

MASARYKOVA UNIVERZITA

FAKULTA INFORMATIKY

# PB016 Umělá inteligence

Sbírka příkladů

Semestr Podzim 2023

Matěj Pavlík, Terézia Mikulová, Ondřej Huvar, Roman Solař

Chyby hlašte na: [469088@mail.muni.cz](mailto:469088@mail.muni.cz)

## Jak se sbírkou pracovat

Sbírka je rozdělena do dvanácti kapitol, které odpovídají tematickým celkům postupně probíraným na přednáškách a cvičení. Každá kapitola je **rozdělena do podkapitol**.

**Před cvičením** si ve sbírce prolistujte příslušnou kapitolu. Pro úspěšné složení odpovědníků potřebujete zhlédnout přednášku a porozumět všem částem kapitoly, které jsou označeny svíslou čarou po levém boku (jako tento odstavec). Pokud pro vás budou surové definice těžko stravitelné, může být pro pochopení jednodušší si přečíst i „omáčku“ okolo, která není zvýrazněná, ale umožňuje pojmům pomocí příkladů lépe porozumět.

**Na cvičení** se budete ještě před implementačními úlohami věnovat příkladům ve sbírce, které jsou označeny hvězdičkou ★. V případě cvičení na logiku se budete příkladům ze sbírky věnovat celou dobu cvičení.

**Po cvičení** můžete své znalosti látky ověřit řešením dalších příkladů ze sbírky, případně rozšířit a doplnit znalosti z přednášky pročitáním doprovodného textu. Podobný postup je doporučen i pro samotnou přípravu na zkoušku.

**Příklad.** Ve sbírce se nacházejí příklady dvou typů – číslované a nečíslované. Nečíslované příklady jsou uvedeny vždy v úvodu k dané podkapitole a slouží k demonstraci či ilustraci probíraného konceptu na konkrétní situaci. Číslované příklady jsou pak uvedeny za vodorovnou dělicí čarou a jsou určeny ke společnému řešení ve cvičení (jsou-li označeny hvězdičkou ★), případně k samostatnému vypracování.

V zelených rámečcích naleznete řešení příkladů. U nečíslovaných příkladů se zobrazují řešení vždy, u číslovaných příkladů hledejte řešení ve verzi sbírky s klíčem.

V modrých rámečcích jsou ve sbírce vyobrazeny definice pojmů.

# Obsah

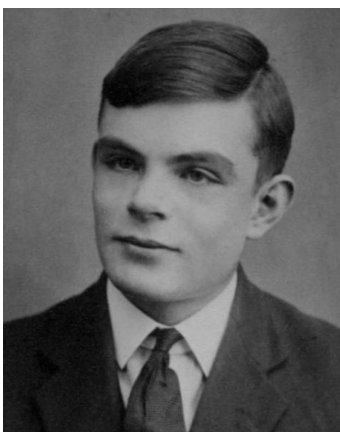
1 Úvod do umělé inteligence	1
2 Prohledávání stavového prostoru	9
3 Dekompozice problému, Problémy s omezujícími podmínkami	27
4 Hry a herní strategie	44
5 Výroková logika	65
6 Predikátová logika	83
7 Důkazové systémy a rezoluce	101
8 Neklasické logiky	128
9 Reprezentace a vyvozování znalostí	149
10 Strojové učení	174
11 Neuronové sítě a hluboké učení	194
12 Zpracování přirozeného jazyka	207

# 1 Úvod do umělé inteligence

Úvodní kapitola nabízí čtenáři sbírky první setkání s umělou inteligencí. Představuje nejprve Turingův test k určení schopnosti stroje chovat se inteligentně a příklady k zamyšlení nad hranicemi a možnostmi oboru umělé inteligence jako takového (v této jediné části sbírky jsou příklady ponechány bez řešení). Další část pak prezentuje pojem *problému* a na řadě příkladů ilustruje, jak může volba vhodné datové struktury či strategie k řešení daného úkolu ovlivnit samotnou možnost úkol vyřešit.

## 1.1 Umělá inteligence a Turingův test

V roce 1950 navrhnul Alan Turing tzv. *Turingův test* (který původně sám nazval „imitační hrou“, viz původní Turingův článek), jež si klade za cíl ověřit schopnost stroje vykazovat inteligentní chování.



Alan Turing ve věku 16 let

Turing v článku nejprve uvádí, že navrhuje uvážit otázku, zda stroje umí myslet. Jelikož však není snadné definovat, co znamená „myslet“, navrhuje nahradit tuto poněkud vágní otázku jinou, a sice zda „Lze sestrojít počítač, který by byl schopen složit Turingův test?“

**Definice 1:** *Turingův test* je navržen jako hra tří hráčů, z nichž jeden je stroj podrobený zkoušce a ostatní dva jsou lidé. V základní variantě Turingova testu jsou do různých místností umístěni A) testovaný stroj, B) člověk, C) rozhodčí. Rozhodčí může komunikovat se zbylými dvěma formou textových zpráv a má za úkol zjistit, který z nich je stroj a který je člověk. Stroj uspěl v Turingově testu, nedokáže-li ho rozhodčí spolehlivě rozeznat od člověka.

Složení Turingova testu počítačem vyžaduje zvládnutí několika oblastí oboru. Pro porozumění zpráv a jejich generování je potřeba *zpracování přirozeného jazyka (NLP)*. Pro uchování údajů a vyvozování závěrů je nutné ovládnout *repräsentaci a vyvozování znalostí*. V neposlední řadě je potřeba metod *strojového učení*, které umožňují učit se a adaptovat se na změny vnějšího prostředí.

Na druhou stranu není pro složení testu nutné zvládnout obory jako *robotická manipulace* či *počítačové vidění*, jelikož veškerá komunikace probíhá výhradně textovými zprávami.

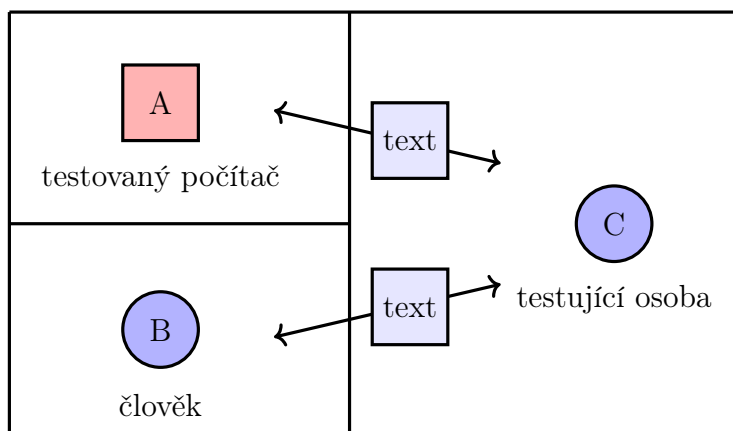


Schéma uspořádání Turingova testu

**Příklad 1.1.1.** Definujte vlastními slovy následující pojmy:

- inteligence,
- umělá inteligence,
- agent,
- racionalnost,
- logické usuzování.

Nalezněte možné námitky proti uvedeným definicím a zpřesněte je.

**Příklad 1.1.2.** Jsou reflexy racionální? Jsou inteligentní? Zdůvodněte.

**Příklad 1.1.3.** Do jaké míry jsou následující počítačové systémy příklady umělé inteligence?

- Čtečka čárových kódů v supermarketu.
- Internetové vyhledávače.
- Hlasové zadávání příkazů telefonu.
- Síťové směrovací algoritmy, které dynamicky reagují na stav sítě.

**Příklad 1.1.4.** Přečtěte si původní článek Alana Turinga *Computing Machinery and Intelligence*. Turing v sekci 6 ve článku rozebírá různé námitky vznesené vůči Turingovu testu. Které z námitek jsou stále platné? Napadnou vás nové námitky, které vzešly z pozdějšího vývoje v oboru?

**Příklad 1.1.5.** Upravíme-li Evansův program pro řešení problémů s geometrickou analogií, aby získal v standardním IQ testu hodnotu 200, jednalo by se o program inteligentnější než člověk? Vysvětlete.

**Příklad 1.1.6.** Alan Turing ve své práci představil seznam věcí, které by stroje nemusely nikdy umět: být milé, být nápadité, být krásné, být přátelské, být iniciativní, mít smysl pro humor, rozeznat správné od špatného, dělat chyby, zamilovat se, vychutnat si jahody se

šlehačkou, okouzlit někoho, poučit se ze zkušenosti, používat správná slova, přemýšlet samy o sobě, mít chování rozmanité jako člověk, provést něco skutečně nového. Kterých se již podařilo dosáhnout? Které jsou v principu dosažitelné počítačem? Které z nich jsou stále problematické, protože vyžadují vědomé mentální stavy?

**Příklad 1.1.7.** Argument čínského pokoje je myšlenkový experiment, který se snaží prokázat, že počítače nemohou mít „mysl“, „chápání“ či „vědomí“, a to nezávisle na tom, jak inteligentní chování vykazují. Nastudujte si o argumentu čínského pokoje více. Znamená vyvrácení argumentu čínského pokoje, že vhodně naprogramované počítače mohou mít mentální stavy? Plyne z přijetí argumentu, že počítače mentální stavy mít nemohou?

## 1.2 Řešení problémů

V této sekci se poprvé setkáme s některými problémy a úvodem do strategií, které lze uplatnit při jejich řešení. K některým z nich se budeme v dalších kapitolách vracet při probírání konkrétních metod řešení problémů (jako například heuristické prohledávání).

Problémem rozumíme zadání množiny konfigurací (např. přiřazení čísel volným políčkům sudoku), mezi nimiž hledáme konfigurace, které splňují konkrétní vlastnosti (např. vlastnost, že je sudoku vyplněno správně). Alternativou k hledání konfigurací splňujících nějakou vlastnost je hledání konfigurací, které jsou v jistém smyslu „nejlepší“. V takovém případě hovoříme o *optimalizačním problému*. Co přesně rozumíme problémem, může být navíc v konkrétním kontextu blíže upřesněno.

**Definice 2:** Při řešení *optimalizačního problému* je cílem nalézt mezi jeho konfiguracemi takovou, která je mezi všemi ostatními nejlepší podle předem daného kritéria.

Plánujeme-li si semestr tak, abychom získali co nejvíce kreditů a zároveň nám zůstalo co nejvíce volného času na své koníčky a kamarády, řešíme optimalizační problém. Naopak snažíme-li se naskládat nákup do ledničky tak, aby se dala zavřít, problém optimalizační neřešíme, protože jsme spokojeni s jakýmkoliv řešením, které nám umožní dovřít dveře ledničky.

Při řešení problémů hraje důležitou roli počet prohledávaných konfigurací. Přirozenou snahou je tento počet co nejvíce zredukovat, čehož lze dosáhnout využitím znalostí o problému a vhodnou volbou datové struktury.

**Příklad.** Uvažte *problém  $n$  dam*. V problému  $n$  dam je rozmístováno  $n$  dam na šachovnici o rozměru  $n \times n$ . Řešením jsou taková rozmístění (všech dam), v nichž se žádná dvojice dam neohrožuje na řádce, ve sloupci ani diagonálně.

- a) Spočítejte počet různých konfigurací problému, tj. počet různých rozmístění dam na šachovnici.

- b) Navrhněte datovou strukturu reprezentující konfigurace problému.
- c) Navrženou datovou strukturu upravte tak, aby svým návrhem omezovala počet různých konfigurací, ale ne počet potenciálních řešení. Vyjděte z povahy problému a promítněte požadavky na řešení problému do návrhu datové struktury. Spočítejte nový počet konfigurací.

- a) Každá dáma může být umístěna na jednu z  $n \times n = n^2$  pozic. Pro  $n$  dam tedy celkově získáme  $\underbrace{n^2 \cdot \dots \cdot n^2}_n = n^{2n}$  rozmístění. Pro 8 dam máme  $8^{16} \approx 2,8 \cdot 10^{14}$  konfigurací.
- b) Na reprezentaci souřadnic jedné dámy lze použít dvojici  $(x, y)$ . Konfiguraci pak reprezentuje pole  $n$  takových dvojic  $[(x_1, y_1), \dots, (x_n, y_n)]$ .
- c) Jistě nelze umístit více dam na jedno políčko šachovnice. Můžeme tedy předpokládat unikátnost souřadnic a místo pole použít množinu souřadnic  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ . Počet konfigurací v takovém případě je  $n^2 \cdot (n^2 - 1) \cdot \dots \cdot (n^2 - n + 1) = \frac{(n^2)!}{(n^2 - n)!}$ . Konkrétně pro 8 dam je to  $64 \cdot 63 \cdot \dots \cdot 57 \approx 1,8 \cdot 10^{14}$  konfigurací.

V návrhu lze jít ještě dále. Z povahy problému víme, že v každém sloupci bude právě jedna dáma – jinak by v nějakém sloupci byly alespoň dvě a ty by se vzájemně ohrožovaly. Dámám v seznamu lze tedy postupně přiřadit  $x$ -ovou souřadnici 1 až  $n$ :  $[(1, y_1), \dots, (n, y_n)]$  a celý seznam zredukovat na seznam pouze  $y$ -ových souřadnic:  $[y_1, \dots, y_n]$ . Pokud nepřipustíme opakování hodnot v seznamu (čili více dam na řádku), omezíme počet možných konfigurací na  $n \cdot (n - 1) \cdot \dots \cdot 1 = n!$ . Pro 8 dam je to  $8! = 40320$  konfigurací.

Příklad demonstruje zásadní vliv promítnutí znalostí o problému do přístupu k jeho řešení. Zatímco naivní řešení problému  $n$  dam vyžaduje pro  $n = 8$  prohledávání prostoru přibližně  $2,8 \cdot 10^{14}$  konfigurací, vhodnou úpravou strategie lze číslo redukovat na 40320. Už pro relativně malá čísla  $n$  tedy volba strategie rozhoduje o tom, zda je problém prakticky řešitelný či ne.

---

**Příklad 1.2.1.** ★ Pro uvedné problémy spočítejte počet různých konfigurací, tj. potenciálních řešení. Pro každý problém navrhněte datovou strukturu pro reprezentaci konfigurací problému.

- a) *Sudoku.* Uvažujte obecné sudoku  $n \times n$  s  $k$  hodnotami zadanými.
- b) *Problém přiřazení.* V problému přiřazení je dána množina  $n$  úkolů  $T$ , množina  $n$  pracovníků  $W$  a cenová funkce  $c : T \times W \rightarrow \mathbb{R}$ , která určuje náklady na vykonání úkolů jednotlivými pracovníky, tj. cena práce pracovníka  $w$  na úkolu  $t$  je dána jako  $c(t, w)$ . Cílem je rozdělit úkoly mezi pracovníky (jeden úkol na jednoho pracovníka) tak, aby celková cena práce byla co nejmenší.
- c) *Problém batohu.* Je zadána množina  $n$  položek  $1, \dots, n$  o hmotnostech  $w_1, \dots, w_n$  a hodnotách  $v_1, \dots, v_n$  a maximální nosnost batohu  $W$ . Cílem je z položek vybrat takové,

aby součet jejich hmotností nepřekročil  $W$  a zároveň byl co nejvyšší.

- a) Zbývá doplnit  $n^2 - k$  políček, každé nějakou z  $n$  různých hodnot. Celkem  $n^{n^2-k}$  konfigurací. Jako datovou strukturu lze použít matici rozměrů  $n \times n$  s číselnými hodnotami. Zadané hodnoty budou nést příznak neměnnosti (technická realizace by závisela na použitém jazyce).
- b) Každému z  $n$  pracovníků je přiřazen nějaký  $n$  úkolů, celkem  $n^n$  konfigurací. Konfigurace lze reprezentovat polem  $A$  délky  $n$  tak, že pracovníkovi  $i$  je přidělen úkol číslo  $A[i]$ . V tomto místě je dobré uvědomit si, že ne všechny konfigurace jsou přípustným řešením problémů, protože může být jeden úkol přiřazen více pracovníků. Vedle ceny přiřazení je tedy potřeba kontrolovat i přípustnost řešení.

Pokud nepřipustíme, aby se hodnoty v poli opakovaly (tj. žádný úkol není přiřazen více pracovníkům), bude pole obsahovat permutace  $n$  čísel, kterých je pouze  $n!$ .

- c) Vybrané položky by bylo možné reprezentovat přímo množinou vybraných položek  $M \subseteq \{1, \dots, n\}$ . Takových množin je  $\sum_{i=0}^n \binom{n}{i}$ .

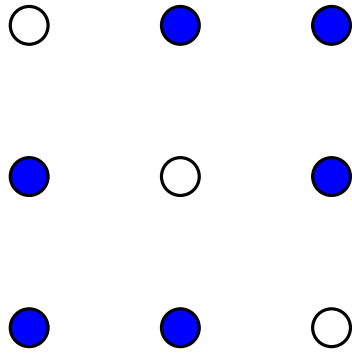
Alternativní možnost reprezentace konfigurací je pole  $n$  booleovských hodnot  $X$ , kde  $X[i]$  je TRUE, právě když je  $i$ -tá položka vybrána. Takových polí je  $2^n$ .

Znamená to, že volba jiné datové struktury v tomto případě změnila počet konfigurací? Po chvíli zamyšlení dojdeme k tomu, že jsou obě reprezentace ekvivalentní – v obou lze reprezentovat všechny konfigurace a lze mezi nimi snadno převádět. Tento fakt nás dovede k (skutečně platné) matematické identitě

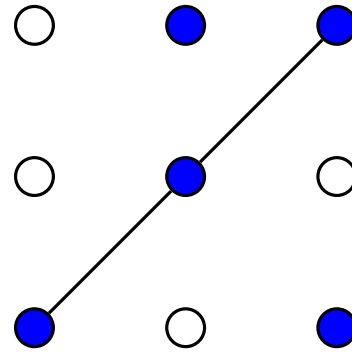
$$\sum_{i=0}^n \binom{n}{i} = 2^n.$$

**Příklad 1.2.2.** Uvažte problém *žádné tři v řadě*. V tomto problému je zadána mřížka rozměrů  $n \times n$  (kde  $n \in \mathbb{N}$ ) a cílem je zjistit, kolik do ní lze umístit bodů tak, aby žádné tři body neležely v jedné přímce.





Na mřížce  $3 \times 3$  lze rozmístit maximálně 6 bodů (zvolené body jsou vyplněny modře).



Uvedené rozmístění nesplňuje požadavek, aby tři body neležely v jedné přímce.

- Spočítejte počet různých konfigurací problému, tj. různých rozmístění bodů (i těch nesplňujících požadavek zadání).
- Navrhněte datovou strukturu reprezentující konfigurace problému.
- Navrženou datovou strukturu upravte tak, aby svým návrhem omezovala počet různých konfigurací, ale ne počet potenciálních řešení. Vyjděte z povahy problému a promítněte požadavky na řešení problému do návrhu datové struktury. Spočítejte nový počet konfigurací.
- \* Proveďte asymptotické srovnání počtu konfigurací v naivním a vylepšeném řešení.
- \* Ještě dále vylepšete návrh datové struktury, aby byl počet konfigurací asymptoticky menší než vzorové řešení části c).

- Na mřížce  $n \times n$  se nachází  $n^2$  bodů, z nichž každý může být vybrán či ne (2 možnosti). Celkem má tedy problém  $2^{n^2}$  konfigurací.
- Konfigurace lze reprezentovat maticí jedniček a nul o rozměru  $n \times n$ .
- Z povahy problému plyne, že v platném řešení na každém řádku mohou být vybrány nejvýše dva body. Datovou strukturu tedy upravíme na matici rozměrů  $2 \times n$ , v níž jsou uloženy číselné souřadnice jednotlivých bodů (pro jednoznačnost ve vzestupném pořadí), případně hodnoty NULL.

Matice reprezentující řešení z levého obrázku výše by vypadala takto.

$$\begin{pmatrix} 1 & 2 \\ 0 & 2 \\ 0 & 1 \end{pmatrix}$$

Na každém řádku může mít matice dvě hodnoty NULL (tedy 0 umístěných bodů), jednu hodnotu NULL (1 umístěný bod) nebo žádnou hodnotu NULL (tedy 2 body), což

je celkem  $1 + n + \binom{n}{2} = \frac{n^2+n+2}{2}$  možností. Jelikož má matice  $n$  řádků, je nový počet konfigurací  $\left(\frac{n^2+n+2}{2}\right)^n$ .

- d)\* Obě funkce rostou velmi rychle, takže budeme porovnávat jejich exponenty po převedení do tvaru exponenciální funkce. Platí

$$2^{n^2} \in 2^{\Theta(n^2)},$$

$$\left(\frac{n^2 + n + 2}{2}\right)^n \approx \left(\frac{n^2}{2}\right)^n = \frac{n^{2n}}{2^n} = \frac{2^{\log n^{2n}}}{2^n} = \frac{2^{2n \log n}}{2^n} = 2^{2n \log n - n} \in 2^{\Theta(n \log n)}.$$

**Příklad 1.2.3.** ★ Které z následujících problémů jsou problémy optimalizační? Zdůvodněte.

- Problém  $n$  dam.
- Sudoku.
- Problém přiřazení.
- Žádné 3 v řadě.

- Není optimalizační. Hledáme konfiguraci, která je platná, ale ne nutně „nejlepší“ mezi všemi.
- Není optimalizační.
- Je optimalizační. Hledáme řešení, pro něž nabývá cena nejmenší hodnoty mezi všemi.
- Je optimalizační. Snažíme se v mřížce vybrat co nejvíce bodů při splnění zadaných kritérií.

**Příklad 1.2.4.** Implementujte řešení problému  $n$  dam využitím naivního přístupu a přístupu omezujícího počet prohledávaných konfigurací. Experimentálně zjistěte časy nalezení všech řešení pro různá malá  $n$  oběma způsoby. Porovnejte s dříve vypočítaným prohledávaným počtem konfigurací. Lze pozorovat úměru mezi časem výpočtu a počtem konfigurací?

Tady bude tabulka a graf jak si vedou naše implementace.

**Příklad 1.2.5.** Implementujte řešení problému žádné 3 v řadě využitím naivního přístupu a přístupu omezujícího počet prohledávaných konfigurací. Experimentálně zjistěte časy nalezení všech řešení pro různá malá  $n$  oběma způsoby. Porovnejte s dříve vypočítaným prohledávaným počtem konfigurací. Lze pozorovat úměru mezi časem výpočtu a počtem konfigurací?

Tady bude tabulka a graf jak si vedou naše implementace.

## 2 Prohledávání stavového prostoru

V předchozí kapitole jsme si vyzkoušeli, jakými způsoby se dá řešit několik klasických problémů, které spadají do oblasti umělé inteligence. Obecně samozřejmě existuje velké množství různých druhů problémů, a tedy i strategií jak je řešit. Často se ale dají využít některé známé univerzální postupy.

V této kapitole si formálně představíme jednu z takových strategií, a to *prohledávání stavového prostoru*. Ukážeme si, jak správně problémy formulovat, představíme si několik různých prohledávacích algoritmů a zaměříme se i na to, v čem se jednotlivé algoritmy liší.

### 2.1 Formulace problému

Abychom problémy mohli řešit (algoritmicky) pomocí prohledávání, je potřeba je nejdříve formálně zadefinovat. Vhodná formální definice problému se pak skládá z několika hlavních komponent, kterými jsou obvykle

- *iniciální stav*,
- *přechodové akce* (a případně jejich nezáporná *cena*),
- *cílová podmínka*.

Zmíněné komponenty nám dohromady *implicitně* zadávají takzvaný *stavový prostor* problému, tedy množinu všech stavů dosažitelných z iniciálního stavu. Tyto stavy můžeme interpretovat jako vrcholy (uzly) grafu, přechodové akce pak určují hrany daného (přechodového) grafu. Díky této interpretaci můžeme řešit různorodé problémy pomocí obecných algoritmů grafového prohledávání. *Řešení problému* je obvykle posloupnost přechodových akcí, která nás dostane z iniciálního stavu do cílového – hledáme tedy cestu v přechodovém grafu. Zajímá-li nás *optimální řešení*, hledáme obvykle takovou cestu, která je nejkratší (nebo nejnějnější) ze všech řešení.

Schopnost implicitně zadat stavový prostor je v oblasti umělé inteligence velmi důležitá. Stavové prostory uvažovaných problémů mohou být enormní (např. pro hru šachy), nebo dokonce nekonečné. Takové stavové prostory obecně nemůžeme reprezentovat explicitně pomocí množiny všech uzlů a hran, ale musíme si vystačit s implicitním zadáním.

**Příklad.** Uvažte *problém  $n$  dam* z minulé kapitoly. Nejdříve zadefinujte problém formálně. Určete

- a) iniciální stav,
- b) přechodové akce (a případně jejich cenu),
- c) cílovou podmínku.

Až budete mít tuto implicitní definici přechodového grafu, zamyslete jak v takovém případě vypadá graf explicitně (jaká je množina vrcholů a množina hran).

Problém můžeme zadefinovat například následovně.

- a) Iniciální stav je prázdná hrací plocha.
- b) Přejchodová akce je přidání dámy na hrací plochu, pokud ta již neobsahuje  $n$  dam. Každá akce má jednotkovou cenu.
- c) Cílová podmínka vyžaduje, aby bylo na ploše všech  $n$  dam tak, aby se neohrožovaly.

Stavový prostor v takovém případě tvoří množina všech rozestavení 0 až  $n$  dam na hrací ploše. Přejchodová hrana je mezi stavy  $s, t$  právě tehdy, když  $s$  obsahuje o jednu dámu méně než  $t$  a zároveň  $t$  má všechny kromě jedné dámy na stejných pozicích jako  $s$ .

---

**Příklad 2.1.1.** Zadefinujte formálně problém *8-posunovačky* z přednášky. Zamyslete se i nad tím, jak bude vypadat přechodový graf.

Stavový prostor bude množina všech možných rozmístění osmi dlaždic a prázdného políčka. Libovolný stav může být (předem) určen jako iniciální. Přejchodové akce jsou „posunutí“ prázdného pole nahoru, dolů, doleva, nebo doprava (je-li to v daném stavu možné). Cílová podmínka vyžaduje, aby rozmístění odpovídalo cílové konfiguraci.

**Příklad 2.1.2.** Aplikujeme-li na číslo 4 sekvenci operací *faktoriál*, *odmocnina* a *dolní celá část*, můžeme získat libovolné přirozené číslo<sup>1</sup>. Například, číslo 5 lze získat jako

$$\lfloor \sqrt{\sqrt{\sqrt{\sqrt{\sqrt{(4!)!}}}}} \rfloor = 5$$

Formulujte tento problém formálně a opět se zamyslete i nad tím, jak bude vypadat přechodový graf. Liší se nějak stavový prostor od předchozích příkladů?

Jako stavový prostor můžeme uvažovat množinu všech nezáporných čísel. Iniciální stav je číslo 4. Přejchodové akce jsou aplikace některé ze tří zmíněných operací. Cílová podmínka vyžaduje, aby stav byl předem zvolené přirozené číslo.

Stavový prostor v tomto příkladě se liší od předchozích tím, že je nekonečný. To může být důležitý faktor při výběru prohledávacích algoritmů.

**Příklad 2.1.3.** Uvažte zjednodušený problém hledání letecké trasy mezi světovými hlavními městy, který musí řešit webové stránky aerolinek. Předpokládejte, že zákazníci hledají takovou (rozumnou) trasu, která je nejlevnější. Zamyslete se, jak byste řádně formulovali tento problém.

---

<sup>1</sup>Donald Knuth, 1964

Každý stav se skládá z lokace (hlavního města) a aktuálního času. Jelikož v některých případech může cena přeletu záviset i na tom, odkud letíme a podobně, budeme ve stavu držet i tyto informace. Iniciální stav (místo a čas) je specifikován uživatelem. Přejížděcí akce jsou využití takového letu z aktuální destinace, který vzlétá v blízké době, nabízí volná místa, a poskytuje dostatek času na přestup. Cena akce odpovídá ceně přeletu. Můžeme však uvažovat i jiné verze problému, které berou v rámci ceny přechodu v potaz i délku letu, časy přestupů apod. Cílová podmínka kontroluje, zda jsme v cílovém stavu určeném uživatelem.

**Příklad 2.1.4.** Uvažujte situaci, kdy se snažíte naplánovat dovolenou po hlavních městech Evropy. Cesta bude začínat i končit v Praze, nejste nijak časově omezeni a máte v úmyslu procestovat právě všechna evropská hlavní města (některá klidně vícekrát). Jak byste řádně formulovali tento problém?

V rámci každého stavu si budeme pamatovat aktuální hlavní město. Spolu s ním si však potřebujeme držet i množinu dosud navštívených hlavních měst. Iniciální stav je  $\langle Praha, \{Praha\} \rangle$ . Stav v průběhu výpočtu může vypadat například  $\langle Amsterdam, \{Praha, Berlín, Amsterdam\} \rangle$ . Cílová podmínka vyžaduje, aby aktuální město bylo *Praha* a všechna hlavní města již byla procestována. Přejížděcí akce korespondují s návštěvou „sousedního“ (na mapě nebo z dopravního hlediska) hlavního města.

## 2.2 Algoritmy prohledávání

Řešením problému prohledávání je posloupnost akcí. Prohledávací algoritmy proto postupně zvažují různé takové možné posloupnosti. Možné posloupnosti akcí začínající v počátečním stavu tvoří takzvaný *prohledávací strom* s počátečním stavem v kořeni. Intuitivně, prohledávací algoritmy postupně expandují dosud neprozkoumané uzly tohoto prohledávacího stromu, který se pak „rozdívá“, až dokud nenarazí na uzel odpovídající cílovému stavu.

Existuje mnoho různých strategií jak prohledávat stavový prostor. Liší se hlavně v tom, jakým způsobem vybrat uzel k expanzi. Výběr vhodného algoritmu závisí na typu problému, našich požadavcích na průběh výpočtu a našich případných dodatečných znalostech o problému. Pokud o problému nemáme žádnou doplňující znalost, která by nám ulehčila práci, obvykle volíme některý z algoritmů *neinformovaného prohledávání*. Naopak, máme-li nějakou dodatečnou informaci, můžeme ji využít například jako základ heuristiky pro některý z algoritmů *informovaného prohledávání*. Obě tyto skupiny si představíme v následujících sekcích.

Jak už jsme naznačili, existuje velké množství strategií, ze kterých můžeme vybírat. Mezi nejdůležitější vlastnosti, podle kterých můžeme nejrůznější algoritmy porovnávat a hodnotit, patří následující čtyři.

### Definice 3:

- Algoritmus je *úplný*, jestliže nalezne řešení vždy, když existuje.
- Algoritmus je *optimální*, pokud platí, že nalezne-li nějaké řešení, je toto řešení nejlepší ze všech (například z pohledu ceny nebo délky cesty).
- *Časová složitost* algoritmu udává maximální čas potřebný k vyřešení problému.
- *Prostorová složitost* algoritmu udává maximální množství paměti potřebné k vyřešení problému.

Poznamenejme, že časovou i prostorovou složitost obecně uvažujeme *asymptotickou*, a to v *nejhorším případě*. V některých případech se však může hodit uvažovat i např. složitost očekávanou. Dále, řešíme-li, zda je algoritmus optimální, je třeba myslet na *metriku*, kterou uvažujeme (často je jasná z kontextu). Některé algoritmy mohou být optimální z pohledu hloubky/délky řešení (tedy uvažujeme délku cesty v grafu), jiné zase z pohledu obecné ceny řešení (důležité pro ohodnocené grafy). Očekáváme-li od algoritmu, že nalezne postupně více řešení, pak optimalitu můžeme intuitivně chápat tak, že pro každá dvě nalezená řešení algoritmus dříve nalezne to lepší.

Při diskusi o složitosti algoritmů je vždy velmi důležité si uvědomit, v závislosti na čem ji vyjadřujeme. Jak bylo uvedeno v předchozí sekci, v oblasti AI prohledávaný stavový prostor typicky reprezentujeme implicitně (pomocí iniciálního stavu a přechodových akcí). K vyjádření složitosti algoritmů pak používáme následující metriky, které vychází z implicitní reprezentace.

### Definice 4:

- *Faktor větvení* (branching factor)  $b$  je maximální počet následníků kteréhokoli uzlu.
- *Hloubka cíle* (goal depth)  $d$  je délka nejkratší cesty z iniciálního stavu do některého cílového uzlu.
- *Maximální hloubka* (maximum depth)  $m$  je délka nejdelší cesty v grafu.

**Příklad.** Uvažte přechodový graf pro *problém  $n$  dam*, který jsme si definovali v ilustrativním příkladu v Podsektci 2.1. Budeme pracovat s konkrétním grafem pro  $n = 8$ . Určete pro něj

- a) faktor větvení  $b$ ,
- b) hloubku cíle  $d$ ,
- c) maximální hloubku  $m$ .

- a) Na prázdnou plochu můžeme umístit dámu kdekoli, tedy iniciální stav má 64 následníků. Žádný jiný stav tolik následníků mít nemůže. Proto  $b = 64$ .
- b) K řešení se nemůžeme dostat před přidáním 8 dam. Zároveň existuje rozmístění 8 dam takové, které splňuje cílovou podmínku. Proto  $d = 8$ .
- c) Maximum dam, které můžeme postupně položit je 8, a tedy  $m = 8$ .

### Příklad 2.2.1.

Zamyslete se a pokuste se určit faktor větvení  $b$ , hloubku cíle  $d$  a maximální hloubku  $m$  i pro přechodový graf problému *8-posunovačky* z příkladu v Podsececi 2.1.

- a) Jelikož máme vždy nejvýše 4 možné tahy, faktor větvení je 4.
- b) Hloubka cíle záleží na zvoleném iniciálním stavu. Polovina iniciálních konfigurací nelze vyřešit vůbec. Zkuste se zamyslet proč. V ostatních případech je dokázáno, že to bude nejhůře 31.
- c) Maximální hloubka opět závisí na iniciálním tahu. Shora lze jistě ohraničit počtem stavů, alespoň pokud ošetříme cykly.

**Příklad 2.2.2.** Uvažme nyní jednoduchou hru piškvorek na hrací ploše velikosti  $3 \times 3$ . Opět určete faktor větvení  $b$ , hloubku cíle  $d$  a maximální hloubku  $m$  i pro stavový prostor tohoto problému.

Zkuste se také zamyslet, kolik existuje možných různých her (posloupností tahů).

V prvním kole je 9 možných tahů, v každém dalším se tento počet snižuje. Faktor větvení je tedy  $b = 9$ . K řešení se můžeme dostat nejdříve již po pátém tahu, tedy  $d = 5$ . Devíti tahy dosáhneme plné hrací plochy, více tahů tedy udělat nelze. Zároveň existuje hra kde lze dosáhnout devíti tahů, tedy  $m = 9$ . Zamyslete se nad maximální hloubkou pro hrací plochu  $2 \times 2$ .

Naivní odhad by mohl být  $9 \times 8 \times \dots \times 1 = 9!$ . Ve skutečnosti však musíme uvažovat že spousta her skončí dříve než devátým tahem.

## 2.3 Neinformované prohledávání

V případě, že o problému nemáme žádnou další informaci (kromě těch z formální definice), jsme obecně nuceni graf prohledávat takzvaně „slepě“. Tyto strategie prohledávají stavový prostor prostým expandováním uzlů a testováním cílové podmínky. Různé algoritmy se liší například v tom, v jakém pořadí uzly expandují. Nejznámějšími strategiemi jsou *prohledávání do šířky* (BFS) a *prohledávání do hloubky* (DFS). S těmito algoritmy jste se jistě hlouběji



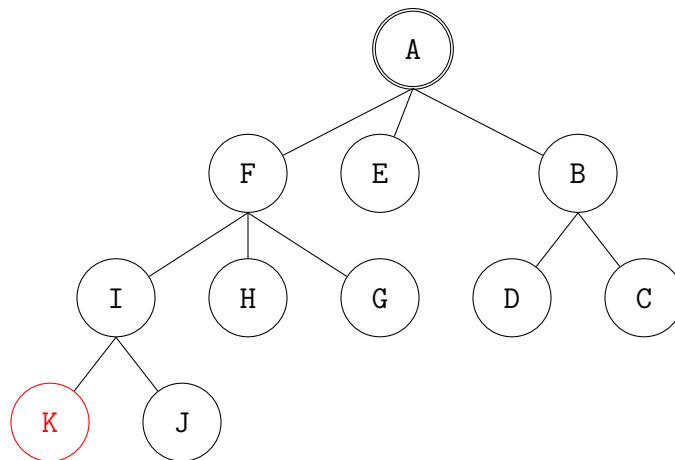
setkali již v rámci některého předchozího předmětu.

Prohledávání do šířky je jednoduchá strategie, kde nejprve expandujeme iniciální uzel grafu, poté všechny jeho následníky, poté všechny jejich následníky, a tak dále. Obecně platí, že veškeré uzly v určité vzdálenosti („úrovni“) od iniciálního uzlu jsou prozkoumány předtím, než expandujeme uzly v úrovni o jedna vyšší.

Prohledávání do hloubky naopak nejdříve vždy expanduje jen prvního následníka každého uzlu. Tímto způsobem intuitivně prozkoumá vždy „aktuálně nejhlubší“ neprozkoumaný uzel. Jakmile expanduje listový uzel, vrací se zpět pomocí *backtrackingu* a prozkoumává stejným způsobem další dosud nenavštívené následníky.

Mezi další algoritmy, se kterými budeme pracovat, patří *prohledávání do hloubky s limitem*, *prohledávání podle ceny* (uniform-cost search) a *prohledávání s postupným prohlubováním* (IDS). Jistě je znáte z přednášky.

**Příklad.** Mějme následující graf, který zachycuje stavový prostor problému. Uvažujte situaci, kdy postupně prohledáváme prostor algoritmy BFS a DFS z iniciálního stavu A, cílovým stavem je K. V jakém pořadí dané algoritmy navštíví jednotlivé stavy? Máte-li na výběr více následníků, postupujte abecedně.

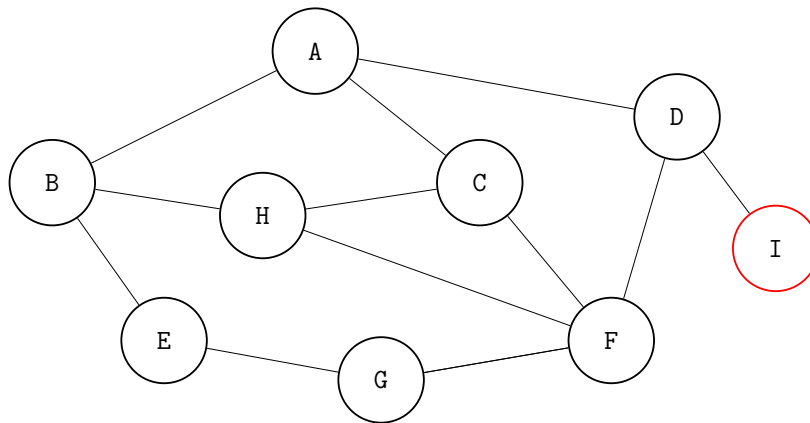


- a) BFS: A - B - E - F - C - D - G - H - I - J - K
- b) DFS: A - B - C - D - E - F - G - H - I - J - K

**Příklad 2.3.1.** Uvažujte stejný strom jako v předchozím příkladě. Jak se změní pořadí uzlů, uvažujeme-li tentokrát, že je iniciálním stavem G? Cílový stav zůstává K.

- a) BFS: G - F - A - H - I - B - E - J - K
- b) DFS: G - F - A - B - C - D - E - H - I - J - K

**Příklad 2.3.2.** V předchozích příkladech jste si vyzkoušeli simulaci algoritmů na stromovém stavovém prostoru. Uvažte nyní následující graf. Určete v jakém pořadí algoritmy BFS a DFS navštíví jednotlivé stavy, uvažujeme-li nejdříve stav A jako iniciální, a poté také stav B jako iniciální. Uvažujte, že výpočet skončí nalezením prvního řešení.



Opět, máte-li na výběr více následníků, postupujte abecedně.

- a) BFS: A - B - C - D - E - H - F - I  
DFS: A - B - E - G - F - C - H - D - I
- b) BFS: B - A - E - H - C - D - G - F - I  
DFS: B - A - C - F - D - I

**Příklad 2.3.3.** Uvažte opět graf z předchozího příkladu. Nechtě tentokrát spustíme prohledávání do hloubky z iniciálního uzlu I a nechtě cílový uzel je nyní F. Uvažujte, že výpočet skončí nalezením prvního řešení.

- a) Je algoritmem nalezené řešení optimální?
- b) Co kdybychom namísto DFS použili DFS s limitem  $l = 4$ ?
- c) A jak by tomu bylo při použití prohledávání s postupným prohlubováním (IDS)?

- a) Nalezené řešení je tvaru I - D - A - B - E - G - F, což jistě není optimální.
- b) Řešení nalezené za použití limitu je tvaru I - D - A - C - F, což opět není optimální.

- c) Konečně, řešení I - D - F je nejkratší cesta z iniciálního do cílového uzlu. Nalezne prohledávání s postupným prohlubováním vždy optimální řešení?

**Příklad 2.3.4.** Mějme opět situaci, kdy prohledáváme výše uvedený graf. Tentokrát nechtě je iniciální stav A a cílový stav F. Jako algoritmus zvolíme DFS s limitem. Není-li řečeno jinak, uvažujte, že výpočet skončí nalezením prvního řešení.

- Jaký je nejvyšší limit, pro který algoritmus nalezne optimální řešení?
- Od jaké hodnoty limitu algoritmus nalezne vždy stejné řešení?
- Uvažujte scénář, kdy prohledáváme celý graf (výpočet neskončí prozkoumáním cílového stavu). Od jaké hodnoty limitu bude výpočet pro DFS s limitem probíhat zcela stejně jako pro klasické DFS?

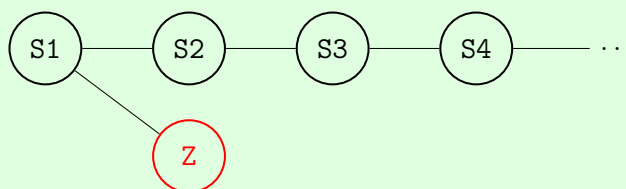
- Jediný takový limit je  $l = 2$ . Pro  $l = 3$  má nalezené řešení tvar A - B - H - F, pro  $l = 1$  algoritmus řešení nenalezne.
- Pro  $l = 4$  má nalezené řešení tvar A - B - E - G - F. Stejně řešení nalezne algoritmus pro každé větší  $l$ .
- Od  $l = 6$ .

**Příklad 2.3.5.** Vymyslete příklady přechodových grafů (případně přímo problémů) takových, že:

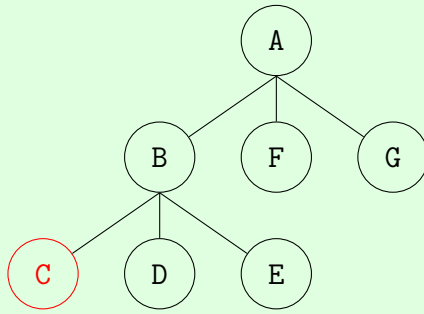
- Prohledávání do hloubky nikdy nenalezne řešení.
- Prohledávání do hloubky nalezne řešení dříve než prohledávání do šířky.
- Prohledávání do hloubky bude procházet uzly ve stejném pořadí jako prohledávání do šířky.

Každou odrážku řešte zvlášť.

- Například následující nekonečný graf, kde S1 je iniciální stav a Z cílový. Uvažujeme, že algoritmus expanduje sousedy podle abecedního pořadí.



- Například následující graf, kde A je iniciální stav a C cílový. Uvažujeme opět, že algoritmy prochází sousedy podle lexikografického pořadí. Pokuste se najít i minimální takový graf.



- c) Můžeme vzít například graf typu *cesta*, kde iniciální stav je „krajní“ vrchol. Napadnou vás i jiná řešení?

**Příklad 2.3.6.** Rozhodněte o pravdivosti následujících tvrzení, své odpovědi odůvodněte. Můžete předpokládat že prohledáváme strom s konečným faktorem větvení, kladnou cenou přechodů a alespoň jedním dosažitelným cílovým uzlem.

- Prohledávání do hloubky s limitem je úplné.
- Prohledávání s postupným prohlubováním má stejnou prostorovou složitost jako prohledávání do hloubky.
- Prohledávání s postupným prohlubováním má horší časovou složitost než prohledávání do šířky.
- Prohledávání do šířky je optimální.

- Neplatí obecně, pouze pokud limit  $l \geq d$ .
- Neplatí obecně, pouze pokud  $m = d$ .
- Neplatí.
- Pouze pro rovnoměrně ohodnocené stromy.

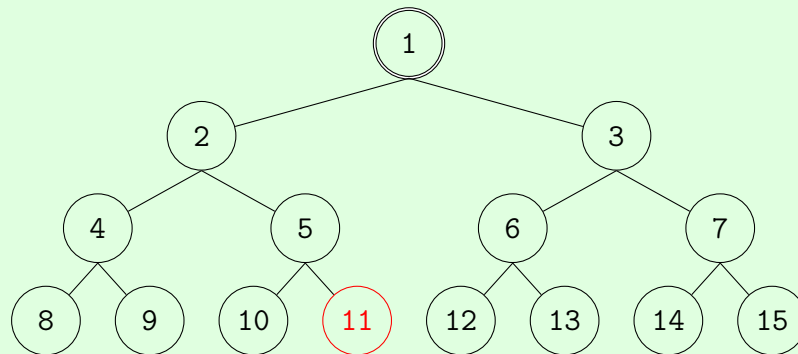
**Příklad 2.3.7.** Popište stavový prostor takový, že prohledávání s postupným prohlubováním má daleko větší složitost než prohledávání do hloubky (jako například  $\mathcal{O}(n^2)$  vs.  $\mathcal{O}(n)$ ).

Uvažme například stavový prostor s faktorem větvení 1, s cílovým uzlem jako listem („na konci“) ve vzdálenosti  $n$ . DFS projde každého následníka jednou a nalezne tedy řešení v  $\mathcal{O}(n)$  krocích. IDS v první iteraci prozkoumá pouze kořen, s každou iterací pak přidá jednoho následníka. Výpočet tedy proběhne v  $1 + 2 + \dots + n \in \mathcal{O}(n^2)$  krocích.

**Příklad 2.3.8.** Uvažte stavový prostor, kde iniciální stav je číslo 1 a každý stav  $k$  má dva následníky – čísla  $2k$  a  $2k + 1$ .

- Načrtněte část stavového prostoru pro stavy 1 až 15.
- Předpokládejte, že cílový stav je 11. Určete v jakém pořadí algoritmy BFS, DFS s limitem  $l = 3$  a prohledávání s postupným prohlubováním navštíví jednotlivé stavy.
- Pokuste se nalézt algoritmus, který vypíše řešení bez jakéhokoli prohledávání. Využijte znalostí o doméně a formulaci problému.

a)



- BFS: 1 2 3 4 5 6 7 8 9 10 11  
 DFS s limitem: 1 2 4 8 9 5 10 11  
 Postupné prohledávání: 1; 1 2 3; 1 2 4 5 3 6 7; 1 2 4 8 9 5 10 11
- Takový algoritmus jistě existuje. Nazvěme akci přechodu ze stavu  $k$  do  $2k$  *Vlevo* a do stavu  $2k + 1$  *Vpravo*. Uvažujme cílový stav jako binární číslo (1011). Jedničky a nuly nám dávají informaci, kterým směrem se vydat. Použijte 0 pro akci *Vlevo* a použijte 1 pro akci *Vpravo*. První jedničku ignorujeme, jelikož začínáme ze stavu 1. Zkuste celý algoritmus zapsat formálně.

## 2.4 Heuristické prohledávání

Máme-li o problému nějakou dodatečnou informaci, často ji můžeme využít k efektivnějšímu prohledávání stavového prostoru. Konkrétně se budeme zabývat strategiemi, které při výběru uzlů k expandování využívají informaci o (odhadu) blízkosti stavů k cíli. Tyto algoritmy obvykle expandují nejdříve ty uzly, které se jeví (v určitém smyslu) jako *nejslibnější*. Souhrně se tyto prohledávací strategie označují jako *best-first search*. Informaci o tom, jak přínosný se jeví daný uzel (jaký je odhad jeho ceny), nám dává takzvaná *ohodnocovací funkce*  $f(n)$ . V průběhu výpočtu si držíme prioritní frontu uzlů uspořádaných dle této ceny a expandujeme stav s aktuálně nejmenší cenou.

Jako komponentu při odhadu ceny většina algoritmů používá *heuristickou funkci*  $h(n)$ . Tato funkce vlastně „reprezentuje“ určitou dodatečnou znalost o doméně problému, a dává jakýsi

odhad na cenu zbývající cesty ze stavu k cíli. Aby byl tento odhad užitečný, některé algoritmy kladou na použité heuristiky podmínky. Jednou z nich je takzvaná *přípustnost*. Ta intuitivně říká, že heuristický odhad ceny cesty z uzlu do cíle nesmí být větší než cena opravdová. V jistém ohledu silnější podmínkou je pak *konzistence* heuristiky.

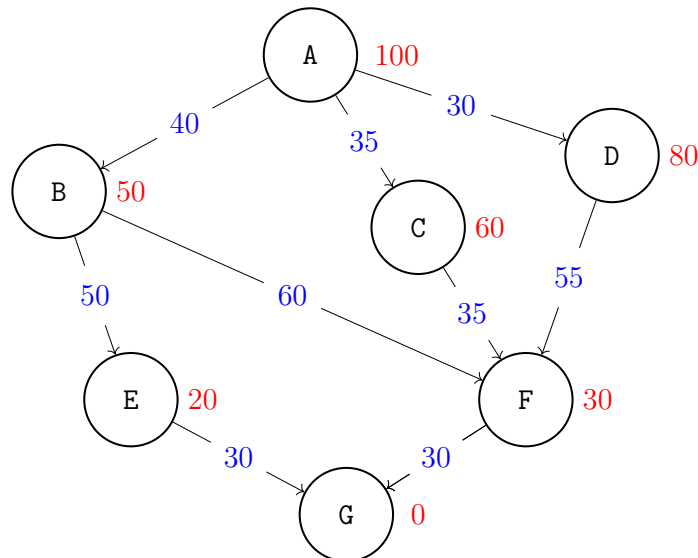
**Definice 5:**

*Ohodnocovací funkce*  $f(n)$  pro každý uzel  $n$  dává celkový odhad jeho ceny. Uzly s nejnižším odhadem jsou expandovány nejdříve.

*Heuristická funkce*  $h(n)$  pro každý uzel  $n$  dává odhadovanou cenu nejlevnější cesty z  $n$  do cílového uzlu. Na heuristiky klademe omezení, že musí být nezáporné, a pokud je  $n$  cílový uzel, pak  $h(n) = 0$ . Pro heuristiky dále platí:

1. Heuristická funkce  $h$  je *přípustná* pokud pro všechny stavy  $n$  platí  $0 \leq h(n) \leq h'(n)$ , kde  $h'(n)$  je skutečná cena cesty ze stavu  $n$  do cíle.
2. Heuristická funkce  $h$  je *konzistentní* (neboli *monotonní*) pokud pro každý stav  $n$  a každého jeho následníka  $m$  platí, že  $h(n) \leq c(n, m) + h(m)$ , kde  $c(n, m)$  je cena přechodu z  $n$  do  $m$ .

V několika dalších demonstrativních příkladech budeme pracovat s následujícím ohodnoceným grafem reprezentujícím stavový prostor.



Pro daný graf platí, že ohodnocení přechodů jsou zobrazena modře, například A-B má cenu 40. Červenou barvou pak jsou pak zaznačeny hodnoty heuristiky pro každý stav, například heuristická hodnota stavu A je 100. Počáteční stav je A, cílový stav je G.

Stručně si představme dvě konkrétní strategie informovaného prohledávání. První je známa jako *greedy best-first search*, neboli *hladové heuristické hledání*. Druhý algoritmus se pak nazývá  $A^*$ . Obě strategie mají společné to, že si v průběhu výpočtu udržují prioritní frontu s uzly k expandování, uspořádanými podle hodnoty  $f(n)$ . V čem se algoritmy liší, je vzorec, pomocí kterého počítají  $f(n)$ . Tento rozdíl se zdá být malý, ale celkově jsou jak úvahy za oběma algoritmy, tak i jejich vlastnosti, velmi odlišné.

Greedy best-first search jednoduše ohodnocuje uzly jen podle hodnoty heuristiky, platí pro něj  $f(n) = h(n)$ . Algoritmu se říká *hladový*, protože se v každém kroku snaží dostat tak „blízko“ k cíli, jak jen to jde. To vede například k tomu, že algoritmus není obecně úplný ani optimální.

$A^*$  je sofistikovanější algoritmus. Kombinuje užitečné vlastnosti *Dijkstrova algoritmu* (doporučujeme si zopakovat) a zároveň rychlost heuristického prohledávání. Při ohodnocování uzlu bere kromě heuristiky v potaz i dosavadní cenu cesty, kterou jsme se do uzlu dostali, označme ji  $g(n)$ . Ohodnocovací funkce má tedy tvar  $f(n) = g(n) + h(n)$ . Použijeme-li heuristiku s vhodnými vlastnostmi, algoritmus je pak garantovaně optimální.

**Příklad.** Uvažujte situaci, kdy prohledáváme výše uvedený přechodový graf pomocí hladového heuristického prohledávání.

- a) V jakém pořadí navštíví algoritmus jednotlivé stavy? Je-li v některém kroku na výběr více možností, postupujte abecedně.
- b) Jak bude vypadat výsledná nalezená cesta? Je toto řešení optimální?

Pořadí stavů je A-B-E-G, což je v tomto případě zároveň i výsledná cesta. Toto řešení ovšem optimální není. Zkuste se zamyslet nad lepší cestou.

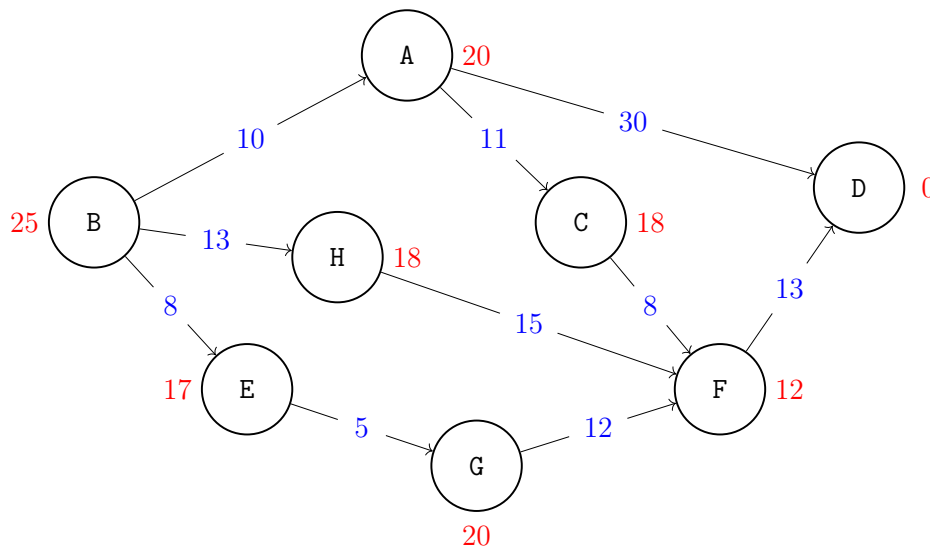
**Příklad.** Nyní uvažujte prohledávání daného grafu pomocí algoritmu  $A^*$ .

- a) Je heuristika zvolená v daném příkladu přípustná? A je konzistentní?
- b) V jakém pořadí navštíví  $A^*$  jednotlivé stavy? Je-li v některém kroku na výběr více možností, postupujte abecedně.
- c) Jak bude vypadat výsledná nalezená cesta? Liší se od cesty nalezené hladovým algoritmem? Je řešení tentokrát optimální?

- a) Heuristika je přípustná, jelikož pro každý stav je hodnota heuristiky menší nebo rovna ceně cesty do cílového stavu. Není ale konzistentní, jelikož  $h(A) > c(A, C) + h(C)$ .
- b) Pořadí stavů při výpočtu je A-B-C-F-G.
- c) Výsledná cesta má tvar A-C-F-G, a tedy se liší od cesty nalezené hladovým algoritmem. Můžete si ověřit, že toto řešení je opravdu optimální.

**Příklad 2.4.1.** ★ Uvažme následující stavový prostor. Iničiální stav je B, cílový stav je D. Ohodnocení přechodů jsou opět zobrazena modře, hodnoty heuristiky červeně. Simulujte výpočet hladového heuristického algoritmu. Je-li více stavů ohodnoceno shodně, postupujte abecedně.

- a) V jakém pořadí navštíví algoritmus jednotlivé stavy?
- b) Jak bude vypadat výsledná nalezená cesta? Je optimální?



Pořadí prohledání stavů je B-E-H-F-D, což se ale liší od nalezené cesty, která má tvar B-H-F-D. Tato cesta není optimální.

Tento příklad připomíná jednu důležitou vlastnost hladového heuristického prohledávání. Zjednodušeně řečeno, algoritmus nemusí v průběhu výpočtu prozkoumávat jen jednu cestu, ale může se také „vracet“ („skákat“), pokud některý dříve objevený uzel má aktuálně nejlepší ohodnocení.

**Příklad 2.4.2.** ★ Opět simulujte prohledávání grafu z předchozího příkladu, tentokrát ale pomocí algoritmu A\*. Je-li více stavů ohodnoceno shodně, postupujte abecedně.

- a) Je heuristika zvolená v daném příkladu přípustná? A je konzistentní?



- b) V jakém pořadí navštíví A\* jednotlivé stavy?
- c) Jak bude vypadat výsledná nalezená cesta?

- a) Tato heuristika je jak přípustná, tak i konzistentní.
- b) B-E-A-H-G-F-D
- c) B-E-G-F-D

**Příklad 2.4.3.** Uvažujte, že věž se může na šachovnici pohybovat o jakýkoli počet políček po přímce, vertikálně nebo horizontálně, ale nemůže přeskakovat ostatní figurky. Je Manhattanská vzdálenost přípustná heuristika pro problém posunutí věže z políčka A na políčko B v co nejmenším počtu tahů? Svou odpověď odůvodněte.

Není přípustná. Manhattanská vzdálenost může nadhodnotit optimální počet zbývajících tahů do cíle, jelikož věž se může posunout o několik políček v jednom tahu. Poznámka – pokud bychom jako cenu cesty místo počtu tahů uvažovali celkový počet projitých políček, pak by byla Manhattanská vzdálenost přípustná.

**Příklad 2.4.4.** Rozhodněte o pravdivosti následujících tvrzení. Odpovědi zdůvodněte. Pokud není řečeno jinak, uvažujte konečný faktor větvení, cenu přechodů vyšší než nějaké kladné  $\epsilon$  a alespoň jeden dosažitelný cílový uzel.

- a) Hodnota přípustné heuristiky nikdy nepřevyšuje zbylou opravdovou cenu (vzdálenost) do cíle.
- b) Algoritmus heuristického prohledávání, jehož fronta je uspořádána dle hodnoty  $f(n) = g(n) + h(n)$  je úplný i optimální, pokud používá přípustnou heuristiku a zároveň ohodnocení uzlů  $f(n)$  monotónně stoupá po jakékoliv cestě do cíle.
- c) Prohledávání podle ceny je jak úplné, tak i optimální, pokud cena cesty nikdy neklesá.
- d) Hladové heuristické prohledávání je jak úplné, tak i optimální, pokud je použitá heuristika přípustná a cena cesty nikdy neklesá.

- a) Pravdivé, plyne z definice přípustné heuristiky.
- b) Ano, protože popsaný algoritmus je A\* a popsaná heuristika je jak přípustná, tak i konzistentní.
- c) Ano, jelikož prohledávání dle ceny je stejné jako A\* prohledávání s konzistentní heuristikou  $h(n) = 0$ .
- d) Ne, protipříklad najdete např. v prvním příkladu.

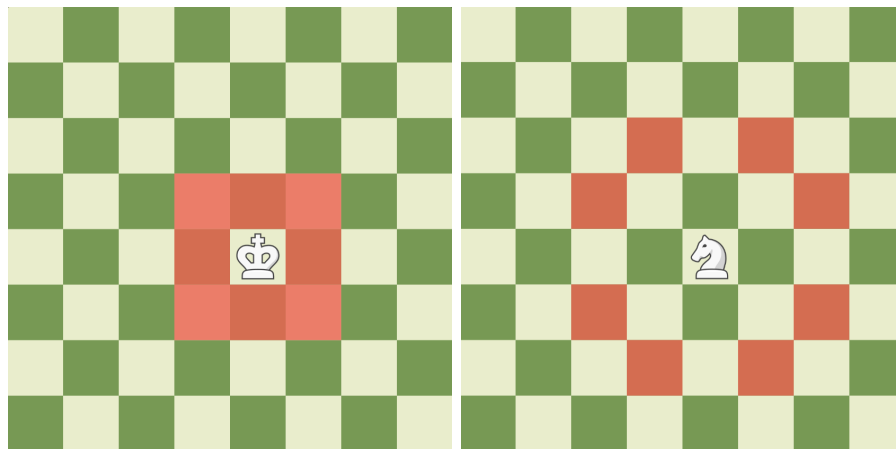
**Příklad 2.4.5.** Která z následujících tvrzení jsou pravdivá? Odpovědi zdůvodněte.

- Prohledávání do hloubky vždy expanduje alespoň tolik uzlů jako  $A^*$  s přípustnou heuristikou.
- $h(n) = 0$  je přípustná heuristika pro 8-posunovačku.
- Prohledávání do šířky je úplné i pokud jsou povoleny přechody s nulovou cenou. Uvažujeme konečný faktor větvení.
- Součet několika přípustných heuristik je opět přípustná heuristika.

- Ne, DFS může (se štěstím) nalézt řešení například hned na první expandované cestě.
- Ano, pro libovolný stav  $n$  nikdy nebude  $h(n) = 0$  vyšší než cena optimální cesty z  $n$  do cílového uzlu.
- Ano, jelikož v tomto případě nulové ceny hran nemají vliv na úplnost algoritmu. Pokud existuje konečné řešení, algoritmus jej najde. Zamyslete se, zda mohou nulové hrany ovlivnit úplnost algoritmu uniform-cost search.
- Ne. Nalezněte protipříklad.

**Příklad 2.4.6.** Ke každému z následujících prohledávacích problémů navrhnete alespoň jednu netriviální heuristiku pro  $A^*$  algoritmus, která bude zároveň přípustná i konzistentní.

- Na šachovnici  $8 \times 8$  potřebujeme posunout figurku krále z horního levého rohu do dolního pravého rohu. Na šachovnici nejsou žádné jiné figurky. Král se může pohybovat na kterékoli z okolních osmi políček. Při návrhu heuristiky můžete použít  $x, y$  k označení vzdálenosti do cílového pole na osách  $x$  a  $y$ .
- Na šachovnici  $8 \times 8$  potřebujeme posunout figurku jezdce (koně) z horního levého rohu do dolního pravého rohu. Na šachovnici nejsou žádné jiné figurky. Jezdec se pohybuje klasicky podle pravidel šachů (viz obrázek). Při návrhu heuristiky opět doporučujeme použít  $x, y$  k označení vzdálenosti do cílového pole na osách  $x$  a  $y$ .



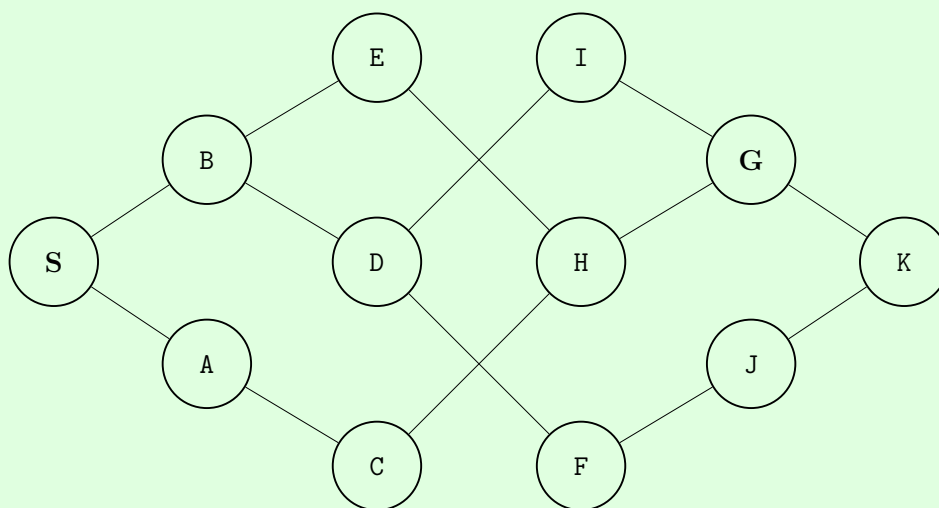
Možné tahy králem a jezdcem na šachovnici.

- a) Možná řešení jsou například  $\max(x, y)$ ,  $(x + y)/2$ , apod.  
 b) Možná řešení jsou například  $\max(x, y)/2$ ,  $(x + y)/3$ , apod.

**Příklad 2.4.7.** Mějme problém pohybu figurky jezdecky na níže zobrazené hrací ploše o rozměrech  $3 \times 4$  z iniciálního pole  $S$  do cílového pole  $G$ . Čísla označují hodnotu heuristiky v daném stavu. Jezdec se pohybuje klasicky podle pravidel šachů (viz minulý příklad). Všechny přechody (pohyby jezdecky) mají jednotkovou cenu.

Nejdříve načrtněte stavový prostor. Poté pro algoritmy DFS, BFS a A\* určete v jakém pořadí navštíví jednotlivé stavy. Předpokládejte, že algoritmy negenerují cesty s cykly a je-li na výběr více následníků, postupují abecedně.

S (3)	H (1)	D (1)	K (1)
I (1)	J (2)	A (2)	E (2)
C (1)	B (2)	G (0)	F (3)



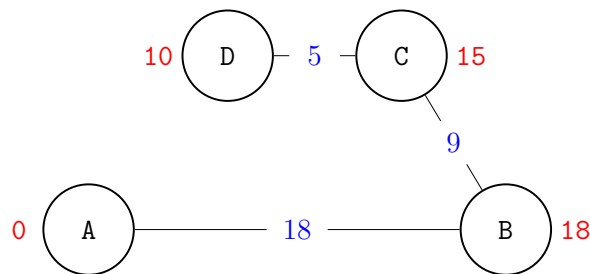
- DFS: S-A-C-H-E-B-D-F-J-K-G
- BFS: S-A-B-C-D-E-H-F-I-G
- A\*: S-A-B-C-D-E-H-G

**Příklad 2.4.8.** Uvažujte prohledávání (konečného) stavového prostoru algoritmem  $A^*$  s konzistentní heuristikou. Necht' existuje jeden dosažitelný cílový uzel  $n_g$  a necht'  $C^*$  je cena optimálního řešení.

- Může se stát, že algoritmus neexpanduje některý uzel  $n_1$ , pro který platí  $f(n_1) < C^*$ ?
- Může algoritmus prohledat jen a pouze uzly  $n$  takové, že  $f(n) < C^*$ ?
- Můžeme před spuštěním výpočtu určit, kolik uzlů  $m$  takových, že  $f(m) > C^*$ , bude expandováno?

- Ne, expanduje všechny takové.
- Algoritmus musí dojít alespoň do  $n_g$ .
- Ano, neexpanduje žádný takový.

**Příklad 2.4.9.** Mějme následující (neorientovaný) graf zadávající stavový prostor. Ohodnocení přechodů jsou jako obvykle zobrazena modře, hodnoty heuristiky červeně, cílový stav je A. Vaším úkolem bude simulovat *tree-search* verzi algoritmu hladového heuristického prohledávání. Pro *tree-search* verzi prohledávacích algoritmů platí, že se neudržuje seznam již navštívených uzlů (neřeší tedy cykly). Simulujte prohledávání ze stavu C. Jaké bude pořadí prohledávaných uzlů?



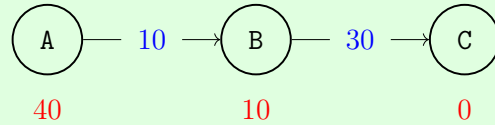
Nejdříve je z fronty (po iniciálním stavu) vybrán D, jelikož má nejnižší odhad ceny do cíle. Ale D je slepá ulička. Řešení by bylo nejdříve navštívit B, to se však nikdy nestane. Prohledáním D se totiž zpět do fronty dostává C, a pak opět D, čímž vzniká nekonečná smyčka. To je dáno tím, že *tree-search* verze nekontroluje, zda už byl uzel navštíven.

Uvažujte jak by dopadla verze algoritmu pro grafové prohledávání (s kontrolou cyklů). Je tato verze obecně úplná?

**Příklad 2.4.10.** Jsou následující tvrzení ohledně heuristik pravdivá? Své odpovědi dokažte.

- Každá přípustná heuristika je i konzistentní.
- Každá konzistentní heuristika je i přípustná.
- Pro libovolný stavový prostor vždy existuje heuristika, která je přípustná i konzistentní zároveň.

a) Ne. Protipříklad je následující. Platí  $h(A) > c(A, B) + h(B)$ .



b) Ano, důkaz lze vést například indukcí.

c) Ano, například  $h(n) = 0$ .

**Příklad 2.4.11.** Rozhodněte a dokažte, zda jsou následující tvrzení o prohledávacích algoritmech pravdivá.

- a) BFS je speciální případ prohledávání podle ceny (uniform-cost search).
- b) Algoritmy DFS, BFS i uniform-cost search jsou speciálními případy best-first search.
- c) Prohledávání podle ceny je speciální případ A\* prohledávání.

- a) Mají-li všechny přechody stejnou cenu, pak  $g(n) = depth(n)$  a uniform-cost search probíhá stejně jako prohledávání do šířky.
- b) DFS se dá chápat jako speciální případ best-first search, kde  $f(n) = -depth(n)$ . BFS je podobně speciální případ best-first search, kde  $f(n) = depth(n)$ . Konečně, uniform-cost search je speciální případ best-first search, kde  $f(n) = g(n)$ .
- c) Uniform-cost search je speciální případ A\* prohledávání s (konzistentní) heuristikou  $h(n) = 0$ .

## 3 Dekompozice problému, Problémy s omezujícími podmínkami

V této kapitole prozkoumáme aplikace prohledávání grafů, kterému jsme se věnovali v kapitole předchozí. Konkrétně si ukážeme, co je AND/OR graf a na jaké široké spektrum problémů ho lze aplikovat – od problémové dekompozice, přes hry až po modelování logických výrazů.

Ve druhé části se budeme zabývat deklarativním přístupem k programování, který umí být překvapivě elegantní a stručný. Zjistíme spolu, že na jeho pozadí se však opět vykonává prohledávání stavového prostoru, které pro mnoho příkladů funguje překvapivě dobře, jindy však narazíme na hranice efektivity.

### 3.1 Dekompozice problému a AND/OR grafy

**Definice 6:** *AND/OR graf* je orientovaný graf s vrcholy typu AND nebo OR (souhrnně zvané vnitřní) a sérií koncových vrcholů  $t_1, t_2, \dots, t_n$ .

Typ vnitřního vrcholu nejčastěji interpretujeme tak, že k vyřešení OR uzlu je potřeba vyřešit kterýkoliv z následníků (potomků). V případě AND uzlu musíme vyřešit všechny následníky (potomky). Koncové vrcholy pak reprezentují dále nedělitelné podproblémy, ať už neřešitelné či se známým řešením. V jiném kontextu lze koncové vrcholy chápat jako splněné či nesplněné.

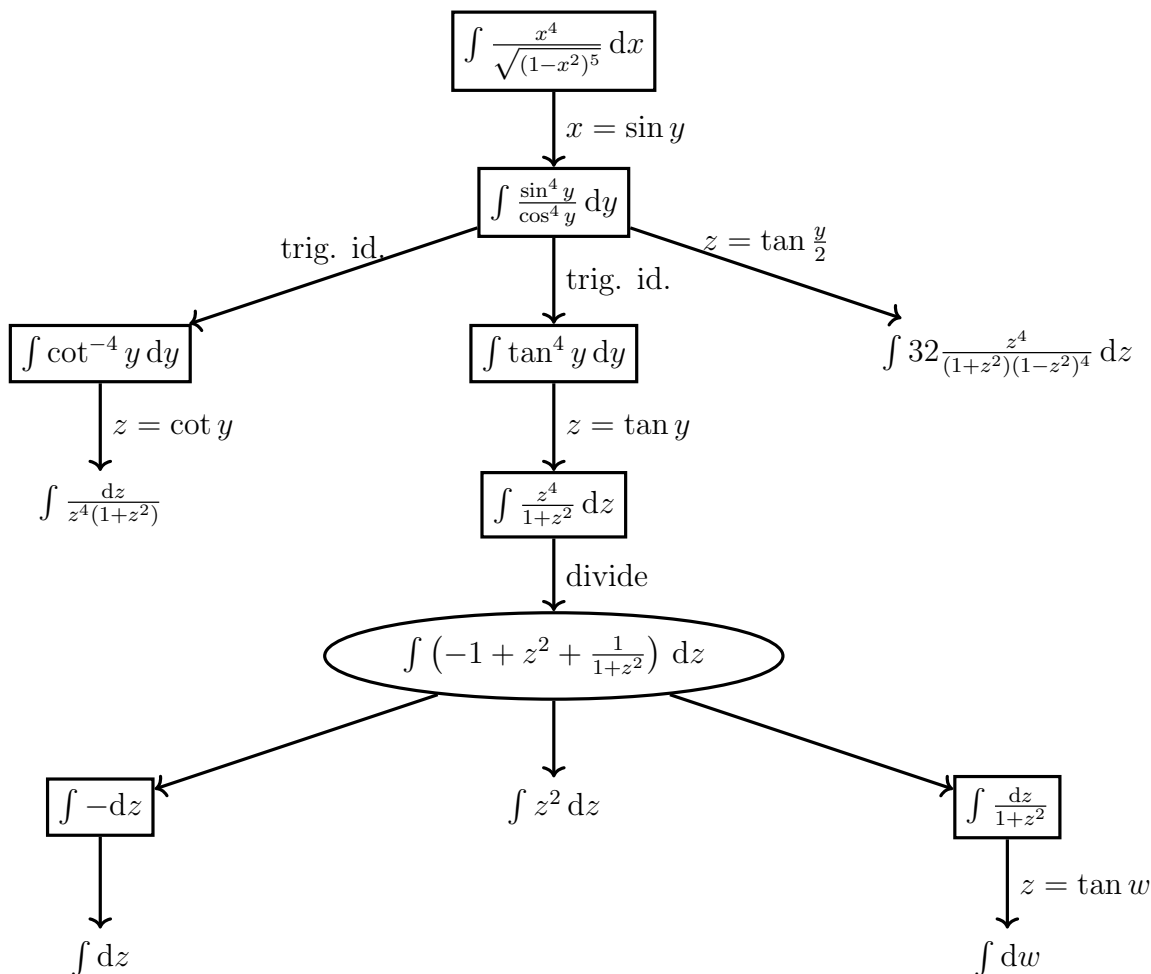
**Příklad.** Rozdělení na vnitřní a koncové uzly v definici není potřeba. Zamyslete se, jak ji upravit tak, abychom s pomocí vnitřních vrcholů mohli modelovat i uzly koncové.

Stačí chápat AND uzel bez následníků jako splněný (řešitelný), zatímco vrchol typu OR bez následníků jako triviálně nesplněný. Listům bychom tedy dali typ AND nebo OR podle toho, zda je chceme mít splněné či nikoliv.

Jako příklad aplikace takového přístupu na problémovou dekompozici (a tedy i příklad AND/OR grafu) si uvedeme strategii řešení neurčitého integrálu. Postup, který si zde ilustrujeme, se v nějaké míře nachází u většiny matematických symbolických řešičů – jako je například WolframAlpha.

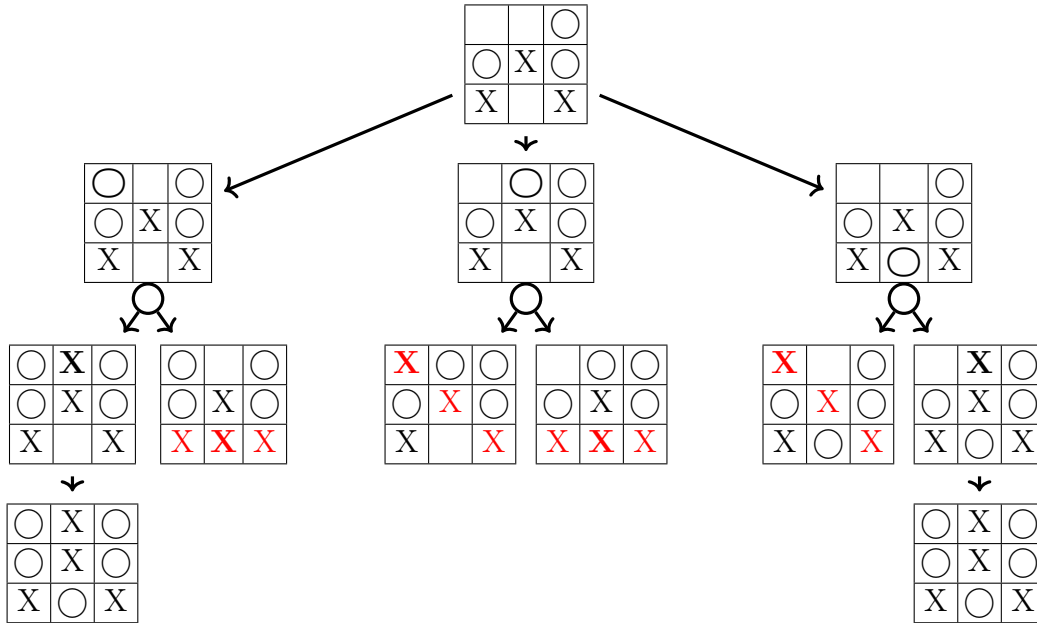
Na následujícím obrázku značíme *AND vrcholy elipsou*, zatímco *OR vrcholy obdélníkem*. Koncové uzly necháváme bez označení. Vrcholy s jediným následníkem mohou být AND i OR bez vlivu na sémantiku (význam) grafu – my v takovém případě volíme konzistentně

obdélníkové ohraničení. OR vrcholy zde tedy reprezentují možné přístupy k řešení, kdežto AND uzly rozdělují problém na sadu podproblémů, které je třeba vyřešit všechny. V některých případech, kde je orientace hran grafu zřejmá (například z rozmístění vrcholů), se šipky v grafu vynechávají.



Častěji budeme vídat AND/OR grafy v kontextu tahových her dvou hráčů – například piškvorek. Omezíme se na hry s perfektní informací, což stručně znamená, že vždy známe úplný aktuální stav hry. Mimo piškvorek toto splňují například šachy či go. Uvědomme si, že tímto způsobem bychom nemohli modelovat třeba poker či kostky.

Na následujícím obrázku používáme pro přehlednost malé kolečko pod vrcholy, které jsou typu AND. Uzly OR a koncové vrcholy jsou neoznačené, neboť jsou rozlišitelné z kontextu. Takové značení můžeme potkat i později tam, kde by prvně dohodnuté znepréhledňovalo nákresy.



V AND/OR grafu se nacházíme v již rozehrané hře piškvorek na hracím poli  $3 \times 3$ . Na tahu je hráč kreslící kolečka a zvažuje všechny své možné tahy. Stačí mu, aby pouze jeden z jeho tahů byl výherní (či alespoň končící remízou) – a tedy vrchol grafu reprezentující aktuální stav hracího pole je typu OR. O úroveň níže zvažujeme soupeřovu odpověď, přičemž předpokládáme, že soupeř je velmi dobrý hráč, který nedělá chyby. Proto potřebujeme, aby *žádná* z jeho reakcí nevedla k naší prohře. Všechny uzly ve druhé úrovni (reprezentující odehrání jednoho našeho tahu) jsou proto typu AND. Při delší hře dvou hráčů se i nadále typy uzlů střídají – jsou typu OR pro náš tah, a naopak typu AND pro tah protihráče.

Analýzou grafu zjistíme, že za předpokladu, že soupeř neudělá chybu, jsme prohráli. Každý z AND uzlů na druhé úrovni je nesplněný, neboť alespoň jeden z jeho následníků značí naši prohru (tři křížky v řadě). Počáteční OR vrchol reprezentující aktuální stav hracího pole je tedy také nesplněný, ačkoliv by bylo možné dosáhnout remízy s nedokonalým soupeřem.

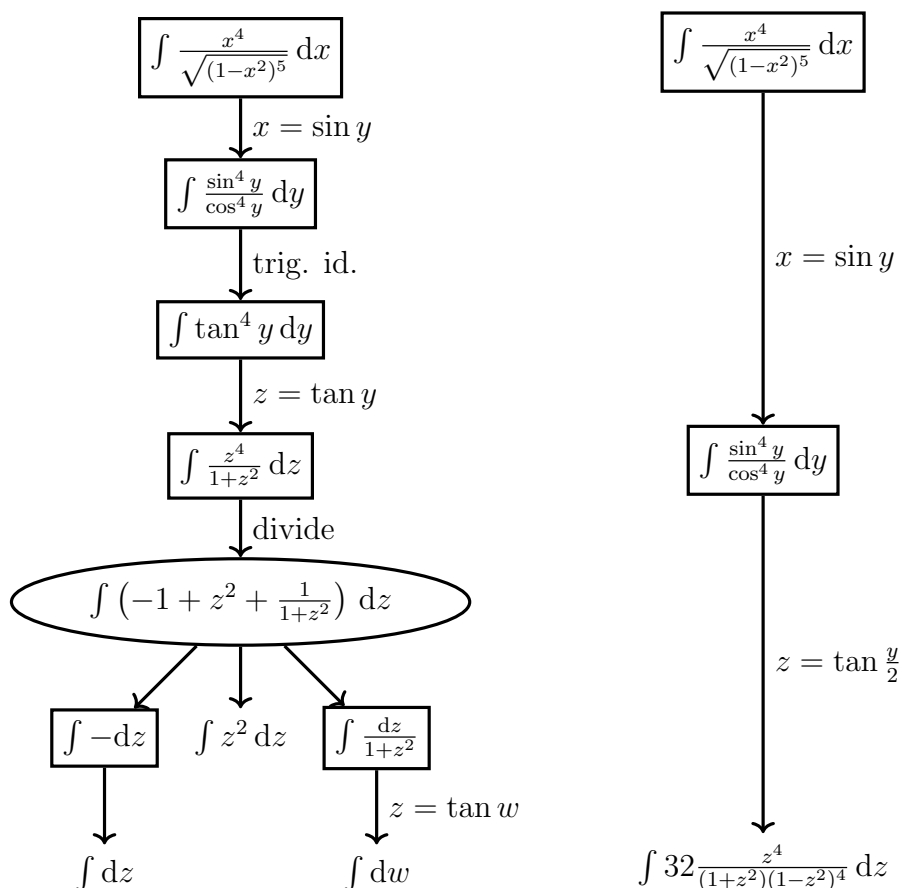
Uvedený příklad je skutečně hodně jednoduchý. V první řadě nevede žádná možnost k naší výhře, v nejlepším případě můžeme doufat jen v remízu. Stojí za zmínku, že typicky se piškvorky nehrají na omezeném hracím poli a i pokud ano, naší naivní metodou vytvořený AND/OR graf by pro běžný čtverečkový papír velikosti A4 obsahoval více vrcholů, než je atomů v celém pozorovatelném vesmíru.



**Definice 7:** Stromem řešení  $T$  problému  $P$  s AND/OR grafem  $G$  je podgraf grafu  $G$ , který je stromem a

- jeho kořen je vrchol reprezentující problém  $P$ ,
- je-li  $N$  vnitřní uzel  $T$  typu AND, pak každý jeho následník v  $G$  je i v  $T$ ,
- je-li  $N$  vnitřní uzel  $T$  typu OR, pak právě jeden z jeho následníků v  $G$  je i v  $T$ .

Pro náš příklad s výpočtem integrálu může strom řešení vypadat například ilustrovanými dvěma způsoby. Praktický význam má však spíše prvně vykreslený strom, kde koncové vrcholy umíme snadno vyřešit.



**Příklad.** Seznamte se s dekompozicí problému Hanojských věží z přednášky, kde problém redukuje na jednoduchý AND/OR graf, kde jsou všechny vnitřní uzly typu AND. Ke kolika fyzickým přesunům disku dojde při počtu kotoučů  $n$  rovno

a) 1,

b) 3,

c) 4?

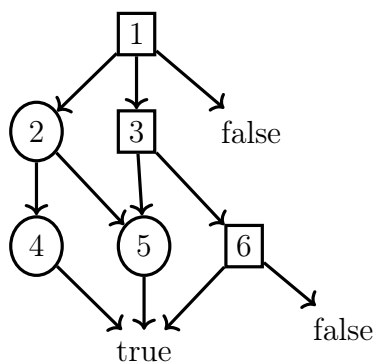
a) 1,

b) 7,

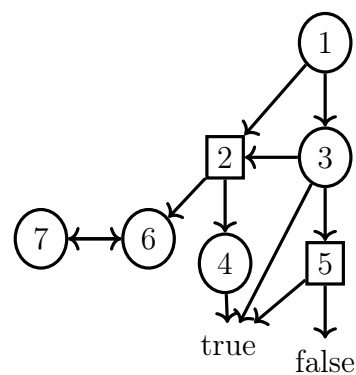
c) 15.

**Příklad.** Rozhodněte, zda je počáteční uzel, značený 1, splněný v následujících AND/OR grafech. Vrchol chápeme jako splněný tehdy, když pro něj existuje v daném grafu strom řešení se všemi koncovými vrcholy nastavenými na *true*. Spočítejte také, kolik takových stromů řešení existuje.

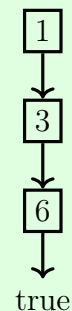
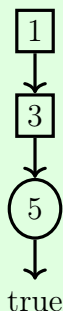
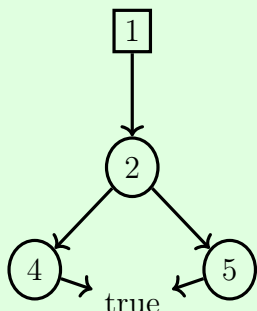
a)



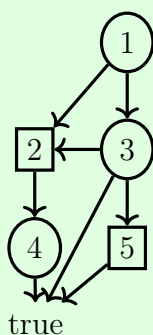
b)



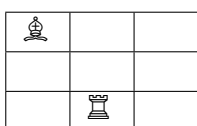
a) Ano, svoji odpověď můžeme potvrdit ukázkou příslušného stromu řešení. My ukážeme všechny 3 možné stromy řešení.



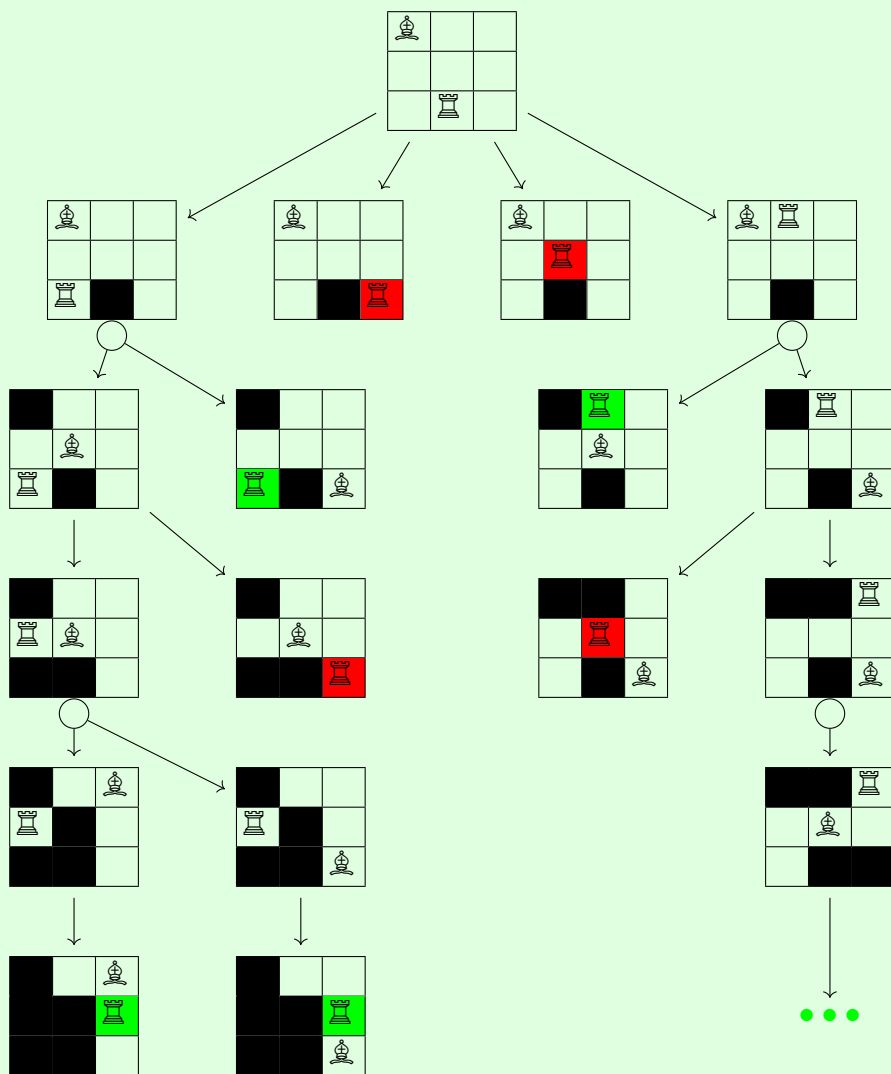
b) Opět ano. Tentokrát je však vyhovující strom řešení pouze jeden. Všimněte si smyčky mezi AND vrcholy 6 a 7, kterou v tomto textu ukazujeme především proto, aby čtenář nezapomněl, že jsou povolené a možné. Učit se s nimi nakládat však nebudeme a konkrétně zde využíváme toho, že ji lze zcela ignorovat díky předchozímu OR uzlu 2.



**Příklad 3.1.1.** Uvažme podivnou hru odehrávající se na hracím poli velikosti  $3 \times 3$ . Hráči se střídají a každý posouvá svou figuru podle pravidel šachu, nesmí však vstoupit na pole, které již bylo v minulosti obsazeno. Podaří-li se hráči sebrat soupeři figuru, vyhrává, naopak je-li hráč na tahu a nemůže se nikam pohnout (neboť dosažitelná pole již byla v minulosti obsazena), prohrává. Úkolem je zanalyzovat situaci, kde proti sobě hraje věž a střelec, první zmíněný je na tahu a počáteční rozmístění je dáno obrázkem níže. Zkonstruuji příslušný AND/OR graf a urči, kdo v takovém případě vyhraje.



Výsledný AND/OR graf je poměrně dlouhý, jednu větev jsme tudíž nevypsali celou - hráči však již v ní nemají možnost volby a každý má k dispozici jen jeden jediný tah, až nakonec hráč s věží vyhraje, protože střelec se nebude mít kam dále pohnout jako první. Výherní pozice značíme zelenou (hráč by vyhrál v příštím kole) a rozmístění, kde bychom nevyhnutelně s dalším soupeřovým tahem prohráli, značíme červeně. Jako obvykle AND uzly poznáme podle kolečka pod daným vrcholem.

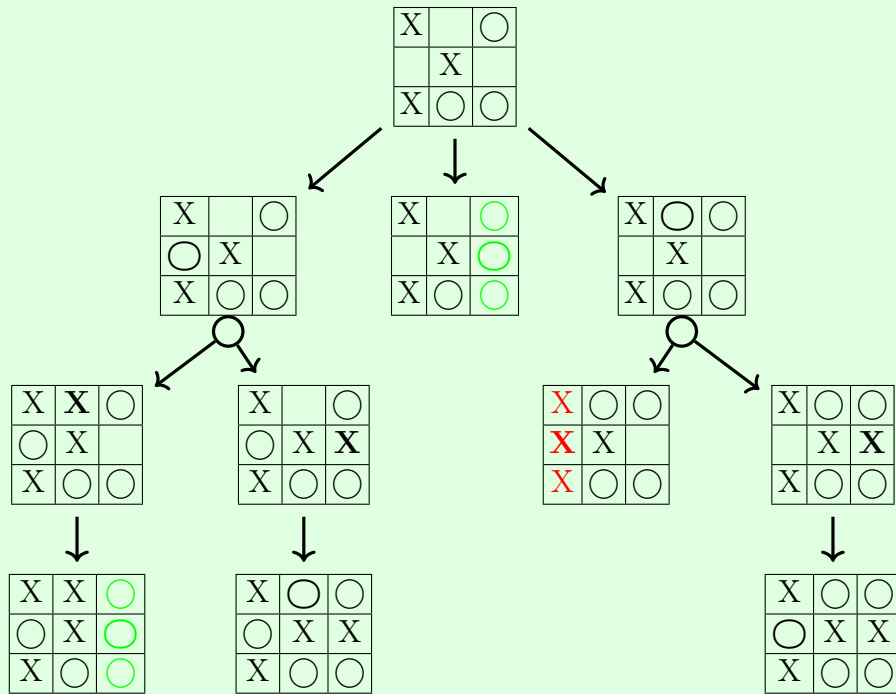


Vidíme tedy, že racionální hráč ovládající věž vyhraje – kdykoliv má možnost volby, může volit například „nejlevější“ strategii. Nakonec takto zvítězí bez ohledu na volby soupeře. V kontextu AND/OR grafu bychom řekli, že cílový (počáteční) vrchol je splněný – je totiž součástí některého stromu řešení, kde jsou všechny koncové vrcholy výherní.

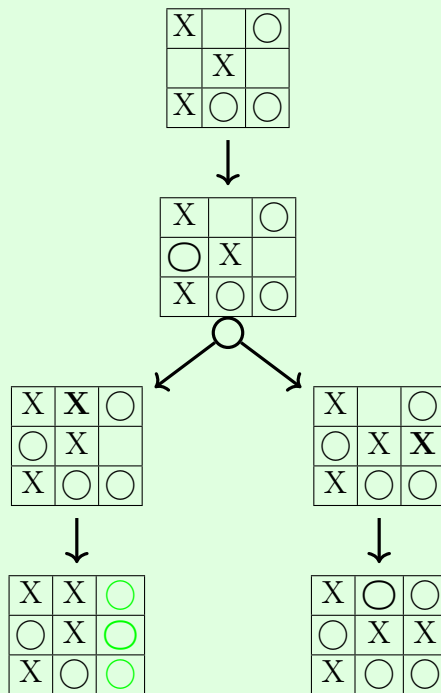
**Příklad 3.1.2.** ★ Uvažte rozehranou partii piškvorek, kde je právě na tahu hráč kreslící kolečka. Sestrojte AND/OR graf a okomentujte stromy řešení tohoto grafu.

X		○
	X	
X	○	○

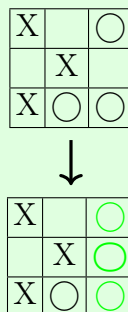
Nejprve si sestrojíme požadovaný AND/OR graf.



Jeden strom řešení vede na remízu, pokud soupeř hraje optimálně, či na výhru, dopustí-li se chyby.



Jiný strom řešení vede k výhře začínajícího hráče.



A poslední (zde neuvedený) končí s racionálním protihráčem prohrou.

### 3.2 Problémy s omezujícími podmínkami

Zlatým grálem programování měl být řešič programů s omezujícími podmínkami, častěji zkracováno jako CSP (constraint satisfaction problem). Je typickým příkladem deklarativního programování, které si můžeme představit tak, že deklaruujeme, co si přejeme, aby bylo splněno, a umělá inteligence to má za úkol zařídit.

Na přednášce jste již jako příklad takového problému viděli problém obarvení grafu, algebrogram a problém N dam. Jako jiný příklad je vhodné uvést typovou inferenci – tedy automatické otypování funkcí a výrazů, které znáte třeba z Haskellu či Pythonu. Podívejme se na tento jednoduchý program.

```
x = f(g(0))
x = f(x)
x = g(x)
```

Jeho analýzou získáme sérii omezení (též podmínek, angl. constraints), jejich konkrétní podoba závisí na uvažovaném typovém systému, zde se pokusíme alespoň o intuitivní náčrt.

- `f` bere jako argument stejný typ, který vrací `g`
- `g` bere argument typu `int`
- `f` vrací stejný typ, který bere
- `g` bere jako argument stejný typ, který vrací `f`

Inteligentní řešič by nám po krátkém zpracování těchto omezení řekl, že obě uvedené funkce jsou typu `int → int`.

Jako názornější příklad uvedeme sudoku.

	2					9	
3		1	9		6	5	2
			8		4		
	9					5	
5			2		3		6
	7					2	
			4		7		
8		2	5		1	7	3
	5					8	

Chceme vytvořit sadu omezení pro naši umělou inteligenci, aby nám poskytla řešení tohoto hlavolamu. Jenže co do našich omezení napsat? První krok je vždy rozmyslet proměnné našeho problému, tedy hodnoty, které bych rád od řešiče získal. Zde se přímo nabízí zavést si sérii proměnných  $x_1, \dots, x_n$ , kde každá reprezentuje hodnotu prázdného políčka a  $n$  je počet prázdných políček. Typicky při deklaraci proměnné zároveň říkáme, jakých můžeme nabývat hodnot – tedy specifikujeme její doménu. Z pravidel sudoku je jasné, že každá proměnná  $x_i$  může nabývat celočíselných hodnot mezi 1 a 9.

Dále víme, že budeme muset vytvořit omezení pro každý sloupec, řádek i čtverec. Pokud si pojmenujeme prázdná políčka popořadě po řádcích, omezení pro první řádek by vypadala jako:  $x_1 \neq 2, x_1 \neq x_2, \dots, x_2 \neq 2, x_2 \neq x_3, \dots$ . Mnoho programovacích jazyků nabízí místo této série 35 podmínek psát pouze něco ve smyslu  $\text{ALL\_DISTINCT}(x_1, x_2, x_3, \dots, x_7)$ , při omezení domén těchto 7 proměnných na  $\{1, 3, 4, 5, 6, 7, 8\}$ . To je nejen výrazně úspornější na zápis, ale především na pozadí schovaný řešič dostane možnost optimalizovat pro toto často se vyskytující omezení. Tento proces bychom opakovali pro každý řádek, sloupec i čtverec a spustili hledání řešení.

Na začátku této sekce jsme zmínili, že takovýto styl programování měl být zlatým grálem informatiky. Není těžké si představit proč – místo psaní dlouhého kódu aplikace bych jen zavedl sérii podmínek jako „když kliknu sem, stane se...“. Takové deklarativní programování by samozřejmě bylo poněkud složitější, než tu ukazujeme, ale to je nakonec imperativní kód ještě mnohem víc. Kde je tedy problém? Pozorný čtenář již jistě uhodl, že v samotném řešiči. Skutečně efektivní umíme sestavit pro určité typy omezení, všude jinde spoléháme na chytré heuristiky při prohledávání všech možných ohodnocení proměnných.

Praktické využití má však dnes CSP při rozvrhování (používá ho i naše fakulta) či konstrukci mikročipů. S jeho pomocí lze modelovat kromě zmíněné typové inference a logických hádanek také maximální tok grafem či lexikální analýzu.

Po nepříliš stručném úvodu jsme připraveni si tyto nové pojmy definovat formálněji.

**Definice 8:** *Problém s omezujícími podmínkami (CSP) je*

- soubor proměnných  $X_1, \dots, X_n$ , každá s neprázdnou doménou  $D_1, \dots, D_n$ ;
- soubor omezení  $C_1, \dots, C_m$ ; každé omezení je podmnožinou  $D_1 \times \dots \times D_n$ ;
- (někdy) účelová funkce  $f : D_1 \times \dots \times D_n \rightarrow \mathbb{R}$ .

Řešením nazveme takovou  $n$ -tici  $(x_1, \dots, x_n) \in D_1 \times \dots \times D_n$ , která splňuje všechna omezení  $C_i$ , tj.  $\forall i. (x_1, \dots, x_n) \in C_i$ .

Má-li CSP více než jedno řešení, může nás zajímat některé konkrétní, potom využíváme účelové funkce  $f$  a hledáme takové řešení, které funkci maximalizuje (či minimalizuje). V některých zdrojích pak takový problém značíme COP (*constraint optimization problem*).

**Příklad.** Uvažme následující program. (Syntax neodpovídá přesně syntaxi z přednášky, ale je dostatečně intuitivní na to, aby nepotřebovala vysvětlení.)

```
X in 1..7
Y in 1..7
X + Y = 7
```

- a) Kolik má takový problém řešení?
- b) Kolik má řešení, přidáme-li omezení `X * Y is even`?

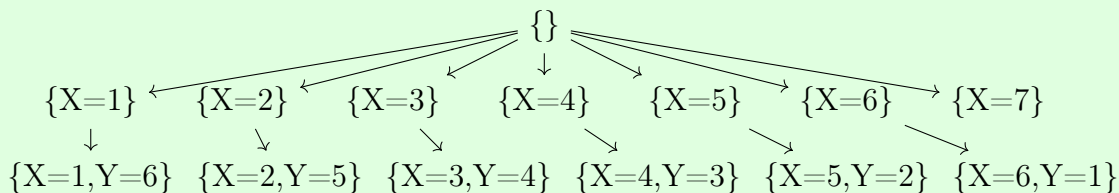


c) Jak vypadá příslušný graf stavů, který prohledáváme?

a) 6 – konkrétně  $\{(X=1, Y=6), (X=2, Y=5), (X=3, Y=4), (X=4, Y=3), (X=5, Y=2), (X=6, Y=1)\}$ .

b) 6, jelikož vždy právě jedna proměnná  $X, Y$  musí být sudá, aby v součtu daly 7. Přidané omezení tedy nemění zadání.

c) Začínáme v prázdném stavu, kde neznáme hodnotu ani jedné proměnné. Z něj se můžeme dostat do stavu, kde hodnota proměnné  $X$  je celé číslo mezi 1 a 7, jelikož jediné tyto stavy jsou v souladu se zadáním. Z každého z těchto stavů, kde známe pouze hodnotu proměnné  $X$  se však již dostaneme v souladu s omezeními v zadání do nanejvýš jediného stavu, kde je dána i hodnota proměnné  $Y$ . Všimněte si, že v jednom uzlu ( $\{X=7\}$ ) naše cesta končí, z něj se nedokážeme dostat při splnění podmínek programu. Přísně vzato v našem grafu existuje mnoho jiných vrcholů – například  $\{X=3, Y=3\}$  – ty nekreslíme, neboť do nich nevede žádná cesta z počátečního stavu, a tak jsou pro naše prohledávání irelevantní.



**Příklad 3.2.1.** ★ Sestavte graf stavů pro následující CSP. Proměnné přiřazujte v sekvenčním pořadí podle jejich deklarace. Popište řešení.

A in 2..4

B in 2..3

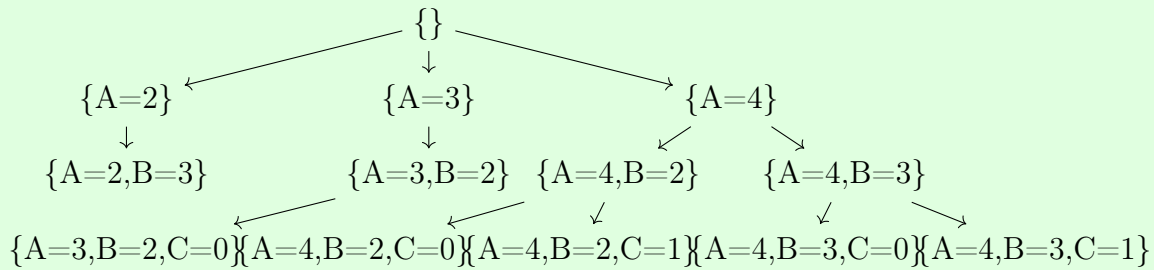
C in 0..6

A - B >= C

A \* (B-1) != B + C

A != B

Nejprve přiřadíme hodnotu proměnné  $A$ , vzhledem k omezením jsou všechny tři hodnoty z domény přípustné. Z nich již jsou povoleny jen některé hodnoty proměnné  $B$ , abychom neporušili podmínku o různé hodnotě  $A$  a  $B$ . Částečný stav  $\{A=2, B=3\}$  je konečný, neboť ho nelze rozšířit bez porušení některého z uvedených omezení. Celkem získáváme 5 různých úplných stavů a máme tedy 5 řešení tohoto problému.



**Příklad 3.2.2.** Převeďte následující algebrogram na CSP. Každé písmenko P v rozepsaném součínu zastupuje prvočíselnou číslici – ne však nutně *stejnou*. Řešení hledat nemusíte. Nezapomeňte specifikovat zavedené proměnné a jejich domény.

$$\begin{array}{r}
 \text{PPP} \\
 \text{PP} \\
 \hline
 \text{PPPP} \\
 \text{PPPP} \\
 \hline
 \text{PPPPP}
 \end{array}$$

Zavedeme si proměnné  $P_1, \dots, P_{18}$ , všechny s doménou  $\{2, 3, 5, 7\}$ . Zvolené značení ilustruje tento náčrt.

$$\begin{array}{r}
 P_1P_2P_3 \\
 P_4P_5 \\
 \hline
 P_6P_7P_8P_9 \\
 P_{10}P_{11}P_{12}P_{13} \\
 \hline
 P_{14}P_{15}P_{16}P_{17}P_{18}
 \end{array}$$

Omezení budou dohromady 3, získaná ze známých pravidel pro součín pod sebou.

$$\begin{aligned}
 (100P_1 + 10P_2 + P_3)P_5 &= 1000P_6 + 100P_7 + 10P_8 + P_9 \\
 (100P_1 + 10P_2 + P_3)P_4 &= 1000P_{10} + 100P_{11} + 10P_{12} + P_{13} \\
 (1000P_6 + 100P_7 + 10P_8 + P_9) + 10(1000P_{10} + 100P_{11} + 10P_{12} + P_{13}) &= \\
 &= 10000P_{14} + 1000P_{15} + 100P_{16} + 10P_{17} + P_{18}
 \end{aligned}$$

Své řešení můžete otestovat například s pomocí pythonovské knihovny `constraint`. Pro zajímavost jediné přípustné řešení vypadá takto.

$$\begin{array}{r}
 775 \\
 33 \\
 \hline
 2325 \\
 2325 \\
 \hline
 25575
 \end{array}$$

**Příklad 3.2.3.** Jako kouzelný čtverec označíme čtvercová matice, kde součet čísel na každém řádku a v každém sloupci je stejný. Jednotlivé buňky mohou nabývat celočíselných hodnot mezi 1 a  $n^2$ , kde  $n$  je dimenze uvažované matice.

Jedním z netriviálních řešení pro  $n = 3$  je tento kouzelný čtverec.

2	9	4
7	5	3
6	1	8

Formulujte tento problém jako CSP pro uvedené  $n = 3$ . Kromě omezení nezapomeňte uvést význam zavedených proměnných a jejich domény.

Jako proměnné zvolíme hodnoty jednotlivých buněk matice.

$x_{1,1}$	$x_{1,2}$	$x_{1,3}$
$x_{2,1}$	$x_{2,2}$	$x_{2,3}$
$x_{3,1}$	$x_{3,2}$	$x_{3,3}$

Doména je pro všechny  $x_{i,j}$  stejná, z definice kouzelného čtverce jde o množinu  $\{1, 2, \dots, n^2 = 9\}$ .

Omezení budeme potřebovat dohromady 5.

$$x_{1,1} + x_{1,2} + x_{1,3} = x_{2,1} + x_{2,2} + x_{2,3}$$

$$x_{1,1} + x_{1,2} + x_{1,3} = x_{3,1} + x_{3,2} + x_{3,3}$$

$$x_{1,1} + x_{1,2} + x_{1,3} = x_{1,1} + x_{2,1} + x_{3,1}$$

$$x_{1,1} + x_{1,2} + x_{1,3} = x_{1,2} + x_{2,2} + x_{3,2}$$

$$x_{1,1} + x_{1,2} + x_{1,3} = x_{1,3} + x_{2,3} + x_{3,3}$$

**Příklad 3.2.4.** Latinský čtverec o velikosti  $n$  je čtvercová matice  $n \times n$ , obsahující v každém řádku (a stejně i sloupci) neopakující se čísla  $1, \dots, n$ . Každý řádek i sloupec je tedy permutace na množině  $\{1, \dots, n\}$ .

Uvádíme konkrétní příklad pro  $n = 3$ .

1	2	3
2	3	1
3	1	2

Formulujte tento problém jako CSP pro uvedené  $n = 3$ . Kromě omezení nezapomeňte uvést význam zavedených proměnných a jejich domény.

Jako proměnné zvolíme hodnoty jednotlivých buněk matice. Doména je pro všechny  $x_{i,j}$  stejná, z definice latinského čtverce jde o množinu  $\{1, 2, 3\}$ .

$x_{1,1}$	$x_{1,2}$	$x_{1,3}$
$x_{2,1}$	$x_{2,2}$	$x_{2,3}$
$x_{3,1}$	$x_{3,2}$	$x_{3,3}$

Omezení bychom mohli vypsát jednotlivě pro každou dvojici buněk zcela vyčerpávajícím způsobem – to je celkem 18 nerovností. Zde ilustrujeme pouze podmínky svazující  $x_{1,1}$ .

$$x_{1,1} \neq x_{1,2}; x_{1,1} \neq x_{1,3}; x_{1,1} \neq x_{2,1}; x_{1,1} \neq x_{3,1}$$

Výhodnější je použít ekvivalent predefinovaného omezení ALL\_DISTINCT námi používaného jazyka. Celkový výčet podmínek potom vypadá takto.

$\text{ALL\_DISTINCT}(x_{1,1}, x_{1,2}, x_{1,3})$

$\text{ALL\_DISTINCT}(x_{2,1}, x_{2,2}, x_{2,3})$

$\text{ALL\_DISTINCT}(x_{3,1}, x_{3,2}, x_{3,3})$

$\text{ALL\_DISTINCT}(x_{1,1}, x_{2,1}, x_{3,1})$

$\text{ALL\_DISTINCT}(x_{1,2}, x_{2,2}, x_{3,2})$

$\text{ALL\_DISTINCT}(x_{1,3}, x_{2,3}, x_{3,3})$

Pro  $n = 3$  má takto definovaný CSP dohromady 12 řešení.

**Příklad 3.2.5.** Zajímavou aplikací CSP je známá Einsteinova hádanka, někdy nazývána zebra. Jejího autora neznáme, často se za něj uvádí Albert Einstein či Lewis Carroll. Týká se pěti sousedů, kteří žijí na ulici v domech v řadě vedle sebe. Každý z nich má jiné povolání, jiný dům, chová jiné zvíře, je jiné národností a preferuje jiné pití. Máme k dispozici několik výroků.

- Angličan žije v červeném domě.
- Španěl chová psa.
- Japonec je malíř.
- Ital pije čaj.
- Nor žije v prvním domě nalevo.
- Majitel zeleného domu pije kávu.
- Zelený dům je bezprostředně napravo od bílého.
- Sochař chová šneky.
- Diplomat žije ve žlutém domě.
- V prostředním domě pijí mléko.
- Nor žije vedle modrého domu.
- Houslista pije ovocný džus.
- Liška žije vedle doktora.
- Kůň je ustájen v domě vedle diplomata.

K úloze patří i otázka. Kdo má zebra a kdo pije vodu?

Vášim úkolem není tuto úlohu vyřešit (ačkoliv to není těžké), ale převést ji na CSP, z jehož řešení bychom uměli odpovédět na zadanou otázku.

Na úloze je důležité zvolit si elegantním způsobem proměnné tak, aby zápis omezení byl co nejsnazší. K tomu si nejprve zavedeme očíslování zmiňovaných domů v řadě zleva doprava čísly 1 až 5. Následně vytvoříme celkem 25 proměnných (s mnemotechnickým pojmenováním), všechny se stejnou doménou  $\{1, 2, 3, 4, 5\}$  – která bude odpovídat právě jednotlivým domům.

Následuje výčet všech zavedených proměnných, všimněte si, část z nich jsme dokázali pojmenovat jen na základě přiložené otázky.

- červený, zelený, bílý, žlutý, modrý;
- Angličan, Španěl, Japonec, Ital, Nor;
- pes, šnek, liška, kůň, zebra;
- malíř, sochař, diplomat, houslista, doktor;
- čaj, káva, mléko, džus, voda

Význam výrazu čaj = 3 je potom „Muž pijící čaj žije v domě číslo 3 – tj. prostředním.“ Nyní nás čeká vytvořit seznam omezení, u kterého oceníme zvolené proměnné a značení.

- Angličan = červený
- Španěl = pes
- Japonec = malíř
- Ital = čaj
- Nor = 1
- zelený = káva
- zelený = bílý + 1
- sochař = šnek
- diplomat = žlutý
- mléko = 3
- |Nor - modrý| = 1
- houslista = džus
- |liška - doktor| = 1
- |kůň - diplomat| = 1

Dobře si ale uvědomme, že toto není kompletní výčet všech omezení. Zároveň je potřeba dodat, že každý obyvatel této neobvyklé ulice má jinou národnost, zvíře, ....

ALL\_DISTINCT(červený, zelený, bílý, žlutý, modrý)

ALL\_DISTINCT(Angličan, Španěl, Japonec, Ital, Nor)

ALL\_DISTINCT(pes, šnek, liška, kůň, zebra)

ALL\_DISTINCT(malíř, sochař, diplomat, houslista, doktor)

ALL\_DISTINCT(čaj, káva, mléko, džus, voda)

Jako součást jediného vyhovujícího řešení dostaneme zebra = 5 a Japonec = 5, zebra tedy vlastní Japonec, obdobně Nor = 1 a voda = 1.

**Příklad 3.2.6.** Zamyslete se, jak formulovat jako CSP problém rozmístění jedné sady šachových figurek po klasické šachovnici velikosti  $8 \times 8$  tak, aby se žádné dvě figurky neohrožovaly. Pěšce nemusíte uvažovat, stačí tedy umístit krále, dámu, 2 střelce, 2 koně a 2 věže.

Obdobně jako u problému N dam si pro každou figuru vytvoříme dvě proměnné, jednu reprezentující řádek, na kterém je umístěna, a druhou příslušný sloupec. Pro krále zavedeme proměnné  $K_r, K_s$ ; pro dámu  $D_r, D_s$ ; střelce  $S_{1,r}, S_{1,c}, S_{2,r}, S_{2,c}$  a obdobně věže a jezdce. Každá z nich může nabývat celočíselné hodnoty 1 až 8.

Omezení je skutečně mnoho, proto je zde nebudeme vypisovat všechna. Pro krále a každou jinou figurku  $X$  vytvoříme následující omezení, které zakazuje, aby figurka  $X$  stála na stejném či vedlejším políčku (vč. diagonály).

$$(K_r - X_r)^2 + (K_s - X_s)^2 > 2$$

Pro dámu a každou jinou figurku  $X$  získáme omezení několik.

$$\begin{aligned} D_r &\neq X_r, D_s \neq X_s \\ (D_r - X_r) &\neq (D_s - X_s), (D_r - X_r) \neq -(D_s - X_s) \end{aligned}$$

Střelce a věže přeskočíme, neboť omezení budou přesně odpovídat těm pro dámu. Místo toho se ještě podíváme na jezdce. Nesmíme zapomenout na požadavek, aby na poli jezdce nestála žádná jiná figurka. Jiné figurky si to sice zajistí samy svými omezeními, prozatím by však nic nebránilo oběma jezdcům stát na stejném poli.

$$\begin{aligned} (J_{1,r} - X_r)^2 + (J_{1,s} - X_s)^2 &> 0 \\ (J_{1,r} - X_r)^2 + (J_{1,s} - X_s)^2 &\neq 5 \end{aligned}$$

Za  $X$  zde opět dosadíme všechny ostatní figurky, včetně  $J_2$ . Stejnou sadu omezení pak budeme muset vytvořit i pro druhého jezdce.

## 4 Hry a herní strategie

V této kapitole naučíme umělou inteligenci řešit deterministické hry dvou hráčů s úplnou informací. Ukážeme si, že jednoduchý algoritmus MINIMAX je pro složitější hry jako jsou piškvorky na dostatečně velké hrací ploše či šachy zcela nevhodný, neboť prohledávaný stavový graf je enormní. Následně spolu MINIMAX zefektivníme při zachování korektnosti, a objevíme tak ALFA-BETA prořezávání.

Nakonec se ještě krátce zamyslíme nad tím, jak bychom námi nalezené postupy mohli aplikovat i na nedeterministické hry dvou hráčů.

### 4.1 MINIMAX

Zabývejme se nyní spolu *tahovou hrou pro dva hráče s perfektní informací*. Dobrým příkladem takové hry, která nás již nějakou dobu touto sbírkou provází, jsou piškvorky. Nejprve se spolu podíváme na hru, kde nám přestává stačit AND/OR graf, a poté přejdeme k MINIMAX grafu.

Pokud chceme modelovat hru s komplikovanějším vyhodnocením než jen výhra či prohra, tedy takovou, kde končíme s nějakým skóre, AND/OR grafy nám to neumožní. Koncové vrcholy totiž mohou být jen splněny či nesplněny, nejsme v nich schopni ukládat číselnou hodnotu skóre a ani s ní následně vyhodnocovat splněnost vnitřních uzlů. Budeme tedy generalizovat myšlenku AND/OR grafů, ovšem omezíme se teď již pouze na stromy.

**Definice 9:** *MINIMAX strom* je strom, kde každý vrchol je buď typu MIN, nebo MAX. Každý list je ohodnocen rozšířeným reálným číslem – tj. prvkem množiny  $\mathbb{R} \cup \{-\infty, \infty\}$ .

MINIMAX strom se vyhodnocuje směrem od listů, dokud nespočteme hodnotu kořene. Hodnota vrcholu typu MAX je rovna maximální hodnotě jeho následníků, hodnota uzlu typu MIN pak odpovídá hodnotě nejmenší napříč všemi následníky.

**Definice 10:** *Algoritmus MINIMAX* je rekurzivní procedura, která ohodnotí každý vrchol MINIMAX stromu.

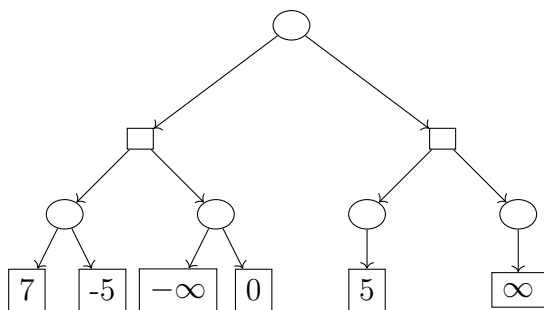
- Hodnota listů je dána jako součást definice MINIMAX stromu.
- Hodnota vrcholu typu MIN je nejmenší hodnota mezi jeho následníky.
- Hodnota vrcholu typu MAX je největší hodnota mezi jeho následníky.

Vypůjčíme si notaci AND/OR grafů a uzly typu MAX budeme značit obdélníkem  $\square$  zatímco MIN vrcholy elipsou  $\circ$ . Jiný častý způsob značení používá pro vrcholy MAX symbol  $\Delta$ , pro MIN uzly  $\nabla$ . Jelikož MINIMAX stromy použijeme pouze pro modelování tahové hry

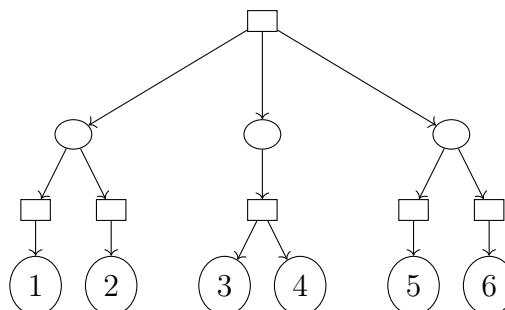
dvou hráčů, budeme se zabývat především těmi stromy, kde se uzly MIN a MAX střídají podle své vzdálenosti od kořene. Dále budeme uvažovat pouze hry s nulovým součtem, což znamená, že obdržím-li v nějaké hře skóre  $s$ , můj soupeř musí získat  $-s$  bodů – každý list nám tak bude stačit ohodnotit jen podle našeho skóre a soupeřovo si snadno domyslíme. Jak se později ukáže, tato myšlenka je velmi důležitá: MINIMAX stromy budou fungovat jen pro hry s nulovým součtem či obecně hry, kde platí, že když maximalizují svoje skóre, zároveň minimalizují to soupeřovo – a naopak.

**Příklad.** Určete hodnotu kořene zadaných MINIMAX stromů.

a)

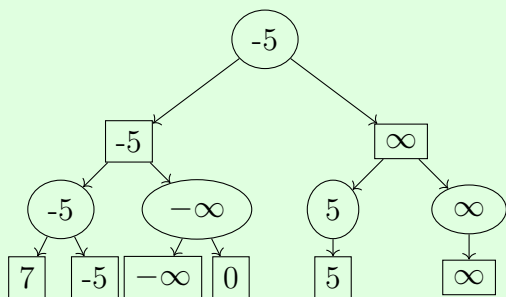


b)

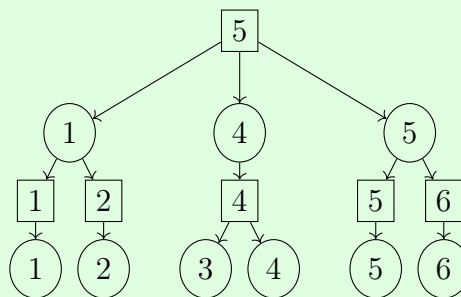


Vyhodnocujeme směrem od listů, uzel po uzlu.

a)



b)



Nyní nám zbývá se podívat, jakým způsobem lze využít tohoto modelu k popisu her. A jako obvykle si při ilustraci pomůžeme piškvorkami na hracím poli  $3 \times 3$ .

Každý list v našem stromu stavů (tedy herních konfigurací) ohodnotíme 0, 1 či -1 podle toho, zda se jedná o remízu, naši výhru či prohru. V této hře budeme první na tahu a kreslíme kolečka. Prohlédněme si tři vybrané listy a jejich přiřazenou hodnotu.



	X	
○	X	○
	X	○

-1

X	○	X
○	X	○
○	X	○

0

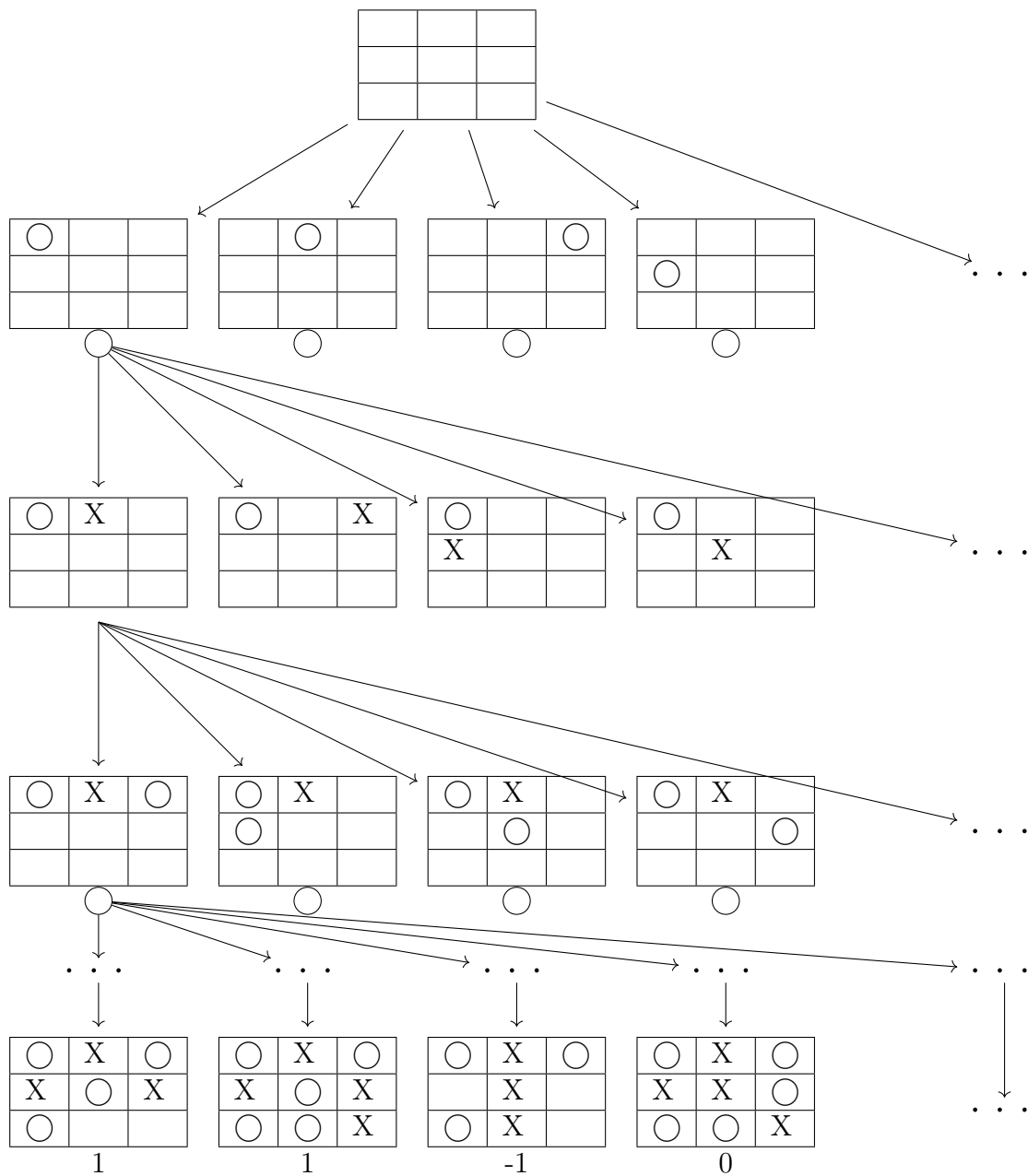
X	X	○
○	○	○
X	X	○

1

Podle našeho ohodnocení je jasné, že kdykoliv budu na tahu já, budu se snažit svoje skóre *maximalizovat*, kdežto protihráč vždy *minimalizovat*. Uzly grafu, kde jsem na řadě já tedy budou typu MAX a ostatní typu MIN.

Uvědomme si, že ne všechny listy jsou ve stejné hloubce (hra může skončit po různém počtu kol). Některé vrcholy se mohou vyskytovat duplicitně, neboť do nich lze „dojít“ různými cestami.

Připomínáme, že vrcholy typu MIN budeme značit malým kolečkem dole, ostatní vrcholy jsou pak typu MAX. Následuje ukázka malé části celého grafu.



**Příklad 4.1.1.** Rozhodněte, zda je následující hry možné modelovat MINIMAX stromy. Pokud ano, zamyslete se nad tím, jak by takové stromy vypadaly.

- šachy
- sudoku
- piškvorky na neomezené hrací ploše
- kámen, nůžky, papír
- tenis

- a) Ano, listy ohodnotíme -1, 1 a 0 podobně jako v piškvorkách. Situace je zde však složitější, protože se v průběhu hry můžeme dostat do té samé konfigurace, ve které jsme už byli někdy dříve. Vytvořit (konečný) graf je však možné jen díky tomu, že profesionální šachy obsahují tzv. pravidlo 50 tahů, které říká, že jestliže se v posledních padesáti kolech nepohl jediný pěšec a nevyhodila se jediná figurka, hra končí remízou. Zde by již mělo být jasné, proč je pravidlo formulováno právě takto – posunutí pěšce ani sebrání figurky nelze vzít zpět a konfigurace se nenávratně mění, čímž znemožníme nekonečnou hru.

S touto myšlenkou nám musí být také jasné, že šachy mají předem určený výsledek, pokud oba soupeři hrají optimálně. Příslušný MINIMAX strom je však natolik velký, že optimální hru dodnes neznáme a šachy jsou tak stále zajímavé a populární. Dnešní umělá inteligence řeší šachy s použitím modifikovaného MINIMAX algoritmu s heuristikami, kdy se nepočítá strom až k listům ale jen do určité předem dané hloubky, ve které se ohodnotí konfigurace na základě vhodné ohodnocovací funkce a zbytek stromu se pak spočte, aby se určil další vhodný tah.

- b) Sudoku jistě není vhodné řešit s pomocí algoritmu MINIMAX – už proto, že to vůbec není tahová hra dvou hráčů – je to však možné. Každý vrchol v našem triviálním stromě by byl typu MAX, neboť je na řadě stále jen „první“ hráč. Jako listy bychom měli všechna možná úplná doplnění hlavolamu, přičemž ten (typicky pouze jeden), který nenarušuje pravidla o sloupcích, řádcích či blocích, bychom ohodnotili 1, kdežto ostatní 0.

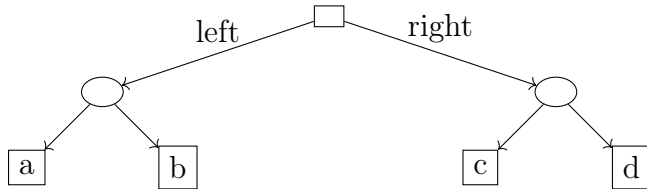
Z konfigurace  $A$  by vedla šipka do konfigurace  $B$  (tj.  $B$  je následník  $A$ ), právě tehdy, když  $B$  vzniklo z  $A$  doplněním jednoho čísla mezi 1 až 9 do některého volného pole.

- c) Pokud se pokusíme vytvořit MINIMAX strom, zjistíme, že je každé jeho patro (tj. množina vrcholů v dané vzdálenosti od kořene) nekonečně velké. Strom má dokonce i nekonečně mnoho nekonečně hlubokých větví – neboť hráči mohou proti sobě hrát donekonečna nekonečně mnoha způsoby. MINIMAX strom pro tuto hru bez zjednodušujících předpokladů tedy neexistuje.
- d) V první řadě je vhodné si uvědomit, že hráči se v tazích nestřídají – ideálně je provedou naráz a nemohou tedy na sebe vzájemně reagovat. Každý stav se tedy skládá z tahu obou hráčů, tj. stav je například (nůžky, kámen), znamenající, že hráč 1 zvolil nůžky a hráč 2 kámen. Každý takový stav (z 9 možných) není problém ohodnotit. Hráč 1 však není schopen ovlivnit, zda skončí po jednom kole ve stavu (nůžky, kámen) či (nůžky, papír), přičemž obě konfigurace pro něj zcela odlišnou hodnotu, a tedy MINIMAX strom není možné sestavit.
- e) Ne, hráč může poslat míč ve kterémkoliv z nekonečně mnoha směrů, hned první „patro“ našeho MINIMAX stromu by tedy mělo nekonečně mnoho uzlů, a to je nemožné.

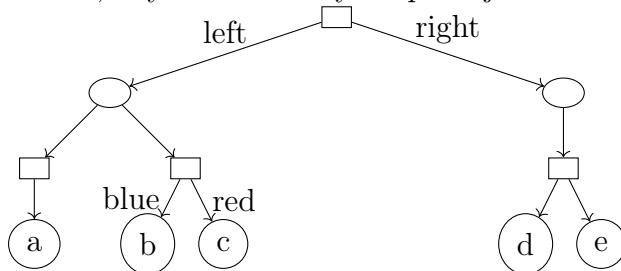
**Příklad 4.1.2.** ★ Doplňte nějaké konkrétní ohodnocení koncových stavů her s naznačeným MINIMAX stromem tak, aby začínající hráč dosáhl nejlepšího možného výsledku hraním

zadaných strategií, předpokládáme-li, že jeho soupeř nedělá chyby. Jaké vztahy obecně musí v takovém případě platit pro hodnoty listů?

a) Výherní strategie nechtě je *left*.

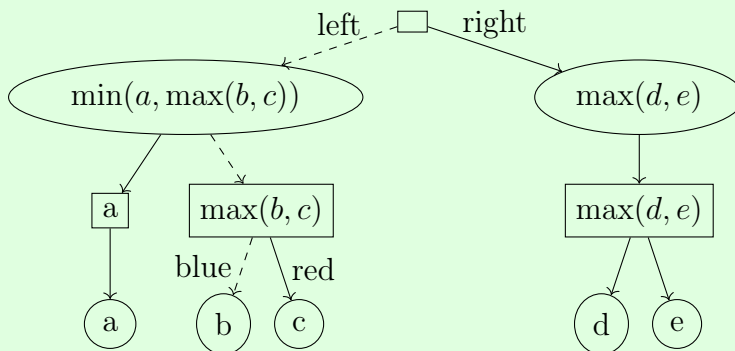


b) Výherní strategie nechtě je volit *left* v prvním kole a *blue* ve druhém. Nezapomeňte zařídit, aby nás dokonalý soupeř s jistotou dovedl k volbě potřebné strategie.



a) Můžeme zvolit například  $a = c = 0$ ,  $b = d = \infty$ . Obecně z grafu vyčteme, že aby se vyplatilo volit strategii *left*, musí maximalizující hráč „vidět“ větší hodnotu v levém uzlu typu MIN. Jinými slovy musí platit  $\min(a, b) \geq \min(c, d)$ .

b) Jako možné konkrétní řešení uvedeme  $a = 1, b = c = d = e = 0$ . Pro obecné řešení si nejdříve rozepíšeme symbolicky hodnoty v uzlech grafu (jen kořen vyhodnocovat není tentokrát potřeba).



Abychom získali jako řešení požadovanou cestu označenou přerušovanými šipkami, musí být splněny následující 3 nerovnice, každá popořadě pro jednu hranu cesty.

$$\begin{aligned} \min(a, \max(b, c)) &\geq \max(d, e) \\ \max(b, c) &< a \\ b &\geq c \end{aligned}$$

Všimněte si ostré nerovnosti ve druhém případě, chceme totiž, aby racionální soupeř

vždy volil tuto variantu, a nám se tak nestalo, že strategie *blue* nebude k dispozici. Nyní už jen získané nerovnice upravíme a ekvivalentně je zapíšeme tak, aby byly čitelnější.

$$\begin{aligned} \max(d, e) &\leq b \\ c &\leq b < a \end{aligned}$$

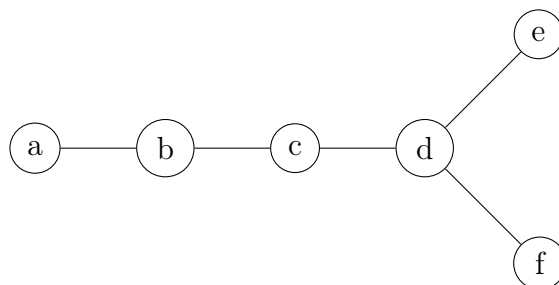
**Příklad 4.1.3.** Dokažte odvážné tvrzení, že MINIMAX stromy rozšiřují AND/OR stromy. Jinými slovy ukažte, že každý AND/OR strom lze chápat jako MINIMAX strom.

Každý koncový vrchol ohodnotíme 1, je-li splněný, v opačném případě 0. Jelikož i listy MINIMAX stromu musí mít nějaký typ, označíme je všechny například za MIN. Vnitřní uzly typu AND pak odpovídají vrcholu typu MIN, jelikož to, že všichni následníci musí být splněni, je v tomto případě ekvivalentní požadavku, aby minimum hodnot následníků bylo 1. Obdobně OR uzly pak odpovídají typu MAX – má-li být splněný alespoň jeden následník, musí být maximum z jejich hodnot rovno 1.

**Příklad 4.1.4.** Ukažte, že obecný MINIMAX strom lze efektivně transformovat na ekvivalentní MINIMAX strom, kde se střídají typy vrcholů – tzn. následník má vždy jiný typ než jeho rodič. Efektivní transformací rozumíme takový algoritmus, která je polynomiální vzhledem k počtu vrcholů vstupního stromu. Ekvivalentní MINIMAX strom musí mít stejné listy se stejnými hodnotami jako strom původní a algoritmus MINIMAX musí napočítat stejnou hodnotu v kořeni.

Stačí si uvědomit, že máme-li vrchol  $V$  a jeho následníka  $N$  se stejnými typy, můžeme (sémanticky ekvivalentně) uzel  $N$  smazat s tím, že jeho potomky „přivěsíme“ pod rodiče  $V$ . Tuto proceduru lze aplikovat opakovaně, dokud se nebudou typy vrcholů střídat podle požadavku.

**Příklad 4.1.5.** Uvažme hru dvou hráčů – pronásledovatele  $P$  a zloděje  $Z$  – na hracím poli zadaném následujícím grafem.

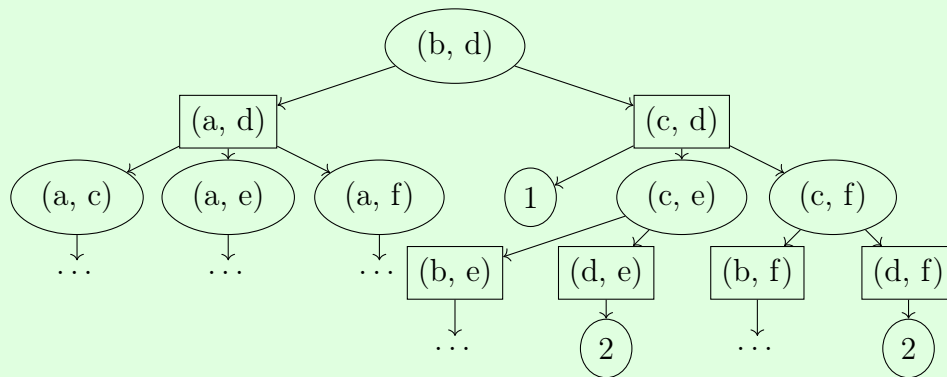


Pronásledovatel je první na tahu a začíná na poli  $b$ , zloděj začíná na poli  $d$ . Každý se během svého tahu posouvá na vedlejší pole, přičemž se po jednom kroku střídají. Hra končí, jestliže se zároveň ocitnou na stejném poli. Zloděj se snaží přežít co nejdéle a hodnota koncové konfigurace je pro něj úměrná tomu, kolik kol odehrál.

- Lze tuto hru modelovat MINIMAX stromem? Lze ji vyřešit?
- Vykreslete a vyřešte strom pro hru omezenou 3 koly – tedy hráč  $P$  i  $Z$  se každý posunou nejvýše třikrát.

Bez omezení na počet kol existují nekonečné strategie (hráči donekonečna střídají ta samá políčka) a MINIMAX strom by tedy byl nekonečně velký (a algoritmus MINIMAX nakonec nevyhnutelně spadl do některé z nekonečných větví).

Intuitivně vidíme, že hra má jasné řešení – racionální pronásledovatel dostihne racionálního zloděje ve dvou kolech. Představme si nekonečný stavový graf pro tuto hru, který níže částečně vykreslujeme.



Tento nekonečný strom by bylo možné vyřešit s prohledáváním do šířky namísto do hloubky (jako to dělá MINIMAX). Kromě toho bychom potřebovali mít při výpočtu nějaké doménové znalosti, tedy například vrchol na obrázku příslušný konfiguraci  $(c, e)$  by se nastavil na hodnotu 2, neboť by našel list s hodnotou 2 a věděl by, že nižšího skóre nedosáhne, neboť se sám nachází již ve druhé úrovni. S těmito vylepšeními bychom skutečně našli ono zjevné řešení této hry.

Strom pro omezení na maximálně 3 kola tu kvůli jeho velikosti neukážeme, prozradíme však, že hodnota v kořenovém uzlu po spuštění MINIMAXu vyjde 2 přesně podle našeho očekávání.

## 4.2 Alfa-beta prořezávání

Řešení deterministické hry s perfektní znalostí můžeme dostat i efektivnější procedurou, než je MINIMAX algoritmus. Zkusme ji spolu objevit.

Až do této chvíle jsme pro zjednodušení tvrdili, že algoritmus MINIMAX vyhodnocuje vrcholy směrem od listů bez toho, abychom podrobněji vysvětlili, proč tomu tak je a jak to

detailně funguje. Podívejme se detailně na skoro-pythonovský (pseudo)kód této procedury.

```
def minimax(node):
    if node.neighbors is empty: # nachazime se v listu
        return node.value
    best_so_far = minimax(node.neighbors[0])
    # vynech nulteho naslednika (sousedu), protoze jsme ho jiz zpracovali
    for neighbor in node.neighbors[1:]:
        value = minimax(neighbor)
        if node.type == "MIN" and value < best_so_far:
            best_so_far = value
        if node.type == "MAX" and value > best_so_far:
            best_so_far = value
    return best_so_far
```

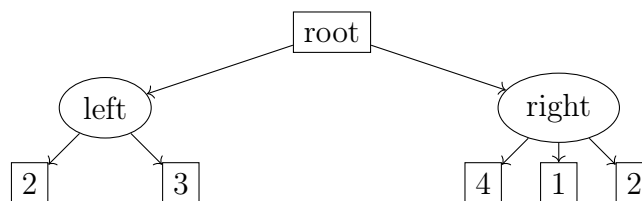
**Příklad.** Jaká je asymptotická časová složitost algoritmu MINIMAX?

Algoritmus probíhá zcela stejně jako prohledávání do hloubky (DFS), v každém uzlu se provedou navíc jen nějaká porovnání s konstantní časovou složitostí. Výsledná asymptotická časová složitost tedy je  $\mathcal{O}(V + E)$ , kde  $V$  je počet vrcholů prohledávaného grafu a  $E$  je počet hran. Toto lze vidět i z toho, že každý uzel a každá hrana se v průběhu výpočtu navštíví právě jedenkrát.

V našem případě uvažujeme pouze souvislé grafy, můžeme tedy výsledek napsat stručněji jako  $\mathcal{O}(E)$ .

Spustíme-li výpočet nad kořenem MINIMAX stromu (tj. zavoláme `minimax(root)`), dostaneme se do for cyklu, kde proceduru rekurzivně zavoláme nad prvním následníkem kořene atp. až se dostaneme k listu. Jedná se tedy o prohledávání do hloubky s několika přidáním výpočty.

Podívejme se teď na jednoduchý příklad.



Budeme předpokládat, že uspořádání seznamu následníků odpovídá tomu na nákrese při čtení zleva doprava. Při vykonávání volání `minimax(root)` se tedy nejprve dostaneme do uzlu `left` a skrze něj okamžitě do listu s hodnotou 2. Dosavadní nejlepší (a tedy v tomto případě nejmenší) hodnota pro vrchol `left` je tedy 2, ale může se pochopitelně změnit při procházení dalšího následníku, ke kterému dochází hned vzápětí. V tomto případě se však nic nezmění a konečně zjišťujeme, že hodnota uzlu `left` je 2.

Kořen tedy skončil zpracování prvního následníka a nejlepší hodnota, kterou zatím viděl, je 2. Protože je typu MAX, víme, že nižší hodnotu už mít nemůže, ale jeho další následník (*right*) by ji mohl ještě zvýšit.

Při vyhodnocování uzlu *right* nejprve vidíme hodnotu 4, poté 1. Zde se na chvíli zastavíme (těsně předtím než bychom zpracovali poslední list). Víme, že konečná hodnota vrcholu *right* nebude vyšší než 1, protože minimalizuje. Na druhou stranu jsme si již před začátkem zpracování tohoto uzlu všimli, že maximalizující kořen *root* nebude mít nižší hodnotu než 2, kterou mu nabízí jeho levý následník. Prohledávání tedy můžeme okamžitě ukončit, jakýkoliv další následník uzlu *right* je nezajímavý a nemůže změnit hodnotu kořene, ačkoliv by pochopitelně ještě mohl snížit hodnotu svého předchůdce.

Ačkoliv jsme si v tuto chvíli ušetřili průchod jediným uzlem, představíme-li si místo listů našeho vzorového příkladu dostatečně velké podstromy, můžeme si hodně pomoci. Tato úspora se neprojeví na nejhorší časové složitosti algoritmu ALFA-BETA, zlepšit se však v průměrném případě nad omezenou vstupní množinou MINIMAX stromů. Hodnotu kořenového uzlu však napočítáme stejnou (což obecně neplatí o ostatních vnitřních uzlech), tedy ALFA-BETA je korektní.

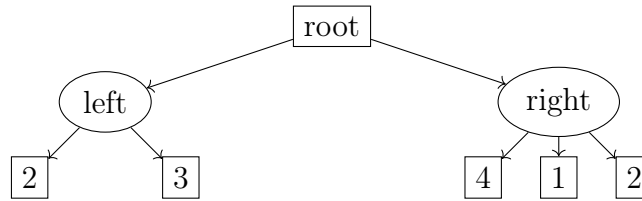
Nyní se již spolu podíváme na implementaci algoritmu ALFA-BETA. Pro přehlednost zcela oddělujeme rutinu pro uzly typu MIN a MAX, z důvodu stručnosti používáme v našem pseudokódu nekonečno INFINITY v inicializaci. Iniciální volání spustíme nad kořenem stromu obdobně jako předtím, tedy příkazem `alphabeta(root, -INFINITY, INFINITY)`.

```
def alphabeta(node, alpha, beta):
    if node.neighbors is empty: # nachazime se v listu
        return node.value
    if node.type == "MAX":
        value = -INFINITY # aby se prepsala pri objevu prvni hodnoty
        for neighbor in node.neighbors:
            value = max(value, alphabeta(neighbor, alpha, beta))
            if value >= beta:
                break
        alpha = max(alpha, value)
    return value
else:
    value = INFINITY
    for neighbor in node.neighbors:
        value = min(value, alphabeta(neighbor, alpha, beta))
        if value <= alpha:
            break
    beta = min(beta, value)
return value
```

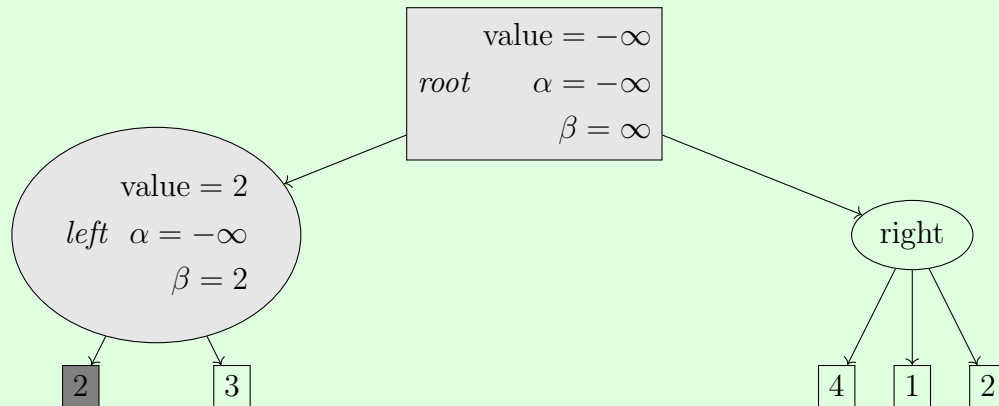
**Příklad.** Odkrojujte si algoritmus nad motivačním příkladem z úvodu této podseky a



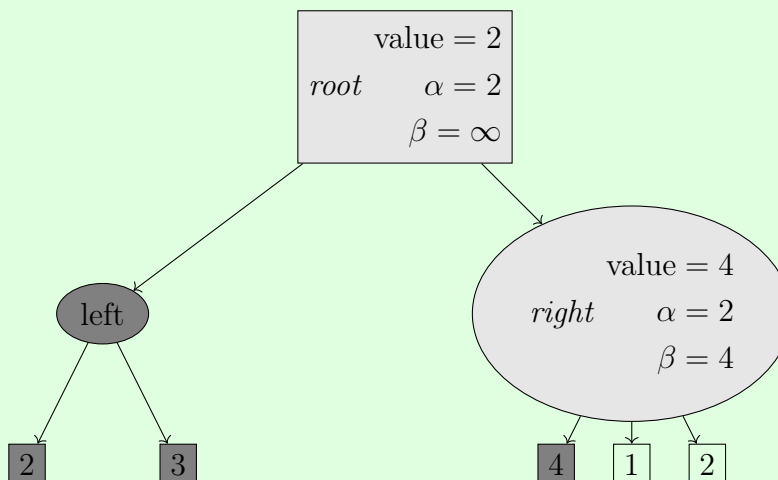
přesvědčte se, že skutečně dojde k ořezání výpočtu tak, jak jsme předpověděli.



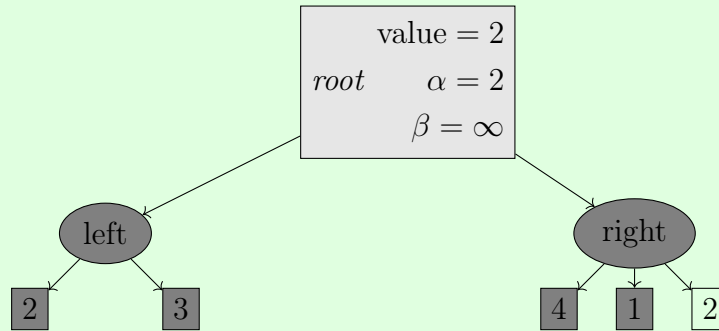
Výpočet začíná voláním  $\text{alphabet}(\text{root}, -\infty, \infty)$ , počáteční hodnoty jsou tedy  $\alpha = -\infty, \beta = \infty$  a začínáme zpracovat kořenový vrchol. Při jeho výpočtu je potřeba najít hodnotu uzlu *left*, který se tedy bude muset vyčíslit (se stejnými hodnotami alfy a bety). Nejprve se objeví list s hodnotou 2, po jeho zpracování ukazují hodnoty proměnných následující ilustrace. Světle šedý uzel je rozpracovaný, tmavě šedý je již úplně zpracovaný.



Následně se obdobně zpracuje list s hodnotou 3, čímž dokončíme výpočet hodnoty vrcholu *left* a vrátíme se s výpočtem zpět do kořene *root*. Ten aktualizuje svoje hodnoty a spustí výpočet nad svým druhým následníkem *right*, přičemž jsme se u levého následníka dostali až k uzlu s hodnotou 3.



Jako další se zpracovává list s hodnotou 1, kdy se konečně projeví smysl zavedených proměnných  $\alpha$  a  $\beta$ . Jelikož se nám hodnota *value* nastaví na 1, což je méně než aktuální hodnota proměnné  $\alpha$  ve vrcholu *right*, výpočet v tomto uzlu se ukončí (a vrátí hodnotu 1 do kořene *root*).



Výpočet v tuto chvíli končí s výslednou hodnotou 2, přičemž poslední list jsme během výpočtu vůbec neviděli.

**Příklad 4.2.1.** Ukažte, že asymptotická časová složitost v nejhorším případě algoritmu ALFA-BETA nad MINIMAX stromy je stejná jako pro algoritmus MINIMAX.

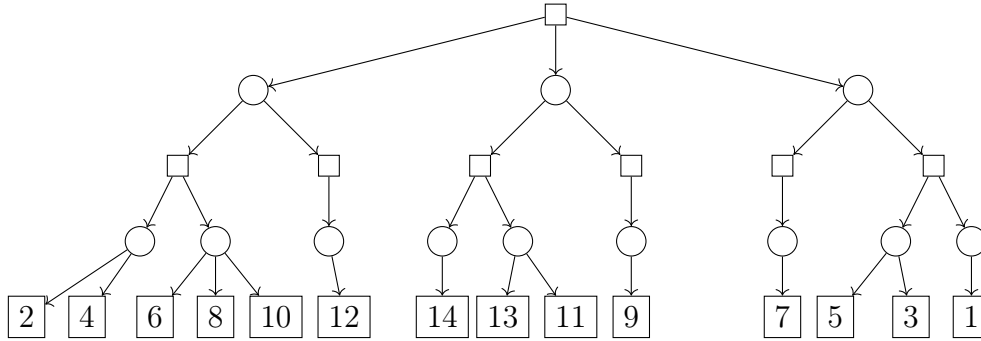
Můžeme dokonce snadno ukázat, že z libovolného stromu dokážeme udělat MINIMAX strom (ohodnocením listů a přidáním typů MIN/MAX všem vrcholům) tak, aby algoritmus ALFA-BETA neprovedl jediný řez. Z toho pak zjevně vyplyne, že v nejhorším případě projdeme celý graf stejně jako v algoritmu MINIMAX.

Slibované zařídíme tak, že všem vrcholům stromu nastavíme stejný typ – například MIN – a listy ohodnotíme zcela libovolnými reálnými čísly. Algoritmus ALFA-BETA pak bude muset projít celý strom, protože k ohodnocení kořene takového stromu je třeba najít minimální hodnotu napříč všemi listy.

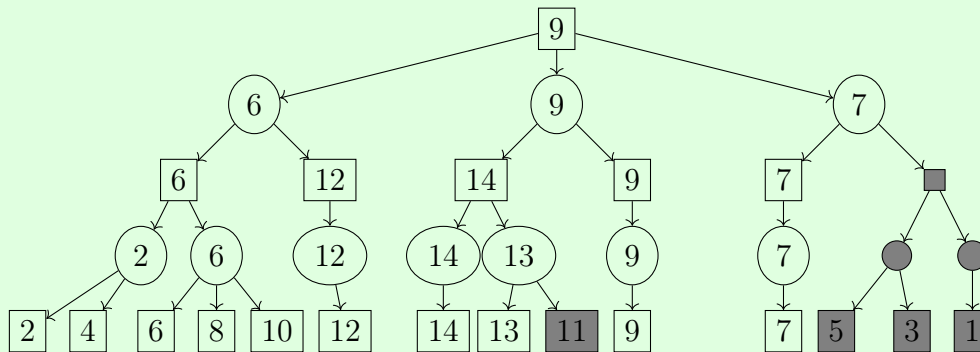
**Příklad 4.2.2. ★** Zadaný strom MINIMAX střídá po úrovních typy svých uzlů a každý jeho list je ve stejné hloubce. Takové v literatuře (a hlavně v praxi) budete vídat nejčastěji, neboť přirozeně vznikají při modelování mnoha zajímavých her. Jako v celém zbytku sbírky předpokládejte i zde, že algoritmus prochází vrcholy v pořadí zleva doprava.

- Vyřešte strom s pomocí algoritmu ALFA-BETA. Především tedy zjistěte výslednou hodnotu kořene a které podstromy budou uřezány, tj. nebudou vůbec navštíveny. Při výpočtu věnujte zvláštní pozornost tomu, abyste porozuměli, jakým způsobem  $\alpha$  odpovídá dolní hranici a  $\beta$  naopak horní.
- Při zachování struktury grafu navrhnete vhodné hodnoty listů tak, aby nedošlo k žádnému prořezávání (a algoritmus tedy navštívil všechny listy).

- c) Při zachování struktury grafu navrhnete vhodné hodnoty listů tak, aby došlo k největšímu možnému prořezávání (a algoritmus tedy navštívil co nejméně uzlů).



- a) Prořezané (nenavštívené) podstromy označíme šedě. Výsledná hodnota v kořeni je 9. Hodnoty ve vnitřních vrcholech nemusí odpovídat skutečné hodnotě při úplném prohledání celého stromu, zapsána je pouze poslední hodnota proměnné `value` z našeho pseudokódu, která by se ještě mohla změnit, kdyby nedošlo k předčasnému ukončení výpočtu (a tedy ořezání stromu). To by však samozřejmě nemělo vliv na výslednou hodnotu kořene.

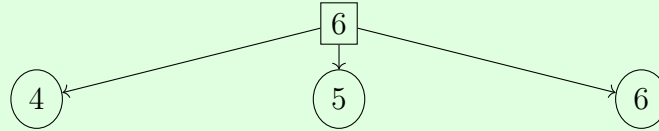


List s hodnotou 11 byl uříznut proto, že v průběhu řešení jeho rodiče jsme zjistili, že jeho největší hodnota bude 13, jelikož minimalizuje. Protože *jeho* rodič (tedy prarodič listu 11) maximalizuje a věděl již, že má přístup k hodnotě 14, hodnota 13 nebo potenciálně nižší ho nezajímala.

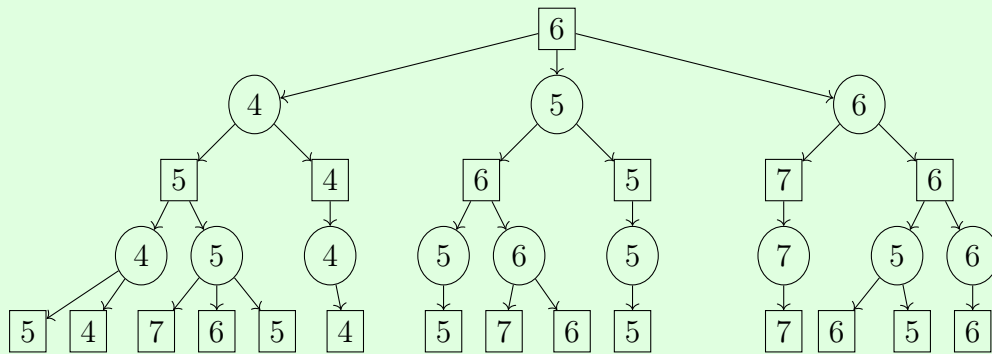
Pravý podstrom byl uříznut ve fázi výpočtu, kdy maximalizující kořen již věděl, že bude mít hodnotu přinejmenším 9. Když se zjistilo, že jeho nejpravější potomek má přístup k listu s hodnotou 7, a tedy vrátí přinejvětším tuto hodnotu, přestal být dále zajímavý a zbytek jeho potomků se neprohledal.

- b) V tomto případě oceníme přístup shora. Zvolme si libovolnou kýženou hodnotu v kořeni a doplňujeme hodnoty potomků tak, abychom ji získali. My volíme hodnotu 6. Nyní je nutné zajistit, aby kořen tuto hodnotu nabył až díky poslednímu potomkovi,

a tím byl nucen ho prozkoumat. První patro našeho stromu tak může vypadat třeba takto.



Nyní stejnou logiku aplikujeme opakovaně vždy o úroveň níže, dokud nedojdeme až k listům. Zde je ukázka možného řešení (pro přehlednost i s vyplněnými hodnotami vnitřních uzlů).

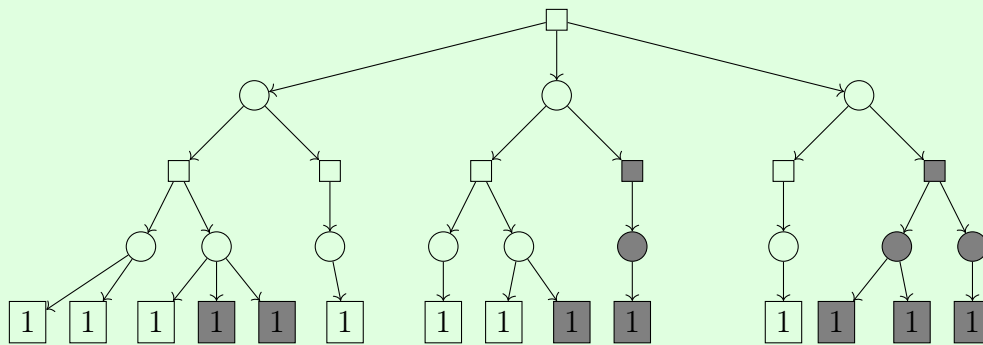


- c) Řešením je například nastavit všechny listy na stejnou hodnotu. Tak se vyhneme navštívení 7 listů, přičemž provedeme celkem 5 prořezání.

Pokud bychom chtěli odargumentovat, že nikde jinde řez provést nejde, stačí u každého navštíveného listu ukázat, že vhodná změna jeho hodnota při zachování celého zbytku stromu změní výsledek (či alespoň sníží celkový počet řezů). To jasně ukazuje, že na jeho hodnotě záleží a je nutné ho navštívit. To je zdlouhavé, neboť je argumentace pro každý list jiná, a my to zde tedy ukazovat nebudeme.

Abychom však byli s důkazem hotovi, museli bychom ukázat, že nám volba různých hodnot v listech nějakým způsobem nepomůže k většímu počtu ořezání. Zde se spokojíme s intuicí – ořezávám, když mi další list nepřinese novou informaci, což lze snadno docílit tím, že budou mít všechny tu stejnou. Kompletní důkaz není složitý, byl by však dlouhý.

Následující ilustrace ukazuje, kde dojde k ořezání (nenavštívené vrcholy jsou značeny šedě) při ilustrační volbě hodnoty listů rovnou 1.

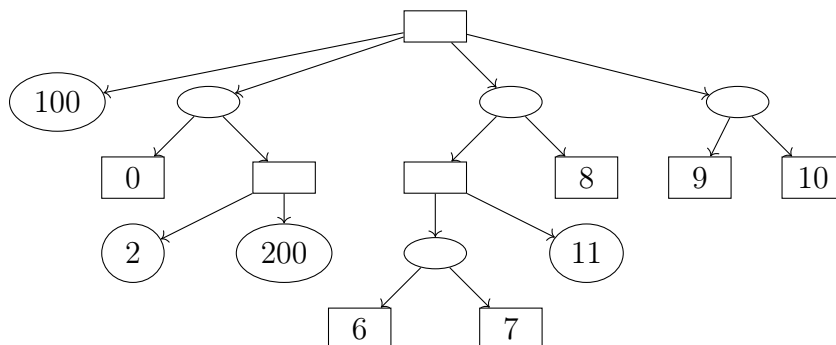


**Příklad 4.2.3.** Při kterých z následujících transformací MINIMAX stromu může dojít ke změně nalezené optimální strategie?

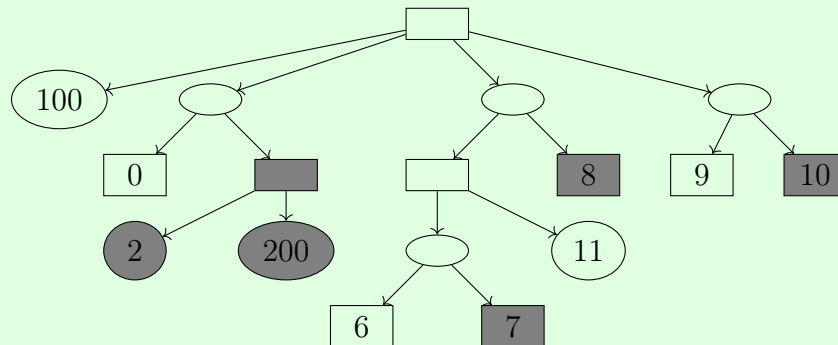
- Ke všem hodnotám listů přičteme stejnou reálnou konstantu  $c$ .
- Všechny hodnoty v listech vynásobíme stejnou konstantou  $c$ .
- Hodnoty ve všech listech se libovolně změní tak, aby mezi nimi zůstalo zachováno jejich původní uspořádání.
- Všechny uzly typu MIN změním na MAX a naopak. Hodnoty v listech pronásobíme  $-1$ .

- Výsledek bude stejný.
- Výsledek bude stejný, je-li konstanta kladná ( $c > 0$ ). Je-li záporná, najdeme místo optimální strategie tu nejhorší.
- Výsledek bude stejný. Již v části *a)* a *b)* jsme si mohli všimnout, že dokud zachováme nerovnosti mezi hodnotami listů, nejlepší strategie zůstává stejná.
- Výsledek bude stejný. Tato transformace odpovídá tomu, že se na hru díváme z pohledu druhého hráče.

**Příklad 4.2.4.** Bez výpočtu určete a zdůvodněte, které podstromy by algoritmus ALFA-BETA v zadaném stromu ořezal. Následníci se prochází v pořadí zadaném obrázkem.



Odřezané podstromy značíme jako obvykle šedou barvou.



Nejlevější šedý podstrom byl odřezán proto, že jeho rodič už ví, že vrátí přinejvětším 0, a to je pro maximalizující kořen nezajímavé – má přístup k hodnotě 100. Obdobným způsobem můžeme argumentovat i u všech ostatních řezů.

Chceme-li se naopak přesvědčit, že naše rozhodnutí projít list s hodnotou 11 bylo správné, uvažme na chvíli, že by ve skutečnosti obsahoval hodnotu 101. Jeho maximalizující prarodič by potom udělal chybu, když se nepodíval na svého druhého potomka s číslem 8 – bez toho by totiž kořenu předal hodnotu 101, a změnil tím výsledek.

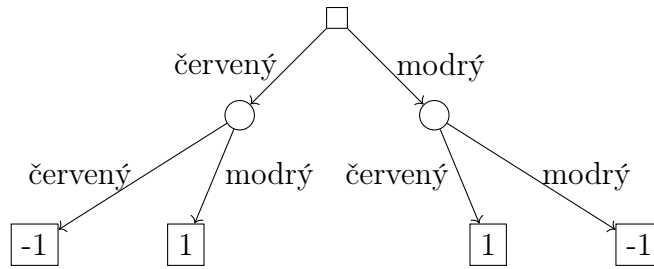
**Příklad 4.2.5.** Představme si MINIMAX strom, který se skládá pouze z uzlů typu MAX. Může při jeho vyhodnocování ALFA-BETA algoritmem dojít k nějakému prořezání?

Může se zdát, že náš algoritmus nikdy žádnou větev neprořeže, neboť hodnota  $\beta$  je v celém průběhu výpočtu rovna  $\infty$ . Vzpomeneme-li si však, že hodnoty v listech jsou rozšířená reálná čísla a mohou též nabýt hodnoty  $\infty$ , zjistíme, že k prořezání v takovém stromě dojde tehdy (a jen tehdy), když se při prohledávání objeví list s hodnotou  $\infty$ .

### 4.3 Nedeterministické hry

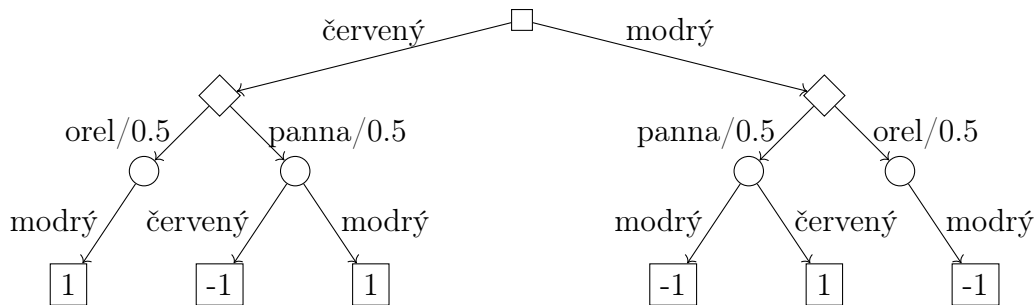
Až do této chvíle jsme tvrdě trvali na tom, aby uvažované hry neobsahovaly náhodu, neboť bychom ji neuměli modelovat. Nyní si ukážeme, že lze vhodně rozšířit MINIMAX stromy tak, aby náhodu v jisté míře postihly a my nad nimi pak mohli provádět smysluplné výpočty a předpovědi. Zjistíme však také, že některé vlastnosti a zákony, které dosud platily, se s tímto rozšířením zcela rozbíjí a mění.

Na začátek uvažme velmi triviální ilustrativní hru. V prvním kole si vezmu modrý, či červený žeton. V dalším kole provede to samé provede soupeř (může si zvolit i žeton stejné barvy). První hráč vítězí, drží-li jiný žeton než jeho soupeř, v opačném případě prohrává. Pro přehlednost přikládáme ještě příslušný MINIMAX strom této jednoduché hry.



Nyní naši hru obohatíme o náhodu. Před tím, než si druhý hráč zvolí svůj žeton, hodíme mincí. Padne-li orel, soupeř přijde o možnost volby – musí si vzít modrý žeton. V opačném případě se nic nemění a hra pokračuje jako předtím.

Pro modelování nedeterministické hry použijeme speciální vrchol (značený kosočtvercem  $\diamond$ ) tam, kde dochází k rozhodnutí na základě náhody. Hrany, které z něj vychází, budou vždy označeny pravděpodobností, se kterou k výběru té konkrétní cesty dojde.



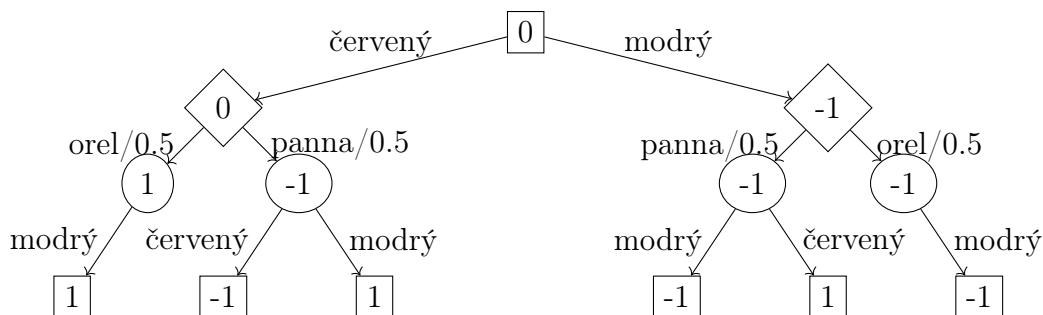
Už tedy chápeme, jak nedeterministické hry modelovat, zajímá nás však, jak se dopočítáme ke zjevnému závěru, že má začínající hráč volit červený žeton. K tomu využijeme upravený algoritmus MINIMAX, který rozšíříme o schopnost počítat hodnotu nového typu uzlu reprezentujícího náhodnou volbu  $\diamond$ .

Pro většinu aplikací je přirozené volit jako hodnotu vrcholu náhody jeho očekávanou hodnotu – tj. hodnotu každého potomka vynásobíme pravděpodobností, že bude vybrán, a výsledky pak sečteme napříč všemi potomky. Poněkud přehlednější zápis říká

$$\text{minimax}(X) = \sum_{n \dots \text{child of } X} \text{minimax}(n) \cdot P(n),$$

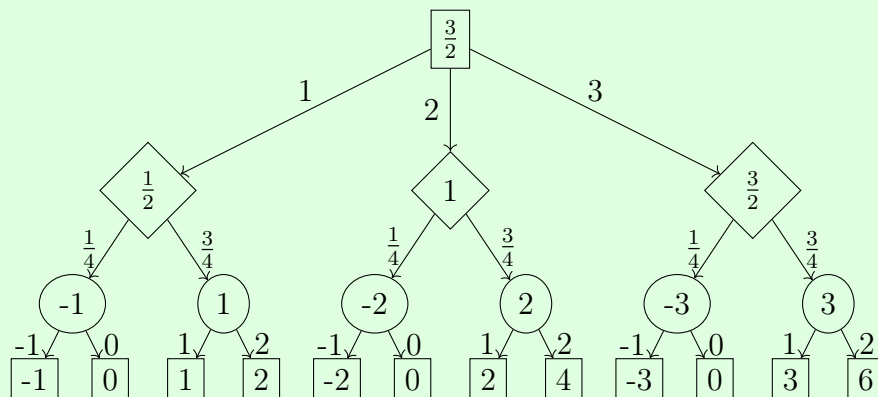
kde  $X$  je uzel reprezentující náhodu,  $\text{minimax}(n)$  je spočtená hodnota potomka  $n$  a  $P(n)$  je pravděpodobnost, že náhodně zvolíme právě potomka  $n$  – tj. číslo na odpovídající hraně grafu.

S touto znalostí již celý graf snadno doplníme. Je vidět, že očekávaný výnos začínajícího hráče je 0 a získá ho za předpokladu, že volí strategii *červený*.



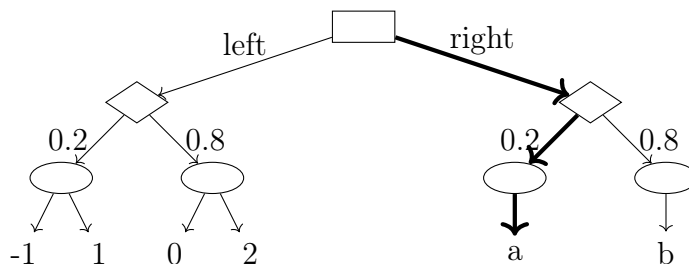
**Příklad 4.3.1.** Sestavte vhodný strom a následně jej vyřešte pro následovně zadanou nedeterministickou hru dvou hráčů.

Začínající hráč si zvolí číslo 1, 2 nebo 3. Následně se hodí 3 mincemi. Padne-li na všech z nich stejná strana, soupeř si smí zvolit číslo -1 nebo 0. V opačném případě vybere číslo 1 či 2. Začínající hráč obdrží od soupeře částku rovnou součinu jejich čísel – v případě záporného součinu naopak musí příslušnou sumu protihráči zaplatit.



Začínající hráč by měl volit strategii 3, racionální soupeř potom zvolí číslo -1 nebo 1 podle toho, co mu výsledek hodu mincí dovolí.

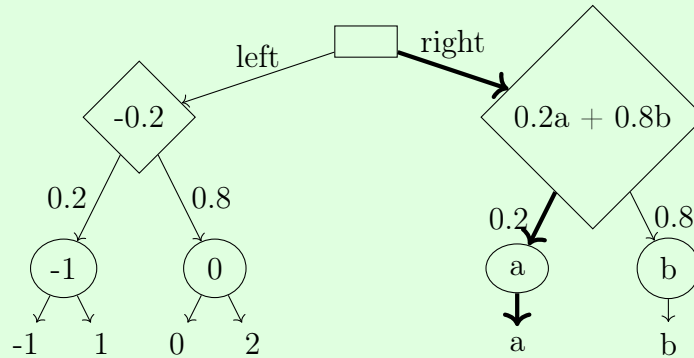
**Příklad 4.3.2.** Uvažme obrázkem zadaný MINIMAX strom rozšířený o vrcholy reprezentující náhodu. Hodnoty dvou listů jsou neznámé, označeny jako  $a$  a  $b$ .





- a) Co musí platit pro hodnoty  $a, b$ , aby optimální strategie odpovídala tučně vyznačené?
- b) Nechť  $b = 1$ . Co musí platit pro hodnotu  $a$ , aby maximalizující hráč volil v prvním kole strategii *left*?
- c) Nechť  $b = -1$ . Co musí platit pro hodnotu  $a$ , aby maximalizující hráč volil v prvním kole strategii *right*?

Nejprve si zadaný strom částečně vyřešíme s pomocí dostupných informací.



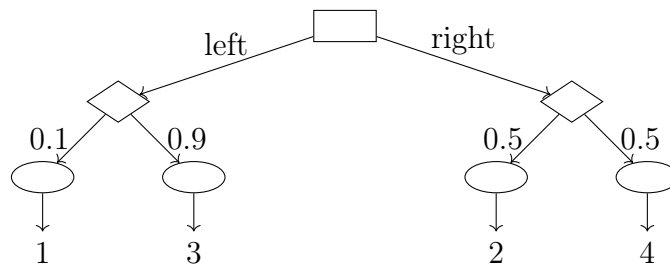
- a) To neumíme zajistit, neboť náhodný uzel nás může poslat také k listu s hodnotou  $b$ .
- b) Zapišeme

$$-0.2 \geq 0.2a + 0.8,$$

neboli zjednodušeně  $-5 \geq a$ .

- c) Po zjednodušení získáme podmínku  $3 \leq a$ .

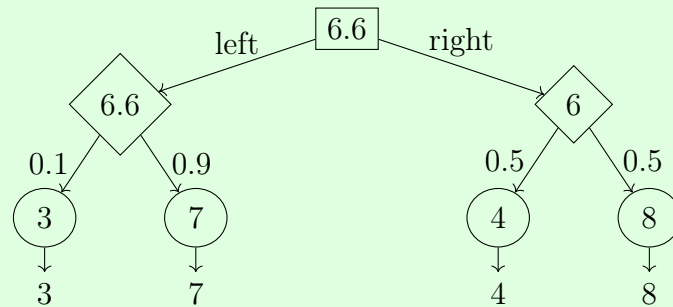
### Příklad 4.3.3.



Nejprve pro obrázkem zadaný rozšířený MINIMAX strom ověřte, že optimální strategie začínajícího hráče je *right*.

Poté ukažte, že lze vhodně transformovat hodnoty listů tak, aby se optimální strategie změnila na *left*. Transformace musí zachovat původní uspořádání mezi listy; tedy jestliže měl list  $A$  před transformací menší hodnota než list  $B$ , bude tomu tak i po ní.

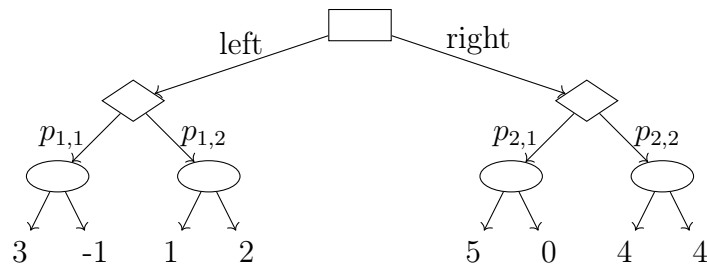
Brzy si všimneme, že libovolná lineární transformace listů nebude fungovat – to dokonce platí obecně, neboť ta vždy zachová optimální strategii. Listy je potřeba přehodnotit nějakým „nestejným“ způsobem. Řešení je nekonečně mnoho, jedno z nich ukazujeme (i s vyřešeným stromem).



Pro srovnání připomínáme, že v sekci ALFA-BETA jsme ukázali, že pro klasické MINIMAX stromy libovolná transformace zachovávající uspořádání listů zachová i optimální strategii.

#### Příklad 4.3.4.

Uvažujeme rozšířený MINIMAX strom s neznámými pravděpodobnostmi větvení.



- Nechť  $p_{1,1} = \frac{2}{7}$ . Co platí pro  $p_{2,1}, p_{2,2}$ , aby optimální strategie začínajícího hráče byla *right*?
- Co pro neznámé pravděpodobnosti platí obecně, má-li být optimální strategie *right*?
- Jak by musely vypadat hodnoty v listech, aby bez ohledu na hodnoty neznámých pravděpodobností byla optimální strategií *right*?

- Je-li  $p_{1,1} = \frac{2}{7}$ , ihned odvodíme i rovnost  $p_{1,2} = \frac{5}{7}$ . V tu chvíli je očekávaná hodnota levého náhodnostního uzlu známá – konkrétně  $\frac{3}{7}$ . Zapišeme hledanou nerovnost  $4p_{2,2} \geq \frac{3}{7}$ .
- V levém náhodnostním uzlu je obecná hodnota  $-1 \cdot p_{1,1} + 1 \cdot (1 - p_{1,1})$ . Po drobných úpravách získáváme konečný vztah  $1 \leq 4p_{2,2} + 2p_{1,1}$ .

c) Je třeba si uvědomit, že každý náhodnostní uzel se vyhodnotí nejvýše na hodnotu svého největšího potomka – a analogicky pro nejnižší hodnotu. Stačí tedy zajistit, aby nejvyšší možná hodnota levého náhodnostního uzlu byla menší než nejnižší hodnota náhodnostního uzlu vpravo. Tedy větší hodnota z obou minimalizačních uzlů vlevo musí vyšší než menší hodnota z minimalizačních uzlů vpravo. Nejjednodušším řešením je tedy pouze nahradit hodnotu uzlu s 0 za 1 (či větší).

## 5 Výroková logika

Výroková logika je základní formalismus, kterým můžeme modelovat jednoduchá tvrzení, jejich pravdivost, vyplývání a jiné vztahy mezi nimi. Staví na pojmu *výrok*, což je elementární tvrzení, o jehož pravdivosti má smysl uvažovat. Výroky reprezentujeme výrokovými proměnnými, které s využitím logických spojek (a, nebo, pokud-pak atp.) můžeme skládat do složitějších formulí a tím modelovat složitější tvrzení (jako např. souvětí).

### 5.1 Syntax

Syntax vyjadřuje, jakým způsobem ve výrokové logice zapisujeme platné formule – zatím však bez toho, abychom jim přiřazovali nějaký význam. První definice popisuje symboly, které používáme...

**Definice 11:** *Abeceda výrokové logiky zahrnuje*

- spočetně mnoho symbolů výrokových proměnných  $p, q, r, \dots$ ,
- symboly pro logické spojky  $\neg, \vee, \wedge, \Rightarrow, \Leftrightarrow, \dots$ ,
- pomocné symboly závorek ( a ).

Ozn.  $\mathcal{P} = \{p, q, r, \dots\}$  množinu symbolů výrokových proměnných.

... další už nám přesně říká, jak vypadá správně utvořená formule ve výrokové logice.

**Definice 12:** *Formule výrokové logiky.*

- Každý symbol  $p \in \mathcal{P}$  je formule.
- Jsou-li  $\varphi, \psi$  fomule, pak rovněž  $\neg(\varphi), (\varphi) \vee (\psi), (\varphi) \wedge (\psi), (\varphi) \Rightarrow (\psi), (\varphi) \Leftrightarrow (\psi), \dots$  jsou formule.

*Závorková konvence* umožňuje vynechávat přebytečné závorky, nedojde-li k porušení sémantické jednoznačnosti formule.

Ozn.  $\mathcal{F}$  množinu formulí výrokové logiky.

Poznáte nyní správně utvořenou formuli výrokové logiky?

---

**Příklad 5.1.1.** Uvažujte formule dle Definice 12. Rozhodněte, která z následujících slov jsou formule (neuvažujte závorkovou konvenci, tj. závorky musí přesně odpovídat definici):

- $w_1 = \neg(\neg p)$
- $w_2 = (p) \wedge (q) \wedge (r)$
- $w_3 = (p) \Rightarrow ((\neg(q)) \vee (\neg(r)))$
- $w_4 = ((p)\neg(q)) \vee (r)$
- $w_5 = (p) = (r)$

- a) Ne. Chybí závorky.
- b) Ne. Chybí závorky, nejednoznačné uzávorkování.
- c) Ano.
- d) Ne. Špatná arita spojek.
- e) Ne. Symbol = není logická spojka.

## 5.2 Sémantika

Sémantika popisuje, jak máme logické formule chápat, přiřazuje celému formalismu konkrétní význam. První důležitý pojem, interpretace, vyjadřuje, zda konkrétní elementární výroky považujeme za pravdivé či nepravdivé.

**Definice 13:** *Interpretace*  $I$  je zobrazení  $I : \mathcal{P} \rightarrow \{0, 1\}$  přiřazující pravdivostní hodnoty 0 (nepravda), 1 (pravda) jednotlivým výrokovým proměnným množiny  $\mathcal{P}$ .

V konkrétní interpretaci elementárních výroků pak můžeme vyhodnotit pravdivostní hodnotu celé formule.

**Definice 14:** *Valuace* (též *vyhodnocení*) příslušící interpretaci  $I$  je zobrazení  $I : \mathcal{F} \rightarrow \{0, 1\}$  přiřazující pravdivostní hodnoty 0 (nepravda), 1 (pravda) jednotlivým formulím z  $\mathcal{F}$ .  
(Jedná se o rozšíření interpretace z atomických formulí na všechny formule podle sémantiky logických spojek. Přesná definice je induktivní a zde ji neuvádíme.)

To s sebou přináší celou řadu pojmů, které blíže popisují vlastnosti dané formule, ať už obecně, či ve vztahu k dalším formulím.

**Definice 15:**

- Formule  $\varphi$  je *pravdivá v interpretaci*  $I$ , jestliže  $I(\varphi) = 1$  (tj. vyhodnotí se v příslušné valuaci na hodnotu 1). V opačném případě je formule *nepravdivá v interpretaci*  $I$ .
- Formule  $\varphi$  je *logicky pravdivá* či *tautologie*, jestliže je pravdivá v libovolné interpretaci.
- Formule  $\varphi$  je *kontradikce*, jestliže je nepravdivá v libovolné interpretaci.
- Formule  $\varphi$  je *splnitelná*, jestliže je pravdivá v nějaké interpretaci. Tato splňující interpretace se nazývá *modelem* formule  $\varphi$ .
- Formule  $\varphi, \psi$  jsou *sémanticky ekvivalentní*, značeno  $\varphi \approx \psi$ , jestliže  $I(\varphi) = I(\psi)$  pro libovolnou interpretaci  $I$ .

Podívejme se na vše zblízka na vzorovém příkladě.

**Příklad.** Mějme formuli výrokové logiky  $\varphi \equiv p \wedge (q \Rightarrow \neg p)$ . Sestavte pravdivostní tabulku formule  $\varphi$  a určete

- a) zda je  $\varphi$  pravdivá v interpretaci  $I(p) = 0, I(q) = 1$ ,
- b) zda je  $\varphi$  kontradikcí či tautologií,
- c) zda je  $\varphi$  splnitelná; případně nalezněte nějaký její model.

Pravdivostní tabulka formule vypadá následujícím způsobem.

ř.	$p$	$q$	$p \wedge (q \Rightarrow \neg p)$
1	0	0	0
2	0	1	0
3	1	0	1
4	1	1	0

Jednotlivé řádky odpovídají různým interpretacím (sloupečky  $p, q$ ) a jim příslušícím valuacím jednotlivých částí (podformulí) formule  $\varphi$  (následující sloupečky). Sloupeček s tučně vyznačenými hodnotami odpovídá valuacím formule  $\varphi$ .

- a) Formule  $\varphi$  je nepravdivá v interpretaci  $I$ , viz ř. 2.
- b) Formule  $\varphi$  není ani kontradikcí, ani tautologií. (Všechny zvýrazněné hodnoty by musely být 0, resp. 1.)
- c) Formule  $\varphi$  je splnitelná, neboť existuje interpretace, v níž je formule pravdivá. Viz ř. 3 tabulky. Modelem formule  $\varphi$  (jediným) je tedy interpretace  $I(p) = 1, I(q) = 0$ .

**Příklad 5.2.1. ★** Pro formuli výrokové logiky  $\psi \equiv (r \Rightarrow p) \vee \neg(q \wedge r)$  sestavte pravdivostní tabulku a rozhodněte, zda formule

- a) je logicky pravdivá,

- b) je splnitelná,
- c) je pravdivá v interpretaci  $I$ , kde  $I(p) = I(q) = 0, I(r) = 1$ ,
- d) je kontradikce či tautologie.

Nalezněte interpretaci  $I$ , kde  $I(q) = 0$ , takovou, že

- e)  $\psi$  je pravdivá v  $I$ ,
- f)  $\psi$  je nepravdivá v  $I$ ,

případně ukažte, že taková interpretace neexistuje.

Pravdivostní tabulka formule vypadá následujícím způsobem.

ř.	$p$	$q$	$r$	$(r \Rightarrow p)$	$\vee$	$\neg$	$(q \wedge r)$
1	0	0	0	0	1	0	1
2	0	0	1	1	0	0	1
3	0	1	0	0	1	0	1
4	0	1	1	1	0	0	1
5	1	0	0	0	1	1	0
6	1	0	1	1	1	1	0
7	1	1	0	0	1	1	1
8	1	1	1	1	1	1	1

- a) Formule  $\psi$  není logicky pravdivá, neboť existuje interpretace, v níž není formule pravdivá. (Viz čtvrtý řádek tabulky.)
- b) Formule  $\psi$  je splnitelná, neboť existuje interpretace, v níž je formule pravdivá. (Viz např. první řádek tabulky.)
- c) Formule  $\psi$  je pravdivá v interpretaci  $I$ . (Viz druhý řádek tabulky.)
- d) Formule  $\psi$  není kotradikce, neboť je splnitelná.
- e) Takovou interpretací je např  $I(p) = I(q) = I(r) = 0$ .
- f) Taková interpretace neexistuje.

**Příklad 5.2.2.** ★ Bez použití pravdivostních tabulek rozhodněte, zda jsou následující formule tautologie.

- a)  $\varphi \equiv (\neg p \Rightarrow (q \wedge \neg q)) \Rightarrow p$
- b)  $\psi \equiv (p \Rightarrow p) \Rightarrow (p \wedge \neg(q \Rightarrow p))$

Obecně: snažíme se nalézt interpretaci, v níž je formule nepravdivá (a tedy dokázat, že se nejedná o tautologii). Vedlejším produktem může být důkaz, že taková interpretace neexistuje, takže formule tautologií je.

- a) Uvažme interpretaci  $I$  takovou, že  $I(\varphi) = 0$ . Pak ze sémantiky implikace nutně  $I(p) = 0$  a  $I(\neg p \Rightarrow (q \wedge \neg q)) = 1$ . Ovšem z  $I(q \wedge \neg q) = 0$  (kontradikce) a  $I(p) = 0$  plyne, že  $I(\neg p \Rightarrow (q \wedge \neg q)) = 0$ , spor. Interpretace  $I$  neexistuje a  $\varphi$  je tautologie.

- b) Hledáme  $I$  s  $I(\psi) = 0$ . Formule  $p \Rightarrow p$  je tautologie, takže musí být  $I(p \wedge \neg(q \Rightarrow p)) = 0$ . To zjevně platí např. pro libovolné  $I$  s  $I(p) = 0$ . Čili  $\psi$  není tautologie.

**Příklad 5.2.3.** Mějme formuli výrokové logiky  $\varphi \equiv p \Rightarrow \neg(\neg q \vee r)$ . Sestavte pravdivostní tabulku formule  $\varphi$  a určete

- zda je  $\varphi$  pravdivá v interpretaci  $I(p) = 0, I(q) = I(r) = 1$ ,
- zda je  $\varphi$  kontradikcí či tautologií,
- zda je  $\varphi$  splnitelná; případně nalezněte nějaký její model.

Pravdivostní tabulka formule vypadá následujícím způsobem.

ř.	$p$	$q$	$r$	$p$	$\Rightarrow$	$\neg$	$(\neg q$	$\vee$	$r)$
1	0	0	0	0	<b>1</b>	0	1	1	0
2	0	0	1	0	<b>1</b>	0	1	1	1
3	0	1	0	0	<b>1</b>	1	0	0	0
4	0	1	1	0	<b>1</b>	0	0	1	1
5	1	0	0	1	<b>0</b>	0	1	1	0
6	1	0	1	1	<b>0</b>	0	1	1	1
7	1	1	0	1	<b>1</b>	1	0	0	0
8	1	1	1	1	<b>0</b>	0	0	1	1

- Formule  $\varphi$  je pravdivá v interpretaci  $I$ , viz ř. 4.
- Formule  $\varphi$  není ani kontradikcí, ani tautologií.
- Formule  $\varphi$  je splnitelná, neboť existuje interpretace, v níž je formule pravdivá. Viz např. ř. 3 tabulky. Modelem formule  $\varphi$  je tedy např. interpretace  $I(p) = I(r) = I(q) = 0$ .

**Příklad 5.2.4.** Mějme formuli výrokové logiky  $\varphi \equiv (p \wedge q) \Leftrightarrow (\neg q \wedge r)$ . *Bez použití pravdivostní tabulky* určete

- zda je  $\varphi$  pravdivá v interpretaci  $I(p) = 0, I(q) = I(r) = 1$ ,
- zda je  $\varphi$  kontradikcí,
- zda je  $\varphi$  tautologií,
- zda je  $\varphi$  splnitelná; případně nalezněte nějaký její model.

- Vyhodnotíme jednotlivé části formule v dané interpretaci:  $I(p \wedge q) = 0, I(\neg q \wedge r) = 0$ , celkem tedy  $I(\varphi) = 1$ .
- Formule  $\varphi$  není kontradikcí, neboť je pravdivá v interpretaci  $I$  z bodu a).



- c) Pokusíme se nalézt interpretaci  $I$ , v níž bude formule  $\varphi$  nepravdivá, abychom vyvrátili, že se jedná o tautologii. Formule bude v  $I$  nepravdivá, pokud se v ní jednotlivé operandy spojky  $\Leftrightarrow$  vyhodnotí na různé hodnoty, čili  $I(p \wedge q) \neq I(\neg q \wedge r)$ . Lehce lze vidět, že tuto vlastnost má např.  $I(p) = I(q) = 0, I(r) = 1$ . Formule  $\varphi$  tedy není tautologií.
- d) Formule  $\varphi$  je splnitelná, neboť existuje interpretace, v níž je formule pravdivá. Modelem je např. interpretace  $I$  z bodu a).

**Příklad 5.2.5.** Udejte příklad formule  $\varphi$  takové, že:

- a)  $\varphi$  obsahuje právě 3 různé výrokové proměnné a je pravdivá právě ve 3 interpretacích,
- b)  $\varphi$  obsahuje právě 3 různé výrokové proměnné, je pravdivá právě ve 3 interpretacích a obsahuje pouze logickou spojku  $\Rightarrow$ .
- c)  $\varphi$  je kontradikce, obsahuje právě 2 různé výrokové proměnné, každou dvakrát.

(Uvažujte interpretaci jako zobrazení přiřazující hodnoty právě výrokovým proměnným vyskytujícím se v  $\varphi$ .)

- a) Např.  $\varphi \equiv p \wedge (q \vee r)$ .
- b) Taková neexistuje.
- c) Např.  $\varphi \equiv (p \wedge \neg p) \vee (q \wedge \neg q)$ .

**Příklad 5.2.6.** Zjistěte, kolik existuje vzájemně neekvivalentních formulí výrokové logiky

- a) obsahujících pouze 3 výrokové proměnné  $p_1, p_2, p_3$  (i vícekrát),
- b) obsahujících pouze  $n$  výrokových proměnných  $p_1, \dots, p_n$  (i vícekrát),
- c) obsahujících pouze  $n$  výrokových proměnných  $p_1, \dots, p_n$  (i vícekrát) pravdivých právě v polovině interpretací.

- a) Máme  $2^3 = 8$  různých interpretací, v každé 2 možné hodnoty valuace formule, celkem tedy  $2^{2^3} = 256$  neekvivalentních formulí.
- b) Máme  $2^n$  různých interpretací, v každé 2 možné hodnoty valuace formule, celkem tedy  $2^{2^n}$  neekvivalentních formulí.
- c) Máme  $2^n$  různých interpretací, z nichž polovina ( $2^{n-1}$ ) má být pravdivá. Možností výběru pravdivých interpretací je  $\binom{2^n}{2^{n-1}}$  a stejně je i vzájemně neekvivalentních formulí.

### 5.3 Normální formy

Normální formy představují speciální způsob zápisu, z kterého lze přímo vyčíst některé sémantické vlastnosti formule. Základní stavební kameny formulí v normální formě jsou literály, z nichž skládáme klauzule, resp. duální klauzule.

#### Definice 16:

- *Literál* je výroková proměnná nebo její negace.
- *Klauzule* je disjunkce literálů.
- *Duální klauzule* je konjunkce literálů.

Spojením klauzulí konjunkcí nebo duálních klauzulí disjunkcí získáme formuli v konjunktivní, resp. disjunktivní normální formě. Když navíc každá klauzule obsahuje všechny uvažované výrokové proměnné, bavíme se o úplné normální formě.

#### Definice 17: Uvažujme výrokové proměnné $p_1, \dots, p_n$ .

- Formule je v *konjunktivní normální formě (KNF)*, jedná-li se o konjunkci klauzulí (s navzájem různými množinami literálů).
- Formule je v *disjunktivní normální formě (DNF)*, jedná-li se o disjunkci duálních klauzulí (s navzájem různými množinami literálů).
- Obsahuje-li navíc každá klauzule (resp. duální klauzule) formule v KNF (resp. DNF) každou z výrokových proměnných  $p_1, \dots, p_n$  právě jednou, hovoříme o *úplné* konjunktivní (resp. disjunktivní) normální formě (ÚKNF, resp. ÚDNF)

Formule  $(p \vee \neg q) \wedge (\neg p \vee \neg q \vee r)$  je v KNF, ale není v ÚKNF, protože první klauzule neobsahuje všechny implicitně uvažované výrokové proměnné  $p, q, r$ . Ekvivalentní formulí v ÚKNF je formule  $(p \vee \neg q \vee r) \wedge (p \vee \neg q \vee \neg r) \wedge (\neg p \vee \neg q \vee r)$ .

Pro převod do úplných normálních forem existuje poměrně přímočarý algoritmus, který je ilustrován na následujícím příkladu.

**Příklad.** Pomocí pravdivostní tabulky převed'te formuli  $p \Leftrightarrow q$  do ÚDNF a ÚKNF.

Vytvoříme pravdivostní tabulku zadané formule a pro interpretace, v nichž je formule pravdivá, přidáme duální klauzuli do ÚDNF; pro interpretace, v nichž je formule nepravdivá, přidáme klauzuli do ÚKNF.

Klauzuli pro danou interpretaci  $I$  tvoříme tak, že pro každou výrokovou proměnnou  $p$ :

- je-li  $I(p) = 0$ , přidáme do klauzule literál  $p$ ,
- je-li  $I(p) = 1$ , přidáme do klauzule literál  $\neg p$ .

V případě duální klauzule pro každou výrokovou proměnnou  $p$ , je-li  $I(p) = 0$ , přidáme literál  $\neg p$ , je-li  $I(p) = 1$ , přidáme literál  $p$ .

$p$	$q$	$p \Leftrightarrow q$	duální klauzule	klauzule
0	0	1	$(\neg p \wedge \neg q)$	
0	1	0		$(p \vee \neg q)$
1	0	0		$(\neg p \vee q)$
1	1	1	$(p \wedge q)$	

ÚDNF:  $(\neg p \wedge \neg q) \vee (p \wedge q)$

ÚKNF:  $(p \vee \neg q) \wedge (\neg p \vee q)$

**Příklad 5.3.1.** ★ Pomocí pravdivostních tabulek nalezněte ÚDNF (úplnou disjunktivní normální formu) a ÚKNF (úplnou konjunktivní normální formu) formule  $\psi \equiv (r \Rightarrow p) \vee \neg(q \wedge r)$  z Příkladu 5.2.1.

ÚDNF:

- $\psi \approx (\neg p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge r) \vee (\neg p \wedge q \wedge \neg r) \vee (p \wedge \neg q \wedge \neg r) \vee (p \wedge \neg q \wedge r) \vee (p \wedge q \wedge \neg r) \vee (p \wedge q \wedge r)$

ÚKNF:

- $\psi \approx (p \vee \neg q \vee \neg r)$

**Příklad 5.3.2.** Použitím ekvivalentních úprav nalezněte KNF následujících formulí.

- ★  $\varphi \equiv (p \Rightarrow q) \Rightarrow r$
- $\eta \equiv (p \Rightarrow q) \Leftrightarrow (p \Rightarrow r)$
- $\psi \equiv \neg(p \Leftrightarrow \neg q) \vee \neg r$

a)  $\varphi \equiv (p \Rightarrow q) \Rightarrow r$

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

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

b)  $\eta \equiv (p \Rightarrow q) \Leftrightarrow (p \Rightarrow r)$

$$\approx ((p \Rightarrow q) \Rightarrow (p \Rightarrow r)) \wedge ((p \Rightarrow r) \Rightarrow (p \Rightarrow q))$$

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

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

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

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

$$\begin{aligned}
&\approx (\neg p \vee \neg q \vee r) \wedge (\neg p \vee q \vee \neg r) \\
&\quad \neg(p \Leftrightarrow \neg q) \vee \neg r && \text{ekvivalence pomocí dvou implikací} \\
&\approx \neg((p \Rightarrow \neg q) \wedge (\neg q \Rightarrow p)) \vee \neg r && \text{eliminace implikace} \\
&\approx \neg((\neg p \vee \neg q) \wedge (q \vee p)) \vee \neg r && \text{de Morgan} \\
&\approx (\neg(\neg p \vee \neg q) \vee \neg(q \vee p)) \vee \neg r && \text{de Morgan} \\
&\approx ((p \wedge q) \vee (\neg q \wedge \neg p)) \vee \neg r && \text{distributivita zleva} \\
&\approx \left( ((p \wedge q) \vee \neg q) \wedge ((p \wedge q) \vee \neg p) \right) \vee \neg r && \text{distributivita zprava} \\
\text{c) } &\approx \left( ((p \vee \neg q) \wedge (q \vee \neg q)) \wedge ((p \wedge q) \vee \neg p) \right) \vee \neg r && \text{eliminace tautologie v konjunkci} \\
&\approx \left( (p \vee \neg q) \wedge ((p \wedge q) \vee \neg p) \right) \vee \neg r && \text{distributivita zprava} \\
&\approx \left( (p \vee \neg q) \wedge ((p \vee \neg p) \wedge (q \vee \neg p)) \right) \vee \neg r && \text{eliminace tautologie v konjunkci} \\
&\approx ((p \vee \neg q) \wedge (q \vee \neg p)) \vee \neg r && \text{distributivita zprava} \\
&\approx (p \vee \neg q \vee \neg r) \wedge (q \vee \neg p \vee \neg r) && \text{formule v ÚKNF}
\end{aligned}$$

**Příklad 5.3.3.** Použitím ekvivalentních úprav nalezněte DNF následujících formulí.

- a)  $\star \psi \equiv (p \Leftrightarrow q) \Rightarrow (r \vee s)$   
b)  $\varphi \equiv (p \Rightarrow q) \Rightarrow r$   
c)  $\eta \equiv (p \vee \neg(\neg r \Rightarrow q)) \wedge (r \Rightarrow p)$

$$\begin{aligned}
\text{a) } &\psi \equiv (p \Leftrightarrow q) \Rightarrow (r \vee s) \\
&\approx ((p \Rightarrow q) \wedge (q \Rightarrow p)) \Rightarrow (r \vee s) \\
&\approx \neg((\neg p \vee q) \wedge (\neg q \vee p)) \vee (r \vee s) \\
&\approx (\neg(\neg p \vee q) \vee \neg(\neg q \vee p)) \vee r \vee s \\
&\approx (p \wedge \neg q) \vee (q \wedge \neg p) \vee r \vee s \\
\text{b) } &\varphi \equiv (p \Rightarrow q) \Rightarrow r \\
&\approx \neg(\neg p \vee q) \vee r \\
&\approx (p \wedge \neg q) \vee r \\
&\quad (p \vee \neg(\neg r \Rightarrow q)) \wedge (r \Rightarrow p) && \text{eliminace implikace} \\
&\approx (p \vee \neg(r \vee q)) \wedge (\neg r \vee p) && \text{de Morgan} \\
&\approx (p \vee (\neg r \wedge \neg q)) \wedge (\neg r \vee p) && \text{distributivita zprava} \\
&\approx \left( (p \wedge (\neg r \vee p)) \vee ((\neg r \wedge \neg q) \wedge (\neg r \vee p)) \right) && \text{distributivita zleva} \\
\text{c) } &\approx \left( ((p \wedge \neg r) \vee (p \wedge p)) \vee ((\neg r \wedge \neg q) \wedge (\neg r \vee p)) \right) && \text{zjednodušení} \\
&\approx \left( ((p \wedge \neg r) \vee p) \vee ((\neg r \wedge \neg q) \wedge (\neg r \vee p)) \right) && \text{distributivita zleva} \\
&\approx \left( ((p \wedge \neg r) \vee p) \vee (((\neg r \wedge \neg q) \wedge \neg r) \vee ((\neg r \wedge \neg q) \wedge p)) \right) && \text{zjednodušení, závorky} \\
&\approx (p \wedge \neg r) \vee p \vee (\neg r \wedge \neg q) \vee (\neg r \wedge \neg q \wedge p) && \text{formule v DNF}
\end{aligned}$$

Z výsledné formule lze dále odstranit přebytečné duální klauzule  $p \wedge \neg r$ ,  $\neg r \wedge \neg q \wedge p$  – místo těch je postačující klauzule  $p$ . Výsledná formule je pak tvaru  $p \vee (\neg r \wedge \neg q)$ .

**Příklad 5.3.4.** Pomocí pravdivostní tabulky převed'te formuli  $r \Rightarrow \neg s$  do ÚDNF.

Vytvoříme pravdivostní tabulku formule a pro každou splňující interpretaci do ÚDNF přidáme odpovídající duální klauzuli:  $(\neg r \wedge \neg s) \vee (\neg r \wedge s) \vee (r \wedge \neg s)$ .

**Příklad 5.3.5.** Pomocí pravdivostní tabulky převed'te formuli  $\neg r \Rightarrow (r \wedge \neg s)$  do ÚKNF.

Vytvoříme pravdivostní tabulku formule a pro každou nesplňující interpretaci do ÚKNF přidáme odpovídající klauzuli:  $(r \vee s) \wedge (r \vee \neg s)$ .

**Příklad 5.3.6.** Ekvivalentními úpravami převed'te formuli  $(\neg q \wedge r) \vee (r \wedge \neg p)$  do ÚDNF.

$$\begin{aligned}
 & (\neg q \wedge r) \vee (r \wedge \neg p) && \text{rozšíření o tautologii} \\
 \approx & ((\neg q \wedge r) \wedge (p \vee \neg p)) \vee (r \wedge \neg p) && \text{distributivita zleva} \\
 \approx & ((\neg q \wedge r \wedge p) \vee (\neg q \wedge r \wedge \neg p)) \vee (r \wedge \neg p) && \text{rozšíření o tautologii} \\
 \approx & ((\neg q \wedge r \wedge p) \vee (\neg q \wedge r \wedge \neg p)) \vee ((q \vee \neg q) \wedge (r \wedge \neg p)) && \text{distributivita zprava} \\
 \approx & ((\neg q \wedge r \wedge p) \vee (\neg q \wedge r \wedge \neg p)) \vee ((q \wedge r \wedge \neg p) \vee (\neg q \wedge r \wedge \neg p)) && \text{odstranění duplicity} \\
 \approx & (\neg q \wedge r \wedge p) \vee (\neg q \wedge r \wedge \neg p) \vee (q \wedge r \wedge \neg p) && \text{formule v ÚDNF}
 \end{aligned}$$

## 5.4 Množiny formulí, splnitelnost, vyplývání

Občas je potřeba pracovat s několika tvrzeními zároveň a zkoumat vztahy mezi nimi. Proto zavádíme následující pojmy.

**Definice 18:** Množina formulí  $\mathcal{T}$  je *splnitelná*, jestliže existuje interpretace  $I$ , v níž jsou pravdivé všechny formule  $\varphi \in \mathcal{T}$ . Potom říkáme, že množina  $\mathcal{T}$  je *pravdivá v interpretaci  $I$*  a tato interpretace se nazývá *modelem* množiny  $\mathcal{T}$ . Není-li množina splnitelná, mluvíme o množině *nesplnitelné*.

Všimněte si, že prázdná množina  $\emptyset$  je splnitelná a je pravdivá v libovolné interpretaci. Dále zavedeme pojem logického vyplývání.

**Definice 19:** Formule  $\varphi$  logicky vyplývá z množiny formulí  $\mathcal{T}$ , zapisováno  $\mathcal{T} \models \varphi$ , právě když je formule  $\varphi$  pravdivá v každém modelu množiny  $\mathcal{T}$ .

Jinak řečeno, formule (závěr) logicky vyplývá z množiny předpokladů, právě když je pravdivá v každé interpretaci, v níž je pravdivá i množina předpokladů.

Vyplývá-li tedy formule  $\varphi$  z prázdné množiny,  $\emptyset \models \varphi$ , jedná se o tautologii (rozmyslete si). Píšeme jenom  $\models \varphi$ . V podobném duchu vynecháváme množinové závorky, je-li množina předpokladů jednoprvková, tedy  $\psi \models \varphi$  namísto  $\{\psi\} \models \varphi$ .

Následující věta dále propojuje logickou spojkou implikace s logickým vyplýváním.

**Věta 20 (o dedukci):** Logické vyplývání  $\{\psi_1, \dots, \psi_n\} \models \varphi$  platí, právě když platí vyplývání  $\{\psi_i, \dots, \psi_{n-1}\} \models \psi_n \Rightarrow \varphi$ .

Následující vzorový příklad ilustruje právě zavedené koncepty.

**Příklad.** Uvažujte množinu formulí  $\mathcal{T} = \{\neg p \Rightarrow r, \neg(q \wedge r), p \vee \neg q\}$ . Sestavte pravdivostní tabulku množiny  $\mathcal{T}$  a určete

- zda je  $\mathcal{T}$  pravdivá v interpretaci  $I(p) = I(q) = I(r) = 0$ ,
- zda je  $\mathcal{T}$  splnitelná; případně nalezněte nějaký její model,
- zda platí logické vyplývání  $\mathcal{T} \models p \wedge \neg r$ ,
- zda platí logické vyplývání  $\mathcal{T} \models r \Rightarrow \neg q$ .

Pravdivostní tabulka množiny  $\mathcal{T}$  vypadá následujícím způsobem.

ř.	$p$	$q$	$r$	$\neg p \Rightarrow r$	$\neg(q \wedge r)$	$p \vee \neg q$	$\mathcal{T}$	$p \wedge \neg r$	$r \Rightarrow \neg q$
1	0	0	0	0	1	1	0	0	1
2	0	0	1	1	1	1	1	0	1
3	0	1	0	0	1	0	0	0	1
4	0	1	1	1	0	0	0	0	0
5	1	0	0	1	1	1	1	1	1
6	1	0	1	1	1	1	1	0	1
7	1	1	0	1	1	1	1	1	1
8	1	1	1	1	0	1	0	0	0

- Množina  $\mathcal{T}$  není pravdivá v interpretaci  $I$ , viz ř. 1.
- Množina  $\mathcal{T}$  je splnitelná, neboť existuje interpretace, v níž je množina pravdivá. Viz např. ř. 2 tabulky. Modelem množiny  $\mathcal{T}$  je tedy např. interpretace  $I(p) = I(q) = 0$ ,  $I(r) = 1$ .

- c) Vyplývání neplatí. Existuje interpretace, v níž je pravdivá množina  $\mathcal{T}$  (předpoklad), ale neplatí závěr. Viz např. ř. 2.
- d) Vyplývání platí. Ve všech interpretacích, v nichž je pravdivá množina  $\mathcal{T}$  (předpoklad), je platný i závěr. Viz ř. 2, 5, 6, 7.

**Příklad 5.4.1.** ★ Určete splnitelnost následujících množin formulí. Je-li množina splnitelná, naleznete nějaký její model. Vyhněte se použití pravdivostních tabulek.

- a)  $\mathcal{T}_1 = \{(p \Rightarrow q) \wedge r, q \wedge r, r \Rightarrow s, p \wedge \neg s\}$
- b)  $\mathcal{T}_2 = \{(p \vee q) \Leftrightarrow r, r, \neg p, q\}$

Aby byla množina  $\mathcal{T}$  logických formulí splnitelná, je nutné, aby existovala interpretace  $I$ , v níž je pravdivá každá formule  $\varphi \in \mathcal{T}$ . (Nestačí, aby každá formule množiny byla splnitelná!) Hledáme tedy splňující interpretace jednotlivých množin.

- a) Formule  $q \wedge r$  vyžaduje, aby pro splňující interpretaci  $I_1$  platilo  $I_1(q) = I_1(r) = 1$ , a formule  $p \wedge \neg s$  vyžaduje  $I_1(p) = 1$  a  $I_1(s) = 0$ . V každé takové interpretaci je ovšem  $I_1(r \Rightarrow s) = 0$ . Množina  $\mathcal{T}_1$  je tedy nesplnitelná.
- b) Formule  $r, \neg p, q$  vyžadují  $I_2(r) = I_2(q) = 1$  a  $I_2(p) = 0$ . V takové interpretaci pak je  $I_2((p \vee q) \Leftrightarrow r) = 1$ .  $I_2$  tedy splňuje všechny formule v množině  $\mathcal{T}_2$ , a tudíž  $\mathcal{T}_2$  je splnitelná a  $I_2$  je jejím modelem.

**Příklad 5.4.2.** ★ Je množina formulí  $\mathcal{T} = \emptyset$  splnitelná? Dokažte.

Ano, prázdná množina je splnitelná. Uvažme libovolnou interpretaci  $I$ . Všechny formule jsou v  $I$  pravdivé, a tudíž  $I$  je modelem  $\mathcal{T}$ .

**Příklad 5.4.3.** ★ Ukažte, že pro každou množinu formulí  $\mathcal{T}$  existuje formule  $\varphi$ , která je pravdivá v interpretaci  $I$ , právě když je  $I$  modelem  $\mathcal{T}$ .

Pro neprázdnou množinu formulí  $\mathcal{T} = \{\varphi_1, \dots, \varphi_n\}$  bude odpovídající formulí  $\varphi \equiv \varphi_1 \wedge \dots \wedge \varphi_n$ . Prázdné množině pak bude odpovídat libovolná tautologie, např.  $\varphi \equiv p \vee \neg p$ .

**Příklad 5.4.4.** Uvažujte množinu formulí  $\mathcal{T} = \{r \wedge \neg q, \neg(\neg r \wedge p), p \Rightarrow r\}$ . Sestavte pravdivostní tabulku množiny  $\mathcal{T}$  a určete

- a) zda je  $\mathcal{T}$  pravdivá v interpretaci  $I(p) = I(q) = I(r) = 1$ ,
- b) zda je  $\mathcal{T}$  splnitelná; případně naleznete nějaký její model,
- c) zda platí logické vyplývání  $\mathcal{T} \models r$ ,

d) zda platí logické vyplývání  $\mathcal{T} \models r \wedge p$ .

Pravdivostní tabulka množiny  $\mathcal{T}$  vypadá následujícím způsobem.

ř.	$p$	$q$	$r$	$r \wedge \neg q$	$\neg(\neg r \wedge p)$	$p \Rightarrow r$	$\mathcal{T}$	$r$	$r \wedge p$
1	0	0	0	0	1	1	0	0	0
2	0	0	1	1	1	1	1	1	0
3	0	1	0	0	1	1	0	0	0
4	0	1	1	0	1	1	0	1	0
5	1	0	0	0	0	0	0	0	0
6	1	0	1	1	1	1	1	1	1
7	1	1	0	0	0	0	0	0	0
8	1	1	1	0	1	1	0	1	1

- Množina  $\mathcal{T}$  není pravdivá v interpretaci  $I$ , viz ř. 8.
- Množina  $\mathcal{T}$  je splnitelná, neboť existuje interpretace, v níž je množina pravdivá. Viz např. ř. 2 tabulky. Modelem množiny  $\mathcal{T}$  je tedy např. interpretace  $I(p) = I(q) = 0$ ,  $I(r) = 1$ .
- Vyplývání platí. Ve všech interpretacích, v nichž je pravdivá množina  $\mathcal{T}$  (předpoklad), je platný i závěr. Viz ř. 2, 6.
- Vyplývání neplatí. Existuje interpretace, v níž je pravdivá množina  $\mathcal{T}$  (předpoklad), ale neplatí závěr. Viz např. ř. 2.

**Příklad 5.4.5.** Uvažujte množinu formulí  $\mathcal{T} = \{(p \Rightarrow q) \vee \neg r, \neg p \Rightarrow s, \neg t \vee p \vee \neg q\}$ . Zadejte formuli  $\varphi$  tak, aby množina  $\mathcal{T} \cup \{\varphi\}$  byla nespjitelná.

Např.  $\varphi \equiv p \wedge \neg p$ .

**Příklad 5.4.6.** Rozhodněte, zda pro libovolnou neprázdnou množinu formulí  $\mathcal{T}$  existuje formule  $\varphi$  taková, že  $\varphi$  není kontradikce a  $\mathcal{T} \cup \{\varphi\}$  je nespjitelná. Své tvrzení dokažte. (Argumentujte formálně s využitím pojmů interpretace, valuace atp.)

Neexistuje, dokážeme sporem. Uvažujme  $\mathcal{T} = \{p \vee \neg p\}$  a necht' je  $\varphi$  formule taková, že  $\varphi$  není kontradikce a  $\mathcal{T} \cup \{\varphi\}$  je nespjitelná. Jelikož  $\varphi$  není kontradikce, existuje interpretace  $I$  taková, že  $I(\varphi) = 1$ . Formule  $p \vee \neg p$  je však pravdivá v libovolné interpretaci, a tedy  $\mathcal{T} \cup \{\varphi\} = \{p \vee \neg p, \varphi\}$  je pravdivá v interpretaci  $I$ , spor s nespjitelností  $\mathcal{T} \cup \{\varphi\}$ .

**Příklad 5.4.7. ★** Uvažujte množinu formulí výrokové logiky  $\mathcal{T}$  a formule  $\varphi, \psi$ . Rozhodněte o platnosti následujících tvrzení:

- Pokud  $\mathcal{T} \models \varphi$  a  $\mathcal{T} \models \psi$ , pak  $\mathcal{T} \models \varphi \vee \psi$ .



- b) Pokud  $\mathcal{T} \models \varphi$  a  $\mathcal{T} \models \psi$ , pak  $\mathcal{T} \models \varphi \wedge \psi$ .
- c) Pokud  $\mathcal{T} \models \varphi \vee \psi$ , pak  $\mathcal{T} \models \varphi$  nebo  $\mathcal{T} \models \psi$ .
- d) Pokud  $\mathcal{T} \models \varphi \wedge \psi$ , pak  $\mathcal{T} \models \varphi$  a  $\mathcal{T} \models \psi$ .
- e) Pokud  $\mathcal{T} \models \varphi$ , pak  $\mathcal{T} \models \varphi \vee \psi$ .
- f) Pokud  $\mathcal{T} \models \varphi$ , pak  $\mathcal{T} \models \varphi \wedge \psi$ .
- g) Pokud  $\mathcal{T} \models \varphi$ , pak  $\mathcal{T} \cup \{\psi\} \models \varphi$ .
- h) Pokud  $\mathcal{T} \cup \{\psi\} \models \varphi$ , pak  $\mathcal{T} \models \varphi$ .

- a) Platí.
- b) Platí.
- c) Neplatí. Např.  $p \vee q \models p \vee q$ , ale  $p \vee q \not\models q$  ani  $p \vee q \not\models p$ .
- d) Platí.
- e) Platí.
- f) Neplatí.
- g) Platí.
- h) Neplatí.

**Příklad 5.4.8.** Ukažte, že logické vyplývání  $\mathcal{T} \models \varphi$  platí, právě když množina  $\mathcal{T} \cup \{\neg\varphi\}$  je nesplnitelná.

*Poznámka: Tato vlastnost odpovídá principu důkazu sporem. Znova ji uvidíme v části zabývající se rezolucí, kde ukazujeme vyplývání pomocí nesplnitelnosti množiny formulí.*

„ $\Rightarrow$ “: Platí-li  $\mathcal{T} \models \varphi$ , znamená to, že pro každou interpretaci  $I$  takovou, že  $I(\mathcal{T}) = 1$ , je  $I(\varphi) = 1$ , a tudíž  $I(\neg\varphi) = 0$ . Tedy v každé interpretaci, která by splnila  $\mathcal{T}$ , bude  $\neg\varphi$  nepravdivá, a tedy je v ní i  $\mathcal{T} \cup \{\neg\varphi\}$  nepravdivá.

„ $\Leftarrow$ “: Je-li  $\mathcal{T} \cup \{\neg\varphi\}$  nesplnitelná, znamená to, že v každé interpretaci  $I$  je  $I(\mathcal{T}) = 0$  nebo  $I(\neg\varphi) = 0$ . To znamená, že v každé interpretaci je buďto  $\mathcal{T}$  nepravdivá nebo  $\varphi$  pravdivá, a tedy  $\mathcal{T} \models \varphi$ .

**Příklad 5.4.9.** S využitím věty o dedukci (a bez použití pravdivostních tabulek) rozhodněte, zda je formule  $\xi \equiv p \Rightarrow (q \Rightarrow (\neg p \Rightarrow r))$  tautologie, kontraktikce, splnitelná formule.

## 5.5 Logické spojky

Se základními logickými spojkami ( $\vee, \wedge, \Rightarrow, \neg, \dots$ ) jsme se již na intuitivní úrovni setkali. Podívejme se na ně ještě jednou zblízka.

**Definice 21:** Sémantika  $n$ -ární logické spojky  $\square$  je dána funkcí  $f_{\square} : \{0,1\}^n \rightarrow \{0,1\}$  následujícím způsobem: Valuace formule  $\psi \equiv \square(\varphi_1, \dots, \varphi_n)$  v interpretaci  $I$  je dána jako  $I(\psi) = f_{\square}(I(\varphi_1), \dots, I(\varphi_n))$ , kde  $I(\varphi_1), \dots, I(\varphi_n)$  jsou valuace podformulí  $\varphi_1, \dots, \varphi_n$  příslušné interpretaci  $I$ .

Intuitivně, pro novou spojku můžeme zadat její sémantiku tabulkou (předepisující funkci  $f_{\square}$ ) a formule s touto spojku pak odpovídající způsobem vyhodnocovat. Všimněte si rovněž, že logické spojky nemusí být nutně unární (jako např.  $\neg$ ) nebo binární (jako např.  $\vee$ ), ale mohou mít libovolnou aritu.

Význam nové spojky ale můžeme definovat i ekvivalencí s jinou formulí. Lze tedy definovat např. nulární falsum (čili konstantní nepravdu) jako  $\perp : \approx p \wedge \neg p$ , či implikaci jako  $\varphi \Rightarrow \psi : \approx \neg \varphi \vee \psi$ .

**Příklad.** Definujte nulární verum  $\top$  (čili konstantní pravdu) ekvivalencí s jinou formulí. Definujte disjunkci tabulkou.

Nulární verum:  $\top : \approx p \vee \neg p$ .  
Disjunkce:

ř.	$\varphi$	$\psi$	$\varphi \vee \psi$
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	1

Ještě zavedeme pojem úplného systému logických spojek.

**Definice 22:** Množina logických spojek tvoří *úplný systém logických spojek*, pokud lze formulemi obsahujícími pouze spojky z dané množiny vyjádřit libovolnou logickou funkci (vnímáme-li formule jako funkce přiřazující zadané interpretaci svou valuaci).

Množina  $\{\neg, \vee, \wedge, \Rightarrow, \Leftrightarrow\}$  tvoří úplný systém logických spojek (důkaz vynecháváme). Ve skutečnosti, jak dále uvidíme, nám stačí pouze spojky  $\{\neg, \Rightarrow\}$ . Zavést ostatní spojky pomocí těchto dvou může v mnoha případech usnadnit argumentaci o formulích (stačí uvažovat o vlastnostech dvou spojek). V praxi (např. při tvorbě elektronických obvodů) se ještě používají úplné systémy logických spojek  $\{\text{NAND}\}$  a  $\{\text{NOR}\}$ <sup>2</sup>, které si vystačí s jedinou spojku.

---

<sup>2</sup> $\varphi \text{NAND} \psi : \approx \neg(\varphi \wedge \psi)$ ,  $\varphi \text{NOR} \psi : \approx \neg(\varphi \vee \psi)$

**Příklad 5.5.1.** Kolik existuje různých vzájemně neekvivalentních  $n$ -árních spojek?

Definující tabulka má  $2^n$  řádků a na každém může být jedna ze dvou hodnot 0, 1. Celkem je tedy  $2^{2^n}$  vzájemně neekvivalentních logických spojek.

**Příklad 5.5.2.** Uvažujte formule výrokové logiky nad abecedou obsahující pouze logické spojky  $\Rightarrow, \neg$ . Zaveďte spojky  $\vee, \wedge, \Leftrightarrow$  jako syntaktické zkratky v tomto systému tak, aby měly obvyklý význam. (Tj. pro slova  $\varphi \vee \psi, \varphi \wedge \psi, \varphi \Leftrightarrow \psi$ , kde  $\varphi, \psi$  jsou správně utvořené formule, uveďte formule v uvažovaném systému, které jsou jimi reprezentovány.)

$$\begin{aligned}\varphi \vee \psi &:\approx \neg\varphi \Rightarrow \psi \\ \varphi \wedge \psi &:\approx \neg(\varphi \Rightarrow \neg\psi) \\ \varphi \Leftrightarrow \psi &:\approx \neg((\varphi \Rightarrow \psi) \Rightarrow \neg(\psi \Rightarrow \varphi))\end{aligned}$$

**Příklad 5.5.3.** Mějme formuli výrokové logiky  $\varphi \equiv \neg(p \wedge (q \square r)) \vee (q \Rightarrow r)$ . Určete (např. tabulkou) sémantiku binární logické spojky  $\square$  tak, aby formule  $\varphi$  byla:

- tautologie,
- kontradikce,
- ekvivalentní formuli  $p \Rightarrow q$ ,

nebo konstatujte, že formule nemůže danou podmínku splnit. Odpovědi zdůvodněte.

Vytvoříme (neúplnou) pravdivostní tabulku formule  $\varphi$  a formule  $p \Rightarrow q$  pro bod c).

ř.	$p$	$q$	$r$	$\neg$	$(p \wedge (q \square r))$	$\vee$	$(q \Rightarrow r)$	$p \Rightarrow q$
1	0	0	0	1	0	0	1	1
2	0	0	1	1	0	0	1	1
3	0	1	0	1	0	0	1	1
4	0	1	1	1	0	0	1	1
5	1	0	0	?	1	?	0	0
6	1	0	1	?	1	?	0	0
7	1	1	0	?	1	?	1	1
8	1	1	1	?	1	?	1	1

- Aby byla  $\varphi$  tautologií, musí v interpretaci  $I(p) = I(q) = 1, I(r) = 0$  být  $I(\varphi) = 1$ , tedy  $I(\neg(p \wedge (q \square r))) = 1$ , a tedy  $I(p \wedge (q \square r)) = 0$ . Jelikož  $I(p) = 1$ , je nutné, aby spojka  $\square$  v interpretaci  $I$  splňovala  $I(q \square r) = 0$ . Definujeme např.  $q \square r :\approx r \wedge \neg r$  (tedy  $\square$  je binární falsum), případně tabulkou

ř.	$q$	$r$	$q \square r$
1	0	0	0
2	0	1	0
3	1	0	0
4	1	1	0

- b) Nelze. Např. v interpretaci  $I(p) = I(q) = I(r) = 0$  je  $I(\varphi) = 1$  nezávisle na sémantice spojky  $\square$ . Viz ř. 1 tabulky.
- c) Nelze. Např. v interpretaci  $I(p) = 1, I(q) = I(r) = 0$  je  $I(\varphi) = 1$  nezávisle na sémantice spojky  $\square$ , ale  $I(p \Rightarrow q) = 0$ . Viz ř. 5 tabulky.

**Příklad 5.5.4.** Kolika způsoby lze definovat sémantiku spojky  $\square$  z bodu a) předchozího příkladu?

Hodnota  $q \square r$  je přesně dána leda pro  $I(q) = 1, I(r) = 0$ . V ostatních 3 interpretacích může být hodnota libovolná, vždy jedna z možností 0, 1. Celkem lze tedy sémantiku spojky korektně dodefinovat  $2^3 = 8$  způsoby.

**Příklad 5.5.5.** Ukažte, že následující množiny tvoří úplné systémy logických spojek. Vyjděte z předpokladu, že  $\{\Rightarrow, \neg\}$  je úplný systém logických spojek. (Tj. vyjádřete formule  $\varphi \Rightarrow \psi$  a  $\neg\varphi$  pomocí formulí obsahující pouze spojky z příslušných množin.)

- {NOR}
- {NAND}

NOR:

$$\begin{aligned} \neg\varphi &\approx \varphi \text{NOR} \varphi \\ \varphi \Rightarrow \psi &\approx \neg\varphi \vee \psi \approx \neg(\neg\varphi \text{NOR} \psi) \approx \neg((\varphi \text{NOR} \varphi) \text{NOR} \psi) \\ &\approx ((\varphi \text{NOR} \varphi) \text{NOR} \psi) \text{NOR} ((\varphi \text{NOR} \varphi) \text{NOR} \psi) \end{aligned}$$

NAND:

$$\begin{aligned} \neg\varphi &\approx \varphi \text{NAND} \varphi \\ \varphi \Rightarrow \psi &\approx \neg\varphi \vee \psi \approx \neg(\varphi \wedge \neg\psi) \approx \varphi \text{NAND} \neg\psi \approx \varphi \text{NAND} (\psi \text{NAND} \psi) \end{aligned}$$

**Příklad 5.5.6.** Uvažujte ternární logickou spojku  $\Delta$  definovanou takto:  $\Delta(\varphi, \psi, \theta) := (\varphi \Rightarrow \psi) \wedge \neg\theta$ . Rozhodněte, zda množina  $\{\Delta\}$  tvoří úplný systém logických spojek, a svou odpověď dokažte. Můžete vyjít z předpokladu, že  $\{\neg, \Rightarrow\}$  je úplný systém logických spojek.

Ano.

**Příklad 5.5.7.** Uvažujte spojku  $\nrightarrow$  definovanou takto:  $\varphi \nrightarrow \psi \approx \neg(\varphi \Rightarrow \psi)$ . Rozhodněte, zda množina  $\{\nrightarrow\}$  tvoří úplný systém logických spojek, a svou odpověď dokažte. Můžete vyjít z předpokladu, že  $\{\neg, \Rightarrow\}$  je úplný systém logických spojek.

Ne.

## 6 Predikátová logika

Predikátová logika přirozeně rozšiřuje logiku výrokovou. Namísto výroků zde hovoříme o predikátech, na které můžeme pohlížet jako na parametrizované výroky, jejichž pravdivost závisí i na hodnotě parametru: např. pravdivost tvrzení „ $x$  je sudé (párne) číslo“ závisí na tom, zda se za hodnotou  $x$  schovává číslo 1 nebo 2.

### 6.1 Syntax

Než začneme zkoumat sémantiku predikátové logiky, podívejme se opět nejprve na její syntax.

**Definice 23:** *Abeceda predikátové logiky* zahrnuje

- symboly pro proměnné  $x, y, z, \dots$ ,
- predikátové symboly  $P, Q, R, \dots$ ,
- funkční symboly  $f, g, h, \dots$  (příp.  $a, b, c, d, \dots$  pro konstanty),
- symboly pro logické spojky  $\neg, \vee, \wedge, \Rightarrow, \Leftrightarrow, \dots$ ,
- pomocné symboly závorek  $( \ )$  a kvantifikátorů  $\forall$  a  $\exists$ .

Ozn.  $\mathcal{V} = \{x, y, z, \dots\}$  množinu všech proměnných.

Pozor, proměnná je v kontextu predikátové logiky něco jiného než výroková proměnná. Výrokovým proměnným ve skutečnosti odpovídají predikáty, které jsou jejich zobecněním. (Podrobněji dále.)

Novým a zcela klíčovým je zde pojem *termu*, který reprezentuje objekt a slouží jako argument pro predikáty.

**Definice 24:** *Term.*

- Každý symbol pro proměnnou je term.
- Je-li  $f$   $n$ -ární funkční symbol a  $t_1, \dots, t_n$  termy, pak také  $f(t_1, \dots, t_n)$  je term.

Nulární funkce, čili konstanty, zapisujeme bez prázdných závorek, tedy  $c$  namísto  $c()$ .

U běžných binárních funkčních symbolů lze používat i infixový zápis bez závorek. U běžných unárních symbolů lze používat i prefixový zápis bez závorek.

Máme-li k dispozici nulární funkční symbol  $c$ , binární funkční symboly  $f$ ,  $+$  a proměnné  $x, y, z, \dots$ , pak např.  $f(x, c)$ ;  $f(f(x, y), z)$ ;  $c$ ;  $+(x, z)$  jsou termy. Poslední uvedený lze navíc

v souladu s výše uvedenou konvencí zapisovat infixově jako  $x + z$ .

V matematickém kontextu tak například matematické výrazy (nikoliv rovnice!) představují termy:  $x + 2$ ,  $7 + 15$ ,  $x^2/0$ . Stále je potřeba mít na mysli, že termy reprezentují **hodnoty** či **objekty**, **nikoliv tvrzení**. Jak tedy reprezentovat tvrzení v predikátové logice?

**Definice 25:** *Formule predikátové logiky.*

- Je-li  $P$   $n$ -ární predikátový symbol a  $t_1, \dots, t_n$  termy, pak  $P(t_1, \dots, t_n)$  je formule.
- Jsou-li  $t_1, t_2$  termy, je  $t_1 = t_2$  formule.
- Jsou-li  $\varphi, \psi$  fomule, pak rovněž  $\neg(\varphi)$ ,  $(\varphi) \vee (\psi)$ ,  $(\varphi) \wedge (\psi)$ ,  $(\varphi) \Rightarrow (\psi)$ ,  $(\varphi) \Leftrightarrow (\psi)$ , ... jsou formule.
- Je-li  $\varphi$  fomule a  $x$  proměnná, pak rovněž  $\exists x(\varphi)$  a  $\forall x(\varphi)$  jsou formule.

U běžných binárních predikátových symbolů lze používat i infixový zápis bez závorek.

Formule vzniklá aplikací predikátového symbolu na termy se nazývá *atomická formule*.

V matematickém kontextu jsou matematické rovnice či nerovnice jako  $x+5 = y-1$ ,  $x^2+1 \leq x$  jednoduchými příklady formulí. Formulí však lze běžně vyjadřovat i složitější tvrzení:

$$\forall x \exists y (x + y = 0).$$

Použité funkční a predikátové symboly pak společně tvoří *jazyk* predikátové logiky  $\mathcal{L}$ , který doplňuje definice termu a formule konkrétním kontextem. Jazyk je v tomto kontextu pouze označení pro množinu funkčních a predikátových symbolů, čili jedná se o jiný pojem než formální jazyk.

**Příklad.** Ve formuli

$$\varphi \equiv \forall x (x \cdot x \geq 0 \Rightarrow \exists y (f(x, y, z) < \sin(0 + x))),$$

kde všechny symboly mají obvyklý (matematický) význam, identifikujte všechny

- a) proměnné,
- b) termy,
- c) funkční a predikátové symboly.

U funkčních a predikátových symbolů určete i jejich aritu.

- a)  $x, y, z$
- b)  $x, x \cdot x, 0, y, z, f(x, y, z), 0 + x, \sin(0 + x)$
- c) Symboly klasifikujeme v tabulce.

symbol	typ	arita	zápis
$\cdot$	funkční	2	infix
$\geq$	predikátový	2	infix
$0$	funkční (konstanta)	0	n/a
$f$	funkční	3	prefix
$\sin$	funkční	1	prefix
$<$	predikátový	2	infix
$+$	funkční	2	infix

Jazyk formule  $\varphi$  je  $\mathcal{L} = \{\cdot, \geq, 0, f, \sin, <, +\}$ .

Proměnné navíc mohou být ve formuli vázány kvantifikátorem (tedy patřit do „scope“ kvantifikátoru), nebo být volné.

**Definice 26:** Výskyt proměnné  $x$  ve formuli  $\varphi$  (výskytem rozumíme její použití v rámci termu, nikoli kvantifikátoru) je *vázaný*, existuje-li podformule  $\psi$ , ozn.  $\psi$ , která obsahuje tento výskyt a začíná  $\exists x$  nebo  $\forall x$ . V opačném případě je výskyt proměnné *volný*.

Formule, která nemá žádný volný výskyt proměnné se nazývá *uzavřená* nebo též *sentence*.

Běžně používané značení  $\varphi(x_1, \dots, x_n)$  znamená, že formule  $\varphi$  má volný výskyt proměnných  $x_1, \dots, x_n$ , ostatní proměnné jsou ve formuli vázané.

**Příklad.** Klasifikujte výskyty proměnných ve formuli  $\forall z(x = y \Rightarrow \exists xP(x, y)) \vee Q(x)$ .

Proměnná  $z$  nemá ve formuli žádný výskyt (za výskyt považujeme, když je proměnná součástí termu, ne kvantifikátoru). Proměnná  $x$  má ve formuli tři výskyty: nejprve volný, potom vázaný a nakonec opět volný. Proměnná  $y$  má ve formuli dva výskyty, oba volné.

**Příklad 6.1.1.** ★ Ve formuli

$$\varphi \equiv 2 \mid x \Rightarrow \exists y(y \cdot 2 = x \vee \sin(x + y) > 1),$$

kde všechny symboly mají obvyklý (matematický) význam, identifikujte (vč. arity, pokud to dává smysl) všechny

- proměnné (vč. jejich výskytů),
- termy,



- c) logické spojky,
- d) funkční a predikátové symboly.

U funkčních a predikátových symbolů určete i jejich aritu.

- a)  $x$  se třemi volnými výskyty,  $y$  se dvěma vázanými výskyty
- b)  $2, x, y, y \cdot 2, x + y, \sin(x + y), 1$
- c)  $\Rightarrow$  binární,  $\vee$  binární
- d) Symboly klasifikujeme v tabulce.

symbol	typ	arita	zápis
2	funkční (konstanta)	0	n/a
	predikátový	2	infix
·	funkční	2	infix
sin	funkční	1	prefix
+	funkční	2	infix
>	predikátový	2	infix
1	funkční (konstanta)	0	n/a

Jazyk formule  $\varphi$  je  $\mathcal{L} = \{2, |, \cdot, \sin, +, >, 1\}$ .

**Příklad 6.1.2.** ★ Uvažujte jazyk  $\mathcal{L}$  predikátové logiky obsahující funkční a predikátové symboly zadané následující tabulkou:

symbol	typ	arita
$f$	funkční	3
$d$	funkční	0
$P$	predikátový	1
$Q$	predikátový	2

Rozhodněte, která z následujících slov jsou **termy** a **formule** jazyka  $\mathcal{L}$ :

- a)  $Q(d, d)$ ,
- b)  $z$ ,
- c)  $f(f(d))$ ,
- d)  $f(d, f(d, d, d), d)$ ,
- e)  $y = x$ ,
- f)  $\forall x(Q(d, d) = x)$ ,
- g)  $\forall y(f(f(x, y, z), d, d))$ ,
- f)  $\forall x(Q(f(f(d, d, d), y, z), f(z, d, y)) \vee P(f(d, z, x)))$ .

- a) Jedná se o formuli.
- b) Jedná se o term.
- c) Nejedná se o term ani formuli, nerespektuje aritu funkčního symbolu  $f$ .
- d) Jedná se o term.
- e) Jedná se o formuli.
- f) Nejedná se o term ani formuli, predikátový symbol  $Q$  použit jako funkční symbol.
- g) Nejedná se o term ani formuli, funkční symbol  $f$  použit jako predikátový symbol.
- h) Jedná se o formuli.

**Příklad 6.1.3.** Uvažujte jazyk  $\mathcal{L}$  predikátové logiky obsahující jeden unární funkční symbol  $f$  a jeden nulární funkční symbol  $c$ . Popište množinu všech termů jazyka  $\mathcal{L}$ .

$$\{f^i(c); i \in \mathbb{N}_0\} \cup \{f^i(x); i \in \mathbb{N}_0, x \text{ je proměnná}\},$$

neboli

$$\{c, f(c), f(f(c)), \dots\} \cup \{x, f(x), f(f(x)), \dots, y, f(y), f(f(y)), \dots, z, \dots\}$$

**Příklad 6.1.4.** Uvažujte jazyk  $\mathcal{L}$  predikátové logiky obsahující funkční a predikátové symboly zadané následující tabulkou:

symbol	typ	arita
$f$	funkční	2
$c$	funkční	0
$P$	predikátový	1
$Q$	predikátový	2

Rozhodněte, která z následujících slov jsou **termy** jazyka  $\mathcal{L}$ :

- a)  $y = x$ ,
- b)  $Q(c, c)$ ,
- c)  $f(x, c)$ ,
- d)  $\forall y(f(f(c, c), y))$ ,
- e)  $z$ ,
- f)  $f(f(c))$ ,
- g)  $f(x, y) \vee f(c, x)$ ,
- h)  $f(f(c, x), f(y, z))$ .

- a) Ne, jedná se o formuli.
- b) Ne, jedná se o formuli.
- c) Ano.
- d) Ne, termy neobsahují kvantifikátory.
- e) Ano.
- f) Ne, nerespektuje aritu funkčního symbolu  $f$ .
- g) Ne, termy neobsahují logické spojky.
- h) Ano.

**Příklad 6.1.5.** Uvažujte jazyk  $\mathcal{L}$  predikátové logiky z předchozího příkladu. Rozhodněte, která z následujících slov jsou **formulemi** jazyka  $\mathcal{L}$ :

- a)  $y = x$ ,
- b)  $\forall x(Q(c, c) = x)$ ,
- c)  $f(x, c)$ ,
- d)  $\forall y(f(f(c, c), y))$ ,
- e)  $Q(f(x, c))$ ,
- f)  $\forall x(Q(f(f(c, c), y), f(z, c)) \vee P(f(c, z)))$ ,
- g)  $\forall x\exists c(f(x, c) = c)$ .

- a) Ano.
- b) Ne, predikátový symbol  $Q$  použit jako funkční symbol.
- c) Ne, jedná se o term.
- d) Ne, funkční symbol  $f$  použit jako predikátový symbol.
- e) Ne, nerespektuje aritu predikátového symbolu  $Q$ .
- f) Ano.
- g) Ne, funkční symbol  $c$  použit jako proměnná v kvantifikátoru.

**Příklad 6.1.6.** Uvažujte jazyk  $\mathcal{L}$  predikátové logiky obsahující jediný funkční symbol  $f$  arity 2 a žádný predikátový symbol. Uveďte příklad formule jazyka  $\mathcal{L}$

- a) s jedinou proměnnou  $x$  s právě dvěma vázanými výskyty a žádným volným výskytem,
- b) s jedinou proměnnou  $x$  s právě jedním vázaným a jedním volným výskytem,
- c) pouze s proměnnými  $x, y$  se dvěma vázanými a žádným volným výskytem, resp. s jedním volným a žádným vázaným výskytem,

d) pouze s proměnnými  $x, y$ , každou s jedním volným a jedním vázaným výskytem.

- a)  $\forall x(x = x)$
- b) Neexistuje.
- c)  $\forall x(x = f(x, y))$
- d)  $(x = y) \vee \forall x \forall y(x = y)$

**Příklad 6.1.7.** Uvažujte jazyk  $\mathcal{L}_1$  predikátové logiky, který obsahuje právě dva funkční symboly:

- nulární symbol (konstantu)  $k$ ,
- unární symbol  $f$

a žádný predikátový symbol. Dále uvažujte jazyk  $\mathcal{L}_2$ , který navíc k těmto dvěma funkčním symbolům obsahuje unární predikátový symbol  $P$ . Rozhodněte o pravdivosti následujících tvrzení a svou odpověď zdůvodněte:

1. Množina  $\mathcal{T} = \{f^i(k) \mid i \in \mathbb{N}_0\} = \{k, f(k), f(f(k)), f(f(f(k))), \dots\}$  jsou právě všechny termy jazyka  $\mathcal{L}_1$ .
2. Neexistuje formule predikátové logiky jazyka  $\mathcal{L}_1$ .
3. Uvažujte formuli  $\varphi \equiv \forall x P(x)$  jazyka  $\mathcal{L}_2$ . Existují podformule  $\psi_1, \psi_2$  formule  $\varphi$  takové, že všechny proměnné formule  $\psi_1$  mají v  $\psi_1$  volný výskyt a všechny proměnné formule  $\psi_2$  mají v  $\psi_2$  vázaný výskyt?

1. Nepravdivé tvrzení. Množina  $\mathcal{T}$  jsou právě všechny **uzavřené** termy jazyka  $\mathcal{L}_1$  (čili termy bez proměnných). Množina všech termů musí navíc nutně zahrnovat i termy s proměnnými.
2. Nepravdivé tvrzení. Formule predikátové logiky jazyka  $\mathcal{L}_1$  existuje, například  $x = x$ . (Implicitně totiž v souladu s Definicí 25 uvažujeme jazyk s rovností.)
3. Pravdivé tvrzení. Máme  $\psi_1 \equiv P(x)$  a  $\psi_2 \equiv \varphi$ .

## 6.2 Sémantika

Podobně jako u výrokové logiky je nyní namístě dodat formulím predikátové logiky konkrétní význam. Poslouží k tomu opět *interpretace*, zde ovšem mající podobu složitější struktury.

**Definice 27:** *Interpretace* (též *realizace*) jazyka  $\mathcal{L}$  je struktura  $I$  zahrnující

- neprázdné *univerzum* (či *doménu*)  $\mathcal{D}_I$ ,
- $n$ -ární relaci  $I(P) \subseteq \mathcal{D}_I^n$  pro každý  $n$ -ární predikátový symbol  $P$  jazyka  $\mathcal{L}$ ,
- zobrazení  $I(f) : \mathcal{D}_I^n \rightarrow \mathcal{D}_I$  pro každý  $n$ -ární funkční symbol  $f$  jazyka  $\mathcal{L}$ .

Univerzum specifikuje množinu objektů, nad kterými pracujeme. Termy pak reprezentují hodnoty z tohoto univerza. Místo jednoduchého přiřazení jedniček a nul výrokovým proměnným jsou predikáty chápány jako relace a jejich pravdivost závisí na hodnotách jejich argumentů.

Abychom mohli vyhodnocovat termy, je potřeba ještě definovat, jak interpretovat proměnné. K tomu slouží valuace.

**Definice 28:** *Valuace* je zobrazení  $V : \mathcal{V} \rightarrow \mathcal{D}_I$  přiřazující proměnným prvky univerza.

Hodnotu termu v interpretaci a valuaci pak upřesňuje následující indukční definice.

**Definice 29:** *Hodnota termu  $t$  v interpretaci  $I$  a valuaci  $V$*  je prvek univerza  $|t|_{I,V} \in \mathcal{D}_I$ , kde

- $|t|_{I,V} = V(x)$  pro  $t = x$ , kde  $x$  je proměnná,
- $|t|_{I,V} = I(f)(|t_1|_{I,V}, \dots, |t_n|_{I,V})$  pro  $t = f(t_1, \dots, t_n)$ , kde  $f$  je  $n$ -ární funkce a  $t_1, \dots, t_n$  jsou termy.

**Příklad.** Uvažujte jazyk  $\mathcal{L}$  s jedním nulárním funkčním symbolem  $c$  a jedním binárním funkčním symbolem  $f$  a jeho interpretaci  $I$ , kde

- univerzum tvoří celá čísla 0, 1 a 2, tj.  $\mathcal{D}_I = \{0, 1, 2\}$ ,
- symbol  $f$  se realizuje jako sčítání modulo 3, tj.  $I(f)(x, y) = (x + y) \bmod 3$ ,
- symbol  $c$  se realizuje jako číslo 1, tj.  $I(c) = 1$ .

Dále uvažujte valuaci  $V$ , která všem proměnným přiřazuje hodnotu 2, tj.  $V(x) = 2$  pro všechna  $x \in \mathcal{V}$ . Určete hodnoty následujících termů v interpretaci  $I$  a valuaci  $V$ :

- $f(c, x)$
- $f(x, f(x, y))$

- Podle druhého bodu definice hodnoty termu je  $|f(c, x)|_{I,V} = I(f)(|c|_{I,V}, |x|_{I,V})$ . Podle druhého, resp. prvního bodu definice máme  $|c|_{I,V} = I(c) = 1$  a  $|x|_{I,V} = V(x) = 2$ , a tedy celkem  $|f(c, x)|_{I,V} = I(f)(1, 2) = (1 + 2) \bmod 3 = 0$ .

- b) Máme postupně  $|f(x, f(x, y))|_{I,V} = I(f)(|x|_{I,V}, |f(x, y)|_{I,V}) = (2 + 1) \bmod 3 = 0$ , protože  $|x|_{I,V} = V(x) = 2$  a  $|f(x, y)|_{I,V} = I(f)(|x|_{I,V}, |y|_{I,V}) = (2 + 2) \bmod 3 = 1$ , neboť  $|x|_{I,V} = |y|_{I,V} = V(x) = 2$ .

Nyní můžeme přistoupit k definování pravdivosti formule v dané interpretaci a valuaci.

**Definice 30:** Zda formule  $\varphi$  je *pravdivá v interpretaci  $I$  a valuaci  $V$* , značeno  $\models_I^V \varphi$ , definujeme induktivně takto:

- $\models_I^V P(t_1, \dots, t_n)$ , právě když  $(|t_1|_{I,V}, \dots, |t_n|_{I,V}) \in I(P)$ ,
- $\models_I^V t_1 = t_2$ , právě když  $|t_1|_{I,V} = |t_2|_{I,V}$ ,
- $\models_I^V \neg\psi$ , právě když není  $\models_I^V \psi$ ,
- $\models_I^V \psi \wedge \xi$ , právě když  $\models_I^V \psi$  a  $\models_I^V \xi$  (a podobně pro zbylé logické spojky; vynecháno),
- $\models_I^V \forall x\psi$ , právě když pro všechny prvky univerza  $d \in \mathcal{D}$  platí  $\models_I^{V'} \psi$ , kde valua  $V'$  vznikla z  $V$  nahrazením obrazu proměnné  $x$  prvkem  $d$ ,
- $\models_I^V \exists x\psi$ , právě když pro nějaký prvek univerza  $d \in \mathcal{D}$  platí  $\models_I^{V'} \psi$ , kde valua  $V'$  vznikla z  $V$  nahrazením obrazu proměnné  $x$  prvkem  $d$ .

**Příklad.** Uvažujte jazyk  $\mathcal{L}$  predikátové logiky obsahující jeden unární funkční symbol  $f$  a jeden binární predikátový symbol  $P$ . Dále uvažujte interpretaci  $I$  jazyka  $\mathcal{L}$ :

- Univerzum tvoří všechna reálná čísla, tj.  $\mathcal{D}_I = \mathbb{R}$ .
- Symbol  $f$  se realizuje jako unární minus, tj.  $I(f)(x) = -x$ .
- Symbol  $P$  se realizuje jako ostré menší než, tj.  $I(P) = \{(x, y) \in \mathbb{R}^2; x < y\}$ .

Nechť  $V$  je valua přiřazující všem proměnným hodnotu 1, tj.  $V(x) = 1$  pro všechna  $x \in \mathcal{V}$ . Rozhodněte, zda následující formule jsou pravdivé v interpretaci  $I$  a valuaci  $V$ :

- a)  $P(x, f(x))$
- b)  $\exists xP(x, f(x))$
- c)  $\exists x\forall yP(x, y)$
- d)  $\exists xP(x, y)$

- a) Máme  $|x|_{I,V} = 1$ ,  $|f(x)|_{I,V} = -1$ . Dle prvního řádku definice je formule  $P(x, f(x))$  pravdivá, právě když  $(|x|_{I,V}, |f(x)|_{I,V}) \in I(P)$ , tj.  $(1, -1) \in \{(x, y) \in \mathbb{R}^2; x < y\}$ , neboli  $1 < -1$ , což zjevně neplatí.

Více neformálně: Termíny  $x$  a  $f(x)$  se vyhodnotí na 1, resp.  $-1$ . Poněvadž se symbol  $P$  realizuje jako ostrá nerovnost a neplatí  $1 < -1$ , je formule nepravdivá.

- b) Ve formuli se vyskytuje pouze proměnná  $x$ , zúžíme tedy argumentaci pouze na ni. Formule je pravdivá, jestliže existuje prvek univerza, jehož dosazením za  $x$  docílíme toho, že je formule  $P(x, f(x))$  pravdivá; takovým prvkem je ovšem např. číslo  $-1$ .
- c) Formule je pravdivá, jestliže existuje prvek, jehož dosazením za  $x$  docílíme toho, že je formule  $\forall y P(x, y)$  pravdivá. Takový prvek nicméně neexistuje, neboť při dosazení libovolného reálného čísla  $d \in \mathcal{D}_I$  za  $x$  budou existovat hodnoty  $y$ , pro které není  $P(x, y)$  pravdivá – např. číslo  $d - 1 \in \mathcal{D}_I$  (a tudíž není  $P(x, y)$  pravdivá pro libovolnou hodnotu  $y$ ).

Pozor na pořadí kvantifikátorů! Aby byla uvažovaná formule pravdivá, musí existovat hodnota  $x$  taková, že již formule  $P(x, y)$  platí pro libovolné  $y$ .

- d) Proměnná  $y$  má v uvažované valuaci hodnotu 1, zjevně existuje hodnota  $x$ , aby formule  $P(x, y)$  byla pravdivá (kupř.  $V(x) = 0$ ), a tedy uvažovaná formule je pravdivá.

Pravdivost formule (pouze) v interpretaci je definována tak, aby byla ekvivalentní doplnění univerzálních kvantifikátorů pro volné proměnné, tedy

$$\models_I P(x, y), \quad \text{právě když} \quad \models_I \forall x \forall y P(x, y) :$$

### Definice 31:

- Formule predikátové logiky je *pravdivá v interpretaci*  $I$ , ozn.  $\models_I \varphi$ , jestliže je pravdivá v interpretaci  $I$  a valuaci  $V$  pro libovolnou valuaci  $V$ .
- Formule predikátové logiky je *logicky pravdivá* či *tautologie*, ozn.  $\models \varphi$ , jestliže je pravdivá v každé interpretaci.

**Příklad 6.2.1.** ★ Ukažte, že existenční kvantifikátor  $\exists$  lze v predikátové logice zavést jako syntaktickou zkratku, tj. ekvivalentními úpravami vyjádřete  $\exists x \varphi$  bez použití symbolu  $\exists$ .

$$\exists x \varphi \approx \neg \forall x \neg \varphi$$

Často se vyskytuje špatné řešení ve stylu „Mějme doménu  $\mathcal{D} = \{a_1, a_2, \dots, a_n\}$ , pak  $\exists x\varphi$  je zkratkou pro  $\varphi(x/a_1) \vee \varphi(x/a_2) \vee \dots \vee \varphi(x/a_n)$ , kde  $\varphi(x/a)$  značí náhradu proměnné  $x$  prvkem  $a$  ve formuli  $\varphi$ .“ Taková odpověď je špatně z několika důvodů:

- Doména není v zadání specifikovaná, takové řešení nefunguje obecně.
- Řešení nepostihuje nekonečné domény jako  $\mathbb{Z}$  či  $\mathbb{R}$  (neumíme vytvořit nekonečně dlouhou formuli).
- $a_1, \dots, a_n$  nejsou termy, ale prvky domény. Neexistuje garance, že v daném jazyce budeme moct vyjádřit každý prvek domény pomocí nějakého termu.

**Příklad 6.2.2.** Nalezněte negace následujících formulí tak, aby se symboly  $\neg$  nacházely výhradně bezprostředně před predikátovými symboly.

- ★  $\forall x(P(x) \vee \exists yQ(y))$ ,
- ★  $\exists x((P(x) \wedge Q(x)) \vee R(x))$ ,
- $\forall x(P(x) \Rightarrow \forall yQ(y))$ ,
- $(\exists xP(x) \vee \forall yQ(y)) \Leftrightarrow \forall x(\exists yR(x, y) \Rightarrow Q(x))$ ,
- $\forall x(P(x) \Rightarrow Q(x)) \wedge \exists x(R(x) \wedge S(x))$ .

- $\neg\forall x(P(x) \vee \exists yQ(y))$   
 $\approx \exists x\neg(P(x) \vee \exists yQ(y))$   
 $\approx \exists x(\neg P(x) \wedge \forall y\neg Q(y))$
- $\neg(\exists x((P(x) \wedge Q(x)) \vee R(x)))$   
 $\approx \forall x\neg((P(x) \wedge Q(x)) \vee R(x))$   
 $\approx \forall x((\neg P(x) \vee \neg Q(x)) \wedge \neg R(x))$
- $\neg(\forall x(P(x) \Rightarrow \forall yQ(y)))$   
 $\approx \exists x\neg(P(x) \Rightarrow \forall yQ(y))$   
 $\approx \exists x(P(x) \wedge \neg\forall yQ(y))$   
 $\approx \exists x(P(x) \wedge \exists y\neg Q(y))$
- $\neg((\exists xP(x) \vee \forall yQ(y)) \Leftrightarrow \forall x(\exists yR(x, y) \Rightarrow Q(x)))$   
 $\approx \neg(\exists xP(x) \vee \forall yQ(y)) \Leftrightarrow \forall x(\exists yR(x, y) \Rightarrow Q(x))$   
 $\approx (\forall x\neg P(x) \wedge \exists y\neg Q(y)) \Leftrightarrow \forall x(\exists yR(x, y) \Rightarrow Q(x))$
- $\neg(\forall x(P(x) \Rightarrow Q(x)) \wedge \exists x(R(x) \wedge S(x)))$   
 $\approx \exists x\neg(P(x) \Rightarrow Q(x)) \vee \forall x\neg(R(x) \wedge S(x))$   
 $\approx \exists x(P(x) \wedge \neg Q(x)) \vee \forall x(\neg R(x) \vee \neg S(x))$



**Příklad 6.2.3.** ★ Uvažte formuli  $\varphi \equiv \forall x(P(x) \Rightarrow \exists y\neg Z(x, y))$ . Pro každou z následujících interpretací rozhodněte, zda je modelem formule  $\varphi$ .

- a)  $\mathcal{D}_I = \{0\}$ ,  $I(P) = \emptyset$ ,  $I(Z) = \emptyset$
- b)  $\mathcal{D}_I = \mathbb{Z}$ ,  $I(P) = \{1, 2, 3, \dots\}$ ,  $I(Z) = \mathbb{Z}^2$
- c)  $\mathcal{D}_I = \mathbb{Z}$ ,  $I(P) = \{1, 2, 3, \dots\}$ ,  $I(Z) = \leq$
- d)  $\mathcal{D}_I = \{0\}$ ,  $I(P) = \{0\}$ ,  $I(Z) = \text{id}^3$

- a) Ano. Jelikož  $I(P) = \emptyset$ , bude formule  $P(x)$  nepravdivá v každé valuaci, a tudíž ze sémantiky implikace bude formule  $P(x) \Rightarrow \psi$  (pro lib.  $\psi$ ) pravdivá v každé valuaci. Celkem tedy  $\varphi$  je pravdivá v  $I$ .
- b) Ne. Jelikož  $I(Z) = \mathbb{Z}^2$ , není  $\neg Z(x, y)$  pravdivá v žádné valuaci. Formule  $P(x) \Rightarrow \exists y\neg Z(x, y)$  je tedy nepravdivá např. ve valuaci přiřazující  $x$  hodnotu 1. (A tedy je nepravdivá celá formule  $\varphi$  univerzálně kvantifikovaná přes proměnnou  $x$ .)
- c) Ano. Pro každou hodnotu proměnné  $x$  splňující  $P(x)$  totiž existuje hodnota proměnné  $y$  taková, že  $Z(x, y)$  je nepravdivá (lze volit  $V(y) = V(x) - 1$ ), a tedy pravá strana implikace  $\exists y\neg Z(x, y)$  je pravdivá.
- d) Ne. V uvažované interpretaci je formule  $\neg Z(x, y)$  pravdivá v těch valuacích  $V$ , kde  $V(x) \neq V(y)$ . V uvažovaném jednoprvkovém univerzu však pro dané přiřazení hodnoty proměnné  $x$  nenalezneme přiřazení proměnné  $y$ , aby  $V(x) \neq V(y)$ . První část implikace je však splňována valuací  $V(x) = 0$ , takže univerzální kvantifikace není pravdivá (a tudíž ani celá formule  $\varphi$ ).

**Příklad 6.2.4.** Uvažte formuli  $\varphi \equiv \exists x(\forall yR(x, y) \Rightarrow Q(x))$ . Pro každou z následujících interpretací rozhodněte, zda je modelem formule  $\varphi$ .

- a)  $\mathcal{D}_I = \{0\}$ ,  $I(R) = \emptyset$ ,  $I(Q) = \emptyset$
- b)  $\mathcal{D}_I = \{0\}$ ,  $I(R) = \{(0, 0)\}$ ,  $I(Q) = \emptyset$
- c)  $\mathcal{D}_I = \mathbb{Z}$ ,  $I(R) = \mathbb{Z} \times \mathbb{Z}$ ,  $I(Q) = \mathbb{N}_0$

- a) Ano. Pro  $V(x) = 0$  je podformule  $\forall yR(x, y) \Rightarrow Q(x)$  pravdivá, jelikož  $I(R) = \emptyset$  (čili nedojde nikdy ke splnění předpokladu implikace).
- b) Ne. Pro  $V(x) = 0$  je totiž formule  $\forall yR(x, y)$  pravdivá (jediná možná hodnota  $V(y) = 0$  a  $(0, 0) \in I(R)$ ), ale  $Q(x)$  je nepravdivá.  $V(x)$  nemůže nebývat jiné hodnoty než 0, a tudíž nelze nalézt hodnotu  $V(x)$ , aby formule  $\forall yR(x, y) \Rightarrow Q(x)$  byla pravdivá.
- c) Ano. Pro  $V(x) = 0$  je podformule  $\forall yR(x, y) \Rightarrow Q(x)$  pravdivá, protože je pravdivá podformule  $Q(x)$  (neboť  $0 \in I(Q) = \mathbb{N}_0$ ).

<sup>3</sup>Relace identity obsahuje dvojice identických prvků,  $\text{id} := \{(x, y) \in \mathcal{D}_I^2; x = y\}$ .

**Příklad 6.2.5.** Uvažte formuli  $\varphi \equiv \forall y(\exists x R(x, y) \Rightarrow Q(y))$ . Pro každou z následujících odrážek popište právě tu  $I(Q)$ , že interpretace  $I$  je modelem formule  $\varphi$ .

- a)  $\mathcal{D}_I = \{0\}$ ,  $I(R) = \{(0, 0)\}$
- b)  $\mathcal{D}_I = \{0\}$ ,  $I(R) = \emptyset$
- c)  $\mathcal{D}_I = \mathbb{Z}$ ,  $I(R) = <$
- d)  $\mathcal{D}_I = \mathbb{Z}$ ,  $I(R) = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid x \text{ je sudé číslo}\}$
- e)  $\mathcal{D}_I = \mathbb{Z}$ ,  $I(R) = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid y \text{ je sudé číslo}\}$

- a)  $I(Q) = \{0\}$
- b)  $I(Q) = \{0\}$  nebo  $I(Q) = \emptyset$
- c)  $I(Q) = \mathbb{Z}$
- d)  $I(Q) = \mathbb{Z}$
- e)  $\{x \in \mathbb{Z} \mid x \text{ je sudé číslo}\} \subseteq I(Q) \subseteq \mathbb{Z}$

**Příklad 6.2.6. ★** Rozhodněte, v jakých interpretacích jsou pravdivé následující formule:

- a)  $\exists x \forall y (P(y) \Rightarrow (x = y))$
- b)  $\exists x (P(x) \wedge \forall y (P(y) \Rightarrow (x = y)))$
- c)  $\forall x \exists y (P(x, x) \Rightarrow (x \neq y \wedge P(x, y)))$

- a) Formule je pravdivá právě v interpretacích  $I$  s libovolným neprázdným univerzem a s  $|I(P)| \leq 1$ .
- b) Formule je pravdivá právě v interpretacích  $I$  s  $|I(P)| = 1$ .
- c) Formule je pravdivá právě v interpretacích  $I$ , kde každý reflexivní prvek  $x$  relace  $I(P)$  je v relaci s nějakým prvkem  $y \neq x$ .

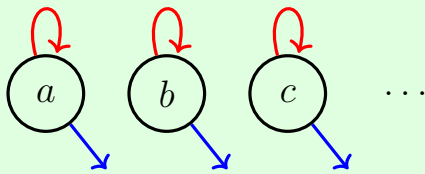
**Příklad 6.2.7.** Rozhodněte, v jakých interpretacích jsou pravdivé následující formule:

- a)  $\forall x \exists y \exists z ((x = y) \vee (x = z)) \wedge (y \neq z)$
- b)  $\forall x \forall y \forall z ((Q(x, y) \wedge Q(y, z)) \Rightarrow Q(x, z))$
- c)  $\forall x (\neg Q(x, x) \wedge \exists y Q(x, y))$
- d)  $\forall x (x \neq f(x) \wedge x = f(f(x)))$

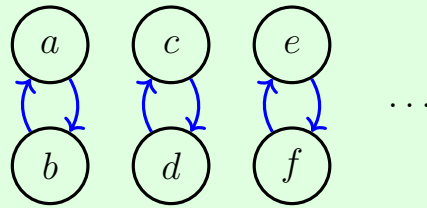
- a) Formule je pravdivá právě v interpretacích alespoň dvouprvkovým univerzem.
- b) Formule je pravdivá právě v interpretacích  $I$ , kde  $I(Q)$  je tranzitivní relace.

- c) Formule je pravdivá právě v interpretacích  $I$ , kde  $I(Q)$  je ireflexivní relace, kde každý prvek je v relaci alespoň s 1 dalším (vychází z něj alespoň 1 šipka v grafu relace).
- d) Formule je pravdivá právě v interpretacích  $I$ , kde každý prvek univerza je v grafu funkce  $I(f)$  zahrnut v nějakém cyklu délky dva. Je-li univerzum  $I$  konečné, má tedy navíc sudou kardinalitu.

Následující obrázky vyobrazují požadované vlastnosti modelů vybraných formulí. Červená barva naznačuje hrany, které v modelu nesmí být, modrá barva naznačuje hrany, které naopak v modelu být musí. V modelu obecně mohou být navíc i jiné než modré hrany. (Jen **pozor**, u funkcí je třeba respektovat, že z každého vrcholu vychází **právě jedna** hrana.)



c) Graf relace  $I(Q)$



d) Graf funkce  $I(f)$

**Příklad 6.2.8.** ★ Uvažujte jazyk  $\mathcal{L}$  predikátové logiky obsahující binární funkční symbol  $+$  a unární predikátový symbol  $K$ . Uvažujte interpretaci (realizaci)  $I$  jazyka  $\mathcal{L}$ , kde univerzum tvoří všechna celá čísla  $\mathbb{Z}$ ,  $+$  se realizuje jako sčítání a  $K(x)$  jako „ $x$  je kladné“. Nalezněte

- fomuli  $\alpha(x)$ , kt. je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x) = 0$ ,
- fomuli  $\beta(x, y)$ , která je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x) < V(y)$ ,
- fomuli  $\gamma(x, y)$ , která je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x) = -V(y)$ ,
- fomuli  $\delta(x)$ , která je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x) = 1$ .

Při definování formulí se můžete odvolat na formule, které jste již zavedli v předchozích bodech.

- Např.  $\alpha(x) \equiv x + x = x$ .
- Např.  $\beta(x, y) \equiv \exists z(K(z) \wedge x + z = y)$ .
- Např.  $\gamma(x, y) \equiv \forall z(\alpha(z) \Rightarrow x + y = z)$ .
- Např.  $\delta(x) \equiv K(x) \wedge \forall y(K(y) \Rightarrow (x = y \vee \beta(x, y)))$ .

**Příklad 6.2.9.** Uvažujte jazyk  $\mathcal{L}$  predikátové logiky obsahující právě dva binární funkční symboly  $*$  a  $+$ . Uvažujte interpretaci (realizaci)  $I$  jazyka  $\mathcal{L}$ , kde univerzum tvoří všechna reálná čísla  $\mathbb{R}$  a funkční symboly se realizují běžným (matematickým) způsobem, tj.  $+$  jako sčítání a  $*$  jako násobení. Nalezněte

- fomuli  $\alpha(x, y, z)$ , kt. je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x) - V(y) = V(z)$ ,

- b) fomuli  $\beta(x)$ , která je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x)$  je 0,
- c) fomuli  $\gamma(x)$ , která je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x)$  je 1,
- d) fomuli  $\delta(x)$ , která je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x)$  je kladné číslo nebo 0,
- e) fomuli  $\varepsilon(x, y)$ , která je pravdivá v  $I$  právě v těch valuacích  $V$ , že  $V(x) \leq V(y)$ .

Při definování formulí se můžete odvolat na formule, které jste již zavedli v předchozích bodech.

- a) Např.  $\alpha(x, y, z) \equiv x = y + z$ .
- b) Např.  $\beta_1(x) \equiv x + x = x$  nebo  $\beta_2(x) \equiv \forall y(x * y = x)$ .
- c) Např.  $\gamma_1(x) \equiv \forall y(y * x = y)$  nebo  $\gamma_2(x) \equiv (x * x = x) \wedge \neg\beta(x)$ .

**Pozor**, v případě  $\gamma_2$  lze jednoduše zapomenout na podmínku, aby  $x$  nebyla nula, ta totiž rovnici  $x^2 = x$  vyhovuje taky.

- d) Např.  $\delta(x) \equiv \exists y(y * y = x)$ .
- e) Např.  $\varepsilon_1(x, y) \equiv \exists z(x + z * z = y)$  nebo  $\varepsilon_2(x, y) \equiv \exists z(\delta(z) \wedge x + z = y)$ .

**Příklad 6.2.10.** ★ Rozhodněte, které z následujících formulí jsou tautologie:

- a)  $\forall x(P(x) \wedge Q(x)) \Leftrightarrow (\forall xP(x) \wedge \forall xQ(x))$
- b)  $\exists x(P(x) \vee Q(x)) \Leftrightarrow (\exists xP(x) \vee \exists xQ(x))$
- c)  $\forall x(A(x) \vee B(x)) \Rightarrow (\forall xA(x) \vee \forall xB(x))$
- d)  $\forall x\forall yR(x, y) \Leftrightarrow \forall y\forall xR(x, y)$
- e)  $(\exists xA(x) \wedge \exists xB(x)) \Rightarrow \exists x(A(x) \wedge B(x))$
- f)  $(\forall xP(x) \Rightarrow \forall xQ(x)) \Rightarrow \forall x(P(x) \Rightarrow Q(x))$

- a) Ano.
- b) Ano.
- c) Ne.
- d) Ano.
- e) Ne.
- f) Ne.

### 6.3 Normální formy

Jelikož rezoluce v predikátové logice vyžaduje práci s formulemi určitých tvarů, je nutné se s těmito tvary seznámit, abychom mohli rezoluci provádět. Podobně jako ve výrokové logice

jim říkáme normální formy. První normální forma, prenexová, má všechny kvantifikátory umístěny před samotným jádrem formule, které je v KNF. Literálem v kontextu predikátové logiky rozumíme libovolnou atomickou formuli nebo její negaci.

**Definice 32:** Uzavřená formule  $\varphi$  se nachází v *prenexové normální formě (PNF)*, je-li tvaru

$$Q_1x_1 \dots Q_nx_n\psi,$$

kde  $Q_i \in \{\forall, \exists\}$ ,  $x_1, \dots, x_n$  jsou proměnné a formule  $\psi$  je v konjunktivní normální formě (zejména tedy neobsahuje žádný kvantifikátor).

Například formule  $\forall x\exists y((P(f(y)) \vee \neg Q(x)) \wedge (\neg P(y) \vee P(x)))$  se nachází v PNF.

Skolemova normální forma oproti PNF ještě navíc vyžaduje, aby veškeré proměnné ve formuli byly všeobecně kvantifikované.

**Definice 33:** Uzavřená formule  $\varphi$  se nachází ve *Skolemově normální formě*, nachází-li se v prenexové normální formě a obsahuje pouze obecné kvantifikátory  $\forall$ .

Například formule  $\forall x((P(f(g(x))) \vee \neg Q(x)) \wedge (\neg P(g(x)) \vee P(x)))$  se nachází ve Skolemově NF, jelikož vznikla skolemizací předchozí formule v PNF. Procesy převodu do PNF a do Skolemovy NF si ukážeme na příkladech.

**Příklad.** Převedte formuli  $\exists x\forall y(P(x, y) \Rightarrow \neg\forall xQ(y, x))$  do PNF.

Algoritmus převodu sestává z několika kroků.

- I. Eliminujeme zbytečné kvantifikátory. (Takové v zadané formuli nejsou.)
- II. Přejmenujeme proměnné tak, aby u každého kvantifikátoru byla jiná proměnná:  
 $\exists x_1\forall y(P(x_1, y) \Rightarrow \neg\forall x_2Q(y, x_2))$
- III. Eliminujeme jiné spojky než  $\vee$ ,  $\wedge$ ,  $\neg$ :  
 $\exists x_1\forall y(\neg P(x_1, y) \vee \neg\forall x_2Q(y, x_2))$
- IV. Přesuneme negaci až před samotné predikáty (s využitím de Morganových zákonů a pravidel pro negaci kvantifikátoru):  
 $\exists x_1\forall y(\neg P(x_1, y) \vee \exists x_2\neg Q(y, x_2))$
- V. Kvantifikátory přesuneme ven z jádra formule:  $\exists x_1\forall y\exists x_2(\neg P(x_1, y) \vee \neg Q(y, x_2))$
- VI. Jádro formule upravíme do KNF pomocí distributivních zákonů. (Již je v KNF.)

Nyní si demonstrujeme převod výsledné formule do Skolemovy NF.

**Příklad.** Převedte formuli  $\exists x_1\forall y\exists x_2(\neg P(x_1, y) \vee \neg Q(y, x_2))$  do Skolemovy NF.

Převod z PNF do Skolemovy NF spočívá v citlivém odstranění existenčních kvantifikátorů. Jejich roli v nově vytvořené formuli zastoupí Skolemovy funkce, které se vyhodnotí na hodnotu, kterou by původně „vybral“ existenční kvantifikátor.

V zadané formuli je volba hodnoty proměnné  $x_1$  pevná (nezávisí na dalších proměnných, protože existenční kvantifikátor je úplně první), nahradíme tedy proměnnou  $x_1$  konstantou  $c$  a kvantifikátor odstraníme ( $x_1 \rightarrow c$ ):

$$\forall y \exists x_2 (\neg P(c, y) \vee \neg Q(y, x_2))$$

Volba hodnoty proměnné  $x_2$  již pevná není, ale může záviset na konkrétní hodnotě  $y$ , protože se obecný kvantifikátor s  $y$  nachází před kvantifikátorem s  $x_2$ . Nahradíme tedy proměnnou  $x_2$  funkcí  $f(y)$  a kvantifikátor odstraníme ( $x_2 \rightarrow f(y)$ ):

$$\forall y (\neg P(c, y) \vee \neg Q(y, f(y)))$$

Nalezená formule je již ve Skolemově NF.

**Příklad 6.3.1.** Převeďte následující formule do prenexové normální formy a proveďte skolemizaci.

- a)  $(\forall x \exists y Q(x, y) \vee \exists x \forall y P(x, y)) \wedge \neg \exists x \exists y \forall z R(x, y, z)$   
b)  $\neg(\forall x \exists y P(x, y) \Rightarrow \exists x \exists y R(x, y)) \wedge \forall x (\neg \exists y Q(x, y))$

a) Při převodu postupujeme dle algoritmu popsaného výše.

- I. Eliminace zbytečných kvantifikátorů je triviální, takové ve formuli totiž nejsou.
- II. Přistoupíme k přejmenování proměnných, aby měl každý kvantifikátor proměnnou s jiným názvem:  $(\forall x \exists y Q(x, y) \vee \exists x_1 \forall y_1 P(x_1, y_1)) \wedge \neg \exists x_2 \exists y_2 \forall z R(x_2, y_2, z)$ .
- III. Formule již obsahuje pouze spojky  $\vee$ ,  $\wedge$ ,  $\neg$ .
- IV. Přesuneme negaci před samotné predikáty:  
 $(\forall x \exists y Q(x, y) \vee \exists x_1 \forall y_1 P(x_1, y_1)) \wedge \forall x_2 \forall y_2 \exists z \neg R(x_2, y_2, z)$ .
- V. Kvantifikátory přesuneme na začátek formule:  
 $\forall x \exists y \exists x_1 \forall y_1 \forall x_2 \forall y_2 \exists z ((Q(x, y) \vee P(x_1, y_1)) \wedge \neg R(x_2, y_2, z))$ .
- VI. Jádro je již v KNF, takže není nutná další úprava.

Výsledná formule v PNF:

$$- \forall x \exists y \exists x_1 \forall y_1 \forall x_2 \forall y_2 \exists z ((Q(x, y) \vee P(x_1, y_1)) \wedge \neg R(x_2, y_2, z))$$

Skolemova NF (nahrazujeme  $y \rightarrow f(x)$ ,  $x_1 \rightarrow g(x)$  a  $z \rightarrow h(x, y_1, x_2, y_2)$ ):

$$- \forall x \forall y_1 \forall x_2 \forall y_2 ((Q(x, f(x)) \vee P(g(x), y_1)) \wedge \neg R(x_2, y_2, h(x, y_1, x_2, y_2)))$$

b) Výsledná formule v PNF:

$$- \forall x \exists y \forall x_1 \forall y_1 \forall x_2 \forall y_2 (P(x, y) \wedge \neg R(x_1, y_1) \wedge \neg Q(x_2, y_2))$$

Skolemova NF (nahrazujeme  $y \rightarrow f(x)$ ):

$$- \forall x \forall x_1 \forall y_1 \forall x_2 \forall y_2 (P(x, f(x)) \wedge \neg R(x_1, y_1) \wedge \neg Q(x_2, y_2))$$

**Příklad 6.3.2.** Uvažujte následující formule:

a)  $\alpha \equiv \forall x \forall y \exists z (x + z = y)$ ,

b)  $\beta \equiv \exists x \forall y (x * y = x)$ ,

c)  $\gamma \equiv \forall x \forall y \exists r (x \leq y \Rightarrow x \leq r \leq y)$

d)  $\delta \equiv \forall x_1 \forall y_1 \forall x_2 \forall y_2 \exists x_3 \exists y_3 (\sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2} = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2})$

e)  $\varepsilon \equiv \forall x \exists y (x > 0 \Rightarrow x = 2^y)$

nad doménou reálných čísel, kde matematické symboly  $+, *, \cdot, \sqrt{\cdot}, \dots$  se interpretují běžným způsobem (tj.  $+$  jako sčítání,  $*$  jako násobení, ...). Proveďte skolemizaci těchto formulí a určete nějakou možnou interpretaci pro každou z nově vzniklých Skolemových funkcí (vč. konstant) tak, aby byla zachována pravdivost formulí.

a)  $\alpha_S \equiv \forall x \forall y (x + (y - x) = y)$ , kde  $-$  se interpretuje běžným způsobem.

b)  $\beta_S \equiv \forall y (0 * y = 0)$ , kde  $0$  se interpretuje běžným způsobem.

c) Např.  $\gamma_S \equiv \forall x \forall y (\neg(x \leq y) \vee x \leq \text{AVG}(x, y) \leq y)$ , kde  $\text{AVG}(x, y)$  se interpretuje jako aritmetický průměr čísel  $x, y$ .

Nezapomeňte se zbavit implikace!

d) Pokud se na problém podíváme z geometrického hlediska, čemuž napovídá i značení proměnných v zadání, formule říká, že pro každou dvojici bodů  $[x_1, y_1] =: A$ ,  $[x_2, y_2] =: B$  existuje bod  $[x_3, y_3] =: C$ , který splňuje  $\|AC\| = \|BC\|$ . Takový bod je například střed úsečky  $AB$ , jehož souřadnice nelezeme právě zprůměrováním jednotlivých složek bodů  $A, B$ .

$$\delta_S \equiv \forall x_1 \forall y_1 \forall x_2 \forall y_2 (\sqrt{(x_1 - f(\vec{x}))^2 + (y_1 - g(\vec{x}))^2} = \sqrt{(x_2 - f(\vec{x}))^2 + (y_2 - g(\vec{x}))^2}),$$

kde  $\vec{x}$  je zkrácený zápis pro  $x_1, y_1, x_2, y_2$  a  $f$  se interpretuje jako průměr  $x$ ových souřadnic, tj.  $f(x_1, y_1, x_2, y_2) = \text{AVG}(x_1, x_2)$  a  $g$  se analogicky interpretuje jako průměr  $y$ ových souřadnic.

e)  $\varepsilon_S \equiv \forall x (\neg(x > 0) \vee x = 2^{f(x)})$ , kde  $f(x)$  se interpretuje např. jako

$$f(x) = \begin{cases} \log_2(x) & x > 0, \\ 0 & \text{jinak.} \end{cases}$$

Zde je velmi důležité nepřehlédnout fakt, že logaritmus je definován pouze na  $\mathbb{R}^+$ , ale naše funkce musí být totální na celé doméně  $\mathbb{R}$ . (Něméně mimo  $\mathbb{R}^+$  na hodnotě nezáleží.)

## 7 Důkazové systémy a rezoluce

Předmětem této kapitoly budou formální systémy, které umožňují vyvozovat závěry na strojové úrovni. Nejprve se seznámíme s rezolucí ve výrokové a predikátové logice a následně s programovacím jazykem Prolog, jehož výpočty jsou na principu rezoluce založeny.

### 7.1 Obecná rezoluce ve výrokové logice

V rámci pojednávání o rezoluční metodě budeme podobně jako dříve pracovat s *klauzulemi*, tedy disjunkcemi literálů. Budeme ovšem používat množinový zápis, kdy klauzuli  $\ell_1 \vee \dots \vee \ell_n$  zapíšeme jako množinu literálů  $\{\ell_1, \dots, \ell_n\}$ . Speciálně zavádíme symbol  $\square$  pro prázdnou množinu literálů, tedy kontradikci.

**Definice 34:** *Rezoluční pravidlo:* Uvažujme dvě klauzule  $\{p, \ell_1, \dots, \ell_n\}$ ,  $\{\neg p, \ell'_1, \dots, \ell'_m\}$ , kde  $\ell_i$ , resp.  $\ell'_i$  jsou literály. Jejich *rezolventou* je klauzule  $\{\ell_1, \dots, \ell_n, \ell'_1, \dots, \ell'_m\}$ .

Rezolventou klauzulí  $\{\neg p, r, \neg t, u\}$  a  $\{s, \neg q, \neg r\}$  je klauzule  $\{\neg p, \neg t, u, s, \neg q\}$ . Klauzule  $\{\neg p, r, \neg t, u\}$  a  $\{\neg r\}$  mají rezolventu  $\{\neg p, \neg t, u\}$ .

Důležitá vlastnost rezolučního pravidla je, že zachovává pravdivost v interpretaci (a tudíž i splnitelnost). Je-li tedy klauzule  $C$  rezolventou nějakých klauzulí  $C_1, C_2$ , vyplývá klauzule  $C$  z množiny  $\{C_1, C_2\}$ . Dokážeme-li z klauzulí  $C_1, C_2$  prázdnou klauzuli  $\square$ , je množina  $\{C_1, C_2\}$  nespílitelná.

Opakovanou aplikací rezolučního pravidla vzniká rezoluční důkaz:

**Definice 35:** *Rezoluční důkaz* klauzule  $C$  z množiny klauzulí  $S$  je konečná posloupnost klauzulí  $C_1, \dots, C_n$ , kde  $C_n = C$  a pro každé  $1 \leq i \leq n$  je  $C_i \in S$  nebo  $C_i$  vzniklo aplikací rezolučního pravidla na klauzule, které mu v posloupnosti předcházely.

Klauzule  $C$  je *rezolučně dokazatelná* z  $S$ , zapisováno  $S \vdash_R C$ , jestliže existuje její rezoluční důkaz z  $S$ .

Rezoluční důkaz prázdné klauzule  $\square$  z množiny klauzulí  $S$  nazýváme *vyvrácením*  $S$ .

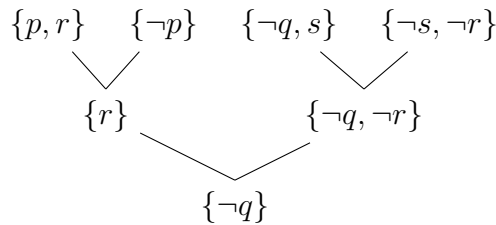
**Příklad.** Uveďte rezoluční důkaz klauzule  $C = \{\neg q\}$  z množiny  $S = \{\{p, r\}, \{\neg s, \neg r\}, \{\neg q, s\}, \{\neg p\}\}$ .

- $C_1 = \{p, r\} \in S$
- $C_2 = \{\neg p\} \in S$
- $C_3 = \{r\}$  rezolventa  $C_1$  a  $C_2$
- $C_4 = \{\neg s, \neg r\} \in S$



- $C_5 = \{\neg q, s\} \in S$
- $C_6 = \{\neg q, \neg r\}$  rezolventa  $C_4$  a  $C_5$
- $C_7 = \{\neg q\} = C$  rezolventa  $C_3$  a  $C_6$

Pro názornější představu je výhodné zapisovat rezoluční důkazy pomocí stromu. Kořenem stromu je vždy výsledná klauzule a potomky každého vnitřního uzlu jsou klauzule, jejichž je daný uzel rezolventou. Listy pak tvoří prvky množiny předpokladů  $S$ . Rezoluční strom klauzule  $\{\neg q\}$  z množiny předpokladů  $\{\{p, r\}, \{\neg s, \neg r\}, \{\neg q, s\}, \{\neg p\}\}$  by mohl vypadat následovně:



Mnohdy (například při strojovém dokazování u dalších typů rezoluce) je výhodnější místo odvození konkrétní klauzule ukázat nespůlnitelnost množiny klauzulí. Následující věta shrnuje, jak lze vyplývání dokázat vyvrácením množiny formulí.

**Věta 36:** Pro množinu formulí  $\mathcal{T}$  a formuli  $\varphi$  platí  $\mathcal{T} \models \varphi$ , právě když  $\mathcal{T} \cup \{\neg\varphi\}$  je nespůlnitelná.

Proto budeme často vyplývání  $\mathcal{T} \models \varphi$  ukazovat tak, že namísto rezolučního důkazu  $\mathcal{T} \vdash_R \varphi$  nalezneme rezoluční vyvrácení  $\mathcal{T} \cup \{\neg\varphi\} \vdash_R \square$ .

Důkaz Věty 36 není zvláště náročný a můžete si ho rozmyslet v rámci Příkladu 5.4.8.

**Příklad 7.1.1.** ★ Obecnou rezolucí ukažte, že

- formule  $(p \vee q) \wedge (\neg q \vee \neg r) \wedge (\neg p \vee s) \wedge (\neg s) \wedge (r \vee s)$  je nespůlnitelná,
- platí logické vyplývání  $\{(\neg s \wedge t) \Rightarrow r, \neg s, \neg s \Rightarrow t\} \models r$ .

- Formuli odpovídá množina klauzulí  $\{\{p, q\}, \{\neg q, \neg r\}, \{\neg p, s\}, \{\neg s\}, \{r, s\}\}$ , jejíž nespůlnitelnost ukážeme obecnou rezolucí – tj. aplikujeme na jednotlivé klauzule rezoluční pravidlo až do odvození prázdné klauzule  $\square$ .

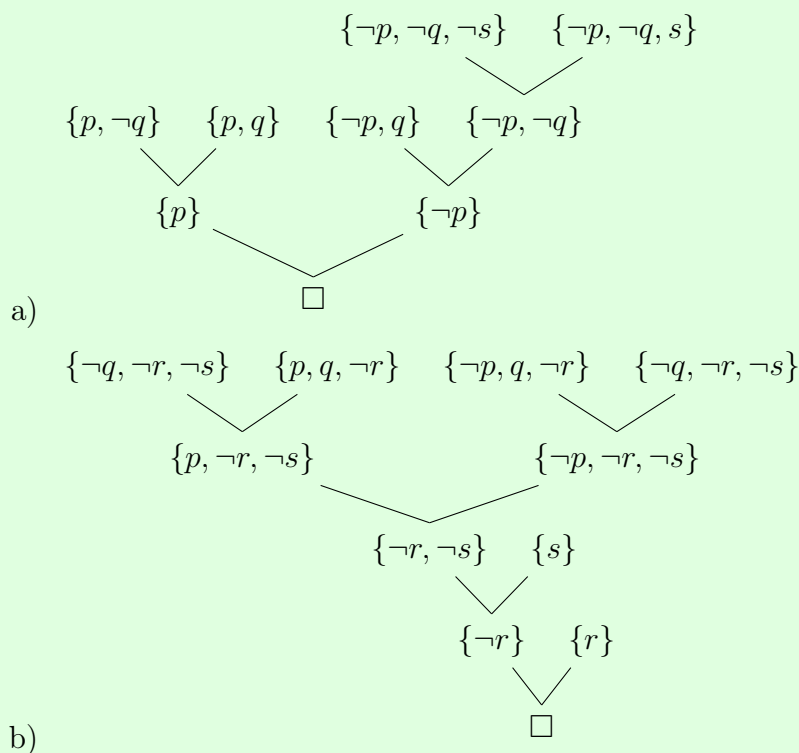
b) I. Převědeme množinu předpokladů na ekvivalentní množinu klauzulí a závěr do tvaru klauzule:  $\{\{s, \neg t, r\}, \{\neg s\}, \{s, t\}\} \models \{r\}$ . Následně odvodíme závěr přímo aplikací rezolučního pravidla.

II. Alternativně lze znegovat závěr a rezolucí ukázat nespelnitelnost množiny klauzulí s přidaným znegovaným závěrem:  $\{\{s, \neg t, r\}, \{\neg s\}, \{s, t\}, \{\neg r\}\}$ .

**Příklad 7.1.2.** Obecnou rezolucí ukažte nespelnitelnost následujících množin klauzulí.

a)  $\{\{\neg p, \neg q, \neg s\}, \{\neg p, \neg q, s\}, \{\neg p, q\}, \{p, \neg q\}, \{p, q\}\}$

b)  $\{\{p, q, \neg r\}, \{r\}, \{\neg p, q, \neg r\}, \{\neg q, \neg r, \neg s\}, \{s\}\}$



**Příklad 7.1.3.** Obecnou rezolucí dokažte následující logická vyplývání. Vyplývání dokažte jak přímo, tak pomocí rezolučního vyvrácení.

a)  $\{\neg p \vee q, \neg r \Rightarrow \neg q\} \models p \Rightarrow r$

b)  $\{(\neg p \wedge r) \Rightarrow q, (\neg p \wedge q) \Rightarrow s, r \Rightarrow \neg p\} \models r \Rightarrow s$

a) I. Převědeme množinu předpokladů na ekvivalentní množinu klauzulí a závěr do tvaru

klauzule:  $\{\{\neg p, q\}, \{r, \neg q\}\} \models \{\neg p, r\}$ . Následně odvodíme závěr přímo aplikací rezolučního pravidla.

$$\begin{array}{c} \{\neg p, q\} \quad \{r, \neg q\} \\ \swarrow \quad \searrow \\ \{\neg p, r\} \end{array}$$

II. Alternativně lze znegovat závěr a rezolucí ukázat nespelnitelnost množiny klauzulí s přidaným znegovaným závěrem. Závěr tvaru  $p \Rightarrow r$  má negaci  $p \wedge \neg r$ , kterou lze vyjádřit dvěma klauzulemi  $\{p\}, \{\neg r\}$ . Vyvrácíme množinu  $\{\{\neg p, q\}, \{r, \neg q\}, \{p\}, \{\neg r\}\}$ .

$$\begin{array}{c} \{\neg p, q\} \quad \{p\} \quad \{r, \neg q\} \quad \{\neg r\} \\ \swarrow \quad \searrow \quad \swarrow \quad \searrow \\ \{q\} \quad \{\neg q\} \\ \swarrow \quad \searrow \\ \square \end{array}$$

b) I. Dokazujeme přímo  $\{\{p, \neg r, q\}, \{p, \neg q, s\}, \{\neg r, \neg p\}\} \models \{\neg r, s\}$ .

$$\begin{array}{c} \{p, \neg r, q\} \quad \{p, \neg q, s\} \\ \swarrow \quad \searrow \\ \{p, \neg r, s\} \quad \{\neg r, \neg p\} \\ \swarrow \quad \searrow \\ \{\neg r, s\} \end{array}$$

II. Vyvrácíme množinu  $\{\{p, \neg r, q\}, \{p, \neg q, s\}, \{\neg r, \neg p\}, \{r\}, \{\neg s\}\}$ .

$$\begin{array}{c} \{p, \neg r, q\} \quad \{p, \neg q, s\} \\ \swarrow \quad \searrow \\ \{p, \neg r, s\} \quad \{r\} \quad \{r\} \quad \{\neg r, \neg p\} \\ \swarrow \quad \searrow \quad \swarrow \quad \searrow \\ \{p, s\} \quad \{\neg p\} \\ \swarrow \quad \searrow \\ \{s\} \quad \{\neg s\} \\ \swarrow \quad \searrow \\ \square \end{array}$$

**Příklad 7.1.4.** Dokažte, že rezoluční pravidlo zachovává pravdivost v interpretaci, tedy že pokud množina klauzulí  $\{\{p, \ell_1, \dots, \ell_n\}, \{\neg p, \ell'_1, \dots, \ell'_m\}\}$ , kde  $\ell_i$ , resp.  $\ell'_i$  jsou literály, je pravdivá v interpretaci  $I$ , pak klauzule  $\{\ell_1, \dots, \ell_n, \ell'_1, \dots, \ell'_m\}$  je rovněž pravdivá v interpretaci  $I$ .

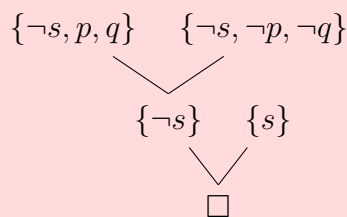
Uvažujme klauzule  $C_1 = \{p, \ell_1, \dots, \ell_n\}$ ,  $C_2 = \{\neg p, \ell'_1, \dots, \ell'_m\}$ , jejich rezolventu  $C = \{\ell_1, \dots, \ell_n, \ell'_1, \dots, \ell'_m\}$  a libovolnou interpretaci  $I$  takovou, že  $I(C_1) = I(C_2) = 1$ . Mohou nastat právě dvě možnosti:

- $I(p) = 1$ . V takovém případě je  $I(\neg p) = 0$ , a tedy, aby  $C_2$  byla pravdivá v  $I$ , musí pro nějaký literál  $\ell'_i \in C_2$  platit  $I(\ell'_i) = 1$ . Ovšem všechny literály  $\ell'_1, \dots, \ell'_m$  jsou zahrnuty i v rezolventě  $C$ , a tudíž  $I(C) = 1$ .
- $I(\neg p) = 1$ . (Analogicky k předchozímu.) Potom  $I(p) = 0$ , a tedy pro nějaký literál  $\ell_i \in C_1$  platí  $I(\ell_i) = 1$ . Potom ovšem rovněž  $I(C) = 1$ .

**Příklad 7.1.5.** Uvažujte upravené rezoluční pravidlo, kde rezolventou klauzulí  $\{p, q, \ell_1, \dots, \ell_n\}$ ,  $\{\neg p, \neg q, \ell'_1, \dots, \ell'_m\}$ , kde  $\ell_i$ , resp.  $\ell'_i$  jsou literály, je klauzule  $\{\ell_1, \dots, \ell_n, \ell'_1, \dots, \ell'_m\}$ . (Tedy rezolvovat lze na dvou proměnných naráz.) Zachovává toto rezoluční pravidlo pravdivost v interpretaci? Svou odpověď dokažte.

Tvrzení neplatí. Uvažme klauzule  $\{p, q\}$ ,  $\{\neg p, \neg q\}$  a interpretaci  $I(p) = 0$ ,  $I(q) = 1$ . Obě klauzule jsou v  $I$  pravdivé, ovšem jejich rezolventa  $\square$  v  $I$  pravdivá není.

Závěrem uvedeného důkazu je, že nelze rezolvovat na více literálech naráz, protože rezoluce musí zachovávat pravdivost v interpretaci. Následující rezoluční strom je tedy neplatný:



Všimněte si, že tímto chybným způsobem „vyvrácená“ množina předpokladů  $\{\{\neg s, p, q\}, \{\neg s, \neg p, \neg q\}, \{s\}\}$  má model např.  $I(s) = I(q) = 1$ ,  $I(p) = 0$ .

## 7.2 SLD rezoluce ve výrokové logice

Kromě obecné rezoluce existují i specifitější typy rezolucí. Jedním z nich je SLD rezoluce, tedy *lineární rezoluce na uspořádaných klauzulích s výběrovým pravidlem*, která se používá ke strojovému dokazování. Lineární rezoluce znamená, že výsledný rezoluční strom je ve tvaru jedné větve.

Výběrové pravidlo specifikuje, na kterém literálu rezolvujeme. Zde budeme používat výběrové pravidlo vybírající první literál v klauzuli – dává smysl pouze v uspořádaných klauzulích (*n-ticích*), nikoliv však v obecných klauzulích (množinách).

**Definice 37:** *Rezoluční pravidlo pro SLD rezoluci:* Uvažujme uspořádané klauzule  $[\neg p_1, \neg p_2, \dots, \neg p_n]$ ,  $[p_1, \neg q_1, \dots, \neg q_m]$ . Jejich SLD rezolventou je usp. klauzule  $[\neg q_1, \dots, \neg q_m, \neg p_2, \dots, \neg p_n]$ .

Rezolventou klauzulí  $[\neg r, \neg p, \neg s]$  a  $[r, \neg t, \neg s]$  je  $[\neg t, \neg s, \neg p, \neg s]$ . Klauzule  $[\neg r, \neg p, \neg s]$  a  $[r]$  pak mají rezolventu  $[\neg p, \neg s]$ .

U strojového dokazování pomocí SLD rezoluce se v souladu s Větou 36 používá vyvrácení množiny klauzulí.

SLD rezoluce je úplná pro speciální množinu klauzulí, které se nazývají Hornovy klauzule.

**Definice 38:** *Hornovy klauzule* jsou klauzule, které obsahují nejvýše jeden pozitivní literál (tedy literál bez negace). Ten v případě uspořádaných klauzulí navíc zapisujeme jako první. Dělíme je na

- *fakta* – klauzule s právě jedním, pozitivním literálem (např.  $[s]$ ,  $[r]$ ),
- *pravidla* – klauzule s právě jedním pozitivním literálem a alespoň jedním negativním literálem (např.  $[s, \neg r, \neg t]$ ,  $[r, \neg q]$ ),
- *cíle* – klauzule obsahující výhradně negativní literály (např.  $[\neg r, \neg t]$ ).

V rámci SLD rezoluce může obecně existovat více způsobů, jak se dobrat výsledku. Kompaktní způsob, který zachycuje prohledávání stavového prostoru a způsoby vyvrácení množiny klauzulí, je SLD strom.

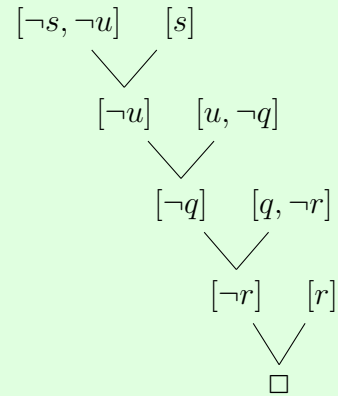
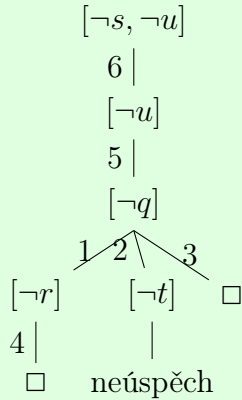
**Příklad.** Pomocí SLD rezoluce vyvraťte množinu  $\{[\neg s, \neg u], [q, \neg r], [q, \neg t], [q], [r], [u, \neg q], [s]\}$ . Nakreslete SLD strom systematického hledání řešení. Nakreslete rezoluční strom pro nejlepší úspěšnou větev SLD stromu.

Rezoluci budeme začínat cílem  $[\neg s, \neg u]$ . Pravidla a fakta seřadíme a očísujeme:

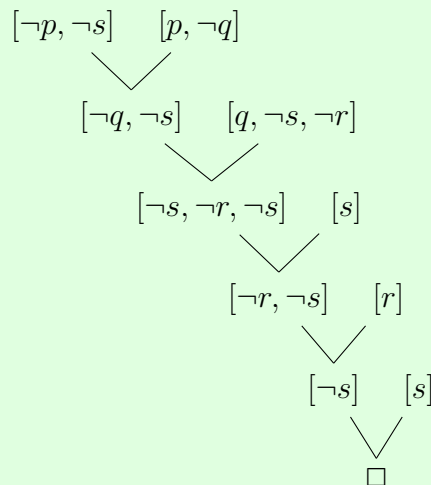
1:  $[q, \neg r]$  2:  $[q, \neg t]$  3:  $[q]$  4:  $[r]$  5:  $[u, \neg q]$  6:  $[s]$

SLD strom zachycuje všechna možná vyhodnocení. Použitá fakta a pravidla jsou naznačena číslem u příslušné hrany SLD stromu. Strom se větví právě v uzlech, kde lze aplikovat více faktů nebo pravidel.

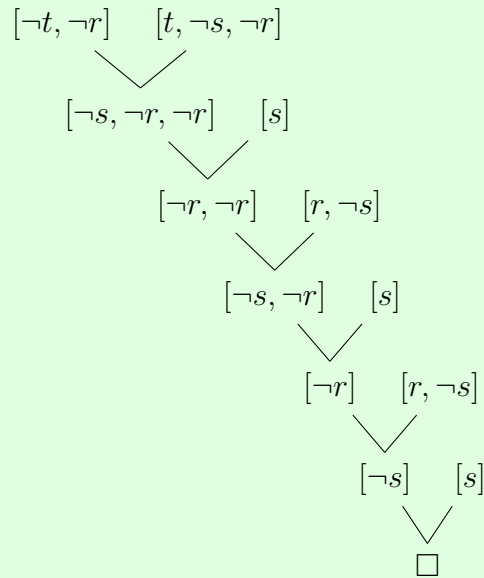
Každá (úspěšná) větev SLD stromu odpovídá jednomu rezolučnímu důkazu. Rezoluční strom pro nejlevější větev je uveden níže.



**Příklad 7.2.1.** ★ SLD rezolucí ukažte nesplnitelnost množiny uspořádaných klauzulí  $\{[q, \neg s, \neg r], [p, \neg q], [s], [r]\} \cup \{[\neg p, \neg s]\}$ .

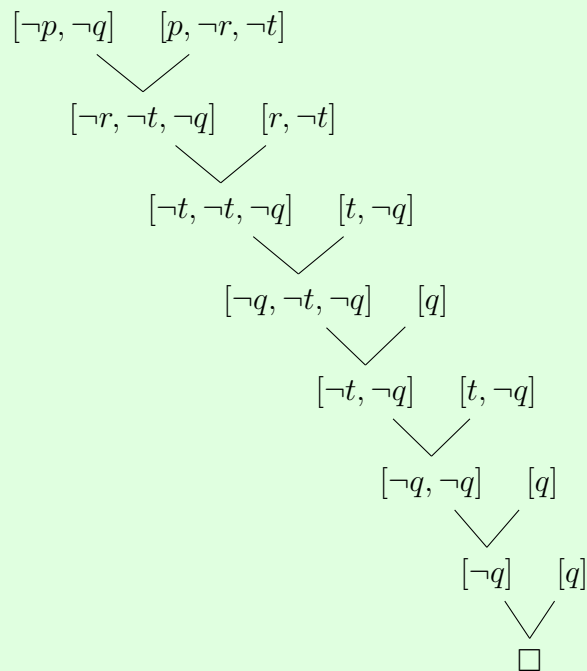


**Příklad 7.2.2.** SLD rezolucí vyvráťte množinu klauzulí  $\{[s], [t, \neg s, \neg r], [r, \neg s]\} \cup \{[\neg t, \neg r]\}$ .



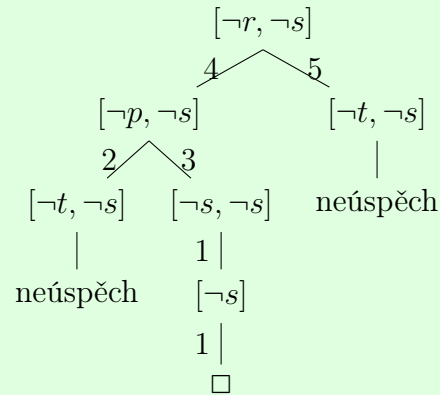
**Příklad 7.2.3.** SLD rezolučním vyvrácením ukažte, že  $\{q, (r \wedge t) \Rightarrow p, t \Rightarrow r, q \Rightarrow t\} \models p \wedge q$ .

Všechny formule uvedené jako předpoklady převedeme na uspořádané klauzule a do množiny přidáme negaci závěru:  $\{[q], [p, \neg r, \neg t], [r, \neg t], [t, \neg q], [\neg p, \neg q]\}$ . Takto vzniklou množinu pak vyvrácíme SLD rezolucí (počínaje klauzulí se samými negativními literály).



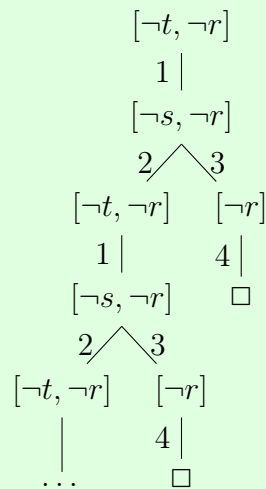
**Příklad 7.2.4.** Vytvořte SLD strom pro vyvrácení následující množiny uspořádaných klauzulí (čísla odpovídají prioritě při výběru daného předpokladu):  $\{1 : [s], 2 : [p, \neg t], 3 : [p, \neg s], 4 : [r, \neg p], 5 : [r, \neg t]\} \cup \{[\neg r, \neg s]\}$ .

Výsledný strom obsahuje dvě neúspěšné větve (posloupnosti použitých předpokladů 42 a 5) a jednu úspěšnou větev (posloupnost předpokladů 4311).



**Příklad 7.2.5.** Vytvořte SLD strom pro vyvrácení následující množiny uspořádaných klauzulí (čísla odpovídají prioritě při výběru daného předpokladu):  $\{1 : [t, \neg s], 2 : [s, \neg t], 3 : [s], 4 : [r]\} \cup \{[\neg t, \neg r]\}$ .

Výsledný strom obsahuje nekonečnou větev (posloupnost použitých předpokladů 1212...) a nekonečně mnoho úspěšných větví (posloupnosti předpokladů 1212...134).





### 7.3 Obecná rezoluce v predikátové logice

Rezoluce v predikátové logice pracuje s formulemi ve Skolemově normální formě (viz kapitulu 6.3 Normální formy). Klausule se opět zapisují jako množiny literálů, formule jako množiny klauzulí. Všeobecné kvantifikátory se v zápisu vynechávají. Formule  $\forall x \forall y ((P(y) \vee \neg Q(y, x)) \wedge \neg P(f(c)))$  se zapíše jako  $\{\{P(y), \neg Q(y, x)\}, \{\neg P(f(c))\}\}$ .

Co když budeme chtít rezolovat klauzule  $\{P(y), \neg Q(y, x)\}$  a  $\{\neg P(f(c))\}$  na predikátu  $P$ ? Povšimněte si, že jako argument predikátu  $P$  je v první klauzuli použit term  $y$ , v druhé klauzuli term  $f(c)$ . Abychom mohli na  $P$  rezolovat, je nutné, aby obě klauzule „hovořily o stejném objektu“, takže je potřeba nejprve sjednotit (neboli *unifikovat*) použité termy.

V uvažovaném příkladě se nabízí nahrazení (neboli *substitute*) proměnné  $y$  termem  $f(c)$ , po kterém bude první klauzule tvaru  $\{P(f(c)), \neg Q(f(c), x)\}$  – všimněte si náhrady  $y$  za  $f(c)$  i v rámci predikátu  $Q$ . Jsme ovšem oprávněni takovou náhradu beztretně provést? Podívejme se, jak vypadá původní klauzule v klasickém zápisu s kvantifikátory:  $\forall x \forall y (P(y) \vee \neg Q(y, x))$  a jak vypadá klauzule po substituci:  $\forall x (P(f(c)) \vee \neg Q(f(c), x))$ . Lze konstatovat, že substituce proměnné za jiný term zachovala pravdivost. Platí-li totiž původní klauzule pro všechna  $y$ , pak platí pro libovolný term, kterým  $y$  nahradíme.

Jelikož substituce zachovává pravdivost, lze ji provádět v rámci aplikace rezolučního pravidla (a občas je to nevyhnutelné). Odvození rezolventy z výše uvedených klauzulí budeme tedy celkem zapisovat následovně:

$$\begin{array}{ccc} \{P(y), \neg Q(y, x)\} & & \{\neg P(f(c))\} \\ & \searrow \quad \swarrow & \\ & y/f(c) & \\ & & \{\neg Q(f(c), x)\} \end{array}$$

Formalizaci obecné rezoluce v predikátové logice zahájíme u pojmu *substitute*.

**Definice 39:** *Substitute* je množina  $\{x_1/t_1, \dots, x_n/t_n\}$ , kde  $x_i$  jsou různé proměnné a  $t_i$  jsou termy.

Speciálně jsou-li všechna  $t_i$  proměnné, hovoříme o *přejmenování proměnných*.

Substitute jsou například množiny  $\{x/y, y/f(c)\}$ ,  $\{y/f(x)\}$ ,  $\{x/y, y/z, z/x\}$ . Poslední uvedená je navíc přejmenováním proměnných.

Substituci  $\Phi = \{x_1/t_1, \dots, x_n/t_n\}$  lze aplikovat na množinu literálů  $S$  tak, že každý výskyt každé proměnné  $x_i$  nahradíme termem  $t_i$ . Aplikaci substitute zapisujeme jako  $S\Phi$ .

**Příklad.** Pro množiny  $S_1 = \{P(x, y), \neg Q(y, f(x))\}$ ,  $S_2 = \{P(f(x)), P(y), P(f(f(c)))\}$  a substitute  $\Phi_1 = \{x/c, y/b\}$ ,  $\Phi_2 = \{x/f(c), y/f(f(c))\}$  stanovte  $S_1\Phi_1$  a  $S_2\Phi_2$ .

$$S_1\Phi_1 = \{P(c, b), \neg Q(b, f(c))\}$$

$$S_2\Phi_2 = \{P(f(f(c)))\}.$$

Substituce  $\Phi_2$  ve vztahu k množině  $S_2$  je v jistém smyslu speciální, všechny literály množiny  $S_2$  totiž substitucí přešly na ten samý literál. Takovým substitucím říkáme *unifikátory*.

**Definice 40:** Substituce  $\Phi$  je *unifikátorem* množiny  $S$ , pokud  $|S\Phi| = 1$ .

Existuje-li pro množinu  $S$  unikátor, hovoříme o *unifikovatelné* množině.

Unifikátorem množiny  $\{P(z, f(z)), P(f(x), f(f(c)))\}$  je substituce  $\{z/f(c), x/c\}$ . Množina  $\{P(z, f(z)), P(f(x), f(c))\}$  není unikovatelná.

Unifikátory budou sloužit k dříve avizovanému sjednocení termů při aplikování rezolučního pravidla. To z technických důvodů aplikujeme na dvojice klauzulí, které nemají společné proměnné. Pokud mají, nejprve proměnné přejmenujeme.

**Definice 41:** *Rezoluční pravidlo pro predikátovou logiku:* Uvažujme dvě klauzule bez společných proměnných  $\{P(\vec{t}_1), \dots, P(\vec{t}_r), L_1, \dots, L_n\}$ ,  $\{\neg P(\vec{t}'_1), \dots, \neg P(\vec{t}'_s), L'_1, \dots, L'_m\}$  a substituci  $\Phi$ , která je unikátorem množiny  $\{P(\vec{t}_1), \dots, P(\vec{t}_r), P(\vec{t}'_1), \dots, P(\vec{t}'_s)\}$ . Jejich *rezolventou* je klauzule  $\{L_1, \dots, L_n, L'_1, \dots, L'_m\}\Phi$ .

Rezolventou klauzulí  $\{P(x), \neg Q(x)\}$  a  $\{Q(c)\}$  je  $\{P(c)\}$  s příslušným unikátorem  $\Phi = \{x/c\}$ . Rezolventou klauzulí  $\{R(x, y), R(c, y), \neg Q(y)\}$  a  $\{\neg R(z, d), P(z)\}$  je klauzule  $\{\neg Q(d), P(c)\}$  (příslušný unikátor  $\Phi = \{x/c, y/d, z/c\}$ ).

Obecnou rezoluci v predikátové logice jinak provádíme stejným způsobem jako v logice výrokové. Pojmy rezolučního důkazu a rezolučního vyvrácení zůstávají nezměněny, pouze využívají rezolučního pravidla pro predikátovou logiku. Navíc na hrany do rezolučního stromu píšeme provedené substituce/unifikace.

**Příklad 7.3.1.** ★ Nalezněte  $S\Phi$  pro zadaná  $S$  a  $\Phi$ .

- $S = \{P(x), Q(y)\}$ ,  $\Phi = \{x/y, y/x\}$
- $S = \{P(c, x), \neg P(x, f(z))\}$ ,  $\Phi = \{f(z)/x, x/d\}$
- $S = \{P(x, y), P(c, z), P(c, c)\}$ ,  $\Phi = \{x/y, y/z, z/c\}$

- $S\Phi = \{P(y), Q(x)\}$
- $S\Phi$  není definováno (viz Definice 39)
- $S\Phi = \{P(y, z), P(c, c)\}$

**Příklad 7.3.2.** ★ Nalezněte (nejobecnější) unifikátory následujících množin literálů ( $a, b, c, d$  jsou konstanty).

- a)  $S = \{P(x), P(y)\}$
- b)  $S = \{P(x), Q(y)\}$
- c)  $S = \{Q(x, x), Q(y, c)\}$
- d)  $S = \{Q(h(x, y), w), Q(h(g(v), a), f(v)), Q(h(g(v), a), f(b))\}$

- a)  $\Phi = \{y/x\}a$
- b) Množina není unifikovatelná.
- c)  $\Phi = \{x/c, y/c\}$
- d)  $\Phi = \{x/g(b), y/a, w/f(b), v/b\}$

**Příklad 7.3.3.** Nalezněte (nejobecnější) unifikátory následujících množin ( $c, d$  jsou konstanty).

- a)  $S = \{P(x, f(x)), P(c, z), P(x, f(y))\}$
- b)  $S = \{P(x, f(x)), P(x, z), P(x, f(y))\}$
- c)  $S = \{P(x, f(y), z), P(g(a), f(w), u), P(v, f(b), c)\}$
- d)  $S = \{R(f(x, g(c)), x), R(z, g(d)), R(f(g(y), w), u)\}$
- e)  $S = \{Q(x, x), Q(y, c), Q(d, z)\}$

- a)  $\Phi = \{x/c, z/f(c), y/c\}$
- b)  $\Phi = \{z/f(x), y/x\}$
- c)  $\Phi = \{x/g(a), v/g(a), y/b, w/b, z/c, u/c\}$
- d)  $\Phi = \{x/g(d), y/d, z/f(g(d), g(c)), u/g(d), w/g(c)\}$
- e) Množina není unifikovatelná.

**Příklad 7.3.4.** Nalezněte rezolventy následujících dvojic klauzulí ( $c$  je konstanta).

- a)  $\{P(x), \neg Q(z)\}, \{\neg R(y), Q(y)\}$
- b)  $\{\neg P(c, x), Q(x)\}, \{P(y, y), \neg R(y, z)\}$
- c)  $\{Q(x), \neg P(c)\}, \{P(y), \neg R(y, x)\}$
- d)  $\{P(f(x), z)\}, \{\neg P(g(x), f(y))\}$
- e)  $\{P(x, f(z)), P(y, y_1)\}, \{\neg P(c, z)\}$

f)  $\{Q(f(f(f(z))), f(f(y))), \{ \neg Q(f(f(x)), x) \}$

a) Rezolventa  $\{P(x), \neg R(y)\}$  s unifikátorem  $\Phi = \{z/y\}$  nebo  $\{P(x), \neg R(z)\}$  s  $\Phi = \{y/z\}$ .

b) Rezolventa  $\{Q(c), \neg R(c, z)\}$  s unifikátorem  $\Phi = \{y/c, x/c\}$ .

c) Ještě před provedením rezoluce musíme ve druhé klauzuli přejmenovat  $x$  na  $x_1$ , aby klauzule obsahovaly jiné proměnné.

Rezolventa  $\{Q(x), \neg R(c, x_1)\}$  s unifikátorem  $\Phi = \{y/c\}$ .

d) Neexistuje vhodný unifikátor, a tudíž ani rezolventa.

e) Ještě před provedením rezoluce musíme ve druhé klauzuli přejmenovat  $z$  na  $z_1$ , aby klauzule obsahovaly jiné proměnné.

Rezolventa  $\square$  s unifikátorem  $\Phi = \{x/c, y/c, z_1/f(z), y_1/f(z)\}$ .

f) Rezolventa  $\square$  s unifikátorem  $\Phi = \{x/f(f(y)), z/f(y)\}$ .

**Příklad 7.3.5. ★** Uvažujte následující tvrzení:

- „Existuje drak.“
- „Draci spí nebo loví.“
- „Když mají draci hlad, nespí.“

Rezolucí dokažte, že z uvedených tvrzení plyne „Když má drak hlad, loví,“ a to dvěma způsoby: (1) přímým odvozením závěru a (2) vyvrácením množiny obsahující předpoklady a negaci závěru. Zejména v řešení

- formalizujte uvedená tvrzení (předpoklady a závěr) v jazyce predikátové logiky,
- určete význam (interpretaci) jednotlivých predikátů,
- proved'te skolemizaci tvrzení a zapište je v klauzulárním tvaru a
- nakonec proved'te samotnou rezoluci.

a) Předpoklady:  $\exists x D(x); \forall x (D(x) \Rightarrow (S(x) \vee L(x))); \forall x ((D(x) \wedge H(x)) \Rightarrow \neg S(x))$ .

Závěr:  $\forall x ((D(x) \wedge H(x)) \Rightarrow L(x))$ .

Negace závěru:  $\exists x (D(x) \wedge H(x) \wedge \neg L(x))$ .

b) Intuitivní interpretaci predikátů zadáme tabulkou.

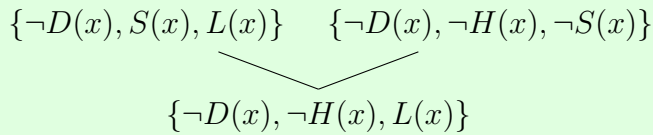
$D(x)$	$x$ je drak
$H(x)$	$x$ má hlad
$L(x)$	$x$ loví
$S(x)$	$x$ spí

c) Předpoklady:  $\{D(a)\}; \{\neg D(x), S(x), L(x)\}; \{\neg D(x), \neg H(x), \neg S(x)\}$ .

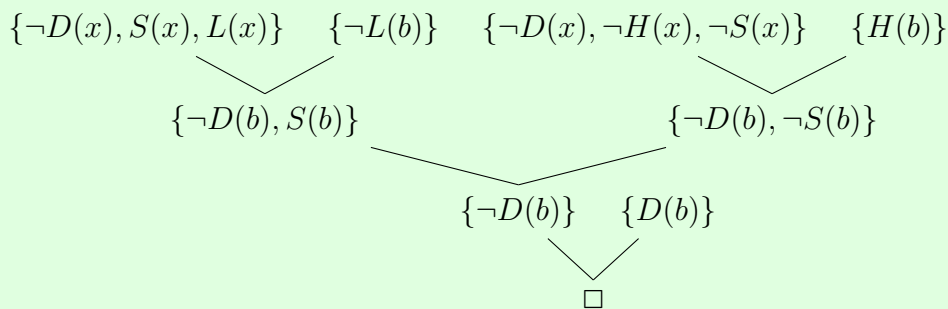
Závěr:  $\{\neg D(x), \neg H(x), L(x)\}$ .

Negace závěru:  $\{D(b)\}, \{H(b)\}, \{\neg L(b)\}$ .

d) (1)



(2)



**Příklad 7.3.6.** Uvažujte následující tvrzení:

- „Jablka nepadají daleko od stromu.“
- „Jím jenom jablka.“
- „Všechno padá daleko od stromu.“

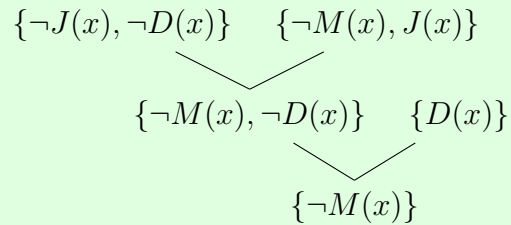
Dokažte pomocí rezoluce, že z těchto tvrzení plyne „Nejím nic,“ a to dvěma způsoby: (1) přímým odvozením závěru a (2) vyvrácením množiny obsahující předpoklady a negaci závěru. Zejména v řešení

- převeďte tvrzení do formální reprezentace,
- určete význam (interpretaci) jednotlivých predikátů,
- provedte skolemizaci jednotlivých formulí a
- nakonec provedte samotnou rezoluci.

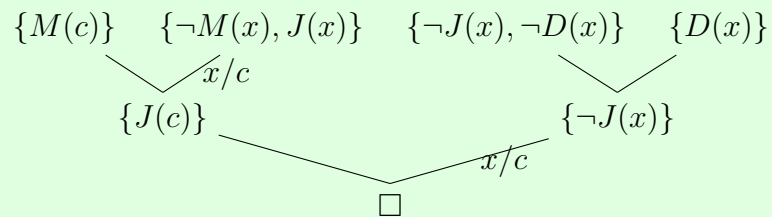
- Máme množinu předpokladů  $\{\forall x(J(x) \Rightarrow \neg D(x)), \forall x(M(x) \Rightarrow J(x)), \forall x D(x)\}$  a závěr  $\forall x \neg M(x)$ .
- Významy predikátů jsou zadány tabulkou.

predikát	výnam
$J(x)$	$x$ je jablko
$D(x)$	$x$ padá daleko od stromu
$M(x)$	jím $x$

- c) Množina předpokladů po skolemizaci je  $\{\{\neg J(x), \neg D(x)\}, \{\neg M(x), J(x)\}, \{D(x)\}\}$ . Závěr po skolemizaci je  $\{\neg M(x)\}$ , negace závěru  $\{M(c)\}$ .
- d) (1) Přímé odvození závěru.



- (2) Vyvrácení množiny s negací závěru.



## 7.4 Prolog

Programovací jazyk Prolog využívá SLD rezoluci k systematickému zjišťování, zda ze zadaných předpokladů plyne závěr. Nejprve se podíváme, jak vypadá program jazyka Prolog, a následně si ukážeme korespondenci mezi Prologem a SLD rezolucí.

**Definice 42:** Program jazyka Prolog (výroková logika) je konečná posloupnost předpokladů  $P_1, \dots, P_n$ , kde předpoklady jsou dvojího typu:

- *Fakta*, která jsou tvaru  $p$ ., kde  $p$  je výroková proměnná. Fakt  $p$ . je ekvivalentní formuli  $p$ .
- *Pravidla*, která jsou tvaru  $p \text{ :- } q, \dots, r$ ., kde  $p, q, \dots, r$  jsou výrokové proměnné. Pravidlo  $p \text{ :- } q, \dots, r$  je ekvivalentní formuli  $p \Leftrightarrow (q \wedge \dots \wedge r)$ .

Nad zadaným programem jazyka Prolog se vyhodnocují dotazy. Dotaz je tvrzení, jehož vyplývání z předpokladů uvedených v programu chceme ověřit.

**Definice 43:** *Dotazy (výroková logika)* jsou tvaru  $?- p, \dots, q.$ , kde  $p, \dots, q$  jsou výrokové proměnné. Dotaz  $?- p, \dots, q.$  je ekvivalentní formuli  $p \wedge \dots \wedge q.$

Ukažme si nyní korespondenci programu Prologu s Hornovými klauzulemi. Fakt  $p.$  lze přímočaře vyjádřit uspořádanou klauzulí  $[p]$ . Pravidlo tvaru  $p :- q, r.,$  ekvivalentní formuli  $p \Leftarrow (q \wedge r),$  lze vyjádřit jako  $p \vee \neg(q \wedge r),$  neboli  $p \vee \neg q \vee \neg r.$  To odpovídá uspořádané klauzuli  $[p, \neg q, \neg r].$  Konečně chceme rezolucí dokázat, že z faktů a pravidel vyplývá závěr vyjádřený dotazem  $?- s, r.$  s významem  $s \wedge r.$  Pro důkaz rezolučním vyvrácením přidáme k pravidlům a faktům negaci formule  $s \wedge r,$  tedy  $\neg s \vee \neg r,$  neboli uspořádanou klauzuli  $[\neg s, \neg r].$  Uvedené klauzule odpovídají faktům, pravidlům a cíli tak, jak jsou definovány pro Hornovy klauzule (viz Definici 38).

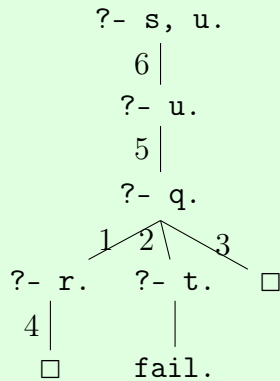
Díky přímé korespondenci s Hornovými klauzulemi lze pro programy Prologu vytvářet SLD rezoluční stromy. Zápis se liší v tom, že místo cíle tvaru  $[-p, \dots, \neg q]$  zapisujeme průběžně vyhodnocovaný dotaz ve tvaru  $?- p, \dots, q.$

Při zadání dotazu nad programem jazyka Prolog probíhá systematické hledání řešení, které odpovídá procházení SLD stromu do hloubky s prioritním výběrem pravidla nebo cíle uvedeného v programu dříve (tj. procházením alternativ zleva doprava). V případě vyvrácení zadaného cíle dojde k vytištění hodnoty `true`.

**Příklad.** Vytvořte SLD strom pro dotaz  $?- s, u.$  nad následujícím programem Prologu:

1.  $q :- r.$
2.  $q :- t.$
3.  $q.$
4.  $r.$
5.  $u :- q.$
6.  $s.$

Rezoluci začínáme dotazem  $?- s, u.$  Z pravidel a faktů programu při vyhodnocování vybíráme postupně shora dolů to, které odpovídá první proměnné dotazu. V případě faktu pak příslušnou proměnnou umažeme, v případě pravidla nahradíme pravou stranou za symbolem  $:-$ . Použitá fakta a pravidla jsou naznačena číslem  $u$  příslušné hrany SLD stromu. Strom se větví právě v uzlech, kde lze aplikovat více faktů nebo pravidel. Prolog při vyhodnocování vždy vstupuje do nejprve do levých větví (tj. těch s nižším číslem použitého pravidla nebo faktu).



Úspěšné větve jsou zakončeny symbolem  $\square$ , neúspěšné symbolem `fail.` V běžných implementacích Prologu se programy, které obsahují proměnnou nedefinovanou ve faktu nebo na levé straně nějakého pravidla (jako např. `t` v tomto programu), ani nespustí. V případě výrokové logiky existence takové proměnné přesně odpovídá existenci neúspěšné větve v SLD stromu.

Všimněte si, že ten samý SLD strom jsme již v jiném hávu viděli v ukázkovém příkladě v podkapitole o SLD rezoluci.

Vedle úspěšných a neúspěšných větví uvedených v příkladu může navíc SLD strom obsahovat i větve nekonečné, kdy Prolog cyklí.

Použití Prologu v predikátové logice je o něco zajímavější. Predikáty zapisujeme malými písmeny (podobně jako výrokové symboly), do závorek uvádíme termy. Proměnné se značí velkými písmeny: `X`, `Y`, `Z`. Syntax Prologu i tvorbu SLD stromů si demonstrováme na příkladě.

**Příklad.** Uvažujte následující program jazyka Prolog:

- 1: `otec(petr, marta).`
- 2: `otec(jan, petr).`
- 3: `otec(petr, jana).`
- 4: `matka(dana, marta).`
- 5: `matka(jitka, petr).`
- 6: `matka(dana, kveta).`
- 7: `rodic(X, Y) :- otec(X, Y).`
- 8: `rodic(X, Y) :- matka(X, Y).`
- 9: `babicka(X, Y) :- matka(X, Z), rodic(Z, Y).`

- a) Zjistěte, zda je Dana rodičem Marty. Nakreslete SLD strom pro zadaný dotaz.
- b) Zjistěte, koho je Jitka babičkou. Nakreslete SLD strom pro zadaný dotaz.

a) Zadaný dotaz je `?- rodic(dana, marta).` SLD strom vypadá takto:



```

?- rodic(dana, marta).
  7, X/dana, Y/marta      8, X/dana, Y/marta
?- otec(dana, marta).    ?- matka(dana, marta).
  |                          4 |
  fail.                      □

```

Všimněte si, že při aplikaci pravidel 7 i 8 došlo k unifikaci proměnných  $X$ ,  $Y$  s konstantami `dana`, `marta`. Úspěšná větev potvrzuje, že Dana je skutečně rodičem Marty.

Při zadání dotazu Prolog vypíše `true`. Protože nalezený výsledek je v nejpravější větvi, další prohledávání nebude umožněno.

b) Zadaný dotaz je `?- babicka(jitka, X)`. SLD strom vypadá takto:

```

?- babicka(jitka, X).
  | 9, P1/jitka, P2/X, P3/Z*
?- matka(jitka, Z), rodic(Z,X).
  | 5, Z/petr
?- rodic(petr, X).
  7, P1/petr, P2/X*      8, P1/petr, P2/X*
?- otec(petr, X).      ?- matka(petr, X).
  1, X/marta / 3, X/jana      |
  □ □                          fail.

```

Proměnné  $P1$ ,  $P2$  a  $P3$  u substitucí s hvězdičkou vznikly přejmenováními proměnných v pravidlech (kvůli konfliktu názvů). Použitá pravidla po přejmenování (v pořadí, v jakém byla použita) byla tvaru

9: `babicka(P1, P2) :- matka(P1, P3), rodic(P3, P2)`.

7: `rodic(P1, P2) :- otec(P1, P2)`.

8: `rodic(P1, P2) :- matka(P1, P2)`.

Z SLD stromu lze vyčíst, že Jitka je babičkou Marty a Jany (dle substituce `X/marta`, resp. `X/jana` ve dvou úspěšných větvích SLD stromu).

Při zadání dotazu Prolog vypíše nejprve `X=marta`, po zadání středníku `X=jana`. Po opětovném zadání středníku vypíše `false` (zbytek stromu neobsahuje úspěšnou větev) a neumožní další prohledávání.

---

**Příklad 7.4.1.** ★ Uvažujte následující program v Prologu:

1. `s`.
2. `t`.

3.  $p :- q.$
4.  $p :- t.$
5.  $q :- s, t.$

a dotaz  $?- p, t.$

- a) Převed'te program a dotaz na odpovídající uspořádané klauzule.
- b) Připomeňte klasifikaci uspořádaných klauzulí využívaných v Prologu. Klasifikujte klauzule z programu a dotazu.
- c) Proveďte vyhodnocení dotazu pomocí SLD rezoluce. Diskutujte obě možnosti zápisu uzlů stromu – tedy klasický zápis s klauzulemi jako u obecné a LI rezoluce a zápis průběžného vyhodnocování dotazu  $?- p, t.$  Prologem.

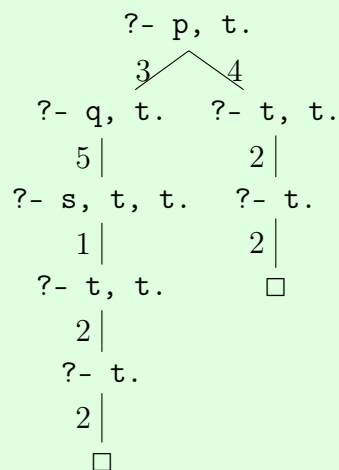
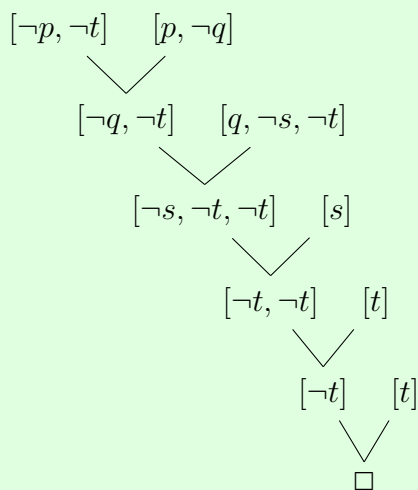
a) Program:  $[s], [t], [p, \neg q], [p, \neg t] [q, \neg s, \neg t].$

Dotaz:  $[\neg p, \neg t].$

- b) V Prologu se využívají Hornovy klauzule, čili klauzule s nejvýše jedním pozitivním literálem (literálem neobsahujícím negaci). Klauzule s právě jedním pozitivním literálem se dělí na *fakta* (žádný negativní literál) a *pravidla* (alespoň jeden negativní literál). Klauzule pouze s negativními literály se nazývají *cíle* (odpovídají dotazům v Prologu).

č. ř.	klauzule	v Prologu	typ
1	$[s]$	$s.$	fakt
2	$[t]$	$t.$	fakt
3	$[p, \neg q]$	$p :- q.$	pravidlo
4	$[p, \neg t]$	$p :- t.$	pravidlo
5	$[q, \neg s, \neg t]$	$q :- s, t.$	pravidlo
-	$[\neg p, \neg t]$	$?- p, t.$	cíl

- c) SLD strom lze zapisovat dvěma způsoby: první způsob je stejný jako u ostatních typů rezoluce. Druhý ukazuje postupné vyhodnocování dotazu, použitá pravidla/fakta jsou naznačena jejich číslem u příslušné hrany. Výhodou tohoto zápisu je možnost zachycení více způsobů vyhodnocení dotazu (strom se větví v místech, kde lze použít více pravidel).

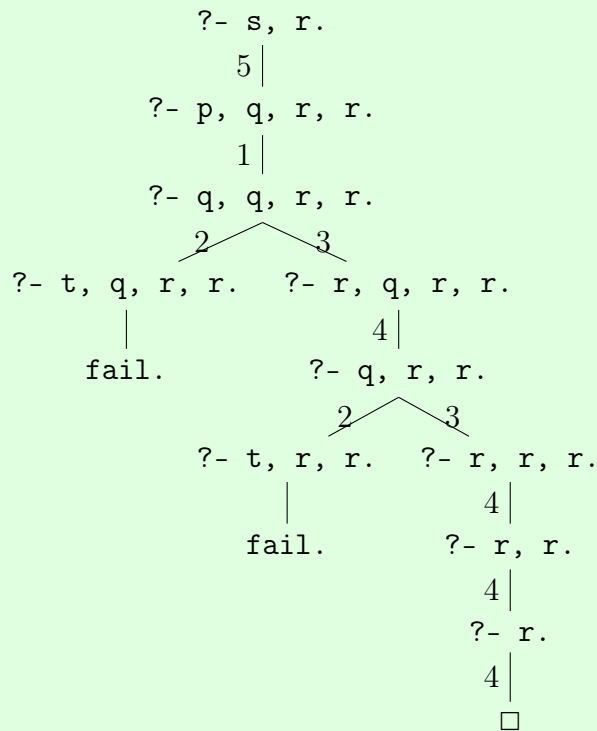


**Příklad 7.4.2.** Uvažujte následující program Prologu:

1.  $p :- q.$
2.  $q :- t.$
3.  $q :- r.$
4.  $r.$
5.  $s :- p, q, r.$

- a) Vytvořte SLD strom pro dotaz  $?- s, r.$
- b) Kolik obsahuje SLD strom úspěšných, resp. neúspěšných větví?
- c) Upravte program tak, aby SLD strom neobsahoval neúspěšné větve.

a) SLD strom vypadá následujícím způsobem:



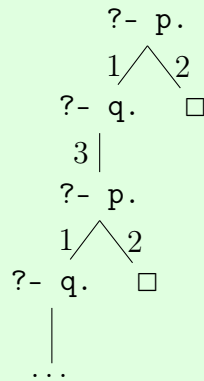
- b) Program obsahuje dvě neúspěšné a jednu úspěšnou větev.
- c) Neúspěšné větve lze odříznout odstraněním řádku 2. z programu.

**Příklad 7.4.3.** Uvažujte následující program Prologu:

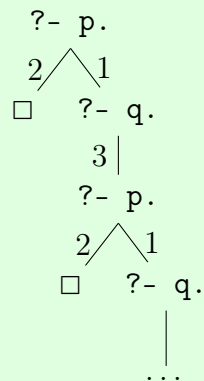
1. p :- q.
2. p.
3. q :- p.

- a) Vytvořte SLD strom pro dotaz ?- p. Jaký výsledek vrátí Prolog po zadání dotazu?
- b) Navrhněte uspořádání pravidel, aby se Prolog nad žádným dotazem nezacyklil.
- c) Kolika způsoby může Prolog v takto upraveném programu dojít k výsledku true.?

- a) Prolog nevrátí žádný výsledek, zacyklí se v nejlevější větvi. (Předpokládáme-li konečný zásobník, skončí s přetečením zásobníku.)



- b) Umístění řádku 1. před řádkem 2. vede k cyklící nejlevější větvi. Prohozením těchto řádků bude nejpravější větev nekonečná. Prolog hledá výsledky nejprve v levých větvích, k zacyklení tedy nedojde.
- c) SLD strom vyhodnocení dotazu `?- p.` nad upraveným programem vypadá takto:



K výsledku `true.` tedy Prolog dojde nekonečně mnoha způsoby. (Vždy posloupností aplikací řádků 1313...132).

**Příklad 7.4.4.** Navrhněte program Prologu (ve výrokové logice) a dotaz, který obsahuje právě jednu úspěšnou větev, právě jednu nekonečnou větev a právě jednu neúspěšnou větev (nacházející se v SLD stromu v tomto pořadí zleva). (*Úkol za bonus \*1: Nalezněte takový program o 3 řádcích.*)

Například dotaz `?- p.` nad programem

```

p.
p :- q.
q :- q, q.

```

```
p :- t.
```

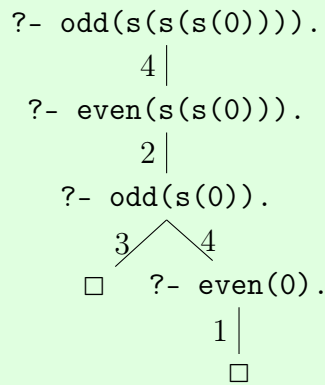
**Příklad 7.4.5.** ★ Uvažujte následující program v Prologu:

1. `even(0)` .
2. `even(s(X)) :- odd(X)` .
3. `odd(s(0))` .
4. `odd(s(X)) :- even(X)` .

a dotaz `?- odd(s(s(s(0))))` .

- a) Zapište SLD strom vyhodnocení dotazu v jazyce Prolog (se všemi možnostmi).
- b) Kolika způsoby může dojít Prolog k závěru, že cíl je pravdivý?
- c) Jak je potřeba upravit program, aby došel Prolog k výsledku vždy jediným způsobem?
- d) Jaké výsledky bude Prolog postupně vracet pro dotaz `?- odd(X)` .?

a) Hledaný SLD strom vypadá následujícím způsobem:



- b) Dvěma.
- c) Odstranit třetí řádek s redundantní informací.
- d) `X = s(0)` ;  
`X = s(0)` ;  
`X = s(s(s(0)))` ;  
`X = s(s(s(0)))` ;  
`X = s(s(s(s(s(0)))))` ;  
...

**Příklad 7.4.6.** Uvažujte následující program jazyka Prolog:

```
1: p(f(f(X))) :- p(X).
```

2:  $p(f(0))$ .

- Vytvořte SLD strom pro dotaz  $?- p(f(f(f(0))))$ . Co vypíše Prolog po zadání dotazu?
- Vytvořte SLD strom pro dotaz  $?- p(f(f(f(f(0)))))$ . Co vypíše Prolog po zadání dotazu?
- Vytvořte SLD strom pro dotaz  $?- p(X)$ . Co vypíše Prolog po zadání dotazu?
- Jak se bude Prolog chovat nad dotazem  $?- p(X)$ , prohodíme-li řádky 1 a 2?

a) Prolog vypíše **true**.

```
?- p(f(f(f(0))))..
    | 1, X/f(0)
?- p(f(0)).
    | 2
    □
```

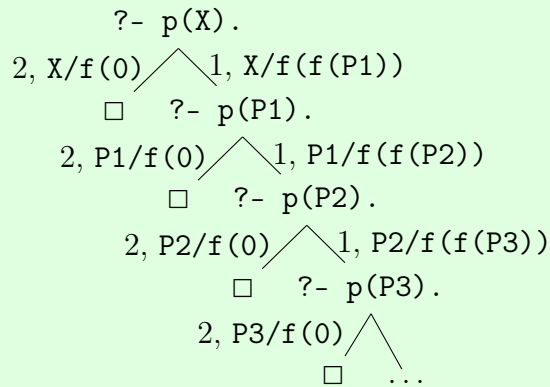
b) Prolog vypíše **false**.

```
?- p(f(f(f(f(0))))).
    | 1, X/f(f(0))
?- p(f(f(0))).
    | 1, X/0
?- p(0).
    |
    fail.
```

c) Prolog nic nevypíše, protože se zacyklí. (Nakonec vypíše chybu o přetečení zásobníku.)

```
?- p(X).
1, X/f(f(P1)) \ 2, X/f(0)
?- p(P1). □
1, P1/f(f(P2)) \ 2, P1/f(0)
?- p(P2). □
1, P2/f(f(P3)) \ 2, P2/f(0)
?- p(P3). □
    \ 2, P3/f(0)
    ... □
```

d) Prolog vypíše  $X=f(0)$  a dále po zadání středníku  $X=f(f(f(0)))$ ,  $X=f(f(f(f(0))))$  atp. Příslušný SLD strom vypadá následovně.

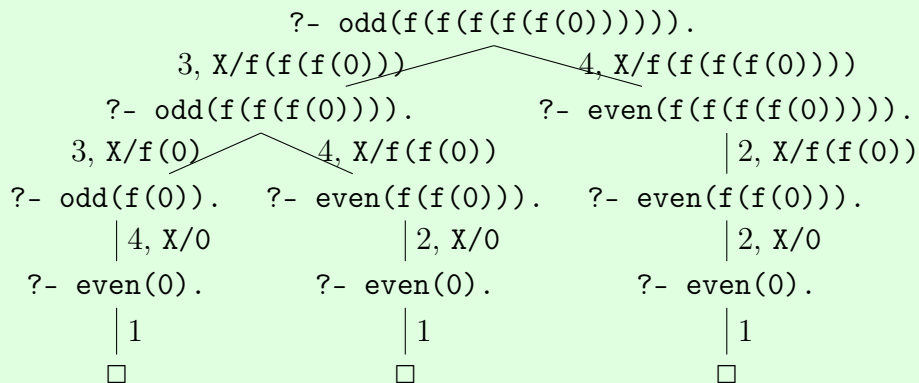


K rekonstrukci hodnot proměnné  $X$  dochází od úspěšného listu směrem ke kořeni – např.  $P1/f(0)$  a poté  $X/f(f(P1))$  dá celkem  $X/f(f(f(0)))$ .

**Příklad 7.4.7.** Uvažujte následující program jazyka Prolog:

- 1: `even(0).`
- 2: `even(f(f(X))) :- even(X).`
- 3: `odd(f(f(X))) :- odd(X).`
- 4: `odd(f(X)) :- even(X).`

Vytvořte SLD strom pro dotaz `?- odd(f(f(f(f(f(0))))))`.

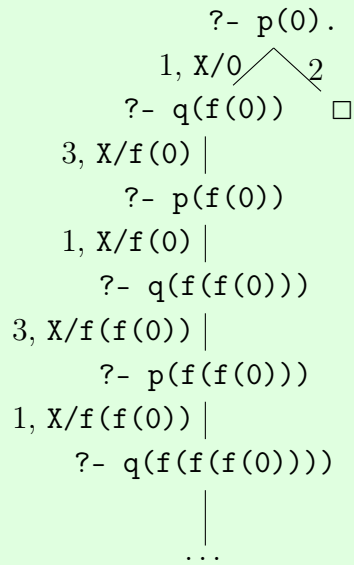


**Příklad 7.4.8.** Uvažujte následující program jazyka Prolog:

- 1: `p(X) :- q(f(X)).`
- 2: `p(0).`
- 3: `q(X) :- p(X).`

Vytvořte SLD strom pro dotaz `?- p(0)`.

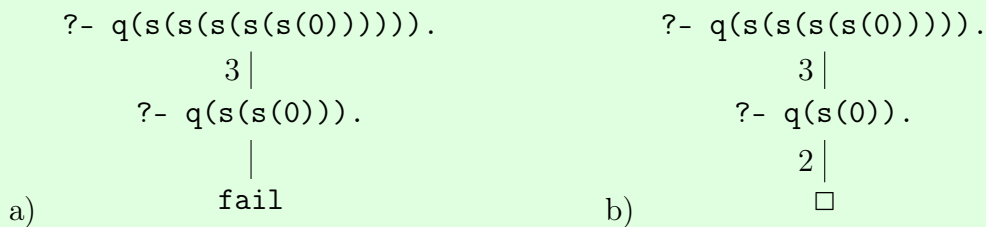




**Příklad 7.4.9.** Uvažujte následující program v Prologu:

1. `q(0).`
2. `q(s(0)).`
3. `q(s(s(s(X)))) :- q(X).`

- a) Zapište SLD strom vyhodnocení dotazu `?- q(s(s(s(s(s(0))))))`. v uvedeném programu.
- b) Zapište SLD strom vyhodnocení dotazu `?- q(s(s(s(s(0)))))`. v uvedeném programu.
- c) Pro jaké dotazy vrací **true**?
- d) Kolika způsoby může dojít Prolog k závěru, že cíl je pravdivý?



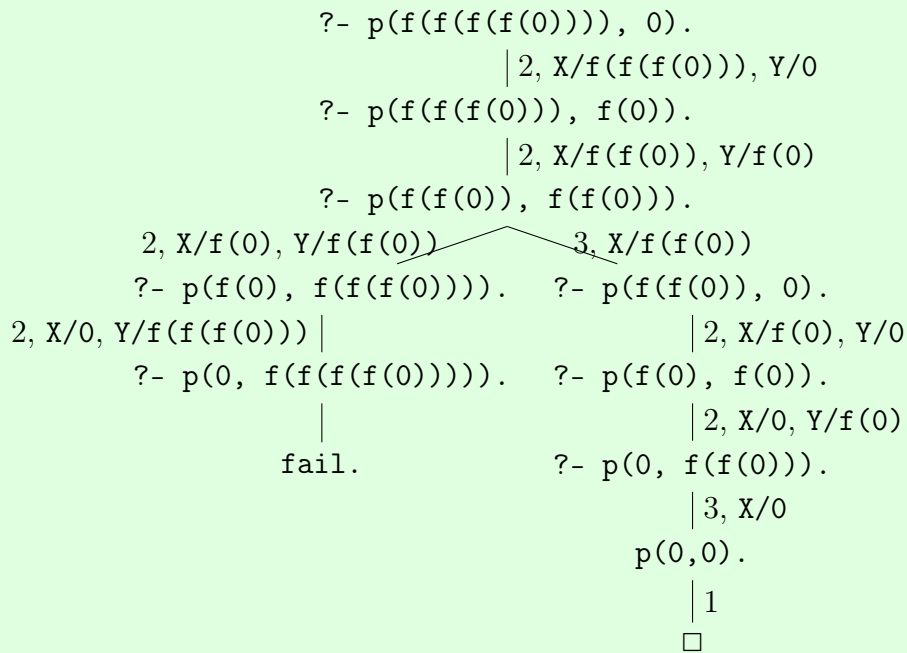
- c) Program vrací **true** pro dotazy tvaru `?- q(s(...s(0)...))`, kde počet `s` dává po dělení třemi zbytek 0 nebo 1.
- d) V tomto programu vždy jediným.

**Příklad 7.4.10.** Uvažujte následující program jazyka Prolog:

- 1:  $p(0,0)$ .
- 2:  $p(f(X), Y) :- p(X, f(Y))$ .
- 3:  $p(X, f(f(0))) :- p(X, 0)$ .

Vytvořte SLD strom pro dotaz  $?- p(f(f(f(f(0))))), 0)$ .

- a) Bude první větev výpočtu úspěšná? Pokud ne, přeuspořádejte řádky v programu tak, aby byla úspěšná první větev výpočtu. Vytvořte SLD strom tohoto upraveného programu pro stejný dotaz.
- b) Jak se přeuspořádání řádků projeví při vyhodnocování dotazu  $?- p(X, 0)$ ?



- a) První větev úspěšná nebude. Prohozením řádků 2 a 3 bude úspěšná hned první větev výpočtu.
- b) V původním programu Prolog vypíše  $X=0$  a po zadání středníku dále nad tímto dotazem cyklí. Po přeuspořádání vrací Prolog výsledky postupně  $X=0$ ,  $X=f(f(0))$ ,  $X=f(f(f(f(0))))$  atd. (Ověřte si SLD stromem.)

## 8 Neklasické logiky

*Neklasické logiky* jsou logiky, které porušují principy *extenzionality* a *dvouhodnotovosti* – principy, na nichž jsou založeny klasické logiky. Tyto principy klasické logiky různým způsobem omezují, například v oblasti reprezentace výrazů přirozeného jazyka. V této kapitole si představíme několik neklasických logik, které se mimo jiné snaží vypořádat právě s těmito omezeními.

### 8.1 Intenze a extenze

U výrazů můžeme rozlišovat mezi tím, co znamenají, a tím, co označují. Uvážíme-li například situaci, že prezident ČR je zároveň nejrychlejším běžcem na světě, pak oba výrazy *prezident ČR* a *nejrychlejší běžec na světě* označují stejnou osobu, ale zároveň mají různé významy. Množina objektů, které výraz označuje, se nazývá *extenze* výrazu. Naopak význam výrazu se nazývá *intenze* výrazu. Intenze je tvořená charakteristikami a vlastnostmi, které daný výraz představuje.

Klasické logiky jsou extenzionální, což znamená, že pracují pouze s extenzemi výrazů a tvrzení. Naproti tomu *intenzionální logiky* se zabírají také intenzemi. Mezi intenzionální logiky patří například transparentní intenzionální logika (TIL) nebo modální logiky.

**Příklad.** Pro výrazy určete jejich extenzi a intenzi:

- a) pes,
- b) Země je kulatá.

- a) Extenze slova pes je to, co výraz označuje. Tvoří ji všechny entity, které splňují podmínky bytí psem, tedy všichni psi.  
Intenze slova pes tvoří vlastnosti a charakteristiky které vyjadřují, co znamená být psem. Například, že pes má srst, je to zvíře, má ocas, čtyři nohy nebo umí štěkat.
- b) Extenze tvrzení je vždy jeho pravdivostní hodnota, v tomto případě je to pravda.  
Intenze tvrzení je jeho samotný obsah – informace, kterou nese.

Výraz může být v tvrzení použit způsobem, kdy záleží pouze na jeho extenzi, nebo způsobem, kdy záleží i na jeho intenzi. V prvním případě říkáme, že je výraz použitý v *extenzionálním kontextu*, ve druhém případě v *intenzionálním kontextu*.

Pokud výraz v tvrzení nelze nahradit libovolným výrazem se stejnou extenzí, aniž by se změnila pravdivostní hodnota (extenze) celého tvrzení, jedná se o intenzionální kontext. Intenzionální kontext můžeme rozpoznat také tehdy, když lze výraz ve větě interpretovat dvěma způsoby (de dicto/de re); jako označení libovolné nspecifikované instance, nebo jako označení konkrétní instance. Například větu *Dáša chce nejmenší kaktus na světě*. lze interpretovat následujícími způsoby:

- *Dáša chce nejmenší kaktus na světě, ať už je to jakákoli rostlina.*
- *Dáša chce konkrétně kaktus Blossfeldia liliputiana, a tento kaktus je také nejmenší kaktus na světě.*

V prvním případě představuje výraz *nejmenší kaktus na světě* blíže neurčený kaktus, který tuto podmínku splňuje, ve druhém případě je výraz použit jako jiné pojmenování pro kaktus Blossfeldia liliputiana.

Konečně se intenzionální kontext vyskytuje i ve větě, jejíž pravdivost nezávisí na tom, zda označovaný objekt vůbec existuje, tedy zda má extenzi. Taková je například věta *Dáša hledá jednorožce.*, která může být pravdivá bez ohledu na to, zda jednorožec existuje, či nikoli. Naopak věta *Dáša má jednorožce.* nemůže být pravdivá, pokud jednorožec neexistuje.

**Příklad.** Výrazy v závorkách substituuje za zvýrazněné výrazy, a rozhodněte, zda je pravdivostní hodnota celého výroku zachována. Poté rozhodněte, zda jsou zvýrazněné výrazy použity v intenzionálním nebo extenzionálním kontextu.

- První člověk na Měsíci** přednesl projev. (Neil Armstrong)
- Honza by chtěl být **prvním člověkem na Měsíci**. (Neil Armstrong)
- Země je kulatá** a Merkur je měsíc. (Slunce je hvězda.)
- Honza ví, že **Země je kulatá**. (Slunce je hvězda.)

- Neil Armstrong pronesl projev.  
Pravdivostní hodnota výroku je zachována. Ta by byla zachována i v případě, že bychom výraz *první člověk na Měsíci* nahradili jakýmkoli výrazem, který označuje osobu Neila Armstronga.  
Jedná se o extenzionální kontext.
- Honza se chce stát Neilem Armstrongem.  
Pravdivostní hodnota tvrzení nemusí být zachována. Honza by chtěl být prvním člověkem na Měsíci, ale to nutně neznamená, že by chtěl být Neilem Armstrongem.  
Jedná se o intenzionální kontext. Alternativně argumentujeme, že větu můžeme interpretovat de dicto i de re.
- Slunce je hvězda a Merkur je měsíc.  
Pravdivostní hodnota tvrzení je zachována. Celé tvrzení bude nepravdivé, i kdybychom vybranou část tvrzení nahradili libovolným jiným tvrzením, které je pravdivé.  
Jedná se o extenzionální kontext.
- Honza ví, že Slunce je hvězda.  
Pravdivostní hodnota tvrzení nemusí být zachována. Honza může vědět, že Země je kulatá, a nevědět, že Slunce je hvězda.  
Jedná se o intenzionální kontext.

**Příklad 8.1.1.** ★ Určete extenzi a intenzi následujících výrazů:

- a) prezident ČR,
- b) Batman,
- c) jednorožec,
- d) Slunce je hvězda.

- a) Extenze výrazu je osoba, která je v současné době prezidentem ČR, tedy Petr Pavel (r. 2023).  
Intenze výrazu je samotný význam výrazu, to, co znamená být prezidentem – prezidentská role.
- b) Extenze výrazu je osoba, která je Batmanem, tedy Bruce Wayne.  
Intenze výrazu může být superhrdina, který bojuje za spravedlnost.
- c) Tento výraz nemá ve skutečném světě extenzi; žádný konkrétní jednorožec neexistuje.  
Intenze vyjadřuje, že se jedná o bájného koně, který má uprostřed čela roh.
- d) Extenze tvrzení je pravda.  
Intenze je informace o tom, že Slunce je hvězda.

**Příklad 8.1.2.** Rozhodněte, zda jsou zvýrazněné výrazy použity v intenzionálním, nebo extenzionálním kontextu.

- a) Albert chce být **nejrychlejším běžcem**.
- b) **Nejrychlejší běžec** doběhl do cíle.
- c) **Sousedův pes** umí udělat salto.
- d) Alenka chce **nejchytřejšího psa na světě**.
- e) Honza hledá **studánku**.
- f) Bořek chce **létající kočku**.
- g) Autorem písně *Like a Rolling Stone* je **Bob Dylan**.
- h) Michal neví, že **autorem písně Like a Rolling Stone je Robert Allen Zimmerman**.

- a) Intenzionální kontext. Albert se nechce stát konkrétním člověkem, který je v současnosti nejrychlejším běžcem. Chce být tím, co představuje pojem nejrychlejší běžec. Alternativně můžeme argumentovat, že větu lze interpretovat jak *de dicto*, tak *de re*.
- b) Extenzionální kontext. Do cíle doběhl nějaký konkrétní člověk, na kterého odkazujeme výrazem nejrychlejší běžec. Kdybychom tento výraz nahradili jakýmkoli jiným výrazem, který odkazuje na danou osobu (jméno, popis), věta by stále platila.
- c) Extenzionální kontext. Výraz ve větě odkazuje na konkrétního psa.

- d) Intenzionální kontext. Větu lze interpretovat jak *de dicto*, tak *de re*. Alenka chce psa Bobíka, a ten je zároveň nejchytřejší pes na světě. Alenka chce vlastnit nejchytřejšího psa na světě, ať už je to jakýkoli pes.
- e) Intenzionální kontext. Honza má žízeň a hledá jakoukoli studánku. Honza hledá konkrétní studánku, pamatuje si, že tady někde byla.
- f) Intenzionální kontext. Větu lze interpretovat jak *de dicto*, tak *de re* a tvrzení může platit i když výraz létající kočka nemá žádnou extenzi.
- g) Extenzionální kontext. Po substituci libovolným výrazem, který označuje Boba Dylana, tvrzení platí.
- h) Intenzionální kontext. Michal nemusí vědět, že výrok je pravdivý, ale může například vědět, že platí, že autorem písně *Like a Rolling Stone* je Bob Dylan.

**Příklad 8.1.3.** ★ Pro zvýrazněné výrazy najděte výrazy se stejnou extenzí, které po substituci změni pravdivostní hodnotu celého tvrzení. Pokud takový výraz neexistuje, zdůvodněte.

- a) Dáša řekla, že **na Marsu žijí Martňané**.
- b) Pokud **na Marsu žijí Martňané**, pak tam jsou budovy.
- c) Policie pátrá po **vandalovi**.
- d) Policie zadržela hledaného **vandala**.
- e) Bořek ví, že  **$1 + 1 = 2$** .

- a) Dáša řekla, že **na Měsíci žijí lidé**. Pravdivostní hodnota nemusí být zachována. Dáša tuhle větu nemusela říct.
- b) Výraz neexistuje. Jedná se o extenzionální kontext. Pro libovolný výraz, který je nepravdivým tvrzením, se pravdivostní hodnota celého tvrzení nezmění.
- c) Policie pátrá po **známém zpěvákovi**. Pravdivostní hodnota nemusí být zachována, policie vůbec nemusí vědět, že vandalem je známý zpěvák a nepátrá konkrétně po něm.
- d) Výraz neexistuje. Jedná se o extenzionální kontext. Policie zatkla konkrétní osobu. Pokud bychom tento výraz nahradili libovolným výrazem odkazujícím na stejnou osobu, například se jménem osoby, pravdivostní hodnota celé věty by zůstala nezměněna.
- e) Bořek ví, že  **$0! = 1$** . Bořek možná ví, že  $1 + 1 = 2$ , ale nemusí ještě vědět, že  $0! = 1$ .

**Příklad 8.1.4.** Uvedená slova použijte v intenzionálním a extenzionálním kontextu.

- a) vítěz turnaje,
- b) tužka,
- c) Mozart.

Řešení by mohlo vypadat například takto:

- a) I: Honza se chce stát vítězem turnaje. E: Vítěz turnaje si podal ruku s poraženým soupeřem.
- b) I: Potřebuji tužku. E: Albert mi půjčil svou tužku.
- c) I: Michal obdivuje Mozarta. E: Mozart složil svou první skladbu.

## 8.2 Modální logiky a možné světy

Modální logiky jsou intenzionální logiky, které se zabývají různými *modalitami* neboli módy pravdy. Mezi tyto modalitty patří:

- *Aletická* modalita – vyjadřuje nutnost – nutně platí, možná platí.
- *Temporální* modalita – vyjadřuje platnost v určitém časovém okamžiku – vždy platilo, bude platit.
- *Deontická* modalita – vyjadřuje, jak by věci měly být – je povinné, je povoleno.
- *Epistemická* modalita – vyjadřuje znalost a přesvědčení – Albert ví, že platí, Albert věří, že platí.

Naproti tomu v klasické logice uvažujeme pouze o tom, zda je tvrzení pravdivé nebo nepravdivé.

Modalitty reprezentujeme pomocí *modálních operátorů*. Ty se od logických operátorů liší tím, že pravdivostní hodnota na jejich výstupu závisí nejen na pravdivostní hodnotě vstupních tvrzení (extenzi), ale také na významu samotných tvrzení (intenzi). Jelikož klasické logiky nepracují s intenzemi výrazů, modalitty správně reprezentovat nedokážou.

Abychom mohli pro nějaký stav světa vyhodnotit pravdivost tvrzení obsahujících modální operátory, musíme vzít v úvahu nejen to, co je a co v dané situaci neplatí, ale také například to, co by mohlo platit, co platilo nebo co si někdo myslí, že platí. Musíme tedy brát v úvahu více než jeden stav světa.

Koncept *možných světů* nám umožňuje modelovat všechny možné stavy světa a vztahy mezi nimi najednou. Takovému modelu říkáme Kripkeho rámec a formálně ho zavedeme v podkapitole 8.3. Každý *svět* v něm představuje jeden možný stav věcí, množina světů představuje všechny možné stavy a *aktuální svět* označujeme svět, který odpovídá skutečnému stavu věcí.

Vztahy mezi světy reprezentujeme pomocí *relace dostupnosti*.

**Definice 44:** *Relace dostupnosti* je binární relace  $S \subseteq W^2$  na množině světů  $W$ . Říkáme, že svět  $w$  je dostupný ze světa  $v$  právě tehdy, když platí  $(v, w) \in S$ .

Obecně platí, že relace dostupnosti představuje určitou formu možnosti pro daný svět. Za možné světy ve světě  $w$  považujeme pouze ty světy, které jsou z  $w$  dostupné.

**Příklad.** Uvažte svět, který obsahuje pouze výroky  $p, q, r$ , a v němž platí  $p \Rightarrow q$ .

- Kolik existuje v popsané situaci možných světů?
- Přidejme podmínku, že ve světě platí také  $p \vee \neg r$ . Kolik světů je možných v takovém případě?

- 6 možných světů. 2 světy nejsou možné, a to světy, kde by platilo  $p, \neg q, r$  a  $p, \neg q, \neg r$ , t.j. z 8 kombinací pravdivostních hodnot výroků může ve světě nastat jenom 6.
- 4 možné světy. Z další podmínky vyplývá, že dva světy, v nichž platí  $\neg p$  a  $r$ , také nejsou možné.

---

**Příklad 8.2.1.** Určete modalitu věty.

- Včera přšelo.
- Hana je přesvědčena, že káva není zdravá.
- Možná má můj pes chřipku.
- Kočky nevědí, že Merkur je planeta.
- Na tomto místě se nesmí skákat do vody.

- Temporální modalita.
- Epistemická modalita.
- Aletická modalita.
- Epistemická modalita.
- Deontická modalita.

**Příklad 8.2.2.** Dáša má dvě rostliny, fialku a zelenec, které zalévá vždy současně. Zelenec uvadne 4 až 6 dní po zalití. Fialka uvadne 10 až 15 dní po zalití. Uvažme zjednodušený svět, který obsahuje pouze tyto dvě rostliny, a v němž je stav rostliny binární – buď je uvadlá, nebo není.

- Uveďte, kolik existuje možných světů.
- Dáša řekla, že rostliny naposledy zalévala před dvěma týdny. Uveďte, kolik možných světů existuje, pokud tuto informaci znáte, a které to jsou.



- a) Tři světy. Pozor, svět, v němž uvadá pouze fialka, není možný.
- b) Dva světy. Z uvedených informací vyplývá, že zelenec je určitě zvadlý a fialka možná. Existují tedy dva možné světy, jeden, ve kterém je uvadlý pouze zelenec, a druhý, ve kterém jsou uvadlé obě rostliny.

**Příklad 8.2.3. ★** V bufetu se prodává káva, bagety a croissanty. Ne vždy ale mají kompletní nabídku. Mějme svět, který reprezentuje stav bufetové nabídky.

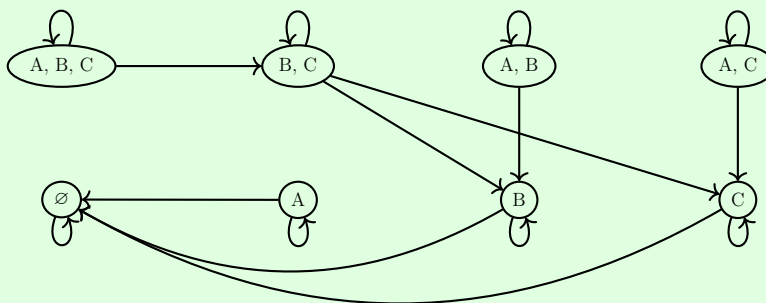
- a) Uveďte, kolik existuje možných světů.
- b) Zjistíte, že v bufetu mají vždy croissanty nebo bagety, ale nemají je současně. Uveďte, kolik možných světů existuje, pokud tuto informaci znáte, a které to jsou.

- a) 8 světů. Pozor, počítáme i svět, kde nemají nic.
- b) 4 světy. Stav světa jsou následující: Mají pouze bagety. Mají pouze croissanty. Mají bagety a kávu. Mají croissanty a kávu.

**Příklad 8.2.4. ★** Do bufetu z předchozího zadání (neberte v úvahu omezení zavedené v bodě b)) přijde student. Chce si koupit kávu, nebo, pokud ji nemají, cokoli jiného. Jak se změní stav bufetu po jeho odchodu? Předpokládejte, že student si koupí pouze jednu věc (nebo, pokud je již všechno vyprodáno, si nekoupí nic) a káva se může vyprodat také.

Možné stavy bufetu reprezentujte pomocí možných světů. Změnu stavu bufetu reprezentujte pomocí relace dostupnosti. Možné světy a relaci graficky znázorněte.

A – mají kávu, B – mají bagety, C – mají croissanty



Stav bufetu po návštěvě studenta buď zůstane stejný (reflexivní relace), nebo si student koupí poslední položku a bufet se přesune z původního stavu do nového stavu, ve kterém tuto položku nemají. Pokud je bufet ve stavu, ve kterém mají kávu, vyprodat se může pouze

káva, v opačném případě se může vyprodat jakákoli položka. Pokud v bufetu již nic nemají, stav bufetu zůstává stejný.

### 8.3 K-logika

*K-logika* je formální systém pracující s modalitou nutnosti, kterou reprezentuje pomocí modálních operátorů *nutně*  $\Box$  a *možná*  $\Diamond$ . K-logika je navíc základní modální logika – její úpravou můžeme zavést logiky dalších modalit.

V této kapitole zavádíme K-logiku s využitím výrokové logiky. Nejprve rozšíříme abecedu výrokové logiky o symboly modálních operátorů:

**Definice 45:** *Abeceda K-logiky* zahrnuje:

- abecedu výrokové logiky,
- symboly  $\Box, \Diamond$ .

Syntax výrokové logiky rozšíříme o pravidlo pro formule s modálními operátory:

**Definice 46:** Je-li  $\varphi$  formule, pak také  $(\Box\varphi)$  a  $(\Diamond\varphi)$  jsou formule.

Dále zavedeme *sémantiku K-logiky*. Interpretace K-logiky, nazývaná *Kripkeho interpretace* neboli *Kripkeho rámec*, nepředstavuje jediný stav světa (jako interpretace v klasických logikách), ale více možných stavů. Každý možný stav je reprezentován jako jeden svět a jemu odpovídající interpretace elementárních výroků.

**Definice 47:** *Kripkeho interpretace*  $C$  (též *Kripkeho rámec*) pro jazyk  $L$  zahrnuje:

- neprázdnou množinu světů  $W$ ,
- relaci dostupnosti  $S \subseteq W^2$ ,
- interpretaci výrokových proměnných  $I_w$  pro každý svět  $w \in W$ .

Relace dostupnosti pro každý možný svět  $w$  určuje, které světy jsou z něj dostupné. Dostupné světy ze světa  $w$  chápeme jako světy, které jsou ve světě  $w$  možné. Interpretace  $I_w$  pro každý možný svět  $w$  určuje, které elementární výroky v něm jsou nebo nejsou pravdivé.

Jelikož různé světy vyhodnocují elementární výroky různě, i složitější formule může v různých světech nabývat různých pravdivostních hodnot.

**Definice 48:** Zda formule  $\varphi$  platí ve světě  $w \in W$ , píšeme  $w \models \varphi$ , definujeme následovně:

- $w \models p$ , právě když atomická formule  $p$  je pravdivá v  $I_w$ ,
- $w \models (\varphi \Rightarrow \psi)$ , právě když  $w \models \varphi$  implikuje  $w \models \psi$  (podobně pro další log. spojky),
- $w \models \Box\varphi$ , právě když  $v \models \varphi$  pro všechny světy  $v \in W$  takové, že  $(w, v) \in S$ ,
- $w \models \Diamond\varphi$ , právě když existuje svět  $v \in W$  takový, že  $(w, v) \in S$  a  $v \models \varphi$ .

Je-li formule atomická, platí ve světě  $w$  právě tehdy, když je pravdivá v  $I_w$ . V případě logických spojek zůstává sémantika stejná jako ve výrokové logice.

Formule  $\varphi$  je *nutně pravdivá* ve světě  $w$ , píšeme  $\Box\varphi$ , právě když je pravdivá ve všech světech dostupných z  $w$ . Formule  $\varphi$  *možná platí* ve světě  $w$ , píšeme  $\Diamond\varphi$ , právě když platí alespoň v jednom světě dostupném ze světa  $w$ .

Všimněte si, že sémantika modálních operátorů  $\Box$  a  $\Diamond$  je podobná sémantice kvantifikátorů  $\forall$  a  $\exists$ . Stejně jako u nich platí i u modálních operátorů, že jeden lze vyjádřit pomocí druhého, tedy:  $\Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi$ .

Nakonec zbývá definovat pravdivost v interpretaci a logickou pravdivost.

**Definice 49:**

- Formule  $\varphi$  je *pravdivá v Kripkeho interpretaci  $C$* , píšeme  $\models_C \varphi$ , jestliže platí ve všech světech  $w \in W$ .
- Formule  $\varphi$  je *logicky pravdivá (tautologie)*, platí-li ve všech Kripkeho interpretacích.

Vyhodnocení K-logické formule v Kripkeho interpretaci si ukážeme na příkladu.

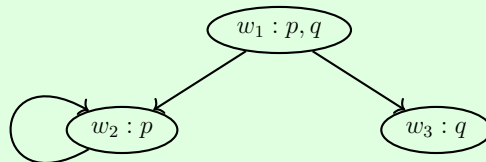
**Příklad.** Uvažte Kripkeho rámeček  $C$ , kde  $W = \{w_1, w_2, w_3\}$ ,  $S = \{(w_1, w_2), (w_1, w_3), (w_2, w_2)\}$  a elementární výroky jsou  $p, q$ . Ve světech platí následující tvrzení:

- $w_1 : p, q$ ,
- $w_2 : p$ ,
- $w_3 : q$ .

Pomocí grafu znázorněte možné světy a relaci dostupnosti. Potom rozhodněte, zda je formule  $\Diamond\Box(\neg q \vee p)$ :

- a) platná ve světě  $w_1$ ,
- b) platná v Kripkeho interpretaci  $C$ ,
- c) logicky pravdivá.

Grafické znázornění možných světů a vztahů mezi nimi vypadá následovně:



Nejprve vytvoříme pravdivostní tabulku, ve které vyhodnotíme formuli pro všechny světy. Ta je užitečná i v případě, že máme vyhodnotit platnost formule pouze v jednom světě. Chceme-li například rozhodnout, zda  $\Box\Diamond\varphi$  platí v nějakém světě, musíme nejdříve vyhodnotit formuli  $\Diamond\varphi$  pro všechny světy, které jsou z tohoto světa dostupné.

	$w_1$	$w_2$	$w_3$
$p$	1	1	0
$q$	1	0	1
$\neg q$	0	1	0
$\neg q \vee p$	1	1	0
$\Box(\neg q \vee p)$	0	1	1
$\Diamond\Box(\neg q \vee p)$	1	1	0

Při vyhodnocování formule  $\Box(\neg q \vee p)$  v předposledním řádku tabulky postupujeme následovně. Ve světě  $w_1$  platí daná formule právě tehdy, když  $\neg q \vee p$  platí ve všech dostupných světech z  $w_1$ , tedy ve  $w_2$  a  $w_3$ . V předchozím řádku tabulky vidíme, že  $\neg q \vee p$  platí ve světě  $w_2$ , ale neplatí ve světě  $w_3$ . Z toho plyne, že formule  $\Box(\neg q \vee p)$  ve světě  $w_1$  neplatí. Ve světě  $w_2$  daná formule platí, protože  $\neg q \vee p$  platí ve světě  $w_2$ . Ve světě  $w_3$  formule platí, protože platí tvrzení, že  $\neg q \vee p$  platí ve všech světech, které jsou z něj dostupné (žádný svět není dostupný z  $w_3$ ).

Formuli  $\Diamond\Box(\neg q \vee p)$  v posledním řádku vyhodnotíme obdobně, avšak nyní s rozdílem, že  $\Box(\neg q \vee p)$  musí platit alespoň pro jeden dostupný svět. To platí pro světy  $w_1$  a  $w_2$ , ale ne pro svět  $w_3$  (žádný takový dostupný svět neexistuje).

Pomocí tabulky pak odpovíme na otázky ze zadání:

- Ano, formule  $\Diamond\Box(\neg q \vee p)$  platí ve světě  $w_1$ .
- Ne, formule neplatí v interpretaci  $C$ , protože neplatí ve světě  $w_3$ .
- Ne, formule není logicky pravdivá, protože neplatí v interpretaci  $C$ .

Formalismus K-logiky lze použít jako základ pro zavedení dalších modálních logik. Světům a relacím dostupnosti pak přiřazujeme konkrétnější význam v závislosti na modální logice, v níž jsou použity. Například v temporální logice představují světy nějaké okamžiky v čase a dostupnost mezi světy představuje změnu stavu světa v čase, v epistemické logice se světy

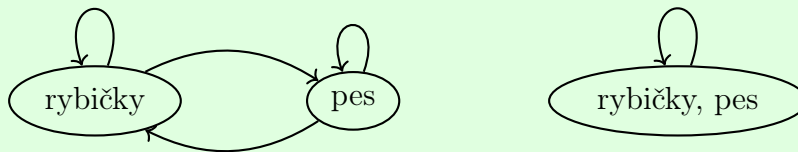
využívají k modelování znalostí jednotlivých agentů.

S pomocí epistemické logiky si ukážeme jedno konkrétnější využití modálních logik. Světy v epistemické logice představují různé, potenciálně reálné stavy světa, relace dostupnosti představují znalosti agentů o daném světě, přičemž pro každého agenta je zavedena vlastní relace. Dostupnost ze světa  $v$  do světa  $w$  prostřednictvím relace agenta  $a$  vyjadřuje, že pokud se agent  $a$  nachází ve světě  $v$ , nemá o tomto světě dostatečné znalosti, aby byl schopen rozlišit, zda se nachází ve světě  $v$  nebo ve světě  $w$ . Modelování konkrétní situace si ukážeme na příkladě.

**Příklad.** Uvažte svět, kde je v jedné místnosti Albert a ve vedlejší místnosti jsou rybičky nebo pes (nebo obojí) přičemž pes štěká pouze v případě, že jsou v místnosti i rybičky. Pomocí Kripkeho rámce zaznačte možné světy a znalosti Alberta v každém světě.

V dané situaci mohou nastat tři různé stavy, tudíž existují tři možné světy. Pokud Albert slyší štěkot, ví, že je ve světě, kde jsou jak rybičky, tak pes, takže ví přesně, ve kterém světě se nachází, a tento svět je nerozlišitelný pouze sám od sebe (značeno pomocí reflexivní relace). Pokud Albert štěkot neslyší, ví, že v místnosti je pouze pes nebo pouze rybičky, ale neví, ve kterém z těchto dvou světů se nachází. Tyto dva světy jsou pro Alberta nerozlišitelné.

Grafická reprezentace Kripkeho rámce je následovná:



**Příklad 8.3.1.** Pomocí K-logiky formálně запиšte následující věty:

- Možná prší nebo možná svítí slunce.
- Možná zároveň prší a svítí slunce.
- Pokud je duha, pak nutně prší a svítí slunce zároveň.

- $\Diamond P \vee \Diamond S$
- $\Diamond(P \wedge S)$
- $D \Rightarrow \Box(P \wedge S)$

**Příklad 8.3.2.** ★ Uvažte Kripkeho rámec  $C$ , kde  $W = \{w_1, w_2, w_3\}$ ,  $S = \{(w_1, w_1)$ ,

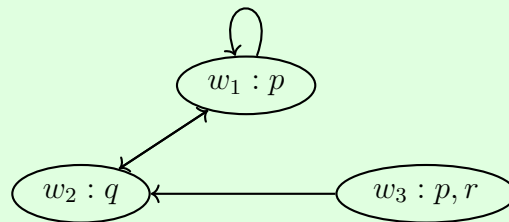
$(w_1, w_2), (w_2, w_1), (w_3, w_2)$ , elementární výroky jsou  $p, q, r$  a ve světech platí následující z nich:

- $w_1 : p$ ,
- $w_2 : q$ ,
- $w_3 : p, r$ .

Bez použití pravdivostní tabulky rozhodněte, zda v dané interpretaci platí následující formule:

- a)  $\Box p$
- b)  $\Diamond \neg q$
- c)  $\neg \Diamond r$
- d)  $p \vee \Diamond p$
- e)  $\Box r \Rightarrow \Diamond r$

Užitečné je znázornit interpretaci formou grafu:



- a) Ne. Neplatí ve světech  $w_1, w_3$ .
- b) Ne. Neplatí ve světě  $w_3$ .
- c) Ano.
- d) Ano.
- e) Ano.

**Příklad 8.3.3.** ★ Uvažte Kripkeho rámec  $C$ , kde  $W = \{w_1, w_2, w_3\}$ ,  $S = \{(w_1, w_3), (w_2, w_2), (w_2, w_1), (w_2, w_3)\}$  a elementární výroky jsou  $p, q$ . Ve světech platí následující tvrzení:

- $w_1 : q, \neg p$ ,
- $w_2 : \neg p, \neg q$ ,
- $w_3 : q, p$ .

Rozhodněte, zda je formule  $\Box(\neg \Diamond q \vee p) \wedge (p \Rightarrow \Box p)$ :

- a) platná ve světě  $w_1$ ,
- b) platná v Kripkeho interpretaci  $C$ ,
- c) logicky pravdivá.

	$w_1$	$w_2$	$w_3$
$p$	0	0	1
$q$	1	0	1
$\Box p$	1	0	1
$\Diamond q$	1	1	0
$\neg \Diamond q$	0	0	1
$\neg \Diamond q \vee p$	0	0	1
$p \Rightarrow \Box p$	1	1	1
$\Box(\neg \Diamond q \vee p)$	1	0	1
$\Box(\neg \Diamond q \vee p) \wedge (p \Rightarrow \Box p)$	1	0	1

- a) Ano.  
b) Ne, formule neplatí v interpretaci  $C$ , protože neplatí ve světě  $w_2$ .  
c) Ne, formule není logicky pravdivá, protože neplatí v interpretaci  $C$ .

**Příklad 8.3.4.** Uvažte Kripkeho rámeček  $C$ , kde  $W = \{w_1, w_2, w_3, w_4\}$ ,  $S = \{(w_2, w_2), (w_3, w_4), (w_4, w_3), (w_4, w_1)\}$ , a elementární výrok  $p$ , který platí ve světech  $w_2$  a  $w_3$ . Pomocí pravdivostní tabulky určete, ve kterých světech platí následující formule:

- a)  $\Box \neg \Box p$   
b)  $\Box \Diamond p$   
c)  $\Diamond \Diamond \Diamond p$

a)  $w_1, w_3$

	$w_1$	$w_2$	$w_3$	$w_4$
$p$	0	1	1	0
$\Box p$	1	1	0	0
$\neg \Box p$	0	0	1	1
$\Box \neg \Box p$	1	0	1	0

b)  $w_1, w_2, w_3$

	$w_1$	$w_2$	$w_3$	$w_4$
$p$	0	1	1	0
$\Diamond p$	0	1	0	1
$\Box \Diamond p$	1	1	1	0

c)  $w_2, w_4$

	$w_1$	$w_2$	$w_3$	$w_4$
$p$	0	1	1	0
$\Diamond p$	0	1	0	1
$\Diamond \Diamond p$	0	1	1	0
$\Diamond \Diamond \Diamond p$	0	1	0	1

**Příklad 8.3.5.** Znegujte následující formule a upravte je tak, aby se symboly negace nacházely pouze přímo před elementárními výroky:

- a)  $\diamond((\neg p \Rightarrow \diamond q) \vee \Box \neg q)$
- b)  $(\Box \neg p \wedge q) \vee \diamond \neg p$
- c)  $\Box \diamond(p \Rightarrow \diamond \Box q)$

- a)  $\neg(\diamond((\neg p \Rightarrow \diamond q) \vee \Box \neg q))$   
 $\approx \Box \neg((\neg p \Rightarrow \diamond q) \vee \Box \neg q)$   
 $\approx \Box(\neg(\neg p \Rightarrow \diamond q) \wedge \neg \Box \neg q)$   
 $\approx \Box((\neg p \wedge \neg \diamond q) \wedge \diamond q)$   
 $\approx \Box(\neg p \wedge \Box \neg q \wedge \diamond q)$
- b)  $(\diamond p \vee \neg q) \wedge \Box p$
- c)  $\diamond \Box(p \wedge \Box \diamond \neg q)$

**Příklad 8.3.6.** ★ Pro každou následujících formulí rozhodněte, zda je tautologie. Pokud ne, určete nějakou podmínku relace dostupnosti tak, aby formule platila ve všech interpretacích, které podmínku splňují.

- a)  $\Box p \Rightarrow p$
- b)  $\Box(p \Rightarrow q) \Rightarrow (\Box p \Rightarrow \Box q)$
- c)  $\Box p \Rightarrow \diamond p$
- d)  $p \Rightarrow \Box \diamond p$
- e)  $\Box p \wedge \Box \neg p$

- a) Ne. Platí, pokud je relace interpretace reflexivní, tedy pokud pro každý svět  $w$  platí  $(w, w) \in S$ .
- b) Ano.
- c) Ne. Platí, pokud je každý svět v relaci s nějakým světem, tedy pokud pro každý svět  $w$  existuje svět  $u$  takový, že  $(w, u) \in S$ .
- d) Ne. Platí, pokud je relace interpretace symetrická, tedy pokud pro všechny světy  $w, u$  platí, že  $(w, u) \in S$  implikuje  $(u, w) \in S$ .
- e) Ne. Platí pouze pro prázdnou relaci dostupnosti.



**Příklad 8.3.7.** Uvažte Kripkeho rámeček  $C$ , kde  $W = \{w_1, w_2, w_3\}$ , elementární výroky jsou  $p, q, r$  a ve světech platí následující z nich:

- $w_1 : p$ ,
- $w_2 : q$ ,
- $w_3 : p, r$ .

Bez použití reflexivních vztahů doplňte relaci dostupnosti Kripkeho rámeček  $C$  tak, aby v ní platili následující formule:

- a)  $\Box p \Rightarrow q$
- b)  $\Box r \wedge \neg \Diamond p$
- c)  $\Diamond p \wedge \Diamond \neg \Box p$

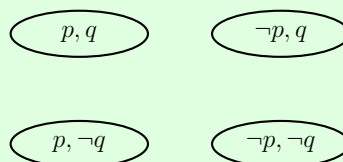
Řešení může vypadat například následovně:

- a)  $\{(w_1, w_2), (w_3, w_2)\}$
- b)  $\emptyset$
- c)  $\{(w_1, w_2), (w_1, w_3), (w_2, w_1), (w_3, w_1), (w_3, w_2)\}$

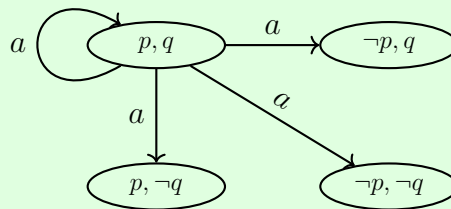
**Příklad 8.3.8.** ★ Na stole jsou dvě neoznačené konvice s kávou, z nichž jedna je bez kofeinu. Alice a Bob si nalili kávu z jedné z nich, aniž by viděli, ze které konvice si nalil kávu ten druhý. Za elementární výroky budeme považovat následující:  $p$  – Alice má kofeinovou kávu.  $q$  – Bob má kofeinovou kávu.

- a) Reprezentujte možné světy a znalosti agentů pomocí Kripkeho rámeček.
- b) Cyril, který kávu dělal, přijde do kuchyně a řekne Bobovi a Alici, ve které konvici byla bezkofeinová káva. Změňte Kripkeho rámeček tak, aby reprezentoval novou situaci.
- c) Ví Alice, ve kterém světě se nachází?
- d) Víte vy, který svět je aktuální?
- e) Alice sdělí Bobovi, jakou má kávu. Změňte Kripkeho rámeček tak, aby reprezentoval novou situaci.
- f) Ví teď Bob, ve kterém světě se nachází?
- g) Víte vy, který svět je aktuální?

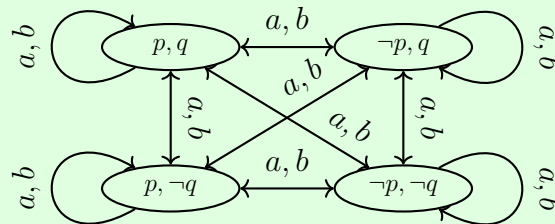
- a) Svět popsany v zadání se nachází v jednom ze 4 různých stavů, existují tedy 4 možné světy. Kripkeho rámeček budeme znázorňovat pomocí grafu. Nejprve přidáme všechny možné světy jako uzly:



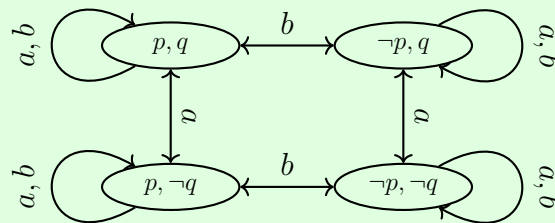
Následně pomocí relace dostupnosti  $S$  zaznamenáme znalosti agentů v každém světě. Uvažujeme takto: Pokud se Alice skutečně nachází ve světě  $p, q$ , které světy by podle ní mohly být skutečné? Tyto světy spojíme se světem  $p, q$  pomocí relace  $S_a$ . V tomto případě Alice neví, jaký druh kávy má ona, ani jaký druh kávy má Bob, takže považuje všechny 4 světy za možné.



Stejným způsobem vyznačíme znalosti Boba a Alice v ostatních světech.



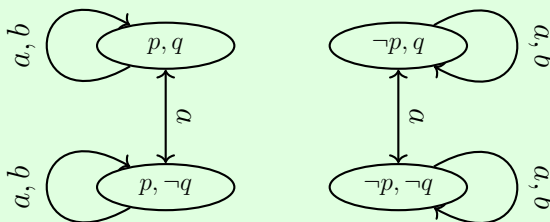
- b) V této situaci agenti vědí, jakou kávu mají oni, ale nevědí, jakou kávu má druhá osoba. To znamená, že pokud je například skutečný svět  $p, q$ , pak Alice považuje svět  $p, \neg q$ , kde má Bob kávu bez kofeinu, za možný, ale světy, kde má ona kávu bez kofeinu, jsou pro ni nemožné. Zrušíme tedy hranu  $a$  ze světa  $p, q$  do těchto světů. Podobně postupujeme i v případě ostatních světů a Bobových znalostí. Model upravíme tak, aby reprezentoval tyto nové znalosti agentů:



c) Ne.

d) Ne.

e) Upravený rámec vypadá následujícím způsobem:



f) Ano. V každém možném stavu světa Bob ví, že se nachází právě v něm.

g) Ne.

## 8.4 Vícehodnotové logiky

*Vícehodnotové logiky* jsou podobné klasickým logikám v tom, že pravdivostní hodnota celého výrazu se skládá z pravdivostních hodnot dílčích výrazů, jsou tedy extenzionální. Nicméně se od nich liší tím, že pracují s více než dvěma pravdivostními hodnotami. Pravdivostní hodnoty se většinou pohybují v intervalu od 0 do 1, přičemž 0 obecně představuje absolutní nepravdu a 1 absolutní pravdu, avšak konkrétní interpretace závisí na oblasti použití logiky.

Mezi vícehodnotové logiky patří tříhodnotová *Łukasiewiczova logika*, která kromě pravdivostních hodnot 0 (nepravda) a 1 (pravda) zavádí hodnotu  $1/2$  (nevím). Následující definice je obdobou definice interpretace z klasické výrokové logiky.

**Definice 50:** *Interpretace  $I$  v Łukasiewiczově logice* je zobrazení  $I : \mathcal{P} \rightarrow \{0, 1/2, 1\}$  přiřazující pravdivostní hodnoty 0 (nepravda), 1 (pravda) a  $1/2$  (nevím) jednotlivým výrokovým proměnným množiny  $\mathcal{P}$ .

Kvůli této změně zavádí Łukasiewiczova logika také novou *valuační funkci*  $val$  (obdobu *val-uace*), která umí pracovat s hodnotou  $1/2$  a zachovává sémantiku logických spojek.

**Definice 51:** Valuační funkce  $val_I$  příslušící interpretaci  $I$  je definována indukcí ke struktuře formule:

- $val_I(p) = I(p)$ , pro atomickou formuli  $p \in \mathcal{P}$ ,
- $val_I(\neg\varphi) = 1 - val_I(\varphi)$ ,
- $val_I(\varphi \wedge \psi) = \min(val_I(\varphi), val_I(\psi))$ ,
- $val_I(\varphi \vee \psi) = \max(val_I(\varphi), val_I(\psi))$ ,
- $val_I(\varphi \Rightarrow \psi) = \min(1, 1 - val_I(\varphi) + val_I(\psi))$ .

V Łukasiewiczově logice nelze implikaci obecně vyjádřit pomocí negace a disjunkce, protože rovnost  $val(p \Rightarrow q) = val(\neg p \vee q)$  neplatí vždy. V případě  $val(p) = 1/2$ ,  $val(q) = 1/2$  je totiž  $val(p \Rightarrow q) = 1$ , ale  $val(\neg p \vee q) = 1/2$ .

V Łukasiewiczově logice je *modelem formule* každá interpretace, v níž je hodnota formule větší než 0.

**Definice 52:** Modelem formule  $\varphi$  v Łukasiewiczově logice je interpretace  $I$  taková, že pro jí příslušící valuační funkci platí  $val_I(\varphi) > 0$ .

**Příklad.** Uvažte formuli  $p \Rightarrow \neg(q \vee p)$  Łukasiewiczovy logiky. Rozhodněte, zda je interpretace  $I(p) = 1/2$ ,  $I(q) = 0$  modelem formule.

Ano, daná interpretace je modelem formule, protože  $val_I(p \Rightarrow \neg(q \vee p)) = 1$ . Formuli vyhodnotíme postupně podle definice funkce  $val_I$ . Mezivýpočty jsou uvedeny v tabulce.

$p$	$\Rightarrow$	$\neg$	$(q$	$\vee$	$p)$
1/2	1	1/2	0	1/2	1/2

Další vícehodnotovou logikou je *fuzzy logika*, která umožňuje vyjádřit míru pravdivosti nejednoznačných výroků, jako je například výrok Petr je mladý. Fuzzy logika nepracuje s diskrétními pravdivostními hodnotami, ale s intervalem hodnot  $[0, 1]$ .

**Definice 53:** Interpretace  $I$  ve fuzzy logice je zobrazení  $I : \mathcal{P} \rightarrow [0, 1]$  přiřazující pravdivostní hodnoty z intervalu  $[0, 1]$  jednotlivým výrokovým proměnným množiny  $\mathcal{P}$ .

Model formule a valuační funkce jsou definovány stejně jako v Łukasiewiczově logice s tím

rozdílem, že valuační funkce pracuje nad spojitou doménou.

**Definice 54:** Modelem formule  $\varphi$  ve fuzzy logice je interpretace  $I$  taková, že pro jí příslušící valuační funkci platí  $val_I(\varphi) > 0$ .

**Příklad.** Uvažte formuli  $(p \vee \neg q) \wedge \neg p$  fuzzy logiky. Rozhodněte, zda je interpretace  $I(p) = 0.2$ ,  $I(q) = 0.9$  modelem formule.

Ano, daná interpretace je modelem formule, protože  $val_I((p \vee \neg q) \wedge \neg p) = 0.2$ . Formuli vyhodnotíme postupně podle definice funkce  $val_I$ . Mezivýpočty jsou uvedeny v tabulce.

$(p$	$\vee$	$\neg$	$q)$	$\wedge$	$\neg$	$p$
0.2	0.2	0.1	0.9	0.2	0.8	0.2

**Příklad 8.4.1.** ★ Uvažte interpretaci Łukasiewiczovy logiky  $I(p) = 1/2$ ,  $I(q) = 0$ ,  $I(r) = 1/2$ . Rozhodněte, zda je modelem následujících formulí:

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

- Ano,  $val_I(\neg(p \wedge r)) = 1/2$ .
- Ano,  $val_I(p \Rightarrow (q \vee \neg r)) = 1$ .
- Ano,  $val_I(\neg(r \Rightarrow q)) = 1/2$ .
- Ne,  $val_I(r \vee \neg r) \wedge q = 0$ .

**Příklad 8.4.2.** Uvažte formuli  $(p \Rightarrow q) \wedge (p \vee \neg q)$  Łukasiewiczovy logiky. Vytvořte pravdivostní tabulku, která zahrnuje pouze interpretace obsahující hodnotu  $1/2$ , a rozhodněte, které z nich jsou modelem formule.

$I(p)$	$I(q)$	$(p$	$\Rightarrow$	$q)$	$\wedge$	$(p$	$\vee$	$\neg$	$q)$
0	1/2	0	1	1/2	1/2	0	1/2	1/2	1/2
1/2	0	1/2	1/2	0	1/2	1/2	1	1	0
1/2	1/2	1/2	1	1/2	1/2	1/2	1/2	1/2	1/2
1/2	1	1/2	1	1	1/2	1/2	1/2	0	1
1	1/2	1	1/2	1/2	1/2	1	1	1/2	1/2

Modelem formule jsou všechny dané interpretace.

**Příklad 8.4.3.** Uvažte interpretaci fuzzy logiky  $I(p) = 0.3$ ,  $I(q) = 0.6$ ,  $I(r) = 0.2$ . Rozhodněte, zda je modelem následujících formulí:

- a) ★  $(\neg p \wedge q) \vee r$
- b) ★  $q \Rightarrow \neg r$
- c)  $\neg r \Rightarrow p$
- d)  $(r \wedge \neg q) \vee \neg(q \wedge p)$

- a) Ano. Hodnota valuace formule je  $\max(\min(1 - 0.3, 0.6), 0.2) = 0.6$
- b) Ano. Hodnota valuace formule je 1.
- c) Ano. Hodnota valuace formule je 0.5.
- d) Ano. Hodnota valuace formule je 0.7.

**Příklad 8.4.4.** Uvažte formuli  $(p \Rightarrow q) \wedge (p \wedge \neg q)$  fuzzy logiky. Nalezněte interpretaci, ve které je valuace formule větší než 0.5.

Pro hledanou interpretaci platí  $\min(\min(1, 1 - I(p) + I(q)), \min(I(p), 1 - I(q))) > 0.5$ . Z toho plyne:

$$\begin{aligned}1 - I(p) + I(q) &> 0.5 \\ I(p) &> 0.5 \\ 1 - I(q) &> 0.5\end{aligned}$$

Po úpravě nerovnic dostaneme následující podmínky interpretace:

- $I(q) < 0.5$
- $0.5 < I(p) < 0.5 + I(q)$

To splňuje například interpretace  $I(p) = 0,6$  a  $I(q) = 0,4$ .

**Příklad 8.4.5.** ★ Nalezněte interpretaci fuzzy logiky, která není modelem ani jedné z následujících formulí:

- $p \wedge \neg q$
- $\neg p \Rightarrow r$
- $q \vee \neg s$

Řešením je například interpretace  $I(p) = 0, I(q) = 0, I(r) = 0, I(s) = 1$ .

**Příklad 8.4.6.** Pro každou z následujících formulí uveďte, které interpretace fuzzy logiky nejsou její modelem.

- a)  $\neg p$
- b)  $p \wedge \neg q$
- c)  $(p \vee q) \Rightarrow r$

- a)  $I(p) = 1$
- b)  $I(p) \in [0, 1], I(q) = 1$  nebo  $I(p) = 0, I(q) \in [0, 1]$
- c)  $I(p) \in [0, 1], I(q) = 1, I(r) = 0$  nebo  $I(p) = 1, I(q) \in [0, 1], I(r) = 0$

**Příklad 8.4.7.** Uvažte interpretaci  $I(p) = 0, I(q) = 0.5, I(r) = x$ . Nalezněte množinu všech hodnot  $x$  takových, že valuace formule  $\neg(q \wedge \neg r) \vee (r \Rightarrow p)$  fuzzy logiky se rovná hodnotě:

- a) 0
- b) 0.5
- c) 0.8
- d) 1

- a)  $\emptyset$
- b)  $\{0.5\}$
- c)  $\{0.2, 0.8\}$
- d)  $\{0, 1\}$

## 9 Reprezentace a vyvozování znalostí

K tomu, aby inteligentní systém správně fungoval, mu potřebujeme dodat dostatek informací o světě, ve kterém pracuje, a naučit ho, jak tyto informace využívat. Zatímco oblast *reprezentace znalostí* se zabývá tím, jak znalosti vyjádřit ve formě, jež je vhodná pro počítačové zpracování, oblast *vyvozování znalostí* se zabývá tím, jak s uloženými znalostmi pracovat a odvozovat z nich znalosti nové. V této kapitole si několik takových přístupů k reprezentaci a vyvozování znalostí představíme.

Důležitou součástí reprezentování znalostí je organizování *objektů* do *kategorií* neboli tříd. Ve fyzické realitě sice přicházíme do kontaktu hlavně s objekty, ale na úrovni myšlení a vyvozování pracujeme ve velkém s kategoriemi. Například pokud si někdo chce koupit stejnou klávesnici, jako má jeho kamarád, nechce ten samý objekt, ale model klávesnice. Stejně tak zařazení objektu do kategorie nám často umožní odvodit o něm nové znalosti. Například vidíme-li v sekci ovoce v obchodě zelený, velký a kulatý objekt, zařadíme jej do kategorie meloun, z čehož můžeme následně odvodit informaci, že vevnitř má červenou dužinu.

V predikátové logice bychom mohli kategorie reprezentovat jako predikát, ale při tomto přístupu nedokážeme správně reprezentovat vlastnosti třídy. Kategorii proto reprezentujeme jako konstantu. Příslušnost do kategorie, která je reprezentována konstantou ověříme použitím predikátu, například  $Member(m, meloun)$ , jenž se vyhodnotí jako pravdivý, pokud je  $m$  instancí kategorie meloun. Takové modelování abstrakce jako objektu se nazývá *reifikace* neboli zvěčnění.

Objekty a třídy organizujeme do *hierarchie tříd*; každý objekt je *instancí* nějaké třídy a třída může být *podtřídou* či *nadtřídou* jiné třídy. Kromě hierarchie tříd se budeme zbývat i tím, jak reprezentovat další znalosti o pojmech. Tyto znalosti nazýváme *fakta*, týkají-li se objektů a *pravidla*, týkají-li se tříd.

### 9.1 Ontologie

*Ontologie* je formální systém, který popisuje svět, případně jeho část, skrze pojmy, vztahy mezi pojmy a axiomy. S použitím ontologie můžeme potom vyjádřit konkrétní znalosti o daném světě. Ontologie můžeme zavést pomocí různých formalismů, v této sbírce budeme používat predikátovou logiku.

Ontologie dělíme na doménové a obecné (upper ontologies). *Obecné ontologie* se snaží popsat celý svět a poskytnout tak nástroj k reprezentaci znalostí z libovolné oblasti. Naproti tomu *doménové ontologie* jsou specifitější, například v zavedených pojmech, a uzpůsobeny k reprezentaci nějaké konkrétní domény.

Jednoduchou ontologii si ukážeme na příkladu.



**Příklad.** Aby student získal zápočet, potřebuje získat alespoň 10 bodů z domácích úkolů a 15 bodů z testu. Zaveďte ontologii, která popisuje uvedenou situaci.

Nejprve pomocí predikátové logiky zavedeme jazyk ontologie, tj. všechny pojmy, které jsou nezbytné k popisu situace. Můžeme používat konstanty, funkce a predikáty.

- $\text{Úkoly}(s, b)$ : predikát, pravdivý, pokud má student  $s$  z domácích úkolů  $b$  bodů.
- $\text{Test}(s, b)$ : predikát, pravdivý, pokud má student  $s$  z testu  $b$  bodů.
- $\text{Zápočet}(s)$ : predikát, pravdivý, pokud byl studentovi  $s$  udělen zápočet.
- $+$ ,  $\leq$ : funkce a predikát, oba se standardní sémantikou.
- 10, 15: konstanty, reprezentují minimální nutný počet bodů.

Dále zavedeme obecně platné axiomy:

- A1:  $\forall s \left( \text{Zápočet}(s) \iff \exists b(10 \leq b \wedge \text{Úkoly}(s, b)) \wedge \exists b(15 \leq b \wedge \text{Test}(s, b)) \right)$

V dalším příkladu si ukážeme, jak lze ontologii použít k reprezentaci konkrétní situace a odvození znalostí.

**Příklad.** Albert získal 12 bodů za domácí úkoly a 20 bodů z testu. Pomocí ontologie z předchozího příkladu reprezentujte situaci a odvoďte, zda Albert dostane zápočet.

Nejprve zavedeme tvrzení specifická pro danou situaci:

- $\text{Úkoly}(\text{Albert}, 12)$
- $\text{Test}(\text{Albert}, 20)$

Z axiomu A1 potom plyne, že platí predikát  $\text{Zápočet}(\text{Albert})$ , a tudíž Albert zápočet získal.

---

**Příklad 9.1.1.** ★ Uvažujme následující neúplný jazyk ontologie, který reprezentuje hru piškvorky:

- $\text{Player}(p)$ : predikát,  $p$  je hráč
- $\text{Mark}(m)$ : predikát,  $m$  je značka
- $\text{Square}(q)$ : predikát,  $q$  je pole
- $p_x, p_o$ : konstanty, hráči „křížek“, resp. „kolečko“
- $q_{11}, q_{12}, \dots, q_{33}$ : konstanty, pole herního plánu
- $s_0$ : konstanta, počáteční situace, situace reprezentuje stav hry
- $\text{WinningPosition}(q_1, q_2, q_3)$ : predikát, pravdivý pro pole, která tvoří výherní pozici
- $\text{Opponent}(p)$ : funkce, reprezentuje protihráče hráče  $p$
- $\text{TurnAt}(s)$ : funkce, reprezentuje hráče, který je na tahu v situaci  $s$

- $\text{Play}(p, q)$ : funkce, reprezentuje akci, že hráč  $p$  označí pole  $q$
- $\text{Poss}(a, s)$ : predikát, je pravdivý, pokud je možné vykonat akci  $a$  v situaci  $s$
- $\text{Result}(a, s)$ : funkce, reprezentuje situaci, jež nastane po provedení akce  $a$  v situaci  $s$

Jazyk doplňte o konstanty, funkce a predikáty, které budou reprezentovat:

- značky, které mohou být umísťovány na pole,
- informaci o tom, která značka patří zadanému hráči,
- informaci o tom, zda hráč v dané situaci vyhrává,
- informaci o tom, která značka je umístěna na zadaném poli v dané situaci.

- $\times, \circ, \sqcup$ : konstanty pro značky „křížek“, „kolečko“, „prázdne políčko“
- $\text{MarkOf}(p)$ : funkce, reprezentuje značku hráče  $p$
- $\text{Wins}(p, s)$ : predikát, pravdivý, pokud v situaci  $s$  vyhrál hráč  $p$
- $\text{MarkAt}(q, s)$ : funkce, reprezentuje značku na poli  $q$  v situaci  $s$

**Příklad 9.1.2.** ★ V ontologii vytvořené v předchozím příkladu chybí axiomy, které by určovaly sémantiku zavedených predikátů a funkcí. Doplňte ontologii o následující axiomy:

- Nikdo jiný kromě  $p_{\times}$  a  $p_{\circ}$  není hráčem.
- Oponent hráče  $p_{\times}$  je  $p_{\circ}$ .
- Značka hráče  $p_{\circ}$  je  $\circ$ .
- Vítězné pozice se skládají pouze z polí, která se nacházejí v jednom řádku, sloupci nebo diagonále (formuli nepište celou, repetitivní části můžete vynechat).
- Axiom určující, kdy platí  $\text{Win}(p, s)$ .
- Axiom určující, kdy platí  $\text{Poss}(a, s)$ .

- $\forall p(\text{Player}(p) \iff p = p_{\times} \vee p = p_{\circ})$
- $\text{Opponent}(p_{\times}) = p_{\circ}$
- $\text{MarkOf}(p_{\circ}) = \circ$
- $\forall q_1, q_2, q_3 (\text{WinningPosition}(q_1, q_2, q_3) \iff$   
 $(q_1 = q_{11} \wedge q_2 = q_{12} \wedge q_3 = q_{13}) \vee (q_1 = q_{21} \wedge q_2 = q_{22} \wedge q_3 = q_{23}) \vee \dots )$   
 Formule pokračuje výčtem všech kombinací  $q_1, q_2, q_3$ , které představují výherní kombinaci.
- $\forall p, s (\text{Wins}(p, s) \iff \exists q_1, q_2, q_3 (\text{WinningPosition}(q_1, q_2, q_3) \wedge$   
 $\text{MarkAt}(q_1, s) = \text{MarkAt}(q_2, s) = \text{MarkAt}(q_3, s) = \text{MarkOf}(p)) )$

f)  $\forall p, q, s (\text{Poss}(\text{Play}(p, q), s) \iff \text{TurnAt}(s) = p \wedge \text{MarkAt}(q, s) = \perp)$

**Příklad 9.1.3.** Využijte nástroj, který pro slovo přirozeného jazyka najde s ním související pojmy ontologie SUMO (pro více informací o nalezených pojmech je rozklikněte) a určete:

- pojmy související se slovem *coffee*,
- pojem *p*, který reprezentuje zimní roční období,
- s čím je pojem *p* ekvivalentní,
- pojem, jehož podtřídou je pojem *p*,
- výraz, který obsahuje pojem *p*, pojem *MediterraneanClimateZone* a pojem *rainySeasonInArea*. Určete význam tohoto výrazu.

- strom (*BotanicalTree*), zrno (*CoffeeBean*), nápoj (*Coffee*), barva (*ColorAttribute*)
- WinterSeason*
- WinterSeason* je ekvivalentní s časovým intervalem od prosince do února.
- WinterSeason* je podtřídou *SeasonOfYear*.
- ( $\implies$

(*climateTypeInArea* ?AREA *MediterraneanClimateZone*)

(*rainySeasonInArea* ?AREA *WinterSeason*))

V oblastech středomořského klimatického pásma je zimní roční období obdobím deštivého počasí.

**Příklad 9.1.4.** Určete význam následujících výrazů ontologie SUMO. Pokud neumíte odvodit význam nějakého pojmu, můžete ho vyhledat zde.

- ★

( $\implies$

(*attribute* ?AREA *MountainousTerrain*)

(*exists*(?MTN)

(*and*

(*instance* ?MTN *Mountain*)

(*located* ?MTN ?AREA))))

- ( $\implies$

(*and*

(*instance* ?SMILE *Smiling*)

(*agent* ?SMILE ?AGENT))

```

(holdsDuring
  (WhenFn ?SMILE )
  (attribute ?AGENT Happiness)))
c) (⇒
  (instance ?O OlympicGames)
  (or
    (exists(?W)
      (and
        (instance ?W WinterSeason)
        (temporalPart
          (WhenFn ?O) ?W)))
    (exists(?S)
      (and
        (instance ?S SummerSeason)
        (temporalPart
          (WhenFn ?O) ?S))))))

```

Řešení podává pro každý bod doslovný a volnější výklad.

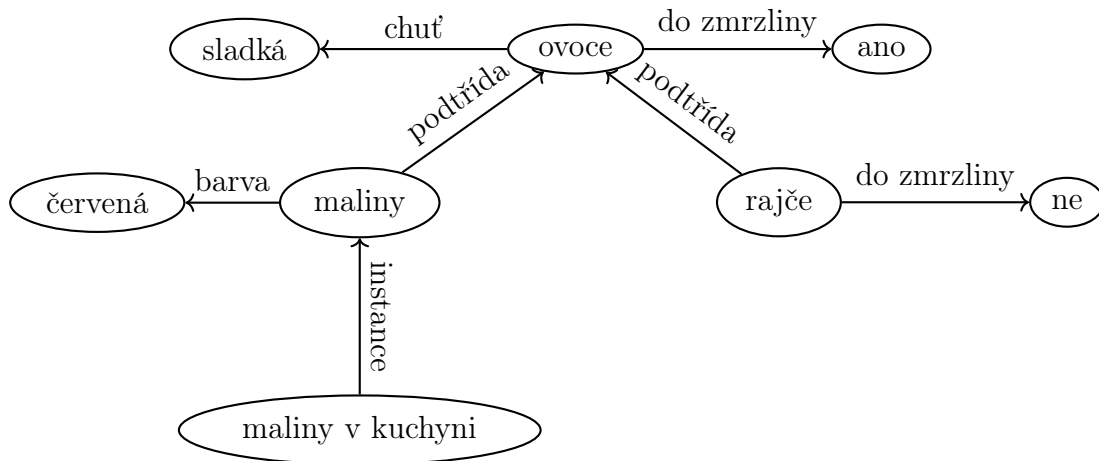
- a) Pokud má objekt AREA atribut *MountainousTerrain*, pak existuje fyzický objekt MTN, který je instancí *Mountain* a nachází se v objektu AREA. / V oblasti s hornatým terénem se nachází hora.
- b) Pokud je proces SMILE instancí *Smiling* a AGENT je agentem tohoto procesu, pak v době, kdy proces SMILE běží, je *Happiness* atributem agenta AGENT. / Agent, který se usmívá, je v tu chvíli šťastný.
- c) Pokud je proces O instancí *OlympicGames*, pak existuje zimní nebo letní období (časový interval W nebo S), jehož součástí je časový interval, během kterého běží proces O. / Olympijské hry se konají v zimě nebo v létě.

## 9.2 Sémantické sítě a rámce

Sémantické sítě reprezentují pojmy a vztahy mezi pojmy. Mají formu grafu, přičemž uzly grafu představují pojmy a hrany vztahy. Nejdůležitější vztahy jsou *instance* a *podtřída*, které tvoří hierarchii pojmů. Další vztahy reprezentují *vlastnosti* pojmů, například *Barva*(červená, beruška), nebo příslušnost části k celku, *Část*(noha, kráva). Vztahy, jež souvisí s hierarchií pojmů zapisujeme ve vertikálním směru, ostatní vztahy zapisujeme horizontálně.

Sémantické sítě podporují dědičnost a mechanismus *vzoru a výjimky*. Pojmy dědí vzorové hodnoty vlastností z nadtříd, přičemž vždy platí ta hodnota, která je k pojmu nejbližší. Vzorovou hodnotu může přepsat výjimka uvedená u uvažovaného pojmu. Princip dědičnosti neplatí nutně v horizontálních vztazích. Například noha, jako část krávy, zdědí barvu krávy, ale nedědí oblíbené jídlo krávy.

**Příklad.** Uvažte následující sémantickou síť.



- Je rajče sladké?
- Je rajče vhodné do zmrzliny?
- Jaké znalosti lze vyvodit ze sítě o pojmu *maliny v kuchyni*?

- Ano.
- Ne.
- Maliny v kuchyni jsou instancí kategorie maliny, jsou ovoce, mají červenou barvu, jsou sladké a vhodné do zmrzliny.

Rámec je univerzální struktura, v níž jsou uloženy veškeré relevantní informace o pojmech. Stejně jako sémantické sítě, rámce reprezentují taxonomii pojmů a jejich vlastností, avšak mají textovou podobu. Pojem je v rámci reprezentovaný pomocí následujících typů hodnot:

- *objekt* – samotný pojem
- *sloty* – vztahy daného pojmu
- *hodnoty slotů* - pojmy, které jsou v uvedeném vztahu s daným pojmem.

Rámce též podporují dědičnost a výjimky. Hvězdičkou jsou označeny sloty, které nesou vzorové hodnoty a mohou být při dědění přepsány výjimkou.

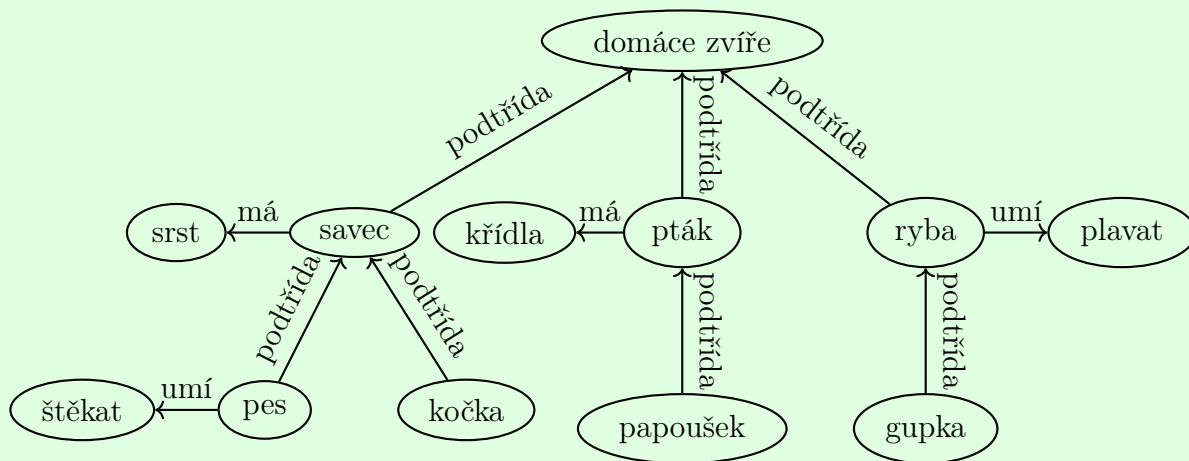
**Příklad.** Pomocí rámce reprezentujte znalosti ze sémantické sítě z předchozího příkladu.

objekty	sloty	hodnoty slotů
ovoce	*chuť:	sladká
	*do_zmrzliny:	ano
rajče	podtřída:	ovoce
	do_zmrzliny:	ne
maliny	podtřída:	ovoce
	*barva:	červená
maliny v kuchyni	instance:	maliny

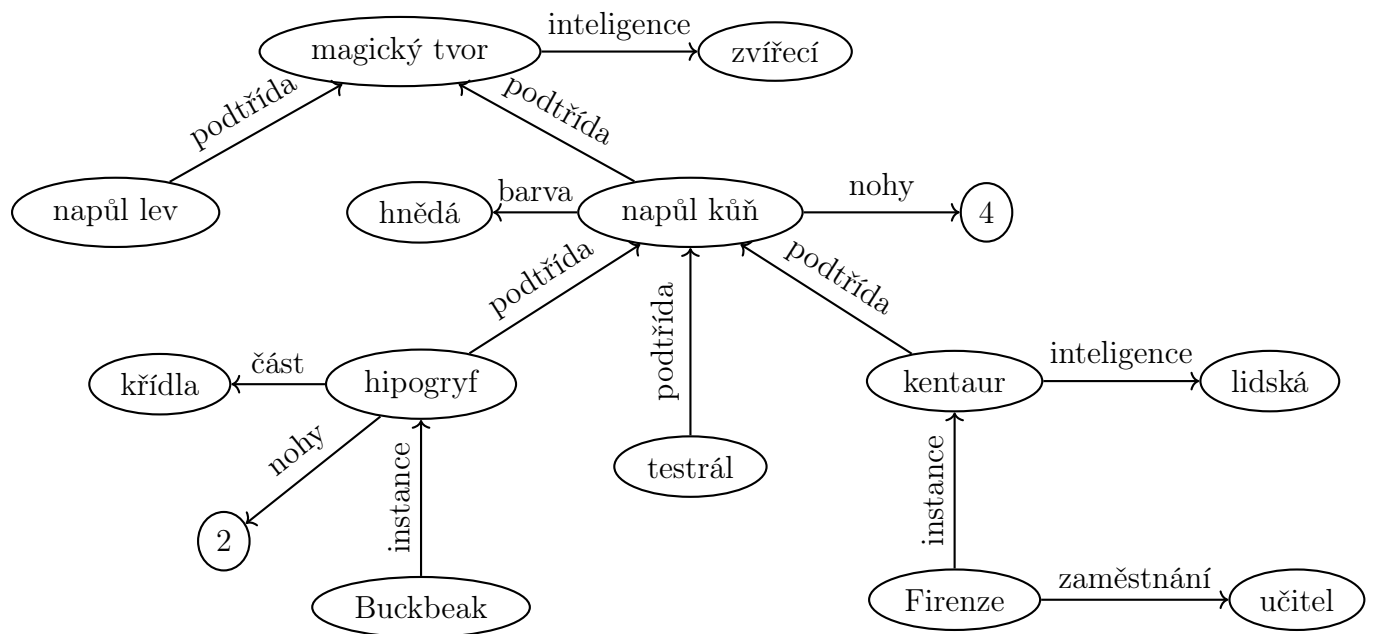
Objekty a hodnoty slotů jsou ekvivalentní s uzly grafu sémantické sítě, sloty jsou ekvivalentní s hranami grafu.

**Příklad 9.2.1.** Vytvořte sémantickou síť, která bude reprezentovat aspoň 4 libovolné kategorie domácích zvířat a jejich vlastnosti.

Například:



**Příklad 9.2.2.** ★ Uvažte následující sémantickou síť, která částečně reprezentuje magické tvory.



Odpovězte na následující otázky:

- Mají hipogryfové lidskou inteligenci?
- Má Firenze lidskou inteligenci?
- Kolik noh mají testrálové?
- Jakou barvu má hipogryf?

- Ne.
- Ano.
- 4
- Hnědou.

**Příklad 9.2.3.** Podle informací vyplývajících ze sémantické sítě z předchozího příkladu popište, co víte o následujících pojmech:

- ★ Buckbeak,
- Firenze,
- napůl lev.

- Buckbeak je hipogryf, napůl kůň, magický tvor, má hnědou barvu, dvě nohy, křídla a zvířecí inteligenci.

- b) Firenze je učitel, kentaur, napůl kůň a magický tvor s lidskou inteligencí. Má čtyři nohy a hnědou barvu.
- c) Napůl lev je magický tvor, který má zvířecí inteligenci.

**Příklad 9.2.4.** ★ Pomocí rámce reprezentujte znalosti o pojmech *napůl kůň* a *Firenze*, které jsou vyznačeny v sémantické síti v Příkladu 9.2.2.

**napůl kůň**

*podtřída:* magický tvor  
*\*barva:* hnědá  
*\*nohy:* 4

**Firenze**

*instance:* kentaur  
*zaměstnání:* učitel

**Příklad 9.2.5.** Uvažte následující rámec:



<b>akční film</b>	<i>podtřída:</i>	film
	<i>*exploze</i>	ano
<b>Drive</b>	<i>instance:</i>	akční film
	<i>exploze:</i>	ne
<b>Skyfall</b>	<i>instance:</i>	akční film
	<i>série:</i>	James Bond
<b>komedie</b>	<i>podtřída:</i>	film
	<i>zábavná:</i>	ano
<b>černá komedie</b>	<i>podtřída:</i>	komedie
	<i>*země původu:</i>	Velká Británie
	<i>humor:</i>	černý
<b>Horší než smrt</b>	<i>instance:</i>	černá komedie
<b>Big Lebowski</b>	<i>instance:</i>	černá komedie
	<i>země původu:</i>	USA
<b>romantická komedie:</b>	<i>podtřída:</i>	komedie
	<i>*konec:</i>	šťastný
<b>Annie Hall</b>	<i>instance:</i>	romantická komedie

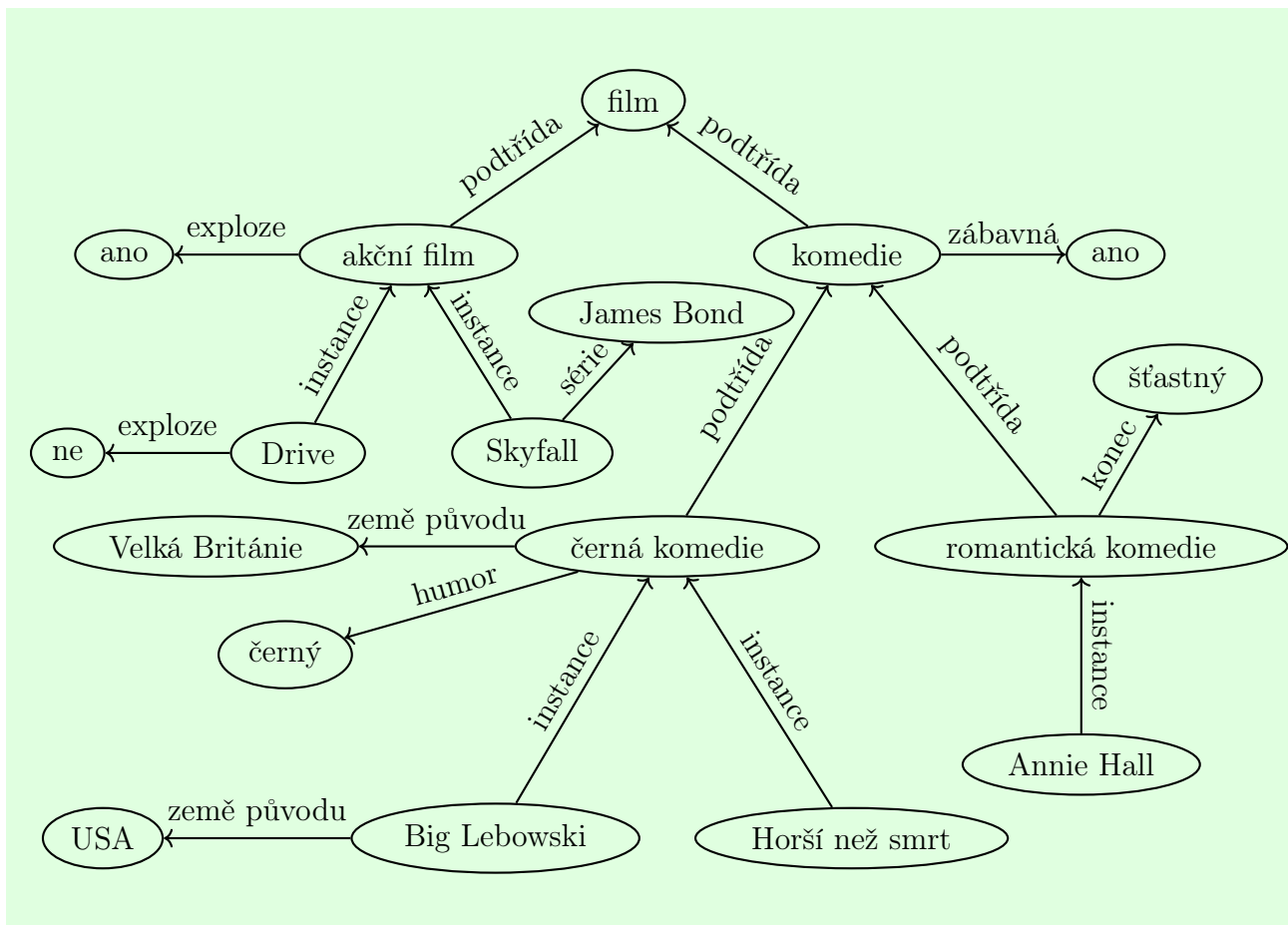
Rozhodněte, zda z rámce vyplývají následující tvrzení:

- Všechny akční filmy obsahují exploze.
- Skyfall obsahuje exploze.
- Akční film má více instancí než černá komedie.
- Horší než smrt je film původem z Británie.
- Annie Hall je film původem z USA.
- Big Lebowski je zábavný film.

- Ne.
- Ano.
- Ne.
- Ano.
- Ne.

f) Ano.

**Příklad 9.2.6.** Rámec z předchozího příkladu převedte na sémantickou síť.



### 9.3 Nejisté znalosti

V běžném životě se často setkáváme s nejistými znalostmi. K reprezentaci neurčitých znalostí se používají například nemonotónní logiky, které předpokládají, že výchozí výroky jsou pravdivé, dokud nejsou vyvráceny. Dalším přístupem je práce s pravděpodobnostmi, které se budeme věnovat v tomto oddílu.

Při definování pravděpodobnosti pracujeme s množinou (všech) elementárních jevů  $\Omega$ , která zahrnuje všechny možné výsledky náhodného pokusu. V případě hodu kostkou by bylo  $\Omega = \{1, 2, \dots, 6\}$ , v případě hodu mincí  $\Omega = \{\text{rub}, \text{líc}\}$ . Každá podmnožina množiny všech elementárních jevů se nazývá jev. V případě hodu kostkou může být jevem například „padné sudé číslo“, čili množina  $\{2, 4, 6\}$ . Množinu všech jevů značíme  $2^\Omega$  (známá též obecně jako

potenční množina či množina všech podmnožin).

**Definice 55:** Uvažme konečnou množinu elementárních jevů  $\Omega$ . *Pravděpodobnost* je funkce  $P : 2^\Omega \rightarrow [0, 1]$  přiřazující jevům hodnoty z intervalu  $[0, 1]$ , která splňuje

- $P(\Omega) = 1$ , čili pravděpodobnost celé množiny jevů je 1,
- $P(A \cup B) = P(A) + P(B)$  pro disjunkttní množiny  $A \subseteq \Omega$ ,  $B \subseteq \Omega$ , čili pravděpodobnost disjunkttních množin je aditivní.

Množina elementárních jevů  $\Omega$  vybavená pravděpodobností  $P$  se nazývá *pravděpodobnostní prostor*  $(\Omega, P)$ .

Intuitivní chápání definice pravděpodobnosti je, že hodnota  $P(A)$  určuje pravděpodobnost, že při náhodném pokusu nastane nějaký z jevů  $\omega \in A$ . Pokud je  $A = \Omega$ , pochopitelně je pravděpodobnost 1, jelikož nějaký z jevů nastat musí. Vyplývá z axiomů, že  $P(\emptyset) = 0$ ? Vyplývá z axiomů, že  $P(\Omega - A) = 1 - P(A)$ ? Rozmyslete si.

**Příklad.** Jsou následující funkce *pravděpodobnosti* nad množinou  $\Omega = \{\text{rub}, \text{líc}\}$ ? U těch, které nejsou, dokažte.

- $P_a = \{\emptyset \mapsto 0, \{\text{rub}\} \mapsto 0.2, \{\text{líc}\} \mapsto 0.7, \{\text{rub}, \text{líc}\} \mapsto 1\}$ ,
- $P_b = \{\emptyset \mapsto 0, \{\text{rub}\} \mapsto -0.2, \{\text{líc}\} \mapsto 1.2, \{\text{rub}, \text{líc}\} \mapsto 1\}$ ,
- $P_c = \{\emptyset \mapsto 0, \{\text{rub}\} \mapsto 0, \{\text{líc}\} \mapsto 1, \{\text{rub}, \text{líc}\} \mapsto 1\}$ ,
- $P_d = \{\emptyset \mapsto 0.3, \{\text{rub}\} \mapsto 0.3, \{\text{líc}\} \mapsto 0.7, \{\text{rub}, \text{líc}\} \mapsto 1\}$ ,

a) Ne, protože  $P_a(\{\text{rub}, \text{líc}\}) = 1 \neq 0.9 = P_a(\{\text{rub}\}) + P_a(\{\text{líc}\})$ .

b) Ne, protože  $P_b(\{\text{rub}\}) = -0.2 \notin [0, 1]$ .

c) Ano.

d) Ne, protože  $P_d(\{\text{rub}\}) = 0.3 \neq 0.6 = P_d(\{\text{rub}\}) + P_d(\emptyset)$ .

Elementární jevy často vykazují nějaké číselné charakteristiky, jako například číslo hozené na kostce či množství srážek v mm. K popisu číselných charakteristik slouží *náhodné proměnné*.

**Definice 56:** *Náhodná proměnná*  $X$  nad pravděpodobnostním prostorem  $(\Omega, P)$  je libovolná (měřitelná) funkce  $X : \Omega \rightarrow \mathbb{R}$ .

Náhodná proměnná tedy přiřazuje každému elementárnímu jevu číselnou hodnotu. Speciálním typem je *booleovská* náhodná proměnná, která své číselné hodnoty omezuje na 0 (nepravda) a 1 (pravda).

**Příklad.** Uvažme pokus, ve kterém 4 krát hodíme mincí. Dále mějme následující výsledek pokusu: líc, líc, rub, líc. Uvažte náhodné proměnné, popisující následující charakteristiky:

a) V každém hoďu padl líc.

b) Počet padlých líců.

Jakých hodnot nabudou tyto proměnné pro uvedený pokus? Jakých hodnot mohou nabývat všeobecně?

a) 0 (nepravda). Náhodná proměnná nabývá hodnot z množiny  $\{0, 1\}$  v závislosti na tom, zda v náhodném pokusu padly jenom líce.

b) 3. Náhodná proměnná nabývá hodnot z množiny  $\{0, 1, 2, 3, 4\}$  v závislosti na počtu padlých líců.

Vedle číselného značení 1 (pravda) a 0 (nepravda) budeme v případě booleovských náhodných proměnných hodnoty takové proměnné  $A$  značit rovněž jako  $a$  a  $\neg a$  (čili název proměnné začínající malým písmenem).

Když máme definovanou náhodnou proměnnou, můžeme uvažovat o pravděpodobnostech, s nimiž bude nabývat jednotlivých hodnot.

**Definice 57:** Uvažujme náhodnou proměnnou  $X$  nad pravděpodobnostním prostorem  $(\Omega, P)$ . *Pravděpodobnost, že proměnná  $X$  nabude hodnoty  $a$ , ozn.  $P(X = a)$ , je definována jako*

$$P(X = a) = P(\{\omega \in \Omega \mid X(\omega) = a\}),$$

neboli pravděpodobnost množiny elementárních jevů, pro něž nabývá náhodná proměnná hodnoty  $a$ .

Pokud je v uvažovaném kontextu zápis jednoznačný,  $P(X = a)$  můžeme značit zkráceně jako  $P(a)$ .

Rozložení pravděpodobnosti mezi jednotlivými hodnotami náhodné proměnné v souladu s uvedenou definicí definuje funkci, která se nazývá *rozdělení pravděpodobností náhodné proměnné*.

**Definice 58:** *Rozdělení (či distribuce) pravděpodobností náhodné proměnné  $X$ , značeno  $P(X)$ , je funkce  $P(X) : \mathbb{R} \rightarrow [0, 1]$ , která přiřadí každé hodnotě  $a \in \mathbb{R}$  pravděpodobnost, že  $X$  nabude hodnoty  $a$  podle Definice 57.*

Rozdělení pravděpodobnosti obsahuje pravděpodobnosti všech hodnot náhodné proměnné. Jelikož náhodná proměnná vždy nabývá nějaké hodnoty, pravděpodobnost, že nabude nějaké

z hodnot, je 1. Proto platí následující tvrzení.

**Věta 59:** Součet hodnot rozdělení pravděpodobnosti náhodné proměnné je roven 1.

**Příklad.** Uvažte booleovskou náhodnou proměnnou Prší ve světě, kde pravděpodobnost toho, že prší, je 0.3. Uveďte, jakým hodnotám se rovnají následující výrazy:

- a)  $P(\neg\text{prší})$ ,
- b)  $P(\text{Prší})$ .

- a) Výraz označuje pravděpodobnost toho, že neprší. Jelikož víme, že náhodná proměnná má pouze dvě hodnoty, potom  $P(\neg\text{prší}) = 1 - P(\text{prší}) = 0.7$ .
- b) Výraz označuje distribuci pravděpodobnosti, tedy

$$P(\text{Prší})(x) = \begin{cases} 0.3 & \text{pro hodnotu prší, neboli } x = 1 \text{ (pravda),} \\ 0.7 & \text{pro hodnotu } \neg\text{prší, neboli } x = 0 \text{ (nepravda),} \\ 0 & \text{jinak, neboli } x \neq 1 \wedge x \neq 0. \end{cases}$$

Distribuce pravděpodobností náhodných proměnných lze navzájem kombinovat. Vznikne tak složená distribuce pravděpodobností.

**Definice 60:** Složené rozdělení pravděpodobnosti náhodných proměnných  $X_1, \dots, X_n$  určuje pravděpodobnost toho, že náhodné proměnné  $X_1, \dots, X_n$  nabudou (v tom samém pokusu) hodnot  $a_1, \dots, a_n$ . Složené rozdělení značíme  $P(X_1 \wedge \dots \wedge X_n)$  nebo  $P(X_1, \dots, X_n)$ . Pravděpodobnost konkrétního přiřazení hodnot značíme obdobně,  $P(a_1 \wedge \dots \wedge a_n)$  nebo  $P(a_1, \dots, a_n)$ , kde  $a_i$  je hodnota náhodné proměnné  $X_i$ .

**Příklad.** Mějme složenou distribuci booleovských náhodných proměnných Prší a Sněží.

- a) Kolik pravděpodobnostních hodnot obsahuje jejich složená distribuce?
- b) Co reprezentuje výraz  $P(\text{sněží}, \neg\text{prší})$ ?

- a) 4. Jsou to  $P(\text{sněží}, \text{prší})$ ,  $P(\neg\text{sněží}, \text{prší})$ ,  $P(\text{sněží}, \neg\text{prší})$  a  $P(\neg\text{sněží}, \neg\text{prší})$ .
- b) Výraz reprezentuje pravděpodobnost, že sněží a neprší zároveň.

*Složené rozdělení pravděpodobnosti všech náhodných proměnných* v uvažovaném světě obsahuje pravděpodobnosti všech možných přiřazení hodnot proměnným. Zahrnuje tedy všechny situace, které mohou v daném světě nastat (z pohledu hodnot náhodných proměnných), nazývané též *atomické události*. Pomocí této distribuce pak lze vypočítat libovolnou pravděpodobnost, a to tak, že sečteme pravděpodobnosti všech atomických událostí, které splňují podmínku jevu, jehož pravděpodobnost nás zajímá.

**Příklad.** Uvažte náhodné proměnné Počasí s hodnotami {oblačno, sněží, slunečno} a Úterý s hodnotami {pravda, nepravda}. Pomocí hodnot složeného rozdělení pravděpodobnosti vyjádřete  $P(\text{oblačno} \vee \text{slunečno})$ .

$$P(\text{oblačno} \vee \text{slunečno}) = P(\text{úterý} \wedge \text{oblačno}) + P(\neg \text{úterý} \wedge \text{oblačno}) + P(\text{úterý} \wedge \text{slunečno}) + P(\neg \text{úterý} \wedge \text{slunečno}).$$

Pravděpodobnost, kterou jsme se zabývali doteď, vyjadřuje pravděpodobnost jevu bez ohledu na další informace, které o světě máme. Při výpočtu podmíněné pravděpodobnosti naopak do výpočtu zapojujeme znalosti o už odehraných událostech.

Například ze statistiky známek z minulého roku student ví, že pravděpodobnost hodnocení  $A$  na zkoušce je  $1/10$ . Pokud se ovšem učil na zkoušku týden a před zkouškou spal 9 hodin, pak může předpokládat, že pravděpodobnost  $A$  je vyšší. Čím více znalostí máme, tím lépe bude hodnota pravděpodobnosti odpovídat realitě.

Máme-li dva jevy  $A$  a  $B$ , podmíněná pravděpodobnost  $P(A | B)$ , čili  $A$  za podmínky  $B$ , je pravděpodobnost jevu  $A$ , pokud nastal jev  $B$ .

**Definice 61:** Uvažme jevy  $A, B \subseteq \Omega$ , kde  $P(B) \neq 0$ . *Podmíněná pravděpodobnost  $A$  za předpokladu  $B$*  je definována vztahem

$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

Nemá-li výskyt jevu  $B$  vliv na pravděpodobnost jevu  $A$ , říkáme, že jevy  $A$  a  $B$  jsou *stochasticky nezávislé*. Pro nezávislé jevy platí, že pravděpodobnost toho, že nastanou oba zároveň, je násobkem pravděpodobnosti jednotlivých jevů.

**Definice 62:** Dva jevy  $A, B \subseteq \Omega$  jsou *stochasticky nezávislé*, pokud

$$P(A \cap B) = P(A) \cdot P(B).$$

Jsou-li jevy  $A, B$  nezávislé, pak platí  $P(A | B) = P(A)$ , neboli zjištěním, zda jev  $B$  nastal, se nedozvíme žádnou další informaci, o pravděpodobnosti jevu  $A$ .

Z pravidla o podmíněné pravděpodobnosti lze odvodit Bayesovu větu:

**Věta 63 (jednoduchá Bayesova věta):** Uvažme dva náhodné jevy  $A, B \subseteq \Omega$ , přičemž  $P(B) \neq 0$ . Pak platí

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}.$$

Bayesovu větu lze dobře aplikovat v případech, kdy uvažujeme o pravděpodobnostech příčiny a následku a známe hodnotu jedné podmíněné pravděpodobnosti. Můžeme tak například vyjádřit pravděpodobnost, že člověk má určitou nemoc, má-li její příznak (přičemž pravděpodobnost příznaku, pokud člověk nemoc má, známe), či pravděpodobnost, že výsledek testu je korektní, vyšel-li pozitivně.

**Příklad.** Víme, že pravděpodobnost slunečného počasí je 0.3 a pravděpodobnost horkého počasí (alespoň 30 °C) je 0.1. Dále víme, že pokud je horko, pravděpodobnost slunečného počasí stoupne na 0.9.

- Jaká je pravděpodobnost, že svítí slunce a je horko?
- Jaká je pravděpodobnost, že svítí slunce a horko není?
- Jaká je pravděpodobnost, že je venku horko, pokud z okna vidíme, že svítí slunce?

Pravděpodobnost slunečného počasí budeme značit  $P(s) = 0.3$ , pravděpodobnost horkého počasí  $P(h) = 0.1$  a pravděpodobnost slunečného počasí za předpokladu, že je horko,  $P(s | h) = 0.9$ .

- Pravděpodobnost, že svítí slunce a je horko označíme jako  $P(s \wedge h)$  a tuto pravděpodobnost vyjádříme pomocí pravidla o podmíněné pravděpodobnosti, tedy  $P(s \wedge h) = P(s | h) \cdot P(h) = 0.9 \cdot 0.1 = 0.09$ .
- Pravděpodobnost, že svítí slunce a horko není, tedy  $P(s \wedge \neg h)$ , vypočítáme s využitím principu složené distribuce pravděpodobnosti. Víme, že  $P(s) = P(s \wedge h) + P(s \wedge \neg h)$ , potom  $P(s \wedge \neg h) = P(s) - P(s \wedge h) = 0.3 - 0.09 = 0.21$ .

- c) Pro výpočet pravděpodobnosti horkého počasí za předpokladu, že je slunečno,  $P(h | s)$  použijeme Bayesovu větu. Platí  $P(h | s) = \frac{P(s|h) \cdot P(h)}{P(s)} = \frac{0.9 \cdot 0.1}{0.3} = 0.3$ . Pravděpodobnost, že je horko, pokud svítí slunce, je tedy 0.3.

**Příklad 9.3.1.** Zaveďte náhodnou proměnnou, která reprezentuje následující charakteristiky:

- Prší.
- Venkovní teplota naměřená digitálním teploměrem.
- Počet hodin, které strávil studiem PB016 absolvent předmětu.

Správné řešení může vypadat například takto:

- Náhodná proměnná Prší, která nabývá hodnot z množiny {pravda, nepravda} podle toho, zda zrovna prší.
- Náhodná proměnná Teplota, která nabývá hodnot z množiny  $\{-20, -19, \dots, 39, 40\}$ , jelikož náš teploměr měří na celé stupně v rozsahu  $-20^\circ\text{C}$  až  $40^\circ\text{C}$ .
- Náhodná proměnná Časová náročnost, která nabývá hodnoty z množiny přirozených čísel podle toho, jak moc se konkrétní student učil.

**Příklad 9.3.2.** Uvažte následující tabulku složených pravděpodobností náhodné proměnné Počasí s hodnotami  $\{0$  (slunečno),  $1$  (déšť),  $2$  (sníh) $\}$  a booleovských proměnných Zácpy a Zpoždění.

	zpoždění		$\neg$ zpoždění	
	zácpy	$\neg$ zácpy	zácpy	$\neg$ zácpy
slunečno	0.15	0.10	0.05	0.30
déšť	0.15	0.05	0.02	0.08
sníh	0.07	0.01	0.01	0.01

Určete hodnoty následujících výrazů.

- ★  $P(\text{zácpy})$
- ★  $P(\text{Zácpy})$
- ★  $P((\text{zácpy} \vee \neg \text{zpoždění}) \wedge \text{slunečno})$
- $P(\text{Zácpy} | \text{déšť})$
- $P(\text{sníh} | \neg \text{zácpy} \vee \text{zpoždění})$

- a) 0.45 ( $0.15 + 0.15 + 0.07 + 0.05 + 0.02 + 0.01 = 0.45$ )



b) Výraz označuje distribuci pravděpodobnosti, tedy

$$P(\text{Zácpy})(x) = \begin{cases} 0.45 & \text{pro hodnotu zácpy, neboli } x = 1 \text{ (pravda),} \\ 0.55 & \text{pro hodnotu } \neg\text{zácpy, neboli } x = 0 \text{ (nepravda),} \\ 0 & \text{jinak, neboli } x \neq 1 \wedge x \neq 0. \end{cases}$$

c)  $0.5$  ( $0.15 + 0.05 + 0.3 = 0.5$ )

d) Výraz označuje distribuci pravděpodobnosti náhodné proměnné Zácpy, pokud prší, tedy

$$P(\text{Zácpy} \mid \text{děšť})(x) = \begin{cases} 0.57 & \text{pro hodnotu zácpy, neboli } x = 1 \text{ (pravda),} \\ 0.43 & \text{pro hodnotu } \neg\text{zácpy, neboli } x = 0 \text{ (nepravda),} \\ 0 & \text{jinak, neboli } x \neq 1 \wedge x \neq 0, \end{cases}$$

přičemž  $0.57 = (0.15 + 0.02)/(0.15 + 0.05 + 0.02 + 0.08)$  a  $0.43 = (0.05 + 0.08)/(0.15 + 0.05 + 0.02 + 0.08)$ .

e)  $0.10$  ( $(0.07 + 0.01 + 0.01) / (0.55 + 0.15 + 0.15 + 0.07) = 0.09 / 0.92 = 0.10$ )

**Příklad 9.3.3.** Uvažte tabulku složeného rozdělení pravděpodobnosti z předchozího příkladu. Určete, kolik hodnot z tabulky potřebujeme použít pro výpočet následujících pravděpodobností.

- a)  $P(\text{děšť} \wedge \neg\text{zpoždění} \wedge \text{zácpy})$
- b)  $P(\text{slunečno} \wedge \text{zpoždění})$
- c)  $P(\text{slunečno} \vee \text{zpoždění})$
- d)  $P(\text{slunečno} \vee \text{děšť})$

a) 1

b) 2

c) 4 (pravděpodobnost lze vyjádřit jako  $1 - P((\text{děšť} \vee \text{sníh}) \wedge \neg\text{zpoždění}) = 1 - (0.02 + 0.08 + 0.01 + 0.01)$ )

d) 4 ( $1 - P(\text{sníh})$ )

**Příklad 9.3.4.** Doplňte následující tabulku složených pravděpodobností o hodnoty  $a$  a  $b$  tak, aby jevy  $X$  a  $Y$  byly na sobě nezávislé.

$X$	$Y$	$P(X \cap Y)$
1	1	$3/5$
1	0	$1/5$
0	1	$a$
0	0	$b$

Součet všech hodnot pravděpodobnosti je 1, proto  $a + b = 1/5$ . Dále spočítáme poměr mezi hodnotami  $a$  a  $b$ . Jelikož jsou jevy nezávislé, platí  $P(X \cap Y) = P(X) \cdot P(Y)$ , z čehož

$$\frac{P(y)}{P(\neg y)} = \frac{P(x)P(y)}{P(x)P(\neg y)} \stackrel{\text{nezáv.}}{=} \frac{P(x, y)}{P(x, \neg y)} = \frac{3/5}{1/5} = 3.$$

Podobně pak

$$\frac{a}{b} = \frac{P(\neg x, y)}{P(\neg x, \neg y)} \stackrel{\text{nezáv.}}{=} \frac{P(\neg x)P(y)}{P(\neg x)P(\neg y)} = \frac{P(y)}{P(\neg y)} = 3.$$

Řešením soustavy  $a + b = 1/5$  a  $a/b = 3$  získáme řešení  $a = 3/20$ ,  $b = 1/20$ .

**Příklad 9.3.5.** Vyjádřete následující pravděpodobnosti:

- $P(a | b)$  pomocí  $P(b)$  a  $P(\neg a \wedge b)$ ,
- $P(a | b)$  pomocí  $P(a \wedge b)$  a  $P(\neg a \wedge b)$ ,
- $P(b)$  pomocí  $P(a | b)$  a  $P(a \wedge b)$ .

a)

$$P(a | b) = 1 - \frac{P(\neg a \wedge b)}{P(b)}$$

b)

$$P(a | b) = \frac{P(a \wedge b)}{P(a \wedge b) + P(\neg a \wedge b)}$$

c)

$$P(b) = \frac{P(a \wedge b)}{P(a | b)}$$

**Příklad 9.3.6.** Dokažte, že

- jsou-li dva jevy  $A$ ,  $B$  nezávislé, pak  $P(A | B) = P(A)$ ,
- jsou-li  $A$ ,  $B$  dva jevy a  $P(B) \neq 0$ , platí Bayesův vzorec

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}.$$

Vyjděte z Definice 61 a Definice 62.

a)

$$P(A | B) \stackrel{\text{def. 61}}{=} \frac{P(A \cap B)}{P(B)} \stackrel{\text{def. 62}}{=} \frac{P(A) \cdot P(B)}{P(B)} = P(A).$$

b)

$$P(A | B) \stackrel{\text{def. 61}}{=} \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B) \cdot P(A)}{P(B) \cdot P(A)} \stackrel{\text{def. 61}}{=} \frac{P(B | A)P(A)}{P(B)}$$

**Příklad 9.3.7.** Následující situace modelujte pomocí náhodných proměnných a odpovězte na otázky.

- a) ★ Mimoszemšťané mohou být přátelští, ale nemusí. 75 % jich přátelských je. Přátelští mimozemšťané přistávají na zemi v 90 % případů přes den, nepřátelští výhradně v noci. Pokud mimozemšťan přistane v noci, jaká je pravděpodobnost, že je přátelský?
- b) Polovina příšer žije na půdě, ostatní ve sklepě. Zatímco 80 % všech příšer je neškodných, všechny příšery, které žijí na půdě, jsou neškodné. Jaká je pravděpodobnost, že příšera, která žije ve sklepě, je neškodná?

- a) Nechť náhodná proměnná  $P$  představuje, zda je mimozemšťan přátelský, a náhodná proměnná  $D$  představuje, zda mimozemšťan přistál přes den. Pak  $P(p) = 0,75$ ,  $P(\neg p) = 0,25$ ,  $P(d | p) = 0,9$ ,  $P(\neg d | p) = 0,1$ ,  $P(d | \neg p) = 0$  a  $P(\neg d | \neg p) = 1$ . Pomocí Bayesova pravidla vyjádříme pravděpodobnost  $P(p | \neg d)$  následovně:

$$\begin{aligned} P(p | \neg d) &= \frac{P(\neg d | p) \cdot P(p)}{P(\neg d)} \\ &= \frac{P(\neg d | p) \cdot P(p)}{P(\neg d | p) \cdot P(p) + P(\neg d | \neg p) \cdot P(\neg p)} \\ &= \frac{0,1 \cdot 0,75}{(0,1 \cdot 0,75) + (1 \cdot 0,25)} \\ &= 0,23 \end{aligned}$$

- b) Nechť náhodná proměnná  $P$  reprezentuje, zda příšera žije na půdě nebo ne (tedy žije ve sklepě), a náhodná proměnná  $N$  reprezentuje, zda je příšera neškodná. Pak  $P(p) = 0,5$ ,  $P(\neg p) = 0,5$ ,  $P(n) = 0,8$ ,  $P(n | p) = 1$  a  $P(\neg n | p) = 0$ . Pravděpodobnost  $P(n | \neg p)$

pak vyjádříme následovně:

$$\begin{aligned}
 P(n \mid \neg p) &= \frac{P(n \wedge \neg p)}{P(\neg p)} \\
 &= \frac{P(n) - P(n \wedge p)}{P(\neg p)} \\
 &= \frac{P(n) - P(n \mid p) \cdot P(p)}{P(\neg p)} \\
 &= \frac{0.8 - 1 \cdot 0.5}{0.5} \\
 &= 0.6
 \end{aligned}$$

**Příklad 9.3.8.** Alenka je odbornice na cvrčky. Většinu druhů dokáže určit podle jejich zvuku. Cvrček zpěvavý však dokáže napodobovat zvuky jiných cvrčků, takže je velmi obtížné ho rozeznat. Navíc je velmi vzácný – tvoří pouze 2 % populace cvrčků. Přesto má Alenka výborné statistiky při rozpoznávání tohoto cvrčka, s pravděpodobností 95 % správně určí, zda se jedná o cvrčka zpěvavého, či nikoli. Pokud se nejedná o cvrčka zpěvavého, určí tak v 96% případech.

- Jaká je pravděpodobnost, že Alenka pozná cvrčka zpěvavého?
- Pokud slyšíme cvrčka, jaká je pravděpodobnost, že Alenka řekne, že jde o cvrčka zpěvavého?

Nechť náhodná proměnná  $A$  představuje, zda Alenka tvrdí, že jde o cvrčka zpěvavého, a náhodná proměnná  $Z$  představuje, zda je cvrček skutečně zpěvavý. Pak  $P(z) = 0,02$ ,  $P(\neg z) = 0,98$ ,  $P(a \wedge z) \vee P(\neg a \wedge \neg z) = 0,95$  a  $P(\neg a \mid \neg z) = 0,96$ ,  $P(a \mid \neg z) = 0,04$ .

a)

$$\begin{aligned}
 P(a \mid z) &= \frac{P(a \wedge z)}{P(z)} \\
 &\stackrel{\text{disj.}}{=} \frac{P(a \wedge z) \vee P(\neg a \wedge \neg z) - P(\neg a \wedge \neg z)}{P(z)} \\
 &= \frac{P(a \wedge z) \vee P(\neg a \wedge \neg z) - P(\neg a \mid \neg z) * P(\neg z)}{P(z)} \\
 &= \frac{0.95 - 0.96 \cdot 0.98}{0.02} \\
 &= \frac{0.0092}{0.02} \\
 &= 0.46
 \end{aligned}$$

b)

$$\begin{aligned} P(a) &= P(a \wedge z) + P(a \wedge \neg z) \\ &= P(a \wedge z) + P(a \mid \neg z) \cdot P(\neg z) \\ &= 0.0092 + 0.04 \cdot 0.98 \\ &= 0.05 \end{aligned}$$

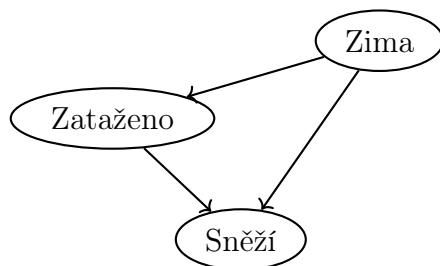
**Příklad 9.3.9.** Alenčin kamarád Bořek tvrdí, že jeho statistiky jsou ještě lepší. Zda se jedná o cvrčka zpěvavého, či nikoli, prý určí správně s pravděpodobností 98 %. K tomu si přisadí Cyril s tvrzením, že on má ještě lepší výsledky. Pokud slyší cvrčka zpěvavého, správně ho prý určí s pravděpodobností 100 %. Bořek ani Cyril nejsou experti na cvrčky, ale o pravděpodobnostech neklamou. Jak je to možné?

Bořek i Cyril mají odlišné, ale jednoduché taktiky. Bořek pokaždé tvrdí, že cvrček není cvrček zpěvavý. Cvrček zpěvavý tvoří pouze 2 % populace všech cvrčků, pravděpodobnost, že cvrček není cvrček zpěvavý, a tedy pravděpodobnost, že Bořek cvrčka určil správně, je proto 98 %. Cyril naopak vždy tvrdí, že slyší cvrčka zpěvavého. Pokud se tedy jedná o cvrčka zpěvavého, má 100% úspěšnost.

## 9.4 Bayesovské sítě

*Bayesovská síť* je datová struktura používaná k reprezentaci pravděpodobnosti nejistých jevů a závislostí mezi nimi. Skládá se z acyklického orientovaného grafu, v němž uzly představují jednotlivé náhodné proměnné a hrany označují, které jevy jsou na sobě závislé. Pokud hrana vede z  $A$  do  $B$ , říkáme, že  $A$  je rodičem  $B$ , což znamená, že náhodná proměnná  $B$  je přímo závislá na náhodné proměnné  $A$ . Pro každou náhodnou proměnnou jsou uvedeny pravděpodobnosti jejích hodnot, které jsou podmíněny tím, jakých hodnot nabývají nadřazené jevy.

Bayesovská síť může vypadat například takto:



Přítom pravděpodobnostní rozdělení jednotlivých proměnných jsou uvedena v tabulce (pravděpodobnosti hodnot *nepravda* nejsou uvedeny, lze je snadno vypočítat):

Zima (Z):

$P(Z)$
0.2

Zataženo (T):

Z	$P(T)$
1	0.6
0	0.5

Sněží (S):

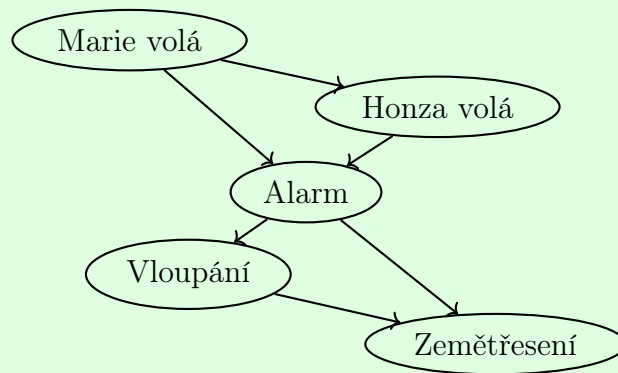
Z	T	$P(S)$
1	1	0.8
1	0	0.01
0	1	0.001
0	0	0

První řádek tabulky Sněží například určuje pravděpodobnost  $P(s | z \wedge t)$ , tedy sněžení za předpokladu, že je zima a zároveň zataženo.

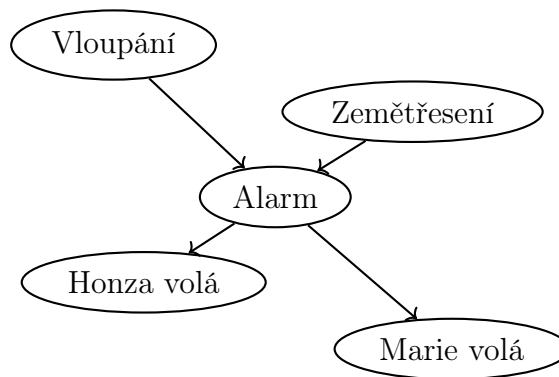
Bayesovskou síť tvoříme postupným přidáváním jevů jako uzlů. Do nového uzlu zavádíme hrany ze stávajících uzlů, které přímo ovlivňují jeho pravděpodobnost. Formálně, pro uzel  $X_i$  hledáme minimální podmnožinu  $\text{Parents}(X_i)$  z množiny již přidávaných uzlů  $X_1, \dots, X_{i-1}$  takovou, že  $P(X_i | \text{Parents}(X_i)) = P(X_i | X_1, \dots, X_{i-1})$ . Poté zopakujeme postup přidáním dalšího uzlu. Pořadí přidávání lze zvolit libovolně, ale přidávání příčin před následky obvykle vede ke kompaktnější síti.

**Příklad.** V domě máme nainstalován alarm, který se spustí v případě vloupání, ale také v případě zemětřesení. Požádali jsme sousedy Marii a Honzu, aby nám zavolali, pokud uslyší náš alarm. Sousédy se mohou splést a zavolat, i když alarm nezazvoní, nebo naopak nezavolat, když zazvoní. Uvažujme náhodné proměnné Vloupání, Zemětřesení, Alarm, Honza volá, Marie volá. Vytvořte síť přidáním uzlů v pořadí M, H, A, V, Z.

1. Přidáme uzel Marie volá. Jelikož v síti nejsou žádné uzly, uzel Marie volá nemá rodiče.
2. Přidáme uzel Honza volá. Pokud volá Marie, pravděpodobně se spustil alarm, a tudíž i pravděpodobnost, že zavolá Honza se zvyšuje – jev Honza volá je v tomto případě přímo závislý na jevu Marie volá. Vedeme tedy hranu z M do H.
3. Přidáme uzel Alarm. Pravděpodobnost, že zvoní alarm se přímo zvyšuje, pokud volá Marie, a ještě více, volá-li i Honza – z uzlů M a H vedeme hranu do A.
4. Přidáme uzel Vloupání. Pokud víme, zda alarm zvonil, informace o tom, zda volala Marie nebo Honza, nám o pravděpodobnosti vloupání nic nového neřekne (předpokládáme, že sousédy nesledují, zda někdo vloupává, pouze volají, pokud slyší alarm). Uzel vloupání je tedy přímo závislý pouze na alarmu.
5. Přidáme uzel Zemětřesení. Pokud víme, že zvonil alarm, pravděpodobnost zemětřesení se zvyšuje. Na druhou stranu, pokud víme, že došlo k vloupání a byl spuštěn alarm, pravděpodobnost je menší (alarm byl pravděpodobně spuštěn lupičem). Pravděpodobnost zemětřesení je tedy přímo ovlivněna spuštěním alarmu i vloupáním.



Z uvedeného příkladu je zřejmé, že dané pořadí přidávání uzlů nebylo zvoleno nejvhodněji a v síti vznikly neintuitivní závislosti. Síť však stále správně reprezentuje situaci. Pokud bychom přidávali uzly v pořadí od příčiny k následku, např. V, Z, A, H, M, vznikla by následující síť, která reprezentuje závislosti intuitivněji a je také kompaktnější:



Bayesovská síť reprezentuje složené rozdělení pravděpodobnosti všech náhodných proměnných, a to často kompaktněji než tabulka složené distribuce. Zároveň, podobně jako za pomoci tabulky, z ní můžeme vyjádřit jakoukoli pravděpodobnost jevu v uvažovaném prostoru.

**Příklad 9.4.1.** ★ Uvažte situaci popsanou v úvodu s náhodnými proměnnými Vloupání, Zemětřesení, Alarm, Honza volá, Marie volá. Vytvořte síť přidáváním uzlů v pořadí Z, H, A, V, M.

Z: Přidáme první uzel.

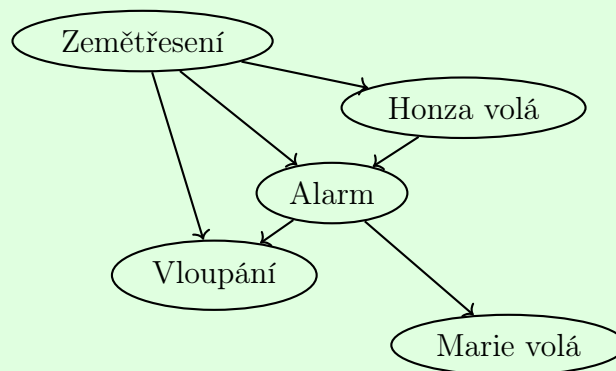
H: Pokud dojde k zemětřesení, rozezní se alarm, a tedy pravděpodobnost, že Honza volá, se zvyšuje.

A: Pravděpodobnost, že zní alarm, se zvyšuje, pokud dojde k zemětřesení, a také v případě,

že volá Honza (mohla nastat situace, kdy Honza volá kvůli alarmu, ale alarm zvoní kvůli vloupání, nikoliv kvůli zemětřesení). Uzel A je tedy závislý na obou jevech.

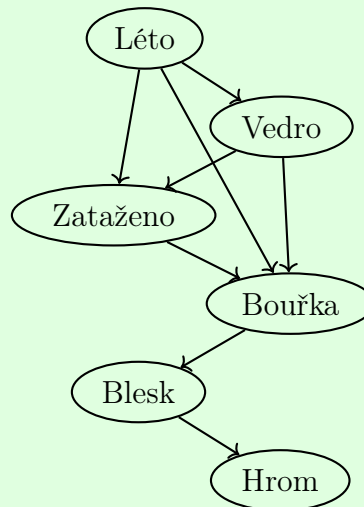
V: Pravděpodobnost vloupání závisí na alarmu, ale také na zemětřesení – pokud víme, že alarm zvoní a zároveň došlo k zemětřesení, je pravděpodobnost vloupání nižší, než když zvoní pouze alarm.

M: Uzel M je přímo závislý pouze na uzlu A.



**Příklad 9.4.2.** Uvažte náhodné proměnné Bouřka, Zataženo, Léto, Hrom, Vedro a Blesk. Vytvořte bayesovskou síť, která bude zachytávat závislosti jevů co nejkompaktněji.

Abychom vytvořili co nejkompaktnější síť, budeme jevy přidávat v pořadí od příčiny k následku (což v tomto případě není zcela jednoznačné, můžete zvolit i jiné pořadí), v tomto řešení konkrétně v pořadí Léto, Vedro, Zataženo, Bouřka, Blesk, Hrom.





## 10 Strojové učení

Metody strojového učení nám umožňují konstruovat systémy, které se svou funkcionalitu postupně učí z dat. Tyto systémy využívají získávané znalosti k tomu, aby vylepšovaly svou výkonnost. Jednou ze základních vlastností takových systémů je schopnost generalizovat, tedy schopnost aplikovat naučené znalosti na doposud neviděná data.

Algoritmy strojového učení můžeme podle způsobu učení rozdělit do třech hlavních kategorií:

- *Učení s učitelem*, kde známe pro tréninková data i očekávané výstupy.
- *Učení bez učitele*, kde se model učí vzory v datech bez dodatečných informací.
- *Zpětnovazebné učení* („reinforcement learning“), kde za provedené akce následuje zpětná vazba v podobě odměny nebo trestu.

Pomocí metod strojového učení můžeme řešit spoustu různých úloh. *Klasifikace* má za cíl přidělit vstupním objektům jednu z konečně mnoha předem určených tříd. Může jít například o rozhodování, zda počasí pro daný den ohodnotit jako *slunečno*, *polojasno*, nebo *deštivo*. Cílem *regrese* je na základě vstupů modelovat spojitou proměnnou, například odhadnout venkovní teplotu v určitém čase. *Shlukování*, které je obvykle příkladem učení bez učitele, se snaží vstupní data rozdělit do několika skupin, které shlukují vzájemně podobné příklady. V této sbírce se budeme věnovat pouze učení s učitelem. Zaměříme se hlavně na úlohu klasifikace, setkáme se ale i s regresí.

Konkrétních metod nebo algoritmů, které lze zařadit pod strojové učení, existuje opět mnoho. V této kapitole se budeme postupně věnovat dvěma základním typům. V první části to bude jednoduchý (ale velmi používaný) model *rozhodovacích stromů*. Druhá část se pak bude týkat algoritmů okolo *lineárních modelů*. Na tyto modely poté navážeme v další kapitole, která představí složitější architektury *neuronových sítí*.

### 10.1 Rozhodovací stromy

Model rozhodovacího stromu nám poskytuje přirozený pohled na klasifikační úlohy. Větve takového stromu reprezentují sekvence testů na atributy vstupu, pomocí kterých postupně dojdeme k výsledku. Není-li řečeno jinak, budeme pro jednoduchost předpokládat, že atributy vstupních příkladů nabývají diskrétních hodnot.

**Definice 64:** *Rozhodovací strom* je speciální typ kořenového stromu, který má dva typy uzlů:

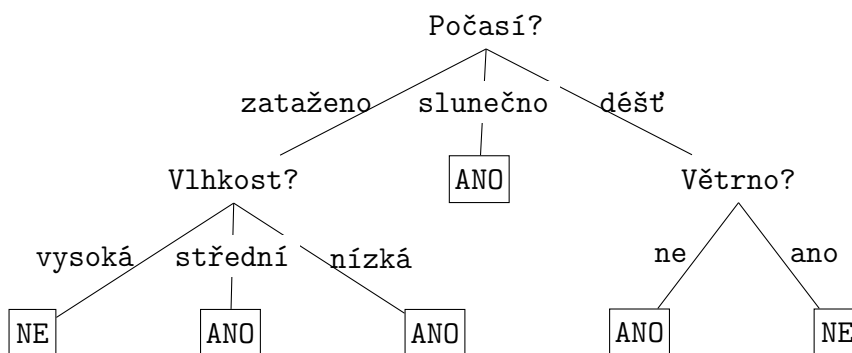
- vnitřní (rozhodovací) uzly,
- listové uzly.

Vnitřní uzel stromu odpovídá testu na hodnotu některého atributu vstupu. Hrany vedoucí z uzlu do jeho potomků poté reprezentují možné hodnoty testovaného atributu. Listové uzly specifikují výslednou třídu.

**Definice 65:** Říkáme, že rozhodovací strom je *konzistentní* s konkrétním datasetem právě když správně oklasifikuje všechny příklady z datasetu.

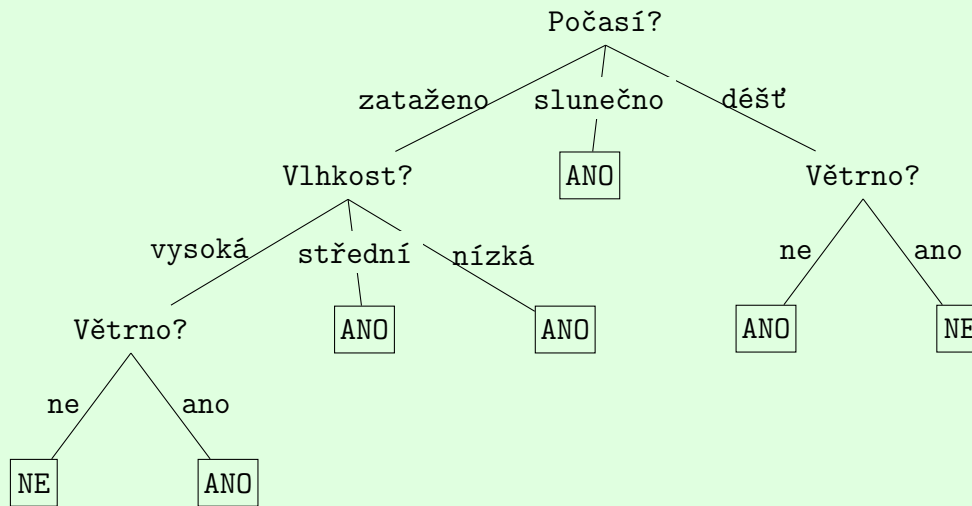
Ilustrujme si klasifikaci pomocí rozhodovacího stromu na elementárním příkladu. Pro jednoduchost uvažujme případ binární klasifikace, kde jsou listové uzly označeny pouze ANO/NE. Obecně ale samozřejmě rozhodovací stromy takto omezeny nejsou.

**Příklad.** Mějme následující strom reprezentující rozhodovací problém, zda jsou venkovní podmínky vhodné pro hraní tenisu (rozhodujeme se pomocí veličin Vlhkost, Počasí, Větrno).



- Klasifikujte pomocí daného stromu následující situace (příklady obsahují atributy v pořadí Vlhkost, Počasí, Větrno):
  - [vysoká, slunečno, ne]
  - [střední, déšť, ano]
  - [vysoká, zataženo, ne]
- Upravte strom přidáním jednoho rozhodovacího uzlu tak, aby výstupem pro příklad [vysoká, zataženo, ano] bylo ANO. Ostatní příklady by nový strom měl klasifikovat stejně jako původní.
- Mohli bychom původní strom použít i ke klasifikaci příkladů z datasetu čtveřic s atributy Vlhkost, Počasí, Větrno a Teplota? Vstupy by tedy kromě původních tří atributů obsahovaly ještě dodatečný atribut Teplota.

- ANO – z prvního rozhodovacího uzlu se dostaneme po hraně slunečno rovnou do listového uzlu
  - NE
  - NE
- Například následující strom:



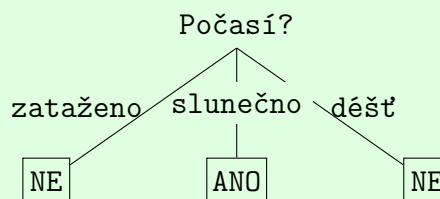
- c) Ano, mohli. Výsledná třída by ale na novém atributu *Teplota* nikdy nezáležela, jelikož se ve stromě podle tohoto atributu nikde nerozhoduje.

**Příklad 10.1.1.** Mějme dataset  $D$  zadaný následující tabulkou (řádky reprezentují jednotlivé příklady, sloupce atributy, závislá proměnná je *Tenis*).

<i>Vlhkost</i>	<i>Počasí</i>	<i>Větrno</i>	<i>Tenis</i>
vysoká	zataženo	ano	NE
střední	slunečno	ne	ANO
nízká	slunečno	ano	ANO
střední	déšť	ne	NE

- a) Vytvořte (ručně) rozhodovací strom, který je konzistentní s tímto datasetem. Snažte se o co nejmenší strom.  
 b) Jak by (konzistentní) strom vypadal, pokud by v  $D$  byl navíc ještě příklad [střední, slunečno, ne] s třídou NE?

- a) Konzistentní je například následující strom:



Pokuste se vytvořit ještě alespoň jeden jiný rozhodovací strom.

- b) Žádný konzistentní strom nemůže existovat, protože máme stejný příklad označený dvěma různými třídami.

**Příklad 10.1.2.** Je pravda, že v žádné větvi rozhodovacího stromu nemá cenu testovat stejný atribut vícekrát? Svou odpověď odůvodněte.

- Máme-li diskrétní atribut, vícenásobné testování na jeho hodnotu v rámci stejné větve stromu by bylo redundantní. Uvažte jednu konkrétní větev  $v$ , na které testujeme atribut  $A$  vícekrát. První test pro  $A$  na této větvi rozdělí všechny instance podle různých hodnot  $A$ . Všechny instance, které po tomto testu větvi  $v$  pokračují, proto musí mít pro  $A$  stejnou hodnotu. Testujeme-li v rámci  $v$  později stejný atribut znova, nemůže dojít k rozdělení zbylých instancí. Druhý test je tedy zbytečný.
- Zajímavé je, že kdybychom uvažili spojité atributy, může mít smysl na jedné větvi testovat jeden atribut vícekrát. Typicky platí, že test na spojitý atribut rozděljuje příklady podle určité *prahové hodnoty* (*threshold*). Pokud testujeme na daný atribut znova, ale s jinou prahovou hodnotou, může dojít k dalšímu rozdělení příkladů.

## 10.2 Učení rozhodovacích stromů

K vytvoření stromu využíváme (jak je tomu ve strojovém učení obvyklé) množinu trénovacích dat. Trénovací data se v tomto případě skládají z dvojic příkladů a jejich očekávaných tříd.

Často lze ovšem vytvořit spoustu různých stromů konzistentních s danými trénovacími daty. Tyto stromy se mohou lišit jak ve své velikosti, tak ve schopnosti generalizace. Cílem obvykle bývá vytvořit co nejmenší konzistentní strom, který co nejlépe generalizuje na nové příklady. Jelikož je ale problém nalezení nejmenšího konzistentního rozhodovacího stromu *NP-úplný*, musíme si vystačit s heuristickými postupy. Tyto postupy budují strom od kořene, a v každém uzlu se snaží najít (v nějakém smyslu) co *nejlepší* rozhodovací atribut. Představíme si postup, který jako metriky k nalezení takového atributu využívá *entropii* a *informační zisk* (*information gain*). Nejdříve si tyto metriky zdefinujeme.

Entropie je obecně mírou neuspořádanosti, nejistoty, nebo informace. V původním smyslu jde o míru nejistoty náhodné proměnné – my budeme uvažovat jako tuto náhodnou proměnnou třídu. Ilustrujme si nejdříve jak vypadají krajní případy pro binární klasifikaci (kde proměnná třídy má dvě hodnoty). Uvažme dataset tvořený příklady pouze jedné třídy. Pokud bychom náhodně vytáhli jeden příklad, už dopředu máme jistotu, jaká bude jeho třída. Proto výsledek tohoto pokusu nenese žádnou informační hodnotu a entropie (míra informace) je nulová. Uvažme naopak dataset, jenž je z poloviny tvořen příklady jedné třídy,

a z druhé poloviny příklady třídy druhé. V tomto případě nemáme dopředu žádnou jistotu, jakou bude mít náhodně zvolený příklad třídu, a proto je entropie nejvyšší (pro binární klasifikaci rovna 1).

**Definice 66:** Uvažujme množinu příkladů  $D$ . Nechť jsou příklady rozděleny do  $n$  tříd a  $P_i$  je pravděpodobnost, že náhodně vybraný příklad bude mít třídu  $i$ . Entropii pro dataset  $D$  spočítáme jako

$$E(\langle P_1, \dots, P_n \rangle) = \sum_{i=1}^n -P_i \cdot \log_2(P_i)$$

Máme-li třídy pouze 2, vzorec se obvykle zapisuje následovně:

$$E(\langle P_1, P_2 \rangle) = -P_1 \cdot \log_2(P_1) - (1 - P_1) \cdot \log_2(1 - P_1)$$

Pro zjednodušení notace budeme někdy uvádět entropii datasetu  $D$  jako  $E(D)$ .

Entropii můžeme využít k výpočtu informačního zisku, jenž nám dává jakýsi heuristický odhad, jak je který atribut „dobrý“. Informační zisk pro určitý atribut nám intuitivně říká, jak moc se liší „neuspořádanost“ množiny před a po rozdělení podle daného atributu – tedy kolik informace získáme testem na hodnotu onoho atributu. Informační zisk je vyšší, pokud se entropie („neuspořádanost“) rozdělením sníží.

**Definice 67:** Uvažujme množinu příkladů  $D$  a atribut  $A$ , který může nabývat celkem  $k$  hodnot.  $A$  tedy rozděluje množinu  $D$  na podmnožiny  $D_1, \dots, D_k$ . Nechť  $p$  je celkový počet pozitivních příkladů v  $D$  a  $n$  počet negativních (uvažujeme binární klasifikaci). Dále nechť  $p_i$  je počet pozitivních příkladů v  $D_i$  a  $n_i$  počet negativních v  $D_i$ .

Informační zisk atributu  $A$  v rámci datasetu  $D$  spočítáme jako

$$Gain(A) = E(\langle \frac{p}{p+n}, \frac{n}{p+n} \rangle) - \sum_{i=1}^k \frac{p_i+n_i}{p+n} \cdot E(\langle \frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i} \rangle)$$

**Příklad.** Mějme dataset daný tabulkou níže (řádky reprezentují jednotlivé příklady, sloupce kromě prvního atributy, závislá proměnná je *Tenis*). Spočítejte

- a) entropii celého setu,
- b) informační zisk pro atribut *Počasi*.

	<i>Vlhkost</i>	<i>Počasí</i>	<i>Větrno</i>	<i>Teplota</i>	<i>Tenis</i>
1	vysoká	zataženo	ano	nižší	NE
2	střední	zataženo	ne	vyšší	ANO
3	nízká	slunečno	ano	nižší	ANO
4	střední	děšť	ne	nižší	NE
5	nízká	slunečno	ano	vyšší	ANO
6	vysoká	děšť	ne	vyšší	NE
7	střední	slunečno	ano	nižší	ANO

Zamyslete se, co nám spočítaná entropie říká o datasetu, a zda-li může být atribut *Počasí* vhodným kandidátem pro rozhodovací uzel.

$$\text{a) } E(\langle P_{ANO}, P_{NE} \rangle) = -\frac{4}{7} \cdot \log_2\left(\frac{4}{7}\right) - \left(\frac{3}{7}\right) \cdot \log_2\left(\frac{3}{7}\right) \approx \frac{4}{7} \cdot 0.807 + \left(\frac{3}{7}\right) \cdot 1.222 \approx 0.985$$

$$\text{b) } \textit{Gain}(\textit{Počasí}) \approx 0.985 - \left(\frac{2}{7} \cdot E(\langle \frac{1}{2}, \frac{1}{2} \rangle) + \frac{3}{7} \cdot E(\langle \frac{3}{3}, \frac{0}{3} \rangle) + \frac{2}{7} \cdot E(\langle \frac{0}{2}, \frac{2}{2} \rangle)\right) \approx 0.985 - \left(\frac{2}{7} \cdot 1 + \frac{3}{7} \cdot 0 + \frac{2}{7} \cdot 0\right) \approx 0.699$$

Původní entropie je vysoká, což znamená, že dataset je poměrně vyvážený. Informační zisk atributu *Počasí* je vysoký, zdá se tedy být vhodným kandidátem. Můžeme si všimnout, že například hodnoty *děšť* i *slunečno* již plně určují výsledek.

Samotná procedura vytváření stromu probíhá rekurzivně od kořene. V každém uzlu vybereme pomocí informačního zisku rozhodovací atribut. Podle různých hodnot vybraného atributu rozdělíme trénovací množinu a rekurzivně pokračujeme. Uzel automaticky prohlásíme za listový, pokud pro něj uvažovaná množina obsahuje jen příklady jedné třídy (jeho label bude tato třída). Uzel také prohlásíme za listový, pokud jsme v dané větvi již otestovali všechny atributy – jeho třída se obvykle zvolí podle převažující třídy mezi příklady (v tomto případě nebude strom kompletně konzistentní).

**Příklad 10.2.1.** ★ Uvažujme opět dataset daný tabulkou z předchozího (úvodního) příkladu. Nalezněte atribut, který by učící algoritmus zvolil jako rozhodovací atribut pro kořen stromu.

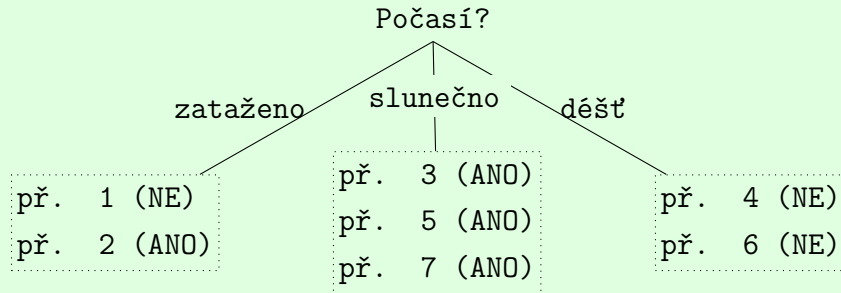
Hledáme takový atribut, jenž má nejvyšší informační zisk. Informační zisk atributu *Počasí* jsme si již spočítali v předchozím příkladu, a to  $\textit{Gain}(\textit{Počasí}) \approx 0.699$ . Spočítáme si i ostatní:

- $\textit{Gain}(\textit{Vlhkost}) \approx 0.985 - \left(\frac{2}{7} \cdot E(\langle \frac{0}{2}, \frac{2}{2} \rangle) + \frac{3}{7} \cdot E(\langle \frac{2}{3}, \frac{1}{3} \rangle) + \frac{2}{7} \cdot E(\langle \frac{2}{2}, \frac{0}{2} \rangle)\right) \approx 0.985 - \left(\frac{2}{7} \cdot 0 + \frac{3}{7} \cdot 0.918 + \frac{2}{7} \cdot 0\right) \approx 0.591$
- $\textit{Gain}(\textit{Větrno}) \approx 0.128$
- $\textit{Gain}(\textit{Teplota}) \approx 0.020$

Nejvyšší informační zisk má atribut *Počasí*, proto zvolíme ten.

**Příklad 10.2.2.** ★ Nyní konečně aplikujte i zbytek učícího algoritmu a vybudujte pomocí něj kompletní rozhodovací strom. Dataset zůstává stejný jako v předchozích příkladech. První krok jste již spočítali v minulém příkladu.

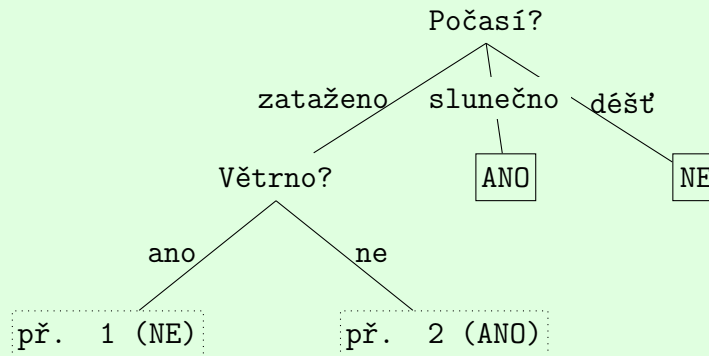
V kořeni rozhodujeme podle atributu *Počasí*. Ten nám rozdělí dataset následovně:



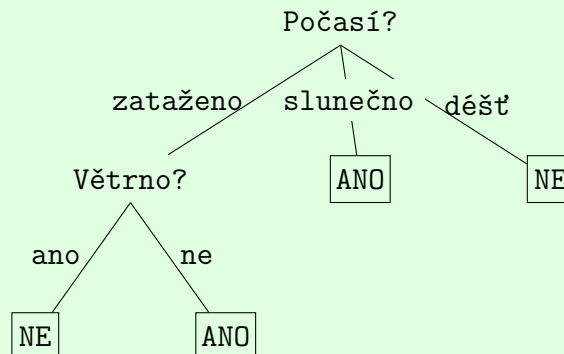
Vidíme, že dva z uzlů můžeme rovnou prohlásit za listy. Zbývá nám pokračovat pouze v levé větvi. Spočítáme informační zisk:

- $Gain(Vlhkost) = 1$
- $Gain(Větrno) = 1$
- $Gain(Teplota) = 1$

Můžeme tedy vybrat libovolný atribut. Vezměme například *Větrno*, jelikož má jen 2 hodnoty. V konkrétní implementaci algoritmu by bylo výběrové kritérium explicitně specifikované.



Oba nově vzniklé uzly můžeme prohlásit za listy. Výsledný strom tedy bude:



**Příklad 10.2.3.** Mějme následující množiny hodnot tříd ( $p/n$ ). Pro každou z nich spočítejte hodnotu entropie. Co jste z těchto výsledků vyzorovali?

- a)  $\{p, p, p, p\}$
- b)  $\{p, p, p, n\}$
- c)  $\{p, p, n, n\}$
- d)  $\{p, n, n, n\}$
- e)  $\{n, n, n, n\}$

- a) 0
- b) 0.81
- c) 1
- d) 0.81
- e) 0

Entropie náhodné proměnné o dvou hodnotách nabývá hodnot od 0 do 1 (včetně). Hodnota 0 znamená, že všechny příklady mají stejnou třídu, a opačně 1 znamená, že máme od obou tříd stejně prvků. Hodnota entropie ale neodpovídá přímo poměru tříd v množině – např. v b) máme poměr pozitivních 0.75, entropie však je 0.81.

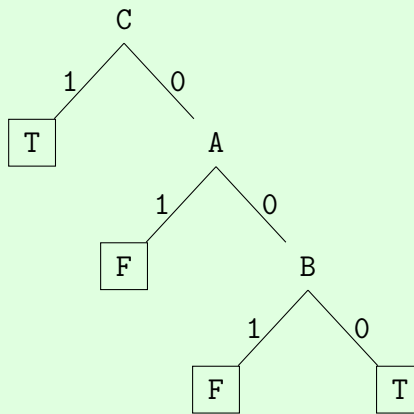
**Příklad 10.2.4.** Uvažujte následující dataset.

$A$	$B$	$C$	$Třída$
0	0	0	T
0	1	0	F
1	0	0	F
1	1	1	T

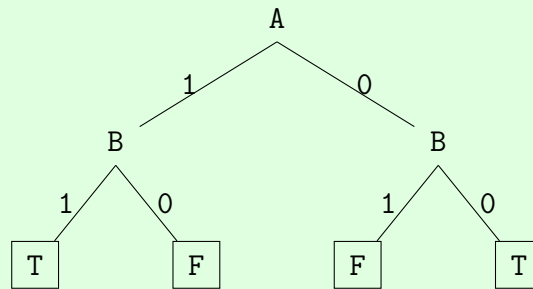
- a) Na tento dataset aplikujte učící algoritmus a vybudujte pomocí něj kompletní rozhodovací strom. Mají-li v některém kroku dva atributy stejný informační zisk, zvolte lexikograficky.
- b) Existuje konzistentní rozhodovací strom s menší hloubkou, než má strom vytvořený pomocí heuristického algoritmu? Pokud ano, vytvořte jej. Pokud ne, odůvodněte proč tomu tak je.

- a) Algoritmus nalezne následující strom:





b) Ano, opravdu existuje. Například strom, který vypadá takto:



Dokážete najít i jiný?

**Příklad 10.2.5.** Necht  $k$ -hodnotový atribut  $A$  rozdělí množinu příkladů  $D$  na podmnožiny  $D_i$ , pro které platí, že obsahují  $p_i$  pozitivních a  $n_i$  negativních příkladů. Ukažte, že pokud je pro všechna  $i$  poměr  $\frac{p_i}{p_i+n_i}$  stejný, atribut má nulový informační zisk.

Označme si  $p = \sum_{i=1}^k p_i$  a  $n = \sum_{i=1}^k n_i$ . Pokud je  $\frac{p_i}{p_i+n_i}$  stejný pro všechna  $i$ , pak musí pro všechna  $i$  platit  $\frac{p_i}{p_i+n_i} = \frac{p}{p+n}$ .

Informační zisk tedy spočítáme jako

$$\begin{aligned} \text{Gain}(A) &= E\left(\left\langle \frac{p}{p+n}, \frac{n}{p+n} \right\rangle\right) - \sum_{i=1}^k \frac{p_i+n_i}{p+n} \cdot E\left(\left\langle \frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i} \right\rangle\right) = \\ &= E\left(\left\langle \frac{p}{p+n}, \frac{n}{p+n} \right\rangle\right) - \frac{p+n}{p+n} \cdot E\left(\left\langle \frac{p}{p+n}, \frac{n}{p+n} \right\rangle\right) = 0 \end{aligned}$$

**Příklad 10.2.6.** V tomto cvičení se podrobněji podíváme na informační zisk. Necht  $T$  je množina  $n$  trénovacích objektů s atributy  $A_1, \dots, A_a$  a  $k$  třídami  $c_1$  až  $c_k$ . Necht

$\{T_i^A | i \in \{1, \dots, m_A\}\}$  je úplný rozklad  $T$  vytvořený rozdělením podle atributu  $A$  (kde  $m_A$  je počet různých hodnot  $A$ ).

- Předpokládejte, že členství ve třídě  $T$  je rovnoměrně rozděleno a nezávislé na hodnotách  $A$ . Vypočítejte entropii pro množinu  $T$ , entropii pro množinu  $T_i^A$  pro  $i \in \{1 \dots m_A\}$ , a také informační zisk pro atribut  $A$ . Interpretujte svůj výsledek.
- Nechť  $A$  je atribut s náhodnými hodnotami, které nekorelují s třídou objektů. Dále nechte  $A$  má dostatek hodnot, takže žádné dvě instance trénovací sady nemají stejnou hodnotu atributu  $A$ . Co se stane v této situaci při budování rozhodovacího stromu? Co je na této situaci problematické?

- a) Předpokládáme rovnoměrné rozdělení členství ve třídě  $T$ , tedy pravděpodobnost  $p_i = \frac{1}{k}$  pro všechny  $1 \leq i \leq k$ . Potom entropii  $T$  spočítáme jako:

$$-\sum_{i=1}^k p_i \log_2 p_i = -k \cdot \frac{1}{k} \cdot \log_2 \frac{1}{k} = -\log_2 \frac{1}{k} = \log_2 k$$

Analogicky, pro každé  $i$ , entropie  $T_i^A$  je  $\log k$ .

Informační zisk atributu  $A$  je:

$$Gain(A) = E(T) - \sum_{i=1}^{m_A} \frac{|T_i^A|}{|T|} \cdot E(T_i^A) = \log_2 k - m_A \cdot \frac{1}{m_A} \cdot \log_2 k = 0$$

Rozdělení podle  $A$  nevede k žádnému informačnímu zisku. Tento výsledek je intuitivní, rozdělení podle takového atributu nepřináší žádný užitek.

- b) Rozdělení podle atributu  $A$  vede přímo k listovým uzlům, které vždy obsahují pouze jediný příklad jedné třídy, tj.  $p_i = 1$  pro jedno  $i$  a  $p_j = 0$  pro všechny  $j \neq i$ . V důsledku toho bude entropie každého uzlu nulová, takže  $Gain(A) = E(T) - 0$  je maximální. Atribut  $A$  bude vybrán jako rozdělující atribut v kořeni a strom je dokončen.

Problém: Strom dosahuje (optimální) nulové chyby na trénovacím datasetu, ale ve skutečnosti je “k ničemu”, protože nedochází k žádné generalizaci a strom si jednoduše zapamatoval trénovací data (příkladný *overfitting*). Pokud je strom aplikován na nová data, která nebyla k dispozici během trénování, lze očekávat velkou chybu.

Taková situace může nastat, pokud je například velikost množiny dat uvažované pro rozdělení velmi malá (ať už při práci s velmi malou trénovací sadou, nebo při rozdělování uzlu hluboko ve stromu). Možným řešením pro druhý případ by mohlo být provádění předběžného prořezávání.

**Příklad 10.2.7.** Uveďte, jaké hlavní výhody a nevýhody rozhodovacích stromů znáte. Uvažujte základní trénovací algoritmus.

Mezi výhody můžeme obecně zařadit například

- interpretovatelnost - rozhodovací strom můžeme jednoduše interpretovat (na rozdíl například od neuronových sítí),
- schopnost řešit nelineární problémy,
- fakt, že metoda neklade přílišná omezení na vstupní data (nevyžaduje číselná data, neklade požadavky na rozložení vstupů)

Mezi nevýhody se naopak dá zařadit to, že

- dochází k častému přeučení se (*overfitting*) na trénovacích datech – dá se řešit různými úpravami trénovacího algoritmu (nebo modelu celkově),
- trénovací proces může být náročný, máme-li větší množství spojitých atributů,
- model může být často nestabilní – malé změny v datech mohou způsobit, že rozhodovací strom bude vypadat úplně jinak.

**Příklad 10.2.8.** Jak už bylo zmíněno v jednom z předchozích příkladů, základní verze učení rozhodovacích stromů je obecně náchylná k tzv. přeučení (*overfitting*), což může ovlivnit schopnost generalizovat. Jaké vás napadnou způsoby, kterými lze často docílit lepší generalizace?

Používané jsou například:

a) Prořezávání stromu (*prunning*) – to může proběhnout:

- V průběhu učení – zastavíme rozšiřování větve v bodě, kdy už například nemáme dostatek dat, nebo poměr tříd překročí určitou hodnotu.
- Po vytvoření stromu, pomocí dosud nepoužitého validačního datasetu (liší se od trénovacích dat). Zvolíme ve stromě některý podstrom, nahradíme jej listem s převažující kategorií, a otestujeme, zda takto upravený strom má menší chybu na validačních datech. Pokud ano, ponecháme ořezanou verzi.

Prořezávání obvykle zvyšuje chybu na učící množině, ale zlepšuje schopnost generalizovat.

b) Natrénování několika menších stromů, které následně rozhodují například pomocí majoritního hlasování. Tato idea je základem pro velmi používané modely jako *random forest*.

### 10.3 Perceptron a lineární klasifikace

V této sekci opustíme rozhodovací stromy a představíme si jiný typ algoritmů. Konkrétně se zaměříme na metody založené na lineárních modelech a představíme si nejjednodušší verze (jednovrstvých) neuronových sítí. Takové modely obvykle pracují nad spojitou doménou, a tedy jejich vstupy bývají vektory (reálných) čísel. Předpokládejme proto dále, že pracujeme s objekty, jejichž atributy mají formu reálných čísel. Toho jde obecně docílit vhodným

předzpracováním dat.

Lineární modely už můžete znát ze statistiky, kde je okolo nich vybudována rozsáhlá teorie. V kontextu tohoto kurzu využijeme myšlenku lineárních modelů k tomu, abychom si pomocí nich zavedli takzvané umělé neurony. Neurony uvnitř fungují jako lineární model s tím, že na výslednou lineární transformaci vstupů pak ještě aplikují takzvanou *aktivační funkci*. Tyto umělé neurony pak tvoří základní stavební bloky neuronových sítí. To, jak můžeme neurony propojovat do nejrůznějších architektur a vytvářet složitější sítě, si podrobněji představíme v další kapitole.

V této sekci se budeme soustředit na jednoduchý binární klasifikátor, *perceptron*. Aby model vstupy rozdělil do dvou tříd, používá jako aktivaci *prahovou funkci*. V  $n$ -rozměrném prostoru vstupů perceptron zadává nadrovinu oddělující příklady různých tříd, jak si konkrétněji ukážeme v následující sekci.

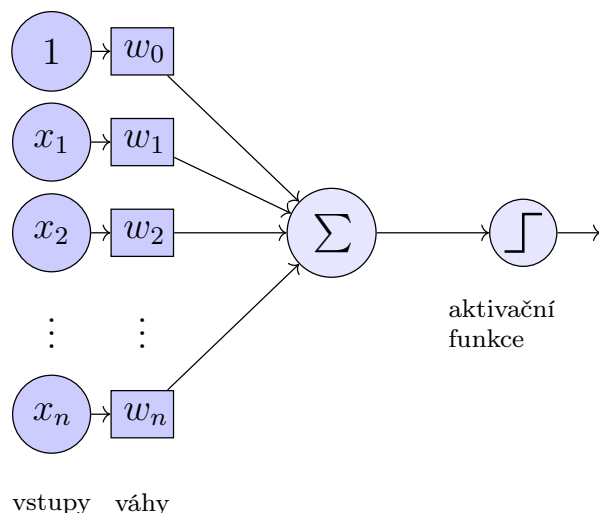
**Definice 68:** Uvažujme vstupní vektor  $\vec{x} = \langle x_1, \dots, x_n \rangle$ . Pak *rozšířený vstupní vektor* je  $\tilde{x} = \langle 1, x_1, \dots, x_n \rangle$ .

Rozšířený vstupní vektor se nám bude hodit pro zkrácený zápis některých výrazů pomocí vektorových operací.

**Definice 69:** Uvažujme příklad  $\vec{x} = \langle x_1, \dots, x_n \rangle$  a vektor vah  $\vec{w} = \langle w_0, w_1, \dots, w_n \rangle$ . Výstup perceptronové jednotky s váhami  $\vec{w}$  pro příklad  $\vec{x}$  je

$$C[\vec{w}](\vec{x}) = \begin{cases} 1 & \text{pokud } w_0 + \sum_{i=1}^n w_i \cdot x_i = \vec{w} \cdot \tilde{x} \geq 0, \\ 0 & \text{jinak.} \end{cases}$$

Funkcionalitu perceptronu si můžeme ilustrovat následujícím diagramem.



**Příklad.** Mějme perceptron s váhovým vektorem  $\vec{w} = \langle 1, 1, 2.5, 0, -1 \rangle$ .

- a) Určete, jak tento model oklasifikuje následující příklady.
- 1)  $\vec{x}_1 = \langle 0, 1, 1, 2 \rangle$
  - 2)  $\vec{x}_2 = \langle -1, 1.5, 3, 4 \rangle$
  - 3)  $\vec{x}_3 = \langle -3, 1, 3, 0.5 \rangle$
- b) Jak bychom museli *co nejméně* změnit některou váhu tak, aby byl výstup pro příklad  $\vec{x}_4 = \langle -2, 1, 2.5, 3 \rangle$  roven 1?

- a) 1) 1, protože  $1 + 1 \cdot 0 + 2.5 \cdot 1 + 0 \cdot 1 + (-1) \cdot 2 = 1.5 > 0$   
 2) 0, protože  $1 + (-1) + 3.75 + 0 + (-4) = -0.25 < 0$   
 3) 1, protože  $1 + (-3) + 2.5 + 0 + (-0.5) = 0$ . Tento bod je zajímavý tím, že leží přímo na „dělicí hranici“ perceptronu. To má pěkný geometrický význam, který si stručně představíme v další sekci.
- b) Pro původní váhy máme  $1 + (-2) + 2.5 + 0 + (-3) = -1.5$ . Potřebujeme aby byla vážená suma rovna (alespoň) 0. Nejmenší změna, kterou toho lze docílit, je změna  $w_4$  z hodnoty  $-1$  na  $-0.5$ . Zamyslete se, proč volíme právě  $w_4$ .

**Příklad 10.3.1.** ★ Pro každou z následujících logických operací definujte perceptron, který ji implementuje. Jako množinu vstupů uvažujte pouze vektory nad  $\{0, 1\}$ .

- a) binární NAND
- b) binární implikace
- c)  $n$ -ární disjunkce

Abychom definovali perceptron, stačí nám definovat jeho vektor vah.

a) Například  $\vec{w} = \langle w_0, w_1, w_2 \rangle = \langle 1.5, -1, -1 \rangle$ .

b) Například  $\vec{w} = \langle w_0, w_1, w_2 \rangle = \langle 0.5, -1, 1 \rangle$ .

c) Například  $\vec{w} = \langle w_0, w_1, \dots, w_n \rangle$ , kde  $w_0 = -1$  a  $w_1 = w_2 = \dots = w_n = 1$ .

**Příklad 10.3.2.** Ručně navrhnete perceptron, který správně klasifikuje následující data.

$x_1$	$x_2$	Třída
0	1	0
2	0	0
1	1	1

Pro nalezení vhodných vah  $w_0, w_1$ , a  $w_2$  musíme zajistit, aby následující nerovnosti byly splněny:

1. Pro příklad (0, 1) třídy 0:  $w_0 + w_2 < 0$
2. Pro příklad (2, 0) třídy 0:  $w_0 + 2w_1 < 0$
3. Pro příklad (1, 1) třídy 1:  $w_0 + w_1 + w_2 \geq 0$

Vhodnou volbou vah, která splňuje všechny tyto podmínky, je například  $w_0 = -1$ ,  $w_1 = 0.4$ , a  $w_2 = 0.7$ .

**Příklad 10.3.3.** Kolik Booleovských funkcí pro dva vstupy jde reprezentovat perceptronem?

Dohromady existuje 16 různých binárních Booleovských funkcí. Jediné dvě z nich, které nejde implementovat perceptronem, jsou XOR a ekvivalence. Je to proto, že pro tyto dvě funkce nelze lineárně separovat příklady různých tříd. Tento fakt bude jasnější po přečtení následující sekce.

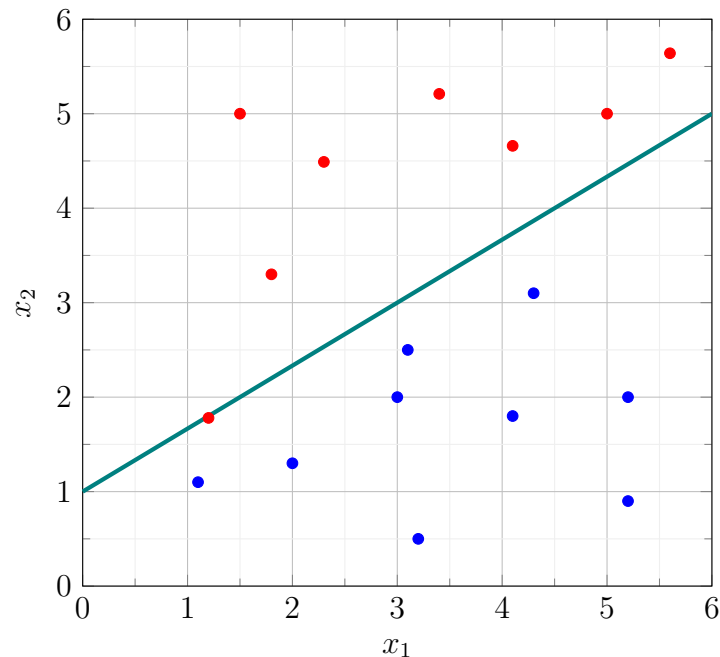
Zkuste si dokázat, že všechny ostatní Booleovské funkce na dvou vstupech opravdu perceptronem reprezentovat jde.

## 10.4 Geometrický význam a perceptronový algoritmus

Perceptron má i svůj geometrický význam, je to takzvaný *lineární separátor*. Váhy perceptronu zadávají v prostoru nadrovinu (v dvourozměrném prostoru je to přímka), která rozděluje daný prostor vstupů na dvě části. Body z jedné části model klasifikuje jako 1, z druhé jako 0.

**Definice 70:** *Dělicí nadrovina* daná perceptronem s váhami  $\vec{w} = \langle w_0, w_1, \dots, w_n \rangle$  je právě množina bodů  $\{\vec{x} = \langle x_1, \dots, x_n \rangle \mid w_0 + \sum_{i=1}^n w_i \cdot x_i = \vec{w} \cdot \vec{x} = 0\}$ .

Pro ilustraci uvažujme případ dvourozměrného prostoru, kde perceptron zadává oddělovací přímku. Příklad grafické reprezentace takového klasifikátoru můžete vidět na následujícím obrázku (body jsou obarveny podle výsledku klasifikace).



**Příklad.** Uvažte model a data reprezentované předchozí ilustrací.

- Kterou barvou jsou v obrázku vykresleny body oklasifikované jako 1?
- Zamyslete se, kolik daný model potřebuje vah, a jaký je jejich geometrický význam.
- Určete z obrázku hodnoty vah pro daný model, víme-li, že  $w_0 = -3$ .

- Červeně. Z definice víme, že body na dělicí přímce jsou oklasifikovány jako 1.
- Celkem máme tři váhy  $\vec{w} = \langle w_0, w_1, w_2 \rangle$ . Váhy  $w_1, w_2$  nám zadávají sklon přímky, zatímco  $w_0$  ovlivňuje posun od počátku. Váze  $w_0$  se proto také někdy říká *bias*.
- Rovnice přímky je  $w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 = 0$ . Víme-li, že  $w_0 = -3$ , pak (pomocí bodů, kterými prochází) dostaneme  $w_1 = -2$  a  $w_2 = 3$ .

Základní idea učení perceptronu je snaha manipulovat s dělicí nadrovinou („otáčet ji“, resp. „posunovat“) tak, aby nakonec správně oddělovala trénovací příklady patřící do různých tříd.

**Definice 71:** Uvažujme tréninkový set  $D = \{(\vec{x}_1, c_1), \dots, (\vec{x}_p, c_p)\}$ , kde  $c_i$  je opravdová třída příkladu  $\vec{x}_i$ . Perceptron  $C$  s váhami  $\vec{w}$  je konzistentní s  $D$  pokud pro všechna  $k$  z  $1, \dots, p$  platí  $C[\vec{w}](\vec{x}_k) = c_k$ .

**Definice 72:** *Online perceptronový algoritmus* je procedura, která iterativně upravuje váhy perceptronu. Postupně prochází přes jednotlivé trénovací příklady, a kdykoliv je příklad špatně oklasifikován, pootočí nadrovinou tak, aby byl příklad blíže správnému poloprostoru.

Uvažujme tréninkový set  $D = \{(\vec{x}_1, c_1), \dots, (\vec{x}_p, c_p)\}$ , kde  $c_i$  je opravdová třída příkladu  $\vec{x}_i$ . Sekvence vektorů vah  $\vec{w}^{(0)}, \vec{w}^{(1)}, \vec{w}^{(2)}, \dots$  je počítána následovně:

- $\vec{w}^{(0)}$  je inicializován náhodně hodnotami okolo 0,
- $\vec{w}^{(t+1)} = \vec{w}^{(t)} - \alpha \cdot (C[\vec{w}^{(t)}](\vec{x}_k) - c_k) \cdot \tilde{x}_k$ ,

kde  $k = (t \bmod p) + 1$ , což vyjadřuje jen to, že příklady jsou uvažovány postupně cyklicky. Konstanta  $0 < \alpha \leq 1$  je tzv. *konstanta učení (learning rate)*.

Pokud se nad algoritmem zamyslíte, pokaždé když narazí na špatně klasifikovaný vektor, tak jej přičte k aktuálnímu vektoru vah nebo ho od něj odečte (po pronásobení  $\alpha$ ).

**Příklad.** Mějme trénovací množinu  $D = \{(\langle -1, 0 \rangle, 1), (\langle 0, 1 \rangle, 1), (\langle 3, 0 \rangle, 0)\}$ . Proveďte 3 kroky perceptronového algoritmu, jestliže  $\vec{w}^{(0)} = \langle 0, 1, -1 \rangle$  a  $\alpha = 1$ .

- $\vec{w}^{(0)} = \langle 0, 1, -1 \rangle$
- $\vec{w}^{(0)} \cdot \tilde{x}_1 = -1$        $C[\vec{w}^{(0)}](\vec{x}_1) = 0$        $\vec{w}^{(1)} = \vec{w}^{(0)} - (0 - 1) \cdot \tilde{x}_1 = \langle 1, 0, -1 \rangle$
- $\vec{w}^{(1)} \cdot \tilde{x}_2 = 0$        $C[\vec{w}^{(1)}](\vec{x}_2) = 1$        $\vec{w}^{(2)} = \vec{w}^{(1)} = \langle 1, 0, -1 \rangle$
- $\vec{w}^{(2)} \cdot \tilde{x}_3 = 1$        $C[\vec{w}^{(2)}](\vec{x}_3) = 1$        $\vec{w}^{(3)} = \vec{w}^{(2)} - (1 - 0) \cdot \tilde{x}_3 = \langle 0, -3, -1 \rangle$

**Příklad 10.4.1.** Platí, že perceptronový algoritmus vždy zkonverguje a nalezne perceptron, který správně klasifikuje všechny příklady z datasetu? Odůvodněte svou odpověď.

Ne. Pokud dataset není lineárně separovatelný, perceptronový algoritmus bude cyklit.

**Příklad 10.4.2.** Mějme trénovací sadu  $D = \{(\langle 3, -1 \rangle, 1), (\langle 2, 1 \rangle, 1), (\langle 0, 3 \rangle, 0)\}$ . Aplikujte perceptronový algoritmus, dokud nenalezne separující nadrovinu. Uvažujte  $\vec{w}^{(0)} = \langle 0, -2, 1 \rangle$  a  $\alpha = 1$ . Na závěr načrtněte dělicí přímku.



- $\vec{w}^{(0)} = \langle 0, -2, 1 \rangle$
- $\vec{w}^{(0)} \cdot \tilde{x}_1 = -7$        $C[\vec{w}^{(0)}](\vec{x}_1) = 0$        $\vec{w}^{(1)} = \vec{w}^{(0)} - (0 - 1) \cdot \tilde{x}_1 = \langle 1, 1, 0 \rangle$
- $\vec{w}^{(1)} \cdot \tilde{x}_2 = 3$        $C[\vec{w}^{(1)}](\vec{x}_2) = 1$        $\vec{w}^{(2)} = \vec{w}^{(1)}$
- $\vec{w}^{(2)} \cdot \tilde{x}_3 = 1$        $C[\vec{w}^{(2)}](\vec{x}_3) = 1$        $\vec{w}^{(3)} = \vec{w}^{(2)} - (1 - 0) \cdot \tilde{x}_3 = \langle 0, 1, -3 \rangle$
- $\vec{w}^{(3)} \cdot \tilde{x}_1 = 6$        $C[\vec{w}^{(3)}](\vec{x}_1) = 1$        $\vec{w}^{(4)} = \vec{w}^{(3)}$
- $\vec{w}^{(4)} \cdot \tilde{x}_2 = -1$        $C[\vec{w}^{(4)}](\vec{x}_2) = 0$        $\vec{w}^{(5)} = \vec{w}^{(4)} - (0 - 1) \cdot \tilde{x}_2 = \langle 1, 3, -2 \rangle$

S těmito váhami algoritmus zkonvergoval, jelikož  $\vec{w}^{(8)} = \vec{w}^{(7)} = \vec{w}^{(6)} = \vec{w}^{(5)}$ .

**Příklad 10.4.3.** Uvažme klasifikátor, který přijímá dvě vstupní hodnoty, obě z domény  $\{0, 1, 2\}$ . Výstupem klasifikátoru je 1, pokud má alespoň jeden ze vstupů hodnotu 2; v opačném případě je výstupem 0. Lze takový klasifikátor implementovat pomocí perceptronu? Pokud ano, sestavte perceptron, který to dokáže; pokud ne, odůvodněte proč ne.

Příklady různých tříd zřejmě nelze lineárně separovat, a tedy pro takovou funkci nemůže existovat perceptron.

**Příklad 10.4.4.** Mějme trénovací sadu  $D = \{(\langle 1, 1 \rangle, 1), (\langle 0, 1 \rangle, 0), (\langle 0, 0 \rangle, 1), (\langle 1, 0 \rangle, 0)\}$ . Uvažujte  $\vec{w}^{(0)} = \langle 0, 0, 0 \rangle$  a  $\alpha = 1$ . Kolik iterací perceptronového algoritmu je třeba, aby našel separující nadrovinu? Nejdříve si problém zkuste načrtnout v rovině.

Algoritmus dělicí nadrovinu nalézt nemůže, protože trénovací set není lineárně separovatelný – v tomto případě se totiž zjevně jedná o problém Booleovské ekvivalence.

**Příklad 10.4.5.** Jste hollywoodský producent. Držíte ve svých rukou nový scénář, podle kterého plánujete natočit film. Před zahájením produkce se však chcete pokusit předpovědět, zda půjde o ziskový film, nebo naprostý propadák. Najmete si tedy dva kritiky A a B, aby scénář přečetli a nezávisle ho ohodnotili na stupnici od 1 do 5 (předpokládejme pouze celočíselné skóre). Jelikož ale kritici mohou být zaujatí, jednoduše průměrovat jejich skóre by nemuselo dávat smysl. Místo toho se pro výslednou predikci (na základě hodnocení kritiků) rozhodnete natrénovat několik perceptronů. Pro trénink uvažme data ohledně zisku předchozích filmů hodnocených kritiky A a B v tabulce níže.

Název filmu	A	B	Zisk
Pellet Power	1	1	Ne
Ghosts!	3	2	Ano
Pac is bac	4	5	Ne
Not a Pizza	3	4	Ano
Endless Maze	2	3	Ano

- (a) Chcete natrénovat perceptron tak, aby generoval  $C = 1$  pokud film přinese zisk,  $C = 0$  v opačném případě. Jak budou vypadat váhy po prvním průchodu perceptronového algoritmu celým datasetem? Příklady procházejte v pořadí daném tabulkou. Uvažujte rychlost učení  $\alpha = 1$  a počáteční váhy  $w_0 = -1$ ,  $w_1 = 0$ ,  $w_2 = 0$ .
- (b) Chcete najít perceptron, který by generoval  $C = 1$ , když je součet skóre obou recenzentů větší než 8, a  $C = 0$  v opačném případě. Je to možné? Pokud ano, jaké jsou váhy  $w_0, w_1$  a  $w_2$ , které to umožní?
- (c) Chcete najít perceptron, který vždy vygeneruje  $C = 1$ , když oba recenzenti souhlasí (když jsou jejich skóre přesně stejné), a vygeneruje  $C = 0$  v opačném případě. Je to možné? Pokud ano, jaké jsou váhy  $w_0, w_1$  a  $w_2$ , které to umožní?

- (a) První příklad v tabulce je klasifikován správně. Druhý příklad je klasifikován nesprávně, a proto váhy aktualizujeme na  $\vec{w} = \langle 0, 3, 2 \rangle$ . Třetí příklad perceptron klasifikuje nesprávně, váhy upravujeme na  $\vec{w} = \langle -1, -1, -3 \rangle$ . Stejně tak je nesprávně oklasifikován příklad číslo čtyři, takže znova upravujeme váhy na  $\vec{w} = \langle 0, 2, 1 \rangle$ . Pátý příklad je klasifikován správně, proto ponecháváme váhy z předchozího kroku.
- (b) Ano, takový perceptron existuje. Stačí, aby váhy splňovaly  $w_1 = w_2$  a  $w_0 = -8w_1$ . Můžeme zvolit například perceptron s váhami  $\vec{w} = \langle -8, 1, 1 \rangle$ .
- (c) Tento problém je aritmetickou verzí problému Booleovské ekvivalence, a proto není lineárně separovatelný. Takový perceptron tedy neexistuje.

## 10.5 Základy lineární regrese

Myšlenku lineárních modelů můžeme samozřejmě využít i k řešení jiných úloh, než jen ke klasifikaci. Asi nejnámější úlohou je regrese, v níž se snažíme hledat funkční závislosti mezi numerickými vlastnostmi. V našem případě budeme uvažovat *lineární regresi* – cílem tedy bude lineárně aproximovat neznámou funkci. Jedná se v podstatě o prokládání trénovačích dat „co nejlepším“ lineárním modelem. Díky tomu poté můžeme predikovat funkční hodnoty i pro dosud neviděné vstupy.

**Definice 73:** Uvažujme  $n$ -rozměrné příklady tvaru  $\vec{x} = \langle x_1, \dots, x_n \rangle$  a vektor vah  $\vec{w} = \langle w_0, w_1, \dots, w_n \rangle$ . Regresní model  $R$  s váhami  $\vec{w}$  je dán jako

$$R[\vec{w}](\vec{x}) = w_0 + \sum_{i=1}^n w_i \cdot x_i = \vec{w} \cdot \tilde{x}$$

To, jak je konkrétní model dobrý (resp. „špatný“), definujeme pomocí chybové funkce (*error function*). Obvykle chybová funkce nějakým způsobem vyjadřuje, jak daleko jsou původní

funkční hodnoty příkladů od těch aproximovaných modelem. Idea je taková, že lepší model by měl mít menší chybu. Mezi nejznámější typy chyb patří například kvadratická chyba (*squared error*) nebo absolutní chyba (*absolute error*), a jejich varianty. V této sekci budeme nejčastěji pracovat s kvadratickou chybou.

**Definice 74:** Uvažujme dataset  $D = \{(\vec{x}_1, f_1), \dots, (\vec{x}_p, f_p)\}$ , kde  $f_i$  je opravdová funkční hodnota příkladu  $\vec{x}_i$ . *Kvadratická chyba* pro model s váhami  $\vec{w} = \langle w_0, w_1, \dots, w_n \rangle$  je

$$E(\vec{w}) = \frac{1}{2} \cdot \sum_{k=1}^p (R[\vec{w}](\vec{x}_k) - f_k)^2 = \frac{1}{2} \cdot \sum_{k=1}^p (\vec{w} \cdot \tilde{x}_k - f_k)^2$$

**Příklad.** Mějme dataset jednorozměrných příkladů a jejich očekávaných funkčních hodnot  $D = \{(43, 41), (44, 45), (45, 49), (46, 47), (47, 44)\}$  a lineární model  $R$  s váhami  $\vec{w} = \langle 9.2, 0.8 \rangle$ .

- Spočítejte pro  $R$  kvadratickou chybu na množině  $D$ .
- Určete pomocí modelu  $R$  aproximaci funkční hodnoty pro  $x = 40$ .

- Spočítejme si nejdříve kvadratickou chybu pro jednotlivé trénovací příklady.

$$(9.2 + 0.8 \cdot 43 - 41)^2 = 6.76$$

$$(9.2 + 0.8 \cdot 44 - 45)^2 = 0.36$$

$$(9.2 + 0.8 \cdot 45 - 49)^2 = 14.44$$

$$(9.2 + 0.8 \cdot 46 - 47)^2 = 1$$

$$(9.2 + 0.8 \cdot 47 - 44)^2 = 7.84$$

$$E(\vec{w}) = \frac{1}{2} \cdot (6.76 + 0.36 + 14.44 + 1 + 7.84) = 15.2$$

- $R[\vec{w}](40) = 9.2 + 0.8 \cdot 40 = 41.2$

Učení regresního modelu je založeno na jiném principu než u perceptronu. Jelikož máme chybovou funkci, můžeme hledat takové váhy, které minimalizují její hodnotu. Řešíme tedy jednoduchý optimalizační problém. Algoritmus, který se k tomu obvykle používá již nejspíše znáte – jedná se o gradientní sestup (*gradient descent*). Tato metoda pomocí odečítání gradientu chybové funkce postupně upravuje váhy tak, aby minimalizovala chybu. Podrobněji se s ní seznámíte buď v následující kapitole, nebo v předmětu IB031.

**Příklad 10.5.1.** Mějme množinu jednorozměrných příkladů a jejich očekávaných funkčních hodnot  $D = \{(1, 0), (4, 2), (5, 3), (6, 3)\}$  a lineární model  $R$  s váhami  $\vec{w} = \langle \frac{2}{3}, \frac{1}{3} \rangle$ .

- Spočítejte pro  $R$  kvadratickou chybu na množině  $D$ .
- Nalezněte nějaký model, který bude mít pro množinu  $D$  menší kvadratickou chybu než  $R$ .

a) Spočítejme si kvadratickou chybu pro jednotlivé trénovací příklady:

$$\left(\frac{2}{3} + \frac{1}{3} \cdot 1 - 0\right)^2 = 1$$

$$\left(\frac{2}{3} + \frac{1}{3} \cdot 4 - 2\right)^2 = 0$$

$$\left(\frac{2}{3} + \frac{1}{3} \cdot 5 - 3\right)^2 = \frac{4}{9}$$

$$\left(\frac{2}{3} + \frac{1}{3} \cdot 6 - 3\right)^2 = \frac{1}{9}$$

$$E(\vec{w}) = \frac{1}{2} \cdot (1 + 0 + \frac{4}{9} + \frac{1}{9}) = \frac{14}{18} \approx 0.78$$

b) Například model s váhami  $w_0 = 0$  a  $w_1 = 0.5$ . Zkuste nalézt model minimalizující kvadratickou chybu.

**Příklad 10.5.2.** Mějme množinu se dvěma příklady  $D = \{(\langle 3, 5 \rangle, 13), (\langle 6, 8 \rangle, 22)\}$  a lineární model  $R$  s váhami  $\vec{w} = \langle 3, 1, 2 \rangle$ .

a) Spočítejte pro  $R$  kvadratickou chybu na množině  $D$ .

b) Aproximujte (predikujte) pomocí modelu  $R$  funkční hodnotu pro  $\vec{x} = \langle 4, 7 \rangle$ .

c) Dokážete najít nějaký model minimalizující na  $D$  kvadratickou chybu?

$$\text{a) } E(\vec{w}) = \frac{1}{2} \cdot ((3 + 1 \cdot 3 + 2 \cdot 5 - 13)^2 + (3 + 1 \cdot 6 + 2 \cdot 8 - 22)^2) = \frac{1}{2} \cdot (9 + 9) = 9$$

$$\text{b) } R[\vec{w}](\langle 4, 7 \rangle) = 3 + 1 \cdot 4 + 2 \cdot 7 = 21$$

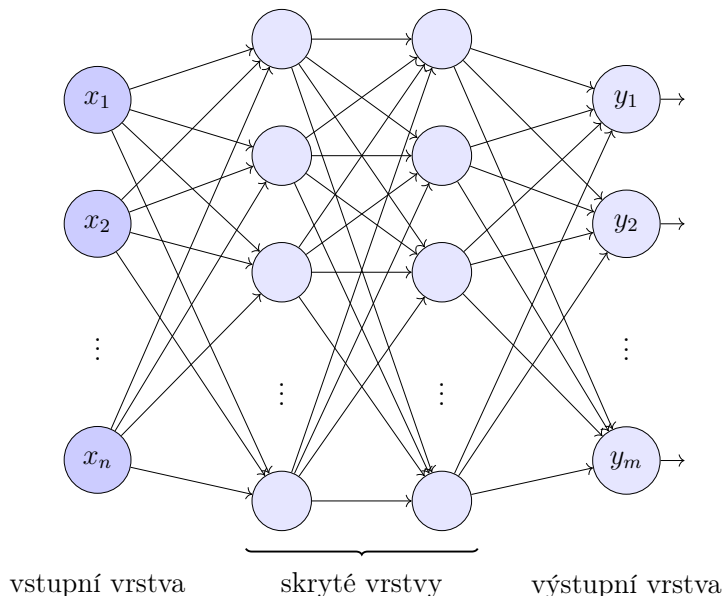
c) Stačí vzít například model, který se od  $R$  liší jen váhou  $w_0 = 0$ . Můžete se přesvědčit, že jím reprezentovaná rovina přesně prochází oběma body.

## 11 Neuronové sítě a hluboké učení

V předchozí kapitole byla poprvé představena jednoduchá jednotka umožňující lineární klasifikaci – perceptron. Sdružením více takových jednotek do vrstev a jejich následným propojením pomocí vazeb vznikají *neuronové sítě*, hlavní předmět této kapitoly.

### 11.1 Struktura a výpočet neuronové sítě

Základní strukturu dopředné neuronové sítě ilustruje následující obrázek.



Na obrázku je znázorněna neuronová síť s  $n$  vstupy, 2 skrytými vrstvami a  $m$  výstupy. Jednotlivé vrstvy jsou úplně propojené.

Výpočet dopředné neuronové sítě na zadaném vstupu probíhá po vrstvách od vstupní vrstvy k výstupní. Nejprve se jednotky vstupní vrstvy inicializují hodnotami vstupu  $\vec{x}$ . Na základě hodnot vah mezi vstupní vrstvou a první skrytou vrstvou se spočítají výstupy jednotlivých jednotek skryté vrstvy. Tento proces se postupně aplikuje na všechny vrstvy, dokud nedojde k dosažení vrstvy výstupní. Výstupy jednotek výstupní vrstvy se pak berou jako výstup neuronové sítě  $\vec{y}$  na vstupu  $\vec{x}$ .

**Definice 75:** *Vážený součet vstupů jednotky  $i$  (též vnitřní potenciál jednotky  $i$ ) je dán jako*

$$\xi_i = -w_{0,i} + \sum_j w_{j,i} \cdot a_j,$$

přičemž  $j$  iteruje přes všechny jednotky předcházející vrstvy,  $w_{0,i}$  je prahová váha jednotky  $i$ ,  $w_{j,i}$  je váha vazby z jednotky  $j$  do jednotky  $i$  a  $a_j$  je výstup jednotky  $j$ .

**Definice 76:** Výstup jednotky  $i$  je definován vztahem

$$a_i = g_i(\xi_i),$$

kde  $g_i$  je aktivační funkce jednotky  $i$ .

**Příklad.** Spočítejte výstup jednotky 4 v druhé skryté vrstvě dopředné neuronové sítě, obsahuje-li první skrytá vrstva jednotky 1, 2, jejichž výstupy jsou  $a_1 = 0,84$ ,  $a_2 = 0,02$ . Váhy hran v síti jsou  $w_{1,4} = -0,15$ ,  $w_{2,4} = 12,4$ , prahová váha jednotky 4 je  $w_{0,4} = 3,8$ . Jednotka 4 používá aktivační funkci sigmoida  $g_4(x) = \frac{1}{1+e^{-x}}$ .

Vážená váha vstupů po odečtení prahové váhy je

$$-w_{0,4} + w_{1,4} \cdot a_1 + w_{2,4} \cdot a_2 = -3,68,$$

a výstup jednotky je

$$a_4 = g_4(-3,68) = 0,025.$$

Lze tedy konstatovat, že její reakce na zadané vstupy je poněkud vlažná.

Pro jaké výstupy jednotek 1, 2 dochází k aktivaci jednotky 4?

V rámci neuronových sítí se používají různé aktivační funkce. Jak uvidíme dále, důležitým požadavkem na aktivační funkci je, aby nebyla lineární; pokud navíc chceme neuronovou síť učit, je vhodné, aby bylo možné aktivační funkci derivovat. Některé z používaných aktivačních funkcí uvádí následující definice.

**Definice 77:** Aktivační funkce.

- Prahová funkce

$$f(x) = \begin{cases} 1 & x \geq 0, \\ 0 & \text{jinak.} \end{cases}$$

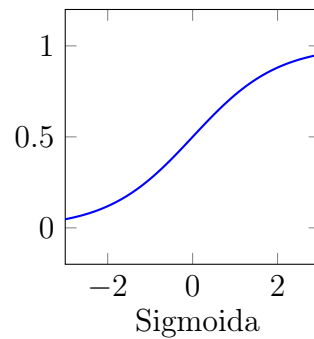
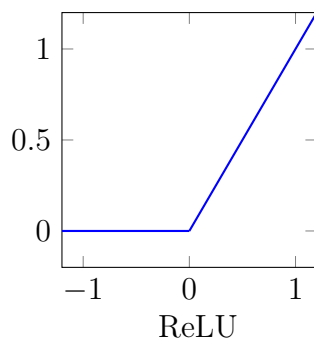
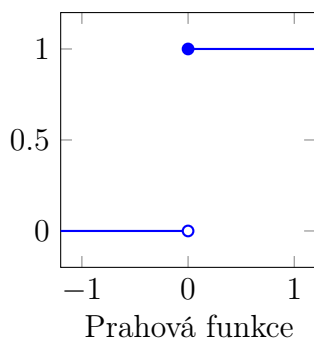
- ReLU

$$f(x) = \begin{cases} x & x \geq 0, \\ 0 & \text{jinak.} \end{cases}$$

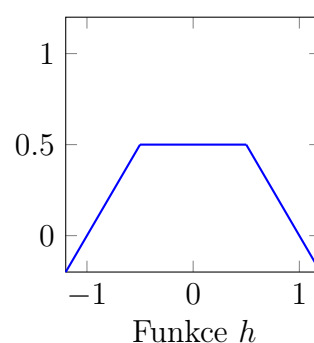
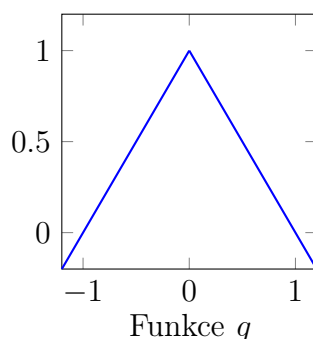
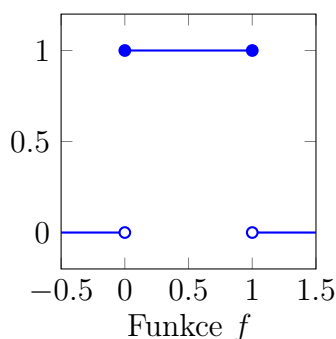
- Sigmoida

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

Grafy aktivačních funkcí jsou znázorněny na následujících obrázcích.



**Příklad 11.1.1.** Uvažte následující funkce  $f$ ,  $g$  a  $h$ .



Aplikací operací skládání funkcí, násobení konstantou, přičítání konstanty a sčítání

- a pomocí prahové funkce vyjádřete funkci  $f$ ,
- a pomocí ReLU vyjádřete funkci  $g$ ,
- a pomocí ReLU vyjádřete funkci  $h$ .

- Označme prahovou funkci  $P(x)$ . Z pohledu na obrázek funkce  $f$  se nabízí nelézt symetrii  $P(x)$  podle přímky  $x = 0.5$  a tu sečíst s původní prahovou funkcí. Symetrii lze získat jako  $P(1 - x)$ . Součtem získáme funkci  $P(x) + P(1 - x)$  s hodnotou 2 na intervalu  $[0, 1]$  a hodnotou 1 na intervalech  $(-\infty, 0)$  a  $(1, \infty)$ .

Celkem tedy vyjádříme funkci  $f$  jako

$$f(x) = P(x) + P(1 - x) - 1.$$

- Označme funkci ReLU jako  $R(x)$ . Pak lze  $g$  vyjádřit jako

$$g(x) = 1 - R(x) - R(-x).$$

- Funkci  $h$  lze vyjádřit jako

$$h(x) = 0.5 - R(0.5 - g(x)).$$

**Příklad 11.1.2.** Uvažte funkci  $F : [0, 10] \rightarrow \mathbb{R}$ . Navrhněte způsob, jak lze pomocí prahové funkce a s využitím operací skládání funkcí, násobení konstantou, přičítání konstanty a sčítání aproximovat funkci  $F$ . Požadavek na aproximaci je, aby pro aproximující funkci  $A$  platilo, že  $A(x) = F(\tilde{x})$  pro nějaké  $\tilde{x}$  takové, že  $|\tilde{x} - x| \leq 1$ , tedy hodnota aproximace se rovná funkční hodnotě v nějakém bodě, který se liší od  $x$  maximálně o 1.

Aproximující funkci vytvoříme sčítáním několika pomocných funkcí  $K[[a, b, c]]$ , které mají konstantní hodnotu  $c$  na intervalu  $[a, b)$  a jsou nulové na zbytku definičního oboru, tj.

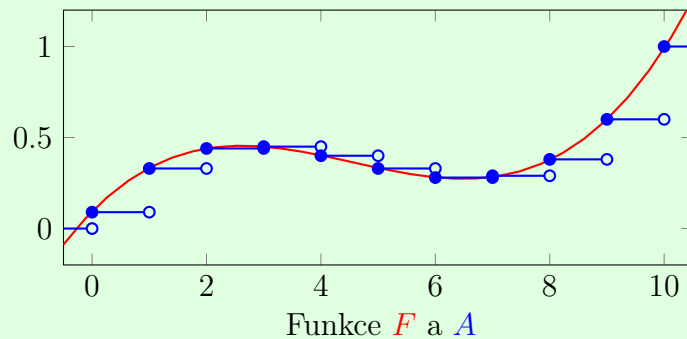
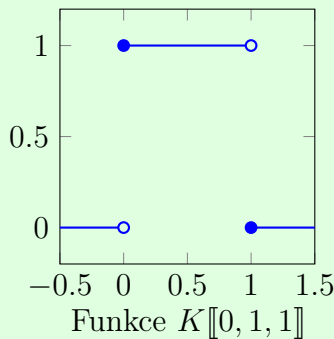
$$K[[a, b, c]](x) = \begin{cases} c & x \in [a, b), \\ 0 & \text{jinak.} \end{cases}$$

Pomocnou funkci lze jednoduše implementovat pomocí prahové funkce  $P$  jako

$$K[[a, b, c]](x) = c \cdot (P(x - a) - P(x - b)).$$

Aproximující funkci pak definujeme jako

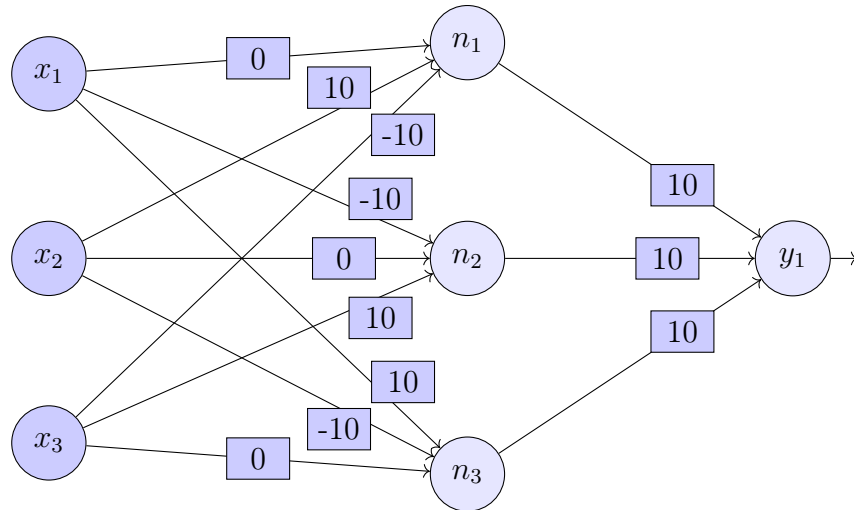
$$A(x) = \sum_{i=0}^{10} K[[i, i + 1, F(i)]].$$



Zvolená aproximující funkce splňuje požadavek ze zadání, což lze ověřit pohledem do obrázku. Pokud bychom chtěli funkci  $F$  aproximovat přesněji, lze k tomu zvolit jemnější dělení intervalu, tedy ne úseky délky 1, ale například 0.1, podle potřeby.

**Příklad 11.1.3.** ★ Uvažujte následující neuronovou síť se vstupy  $x_1, x_2, x_3$ , skrytou vrstvou s jednotkami  $n_1, n_2, n_3$  a výstupní vrstvou s jednotkou  $y_1$ .





Čísla na obrázku znázorňují hodnoty vah jednotlivých vazeb. Aktivační funkce všech jednotek je sigmoida  $\sigma$  a prahová váha je 5.

- Jaké budou výstupy sítě pro vstupy  $(0, 0, 0)$ ,  $(1, 0, 1)$ ,  $(1, 1, 1)$ ,  $(-1, 0, 1)$ ,  $(0.5, 0.4, 0.6)$ ?
- Napište jednoduchý program, který pro zadaný vstup spočítá výstup sítě.
- Co síť počítá? Nalezněte co nejstručnější slovní charakteristiku.
- Stačila by pro takový výpočet neuronová síť bez skrytých vrstev?

- a) Spočítáme vážený součet vstupů  $\xi_1$  jednotky  $n_1$  při vstupu  $\vec{x} = (0, 0, 0)$ . Vážený součet vstupů je dán součtem jednotlivých vstupních hodnot pronásobených vahami vstupujících vazeb sníženým o prahovou váhu, tedy  $\xi_1 = \sum_i w_i \cdot x_i - w_0 = 0 \cdot 0 + 10 \cdot 0 + -10 \cdot 0 - 5 = -5$ . Hodnota na výstupu je dána aplikací aktivační funkce na vážený součet vstupů, tedy  $\sigma(\xi_1) = \frac{1}{1+e^{-\xi_1}} = 0,0067$ . Podobným způsobem se spočítá vážený součet vstupů a výstupy ostatních jednotek skryté vrstvy. Výstupy se následně použijí k výpočtu váženého součtu vstupů jednotky  $y_1$  a jejího výstupu, který je zároveň výstupem sítě jako celku.

vstup	$\xi_1$	$\xi_2$	$\xi_3$	$\sigma(\xi_1)$	$\sigma(\xi_2)$	$\sigma(\xi_3)$	$\xi_{y_1}$	$\sigma(\xi_{y_1})$ výstup
$(0, 0, 0)$	-5	-5	-5	0,007	0,007	0,007	-4,80	0,008
$(1, 0, 1)$	-15	-5	+5	0,000	0,007	0,993	+5,00	0,993
$(1, 1, 1)$	-5	-5	-5	0,007	0,007	0,007	-4,80	0,008
$(-1, 0, 1)$	-15	+15	-15	0,000	1,000	0,000	+5,00	0,993
$(0, 5; 0, 4; 0, 6)$	-7	-4	-4	0,001	0,018	0,018	-4,63	0,010

- c) Síť rozpoznává, zda se výrazněji liší jednotlivé hodnoty vstupního vektoru. V případě, kdy se hodnoty liší přibližně o 1 či více, je již výstup neuronové sítě roven téměř jedničce. Čím více si jsou hodnoty podobné, tím je výsledná hodnota blíže nule.

Jednotlivé jednotky skryté vrstvy  $n_1, n_2, n_3$  reagují na pozitivní hodnotu rozdílu mezi vstupními hodnotami, tj.  $x_2 - x_3$ ,  $x_3 - x_1$ , resp.  $x_1 - x_2$ . Aktivaci alespoň jedné z jednotek pak zachycuje výstupní jednotka  $y_1$ .

Škálování vah vazeb a prahové váhy by bylo možno docílit větší či menší citlivosti sítě. Zvolené škálování funguje dobře pro vstupy, jejichž hodnoty jsou z intervalu  $[-1, 1]$ .

- d) Ne. Neuronová síť s jedním výstupem bez skrytých vrstev má pouze jedinou buňku. Její vnitřní potenciál je lineární funkcí, a tudíž při použití libovolné aktivační funkce budou v (trojrozměrném) prostoru vstupů vždy vznikat roviny, pro jejichž body je vnitřní potenciál buňky, a tedy i výstup sítě, vždy stejný.

Jelikož pozitivní případy, tedy přibližně stejné hodnoty  $x_1, x_2, x_3$ , tvoří v prostoru válec kolem přímky  $\{(t, t, t) \mid t \in \mathbb{R}\}$ , nelze je v takové síti oddělit od případů negativních.

**Příklad 11.1.4.** Dokažte, že každou booleovskou funkci  $F$ , tj. funkci tvaru  $F : \{0, 1\}^n \rightarrow \{0, 1\}$ , lze vyjádřit pomocí neuronové sítě s jednou skrytou vrstvou, kde každý neuron má prahovou funkci jako aktivační funkci.

Uvažme všechny takové hodnoty  $\vec{x}$ , že  $F(\vec{x}) = 1$ . Takových hodnot je konečně mnoho, nejvýše  $2^n$ , označme je  $\vec{x}_1, \dots, \vec{x}_N$ . V první vrstvě navrhované sítě se bude pro každý vstup  $\vec{x}_i$  nacházet jednotka  $i$ , která se aktivuje právě na vstupu  $\vec{x}_i$ .

Pro hodnotu  $\vec{x}_i = (x_{i,1}, \dots, x_{i,n})$  budou hodnoty vah jednotky  $i$

$$w_{i,0} = \sum_{j=1}^n x_{i,j}$$

a

$$w_{i,j} = \begin{cases} 1 & \text{pokud } x_{i,j} = 1, \\ -1 & \text{jinak.} \end{cases}$$

Ověřte si, že se jednotka  $i$  aktivuje, právě když je na vstupu  $\vec{x}_i$ .

K vytvoření celé sítě reprezentující funkci  $F$  nyní stačí vytvořit druhou vrstvu, která bude sestávat z jediné jednotky, implementující operaci OR mezi všemi jednotkami první vrstvy. Postup implementace operace OR jsme diskutovali v předchozí kapitole v rámci Příkladu 10.3.1.

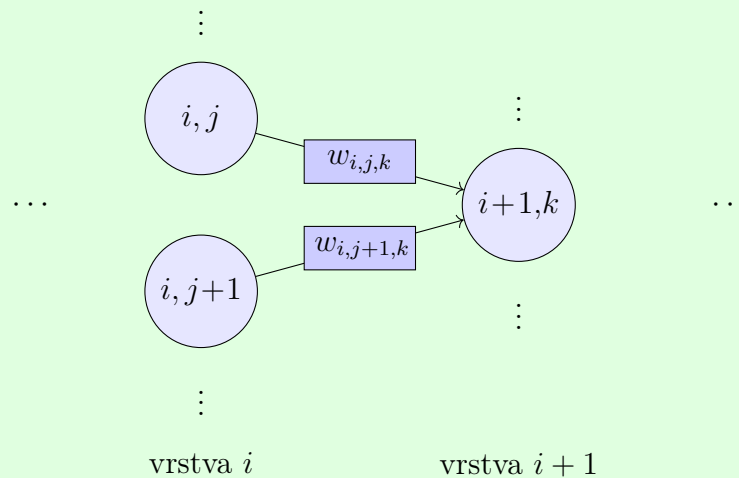
**Příklad 11.1.5.** Uvažujte vícevrstvou neuronovou síť s jedním výstupem, kde každá vnitřní jednotka  $i$  má aktivační funkci  $f_i(x) = A_i x + B_i$  (kde  $A_i, B_i$  jsou reálné konstanty) a výstupní jednotka je aktivována prahovou funkcí.

- Má neuronová síť stejnou vyjadřovací sílu jako neuronová síť, která pro vnitřní jednotky používá aktivační funkci  $\sigma$  (sigmoidu)? Vyjadřovací silou rozumíme, jaké různé funkce lze neuronovou sítí reprezentovat. Zdůvodněte.
- Má neuronová síť větší vyjadřovací sílu než perceptron? Pokud ano, nalezněte příklad funkce, kterou lze sítí vyjádřit, ale perceptronem ne. Pokud ne, ukažte, jak takovou

sít převést na perceptron počítající stejnou funkci.

- a) Ne, nemá. Každou hodnotu včetně výstupu  $y$  v takové síti lze totiž vyjádřit jako lineární funkci vstupu  $\vec{x}$ . Každá další vrstva pouze transformuje (lineárně) předchozí vrstvy, které jsou samy lineární transformací vstupu. Konkrétní hodnoty viz další odrážka. V případě použití sigmoidy  $\sigma$  jakožto aktivační funkce je linearita odstraněna a při dostatečném počtu skrytých vrstev, lze reprezentovat libovolnou funkci.
- b) Jak bylo uvedeno v předchozí odrážce, výstup každé síti s výhradně lineárními aktivačními funkcemi je lineární funkce vstupů. Podívejme se tedy, jak takovou síť převést na perceptron, čímž získáme konstruktivní důkaz tohoto tvrzení.

Uvažme neuronovou síť s  $n - 1$  skrytými vrstvami  $1, \dots, n - 1$ , výstupní vrstvou  $n$  a vstupní vrstvou  $0$ , kde vrstva  $i$  obsahuje  $N_i$  jednotek  $i, 1; \dots; i, N_i$  s aktivačními funkcemi  $f_{i,j}(x) = A_{i,j}x + B_{i,j}$  pro  $i \geq 1$ . Váha spoje z  $j$ -té jednotky vrstvy  $i$  do  $k$ -té jednotky vrstvy  $i + 1$  je  $w_{i,j,k}$ . Dle zadání má výstupní vrstva jedinou jednotku, ale konstrukci by šlo jednoduše zobecnit.



Tady bude řešení zbytku příkladu.

## 11.2 Učení neuronové sítě

V této sekci pronikneme hlouběji do samotné podstaty vícevrstvých neuronových sítí. Konkrétně se zaměříme na *algoritmus zpětného šíření chyby*, který umožňuje trénování neuronových sítí na konkrétní sadě dat.

Učící sada obsahuje dvojice tvaru

$$(\vec{x}, \vec{y}_{\text{exp}}),$$

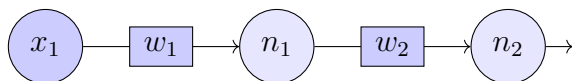
kde  $\vec{x}$  je vstup neuronové sítě a  $\vec{y}_{\text{exp}}$  je očekávaný výstup. Učení probíhá na principu úpravy hodnot vah v síti, aby došlo k minimalizaci chybové funkce

$$E = E(\vec{y}, \vec{y}_{\text{exp}}),$$

kde  $\vec{y}$  je skutečný výstup sítě při vstupu  $\vec{x}$ . Chybová funkce musí být navržena tak, aby její minima odpovídala minimálnímu rozdílu mezi skutečným a očekávaným výstupem.

Příkladem chybové funkce je kvadratická chyba, kterou jsme představili v předchozí kapitole. Pro použití v neuronových sítích je třeba definici mírně upravit, aby zvládla pojmout vícerozměrný výstup.

**Příklad.** Uvažujte následující jednoduchou neuronovou síť s jednou skrytou vrstvou.



Symbole  $\xi_1, \xi_2$  označují vážené součty vstupů jednotky 1, resp. 2; symbole  $a_1, a_2$  pak jejich výstupy a  $g_1, g_2$  jejich aktivační funkce. Chybová funkce je

$$E = E(y, y_{\text{exp}}) = E(a_2, y_{\text{exp}}).$$

- Spočítejte parciální derivace  $\partial E / \partial w_2$  a  $\partial E / \partial w_1$ .
- Jaký je význam spočítaných parciálních derivací při učení neuronové sítě?

a) Při výpočtu využijeme toho, že platí

$$a_2 = g_2(\xi_2) \quad \text{a} \quad \xi_2 = a_1 \cdot w_2,$$

což plyne z definice výstupu jednotky, resp. definice váženého součtu vstupů jednotky. Jelikož  $E$  je funkcí  $a_2$  a  $a_2$  je funkcí  $\xi_2$ , lze s využitím řetězového pravidla rozepsat

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial a_2} \frac{\partial a_2}{\partial w_2} = \frac{\partial E}{\partial a_2} \frac{\partial a_2}{\partial \xi_2} \frac{\partial \xi_2}{\partial w_2}.$$

S využitím vztahů výše můžeme rozepsat poslední dva součinitele jako

$$\frac{\partial a_2}{\partial \xi_2} = g_2' \quad \text{a} \quad \frac{\partial \xi_2}{\partial w_2} = a_1$$

a celkově získáme

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial a_2} \cdot g_2' \cdot a_1 = \frac{\partial E}{\partial y} \cdot g_2' \cdot a_1.$$

Jednotlivé členy lze v konkrétním případě již získat snadno: parciální derivaci chybové funkce podle skutečného výstupu sítě lze spočítat z konkrétní podoby chybové funkce, derivaci aktivační funkce spočítáme opět z jejího tvaru (proto je dobré, aby měla derivaci, jak bylo uvedeno dříve při zavádění pojmu aktivační funkce) a hodnotu  $a_1$  získáme prostým provedením výpočtu sítě nad daným vstupem.

Obdobným způsobem spočítáme

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_2} \frac{\partial a_2}{\partial w_1} = \frac{\partial E}{\partial a_2} \frac{\partial a_2}{\partial \xi_2} \frac{\partial \xi_2}{\partial w_1}.$$

Jelikož  $\xi_2 = a_1 \cdot w_2$  a  $a_1$  dále závisí na  $w_1$ , pokračujeme ve výpočtu:

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_2} \frac{\partial a_2}{\partial \xi_2} \frac{\partial \xi_2}{\partial a_1} \frac{\partial a_1}{\partial w_1} = \frac{\partial E}{\partial a_2} \frac{\partial a_2}{\partial \xi_2} \frac{\partial \xi_2}{\partial a_1} \frac{\partial a_1}{\partial \xi_1} \frac{\partial \xi_1}{\partial w_1} = \frac{\partial E}{\partial a_2} \cdot g'_2 \cdot w_2 \cdot g'_1 \cdot x_1.$$

- b) Cílem učicího algoritmu je pomocí úprav vah vazeb v síti minimalizovat chybu, tj. rozdíl mezi výstupy, které neuronová síť dává, a očekávanými výstupy. Parciální derivace chybové funkce podle váhy vazby vyjadřuje míru, s jakou se mění velikost chyby při změně váhy.

Je-li například  $\partial E / \partial w_1 = 2$ , znamená to, že při zvýšení váhy  $w_1$  o 1 se chyba  $E$  zvýší přibližně o 2. V takovém případě tedy chceme váhu snížit, což povede ke snížení chyby.

Učicí algoritmus tedy v každé iteraci upraví váhy podle předpisu

$$(w_1, w_2) \leftarrow (w_1, w_2) - \alpha \cdot \left( \frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2} \right),$$

kde  $\alpha > 0$  je tzv. *učicí konstanta (learning rate)*, která reguluje rychlost učení a s níž jsme se setkali již v předchozí kapitole.

Parciální derivace  $\partial E / \partial w_i$  chybové funkce podle váhy hrany  $w_i$  tedy určuje, jaký má změna váhy vliv na výslednou chybu výpočtu.

Sdružíme-li všechny parciální derivace do vektoru

$$\nabla E = \left( \frac{\partial E}{\partial w_1}, \dots, \frac{\partial E}{\partial w_n} \right),$$

získáme tzv. *gradient chyby*, který vyjadřuje, kterým „směrem“ (v rámci vícerozměrného prostoru všech možných vah hran) chyba nejrychleji roste. Při postupu opačným směrem, tedy odečtením násobku gradientu od vah sítě, dosahuje síť snížení chyby.

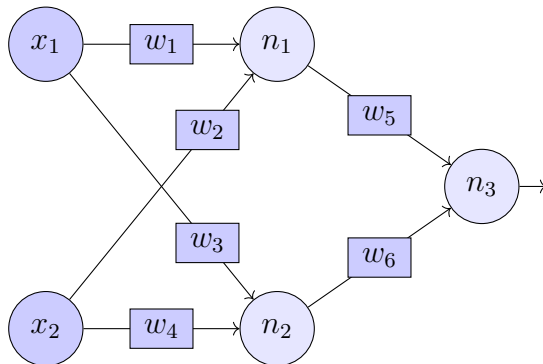
**Definice 78:** *Online algoritmus pro učení neuronové sítě* je procedura, která iterativně upravuje váhy v neuronové síti. Postupně prochází přes jednotlivé trénovací příklady a odečte od hodnot vah o gradient chyby pronásobený učicí konstantou.

Uvažujme tréninkový set  $D = \{(\vec{x}_1, \vec{y}_1), \dots, (\vec{x}_p, \vec{y}_p)\}$ , kde  $\vec{y}_i$  je očekávaný výstup při vstupu  $\vec{x}_i$ . Sekvence vektorů vah  $\vec{w}^{(0)}, \vec{w}^{(1)}, \vec{w}^{(2)}, \dots$  je počítána následovně:

- $\vec{w}^{(0)}$  je inicializován náhodně hodnotami okolo 0,
- $\vec{w}^{(t+1)} = \vec{w}^{(t)} - \alpha \cdot \nabla E(\vec{y}^{(t)}, \vec{y}_k)$ ,

kde  $k = (t \bmod p) + 1$  a  $\vec{y}^{(t)}$  je výsledek vrácený sítí v iteraci  $t$  na vstupu  $\vec{y}_k$ .

**Příklad 11.2.1.** Uvažujte následující neuronovou síť s jednou skrytou vrstvou.



Symbole  $\xi_1, \xi_2, \xi_3$  označují vážené součty vstupů jednotky 1, 2, resp. 3; symboly  $a_1, a_2, a_3$  pak jejich výstupy a  $g_1, g_2, g_3$  jejich aktivační funkce. Prahové váhy neuvažujeme. Chybová funkce je

$$E = E(y, y_{\text{exp}}) = E(a_3, y_{\text{exp}}).$$

- Spočítejte  $\partial E / \partial w_5$ .
- Spočítejte  $\partial E / \partial w_1$ .
- Jaký je obecný předpis pro  $\partial E / \partial w_i$  v obecné dopředné neuronové síti?

a) Platí

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial a_3} \frac{\partial a_3}{\partial \xi_3} \frac{\partial \xi_3}{\partial w_5}.$$

Ze struktury sítě máme  $\partial a_3 / \partial \xi_3 = g'_3$  a  $\xi_3 = a_1 \cdot w_5 + a_2 \cdot w_6$ , a tudíž

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial a_3} \cdot g'_3 \cdot \frac{\partial}{\partial w_5} (a_1 \cdot w_5 + a_2 \cdot w_6) = \frac{\partial E}{\partial a_3} \cdot g'_3 \cdot a_1.$$

b) Platí

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_3} \frac{\partial a_3}{\partial \xi_3} \frac{\partial \xi_3}{\partial w_1} = \frac{\partial E}{\partial a_3} \cdot g'_3 \cdot \frac{\partial}{\partial w_1} (a_1 \cdot w_5 + a_2 \cdot w_6) = \frac{\partial E}{\partial a_3} \cdot g'_3 \cdot \left( \frac{\partial a_1}{\partial w_1} \cdot w_5 + \frac{\partial a_2}{\partial w_1} \cdot w_6 \right).$$

Ze struktury sítě máme  $\xi_1 = x_1 \cdot w_1 + x_2 \cdot w_2$ , a tedy

$$\frac{\partial a_1}{\partial w_1} = \frac{\partial a_1}{\partial \xi_1} \frac{\partial \xi_1}{\partial w_1} = g'_1 \cdot \frac{\partial}{\partial w_1} (x_1 \cdot w_1 + x_2 \cdot w_2) = g'_1 \cdot x_1.$$

Podobně  $\xi_2 = x_1 \cdot w_3 + x_2 \cdot w_4$ , takže

$$\frac{\partial a_2}{\partial w_1} = \frac{\partial a_2}{\partial \xi_2} \frac{\partial \xi_2}{\partial w_1} = g'_2 \cdot \frac{\partial}{\partial w_1} (x_1 \cdot w_3 + x_2 \cdot w_4) = 0.$$

Proč v druhém případě vyšla derivace nulová? Všimněme si, že derivace  $\partial a_2 / \partial w_1$  vyjadřuje, o kolik se změní hodnota výstupu jednotky  $n_2$  při jednotkové změně hodnoty hrany  $w_1$ . Pohledem na náčrt sítě si však lze všimnout, že hrana  $w_1$  hodnotu  $n_2$  nijak neovlivňuje, a tudíž je derivace nulová dle očekávání.

Spojením výsledků získáme celkem

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial a_3} \cdot g'_3 \cdot (g'_1 \cdot x_1 \cdot w_5 + 0) = \frac{\partial E}{\partial a_3} \cdot g'_3 \cdot w_5 \cdot g'_1 \cdot x_1.$$

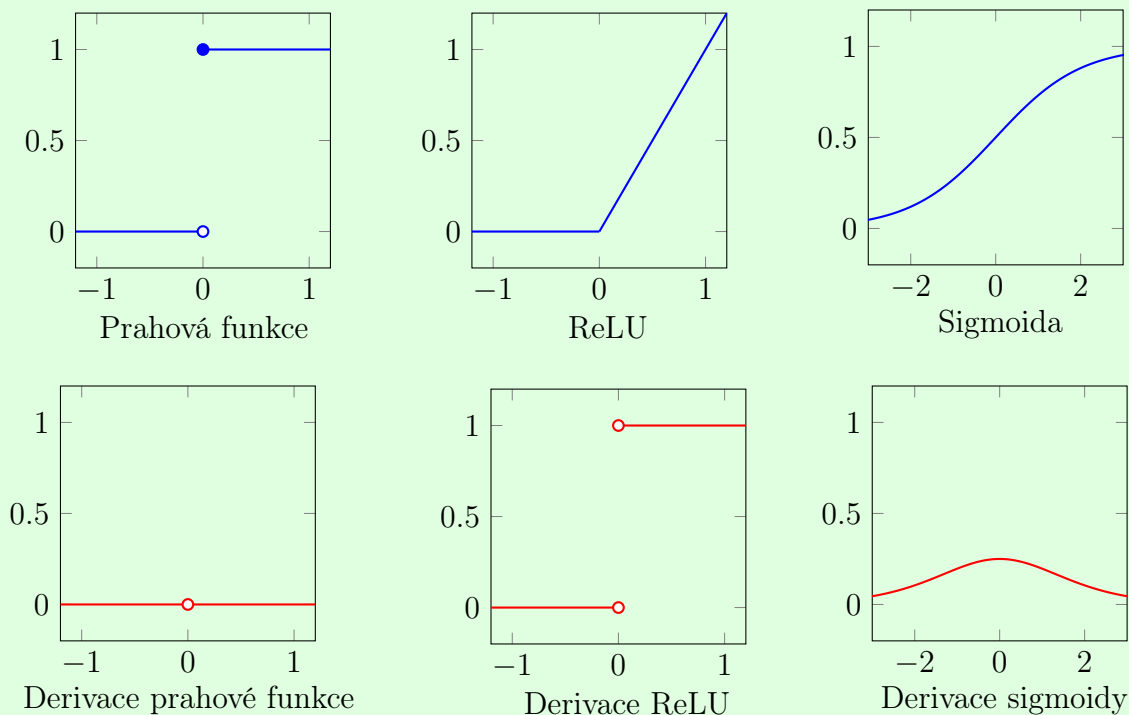
c) **Tady bude řešení.**

**Příklad 11.2.2.** Nalezněte derivace následujících aktivačních funkcí:

- prahová funkce,
- ReLU,
- sigmoida.

- Prahouá funkce je konstantní na intervalech  $(-\infty, 0)$ ,  $[0, \infty)$ , což by napovídalo tomu, že derivace je 0. Tak tomu skutečně je s výjimkou bodu 0, kde derivace není definována kvůli skoku hodnoty. Přesněji řečeno, levá derivace je  $\infty$  a pravá je 0, tudíž derivace není definována.
- ReLU je konstantní na intervalu  $(-\infty, 0)$  s derivací 0 a lineární na  $(0, \infty)$  s derivací 1. V bodě 0 opět derivace definována není, levá derivace je 0 a pravá derivace je 1.
- Zderivovat sigmoidu je poměrně zdlouhavé, proto zde uvedeme pouze výsledek, čili  $\sigma'(x) = \sigma(x)(1 - \sigma(x))$ . Podrobný postup lze nalézt například zde.

Grafy aktivačních funkcí (modře) a jejich derivací (červeně) jsou znázorněny na následujících obrázcích.



**Příklad 11.2.3.** Diskutujte význam učící konstanty  $\alpha$  při učení neuronové sítě. Jaký má vliv na průběh učení (pozitivní i negativní), je-li hodnota konstanty velmi velká, resp. velmi malá?

Příliš malá učící konstanta může způsobit, že proces trvá déle, než by bylo nutné. V případě zvolení příliš vysoké učící konstanty může docházet k „přeskakování“ optima, následkem čehož učení nemusí konvergovat.

### 11.3 Zpracování obrazu a klasifikace

Při klasifikaci do  $n$  kategorií je vhodné získat výstup ve formě  $n$ -tice pravděpodobností. K tomu slouží funkce *softmax*, která zvýrazní rozdíly mezi jednotlivými výstupy pomocí exponenciální funkce a zároveň provede normalizaci, aby součet členů  $n$ -tice byl roven jedné.



**Definice 79:** Funkce *softmax* transformuje  $n$ -tici reálných čísel na  $n$ -tici pravděpodobností následujícím způsobem

$$\text{softmax}(x_1, \dots, x_n) = \left( \frac{e^{x_1}}{\sum_{i=1}^n e^{x_i}}, \frac{e^{x_2}}{\sum_{i=1}^n e^{x_i}}, \dots, \frac{e^{x_n}}{\sum_{i=1}^n e^{x_i}} \right).$$

Zamyslete se, zda může funkce softmax vrátit pro nějakou kategorii nulovou pravděpodobnost.

### Příklad 11.3.1.

- Bez použití kalkulačky odhadněte výsledek aplikace funkce softmax na následující vstupy:  $(1, 3, 2)$ ,  $(1, 1, 1)$ ,  $(2, 0, 1)$ ,  $(-5, 0, 0)$ .
- Odhad z předchozí části ověřte výpočtem.

- Softmax počítá mocniny čísla  $e$  (přibližně 2.7) ze zadaného vstupu. Ty pak normuje, aby jejich součet byl 1. V případě vstupu  $(1, 3, 2)$  můžeme pomocí aproximace  $e \approx 3$  odhadnout poměr složek výsledného vektoru na  $1 : 9 : 3$ . Po znormování získáme přibližně  $\text{softmax}(1, 3, 2) \approx (1/10, 3/4, 1/4) = (0.1, 0.75, 0.25)$ . V případě vstupu  $(1, 1, 1)$  lze výsledek hned určit přesně jako  $\text{softmax}(1, 1, 1) = (1/3, 1/3, 1/3)$ . Pro vstup  $(2, 0, 1)$  odhadneme podobně jako v prvním případě, poměr výsledných složek bude  $9 : 1 : 3$ , čili  $\text{softmax}(2, 0, 1) \approx (3/4, 1/10, 1/4) = (0.75, 0.1, 0.25)$ . Poslední vstup  $(-5, 0, 0)$  dá vzniknout složkám v poměru přibližně  $1 : 250 : 250$ , takže  $\text{softmax}(-5, 0, 0) \approx (1/500, 1/2, 1/2) = (0.002, 0.5, 0.5)$ .
- Následující tabulka shrnuje provedené odhady a skutečné hodnoty.

$\vec{x}$	odhad	$\text{softmax}(\vec{x})$
$(1, 3, 2)$	$(0.100, 0.750, 0.250)$	$(0.090, 0.665, 0.245)$
$(1, 1, 1)$	$(0.333, 0.333, 0.333)$	$(0.333, 0.333, 0.333)$
$(2, 0, 1)$	$(0.750, 0.100, 0.250)$	$(0.665, 0.090, 0.245)$
$(-5, 0, 0)$	$(0.002, 0.500, 0.500)$	$(0.003, 0.498, 0.498)$

## 12 Zpracování přirozeného jazyka

Na jazyky formální i přirozené nahlížíme jako na množiny platných vět. Na rozdíl od formálních jazyků (připomeňme třeba jazyk formulí výrokové logiky) s sebou však jazyky přirozené přinášejí řadu komplikací.

Je například poměrně jednoduché určit, že  $p \vee \vee \vee$  není platnou formulí výrokové logiky, protože nevyhovuje definici. Je ovšem věta „To byl teda krutopřísněj vejlet.“ platnou větou češtiny? Odpověď není zcela jednoznačná a s ohledem na časový vývoj jazyka a nářeční nebo stylistickou rozmanitost přirozeného jazyka nelze ani přesně definovat, co vše do jazyka ještě patří a co už ne.

Slova přirozeného jazyka mohou být víceznačná – např. slovo *list* může znamenat kus papíru či část rostliny. Přirozený jazyk je navíc nejasný – kdo je činitelem ve větě „Přišel.“? Existuje navíc mnoho dalších komplikací (viz přednášku), které ze zpracování přirozeného jazyka činí nelehkou úlohu.

V této kapitole jsou představeny některé z metod, které se při zpracování přirozeného jazyka používají.

### 12.1 Předzpracování dat

Textová data, se kterými potřebujeme pracovat, jsou mnohdy dostupná pouze ve velmi neotesané podobě, např. ve formě webových stránek či hrubého textu s nadbytečnými údaji. Cílem předzpracování může být očištění dat od nadbytečných informací (např. různých anotací), jejich segmentace či normalizace, tedy celkově převedení dat do formy, na niž lze již aplikovat samotné metody zpracování přirozeného jazyka.

---

**Příklad 12.1.1.** Na adrese [https://web2.mlp.cz/koweb/00/04/34/55/03/bila\\_nemoc.html](https://web2.mlp.cz/koweb/00/04/34/55/03/bila_nemoc.html) si stáhněte text knihy Karla Čapka Bílá nemoc ve formátu HTML. Formát HTML je textový formát, který obsahuje značky pro internetový prohlížeč, které umožňují dokument zobrazit ve formě webové stránky.

- a) Napište skript, který z dokumentu extrahuje pouze text díla, konkrétně ve formě seznamu jednotlivých replik jako řetězců. Výstup tedy bude začínat následujícím způsobem.

```
["Mor je to, mor. V naší ulici už je v každém domě ...",  
 "Žádný mor, malomocenství. Bílá nemoc tomu říkají,...",  
 "Kriste panebože - Kriste panebože - Kriste panebože -",  
 ...  
]
```

Ve výsledných řetězcích je vypuštěna informace o konkrétní postavě, která danou repliku říká, neboť se jedná o značně repetitivní informaci.

- b) Přidejte do skriptu funkce, které provádějí segmentaci textu na věty, resp. tokenizaci na jednotlivá slova. Slova získávejte očištěná od interpunkčních znamének.
- c) Vytvořte slovník s frekvencemi jednotlivých slov v díle. Je vhodné mít metodu, která pro každé slovo vrátí jeho základní tvar. Např. slovo „ale“ se v textu může vyskytovat i ve formách „Ale“ nebo „ALE“, ale všechny tyto tvary by se ve slovníku měly započítat pod položku „ale“.
- d) Napište funkci, která vypíše seznam  $n$  nejčastějších slov, která se v díle (nebo jiném zadaném textu) vyskytují.

d) Tady bude statistika nejčastějších slov.

## 12.2 Gramatiky

Gramatiky představují jeden ze způsobů, kterým lze popsat jazyk. Gramatika udává předpis, jakým lze z počátečního symbolu, tzv. kořene gramatiky, postupnou aplikací pravidel gramatiky odvodit platné věty jazyka gramatiky. Speciálním typem gramatik jsou gramatiky bezkontextové, kterým budeme v této sekci věnovat speciální pozornost.

**Definice 80:** *Bezkontextová gramatika sestává z*

- množiny terminálních symbolů (*slov jazyka*),
- množiny neterminálních symbolů (*syntaktických kategorií*),
- speciálního neterminálního symbolu  $S$  reprezentujícího celou větu jazyka (též nazýván *kořen gramatiky*),
- souboru *přepisovacích pravidel tvaru*

neterminál  $\rightarrow$  libovolný řetězec.

Běžná konvence je psát neterminály s prvním písmenem velkým a veškerá pravidla, která vycházejí ze stejného neterminálu sdružit pomocí symbolu svislítka, tj. např.

$$S \rightarrow abbb,$$

$$S \rightarrow cac$$

píšeme

$$S \rightarrow abbb \mid cac.$$

Gramatika definuje jazyk jako množinu vět sestávajících pouze ze slov jazyka, které lze vygenerovat z kořene gramatiky  $S$  postupnou aplikací pravidel gramatiky. Odvození věty v gramatice navíc definuje tzv. *syntaktický strom*. Samotná pravidla gramatiky se často rozdělují na pravidla (pouze mezi syntaktickými kategoriemi) a lexikon (seznamy slov příslušících jednotlivým kategoriím).

**Příklad.** Uvažte gramatiku s následujícími pravidly:

$$\begin{aligned}S &\rightarrow NP VP, \\NP &\rightarrow Noun \mid Adj NP, \\VP &\rightarrow Verb\end{aligned}$$

a následujícím lexikonem:

$$\begin{aligned}Noun &\rightarrow dítě \mid člověk \mid kapsa, \\Adj &\rightarrow starý \mid cestující \mid nové, \\Verb &\rightarrow píše \mid sedí \mid mluví.\end{aligned}$$

a) Rozhodněte, které z následujících vět lze v gramatice vygenerovat.

1. „cestující sedí“
2. „nové nové kapsa píše“
3. „starý člověk mluví“

b) Nalezněte odvození vět v gramatice.

- a) Gramatika generuje věty, které začínají jmennou frází (*NP*, neboli *noun phrase*), po níž bezprostředně následuje slovesná fráze (*VP*, neboli *verb phrase*). Jmenná fráze sestává z podstatného jména (*Noun*), jemuž předchází libovolný počet přídavných jmen (*Adj*). Slovesná fráze sestává z jediného slovesa.

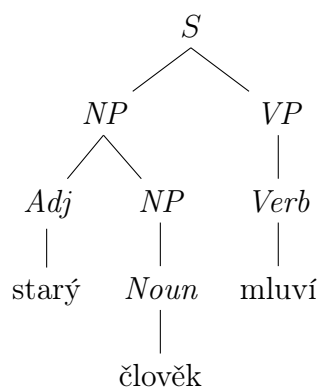
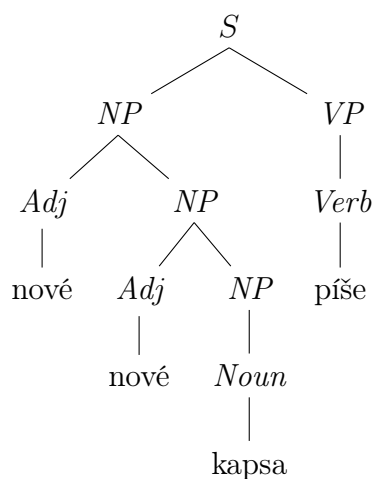
Větu 1 nelze v gramatice vygenerovat, jelikož neobsahuje podstatné jméno. Věty 2 i 3 splňují popis výše a v gramatice je vygenerovat lze.

b)

$S \Rightarrow NP VP$   
 $\Rightarrow Adj NP VP$   
 $\Rightarrow nové NP VP$   
 $\Rightarrow nové Adj NP VP$   
 $\Rightarrow nové nové NP VP$   
 $\Rightarrow nové nové Noun VP$   
 $\Rightarrow nové nové kapsa VP$   
 $\Rightarrow nové nové kapsa Verb$   
 $\Rightarrow nové nové kapsa píše$

$S \Rightarrow NP VP$   
 $\Rightarrow Adj NP VP$   
 $\Rightarrow starý NP VP$   
 $\Rightarrow starý Noun VP$   
 $\Rightarrow starý člověk VP$   
 $\Rightarrow starý člověk Verb$   
 $\Rightarrow starý člověk mluví$

Syntaktické stromy vět z příkladu vypadají následovně.



Z příkladu lze vypožorovat několik skutečností. Zaprvé, mohou existovat věty přirozeného jazyka, které nejsou gramatikou popsány – viz věta 1. Zadruhé, gramatika může generovat věty, které do popisovaného přirozeného jazyka nepatří – viz věta 2, kde je porušena shoda v rodě mezi podstatným jménem *kapsa* (ženský rod) a přídavným jménem *nové* (střední rod).

Cílem návrhu gramatik je minimalizovat počet vět, které gramatika generuje a nepatří do jazyka, a zároveň maximalizovat počet vět, které gramatika generuje a do jazyka patří. Kvalitu gramatiky kvantitativně popisují pojmy *pokrytí* a *přesnost*.

**Definice 81:** Uvažujme zamýšlený jazyk  $L$  a gramatiku  $G$  generující jazyk  $L(G)$ .

- *Pokrytí* gramatiky  $G$  je  $\frac{|L \cap L(G)|}{|L|}$ , tedy podíl vět jazyka  $L$ , které lze v gramatice vygenerovat.
- *Přesnost* gramatiky  $G$  je  $\frac{|L \cap L(G)|}{|L(G)|}$ , tedy podíl vět generovaných gramatikou, které patří do zamýšleného jazyka.

**Příklad 12.2.1.** ★ Uvažte gramatiku s následujícími pravidly:

$$\begin{aligned} S &\rightarrow NP VP \mid VP, \\ NP &\rightarrow Noun \mid NP Conj NP, \\ VP &\rightarrow NP Esse \end{aligned}$$

a následujícím lexikonem:

$$\begin{aligned} Noun &\rightarrow \text{Romulus} \mid \text{Remus} \mid \text{Danubius} \mid \text{fratellus} \mid \text{fratelli} \mid \text{fluvius}, \\ Esse &\rightarrow \text{sum} \mid \text{est} \mid \text{sunt} \mid \text{eram} \mid \text{erat} \mid \text{erant}, \\ Conj &\rightarrow \text{et}. \end{aligned}$$

a) Rozhodněte, které z následujících vět lze v gramatice vygenerovat.

1. „Romulus et Remus fratelli erant“
2. „Remus et Danubius et Romulus sum“
3. „Danubius est fluvius“

b) Nalezněte ke každé větě její syntaktický strom, pokud existuje.

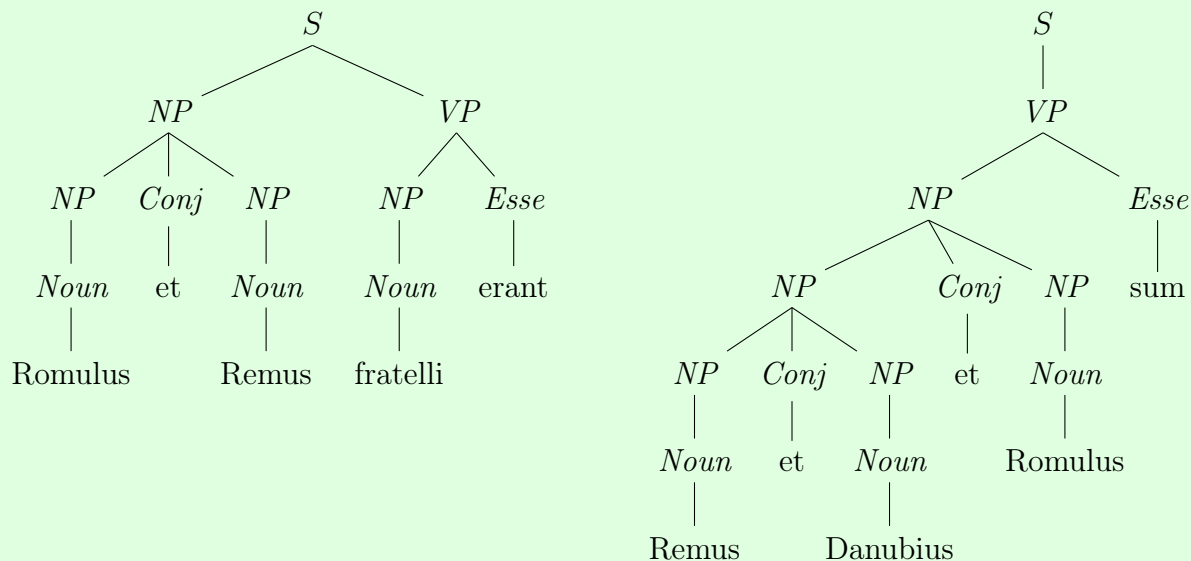
a) Gramatika tvoří věty dodržující schéma

$$[\text{jmenná fráze}], \text{jmenná fráze}, \text{sloveso být},$$

kde jmenná fráze se může skládat z několika podstatných jmen pospojovaných spojkou „et“. První i druhá věta (v překladu „Romulus a Remus bratři byli“ a „Remus a Dunaj a Romulus jsem“) toto schéma dodržují a lze je tedy v gramatice vygenerovat, byť druhá věta nedává smysl.

Třetí věta (v překladu „Dunaj je řeka“) schéma nedodržuje kvůli umístění slovesa *být*, takže ji v gramatice vygenerovat nelze.

b) Syntaktické stromy uvedených vět vypadají následovně.



**Příklad 12.2.2.** Navrhněte gramatiku pro jednoduché české (resp. slovenské) věty v minulém čase, které dodržují následující schéma:

[určení času / podmět], přísudek v minulém čase, [předmět],

tedy například „včera jsme koupili maso“, „já jsem běžel“, „hráli jste hru“. Nezapomeňte, že pořadí pomocného slovesa *být* a významového slovesa závisí na přítomnosti jiného slova na začátku věty (věta nikdy nezačíná pomocným slovesem). Ošetřete shodu podmětu s přísudkem

- přímo návrhem gramatiky,
- přidáním (jednoduše implementovatelných) testů na shodu.

Tady bude řešení.

**Příklad 12.2.3.** Uvažme gramatiku  $G$  s následujícími pravidly

$$S \rightarrow AAA \mid BA \mid AB \mid C,$$

$$A \rightarrow a,$$

$$B \rightarrow bA \mid Ab,$$

$$C \rightarrow BA \mid AB \mid cB \mid Bc$$

a jazyk  $L$  vět délky 3 sestavených ze slov  $a$ ,  $b$ ,  $c$ , které obsahují slovo  $b$  právě jednou.

- Je gramatika jednoznačná, neboli existuje pro každou větu jazyka  $L(G)$  odvoditelnou v  $G$  právě jeden syntaktický strom?

- b) Jaké je pokrytí gramatiky  $G$  vzhledem k zamýšlenému jazyku  $L$ ?  
 c) Jaká je přesnost gramatiky  $G$  vzhledem k zamýšlenému jazyku  $L$ ?

a) Není, například pro větu  $aab$  existují dva syntaktické stromy:



- b) Pokrytí je podíl vět zamýšleného jazyka, které lze v gramatice vygenerovat. Zamýšlený jazyk má celkem 12 vět – 3 pozice slova  $b$  a dvakrát dvě volby mezi slovy  $a$  a  $c$ . Z vět zamýšleného jazyka nepatří do jazyka gramatiky pouze věty obsahující na prostřední pozici slovo  $c$  –  $bcc$ ,  $bca$ ,  $ccb$ ,  $acb$  – a věty obsahující  $c$  dvakrát – mimo již zmíněných ještě  $cbc$ . Celkově je tedy pokrytí  $7/12$ .
- c) Přesnost je podíl vět generovaných gramatikou, které patří do zamýšleného jazyka. Jediná věta z  $L(G)$ , které nepatří do  $L$ , je  $aaa$ . Celkový počet vět v  $L(G)$  je 8. Přesnost gramatiky je tedy  $7/8$ .

### 12.3 Syntaktická analýza

Gramatiku popisující jazyk lze použít k provádění *syntaktické analýzy*. Výsledkem syntaktické analýzy je syntaktický strom, který reflektuje strukturu věty a může sloužit jako podklad pro další zpracování.

**Příklad 12.3.1.** Knihovna NLTK (ze cvičení) umožňuje provádění metod zpracování přirozeného jazyka jako předzpracování textu, značkování či syntaktickou analýzu. Napište gramatiku pro knihovnu NLTK, která bude schopná rozeznat slovesný čas vět v anglickém jazyce. Značky, které slovům NLTK přidělí, nesou vedle slovesného druhu i informaci o tvaru slovesa (gerundium, minulý čas apod.), které lze za tímto účelem využít. Gramatika by měla umět určit alespoň následující časy.



I speak	<i>present simple</i>
You are speaking	<i>present continuous</i>
He has spoken	<i>present perfect</i>
We spoke	<i>past simple</i>
They were speaking	<i>past continuous</i>
I had spoken	<i>past perfect</i>

Tady bude řešení.