typům objektů. Vybrat kombinaci r z n s opakováním znamená umístit do těchto přihrádek celkem r kuliček. Počet kuliček v i-té přihrádce můžeme totiž chápat jako počet objektů typu i, které jsme vybrali. Hledaný počet kombinací $\overline{C}(n,r)$ je tedy stejný jako počet umístění r kuliček do n přihrádek.

Abychom určili počet takových umístění, budeme každé umístění reprezentovat posloupností nul (reprezentují přepážky mezi přihrádkami) a jedniček (reprezentují kuličky). Např. pro $n = 4$ a $r = 6$ řetězec 101100111 reprezentuje umístění, kdy je v první přihrádce 1 kulička, ve druhé 2 kuličky, ve třetí 0 kuliček, ve čtvrté 3 kuličky. Tedy: první jednička reprezentuje 1 kuličku v první přihrádce; následující nula reprezentuje přepážku; následující dvě jedničky reprezentují 2 kuličky ve druhé přihrádce; následující nula reprezentuje přihrádku mezi druhou a třetí přihrádkou; pak nenásleduje žádná jednička (tj. třetí přihrádka neobsahuje žádnou kuličku), ale hned další nula reprezentující přepážku mezi třetí a čtvrtou přihrádkou; následují tři jedničky reprezentující 3 kuličky ve čtvrté příhrádce. Protože máme $n-1$ přepážek a r kuliček, je každé umístění reprezentováno řetězcem délky $n+r-1$, ve kterém je $n-1$ nul a r jedniček. Každý takový řetězec je určen tím, na kterých jeho $n-1$ pozicích z 1.,...,$(n+r-1)$-té pozice jsou nuly (na ostatních pozicích jsou totiž jedničky). Výběr $n-1$ pozic pro nuly z celkového počtu $n+r-1$ pozic je kombinace $n-1$ z $n+r-1$, a těch je podle věty 5.24 $\binom{n+r-1}{n-1}$. □

Příklad 5.33. Vraťme se k zákuskům (viz výše). Výběr 6 zákusků ze 4 druhů zákusků je kombinace 6 z 4 s opakováním. Těch je podle Věty 5.32 $\overline{C}(n,r) = \binom{n+r-1}{n-1} = \binom{4+6-1}{4-1} = \binom{9}{3} = \frac{9!}{6!3!} = 84$.

Poznámka 5.34. Zastavme se u pojmů permutace s opakováním, variace s opakováním a kombinace s opakováním. Ve všech případech máme vlastně objekty rozděleny do několika typů. Zatímco však u permutací s opakováním je objektů každého typu předepsaný počet a tyto počty mohou být pro různé typy různé, u variací i kombinací s opakováním je objektů každého typu neomezeně mnoho.

5.3.4 Další výběry

Permutace, variace a kombinace jsou základní typy výběrů. Ukázali jsme si základní typy úvah, které vedou ke stanovení jejich počtu. Prakticky se však můžeme setkat s příklady složitějšími, ve kterých úvahy o permutacích, variacích a kombinacích můžeme využít.

Příklad 5.35. Ligu hraje 14 týmů. Výsledek ligy je dán tím, které týmy obsadí 1., 2. a 3. místo a které 2 týmy sestoupí do nižší soutěže. Kolik je možných výsledků ligy?

Výsledek ligy je dán výběrem týmů na 1.-3. místě a výběrem týmů, které sestupují. Týmy na 1.-3. místě jsou tři a vybíráme je z 14 týmů, přitom na pořadí výběru záleží. Jde tedy o variace 3 z 14 a je jich $V(14,3)$. Po jejich výběru vybereme ze zbývajících 11 týmů dva, které sestoupí. Zde na pořadí nezáleží. Jde tedy o kombinace 2 z 11 a je jich $\binom{11}{2}$. Podle pravidla součinu je celkově $V(14,3) \cdot \binom{11}{2}$ možných výsledků ligy.

Můžeme ale také postupovat obráceně, tj. nejdřív vybrat ze 14 dva sestupující týmy a pak ze zbylých 12 vybrat 3 medailisty. Tak dostaneme $\binom{14}{2} \cdot V(12,3)$ možností. Výsledek je ale stejný jako u první úvahy, protože

$$\begin{aligned}\binom{14}{2} \cdot V(12,3) &= \frac{14 \cdot 13}{2} \cdot 12 \cdot 11 \cdot 10 = \frac{14 \cdot 13 \cdot 12 \cdot 11 \cdot 10}{2} = \\ &= 14 \cdot 13 \cdot 12 \cdot \frac{11 \cdot 10}{2} = V(14,3) \cdot \binom{11}{2}.\end{aligned}$$

Příklad 5.36. Kolika různými způsoby lze kolem kulatého stolu se 6 židlemi posadit 6 osob? Přitom dvě posazení, která se liší jen pootočením, považujeme za shodná.

Označmě osoby A, B, C, D, E, F. Kdyby i dvě posazení lišící se pootočením, byla považována za různá, pak by počet všech posazení byl stejný jako počet všech permutací 6 objektů, tj. $P(6) = 6!$. Kruhové uspořádání kolem stolu by totiž nehrálo roli. Kolem stolu je 6 míst, můžeme jim říkat 1., 2., ..., 6. místo. Otázka by pak byla, kolika způsoby můžeme umístit 6 osob na 6 míst, tj. vlastně kolika způsoby lze uspořádat 6 osob. Odpověď je pak zjevně $P(6)$. Považujeme-li však posazení za shodná, právě když lze z jednoho do druhého přejít pootočením, bude celkový počet posazení menší. Dojdeme k němu např. následovně. Kruhové posazení kolem stolu "roztrhneme" a zapíšeme lineárně. Např. ABCFDE je zápis, kdy A sedí na 1. židli, ..., E sedí na 6. židli. Postupným otáčením tohoto posazení o 0 až 5 židlí dostaneme celkem 6 jeho zápisů: ABCFDE, BCFDEA, CFDEAB, FDEABC, DEABCF, EABCFD. Celkový počet zápisů je tedy 6 krát větší než počet posazení. Protože zápisů je $P(6)$, je hledaný počet posazení $\frac{P(6)}{6} = \frac{6!}{6} = 5!$.

Příklad 5.37. Kolik existuje posloupností n nul a k jedniček, ve kterých žádné dvě jedničky nejsou vedle sebe?

Představme si posloupnost n nul. Na začátku, mezi nulami a na konci této posloupnosti je celkem $n+1$ míst (např. pro posloupnost 000 jsou to místa _0_0_0_). Libovolnou

posloupnost n nul a k jedniček, která splňuje požadované podmínky, tak, že na vzniklých $n+1$ míst umístíme k jedniček. Takových možností je právě tolik, kolika způsoby můžeme z $n+1$ míst (mezi nulami) vybrat k míst (na každé z nich dáme jedničku), tedy právě $\binom{n+1}{k}$. Počet hledaných posloupností je tedy $\binom{n+1}{k}$.

Shrnutí

Kombinatorika se zabývá zjišťováním počtu možností, které mnohou nastat za předem daných podmínek. Základní kombinatorická pravidla jsou pravidlo součtu a pravidlo součinu. Pomocí nich se dají určit např. počty možností různých typů výběrů. Mezi základní typy výběrů patří permutace, variace a kombinace. Permutace n prvků je jejich libovolné uspořádání. Variace k prvků z n prvků je libovolný výběr k prvků z n prvků, ve kterém na pořadí vybíraných prvků záleží. Kombinace k prvků z n prvků je libovolný výběr k prvků z n prvků, ve kterém na pořadí vybíraných prvků nezáleží. Variace a kombinace s opakováním jsou podobné výběry, ve kterých se vybírané prvky mohou opakovat.

Pojmy k zapamatování

- pravidla součtu a součinu,
- permutace a permutace s opakováním,
- variace a variace s opakováním,
- kombinace a kombinace s opakováním.

Kontrolní otázky

1. *Co říká pravidlo součtu? Co říká pravidlo součinu?*
2. *Čím se liší permutace a variace? Čím se liší variace a kombinace?*
3. *Čím se liší permutace a permutace s opakováním? Čím se liší variace a variace s opakováním? Čím se liší kombinace a kombinace s opakováním?*
4. *Čím se liší aspekt opakování u permutací s opakováním, variací s opakováním a kombinací s opakováním?*

Cvičení

1. Kolik existuje v soustavě o základu n nezáporných čísel, které mají právě k číslic?
2. Kolik různých slov lze získat z akronymu WYSIWYG?
3. Dokažte matematickou indukcí, že $\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$ pro všechna $n \in N$ a $k = 0, 1, \ldots, n$.
4. Definujme indukcí $P^1(X) = P(X)$ a pro $n > 1$ $P^n(X) = P(P^{n-1}(X))$. Je-li množina X konečná, kolik prvků má $P^n(X)$?
5. Kolik má n-prvková množina m-prvkových podmnožin ($m < n$)?
6. Kolik existuje funkcí m prvkové do n prvkové množiny?
7. Kolik existuje injektivních funkcí z m prvkové do n prvkové množiny?
8. Kolik existuje n-árních operací na m-prvkové množině? Kolik z nich je injektivních? Kolik jich je surjektivních?
9. Krotitel má do arény přivést za sebou jdoucích 5 lvů a 4 tygry. Přitom žádní dva tygři nesmí jít bezprostředně za sebou (musí mezi nimi být lev). Kolika způsoby to lze provést? Na pořadí tygrů i lvů záleží.

10. Rozeberte předchozí cvičení pro případ n lvů a k tygrů.
11. Na poličce je 12 knih. Kolika způsoby lze vybrat 5 z nich tak, aby žádné dvě z vybraných nestály vedle sebe? Jak je to při výběru k knih z n?

Úkoly k textu

1. U Příkladů 5.1, 5.2, 5.3 zdůvodněte použité kombinatorické úvahy.
2. Vraťme se k Příkladu 5.3. Jaký je největší počet k kódových posloupností binárního kódu délky n, který umožňuje opravu až t-násobných chyb? t-násobnou chybou vznikne z daného slova slovo, které se od daného liší právě v t pozicích. Příklad 5.3 tedy dává odpověď pro $t = 1$. [Odpověď: $\frac{2^n}{1+n+\binom{n}{2}+\cdots+\binom{n}{r}}$.]
3. Vraťme se k Příkladu 5.8. Navrhněte různé tvary registračních značek a pro každý tvar spočítejte odpovídající počet značek, které lze přidělit. Jaké pravidlo pro návrh tvaru registračních značek plyne?
4. Zdůvodněte podrobně Větu 5.13.
5. Zdůvodněte podrobně Větu 5.16.
6. Zdůvodněte podrobně Větu 5.24.

   Promyslete si a zdůvodněte důkaz Věty 5.32.

Řešení

1. $(n-1)\cdot n^{k-1}$.
   Návod: Jako první číslici lze použít $n-1$ číslic (nelze 0), na každou z dalších $k-1$ pozic pak n číslic. Podle principu součinu to lze celkem $(n-1)\cdot n^{k-1}$ způsoby.
2. 1260.
   Návod: Je to $P(2,2,1,1,1)$.
3. Dokažte matematickou indukcí, že $\binom{n}{k}+\binom{n}{k-1}=\binom{n+1}{k}$ pro všechna $n \in N$ a $k = 0, 1, \ldots, n$.
4. Označme $k = |X|$. Pak $|P^1(X)| = 2^k$, $|P^2(X)| = 2^{2^k}$, atd. Obecně je $|P^n(X)| = 2^{2^{\cdot^{\cdot^{\cdot^{k}}}}}$ (dvojka je tam k krát).
5. $\binom{n}{m}$, je to právě počet kombinací m z n.
6. $\overline{V}(n,m) = n^m$.
   Návod: Mějme $X = \{x_1, \ldots, x_m\}$, $Y = \{y_1, \ldots, y_n\}$. Libovolná funkce f je dána uspořádanou m-ticí $\langle f(x_1), \ldots, f(x_m)\rangle$ hodnot $f(x_i) \in Y$. Výběr každé takové m-tice je variace m z n s opakováním. Těch je $\overline{V}(n,m) = n^m$.
7. Pro $m \leq n$ existuje $V(n,m)$ injektivních funkcí, pro $m > n$ žádná.
   Návod: Viz předchozí cvičení, jde o variace bez opakování.
8. m^{m^n}.
   Návod: Pro $|X| = m$ je to počet zobrazení množiny X^n do množiny X. Protože $|X^n| = m^n$, je jich n^{m^n} (viz předchozí cvičení).
9. Existuje 43200 způsobů.
   Návod: Lvy lze rozmístit $P(5) = 5! = 120$ způsoby. Zbývá 6 míst pro umístění tygrů (na začátku, mezi lvy a na konci). Do nich lze tygry umístit $V(6,4) = 360$ způsoby. Podle pravidla součinu existuje celkem $120 \cdot 360 = 43200$ způsobů.

10. Pro $k \leq n+1$ existuje $P(n) \cdot V(n+1,k)$ způsobů. Pro $k > n+1$ takový způsob neexistuje.
11. Existuje 56 možností. V obecném případě existuje $\binom{n+k-1}{k}$ možností (pokud $2k-1 \leq n$, jinak žádná možnost neexistuje).
    Návod: Každý každý takový výběr k knih z n knih můžeme reprezentovat posloupností k jedniček (na pozicích vybraných knih) a $n-k$ nul (na pozicích nevybraných knih), ve které se nevyskytují sousedící jedničky (vybrané knihy nestojí vedle sebe). Těch je podle Příkladu 5.37 $\binom{n-k+1}{k}$.

Studijní cíle: Po prostudování kapitoly 5.4 by student měl být znát princip inkluze a exkluze a umět ho použít.

Klíčová slova: princip inkluze a exkluze

5.4 Princip inkluze a exkluze

V nabídce volitelných předmětů je němčina a angličtina. Němčinu si zvolilo 15 studentů, angličtinu 30 studentů. 5 studentů si zvolilo němčinu i angličtinu. Kolik studentů si jako volitelný předmět vybralo cizí jazyk (tj. němčinu nebo angličtinu)? Označme N a A po řadě množiny studentů, kteří si zapsali němčinu a angličtinu. Sečteme-li $|N|$ (počet těch, kteří si zapsali němčinu) a $|A|$ (počet těch, kteří si zapsali angličtinu), počítáme dvakrát ty, kteří si zapsali němčinu i angličtinu (těch je $|N \cap A|$). Ty tedy musíme od $|N| + |A|$ odečíst. Počet $|N \cup A|$ těch, kteří si zapsali němčinu nebo angličtinu je tedy

$$|N \cup A| = |N| + |A| - |N \cap A| = 15 + 30 - 5 = 40.$$

Jiný příklad: Na jisté univerzitě je 56 učitelů členy americké informatické společnosti ACM (Association for Computing Machinery). Členové ACM si mohou přikoupit členství v některé z tzv. special interest group (SIG, SIG jsou součásti ACM). Ze zmíněných 56 učitelů jich je 20 členy SIGMOD (Special Interest Group on Management of Data), označme jejich množinu A_1; 15 členy SIGIR (Special Interest Group on Information Retrieval), označme jejich množinu A_2; 20 členy SIGKDD (Special Interest Group on Knowledge Discovery in Data), označme jejich množinu A_3. Dále je známo, že 10 jich je členy SIGMOD i SIGIR, 8 jich je členy SIGMOD i SIGKDD, 7 jich je členy SIGIR i SIGKDD, 4 jsou členy SIGMOD, SIGIR i SIGKDD. Kolik z 56 členů ACM je členem některé z SIGMOD, SIGIR, SIGKDD? Ptáme se vlastně, kolik prvků má množina $A_1 \cup A_2 \cup A_3$, přitom známe $|A_1|$, $|A_2|$, $|A_3|$, $|A_1 \cap A_2|$, $|A_1 \cap A_3|$, $|A_2 \cap A_3|$ a $|A_1 \cap A_2 \cap A_3|$. Pokud bychom pouze sečetli $|A_1| + |A_2| + |A_3|$, pak jsou dvakrát započítáni ti z $A_1 \cap A_2$, z $A_1 \cap A_3$ a z $A_2 \cap A_3$, a dokonce třikrát jsou započítáni ti z $A_1 \cap A_2 \cap A_3$. To svádí k tomu říci, že $|A_1 \cup A_2 \cup A_3|$ se rovná

$$|A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| - 2|A_1 \cap A_2 \cap A_3|.$$

To je ale špatně. Odečteme-li totiž od $|A_1| + |A_2| + |A_3|$ počty $|A_1 \cap A_2|$, $|A_1 \cap A_3|$ i $|A_2 \cap A_3|$, odečítáme v každém z $|A_1 \cap A_2|$, $|A_1 \cap A_3|$ a $|A_2 \cap A_3|$ i počet těch, kteří jsou v $A_1 \cap A_2 \cap A_3$ (nakreslete si obrázek). Tedy od $|A_1| + |A_2| + |A_3|$ jsme odečetli $3|A_1 \cap A_2 \cap A_3|$. K tomu jsme pak ještě odečetli $2|A_1 \cap A_2 \cap A_3|$. Celkově jsme tedy od $|A_1| + |A_2| + |A_3|$ odečetli $|A_1 \cap A_2 \cap A_3|$ pětkrát a měli jsme to odečíst jen dvakrát. Správný výsledek tedy je

$$\begin{aligned} |A_1 \cup A_2 \cup A_3| &= |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + \\ &\quad + |A_1 \cap A_2 \cap A_3| = 20 + 15 + 20 - 10 - 8 - 7 + 4 = 24. \end{aligned}$$

Úvahy ukázané na výše uvedených příkladech jsou předmětem tzv. principu inkluze a exkluze (tj. zapojování a vylučování).

Věta 5.38 (princip inkluze a exkluze)**.** *Pro množiny $A_1, \ldots, A_n$ platí*

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} |\bigcap_{i \in I} A_i|.$$

Zastavme se nejdřív nad tím, co princip inkluze a exkluze říká. Na levé straně rovnosti je počet prvků, které patří do sjednocení $A_1 \cup \cdots \cup A_n$, tj. alespoň do jedné z $A_1, \ldots, A_n$. Na pravé straně je součet výrazů $(-1)^{|I|+1}| \cap_{i \in I} A_i|$, kde I probíhá přes všechny neprázdné podmnožiny množiny $\{1, \ldots, n\}$. $| \cap_{i \in I} A_i|$ je počet prvků průniku množin, jejichž index patří do I, tj. např. pro $I = \{2, 3, 5\}$ je to $|A_2 \cap A_3 \cap A_5|$. Výraz $(-1)^{|I|+1}$ je roven 1, pokud I obsahuje lichý počet prvků, a je roven -1, pokud I obsahuje sudý počet prvků. Tedy: v součtu na pravé straně jsou počty prvků všech možných průniků (jednočlenných, dvoučlenných,..., až po n-členný) utvořené z $A_1, \ldots, A_n$, přitom počet prvků daného průniku je se znaménkem + pro průniky lichého počtu množin a se znaménkem − pro průniky lichého počtu množin. Zkontrolujte, že vzorec pro $n = 2$ i $n = 3$ dává právě dva vzorce, ke kterým jsme došli i příkladů s volitelnými předměty a s členstvím v SIG ACM. Přejděme nyní k důkazu Věty 5.38.

Důkaz. Vezměme libovolný prvek u z $A_1 \cup \cdots \cup A_n$ a porovnejme, jakým číslem p "přispívá" na levé a na pravé straně dokazované rovnosti. Na levé straně přispívá zřejmě jedničkou (tedy $p = 1$).

Pro pravou stranu je to složitější. Prvek u patří právě do, řekněme, k množin z množin $A_1, \ldots, A_n$. Můžeme předpokládat, že to jsou množiny $A_1, \ldots, A_k$ (kdyby ne, množiny přeznačíme). Pak u patří do nějakého průniku $\bigcap_{i \in I} A_i$ z pravé strany, právě když je to průnik nějakých q množin vybraných z $A_1, \ldots, A_k$ pro nějaké $q \leq k$. Je-li to průnik lichého počtu množin, u do odpovídajícího výrazu $(-1)^{|I|+1}|\bigcap_{i \in I} A_i|$ přispívá číslem 1; je-li to průnik sudého počtu množin, u do výrazu $(-1)^{|I|+1}|\bigcap_{\in I} A_i|$ přispívá číslem -1. Z $A_1, \ldots, A_k$ lze vytvářet jednoprvkové, ..., k-prvkové průniky. Počet q-prvkových průniků je přitom $\binom{k}{q}$. Vidíme tedy, že u přispívá celkem na pravou stranu číslem

$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \cdots + (-1)^{k+1}\binom{k}{k}.$$

Jaká je hodnota tohoto součtu? Vezměme binomickou větu a dosaďme do (21) $x = -1$. Dostaneme

$$\begin{aligned} 0 = 0^k = (1-1)^k = (1+x)^k = \sum_{i=0}^{k} \binom{k}{i} x^i = 1 + \sum_{i=1}^{k} (-1)^i \binom{k}{i} = \\ = \quad 1 - \left(\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \cdots + (-1)^{k+1}\binom{k}{k} \right). \end{aligned}$$

Odtud tedy vidíme, že hledaná hodnota součtu je 1. Prvek u tedy i na pravou stranu přispívá jedničkou ($p = 1$). Protože x byl libovolný, výrazy na levé a pravé straně dokazované rovnosti mají stejné hodnoty. □

Příklad 5.39. Kolik je přirozených čísel mezi 1 a 100 (včetně 1 i 100), která nejsou dělitelná ani dvěma, ani třemi nebo pěti? Princip inkluze a exkluze můžeme použít následovně. Označme

$$A_1 = \{n \in \mathbb{N} \mid 1 \leq n \leq 100,\ n \text{ je dělitelné } 2\}, \tag{23}$$

$$A_2 = \{n \in \mathbb{N} \mid 1 \leq n \leq 100,\ n \text{ je dělitelné } 3\}, \tag{24}$$

$$A_3 = \{n \in \mathbb{N} \mid 1 \leq n \leq 100,\ n \text{ je dělitelné } 5\}. \tag{25}$$

Pak přirozená čísla mezi 1 a 100 (včetně 1 i 100), která nejsou dělitelná ani dvěma, ani třemi nebo pěti, jsou právě prvky množiny $A = \overline{A_1} \cap \overline{A_2} \cap \overline{A_3}$. Protože

$$\overline{A_1} \cap \overline{A_2} \cap \overline{A_3} = \overline{A_1 \cup A_2 \cup A_3},$$

je $|A| = |\overline{A_1 \cup A_2 \cup A_3}| = 100 - |A_1 \cup A_2 \cup A_3|$. Podle principu inkluze a exkluze je

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

Zbývá tedy určit $|A_1|$, $|A_2|$, $|A_3|$, $|A_1 \cap A_2|$, $|A_1 \cap A_3|$, $|A_2 \cap A_3|$, $|A_1 \cap A_2 \cap A_3|$, což je snadné. Ukažme to na příkladu množiny $A_1 \cap A_2$. $A_1 \cap A_2$ je množina přirozených čísel mezi 1 a 100, která jsou dělitelná dvěma i třemi. To jsou ale právě ta čísla, která jsou dělitelná 6 (číslo je dělitelné 6, právě když je dělitelné 2 i 3). Těch je $\lfloor \frac{100}{6} \rfloor = 16$ (dolní celá část čísla $\frac{100}{6}$), tj. $|A_1 \cap A_2| = 16$. Podobně dostaneme $|A_1| = 50$, $|A_2| = 33$, $|A_3| = 20$, $|A_1 \cap A_3| = 10$, $|A_2 \cap A_3| = 6$, $|A_1 \cap A_2 \cap A_3| = 3$. Dosazením pak dostaneme $|A| = 100 - |A_1 \cup A_2 \cup A_3| = 26$.

Shrnutí

Princip inkluze a exkluze je často používaným kombinatorickým principem. Udává počet prvků sjednocení několika množin pomocí počtu prvků průniků jednotlivých množin.
Počítání pravděpodobností jednoduchých jevů patří mezi základní aplikace kombinatorického počítání. Pravděpodobnost jevu je dána podílem počtu možností příznivých danému jevu ku počtu všech možností. Kombinatorické úvahy se uplatní při určování počtu množností.

Pojmy k zapamatování

- princip inkluze a exkluze.

Kontrolní otázky

1. *Co říká princip inkluze a exkluze?*
2. *Jak se zjednoduší vzorec z principu inkluze a exkluze, jsou-li množiny $A_1, \ldots, A_n$ po dvou disjunktní?*

Cvičení

1. Určete počet přirozených čísel mezi 1 a 2000 (včetně 1 i 2000), která nejsou dělitelná ani 2, ani 3, ani 5.
2. Určete počet přirozených čísel mezi 1 a 2000 (včetně 1 i 2000), která nejsou dělitelná ani 2, ani 3, ani 5, ani 7.
3. Určete počet přirozených čísel mezi 1 a 2000 (včetně 1 i 2000), která nejsou dělitelná ani 2, ani 3, ani 5, ale jsou dělitelná 7.
4. Kolika způsoby můžeme rozmístit 15 různých knih do 5 přihrádek tak, aby v každé přihrádce byla aspoň jedna kniha, ale nejvýše 4 knihy?

Úkoly k textu

1. U Příkladů 5.1, 5.2, 5.3 zdůvodněte použité kombinatorické úvahy.
2. Vraťme se k Příkladu 5.3. Jaký je největší počet k kódových posloupností binárního kódu délky n, který umožňuje opravu až t-násobných chyb? t-násobnou chybou vznikne z daného slova slovo, které se od daného liší právě v t pozicích. Příklad 5.3 tedy dává odpověď pro $t = 1$. [Odpověď: $\frac{2^n}{1+n+\binom{n}{2}+\cdots+\binom{n}{r}}$.]
3. Vyřešte podrobně Příklad 5.39.
4. Na základě Příkladu 5.39 dokažte následující tvrzení: Jsou-li podmnožiny $A_1, \ldots, A_n$ nějaké k-prvkové množiny X, pak počet prvků množiny X, které nepatří ani jedné z množin $A_1, \ldots, A_n$ je $k - \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,n\}} (-1)^{|I|+1} |\bigcap_{i \in I} A_i|$.

Řešení

1. 534.
2. 458.
3. 76.
4. $15!\dot{(}\binom{14}{10} - \binom{5}{1}\binom{10}{6} + \binom{5}{2}\binom{6}{2})$.

6 Pravděpodobnost a statistika

Studijní cíle: Po prostudování kapitoly by student měl znát základní pojmy teorie pravděpodobnosti, zejména pro případ konečných a diskrétních pravděpodobnostních prostorů. Student by měl umět v základních úlohách o určování pravděpodobností samostatně provést správnou kombinatorickou úvahu.

Klíčová slova: pravděpodobnost, pravděpodobnostní prostor, náhodná veličina, střední hodnota, rozptyl, směrodatná odchylka, kvantil

6.1 Co a k čemu je pravděpodobnost a statistika

Pravděpodobnost jako matematická disciplína se zabývá se určováním pravděpodobností náhodných jevů a spolu se statistikou je jednou z nejužitečnějších oblastí matematiky. Příklady problémů, které pravděpodobnost řeší, jsou:

- Jaká je pravděpodobnost, že při hodu kostkou padne číslo 4?
- Jaká je pravděpodobnost, že při hodu kostkou padne sudé číslo?
- Jaká je pravděpodobnost, že náhodně vybraný člověk má vysoký tlak za předpokladu, že jeho index tělesné hmotnosti je > 30 kg/m^2?
- Jaká je očekávaná hodnota výšky muže v České republice?

Teorie pravděpodobnosti vznikla spolu s kombinatorikou při analýze hazardních her. První úvahy o pravděpodobnosti však lze nalézt už v arabské matematice při studiu kryptografie (8–13. stol. n. l.). O rozvoj teorie pravděpodobnosti se v 16. a 17. století n. l. zasloužili zejména Gerolamo Cardano, Pierre de Fermat, Blaise Pascal a Christian Huygens, který napsal první knihu o pravděpodobnosti (*De Ratiociniis in Ludo Aleae*, 1657). Moderními úvahami o pravděpodobnosti o něco později přispěl Pierre de Laplace, který je shrnul v knize *Théorie analytique des probabilités* (1812). Současný matematický přístup k pravděpodobnosti vytvořil v knize *Grundbegriffe der Wahrscheinlichkeitsrechnung* (1933).

Pravděpodobnost a statistika nás učí, jak kvantitativně vyhodnocovat data. Každý přírodovědec nebo technik by měl proto základy této disciplíny ovládat. Kromě toho, že nám dává výše zmíněné dovednosti, má řadu použití v informatice – na pravděpodobnosti je např. založen pojem složitost algoritmu v průměrném případě, je klíčová pro tzv. pravděpodobnostní algoritmy, stojí na ní různé metody strojového učení apod.

6.2 Pravděpodobnost

Při studiu pravděpodobnosti se musíme zamyslet nad pojmy, které přirozeně při úvanách o pravděpodobnosti používáme. Mezi otázky, které musíme z tohoto pohledu zodpovědět, např. patří:

- Jaký je obecný rámec úvah o pravděpodobnosti?
- Pravděpodobnost čeho vlastně určujeme?
- Co je to vlastně pravděpodobnost?

6.2.1 Intuitivní přístup a klasická definice pravděpodobnosti

Klasická definice vychází z intuitivního pojetí pravděpodobnosti.

Začneme intuitivním pohledem na pravděpodobnost. Ten se odráží v tzv. klasické (někdy Laplaceově) „definici“ pravděpodobnosti, která říká:[18]

$$\text{pravděpodobnost} = \frac{\text{počet příznivých výsledků}}{\text{počet všech výsledků}}$$

Přitom musí platit, že všechny j výsledky jsou stejně pravděpodobné. Klasickou definici osvětlíme jednoduchým příkladem:

Příklad 6.1. Jaká je pravděpodobnost, že při hodu kostkou padne sudé číslo?

Možné výsledky jsou $1, 2, \ldots, 6$; příznivé výsledky jsou $2, 4, 6$. Tedy

$$\text{pravděpodobnost } = \frac{3}{6} = 0.5.$$

Počítání pravděpodobností je jednou z užitečných aplikací kombinatoriky.

Počítáme tedy pravděpodobnosti tzv. jevů (říkáme také událostí; např. že padne sudé číslo) a uvažujeme všechny možné výsledky a výsledky, které jsou příznivé danému jevu. Při určování počtu všech možných výsledků a příznivých výsledků často používáme různé kombinatorické úvahy (to uvidíme v dalších příkladech). Tak lze vyřešit mnoho jednoduchých příkladů.

Průvodce studiem

Předpoklad klasické definice pravděpodobnosti, že všechny možné výsledky jsou stejně pravděpodobné, je zásadní. Představme si, že házíme kostkou, u které mají jednička a trojka pravděpodobnost $1/4$, a ne $1/6$ (např. proto, že kostka je z nehomogenního materiálu), zatímco ostatní výsledky mají pravděpodobnost $1/8$. Snadno dojdeme k závěru, že pak je pravděpodobnost, že padne sudé číslo rovna $3/8 = 0.375$, zatímco podle klasické definice pravděpodobnosti je to 0.5.

Když podrobněji rozebereme, o co v klasické definici pravděpodobnosti P daného jevu jde, dojdeme k následujícímu pohledu:

$$\text{pravděpodobnost} = \frac{|A|}{|E|},$$

kde

- A je množina všech výsledků příznivých danému jevu,
- E je množina všech možných výsledků.

Pro výpočet klasické pravděpodobnosti je tedy třeba:

1. určit množinu E všech elementárních jevů (výsledků) a ověřit, že jsou všechny stejně pravděpodobné,
2. určit množinu $A \subseteq E$ výsledků příznivých danému jevu,
3. určit počet prvků množiny E, tj. určit $|E|$,
4. určit počet prvků množiny A, tj. určit $|A|$.

[18] Uvozovky píšeme proto, že přísně vzato nejedná o přesnou definici, spíše o představu.

V krocích 1. a 2. si koncepčně ujasníme situaci (1. a 2. odpovídá nalezení správného pohledu na věc), v krocích 3. a 4. obvykle provedeme určité kombinatorické úvahy.

Uveďme několik dalších jednoduchých příkladů.

Příklad 6.2. Házíme modrou a červenou kostkou. Jaká je pravděpodobnost, že na modré kostce padne sudé a na červené liché číslo?

Na sitaci se můžeme dívat takto: Výsledek, tj. elementární jev, je dán dvojicí čísel $\langle m, c\rangle$, kde $m, c \in \{1, 2, 3, 4, 5, 6\}$ a m a c jsou po řadě výsledek na modré a červené kostce. Tedy máme

$$E = \{\langle m, c\rangle \mid m, c \in \{1, 2, 3, 4, 5, 6\}\}.$$

Elementární jev $\langle m, c\rangle$ je příznivý události „na modré kostce padne sudé a na červené liché číslo“, právě když $m \in \{2, 4, 6\}$ a $c \in \{1, 3, 5\}$. Tedy

$$A = \{\langle m, c\rangle \mid m \in \{2, 4, 6\},\ c \in \{1, 3, 5\}\}.$$

Vidíme, že $|E| = 6 \cdot 6 = 36$ (přímo podle pravidla součinu) a že $|A| = 3 \cdot 3 = 9$ (podle pravidla součinu). Tedy hledaná pravděpodobnost $P(A)$ je $P(A) = \frac{|A|}{|E|} = \frac{9}{36} = 0.25$.

Příklad 6.3. Házíme dvěma kostkami, které jsou nerozlišitelné. Jaká je pravděpodobnost, že aspoň na jedné z nich padne dvojka?

Vezměme opět $E = \{\langle m, c\rangle \mid m, c \in \{1, 2, 3, 4, 5, 6\}\}$. Zajímá nás teď jev $A = \{\langle m, c\rangle \in E \mid m = 2 \text{ nebo } c = 2\}$ a jeho počet prvků. K němu můžeme dojít tak: Pro jevy $A_1 = \{\langle m, c\rangle \in E \mid m = 2\}$ (na modré padne dvojka) a $A_2 = \{\langle m, c\rangle \in E \mid c = 2\}$ (na červené padne dvojka) zřejmě platí $A = A_1 \cup A_2$. Protože $A_1 \cap A_2 = \{\langle 2, 2\rangle\}$, je podle principu inkluze a exkluze

$$|A| = |A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| = 6 + 6 - 1 = 11.$$

Pravděpodobnost, že aspoň na jedné z kostek padne dvojka je tedy $P(A) = \frac{|A|}{|E|} = \frac{11}{36}$.

Příklad Z balíčku 32 mariášových karet vybereme 4 karty.

- Jaká je pravděpodobnost, že vybereme 4 esa?
- Jaká je pravděpodobnost, že vybereme 2 esa?

Výsledek $= \{k_1, k_2, k_3, k_4\}$. Všech možností je $\binom{32}{4} = \frac{32 \cdot 31 \cdot 30 \cdot 29 \cdot}{4!} = 35960$.

- 4 esa:
  - Příznivý výsledek = všechny karty k_i jsou esa.
  - Existuje 1 příznivý výsledek. Tedy

$$\text{pravděpodobnost } = \frac{1}{\binom{32}{4}} = \frac{1}{35960} \approx 0.0000278$$

– 2 esa

  – Příznivý výsledek = právě dvě karty k_i jsou esa.
  – Jsou celkem 4 esa. Z nich lze vybrat 2 esa $\binom{4}{2} = \frac{4 \cdot 3}{2!} = 6$ způsoby.
    Zbylé 2 karty (ne esa) vybrat z 28 ostatních, což lze $\binom{28}{2} = \frac{28 \cdot 27}{2!} = 378$ způsoby.
    Dle pravidla součinu lze příznivý výsledek $\{k_1, k_2, k_3, k_4\}$ vybrat

$$\binom{4}{2} \cdot \binom{28}{2} = 6 \cdot 378 = 2268 \text{ způsoby}$$

    Tedy

$$\text{pravděpodobnost } = \frac{2268}{35960} \approx 0.063$$

6.2.2 Kolmogorovova definice pravděpodobnosti

Nyní popíšeme rigoróznější a obecnější pohled na pojem pravděpodobnost. Pochází od sovětského matematika Andreje Kolmogorova a tvoří základ současného pojetí pravděpodobnosti.

Kolmogorovovo pojetí pravděpodobnosti lze popsat následovně:

– je dán „náhodný pokus“, např.:

  – hod kostkou,
  – výběr člověka z populace ČR,
  – spuštění algoritmu na vybraném vstupu, apod.;

– k pokusu patří množina Ω možných výsledků (tzv. elementárních jevů), např.:

  – hod kostkou: $\Omega = \{1, 2, 3, 4, 5, 6\}$,
  – výběr člověka: $\Omega = \{$ člověk 1, …, člověk $10^7\}$,
  – spuštění algoritmu: $\Omega = \{$vstup 1, …, vstup $n\}$;

– určujeme pravděpodobnost tzv. jevů, např.:

  – hod kostkou: jev „sudé číslo“ $= \{2, 4, 6\}$,
  – výběr člověka: je „žena“ $= \{$ člověk c v ČR $\mid c$ je žena$\}$,
  – spuštění algoritmu α: jev „skončí v čase t“ $=\{$ vstup $\mid t_\alpha(\text{vstup}) \leq t \}$,
  – obecně: jev A = množina výsledků, tedy $A \subseteq \Omega$;

– $\mathcal{A}$ = množina uvažovaných (tzv. měřitelných) jevů, např. $\mathcal{A} = 2^\Omega$

– pravděpodobnost (přesněji pravděpodobnostní míra) je funkce $P : \mathcal{B} \to [0, 1]$ splňující jisté vlastnosti.

Základním pojmem v Kolmogorovově přístupu je pojem pravděpodobnostní prostor. Ten tvoří základ pro všechny úvahy o pravděpodobnosti: každé použití pravděpodobnosti a statistiky vyžaduje nejdříve zvolit vhodný pravděpodobnostní prostor. Jeho volbu lze nahlížet jako „napasování“ teoretických pojmů na realitu konkrétní úlohy, kterou máme řešit. Tato volba není jednoznačná (je příkladem toho, že v matematice se lze správného výsledku dobrat různými postupy). Zatím bez uvedení úplných podmínek uveďme, že pravděpodobnostní prostor je trojice $\langle \Omega, \mathcal{A}, P \rangle$, kde

– Ω je neprázdná množina tzv. elementárních jevů (výsledky pokusu)

– $\mathcal{A} \subseteq 2^\Omega$ je množina tzv. jevů

– $P : \mathcal{A} \to [0,1]$ je pravděpodobnostní míra
pro jev $A \in \mathcal{A}$ je $P(A) \in [0,1]$ pravděpodobnost, že nastane jev A

Příklad 6.4. Uvažujme hod kostkou. To je z hlediska Kolmogorovovy definice náhodný pokus. Nejpřirozenější způsob, jak definovat pravděpodobnostní prostor popisující hod kostkou je tento:

– $\Omega = \{\omega_1, \ldots, \omega_6\}$, kde $\omega_i =$ „padne číslo i“

– $\mathcal{A} = 2^\Omega$, všech 64 možných jevů A, např.

  – $A = \{\omega_1, \omega_3, \omega_5\}$ je jev „padne liché číslo“
  – $A = \{\omega_3\}$ je jev „padne 3“

– $P : \mathcal{A} \to [0,1]$ je dána následovně:

  – pro ω_i je $P(\{\omega_i\}) = 1/6$
  – pro ostatní $A \subseteq \Omega$ je $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$, např.

$$\text{pravděpodobnost}(\text{„padne liché číslo“}) = P(\{\omega_1, \omega_3, \omega_5\}) = P(\{\omega_1\}) + P(\{\omega_3\}) + P(\{\omega_5\}) = 1/6 + 1/6 + 1/6 = 1/2$$

$$\text{pravděpodobnost}(\text{„padne 2 nebo 5“}) = P(\{\omega_2, \omega_5\}) = P(\{\omega_2\}) + P(\{\omega_5\}) = 1/6 + 1/6 = 1/3$$

Kolmogorovova definice pravděpodobnosti

Definice 6.5. *Pravděpodobnostní prostor* je uspořádaná trojice $\langle \Omega, \mathcal{A}, P \rangle$, kde

– $\langle \Omega, \mathcal{A} \rangle$ je *σ-algebra* na Ω, tj. $\Omega \neq \emptyset$, $\emptyset \neq \mathcal{A} \subseteq 2^\Omega$ a platí:

  (a) je-li $A \in \mathcal{A}$, pak $\overline{A} \in \mathcal{A}$;
  (b) jsou-li $A_1, A_2, \cdots \in \mathcal{A}$, pak $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

– P je *pravděpodobnostní míra*, tj. P je zobrazení přiřazující každé množině $A \in \mathcal{A}$ reálné číslo $P(A)$, které splňuje:

  (a) $P(A) \geq 0$ pro každý $A \in \mathcal{A}$,
  (b) $P(\Omega) = 1$,
  (c) $P(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty P(A_i)$ pro každou posloupnost jevů $A_1, A_2, \ldots$, které jsou po dvou disjunktní, tj. $A_i \cap A_j = \emptyset$ pro $i \neq j$.

Prvky $\omega \in \Omega$ se nazývají *elementární jevy* a představují výsledky náhodného pokusu. Množiny $A \in \mathcal{A}$ se nazývají *jevy*, někdy také měřitelné jevy, a jsou to podmnožiny množiny Ω, ale ne každá podmnožina množiny Ω musí být jevem. Jev je tedy množina A sestávající z nějakých výsledků pokusu, o nichž říkáme, že jsou jevy A příznivé. Pro jev A se číslo $P(A)$ nazývá *pravděpodobnost jevu A*.

Pravděpodobnostní prostor se nazývá *diskrétní*, pokud je množina Ω je konečná nebo spočetná. Jiné pro jednoduchost uvažovat nebudeme.

Poznámka 6.6. σ-algebra $\langle \Omega, \mathcal{A} \rangle$ tedy popisuje, všechny možné výsledky daného náhodného pokusu, dále pak obsahuje množinu $\mathcal{A}$ těch jevů, pro které je definována jejich pravděpodobnost (tj. lze měřit jejich pravděpodobnost, v principu mohou existovat „divné“ jevy, pro než pravděpodobnost definovat smysluplně nelze). Podmínka (a) říká, že když A je měřitelný jev, pak ji jeho doplněk $\overline{A}$ je měřitelný; (b) říká, že sjednocení spočetné množiny měřitelných jevů je měřitelný jev.

Podmínka (a) z definice pravděpodobnostní míry říká, že pravděpodobnost je nezáporné číslo. Sama množina Ω patří mezi měřitelné jevy (tj. $\Omega \in \mathcal{A}$); je to tzv. *jistý jev* (každý výsledek je mu příznivý). Jejím doplňkem je $\emptyset$, která je tedy také jevem, tzv. *nemožným jevem* (žádný výsledek mu není příznivý). Podmínka (b) říká, že pravděpodobnost jistého jevu je 1. Podmínky (a) a (b) mají technický význam a zajišťují, že pravděpodobnost $P(A)$ každého jevu je z intervalu $[0, 1]$. Podmínka (c) se nazývá aditivita pravděpodobnosti. Pro dva jevy, A_1 a A_2, říká, že pokud $A_1 \cap A_2 \neq \emptyset$ (nemají společný výsledek), pak je pravděpodobnost $P(A_1 \cup A_2)$ rovna součtu $P(A_1)+P(A_2)$ jejich pravděpodobností (rozmyslete si tuto vlastnost na příkladu hodu kostkou).

Příklad 6.7. Uvažujme hod kostkou (viz výše). Ten lze nahlížet jako náhodný pokus, který je popsán pravděpodobnostním $\langle \Omega, \mathcal{A}, P \rangle$, kde:

- $\Omega = \{\omega_1, \ldots, \omega_6\}$,
- $\mathcal{A} = 2^\Omega$,
- $P(A) = {}^{|A|}/_{6}$.

Tedy prostor má 6 elementárních jevů, jevem je každá podmnožina $A \subseteq \Omega$ a pravděpodobnost $P(A)$ jevu A je definována v souladu s klasickou definicí.

Příklad 6.8. Důležité příklady diskrétních pravděpoodobnostních prostorů:

- Konečný s rovnoměrným rozdělením pravděpodobnosti:
  - $\Omega = \{\omega_1, \ldots, \omega_n\}$
  - $\mathcal{A} = 2^\Omega$
  - $P(A) = {}^{|A|}/_{n}$
  - Odpovídá klasické (Laplaceově) definici pravděpodobnosti.
- Konečný s obecným rozdělením pravděpodobnosti:
  - $\Omega = \{\omega_1, \ldots, \omega_n\}$
  - $\mathcal{A} = 2^\Omega$
  - pro každé ω_i dáno $p_i \in [0, 1]$ t.ž. $\sum_{i=1}^{n} p_i = 1$
  - $P(A) = \sum_{\omega_i \in A} p_i$
  - pro $p_i = {}^{1}/_{n}$ jde o konečný s rovnoměrným rozdělením
- Spočetný:
  - $\Omega = \{\omega_1, \omega_2, \ldots\}$, $\mathcal{A} = 2^\Omega$,
  - pro každé ω_i dáno $p_i \in [0, 1]$ t.ž. $\sum_{i=1}^{\infty} p_i = 1$
  - $P(A) = \sum_{\omega_i \in A} p_i$

Existují další typy pravděpodobnostních prostorů (zejm. spojité).

Věta 6.9. *Pro jevy libovolné A a B v pravděpodobnostním prostoru platí:*

(a) $P(\overline{A}) = 1 - P(A)$

(b) $P(\emptyset) = 0$

(c) je-li $A \subseteq B$*, pak* $P(A) \leq P(B)$

(d) $0 \leq P(A) \leq 1$

(e) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$*,*

Důkaz. (a) Jevy A a $\overline{A}$ jsou disjunktní a platí tedy $1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A})$, odkud tvrzení ihned plyne.

(b) Dle (a) je $P(\emptyset) = P(\overline{\Omega}) = 1 - P(\Omega) = 1 - 1 = 0$.

(c) Když $A \subseteq B$, pak $B = A \cup B - A$. Protože jevy A a $B - A$ jsou disjunktní, je $P(B) = P(A) + P(B - A)$. Protože $P(B - A) \geq 0$, je $P(B) \geq P(A)$.

(d) $0 \leq P(A)$ je podmínka (a) z definice 6.5. Protože $A \subseteq \Omega$, plyne z (c), že $P(A) \leq P(\Omega)$. Protože podle definice 6.5 je $P(\Omega) = 1$, je tvrzení dokázáno.

(e) Je $A = (A-B) \cup (A \cap B)$, $B = (B-A) \cup (A \cap B)$ a $A \cup B = (A-B) \cup (A \cap B) \cup (B-A)$, přitom $A - B$, $B - A$ a $A \cap B$ jsou po dvou disjunktní. Z aditivity P tedy plyne $P(A \cup B) = P((A - B) \cup (A \cap B) \cup (B - A)) = P(A - B) + P(A \cap B) + P(B - A)$. Platí tedy

$$\begin{aligned} P(A \cup B) &= P(A - B) + P(A \cap B) + P(B - A) = \\ &= P(A - B) + P(A \cap B) - P(A \cap B) + P(A \cap B) \\ &\quad + P(B - A) + P(A \cap B) - P(A \cap B) = \\ &= P((A - B) \cup A \cap B) + P((B - A) \cup (A \cap B)) - P(A \cap B) = \\ &= P(A) + P(B) - P(A \cap B). \quad \square \end{aligned}$$

Příklad 6.10. Balíček mariášových karet obsahuje karty 4 barev (červené, zelené, žaludy, kule) a 8 hodnot (od sedmičky po eso), celkem 32 karet. Náhodně vytáhneme tři karty. Jaká je pravděpodobnost, že:

(a) jako druhá karta bude vytaženo červené eso;

(b) jako druhá karta bude vytažena červená nebo žaludy;

(c) budou vytažena 3 esa;

(d) bude vytaženo aspoň jedno eso?

(a): Potřebujeme vzít v úvahu pořadí karet. Za elementární jevy (výsledky, prvky množiny Ω) tedy považujme uspořádané trojice $\langle k_1, k_2, k_3 \rangle$, kde k_i označuje i-tou vybranou kartu. Takových trojic je $32 \cdot 31 \cdot 30$ (jde o variace 3 z 32), tj. máme $|\Omega| = 32 \cdot 31 \cdot 30$. Výsledky příznivé danému jevu A (druhá karta bude červené eso) jsou právě trojice, pro které k_2 je červené eso. Těch je $31 \cdot 30$ (dle pravidla součinu: k_1 lze zvolit 31 způsoby, k_3 pak nezávisle na tom 30 způsoby). Pravděpodobnost výběru každé trojice je stejná: $\frac{1}{32 \cdot 31 \cdot 30}$. Jde tedy o klasickou pravděpodobnost, a proto $P(A) = \frac{|A|}{|\Omega|} = \frac{31 \cdot 30}{32 \cdot 31 \cdot 30} = \frac{1}{32}$.

Jinou úvahou: Je zřejmé, že pro každou kartu je pravděpodobnost p, že bude vytažena jako druhá, stejná. Protože karet je 32 a protože součet těchto pravděpodobností je 1, je $p = \frac{1}{32}$.

(b): Zvolme Ω stejně jako v (a). Jev A, tj.

$$A = \{\langle k_1, k_2, k_3\rangle \in \Omega \mid k_2 \text{ je červená nebo žaludy}\},$$

má $(8+8)\cdot 31\cdot 30$ prvků. Totiž, na pozici k_2 lze zvolit libovolnou z 8 červených nebo libovolnou z 8 žaludských karet. Zbylé karty lze po volbě karty k_2 zvolit, stejně jako v (a), $31\cdot 30$ způsoby. Je zřejmé, že tímto způsobem dostaneme všechny trojice, které jsou příznivé jevu A. Dostáváme tedy $P(A) = \frac{|A|}{|\Omega|} = \frac{16\cdot 31\cdot 30}{32\cdot 31\cdot 30} = \frac{1}{2}$.

Pokud bychom příznivé výsledky popisovali od volby první karty, mohli bychom postupovat takto. k_1 lze zvolit 32 způsoby. Nyní je třeba rozlišit, zda k_1 je, nebo není jednou z červených nebo žaludských karet. Pokud je, tj. k_1 je jednou z 16 červených nebo žaludských, lze k_2 zvolit 15 způsoby (zbývá 15 červených nebo žaludských), poté pak zvolit k_3 jedním ze 30 zbývajících způsobů; tím dostáváme $16\cdot 15\cdot 30$ výsledků. Pokud není, tj. k_1 je jednou z 16 zelených nebo kulových, pak lze k_2 zvolit 16 způsoby, poté k_3 30 způsoby; dostáváme $16\cdot 16\cdot 30$ dalších způsobů. Tyto způsoby jsou navzájem různé, celkem tedy máme $16\cdot 16\cdot 30 + 16\cdot 15\cdot 30 = 16\cdot(16+15)\cdot 30 = 16\cdot 31\cdot 30$ způsobů. Došli jsme tedy ke stejnému počtu výsledků příznivých jevu A, ovšem bylo to složitější. Vidíme tedy, že je důležité se vhodným způsobem na situaci podívat.

(c) Zvolme Ω opět jako v (a). Protože jsou celkem 4 esa, je celkem $4\cdot 3\cdot 2$ trojic, ve kterých jsou 3 esa, tj. výsledků příznivých danému jevu A. Je tedy $P(A) = \frac{4\cdot 3\cdot 2}{32\cdot 31\cdot 30} \approx 0.0008$.

Jiný pohled. Považujme za výsledek množinu tří vybraných karet, tj. nikoli uspořádanou trojici $\langle k_1, k_2, k_3\rangle$ jako výše, ale neuspořádanou trojici neboli tříprvkovou množinu $\{k_1, k_2, k_3\}$. Takových trojic je $\binom{32}{3}$. Trojic obsahujících jen esa je $\binom{4}{3}$. Konkrétně máme $\Omega = \{\{k_1, k_2, k_3\} \mid k_1 \neq k_2 \neq k_3 \neq k_1\}$ a

$$A = \{\{e_1, e_2, e_3\}, \{e_1, e_2, e_4\}, \{e_1, e_3, e_4\}, \{e_2, e_3, e_4\}\},$$

kde $e_1, \ldots, e_4$ označují čtyři esa. Hledaná pravděpodobnost je tedy

$$P(A) = \frac{\binom{4}{3}}{\binom{32}{3}} = \frac{\frac{4!}{3!}}{\frac{32!}{29!3!}} = \frac{4\cdot 3\cdot 2}{32\cdot 31\cdot 30},$$

a to je stejná hodnota, ke které jsme došli prvním způsobem.

(d) Tento příklad je jedním z častých případů, kdy je snadnější určit pravděpodobnost jevu $\overline{A}$, který je komplementární k danému jevu A. Zvolme Ω jako v (a). $\overline{A}$ je jev, který nastane, když nepadne žádné eso, tj. každá karta v trojici $\langle k_1, k_2, k_3\rangle$ bude některou z 28 karet, které nejsou esa. Takových trojic je $28\cdot 27\cdot 26$. Dostáváme tedy $P(\overline{A}) = \frac{28\cdot 27\cdot 26}{32\cdot 31\cdot 30} \approx 0.66$ a $P(A) = 1 - P(\overline{A}) \approx 0.34$.

Ke stejnému výsledku opět dojdeme, budeme-li za elementární jevy brát tříprvkové množiny $\{k_1, k_2, k_3\}$. Trojic neobsahujících esa je $\binom{28}{3}$, tedy $P(\overline{A}) = \frac{\binom{28}{3}}{\binom{32}{3}}$, což po úpravách dává opět $\frac{28\cdot 27\cdot 26}{32\cdot 31\cdot 30}$.

Pokud bychom chtěli určit pravděpodobnost jevu A přímo, mohli bychom postupovat takto. Za výsledky berme opět neuspořádané trojice. Jev A (aspoň jedno eso) lze chápat jako sjednocení tří navzájem disjunktních jevů, A_1 (právě jedno eso), A_2 (právě dvě esa), A_3 (tři esa). Počet trojic v A_1 je $4\cdot\binom{28}{2}$, protože trojice s právě jedním esem dostaneme tak, že zvolíme eso ($\binom{4}{1} = 4$ způsoby), zvolíme neuspořádanou dvojici karet, která jsou esy ($\binom{28}{2}$ způsobů), a eso do této dvojice přidáme. Podle pravidla součinu tak získáme $4\cdot\binom{28}{2}$ různých trojic, tedy $|A_1| = 4\cdot\binom{28}{2}$. Trojic s právě dvěma esy je podle podobné úvahy (zvolíme dvojici es a pak znývající kartu) $\binom{4}{2}\cdot\binom{28}{1} = \binom{4}{2}\cdot 28$. Trojic s třemi esy je $\binom{4}{3}$. Je tedy

$$\begin{aligned} |A| &= |A_1 \cup A_2 \cup A_3| = |A_| \cup |A_2| \cup |A_3| = \\ &= 4\cdot\binom{28}{2} + \binom{4}{2}\cdot 28 + \binom{4}{3} = 4\cdot 378 + 6\cdot 28 + 4 = 1684. \end{aligned}$$

Je tedy

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1684}{\binom{32}{3}} = \frac{1684}{4960} \approx 0.34.$$

Došli jsme tedy ke stejné hodnotě, ale složitěji.

Pozor na chyby:

Příklad 6.11. Házíme dvěma kostkami. Jaká je pravděpodobnost, že padne součet 8? Jak určit množinu Ω elementárních jevů? Ukážeme tři možnosti.

(a) Možnosti výsledného součtu jsou 2 (padnou dvě jedničky), 3 (jednička a dvojka), $\ldots$, 12 (dvě šestky). Tato úvaha nás vede k volbě $\Omega = \{2, \ldots, 12\}$. Ω má 11 prvků.

(b) Za elementární jev můžeme považovat (neuspořádanou) dvojici $\{i, j\}$, která vyjadřuje, že na jedné kostce padne i, na druhé j a my nerozlišujeme, na které. Tedy $\Omega = \{\{1,1\}, \{1,2\}, \{1,3\}, \ldots, \{6,6\}\}$. Ω má 21 prvků (15 dvojic, které je možné vybrat z 6 možností: $15 = \binom{6}{2}$; 6 jednoprvkových množin $\{1,1\}, \ldots, \{6,6\}$).

(c) Za elementární jev budeme považovat dvojici čísel $\langle i, j\rangle$, která vyjadřuje, že na první kostce pade i a na druhé j. Pak $\Omega = \{\langle 1,1\rangle, \langle 1,2\rangle, \langle 2,1\rangle, \ldots, \langle 6,6\rangle\}$ a Ω má 36 prvků.

O žádné z těchto možností nelze říct, že to je „ta správná“. Záleží na tom, jak budeme se zvolenou množinou Ω pracovat. Pojďme tedy spočítat pravděpodobnost, že padne součet 8. Použijme Laplaceovo pravidlo, tj. stanovme zmíněnou pravděpodobnost, p, jako podíl počtu výsledků příznivých jevu „padne součet 8“ a počtu všech výsledků, tj. počtu prvků množiny Ω.

Při volbě (a) vychází $p = \frac{1}{11} \approx 0.091$, protože z 11 elementárních jevů v Ω je jen jeden příznivá, totiž 8. Při volbě (b) jsou příznivé $\{2,6\}$, $\{3,5\}$, a $\{4,4\}$, tedy $p = \frac{3}{21} \approx 0.143$. Při volbě (c) jsou příznivé $\langle 2,6\rangle$, $\langle 6,2\rangle$, $\langle 3,5\rangle$, $\langle 5,3\rangle$, $\langle 4,4\rangle$. Dostáváme tedy $p = \frac{5}{36} \approx 0.139$.

Jak je to možné? Úvahy týkající se případů (a) a (b) jsou totiž chybné. Problém je v tom, že jednotlivé elementární jevy v množinách Ω nejsou stejně pravděpodobné, jak to vyžaduje Laplaceovo pravidlo. Uvažme například elementární jevy $\{2,6\}$ a $\{4,4\}$ v případu (b). $\{2,6\}$ představuje ve skutečnosti dva možné výsledky: na první kostce padne 2, na druhé 6, druhý výsledek je, že na první kostce padne 6, na druhé 2; $\{4,4\}$ představuje jen jeden takový výsledek: na první i na druhé kostce padne 4. $\{2,6\}$ má tedy dvakrát větší pravděpodobnost než $\{4,4\}$, a použít Laplaceovo pravidlo tedy nelze.

Při volbě množiny elementárních jevů je tedy třeba přihlížet k úloze kterou máme řešit. Jako rozumná se však ukazuje zásada volit jako elementární jevy neredukovatelné, neagregované entity, jak jsme to udělali v případu (c) Příkladu 6.11.

Příklad 6.12. V dodávce zboží je 85 výrobků bezchybných a 15 vadných. Náhodně vybereme 10 výrobků. Jaká je pravděpodobnost, že mezi nimi bude aspoň jeden vadný? Protože na pořadí mezi 10 vybranými výrobky nezáleží, vezměme za Ω množinu všech 10-tiprvkových podmnožin množiny $\{v_1, \ldots, v_{100}\}$ (výrobky v dodávce). Jejich počet je roven počtu kombinací 10 z 100, tj. počtu způsobů, jak vybrat 10 prvků ze 100, což je $\binom{100}{10}$. Místo určení pravděpodobnosti daného jevu A, tj. „aspoň jeden vadný“, je v tomto případě snazší určit pravděpodobnost jevu $\overline{A}$, tj. „žádný vadný“. $\overline{A}$ obsahuje výsledky, tj. 10-ti prvkové podmnožiny, ve kterých jsou všechny výrobky bezvadné, a těch je $\binom{85}{10}$. Je tedy

$$P(\overline{A}) = \frac{\binom{85}{10}}{\binom{100}{10}} = \frac{\frac{85!}{75!10!}}{\frac{100!}{90!10!}} = \frac{85!90!}{75!100!} \approx 0.18$$

. Tedy $P(A) = 1 - P(\overline{A}) \approx 0.82$

Příklad 6.13. V dodávce je n výrobků, z nich je k bezchybných a $n-k$ vadných. Jaká je pravděpodobnost, že z p náhodně vybraných výrobků bude právě r vadných?

Za elementární jev považujme množinu p vybraných výrobků. Takových množin je $\binom{n}{p}$. Protože r vadných lze vybrat $\binom{n-k}{r}$ způsoby a zbývajících $p-r$ bezchybných lze vybrat $\binom{k}{p-r}$ způsoby, existuje $\binom{n-k}{r} \cdot \binom{k}{p-r}$ elementárních jevů, které jsou obsaženy v daném jevu A (r vadných a $p-r$ bezchybných). Je tedy

$$P(A) = \frac{\binom{n-k}{r}\binom{k}{p-r}}{\binom{n}{p}}.$$

6.2.3 Náhodné veličiny a jejich charakteristiky

Pojem náhodné veličiny je jedním z nejdůležitějších pojmů teorie pravděpodobnosti. Představme si, že náhodný pokus spočívá v náhodném výběru muže v České republice. Označíme-li $\Omega = \{\omega_1, \ldots, \omega_k\}$ množinu všech mužů v České republice, lze tento výběr popsat pravděpodobnostním prostorem, ve kterém množinou elementárních jevů (možných výsledků) je Ω a pravděpodobnost výběru každého muže ω_i je $P(\{\omega_i\}) = 1/k$. V této situaci nás může zajímat například zajímat výška mužů nebo jiná veličina, kterou můžeme pozorovat (hmotnost, věk, roční příjem apod.). Vznikají přirozené otázky:

- Jaká je pravděpodobnost, že náhodně vybraný muž má výšku aspoň 180 cm?
- Jaká je očekávaná hodnota výšky náhodně vybraného muže?
- Jaká je nejmenší a největší výška může a jak jsou hodnoty výšky mužů mezi nimi jsou rozptýleny?

Výšku mužů lze chápat jako funkci $X : \Omega \to \mathbb{R}$, která muži $\omega \in \Omega$ přiřadí jeho výšku $X(\omega)$, např. tedy $X(\omega) = 182$. Výška mužů se tedy v tomto pohledu jeví jako náhodná veličina: výška vybraného muže je náhodná, protože tento muž je vybrán náhodně.

Definice Náhodná veličina na konečném nebo diskrétním pravděpodobnostním prostoru $\langle \Omega, 2^\Omega, P \rangle$ je funkce

$$X : \Omega \to \mathbb{R}.$$

Poznámka 6.14. Naše definice je speciálním případem obecné definice náhodné veličiny.

Příklad 6.15. (triviální) Hod kostkou, $\Omega = \{\omega_1, \ldots, \omega_6\}$, $P(\{\omega_i\}) = 1/6$. Pak funkce X definovaná

$$X(\omega_i) = i$$

je náhodná veličina na $\langle \Omega, 2^\Omega, P \rangle$.

Příklad 6.16. Experiment skončí úspěchem s pravděpodobností p a neúspěchem s pravděpodobností $1-p$. Provedeme ho $3\times$ po sobě; provedení jsou na sobě nezávislá. Uvažujme $p = 1/2$.

Pro tuto situaci uvažujme pravděpodobnostní prostor $\langle \Omega, 2^\Omega, P \rangle$, kde

- $\Omega = \{\langle 0,0,0\rangle, \langle 0,0,1\rangle, \langle 0,1,0\rangle, \ldots, \langle 1,1,1\rangle\}$,
  $P(\{\omega\}) = 1/|\Omega| = 1/8$.

Zobrazení $X : \Omega \to \mathbb{R}$

$$X(\langle a, b, c\rangle) = a + b + c$$

je náhodná veličina, která jejíž hodnotou je počet úspěšných experimentů v sérii $\langle a, b, c\rangle$. Platí např.

$$X(\langle 0, 0, 0\rangle) = 0, X(\langle 0, 1, 0\rangle) = 1, X(\langle 1, 0, 1\rangle) = 2.$$

Možné hodnoty veličiny X jsou 0, 1, 2 a 3, tj. $X(\Omega) = \{0, 1, 2, 3\}$.

Uvažujme nejdřív otázku:

$$\text{Jaká je pravděpodobnost, že veličina } X \text{ má hodnotu } x? \tag{26}$$

Jaká je tedy například pravděpodobnost, že náhodná veličina X z příkladu 6.16 má hodnotu 1? Podobného charakteru jsou otázky:

- Jaká je pravděpodobnost, že počet chyb náhodně vybraného výrobku je 2?
- obecněji: Jaká je pravděpodobnost, že hmotnost h náhodně vybraného muže je $75\,\text{kg} \leq h \leq 85\,\text{kg}$?

Rozeberme nyní otázku (26). Lze ji formulovat tak, aby se stala standardní otázkou, tj. takto?:

$$\text{Jaká je pravděpodobnost jevu } A? \tag{27}$$

Uvědomme si, že naši otázku lze ekvivalentně vyslovit takto: jaká je pravděpodobnost, že výška vybraného muže je x? Nyní je zřejmé, že jev A, o jehož pravděpodobnost jde, je

$$A = \{\omega \in \Omega \mid X(\omega) = x\},$$

tedy A je množina těch výsledků ω, jejichž hodnota $X(\omega)$ veličiny X je rovna x. Došli jsme k tomu, že otázka (26) je vlastně otázkou:

$$\text{Jaká je pravděpodobnost } P(\{\omega \in \Omega \mid X(\omega) = x\})? \tag{28}$$

Tuto pravděpodobnost značíme také $P(X = x)$.

Příklad 6.17. Vraťme se k příkladu 6.16. Jaká je pravděpodobnost, že ze 3 pokusů budou právě dva úspěšné? Tedy jaká je $P(X = 2)$?

Platí: jev „$X = 2$“ $= \{\omega \in \Omega \mid X(\omega) = 2\} = \{\langle 0, 1, 1\rangle, \langle 1, 0, 1\rangle, \langle 1, 1, 0\rangle\}$.

Proto $P(X = 2) = 3 \cdot 1/8 = 3/8$.

Vidíme tedy, že je-li X náhodná veličina na pravděpodobnostním prostoru $\langle \Omega, 2^\Omega, P\rangle$, můžeme určovat pravděpodobnosti, které se týkají hodnot náhodné veličiny X. Podobně jako jsme určovali pravděpodobnost $P(X = x)$, můžeme určovat pravděpodobnost $P(X \leq x)$ nebo $P(X \geq x)$.

Příklad 6.18. Uvažujme opět příklad 6.16. Pravděpodobnost, že ze tří po sobě provedených pokusech budou aspoň dva úspěšné, je

$$P(X \geq 2) = P(\{\langle 0, 1, 1\rangle, \langle 1, 0, 1\rangle, \langle 1, 1, 0\rangle, \langle 1, 1, 1\rangle\}) = 4 \cdot 1/8 = 1/2.$$

Jaká je pravděpodobnost, že ze tří pokusů bude právě jeden úspěšný nebo právě tři úspěšné? Jaká je tedy $P(X \in \{1, 3\})$?

Protože

$$P(X \in \{1, 3\}) = P(\omega \in \Omega \mid X(\omega) \in \{1, 3\}),$$

je

$$P(X \in \{1, 3\}) = P(\{\omega \mid X(\omega) = 1 \text{ nebo } X(\omega) = 3\}) =$$
$$P(\{\langle 0, 0, 1\rangle, \langle 0, 1, 0\rangle, \langle 1, 0, 0\rangle, \langle 1, 1, 1\rangle\}) = 4 \cdot 1/8 = 1/2.$$

Střední hodnota náhodné veličiny

Uvažujeme-li množinu $X(\Omega)$ všech možných hodnot náhodné veličiny X, můžeme se ptát, jaká je očekávaná hodnota výsledku. Jsou-li například tři možné výsledky, ω_1, ω_2 a ω_3, mají hodnoty $X(\omega_1) = 2$, $X(\omega_2) = 6$, $X(\omega_3) = 10$, pak pokud jsou stejně pravděpodobné, je intuitivně jasné. že očekávanou hodnotou výsledku by měla být hodnota 6. S pravděpodobností $1/3$ totiž výsledek bude mít hodnotu 2, 6 nebo 10. Průměrná hodnota výsledku je tedy $\frac{2+6+10}{3} = 6$. Tuto průměrnou hodnotu je rozumné považovat za očekávanou hodnotu výsledku náhodného pokusu. Jiný pohled na výpočet průměrné hodnoty je tento: $6 = 1/3 \cdot 2 + 1/3 \cdot 6 + 1/3 \cdot 10$, tedy očekávaná hodnota je

$$6 = 2 \cdot P(X = 2) + 6 \cdot P(X = 6) + 10 \cdot P(X = 10).$$

Očekávanou hodnotu tedy podle tohoto přístupu získáme tak, že každou možnou hodnotu x z množiny $X(\Omega) = \{2, 6, 10\}$ vynásobíme její pravděpodobností $P(X = x)$ a výsledné součiny $x \cdot P(X = x)$ sečteme. To vede k následující definici.

Definice *Střední hodnota* (také očekávaná hodnota) náhodné veličiny X se značí $E(X)$ a je definována následovně

$$E(X) = \sum_{x \in X(\Omega)} x \cdot P(X = x).$$

I v případě, že pravděpodobnosti výsledků nejsou stejné, je uvedený přístup intuitivně rozumný. Když např. $P(\{\omega_1\}) = 1/2$, $P(\{\omega_2\}) = 1/4$ a $P(\{\omega_1\}) = 1/4$, pak uvedený vzorec dá

$$X(\omega_1) \cdot P(X = 2) + X(\omega_2) \cdot P(X = 6) + X(\omega_3) \cdot P(X = 10) = 2 \cdot 1/2 + 6 \cdot \inf 14 + 10 \cdot 1/4 = 5.$$

- Pokud je množina hodnot veličiny X konečná, tj. $X(\Omega) = \{x_1, \ldots, x_n\}$, je
$$E(X) = x_1 \cdot P(X = x_1) + \cdots + x_n \cdot P(X = x_n).$$
- Pokud je navíc pravděpodobnostní rozdělení rovnoměrné, tj. $P(X = x_i) = 1/n$, pak
$$E(X) = x_1 \cdot 1/n + \cdots + x_n \cdot 1/n = \frac{\sum_{i=1}^{n} x_i}{n},$$
tedy $E(X)$ je aritmetický průměr hodnot x_i.
- Obecně: $E(X)$ vyjadřuje, s přihlédnutím k pravděpodobnosti hodnot, očekávanou hodnotu výsledku.
- $E(X)$ nemusí být rovna žádné z hodnot, které X nabývá (to jsme viděli v právě uvedeném příkladu).

Příklad 6.19. $\Omega = \{\omega_1, \ldots, \omega_4\}$ s rovnoměrným rozdělením, tj. $P(\{\omega_i\}) = 1/4$ pro každé i.

X je dána takto:

$$X(\omega_1) = 1,\ X(\omega_2) = 10,\ X(\omega_3) = 1,\ X(\omega_4) = 12.$$

Tři možné hodnoty, $x_1 = 1$, $x_2 = 10$ a $x_3 = 12$ a je

Je $P(X = 1) = P(\{\omega_1, \omega_3\}) = 1/2$, $P(X = 10) = P(\{\omega_2\}) = 1/4$ a $P(X = 12) = P(\{\omega_4\}) = 1/4$. Tedy

$$\begin{aligned} E(X) &= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + x_3 \cdot P(X = x_3) = \\ &= 1 \cdot P(X = 1) + 10 \cdot P(X = 10) + 12 \cdot P(X = 12) = \\ &= 1 \cdot 1/2 + 10 \cdot 1/4 + 12 \cdot 1/4 = 6. \end{aligned}$$

Příklad 6.20. Jaká je $E(X)$ v příkladu 6.16?

Je $\Omega = \{\langle 0,0,0\rangle, \langle 0,0,1\rangle, \langle 0,1,0\rangle, \ldots, \langle 1,1,1\rangle\}$. Je $X(\langle a,b,c\rangle) = a+b+c$ a $X(\Omega) = \{0,1,2,3\}$.

$$\begin{aligned} E(X) &= 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3) = \\ &= 0 \cdot 1/8 + 1 \cdot 3/8 + 2 \cdot 3/8 + 3 \cdot 1/8 = \\ &= 12/8 = 1.5. \end{aligned}$$

Příklad 6.21. (časová složitost $T(n)$ algoritmu A v průměrném případě jako střední hodnota)

Obvyklá definice $T(n)$: nechť $I_1, \ldots, I_k$ jsou všechny vstupy velikosti n, pak

$$T(n) = \frac{t_1 + \cdots + t_k}{k}, \quad \text{kde } t_i \text{ je počet kroků } A \text{ (trvání) pro zpracování vstupu } I_i.$$

Definice pojmu časová složitost v průměrném případě pomocí pojmu střední hodnota.

Pro danou velikost n vstupu uvažujme množinu $\Omega_n = \{I_1, \ldots, I_k\}$ všech vstupů algoritmu A, které mají velikost n. Předpokládejme navíc, že pravděpodobnost, že na vstupu je I_j je $p(I_j)$; tedy $\sum_{i=1}^{k} p(I_i) = 1$. Definujme náhodnou veličinu X_n:

$$X_n(I_i) = t_i$$

Pak časová složitost v průměrném případě je funkce $T : \mathbb{N} \to \mathbb{R}$ definovaná

$$T(n) = E(X_n), \quad \text{tedy } T(n) = \sum_i [\text{p-st vstupu } I_i] \cdot [\text{trvání pro } I_i].$$

Pokud $p(I_i) = 1/k$ (stejně pravděpodobné vstupy), dostaneme obvyklou definici.

Věta 6.22. *Nechť X a Y jsou náhodné veličiny na konečném nebo diskrétním $\langle \Omega, 2^\Omega, P\rangle$ a $c \in \mathbb{R}$. Uvažujme funkce $X+Y$ a $c \cdot X$ definované $(X+Y)(\omega) = X(\omega)+Y(\omega)$ a $(c \cdot X)(\omega) = c \cdot X(\omega)$. Pak*

(a) $X + Y$ je náhodná veličina

(b) $E(X+Y) = E(X) + E(Y)$.

(c) $c \cdot X$ je náhodná veličina.

(d) $E(c \cdot X) = c \cdot E(X)$.

Důkaz. (a) a (c): triviální (za předpokladů je každá funkce $f : \Omega \to \mathbb{R}$ náhodná veličina).

(b) a (d): Uvědomme si, že $E(Z) = \sum_{z \in Z(\Omega)} z \cdot P(Z=z) = \sum_{\omega \in \Omega} Z(\omega) \cdot P(\{\omega\})$.

Pak tedy

$E(X+Y) = \sum_z z \cdot P(X+Y=z) = \sum_{\omega\in\Omega}(X+Y)(\omega) \cdot P(\{\omega\}) = \sum_{\omega\in\Omega} X(\omega) \cdot P(\{\omega\}) + \sum_{\omega\in\Omega} Y(\omega) \cdot P(\{\omega\}) = E(X) + E(Y)$.

$E(cX) = \sum_z z \cdot P(cX = z) = \sum_{\omega\in\Omega}(cX)(\omega) \cdot P(\{\omega\}) = c \cdot \sum_{\omega\in\Omega}(cX)(\omega) \cdot P(\{\omega\}) = c \cdot E(X)$. □

Rozptyl a směrodatná odchylka náhodné veličiny

Střední hodnota náhodné veličiny nám dává užitečnou, ale jen omezenou informaci. To platí i pro průměrnou hodnotu, která je speciálním případem střední hodnoty: Jeden člověk sní celé kuře, druhý nic; v průměru měl každý půl kuřete.

Příklad 6.23. Existují tři vstupy, I, J a K pro algoritmus A. jejich zpracování trvá 10, 1000 a 1990 kroků. V průměru tedy trvá zpracování vstupu 1000 kroků.

(b) Zpracování vstupů I, J a K trvá 990, 1000 a 1010 kroků. V průměru tedy 1000 kroků.

Vidíme tedy, že průměr (a střední hodnota) poskytují jen omezenou informaci o veličině X. Hodnoty X (počet snědených kuřat, počet vykonaných kroků) mohou být kolem střední hodnoty $E(X)$ různě rozptýleny. K vyjádření toho, jak moc jsou rozptýleny, slouží tzv. rozptyl, který je další užitečnou charakteristikou náhodných veličin.

Definice 6.24. *Rozptyl* (také *variance*) $\operatorname{var} X$ (také $\mu_2(X)$) náhodné veličiny X je definován vztahem

$$\operatorname{var} X = E((X - E(X))^2).$$

Směrodatná odchylka σ náhodné veličiny X je druhá odmocnina rozptylu, tj.

$$\sigma = \sqrt{\operatorname{var} X}.$$

- $\operatorname{var} X$ vyjadřuje, jak moc jsou hodnoty X rozptýleny kolem $E(X)$. Čím je větší, tím jsou více rozptýleny.
- Jak chápat vzorec pro $\operatorname{var} X$? X je náhodná veličina, $E(X)$ je číslo, $X - E(X)$ je náhodná veličina, $(X - E(X))^2$ je náhodná veličina, rozptyl $E((X - E(X))^2)$ je tedy střední hodnota veličiny $(X - E(X))^2$.

Věta 6.25. *Pro náhodnou veličinu X a konstanty $a, b \in \mathbb{R}$ platí:*

(a) $\operatorname{var}(b) = 0$ *(b jako konstantní funkce je náhodnou veličinou),*

(b) $\operatorname{var}(aX) = a^2 \operatorname{var}(X)$,

(c) $\operatorname{var}(X + b) = \operatorname{var}(X)$,

(d) $\operatorname{var}(aX + b) = a^2 \operatorname{var}(X)$,

Důkaz. Stačí (d), ostatní plynou z (d):

$$\begin{aligned} \operatorname{var}(aX + b) &= E([aX + b - E(aX + b)]^2) = E([aX + b - aE(X) - b]^2) = \\ &= E([aX - aE(X)]^2) = E(a^2[X - E(X)]^2) = \\ &= a^2 E([X - E(X)]^2) = a^2 \operatorname{var}(X). \end{aligned}$$

□

Poznámka 6.26. Pro náhodnou veličinu X má náhodná veličina Z daná $Z = \frac{X - E(X)}{\sqrt{\operatorname{var} X}}$ vlastnosti (důkaz dosazením):

$$E(Z) = 0 \quad \text{a} \quad \operatorname{var} Z = 1$$

Příklad 6.27. Uvažujme $\Omega = \{\omega_1, \ldots, \omega_4\}$, $P(\{\omega_i\}) = 1/4$, a náhodnou veličinu danou

$$X(\omega_1) = 0, X(\omega_2) = 2, X(\omega_3) = 4, X(\omega_4) = 6.$$

Je zřejmé, že

$$P(X = 0) = 1/4, P(X = 2) = 1/4, P(X = 4) = 1/4, P(X = 6) = 1/4,$$

a tedy

$$E(X) = 0 \cdot 1/4 + 2 \cdot 1/4 + 4 \cdot 1/4 + 6 \cdot 1/4 = 3.$$

Označíme-li $x_1 = 0, \ldots, x_4 = 6$, pak

$$\begin{aligned} \operatorname{var} X &= E((X - E(X))^2) = \sum_{i=1}^{4} (x_i - 3)^2 \cdot 1/4 \\ &= [(0-3)^2 + (2-3)^2 + (4-3)^2 + (6-3)^2] \cdot 1/4 \\ &= [9 + 1 + 1 + 9] \cdot 1/4 = 5. \end{aligned}$$

Uvažujme nyní veličinu Y danou (porovnej s X)

$$Y(\omega_1) = 1, Y(\omega_2) = 2, Y(\omega_3) = 4, Y(\omega_4) = 5.$$

Pak

$$P(Y = 1) = 1/4, P(Y = 2) = 1/4, P(Y = 4) = 1/4, P(Y = 5) = 1/4.$$

Dostaneme $E(Y) = 3$. Označíme-li $y_1 = 1$, $y_2 = 2$, $y_3 = 4$, $y_4 = 5$, pak

$$\begin{aligned} \operatorname{var} Y &= \sum_{i=1}^{4} (y_i - 3)^2 \cdot 1/4 \\ &= [(1-3)^2 + (2-3)^2 + (4-3)^2 + (5-3)^2] \cdot 1/4 = \\ &= [4 + 1 + 1 + 4] \cdot 1/4 = 2.5. \end{aligned}$$

Tedy $\operatorname{var} Y < \operatorname{var} X$ v souladu s intuicí: Hodnoty X jsou více rozptýleny.

Dále: pro X je $\sigma = \sqrt{3}$, pro Y je $\sigma = \sqrt{2.5}$.

- Střední hodnota a rozptyl jsou nejčastěji používané charakteristiky.
- Jsou to speciální případy charakteristik nazývaných momenty.
  ($E(X)$ je 1. obecný moment, $\operatorname{var} X$ je 2. centrální moment.)
- Mají uplatnění i v situacích, kde hodnoty nemají typickou pravděpodobnostní interpretaci, viz následující příklad.

Příklad 6.28. Jsou dány hodnoty $x_1, \ldots, x_n$. Jaká je průměrná hodnota a jaký je rozptyl?

Napojení na pravděpodobnostní pojmy:

x_i považujeme za hodnoty náhodné veličiny X; $P(X = x_i) = \frac{count(x_i)}{n}$, kde $count(x_i)$ je počet výskytů x_i v $x_1, \ldots, x_n$ (mohou se opakovat).

$E(X)$ a $\operatorname{var} X$ se pak rutinně spočítá.

Hodnoty x_i jsou např.: zadluženost firem, délka života obyvatel atd.

Kvantily a modus náhodné veličiny

Definice 6.29. Pro $q \in (0,1)$ je $100q\%$-*kvantil*, někdy q-*kvantil*, diskrétní náhodné veličiny X hodnota $x_q \in \mathbb{R}$, pro kterou je

$$P(X < x_q) \leq q \quad \text{a} \quad P(X \leq x_q) \geq q$$

Význam:

- Při rovnoměrném rozdělení pravděpodobnosti: 25% všech hodnot je $\leq x_q$.
- Obecně: Pravděpodobnost, že hodnota je $\leq x_q$, je q.

Kvantily nejsou určeny jednoznačně (q-kvantilem může být interval hodnot), uvidíme.

Speciální názvy:

- *medián* je 0.5-kvantil,
- *dolní kvartil* je 0.25-kvantil, *horní kvartil* je 0.75-kvantil,
- k. *decil* je $k/10$-kvantil, k. *percentil* je $k/100$-kvantil.

Definice 6.30. *Modus* náhodné veličiny je hodnota $\hat{x}$ taková, že pro každé $x \in \mathbb{R}$ je

$$P(X = \hat{x}) \geq P(X = x).$$

Modus je, zhruba řečeno, nejčastější hodnota náhodné veličiny.

Příklad 6.31. Hod kostkou, $\Omega = \{\omega_1, \ldots, \omega_6\}$, $X(\omega_i) = i$.

- $x_{0.5}$ (medián) je každé číslo z intervalu $[3, 4)$.
- $x_{0.25}$ (dolní kvartil) neexistuje.
- $x_{0.75}$ (horní kvartil) neexistuje.
- $x_{1/3}$ je každé číslo z intervalu $[2, 3)$.
- Modus je každá z hodnot $1, \ldots, 6$.

Příklad 6.32. Uvažujme $\Omega = \{\omega_1, \ldots, \omega_4\}$, $P(\{\omega_i\}) = 1/4$ a náhodnou veličinu danou

$$X(\omega_1) = 2, X(\omega_2) = 2, X(\omega_3) = 4, X(\omega_4) = 6.$$

Pak

$$p(X = 2) = 1/2, \quad p(X = 4) = 1/4, \quad p(X = 6) = 1/4.$$

- Mediánem je každá hodnota z intervalu $[2, 4)$.
- Dolní kvartil neexistuje.
- Horní kvartil je každá hodnota z $[4, 6)$.
- Modus je jediný, je jím hodnota 2.

Shrnutí

Teorie pravděpodobnosti je jednou z nejužitečnějších oblastí matematiky. Počítání pravděpodobností jednoduchých jevů patří mezi základní aplikace kombinatorického počítání. Pravděpodobnost jevu je dána podílem počtu možností příznivých danému jevu ku počtu všech možností. Kombinatorické úvahy se uplatní při určování počtu možností.
Pojem pravděpodobnostního prostoru představuje formální rámec pro všechny úvahy o pravděpodobnosti. Číselné charakteristiky výsledků náhodných pokusů jsou formalizovány náhodnými veličinami. Mezi jejich nejdůležitější charakteristiky patří střední hodnota, rozptyl a kvantily.

Pojmy k zapamatování

- elementární jev, jev,
- pravděpodobnost,
- pravděpodobnostní prostor,
- náhodná veličina
- střední hodnota, rozptyl, kvantily

Kontrolní otázky

1. *Jaký je rozdíl mezi pojmy jev a elementární jev?*
2. *Co je to pravděpodobnost jevu a jak je definována?*

Cvičení

1. Házíme dvěma kostkami. Máme si vsadit na číslo, které vzejde jako součet výsledků na jednotlivých kostkách. Na jaké číslo vsadíme?
2. Házíme třikrát po sobě kostkou. Jaká je pravděpodobnost, že výsledek při druhém i při třetím hodu je větší než výsledek při prvním hodu?
3. Házíme třikrát po sobě kostkou. Jaká je pravděpodobnost, že výsledek při druhém hodu je větší než výsledek při prvním hodu a že výsledek při třetím hodu je větší než výsledek při druhém hodu?

Úkoly k textu

1. Vysvětlete podrobně chybu popsanou v odstavci na Příkladem 6.3.

Řešení

1. 6, 7 nebo 8.
2. 55/216.
3. 5/54.

7 Nekonečno

Studijní cíle: Po prostudování kapitoly 7 by student měl rozumět základním pojmům týkajících se nekonečných množin, zejména spočetným a nespočetným množinám. Měl by umět prokázat spočetnost a nespočetnost množin.

Klíčová slova: nekonečno, spočetná množina, nespočetná množina, diagonální metoda

7.1 Proč se zabýváme nekonečnem?

Nekonečno je jedním z nejzáhadnějších jevů. Mohlo by se zdát, že v prakticky motivovaných úlohách vystačíme s konečnými množinami, a že se tedy nekonečnem zabývat nemusíme. Množin, se kterými při řešení praktických problémů pracujeme, jsou ale často nekonečné. Například různé úvahy o algoritmech vedou na nekonečné množiny. Potřebujeme tedy vědět, jak se s nekonečnými množinami pracuje.

V této souvislosti se nabízí řada přirozených otázek. Například ty následující. Jak měřit velikost množin, zejména nekonečných? Co znamená, že dvě množiny jsou stejně velké? Která množina je větší, $\mathbb{N}$ nebo $\mathbb{Q}$? $\mathbb{Q}$ nebo $\mathbb{R}$? V dalším textu se na tyto a další otázky pokusíme odpovědět.

7.2 Konečné a nekonečné množiny

Podle intuice je množina A konečná, pokud lze její prvky „očíslovat“ čísly $1, \ldots, n$. Z toho vychází následující definice.

Definice 7.1. Množina A je konečná, pokud existuje $n \in \mathbb{N}$ a bijekce

$$f : \{1, 2, \ldots, n\} \to A.$$

Množina A je nekonečná, pokud není konečná.

Je-li A konečná, číslo n z definice se nazývá počet prvků (také mohutnost, velikost) A, značí se $|A|$.

Příklad 7.2. $\{2, 3, 5, 7\}$ je konečná, protože např. zobrazení $f : \{1, 2, 3, 4\} \to \{2, 3, 5, 7\}$ definované

$$f(1) = 2, \quad f(2) = 3, \quad f(3) = 5, \quad f(4) = 7$$

je bijekce. Tedy $|\{2, 3, 5, 7\}| = 4$.

Příklad 7.3. Množina S všech kladných sudých čísel je nekonečná.

Toto tvrzení je celkem zřejmé. Dokázat lze následovně: Kdyby $f : \{1, \ldots, n\} \to S$ byla bijekce, vezměme $m =$ největší z $f(1), \ldots, f(n)$. Pak $2(m + 1)$ je sudé, ale protože $2(m + 1) > m$, není obrazem žádného i, což je spor s tím, že f je bijekce.

Příklad 7.4. Množina $\mathbb{R}$ je nekonečná. To lze dokázat podobně jako v předchozím příkladu.

7.3 Spočetné množiny

Volně řečeno, množina A je spočetná, pokud lze její prvky seřadit do posloupnosti (konečné nebo nekonečné).

Definice 7.5. Množina A je spočetná, pokud je

konečná,

nebo existuje bijekce $f : \mathbb{N} \to A$ (pak je tzv. nekonečná spočetná, popř. spočetně nekonečná).

Tedy A je spočetná, právě když existuje posloupnost

$$a_1, \ldots, a_n \quad \text{nebo} \quad a_1, a_2 \ldots$$

ve které se vyskytují všechny prvky z A (f je surjekce) a neopakují se v ní (f je injekce).

Příklad 7.6 (nekonečné spočetné množiny)**.** – $\mathbb{N}$

Základní nekončená spočetná množina: identita $f(n) = n$ je bijekce $f : \mathbb{N} \to \mathbb{N}$.

– $\{2, 4, 6, 8, \ldots\}$

Zobrazení $f : \mathbb{N} \to \{2, 4, 6, 8, \ldots\}$ dané $f(n) = 2n$ je bijekce.

– $\mathbb{N} \cup \{0\}$

Zobrazení $f : \mathbb{N} \to \mathbb{N} \cup \{0\}$ dané $f(n) = n - 1$ je bijekce.

– $\mathbb{Z}$

Požadovaná posloupnost z prvků $\mathbb{Z}$ je např.

$$0, -1, 1, -2, 2, -3, 3, \ldots$$

Jinými slovy, bijekce $f : \mathbb{N} \to \mathbb{Z}$ z definice je

$$f(1) = 0, f(2) = -1, f(3) = 1, f(4) = -2, f(5) = 2, \ldots$$

(Napište explicitní vzorec pro $f(n)$.)

– $\{k, k + 1, k + 2, \ldots\}$ pro libovolné $k \in \mathbb{Z}$

Zobrazení $f : \mathbb{N} \to \{k, k + 1, k + 2, \ldots\}$ dané $f(n) = n + (k - 1)$ je bijekce.

Zastavme se u posledního příkladu, tedy toho, že množina $\{k, k+1, \ldots\}$ je spočetná pro libovolné $k \in \mathbb{Z}$. Pro prokázání spočetnosti této množiny se nabízí následující argument: $\{k, k + 1, \ldots\} \subseteq \mathbb{Z}$ a $\mathbb{Z}$ je spočetná, proto je i $\{k, k + 1, \ldots\}$ spočetná. Tento argument je správný. Platí totiž, že každá nekonečná podmnožina spočetné množiny je spočetná.

Právě zmíněné tvrzení, které dokážeme později, dává odpověď na netriviální otázku. Definice spočetné množiny nevylučuje množiny, které jsou větší než konečné, ale menší než spočetné. Zmíněné tvrzení říká, že takové neexistují.

Příklad 7.7. *Abecedou*. rozumíme libovolnou neprázdnou množinu, jejíž prvky nazýváme symboly abecedy. Příkladem často používané abecedy je $\{0, 1\}$, někdy nazývaná binární abeceda, jejímiž jedinými symboly jsou 0 a 1. *Řetězec* (někdy slovo) nad abecedou Σ je libovolná konečná posloupnost symbolů abecedy. Následující posloupnosti jsou řetězce nad abecedou $\{0, 1\}$:

$$01, 11, 10, 0, 111, 01010101.$$

Mezi řetězce se počítá i tzv. prázdný řetězec, ε, který neobsahuje žádný symbol. Množina všech řetězců nad abecedou Σ se označuje Σ^*.

Následující posloupnost je nekonečná, obsahuje všechny řetězce z $\{0, 1\}^*$ a prvky se v ní neopakují:

$$\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, 0000, 0001, \ldots$$

Prokazuje tedy, že množina $\{0, 1\}^*$ všech řetězců nad abecedou $\{0, 1\}$ je nekonečná spočetná množina.

Uspořádání řetězců

$$\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, 0000, 0001, \ldots$$

je obecně důležité (tzv. shortlex uspořádání). Je definováno následovně:

1. kratší řetězce jsou před delšími,
2. stejně dlouhé řetězce jsou uspořádány lexikograficky.

Podobně lze seřadit do posloupnosti všechna slova nad libovolnou konečnou abecedou. Platí tedy:

Věta 7.8. *Množina Σ^* všech řetězců nad abecedou $\Sigma = \{a_1, \ldots, a_n\}$ je spočetná.*

Např. pro $\Sigma = \{a, b, \ldots, z\}$ je $\Sigma^* = \{\varepsilon, a, computer, world, ssca, \ldots\}$.

Věta 7.8 má důležitý důsledek: Je-li možné nějaké objekty kódovat (reprezentovat, popsat) jako řetězce nad nějakou konečnou abecedou Σ, je množina těchto kódů (popisů) spočetná, a tedy i množina daných objektů je spočetná.

Skutečně, je-li O množina uvažovaných objektů, lze kódování objektů z O chápat jako prosté zobrazení $c : O \to \Sigma^*$. Pro množinu

$$c(O) = \{c(o) \mid o \in O\}$$

všech kódů pak platí $c(O) \subseteq \Sigma^*$ a vzhledem ke spočetnosti Σ^* je i $c(O)$ spočetná.

Příklad 7.9. Z právě provedené úvahy plyne, že množina všech zdrojových kódů v jazyku C (Python, Lisp, ...) je spočetná. Každý zdrojový kód daného programovacího jazyka lze totiž chápat jako řetězec nad abecedou Σ symbolů, které lze ve zdrojovém kódu používat. Množina všech zdrojových kódů daného jazyka je podmnožinou množiny Σ^*, tedy je podmnožinou spočetné množiny, a je proto podle výše uvedené vlastnosti sama spočetná.

Odbočka: lexikografické a shortlex uspořádání

Předpokládejme, že je dána abeceda (tj. konečná množina symbolů) $\Sigma = \{a_1, \ldots, a_n\}$, na ní lineární uspořádání $\leq$, dále že $a_1 < \cdots < a_n$. Délka $|u|$ řetězce $u \in \Sigma^*$ je počet znaků v u, tedy např. $|010| = 3$, $|0| = 1$, $|\varepsilon| = 0$.

Lexikografické (slovníkové) uspořádání $\leq_l$ na Σ^* je definováno následovně: pro $u = u_1 \cdots u_p, v = v_1 \cdots v_q \in \Sigma^*$ je

$$u \leq_l v \quad \text{p. k.} \quad u = v \text{ nebo pro nějaké } i \text{ je } u_1 \cdots u_{i-1} = v_1 \cdots v_{i-1} \text{ a } u_i < v_i$$

Shortlex uspořádání $\leq_s$na Σ^* je definováno následovně: pro $u = u_1 \cdots u_p, v = v_1 \cdots v_q \in \Sigma^*$ je

$$u \leq_s v \quad \text{p. k.} \quad |u| < |v| \text{ nebo } (|u| = |v| \text{ a } u \leq_l v).$$

Rozdíl mezi $\leq_l$ a $\leq_s$ spočívá v následujícím:

- $\leq_s$ je stejného typu jako přirozené uspořádání $\mathbb{N}$ ($\leq_s$ vytváří posloupnost),
- $\leq_l$ není stejného typu jako přirozené uspořádání $\mathbb{N}$ (netvoří takovou posloupnost):
  - např. množina $A = \{1, 01, 001, 0001, \ldots\}$ nemá nejmenší prvek vzhledem k $\leq_l$,
  - každá podmnožina $A \subseteq \mathbb{N}$ ale nejmenší prvek má.

Odbočka: algoritmický pohled

Spočetné množiny lze chápat jako množiny, o kterých lze uvažovat jako o množinách, jejichž prvky mohou být postupně vypisovány nějakým algoritmem (který případně pracuje nekonečně dlouho).

Algoritmus vypisující $\{1, \ldots, n\}$ (konečná spočetná):

for $i \leftarrow 1$ to n do print(i)

Algoritmus vypisující$\{1, 2, \ldots\}$ (nekonečná spočetná):

$i \leftarrow 1$
while *true* do
  print(i)
  $i \leftarrow i + 1$

Takové množiny se nazývají *algoritmicky vyčíslitelné* (nebo *rekurzívně vyčíslitelné*). Každá algoritmicky vyčíslitelná množina je tedy spočetná. V dalším ale ukážeme, že existuje spočetná množina, která není algoritmicky vyčíslitelná.

Příklad 7.10 (nekonečné spočetné množiny, pokračování)**.** – $\mathbb{Q}$

Uvědomme si, že

- čísla $r \in \mathbb{Q}$ lze chápat (vyjádřit) jako zlomky $\frac{p}{q}$, kde $p \in \mathbb{Z}$ a $q \in \mathbb{N}$;
- připustíme-li pro $r \neq 0$ pouze $\frac{p}{q}$, kde p a q jsou nesoudělná, a pro $r = 0$ pouze zlomek $\frac{0}{1}$, je každé $r \in Q$ vyjádřeno právě jedním zlomkem $\frac{p}{q}$ (takové $\frac{p}{q}$ jsou v tzv. kanonickém tvaru).

První argument prokazující spočetnost spočívá v následujícím (promyslete ho detailně):

- nakreslíme množinu $\mathbb{Z} \times \mathbb{N}$,
- její prvky $\langle p, q \rangle$ lze seřadit do posloupnosti,
- z ní vybrat jen $\langle p, q \rangle$ takové, že $\frac{p}{q}$ je v kanonickém tvaru,
- vybraná podposloupnost tedy obsahuje právě všechna racionální čísla a ta se v ní neopakují.

Později si ukážeme další argument.

7.4 Nespočetné množiny

Definice 7.11. Množina se nazývá nespočetná, pokud není spočetná. a tou množinou).

Nespočetná množina je tedy nekonečná, ale neexistuje bijekce mezi $\mathbb{N}$ a touto množinou. Tedy nespočetná množina je „ještě větší“ než $\mathbb{N}$.

Existují ale nespočetné množiny?

Věta 7.12. *Množina $2^{\mathbb{N}}$ všech podmnožin množiny $\mathbb{N}$ je nespočetná.*

Důkaz. Provedeme tzv. diagonální metodou. Důkaz se vede sporem.

Předpokládejme, že $2^{\mathbb{N}}$ je spočetná. Pak existuje bijekce $f : \mathbb{N} \to 2^{\mathbb{N}}$, $f(i) = A_i$, a tedy posloupnost $A_1, A_2, \ldots$ všech podmnožin množiny $\mathbb{N}$. Sestrojíme nyní množinu $B \subseteq \mathbb{N}$, která je různá od každé z $A_1, A_2, \ldots$, to bude spor.

Každou množinu A_i reprezentujeme řádkem s hodnotami 0 a 1. Například

	1	2	3	4	$\cdots$	j	$\cdots$
A_i	0	0	1	0	$\cdots$	a_{ij}	$\cdots$

reprezentuje množinu A_i, pro kterou $1 \notin A_i$, $2 \notin A_i$, $3 \in A_i$, $4 \notin A_i$ a obecně

$$j \in A_i, \text{ pokud } a_{ij} = 1, \quad j \notin A_i, \text{ pokud } a_{ij} = 0.$$

Uvažujme následující schéma a jeho hlavní diagonálu:

	1	2	3	4	$\cdots$
A_1	$\boldsymbol{a_{11}}$	a_{12}	a_{13}	a_{14}	$\cdots$
A_2	a_{21}	$\boldsymbol{a_{22}}$	a_{23}	a_{24}	$\cdots$
A_3	a_{31}	a_{32}	$\boldsymbol{a_{33}}$	a_{34}	$\cdots$
A_4	a_{41}	a_{42}	a_{43}	$\boldsymbol{a_{44}}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Definujme množinu B jako množinu, jejíž řádek b vznikne „negací“ hlavní diagonály. Tedy pro každé $i \in \mathbb{N}$ definujeme $b_i = 1 - a_{ii}$ neboli

$$i \in B, \text{ právě když } i \notin A_i.$$

Pak B je podmnožinou $\mathbb{N}$, ale $B \neq A_1$, $B \neq A_2$, …, protože B a A_i se liší v prvku i. □

Věta 7.13. *Množina 2^A všech podmnožin libovolné nekonečné spočetné množiny A je nespočetná.*

Důkaz. Diagonální metodou jako pro důkaz předchozí věty pro $A = \mathbb{N}$. □

Z uvedených tvrzení plyne, že řada důležitých množin je nespočetných.

Definice 7.14. *Formální jazyk* na konečnou abecedou Σ je libovolná množina $L \subseteq \Sigma^*$.

Řetězce $w \in L$ (slova jazyka L) chápeme jako správně utvořené (podle pravidel jazyka), řetězce $w \in \Sigma^* - L$ chápeme jako nesprávně utvořené. Následující množiny jsou formální jazyky:

- $\{b_n \cdots b_1 0 \mid b_i \in \{0,1\}\}$, tj. $\{0, 00, 10, 000, 010, \dots\}$, je formální jazyk nad $\{0,1\}$. Sestává ze slov představujících zápisy sudých nezáporných čísel ve dvojkové soustavě.
- $\{x, y, (x+y), (x*x), (x*(y+y)), \dots\}$ je jazyk nad $\{x, y, (,), *, +\}$, jehož slova jsou správně utvořené aritmetické výrazy nad danými proměnnými a symboly operací.

Věta 7.15. *Množina všech formálních jazyků nad libovolnou konečnou abecedou je nespočetná.*

Důkaz. Důsledek věty 7.13 a dříve uvedeného tvrzení, že množina Σ^* všech řetězců nad Σ je spočetná. □

Předchozí věta má důležitý důsledek, který se týká možnosti generovat všechny řetězce daného formálního jazyka L pomocí nějakého algoritmu (tedy algoritmické vyčíslitelnosti). Některé formální jazyky jsou algoritmicky vyčíslitelné (tedy všechna jejich slova lze vypisovat algoritmem, viz výše). Mezi ně patří všechny konečné jazyky jako např. $L = \{0, 1, 00, 11\}$. Je zřejmé, že následující nekonečné jazyky jsou také algoritmicky vyčíslitelné:

- $\{1, 11, 111, 1111, \dots\}$,
- $\{01, 0011, 000111, \dots\}$,
- $\{1, 0110, 0011100, 0001111000, \dots\}$,
- $\{u \in \{0,1\}^* \mid u \text{ obsahuje stejný počet 0 a 1}\}$

Existují ale formální jazyky, které nelze vypisovat pomocí algoritmů. Jsou tedy tak vnitřně složité, pomyslná pravidla pro jejich generování nelze popsat žádným algoritmem. Popisuje to následující věta.

Věta 7.16. *Existuje množina řetězců nad $\{0,1\}$, která není algoritmicky vyčíslitelná.*

Důkaz. Uvažujme nějaký programovací jazyk, ve kterém lze zapsat libovolný algoritmus. Takové jazyky se nazývají *turingovsky úplné* a patří mezi ně např. jazyk C, Python, Java nebo Lisp. Každý algoritmus lze tedy reprezentovat zdrojovým kódem ve zvoleném programovacím jazyku, tedy řetězcem symbolů nad nějakou abecedou Σ (zdrojový kód je řetězec symbolů). Protože množina řetězců nad Σ je dle věty 7.8 spočetná, je spočetná i množina všech algoritmů.

Množina všech formálních jazyků nad $\{0,1\}$ (tedy množin řetězců nad $\{0,1\}$) je dle věty 7.15 nespočetná.

Algoritmů je tedy spočetně mnoho, zatímco množin řetězců nad $\{0,1\}$ je nespočetně mnoho. Existuje tedy množina $A \subseteq \{0,1\}^*$, pro kterou neexistuje algoritmus, který by ji vypisoval. (Přesněji: neexistuje přiřazení, které by každé množině řetězců nad $\{0,1\}$ přiřazovalo odpovídající algoritmus; takové přiřazení by muselo být prostým zobrazením nespočetné množiny do spočetné množiny; jak uvidíme níže, takové zobrazení neexistuje.) □

Poznámka 7.17. Algoritmicky vyčíslitelných množin řetězců nad $\{0,1\}$ je dokonce více než těch, které nejsou algoritmicky vyčíslitelné. Algoritmicky vyčíslitelných je totiž spočetně mnoho a když je odstraníme nespočetné množiny všech množin řetězců nad $\{0,1\}$, zbyde stále nespočetná množina. Jejími prvky jsou právě ty množiny řetězců nad $\{0,1\}$, které nejsou algoritmicky vyčíslitelné (rozdíl nespočetné a spočetné množiny je totiž, jak uvidíme níže, nespočetná množina).

Ukázali jsme, že množiny $\mathbb{N}$, $\mathbb{Z}$ i $\mathbb{Q}$ jsou spočetné.

Věta 7.18. *Množina $\mathbb{R}$ je nespočetná.*

Důkaz. (původním Cantorovým argumentem): Stačí ukázat, že otevřený interval $(0,1)$ není spočetná množina (zkuste zdůvodnit proč).

Každé reálné $r \in (0,1)$ lze vyjádřit pomocí desetinného rozvoje

$$r = 0, r_1 r_2 r_3 \cdots .$$

Předpokládejme, že místo konečných rozvojů uvažujeme jen nekonečné rozvoje s periodou 9. Tj. místo $0,378$ uvažujeme $0,377999\cdots$ apod. Pak má každé číslo právě jeden rozvoj.

Kdyby bylo možné čísla z $(0,1)$ seřadit do posloupnosti $r_1, r_2, \ldots$, mohli bychom uvažovat číslo $s = 0, s_1 s_2 \cdots$ definované pomocí diagonály schématu

	1	2	3	$\cdots$
r_1	$\boldsymbol{r_{11}}$	r_{12}	r_{13}	$\cdots$
r_2	r_{21}	$\boldsymbol{r_{22}}$	r_{23}	$\cdots$
r_3	r_{31}	r_{32}	$\boldsymbol{r_{33}}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

předpisem: pro $i \in \mathbb{N}$ je s_i libovolné z $\{1, \ldots, 9\} - \{r_{ii}\}$.

Pak s má uvažovaný rozvoj a $s \neq r_1$, $s \neq r_2$, …, což je spor s předpokladem. □ □

7.5 Další vlastnosti spočetných množin

Nyní uvedeme další užitečné vlastnosti spočetných množin.

Věta 7.19. *Pro libovolnou nekonečnou množinu A existuje injektivní zobrazení $f : \mathbb{N} \to A$.*

(Tedy A obsahuje nekonečnou spočetnou podmnožinu; totiž $f(\mathbb{N})$.)

Důkaz. $f : \mathbb{N} \to A$ budeme definovat indukcí:

1. Definujme $f(1)$ jako libovolný prvek z A.
2. Předpokládejme, že pro $n \in \mathbb{N}$ jsou definovány $f(1), \ldots, f(n)$ a že jsou po dvou různé.

   Pak $\{f(1), \ldots, f(n)\} \subseteq A$ je konečná, je různá od A, a tedy existuje $a \in A - \{f(1), \ldots, f(n)\}$.

   Položme $f(n+1) = a$.

Tím jsme definovali injektivní zobrazení $f : \mathbb{N} \to A$. (Že jsme skutečně definovali zobrazení plyne z principu definice matematickou indukcí; viz kapitolu 8). □

Uvedeme nyní tvrzení, která se často k prokázání spočetnosti množin používají.

Věta 7.20. *Pokud A je spočetná a $B \subseteq A$, pak B je spočetná.*

Důkaz. Je-li B konečná, je dle definice spočetná.

Je-li B nekonečná, existuje dle věty výše injekce $f : \mathbb{N} \to B$. Uvažujme $f(\mathbb{N})$. Pak:

$$f(\mathbb{N}) \subseteq B \subseteq A,$$

přitom existuje bijekce množiny A na $f(\mathbb{N})$.

(Bijekci $g : A \to f(\mathbb{N})$ dostaneme např. jako $g = h^{-1} \circ f$, kde $h : \mathbb{N} \to A$ je bijekce.)

Tvrzení nyní plyne z následujícího lemma: □

Lemma 7.21. *Existuje-li bijekce množiny A_1 na A_3 a je-li $A_3 \subseteq A_2 \subseteq A_1$, pak existuje bijekce množiny A_1 na A_2.*

Důkaz. Nebudeme uvádět. □

Věta 7.22. *Kartézský součin $A \times B$ spočetných množin A a B je spočetná množina.*

Důkaz. Jsou-li A a B konečné, je $A \times B$ konečná (a tedy spočetná).

Je-li jedna nekonečná (např. $A = \{a_1, a_2, \dots\}$) a druhá konečná ($B = \{b_1, \dots, b_k\}$), lze prvky $\langle a_i, b_j \rangle \in A \times B$ uspořádat do posloupnosti

$$\langle a_1, b_1 \rangle, \langle a_1, b_2 \rangle, \dots, \langle a_1, b_k \rangle, \langle a_2, b_1 \rangle, \dots, \langle a_2, b_k \rangle, \langle a_3, b_1 \rangle, \dots$$

(bijekce $f : \mathbb{N} \to A \times B$ je $f(n) = \langle a_i, b_j \rangle$, kde $i = \lceil n/k \rceil$ a $j = ((n-1) \bmod k) + 1$)

Jsou-li obě množiny nekonečné, $A = \{a_1, a_2, \dots\}$ a $B = \{b_1, b_2, \dots\}$, definujme $f : A \times B \to \mathbb{N}$ takto:

$$f(a_i, b_j) = 2^i \cdot 3^j.$$

Protože 2 a 3 jsou prvočísla, je $2^i \cdot 3^j = 2^k \cdot 3^l$, právě když $i = k$ a $j = l$, tedy f je injekce.

Máme tedy bijekci $A \times B$ na $f(A \times B)$.

$f(A \times B) \subseteq \mathbb{N}$ a $\mathbb{N}$ je spočetná, proto je věty 7.20 i $f(A \times B)$ spočetná. Protože f je bijekce, je i $A \times B$ je spočetná. □

Věta 7.23. *Jsou-li $A_1, \dots, A_n$ spočetné množiny, je i $A_1 \times \cdots \times A_n$ spočetná množina.*

Důkaz. Matematickou indukcí s použitím předchozí věty.

1. pro $n = 1$ zřejmé.

2. Předpokládejme, že tvrzení platí pro n a uvažujme $A_1 \times \cdots \times A_{n+1}$.

Pak $A_1 \times \cdots \times A_n$ je spočetná a dle předchozí věty je i $(A_1 \times \cdots \times A_n) \times A_{n+1}$ spočetná.

Je zřejmé, že $f : A_1 \times \cdots \times A_n) \times A_{n+1} \to A_1 \times \cdots \times A_{n+1}$ definovaná

$$f(\langle \langle a_1, \dots, a_n \rangle, a_{n+1} \rangle) = \langle a_1, \dots, a_{n+1} \rangle$$

je bijekce.

Tedy i $A_1 \times \cdots \times A_{n+1}$ je spočetná. □

Věta 7.24. *Sjednocení spočetného systému spočetných množin je spočetná množina.*

Ukažme nyní, že některé dříve uvedená tvrzení o spočetnosti množin lze snadno získat jako důsledky právě uvedených tvrzení.

Příklad 7.25. Víme, že množina $\mathbb{Q}$ je spočetná. Ukažme to jinak, a sice následovně:

- rozložme $\mathbb{Q} = \mathbb{Q}^- \cup \{0\} \cup \mathbb{Q}^+$;
- ukážeme, že $\mathbb{Q}^+ = \{\frac{p}{q} \in \mathbb{Q} \mid p, q \in \mathbb{N} \text{ nesoudělná } \}$ je spočetná:
  - zobrazení $f : \mathbb{Q}^+ \to \mathbb{N} \times \mathbb{N}$ definované $f(\frac{p}{q}) = \langle p, q \rangle$ je bijekce množiny $\mathbb{Q}^+$ na podmnožinu $f(\mathbb{Q}^+)$ spočetné množiny $\mathbb{N} \times \mathbb{N}$, proto je $\mathbb{Q}^+$ je spočetná;
- $\mathbb{Q}^-$ je spočetná ($x \mapsto -x$ je bijekce $\mathbb{Q}^-$ na $\mathbb{Q}^+$) a $\{0\}$ je spočetná,
- tedy i $\mathbb{Q} = \mathbb{Q}^- \cup \{0\} \cup \mathbb{Q}^+$ je spočetná.

Příklad 7.26. Ukážeme, že množina Σ^* všech řetězců nad spočetnou abecedou Σ je spočetná (spočetnost množiny Σ^* pro konečnou Σ jsme už prokázali):

- Každý řetězec $a_1 \cdots a_n \in \Sigma^*$ lze jednoznačně reprezentovat n-ticí $\langle a_1, \dots, a_n \rangle \in \Sigma^n$.

  (prázdný řetězec ε pak prvkem $\emptyset$; poznamenejme, že $\Sigma^0 = \{\emptyset\}$)

– Zřejmě tedy existuje bijekce množiny Σ^* na $\bigcup_{n=0}^{\infty} \Sigma^n$.
($\varepsilon \mapsto \emptyset$ a $a_1 \cdots a_n \mapsto \langle a_1, \ldots, a_n \rangle$)

– $\bigcup_{n=0}^{\infty} \Sigma^n$ je spočetná (je to sjednocení spočetného systému spočetných množin).

– Tedy i Σ^* je spočetná.

7.6 Jak porovnávat množiny podle jejich velikosti?

Pro konečné množiny A a B platí (ověřte):

– $|A| = |B|$, právě když existuje bijekce $f : A \to B$.

– $|A| < |B|$, právě když existuje injekce $f : A \to B$ a neexistuje injekce $g : B \to A$.

Pomocí pokročilejšího aparátu teorie množin lze dokázat, že pro obecné množiny A a B nastane právě jedna z možností:

1. Existuje bijekce $f : A \to B$.
   Pak se A a B považují za stejně velké (jsou tzv. ekvivalentní; mají stejnou mohutnost).
2. Existuje injekce $f : A \to B$ a neexistuje injekce $g : B \to A$.
   Pak se A považuje za menší než B (A má menší mohutnost než B).
3. Existuje injekce $f : B \to A$ a neexistuje injekce $g : A \to B$.
   Pak se A považuje za větší než B (A má větší mohutnost než B).

Kardinální čísla

Kardinální číslo (kardinalita) množiny A je objekt, označujeme ho $|A|$, přiřazený množině A tak, že $|A| = |B|$, právě když A a B jsou ekvivalentní. Pojem kardinální číslo jsme nezavedli přesně (neřekli jsme, co je tím přiřazeným objektem).

Poznámka 7.27. Přesnou definici pojmu kardinální číslo nebudeme uvádět (zájemce odkazujeme na literaturu o teorii množin). Pro konečnou A lze $|A|$ ztotožnit s počtem prvků množiny A (pro ten jsme zavedli symbol $|A|$). Pro nekonečnou množinu složitější. Pro zajímavost uveďme, že Georg Cantor definoval $|A|$ jako třídu všech množin ekvivalentních s A (to je ale problematické z hlediska axiomatických teorií množin). V axiomatickém systému teorie množin ZFC se $|A|$ definuje jako nejmenší tzv. ordinální číslo ekvivalentní s A.

Pro možnosti 1., 2. a 3. výše pak píšeme $|A| = |B|$, $|A| < |B|$ a $|A| > |B|$.

Poznamenejme, že se běžně používá toto označení:

– $\aleph_0 = |\mathbb{N}|$, tj. $\aleph_0$ je mohutnost množiny $\mathbb{N}$,

– $\aleph_1 = |\mathbb{R}|$, tj. $\aleph_1$ je mohutnost množiny $\mathbb{R}$.

Z předchozího víme, že $\aleph_0 < \aleph_1$.

Uvedeme nyní dvě důležitá, tvrzení tvrzení, která se při úvahách o porovnávání velikosti množin používají.

Věta 7.28 (Cantorova-Bernsteinova)**.** *Pokud pro množiny A a B existuje injekce $f : A \to B$ a injekce $g : B \to A$, pak existuje bijekce množiny A na množinu B.*

Důkaz. Vynecháme. □

Cantorova-Bernsteinova (také Cantorova-Schröderova-Bernsteinova) věta dokresluje výše uvedené možnosti 1., 2. a 3. Pokud tedy existuje injekce A do B i injekce B do A, je to případ 1 (existuje bijekce A na B). Větu formuloval Georg Cantor. V roce 1897 ji pak dokázal Felix Bernstein.

Věta 7.29 (Cantorova)**.** *Pro každou množinu A platí $|A| < |2^A|$.*

Cantorova věta tedy říká, že existuje injekce $f : A \to 2^A$, ale neexistuje injekce $g : 2^A \to A$.

Před jejím důkazem uveďme následující:

- S ohledem na to, že $a \mapsto \{a\}$ je injekce A do 2^A, a s ohledem na Cantorovu-Bernsteinovu větu lze Cantorovu větu ekvivalentně formulovat takto:

  Neexistuje bijekce $f : A \to 2^A$.
- Pro konečné množiny známe, protože pak $|2^A| = 2^{|A|}$.
- Ukazuje, že existují větší a větší množiny, tedy existuje nekonečná hierarchie nekonečen.
- Víme, že $|\mathbb{N}| < |\mathbb{R}|$. Ale ještě větší než $\mathbb{R}$ je množina $2^{\mathbb{R}}$. Ještě větší je $2^{2^{\mathbb{R}}}$ atd.
- S tím souvisí slavná tzv. hypotéza kontinua:

  Neexistuje množina A, pro kterou $|\mathbb{N}| < |A| < |\mathbb{R}|$.

Důkaz Cantorovy věty. Protože $a \mapsto \{a\}$ je injekce A do 2^A, stačí dokázat, že neexistuje surjekce $f : A \to 2^A$. To dokážeme sporem.

Předpokládejme, že $f : A \to 2^A$ je surjekce. Uvažujme množinu $D \in 2^A$ definovanou následovně:

$$a \in D, \text{ právě když } a \notin f(a).$$

(např. pro $A = \mathbb{N}$, $a = 2$, pro $f(a) = \{1, 3, 5\}$ je $2 \notin D$, pro $f(a) = \mathbb{N}$ je $a \in D$)

f je surjekce, tedy existuje $d \in A$ tak, že $f(d) = D$.

Pak

$$d \in D, \text{ právě když } d \notin f(d), \text{ p.k. } d \notin D,$$

což je spor. □

Myšlenka důkazu je obměnou dříve uvedené diagonální metody. Srovnejte tento důkaz s předchozím důkazem tvrzení, že $2^{\mathbb{N}}$ je nespočetná (z toho také plyne $|\mathbb{N}| < |2^{\mathbb{N}}|$; proč?).

Shrnutí

Nekonečné množiny vystupují v mnoha úvahách o základních otázkách matematiky a informatiky. Uvedli jsme základní pojmy související s nekonečnými množinami, důležité příklady a základní vlastnosti nekonečných množin.

Pojmy k zapamatování

- konečná množina, nekonečná množina
- spočetná množina
- nespočetná množina
- kardinalita množiny

Kontrolní otázky

1. *Existují větší množiny než je množina* $\mathbb{N}$*?*
2. *Existují větší množiny než je množina* $\mathbb{R}$*?*
3. *Existuje největší množina?*

8 Indukce a rekurze

Studijní cíle: Student se seznámí se základy indukce a rekurze, zejména s důkazy matematickou indukcí, definicemi matematickou indukcí a strukturální indukcí.

Klíčová slova: rekurze, rekurzívní definice, princip indukce, matematická indukce, strukturální indukce

Indukce a rekurze jsou důležité jevy, které se přirozeně objevují v mnoha úvahách v matematice a informatice. Už na střední škole se probírá důkaz matematickou indukcí, při programování se setkáváme s rekurzívními algoritmy, mnohé pojmy mají rekurzívní definice. S rekurzí se setkáváme i v jiných oblastech, například s rekurzívními obrazci v umění nebo s rekurzívními motivy v architektuře.

Indukce a rekurze jsou důležité a navzájem provázané jevy. S nadsázkou lze říct, že tvoří dvě strany jedné mince. Na otázku, zda jde o rekurzi nebo o indukci je někdy těžko odpovědět, protože odpověď je často otázka pohledu. Zatímco pro indukci je charakteristický postup od menšího (jednoduššího) k většímu (složitějšímu), tedy přístup „zdola nahoru“ (bottom-up), pro rekurzi je to naopak, tedy přístup „shora dolů“ (top-down).

8.1 Úvodní příklady

Uvažujme následující proceduru pro výpočet faktoriálu čísla n:

```
f(n)
1  if n = 1 then return 1
2  else return n * f(n − 1)
```

Protože se v těle této procedury pro výpočet $\mathsf{f}(n)$ využívá f (na řádku 2), Řádek 1 obsahuje tzv. ukončující podmínku (také limitní podmínku); bez ní se procedura „zacyklí“. Není třeba podrobněji vysvětlovat, že podle této procedury je

$$\mathsf{f}(4) = 4 * \mathsf{f}(3) = 4 * 3 * \mathsf{f}(2) = 4 * 3 * 2 * \mathsf{f}(1) = 4 * 3 * 2 * 1.$$

Tuto proceduru lze jinak popsat následovně:

1. pro $n = 1$ je $f(n) = 1$
2. pro $n > 1$ je $f(n) = n * f(n-1)$

nebo ještě jinak:

$$f(n) = \begin{cases} 1, & \text{pokud } n = 1, \\ n * f(n-1), & \text{pokud } n > 1. \end{cases}$$

Na uvedenou proceduru se lze dívat jako na proceduru, která vychází z induktivní definice:

- faktoriál definujeme nejprve pro základní prvky (pro $n = 1$),
- pro složitější prvky ($n > 1$) definujeme faktoriál pomocí toho, co jsme definovali pro jednodušší prvky ($f(n-1)$).

Uvažme nyní následující proceduru pro výpočet faktoriálu:

```
f(n)
1  if n > 1 then return n * f(n − 1)
2  else return n * f(n − 1)
```

Od první uvedené procedury definice se liší jen v pořadí podmínek (a následných příkazů) pro hodnotu n. Zatímco první procedura má induktivní charakter (zdola nahoru), právě uvedená procedura popisuje faktoriál přístupem shora dolů.

Podívejme se na další úvodní příklady.

Příklad 8.1. Definice mocniny čísla a^n:

```
power(a, n)
1  if n = 0 then return 1
2  return a * power(a, n − 1)
```

1. pro $n = 0$ je $a^n = 1$
2. pro $n > 1$ je $a^n = a * a^{n-1}$

Všimněme si, že stejné schéma má definice mocniny R^n relace R.

Příklad 8.2. Induktivní definice množiny L všech lichých čísel:

1. pro $1 \in L$
2. pokud $n \in L$, pak $n + 2 \in L$
   (nebo úplně: pro každé $n > 1$: pokud $n \in L$, pak $n + 2 \in L$)

Matematickou indukcí lze např. dokázat, že každé $n \in L$ je liché.

Příklad 8.3. Induktivní definice množiny $A \subseteq \mathbb{N}_0 \times \mathbb{N}_0$:

1. pro $\langle 0, 0\rangle \in A$
2. pokud $\langle m, n\rangle \in A$, pak $\langle m + 5, n + 1\rangle \in A$

Je $S = \{\langle 0, 0\rangle, \langle 5, 1\rangle, \langle 10, 2\rangle, \langle 15, 3\rangle, \dots\}$

Zobecněnou matematickou indukcí lze např. dokázat, že pro každé $\langle m, n\rangle \in A$ je $m+n$ dělitelné 3.

Příklad 8.4. Definice formule výrokové logiky na množinou $V = \{p, q, r, \dots\}$ výrokových symbolů:

1. každý výrokový symbol p je formule (tzv. atomická formule);
2. jsou-li φ a ψ formule, jsou i výrazy

   $\neg\varphi$,
   $(\varphi \wedge \psi)$,
   $(\varphi \vee \psi)$,

$(\varphi \to \psi)$,
$(\varphi \leftrightarrow \psi)$

formule (tzv. složené formule).

I tato definice je příkladem induktivní definice. V tomto případě je definována jistá syntaktická struktura (množina formulí). Je to příklad tzv. strukturální indukce, která je zobecněním matematické indukce.

Příklad 8.5. Sierpinského trojúhelníky $S(n)$:

$S(1)$ $S(2)$ $S(3)$ $S(4)$ $S(5)$

Příklad 8.6. Drosteho efekt (Drosteho kakao, 1904)

8.2 Matematická indukce a důkaz matematickou indukcí

Budeme se nyní zabývat matematickou indukcí, která je všeobecně známým příkladem použití indukce, resp. matematickou indukcí v její základní podobě (více uvedeme později).

Matematická indukce umožňuje dokazovat tvrzení tvaru

$$\text{pro každé přirozené číslo } n \text{ platí } V(n),$$

kde $V(n)$ je nějaké tvrzení, které závisí na n. Příkladem takového tvrzení je:

$$\text{Pro každé přirozené číslo } n \text{ platí } 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Základem dokazování matematickou indukcí je následující tvrzení.

Věta 8.7 (princip indukce)**.** *Nechť je pro každé $n \in \mathbb{N}$ dáno tvrzení $V(n)$. Předpokládejme, že platí*

(a) $V(1)$ (indukční předpoklad),

(b) pro každé $n \in \mathbb{N}$: z $V(n)$ plyne $V(n+1)$ (indukční krok).

Pak $V(n)$ platí pro každé $n \in \mathbb{N}$.

Princip indukce je jednou ze základních vlastností přirozených čísel. Lze ho dokázat z následující vlastnosti $\mathbb{N}$:

$$\text{Každá neprázdná podmnožina } K \subseteq \mathbb{N} \text{ má nejmenší prvek.} \tag{29}$$

Tato vlastnost se zdá být intuitivně zřejmá. Jak ale víme, že každá neprázdná podmnožina $\mathbb{N}$ má nejmenší prvek? To vede k otázce, co jsou vlastně přirozená čísla. Tím se zatím dále zabývat nebudeme. Pro naše potřeby pouze zmiňme, že množina přirozených čísel se v moderní matematice definuje axiomaticky pomocí tzv. Peanových axiomů.

Důkaz principu indukce z vlastnosti (29). Sporem: Předpokládejme, že princip indukce neplatí, tj. existuje tvrzení $V(\cdot)$, splňující

(a) $V(1)$,

(b) pro každé $n \in \mathbb{N}$: z $V(n)$ plyne $V(n+1)$,

ale pro nějaké $n' \in \mathbb{N}$ tvrzení $V(n')$ neplatí.

Označme

$$K = \{m \in \mathbb{N} \mid V(m) \text{ neplatí}\}$$

K je neprázdná (neboť $n' \in K$).

K má tedy nejmenší prvek ka ten je různý od 1 (protože $V(1)$ platí).

Pak tedy $k-1 \notin K$, tedy $V(k-1)$ platí.

Z indukčního kroku plyne, že platí i $V(k)$, tedy $k \notin K$, což je spor s $k \in K$. □

Ve zbytku této části ukážeme několik příkladů použití důkazu matematickou indukcí.

Příklad 8.8. Dokažme, že pro každé $n \in \mathbb{N}$ je

$$\text{pro každé } n \in \mathbb{N} \text{ je } 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

Tedy $V(n)$ je tvrzení $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$.

(a) Indukční předpoklad: $V(1)$ je tvrzení $1 = \frac{1\cdot(1+1)}{2}$, což platí.

(b) Indukční krok: dokázat, že z $V(n)$ plyne $V(n+1)$.

Tedy dokázat, že z $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$ plyne $1 + 2 + \cdots + (n+1) = \frac{(n+1)(n+2)}{2}$

$$1 + \cdots + n + (n+1) = (1 + \cdots + n) + (n+1) = \text{(dle indukčního předpokladu } V(n))$$
$$\frac{n(n+1)}{2} + n + 1 = \frac{n(n+1) + 2(n+1)}{2} = \frac{n^2 + 3n + 2}{2} = \frac{(n+1)\cdot(n+2)}{2}.$$

Příklad 8.9. Dokažte indukcí, že $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$.

$V(n)$ je tvrzení $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$.

$V(1)$ platí, protože je to tvrzení $1^2 = \frac{1(1+1)(2+1)}{6}$.

Předpokládejme, že platí $V(n)$ a dokažme $V(n+1)$, tj. dokažme

$$\sum_{k=1}^{n+1} k^2 = \frac{(n+1)(n+2)(2(n+1)+1)}{6}.$$

Je

$$\begin{aligned}\sum_{k=1}^{n+1} k^2 &= \sum_{k=1}^{n} k^2 + (n+1)^2 = \frac{n(n+1)(2n+1)}{6} + \frac{6(n+1)^2}{6} \\ &= \frac{2n^3+9n^2+13n+6}{6} = \frac{(n+1)(n+2)(2(n+1)+1)}{6}.\end{aligned}$$

Následující příklad ukazuje, že při používání důkazu matematickou indukcí musíme být obezřetní.

Příklad 8.10. Dokažte, že pro každou posloupnost n prvků $a_1, \ldots, a_n$ platí, že všechny prvky v ní jsou stejné.

Důkaz matematickou indukcí:

Indukční předpoklad: Pro číslo 1 je tvrzení triviálně splněno.

Indukční krok: Předpokládejme, že tvrzení platí pro k prvků. Uvažujme posloupnost libovolných $k+1$ prvků $a_1, \ldots, a_{k+1}$.

Pak $a_1, \ldots, a_k$ je posloupnost k prvků a $a_2, \ldots, a_{k+1}$ je posloupnost k prvků, a podle předpokladu tedy $a_1 = \cdots = a_k$ a $a_2 = \cdots = a_{k+1}$. Odtud plyne $a_1 = \cdots = a_{k+1}$.

Tvrzení je ovšem zjevně nepravdivé. Kde je tedy v předloženém důkazu chyba?

Neplatí, že z $V(1)$ plyne $V(2)$ (z toho, že v posloupnosti a_1 jsou prvky stejné a že v posloupnosti a_2 jsou prvky stejné, plyne, že v posloupnosti a_1, a_2 jsou všechny prvky stejné). Indukční krok tedy nelze ověřit.

Odbočka: o přirozených číslech

- S přirozenými čísly pracujeme jako se známou strukturou, tj. $\langle \mathbb{N}, +, \cdot, \leq \rangle$.

  Předpokládáme vlastnosti, které známe.

- Otázky jako „proč platí princip indukce?“ nebo „proč má každá $A \subseteq \mathbb{N}$ nejmenší prvek“ vedou k otázce:

Co jsou vlastně přirozená čísla?

- Jsou budována axiomaticky, tj. je to struktura splňující jisté axiomy.
- Za základní (nedefinovatelné) jsou považovány:
  - Konstanta 1.
    (alternativně 0 míst 1; pak je i 0 považována za přirozené číslo; je to otázka vkusu a technické výhodnosti)

  - Unární operace S. $S(x)$ je interpretován jako následovník čísla x.
    (tedy $S(1) = 2$, $S(2) = 3$, $S(3) = 4$, ...)
- Přirozená čísla jsou pak definována jako struktura $\langle \mathbb{N}, 1, S \rangle$, kde $1 \in \mathbb{N}$ a $S : \mathbb{N} \to \mathbb{N}$, splňující tzv. Peanovy axiomy.

Následující výčet uvádí některé z Peanových axiomů:

- ...
- pokud $m \neq n$, pak $S(m) \neq S(n)$
- neexistuje $n \in N$ tak, že $S(n) = 1$
  (1 není následovníkem žádného přirozeného čísla)
- Pokud $K \subseteq \mathbb{N}$ je množina splňující
  - $1 \in K$
  - pro každé $n \in \mathbb{N}$: pokud $n \in K$, pak $n + 1 \in K$,

  pak $K = \mathbb{N}$.

Poznamenejme, že další operace a relace (např. $+$, $\cdot$, $\leq$) jsou definované jako odvozené. Operace $+$ je např. definována následovně:

- $n + 1 = S(n)$
- $n + S(m) = S(n + m)$

Všimněme si, že poslední uvedený Peanův axiom

- Pokud $K \subseteq \mathbb{N}$ je množina splňující
  - $1 \in K$
  - pro každé $n \in \mathbb{N}$: pokud $n \in K$, pak $n + 1 \in K$,

  pak $K = \mathbb{N}$.

říká právě to, co princip indukce (věta 8.7). Skutečně: položíme-li v tom principu $K = \{n \mid V(n) \text{ platí}\}$, pak tvrzení principu je shodné s tvrzením axiomu.

Tedy:

- Při tomto pohledu (axiomatickém) na přirozená čísla tedy není třeba princip důkazu indukcí dokazovat (je to axiom).
- Proč jsme ho tedy dokazovali?
  - Protože s přirozenými čísly pracujeme intuitivně, jako se strukturou splňující známé vlastnosti.
  - Jednou z nich je: Každá $A \subseteq \mathbb{N}$ má nejmenší prvek (vzhledem k $\leq$).
    Jinými slovy: $\langle \mathbb{N}, \leq \rangle$ je dobře uspořádaná.
  - Tu jsme použili v důkazu principu.

Důležité jsou v tomto kontextu následující tvrzení (první z nich už známe):

- Každá $A \subseteq \mathbb{N}$ má nejmenší prvek $\Rightarrow$ princip indukce.
- Princip indukce $\Rightarrow$ každá $A \subseteq \mathbb{N}$ má nejmenší prvek.

Varianty důkazu matematickou indukcí

Indukce nemusí začínat číslem 1. Místo $K = \{1, 2, \dots\}$ můžeme vlastnost V indukcí dokázat pro $K = \{4, 5, 6, \dots\}$, $K = \{-3, -2, -1, 0, 1, \dots\}$ apod.

Věta 8.11 (začátek v k)**.** *Nechť $k \in \mathbb{Z}$, $K = \{k, k+1, k+2, \dots\}$ a pro každé $n \in K$ je dáno tvrzení $V(n)$. Předpokládejme, že platí*

(a) $V(k)$ (indukční předpoklad),

(b) pro každé $n \in K$: z $V(n)$ plyne $V(n+1)$ (indukční krok).

Pak $V(n)$ platí pro každé $n \in K$.

Důkaz. Plyne ze základního principu indukce:

Uvažujme tvrzení $W(n)$ pro $n = 1, 2, 3, \dots$ definované:

$$W(n) = V(n + (k-1)).$$

Pak $W(1)$ je $V(k)$, $W(2)$ je $V(k+1)$, atd.

Podmínky (a) a (b) výše jsou pak podmínky $W(1)$ a $W(n) \Rightarrow W(n+1)$ ze základního principu.

Dle základního principu tedy platí $W(1)$, $W(2)$, ...

Tedy platí $V(1)$, $V(2)$, ... □

Příklad 8.12. Zobecněte vzorec

$$1 + 2 + \cdots + n = \frac{n \cdot (n+1)}{2}$$

pro

$$k + (k+1) + \cdots + n$$

Zobecnění se nejdřív pokusíme odhadnout. V uvedeném vzorci $1+2+\cdots+n = \frac{n\cdot(n+1)}{2} = \frac{n\cdot(n+1)}{2}$ je n počet sčítanců a $n+1$ je součet prvního a posledního sčítance. Ověříme tedy, zda platí

$$k + (k+1) + \cdots + n = \frac{(n-k+1)\cdot(n+k)}{2}.$$

Např. pro $k = 2$ a $n = 6$ je

$$k + (k+1) + \cdots + n = 2 + 3 + 4 + 5 + 6 = 20 = \frac{5 \cdot (6+2)}{2},$$

pro $k = -2$ a $n = 1$ je

$$k + (k+1) + \cdots + n = -2 - 1 + 0 + 1 = -2 = \frac{4 \cdot (-1)}{2}.$$

Důkaz indukcí v počátkem v $k = -2$ vypadá takto:

Indukční předpoklad: Pro $n = k$ vzorec zřejmě platí.

Indukční krok: předpokládejme, že vzorec platí pro n. Pro $n+1$ s využitím vzorce pro n dostaneme

$$\begin{aligned} k + \cdots + n + (n+1) &= \frac{(n-k+1)\cdot(n+k)}{2} + (n+1) \\ = \quad \frac{(n-k+1)\cdot(n+k) + 2n + 2}{2} &= \frac{n^2 + nk - nk - k^2 + n + k + 2n + 2}{2}. \end{aligned}$$

Úpravou pravé strany vzorce pro součet $k + \cdots + n + (n+1)$ však dostaneme stejnou hodnotu:

$$\frac{((n+1)-k+1)\cdot((n+1)+k)}{2} = \frac{(n-k+2)\cdot((n+k+1)}{2}$$
$$= \frac{n^2+nk+n-nk-k^2-k+2n+2k+2}{2} = \frac{n^2+nk-nk-k^2+n+k+2n+2}{2}.$$

Věta 8.13 (více předpokladů, tzv. silný princip indukce)**.** *Nechť je pro každé $n \in \mathbb{N}$ dáno tvrzení $V(n)$. Předpokládejme, že platí*

(a) $V(1)$ (indukční předpoklad),

(b) pro každé $n \in K$: z $V(1), \ldots, V(n)$ plyne $V(n+1)$ (indukční krok).

Pak $V(n)$ platí pro každé $n \in \mathbb{N}$.

Důkaz. Snadné ze základního principu indukce:

Uvažujme nové tvrzení W, které je pro každé $n \in \mathbb{N}$ definováno následovně:

$$W(n) \text{ platí, právě když platí } V(1), \ldots, V(n).$$

Dokážeme-li tedy výše uvedené body (a) a (b), dokážeme vlastně, že platí

(a) $W(1)$,

(b) pro každé $n \in K$: z $W(n)$ plyne $W(n+1)$.

Dle principu indukce tedy platí $W(n)$ pro každé $n \in \mathbb{N}$. To ale znamená, že i $V(n)$ platí pro každé $n \in \mathbb{N}$. □

Silný princip indukce platí pro začátek v libovolné hoodnotě $k \in \mathbb{N}$ (ne nutně $k = 1$); zformulujte a dokažte ho sami.

Naopak zřejmě platí, že základní princip indukce plyne ze silného principu.

Příklad 8.14. Dokažte, že každé $n \in \mathbb{N}$ má prvočíselný rozklad.

Silným principem indukce se začátkem v $k = 2$.

indukční předpoklad: $V(2)$

Platí, protože 2 je prvočíslo, tedy $2 = 2^1$ je požadovaný rozklad.

indukční krok: z $V(2)$, …, $V(n)$ plyne $V(n+1)$

Buď je $n+1$ prvočíslo, pak rozklad je $n+1 = (n+1)^1$,

nebo n je složené, tedy $n = r \cdot s$, kde $2 \leq r, s < n$.

Pak dle $V(r)$ a $V(s)$ existují rozklady $r = p_1^{n_1} \cdots p_k^{n_k}$ a $s = q_1^{m_1} \cdots q_l^{m_l}$, a tedy požadovaný rozklad je

$$n+1 = p_1^{n_1} \cdots p_k^{n_k} \cdot q_1^{m_1} \cdots q_l^{m_l}.$$

8.3 Definice matematickou indukcí

Vraťme se k výše uvedené definici faktoriálu:

1. pro $n = 1$ je $f(n) = 1$
2. pro $n > 1$ je $f(n) = n * f(n-1)$

Intuitivně je jasné, že tímto způsobem je jednoznačně definována jistá funkce (té pak říkáme faktoriál). Z čeho ale plyne že funkce splňující podmínky 1 a 2 z uvedené definice existuje a je určena jednoznačně?

Základem je následující věta, kterou uvedeme bez důkazu.

Věta 8.15. *Nechť je dána množina V, prvek $a \in V$ a funkce $G : \mathbb{N} \times V \to V$. Pak existuje právě jedna funkce*

$$F : \mathbb{N} \to V,$$

pro kterou platí

- $F(1) = a$,
- *pro každé $n \in N$:* $F(n+1) = G(n, F(n))$.

Příklad 8.16. Pro definici faktoriálu $f(n)$ výše je:

$$F = f, \quad V = \mathbb{N}, \quad a = 1, \quad G(m,n) = (m+1) \cdot n.$$

Příklad 8.17. Uvažujme předpis:

1. pro $n = 0$ je $f(n) = 1$
2. pro $n > 0$ je $f(n+1) = -f(n)$

Dle věty (resp. její modifikace pro $\mathbb{N}_0$) je jím jednoznačně určena funkce $f : \mathbb{N}_0 \to \mathbb{Z}$.

($V = \mathbb{Z}$, $a = 1$, $G(m,n) = -F(n)$)

Snadno se vidí, že $f(n) = (-1)^n$.

Příklad 8.18 (zobecněná definice indukcí, dle modifikace uvedené věty)**.** Uvažujme předpis:

1. $f(1) = 3$, $f(2) = 5$,
2. $f(n+1) = 2f(n) - f(n-1)$.

Dle věty (resp. její modifikace pro více předpokladů) je jím jednoznačně určena funkce $f : \mathbb{N} \to \mathbb{Z}$.

Snadno se dokáže, že $f(n) = 2n + 1$.

8.4 Strukturální indukce

Strukturální indukce je zobecněním matematické indukce. Místo množiny $\mathbb{N}$, se kterou pracuje matematická indukce, pracuje strukturální indukce s množinou T jistých objektů. Základní myšlenky strukturální indukce jsou následující:

- T je zpravidla množina řetězců utvořených podle induktivních pravidel. Například formule:
  - (bazické/atmoické) $p \in T$ pro každý výrokový symbol p;
  - (složené) pokud $\varphi, \psi \in T$, pak $\neg\varphi \in T$, $(\varphi \wedge \psi) \in T$, $(\varphi \vee \psi) \in T$, $(\varphi \to \psi) \in T$ a $(\varphi \leftrightarrow \psi) \in T$.
- Důkaz strukturální indukcí: Že tvrzení V platí pro všechny prvky $t \in T$ se dokáže takto (uvedeme to na výše uvedeném příkladu formulí):
  - indukční předpoklad: V platí pro všechny atomické (bázické) $t \in T$
  - indukční krok: pokud V platí pro φ a ψ, pak V platí i pro $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \to \psi)$ a $(\varphi \leftrightarrow \psi)$.

Příklad 8.19. Dokažte strukturální indukcí, že v každé formuli výrokové logiky je počet levých i pravých závorek stejný.

Označme $l(\varphi)$ a $r(\varphi)$ počty levých a pravých závorek ve φ.

indukční předpoklad: $l(p) = 0 = r(p)$ zřejmě platí.

indukční krok: (uvedeme jen pro např. pro $\wedge$):

pookud $l(\varphi) = r(\varphi)$ a $l(\psi) = r(\psi)$

pak zřejmě

$$l((\varphi \wedge \psi)) = 1 + l(\varphi) + l(\psi) = 1 + r(\varphi) + r(\psi) = r((\varphi \wedge \psi)).$$

Definice strukturální indukcí

Definice strukturální indukcí je zobecněním definice matematickou indukcí. Chceme definovat nějaký objekt pro každý prvek z množiny T. To uděláme následovně:

- definujeme pro bazické prvky T,
- definujeme pro složené prvky T.

Příklad 8.20. Strukturální indukcí budeme definovat počet $\#\varphi$ výskytů výrokových proměnných ve formuli φ. Tedy chceme, aby např. $\#(p \wedge (\neg p \vee q)) = 3$.

Definice strukturální indukcí vypadá následovně:

- $\#p = 1$,
- $\#(\neg\varphi) = \#\varphi$,
  $\#(\varphi \wedge \psi) = \#\varphi + \#\psi$,
  ...

V prvním bodě jsme tedy definovali $\#\varphi$ pro bazické prvky (tedy pro výrokové symboly p); ve druhém bodě pak pro složené prvky z T (tedy pro složené formule, tj. pro $\neg\varphi$, $\varphi \wedge \psi$, atd.).

Následující výčet uvádí další množiny T struktur, pro které lze použít strukturální indukci:

- aritmetické výrazy nad $X = \{x, y, \dots\}$ a funkčními symboly $+$ a $\cdot$
  - pokud $t \in X$, pak $t \in T$ (atomický),
  - pokud $t_1, t_2 \in T$, pak $(t_1 + t_2) \in T$ a $(t_1 \cdot t_2) \in T$ (složený)
- rekurzívně definované seznamy,
- rekurzívně definované stromy,
- ...
- přirozená čísla jsou speciálním případem:
  - $1 \in T$ (první číslo),
  - pokud $t \in T$, pak $S(t) \in T$ (následovník)

Shrnutí

Indukce patří k důležitým způsobům definice a a dokazování vlastností různých rekurzívních struktur. Tyto struktury mají rekurzívní charakter. Zobecněním klasické matematické indukce je strukturální indukce.

Pojmy k zapamatování

- matematická indukce
- definice indukcí
- strukturální indukce

Kontrolní otázky

1. *Jaké znáte varianty základního principu indukce?*
2. *Jak lze princip důkazu matematickou indukcí dokázat? Co jsou Peanovy axiomy?*

Cvičení

1. Dokažte indukcí, že součet prvních n lichých čísel je n^2, tj.
$$1 + 3 + 5 + \cdots + (2n - 1) = n^2$$
2. Dokažte indukcí, že $\sum_{k=1}^{n} k^3 = [\frac{n(n+1)}{2}]^2$.
3. Dokažte indukcí, že pro $n \in \mathbb{N}$ je $2^{n+2} + 3^{2n+1}$ dělitelné 7.
4. Dokažte, že pro $n \in \mathbb{N}$ je $(1 + \frac{1}{3})^n \geq 1 + \frac{n}{3}$.

Úkoly k textu

1. úkol k textu 1

Řešení

1. Jednoduché, standardně použitím jednoduchých úprav.
2. Jednoduché, standardně použitím jednoduchých úprav.
3. Jednoduché, standardně použitím jednoduchých úprav.
4. Jednoduché, standardně použitím jednoduchých úprav.

Reference

[Goo98] Goodaire E. G., *Parmenter M. M.: Discrete Mathematics with Graph Theory.* Prentice-Hall, Inc., 1998.

[Gri99] Grimaldi R.: *Discrete and Combinatorial Mathematics. An Applied Introduction. 4th ed.* Addison Wesley, Reading, MA, 1999.

[KlYu95] Klir G. J., Yuan B.: *Fuzzy Sets and Fuzzy Loic. Theory and Applications.* Prentice Hall, Upper Saddle River, NJ, 1995.

[MaNe00] Matoušek J., Nešetřil J.: *Kapitoly z diskrétní matematiky.* Karolinum, Praha, 2000.

[Mau91] Maurer S. B., Ralston A.: *Discrete Algorithmic Mathematics. Second edition.* A K Peters, Natick, MA, 1998.

[PrYe73] Preparata F. P., Yeh R. T.: *Introduction to Discrete Structures. For Computer Science and Engineering.* Addison Wesley, Reading, MA, 1973.

[Soch01] Sochor A.: *Klasická matematická logika.* Karolinum, Praha, 2001 (v prodeji, velmi dobře psaná s řadou doplňujících informací).

[Šve02] Švejdar V.: *Logika, neúplnost a složitost.* Academia, Praha, 2002.

[Vau85] Vaught R. L.: *Set theory: An introduction.* Birkhäuser, Boston, 1985.

[Vil77] Vilenkin N. J.: *Kombinatorika.* SNTL, Praha, 1977.